

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

The Relative Timing of Human Migration and Land-Cover and Land-Use Change — An Evaluation of Northern Taiwan from 1990 to 2015

Permalink

<https://escholarship.org/uc/item/8t5432st>

Author

Shih, Hsiao-chien

Publication Date

2020

Peer reviewed|Thesis/dissertation

SAN DIEGO STATE UNIVERSITY AND
UNIVERSITY OF CALIFORNIA

Santa Barbara

The Relative Timing of Human Migration and Land-Cover and Land-Use Change — An
Evaluation of Northern Taiwan from 1990 to 2015

A Dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Geography

by

Hsiao-chien Shih

Committee in charge:

Professor Douglas A. Stow, Chair

Professor John R. Weeks

Professor Dar A. Roberts

Professor Konstadinos G. Goulias

June 2020

The dissertation of Hsiao-chien Shih is approved.

Konstadinos G. Goulias

Dar A. Roberts

John R. Weeks

Douglas A. Stow, Committee Chair

May 2020

The Relative Timing of Human Migration and Land-Cover and Land-Use Change —

An Evaluation of Northern Taiwan from 1990 to 2015

Copyright © 2020

by

Hsiao-chien Shih

Dedicated to my grandparents, my mother, and Yi-ting.

ACKNOWLEDGEMENTS

This study was funded by Yin Chin Foundation of U.S.A., STUF United Fund Inc., the Long Jen-Yi Travel fund, William & Vivian Finch Scholarship, and a doctoral stipend through San Diego State University.

I would like to thank for the committee members of my dissertation, Drs. Stow, Weeks, Roberts, and Goulias along with other professors. They helped me conceptualize the ideas and revise the content of the dissertation, and I cannot finish this dissertation without their help. I am really appreciated Dr. Kou-chen Chang's suggestion of land-use data collection and Dr. Carvalho's advice on time series analysis.

VITA OF HSIAO-CHIEN SHIH
May 2020

EDUCATION

Bachelor of Arts in Geography, National Taiwan Normal University, June 2010
Master of Science in Geography, San Diego State University, May 2015
Doctor of Philosophy in Geography, San Diego State University and University of California, Santa Barbara, June 2020 (expected)

PROFESSIONAL EMPLOYMENT

2013-20: Teaching Associate, Department of Geography, San Diego State University

PUBLICATIONS

Shih, H., Stow, D. A., & Tsai, Y. H. (2019). Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. *International journal of remote sensing*, 40(4), 1248-1274.

Toure, S., Stow, D., Shih, H., Weeks, J., Lopez-Carr, D. (2018). Land cover and land use change analysis using multi-spatial resolution data and object-based image analysis. *Remote Sensing of Environment*, 210, 259-268.

Coulter, L. L., Stow, D. A., Tsai, Y. H., Ibanez, N., Shih, H., Kerr, A., ... & Mensah, F. (2016). Classification and assessment of land cover and land use change in southern Ghana using dense stacks of Landsat 7 ETM+ imagery. *Remote Sensing of Environment*, 184, 396-409.

Stow, D. A., Weeks, J. R., Shih, H., Coulter, L. L., Johnson, H., Tsai, Y. H., ... & Mensah, F. (2016). Inter-regional pattern of urbanization in southern Ghana in the first decade of the new millennium. *Applied Geography*, 71, 32-43.

Shih, H., Stow, D. A., Weeks, J. R., & Coulter, L. L. (2016). Determining the type and starting time of land cover and land use change in southern Ghana based on discrete analysis of dense Landsat image time series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 2064-2073.

Toure, S., Stow, D., Shih, H., Coulter, L., Weeks, J., Engstrom, R., & Sandborn, A. (2016). An object-based temporal inversion approach to urban land use change analysis. *Remote Sensing Letters*, 7(5), 503-512.

Stow, D. A., Shih, H., & Coulter, L. L. (2014). Discrete classification approach to land cover and land use change identification based on Landsat image time sequences. *Remote sensing letters*, 5(10), 922-931.

Chang, K., Tian, Y., & Shih, H. (2012). Using Multi-temporal and PCA&NDVI to Improve the Accuracy and Integrity of Land Cover Classification, *Journal of Geographical Research*, (57), 49-60.

AWARDS

The Honorable Mention Award, 5th National High School Geography Olympiad Competition, Kaohsiung, 2006

The Best Student Poster, 7th Taipei International Digital Earth Symposium, Taipei, 2010

National Science Council funded Undergraduate Research Project, Taipei, 2010

Alvena Storm Memorial Scholarship, San Diego State University, 2015

Presidential Graduate Research Fellowship, San Diego State University, 2015

William & Vivian Finch Scholarship, San Diego State University, 2016

Southern California Scholarship, Yin Chin Foundation of U.S.A., STUF United Fund Inc., Los Angeles, 2017

Long Jen-Yi Travel fund, San Diego State University, 2018

Department Citizenship award, San Diego State University, 2019

FIELDS OF STUDY

Remote Sensing, Image processing and analysis, Geographic Information Science, Statistics, Machine learning, Demographical research

ABSTRACT

The Relative Timing of Human Migration and Land Cover and Land Use Change —
An Evaluation of Northern Taiwan from 1990 to 2015

by

Hsiao-chien Shih

Urban land expansion can be driven by and drive population growth, but determining cause-effect relationships of urban land expansion and population growth is challenging due to the temporal resolution of decadal censuses for most countries. The relative timing of urban land expansion and population change was explored based on a case study for north Taiwan from 1990 to 2015. Data on urban land expansion were derived from a dense time series of Landsat satellite imagery, and population change from annual population registers at the district level.

A novel algorithm for identifying the starting time of urban expansion was developed and tested based on dense time series of Vegetation-Impervious-Soil (V-I-S) proportion maps derived from Landsat surface reflectance imagery. The time of urbanization estimated by logistic regression was determined from the Impervious cover time series. The identified location and estimated time for newly urbanized lands were generally accurate, with 80% of urban expansion estimated within ± 2.4 years.

The random forest (RF) classification routine was applied to Landsat image inputs for detailed urban land-use classification. Input features were extracted from a 2015 Landsat image dataset, and the top 10 most contributing features for RF classifiers were the six spectral wavebands, gray level co-occurrence matrix (GLCM) homogeneity of V-I-S, and Vegetation temporal variation. The Landsat image was tested with different input features, and the land-use maps derived from the top 10 features yielded the most accurate map. Areal extents of annual urban land uses (Residential, Employment, and Transportation corridor categories) were derived by applying a conditional statement based on the urbanization date of the urban expansion map and overlaying on the 2015 land-use map.

The relative timing and general relationship between urban land-use and population dynamics was explored at the district level. Linear regression was run to understand the general relationship between population and land use. The relative timing between population growth and land-use change was estimated with lagged correlation. Residential land use and change was most related to population and change. Population growth occurred 2.5 years later than Residential land expansion based on the median time lag.

TABLE OF CONTENTS

Chapter 1. Introduction.....	1
1.1.Overview.....	1
1.2. Study Area and Study Period.....	5
Chapter 2. Estimating the Starting Time and Identifying the Type of Urbanization based on Dense Time Series of Landsat-Derived Vegetation-Impervious-Soil (V-I-S) Maps	8
2.1. