Modeling and Assessment of Urban Environmental Issues: Land Use Change in Japan and Air Pollution in China

杜, 国棟

Modeling and Assessment of Urban Environmental Issues: Land Use Change in Japan and Air Pollution in China

Guodong Du

June 5, 2018

Modeling and Assessment of Urban Environmental Issues: Land Use Change in Japan and Air Pollution in China

Guodong Du

A Thesis Submitted in Partial Fulfillment of the Requirement for

Degree of Doctor of Engineering

Department of Urban and Environmental Engineering

Graduate School of Engineering

Kyushu University

Japan

June 5, 2018

Kyushu University

Graduate School of Engineering

Guodong Du

The undersigned hereby certify that they have read and recommended to be the Graduate School of Engineering fro the acceptance of this thesis entitled by

Modeling and Assessment of Urban Environmental Issues: Land Use Change in Japan and Air Pollution in China

by Guodong Du in partial fulfillment of the requirements for the Degree of Doctor of Engineering

---

Shunsuke Managi, Ph.D

Professor, Department of Urban and Environmental Engineering

Academic advisor

Kenichi Tsukahara, Ph.D

Professor, Department of Urban and Environmental Engineering

Akihito Ozaki, Ph.D

Professor, Department of Architecture and Urban Design

Abstract

The development and sustainability of the urban areas or cities play important roles in the socio-economic development, but are also subject to various environmental issues. In order to achieve sustainable development, the urban environmental issues have received great attentions from both urban planners and policy-makers; the priority issue vary with the economy levels of different areas. Japan is classified as a highly developed country, and Japan cities prioritize the achievement of sustainable and stable land use changes (LUC) according to the voluntary national review 2017 reported by United Nations. On the other hand, the cities in the fast developing country, such as Chinese cities tend to prioritize the mitigation of air pollution problem, which is mainly caused by the intensive economy development.

This study aims to provide supportive technique and evidence that would assist the promotion of the sustainable development of urban area in Japan and China. LUC process in a highly developed urban area is more sophisticated than that in fast developing area in terms of the complexity of spatio-temporal change pattern of land use (LU). Hence, to support urban land use planning, more advanced LUC modeling techniques are needed to capture the LUC pattern and to forecast the future urban LU. On the other hand, Chinese government has invested heavily to abate the air pollution problems but without the satisfactory outcomes. Estimating residents' monetary valuation of air pollution could help to improve the effectiveness of air pollution control policy making by allowing for cost-benefit analyses. Given the abovementioned context, this study explores the enhancement of LUC modeling in the Greater Tokyo Area by incorporating advanced machine learning (ML) and deep learning (DL) techniques to existing modeling approach; this study also assesses the impact of air pollution on Chinese people' well-being and estimates the monetary value of air pollution by using subjective well-being (SWB) approach.

As for the LUC modeling in the Greater Tokyo Area, this study explores the en-

hancement method from three perspectives: 1) enhances the stochastic modeling with tree-based ensemble algorithms, including bagged decision tree (bagged DT), bagged gradient boosting decision tree (bagged GBDT), random forests (RF) and extremely randomized trees (ERT); 2) enhances the spatial modeling with convolutional-based deep learning methods, including convolutional neural networks (CNN) and convolutional denoising autoencoders (CDAE); 3) enhances the temporal modeling with recurrent neural networks (RNN), including simple RNN and three RNN variants with gated architecture.

The results show that the LUC modeling in Greater Tokyo Area benefits from incorporating certain degree of randomness given that the ERT model, which has the highest degree of randomness among the four tree-based models, significantly outperforms the other models by 5%∼30%. The results also provide the evidence of convolutional-based models' ability to enhance the conventional LUC models by extracting and supplying useful spatial features from the satellite images, given that both convolutional-based models outperform a multi-layer perceptron (MLP) model which uses only conventional geographical features by 15%∼30%. Moreover, usage of RNN to model the spatio-temporal dynamics of LUC process yields reliable LU forecasts; the higher performance of RNN variants with gated architecture also indicates that modeling long-term temporal dependency of LUC process can further improve the modeling performance.

In the examination of the impact of air pollution on the residents and the analysis uses the combined data of 1) SWB data and the other individual characteristics from an original Internet survey conducted in China during January and February in 2016, and 2) air pollution data collected from official statistical yearbook or measurement of monitoring sites. This study uses regression analyses to determine the relationships between SWB and the air pollution variables; the estimated coefficients from the regression analyses are then used to estimate the monetary value of air pollution for Chinese residents living in Northeastern region which is a declining heavy industrial area, as well as Beijing city and Shanghai city which are the largest and yet growing cities in China.

The results of regression analyses show the statistically significant negative effect of air pollution on Chinese people's subjective well-being. Nevertheless, the magnitude of impact of air pollution and the estimated monetary values of air pollution varies with regions and across the cities. Northeastern Chinese residents place nearly two times higher monetary value on air pollution compare to the Beijing and Shanghai residents. Furthermore, the estimated monetary values also vary with the time-specification of air pollution variables, air pollutants, and also with respect to subjective health evaluation and household characteristics such as household income, subjective health condition, etc.

This study contributes to both the research of LUC modeling and air pollution assessment. In the filed of LUC modeling, this study reveals the positive effect of stochastic mechanism in tree-based algorithms for modeling the LUC process in highly-developed metropolitan area; moreover, it is the first time to introduce and to identify the great potentials of DL techniques for LUC modeling. By using these approaches, more reliable LU prediction could be generated and used to better support the decision-making of strategic urban planning in Greater Tokyo Area. On the other hand, in the filed of air pollution assessment, this study provides the up-to-date assessment results of the direct and interacted effect of air pollution on Chinese people's well-being with taking temporal, spatial and personal factors into account, which allows for specific suggestions to Chinese policy-makers.

This study presents useful findings for two important and highly prioritized urban environmental issues by examining the Japan and China areas. Nonetheless, the future work should consider the connection between the two environmental issues, land use change and air pollution, by combining the findings of this study to build a modeling and assessment framework.

Keywords

Urban environmental issues, Land use change, Tree-based ensemble algorithms, neural networks, subjective well-being, air pollution, monetary valuation

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my advisor Shunsuke Managi for the continuous support of my Ph.D. study. I have been extremely honored to be his Ph.D student and to have his guidance, encouragement and advices. His constructive suggestions have been of great help for completing this research and organizing the structure of this thesis. I am also thankful for the excellent example he has provided as a successful researcher and professor.

I am grateful to all of those with whom I have gad the pleasure to work during these three years. I would like to pay special thankfulness to my co-author, Kong Joo Shin, for her willingness and enthusiasm to offer insights and reviews to this research. Her valuable inputs significantly improved the quality of this research, and her sympathetic attitude helped me to work in time. I am grateful to Chiaki Matsunaga for her assistance and guidance to the graduation procedure and other administrative matters. I also wish to acknowledge all the helps provided by other researchers Shinya Ikeda, Wataru Nozawa, Tetsuya Tamaki, Mihoko Wakamatsu, Rintaro Yamaguchi, Fukai Hiroki, Akinori Kitsuki and Moinul Islam. Past and present graduate students that I have had the pleasure to work with or alongside of are doctor students Hiroki Onuma, Toshihiko Kitamura, Yogi Sugiawan, Liang Yuan, Jun Xie, Chi Zhang, Binqi Zhang, Toliver Clarence, Moegi Igawa, Junya Kumagai and Thierry Coulibaly; Masters program students Kei Takahashi, Naoto Tada, Ryota Nonaka, Takanori Okada, Qiuyi Chen, Ebrahim Aly and Rafid Mahful. The kind supports received from Naoko Endo and Mayo Miyaishi are also appreciated.

I would like to thank the members of my dissertation committee, Kenichi Tsukahara and Akihito Ozaki, for generously offering their time, support and guidance to this thesis. Their valuable comments greatly improved the quality of this thesis.

Lastly, I would like to thank my family for all their love and encouragement. For my parents and my parents in law who supported me and my wife in all our pursuits.

For my pets who relieve me of stress and depression. Most of all for my loving and patient wife who is also my best friend, co-author and colleague, and whose support and understanding are most appreciated.

Contents

List of Figures

List of Tables

Introduction

Urban areas changes and grows with the social transformation process, continuously shaping the earth surface. According to the UN report, 50 percent of the world's population was living in urban area in 2014, and 66 percent of the world's population is projected to live in urban area by 2050. The development and sustainability of the cities or urban area play important role in socio-economic development at the era of globalization (Keivani, 2010). It serves as centers for finance and producer services, and provide the essential elements of residents' well-being and cultural development. Cities are also serve as the centers of political power and administration. The policy and developmental agendas of cities have profound influences on the policy making of the surrounding regions, and the entire nation.

However, the concentration of population, power and resources exposes the cities to various economic-environmental issues, such as urban sprawl, urban vegetation, air pollution, greenhouse gas emission, etc. The interactions of these environmental issues complicate the analyses of the impacts and the consideration of resolutions. For example, urban sprawl can lead to loss of urban vegetation, and loss of urban vegetation can further cause deterioration of air quality and loss of ecosystem diversity; urban sprawl also can influence a transportation pattern, and an ill-planned transportation pattern can increase the usage of private vehicle and the occurrence of traffic congestion, which lead to the increase of air pollutants emission.

Since the report from the UN World Commission on Environment and Development (the Brundtland Commission) was published in 1987, the concept of sustainable development has been widely adopted by the administrators and planners as the principle to deal with aforementioned urban problems (Hassan and Lee, 2015). The prioritized sustainable development goals of countries and cities vary with varying economy conditions. Consequently, the focuses of environmental analyses in different countries also vary significantly. Japan, one of the most developed country, has already effectively controlled various pollution problems, and has turned its focus toward the sustainable development goals such as sustainable and resilient land use according to the voluntary national review 2017 reported by UN [1]. In particular, the Greater Tokyo Area in Japan, which is the world's largest metropolitan area with approximately 37 million citizens, has a special responsibility to provide a leadership and provide the success case of the sustainable urban development.

The Greater Tokyo Area faces major challenge; the area already has high population density but the population is still growing. Such situation may lead to the spontaneous expansion of urban area, loss of urban vegetation and biodiversity, etc. Strategic urban planning could help to ensure the sustainable development of the Greater Tokyo Area. However, the urban planning for the mega metropolitan area is extremely complicated and challenging. Land use change (LUC) modeling is an effective approach to understand the complex urban system and to model the LUC process. It provides essential LU forecast for predicting the possible environmental outcomes caused by the current urban development plan, and by doing so could help to support the decision-making of urban planning.

On the other hand, in the developing economy, China is well-known for suffering from severe air pollution problem mainly caused by the intensive economy development. Air pollution problem has attracted broad attentions from Chinese government and pub-

[1] Refer to https://sustainabledevelopment.un.org/memberstates/japan for the whole list of the prioritized sustainable development goals of Japan

lic, and has been regarded as one of the top-concerned environmental issues in China. Chinese government has published several temporary air pollution control policies and invested numerous money and efforts to mitigate air pollution in the past several years. Nonetheless, the reduction target has not been met; for instance, despite the air pollution mitigation effort, the city of Being have not reached the official 5-years target. Improving the applicability and cost-efficiency of air pollution control policy need the impact evaluation of air pollution. In particular, monetary valuation of air pollution can provide the bases for the effective cost-benefit analyses.

In the field of environmental studies, many studies have conducted methodological or application research of LUC modeling and monetary valuation of public goods. However, the existing literature has limitations. In terms of LUC modeling, previous studies mainly focus on the LUC modeling in fast-developing regions where are mainly located in the developing world, and rarely consider the LUC modeling in highly-developed area. In the developed areas, the frequency of LUC is relatively mild but the LUC process is more complicated. In contrast with fast-developing urban areas where are dominated by urban expansion, most developed mega cities are simultaneously experiencing urban expansion, urban decay and urban renewal. These processes are usually regional-specific and are driven by various socio-economic forces. For example, the urban expansion usually occurs at suburbs where the housing price is relatively cheap, while urban decay and urban renewal usually occur at downtown where is declining because of either the transfer of urban center or the deterioration of living/commercial environment. These distinct characteristics of LUC process in developed urban area may decrease the applicability of LUC models developed for developing urban areas. Therefore, in order to assist the urban planning in developed urban areas by LUC modeling, the existing LUC models need to be further enhanced and extended.

In terms of the assessment of air pollution, subjective well-being (SWB) approach is an emerging approach to evaluate the impact of air pollution and estimate the monetary value of air pollution. Due to its effectiveness, it has been widely adopted by researchers.

However, previous SWB studies mainly focuses on developed areas mainly due to the data availability, which is contrary to the LUC modeling studies. The data availability in China is even worse, because of the various limitations of collecting individual data through survey in China. Furthermore, the survey data used in previous studies that focus on China are relatively out-of-dated; the latest Chinese survey data used in existing literature were collected in 2014 (Xu and Li, 2016). Given the rapid changing atmospheric and policy environment, a SWB study with using more up-to-dated survey data is necessary for providing effective suggestions for policy-makers.

In order to provide supportive technique and evidence to promote the sustainable development of urban area in Japan and China, this study aims to enhance the LUC modeling in Greater Tokyo Area. As a highly developed metropolitan area, Greater Tokyo Area has several distinct characteristics compared with fast-developing urban area, including 1) complicated LUC pattern, i.e., urban expansion, urban decay and urban renewal could occur at the same time; 2) driven by spontaneous behavior rather than explicit urban development plan; 3) slow development and low LUC frequency. These distinct characteristics pose technical challenges for the LUC modeling in Greater Tokyo Area, such as the complex LU transition rules and data imbalance problem. Existing LUC models are mainly developed for fast-developing urban area, such as Tehran (Tayyebi et al., 2011) and Guangzhou (Li et al., 2015a), and may not be able to effective tackle the technical challenges of LUC modeling in Greater Tokyo Area.

Cellular automata (CA) is the most prevalent modeling approach in contemporary studies. It is a simulation method defining and stacking a series of if-then transition rules to model the LUC process. CA is commonly transformed into more sophisticated variants or combined with other approaches to provide reliable simulation for complex LUC modeling tasks. One popular approach is the integration of machine learning (ML) methods and CA. ML, or statistical learning, is a field of statistics and computer science that gives computer systems the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed (Samuel, 1959). In

the integrated model, ML is used to predict the LU transition probability by assimilating spatio-temporal data, and CA is used to simulate the LU pattern by defining and improving the LU transition rules through trail-and-error tests. Although a variety of ML techniques have been applied in LUC modeling, modern ML techniques, particular ensemble models and deep learning (DL) techniques, are rarely used in existing literature.

This study extends the existing LUC modeling framework by incorporating tree-based algorithms and deep learning techniques. In particular, tree-based algorithms are used to enhance the stochastic modeling of LUC process; convolutional neural networks (CNN) is used to extract hidden spatial features from satellite images to capture the neighborhood characteristics; recurrent neural networks (RNN) is used to enhance the modeling of spatio-temporal dynamics of LUC process. Table 1 summarizes the ML/DL and CA methods used in this study.

Although the study focuses on the LUC modeling of Greater Tokyo Area, the specific study area varies among different sub-studies due to the different incentives and the limitations of data availability and computational power. The study areas are the whole Greater Tokyo Area, Saitama prefecture and Tsukuba city for the three sub-studies that focus on tree-based algorithms, CNN and RNN, respectively. With respect to the incentive, the whole Greater Tokyo Area intrinsically has complicated LU transition rules due to the massive and complex urban system, which is suitable for examining the capability of tree ensemble methods to tackle complex LUC modeling tasks; Saitama prefecture has complicated LU pattern with intensive interspersion of built-up, agriculture and forest, which is suitable for examining the benefits of incorporating spatial features extracted by CNN; Tsukuba city undertakes slow, continuous and stable expansion in the last decades, which is suitable for examining the benefit of capturing temporal dependency by RNN.

With respect to the limitation, the LU data for sub-study of tree-based algorithms were obtained from Ministry of Land, Infrastructure, Transport and Tourism of Japan, while the LU data for the other two sub-studies were classified from satellite images.

Table 1: Brief summary of methods used in LUC modeling in this study

Tree-based algorithms	Decision tree	Basis of tree ensemble methods
	Gradient boosting decision tree	Tree ensemble methods with using boosting; used for reducing the error of bias
	Bagged trees Random forest Extremely randomized trees	Tree ensemble methods with using bagging; used for reducing the error of variance; three methods differ in the varying degree of stochastic mechanism
Neural networks	Multi-layer perceptron	Fully connected feed-forward neural network with multiple hidden layers
	Convolutional neural network	Feed-forward neural network that uses convolution at least for one layer; specifically designed for hierarchical feature extraction from data that have grid-like topology (e.g. images)
	Convolutional denoising autoencoders	Feed-forward neural network in an unsupervised learning approach; used for learning hidden representations from input data
	Simple recurrent neural network	Basic form of recurrent neural networks, which is used for processing sequential data (e.g. time-series data)
	Long short-term memory Long short-term memory with peephole connection Gated recurrent unit	Variants of recurrent neural networks that introduces advanced gated architecture, granting stronger capability of modeling long-term temporal dependency in sequential data
Cellular automata	DINAMICA-variant	DINAMICA is a popular patch-based cellular automata framework; it is specifically modified for adapting to the modeling framework in this study

Given the relatively low accuracy of LU classification from satellite images for large area, the study areas of the two sub-studies, which focus on deep learning techniques, cannot cover the whole Greater Tokyo Area. On the other hand, deep learning methods requires massive computational power. For instance, CNN models have millions of parameters; RNN models have complex derivative computation which requires large memory. Moreover, the whole spatio-temporal datasets usually have sizes of $5 \sim 15$ GB, which further increase the computational power requirement. This study uses GPU computational accelerating technique for deep learning methods with using a NVIDIA GTX 1080 card with 8 GB memory, however, the computational power is still restricted.

In the line of LUC modeling studies, it is the first time to address the benefit of

stochastic mechanism for improving the accuracy of LU transition probability prediction; it is also the first time to introduce DL into LUC modeling. This study contributes the LUC modeling studies by identifying the performance improvement from incorporating tree ensemble methods and deep learning techniques and by providing effective modeling approaches for highly developed urban area through the case of Greater Tokyo Area.

In order to support the policy-making of air pollution control measures in China, this study also assesses the impact of air pollution by estimating the monetary value of air pollution in urban area of China by using subjective well-being (SWB) data and individual characteristics data from an Internet survey during January and February 2016. Survey data is combined with objective air pollution data, and the regression analysis is employed to evaluate the impact of air pollution on people's well-being and to estimate the monetary values of air pollution for Chinese people. This study focuses on northeast region, Beijing and Shanghai, because these regions are relatively suffering more from air pollution due to the geographical location and/or intensive economic activities. Northeast region represents a heavy industrial area where is under declining economy, while Beijing and Shanghai are the two largest cities where are still undergoing fast-development in China.

In previous studies, in addition to stated-preference and revealed-preference approaches, the subjective well-being (SWB) approach is gaining popularity in the field of environmental economics, which emphasizes the environmental impact on people?s subjective evaluation of their own well-being. Self-reported well-being is regarded as a robust empirical approximation of overall utility. Along with determinants such as income and other demographic factors, the impacts of various dimensions of environmental quality have been investigated by examining the relationship between environmental quality and self-reported well-being.

The SWB analyses in this study use the latest survey data collected in the beginning of 2016, allowing for the up-to-date policy implication for air pollution control in China.

Compared with previous studies, this study not only analyzes the impact of air pollution on people?s well-being, but also analyzes the interacted effects of air pollution and other individual characteristics. Moreover, this study also uses geographical information system (GIS) tools to disaggregate the air pollution data into individual level and provide enhanced monetary valuation of air pollution for Beijing and Shanghai residents. Table 2 summarizes the methods used in SWB analyses.

Table 2: Brief summary of methods used in SWB analyses in this study

Categories		Description
Regression analyses	Ordinary least square regression	A common method for estimating unknown parameters in a linear model
	Ordered probit	A specific method for estimating unknown parameter in a linear model with ordinal dependent variable that has more than two outcomes
Spatial interpolation	Ordinary Kriging interpolation	A popular geostatistical interpolation method, considering both the spatial distance and spatial autocorrelation

The thesis is organized as follows. This thesis has two parts: LUC modeling in Greater Tokyo Area and monetary valuation of air pollution. In Part 1, Chapter 1 introduces the background and existing literature regarding the LUC modeling studies; Chapter 2 presents the study of incorporating tree-based algorithms; Chapter 3 presents the study of incorporating CNN; Chapter 4 presents the study of using RNN. In Part 2, Chapter 5 introduces the background and existing literature regarding the SWN studies focusing on air pollution; Chapter 6 presents the study conducted in northeast part of China; Chapter 7 presents the study conducted in Beijing and Shanghai. Finally, Chapter 8 concludes.

Part I

Land use change modeling in Greater Tokyo Area

Chapter 1

Background

Throughout the history of civilization, human beings have continuously shaped the natural environment to satisfy the demand of human society development by turning wildness into lands with explicit socio-economic functions (e.g., lands used for residential, farming, forestry or industrial purposes). The changes of earth observations are described as the land use and land cover (LULC) change process. Land cover (LC) describes the overlays or current covers of the ground, such as the vegetation, bare sold, hard surface, etc (Di Gregorio and Jansen, 1997), and land use (LU) describes the ways that human beings make use of and manage the land and its resources. In comparison, LC is easily observable and directly describes the earth observation, while LU is difficult to observe and to represent the function of land in terms of the human living. Given the intrinsic connection between LU and human activities, the analyses of LU and land use change (LUC) are more frequently adopted to study the relationship between human society and the natural environment.

LUC process has profound impacts on the economic development and social process. The land use is one of the three major factors of production in classic economics along with labor and capital; land use is the backbone of agricultural economies and offers

both benefits and challenges for the economic development and social progress (Wu et al., 2008). For example, the LUC from forest to agriculture guarantees the food supply for the growing population, but it may also lead to ecosystem degradation (Lubowski et al., 2006); the LUC from agriculture to built-up promotes urbanization, but it may affect the living of rural people, particularly those living at the urban fringe (Lisansky, 1986).

LUC is also generally considered to be the single most important factor that affects the ecosystem health (Hunsaker and Levine, 1995). LUC alters the fluxes of mass and energy in the ecosystem, which has consequences for ecological structure, functioning and the flow of ecological goods and services (Bockstael et al., 2000). Out of the various possible consequences, pollution and climate change are the two most noteworthy problems. LUC from natural lands to agriculture or built-up usually increases the discharges of nutrients, toxics, or the other chemical substances generated by the irrigation or industrial production into water bodies, and also increases the emission of air pollutants generated by the household, transportation or industrial production. Moreover, LUC affects the climate change in various ways: deforestation (Le Quéré et al., 2009), the changes of atmospheric conditions (Pielke et al., 1998), burning of fossil fuels. Previous researches shows that cities, which bear the most intensive human activities account for about 80% of the world's carbon emissions (Wu, 2008).

Given the great influence of LU on human society and ecosystem, the management of LU has attracted broad attention from the governments and organizations. The United Nations reported that the management of LU is critical to achieve the 2030 agenda of sustainable development in the 2017 report [1]. In particular, the report emphasizes the importance of LU management to achieve the Goal 11 (sustainable cities and communities) and the Goal 15 (life on land), which focus on the urbanization and sustainable usage of natural resources, respectively.

[1] http://www.undp.org/content/undp/en/home/presscenter/pressreleases/2017/09/11/better-land-use-and-management-critical-for-achieving-agenda-2030-says-a-new-report.html

LU planning is an essential tool for managing the land use. It refers to the decision-making process in which a society decides where socio-economic activities, such as commerce, agriculture and housing should take place. By prioritizing and restricting the existence of certain LU for specific area, and also controlling the driving forces of certain LU types, LU planning grants to the possibility to actively control the intensity, location and timing of LUC to some extent; it allows to search the optimal LUC scenario to maximize the benefits and to minimize the negative impacts of LUC. LU planning has been used as the managing tool by the national and local governments. The purpose and focus of LU planning have evolved from the sole management and control strategy of urbanization to a combination of strategic and environmental planning that consider human, animal and vegetation life (Walters, 2007).

Although LU planning is primarily an economics and management problem, due to the complexity of urban LUC process, the LU planning needs insight from geographical and environmental studies to produce reliable and appropriate planning scenario, and LUC modeling provides an effective tool. LUC modeling uses mathematical methods to simulate and/or forecast the LU pattern by assimilating various driving factors of LUC process. LUC modeling is mainly used to provide the quantitative evidence for LU planning by either forecasting the expected LU pattern given the historical trend, or to examine the possible LU pattern under different development scenarios by combined with scenario-based approach. The development of urban LUC modeling actually has a longer history than LU planning. The primitive urban LUC modeling problem had been proposed and studied by von Thunen's classical model of agricultural location in 1826. In the past two centuries, various modeling theory and techniques have been developed and continuously upgraded. The contemporary urban LUC modeling theory, which has been proposed and quickly adopted by the researchers since 1980s, views cities as self-organizing systems that exists in a constant exchange of goods and energy within its territory.

Of the various models that are developed based on the self-organizing system theory,

cellular automata (CA) is the simplest but most popular model. CA is a special type of automata that are arranged in regularly tessellated space (usually in 2D grid). Each cell in the tessellated space holds a state and interacts with its neighboring cells. The information flow between cells are controlled by neighborhood rules, i.e., the state of a certain cell would change if the characteristics of the neighborhood meet certain conditions. Although the mechanism is simple, CA is demonstrated to be able to simulate complex system by stacking a series of transition rules. Moreover, CA is naturally compatible with the geographical information system (GIS) dataset and techniques, which is rapidly developing since 1990s because of the tessellated space design. The integration of CA and GIS has greatly enhanced the strength of CA to describing the LUC process.

However, when dealing with relatively complex urban system, CA can hardly capture all the driving factors as neighborhood rules and cannot efficiently handle the geospatial data. Hence, various variants of CA are developed to overcome this limitation. Transition rules (neighborhood rules) are the core component of a CA model, which represent the logic of the LUC process and hence determine the spatial dynamics of the system (White and Engelen, 2000). The variants of CA mainly enhance the certain aspects of CA by modifying, transforming or extending the mechanism of constructing neighborhood rules. Some classic variants includes 1) constrained CA, which took into account the constraints of the qualities of the lands, the effects of neighboring LU activities, and the aggregate level of demand for each LU by integrating a macro-scale socio-economic constraint model; 2) the SLEUTH model, which incorporates six driving factors (slope, land cover, exclusion, urbanization, urbanization and hill shade) and define four types of urban growth (diffusion, breed, spread, road gravity and slope); 3) integrated statistical and CA model, which uses statistical learning model to predict the transition probability, and then uses CA model to simulate the LU pattern based on the transition probability map. In terms of their designs, the SLEUTH model builds a modeling system based on the urban development theory, and then feed empirical data into the modeling system for calibration; both constraint CA and integrated statistical and CA model incorporate

technique used in the other fields to enhance the CA, but the difference is that constraint CA adopts a hierarchical structure while integrated statistical and CA model resembles a loosely connected pipeline.

Statistical learning, or machine learning, is a field of statistics and computer science that gives computer systems the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed (Samuel, 1959). In terms pf data analytics, machine learning is used to devise complex models and algorithms that lend themselves to prediction; this usage of machine learning is also known as predictive analytics. Some classic algorithms includes decision tree (DT), logistic regression (LR), support vector machine (SVM), neural networks (NN), naive Bayes, Bayesian networks, etc. Over the last few decades, most of these algorithms have been applied in LUC modeling as the form of either standalone application or integrated model with CA and/or other mathematical models. In previous studies, the models have been applied in various cities and regions such as São Paulo (Almeida et al., 2008), Missouri State (Liu and Seto, 2008), the Beijing-Tianjin-Tangshan Metropolitan Area (Kuang, 2011), and Athens (Grekousis et al., 2013).

Out of the three classic variants, integrated statistical and CA approach may have the highest flexibility and the highest potential of further improvement. The variety and scalability of statistical learning methods allow for flexible model integration to accommodate the varying incentives and focuses of different LUC modeling tasks. For example, the integrated model of CA and LR or DT can provide both reliable LU pattern prediction and interpretation on the effects of different driving factors. On the other hand, LR and DT can be replaced by SVM or NN when the predictive power of LUC model is considered as an important determinant of the model selection for certain LUC modeling tasks.

Moreover, statistical learning is rapidly developing in recent years, particularly in terms of the development and spreading of ensemble models, and also the deep neural

networks. These methods have been applied in other fields that have close relationship with LUC modeling such as remote sensing, and are demonstrated to be able to yield better performance than the classic statistical learning methods. Therefore, these more advanced methods may further enhance the existing modeling techniques.

Compared with most of other LUC approaches, the integrated statistical and CA approach is mainly data-driven and has relatively less solid theoretical foundation with respective to the LUC and urban development theory. In recent years, the spatial data with good quality is growing and becoming easier to access (e.g. Landsat imagery, OpenStreetMap); the power of statistical approaches would be further amplified if the rich spatial data can be appropriately utilized. Furthermore, by assimilating spatial data that captures driving factors of LUC process including accessibility, neighborhood characteristics, elevation, slope, etc. The statistical learning models actually can be viewed as a approximation to the LUC models that are based on classic urban development theory such as SLEUTH with using empirical approach.

Although the LUC modeling is gaining popularity in various fields (e.g., geosciences, urban planning, ecological modeling, etc.) and the number of publications is continuously increasing in the recent years, there are several limitations. In terms of the incentive of LUC modeling studies, the methodological studies of LUC modeling generally receive less attention than the application studies. The existing LUC models are proven to be able to provide reliable support for some application cases, such as the urban expansion modeling at fast-developing area or the scenario-based LUC modeling. However, these applications either have relatively simple transition rules to be modeled or have fewer requirements for the predictive accuracy. In order to deal with LUC modeling in a more complex system such as the highly-developed urban areas, methodological research on developing LUC models with higher predictive power is necessary. The existing studies focusing on the integrated statistical approach mainly improve the predictive performance of LUC models by combining or modifying classic statistical learning methods such as LR, rather than incorporating more advanced statistical learning methods such as ensemble models and

deep learning. In addition, most methodological studies mainly focus on the suitability and predictive performance of the developed LUC model and lack an analytic description on the characteristics of certain LUC modeling problems.

In order to address the aforementioned limitations, this study conducts a methodological research on the LUC modeling with using advanced statistical learning methods and CA in a highly developed metropolitan area – the Greater Tokyo Area. Rather than to solely examine the suitability of specific statistical learning methods, this study aims to utilize these statistical learning methods to provide an insight into the characteristics of LUC modeling In particular, this study explores the effect of stochastic modeling and the enhancements on temporal and spatial modeling.

To explore the effect of stochastic modeling, this study develops four tree-based models with the mutual basic design but varying degree of randomness, namely bagging trees (BT), bagged gradient boosting decision tree (bagged GBDT), random forests (RF) and extremely randomized trees (ERT), to simulate the multiple LUC process in the Greater Tokyo Area with considering a total of 18 LU transition types. By examining and comparing their predictive performances of the four tree-based models, this study discusses the benefit of stochastic mechanism of learning algorithm for improving the predictive performance of LUC models. In addition, the explanatory abilities of different driving factor categories for the complicated multiple LUC processes are demonstrated using the results generated from tree-based models.

In order to enhance the spatial and temporal modeling of LUC process, this study develops the convolution neural networks (CNN) based LUC models and recurrent neural networks (RNN) based LUC models, respectively. Both models incorporate advanced deep learning methods into LUC modeling, but the mechanisms of incorporation are distinctively different; the CNN based models extract spatial information directly from the satellite images to improve the performance of transition probability estimation, and then determine the transition rules by combining with CA. On the other hand, the RNN based

models are standalone models, which are capable of capturing the long-term temporal variation from a time series LU data covering a time span of 17 years from 2000 to 2016. The convolutional-based models are developed for modeling the LU transition between agriculture, forest and built-up in Saitama prefecture; and the RNN models are developed for modeling the LU transition from non-built-up to built-up in Tsukuba city. This study also adopts the comparative approaches to show the improvement from using the deep learning techniques and to provide insights on how the deep NN improves the transition rules determination. In terms of the CNN based approach, this study developed a hybrid CNN model and a convolutional denoising autoencoder (CDAE) model to show the different spatial feature extraction processes. Also, in terms of the RNN based approach, this study developed four RNN models that belong to two categories of RNN variants: 1) simple RNN model, which is the basic variant of RNN, 2) RNN variants with gated architecture: long short-term memory (LSTM) model, long short-term memory (LSTM) with peephole model, and gated recurrent unit (GRU) model. These models are used to demonstrate the importance of being able to learning long-term temporal dependency for LUC modeling.

Chapter 2

Modeling with tree-based algorithms in Greater Tokyo Area

2.1 Motivation

Although a wide variety of land use change (LUC) models were developed and utilized in previous studies, the majority of these studies only focused on the modeling of urbanization / urban expansion / urban growth / urban sprawl (e.g. Wang and Mountrakis, 2011; Al-sharif and Pradhan, 2015; Berberoğlu et al., 2016). The previous studies simplified the urban dynamics into a plain binary transition process from non-built-up to built-up lands. The modeling of binary transitions is insufficient to reflect real-world LUC processes, and it cannot support the analyzes of urban phenomena such as urban renewal and urban decay. To address these issues, multiple LUC modeling, which enables the consideration of the transitions between various natural and built-up lands, is required.

Certain studies used multiple LUC modeling in urban areas. Most previous studies have focused on transitions from natural land use types to built-up types (e.g. Li and Yeh, 2002; Camacho Olmedo et al., 2013). However, there is a growing number of studies

that model the transitions among built-up types to expand the coverage of transition types (Almeida et al., 2003, 2008; Zheng et al., 2015). In addition, previous studies mainly focused on fast developing urban areas such as Adana in Turkey (Berberoğlu et al., 2016), Guangzhou (Chen et al., 2014) and Shenzhen (Chen et al., 2016) in China. Thus, limited work is available for highly developed areas with characteristics significantly different from those of fast developing areas (e.g. vast built-up areas, slow urban growth, land shortage problems and intensive redevelopment activities). These characteristics could lead to a distinct LUC process, but there is little evidence of LUC work focused on developed cities in the literature.

The lack of multiple LUC modeling in a highly developed urban system may be largely due to the following difficulties: 1) Compared with a fast-developing urban system, a highly developed urban system has no dominating driving forces, such as demands imposed by rapid economic growth or urban expansion plans, that can largely explain the LUC process. Instead, both driving forces and spatial patterns are relatively more diverse, which could impose a great challenge in the analysis of the relationship between driving factors and LUC (Irwin and Geoghegan, 2001). 2) The complex transition rule sets require larger and sophisticated spatial variable sets and an effective modeling framework that can handle high-dimensional datasets (Li and Yeh, 2002). 3) Finally, there is a lack of previous knowledge on the explanatory power of spatial variables for different land use transitions.

A variety of statistical learning methods, including logistic regression (LR) (Munshi et al., 2014), neural networks (NN) (Li and Yeh, 2001), support vector machine (SVM) (Yang et al., 2008), decision tree (DT) (Li and Gar-On Yeh, 2004), and multi-criteria evaluation (MCE) (Camacho Olmedo et al., 2013), have been integrated with cellular automata (CA) to analyze LUC in the literature. The predictive ability and interpretability are two major model selection criteria; however, in practice, it is rare for a method to perform well under both criteria. LR and NN are the two most prevalent methods; LR yields easily interpretable results but is not capable of handling complex or large-scale

problems because of the linear design (Li and Yeh, 2002), whereas NN has a strong predictive ability and non-linear design but is barely interpretable (Pijanowski et al., 2002; Guan et al., 2005). DT is also a non-linear model similar to NN but does not possess as remarkable a predictive ability as NN, and its interpretation is not as straightforward as LR, which may explain the relatively small number of applications in LUC modeling (e.g. Li and Gar-On Yeh, 2004; Al-sharif and Pradhan, 2015).

According to statistical studies, the predictive performance of DT can be improved by incorporating ensemble methods, such as bagging and boosting, thereby yielding a variety of tree-based ensemble algorithms that have been proven to be competitive with NN. These tree-based ensemble methods may be a solution to the dilemma concerning predictive ability and interpretability. This possibility can be identified by answering two questions: 1) *Do tree-based methods perform better than strong predictors, such as NN?* 2) *Among the various tree-based methods, with their respective distinct designs, which method is the most suitable for LUC modeling and why?*. Although a few studies used tree-based methods and reported satisfactory predictive performances (Li et al., 2014, 2015a; Kamusoko and Gamba, 2015), the two questions remain unanswered because of the selective application and a lack of a systematic evaluation in these studies.

To fully explore the potential of tree-based methods, this study combines a CA model and 4 tree-based models (bagged trees (BT), random forest (RF), extremely randomized trees (ERT) and bagged gradient boosting decision trees (bagged GBDT)) to simulate the LUC in the Greater Tokyo Area from 2009 to 2014. This study compares the predictive performances of the tree-based models between themselves and with the results obtained under the NN method by using both the area under the receiver operating characteristics (AUC-ROC) curve and the area under the precision-recall (AUC-PR) curve. In addition, with the variable importance evaluation embedded in tree-based algorithms, this study provides an interpretation of the effects of different driving factors. The findings of this study provide insights and evidence regarding model selection and LUC in a highly developed urban system.

2.2 Methodology

2.2.1 Transition probability modeling

2.2.1.1 Tree-based algorithms

DT (decision tree) is an inductive classification method in the form of an inverted tree structure. This method recursively partitions the learning data using a set of "if-then" rules and seeks to obtain the "best" split at each step. DT contains internal nodes that would be further split and leaf nodes that would not, therein connecting them with branches, which represent conjunctions of variables. The recursion is completed once the subset at a node has the same class label or when some pre-set stopping rules are met.

This study uses the CART (classification and regression tree) algorithm (Breiman et al., 1984) to construct the DT models. CART uses the Gini impurity as the metric to measure the quality of a split, which is defined as:

$$I_G(t) = \sum_{i=1}^{c} p(i|t)(1 - p(i|t)) = 1 - \sum_{i=1}^{c} p(i|t)^2 \tag{2.1}$$

where I_G denotes the Gini impurity at a particular node t, $p(i|t)$ is the proportion of the observations that belong to class i for node t, and c is the number of classes for node t. Intuitively, the Gini impurity can be understood as a criterion to minimize the probability of misclassification. CART selects the split rule that maximizes the Gini impurity at the child node of t as the best split rule.

Due to its classification criteria, DT is intrinsically sensitive to the input data structure; hence, the results obtained using DT are unstable and prone to over-fitting (Caruana and Niculescu-Mizil, 2006). On the other hand, practical algorithms are based on heuristic algorithms, such as greedy algorithms, where locally optimal decisions are made at each node, and cannot guarantee that globally optimal decisions will be returned (Ben-

Gal et al., 2014). These two issues represent variance (the error from the sensitivity to changes in a dataset) and bias (the error from incorrect model hypothesis) problems and are usually addressed using boosting or bagging.

The boosted tree method was developed to achieve bias reductions by incrementally building an ensemble by learning each new instance to emphasize the training instances previously mis-classified (Quinlan et al., 1996). This study uses a typical algorithm, GBDT (gradient boosting decision trees), which constructs additive regression models by sequentially fitting single DT models to minimize the current pseudo-residuals by least squares at each iteration (Friedman, 2002). GBDT is well recognised for its outstanding predictive power, but is so vulnerable to noise that its predictions are sometimes not even competitive to single DT when the noise is high (Opitz and Maclin, 1999). On the other hand, bagging aims to improve the accuracy through variance reduction (Breiman, 1996). BT (bagged trees) forms multiple versions of DT models by repetitively fitting them to subsample sets drawn from bootstrap sampling (random sampling with replacement), and then, it aggregates their predictions by averaging.

RF (random forests) (Breiman, 2001) is a widely applied learning algorithm in various fields. This method is an enhanced version of the standard bagged trees method, with additional randomness imposed at the split selection step. It also trains multiple CART models from bootstrap replicas of the samples, but it derives the optimal split by searching a random subset of candidate variables at each node. RF generally outperforms the standard bagged trees method and is more robust than boosting trees with respect to noise (Caruana and Niculescu-Mizil, 2006).

In ERT (extremely randomized trees) or ET (extra-trees) (Geurts et al., 2006), the randomization is extended compared with RF in that both the variable and the cut point are selected at random when splitting a node. This method is based on the rationale that the extreme randomization of the cut point and variable combined with ensemble schemes should be able to further reduce the model's dependence on the data structure and hence

improve the model's generalization performance. The standard ERT algorithm published by Geurts et al. (2006) uses the whole learning sample rather than a bootstrap replica to grow trees to compensate for the accuracy loss caused by randomness. However, this study discards this change to further increase the randomness of the algorithm to provide a matched comparison with the results obtained using BT and RF.

To mitigate the impact of noise produced by the irrelevant variables, this study performed variable selection using the variable importance evaluation embedded in those tree-based learning algorithms. The evaluation is based on the idea that the relative rank (i.e., depth) of a variable used as an internal node explains the relative importance of the variable. The importance score of a variable is defined as the normalized total reduction of the Gini impurity produced by that particular variable. This study sorted the obtained variable importance in descending order, calculated the accumulation, and selected the variables within 95% of the total accumulation.

This study also performed hyperparameters optimization using a grid search method to avoid over-fitting and improve the model's performance. The hyperparameters of DT include the tree depth, minimum samples in a leaf node, minimum samples for a split and minimum impurity for a split. In addition to these basic hyperparameters, this study further tuned the tree numbers and bootstrap sampling ratio for RF and ERT and the tree numbers and learning rate for GBDT.

2.2.1.2 Multi-layer perceptron

MLP (multi-layer perceptron) is a basic feed-forward neural network and consists of one input layer, one or more hidden layers and one output layer. These layers are fully connected by a set of weights, which are learned and updated by the back-propagation algorithm. MLP's ability to learn and generalize depends on its architecture (number of hidden layers and nodes) and on the hyperparameters (learning rate, etc.).

In this study, this study uses cross-entropy as the loss function, ReLU (rectified linear unit) as the activation function and the mini-batch gradient descent algorithm as the optimizer to train the model. A set of hyperparameters were tuned in the training process to achieve the optimal generalization performance, including the number of hidden layers, learning rate, training epoch, mini-batch size, momentum, learning rate decay, L2 regularization and dropout ratio. In addition, each mini-batch was specifically designed to have the same ratio between samples with changed or unchanged state of land use over time as the original dataset, while training.

2.2.1.3 Bootstrap sampling and aggregating

The dilemmas of spatial autocorrelation and sample representativeness are tricky issues in spatial modeling (Hirzel and Guisan, 2002; Munroe et al., 2004). Previous studies usually adopted either random sampling which can well represent the population (Xie et al., 2005; Huang et al., 2009; Chen et al., 2014), or stratified random sampling which seeks to achieve a balance between spatial autocorrelation and sample representativeness (Arsanjani et al., 2012; Mozumder et al., 2016). However, these methods may still lead to a poor representative sample when applied to the multiple LUC problem, in which the spatial heterogeneity tends to be much higher than in typical cases of binary LUC.

This study uses an approach based on the idea of bagging, which seeks to reduce the sampling error by incorporating the whole spatial dataset into the model training in an ensemble approach. The procedure is as follows: 1) use bootstrap sampling to repetitively split the whole sample set into a training set and a holdout set at a ratio of 0.35:0.65, 2) learn a basic predictor on the training set and subsequently obtain the prediction of the holdout set at each iteration, and 3) aggregate the predictions of the holdout set by performing averaging when the iteration is finished. The standard bagging method is used to address the disadvantages of DT algorithms. This approach has a similar procedure as the bagging method but possesses the advantages of being able to reduce the possible

sampling error using bootstrap sampling and aggregating.

In practice, this study uses DT, GBDT and MLP as the basic predictors and obtain bagged DT, bagged GBDT and bagged MLP models. Since RF and ERT are already bagging ensembles of DT, this study directly uses them to implement the bootstrap sampling and aggregating approach rather than wrapping them into other bagging schemes. Finally, five models (DT, RF, ERT, bagged GBDT and bagged MLP) were developed for the transition probability prediction.

2.2.2 CA model

The CA model used in this study is based on the well-developed CA model DINAMICA (Soares-Filho et al., 2002), which has been widely applied in ecological and urban LUC modeling (e.g. Almeida et al., 2003; Pérez-Vega et al., 2012; Rossetti et al., 2013). DINAMICA defines two main vicinity-based transitional functions, expander and patcher, to simulate the land use patch dynamics in a stochastic multi-step approach. The expander function is dedicated to the expansion or contraction of the previous patches of a certain land use class, and the patcher function is designed to generate new patches. The two processes are merged using the following calculation:

$$Q_{ij} = r \times expander + s \times patcher \tag{2.2}$$

where Q_{ij} is the total number of transitions from land use class i to j; r and s are the percentages performed by the expander and patcher functions, respectively; and $r + s = 1$. The patch size is drawn from a log-normal distribution, and the patch shape or compactness is determined by a parameter named isometry.

In this study, the simulation was iterated over the modeling period on a yearly basis. The total number of transitions is determined by a simple operation of cross-tabulation based on initial and final land use maps and is then assigned to each iteration in an

average-based manner. The simulated map is updated by each iteration according to the results obtained by the expander and patcher functions. Both functions use a stochastic selection mechanism to select seeds (the centre cell of a transition patch), which prioritize high transition probabilities over low transition probabilities with a certain degree of randomness, where the randomness strength can be adjusted.

This study performed some modifications to the expander function to adapt it to the LUC problem of this study. The modified expander function is defined as:

$$\begin{aligned} &if \quad n_j > 3 \quad or \quad P(ij)(xy) > t \quad then \quad P'(ij)(xy) = P(ij)(xy) \\ &else \quad P'(ij)(xy) = p(ij)(xy) \times \sqrt{\frac{n_j}{4}} \end{aligned} \tag{2.3}$$

where $P(ij)(xy)$ denotes the transition probability from land use class i to j, t denotes a preset threshold, and n_j denotes the number of cells of land use class j occurring in a 3×3 window.

The expander function can be regarded as a penalty mechanism on cells that have relatively few neighbors of land use j. The penalty here is specifically designed to be lighter than that in the original expander function by taking the square of the original coefficient $n_j/4$ and establishing a safe zone based on a preset threshold. The rationale behind this penalty abatement is that 1) the impact of neighboring land use has already been controlled in the transition probability prediction by employing land use enrichment factors, and 2) misclassification mainly occurs at cells with low certainty (low transition probability) rather than cells with high certainty (high transition probability) according to previous studies (Gong et al., 2015; Li et al., 2015a).

2.2.3 Validation

2.2.3.1 Validation of statistical learning models

This study uses the receiver operating characteristics (ROC) curve and the precision-recall (PR) curve, specifically the corresponding areas under the curves (AUC-ROC and AUC-PR), to evaluate the predictive performances of the tree-based learning methods. Both metrics are calculated based on confusion matrix.

ROC is a well-documented model assessment tool. It depicts the trade-off between the false-positive rate (FPR) and the true-positive rate (TPR) with varying thresholds, and the AUC-ROC is used as a quantitative measure for assessing the classification performance. Specifically, an AUC-ROC value of 0.5 is the random baseline, and values below 0.5 indicate a systematically incorrect model (Jansen and Veldkamp, 2012). ROC is frequently used to evaluate the quality of the transition probability (Pontius and Schneider, 2001). However, while ROC describes the performance of both negative and positive classes, which correspond to unchanged and changed areas in the context of LUC modeling, the main interest is usually the simulation performance of changed areas (Pérez-Vega et al., 2012; Gong et al., 2015). Moreover, AUC-ROC can present an overly optimistic view of an algorithm's performance if the data are highly imbalanced (Davis and Goadrich, 2006).

Hence, this study uses AUC-PR to address the above-mentioned problems. AUC-PR is a sequence of precision and recall values with varying thresholds. It provides a more specific assessment of a model's ability to predict changed areas and exclude the influence of data imbalance. It should be noted that although AUC-ROC and AUC-PR are closely related, the algorithms that optimize AUC-ROC are not guaranteed to optimize AUC-PR (Davis and Goadrich, 2006).

2.2.3.2 Validation of CA simulation

This study performs both cell-to-cell and vicinity-based assessments to evaluate the simulation quality using the overall accuracy and Cohen's Kappa statistics for the cell-to-cell assessment and fuzzy Kappa statistics for the vicinity-based assessment. This study uses the fuzzy Kappa over other metrics (e.g. the multi-resolution fitting procedure (Costanza, 1989) and the fuzzy similarity Hagen (2003)) because of its intrinsic connection to classic Kappa, which enables comparison between the cell-to-cell and vicinity-based performances.

The classic Kappa statistics describe the agreement between two observation categories by excluding the agreement due to chance (Bennett et al., 2013) and is defined as:

$$K = \frac{p_0 - p_e}{1 - p_e} \tag{2.4}$$

where p_0 is the relative observed agreement and p_e is the hypothetical probability of chance agreement. In this binary case

$$p_0 = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.5}$$

$$p_e = \frac{TP + FP}{TP + TN + FP + FN} \times \frac{TP + FN}{TP + TN + FP + FN} + \tag{2.6}$$

$$\frac{TN + FP}{TP + TN + FP + FN} \times \frac{TN + FN}{TP + TN + FP + FN} \tag{2.7}$$

where TP denotes true positive, FP denotes false positive, TN denotes true negative, FN denotes false negative. In terms of the implication of the magnitude of the Kappa coefficient, a value above 0.8 represents a strong agreement between two maps, a value between 0.6 and 0.8 represents a somewhat strong agreement, a value between 0.4 and 0.6 represents a moderate agreement, and a value below 0.4 represents a weak agreement (Landis and Koch, 1977).

Fuzzy Kappa has the same rationale as classic Kappa, but it introduces fuzzy set

theory to account for the category similarity and neighborhood similarity by attributing some degree of agreement to similar land use categories (e.g. low-density and high-density low-rise buildings) at corresponding locations or in a neighborhood. This study only considers the neighborhood similarity in this study because the category similarity may exaggerate the performance of a simulation. A Gaussian distance decay function is used to specify the agreement level with respect to distance within a neighborhood. In addition, This study uses a two-way similarity approach recommended by Hagen (2003) to avoid over-estimation caused by the overpowering influence of the similarity between neighborhoods. This approach would significantly reduce the degree of agreement if the centre cells of two maps do not belong to the same category. For a more detailed demonstration and mathematical description of fuzzy Kappa and the two-way similarity, refer to Hagen (2003).

2.2.4 Model framework

Figure 2.1 shows the model framework. The multiple LUC modeling is disaggregated by the initial land use classes and then divided into several sub-tasks with an identical land use class within each sub-task. These sub-tasks are separately performed with the integrated model. The outputs are aggregated to produce the final land use map at the end of the process.

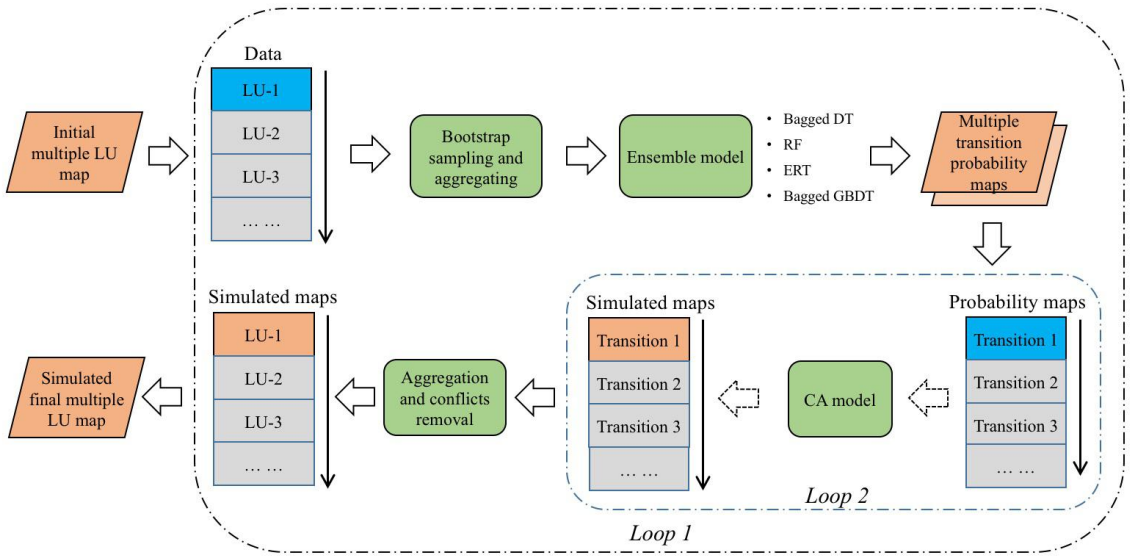


Figure 2.1: Modeling framework

Cells with a certain land use class could undergo different transitions, leading to a multi-classification problem in statistical learning. The One-versus-Rest (OvR) approach, which trains a single predictor per class, with the samples of that class as positive samples and all other samples as negative samples, is used to address this issue. However, the OvR approach has a shortcoming in that it may aggravate the data imbalance problem (i.e., observations of some classes are significantly fewer in number than those of other classes). To avoid this problem, this study uses a class weight adjustment, which assigns higher weights to minority classes according to their proportions and hence ensures that minority classes obtain greater focus by the predictor.

Four tree-based learning methods are used to achieve the transition probability prediction, but only the method with the highest predictive performance would be used to produce the transition probability maps for further simulation.

Given that the CA model can only simulate one transition at a time, the already

disaggregated multiple LUC processes are further divided into binary transitions. The binary transitions are separately simulated by the CA model, and then, the simulation results are aggregated. To resolve conflicting locations that may be identified during the aggregation process, this study 1) removes the contenders that are not selected as seeds in the corresponding expander or patcher simulating process and 2) selects the survivor with the highest transition probability.

2.3 Implementation

2.3.1 Study area and land use data

The study area is the Greater Tokyo Area in Japan, which is a highly developed metropolitan area. This area consists of Tokyo city and its three surrounding prefectures; the area is approximately 18,500 km^2 and contains a population of approximately 38 million. This area is the second largest single metropolitan area in the world in terms of built-up or urban function landmass, approximately 8,500 km^2. The Greater Tokyo Area experienced rapid LUC until the 1980s; subsequently, the change rate decreased dramatically after the asset bubble burst. The LUC rate has remained low since then.

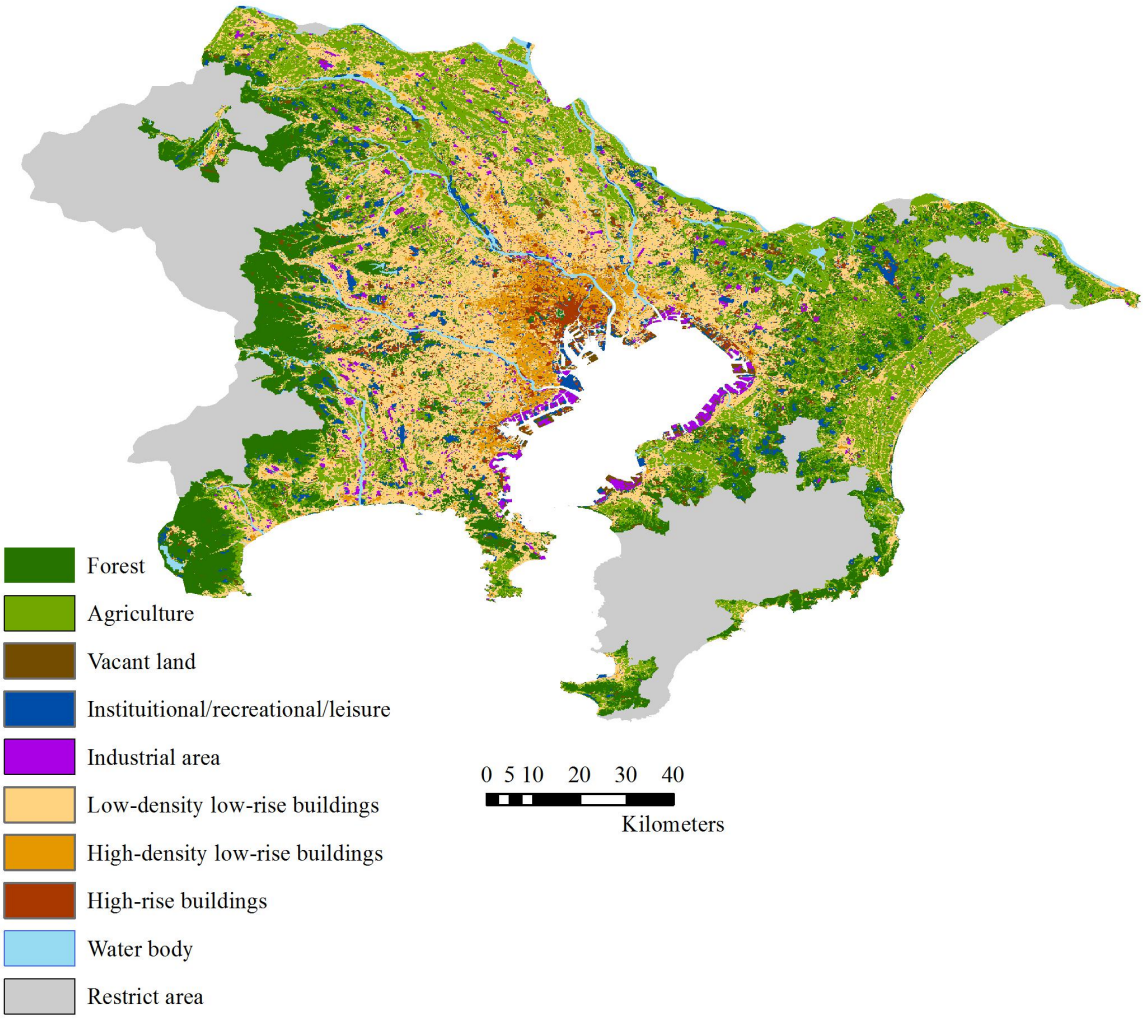


Figure 2.2: Land use map of Greater Tokyo Area in 2009

Table 2.1: Land use summaries of 2009 and 2014

	2009	2014
Agri.	362,134 (19.64%)	345,382 (18.73%)
Forest	287,771 (15.61%)	284,075 (15.41%)
Vacant	47,489 (2.57%)	43,705 (2.37%)
High-rise bldgs.	32,316 (1.75%)	47,985 (2.60%)
Industrial	30,987 (1.68%)	29,175 (1.58%)
Low-density and low-rise bldgs.	443,373 (24.04%)	447,311 (24.25%)
High-density and low-rise bldgs.	52,284 (2.83%)	52,577 (2.85%)
IRL	66,225 (3.59%)	66,941 (3.63%)
Water body	45,952 (2.49%)	52,330 (2.83%)
Restricted area	475,403 (25.78%)	474,453 (25.73%)

This study uses finely classified land use maps of 2009 and 2014 from a geographical project of the Ministry of Land, Infrastructure and Transportation of Japan [1]. The land use maps have a resolution of 100×100 m, which is acceptable for a large-scale study. This study aggregated the 17 land use classes in the original land use maps into 10 classes by merging similar categories. The 10 land use classes are agriculture, forest, vacant land, institutional/recreational/leisure (IRL) lands, industrial area, low-density low-rise buildings, high-density low-rise buildings, high-rise buildings, water body and restricted area (see Figure 2.2 for the spatial distribution of LU categories and Table 2.2 for the definition of LU categories). This study excludes restrict area and water body from consideration, and are left with 8 land use classes for analysis. In the Greater Tokyo Area, high-rise building areas are mainly located in the central area. Low-density and low-rise building areas, agriculture and forest are scattered across the non-central area.

[1] Online access: http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L03-b-u.html

Table 2.2: Definition of land use categories in Greater Tokyo Area for 2009 and 2014

LU category	Description
Agriculture	Lands used for growing crops including wet paddy filed, dry paddy field, swamp paddy field, fields used for growing wheat, upland rice, vegetables, fruits, tea, wax tree, paper mulberry, hemp palm, etc., as well as grassland and lawn
Forest	Lands where perennial plants are densely distributed
Vacant land	Areas including vacating man-made open space, wasteland, cliff, wetland, rocky land, mining land, etc.
High-rise and high-density buildings	Residential, commercial and other areas where mansions with four stories and above are densely distributed
Low-rise and high-density buildings	Residential areas where residential buildings with three stories and lower are densely distributed
Low-rise and low-density buildings	Residential areas where residential buildings with three stories and lower are loosely distributed
Industrial area	Lands where buildings used for manufacturing production are distributed
Institutional / recreational / leisure lands	Areas including playground, air port, racecourse, baseball ground, schools, well-maintained park and green area (water body excluded), golf course, etc.
Water body	Areas including river and river bed, artificial lake, natural lake, pond, fish farm, etc. where are filled with water for most of the time
Restricted area	Areas where LU information is not available

Notes:
The definitions of LU categories are based on the information provided by National Land Information Division, National and Regional Policy Bureau of Japan
(http://nlftp.mlit.go.jp/ksj/gml/codelist/LandUseCd-09-u.html)

Table 2.3: Transition matrix from 2009 to 2014

2009 \ 2014	Agri.	Forest	Vacant	High-rise bldgs.	Industrial	Low-density and low-rise bldgs.	High-density and low-rise bldgs.	IRL
Agri	.	7522 (2.10%)				20774 (5.70%)		
Forest	9083 (3.20%)					6623 (2.30%)		
Vacant		2267 (4.80%)		1066 (2.20%)	701 (1.5%)	4200 (8.80%)		2110 (4.40%)
High-rise bldgs.						1727 (5.30%)	1187 (3.70%)	
Industrial			1248 (3.90%)	605 (2%)				1426 (4.60%)
Low-density and low-rise bldgs.				14055 (3.10%)				
High-density and low-rise bldgs.				2347 (4.50%)				
IRL		2066 (3.10%)		3841 (5.80%)				

Of the various possible land use transitions that could occur from the 8 land use types, this study excluded transition categories with a frequency of less than 1% of the total number of transitions. The remaining 18 transitions are summarized in Tables 2.1 and 2.3. LU transitions seem to be rare events, as most transition frequencies do not exceed 5%, yielding a highly imbalanced dataset (i.e., the unchanged land use class has far more observations than the changed land use class). Typical urban phenomena are reflected in the land use transitions. For instance, transitions from agriculture/forests to low-density and low-rise buildings reflect urban expansion, transition from low-rise buildings to high-rise buildings reflects urban renewal, and transition from IRL to forests reflects urban decay.

2.3.2 Driving factors of LUC

This study constructed a variable set with approximately 100 dimensions. The variables can be categorized into neighborhood land use enrichment, accessibility, physical factors, and socio-economic and policy factors (see Table 2.4). The spatial data were collected from the Ministry of Land, Infrastructure and Transportation of Japan, OpenStreetMap and the SRTM3 (Shuttle Radar Topography Mission 3) database.

The land use enrichment factor is a measure developed by Verburg et al. (2004) and extended by (Liao et al., 2016) to characterize the over- or under-representation of land use types in a neighborhood. This measure is defined as:

$$F_{l,d,i} = \frac{n_{l,d,i}/n_{d,i}}{N_l/N} \tag{2.8}$$

where $F_{l,d,i}$ is the enrichment of neighborhood d of location i with land use type l, $n_{l,d,i}$ is the number of cells of land use type l in the neighborhood d of the centre cell i, $n_{d,i}$ is the total number of cells in neighborhood d, N_l is the number of cells with land use type l, and N is the total number of cells in the map. This study calculated the enrichment

factors in Moore neighborhoods of varying sizes with radii of 1, 3, 7, 20, and 55 cells to capture the characteristics of both small and large neighborhoods.

This study also considered a variety of socio-economic and policy factors which have been demonstrated to play important roles in LUC modeling (Guan et al., 2005; Puertas et al., 2014; Xu et al., 2016). The socio-economic factors include *population density*, *land prices*, *GDP*, *floor area ratio (FAR)* and *administrative district dummies*, and the policy factors include *development planning* and *LU zoning*. *Land prices* were interpolated from a vector map with a sequence of survey locations and land price data using the ordinary Kriging method. *FAR* represents the free population capacity characterizing a building in an area and is calculated as the difference between the permitted maximum FAR and the current FAR.

The Greater Tokyo Area is composed of various areas, from central capital area to more remote suburbs that vary greatly in city orientation, development level, and demographic characteristics. This study uses the variation in socio-economical factors to capture the spatial variation within the Greater Tokyo Area; *administrative district dummy* and *development planning* are used to control for the difference in city orientation, *GDP* and *land price* reflect the variation in development level, and *population density* helps to control varying demographic characteristics.

2.3.3 Programming environment

The whole modeling process was implemented in Python. In addition to the basic scientific computation libraries of Numpy, Pandas and Scipy, Arcpy (ESRI Inc.), Pykrige (Murphy, 2014) and Pysal (Rey and Anselin, 2010) were used for the geographical data processing and spatial analysis, Scikit-learn (Pedregosa et al., 2011) was used to develop the tree-based models, and Tensorflow (Abadi et al., 2016) was used to develop the MLP models. The full implementation in Python yields an efficient workflow and allows for

great flexibility in making modifications to the methods developed in previous studies for adapting to specific LUC modeling problems.

2.4 Results

2.4.1 Assessment of transition probability prediction models

The application and examination of various statistical methods are the main focus of this study. This study provides 1) a comparison of different statistical learning algorithms as basic predictors of the transition probability and 2) the combination of these algorithms with the bootstrap sampling and aggregating approach. In the calibration stage, the dataset is split into a training set and a test set at a ratio of 0.8:0.2. All models were built upon the training set and evaluated on the test set. The training and test sets retain the same ratio of samples, with unchanged or changed states of land use over time, as the raw dataset. This avoids an over-estimated model performance due to the data imbalance. DT, GBDT and MLP were used as basic predictors and were trained on the same sub-sample, at a proportion of 35%. The outputs were binarized to calculate the AUC-PR and AUC-ROC curves. Table 2.5 summarizes the results.

Overall, the AUC-ROC curves of all the models are above 0.5, confirming the validity of these models. DT has substantially lower AUC-ROC curves than MLP but exhibits a better performance than MLP with respect to AUC-PR. Given that the PR curve and ROC curve share a mutual index of TPR (recall) and differ in the use of precision and FPR, the possible reasons for this result are that DT tends to sacrifice overall accuracy to achieve high precision, whereas MLP better focuses on overall accuracy, and / or that DT is more effective in resolving the data imbalance problem. These tendencies should be associated with the different objective functions and optimization strategies; however, an explicit mathematical explanation is beyond the scope of this study. The

predictive performance of DT is greatly improved when using boosting, especially in terms of AUC-ROC. GBDT outperforms MLP on both metrics, but its performance seems to be problem specific, i.e., in a few cases, such as transitions beginning from IRL, the AUC-PR of GBDT is lower than those of DT. The possible large amount of noise in these problems may be responsible for this phenomenon.

Next, this study combined the basic predictors with bagging-based ensemble methods. Table 2.6 shows substantial improvements in the prediction metrics for the basic predictors when integrated with bagging-based methods. From DT to BT, the average and standard deviation of AUC-ROC over various transition problems improved from 0.64 ± 0.07 to 0.91 ± 0.06, and those of AUC-PR changed from 0.32 ± 0.15 to 0.57 ± 0.18. Integrating GBDT with a bagging algorithm does not induce as significant of improvements as does integrating DT with a bagging algorithm; the bagged GBDT achieves AUC-ROC of 0.92 ± 0.06 and AUC-PR of 0.58 ± 0.24, which are similar to those obtained under BT. This result demonstrates that the advantage of GBDT over DT could be canceled out when being incorporated into a bagging scheme.

Table 2.4: Spatial variables

Categories	Factors
neighborhood characteristics (with varying neighborhood sizes)	neighborhood enrichment of forest
	neighborhood enrichment of agriculture
	neighborhood enrichment of vacant land
	neighborhood enrichment of institutional land
	neighborhood enrichment of recreational/leisure land
	neighborhood enrichment of low-density and low-rise buildings
	neighborhood enrichment of high-density and low-rise buildings
	neighborhood enrichment of high-rise buildings
	neighborhood enrichment of water body
	neighborhood enrichment of restrict area
Accessibility	*Proximity to roads and railways*
	Distance to motorway
	Distance to highway
	Distance to primary road
	Distance to secondary road
	Distance to tertiary road
	Distance to railway
	Distance to subway
	Proximity to regional centres
	Distance to metropolitan area centre
	Distance to city centre
	Distance to district centre
	Distance to town centre
	Proximity to other geographical features
	Distance to river
	Distance to coast line
	Distance to airport
	Distance to university
	Distance to hospital
Physical factors	Coordinates
	Elevation
	Slope
	Hill shade
	Soil type
Socio-economical factors	Population density
	Land prices
	GDP
	Floor area ratio
	Administrative district dummies (categorical, 4 categories)
Policy factors	Urban development plan (categorical, 6 categories)
	Land use zoning (categorical, 13 categories)

Table 2.5: Single predictor comparison

	DT		GBDT		MLP	
	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR
Forest to agri.	0.559	0.142	0.81	0.166	0.719	0.116
Forest to low-density and low-rise bldgs.	0.574	0.172	0.876	0.173	0.822	0.167
Agri. to forest	0.566	0.130	0.868	0.093	0.734	0.086
Agri. to low-density and low-rise bldgs.	0.588	0.196	0.783	0.171	0.752	0.162
Vacant to IRL	0.721	0.483	0.878	0.560	0.857	0.479
Vacant to forest	0.659	0.361	0.867	0.384	0.807	0.312
Vacant to low-density and low-rise bldgs.	0.604	0.223	0.855	0.295	0.823	0.213
Vacant to industrial	0.661	0.411	0.849	0.440	0.782	0.377
Vacant to high-rise bldgs.	0.666	0.332	0.951	0.349	0.938	0.275
IRL to forest	0.582	0.241	0.927	0.221	0.858	0.243
IRL to high-rise bldgs.	0.609	0.231	0.915	0.156	0.932	0.13
Industrial to Low-density and low-rise bldgs.	0.614	0.269	0.863	0.310	0.770	0.200
Industrial to vacant	0.645	0.322	0.803	0.292	0.807	0.268
Industrial to IRL	0.848	0.720	0.964	0.838	0.885	0.784
Low-density and low-rise bldgs. to high-rise bldgs.	0.668	0.361	0.825	0.230	0.807	0.321
High-density and low-rise bldgs. to high-rise bldgs.	0.701	0.424	0.881	0.408	0.817	0.466
High-rise bldgs. to low-density and low-rise bldgs.	0.631	0.323	0.821	0.311	0.792	0.215
High-rise bldgs. to high-density and low-rise bldgs.	0.713	0.460	0.939	0.451	0.865	0.419

Table 2.6: Bagged predictors comparison

	BT		RF		ERT		Bagged GBDT		Bagged MLP	
	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR
Forest to agri.	0.845	0.179	0.896	0.241	0.907	0.257	0.864	0.168	0.87	0.197
Forest to low-density and low-rise bldgs.	0.877	0.255	0.934	0.326	0.94	0.331	0.91	0.242	0.892	0.259
Agri. to forest	0.837	0.175	0.892	0.236	0.899	0.249	0.864	0.153	0.869	0.173
Agri. to low-density and low-rise bldgs.	0.745	0.266	0.792	0.331	0.851	0.335	0.743	0.24	0.756	0.263
Vacant to IRL	0.957	0.777	0.96	0.778	0.975	0.821	0.969	0.807	0.935	0.662
Vacant to forest	0.937	0.646	0.941	0.648	0.954	0.685	0.954	0.695	0.923	0.566
Vacant to low-density and low-rise bldgs.	0.934	0.462	0.935	0.471	0.951	0.505	0.942	0.498	0.909	0.383
Vacant to industrial	0.928	0.659	0.934	0.675	0.946	0.716	0.932	0.704	0.934	0.544
Vacant to high-rise bldgs.	0.964	0.617	0.97	0.632	0.983	0.677	0.976	0.634	0.962	0.549
IRL to forest	0.898	0.646	0.912	0.67	0.926	0.698	0.886	0.612	0.907	0.651
IRL to high-rise bldgs.	0.89	0.667	0.917	0.678	0.932	0.723	0.901	0.613	0.912	0.639
Industrial to Low-density and low-rise bldgs.	0.923	0.944	0.936	0.964	0.944	0.989	0.934	0.93	0.906	0.883
Industrial to vacant	0.968	0.54	0.975	0.55	0.988	0.596	0.963	0.491	0.936	0.476
Industrial to IRL	0.968	0.866	0.975	0.886	0.989	0.916	0.984	0.889	0.977	0.835
Low-density and low-rise bldgs. to high-rise bldgs.	0.941	0.614	0.959	0.638	0.971	0.686	0.957	0.668	0.947	0.613
High-density and low-rise bldgs. to high-rise bldgs.	0.937	0.642	0.949	0.664	0.964	0.699	0.947	0.658	0.937	0.604
High-rise bldgs. to low-density and low-rise bldgs.	0.926	0.566	0.923	0.582	0.936	0.618	0.933	0.595	0.87	0.499
High-rise bldgs. to high-density and low-rise bldgs.	0.964	0.793	0.966	0.813	0.977	0.84	0.967	0.844	0.96	0.724

Among the three bagging ensemble models of DT, BT and RF achieve similar performances, whereas ERT significantly outperforms them, with AUC-ROC of 0.95 ($\pm$ 0.04) and AUC-PR of 0.63 ($\pm$ 0.22). This result indicates that the randomization mechanism of the bagging-based algorithm may affect the performance of the transition probability prediction, but the impact depends on the specifics of the randomization type. Compared with BT, which does not incorporate any randomization in the modeling process, RF introduces a certain degree of randomization by randomly selecting variables for splits to reduce the model's reliance on a specific explanatory variable set; however, this randomization method does not seem to have a significant effect on improving the predictive performance. On the other hand, ERT further incorporates randomization in the cut-point selection, which is approximately equivalent to the random determination of the transition rules for each basic predictor. This particular randomization method helps ERT to achieve the highest predictive performance, which highlights the effectiveness of the random determination of transition rules for representing the stochastic process of LUC.

2.4.2 Assessment of CA simulation

The land use simulation by the expander and patcher functions is a stochastic process, which implies different outputs for each simulation. This study ran the simulation with the same configuration 10 times and obtained the final results by plurality voting. Table 2.7 presents the assessments of the simulation results.

Overall, the Kappa coefficients ranged from 0.41-0.92, which indicates a large variation among the transition types. According to the implication of the magnitude of the Kappa coefficient (Landis and Koch, 1977), the modeling of transitions between forest, agriculture and low-density and low-rise buildings and of transitions from industrial to vacant land and IRL only achieves moderate agreement; the modeling of transitions beginning from vacant land, transitions from IRL to high-rise buildings and transitions from

Table 2.7: Assessment of simulated results

	Overall accuracy	Kappa	Fuzzy Kappa			
			3×3	5×5	7×7	11×11
Forest to agri.	0.958	0.41	0.532	0.575	0.61	0.612
Forest to low-density and low-rise bldgs.	0.971	0.421	0.532	0.581	0.612	0.648
Agri. to forest	0.975	0.409	0.454	0.497	0.523	0.56
Agri. to low-density and low-rise bldgs.	0.959	0.613	0.607	0.621	0.635	0.655
Vacant to IRL	0.997	0.927	0.959	0.97	0.975	0.98
Vacant to forest	0.997	0.898	0.937	0.95	0.955	0.961
Vacant to low-density and low-rise bldgs.	0.977	0.852	0.883	0.9	0.909	0.922
Vacant to industrial	0.98	0.784	0.894	0.932	0.947	0.961
Vacant to high-rise bldgs.	0.993	0.922	0.961	0.974	0.98	0.986
IRL to forest	0.976	0.797	0.803	0.841	0.857	0.88
IRL to high-rise bldgs.	0.996	0.916	0.949	0.962	0.965	0.968
Industrial to Low-density and low-rise bldgs.	0.987	0.843	0.857	0.873	0.884	0.9
Industrial to vacant	0.964	0.554	0.574	0.59	0.62	0.664
Industrial to IRL	0.974	0.53	0.55	0.578	0.598	0.629
Low-density and low-rise bldgs. to high-rise bldgs.	0.988	0.782	0.8	0.844	0.871	0.905
High-density and low-rise bldgs. to high-rise bldgs.	0.972	0.642	0.665	0.684	0.694	0.71
High-rise bldgs. to low-density and low-rise bldgs.	0.962	0.615	0.66	0.687	0.703	0.727
High-rise bldgs. to high-density and low-rise bldgs.	0.976	0.632	0.647	0.659	0.665	0.675

industrial to low-density and low-rise buildings achieves strong agreement.

Fuzzy Kappa explores the possibility that the land use simulation may successfully locate an approximate position of LUC but fail to predict the state of the exact cell. The coefficients increase by 0.01-0.12 from Kappa to fuzzy Kappa with a neighborhood size of 3×3 for land use transitions. In particular, transitions beginning from forest and transitions from vacant land to industrial areas show the largest increase, approximately 0.1. Although the misclassification of exact cells is common when using the expander and patcher functions, these differences are sometimes large enough to imply the incapability of transition probability models to distinguish the sophisticated variations characterizing these particular transitions.

Figure 2.3 shows a comparison of the simulated and actual maps as well as an enlarged map of the central urban area. Figure 2.4 presents the visualization of simulation errors. According to Figure 2.4, incorrectly estimated cells are mainly distributed in the central and northwest parts of Greater Tokyo Area and have very sparse spatial distributions. Error type 2 is the case in which the model fails to predict the future change in land use types, and this error occurs in approximately 2/3 of all cells with any error. Error type 3 is the case in which the model successfully predicts the future change in land use but fails to predict the correct destination, and this error occurs in approximately 3% of all cells with any error. The large difference in the number of cells with error type 2 and error type 3 indicates that predicting the occurrence of land use change is more difficult than predicting the type of transition.

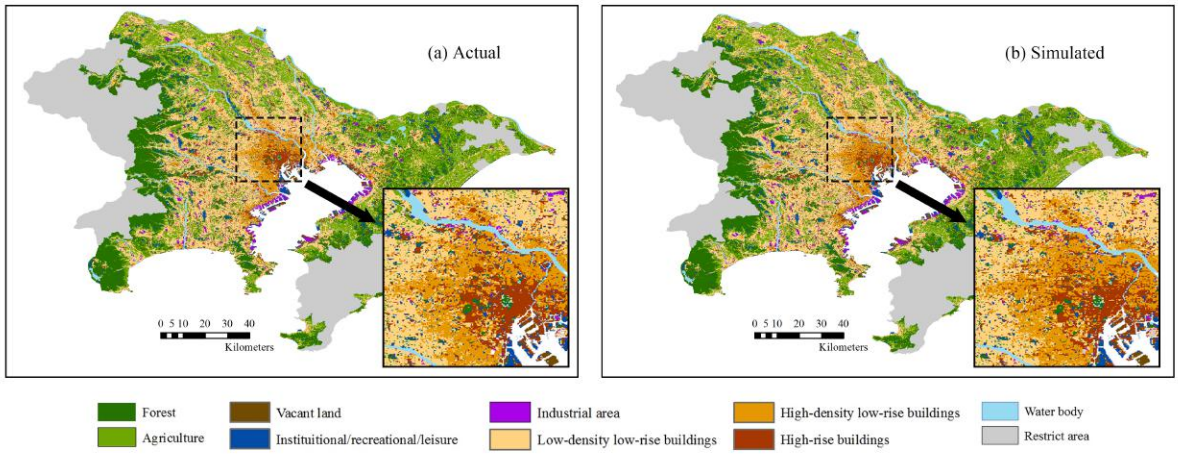


Figure 2.3: Visualization of actual map and simulated map in 2014

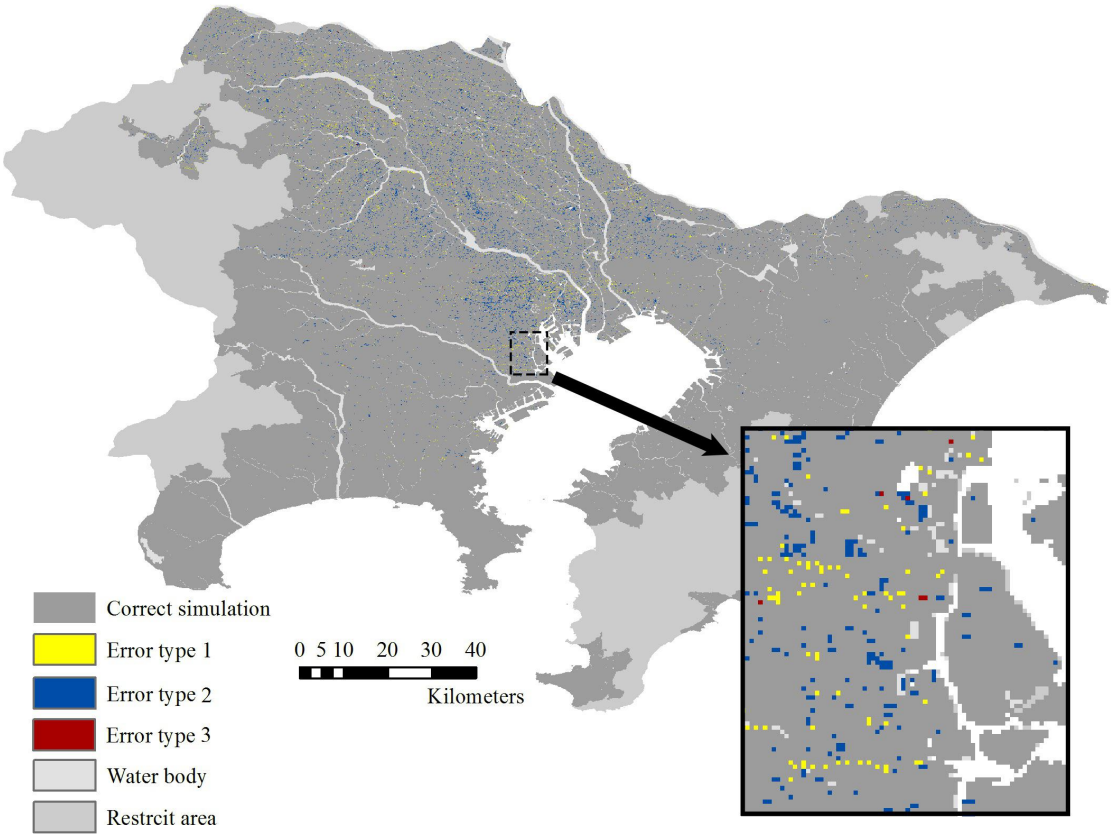
2.5 Discussion

2.5.1 Decomposition of ERT model

Our results indicate that using the ERT algorithm to enhance DT and address the spatial sampling problem has proven to be an effective approach; however, the rationale for what observed as the results needs further interpretation. ERT and other bagging ensemble models would reduce two sources of variance: the variance inherent to the DT algorithm and the variance arising from the sampling. For simplicity, the former is denoted as DT enhancement, and the latter is denoted as sampling correction.

To decompose the variance reduction process, this study used the ERT models as basic predictors and trained the bagged ERT models using the bootstrap sampling and aggregating approach. These bagged ERT models were compared with single DT and single ERT models, which were trained on a sub-sample to calculate the variance reductions. It is presumed that the variance reductions from the single DT to single ERT models are accounted for by DT enhancement, and the variance reductions from the single ERT to bagged ERT models are accounted for by sampling correction.

In Table 2.8, the first step (from single DT to single ERT) and the second step (from single ERT to bagged ERT) present different trends with respect to the three metrics. The first step significantly reduces the total error; the second step also reduces the total error but much less. The total error decrease is supposed to equal the reduction in the variance, as theoretically bagging algorithms only address error variance. Therefore, the results indicate that both DT enhancement and sampling correction have reduced the variance of the transition probability prediction. Overall, the contribution of the DT algorithm enhancement is much higher. Combining the results of AUC-ROC and AUC-PR, it can be seen that both metrics significantly increase through two steps; the proportional increase extents of both metrics by DT enhancement are similar, whereas with respect to sampling correction, the proportional increase extent of AUC-PR is larger than that of AUC-ROC. This result may imply different effects of DT enhancement and sampling correction in the variance reduction based on the different characteristics of the ROC and PR curves. DT enhancement improves the performance of both positive (changed cells) and negative classes (unchanged cells) with no preference, whereas sampling correction mainly focuses on improving the positive class (changed cells).



Error type 1: land use does not actually change but changes in the results of simulation
Error type 2: land use actually changes but does not change in the results of simulation
Error type 3: land use actually changes and also changes in the results simulation, but the predicted change is inaccurate

Notes:
Error type 1: land use does not actually change but changes in the results of simulation
Error type 2: land use actually changes but does not change in the results of simulation
Error type 3: land use actually changes and also changes in the results simulation, but the predicted change is inaccurate

Figure 2.4: Visualization of correction and error

Table 2.8: Improvement evaluation from using bagging-based algorithms

	Total error			AUC-ROC			AUC-PR		
	Single DT	Single ERT	Bagged ERT	Single DT	Single ERT	Bagged ERT	Single DT	Single ERT	Bagged ERT
Forest to agri.	0.056	0.028	0.027	0.555	0.851	0.909	0.144	0.202	0.257
Forest to low-density and low-rise bldgs.	0.039	0.02	0.019	0.577	0.893	0.944	0.175	0.322	0.332
Agri. to forest	0.037	0.019	0.018	0.564	0.803	0.892	0.124	0.244	0.245
Agri. to low-density and low-rise bldgs.	0.093	0.049	0.046	0.592	0.832	0.842	0.191	0.278	0.334
Vacant to IRL	0.043	0.021	0.018	0.718	0.902	0.977	0.478	0.672	0.811
Vacant to forest	0.029	0.015	0.014	0.662	0.879	0.952	0.365	0.525	0.686
Vacant to low-density and low-rise bldgs.	0.098	0.051	0.046	0.6	0.858	0.941	0.225	0.337	0.505
Vacant to industrial	0.017	0.009	0.008	0.664	0.877	0.941	0.415	0.547	0.706
Vacant to high-rise bldgs.	0.025	0.013	0.012	0.666	0.915	0.982	0.331	0.43	0.681
IRL to forest	0.027	0.015	0.013	0.587	0.823	0.918	0.244	0.349	0.697
IRL to high-rise bldgs.	0.016	0.008	0.008	0.614	0.826	0.929	0.231	0.26	0.728
Industrial to Low-density and low-rise bldgs.	0.025	0.013	0.012	0.843	0.985	0.994	0.716	0.88	0.921
Industrial to vacant	0.092	0.054	0.049	0.617	0.852	0.934	0.277	0.376	0.586
Industrial to IRL	0.049	0.024	0.022	0.642	0.812	0.985	0.319	0.358	0.591
Low-density and low-rise bldgs. to high-rise bldgs.	0.04	0.02	0.018	0.663	0.894	0.961	0.362	0.518	0.677
High-density and low-rise bldgs. to high-rise bldgs.	0.053	0.027	0.025	0.699	0.913	0.958	0.419	0.58	0.702
High-rise bldgs. to low-density and low-rise bldgs.	0.066	0.035	0.033	0.629	0.838	0.932	0.331	0.357	0.617
High-rise bldgs. to high-density and low-rise bldgs.	0.04	0.022	0.02	0.712	0.955	0.976	0.461	0.593	0.831

2.5.2 Driving factors of LUC

Tree-based learning algorithms provide not only high-quality transition probability prediction but also insights into the complex urban dynamics. Using the variable importance evaluation, this study examines the explanatory powers of various driving factors. This study considers five driving factor categories: neighborhood effect, accessibility, physical factors, socio-economic factors and policy. To address the different influences of the end land use of a specific transition and other land use types, this study further divides the neighborhood effect into main and auxiliary neighborhood effects, which account for the neighborhood enrichment of end land use type and other land use types, respectively. The average importance score of variables that belong to the same category of driving factor is calculated to represent the average importance of the driving factor. Figure 2.5 presents the results.

The main neighborhood effect is the strongest explanatory factor, but its explanatory factor varies significantly with transition type. In contrast, the impacts of the auxiliary neighborhood characteristics are relatively small but consistent over transition type. Accessibility and physical factors, which are commonly used in LUC modeling studies, have modest explanatory powers. Socio-economic and policy factors are relatively less important but may play critical roles in certain land use transition processes such as land use transitions from certain sources to high-rise buildings (see T7, T11, T15 and T16 in Figure 2.5). This finding is plausible given that high-rise buildings are mainly high-grade residential or commercial buildings, and LUC occurs when the new land use type is able to provide greater value, especially economic value in a free market (Zheng et al., 2015). On the other hand, transitions related to industrial areas are heavily constrained by policy factors, particularly the land use zoning policy (see T8 and T12 in Figure 2.5). This result may be because industrial areas are subject to relatively strict government regulations due to their environmental impact and social concerns. In contrast, transitions between forests, agriculture and low-density and low-rise buildings are not explained by

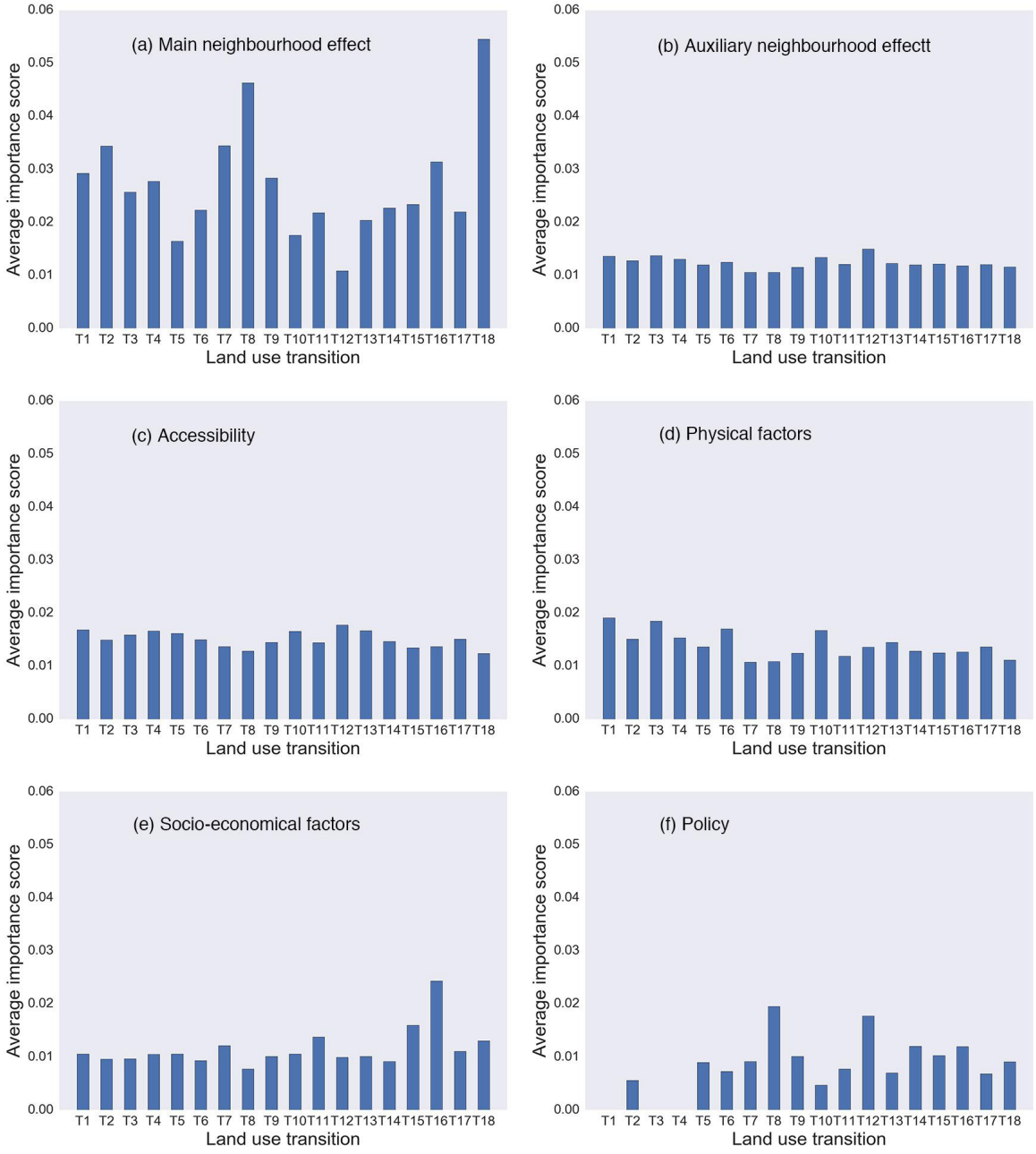


Figure 2.5: Importance evaluation of driving factors

policy factors, except for the transition from forest to low-density and low-rise buildings; this indicates relatively soft regulations of these behaviors.

This study further explores the main and auxiliary neighborhood effects by examining the importance scores of land use enrichment variables with varying neighborhood size (see Figure 2.6). In general, the importance scores of the main neighborhood effect decrease with the neighborhood size, whereas the importance scores of the auxiliary neighborhood effect increase with the neighborhood size. This result may reflect the distinct rationale behind the main and auxiliary neighborhood effects; the former follows the expanding rule whereby a nearby neighborhood has a greater impact, whereas the latter represents the influence of more general characteristics of the local area.

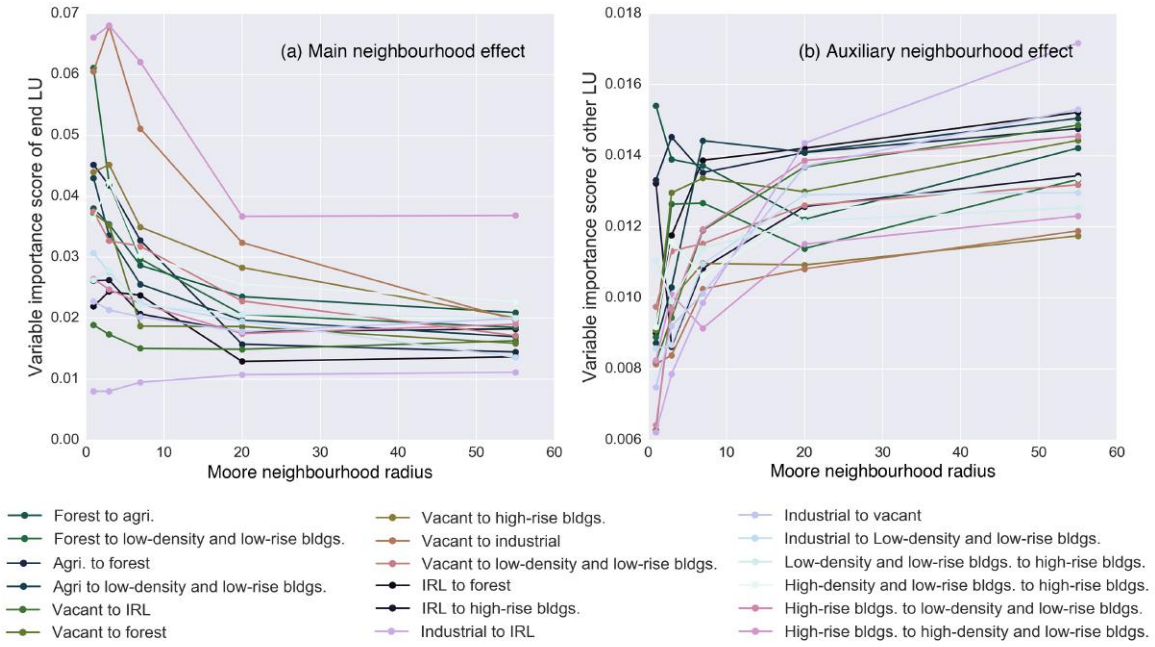


Figure 2.6: Importance of main and auxiliary neighborhood effects with varying neighborhood sizes

2.5.3 Varying modeling performances

Table 2.7 and 2.8 show a significant modeling performance variation among different types of land use transition. Such variation may result from variations in the degree of spatial autocorrelation and/or variations in the land use transition characteristics. For instance, in our results, the transition to IRL mainly occurs around the harbor. Hence this transition has the highest spatial autocorrelation compared to the other two transitions beginning from industrial areas that occur in more dispersed areas, and this transition exhibits a higher modeling performance.

This study provides two examples in which specific characteristics of a transition lead to significantly high or low performances. Usually, LUC modeling should consider the two processes: discarding the functionality of the current land use and accepting the functionality of a new usage. However, transitions from vacant land to other land uses only consider the latter process, because although vacant land has certain economical value depending on factors such as location and ownership, it has almost no industrial or ecological contribution.

This particular characteristic may lead to a relatively simple and straightforward transition rule that can be more easily addressed by statistical models and produce the exceptionally high modeling performances. On the contrary, the modeling of transitions between forest, agriculture and low-density and low-rise buildings achieves relatively poor performances compared to the other transitions because these land use types are mainly privately owned; low-density and low-rise buildings mainly consist of residential buildings, and more than half of the forest areas in Japan are privately owned. Therefore, individual decision making seems to play an important role in determining the transition rules between the three land use types, and its absence in the LUC modeling could undermine the predictive performances.

2.6 Summary

This study used an integrated approach based on a tree-based model and a CA model to simulate multiple LUC processes in the Greater Tokyo Area, focusing on the application of tree-based methods and the interpretation of the urban dynamics. This study used 4 tree-based models (BT, RF, ERT and bagged GBDT) to predict the transition probability of LUC processes and compared the predictive performances of these methods with MLP and among themselves. All the tree-based models outperformed NN, and among them, ERT showed the best predictive performance in this multiple LUC modeling task. The results indicate that bagging provides a greater performance improvement for DT than boosting, and the introduction of randomization into the bagging algorithm improves the predictive performance. The results of this study indicate that the integration of tree-based methods and CA is an effective approach for multiple LUC modeling.

In addition to the outstanding predictive ability, tree-based models also provide insights into the driving factors of LUC processes and help us to understand the complex urban dynamics. In terms of this study, the land use neighborhood characteristics have the strongest impact on LUC in the target area, but their explanatory powers in various LUC processes vary. Moreover, the impact depends on the neighborhood land use type and the neighborhood size; the neighborhood effect of end land use generally decreases with the neighborhood size, whereas the neighborhood effects of other land uses generally increase with the neighborhood size. The significant influence of a large neighborhood was reported in previous studies (Hagoort et al., 2008; van Vliet et al., 2013b), and the findings of this study further elucidated the additional details of the relationship between the strength of the neighborhood effect, the neighborhood size and the land use type. This study also found that socio-economic factors, particularly land prices, have a strong influence on the transitions from various land use types to high-rise buildings, indicating their important roles in urban redevelopment. Furthermore, land use zoning policies act as strong constraints on the LUCs associated with industrial areas but have little

influence on the changes between forests, agricultural lands and low-density and low-rise buildings.

The approach developed in this study, which combined the predicted transition probability and driving factor analysis, can provide comprehensive evidence for generating detailed and effective zoning legislation for various urban land use types in metropolitan. In addition, the approach can be used to generate the multiple land use maps, which are useful inputs for various environmental modeling tasks, such as air pollution modeling and carbon footprint analysis, where the differentiation of natural land use and various built-up types is essential.

Chapter 3

Modeling with convolutional neural networks in Saitama prefecture

3.1 Motivation

Cellular automata (CA) simulates the complex transitional rules by stacking simple neighborhood rules (White and Engelen, 1993). Given its simple but effective mechanism, CA has become the most prevalent approach in LUC modeling over the last decade (Aburas et al., 2016). CA's effectiveness also indicates the important role of neighborhood rules in LUC modeling; CA variants can enhance CA's performance by modifying, transforming or extending the mechanisms of neighborhood rule construction (Santé et al., 2010; Chaudhuri and Clarke, 2013).

Patch-based CA adopts a patch-based simulation strategy rather than a cell-based strategy (Li et al., 2013; Chen et al., 2014, 2016; Li et al., 2017). It simulates the behavior of LU patches (i.e., homogeneous cells that are spatially connected) to generate overall LU patterns, and this simulation process can be referred to as a mechanism that binds and regularizes the transitional rules of cells that are located in the same neighborhood.

Other CA variants combine the CA with statistical learning methods, in which neighborhood characteristics are usually incorporated to estimate the LU transition probability (e.g. Li and Yeh, 2002; Yang et al., 2008; Al-sharif and Pradhan, 2015; Du et al., 2018). In the integrated modeling system, previous studies show that the accuracy of the intermediate transition probability map greatly influences the final simulated performance (Camacho Olmedo et al., 2013). To capture precise neighborhood characteristics, Verburg et al. (2004) designed LU enrichment metrics to measure the relative abundance of LU categories in the neighborhood. Liao et al. (2016) extended the LU enrichment by assigning various distance-based influence weights. Other studies apply landscape metrics, which were originally used to analyze ecological issues, to the LUC modeling. Several typical categories of landscape metrics are used in LUC modeling studies: area metrics (e.g., largest patch index (Herold et al., 2003)), shape metrics (e.g., perimeter-area ratio (Chen et al., 2016)), aggregation metrics (e.g., landscape shape index (Verstegen et al., 2014), contagion (Herold et al., 2003), percentage of like adjacencies (Roy Chowdhury and Maithani, 2014)), and isolation metrics (e.g., landscape similarity index (Li et al., 2015a), Euclidean nearest neighbor distance (Chen et al., 2016)). However, these approaches have two major limitations. First, they are limited in terms of their ability to capture complex spatial features (e.g., spatial pattern). Most metrics are designed to capture simple features such as quantity, ratio, area or edge. Moreover, the composite metrics are mainly designed to capture specific aspects of neighborhood characteristics. For instance, contagion specifically represents the aggregation/interspersion degree of neighborhood patches. Finally, these approaches derive spatial features from classified LU maps, which are relatively more homogeneous and have less spatial variance compared with the original satellite images.

Convolutional neural networks (CNN), a classic deep-learning method, may be the solution for overcoming the abovementioned limitations. CNN is well-known for its ability to process image data and extract hierarchical features (LeCun et al., 2015). It learns low-level spatial structures (e.g., edges) from its first convolutional layer and gradually

stacks and extracts complex hierarchical spatial features as 'the model goes deeper'. CNN is used to solve various image processing tasks, including image classification, object detection/tracking, semantic segmentation, etc., and has been applied in various fields, including computer vision (e.g. Krizhevsky et al., 2012; Cox and Dean, 2014), remote sensing (e.g. Maggiori et al., 2016; Wang et al., 2016; Long et al., 2017), medical image analysis (e.g. Li et al., 2014; Qayyum et al., 2017), etc. In particular, CNN has recently gained popularity in remote-sensing studies Nogueira et al. (2017), which is closely related to LUC modeling studies. Makantasis et al. (2015) classified hyperspectral images using a CNN with only two layers and achieved state-of-the-art performance.

Moreover, deep learning essentially refers to multi-layered interconnected neural networks; its basic form has been used in LUC modeling since the early 2000s (Li and Yeh, 2001, 2002). Previous researchers have applied neural networks in various ways: standalone application (e.g. Liu and Seto, 2008; Wang and Mountrakis, 2011), integration with CA and/or other statistical methods (e.g. Guan et al., 2005; Grekousis et al., 2013; Li et al., 2015a), etc. These studies found that neural networks can result in reliable LU predictions. Nevertheless, other than the multi-layer perceptron (MLP), powerful neural network variants with advanced architectures are rarely used in LUC modeling studies.

This study develops an integrated modeling framework that consists of a hybrid CNN model and a DINAMICA-based CA model to simulate the LUC process of the Saitama prefecture, which is located at the western side of Japan's Greater Tokyo Area. The hybrid CNN model estimates the LU transition probability based both on spatial features learned from satellite images and on manually designed geographical features. The DINAMICA-based CA model simulates the LU pattern by referring to the generated transition probability map. This study identifies the improvement in predictive performance from incorporating CNN by comparing the accuracies of the transition probability maps, which are estimated using the hybrid CNN model and an MLP model that accepts only geographical features. The area under receiver operating characteristic curve (AUC-ROC) and the area under precision-recall curve (AUC-PR) are employed to eval-

uate the estimation accuracy. In addition, this study develops a convolutional denoising autoencoder (CDAE) model, which learns latent spatial features from satellite images in an unsupervised approach, as an alternative to the supervised CNN model. This study contributes to the existing literature by 1) identifying the benefit of utilizing satellite images data through convolutional-based deep learning techniques for LUC modeling and 2) elucidating the strengths of the supervised and unsupervised approaches.

3.2 Methodology

3.2.1 Neural network models

3.2.1.1 Geo-net

As the reference model, this study develops a neural network model, which includes a set of conventional geographical feature and excludes features linked to satellite images. This model is compared with the hybrid CNN model, which includes both geographical features and features linked to satellite images. Specifically, the reference model is an MLP with ReLU as the non-linear activation function and Sigmoid as the classifier. This study constructs and uses the commonly adopted geographical features, which are land-use enrichment (Verburg et al., 2004), proximity factors, land price, population density and physical factors, covering the neighborhood characteristics, accessibility, socio-economical and physical factors for the cell of interest. Table 3.1 describes the geographical features.

3.2.1.2 Conv-net

CNN is a special class of neural networks that uses convolutional operations in place of matrix multiplication in the hidden layers. A typical hidden CNN layer consist of three parts: 1) a convolutional layer that performs several parallel convolutions, 2) non-linear

Table 3.1: Geographical features used in LUC models

Category	Description
LU enrichment	The relative abundance of certain LU category in the neighborhood calculated by $Enrichment_{i,k,d} = \frac{n_{k,d,i}/n_{d,i}}{N_k/N}$, where $Enrichment_{i,k,d}$ is the enrichment of neighborhood d of location i with land use type k, n denotes the number of cells in the neighborhood, and N denotes the number of cells in the whole map
LU enrichment of forest	
LU enrichment of agri.	
LU enrichment of built-up	
LU enrichment of water body	
Accessibility	
Distance to major roads	The nearest Euclidean distance from the given cell to certain geographical objects
Distance to railway stations	
Distance to urban center	
Socio-economical factors	
Land price	The estimated land price of given cell; the raw data is provided by the Ministry of Infrastructure, Land and Tourism of Japan
Population density	The population density of given cell, provided by the Ministry of Infrastructure, Land and Tourism of Japan
Physical factors	
Elevation	The elevation of given cell, provided by SRTM (Shuttle Radar Topographic Mission) database
Coordinates	The coordinates of given cell

Notes:
The land price map is interpolated from a polypoint land price map, in which each record is collected by field survey, by using ordinary Kriging interpolation method.

activation layer, 3) pooling layer that replaces the output of the net at a certain location with a summary statistical of the nearby outputs. Compared with the conventional fully connected networks, CNN can be regarded as a locally connected network, which allows each hidden unit to connect to a small subset of the input units. Specifically, when processing images, each hidden unit will connect only to a small contiguous region of pixels in the input. This architecture grants CNN higher computational efficiency and the ability to capture the local pattern. The pooling operation further grants CNN an invariance in small local translation, which is a particularly useful property for tasks when identifying whether the existence of some features is more important than their location (e.g., object detection).

Various meta-architectures of CNN have been developed in these years. Although their performances have been evaluated using ImageNet classification or similar classification tasks, their suitabilities to the LUC modeling problem still require examination, due to LUC modeling's distinct characteristics. Specifically, 1) ordinary satellite images, such as Landsat images, have relatively low resolution, vague edges and barely distinguishable objects; 2) local satellite image patches may contain redundant information because of the high spatial autocorrelation; and 3) the LUC modeling's desired features may differ.

This study builds the CNN architecture using as reference the designs of three classic meta-architectures: Alex-net (Krizhevsky et al., 2012), which has a relatively large convolutional filter size; VGG (Simonyan and Zisserman, 2014), which has a small kernel size but a deep architecture; and ResNet (He et al., 2016), which constructs residual blocks to facilitate a better gradient flow. The architectures are finally determined according to the results of trial-and-error experiments. Table 3.2 presents the model architectures for modeling three LU transitions. These architectures include several noteworthy aspects: 1) this study uses a stride of 3 rather than 2 while downsampling, given the input satellite image that represents the neighborhood usually has an odd-numbered size number to guarantee that the cell of interest is centrally located; 2) although the three models' architectures are determined independently, their convolutional architectures turn out to be identical, which may indicate that CNN serves the same function even when the target transitional rules are different; 3) the pooling layers are used only next to the first two convolutional layers because the posing information, such as the direction to the center, may become important at a higher hierarchical level; and 4) the number of filters in our study is large for a binary classification task, possibly because of the high variance of spatial patterns.

According to the neighborhood effect, the influence of a neighboring cell decreases with its distance to the central cell. To introduce this mechanism into conv-net, this study specifically designs a regularization layer named the spatial-weight layer. It is

Table 3.2: Architectures of the conv-nets

Forest to agri.	Agri. to forest	Agri. to built-up
conv-128	conv-128	conv-256
spatial weight		
max pooling		
conv-256	conv-256	conv-512
spatial weight		
max pooling		
conv-512	conv-512	conv-1024
spatial weight		
conv-1024	conv-1024	conv-2048
spatial weight		
global average pooling		
dense-1036 dense-400 dense-80 dense-7	dense-1036 dense-400 dense-80 dense-7	dense-2048 dense-800 dense-300 dense-120 dense-60
Sigmoid		

Notes:
1. Conv-N denotes a convolutional block composing of a convolutional layer with 3×3 filter size and N of filters, a Batchnorm layer and a ReLU layer.
2. Max pooling layer has a kernel size of 3×3 and a stride of 3.
3. Dense-N denotes fully connected (dense) layer with N of hidden units.
4. Input image tensor has the shape of $(7 \times 26 \times 26)$.

described as

$$y = x * SW \tag{3.1}$$

$$SW_{i,j,d} = e^{a_d dist_{i,j}} + b_d \tag{3.2}$$

y and x are the output and input, respectively. SW is the spatial weight. i, j, d denote the location in an image tensor, $dist_{i,j}$ is the Euclidean distance from location(i, j) to the center, a_d and b_d are the trainable parameters. This spatial weight layer can be regarded as imposing a distance-based prior on the spatial feature map, which exponentially reduces the influence of spatial features that are far from the center.

The CNN model is placed parallel to an MLP model that accepts the geographical

features as input. Both networks are then connected to an MLP classifier composed of several fully connected hidden layers (see Figure 3.1). This hybrid CNN model is trained as a whole.

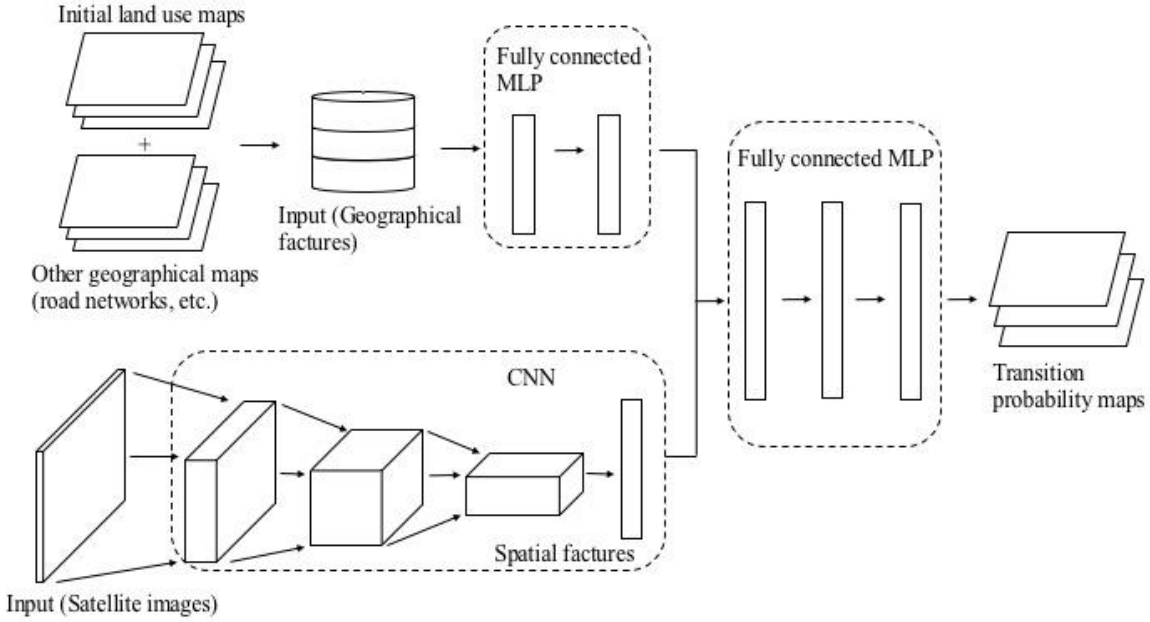


Figure 3.1: Structure of conv-net

3.2.1.3 CDAE-net

An autoencoder (AE) is an unsupervised learning algorithm that copies the input to the output. It is essentially a neural network; thus, it can be trained by backpropagation. An AE consists of two parts: the encoder and the decoder. The encoder maps the input into hidden representations, and the decoder reconstructs the input from the hidden representations. The general form of an AE is

$$h = s(Wx + b) \tag{3.3}$$

$$y = s(W'h + b') \tag{3.4}$$

x is the input; h is the latent representation; y is the output; s is non-linear activation function such as Sigmoid; W and W' are the encoder and decoder weights, respectively; and b and b' are the encoder and decoder biases, respectively. AE usually has an under-complete architecture, in which the dimension of h is smaller than x, to extract useful features from input rather than just learn an identity function.

A denoising autoencoder (DAE) (Vincent et al., 2008) is a variant of AE that is designed to capture more robust features by reconstructing the input from a corrupted version of it. The general form of a DAE is

$$h = s(W\tilde{x} + b) \tag{3.5}$$

$$y = s(W'h + b') \tag{3.6}$$

$\tilde{x}$ is a copy of x that has been corrupted by some form of noise. The noise injection forces the DAE to capture the statistical dependencies between the inputs by causing the DAE to undo the effect of the corruption process.

In this study, this study incorporates the convolutional operation into a DAE and develop a convolutional denoising autoencoder (CDAE) model to tackle the possible data problems that cannot be effectively addressed by CNN, namely, the redundant spatial information and the satellite image noise. The CDAE model is loosely combined with an MLP classifier to estimate the transition probability. Specifically, original input from satellite images is fed into the CDAE model to produce the latent representation, and the latent representation is then fed into an MLP classifier together with geographical features.

Figure 3.2 illustrates the architecture of this CDAE-net, and Table 3.3 presents the architectures of its encoder and classifier. For each convolutional or pooling layer in the encoder, the decoder has a deconvolutional layer (transposed convolutional layer) or an unpooling layer with the same configuration at the corresponding location, and the en-

coder and decoder weights are untied. Although DAE does not need to be undercomplete, this study uses a 'bottleneck' hidden layer with 567 hidden units, to extract the most salient features.

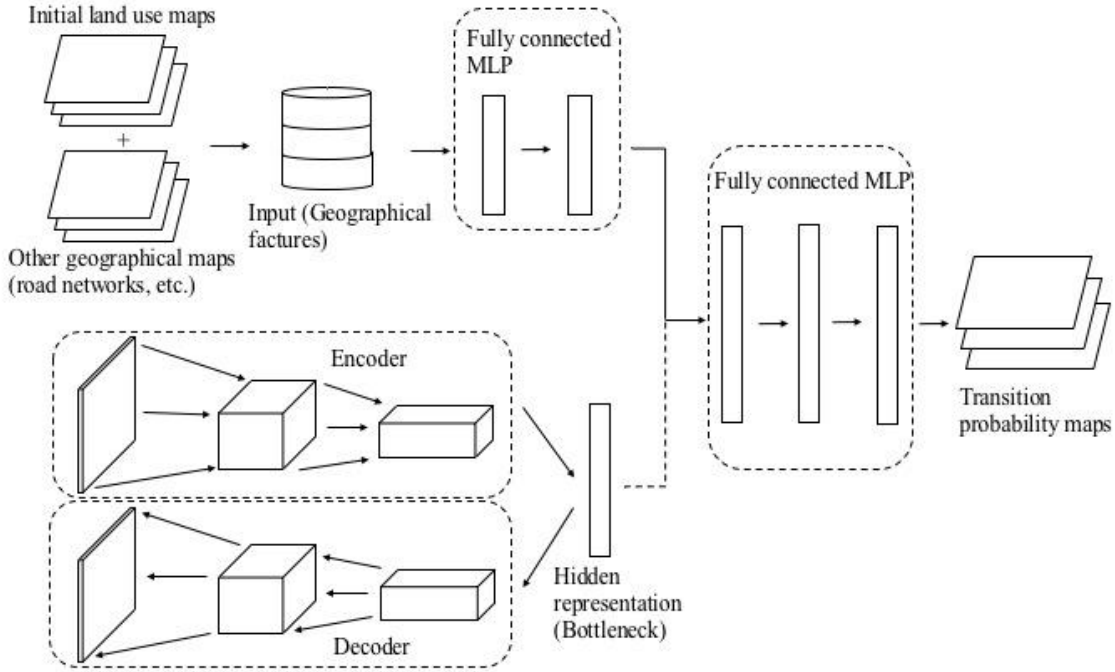


Figure 3.2: Structure of cdae-net

CDAE-net and conv-net differ in two major ways in terms of the architecture: 1) CDAE-net and classifier are separately trained independently; 2) CDAE-net have no global pooling layer at the end of its architecture, and its output 3D tensor is raveled and directly fed into the fully connected classifier.

3.2.1.4 Model training

As the main regularization method, this study adds Gaussian noise to the gradient (Neelakantan et al., 2015a), which is demonstrated effective for training deep networks, and discard dropout because our experiments show that Batchnorm eliminates its need. This study uses the image jitter method proposed by Krizhevsky et al. (2012) to produce train-

Table 3.3: Architectures of the CDAE-nets

Forest to agri.	Agri. to forest	Agri. to built-up
conv-64		
max pooling		
conv-128		
max pooling		
dense-579 dense-579 dense-579	dense-579 dense-579 dense-579	dense-579 dense-579 dense-579 dense-100
Sigmoid		

Notes:
1. Conv-N denotes a convolutional block composing of a convolutional layer with 3×3 filter size and N of filters, a Batchnorm layer and a ReLU layer.
2. Max pooling layer has a kernel size of 3×3 and a stride of 3.
3. Dense-N denotes fully connected (dense) layer with N of hidden units.
4. Input image tensor has the shape of $(7 \times 81 \times 81)$.
5. For each convolutional layer in the encoder, there is a corresponding deconvolutional layer in the decoder; for each each max pooling layer, there is a corresponding unpooling layer in the decoder.

ing samples with varying degrees of illumination to improve the model's generalization performance; however, this study does not use image flip to preserve the pose information. This study uses binary cross-entropy as the loss function for training classifier and mean square error as the loss function for training CDAE. This study uses stochastic gradient descend as the optimization method, and use the parameter initialization method suggested by Glorot and Bengio (2010), which shortens the convergence time by approximately 0.8 compared with the initialization method based on random sampling from Gaussian distribution. For each neural network model, this study performs fine-tuning on a set of hyperparameters, including the learning rate, the learning rate decay frequency, momentum, and the Gaussian noise coefficient.

3.2.2 Cellular automata

The CA model is a variant of DINAMICA, which is the same with the CA model used in section 2.2.2. The two processes are merged using the following calculation:

$$Q_{ij} = r \times expander + s \times patcher \tag{3.7}$$

where Q_{ij} is the total number of transitions from land-use class i to j; r and s are the percentages performed by the expander and patcher functions, respectively; and $r + s = 1$.

The expander function is defined as follows:

$$if \quad n_j > 3 \quad or \quad P(ij)(xy) > t \quad then \quad P'(ij)(xy) = P(ij)(xy) \tag{3.8}$$

$$else \quad P'(ij)(xy) = p(ij)(xy) \times \sqrt{\frac{n_j}{4}} \tag{3.9}$$

$P(ij)(xy)$ denotes the transition probability from land-use class i to j, t denotes a preset threshold, and n_j denotes the number of cells of land-use class j occurring in a 3×3 window. Both processes use a stochastic selection mechanism to select seeds (the center cell in a transition patch). This mechanism selects the seeds while prioritizing high over low transition probabilities with a certain degree of randomness. The patch size is drawn from a lognormal distribution, and the patch shape or compactness is determined by a parameter called isometry.

The transition quantity is determined based on transition probability maps. Specifically, this tudy manually sets a threshold for transition probability, and the quantity of cells that have transition probability above the threshold is used as the total quantity for CA simulation.

3.2.3 Evaluation metrics

This study uses AUC-ROC and AUC-PR to assess the predictive performances of the neural network models. Their advantages and differences are demonstrated in section 2.2.3.2.

This study uses overall accuracy, quantity/allocation disagreement, Cohen's Kappa statistic, Kappa simulation and fuzzy Kappa simulation to evaluate the agreement between a simulated LU map and an actual LU map. Quantity/allocation disagreement and Kappa statistic are commonly used cell-to-cell evaluation metrics. Quantity disagreement is defined as the difference between two maps in terms of the quantity of land-use category mismatch. Allocation disagreement is the difference between two maps in terms of the mismatch of each land-use category's spatial allocation. Kappa statistic is a classic map comparison method, which excludes the proportion of agreement by chance (refer to section 2.2.3.2 for details). Nevertheless, the two metrics cannot eliminate the influence of LU persistence, which may lead to over-estimation of the predictive performance of LUC model. In order to deal with this limitation, this study employs Kappa simulation and fuzzy Kappa simulation. Both metrics exclude the influence of LU persistence by incorporating the initial land use map, but Kappa simulation is a cell-to-cell metrics, while fuzzy Kappa simulation is a vicinity-based metrics, which introduces fuzzy set theory to account for the LU category similarity and neighborhood similarity. This study only considers the neighborhood similarity in this study given the limited number of LU categories, and use a Gaussian distance decay function to specify the agreement level with respect to distance within a neighborhood. Taylor and Hagen (2003) and van Vliet et al. (2013a) provides demonstrations and technical details of fuzzy Kappa and fuzzy Kappa simulation, respectively.

3.3 Implementation

Our study area is the Saitama prefecture of Japan, which is located at the western side of the Greater Tokyo area. It covers an area of 3,798 km^2 and has a population size of 7,237 thousand. Most parts of Saitama can be regarded as Tokyo suburbs, and Saitama's urban area is constantly expanding due to immigration to the Greater Tokyo Area.

This study collects Global Land Survey (GLS) satellite image datasets for 2000, 2005 and 2010. GLS datasets have eight bands, within which band 8 (Panchromatic) has a resolution of 15 meters, and the other seven bands have resolutions of 30 meters each. All bands except band 8 are resampled into 15 meters and combined for LU classification. This study classifies four LU categories (water, agriculture, forest and built-up) using the supervised classification algorithm provided in ERDAS IMAGINE V2016 (Hexgon Geospatial, U.S.).

Supervised LU classification algorithm needs a prior knowledge of the LU distribution in the study area. More specifically, a set of pixels in a specific area is selected as training samples and defined with LU categories manually. Then, the algorithm is trained and used to recognize the LU categories for pixels in other parts of the satellite image based on this manually defined samples.

In this study, the criteria for selecting the sub-areas, which is used for providing prior knowledge, is: 1) sub-area should cover all four target LU categories (forest, agriculture land, residential land and water, see Table 3.5 for the definition of four LU categories); 2) given agriculture lands includes many different kinds of crops, which usually show colors with subtle differences on the satellite images, sub-area should have various agriculture lands with considerable variation of colors. Based on this criteria, the selection was conducted by visually interpreting the Landsat satellite images and the corresponding Google Earth historical satellite images.

Within the selected sub-areas, a set of pixels with varying colors is chosen as training

samples. The set of pixels covers most of colors with subtle difference and accounts for approximately 1/30 of all the pixels in these areas. The reference LU categories of the training samples are defined by visual interpretation, i.e., by comparing the Landsat satellite images with Google Earth historical satellite images that have finer resolution with approximately 5 meter, and visually determine the LU categories based on the geographical object shown on Google Earth historical satellite images. In the training process, the computer system continuously makes comparison between a seed pixel in the training samples and the contiguous pixels based on the similarity of spectral information of each pixel. Once the LU category of a contiguous pixel is accepted, the contiguous pixel is included as a training sample. Then the computer system will move on to recognize the neighboring pixels of the newly accepted pixel until every pixel in the area is recognized.

This study uses Mahalanobis distance method as supervised decision rule for the classification. It is one of the most popular method used for processing remote sensing data. Mahalanobis distance method calculates the spectral distance between the measurement vector (contains the measures of each spectral band) for the contiguous pixel and the mean vector for each training samples. The equation for the Mahalanobis distance for a target contiguous pixel, which needs to be classified, is

$$D = (X - M_c)^T Cov_c^{-1} (X - M_c) \tag{3.10}$$

where D denotes Mahalanobis distance, c is a particular class (LU category), X is the measurement vector of target pixel, M_c is the mean vector of the training sample of class (LU category) c, Cov denotes the covariance matrix of the pixel in the training samples with class (LU category) of c. The target pixel is assigned to the LU category with the lowest D.

Table 3.4 presents the confusion matrices of LUC, and Figure 3.3 shows LU maps of the three years. This study excludes the LU transitions that have transition rates below 1% and are left with transition from forest to agriculture, transition from agriculture to

forest, and transition from agriculture to built-up as the modeling objects.

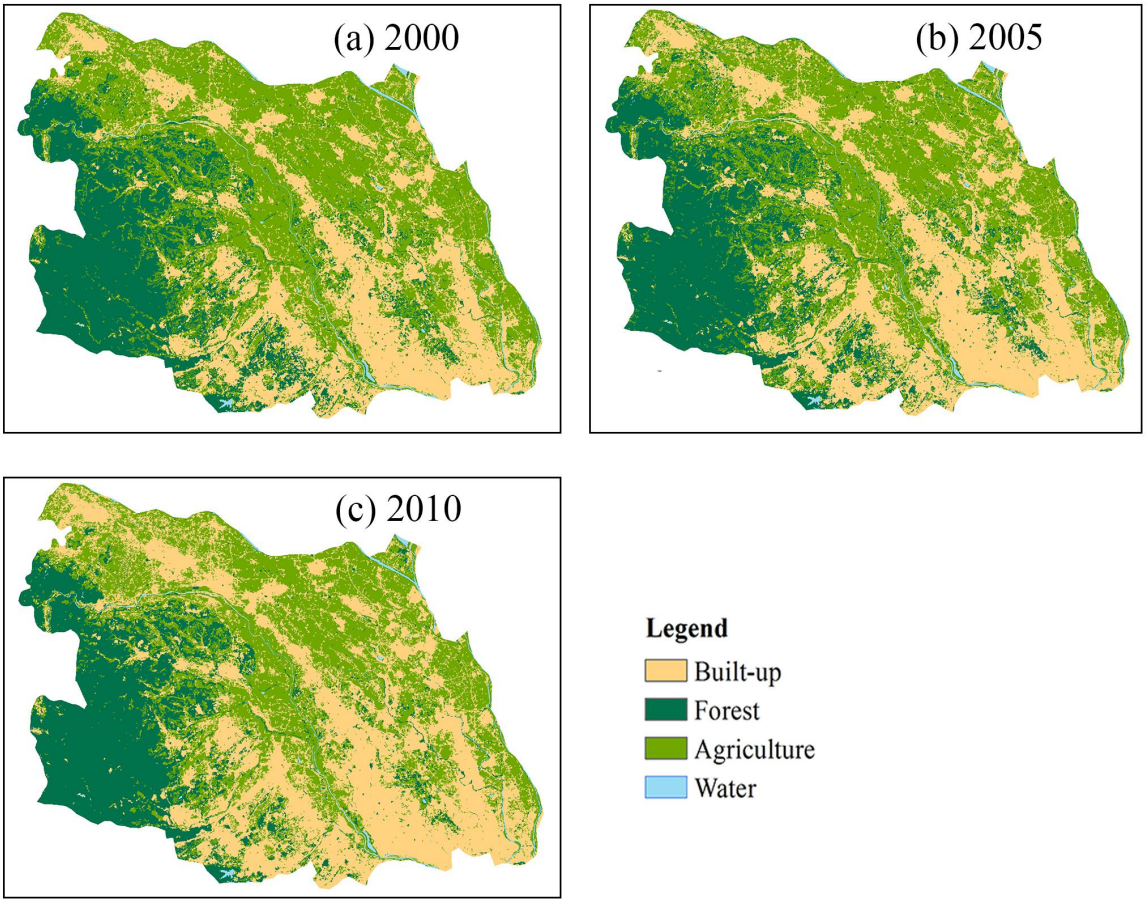


Figure 3.3: Actual LU maps in Saitama prefecture of Greater Tokyo Area for 2000, 2005 and 2010

Given the possibility of varying transitional rules being present among the three LU transition types, This study separately develops transition probability estimation models and CA models for each of them. Geographic features are derived based on spatial data collected from the Ministry of Infrastructure, Land and Tourism of Japan, except for LU enrichment, which is directly calculated from LU maps. Satellite image patches of a certain size are cropped from the satellite images and used as input for the conv-net or CDAE model. The input image size is determined based on previous evidence on the effect of neighborhood size and trial-and-error experiments. The image input size is 27×27 for conv-net and 81×81 for CDAE-net. In addition, the satellite input image

Table 3.4: Confusion matrices from 2000 to 2005 and from 2005 to 2010

		2005			
		Built-up	Forest	Agri.	Water body
2000	Built-up	4090795 (98.99%)	13672 (0.33%)	20508 (0.50%)	7522 (0.18%)
	Forest	22787 (0.72%)	2693884 (85.10%)	435253 (13.75%)	13672 (0.43%)
	Agri.	419344 (7.48%)	575959 (10.27%)	4601687 (82.06%)	10937 (0.20%)
	Water body	478 (0.39%)	888 (0.72%)	683 (0.55%)	121397 (98.34%)
		2010			
		Built-up	Forest	Agri.	Water body
2005	Built-up	4601286 (98.30%)	33151 (0.71%)	37887 (0.81%)	8530 (0.18%)
	Forest	20894 (0.66%)	2564788 (80.75%)	565762 (17.81%)	24863 (0.78%)
	Agri.	605314 (11.99%)	345704 (6.85%)	4085922 (80.93%)	11839 (0.23%)
	Water body	947 (0.77%)	591 (0.48%)	591 (0.48%)	121397 (98.28%)

Notes:
1. The confusion matrix is presented as num of cells (the percentage).
2. The percentage is calculated by $(num_t - num_{t-1})/num_{t-1}$ where num denotes the number of cells and t denotes the time.

Table 3.5: Definition of land use categories in Saitama prefecture for 2000, 2005 and 2010

LU category	Description
Agriculture	Lands used for growing crops including wet paddy filed, dry paddy field, swamp paddy field, fields used for growing wheat, upland rice, vegetables, fruits, tea, wax tree, paper mulberry, hemp palm, etc., as well as grassland and lawn
Forest	Lands where perennial plants are densely distributed
Built-up	Lands where residential buildings, commercial buildings, etc. are densely distributed
Water body	Areas including river and river bed, artificial lake, natural lake, pond, fish farm, etc. where are filled with water for most of the time

Notes:
The definitions of LU categories are based on the information provided by National Land Information Division, National and Regional Policy Bureau of Japan
(http://nlftp.mlit.go.jp/ksj/gml/codelist/LandUseCd-09.html)

has seven bands, excluding band 6 (thermal) due to its low spatial variation.

The LUC models are trained on the 2000 and 2005 datasets, validated on a subset of data for 2005 and 2010, and tested on the whole dataset for 2005 and 2010. To minimize the spatial autocorrelation between the validation set and the test set to facilitate an

unbiased model evaluation, the validation set is extracted from a sub-region of Saitama covering approximately 15% of the total area. In terms of sampling, previous studies have normally used random or stratified sampling to avoid the influence of spatial autocorrelation. However, the mini-batch learning criterion of a neural network naturally mitigates the influence of spatial autocorrelation to some extent. This study uses a bootstrap over-sampling strategy in this study. Specifically, a mini-batch of data is randomly sampled from the dataset with replacement, and samples belonging to negative and positive labels have the same proportion within a mini-batch. Over-sampling could make the model prone to over-fitting. To address this issue, this study considers the fine-tuning of the Gaussian noise coefficient. More discussion on the over-sampling is provided by Batista et al. (2004).

3.4 Results

3.4.1 Evaluation on the modeling performances

Table 3.6 summarizes the evaluation results for geo-net, conv-net, CDAE-net. AUC-ROC and AUC-PR are calculated based on the transition probability estimated by the three models. For modeling all three LU transitions, conv-net and CDAE-net consistently outperform the geo-nets for all the evaluation metrics. Particularly, conv-net and CDAE-net outperform the geo-nets by approximately 0.02$\sim$0.10 in terms of AUC and approximately 0.053$\sim$0.15 in terms of the Kappa statistic. These results confirm that the use of convolutional-based neural networks to extract spatial features from satellite images improves LUC modeling performance.

The results indicate the validity of all the three models to learn pattern from the spatial data and to determine the transition rules, given that the AUC-ROCs of the three models are much larger than 0.7. However, the results exhibit substantially different pre-

Table 3.6: Performance evaluation of the estimated transition probabilities

LU transition	Model	AUC-ROC	AUC-PR
Forest to agri.	Geo-net	0.886	0.595
	Conv-net	0.944	0.714
	CDAE-net	0.912	0.675
Agri. to forest	Geo-net	0.863	0.397
	Conv-net	0.905	0.493
	CDAE-net	0.880	0.415
Agri. to built-up	Geo-net	0.659	0.215
	Conv-net	0.694	0.239
	CDAE-net	0.714	0.268

dictive performances for the three LU transitions. The AUC-PR obtained from the LUC models for transitions from forest to agriculture, agriculture to forest, and agriculture to built-up are 0.69, 0.42 and 0.26, respectively; the highest AUC-ROC for the three transitions are 0.92, 0.89, 0.72, respectively. According to these results, the modeling performance for the transition from agriculture to built-up is significantly worse than the other two transitions.

Two possible reasons explain the lower performance of agriculture to built-up. First, in the LU classification, agriculture and built-up are relatively difficult to differentiate because they usually have interspersing spatial distributions, particularly in the suburbs and frequently appear as similar colors on the satellite images. According to the results of the LU classification assessment provided by the supervised classification algorithm, the classification accuracies of agriculture and built-up (three-years average of 85 % and 83%, respectively) are significantly lower than forest (three-years average of 89%). Therefore, the LUC modeling of the transition from agriculture to built-up may suffer more from the data noise problem than the other two transitions, which then leads to the relatively poor performance. Second, the lower transition performance may be driven by the lack of information regarding individual decision-making regarding the transitions from agriculture to built-up in study area. Given that the Saitama area has no intensive urban development plan built-up and agriculture are commonly privately owned, the

individual decision-making factor would act as an important determinant of the LUC process of agriculture to built-up. However, individual decision-making cannot be captured through spatial features, including satellite images; this information needs to be collected separately (Du et al., 2018).

The results indicate that, due to the capability of data denoising, CDAE-net outperforms conv-net when the image data is noisy. For the transition from agriculture to built-up, CDAE-net outperforms conv-net by approximately 0.03 in terms of AUC-PR. However, conv-net outperforms CDAE-net when data quality is better. For transitions between agriculture and forest, conv-net outperforms CDAE-net by approximately 0.01∼0.02 in terms of AUC-PR; this result may be induced by CDAE-net's data compression process, which inevitably discards some useful spatial information. This compression process would be preferred when data are noisy but it may lower the performance otherwise.

3.4.2 Land-use simulations

This study simulates the LU maps for 2010 by using the DINAMICA-based CA based on the transition probability maps produced by the geo-net, conv-net and CDAE-net, respectively. In addition, this study simulates LU map based on the transition probability of agriculture to forest and forest to agriculture predicted by the conv-net and the transition probability of agriculture to built-up predicted by the CDAE-net. Figure 3.4 shows the comparison between simulated and actual LU maps for 2010, and Table 3.7 presents the evaluation metrics of the simulated maps, including accuracy, quantity/allocation disagreement, Kappa statistic, Kappa simulation and fuzzy Kappa simulation with neighborhood sizes of 3×3, 7×7 and 11×11.

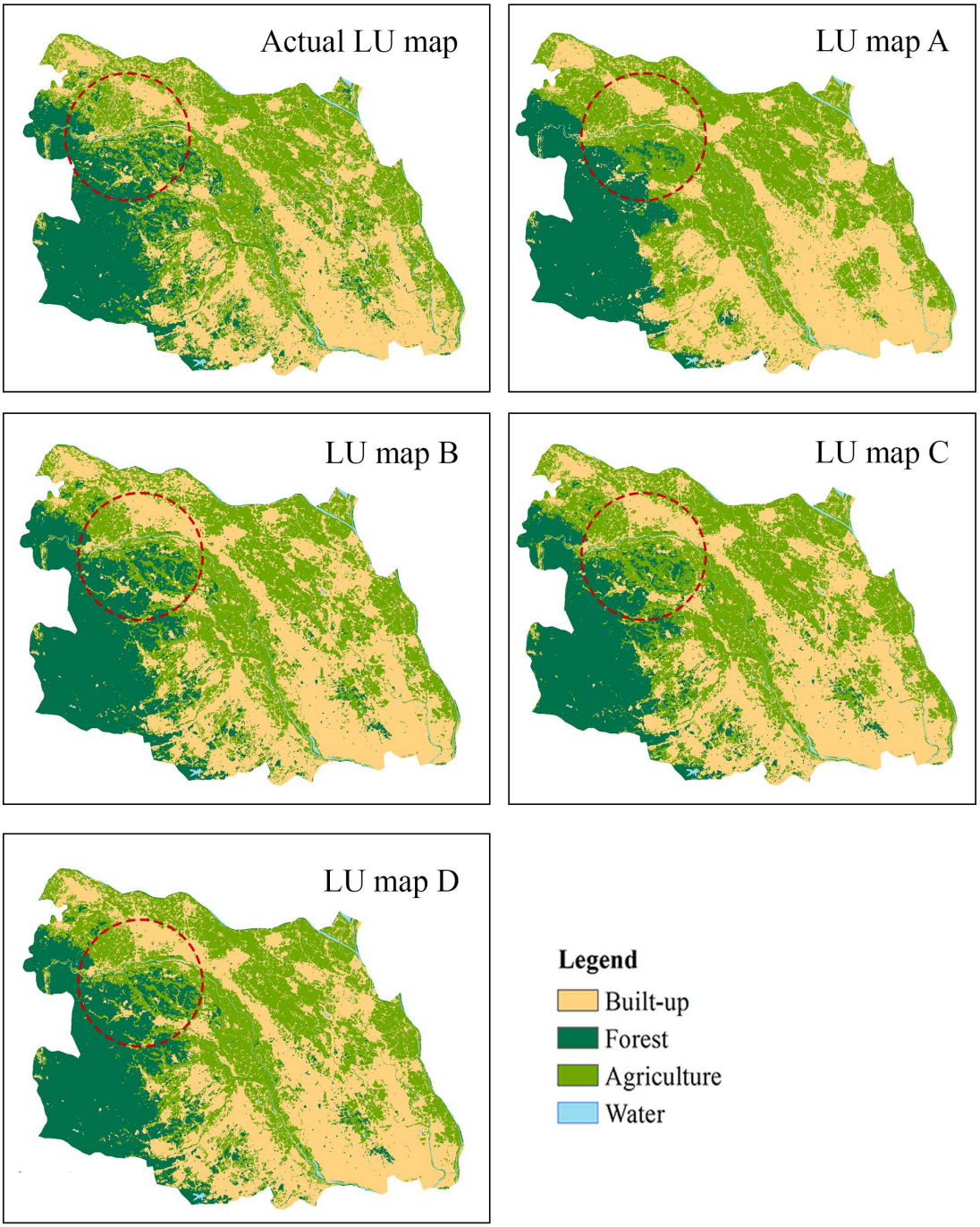


Figure 3.4: Actual and simulated LU maps for 2010

Notes:

1. Land use maps of A, B and C are simulated by using CA based on the transition probability predicted by the geo-net, conv-net and CDAE-net, respectively.
2. Land use map of D are simulated based on the transition probability of agriculture to forest and forest to agriculture predicted by the conv-net and the transition probability of agriculture to built-up predicted by the CDAE-net.

The order of the performances of the three LUC models is conv-net + CA > CDAE-net + CA > geo-net + CA; this result is consistent with the difference of transition probability prediction performances. The LU map D in Table 3.7 combines the transition probability predictions with the highest accuracy, which also exhibits the highest simulation performance. The values of Kappa simulation and fuzzy Kappa simulation are significantly lower than the values of Kappa statistic; this result is plausible given the exclusion of the influence of LU persistence. The significant difference between the values of Kappa statistic and Kappa simulation indicates that the Kappa simulation is stricter metrics than the Kappa statistic. In terms of the fuzzy Kappa simulation, the value of fuzzy Kappa simulation obtained from the geo-net + CA increases much less with the increase of neighborhood size when compared with the increase of values obtained from the conv-net + CA or the CDAE-net + CA. Compared with Kappa simulation, fuzzy Kappa simulation provides additional evaluation aspect of LUC model's capability to yield the 'near hits' (i.e. the LUC model does not precisely allocate the LU transition to the target cell but allocates the LU transition to the cells within the neighborhood of the target cell). This result indicates the advantage of convolutional-based models over conventional LUC models, which use only geographical features, when modeling the spatial pattern of LUC process.

Table 3.7: Performance evaluation of the simulated LU maps

LU map	Accuracy	Disagreement		Kappa statistic	Kappa simulation	Fuzzy Kappa simulation		
		Quantity	Allocation			3×3	7×7	11×11
A	0.897	0.016	0.086	0.812	0.337	0.338	0.340	0.345
B	0.908	0.011	0.081	0.830	0.391	0.414	0.455	0.484
C	0.901	0.013	0.078	0.822	0.376	0.399	0.437	0.455
D	0.914	0.005	0.080	0.842	0.412	0.435	0.474	0.501

Notes:
1. Land use maps of A, B and C are simulated by using CA based on the transition probability predicted by the geo-net, conv-net and CDAE-net, respectively.
2. Land use map of D are simulated based on the transition probability of agriculture to forest and forest to agriculture predicted by the conv-net and the transition probability of agriculture to built-up predicted by the CDAE-net.

3.5 Discussion

3.5.1 Model visualization

Although the explicit mechanism inside deep neural networks cannot be elucidated, some visualization techniques shed light on how conv-net and CDAE-net process the satellite images. Figure 3.5 visualizes the activation maps from the first convolutional layers of conv-net and CDAE-net. The activation maps from conv-net and CDAE-net exhibit substantially different spatial patterns. This difference may be explained by the different purposes that the two models serve: conv-net seeks to capture the pattern that helps to explain the objective function, while CDAE-net seeks to capture the latent spatial features that help to explain the spatial variations in input images. Therefore, in conv-net, patterns with high activation values exhibit irregular shapes, but in CDAE-net, patterns with high activation values resemble the skeleton of geographical objects.

This study uses t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) to visualize the distribution of spatial features extracted by the convolutional models. This study randomly selects a total of 2000 samples from the test set and feed them into conv-net and CDAE-net to obtain spatial features from the final convolutional layers. The spatial features are then embedded into 2D vectors by t-SNE. Figure 3.6 shows the results of t-SNE. Theoretically, the distribution of spatial images in a 2D space may reflect the CNN effect, given that CNN would gradually transform the satellite image into linearly separable representations. As shown in Figure 3.6, the degree of sample aggregation with the same label is consistent with the accuracy of transition probability estimation; more samples with the same label aggregate together denote a higher predictive performance of the corresponding transition. For example, the transition from forest to agriculture has the highest predictive performance and the most visually separate spatial feature distribution. On the other hand, the transition from agriculture to built-up has the lowest predictive performance and barely separate spatial

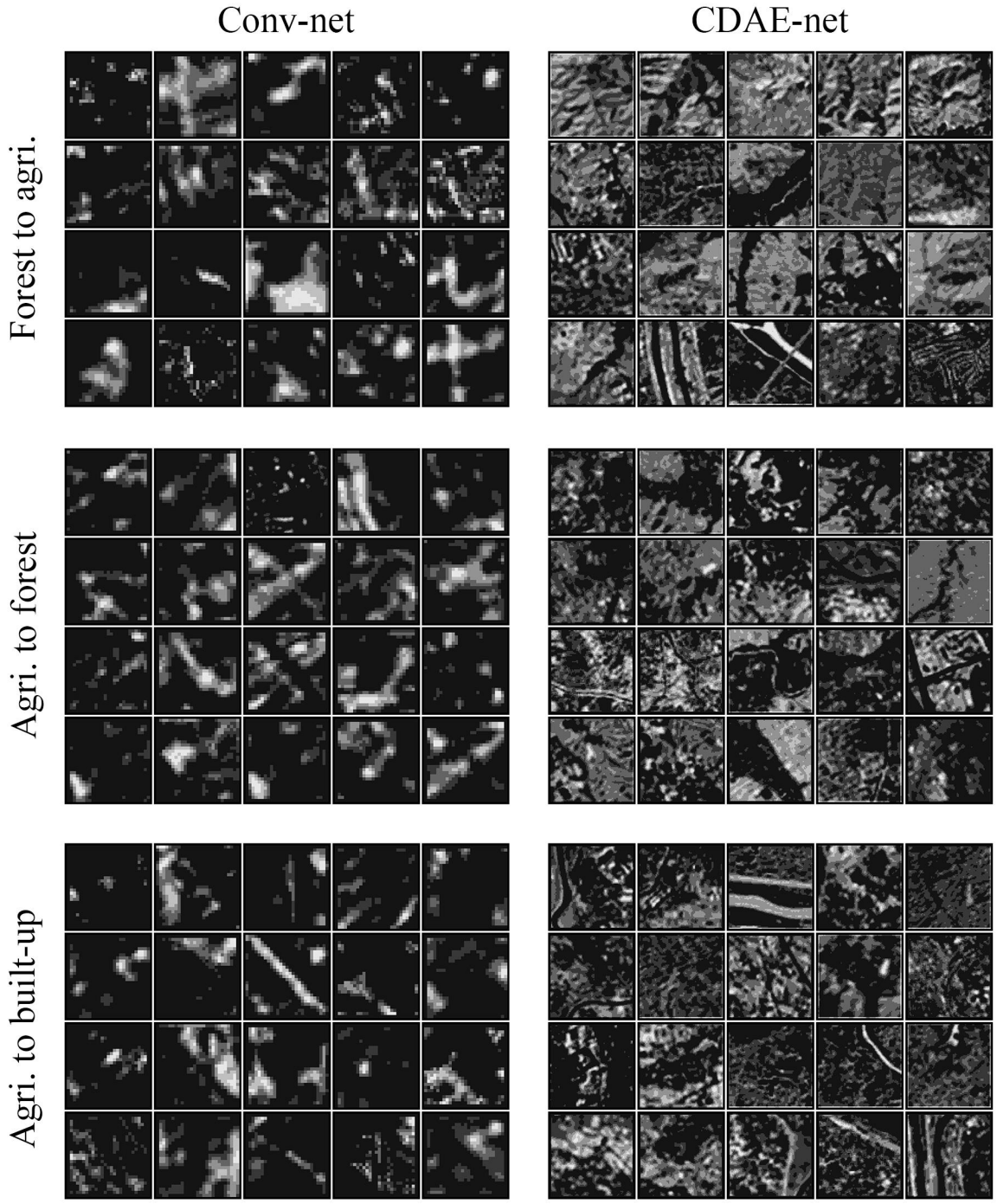


Figure 3.5: Visualization of outputs from the first convolutional layers of the conv-nets and the CDAE-nets

feature distribution. Given that CDAE-net uses an unsupervised learning method, their embedded feature distributions are obviously different from those of conv-net. In terms of the transition from forest to agriculture, although the samples cluster into groups in both distributions, the clusters from CDAE-net appear to be more scattered than those from conv-net. To process the spatial features generated from the CDAE-net model, the classifier may require a higher level of non-linearity, which may partially explain the phenomenon in which the CDAE-net classifier generally has more hidden layers than conv-net's.

3.5.2 Model architecture

This study analyzes the convolutional filter, the spatial weight layer and pooling within conv-net's architecture to examine its effect on the predictive performance of transition probability. Table 3.8 presents the architectures of the baseline models for three transitions and their variants. For simplicity, given that the baseline models for three transitions have similar architectures as shown in Table 3.8, this study uses the same architecture for the three transitions in the analyses. This study also omits the classifier architectures, which are the same as those in Table 3.2. All the models are independently developed and trained to facilitate an unbiased comparison. This study finds that the model's performance is sensitive to the weight and bias initialization; this trait causes some differences in the evaluation results compared with the evaluation results shown in Table 3.6. Table 3.9 presents the corresponding results of analyses of varying filter size, spatial weight layer and pooling.

The results comparison on the varying filter sizes show that the predictive performances decrease significantly as the filter size increases. In terms of the transition from forest to agriculture, the AUC-PR decreases by approximately 11% from 0.69 to 0.61, indicating that conv-net's predictive performance is sensitive to the filter size. Furthermore, the improved performance gained by the smaller filter size implies that the large

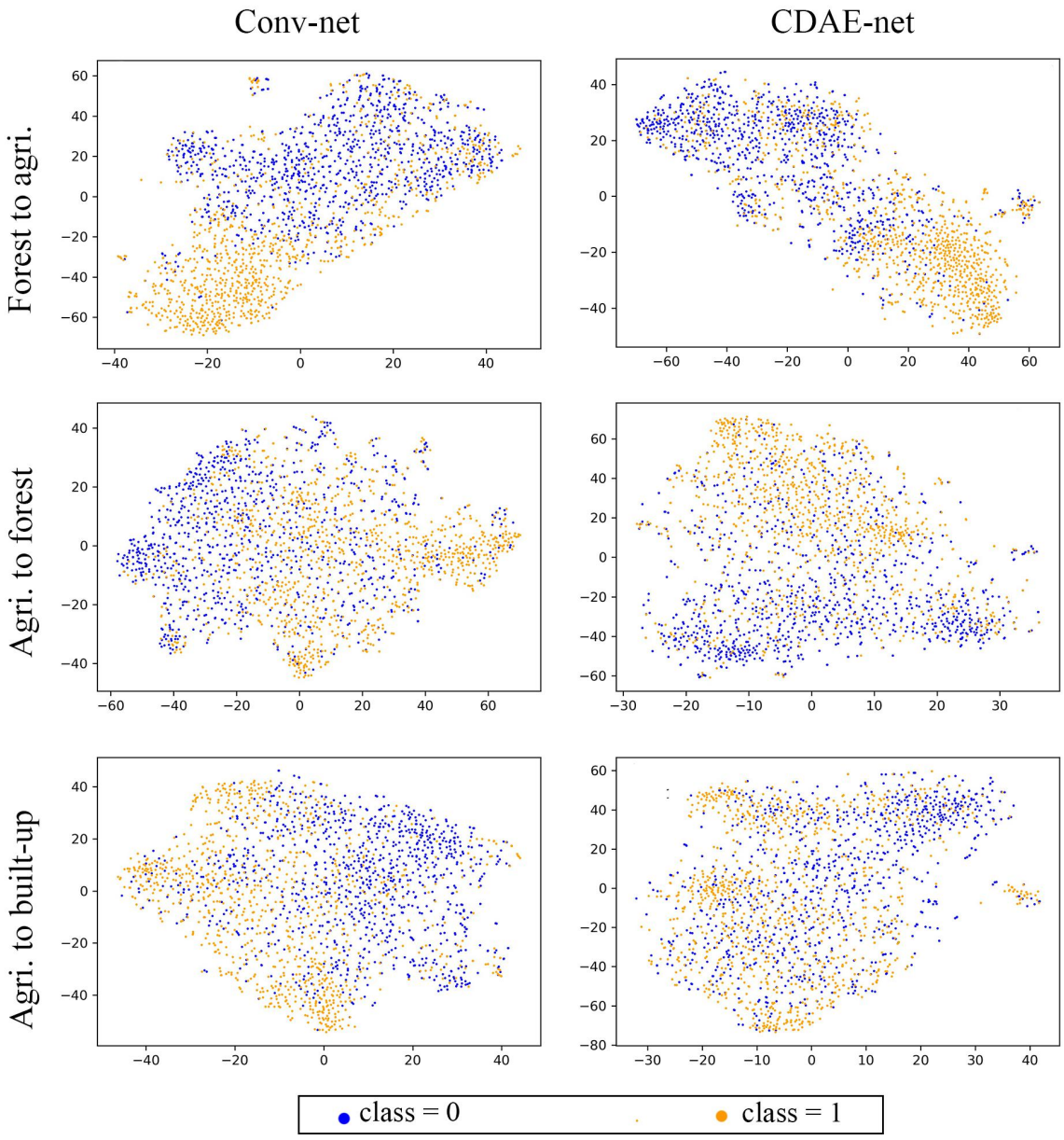


Figure 3.6: Results of t-SNE for the spatial features that are extracted from satellite images by the conv-nets and CDAE-nets

filter size is not necessary to address the redundancy of spatial information in the satellite images.

The spatial weight layer analysis is designed to identify the benefit of incorporating the specific regularization on the spatial features. The results show that AUC-ROC and

Table 3.8: The architectures of baseline models and their variants used for sensitivity analyses

Baseline models	Variants: varying filter size		Variant: no spatial weight layer	Variant: using average pooling
	5×5	7×7		
conv-128 spatial weight max pooling	conv-128 spatial weight max pooling	conv-128 spatial weight max pooling	conv-128 max pooling	conv-128 spatial weight average pooling
conv-256 spatial weight max pooling	conv-256 spatial weight max pooling	conv-256 spatial weight max pooling	conv-256 max pooling	conv-256 spatial weight average pooling
conv-512 spatial weight	conv-512 spatial weight	conv-512 spatial weight	conv-512	conv-512 spatial weight
conv-1024 spatial weight	conv-1024 spatial weight	conv-1024 spatial weight	conv-1024	conv-1024 spatial weight
global average pooling				

Notes:
1. The same convolutional architecture is used for the LUC models of all three transitions.
2. The architectures of classifiers are the same as those in Table 3.2 and thus are omitted.
3. Conv-N denotes a convolutional block composing of a convolutional layer with 3×3 filter size and N of filters, a Batchnorm layer and a ReLU layer.
4. Pooling layer has a kernel size of 3×3 and a stride of 3.

AUC-PR gradually increase by approximately 0.3 from the variant model without the spatial weight layer on the baseline models; this result indicates the effectiveness of the spatial weight layer in improving the predictive performance. However, the spatial weight layer is specifically designed as a regularization method for spatial features extracted from the satellite images, and its effect depends on the representative degree of the spatial features; for example, the performance improvement for the transition from agriculture to built-up is much smaller than forest to agriculture.

To examine the pooling effect, this study presents the binary cross-entropy losses of the training and test sets instead of the AUC to reflect the model capacity and generalization performances. The results show that the models using max pooling exhibit higher training and lower test losses than the models using average pooling; models using max pooling have smaller capacity but better generalization ability, while models using average pooling

Table 3.9: Results of the sensitivity analyses with respect to filter size, spatial weight layer and pooling

(a) Varying filter size				
		Forest to agri.	Agri. to forest	Agri. to built-up
Baseline (3×3)	AUC-ROC	0.941	0.904	0.694
	AUC-PR	0.711	0.478	0.238
Variant (5×5)	AUC-ROC	0.923	0.897	0.671
	AUC-PR	0.698	0.447	0.223
Variant (7×7)	AUC-ROC	0.901	0.860	0.648
	AUC-PR	0.625	0.419	0.209
(b) Spatial weight layer				
		Forest to agri.	Agri. to forest	Agri. to built-up
Baseline (with spatial weight)	AUC-ROC	0.939	0.906	0.693
	AUC-PR	0.712	0.470	0.239
Variant (no spatial weight)	AUC-ROC	0.924	0.906	0.684
	AUC-PR	0.688	0.447	0.232
(c) Max v.s. average pooling				
		Forest to agri.	Agri. to forest	Agri. to built-up
Baseline (max pooling)	Training loss	0.317	0.522	0.423
	Test loss	0.388	0.603	0.512
Variant (average pooling)	Training loss	0.295	0.508	0.412
	Test loss	0.410	0.624	0.527

Notes:
Training loss and test loss are the binary cross entropy losses of training set and test set, respectively.

have larger capacity but limited generalization ability. If pooling is considered as a prior on the spatial features, max pooling imposes a stronger prior than average pooling. This result is because max pooling replaces the values within a kernel with their maximum value rather than with the average. Hence, a stronger prior may be more beneficial for filtering out the useful features from the satellite images.

3.6 Summary

This study applies CNN to enhance the performance of LUC modeling. This study developed two convolutional-based models, conv-net and CDAE-net, to estimate three types of LU transition probabilities: forest to agriculture, agriculture to built-up and agriculture to forest. The results show that both conv-net and CDAE-net improve the accuracy of transition probability estimation compared with the MLP estimator, which has conventional geographical features as its sole input. Moreover, conv-net and CDAE-net achieve similar predictive performances of the estimation transition probabilities between forest and agriculture. On the other hand, CDAE-net significantly outperforms conv-net when estimating transition probability from agriculture to built-up. This result may be explained by CDAE-net's relatively more effective task handling performance for relatively complicated transitional rules and/or data with higher noise because it can learn latent representations and the denoising design.

This study's results provide several useful findings on convolutional-based model architecture. 1) Shallow architecture is sufficient for the LUC modeling task in this study. Conv-net and CDAE-net have only four and two convolutional layers, respectively; the layers are rather shallow compared with commonly used layers in computer vision studies. 2) The LUC models learn different transitional rules per the LUC process, and the model architectures could vary. Although the classifier architectures are different, their convolutional architectures are very similar for the transitions considered in this analysis. This observation indicates that the spatial features are extracted using similar learning processes. Hence, the extracted spatial features from the satellite images have similar degrees of complexity. 3) The spatial weight layer, which is specifically designed to apply distance-decay regularization on spatial features, effectively improves conv-net's predictive performance.

Chapter 4

Modeling with recurrent neural works in Tsukuba city

4.1 Motivation

Most of these variants of CA intrinsically treat the LUC process as a Markov process, meaning that the land use of the next time step only depends on the status of prior time step, although there are a few exceptions, such as SLEUTH (Clarke and Gaydos, 1998), LEI-CA (Liu et al., 2014), and survival analysis (SA)-patch-CA (Chen et al., 2016). One typical variant is the MC-CA model, in which the MC model is used to estimate the temporal variation of the LUC process to predict the future LUC demand and CA is used to predict or simulate the spatial pattern. The MC-CA model has been reported to yield accurate land use predictions in various LUC modeling tasks, particularly when statistical learning methods, such as logistic regression and neural networks, are incorporated (Arsanjani et al., 2012; Razavi, 2014).

However, using the Markov-based approach presents a major concern because the land use status of an area may depend on the historical characteristics of the area and

its neighboring areas. Hence, whether the LUC process is a Markov process is debatable. For example, 1) an area would retain its land use type for years after a change because of the persistence characteristics of land use, and 2) a non-built-up area located in a region that has been undergoing active development for a long time is more likely to change to a built-up area. Based on these two situations, the inability of Markov-based models to incorporate historical information would likely lead to inaccurate estimations. Nevertheless, the long temporal dependency issue of LUC modeling has been rarely addressed in previous studies.

Recurrent neural networks (RNNs) are state-of-the-art neural network models that are naturally suited to modeling time series data and other sequential data (Lipton et al., 2015). The fundamental principle underlying RNNs is to utilize sequential information for future predictions. With the recent development of computational schemes and deep network training techniques, RNNs have been gaining attention in various fields of environmental study (e.g. land cover detection (Lyu et al., 2016), air quality modeling (Biancofiore et al., 2015), and meteorological modeling (Le et al., 2017)). Nonetheless, RNNs have not been applied for LUC modeling.

This study examines the potential use of RNNs for modeling the spatio-temporal variation of LUC process. This study explores the benefit of modeling long-term dependency of LUC process by examining the predictive performances of two RNNs categories, simple RNN and RNN with gated architecture (long short-term memory (LSTM), LSTM-peephole and gated recurrent units (GRU)), which vary in their ability of modeling temporal dependency. The RNN models are developed by using 17 years of spatio-temporal data with high temporal resolution, and use the constructed RNN models to forecast a continuous LUC process.

4.2 Methodology and data

4.2.1 RNN models

4.2.1.1 RNN family

RNNs are essentially feed-forward neural networks with connections that span adjacent time steps, thereby introducing a notion of time to the model (Lipton et al., 2015). At time t, nodes with recurrent connections receive inputs from the current time step x_t and the hidden state h_{t-1} in the network's previous state. RNNs take the following general form:

$$h_t = f(h_{t-1},\, x_t, \theta) \tag{4.1}$$

where f is an activation function and θ is a parameter that is usually shared over time. h_{t-1} contains and accumulates information through previous time steps.

The simple RNN is a variant that was developed early and is described using the following equation:

$$h_t = tanh(W_t h_{t-1} + W_x x_t + b) \tag{4.2}$$

where tanh is the non-linear activation function, W_t and W_x are weights and b is bias, and these parameters are learnable during model training. The basic structure of the simple RNN is illustrated in Figure 4.1a. The simple RNN is known to suffer from the vanishing gradient problem, particularly with long sequential tasks (Graves, 2013). Of the several techniques that have been developed to address the vanishing gradient problem, adding a gated architecture to recurrent units (e.g. LSTM) is the most prevalent approach.

Gated RNNs adopt a completely different mechanism based on the idea of creating paths through time with derivatives that neither vanish nor explode. LSTMs implement this idea by introducing self-loops to recurrent units, which create memory cells with a core that stores cell states and several gates that control the information flow. LSTMs are

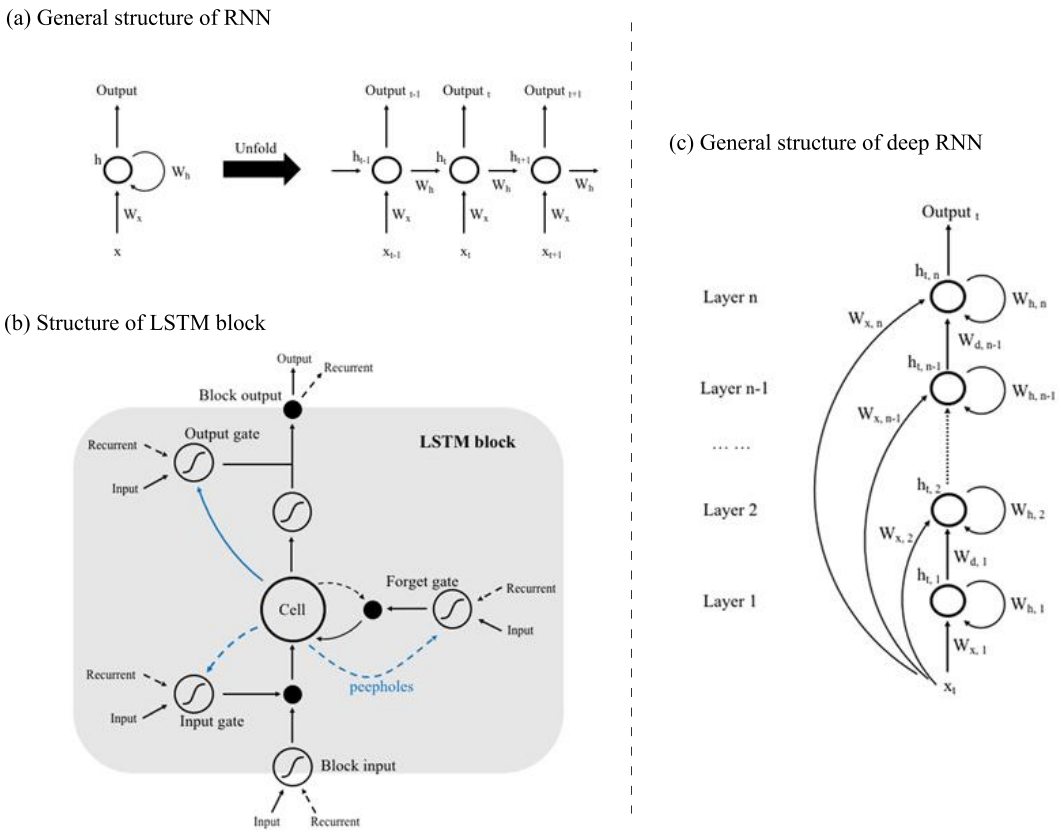


Figure 4.1: Illustration of general structure of simple RNN (a), structure of LSTM block with peephole connection (b) and general structure of deep RNN (c)

successful at modeling long-term dependency and have become the most popular RNN architecture (Greff et al., 2017). Among the various LSTM variants, this study used a classical but relatively complex architecture, LSTM with peephole, as the baseline LSTM model. This variant is described as follows:

$$z_t = tanh(W_{zh}h_{t-1} + W_{zx}x_t + b) \qquad block\ \ input \qquad (4.3)$$
$$i_t = sigmoid(W_{it}h_{t-1} + W_{ix}x_t + W_{ic}c_{t-1} + b) \qquad input\ gate \qquad (4.4)$$
$$f_t = sigmoid(W_{ft}h_{t-1} + W_{fx}x_t + W_{fc}c_{t-1} + b) \qquad forget\ gate \qquad (4.5)$$
$$c_t = f_t h_{t-1} + i_t z_t \qquad cell\ state \qquad (4.6)$$
$$o_t = sigmoid(W_{ot}h_{t-1} + W_{ox}x_t + W_{oc}c_t + b) \qquad output\ gate \qquad (4.7)$$
$$h_t = o_t tanh(s_t) \qquad block\ output \qquad (4.8)$$

Figure 4.1b illustrates an LSTM-peephole block. In addition to the LSTM-peephole model, this study uses two alternative variants: LSTM and GRU. LSTM is described as follows:

$$z_t = tanh(W_{zh}h_{t-1} + W_{zx}x_t + b) \quad block\ \ input \quad (4.9)$$
$$i_t = sigmoid(W_{it}h_{t-1} + W_{ix}x_t + b) \quad input\ \ gate \quad (4.10)$$
$$f_t = sigmoid(W_{ft}h_{t-1} + W_{fx}x_t + b) \quad forget\, gate \quad (4.11)$$
$$c_t = f_t h_{t-1} + i_t z_t \quad cell\, state \quad (4.12)$$
$$o_t = sigmoid(W_{ot}h_{t-1} + W_{ox}x_t + b) \quad output\, gate \quad (4.13)$$
$$h_t = o_t tanh(s_t) \quad block\, output \quad (4.14)$$

GRU is described as follows

$$z_t = tanh(W_{ut}h_{t-1} + W_{ux}x_t + b) \quad update\ \ gate \quad (4.15)$$
$$r_t = tanh(W_{rt}h_{t-1} + W_{rx}x_t + b) \quad reset\ \ gate \quad (4.16)$$
$$h_t = (1 - z_t)h_{t-1} + z_t tanh(W_{hx}x_t + W_{ht}(r_t h_{t-1}) + b) \quad block\ \ output \quad (4.17)$$

LSTM-peephole and LSTM have almost the same architecture, but LSTM has no peephole connection. The peephole connection allows access to the raw information stored as a cell state, thereby guaranteeing the ability of the memory cell to always learn from the past, even when the output gate is closed. GRU removes the peephole connection and the output activation function and use only two gates (reset and update) to control the forget vector and the cell-state update. Although GRU greatly simplifies the LSTM architecture, they preserve the essential functionality of the original LSTMs. GRU has been documented to achieve similar performance as LSTMs with less computational cost (Chung et al., 2014).

4.2.1.2 Deep RNNs

RNNs are intrinsically deep in the temporal dimension but shallow in the spatial dimension. Extending their depth in the spatial dimension may enhance RNNs' abilities to transform sequential inputs into high-dimensional representations and also to learn useful information from the representations (Hefron et al., 2017). Graves et al. (2013) demonstrated that increasing the depth of RNNs improve the RNNs' predictive performances more than adding memory cells. In this study, this study further develops deep simple RNN, deep LSTM, deep LSTM-peephole and deep GRU models by stacking recurrent layers in a sequence-to-sequence manner. Figure 4.1c illustrates the basic structures of deep RNN models. Each recurrent layer is composed of certain number of recurrent cells; outputs from lower recurrent layers together with the external inputs are used as the inputs of higher recurrent layers. These deep RNN models are used as supplementary models to the single-layer RNN models in order to explore the applicability of RNN models with higher capacity for the LUC modeling task in this study.

4.2.2 Study area, data and spatial features

4.2.2.1 Study area and data

The study area is the city of *Tsukuba*, which is located in the northern part of the Greater Tokyo Area in Japan, covers a total area of 283.72 square kilometers, and has an estimated population of 223 thousands. Tsukuba's urban area has continuously expanded over the past two decades; the expansion has been driven by the government development policy to build the city as a center of research and technological advancement. This study conducts land use classifications of the city's Landsat satellite images (Landsat 7 and 8) from 2000 to 2016 using a supervised classification provided by ERDAS IMAGINE 2016 V16.1 (Hexagon Geospatial, U.S.). Section 3.3 provides the description of the basic theory and general procedure of supervised LU classification. Table 4.1 presents the

definitions of the LU categories of the LU maps. This study has obtained land use maps with $30 \times 30\, m^2$ resolution with the following area types: non-built-up, built-up and water body. Figure 4.2 presents the land use maps of year 2001, 2006, 2011 and 2016.

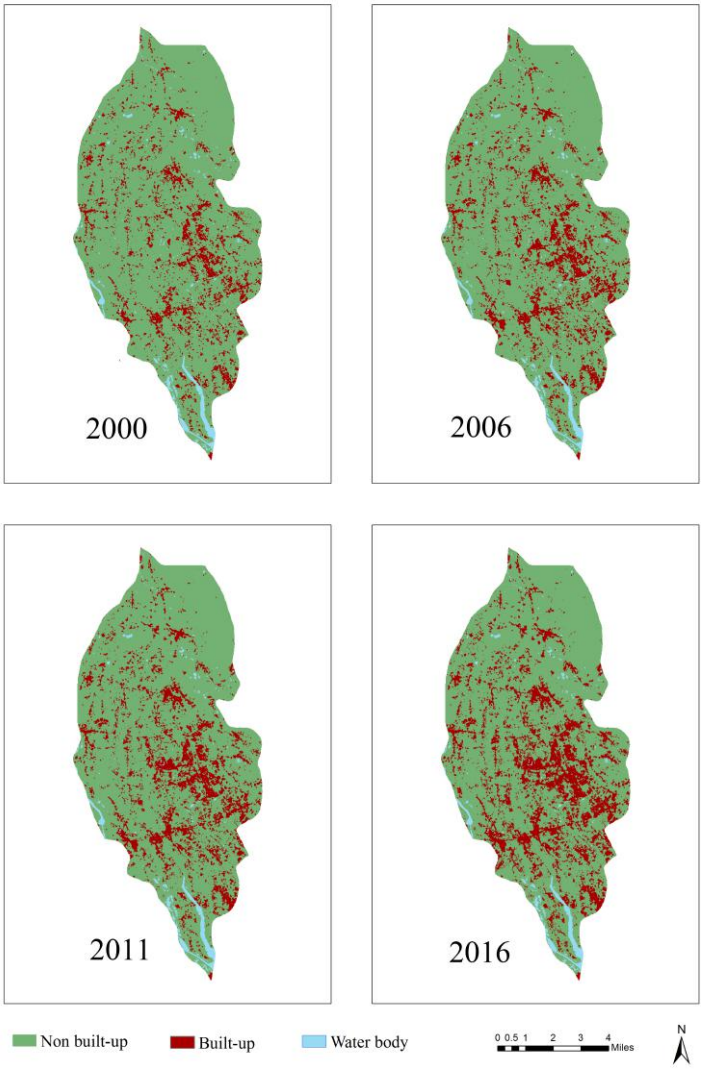


Figure 4.2: Land use maps of the city of *Tsukuba* for 2001, 2006, 2011 and 2016

4.2.2.2 spatial features

This study constructs a spatial feature set composed of neighborhood characteristics, geometric properties, proximity factors and physical factors (see Table 4.2). Out of

Table 4.1: Definition of land use categories in Tsukuba city from 2000 to 2016

LU category	Description
Built-up	Lands where residential buildings, commercial buildings, etc. are densely distributed
Non-built-up	Lands where has no densely distributed residential buildings, commercial buildings, etc.
Water body	Areas including river and river bed, artificial lake, natural lake, pond, fish farm, etc. where are filled with water for most of the time

Notes:
The definitions of LU categories are based on the information provided by National Land Information Division, National and Regional Policy Bureau of Japan
(http://nlftp.mlit.go.jp/ksj/gml/codelist/LandUseCd-09.html)

these four features categories, neighborhood characteristics and geometric properties are derived directly from the LU maps; in particular, the neighborhood characteristics and geometric properties for subsequent time step are computed based on the LU map of the prior time step. The proximity factors and physical factors are derived from collected spatial data such as road network maps and digital elevation maps.

Although previous studies have developed specific metrics to capture the neighborhood characteristics, this study uses the LU categories in a Moore neighborhood with a size of 3×3 as the input for RNN models (i.e. the LU categories in 3×3 neighborhood are raveled into a 1-d vector and are feed into the RNN models) in order to allow RNN models to access a more detailed neighborhood information. As the spatial feature, the neighborhood LU categories have intrinsic spatial autocorrelation. Nevertheless, this issue can be handled by the RNN models because of their non-linear design.

This study uses geometric properties at both the cell and patch level to capture certain spatial patterns in the neighborhood. Cell-level metrics include *distance to patch centroid*, *statistics of distance to patch edges* and *statistics of distance to neighboring patches*, and patch-level metrics include *patch area*, *patch perimeter*; *patch equivalent diameter*; *patch eccentricity*, *major axis length*. Cell-level metrics serve a similar purpose as neighborhood characteristics but cell-level metrics focus more extensively on speci-

Table 4.2: Description of spatial features used for modeling the LUC process

	Description
Neighorhood characteristics	Land use classes (non built-up, built-up or water body) of neighboring cells within certain size
geometric properties	
Cell-level	
Distance to patch centroid	Euclidean distance between target cell and the centroid of the local patch
Statistics of distances to patch edges	Mean, std. dev., and minimum of the distances between target cell and the edge of local patch
Statistics of distances to neighboring patches	Mean, std. dev., and minimum of the distances between target cell and the edge of neighboring patches
Patch-level	
Patch area	The area of local patch
Patch perimeter	The perimeter of local patch
Patch equivalent diameter	The diameter of a circle with the same area as the local patch
Patch major axis length	The length of the major axis of the ellipse that has the same normalized second central moments as the local patch
Patch eccentricity	Eccentricity of the ellipse that has the same second-moments as the local patch. The eccentricity is the ratio of the focal distance (distance between focal points) over the major axis length
Proximity	
Distance to highway	The nearest Euclidean distance between target cell and highway
Distance to major roads	The nearest Euclidean distance between target cell and major roads
Distance to railway or subway	The nearest Euclidean distance between target cell and railway or subway
Physical factor	
Elevation	Elevation of target cell
Coordinates	Coordinates of target cell

Notes:
1. Neighborhood characteristics and geometric properties are calculated based on the land use maps.
2. Proximity factors are calculated based on the road and railway networks data, which are collected from the Ministry of Land, Infrastructure, Transport and Tourism of Japan for 2000 and 2005 and OpenStreetMap database for 2009 to 2016.
3. Elevation data are collected from SRTM (Shuttle Radar Topographic Mission) database.

fying the relative location of the patch's target cell and also the neighboring patches. Essentially, these geometric properties have similar roles as the landscape metrics used in previous studies of LUC modeling: both spatial features are used to capture the spatial patterns in the neighborhood. Moreover, many landscape metrics are calculated based on geometric properties, such as patch area and perimeter. However, compared with the landscape metrics, the geometric properties are simpler in terms of computational. Moreover, geometric properties are mainly designed to describe the geometric location and shape rather than to reflect certain ecological status. This study uses the geometric properties instead of the composite landscape metrics mainly because the computation of geometric properties is relatively more efficient. The neighborhood size used for computing geometric properties is 18×18, which is determined based on trial-and-error tests to achieve a balance between predictive performance and computational cost.

4.2.3 Implementation

4.2.3.1 Modeling framework

Figure 4.3 presents the modeling framework. The cells that are located in the area of no built-up area in the initial time (year 2000) are defined as modeling area, and are used as samples for LUC modeling. The other cells are defined as persistent area. This study splits the spatio-temporal data into three sets: 1) training set composed of data from 2000 to 2010, which are used to train RNN models; 2) validation set composed of data for 2011, which are used to examine the generalization performance of RNN models while training; 3) test set composed of data from 2012 to 2015, which are used to test the forecasting performances of RNN models after the model training is completed.

The RNN models estimate the LU categories in a deterministic approach. For each cell at each time step, the RNN models continuously yield probability-like predictions of LU categories (non-built-up and built-up), which are produced from a Softmax function

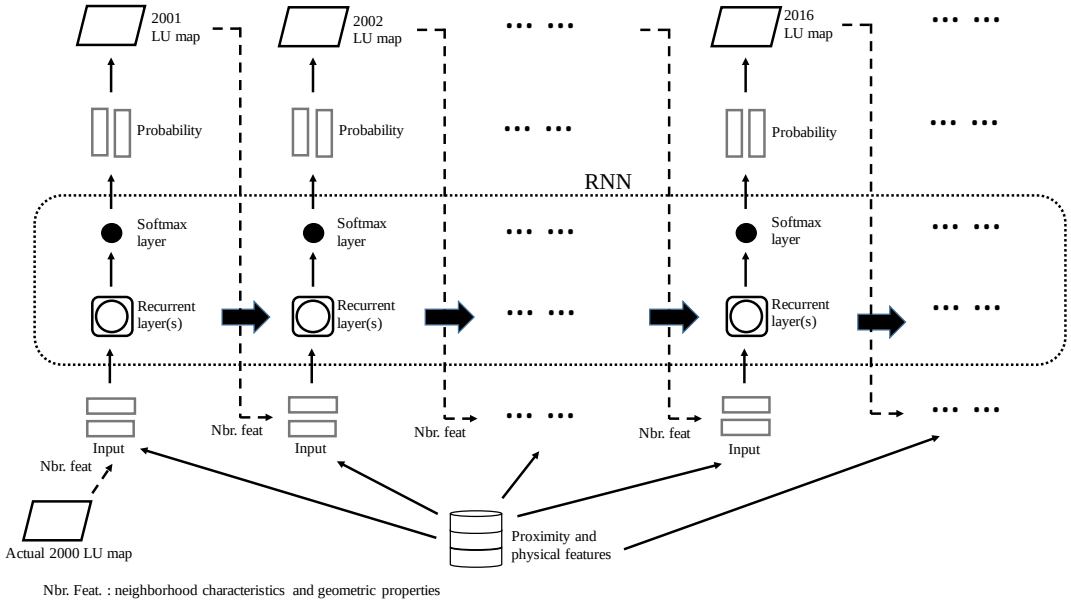


Figure 4.3: Modeling framework

layer. Then RNN models determine the final predictions of LU categories by selecting the LU category with higher probability. Based on this process, both the LU transition quantity and allocation are automatically determined by the RNN models.

The RNNs are deterministic approach which make prediction by assimilating the information in previous time steps, while CA is essentially a stochastic approach. The randomness introduced by CA may accumulate over time, hence may undermine the predictive performance of RNN models; particularly the predictive performance for time steps at the end of the temporal sequence may be undermined.

This modeling framework is coded and implemented in Python. The RNN models are developed based on the framework of PyTorch, which is a Python machine learning library. Proximity factors and physical factors are processed and generated by ArcPy (ESRI Inc., U.S.); geometric properties are processed and generated based on SciPy.

4.2.3.2 Model training

In terms of sampling, most previous studies have used random or stratified sampling to avoid the influence of spatial autocorrelation. However, the mini-batch learning criterion of a neural network naturally mitigates the influence of spatial autocorrelation to certain extent. Furthermore, this study uses a bootstrap over-sampling strategy in this study to manage the data imbalance problem (i.e., the cells with unchanged LU category are much more than the cells with changed category). In particular, a mini-batch of data is randomly sampled from the dataset with replacement, and samples that belongs to negative and positive labels have the proportion of approximately 7:3 within a mini-batch.

Due to the small study area, the sample size used in this study is limited; this may induce over-fitting problem of the RNN models. In order to avoid the over-fitting problem, this study applies various generalization methods. This study uses both dropout (Zaremba et al., 2014) and adding Gaussian noise to the gradient (Neelakantan et al., 2015b), which are shown to effectively resolve the issues in previous studies as regularization methods. This study also adopts a strict early stopping scenario to prevent the over-fitting (i.e., the model would automatically stop training if the moving average validation loss does not decrease by 5% on the basis of the moving average validation loss of prior step.) In addition, the fine-tuned RNN models turn out to have low dimensional hidden space; the number of hidden units is ranging between $1 \times number\,of\,features \sim 5 \times number\,of\,features$. In particular, the number of hidden units decreases with the increase of the RNN model depth.

This study uses Batch Normalization (Ioffe and Szegedy, 2015; Cooijmans et al., 2016), a technique shown to work well with deep neural networks to improve the performance of RNN training. This study uses a gradient clip with a threshold of 1 to prevent gradient exploding. This study uses cross-entropy as the loss function, and stochastic gradient descent as the optimization algorithm. By using both grid search and random search,

this study fine-tunes the hyperparameters of RNN models, including number of hidden units, learning rate, momentum, learning rate decay rate, Gaussian noise coefficient, dropout rate and over-sampling ratio.

4.2.4 Performance evaluation metrics

Given that the purpose of developing the RNN models is to forecast the LU maps rather than to estimate the LU transition probability maps, this study does not use the probability-based evaluation metrics such as receiver operating characteristic curve. Instead, this study uses classification evaluation metrics (accuracy and F1 score) and map comparison metrics (Kappa simulation and fuzzy Kappa simulation) to evaluate the predictive performances of RNN models.

In order to eliminate the influence of LU persistence and avoid the over-estimation of predictive performances, classification evaluation metrics are calculated based on the predicted results and actual LU in modeling area instead of the whole study area. Both accuracy and F1 score are cell-to-cell metrics, and are defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.18}$$

$$F1\,score = 2 \times \frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \tag{4.19}$$

where TP denotes true positive, FP denotes false positive, TN denotes true negative, FN denotes false negative. Compared with accuracy, F1 score measures the classification performance of positive label, which enables specific evaluation of the model's capability to predict the transitions from non-built-up to built-up land.

Map comparison metrics evaluate the agreement of predicted and actual LU maps. Kappa simulation and fuzzy Kappa simulation are developed based on Kappa statistic, which further eliminate the influence of LU persistence. Refer to section 3.2.3 for details.

4.3 Results and discussion

4.3.1 The predictive performances of RNN models

4.3.1.1 Analysis of single-layer RNN models with varying architectures

Table 4.3 presents the predictive performance evaluation results of simple RNN, LSTM, LSTM-peephole and GRU on test data (2012 $\sim$ 20116). van Vliet et al. (2013a) demonstrate that a fuzzy Kappa simulation with value higher than 0 indicate that a model explains LUC better than random guess. Table 4.3 shows that all RNN models achieve fuzzy Kappa simulations with values higher than 0.45, which indicates the validity of using RNNs to model the spatio-temporal LUC process. The RNN models with gated architecture (LSTM, LSTM-peephole and GRU models) greatly outperform the simple RNN model. Given that RNNs with gated architecture are more capable of capturing and modeling the long-term temporal dependency in the data compared with simple RNN, RNN models with greater ability to model temporal dependency can model the spatio-temporal dynamics of LUC process better.

Among the three RNN models with gated architecture, LSTM and LSTM-peephole models outperforms GRU model; this result indicates the better performance of LSTM architecture for this specific LUC modeling task. Moreover, the LSTM and LSTM-peephole models achieve similar predictive performances in terms of cell-to-cell metrics: accuracy, F1 score and Kappa simulation. However, the LSTM-peephole model slightly outperforms the LSTM model in terms of vicinity-based metrics. This result indicates that LSTM-peephole model yields relatively more 'near-hits' (i.e. correctly predict the LU category of neighboring cells rather than the central cell itself). Therefore, LSTM-peephole model is considered to have the highest predictive performance for this specific LUC modeling task.

Figure 4.4 presents the comparison of actual and predicted LU maps for 2016, and

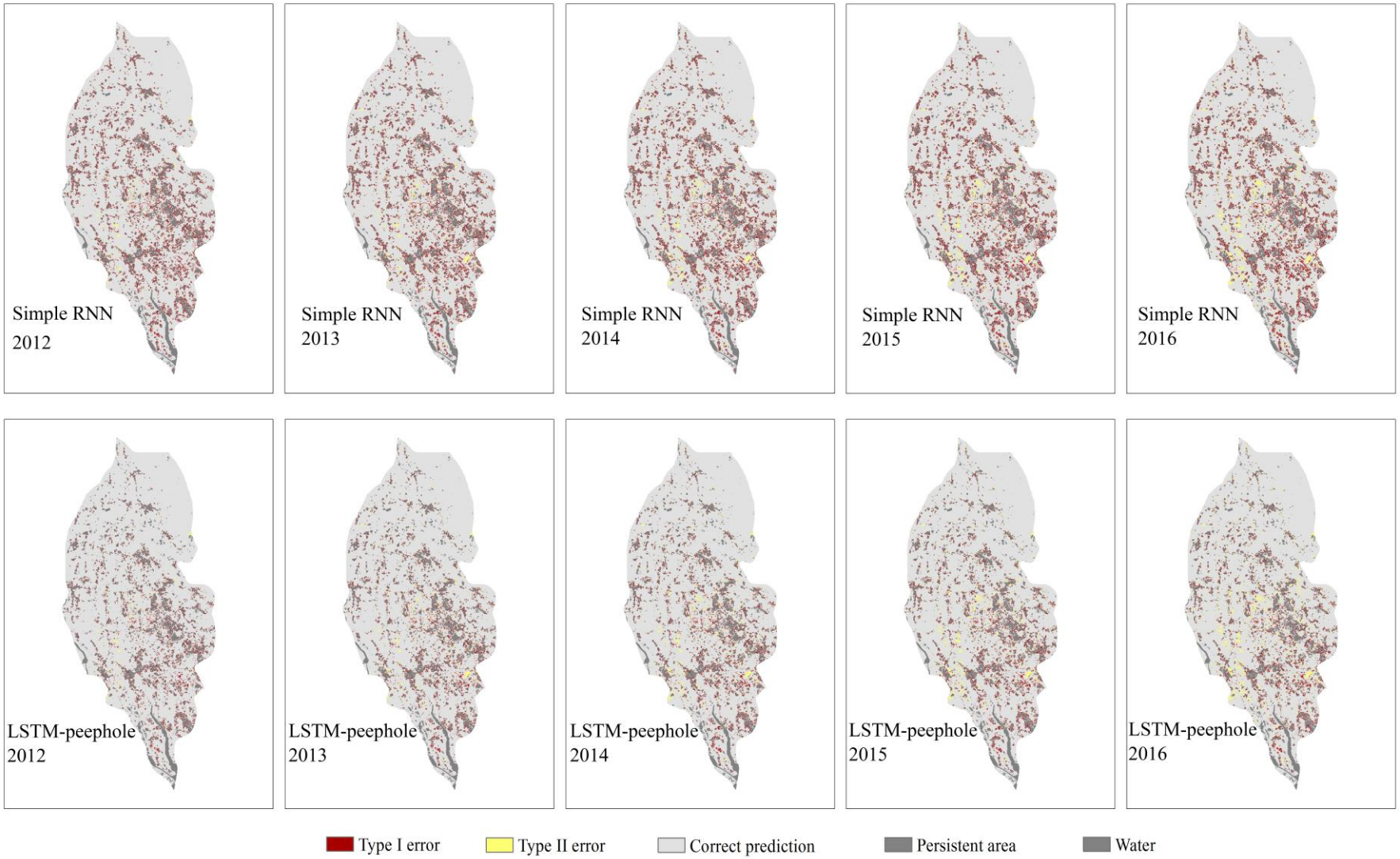
Figure 4.5 presents the spatial distribution maps of prediction errors produced by simple RNN and LSTM-peephole models. Both simple RNN and LSTM-peephole models over-estimate the quantity of LU transitions from non-built-up to built-up, given that the cells with Type I error (false positive) is significantly larger than the cells with Type II error (false negative). Furthermore, in terms of both simple RNN and LSTM-peephole, the cells with Type I error are mainly distributed around the already existed built-up land at the initial modeling time (year of 2000), while the cells with Type II error are mainly distributed at locations isolated from the already existed built-up land. This result indicates that the RNN models is capable of predicting the expansion of built-up land, but is relatively poor at predicting the emergence of new built-up land. In addition, compared with simple RNN model, LSTM-peephole model yields much less Type I error; this implies that modeling longer temporal dependency reduces the incorrect rejection of the hypothesis that LU category does not change, and by doing so improves the predictive performance for modeling LUC process. Table 4.3 shows that the prediction accuracy decreases over time for all RNN models. It should be noted that this phenomenon does not imply a modeling problem caused by over-fitting. Instead, this phenomenon seems to be observed due to the error propagation over time. According to the mechanism of RNN models, the prediction of LU for the subsequent time step is based on the hidden representations in previous time steps and also the neighborhood features and geometric properties calculated based on the prediction of LU in the prior time step. Hence, the errors of LU predictions in previous time steps would affect the accuracy of LU prediction for the subsequent time step, and leads to the observed phenomenon of relatively low accuracy of subsequent time step compared to that of the prior time step. A simple example is that a cell that is incorrectly classified into built-up land at the prior time step would generally retain the misclassified LU category at the subsequent time step; Figure 4.5 provides the visualization.

Figure 4.4: Actual and predicted LU maps for 2016

Table 4.3: Results of evaluation metrics calculated from the prediction results of RNN models from 2012 to 2016

		Accuracy	F1 score	Kappa simulation	Fuzzy Kappa simulation 3×3	5×5	9×9
simple RNN	2012	0.921	0.583	0.507	0.593	0.641	0.711
	2013	0.911	0.577	0.497	0.585	0.635	0.706
	2014	0.902	0.570	0.485	0.572	0.623	0.695
	2015	0.895	0.564	0.485	0.561	0.620	0.689
	2016	0.888	0.556	0.476	0.549	0.614	0.686
LSTM	2012	0.961	0.714	0.646	0.692	0.757	0.811
	2013	0.949	0.693	0.627	0.671	0.730	0.800
	2014	0.940	0.680	0.611	0.653	0.715	0.787
	2015	0.933	0.662	0.596	0.638	0.692	0.766
	2016	0.923	0.649	0.579	0.617	0.678	0.742
LSTM-peephole	2012	0.954	0.714	0.643	0.697	0.761	0.816
	2013	0.944	0.689	0.629	0.683	0.740	0.798
	2014	0.937	0.673	0.609	0.664	0.722	0.782
	2015	0.931	0.660	0.592	0.647	0.706	0.768
	2016	0.924	0.657	0.586	0.637	0.694	0.753
GRU	2012	0.953	0.694	0.604	0.646	0.713	0.778
	2013	0.939	0.675	0.596	0.631	0.690	0.759
	2014	0.929	0.652	0.576	0.615	0.673	0.747
	2015	0.920	0.634	0.565	0.695	0.659	0.723
	2016	0.909	0.622	0.547	0.581	0.642	0.709

Notes:
1. All evaluation metrics exclude the influence of LU persistence in different approaches, refer to section3.4 for details.
2. The calculations of Kappa simulation and fuzzy Kappa simulation uses the land use map of 2000 as initial map and compares the predicted and actual land use maps. Fuzzy Kappa simulations are calculated based on three different neighborhood size of neighborhood membership function: 3×3, 5×5, and 9×9.



Notes: Type I error: false positive, incorrect rejection of the hypothesis that LU category does not change; Type II error: false negative, failure to reject the hypothesis that LU category does not change.

Figure 4.5: Spatial distribution maps of errors of prediction results generated by simple RNN and LSTM-peephole models from 2012 to 2016

4.3.1.2 Analysis of varying sequential length

In order to further explore the benefit of modeling temporal dependency, this study examines the predictive performances of LSTM-peephole models with varying sequential length of training set. The LSTM-peephole models are independently fine-tuned to facilitate an unbiased examination. Table 4.4 presents the results. According to the values of all evaluation metrics, the predictive performances of LSTM-peephole models decrease with the decrease of sequential length of training set. The decrease of sequential length leads to the loss of temporal information and temporal relationship that could be learned and used by the LSTM-peephole model. This result shows the benefit of incorporating rich temporal information of LU to model the LUC process.

Given that this study is a case study rather than a comparative study, this study does not develop a LUC model that is trained with data in two time steps to specifically compare with the RNN models. However, the analysis of varying sequential length provides a special case, which enables a rough comparison to show the predictive performance improvement by the inclusion of the temporal dependency in modeling of RNN models. When the sequential length decreases to 2, the LSTM-peephole model is trained using the data in 2009 and 2010, is validated using the data in 2011, and then is tested using the data from 2012 to 2016. In this case, the LSTM-peephole model cannot learn any further useful temporal dependency from the training data given that only the variation between two time periods are available. Compared with this limited LSTM-peephole model, the baseline LSTM-peephole model achieves F1 score and Kappa simulation that are approximately 0.16 and 0.17 higher, respectively.

4.3.1.3 Analysis of deep LSTM-peephole models

According to the results in Table 4.3, single-layer LSTM-peephole model is shown to have the highest predictive performance. This study further develops deep LSTM-peephole

Table 4.4: Results of evaluation metrics calculated from the prediction results of LSTM-peephole model with varying sequential length of training set for 2016

	Accuracy	F1 score	Kappa simulation	Fuzzy Kappa simulation		
				3×3	5×5	9×9
11 (Baseline)	0.924	0.657	0.586	0.637	0.694	0.753
10	0.917	0.642	0.553	0.628	0.688	0.75
9	0.916	0.641	0.55	0.636	0.677	0.745
8	0.889	0.591	0.523	0.577	0.624	0.701
7	0.900	0.608	0.521	0.579	0.62	0.694
6	0.889	0.589	0.494	0.55	0.598	0.673
5	0.870	0.562	0.472	0.516	0.548	0.636
4	0.882	0.576	0.458	0.497	0.545	0.633
3	0.871	0.548	0.431	0.477	0.534	0.614
2	0.866	0.542	0.429	0.473	0.534	0.608

Notes:
1. All evaluation metrics are calculated from the LU prediction for 2016. 2. All evaluation metrics exclude the influence of LU persistence in different approaches, refer to section3.4 for details.
3. The calculations of Kappa simulation and fuzzy Kappa simulation uses the land use map of 2000 as initial map and compares the predicted and actual land use maps. Fuzzy Kappa simulations are calculated based on three different neighborhood size of neighborhood membership function: 3×3, 5×5, and 9×9.
4. Sequential length represents the time span of annual data in training set. In baseline model, training set contains data of 11 years from 2000 to 2010.
5. All results are obtained from LSTM-peephole models with independently fine-tuned hyper-parameters.
6. The model with sequential length of 11 is the baseline model.

models with varying model depth (depths of 1, 3, 5, and 8) to examine the applicability of RNN models with higher capacity. The deep LSTM-peephole models are independently fine-tuned; Table 4.5 presents the performance evaluation results. The results show that the single-layer LSTM-peephole model outperforms the other models with more layers. A possible explanation is that the deep models are unnecessarily complex for this particular LUC modeling task. Hence, the models have relatively low generalization performances even though strict regularization methods are implemented. Another possible explanation is that deep RNN models are intrinsically difficult to train because of the complex gradient flow. Consequently, the relatively poor training performances of deeper RNN models lead to the relatively poor generalization performances. Regardless of the specific reasons, the result indicates that single-layer LSTM-peephole model is sufficient to resolve this

particular LUC modeling task.

Table 4.5: Results of evaluation metrics calculated from the prediction results of deep LSTM-peephole model with varying model depth for 2016

	Accuracy	F1 score	Kappa simulation	Fuzzy Kappa simulation		
				3×3	5×5	9×9
1 (baseline)	0.924	0.657	0.586	0.637	0.694	0.753
3	0.912	0.634	0.554	0.588	0.652	0.697
5	0.910	0.627	0.549	0.582	0.645	0.696
8	0.907	0.606	0.500	0.574	0.618	0.684

Notes:
1. All evaluation metrics are calculated from the LU prediction for 2016. 2. All evaluation metrics exclude the influence of LU persistence in different approaches, refer to section3.4 for details.
3. The calculations of Kappa simulation and fuzzy Kappa simulation uses the land use map of 2000 as initial map and compares the predicted and actual land use maps. Fuzzy Kappa simulations are calculated based on three different neighborhood size of neighborhood membership function: 3×3, 5×5, and 9×9.
4. All results are obtained from LSTM-peephole models with independently fine-tuned hyper-parameters.
5. The model with depth of 1 (single layer) is the baseline model.

4.3.2 Predictions

Given that the single-layer LSTM-peephole model achieve the highest predictive performance for this specific modeling task, this study uses the single-layer LSTM-peephole model to forecast the future LU maps of the city of *Tsukuba*. Figure 4.6 shows the LU maps from 2012 to 2026. The LU prediction from 2012 to 2017 are also presented in order to show the continuous LUC process and the variation of LU pattern that are predicted by the LSTM-peephole model. According to the Figure 4.6, the built-up land in the city of *Tsukuba* will continue expanding over the next decade. Figure 4.7 depicts the time trend of the area of built-up land in previous years and in the future years. Given that the LSTM-peephole model is trained using the historical data, it can be interpreted that the forecasted expanding speed of built-up land is similar with the actual speed observed in previous years.

However, Figure 4.6 also uncovers a limitation of the LSTM-peephole model. The built-up lands have expanded significantly in the bottom area of the city of *Tsukuba*, which are marked by red circles in the maps. In reality, large-scale land use transitions usually take a few years to complete and do not occur as a sudden transition within a year or less. This mismatch indicates the model's inability to predict the precise timing of the LUC process. This problem may be partially caused by the direct use of discrete number changes to represent the LUC process. Given that the inputs do not introduce any explicit timing indicator into the LSTM-peephole model, the model can only learn an implicit notation of time from the temporal variation of LU through its internal mechanism, which is rather insufficient to learn the precise timing of LUC process.

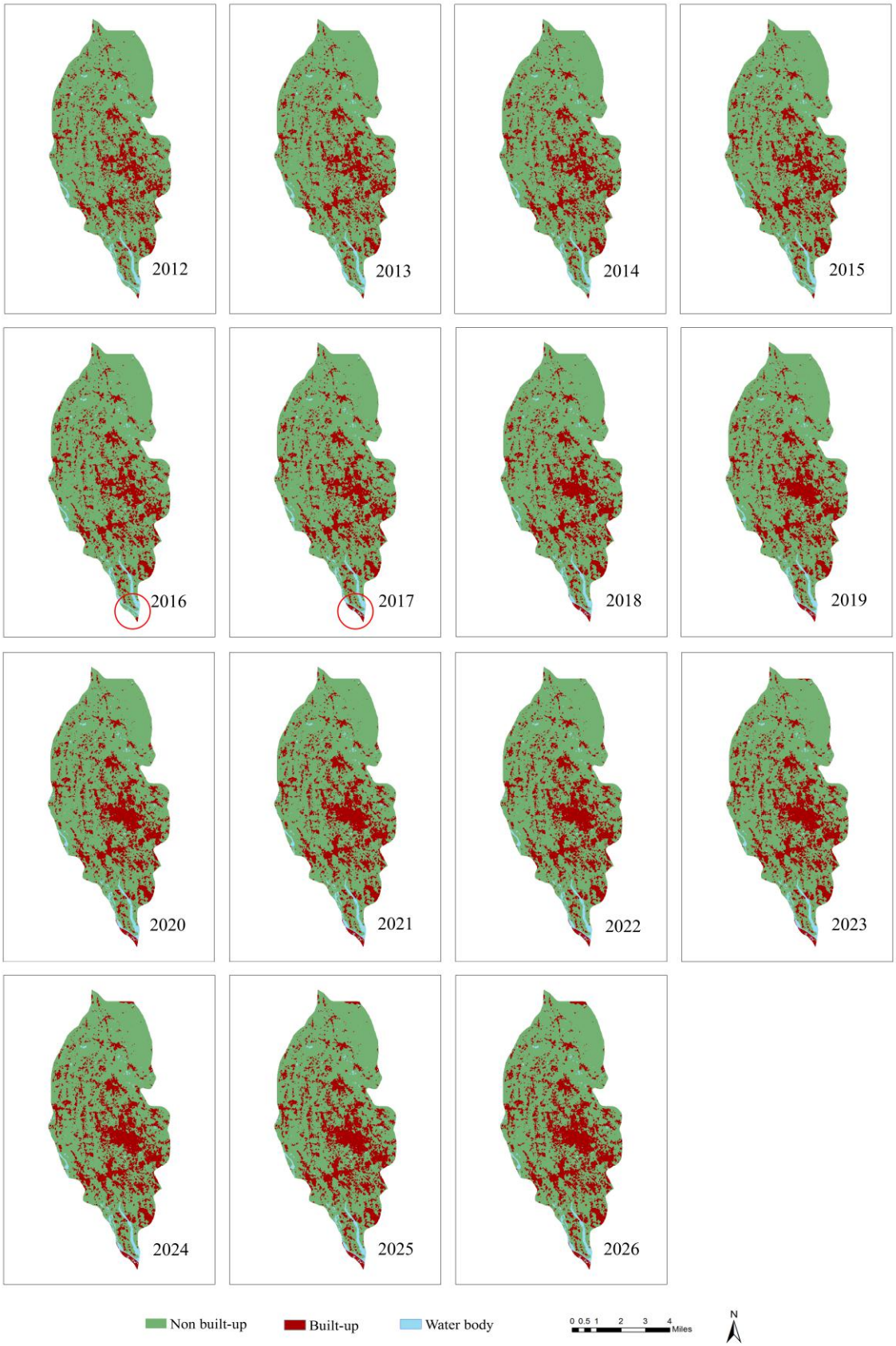


Figure 4.6: Land use maps in the city of *Tsukuba* produced by LSTM-peephole model from 2012 to 2027

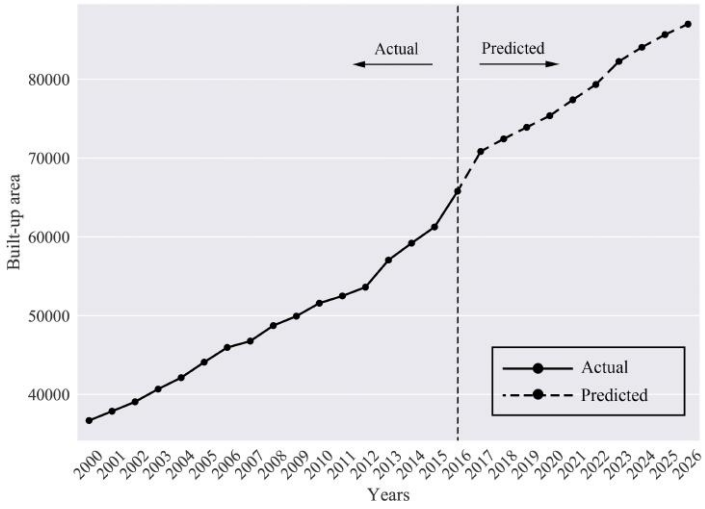


Figure 4.7: Time trend of built-up area in the city of *Tsukuba* from 2000 to 2027

Notes: Solid line shows the actual built-up areas in the period between 2000 to 2016. Dash line between 2017 to 2027 indicates the predicted built-up areas forecasted by LSTM-peephole model.

4.4 Summary

This study applied RNNs to model the spatio-temporal dynamics of LUC process in the city of *Tsukuba*, Japan. Four RNN models (simple RNN, LSTM, LSTM-peephole and GRU models) are trained, validated and tested on annual LU and the other spatial data of the city from 2000 to 2016. The predictive performances of the RNN models are evaluated using both classification metrics (accuracy and F1 score), and map comparison metrics (Kappa simulation and fuzzy Kappa simulation); both evaluation metrics free from the influence of LU persistence on the predictive performance evaluation. According to the implication of fuzzy Kappa simulation van Vliet et al. (2013a), all the RNN models perform significantly better than a random model (predict the LUC by random gauss) given that the fuzzy Kappa simulation values of each model for per test year (2012, 2013, 2014, 2015, 2016) are significantly higher than 0 (the lowest value is approximately 0.48); these values indicate the validity of the approach that uses RNNs to model the

spatio-temporal dynamics of LUC process. In particular, out of the four models, the LSTM and LSTM-peephole models significantly outperform the other two models, and the LSTM-peephole model slightly outperforms the LSTM model.

Our results also indicate that the ability to model the temporal dependency of LUC process greatly influences the predictive performances of modeling LUC process with RNN models. LSTM, LSTM-peephole and GRU models, which have advanced gated architecture compared with simple RNN, significantly outperform the simple RNN model. This result indicates that the ability of model to further take longer temporal dependency in to account improves the RNN model's predictive performance. Moreover, the predictive performance of LSTM-peephole model gradually decreases with the decrease of temporal sequential length of training set. Given that the temporal sequential length of training set represents the richness of learnable temporal relationships in the spatio-temporal data, this result indicates that the predictive performance of LSTM-peephole model benefits from modeling longer temporal dependency.

Part II

Air pollution and subjective well-being

Chapter 5

Background

Understanding people's subjective perception of environmental problems is crucial in the field of environmental economics and environmental impact assessment. Various approaches have been used to evaluate the impact of environmental problems. Among these approaches, a growing amount of literature uses the subjective well-being (SWB) approach, which emphasizes the environmental impact on people's subjective evaluation of their own well-being.

According to previous studies, the SWB approach provides an effective quantitative evaluation of the environment as public goods and enables us to analyze the effect of environmental problems on people's welfare. Subjective evaluation allows us to incorporate people's environmental concerns in addition to the stated-preference (e.g. Wang and Mullahy, 2006) or revealed-preference approaches (e.g. Kim et al., 2003) that have traditionally been used by economists to incorporate subjectivity into such evaluations. Self-reported well-being is regarded as a robust empirical approximation of overall utility; thus, it is meaningful to use subjective well-being (SWB) for a direct evaluation of environmental quality (e.g. Welsch, 2009).

Previous studies have shown that in addition to income and various personal and

household characteristics, environmental conditions have a statistically significant impact on SWB, and the SWB approach provides an effective quantitative evaluation of the environment, e.g., climate change (e.g. Sekulova and van den Bergh, 2013), airport noise (e.g. van Praag and Baarsma, 2005), ecosystem diversity airport noise (e.g. Ambrey et al., 2014), flood disaster (e.g. Van Ootegem and Verhofstadt, 2016) and air pollution (e.g. Welsch, 2002, 2007; Ferreira et al., 2013; Levinson, 2012).

There is a line of studies that analyze the impact of air pollution on people's subjective-welling across the world at different scale. Welsch (2006) analyzed the impacts of a series of air pollutants including NO_2, total suspended particulates (TSP) and Lead (Pb) in 10 European countries and found that NO_2 and Pb have a statistically significant negative effect on the life satisfaction measure. Ferreira et al. (2013) combined the SWB approach with a Geographic Information System (GIS) technique to assess the impact of the SO_2 concentration on life satisfaction with regional level data from 23 European countries. Levinson (2012) used the level of PM_{10} as a proxy for air quality and evaluated its impact on the happiness rating in the U.S. In addition to country-scale analyses, regional or city level analyses are also available. MacKerron and Mourato (2009) conducted an Internet survey and collected self-reported life satisfaction (LS) data in London and analyzed the effect of NO_2 on the level of LS. Similarly,Ambrey et al. (2014) used the LS approach to estimate the cost of air pollution (PM_{10}) in Southeast Queensland in Australia. Moreover, Cuñado and de Gracia (2013) evaluated the roles of both air pollution (PM_{10} and NO_2) and climate change to explain the regional differences in life satisfaction among Spanish regions. Overall, the results of previous studies show that in most cases, air pollution has a significant negative impact on people's SWB.

However, previous studies mainly conduct analyses with a particular focus on the impact of environmental quality are mostly concentrated in developed countries mainly because of data availability. There is a set of SWB studies for China, but most studies focus on the impact of individual attributes and social issues (See (Bian et al., 2015) for overview). Nonetheless, several recent studies have examined the effect of environmental

quality on Chinese people's SWB. Smyth et al. (2008, 2011) examined the relationship between air pollution and SWB mainly in urban areas of China using originally collected survey data on SWB. Smyth et al. (2008) used a 2003 survey with 8,890 valid responses from 30 major Chinese cities and found that respondents living in cities with relatively high SO_2 emission levels reported significantly lower SWB. In 2007, Smyth et al. (2011) conducted a survey with 2,741 participants in six Chinese cities and found that the atmospheric pollution (SO_2 and suspended particle concentration) had a significant negative effect on the originally constructed personal well-being index. Using happiness data collected in 2012, Li et al. (2014) examined the effect of estimated perceived risk from air pollution on happiness in mining areas of China. Their results showed that air pollution significantly lowered people's happiness and suggested that air pollution reduction is an important policy measure to improve people's happiness. Xu and Li (2016) also reported negative effects of air pollution on happiness based on happiness measures from the World Values Survey 2007 and subjective air pollution perceptions.

In this field of subjective well-being analyses, there is a trend in recent years to disaggregate SWB and air pollution data for allowing a precise assessment for individuals. Some studies have incorporated advanced techniques such as GIS or atmospheric modeling techniques to match individual survey data and location-specific air pollution data. Ferreira et al. (2013) used a spatial interpolation method (inverse distance weighting) to generate individual-level SO_2 concentrations for respondents in 23 European countries to analyze the pollutant's relationship with people's life satisfaction. Levinson (2012) used a weighted-distance interpolation method to estimate individual-level PM_{10} concentration in the U.S. Orru et al. (2016) also used a Eulerian air quality dispersion model to generate PM_{10} data for 30 nations across Europe. Based on air pollution datasets created by various estimation techniques, these studies found that robust significant negative effects of air pollution were reported. This study also incorporates GIS techniques to match location-specific pollution data to individual survey data.

Impact assessment of air pollution in the urban cities are important to assess given

that more than 80% of people worldwide, who live in urban areas are exposed to air quality levels that exceed WHO limits. In particular, Chinese urban areas have received much attention of policy makers and researchers due to alarming pollution level. In 2014, Chinese president Jinping Xi said at an official government press conference, 'air quality has directly affected Chinese people's happiness.' In recent years, Beijing has frequently experienced heavy haze in the winter (Han et al., 2015; Zhang et al., 2016). In January 2013, the daily $PM_{2.5}$ concentration frequently exceeding the recording range of the monitoring instruments. The government enacted a series of regulations and invested extensively in air pollution abatement. Similarly, in Shanghai, due to the dramatic increases in energy consumption and pollutant emissions caused by recent rapid urbanization, air quality and visibility have been deteriorating, and serious haze episodes have become more numerous (Gao et al., 2011; Wang et al., 2012). The government has begun to express concern over the impact of air pollution on residents' well-being.

This study aims to analyze the impact of air pollution issue on the subjective well-being of urban residents in China. In particular, this study focuses on the North part of China, where is suffering from the air pollution issue most. This study has two stages: firstly, this study conducts analyses in the Northeast China, where is a heavy industrial area under declining economy; secondly, this study conducts a comparative analyses in Beijing and Shanghai, which are the largest cities in Northern and Southern China, respectively. Both analyses collect subjective well-being measures from an original Internet survey that took place during the January and February, 2016. The analyses in Northeast China use aggregated air pollution data because of limitation of data availability in the area; while the analyses in Beijing and Shanghai use disaggregated air pollution data, which is produced from Ordinary Kriging interpolation method. The relationship between air pollution and the subjective well-being is examined by regression analyses. Furthermore, the monetary value of air pollution is also estimated based on the results of regression analyses.

Chapter 6

The impact of air pollution on subjective well-being in Northeast region of China

6.1 Motivation

This study builds on the previous empirical analyses on air pollution and SWB by examining the impact of $PM_{2.5}$ on life satisfaction using an original survey conducted in the northeast region of China in January and February 2016. Air pollution has been and continues to be a serious environmental problem for China, and public attention and concern has surged exponentially surged in the past few years. One of the main reasons is worsening air quality. Frequent heavy hazes have been observed, particularly in northern and eastern China in winter. The recent hazes with high concentrations of $PM_{2.5}$ are characterized by experts as 'extremely severe and persistent' (Huang et al., 2016; Zhang et al., 2017). In recent years, policy makers have recognized the damages of air pollution and have taken action to control them. The Chinese State Council issued the Action

Plan of Air Pollution Prevention and Control in 2013, which urged local governments to set 5-year pollution reduction targets.

Among air pollutants, $PM_{2.5}$ has received particular attention from policy makers and the general population (Zhao et al., 2013; Li et al., 2015a). Since the U.S. embassy in Beijing began recording and publishing daily $PM_{2.5}$ levels in 2011, the recording practice has spread across China, and timely data are made public through Internet websites and mobile applications. With the public's access to information, Chinese government added $PM_{2.5}$ as major pollutant for regular monitoring in the updated 2012 China's Ambient Air Quality Standards. Hence, our survey reflects the recent increase in attention to $PM_{2.5}$ and general air pollution in China. In addition to the standard analysis of pollution's impact on SWB, this study considers the possibility of varying effects of $PM_{2.5}$ concentration across different demographic groups by examining the interaction effects between pollution measures and respondents' subjective health evaluation, whether they have young children and their environmental awareness. Furthermore, using the results of regression analyses, this study calculates the monetary value (MV) or willingness-to-pay (WTP) or of air pollution for a reduction in the $PM_{2.5}$ concentration.

6.2 Data and variables

6.2.1 Study area

In this analysis, this study focuses on the northeast region of China, which covers 787 thousand square kilometers (12% of China's total area) and contains 109 million people (8% of China's total population). The region consists of 3 provinces with 35 prefecture-level cities. Because China is a massive country that contains different cultural entities, the uniform culture of this given area allows us to analyze a sample of respondents who share similar definitions of life satisfaction or happiness. This particular region is

distinctly different from other parts of China in terms of the shared history, culture, customs and native dialects within the region. The residents of this region are known to have relatively strong senses of identity based on an attachment to the region, and the people have often been referred to as "Northeast men".

Welsch (2009) argued that the level of happiness was comparable if people shared a common opinion of what happiness is. Moreover, environmental issues such as air pollution are localized in a physical, social and cultural context in which the individuals live, work and interact with one another (Bickerstaff and Walker, 2001). Thus, regional analysis allows for a more accurate extraction of the impact of a particular determinant on SWB without the need to control the variations in the sources that must be considered if this study extends the study area.

6.2.2 Survey

Recent studies on SWB in China have used the Chinese General Social Survey (CGSS) as common data to study the SWB-related questions posed in these studies (Huang et al., 2016; Zhao, 2012). In this study, this study uses an original Internet survey administered in Northeast China in January and February 2016. Compared to previous surveys used in SWB studies in China, this survey has expanded coverage of questions on people's environmental awareness and behavior as well household characteristics that are relevant to studying the impact of environmental quality on SWB. Web-based surveys have an advantage in avoiding interviewer bias caused by arbitrary factors (such as the appearance or gender of interviewers) in responses to sensitive questions, such as household income, employment status and MV (Welsch, 2009).

Out of 1,208 respondents, this study eliminates those who did not provide an answer regarding household income; consequently, 1,002 respondents remain for the analysis. Table 6.1 provides a comparison between the demographic distribution of our sample

and other available data. First, this study compares the age and gender distributions of the survey data with official statistics from 2015 Provincial Statistical Yearbooks for the Northeast region. The age distributions in Table 6.1 indicate that our sample is skewed toward a younger population. Because this study expects to experience some difficulty collecting responses from the age group over 60, this study substitutes the insufficient share of respondents over age 60 with respondents in the age group of 50∼59. Hence, this study has a higher proportion of respondents in the age group of 36∼60. In terms of gender, the ratio in our sample is close to the ratio in statistical data.

Table 6.1: Socio-demographic characteristics of respondents

	This survey (N=1002)	Comparator[a]
Age[b]		
20 ∼ 35	32%	28%
36 ∼ 50	33%	52%[c]
51 ∼ 60	31%	
61 ∼	4%	20%
Gender		
Male	53%	51%
Female	47%	49%
Gross household income[d] (CNY)		
Mean (Std. Dev.)	108,561 (78,559)	71,970 (71,261)
0 ∼ 10,000	6.5%	3.2%
10,001 ∼ 25,000	4.9%	13.1%
25,001 ∼ 75,000	22.6%	54.4%
75,001 ∼ 150,000	45.4%	24.4%
150,001 ∼ 300,000	19.3%	3.4%
300,000 ∼	2.5%	11.3%

Notes:

[a] Age and gender distributions of Northeast China are obtained from the 2015 Statistical Yearbooks of Heilongjiang, Jilin and Liaoning Provinces. Gross household income distributions are derived from Chinese General Social Survey (CGSS, 2013). Responses in Northeast area (758 samples) are extracted from the original CGSS dataset.

[b] Given that there is almost no respondent under 20 years old in this survey, for presenting a comparative result, population under 20 years old are excluded from the original statistics dataset.

[c] 2015 Provincial Statistical Yearbooks only provide the share of population aging between 36 ∼ 60.

[d] CGSS 2013 only has annual gross household income of 2011, while our survey collects gross household income of 2015. To deal with this temporal difference, this study collected yearly income increase rates between 2011 and 2015 from Statistical Yearbooks, and then estimate the household income of 2015 by applying each year's income increase rate to income of 2011.

Since official yearbooks have individual annual disposal income data and this study collected pre-tax annual household income, this study compares our income distribution with CGSS data that are used in previous studies. This study extracts the 2011 income data of our target region and then predict income levels in 2015 by adjusting the increase in income, calculated from statistical yearbooks. The average annual gross household income of our respondents is approximately 36,000 CNY (5,500 USD) higher than that in the CGSS. In both data sets, observations are available in income categories, and the majority of respondents have a household income level between 25,000 CNY $\sim$ 150,000 CNY. However, over half of respondents in CGSS have a household income between 25 and 75 thousand CNY, while nearly half of our respondents have an income between 75,000 and 150,000 CNY. The relatively high income may be partially due to insufficient responses from elders who have retired, as mentioned above, and the comparison implies that the results of our study place a heavier emphasis on the attitude of the urban population in the labor force.

6.2.3 Subjective well-being measure

Since SWB-related questions are vulnerable to the context in which they are asked, the SWB questions were placed at the beginning of our questionnaire. Previous studies have used both life satisfaction and happiness as a measure of SWB. Although these measures are sometimes used interchangeably, this study uses life satisfaction (*manyi*) rather than happiness (*xingfu*) in our analysis based on the distinction between the two in the Chinese language. The Chinese term *"xingfu"* is usually used as a translation for happiness, which is the closest descriptor of an overall evaluation of one's life, with particular emphasis on interpersonal relationships and other emotional factors. On the other hand, *"manyi"* emphasizes satisfaction with the relative standard of living or material comforts (Chen et al., 2015). Because this study primarily focused on the impact of air pollution, it is more appropriate to use life satisfaction (*manyi*).

This study asked, "Overall, how satisfied are you with your life?" and respondents were given 5 options on the scale: 1) completely unsatisfied, 2) somewhat unsatisfied, 3) neither satisfied nor unsatisfied, 4) slightly satisfied, and 5) Completely satisfied. Figure 6.1 shows the distribution of the life satisfaction rating for the respondents. The average score was 3.85 and the median score was 4. Approximately 54 percent of the respondents chose "slightly satisfied" (a score of 4), which is consistent with evidence from a psychological study that Chinese respondents tend to self-report moderate scores in psychometric tests because modesty is highly valued in the Chinese culture (Lau et al., 2005). Nevertheless, this study finds variation in reported life satisfaction. The distribution is right-skewed toward higher satisfaction scores, but within a score of 3 to 5, it is normally distributed.

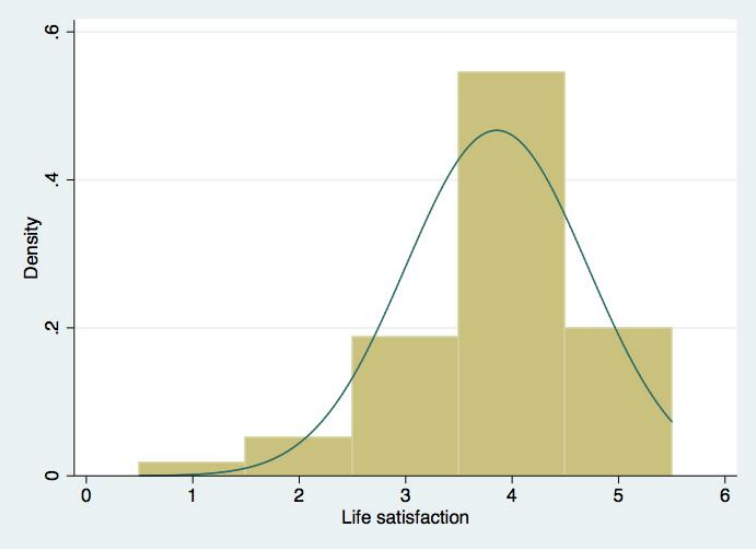


Figure 6.1: Distribution of Life satisfaction rating (N=1002)

6.2.4 $PM_{2.5}$ measures

In this study, this study uses hourly $PM_{2.5}$ data published by China's National Environmental Monitoring Center for 34 cities in Northeast China from January 2015 to February 2016. Most previous studies that analyze the effect of air pollution on SWB have used the annual average concentration level of air pollutants as the pollution measure (e.g. MacK-

erron and Mourato, 2009; Smyth et al., 2011; Levinson, 2012; Ferreira et al., 2013). The time period selected may be limited by the available data, and previous studies are not concerned with whether the air pollution effect varies depending on time specification. In this study, in addition to the commonly used annual average air pollution measures, this study uses the one-month and three-month averages of $PM_{2.5}$ concentration. The results of the pollution measures can be used to check the robustness of pollution effects, and they show which time specification people would react to most strongly. To best proxy for the pollution concentration level felt by the respondents, this study calculates the average measures based on the date on which each respondent responded to the survey. Given that the survey was conducted over a two-month period, respondents living in the same city had varying pollution measures depending on when they took the survey.

Figure 6.2 shows the maps of the city-level air pollution data in Northeast China with different average $PM_{2.5}$ concentration measures. With all three measures, relatively high $PM_{2.5}$ concentrations were recorded in the densely populated central and southern parts of the region where most of Northeast China's heavy industrial factories are located. In contrast, the northern part of the region, with a smaller population and fewer industries, is less polluted. In extreme cases, the pollution level of $PM_{2.5}$ is more than 4 times larger than 35 $\mu g/m^3$, which is the official standard established by China's Ministry of Environmental Protection (China's Ambient Air Quality, GB3095-2012).

In addition, the severity of the $PM_{2.5}$ pollution varies with time specifications; the three-month > one-month > one-year concentration averages with concentration ranges are 30 $\mu g/m^3 \sim$ 130 $\mu g/m^3$, 25 $\mu g/m^3 \sim$ 80 $\mu g/m^3$, and 25 $\mu g/m^3 \sim$ 70 $\mu g/m^3$, respectively. The order of severity may partly be due to our survey period, which was during winter, the season with the most serious pollution problem because of fossil fuel combustion used in heating and because of the stationary meteorological condition (Li et al., 2015b). The air pollution during winter usually reaches its peak in December and then begins to decrease in January (Rohde and Muller, 2015).

6.2.5 other control variable

Table 6.2 summarizes all the variables used in the analyses including all the other control variables. this study controls the individual and household characteristics that have been frequently used in previous studies to analyze the determinants of SWB. This study uses the following individual characteristics as control variables: age, gender, subjective health condition, education, number of (young) children, marital status, and modes of commuting (Huang et al., 2016; Qian and Qian, 2015). Age and age-squared are both included in the estimation to test the U-shaped effect of age on SWB that has been previously found by scholars (Lange, 2010; Rao et al., 2016). To control for the personality of respondents, this study uses aggregated personality measures calculated using 10 personality-related questions. Using factor analysis, this study identifies three personality traits: passive, stable and outgoing. This study uses the predicted values for each trait to control for variation in personalities among respondents.

Table 6.2: Variable description

Variable	Description	Mean (Std. Dev.)
Subjective well-being		
Life satisfaction	"Overall, how satisfied are you with your life?" (1-5 scale)	3.85 (0.85)
Air pollution		
$PM_{2.5}$-one month average	Average $PM_{2.5}$ concentration in one month prior to the day when respondent took the Internet survey	66.36 (16.56)
$PM_{2.5}$-three months average	Average $PM_{2.5}$ concentration in three months prior to the day when respondent took the Internet survey	95.81 (21.90)
$PM_{2.5}$-one year average	Average $PM_{2.5}$ concentration in one year prior to the day when respondent took the Internet survey	60.13 (11.69)
Demographic and personal indicators		

Income	Respondent's household gross income, which is the mid-point of self-reported income range (i.e. if the respondent selected 108,000-119,999 CNY, then that would be 114,000 CNY)	108,561 (78,559)
Poor subjective health	Dummy: 1 if the respondent chose his/her health state as "poor" or "very poor", 0 otherwise	0.047 (0.21)
Age	The age of respondent	42.17 (12.67)
Age-squared	Square of centered age	160.36 (150.74)
Female	Dummy: 1 if the respondent is female, 0 otherwise	0.45 (0.50)
College graduates	Dummy: 1 if the respondent has graduated from university or college, 0 otherwise	0.58 (0.49)
Marital status		
Single	Dummy: 1 if the respondent has never married, 0 otherwise	0.20 (0.40)
Married (reference)	Dummy: 1 if the respondent has been married, 0 otherwise	0.77 (0.42)
Divorced or widowed	Dummy: 1 if the respondent has been divorced or widowed	0.28 (0.17)
Unemployed	Dummy: 1 if the respondent has no occupation	0.05 (0.20)
Household with children	Dummy: 1 if the respondent has one or more children, 0 otherwise	0.74 (0.44)
Household with young children	Dummy: 1 if the respondent has one or more children under 6 years old, 0 otherwise	0.19 (0.39)
Commuting		
By cars	Dummy: 1 if the respondent usually go to work or school by cars, 0 otherwise	0.39 (0.49)
By public transmit	Dummy: 1 if the respondent usually go to work or school by bus or subway, 0 otherwise	0.44 (0.49)
By motorcycle, bicycleor walking (reference)	Dummy: 1 if the respondent usually go to work or school by motorcycle, bicycle or walking, 0 otherwise	0.33 (0.47)
Personality		
Passive	This study presents ten typical personalities in the survey and asked respondents to rate the consistency with their personalities. This study then aggregates the ten indicators into three factors via factor analysis.	0(0.78)

Stable	0 (0.77)	
Outgoing	0 (0.61)	
Environmental activities	Dummy: 1 if the respondent has experience of participating in environmental activities in last year, 0 otherwise	0.94 (0.24)
Spending on environmental activities	The average share of money spending on environmental activities out of monthly household income, which is the mid-point of reported value range	6.43 (7.75)
Regional controls		
City size		
Large (reference)	Dummy: population above 700,000	0.53 (0.49)
Medium	Dummy: population between 300,000 and 700,000	0.25 (0.43)
Small	Dummy: population below 300,000	0.21 (0.40)
Urban green coverage	The share of the of various urban public green lands and suburban scenic areas out of total area	39.73 (6.12)

This study also uses the respondents' participation in environmental activities and their spending on environmental activities as a proxy for people's environmental awareness. Liao et al. (2015) used participation in community environmental protection as a proxy to analyze the relationship between environmental awareness and SWB. Furthermore, to prevent omitted variable bias, this study controls for two possible geographical variations that may affect overall life satisfaction. To capture the possible impact of urbanization on life satisfaction, this study controls for the city size by classifying cities into "large", "medium" or "small" based on their population sizes (Jiang et al., 2012). Smyth et al. (2011) found that parkland areas are positively correlated with personal well-being. Hence, this study controls for the urban green coverage, which is the share of the various urban public green land spaces and suburban scenic areas within the total area. Both measures are obtained from the 2015 Provincial Statistical Yearbook of Northeast China.

6.3 Methodology

This study uses the following estimation equation to evaluate the impact of $PM_{2.5}$ concentration on the life satisfaction of the respondents.

$$LS_i = \beta_1 PM_{2.5,i} + \beta_2 log(Income) + \beta_3 Z_i + \beta_4 C_i + \epsilon \tag{6.1}$$

LS is the self-reported life satisfaction. $PM_{2.5\,i}$ is an air pollution measure, and three time-varying concentration measures are the main explanatory variables in the analysis. $log(Income)$ is the logarithm of the self-reported household income, and Z is a vector of the demographic and personal variables that are described in the previous section. C is a vector of the city-specific indicators including city size and urban green coverage. i refers to the respondents and is a unit of analysis. β is the regression coefficient of the variables.

This study further expanded the equation by introducing interaction variables between the $PM_{2.5}$ concentration measures and health (poor subjective health, household with children and household with young children) and environmental awareness (environmental activities and spending on environmental activities).

$$LS_i = \beta_1 PM_{2.5,i} + \beta_2 log(Income) + \beta_3 Z_i + \beta_4 C_i + \beta_I I_i + \epsilon \tag{6.2}$$

Ii refers to the interaction variables and β_I refers to the regression coefficient of the interacted variables. The interacted variables are centered to eliminate collinearity. Correlated independent variables cause problems in estimating regression coefficients. This study checks the variable inflation factor (VIF) to determine whether multi-collinearity will be a problem in the estimation. Additionally, to eliminate the collinearity between these age and age-squared variables, age is centered and then squared.

In previous studies, ordinary least squares (OLS) and ordered probit models have been

used as common empirical methods (Ferrer-i Carbonell and Gowdy, 2007; Luechinger, 2009). Although the use of an ordered probit model seems to be more appropriate because the SWB measures are ordinal, Ferrer-i Carbonell and Frijters (2004) found that both approaches could provide robust and similar results. This claim is supported by the results of Van Den Berg and Ferrer-I-Carbonell (2007); MacKerron and Mourato (2009); Levinson (2012) and others. Therefore, this study uses OLS as the main regression algorithm for the interpretation.

In addition, this study estimates the implicit MV with estimated coefficients, which indicates the trade-off between household income and $PM_{2.5}$ concentrations while holding people's life satisfaction constant. First, this study calculates the MV based on Eq. 6.2, without considering the interaction variables.

$$MV = -income_i \frac{\beta_1}{\beta_2} \tag{6.3}$$

Then, this study adds each interaction variable to Eq. 6.3 if the estimated regression coefficient is statistically significant.

$$MV = -income_i \frac{\beta_1 + \beta_I I_i}{\beta_2} \tag{6.4}$$

6.4 Empirical results and discussion

Table 6.3 shows the results of Eq. 6.2, and Table 4 provides the MV that is calculated based on the estimation results in Table 6.3. Table 6.4 provides the results of the OLS estimations of the baseline model in Eq. 6.1 and the ordered probit estimation of Eq. 6.2. It also includes the results of the measurement with the 5-point happiness scale to show the robust effect of $PM_{2.5}$ concentration levels on subjective well-being.

6.4.1 The impact of $PM_{2.5}$ on life satisfaction

As shown in Table 6.3, all three $PM_{2.5}$ measures have statistically significant negative impacts on subjective life satisfaction. This finding is consistent with the results of previous studies that find that air pollution negatively impacts SWB in China (e.g., Smyth et al., 2008; 2011). Given that the magnitudes of the $PM_{2.5}$ measures are different, this study uses the Wald test to determine whether there are statistically significant differences between the estimated coefficients of the $PM_{2.5}$ measures. This study finds statistically significant differences between the two measures with shorter time specifications. Thus, the negative effect of the annual average $PM_{2.5}$ concentration on life satisfaction is stronger than the negative effects of the one-month and three-month averages. The order of the impact does not correspond to the order of actual severity; three-month averages are significantly higher than one-month or annual averages. Thus, it may be reasonable to assume that the severity of air pollution does not directly affect people's recognition of the impact of air pollution on their everyday life, at least where the level of air pollution is relatively high year round, such as in Northeast China.

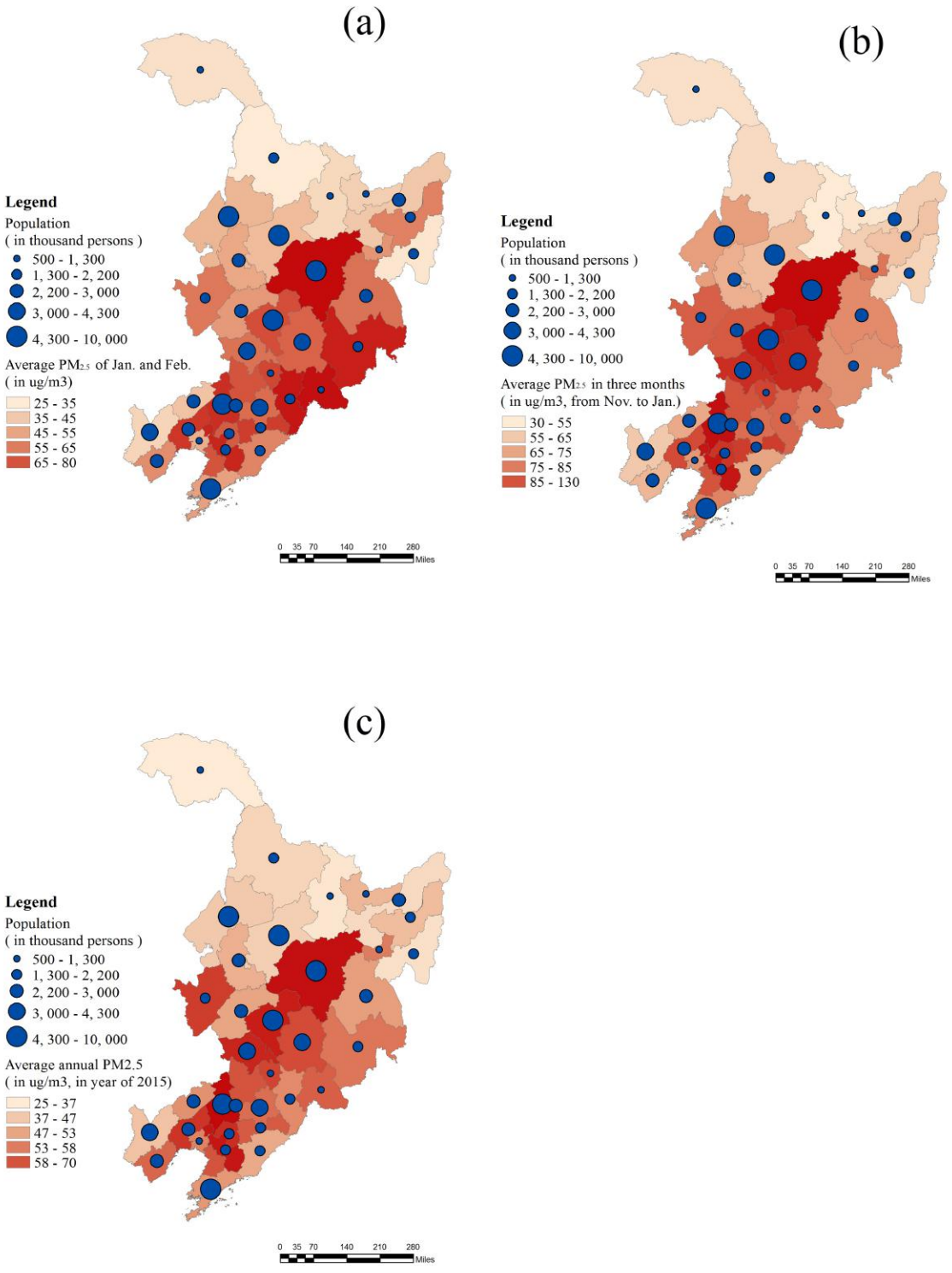


Figure 6.2: Spatial distribution of $PM_{2.5}$. average: (a) $PM_{2.5}$-one-month, (b) $PM_{2.5}$-three-months, (c) $PM_{2.5}$-annual

Table 6.3: The main models with using pollutant's concentration of "today" as air pollution indicators and the LS as dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$PM_{2.5}$-one month average		-0.00307*			-0.00339**		
		(0.00157)			(0.00158)		
$PM_{2.5}$-three months average			-0.00397***			-0.00467***	
			(0.00132)			(0.00134)	
$PM_{2.5}$-annual average				-0.00671***			-0.00750***
				(0.00242)			(0.00243)
Log (income)	0.0594**	0.0586**	0.0567**	0.0573**	0.0633**	0.0561**	0.0540*
	(0.0282)	(0.0283)	(0.0283)	(0.0283)	(0.0285)	(0.0282)	(0.0283)
Poor subjective health	-1.069***	-1.068***	-1.078***	-1.076***	-1.092***	-1.116***	-1.106***
	(0.115)	(0.115)	(0.114)	(0.114)	(0.115)	(0.116)	(0.115)
Age	0.0118***	0.0106***	0.0109***	0.0110***	0.00957***	0.0104***	0.0107***
	(0.00309)	(0.00311)	(0.00308)	(0.00308)	(0.00315)	(0.00307)	(0.00308)
Age-squared	0.000554***	0.000527***	0.000517***	0.000534***	0.000571***	0.000536***	0.000552***
	(0.000190)	(0.000190)	(0.000189)	(0.000189)	(0.000191)	(0.000189)	(0.000189)
Female	0.0776	0.0798	0.0834*	0.0835*	0.0777	0.0868*	0.0931*
	(0.0482)	(0.0485)	(0.0484)	(0.0484)	(0.0485)	(0.0481)	(0.0482)
College graduates	0.177***	0.175***	0.166***	0.162***	0.173***	0.164***	0.154***
	(0.0509)	(0.0508)	(0.0508)	(0.0510)	(0.0507)	(0.0505)	(0.0509)
Single	-0.143	-0.147	-0.144	-0.151	-0.160	-0.144	-0.150
	(0.122)	(0.122)	(0.122)	(0.122)	(0.122)	(0.121)	(0.121)
Separated or widowed	0.0773	0.0760	0.0771	0.0717	0.0849	0.0799	0.0726
	(0.152)	(0.151)	(0.151)	(0.151)	(0.151)	(0.150)	(0.150)
Unemployed	-0.402***	-0.404***	-0.384***	-0.386***	-0.388***	-0.379***	-0.379***
	(0.122)	(0.121)	(0.121)	(0.121)	(0.121)	(0.121)	(0.121)
Household with children	-0.125	-0.119	-0.122	-0.120	-0.0932	-0.112	-0.107
	(0.114)	(0.113)	(0.113)	(0.113)	(0.114)	(0.112)	(0.113)

Household with young	0.134*	0.116	0.119	0.119	0.109	0.123*	0.128*
Children	(0.0735)	(0.0736)	(0.0733)	(0.0733)	(0.0737)	(0.0728)	(0.0729)
Commuting by car	0.108*	0.104*	0.119**	0.121**	0.102*	0.128**	0.137**
	(0.0553)	(0.0555)	(0.0557)	(0.0559)	(0.0557)	(0.0557)	(0.0561)
Commuting by public	-0.0850*	-0.0764	-0.0724	-0.0707	-0.0714	-0.0631	-0.0622
transportation	(0.0502)	(0.0502)	(0.0501)	(0.0502)	(0.0502)	(0.0499)	(0.0500)
Personality- passive	-0.0246	-0.0276	-0.0327	-0.0311	-0.0204	-0.0203	-0.0155
	(0.0343)	(0.0343)	(0.0343)	(0.0343)	(0.0343)	(0.0342)	(0.0342)
Personality- stable	0.0219	0.0225	0.0193	0.0212	0.0160	0.0169	0.0259
	(0.0315)	(0.0314)	(0.0314)	(0.0314)	(0.0315)	(0.0314)	(0.0314)
Personality- outgoing	0.199***	0.200***	0.198***	0.200***	0.193***	0.196***	0.204***
	(0.0408)	(0.0407)	(0.0406)	(0.0406)	(0.0406)	(0.0403)	(0.0404)
Environmental activities	0.356***	0.361***	0.361***	0.366***	0.335***	0.354***	0.330***
	(0.105)	(0.105)	(0.105)	(0.105)	(0.107)	(0.105)	(0.107)
Spending on	0.00785**	0.00819**	0.00819**	0.00843**	0.00822**	0.00891***	0.00993***
environmental activities	(0.00328)	(0.00329)	(0.00328)	(0.00328)	(0.00328)	(0.00326)	(0.00328)
City size-small		-0.0870	-0.151	-0.122	-0.0725	-0.120	-0.0903
		(0.0892)	(0.0933)	(0.0907)	(0.0891)	(0.0929)	(0.0904)
City size-medium		0.0464	-0.0160	-0.00500	0.0551	-0.00124	0.00735
		(0.0617)	(0.0673)	(0.0668)	(0.0617)	(0.0671)	(0.0667)
Urban green coverage		0.00536	0.00497	0.00553	0.00476	0.00411	0.00408
		(0.00397)	(0.00396)	(0.00396)	(0.00398)	(0.00395)	(0.00396)
Interactions[a]							
With poor subjective					-0.0173**	-0.0128**	-0.0188**
health					(0.00721)	(0.00552)	(0.00926)
With household with					0.000292	0.00156	0.00687
children					(0.00326)	(0.00241)	(0.00471)
With household with					-0.00870**	-0.00948***	-0.0118**
young children					(0.00412)	(0.00332)	(0.00590)
With environmental					-0.00396	-0.000870	-0.00579

activities					(0.00559)	(0.00374)	(0.00726)
With spending on					-6.23E-05	-0.000412***	-0.000870***
environmental activities					(0.000197)	(0.000155)	(0.000292)
Constant	2.229***	2.271***	2.491***	2.470***	2.351***	2.430***	2.440***
	(0.351)	(0.398)	(0.406)	(0.409)	(0.354)	(0.350)	(0.350)
R-squared	0.248	0.255	0.259	0.258	0.264	0.275	0.273

Notes:

1. N=1002. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1.

2. Column (1) is baseline model only with survey variables. Column (2) $\sim$ (4) are models without interactions. Column (5) $\sim$ (7) are the full model with interactions.

[a] Interaction terms between $PM_{2.5}$ indicators and other control variables.

Table 6.4: The main models with using pollutant's concentration of "today" as air pollution indicators and the LS as dependent variables

	(1)	(2)	(3)	(4)	(5)	(6)
	Happiness regression (OLS)			LS regression (ordered probit)		
$PM_{2.5}$-one month average	-0.00564**			-0.00479***		
	(0.00236)			(0.00177)		
$PM_{2.5}$-three months average		-0.00757***			-0.00379**	
		(0.00203)			(0.00151)	
$PM_{2.5}$-annual average			-0.0125***			-0.00796***
			(0.00368)			(0.00274)
Log (income)	0.0867**	0.0745*	0.0708*	-0.0172	-0.0215	-0.0241
	(0.0422)	(0.0421)	(0.0421)	(0.0320)	(0.0319)	(0.0319)
Poor subjective health	-1.331***	-1.372***	-1.353***	0.0211	0.0198	0.0322
	(0.172)	(0.176)	(0.174)	(0.129)	(0.131)	(0.130)
Age	0.0170***	0.0188***	0.0190***	0.00843**	0.00976***	0.00981***

	(0.00476)	(0.00471)	(0.00470)	(0.00353)	(0.00347)	(0.00347)
Age-squared	0.000881***	0.000835***	0.000855***	-5.14E-05	-7.18E-05	-7.31E-05
	(0.000286)	(0.000285)	(0.000284)	(0.000214)	(0.000213)	(0.000213)
Female	0.114	0.128*	0.139*	-0.0183	-0.0148	-0.00989
	(0.0725)	(0.0726)	(0.0726)	(0.0543)	(0.0544)	(0.0543)
College graduates	0.276***	0.263***	0.243***	0.0717	0.0656	0.0590
	(0.0757)	(0.0760)	(0.0763)	(0.0569)	(0.0571)	(0.0573)
Single	-0.215	-0.189	-0.197	0.0802	0.0876	0.0837
	(0.181)	(0.181)	(0.181)	(0.137)	(0.137)	(0.136)
Separated or widowed	0.152	0.144	0.129	0.143	0.142	0.145
	(0.229)	(0.228)	(0.228)	(0.170)	(0.170)	(0.169)
Unemployed	-0.526***	-0.522***	-0.522***	0.00720	0.00601	0.00630
	(0.177)	(0.178)	(0.179)	(0.136)	(0.136)	(0.137)
Household with children	-0.213	-0.248	-0.239	-0.197	-0.216*	-0.208
	(0.170)	(0.170)	(0.170)	(0.128)	(0.127)	(0.127)
Household with young	0.226**	0.253**	0.257**	0.109	0.126	0.126
Children	(0.112)	(0.112)	(0.112)	(0.0826)	(0.0823)	(0.0821)
Commuting by car	0.164**	0.209**	0.222***	-0.0225	-0.0118	5.07E-05
	(0.0836)	(0.0845)	(0.0849)	(0.0625)	(0.0630)	(0.0632)
Commuting by public	-0.105	-0.0927	-0.0917	0.0797	0.0813	0.0862
transportation	(0.0750)	(0.0753)	(0.0754)	(0.0562)	(0.0564)	(0.0564)
Personality- passive	-0.0416	-0.0434	-0.0359	0.0350	0.0320	0.0328
	(0.0513)	(0.0515)	(0.0515)	(0.0385)	(0.0386)	(0.0386)
Personality- stable	0.0287	0.0292	0.0429	0.0299	0.0303	0.0337
	(0.0466)	(0.0468)	(0.0468)	(0.0353)	(0.0355)	(0.0354)
Personality- outgoing	0.281***	0.288***	0.299***	-0.0478	-0.0444	-0.0405
	(0.0607)	(0.0608)	(0.0608)	(0.0456)	(0.0455)	(0.0455)
Environmental activities	0.497***	0.536***	0.498***	-0.108	-0.0971	-0.105
	(0.158)	(0.156)	(0.158)	(0.120)	(0.119)	(0.120)
Spending on environmental	0.0134***	0.0150***	0.0167***	0.00194	0.00193	0.00254

activities	(0.00491)	(0.00494)	(0.00497)	(0.00368)	(0.00369)	(0.00370)
City size-small	-0.100	-0.177	-0.134	-0.0161	-0.0436	-0.0331
	(0.132)	(0.140)	(0.135)	(0.0999)	(0.105)	(0.102)
City size-medium	0.103	0.00905	0.0177	-0.0548	-0.0880	-0.0979
	(0.0929)	(0.102)	(0.101)	(0.0692)	(0.0758)	(0.0751)
Urban green coverage	0.00753	0.00669	0.00685	0.00216	0.00187	0.00200
	(0.00592)	(0.00594)	(0.00593)	(0.00446)	(0.00446)	(0.00446)
Interaction with poor	-0.0184*	-0.0133	-0.0185	-0.00556	-0.00392	-0.00273
health	(0.0107)	(0.00823)	(0.0136)	(0.00807)	(0.00623)	(0.0104)
Interaction with household	-0.00160	0.000503	0.00647	0.00379	0.00440	0.0105**
with children	(0.00483)	(0.00361)	(0.00703)	(0.00365)	(0.00273)	(0.00530)
Interaction with household	-0.0144**	-0.0150***	-0.0187**	-0.00615	-0.00402	-0.00703
with young children	(0.00631)	(0.00510)	(0.00900)	(0.00461)	(0.00374)	(0.00663)
Interaction with environmental	-0.00937	-0.00312	-0.0116	-0.00698	-0.00352	-0.00663
activities	(0.00818)	(0.00547)	(0.0106)	(0.00626)	(0.00422)	(0.00816)
Interaction with spending on	-6.17E-05	-0.000625***	-0.00138***	0.000126	2.62E-05	-0.000125
environmental activities	(0.000292)	(0.000233)	(0.000446)	(0.000221)	(0.000175)	(0.000329)
Constant				3.559***	3.570***	3.589***
				(0.397)	(0.396)	(0.395)
R-squared				0.033	0.031	0.034
Pseudo R-squared	0.1160	0.1223	0.1210			

Notes:

1. N=1002. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1.

2. Models in column (1) $\sim$ (3) use happiness measures as dependent variable and estimated by OLS, while models in column (4) $\sim$ (6) use life satisfaction (LS) as dependent variable and estimated by ordered probit.

The coefficient of each $PM_{2.5}$ measure indicates the effect of a 1 $\mu g/m^3$ change in $PM_{2.5}$ concentration on life satisfaction. Given that mean annual average concentration for the respondents at the time of survey was approximately 60 $\mu g/m^3$, a complete reduction to 0 $\mu g/m^3$ improves LS by 0.45 points when this study uses the coefficient of annual average (-0.0075). If this study uses the mean value of the three-month average, 96 $\mu g/m^3$, the complete reduction improves LS by 0.72 points. The upper limit suggested by official guideline of the World Health Organization (WHO) is 10 $\mu g/m^3$ for an annual average $PM_{2.5}$ concentration. Reducing the current annual averages and average levels during winter season to 10 $\mu g/m^3$ would improve LS by 0.38 and 0.64 points, respectively. Because this study uses a 5-point-scale life satisfaction measure, these numbers suggest a significant impact. In comparison to other events, the positive effect of finding employment on LS is approximately 0.38 points; this effect has almost the same magnitude as the effect on LS of reducing pollution to meet the annual average guidelines of the WHO (10 $\mu g/m^3$ for average annual $PM_{2.5}$ concentration).

Table 6.4 shows the results of alternative specifications and regression methods. Regardless of whether this study uses OLS or ordered probit as the estimation method, and regardless of whether this study uses life satisfaction or happiness as the dependent variables, this study finds robust negative impact of $PM_{2.5}$ on subjective well-being.

6.4.2 The effect of subjective health, children and the interaction effects with $PM_{2.5}$ measures

According to the results in Table 6.3, the interaction terms of the $PM_{2.5}$ measures and poor subjective health are negative. Because $PM_{2.5}$ concentrations and poor subjective health have a negative impact on life satisfaction, the negative coefficient of the interaction variable indicates that poor subjective health evaluation enhances the negative effect of $PM_{2.5}$ concentration, and vice versa. This result is different from the results in the U.S. provided by Levinson (2012), who showed that the effect of air pollution on SWB did not

vary based on self-reported health conditions. The significant interaction effect between health evaluation and pollution measures is partially explained by the rising awareness of the population regarding the health impact of air pollution (Chan et al., 2015). A possible interpretation of the significant negative interaction effect between pollution levels and subjective health evaluation is that people who perceive their health condition as poor are more concerned about further health deterioration due to the potential harm caused by air pollution than people who perceive a more satisfactory health condition.

In addition, this study finds a statistically significant negative impact on the interaction variable between $PM_{2.5}$ measures and the dummy variable household with young children. According to an interview conducted by Bickerstaff and Walker (2001) in Birmingham, U.K., people are more sensitive to the negative impact of air pollution on other people's health rather than on their own health, particularly the health of their family members. Our findings provide empirical evidence to support what Bickerstaff and Walker (2001) found with the Chinese data. Moreover, parents seem to generally think that young children are more susceptible than adults to the negative impact of air pollution. This finding explains why the interaction between $PM_{2.5}$ and children without an age restriction does not have a statistically significant negative impact.

In terms of magnitude of these interaction effects, given the current average annual concentration of 60 $\mu g/m^3$, the average negative enhancement effect of $PM_{2.5}$ is -0.7 points for people with young children and -1.1 points for people with poor subjective health evaluation. The LS of people with a poor subjective health evaluation are on average -1.1 point lower than those with a good subjective health evaluation, and the current pollution level doubles the negative impact of health concerns on LS. Since this study does not find a very robust effect of having young children, the negative effect of the interaction variable suggests that severe pollution may create an additional burden for parents.

Table 6.5 shows the results of the MV calculations using Eqs. 6.3 and 6.4. This study

calculates the average MV for groups with a varying combination of subjective health evaluation and whether they have young children under the age of 6. The USD/CNY values in Table 6.5 indicate average payment that respondents in particular groups are willing to pay for a 1-$\mu g/m^3$ reduction in the average $PM_{2.5}$ concentration. Almost three quarters of respondents fall into category of having reportedly good health and not having young children in their household; on average, respondents in this category (baseline group) put approximately 2,300 USD (15,000CNY) for a 1 $\mu g/m^3$ reduction in the average annual $PM_{2.5}$ concentration. People reported their health to be poor and/or have young children have a higher MV for pollution reduction. The average MV of the group with poor subjective health is almost 3.5 times higher than the MV of the baseline group. Similarly, the MV of respondents with young children is more than double the MV of the baseline group.

Table 6.5: Monetary value (MV) of different groups

Respondents' characteristics	$PM_{2.5}$-one-month	$PM_{2.5}$-three-months	$PM_{2.5}$-one-year
Good health and without young children (N=747)	USD 894 (CNY5,813)	USD 1,390 (CNY9,037)	USD 2,319 (CNY15,077)
Poor health and without young children (N=43)	USD 5,242 (CNY34,083)	USD 5,021 (CNY32,637)	USD 7,859 (CNY51,089)
Good health and with young children (N=188)	USD 2,753 (CNY17,899)	USD 3,676 (CNY23,896)	USD 5,275 (CNY34,293)
Poor health and with young children (N=5)	USD 7,102 (CNY46,169)	USD 7,307 (CNY47,497)	USD 10,816 (CNY70,304)

Notes:
1. Groups are classified according to the corresponding dummy variables (poor health and household with young children).
2. MV in CNY is converted into USD using an exchange rate of 6.50.

While the relative difference in MV for pollution reduction is clear across groups, it is difficult to assess whether the monetary evaluation is particularly high or low given that the monetary evaluations do not specify the time frames in which payment would be issued. The average MV values are the absolute monetary value that people are willing to place for a 1-$\mu g/m^3$ reduction in average $PM_{2.5}$ concentration and does not indicate that

people are willing to pay that amount per year. Nevertheless, the amounts are significant and reflect people's concern for the deterioration of air quality. Moreover, the results indicate that parents of young children or people with ill health are more sensitive to the change in pollution concentrations and would benefit more from pollution reduction efforts.

6.4.3 Environmental awareness

Our results show that environmental activities and spending on environmental activities have a significant positive impact on people's overall life satisfaction. These results are consistent with previous findings on environmental awareness (e.g. Ferrer-i-Carbonell and Gowby, 2007; Sekulova et al., 2013). This study also finds that the interaction variables between environmental activities and $PM_{2.5}$ measures are not statistically significant, whereas the interaction between spending on environmental activities and $PM_{2.5}$ measures has a significantly negative impact on overall life satisfaction. This result indicates that people who spend more on environmental activities are more vulnerable to the negative impact of air pollution.

This study can interpret the negative interaction impact of environmental activities and $PM_{2.5}$ on life satisfaction from the perspective of "psychological anticipation". People's spending on environmental activities can be considered an investment in their environment, and people who invest more in a better environment may naturally have a higher expectation level for an improved environment. Therefore, these people are more likely to experience a negative sentiment when their living environment does not improve as expected. This study also denotes that this "psychological anticipation" may differ depending on the length of time because this study observes different results that depend on the $PM_{2.5}$ measures, which differ during the time period. Column (1) of Table 6.3 shows no significant effect, whereas the coefficients in Columns (2) and (3) have a statistically significant impact, and the estimated coefficients in Column (2) are smaller than

in Column (3). These results may indicate that people expect their investment to have a short-term impact, but as time passes, their expectation of a return grows, and they become increasingly unhappy when the expected return is not received.

Figure 6.3 shows the plot of marginal impact of environmental spending on overall life satisfaction. Spending on environmental activities has a positive impact when the $PM_{2.5}$ concentration is relatively low; however, once the level passes the cut-off concentration level of 21 $\mu g/m^3$ for the $PM_{2.5}$ three-month average and 11 $\mu g/m^3$ for $PM_{2.5}$ annual average, additional spending on environmental activities may actually reduce people's overall life satisfaction.

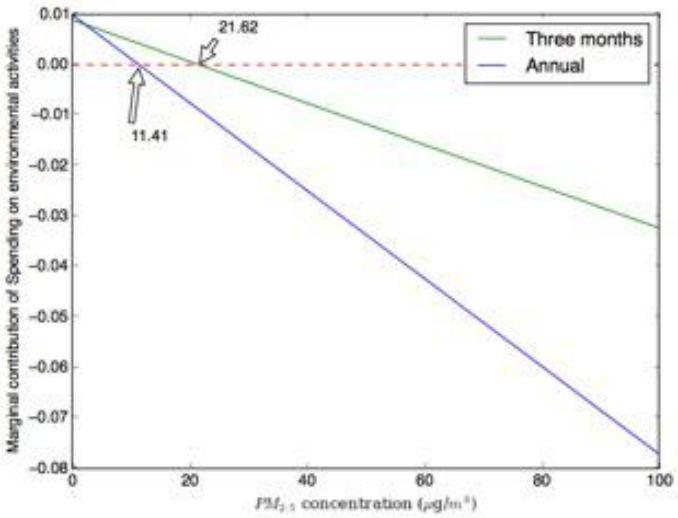


Figure 6.3: Marginal life satisfaction contribution of spending on environmental activities

6.4.4 Demographic variables and area characteristics

Most of the demographic variables that this study controls in the estimation model, including household income, subjective health condition, college graduates and unemployed, have a statistically significant impact. The respondents with a higher income and higher education are more satisfied with their life, and people in poor health and people who are unemployed have a lower life satisfaction level. Age and age-squared both

affect life satisfaction. The positive signs of age-squared indicate a U-shaped effect from age. Gender impacts life satisfaction in a certain model, but the effect is inconsistent; therefore, it is difficult to conclude a robust impact. This result only partially supports previous findings that women tend to have a higher SWB than men (Ferreira et al., 2013; Wang et al., 2015). Similarly, a previous study emphasized that family has a particular significance for Chinese society (Zhao, 2012), but this study does not find a consistent impact of other demographic variables for which previous studies have found a significant effect on life satisfaction, namely, marital status and whether the respondents have children. In contrast, this study found a positive effect for having children below the age of 6 years. Our results do not suggest that marriage or having children are unimportant factors for Chinese people in this target region, but it suggests that these variables do not explain the variation in the life satisfaction among people in the target region.

In addition, our results indicate that using a car as a commuting mode has a statistically significant positive impact on life satisfaction, while using public transportation for commuting does not have this impact. This finding may be explained by the fact that Chinese society values cars as a symbol of wealth and social status, particularly in less developed regions such as our target region in the northeastern provinces where car ownership is not as common as in richer major cities. Sekulova and van den Bergh (2013) also found a positive impact of car ownership on people's life satisfaction in Spain, possibly due to the relative freedom in mobility that cars can provide. Moreover if roads and public transportation are both congested, then it seems that people would rather choose road traffic than the congestion inside of trains and buses.

Finally, regional control variables do not have a statistically significant effect on life satisfaction as this study expected. This outcome may be explained by the cultural uniformity and industrial structure of Northeast China, which was previously described.

6.5 Summary

The issues of environmental degradation and air pollution are pressing concerns in China. This is particularly true in urban areas that are experiencing the exponential growth of energy consumption, accompanied by rapid urban expansion and overall economic growth during the past decade. According to a report from the Asian Development Bank, less than 1% of 500 Chinese cities meet the air quality guidelines (10 $\mu g/m^3$ for an average annual $PM_{2.5}$ concentration) recommended by the WHO. Along with corruption, income inequality and unemployment, the environmental problem has become a major cause of social unrest in China (Zhang, 2007). Exponential increases in the availability of information about the pollution level and its negative effects on people have raised the importance of how people feel about pollution's impact on their everyday life.

This study uses the life satisfaction approach to evaluate the effect of air pollution in the northeast region of China, which is one of the most severely polluted regions in China. This study uses self-reported SWB data from an original Internet survey conducted in 2016, which allows us to provide the latest empirical evidence on SWB in China. This study combines detailed $PM_{2.5}$ data at the city level and combine it with survey to analyze the impact of air quality on life satisfaction. Our results are consistent with the evidence from previous studies that air pollution has a significant negative impact on people's life satisfaction. The results also show that the effect of air pollution on subjective well-being is affected by individual characteristics. In particular, a poor subjective evaluation of health condition and having young children result in significant increase of participants' MV for air pollution reduction. The effect of air pollution levels also depends on whether people spend on environmental activities.

In recent years, the Chinese government has seriously addressed the issue of air pollution by introducing various measures to remedy the situation. Chinese policy makers have also acknowledged that the way people perceive environmental problems does matter. In an official government press release in 2014, Chinese president Jinping Xi stated,

"air quality has directly affected the Chinese people's happiness." Although our study is limited in coverage, there are areas that share similar characteristics with our target region. In China, there are industrial regions such as those in Middle Yangtze and Sichuan where mining and heavy manufacturing industries are concentrated. Moreover, outside of China, industrial regions in developing areas such as the Chota Nagpur Plateaus region in India, La Plata urban districts in Argentina, and Belo Horizonte in Brazil share similar industrial characteristics and environmental problems. Thus, the implications of our empirical results are somewhat indicative of what is happening in these areas.

According to our empirical results, the government may increase the overall welfare of its people by paying more attention to particular groups that are vulnerable to the negative effects of air pollution. In particular, for people who have a poor subjective health evaluation and/or live with young children under 6 years old, government agencies may take specific measures to relieve people's anxiety regarding the potential health risk of air pollution by providing outlets for their concerns and by communicating preventative measures that people can take to mitigate damages from air pollution.

Chapter 7

The impact of air pollution on the subjective well-being in Beijing and Shanghai

7.1 Motivation

This study evaluates the impact of one of the most prevalent and prioritized environmental problems, air pollution that is responsible for an estimated seven million deaths annually, or one in eight premature deaths every year. Previous studies have shown robust negative impact of air pollution on SWB. However, empirical studies that consider variability in air pollution affect both by pollutants and cities are rather limited. Despite the fact that air pollution levels are determined by the concentrations of a complex mixture of air pollutants, most empirical studies use a specific pollutant as a proxy for overall air pollution level (e.g. Levinson, 2012; Ferreira et al., 2013). Also, few studies have performed city-level comparative analyses. Hence, this study provides a case application that examine the impacts of four major pollutants on SWB, i.e., the Sulfur Dioxide

(SO_2), Nitrogen Dioxide (NO_2), coarse particles with a diameter between 2.5 and 10 μm (PM_{10}) and fine particles with a diameter of 2.5 μm or less ($PM_{2.5}$) in two major Chinese cities: Beijing and Shanghai.

Most related studies use aggregated air pollution datasets at regional or local levels, which are sufficient for yielding robust results. However, as Brereton et al. (2008) suggested, the explanatory power of the subjective well-being function can be increased if location-specific factors are taken into account. Moreover, previous studies have noted that a limited availability of pollution data at the local or regional level restricted their analyses to either the national level or a particular, localized area where richer data were available (Rehdanz and Maddison, 2005; Welsch, 2006). These remarks indicate the importance of local analyses with detailed pollution data.

This study uses the data from an original Internet social survey conducted in 2016 to analyze the impact of air pollution on the life satisfaction (LS) of Beijing and Shanghai residents. These data are matched with pollutant estimates for the locations of the respondents within each city. Given that air pollution is a localized phenomenon and that air pollution levels have spatial variation, particularly in big cities such as Beijing and Shanghai, this study uses geographic information system (GIS) interpolation techniques to estimate residents' exposure to each pollutant using air pollution data from monitoring sites in the cities. This study also uses different time specifications in order to further evaluate the variability in impacts of air pollution across the residents of Beijing and Shanghai. Pollution exposure on the day of the survey is used as the main indicator, and the impact of the changes in pollution exposure on the days leading up to the survey date is used to examine whether resident well-being is sensitive to the pollution level in the recent past.

7.2 Data and methods

7.2.1 Subjective well-being survey

7.2.1.1 Survey

In this study, this study uses an original Internet survey conducted during January and February 2016 in Beijing and Shanghai. The survey questionnaire was designed to collect self-reported overall life satisfaction levels as well as other personal and household characteristics. This study uses Internet surveys have an advantage in avoiding interviewer bias caused by arbitrary factors, such as the appearance or gender of interviewers, in responses to sensitive questions such as household income, employment and marriage status (Welsch, 2009). Nevertheless, denote that Internet surveys may also suffer from a possible selection bias; people with relatively high life satisfaction could be more willing to participate in the survey as people with relatively low life satisfaction may feel less comfortable with answering questions regarding their living condition and SWB. This study further discusses validity of the collected data by examining the distribution of dependent variable and the main individual and household characteristics below. Out of 1,022 (Beijing) and 957 (Shanghai) observations with an accurate zip code, which was necessary to merge the survey data with pollution data, this study excluded 64 and 76 responses that were collected during the Spring Festival holiday in Beijing and Shanghai, respectively. 958 observations were left for Beijing and 881 observations were left for Shanghai.

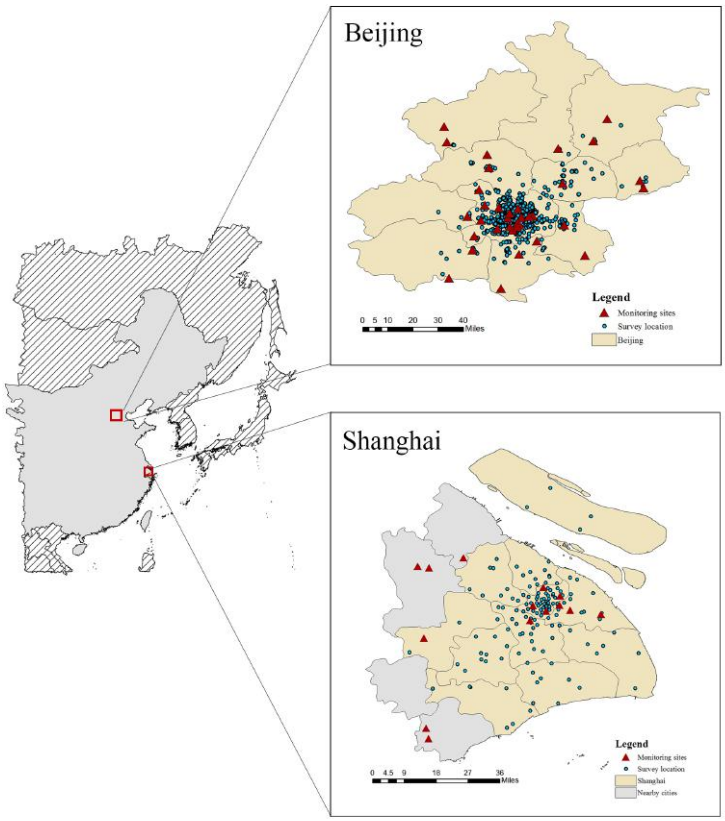


Figure 7.1: Location of respondents and monitoring stations in Beijing and Shanghai

Beijing and Shanghai cover 16,807 km^2 and 6,340 km^2 and have 21.7 million and 24.2 million inhabitants, respectively. According to the economical label and urban development plans, districts in both cities can be classified into three categories: central districts, rapidly developing districts and suburbs. Over a half of the inhabitants of both cities (59% for Beijing and 53% for Shanghai) live in the central and rapidly developing districts. Figure 7.1 shows the locations of the Beijing and Shanghai respondents, which are geo-coded with reported residential addresses and zip codes; this figure shows that 836 of 958 respondents and 743 of 881 respondents reside in the central and rapidly developing areas in Beijing and Shanghai, respectively. Hence, the results of this study heavily represent the urban population of Beijing and Shanghai.

Table 7.1: Characteristics of respondents with comparison

	Beijing		Shanghai	
	This survey	Comparative survey[a]	This survey	Comparative survey[a]
Age[b]				
20 ∼ 34	34%	36%	32%	24%
35 ∼ 54	58%	48%	54%	45%
55 ∼	8%	16%	14%	31%
Gender				
Male	53%	51%	49%	49%
Female	47%	49%	51%	51%
Household income (CNY)[c]				
Mean	174,106	142,211	161,501	169,620
0 ∼ 50,000	11%	7%	10%	12%
50,001 ∼ 100,000	10%	22%	9%	29%
100,001 ∼ 150,000	27%	42%	25%	22%
150,001 ∼ 200,000	24%	12%	16%	13%
200,001 ∼ 250,000	15%	10%	22%	5%

Notes:
[a] Age and gender distributions are obtained from the 2015 Statistical Yearbooks of Beijing and Shanghai. Gross household income distributions are derived from Chinese General Social Survey (CGSS, 2013) with a sample size of 560 for Beijing and 531 for Shanghai.
[b] Given that there is almost no respondent under 20 years old in this survey, for presenting a comparative result, population under 20 years old are excluded from the original statistics dataset when calculating the ratio.
[c] CGSS 2013 only has annual gross household income of 2011, but our survey collects gross household income of 2015. To deal with this temporal difference, this study collected yearly income increase rates between 2011 and 2015 from Statistical Yearbooks, and then estimate the household income of 2015 by applying each year's income increase rate to income of 2011.

Table 7.1 provides a comparison of basic demographic characteristics between our sample and the data from statistical yearbooks and the Chinese General Social Survey (CGSS 2013). This study compares ages and gender distributions with official statistics from the 2015 Statistical Yearbooks of Beijing and Shanghai. Our sample has relatively fewer respondents who are older than 55 years of age due to the difficulty of reaching this age group through an Internet survey.

Given the lack of official statistical data for gross annual household income, this study compares our income distribution with CGSS data that were used by similar previous

SWB studies (e.g. Wang et al., 2015; Qian and Qian, 2015). The average annual household income of respondents in Beijing is approximately 174,000 CNY (26,000 USD), which is higher than the average income shown in the CGSS in Beijing by approximately 31,000 CNY (5,000 USD). Our data include samples from the lower-income distribution; thus, the skewness is not extreme. Moreover, the average annual household income in Shanghai is approximately 161,000 CNY (25,000 USD), which is marginally lower than the average income in the CGSS.

7.2.1.2 Life satisfaction measures

This study uses an overall LS measure as a dependent variable in the analysis. Given that subjective well-being-related questions are particularly vulnerable to the context and timing in which they are asked, they were placed at the beginning of our questionnaire in order to avoid short-term bias from questions that were asked prior to the SWB-related questions. This study asked the following question: "Please imagine a ladder with steps numbered from 0 to 10 at the top. The top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?" The respondents chose a number from an 11-point scale from 0 (worst possible life) to 10 (best possible life). Figure 7.2 shows the distribution of self-reported LS rating, with a mean of 6.81 and a median of 7. The ratings are normally distributed and slightly skewed toward the right. The similar skewness of life satisfaction distribution (toward the right) has been frequently reported in related previous studies on various countries including the studies on China (e.g. MacKerron and Mourato, 2009; Gao et al., 2014; Proto and Rustichini, 2015; Wu and Zhu, 2016). Thus, the skewness toward the right seems to be the natural characteristics of LS scores rather than the result of selection bias caused by the collection method.

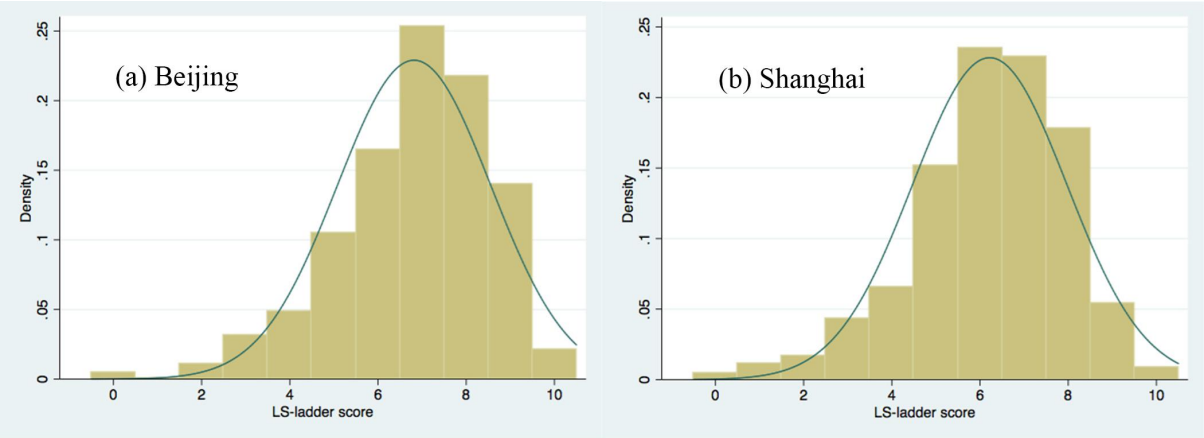


Figure 7.2: Distribution of self-reported life satisfaction rating

7.2.1.3 Household and personal characteristics

Using our survey, this study constructed various variables of household and personal characteristics. This study controls for determinants that have commonly been used in previous SWB studies: annual household income (household income), age (age and age-squared), gender (female), education degree (college graduate), marital status (single, married, divorced or widowed), household members (household with children and household with young children) and employment status (unemployed). In addition to controlling for household income, this study controls for two asset-related factors: car-commuting and residence type (rented apartment/privately owned apartment/rented single family house/privately owned single family house). Previous studies have shown that current mood and context may cause fluctuations in people's answers to subjective well-being from day to day. Therefore, this study controls for the self-reported health evaluation (poor subjective health) and a series of psychological factors, including the importance of being trusted by others (trust), mood (enjoy and sad) and personality (stable, passive and outgoing). The three personality indicators are derived from a factor analysis of the responses to a series of personality questions with three options, i.e., disagree, neither agree nor disagree, and agree. this study also controls for congestion time and neighborhood safety, both of which have been shown to have impacts on people's life satisfaction,

particularly in big cities (Bergstad et al., 2011; Olsson et al., 2013; Sekulova and van den Bergh, 2013).

7.2.1.4 Control of spatial and temporal variation

This study controls for several spatial and temporal variations in the data that are not captured by the household and personal variables. For spatial control, this study employs the distance of the respondent's home address to the nearest central business district (CBD) to capture unobserved characteristics that are likely to be associated with factors such as accessibility and residential size, which might affect overall SWB. Shanghai officially defines six CBDs, whereas Beijing has one large CBD. Thus, for Beijing, this study calculates the distance to areas within the city that have roles equivalent to those of CBDs, including administrative, technology and financial centers, and identify them as Beijing's CBDs in this analysis.

Given that our survey was conducted over a 40-day period, potential temporal variation should be controlled for in the analysis. This study investigated several potential variables, including date dummies, dummies for the periods pre- and post-spring festival, dummies for weekday and weekend, as well as weather and temperature. This study presents here the results of estimation model that include weather and temperature, which are reported to be important determinants of LS in previous studies (MacKerron and Mourato, 2009; Connolly, 2013); the remaining variables have either very high variance inflation factor (VIF) values or no statistically significant impact on SWB.

7.2.2 Air pollution exposure of residents

7.2.2.1 Air pollution data from monitoring sites

In this analysis, this study uses aggregated data of hourly monitored air pollution data that are collected at thirty-five stations in Beijing and nine stations in Shanghai as well as four stations in nearby cities (Figure 7.1). At each station, automatic monitoring systems measure the ambient concentrations of SO_2, NO_2, CO, O3, PM_{10} and $PM_{2.5}$ based on China's Environmental Protection Standards (HJ 193-2013 and HJ 655-2013). The real-time concentration values from monitoring stations are automatically transferred to the China National Environmental Monitoring Center and then published with validation from the Technical Guidelines on Environmental Monitoring Quality Management (HJ 630-2011).

7.2.2.2 Spatial interpolation

This study uses the ordinary Kriging (OK) method, a spatial interpolation technique, which is popularly used in the field of epidemiological and atmospheric studies (e.g. O'Leary and Lemke, 2014; Arifin et al., 2015), to match air pollution data with geo-coded survey data at the individual level. The SWB study by (Ferreira et al., 2013) used inverse distance weighting (IDW) as the interpolation method. IDW is generally regarded as a deterministic interpolation method, which makes predictions directly based on the distance to surrounding measured values; while OK method, as well as the other kriging methods, are regarded as a geostatistical method that further incorporate the statistical relationship among the measured points (Jerrett et al., 2005). OK method is more advanced than IDW given that OK method considers both the distance between a location of interest and the sample locations, and also the distance between sample locations. In particular, taking the latter distance effectively eliminates the deleterious effect of clustering in samples on the quality of the interpolated estimates. Li and Heap

(2011) provides further details regarding comparison of various interpolation methods, including the OK method and IDW.

The general form of kriging methods can be written as follows:

$$\hat{z}(s_0) = \sum_{i=1}^{n} w_i z(s_i) + \epsilon(s_0) \tag{7.1}$$

Z represents the pollutants. $\hat{z}(s_0)$ is the concentration at unobserved location s_0 and is estimated using a weighted average of known neighborhood samples $z(s_i)$. w_i is the optimal weight assigned to neighborhood $z(s_i)$. $\epsilon(s_0)$ is a random field with covariogram function $C(h)$ and variogram (semi-variogram) $\gamma(h)$. w_i should satisfy two objectives in the geo-statistical formulation: (1) unbiased, which indicates $\sum_{i=1}^{n} w_i = 1$, and (2) minimal variance of estimation, which is expressed as $Var(\epsilon(s_0) = 0$. The weights are determined by solving this minimization problem using variogram $\gamma(h)$, which is defined as

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} \{z_(s_i) - z(s_i + h)\} \tag{7.2}$$

where n is the number of pairs of sample points of observations, and h is the spatial lag distance.

Out of the various variants of Kriging methods, the OK method is a relatively simple variant. However, OK method is considered to be suitable for interpolating the air pollution data of this study. CoKriging is often used to take advantage of the covariance of two or more regionalized variables. However, in this study, the correlations between the pollutants are not strong (feature-space correlation coefficients are generally lower than 0.55) and/or are unstable across time; moreover, the other pollutants that could be used as co-variables do not have a more dense distribution. Thus, coKriging may not be able to enhance the predictive performance in the setting of this study.

Simple Kriging (SK) and universal Kriging (UK) are two commonly used variants other than OK. The three variants differ in their assumptions in terms of the trend

$\mu(z) = E[z(s)]$: OK assumes $\mu(z)$ to be unknown constant that should be estimated; SK assumes that is known, which can be directly derived from the data; UK assumes that $\mu(z)$ is a linear function of the spatial coordinates. This study uses OK rather than SK or UK because the assumption of SK seems to be too strict for our datesets, and the results of trend analyses do not show explicit and/or strong trend in the datasets. Furthermore, this study also conducted a preliminary examination on the predictive performances of the three Kriging methods by using 10 randomly chosen data sets of different survey dates, and found that OK models slightly outperform the other two models in terms of both accuracy and stability.

For each respondent, this study use the OK method to estimate the average concentration of each pollutant at the given residential address. This technique is implemented using the Python package (ArcPy) provided by ArcGIS 10.3 (ESRI Inc., U.S.). Note that while conducting OK interpolation, this study observed substantially high NO_2 concentrations at mobile monitoring sites compared with concentrations at stationary sites. This spatial heterogeneity, which is common in urban areas, may lead to an upward bias in exposure estimation, especially for residences near the mobile sites. Therefore, this study excludes the 5 mobile monitoring sites for the interpolation process of NO_2 in Beijing.

The published hourly data of four air pollutants (SO_2, NO_2, PM_{10} and $PM_{2.5}$) during January and February 2016 are aggregated, and the following three steps are used to generate the air pollution exposure. First, using the hourly monitoring air pollution data, this study calculates the daily average and moving weekly average concentration at each monitoring site for each day between Jan. 12 and Feb. 29 except the days of the spring festival period, providing 48 days for both pollution indicators. Second, this study applies the OK method to construct spatial distribution maps for each day and to estimate the air pollutant concentrations at geo-coded respondents' residential locations. Lastly, this study matches the estimated pollution and a respondent's survey data level using the recorded survey dates. In addition, the air pollution exposures of respondents

on the day before the survey date is estimated based on the spatial distribution maps interpolated using the daily pollution measures.

In particular, this study does not use a uniform variogram model for all OK models given that the spatial relationship between monitoring sites for air pollutants may change over time with the change of meteorological condition and emission sources. This study constructs OK models with different variogram models (including linear, Gaussian, exponential and spherical) for each day during survey period. Then this study fits the models with the monitored air pollution data, and select the model with the lowest RMSE (root mean square error) calculated from leave-one-out cross-validation. While the optimal OK models for a date have different variogram models, this study also finds that the selected OK models in nearby dates tend to be the same variogram model; this suggests the existence of some time trends. Figure 7.3 $\sim$ 7.6 present the examples of interpolation maps of daily and weekly average concentration of the four pollutants in Beijing and Shanghai.

7.2.2.3 Estimated air pollution exposure

Table 7.2 presents the distribution of the estimated air pollution exposure in Beijing and Shanghai during the analysis period. The average concentrations of SO_2, NO_2, PM_{10} and $PM_{2.5}$ are 24.2 $\mu g/m^3$, 54.6 $\mu g/m^3$, 63.6 $\mu g/m^3$, and 57.3 $\mu g/m^3$ in Beijing and 27.0 $\mu g/m^3$, 61.0 $\mu g/m^3$, 100.2 $\mu g/m^3$, and 75.8 $\mu g/m^3$ in Shanghai. Although the pollution problem in Beijing is more well known, the overall air quality in Shanghai was significantly worse than that of Beijing during the survey period. The standard deviations for the studied pollutants are similar in the two cities, although Beijing's pollution level range is generally wider.

To analyze the temporal and spatial variations in air pollution level in our sample, this study calculates the average concentration and standard deviation per day. Figure reffig, temporal presents the temporal and spatial variations. Spatial variation on a particular date is represented by the symbol and error bar, which denotes the average

Table 7.2: Summary statistics of interpolated air pollutants exposures for our survey

	Beijing (N: 953)				Shanghai (N: 881)			
	SO_2	NO_2	PM_{10}	$PM_{2.5}$	SO_2	NO_2	PM_{10}	$PM_{2.5}$
Mean	24.2	54.6	63.6	57.3	27	61	100.2	75.8
Std. Dev.	16.1	27.1	31.7	42.7	11.7	22.6	44.2	45.4
5%	5.6	18.1	24.2	10.2	10.9	23.5	41.2	24
25%	11.3	31.5	41.6	20.4	17.5	42.6	69.9	33.9
50%	19.8	54.6	53.3	38.2	22.3	64	87.5	53.6
75%	36.7	75.7	86.6	90.3	37.7	77.4	125.9	118.8
95%	52.4	96.6	122.8	138.7	47	89.3	184.3	150.7

concentration ± 2 standard deviation. Temporal variation is indicated by the time trend of concentration. There are considerable variations both temporally and spatially in our data; however, the magnitude of spatial variation across dates of the survey period vary because the variation on a given date is heavily dependent on the number of observations collected that day. Compared with Shanghai, Beijing appears to have larger spatial variations. This trend might be partly due to the greater number of monitoring stations in Beijing, which allows us to construct a more sophisticated spatial distribution map of air pollution via spatial interpolation analysis. Another possible explanation for this trend is that districts and areas within Beijing are defined with more distinct functionalities compared with the definitions of regions in Shanghai (Gaubatz, 1999). In particular, Beijing has more imbalanced spatial distribution of manufacturing industries, which might result in area-specific emission sources and pollution distributions.

7.2.3 Empirical model

This study uses OLS as the main regression algorithm for the interpretation. This study specifies the function of LS as follows:

$$LS = f\{income, environment, personalcharacteristics, areacharacteristics\} \quad (7.3)$$

Our dependent variable is self-reported LS rating. Income is the reported gross annual household income (the log of the value is used herein). Environment contains the estimated air pollutant concentrations of SO_2, NO_2, PM_{10} and $PM_{2.5}$ at various times as well as temporal changes estimated per respondent. Personal characteristics is a set of control variables including age, gender, subjective health condition, employment status, marriage status, household members, mood, personality and commuting-related factors. Spatial and temporal control variables include distance to the CBD from the residence, weather and temperature of survey data. Table 7.3 presents the descriptions and statistics of all the employed variables.

In addition, using the estimated coefficients, this study estimates the monetary value of air pollution, which indicates the trade-off between household income and pollutant concentrations when holding people's LS constant. This study calculates the monetary valuation (MV) of pollution based on the OLS estimation without taking interaction variables into account.

$$MV = -income_i \frac{\beta_{air}}{\beta_{income}} \tag{7.4}$$

β_{air} is the estimated coefficient of each pollutant, and β_{income} is the coefficient of income.

Table 7.3: Variable description

Variable	Description	Mean (Std. Dev.) Beijing	Shanghai
Subjective well-being			
Life satisfaction	"Please imagine ladder with steps numbered from 0-10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?" 0 (worst possible life) - 10 (best possible life) scale	6.82 (1.74)	6.22 (1.75)
Air pollutants[a]			
SO_2	Ambient concentration ($\mu g/m^3$), interpolated from monitoring concentration reading by using Ordinary Kriging spatial interpolation method	24.23 (16.15)	26.98 (11.71)

NO_2		54.62	61.01
		(27.77)	(22.60)
PM_{10}		63.60	100.27
		(31.72)	(44.22)
$PM_{2.5}$		57.38	75.81
		(42.71)	(45.36)
Demographic and personal indicators			
Income	Respondent's household gross income (thousand CNY), which is the mid-point of self-reported income range (i.e. if the respondent selected 108,000-119,999 CNY, then that would be 114,000 CNY)	174,106 (105,132)	161,501 (82,441)
Age	The reported age of respondent	40.63 (11.57)	40.81 (11.95)
Age-squared	The square of reported age	1,784.64 (941.03)	1808.51 (1,037.26)
Female	Dummy: 1 if the respondent is female, 0 otherwise	0.47 (0.49)	0.51 (0.50)
College graduate	Dummy: 1 if the respondent has graduated from university or college, 0 otherwise	0.88 (0.32)	0.86 (0.34)
Poor health	Dummy: 1 if the respondent chose his/her health state as "poor" or "very poor", 0 otherwise	0.05 (0.21)	0.06 (0.25)
Marital status			
Single	Dummy: 1 if the respondent has never married, 0 otherwise	0.14 (0.45)	0.19 (0.39)
Married	Dummy: 1 if the respondent has been married, 0 otherwise	0.83 (0.36)	0.78 (0.41)
Divorced or widowed	Dummy: 1 if the respondent has been divorced or widowed, 0 otherwise	0.01 (0.11)	0.03 (0.18)
Household with children	Dummy: 1 if the respondent has one or more children, 0 otherwise	0.77 (0.41)	0.79 (0.56)
Household with young children	Dummy: 1 if the respondent has one or more children under 6 years old, 0 otherwise	0.51 (0.50)	0.26 (0.43)
Unemployed	Dummy: 1if the respondent currently has no full-time job, 0 otherwise	0.02 (0.12)	0.02 (0.15)
Residential categories			
Rented apartment	Dummy: 1 if the respondent is living in a rented apartment, 0 otherwise	0.084 (0.24)	0.072 (0.26)
Owned apartment	Dummy: 1 if the respondent is living in a privately owned apartment, 0 otherwise	0.78 (0.41)	0.83 (0.37)

Rented house	Dummy: 1 if the respondent is living in a rented single family house, 0 otherwise	0.011 (0.11)	0.016 (0.12)
Owned house	Dummy: 1 if the respondent is living in a privately owned single family house, 0 otherwise	0.11 (0.32)	0.067 (0.25)
Neighborhood safety	Dummy: 1 if the respondent regard his residential neighborhood is somewhat or very safe	0.87 (0.32)	0.88 (0.32)
Trust	Dummy: 1 if respondent think to be able to trust people/organizations is very or somewhat important, 0 otherwise	0.94 (0.22)	0.91 (0.29)
Enjoy	"How often have you fell enjoyment today?" 1 (not at all) - 4 (often) scale	3.44 (0.69)	3.22 (0.76)
Sad	"How often have you feel sad today?" 1 (not at all) - 4 (often) scale	2.01 (0.82)	1.92 (0.83)
Personality			
Stable	Ten different personality descriptions were presented in the survey and the respondent was asked to rank whether these descriptions are applied to them. Then, factor analysis is used to find the most important factors from observed data, as a result, three personalities are extracted	-0.0025 (0.74)	0.06 (0.78)
Passive		-0.0058 (0.68)	0.16 (0.77)
Outgoing		0.0015 (0.64)	-0.03 (0.67)
Commuting means			
By cars	Dummy: 1 if the respondent usually go to work or school by cars, 0 otherwise	0.46 (0.49)	0.35 (0.48)
By public transmit	Dummy: 1 if the respondent usually go to work or school by bus or subway, 0 otherwise	0.53 (0.49)	0.55 (0.49)
By motorcycle, bicycle or walking	Dummy: 1 if the respondent usually got to work or school by motorcycle, bicycle or walking, 0 otherwise	0.67 (0.46)	0.31 (0.46)
Congestion time	The average congestion hours that the respondent experienced in one month, calculated by reported commuting hours with congestion minus that without congestion	0.41 (1.16)	0.34 (0.56)
Knowledge level of air pollution	Self-reported knowledge level of air pollution: "please select an option that appropriately describes your level of knowledge for air pollution" 1 (do not have any knowledge) - 5 (very knowledgeable)	4.48 (0.82)	3.94 (0.75)
Spatial and temporal controls			
Sunny	Dummy: 1 if the survey date was sunny, 0 otherwise	0.71 (0.45)	0.39 (0.49)
Temperature	The average temperature of the survey date (centigrade)	-3.34 (3.46)	3.15 (3.36)

Distance to CBD	The distance of respondent's home to nearest CBD (km)	7.24 (8.13)	7.87 (8.94)

Notes: [a] Air pollutants' concentrations with various time-specifications were estimated and matched with respondents' survey dates, including 1) "today" which denotes the concentration of the day when the respondent took the survey; 2) "yesterday" which denotes the concentration of one day before the survey date; 3) "week" which denotes the average concentration in one week before the survey date. In this study, the time-specification of "today" is used as the main pollution indicator, its average and standard deviation over all respondent are shown in the right columns.

7.3 Empirical results and discussion

7.3.1 The effect of air pollution on life satisfaction

This study uses the estimated average air pollution exposure for every respondent at various time specifications, including the day of the survey, the day before the survey, and a one-week moving average prior to the survey date, to examine the impact of air pollution on LS. This study uses the average air pollution exposure on the survey date as the primary pollution indicator. Table 7.4 and Table 7.5 show the baseline and full regression results for the main indicator. The data indicates that there are both spatial temporal variation of pollution. This study does not consider the spatial and temporal variation of air pollution separately; however, this does not undermine the purpose of the analysis to measure the impact of variation in pollution on variation in individual well-being.

The results of the baseline models in Table 7.4 indicate that all four pollutants have a significant negative impact on LS in Beijing, although only NO_2 has a significant negative impact on LS in Shanghai. There may be a concern regarding low R-squared of the model in which his study only includes pollutants as independent variables. Denote that this observation should not be a critical concern in our analysis and discussion given the following two points: 1) R-squared does not determine whether the coefficient estimates

Table 7.4: Baseline models with using pollutant's concentration of "today" as the only independent variable and the LS as dependent variable

Beijing				
	1	2	3	4
SO_2	-0.0107***			
	(0.00346)			
NO_2		0.00675***		
		(0.00205)		
PM_{10}			-0.00437**	
			(0.00176)	
$PM_{2.5}$				-0.00339***
				(0.00131)
Constant	7.089***	7.199***	7.108***	7.024***
	(0.101)	(0.125)	(0.125)	(0.0933)
Observations	958	958	958	958
R-squared	0.01	0.011	0.006	0.007
	Shanghai			
	5	6	7	8
SO_2	-0.000576			
	(0.00504)			
NO_2		-0.00781***		
		(0.0026)		
PM_{10}			0.00086	
			(0.00133)	
$PM_{2.5}$				0.000541
				(0.0013)
Constant	6.236***	6.697***	6.134***	6.179***
	(0.148)	(0.169)	(0.146)	(0.115)
Observations	881	881	881	881
R-squared	0	0.01	0	0

Notes: Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1.

are bias, 2) the interpretations of the statistically significant variables are unchanged by regardless of R-squared; the significant coefficients still represent the mean change in the response for one unit of change in the predictor while holding all the other independent variables in the model constant. This study uses statistical significance and the coefficient to make comparative discussion between the two cities.

Table **??** presents the results of the full models with control variables. In Beijing, the regression coefficients of all four pollutants remain negative and statistically significant.

In Shanghai, PM_{10} and $PM_{2.5}$ remain insignificant, although SO_2 has a statistically significant impact at the 5% level in the full model. These results are consistent with the results of multi-country analyses in Welsch (2006); namely, NO_2 has a significant impact on LS. Nevertheless, in terms of the magnitude of the impacts, i.e., the coefficients, the absolute magnitude of the impact is smaller in Shanghai than in Beijing for all of the pollutants except SO_2. The results generally indicate relatively stronger awareness of air pollution by Beijing residents compared to Shanghai residents.

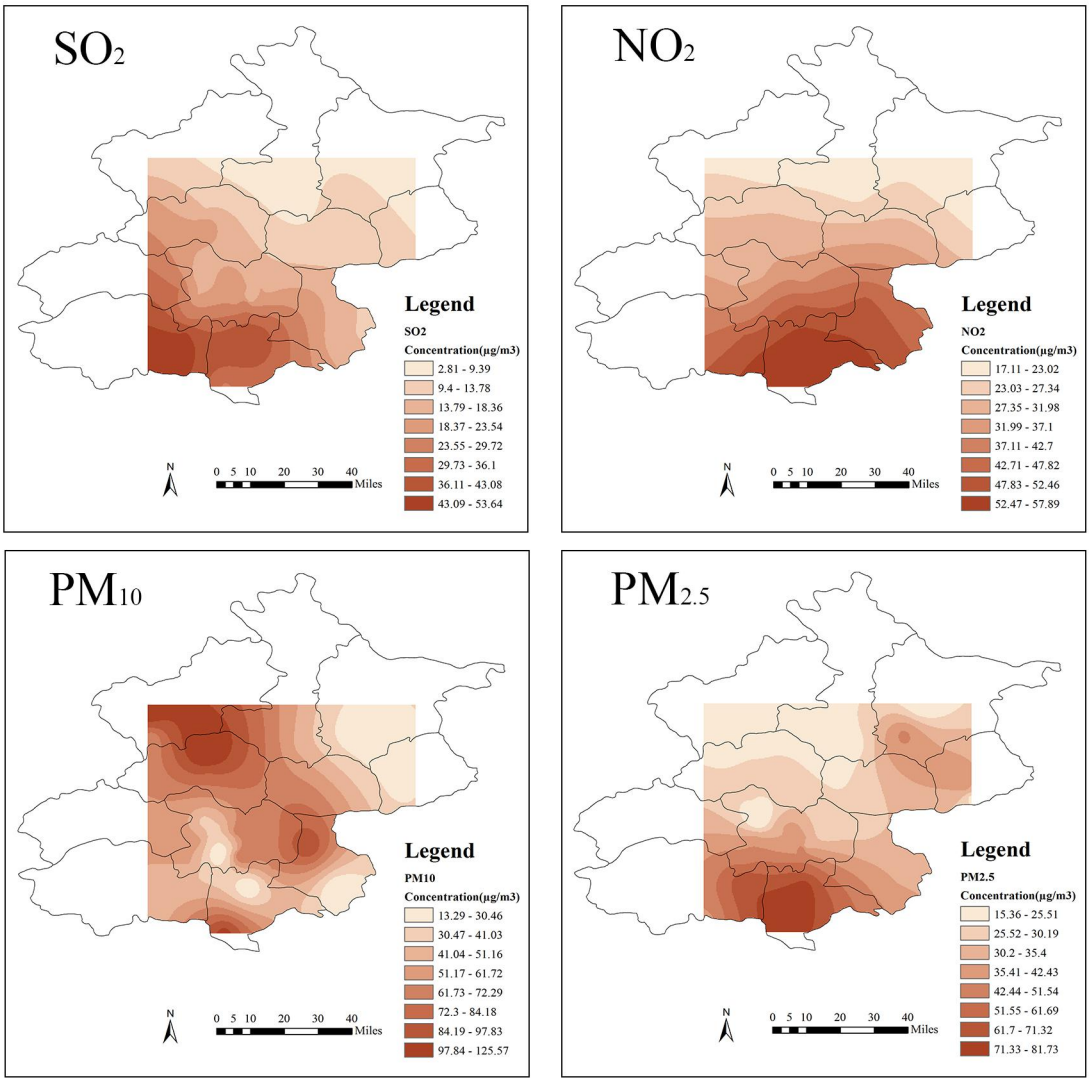


Figure 7.3: Interpolation maps of daily average concentration of four pollutants in Beijing for Jan. 26, 2016

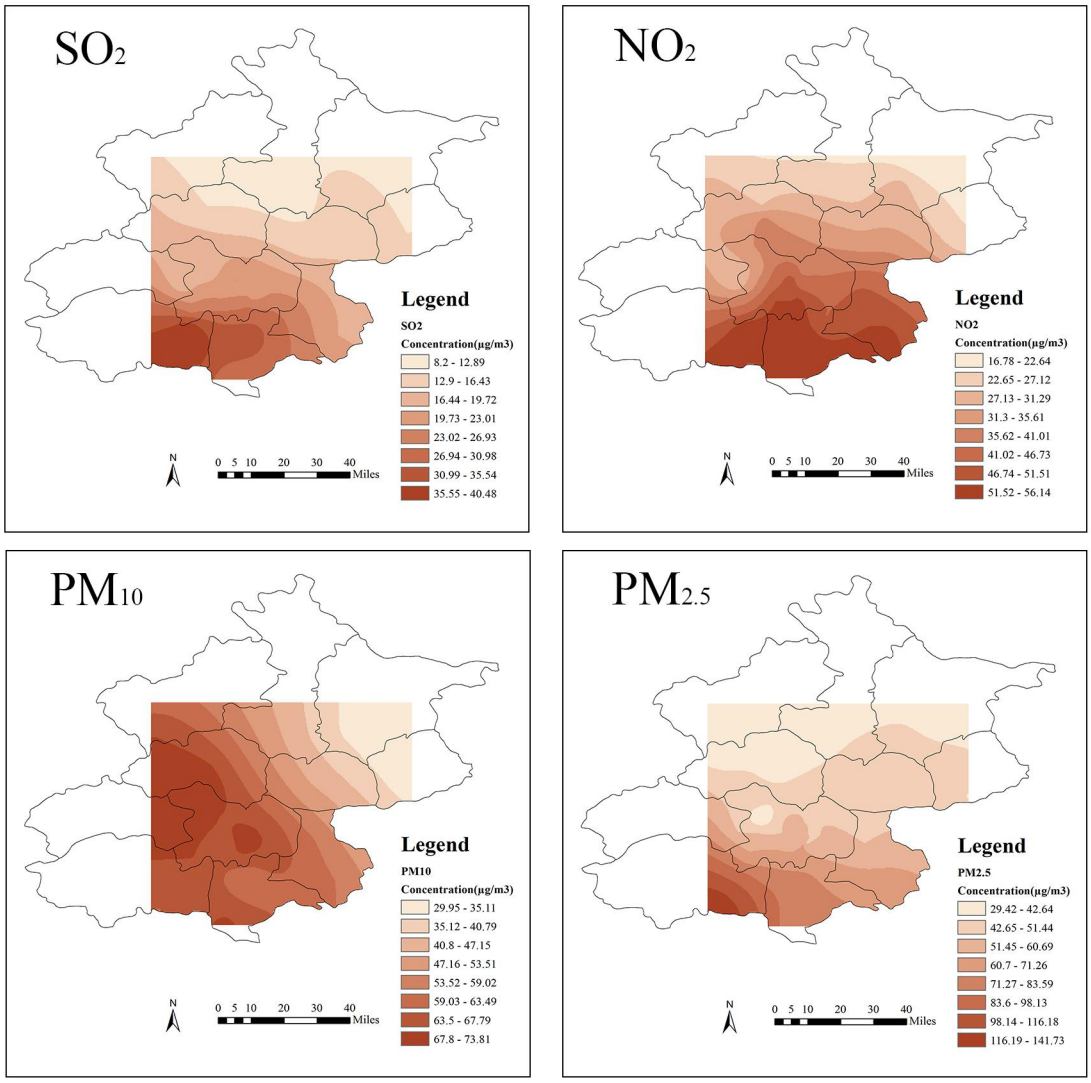


Figure 7.4: Interpolation maps of weekly average concentration of four pollutants in Beijing for Jan. 26, 2016

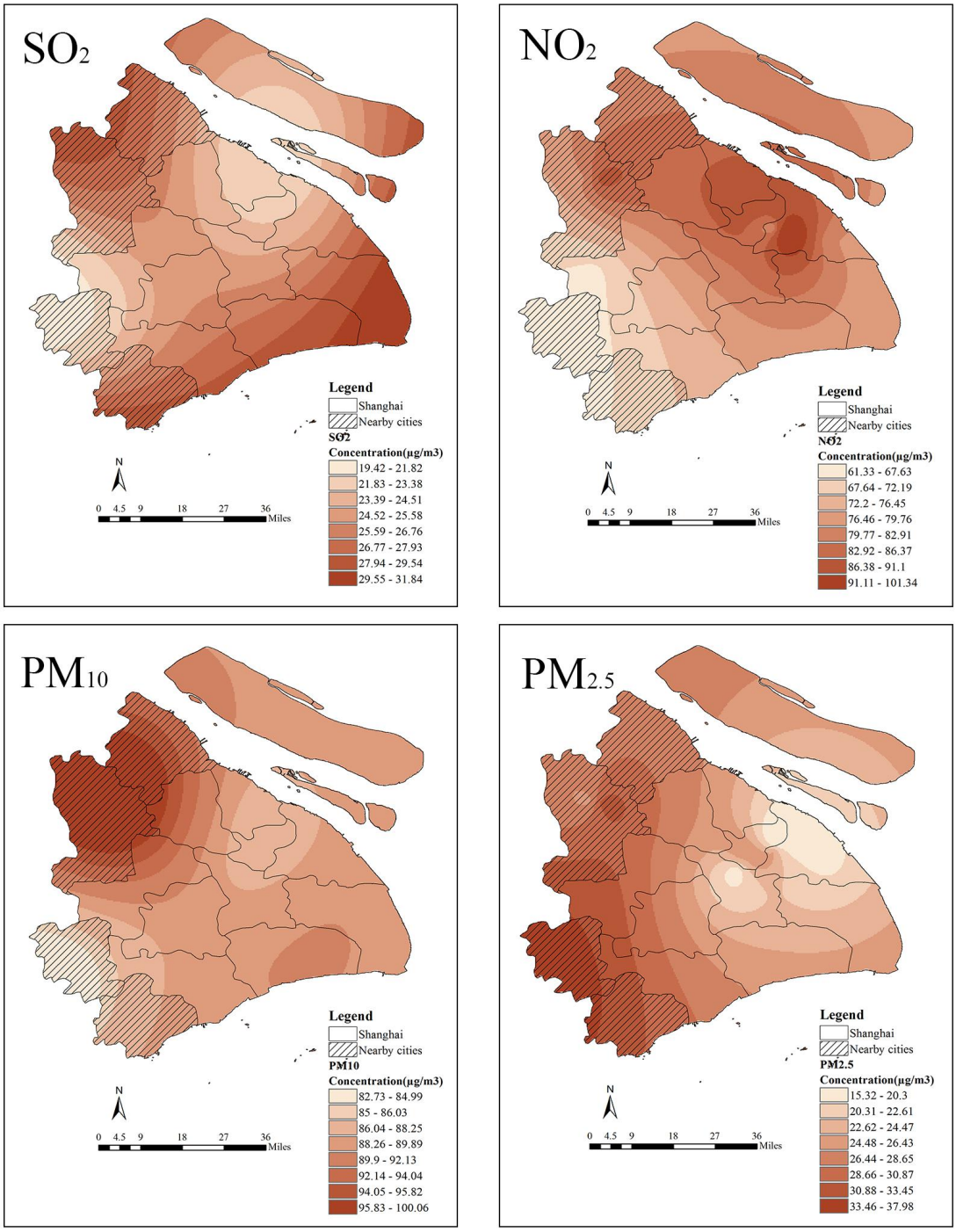


Figure 7.5: Interpolation maps of daily average concentration of four pollutants in Shanghai for Jan. 26, 2016

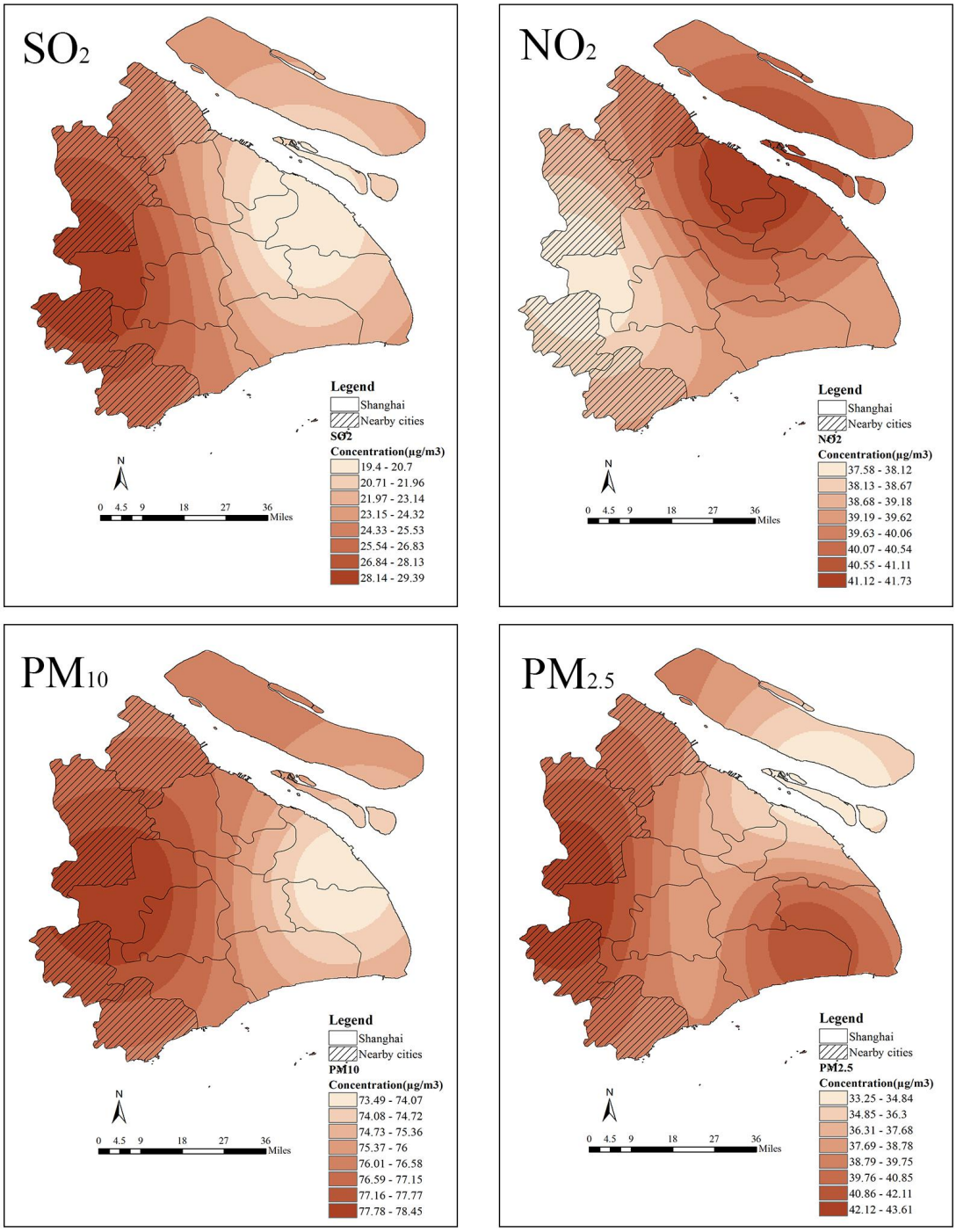


Figure 7.6: Interpolation maps of weekly average concentration of four pollutants in Shanghai for Jan. 26, 2016

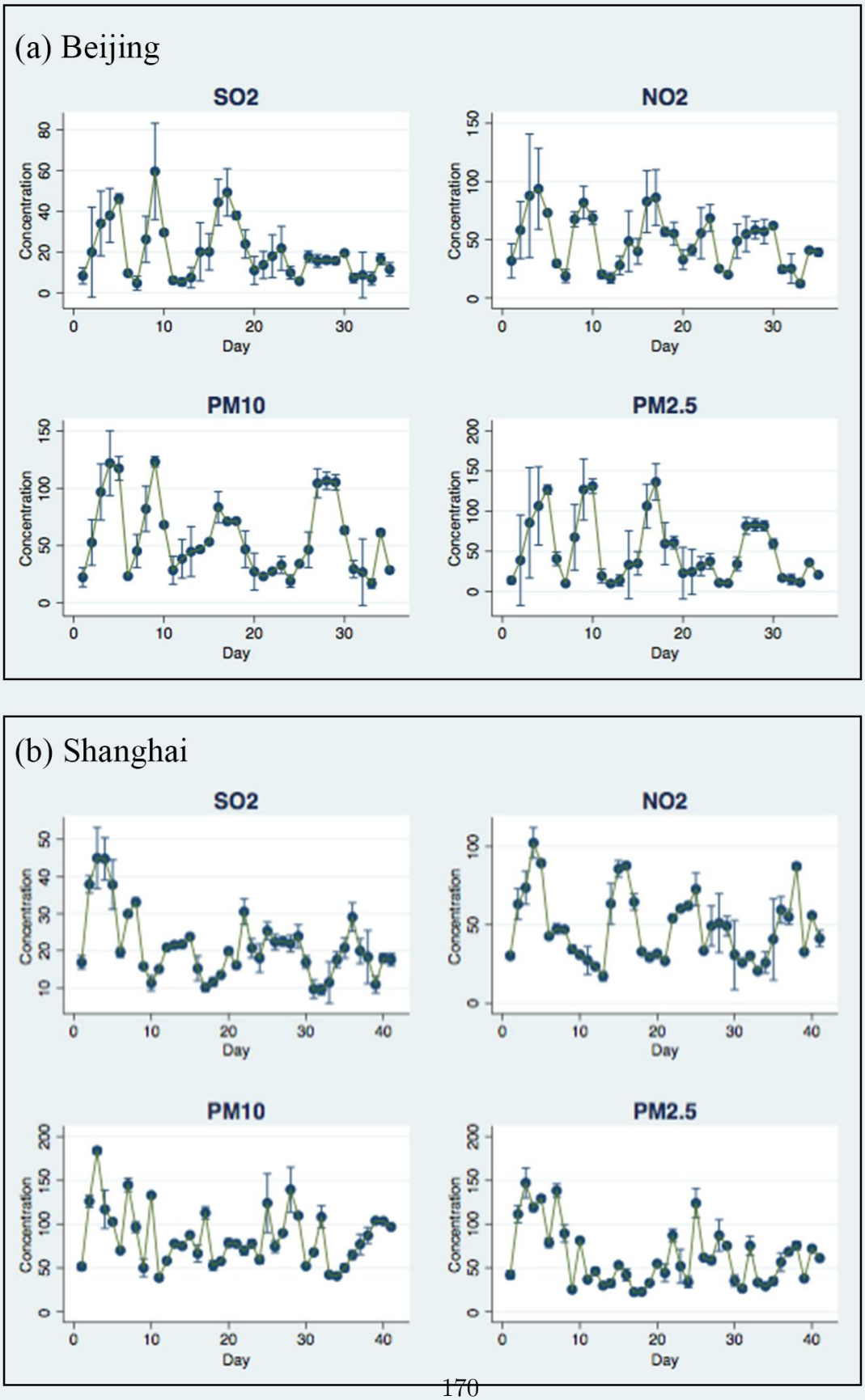


Figure 7.7: Temporal variations of interpolated air pollution exposures in Beijing and Shanghai

Table 7.5: Main models with using pollutant's concentration of "today" as air pollution indicators and the LS as dependent variables

	Beijing				Shanghai			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SO_2	-0.0109***				-0.0125**			
	(0.00328)				(0.00562)			
NO_2		-0.00612***				-0.00513**		
		(0.00202)				(0.00236)		
PM_{10}			-0.00475***				-0.00184	
			(0.00176)				(0.00156)	
$PM_{2.5}$				-0.00283**				-0.00245
				(0.00122)				(0.00176)
Log (income)	0.143**	0.143**	0.139**	0.140**	0.139**	0.146**	0.139**	0.138**
	(0.0601)	(0.0602)	(0.0602)	(0.0603)	(0.0641)	(0.0642)	(0.0642)	(0.0642)
Age	-0.00415	-0.00433	-0.00256	-0.00284	-0.00473	-0.00652	-0.00485	-0.00506
	(0.00640)	(0.00642)	(0.00637)	(0.00639)	(0.00632)	(0.00639)	(0.00634)	(0.00635)
Age-squared	-0.000227	-0.000228	-0.000240	-0.000185	0.000638*	0.000646*	0.000660*	0.000669**
	(0.000421)	(0.000422)	(0.000423)	(0.000422)	(0.000339)	(0.000339)	(0.000339)	(0.000339)
Female	0.0844	0.0904	0.0936	0.0857	0.255**	0.264**	0.252**	0.247**
	(0.0985)	(0.0987)	(0.0989)	(0.0989)	(0.104)	(0.105)	(0.105)	(0.104)
Collage graduate	-0.214	-0.201	-0.216	-0.216	0.215	0.206	0.209	0.208
	(0.155)	(0.155)	(0.155)	(0.155)	(0.158)	(0.158)	(0.158)	(0.158)
Poor health	-1.269***	-1.303***	-1.287***	-1.282***	-1.010***	-0.968***	-0.987***	-0.997***
	(0.234)	(0.234)	(0.235)	(0.235)	(0.218)	(0.218)	(0.218)	(0.219)
Single	0.0243	0.0245	0.0370	0.0394	-0.194	-0.211	-0.203	-0.209
	(0.216)	(0.216)	(0.217)	(0.217)	(0.190)	(0.190)	(0.191)	(0.191)
Divorced or widowed	-0.469	-0.470	-0.496	-0.480	-0.207	-0.141	-0.183	-0.208
	(0.430)	(0.431)	(0.432)	(0.432)	(0.297)	(0.298)	(0.297)	(0.298)

Household with	0.198	0.198	0.199	0.215	0.240*	0.245*	0.243*	0.245*
children	(0.191)	(0.191)	(0.191)	(0.191)	(0.131)	(0.131)	(0.131)	(0.131)
Household with	-0.159	-0.160	-0.162	-0.151	0.209	0.203	0.220	0.211
young children	(0.145)	(0.146)	(0.146)	(0.146)	(0.143)	(0.144)	(0.144)	(0.144)
Unemployed	-0.134	-0.145	-0.153	-0.148	0.0159	0.0471	0.0240	0.0340
	(0.419)	(0.420)	(0.420)	(0.420)	(0.338)	(0.338)	(0.339)	(0.339)
Resident in rented apartment	-0.918***	-0.934***	-0.927***	-0.941***	-0.527**	-0.543**	-0.522**	-0.511*
	(0.235)	(0.235)	(0.235)	(0.235)	(0.262)	(0.262)	(0.263)	(0.263)
Resident in owned apartment	-0.170	-0.182	-0.186	-0.182	-0.207	-0.226	-0.217	-0.205
	(0.140)	(0.140)	(0.140)	(0.140)	(0.192)	(0.192)	(0.192)	(0.193)
Resident in rented house	-0.0220	-0.0465	-0.0402	-0.0241	-0.143	-0.197	-0.210	-0.183
	(0.464)	(0.465)	(0.465)	(0.466)	(0.444)	(0.443)	(0.444)	(0.444)
Neighborhood safety	0.489***	0.486***	0.502***	0.507***	0.580***	0.576***	0.569***	0.569***
	(0.150)	(0.150)	(0.150)	(0.150)	(0.165)	(0.165)	(0.165)	(0.165)
Trust	0.982***	0.964***	0.980***	0.973***	0.112	0.132	0.120	0.117
	(0.215)	(0.215)	(0.216)	(0.216)	(0.175)	(0.175)	(0.175)	(0.175)
Enjoy	0.452***	0.440***	0.441***	0.445***	0.591***	0.588***	0.595***	0.589***
	(0.0683)	(0.0685)	(0.0686)	(0.0686)	(0.0716)	(0.0717)	(0.0718)	(0.0719)
Sad	-0.161**	-0.157**	-0.152**	-0.154**	-0.345***	-0.345***	-0.351***	-0.351***
	(0.0627)	(0.0627)	(0.0627)	(0.0628)	(0.0659)	(0.0659)	(0.0660)	(0.0660)
Personality-stable	-0.00547	-0.00993	-0.0171	-0.0109	0.0658	0.0626	0.0606	0.0641
	(0.0694)	(0.0693)	(0.0693)	(0.0696)	(0.0669)	(0.0669)	(0.0670)	(0.0670)
Personality-passive	-0.132*	-0.129*	-0.142*	-0.134*	0.0442	0.0446	0.0456	0.0490
	(0.0775)	(0.0776)	(0.0776)	(0.0777)	(0.0725)	(0.0725)	(0.0727)	(0.0726)
Personality-outgoing	0.430***	0.433***	0.425***	0.430***	0.242***	0.237***	0.244***	0.244***
	(0.0822)	(0.0824)	(0.0824)	(0.0825)	(0.0808)	(0.0808)	(0.0809)	(0.0809)
Car-commuting	0.0610	0.0651	0.0723	0.0607	0.515***	0.504***	0.515***	0.516***
	(0.110)	(0.110)	(0.110)	(0.110)	(0.121)	(0.121)	(0.121)	(0.121)
Public transit- commuting	-0.134	-0.129	-0.139	-0.132	0.0883	0.0822	0.0877	0.0873
	(0.102)	(0.103)	(0.103)	(0.103)	(0.110)	(0.110)	(0.110)	(0.110)

Congestion time	-0.138***	-0.136***	-0.137***	-0.136***	-0.0645	-0.0618	-0.0641	-0.0664
	(0.0425)	(0.0425)	(0.0426)	(0.0426)	(0.0909)	(0.0909)	(0.0911)	(0.0911)
Knowledge level of air pollution	-0.334***	-0.332***	-0.332***	-0.334***	-0.147**	-0.154**	-0.147**	-0.147**
	(0.0613)	(0.0614)	(0.0614)	(0.0615)	(0.0702)	(0.0702)	(0.0703)	(0.0703)
Sunny	0.183	0.142	0.179	0.126	0.125	-0.0541	0.0426	0.104
	(0.128)	(0.125)	(0.131)	(0.126)	(0.140)	(0.110)	(0.144)	(0.165)
Temperature	0.00638	0.00937	0.00873	-0.000828	-0.0131	-0.00390	-0.00993	-0.00730
	(0.0167)	(0.0174)	(0.0178)	(0.0166)	(0.0151)	(0.0155)	(0.0152)	(0.0154)
Distance to CBD	-1.284***	-1.385***	-1.289***	-1.169**	-0.0625	0.00526	-0.0554	-0.0726
	(0.465)	(0.467)	(0.466)	(0.467)	(0.458)	(0.458)	(0.460)	(0.460)
Constant	5.902***	6.039***	5.980***	5.786***	3.111***	3.082***	2.979***	2.981***
	(0.565)	(0.577)	(0.576)	(0.565)	(0.869)	(0.868)	(0.870)	(0.868)
Observations	958	958	958	958	878	878	878	878
R-squared	0.330	0.329	0.327	0.326	0.321	0.321	0.319	0.319

The coefficients indicate the change in LS rating with a 1-$\mu g/m^3$ change in a pollutant's concentration, and the magnitudes of impact vary across pollutants. This study calculates the possible LS improvement from reducing air pollutant concentration to a lower health risk level based on the Chinese and the U.S. air quality index (AQI); the results are presented in Table 7.6. In AQI standards, air pollution concentrations are categorized into different levels that correspond to certain health risk levels. The U.S. and Chinese standards are similar for all pollutants except the thresholds of $PM_{2.5}$.

Table 7.6: Upper limit of pollutant concentrations in China and the U.S. AQI standards

China				
AQI category	SO_2	NO_2	PM_{10}	$PM_{2.5}$
Good	50	40	50	35
Moderate	150	80	150	75
Unhealthy for sensitive groups	475	180	250	115
Unhealthy	800	280	350	150
Very unhealthy	1600	565	420	250
Hazardous	2100	750	500	350
Severe	2620	940	600	500
U.S.				
AQI category	SO_2	NO_2	PM_{10}	$PM_{2.5}$
Good	–	–	54	12
Moderate	–	–	154	35.4
Unhealthy for sensitive groups	–	–	254	55.4
Unhealthy	–	–	354	150.4
Very unhealthy	–	–	424	250.4
Hazardous	–	–	604	500.4

Notes:
1. Unit: $\mu g/m^3$. Breakpoints of each category are the 24-hours average concentrations.
2. The breakpoints of SO_2 and NO_2 in U.S. AQI standard are given as one-hour average concentration, which cannot be directly quantitatively compared with the counterparts in China's standard that are 24-hours average. Therefore, they are blank here.

Current average pollution levels in our samples correspond to the following health risk levels in China's AQI standards: SO_2 concentrations are in the "Good" category for both cities (24 $\mu g/m^3$ and 26 $\mu g/m^3$ for Beijing and Shanghai, respectively); NO_2 and PM_{10} concentrations are in the "Moderate" category (54 $\mu g/m^3$ and 61 $\mu g/m^3$ for NO_2 and

63 $\mu g/m^3$ and 100 $\mu g/m^3$ for PM_{10} for Beijing and Shanghai, respectively) according to both Chinese and U.S. standards; $PM_{2.5}$ is at the higher end of the "Moderate" category according to the Chinese standard and in the "Unhealthy" category according to the U.S. standard (57 $\mu g/m^3$ and 75 $\mu g/m^3$ for Beijing and Shanghai, respectively). If the average levels of NO_2, PM_{10} and $PM_{2.5}$ in Beijing were to decrease to the upper-limit concentration of the "Good" category according to China's standard, the LS of Beijing residents would improve on average by approximately 0.09, 0.06, and 0.09 points, respectively. If the level of $PM_{2.5}$ in Beijing is to decrease to the upper-limit concentration of the "Good" category according to U.S. standard, LS would improve on average by approximately 0.12 points. Thus, reducing the levels of each pollutant to the "Good" category of the AQI standards would improve LS by 0.24∼0.27.

Based on the corresponding AQI standards, $PM_{2.5}$ pollution is much more serious than the other three pollutants; $PM_{2.5}$ is relatively more detrimental to people's health than other pollutants (Dockery et al., 1993; Muller and Mendelsohn, 2007; Beelen et al., 2014). However, the magnitude of the impacts on LS does not appear to reflect the relative health risks from pollution exposure. In Beijing, the coefficients for $PM_{2.5}$ and PM_{10} are smaller than those of SO_2 and NO_2, and they do not impact the LS of Shanghai residents. One explanation for this phenomenon is that SO_2 has been widely discussed for a relatively long period of time in China, whereas $PM_{2.5}$ and PM_{10} have started to receive attention more recently because of the exponential increase in the occurrence of haze episodes. People may be more sensitive to industrial pollution and emissions from cars and may also blame SO_2 and NO_2 for haze and health concerns that are actually caused by $PM_{2.5}$ and PM_{10} due to a lack of knowledge and understanding of the differences between pollutants.

In addition, according to the results provided in Table 7.5, the subjective knowledge level of air pollution has statistically significant negative relationship with the LS of both Beijing and Shanghai residents. This result indicates that people, who have reported higher knowledge level of air pollution, have relatively lower life satisfaction in both cities.

The estimated coefficients of knowledge level of air pollution in Beijing and Shanghai are approximately -0.334 and -0.147, respectively. The difference between the coefficients indicates that the LS scores of Beijing residents are more negatively affected by the reported knowledge level of air pollution compared to that of Shanghai residents. Given that the variable of knowledge level of air pollution is self-reported subjective knowledge level rather than the object knowledge level, this variable may be closely associated with people's awareness to air pollution. Hence, the regression result implies that the awareness of air pollution may decrease people's LS, and this effect is particularly serious in Beijing because the issue of air pollution is more politicized and widely reported in Beijing than in Shanghai.

7.3.2 The impact of temporal changes in air pollution levels

In addition to analyzing pollution levels based on the survey data, this study examines the impact of temporal changes in pollutant concentrations on LS. In particular, this study are interested in whether people are negatively affected by worsening air pollution conditions. First, this study takes the difference between the pollution concentration on the survey date and both the concentration from the day before and the average concentration from one week prior to the survey date. A positive difference indicates that the air pollution condition worsened on the survey date compared to the day before or the average pollution over the previous 7 days. This study then separately uses the generated two variables to rerun the full model. The regression results for the air pollution indicators are shown in Table 7.7.

Table 7.7: Full models with using the difference of concentrations with various time-specification as air pollution indicators and LS as dependent variable

	the difference of concentrations of survey date and day before as the pollution indicator				the difference of concentrations of survey date and one week's average prior to the survey date as the pollution indicator			
Beijing	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SO_2	-0.00556				-0.00863**			
	(0.00345)				(0.00348)			
NO_2		-0.00234				-0.00485**		
		(0.00218)				(0.00246)		
PM_{10}			-0.00199				-0.00533***	
			(0.00177)				(0.00188)	
$PM_{2.5}$				-0.000648				-0.00275**
				(0.00117)				(0.00129)
Constant	4.644***	4.642***	4.605***	4.595***	4.597***	4.646***	4.566***	4.598***
	(0.897)	(0.899)	(0.897)	(0.898)	(0.895)	(0.896)	(0.894)	(0.895)
Observations	958	958	958	958	958	958	958	958
R-squared	0.303	0.302	0.302	0.301	0.305	0.304	0.307	0.304
Shanghai	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
SO_2	-0.00503				-0.0106			
	(0.00667)				(0.00655)			
NO_2		0.00118				-0.00353		
		(0.00302)				(0.00234)		
PM_{10}			-0.000258				-0.00182	
			(0.00111)				(0.00170)	
$PM_{2.5}$				0.000168				-0.00279
				(0.00169)				(0.00196)
Constant	3.594***	3.569***	3.574***	3.568***	3.510***	3.549***	3.510***	3.522***
	(1.016)	(1.016)	(1.016)	(1.019)	(1.015)	(1.015)	(1.017)	(1.016)
Observations	878	878	878	878	878	878	878	878
R-squared	0.314	0.314	0.314	0.314	0.316	0.316	0.315	0.315

Moreover, this study observes that Beijing residents are affected by the difference in pollution between the survey date and the one-week average. This result indicates that although pollution exposure on the survey day is important for LS, the residents appear to compare the current pollution level with the levels in the previous days: if the pollution level is relatively worse, the residents feel worse, and vice versa. In contrast, in Shanghai, temporal changes in pollution levels do not affect respondents' LS, possibly because Shanghai residents have lower awareness of air pollution than do Beijing residents and are generally disinterested in the politicized pollution issues. Furthermore, Shanghai residents are comparatively wealthier than Beijing residents, and money can buy insulation from pollution exposure; this difference might also explain the lack of a significant effect of pollution exposure or changes in pollution level on LS in Shanghai. However, the data do not support such a mechanism, as this study found no significant impact of an interaction between estimated pollution level and any variable potentially associated with insulation, such as household income, car as commuting mode and privately owned residence.

7.3.3 The monetary valuation of air pollution

Table 7.8 shows the estimated average monetary values for different income groups for each pollutant and the statistical significance of the regression coefficients for estimated pollution concentration and household income in the main models. The results of the main regression models show that household income has a positive impact on the LS of the residents of both cities. The precision of the magnitude of the coefficients is important, as a biased coefficient would affect the accuracy of the average monetary valuation. In particular, the over-estimation of the income coefficient due to insufficient control of asset/wealth-related factors is a concern. As expected, the income coefficients in the estimation models without controlling for asset/wealth-related variables such as car-commuting and residential type dummies are approximately one forth higher than

those in the models in which these variables are controlled. By controlling for other determinants that might be associated with income, such as health and living in a safe neighborhood, this study heavily mitigates the possible upward bias of income coefficient.

Table 7.8: Estimated monetary valuation of air pollution

		Beijing	Shanghai
SO_2	95% lower bound of income	US$ 175 (CNY 1,143)	US$ 207 (CNY 1,348)
	Average income	US$ 1,899 (CNY 12,348)	US$ 2,234 (CNY 14,523)
	95% higher bound of income	US$ 4,573 (CNY 29,727)	US$ 4,012 (CNY 26,079)
NO_2	95% lower bound of income	US$ 98 (CNY 641)	US$ 81 (CNY 527)
	Average income	US$ 1,066 (CNY 6,933)	US$ 873 (CNY 5,674)
	95% higher bound of income	US$ 2,567 (CNY 16,690)	US$ 1,567 (CNY 10,189)
PM_{10}	95% lower bound of income	US$ 78 (CNY 512)	–
	Average income	US$ 851 (CNY 5,535)	–
	95% higher bound of income	US$ 2,050 (CNY 13,327)	–
$PM_{2.5}$	95% lower bound of income	US$ 46 (CNY 303)	–
	Average income	US$ 503 (CNY 3,274)	–
	95% higher bound of income	US$ 1,212 (CNY 7,883)	–

Notes:
1. Exchange rate: 1US$ = 6.5 CNY
2. Given that air pollution is negative amenity, the estimated MV has negative value in people's utility. It can be also considered as a dollar value that residents are willing to pay for pollution reduction.

These estimated monetary values can be considered as proxies for the amount that people are willing to pay for a 1 $\mu g/m^3$ concentration decrease (vice versa) per year. According to the estimated results, on average, the monetary values of SO_2, NO_2, PM_{10}, and $PM_{2.5}$ for Beijing residents are 1,899 USD (12,348 CNY), 1,066 USD (6,933 CNY), 851 USD (5,535 CNY), and 503 USD (3,274 CNY), respectively. The monetary values of SO_2 and NO_2 for Shanghai residents are 2,234 USD (14,523 CNY) and 873 USD (5,674 CNY), respectively. Hence, Beijing residents place almost the same value on SO_2 reduction as do

Shanghai residents but approximately 1.5 times more value on NO_2 reduction, on average. Although this study uses different data and methodology, the estimated monetary values seem to be comparable to the ones that found in previous studies. Estimated average monetary valuation of NO_2 for 10 European countries reported by Welsch (2006) was 854 USD per unit reduction; this valuation is very close to the valuation in Shanghai. Similarly, Levinson (2012) reported 1,057 USD per unit reduction of PM_{10} in the US, which is also close to the evaluation in Beijing. Zhang et al. (2017) reported a substantially lower average willingness-to-pay of 42 USD (258 CNY) for 1 $\mu g/m^3$ reduction in $PM_{2.5}$ for average Chinese; given that Beijing residents have a much higher income level and relatively more aware and sensitive about air pollution condition than the residents in other parts of China.

7.3.4 The impacts of other determinants

There are several notable differences in significance of individual and household characteristics across cities that allow us to examine different characteristics of the residents in the two cities. Household income has a positive impact on LS for the residents in both cities, but the magnitude of the impact is larger in Shanghai. Commuting by car raises the LS only in Shanghai. Beijing residents value other people's trust, whereas trust does not appear to affect the LS of Shanghai residents. The subjective well-being of Shanghai residents appears to be more sensitive to current mood.

These results appear to support the general characterization of "Shanghainese" as money-oriented and striving for higher living standards. The effect of age is also different for the two cities; age affects the LS for Beijing, whereas the U-shaped effect of age, reported in previous studies, appears to hold in Shanghai. With respect to gender, females in Shanghai have higher LS than do males, where there is no effect of gender in Beijing. Furthermore, this study observes different impacts of distance to the CBD between the cities: The LS of Beijing residents decreases with increasing distance between residence

and the CBD, whereas that of Shanghai residents is unaffected by the distance to the CBD. This result may be partially explained by the relatively slow development of the Beijing suburb areas compared to the Shanghai suburbs, which exhibit characteristics of suburbs in developed countries, with a high concentration of upper-income-class residents.

There are also determinants that have similar impacts in both cities. A college degree, marriage, and having children/young children (under 6 years old) do not explain the variation in LS and do not follow the empirical evidence from previous studies (Welsch, 2009,?). These results can be partly explained by the absence of a premium on higher education because it has become more common to attend a university, while having a degree does not guarantee a better job or a higher social status in the recent business environment. Moreover, the high stress and environment of the extremely competitive job market and high living expenses, marriage, and children in these cities may have cancelled out the possible benefits of having a family, which is partially supported by the increasing divorce rates in the two cities. In contrast, individuals with a good subjective health condition, with outgoing personalities or living in safe communities have relatively higher LS in both cities. This study found no significant effect of temperature or weather in either city.

7.3.5 Robustness check

This study provides three robustness checks: regression analyses with an alternative air pollution exposure estimation technique, an alternative regression method, and an alternative LS indicator (Table 7.9).

Table 7.9: Results of robustness check

a. Robustness check of regression models with using air pollutants concentration of "today" generated from nearest sites approach

	Beijing				Shanghai			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SO_2	-0.00653**				-0.0132**			
	(0.00262)				(0.00547)			
NO_2		-0.00278*				-0.00450**		
		(0.00149)				(0.00227)		
PM_{10}			-0.00211**				-0.00142	
			(0.00101)				(0.00152)	
$PM_{2.5}$				-0.00211**				-0.00249
				(0.00101)				(0.00173)
Log (income)	0.148**	0.141**	0.136**	0.136**	0.142**	0.146**	0.139**	0.139**
	(0.0603)	(0.0604)	(0.0604)	(0.0604)	(0.0640)	(0.0642)	(0.0642)	(0.0642)
Constant	5.744***	5.815***	5.785***	5.785***	3.095***	3.069***	2.955***	2.971***
	(0.562)	(0.572)	(0.567)	(0.567)	(0.867)	(0.869)	(0.870)	(0.867)
R-squared	0.326	0.324	0.325	0.325	0.322	0.321	0.318	0.319

b. Robustness check of the main model with using ordered probit approach

	Beijing				Shanghai			
SO_2	-0.00801***				-0.00873**			
	(0.00235)				(0.00393)			
NO_2		-0.00457***				-0.00385**		
		(0.00144)				(0.00165)		
PM_{10}			-0.00391***				-0.00128	
			(0.00126)				(0.00109)	
$PM_{2.5}$				-0.00218**				-0.00163

				(0.000874)				(0.00122)
Log (income)	0.0918**	0.0917**	0.0883**	0.0897**	0.0835*	0.0884**	0.0834*	0.0830*
	(0.0430)	(0.0430)	(0.0430)	(0.0430)	(0.0449)	(0.0449)	(0.0449)	(0.0449)
Pseudo R-squared	0.0829	0.0825	0.0827	0.0812	0.0893	0.0898	0.0890	0.0889
c. Robustness check of the main model with using alternative LS indicator								
			Beijing				Shanghai	
SO_2	-0.00621*				-0.00859			
	(0.00324)				(0.00562)			
NO_2		-0.00295*				-0.00587**		
		(0.00199)				(0.00235)		
PM_{10}			-0.00179				-0.0000590	
			(0.00174)				(0.00156)	
$PM_{2.5}$				-0.00124				-0.000916
				(0.00121)				(0.00175)
Log (income)	0.0942	0.0939	0.0920	0.0926	0.101	0.109*	0.101	0.101
	(0.0594)	(0.0595)	(0.0595)	(0.0595)	(0.0640)	(0.0639)	(0.0641)	(0.0641)
Constant	6.297***	6.347***	6.246***	6.214***	3.878***	3.960***	3.709***	3.751***
	(0.878)	(0.885)	(0.881)	(0.878)	(0.868)	(0.865)	(0.868)	(0.866)
R-squared	0.358	0.357	0.356	0.356	0.406	0.409	0.404	0.405

To validate the performance of estimating air pollution exposure using the OK method, this study calculates the pollution exposure indicator for the survey date by simply matching the pollution level of the nearest monitoring station from the geo-coded residents' locations. As show in Table 7.9, the regression results for the pollutants with alternative pollution indicators are similar to the main results. The pollution indicators have higher statistical significance in the main result in which the pollution level is estimated using the OK method; hence, the original results are more robust. Therefore, this study recommends the use of a more sophisticated spatial interpolation method in SWB studies to improve the performance of empirical analyses, especially when rich spatial objective data, such as pollution data from various monitoring sites, are available.

Although it was discussed in Section 5 that OLS is the preferred method according to previous studies, this study examines the impacts of air pollution on LS using an ordered probit model. The signs and statistical significances are consistent with the results from OLS for both pollution indicators and other control variables, which indicate the robustness of our regression results and the validity of the comparative results of Beijing and Shanghai that well-being of Beijing residents are more susceptible to air pollution.

Lastly, this study uses an alternative SWB indicator to examine whether the effect of air pollutions found in the main results still hold. The alternative measure is the response to the following question: "All things considered, how satisfied are you with your life as a whole these days?" Respondents answered by choosing from a 10-point scale ranging from 1 (very dissatisfied) to 10 (very satisfied). Overall, the impacts of pollutants are less significant compared to the results of main results that use Cantril ladder as the SWB measure and the results do not indicate clear trend of statistically negative impacts of air pollution, especially for the Beijing residents.

Nonetheless, the results of alternative SWB indicator does not necessary undermine the above-mentioned main results given the difference in the context of question. The

most distinctive feature of Cantril ladder measure is that it does not use "happiness" or "life satisfaction" within the question, while the alternative measures ask to rate "life satisfaction" directly. According to the previous studies, SWB questions that include the "happiness" or "life satisfaction" could direct respondents to use narrow definition of well-being compared with more general questions without particular words to that describe the state of well-being. Moreover, usage of "happiness" or "life satisfaction" in the question tend to induce respondents to provide relatively more subjective and sentimental evaluations of their life compared to the questions without those words under certain cultural and language circumstances. Especially in Chinese language, the respondents are more likely to be sensitive to the usage of particular wording of well-being (Chui and Wong, 2016). Given that air pollution is rather objective daily problems, the Cantril ladder measure that this study uses in the main analysis fits the purpose of examining its impact on SWB.

7.4 Summary

This paper has examined the role of environmental quality in determining people's well-being by empirical analyzing the impacts of various air pollutants in Beijing and Shanghai using self-reported life satisfaction data from an original Internet survey conducted in 2016. This study combines the survey data and air pollution data for four major pollutants (SO_2, NO_2, PM_{10} and $PM_{2.5}$) collected from 35 and 12 monitoring sites in Beijing and Shanghai, respectively. This study uses a GIS interpolation technique to estimate the residents' pollution exposure to provide more reliable air pollution effects compared to matching respondents' data with pollution level from the nearest monitoring site from their residences.

The results of empirical analysis highlight the importance of air pollution as a determinant of people's life satisfaction and people's perception of pollutants may digress

from the actual severity or threat. Also, our results show that the pollution environment and people's perceptions vary across cities; Both SO_2 and NO_2 negatively affect the LS of residents in both cities but the magnitude of the impacts for SO_2 and NO_2 are comparatively smaller in Shanghai, and PM_{10} and $PM_{2.5}$ negatively impact Beijing residents but do not have a statistically significant impact on LS for Shanghai residents. The residents in both cities have similar valuation for 1-$\mu g/m^3$ of SO_2, whereas for NO_2, the 1-$\mu g/m^3$ value of Beijing residents is approximately 1.5-fold that of Shanghai residents. This finding indicates that pollution reduction would more effectively improve the LS of Beijing residents. Moreover, the LS of Beijing residents is also sensitive to changes in the pollution level leading up to the survey date, while Shanghai residents are unaffected by the pollution level on previous days. When the pollution level exceeds the previous days' level or 7-day averages, it has a negative effect on Beijing residents but no statistically significant impact on Shanghai residents.

The empirical evidence indicates relative interest and concern for pollution by Beijing residents compared to their counterparts in Shanghai, although the pollution level for the period of analysis is either evenly matched or slightly higher in Shanghai. While these two cities are the largest Chinese cities and are globally recognized, our results support the general characterization of the two rival cities: Beijing as political center and Shanghai as a financial and business center. As the pollution issue becomes more politicized, the concern and sensitivity of Beijing residents may increase, while the everyday well-being of Shanghai residents may remain relatively less affected by pollution. For example, when this study asked respondents to report their knowledge on air pollution, approximately 65% of Beijing respondents responded 'very knowledgeable', whereas only approximately 20% of Shanghai respondents did so. Although self-reported knowledge level does not necessarily accurately correspond to actual knowledge level, the findings provide some support for a difference between Beijing and Shanghai residents in the attention given to the pollution issue.

In addition, the regression results indicate that the residents, who reported higher

knowledge regarding air pollution, seem to have relatively lower LS. Given that the knowledge variable in the analysis is subjective and is on air pollution as a whole, the variable may reflect people's general awareness rather than their actual knowledge level. Hence, while it is possible to consider the case in which official policies and activities could effectively modify people's subjective awareness of air pollution to increase their life satisfaction, this study needs to be cautious to claim the impacts of additional or suppressing information; further analyses would be required to confirm the impact of information availability and control about air pollution on LS.

Chapter 8

Conclusion

In order to provide supportive technique and evidence for promoting the sustainable development of urban area in Japan and China, this study enhanced and extended the land use change (LUC) modeling framework in a highly-developed metropolitan area – the Greater Tokyo Area, and also estimated the monetary values of air pollution in Beijing and Shanghai, and northeast region in China.

This study enhanced the LUC modeling from three perspectives: 1) enhances the stochastic modeling with tree-based ensemble algorithms; 2) enhances the spatial modeling with convolutional neural networks (CNN); 3) enhances the temporal modeling with recurrent neural networks (RNN). LUC models incorporating tree-based algorithms were developed and evaluated in Greater Tokyo Area, which has complicated LU transition rules; LUC models incorporating CNN were developed and evaluated in Saitama prefecture, which has complex and scattered spatial pattern of LU; LUC models incorporating RNN were developed and evaluated in Tsukuba City, which is undergoing continuous and stable urban expansion in the past decades.

This study integrates cellular automata (CA) with four tree-based ensemble algorithms, which are bagged decision tree (bagged DT), bagged gradient boosting decision

tree (bagged GBDT), random forests (RF) and extremely randomized trees (ERT). The four algorithms differ in the degree of randomness in the mechanism. The results show that all four tree ensemble models outperform a multi-layer perceptron (MLP) model, which is a popular in previous LUC modeling studies, by approximately 5% $\sim$ 30%. The results also show that tree ensemble models with higher degree of randomness achieve better predictive performances. In particular, ERT, which incorporates the highest degree of randomness among the four algorithms, achieves the best predictive performance. These results indicate that introducing randomness into LUC modeling by tree-based ensemble algorithms improves the predictive performance of LUC in Greater Tokyo Area.

In terms of the LUC modeling with CNN, this study developed integrated convolutional neural networks (conv-net) and integrated convolutional denoising autoencoder and MLP (cdae-net) to enhance the prediction of LU transition probability. The conv-net and cdae-net combine manually designed geographical features and hidden spatial features extracted from satellite images. According to the comparison results of conv-net / cdae-net and a LUC model that uses only geographical features (geo-net), both conv-net and cdae-net significantly outperform the geo-net in terms of prediction accuracy of both transition probability and final LU map by approximately 15 $\sim$ 30%. This result indicates that incorporating spatial features by convolutional-based method can improve the predictive performance of conventional LUC modeling approach. Moreover, the predictive performances of conv-net and cdae-net depend on the data; conv-net generally outperforms cdae-net, while cdae-net outperforms conv-net when data is noisy.

In terms of the LUC modeling with RNN, this study constructed two categories of RNN models, which are simple RNN which is the basic RNN variant, and RNN variants with gated architecture (long short-term memory (LSTM), LSTM with peephole connection (LSTM-peephole) and gated recurrent unit (GRU)). Results show that RNN variants with gated architecture (LSTM, LSTM-peephole, GRU) significantly outperform simple RNN. Given that RNN variants with gated architecture have stronger capability of modeling long-term temporal dependency than simple RNN, these results indicate the

positive effect of modeling long-term dependency for improving the performance of LUC modeling. In addition, out of the three RNN variants with gated architecture, LSTM-peephole exhibits the highest predictive performance.

Given the above results, this study provided enhanced approaches for modeling the LUC process in Greater Tokyo Area, and showed the effectiveness of these approaches for improving the modeling performances. By using these approaches, more reliable LU prediction could be generated and used to better support the decision-making of strategic urban planning in Greater Tokyo Area. Moreover, the modeling approaches developed in Greater Tokyo Area (including the whole area, Saitama prefecture and Tsukuba city) could be applied in other highly developed metropolitan areas, which share similarities with Greater Tokyo Area, such as New York or London metropolitan area. In addition, the modeling approaches in this study were designed and developed for addressing specific characteristics of modeling area (i.e., complex LU transition rules, complex spatial pattern of LU distribution, or long-term slow urban expansion). These approaches can be flexibly combined according to the characteristics of specific task.

With respect to the air pollution assessment, this study uses SWB approach to assess the impact of air pollution on people's well-being and to estimate the monetary value of air pollution. SWB measures and other individual characteristics are collected from an original Internet survey conducted in China during January and February in 2016. Air pollution data are collected from official statistical yearbook or measurement of monitoring sites.The results of regression analyses show the statistically significant negative effect of air pollution on Chinese people's well-being. The magnitude of impact of air pollution varies with the regions. In particular, $PM_{2.5}$ has statistically significant effects on the well-being of residents in Northeastern region and Beijing, but does not seem to have robust impact on the well-being of Shanghai residents. Furthermore, the well-being of Beijing residents are more affected by the NO_2 compared with Shanghai residents, but are less affected by NO_2 compared with Shanghai residents.

Based on these regression results, the estimated average monetary value of $PM_{2.5}$ with time specification of one month is approximately USD 900 in Northeastern region of China; the estimated average monetary value of $PM_{2.5}$ with time specification of one day is approximately USD 500 in Beijing. These results indicate that, although the northeastern Chinese residents are relatively less wealthy, they are willing to pay more to compensate for their loss of SWB caused by the same unit of air pollution when compared with Beijing residents.

The assessment results provide several policy implications for Chinese policy-makers. Firstly, although air pollution problem in big cities such as Beijing and Shanghai received great attention from the public and government, the results of this study show that northeast Chinese residents suffer more from air pollution than Beijing and Shanghai residents. This finding also indicates that the air pollution mitigation in Northeast China has higher marginal benefits compared with that in Beijing or Shanghai. Thus, Chinese government is suggested to pay higher attention and investment to the air pollution mitigation in less-developed area such as northeast region of China.

Second, the magnitude of negative impact of air pollution varies with the individual characteristics, particularly subjective health condition, household with young children and expenditure in environmental activities. Individuals' worries to their and their children's health condition may lead to the aggravation of negative impact of air pollution; on the other hand, individuals who have spent a lot money on environmental activities may be troubled by the fact that air quality does not significantly become better regardless of their efforts either by voluntary or by mandatory. Therefore, Chinese government is suggested to take measures to relieve individuals' anxiety on their and their children's health by propagation or education. In addition, government is also suggested to curb the trend of financial contribution in environmental activities.

Lastly, different air pollutants have negative impacts on individuals' well-being with varying magnitudes. However, it is not suggested to over-interpret this result. For in-

stance, the finding, which is that PM_{10} and $PM_{2.5}$ have no statistically significant negative impact on Shanghai residents' well-being, does not imply that the pollution control of PM_{10} and $PM_{2.5}$ is unnecessary in Shanghai. Instead, government is suggested to 1) investigate whether individuals hold biased recognition to the effects of different air pollutants; 2) take the varying effects of air pollutants on people's well-being into account when proposing air pollution mitigation goals or policy.

This study still has several limitations, and can be further improved in future work. Although this study has developed approaches to enhance the LUC modeling in a complex urban system, due to the limitations of spatial availability and computation power, LUC models were developed in different urban areas, which fails to allow an unbiased examination and comparison of varying modeling techniques and strategies. In addition, this study present useful findings for two important and highly prioritized urban environmental issues by examining the Japan and China areas. Nonetheless, the future work should consider the connection between the two environmental issues, land use change and air pollution, by combining the findings of this study to build a modeling and assessment framework. Updating the modeling framework would allow for a modeling flow to forecast the future LU map, then to estimate the future spatial distribution of air pollution, and finally to assess the possible impact of current urban planning scenario on people's well-being.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Aburas, M. M., Ho, Y. M., Ramli, M. F., Ash'aari, Z. H., 2016. The simulation and prediction of spatio-temporal urban growth trends using cellular automata models: A review. International Journal of Applied Earth Observation and Geoinformation 52, 380–389.
URL http://dx.doi.org/10.1016/j.jag.2016.07.007

Al-sharif, A. A., Pradhan, B., 2015. A novel approach for predicting the spatial patterns of urban expansion by combining the chi-squared automatic integration detection decision tree, Markov chain and cellular automata models in GIS. Geocarto International 30 (8), 858–881.
URL http://www.tandfonline.com/doi/full/10.1080/10106049.2014.997308

Almeida, C. M., Batty, M., Monteiro, A. M. V., Câmara, G., Soares-Filho, B. S., Cerqueira, G. C., Pennachin, C. L., 2003. Stochastic cellular automata modeling of urban land use dynamics: Empirical development and estimation. Computers, Environment and Urban Systems 27 (5), 481–509.

Almeida, C. M., Gleriani, J. M., Castejon, E. F., SoaresFilho, B. S., 2008. Using neural networks and cellular automata for modelling intraurban landuse dynamics. International Journal of Geographical Information Science 22 (9), 943–963.

Ambrey, C. L., Fleming, C. M., Chan, A. Y.-C., 2014. Estimating the cost of air pollution in South East Queensland: An application of the life satisfaction non-market valuation approach. Ecological Economics 97, 172–181.
URL http://www.sciencedirect.com/science/article/pii/S0921800913003479

Arifin, S., Arifin, R., Pitts, D., Rahman, M., Nowreen, S., Madey, G., Collins, F., 2015. Landscape Epidemiology Modeling Using an Agent-Based Model and a Geographic

Information System. Land 4 (2), 378–412.
URL http://www.mdpi.com/2073-445X/4/2/378/

Arsanjani, J. J., Helbich, M., Kainz, W., Boloorani, A. D., 2012. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. International Journal of Applied Earth Observation and Geoinformation 21 (1), 265–275.

Batista, G. E. A. P. A., Prati, R. C., Monard, M. C., 2004. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter 6 (1), 20.
URL http://portal.acm.org/citation.cfm?doid=1007730.1007735

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z. J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Oudin, A., Forsberg, B., Modig, L., Havulinna, A. S., Lanki, T., Turunen, A., Oftedal, B., Nystad, W., Nafstad, P., De Faire, U., Pedersen, N. L., Östenson, C. G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K. T., Overvad, K., Ellermann, T., Eeftens, M., Peeters, P. H., Meliefste, K., Wang, M., Bueno-De-Mesquita, B., Sugiri, D., Krämer, U., Heinrich, J., De Hoogh, K., Key, T., Peters, A., Hampel, R., Concin, H., Nagel, G., Ineichen, A., Schaffner, E., Probst-Hensch, N., Künzli, N., Schindler, C., Schikowski, T., Adam, M., Phuleria, H., Vilier, A., Clavel-Chapelon, F., Declercq, C., Grioni, S., Krogh, V., Tsai, M. Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Katsoulis, M., Trichopoulou, A., Brunekreef, B., Hoek, G., 2014. Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. The Lancet 383 (9919), 785–795.

Ben-Gal, I., Dana, A., Shkolnik, N., Singer, G., 2014. Efficient construction of decision trees by the dual information distance method. Quality Technology & Quantitative Management 11 (1), 133–147.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling and Software 40, 1–20.

Berberoğlu, S., Akın, A., Clarke, K. C., 2016. Cellular automata modeling approaches to forecast urban growth for adana, turkey: A comparative approach. Landscape and Urban Planning 153, 11–27.

Bergstad, C. J., Gamble, A., Gärling, T., Hagman, O., Polk, M., Ettema, D., Friman, M., Olsson, L. E., 2011. Subjective well-being related to satisfaction with daily travel. Transportation 38 (1), 1–15.

Bian, Y., Zhang, L., Yang, J., Guo, X., Lei, M., 2015. Subjective Wellbeing of Chinese People: A Multifaceted View. Social Indicators Research 121 (1), 75–92.

Biancofiore, F., Verdecchia, M., Di Carlo, P., Tomassetti, B., Aruffo, E., Busilacchio, M., Bianco, S., Di Tommaso, S., Colangeli, C., 2015. Analysis of surface ozone using a recurrent neural network. Science of the Total Environment 514, 379–387.

Bickerstaff, K., Walker, G., 2001. Public understandings of air pollution: The 'localisation' of environmental risk. Global Environmental Change 11 (2), 133–145.

Bockstael, N. E., Irwin, E. G., et al., 2000. Economics and the land use-environment link. The international yearbook of environmental and resource economics 2001, 1–54.

Breiman, L., 1996. Bagging predictors. Machine learning 24 (2), 123–140.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., 1984. Classification and regression trees. CRC press.

Brereton, F., Clinch, J. P., Ferreira, S., 2008. Happiness, geography and the environment. Ecological Economics 65 (2), 386–396.

Camacho Olmedo, M. T., Paegelow, M., Mas, J. F., 2013. Interest in intermediate soft-classified maps in land change model validation: suitability versus transition potential. International Journal of Geographical Information Science 27 (12), 2343–2361. URL http://www.tandfonline.com/doi/abs/10.1080/13658816.2013.831867

Caruana, R., Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd international conference on Machine learning. ACM, pp. 161–168.

Chaudhuri, G., Clarke, K. C., 2013. The SLEUTH Land Use Change Model: A Review. The International Journal of Environmental Resources Research 1 (1), 89.

Chen, J., Davis, D. S., Wu, K., Dai, H., 2015. Life satisfaction in urbanizing China: The effect of city size and pathways to urban residency. Cities 49, 88–97.

Chen, Y., Li, X., Liu, X., Ai, B., 2014. Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation

strategy. International Journal of Geographical Information Science 28 (2), 234–255.
URL http://www.tandfonline.com/doi/abs/10.1080/13658816.2013.831868

Chen, Y., Li, X., Liu, X., Ai, B., Li, S., 2016. Capturing the varying effects of driving forces over time for the simulation of urban growth by using survival analysis and cellular automata. Landscape and Urban Planning 152, 59–71.
URL http://dx.doi.org/10.1016/j.landurbplan.2016.03.011

Chui, W. H., Wong, M. Y., 2016. Gender Differences in Happiness and Life Satisfaction Among Adolescents in Hong Kong: Relationships and Self-Concept. Social Indicators Research 125 (3), 1035–1051.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Clarke, K. C., Gaydos, L. J., 1998. Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. International Journal of Geographical Information Science 12 (7), 699–714.

Connolly, M., 2013. Some Like It Mild and Not Too Wet: The Influence of Weather on Subjective Well-Being. Journal of Happiness Studies 14 (2), 457–473.

Cooijmans, T., Ballas, N., Laurent, C., Gülçehre, Ç., Courville, A., 2016. Recurrent batch normalization. arXiv preprint arXiv:1603.09025.

Costanza, R., 1989. Model goodness of fit: a multiple resolution procedure. Ecological modelling 47 (3-4), 199–215.

Cox, D. D., Dean, T., 2014. Neural networks and neuroscience-inspired computer vision.

Cuñado, J., de Gracia, F. P., 2013. Environment and Happiness: New Evidence for Spain. Social Indicators Research 112 (3), 549–567.

Davis, J., Goadrich, M., 2006. The relationship between precision-recall and roc curves. In: Proceedings of the 23rd international conference on Machine learning. ACM, pp. 233–240.

Di Gregorio, A., Jansen, L. J., 1997. A new concept for a land cover classification system. In: Proceedings of the Earth Observation and Environmental Information 1997 Conference. Alexandria, Egypt. pp. 13–16.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., Speizer, F. E., 1993. An Association between Air Pollution and Mortality in Six U.S. Cities. New England Journal of Medicine 329 (24), 1753–1759.
URL http://www.nejm.org/doi/abs/10.1056/NEJM199312093292401

Du, G., Shin, K. J., Yuan, L., Managi, S., 2018. A comparative approach to modelling multiple urban land use changes using tree-based methods and cellular automata: the case of greater tokyo area. International Journal of Geographical Information Science 32 (4), 757–782.

Ferreira, S., Akay, A., Brereton, F., Cuñado, J., Martinsson, P., Moro, M., Ningal, T. F., 2013. Life satisfaction and air quality in Europe. Ecological Economics 88, 1–10.

Ferrer-i Carbonell, A., Frijters, P., 2004. How important is methodology for the estimates of the determinants of happiness? Economic Journal 114 (497), 641–659.

Ferrer-i Carbonell, A., Gowdy, J. M., 2007. Environmental degradation and happiness. Ecological Economics 60 (3), 509–516.

Friedman, J. H., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis 38 (4), 367–378.

Gao, L., Jia, G., Zhang, R., Che, H., Fu, C., Wang, T., Zhang, M., Jiang, H., Yan, P., 2011. Visual range trends in the Yangtze River Delta Region of China, 1981-2005. Journal of the Air and Waste Management Association 61 (8), 843–849.

Gao, S., Meng, X., Zhang, L., 2014. Fiscal Decentralization and Life Satisfaction: Evidence from Urban China. Social Indicators Research 119 (3), 1177–1194.

Gaubatz, P., 1999. China's urban transformation: Patterns and processes of morphological change in Beijing, Shanghai and Guangzhou. Urban Studies 36 (9), 1495–1521.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning 63 (1), 3–42.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) 9, 249–256.
URL http://machinelearning.wustl.edu/mlpapers/paper{_}files/AISTATS2010{_}GlorotB10.pdf

Gong, Z., Thill, J. C., Liu, W., 2015. ART-P-MAP Neural Networks Modeling of Land-Use Change: Accounting for Spatial Heterogeneity and Uncertainty. Geographical Analysis 47 (4), 376–409.

Graves, A., 2013. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.

Graves, A., Mohamed, A.-R., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (6), 6645–6649.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., Schmidhuber, J., 2017. Lstm: A search space odyssey. IEEE transactions on neural networks and learning systems.

Grekousis, G., Manetos, P., Photis, Y. N., 2013. Modeling urban evolution using neural networks, fuzzy logic and GIS: The case of the Athens metropolitan area. Cities 30 (1), 193–203.
URL http://dx.doi.org/10.1016/j.cities.2012.03.006

Guan, Q., Wang, L., Clarke, K. C., 2005. An Artificial-Neural-Network-based, Constrained CA Model for Simulating Urban Growth. Cartography and Geographic Information Science 32 (4), 369–380.
URL http://www.tandfonline.com/doi/abs/10.1559/152304005775194746

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

Hagoort, M., Geertman, S., Ottens, H., 2008. Spatial externalities, neighbourhood rules and ca land-use modelling. The Annals of Regional Science 42 (1), 39–56.

Han, L., Zhou, W., Li, W., Meshesha, D. T., Li, L., Zheng, M., 2015. Meteorological and urban landscape factors on severe air pollution in Beijing. Journal of the Air and Waste Management Association 65 (December), 782–787.
URL http://dx.doi.org/10.1080/10962247.2015.1007220

Hassan, A. M., Lee, H., 2015. Toward the sustainable development of urban areas: An overview of global trends in trials and policies. Land Use Policy 48, 199–212.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778.
URL http://ieeexplore.ieee.org/document/7780459/

Hefron, R. G., Borghetti, B. J., Christensen, J. C., Kabban, C. M., 2017. Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. Pattern Recognition Letters 94, 96–104.

Herold, M., Goldstein, N. C., Clarke, K. C., 2003. The spatiotemporal form of urban growth: Measurement, analysis and modeling. Remote Sensing of Environment 86 (3), 286–302.

Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological modelling 157 (2), 331–341.

Huang, B., Zhang, L., Wu, B., 2009. Spatiotemporal analysis of rural–urban land conversion. International Journal of Geographical Information Science 23 (3), 379–398.

Huang, J., Wu, S., Deng, S., 2016. Relative Income, Relative Assets, and Happiness in Urban China. Social Indicators Research 126 (3), 971–985.

Hunsaker, C. T., Levine, D. A., 1995. Hierarchical approaches to the study of water quality in rivers. BioScience 45 (3), 193–203.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. pp. 448–456.

Irwin, E. G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment 85 (1), 7–24.

Jansen, L. J., Veldkamp, T., 2012. Evaluation of the variation in semantic contents of class sets on modelling dynamics of land-use changes. International Journal of Geographical Information Science 26 (4), 717–746.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models.

Jiang, S., Lu, M., Sato, H., 2012. Identity, Inequality, and Happiness: Evidence from Urban China. World Development 40 (6), 1190–1200.

Kamusoko, C., Gamba, J., 2015. Simulating Urban Growth Using a Random Forest-Cellular Automata (RF-CA) Model. ISPRS International Journal of Geo-Information 4 (2), 447–470.

Keivani, R., 2010. A review of the main challenges to urban sustainability. International Journal of Urban Sustainable Development 1 (1-2), 5–16.

Kim, C. W., Phipps, T. T., Anselin, L., 2003. Measuring the benefits of air quality improvement: A spatial hedonic approach. Journal of Environmental Economics and Management 45 (1), 24–39.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems, 1–9.

Kuang, W., 2011. Simulating dynamic urban expansion at regional scale in Beijing-Tianjin-Tangshan Metropolitan Area. Journal of Geographical Sciences 21 (2), 317–330.

Landis, J. R., Koch, G. G., 1977. The measurement of observer agreement for categorical data. biometrics, 159–174.

Lange, T., 2010. Culture and life satisfaction in developed and less developed nations. Applied Economics Letters 17 (9), 901–906.

Lau, A. L. D., Cummins, R. A., McPherson, W., 2005. An investigation into the cross-cultural equivalence of the Personal Wellbeing Index. Social Indicators Research 72 (3), 403–430.

Le, J. A., El-Askary, H. M., Allali, M., Struppa, D. C., 2017. Application of recurrent neural networks for drought projections in California. Atmospheric Research 188, 100–106.

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., et al., 2009. Trends in the sources and sinks of carbon dioxide. Nature geoscience 2 (12), 831.

LeCun, Y. A., Bengio, Y., Hinton, G. E., 2015. Deep learning. Nature 521 (7553), 436–444.

Levinson, A., 2012. Valuing public goods using happiness data: The case of air quality. Journal of Public Economics 96 (9-10), 869–880.

Li, J., Heap, A. D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors.

Li, X., Gar-On Yeh, A., 2004. Data mining of cellular automata's transition rules. International Journal of Geographical Information Science 18 (8), 723–744.
URL http://www.tandfonline.com/doi/abs/10.1080/13658810410001705325

Li, X., Gong, P., Yu, L., Hu, T., 2017. A segment derived patch-based logistic cellular automata for urban growth modeling with heuristic rules. Computers, Environment and Urban Systems 65, 140–149.

Li, X., Lin, J., Chen, Y., Liu, X., Ai, B., 2013. Calibrating cellular automata based on landscape metrics by using genetic algorithms. International Journal of Geographical Information Science 27 (3), 594–613.
URL http://www.tandfonline.com/doi/abs/10.1080/13658816.2012.698391

Li, X., Liu, X., Gong, P., 2015a. Integrating ensemble-urban cellular automata model with an uncertainty map to improve the performance of a single model. International Journal of Geographical Information Science 29 (5), 762–785.
URL http://www.tandfonline.com/doi/full/10.1080/13658816.2014.997237

Li, X., Yeh, A. G., 2001. Calibration of cellular automata by using neural networks for the simulation of complex urban systems. Environment and Planning A 33 (8), 1445–1462.

Li, X., Yeh, A. G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science 16 (4), 323–343.
URL http://www.tandfonline.com/doi/abs/10.1080/13658810210137004

Li, Y., Zhao, H., Wu, Y., 2015b. Characteristics of particulate matter during haze and fog (Pollution) episodes over Northeast China, autumn 2013. Aerosol and Air Quality Research 15 (3), 853–864.

Li, Z., Folmer, H., Xue, J., 2014. To what extent does air pollution affect happiness? The case of the Jinchuan mining area, China. Ecological Economics 99, 88–99.

Liao, J., Tang, L., Shao, G., Su, X., Chen, D., Xu, T., 2016. Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. Environmental Modelling and Software 75, 163–175.
URL http://dx.doi.org/10.1016/j.envsoft.2015.10.014

Liao, P.-s., Shaw, D., Lin, Y.-m., 2015. Environmental Quality and Life Satisfaction: Subjective Versus Objective Measures of Air Quality. Social Indicators Research 124 (2), 599–616.
URL http://link.springer.com/10.1007/s11205-014-0799-z

Lipton, Z. C., Berkowitz, J., Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

Lisansky, J., 1986. Farming in an urbanizing environment: Agricultural land use conflicts and right to farm. Human Organization 45 (4), 363–371.

Liu, W., Seto, K. C., 2008. Using the ART-MMAP neural network to model and predict urban growth: A spatiotemporal data mining approach. Environment and Planning B: Planning and Design 35 (2), 296–317.

Liu, X., Ma, L., Li, X., Ai, B., Li, S., He, Z., 2014. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. International Journal of Geographical Information Science 28 (1), 148–163.

Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing 55 (5), 2486–2498.

Lubowski, R. N., Vesterby, M., Bucholtz, S., Baez, A., Roberts, M. J., et al., 2006. Major uses of land in the united states, 2002. Economic information bulletin 14.

Luechinger, S., 2009. Valuing Air Quality Using the Life Satisfaction Approach. The Economic Journal 119, 482–515.

Lyu, H., Lu, H., Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sensing 8 (6), 1–22.

Maaten, L. v. d., Hinton, G., 2008. Visualizing data using t-sne. Journal of machine learning research 9 (Nov), 2579–2605.

MacKerron, G., Mourato, S., 2009. Life satisfaction and air quality in London. Ecological Economics 68 (5), 1441–1453.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2016. Fully convolutional neural networks for remote sensing image classification. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). pp. 5071–5074.
URL http://ieeexplore.ieee.org/document/7730322/

Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N., 2015. Deep Supervised Learning for Hyperspectral Data Classification through Convolutional Neural Networks. IGARSS 2015. 2015 IEEE International Geoscience and Remote Sensing Symposium. Proceedings \, 4959–4962.

Mozumder, C., Tripathi, N. K., Losiri, C., 2016. Comparing three transition potential models: A case study of built-up transitions in North-East India. Computers, Environment and Urban Systems 59, 38–49.

Muller, N. Z., Mendelsohn, R., 2007. Measuring the damages of air pollution in the United States. Journal of Environmental Economics and Management 54 (1), 1–14.

Munroe, D. K., Southworth, J., Tucker, C. M., 2004. Modeling spatially and temporally complex land-cover change: The case of western honduras. The Professional Geographer 56 (4), 544–559.

Munshi, T., Zuidgeest, M., Brussel, M., van Maarseveen, M., 2014. Logistic regression and cellular automata-based modelling of retail, commercial and residential development in the city of Ahmedabad, India. Cities 39, 68–86.

Murphy, B., 2014. Pykrige: Development of a kriging toolkit for python. In: AGU Fall Meeting Abstracts.

Neelakantan, A., Vilnis, L., Le, Q. V., Sutskever, I., Kaiser, L., Kurach, K., Martens, J., 2015a. Adding gradient noise improves learning for very deep networks. arXiv preprint arXiv:1511.06807.

Neelakantan, A., Vilnis, L., Le, Q. V., Sutskever, I., Kaiser, L., Kurach, K., Martens, J., 2015b. Adding gradient noise improves learning for very deep networks. arXiv preprint arXiv:1511.06807.

Nogueira, K., Penatti, O. A., dos Santos, J. A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition 61, 539–556.

O'Leary, B. F., Lemke, L. D., 2014. Modeling spatiotemporal variability of intra-urban air pollutants in Detroit: A pragmatic approach. Atmospheric Environment 94, 417–427.

Olsson, L. E., Gärling, T., Ettema, D., Friman, M., Fujii, S., 2013. Happiness and Satisfaction with Work Commute. Social Indicators Research 111 (1), 255–263.

Opitz, D. W., Maclin, R., 1999. Popular ensemble methods: An empirical study. J. Artif. Intell. Res.(JAIR) 11, 169–198.

Orru, K., Orru, H., Maasikmets, M., Hendrikson, R., Ainsaar, M., 2016. Well-being and environmental quality: Does pollution affect life satisfaction? Quality of Life Research 25 (3), 699–705.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12 (Oct), 2825–2830.

Pérez-Vega, A., Mas, J. F., Ligmann-Zielinska, A., 2012. Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest. Environmental Modelling and Software 29 (1), 11–23.
URL `http://dx.doi.org/10.1016/j.envsoft.2011.09.011`

Pielke, R. A., Avissar, R., Raupach, M., Dolman, A. J., Zeng, X., Denning, A. S., et al., 1998. Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. Global change biology 4 (5), 461–475.

Pijanowski, B. C., Brown, D. G., Shellito, B. A., Manik, G. A., 2002. Using neural networks and gis to forecast land use changes: a land transformation model. Computers, environment and urban systems 26 (6), 553–575.

Pontius, R. G., Schneider, L. C., 2001. Land-cover change model validation by an roc method for the ipswich watershed, massachusetts, usa. Agriculture, Ecosystems & Environment 85 (1), 239–248.

Proto, E., Rustichini, A., 2015. Life satisfaction, income and personality. Journal of Economic Psychology 48, 17–32.

Puertas, O. L., Henríquez, C., Meza, F. J., 2014. Assessing spatial dynamics of urban growth using an integrated land use model. application in santiago metropolitan area, 2010–2045. Land use policy 38, 415–425.

Qayyum, A., Anwar, S. M., Awais, M., Majid, M., 2017. Medical image retrieval using deep convolutional neural network. Neurocomputing 266, 8–20.

Qian, Y., Qian, Z., 2015. Work, Family, and Gendered Happiness Among Married People in Urban China. Social Indicators Research 121 (1), 61–74.

Quinlan, J. R., et al., 1996. Bagging, boosting, and c4. 5. In: AAAI/IAAI, Vol. 1. pp. 725–730.

Rao, Y., Mei, L., Zhu, R., 2016. Happiness and Stock-Market Participation: Empirical Evidence from China. Journal of Happiness Studies 17 (1), 271–293.

Razavi, B. S., 2014. Predicting the Trend of Land Use Changes Using Artificial Neural Network and Markov Chain Model (Case Study : Kermanshah City). Research Journal of Environmental and Earth Sciences 6 (4), 215–226.

Rehdanz, K., Maddison, D., 2005. Climate and happiness. Ecological Economics 52 (1), 111–125.

Rey, S. J., Anselin, L., 2010. Pysal: A python library of spatial analytical methods. Handbook of applied spatial analysis, 175–193.

Rohde, R. A., Muller, R. A., 2015. Air pollution in China: Mapping of concentrations and sources. PLoS ONE 10 (8).

Rossetti, L. A., Sergio dos Anjos, F. P., De Almeida, C. M., 2013. Cellular automata-based spatial dynamic modeling for analyzing urban land use change. In: Urban Remote Sensing Event (JURSE), 2013 Joint. IEEE, pp. 1–4.

Roy Chowdhury, P. K., Maithani, S., 2014. Modelling urban growth in the Indo-Gangetic plain using nighttime OLS data and cellular automata. International Journal of Applied Earth Observation and Geoinformation 33 (1), 155–165.
URL http://dx.doi.org/10.1016/j.jag.2014.04.009

Samuel, A. L., 1959. Some studies in machine learning using the game of checkers. IBM Journal of research and development 3 (3), 210–229.

Santé, I., García, A. M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landscape and Urban Planning 96 (2), 108–122.

Sekulova, F., van den Bergh, J. C. J. M., 2013. Climate change, income and happiness: An empirical study for Barcelona. Global Environmental Change 23 (6), 1467–1475.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition, 1–14.
URL http://arxiv.org/abs/1409.1556

Smyth, R., Mishra, V., Qian, X., 2008. The Environment and Well-Being in Urban China. Ecological Economics 68 (1-2), 547–555.

Smyth, R., Nielsen, I., Zhai, Q., Liu, T., Liu, Y., Tang, C., Wang, Z., Wang, Z., Zhang, J., 2011. A study of the impact of environmental surroundings on personal well-being in urban China using a multi-item well-being indicator. Population and Environment 32 (4), 353–375.

Soares-Filho, B. S., Coutinho Cerqueira, G., Lopes Pennachin, C., 2002. DINAMICA - A stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecological Modelling 154 (3), 217–235.

Taylor, P., Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. International Journal of Geographical Information (June), 37–41.

Tayyebi, A., Pijanowski, B. C., Tayyebi, A. H., 2011. An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran. Landscape and Urban Planning 100 (1-2), 35–44.

Van Den Berg, B., Ferrer-I-Carbonell, A., 2007. Monetary valuation of informal care: The well-being valuation method. Health Economics 16 (11), 1227–1244.

Van Ootegem, L., Verhofstadt, E., 2016. Well-being, life satisfaction and capabilities of flood disaster victims. Environmental Impact Assessment Review 57, 134–138.

van Praag, B. M., Baarsma, B. E., 2005. Using happiness surveys to value intangibles: The case of airport noise. Economic Journal 115 (500), 224–246.

van Vliet, J., Hagen-Zanker, A., Hurkens, J., van Delden, H., 2013a. A fuzzy set approach to assess the predictive accuracy of land use simulations. Ecological Modelling 261-262, 32–42.
URL http://dx.doi.org/10.1016/j.ecolmodel.2013.03.019

van Vliet, J., Naus, N., van Lammeren, R. J., Bregt, A. K., Hurkens, J., van Delden, H., 2013b. Measuring the neighbourhood effect to calibrate land use models. Computers, Environment and Urban Systems 41, 55–64.

Verburg, P. H., de Nijs, T. C. M., van Eck, J. R., Visser, H., de Jong, K., 2004. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems 28 (6), 667–690.

Verstegen, J. A., Karssenberg, D., van der Hilst, F., Faaij, A. P., 2014. Identifying a land use change cellular automaton by Bayesian data assimilation. Environmental Modelling & Software 53, 121–136.
URL http://linkinghub.elsevier.com/retrieve/pii/S1364815213002909

Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A., 2008. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on Machine learning - ICML '08. pp. 1096–1103.
URL http://portal.acm.org/citation.cfm?doid=1390156.1390294

Walters, D. R., 2007. Designing Community: charrettes, master plans and form-based codes. Routledge.

Wang, H., Mullahy, J., 2006. Willingness to pay for reducing fatal risk by improving air quality: A contingent valuation study in Chongqing, China. Science of the Total Environment 367 (1), 50–57.

Wang, J., Mountrakis, G., 2011. Developing a multi-network urbanization model: A case study of urban growth in Denver, Colorado. International Journal of Geographical Information Science 25 (2), 229–253.
URL http://www.tandfonline.com/doi/abs/10.1080/13658810903473213

Wang, L., Scott, K. A., Xu, L., Clausi, D. A., 2016. Sea Ice Concentration Estimation during Melt from Dual-Pol SAR Scenes Using Deep Convolutional Neural Networks: A Case Study. IEEE Transactions on Geoscience and Remote Sensing 54 (8), 4524–4533.

Wang, P., Pan, J., Luo, Z., 2015. The Impact of Income Inequality on Individual Happiness: Evidence from China. Social Indicators Research 121 (2), 413–435.

Wang, T., Jiang, F., Deng, J., Shen, Y., Fu, Q., Wang, Q., Fu, Y., Xu, J., Zhang, D., 2012. Urban air quality and regional haze weather forecast for Yangtze River Delta region. Atmospheric Environment 58, 70–83.

Welsch, H., 2002. Preferences over Prosperity and Pollution: Environmental Valuation based on Happiness Surveys. Kyklos 55 (4), 473–494.
URL http://doi.wiley.com/10.1111/1467-6435.00198

Welsch, H., 2006. Environment and happiness: Valuation of air pollution using life satisfaction data. Ecological Economics 58 (4), 801–813.

Welsch, H., 2007. Environmental welfare analysis: A life satisfaction approach. Ecological Economics 62 (3-4), 544–551.

Welsch, H., 2009. Implications of happiness research for environmental economics.

White, R., Engelen, G., 1993. Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns. Environment and Planning A 25 (8), 1175–1199.
URL http://journals.sagepub.com/doi/10.1068/a251175

White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, environment and urban systems 24 (5), 383–400.

Wu, J., et al., 2008. Land use changes: Economic, social, and environmental impacts. Choices 23 (4), 6–10.

Wu, J. J., 2008. Making the case for landscape ecology an effective approach to urban sustainability. Landscape journal 27 (1), 41–50.

Wu, Y., Zhu, J., 2016. When Are People Unhappy? Corruption Experience, Environment, and Life Satisfaction in Mainland China. Journal of Happiness Studies 17 (3), 1125–1147.

Xie, C., Huang, B., Claramunt, C., Chandramouli, C., 2005. Spatial logistic regression and gis to model rural-urban land conversion. In: Proceedings of PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: frameworks, models and applications. University of Toronto, pp. 12–15.

Xu, J., Li, J., 2016. Tax Payment, Social Contribution for Pollution Prevention and Happiness. PROBLEMY EKOROZWOJU 11 (1), 59–64.

Xu, X., Du, Z., Zhang, H., 2016. Integrating the system dynamic and cellular automata models to predict land use and land cover change. International Journal of Applied Earth Observation and Geoinformation 52, 568–579.
URL http://dx.doi.org/10.1016/j.jag.2016.07.022

Yang, Q., Li, X., Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. Computers & Geosciences 34 (6), 592–602.
URL http://www.sciencedirect.com/science/article/pii/S0098300407002087

Zaremba, W., Sutskever, I., Vinyals, O., 2014. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

Zhang, H., Wang, S., Hao, J., Wang, X., Wang, S., Chai, F., Li, M., 2016. Air pollution and control action in Beijing.

Zhang, X., Zhang, X., Chen, X., 2017. Valuing Air Quality Using Happiness Data: The Case of China. Ecological Economics 137, 29–36.

Zhang, Z. X., 2007. China is moving away the pattern of "develop first and then treat the pollution". Energy Policy 35 (7), 3547–3549.

Zhao, H., Che, H., Zhang, X., Ma, Y., Wang, Y., Wang, H., Wang, Y., 2013. Characteristics of visibility and particulate matter (PM) in an urban area of Northeast China. Atmospheric Pollution Research 4 (4), 427–434.
URL http://linkinghub.elsevier.com/retrieve/pii/S1309104215303640

Zhao, W., 2012. Economic inequality, status perceptions, and subjective well-being in China's transitional economy. Research in Social Stratification and Mobility 30 (4), 433–450.

Zheng, H. W., Shen, G. Q., Wang, H., Hong, J., 2015. Simulating land use change in urban renewal areas: A case study in Hong Kong. Habitat International 46, 23–34.
URL http://dx.doi.org/10.1016/j.habitatint.2014.10.008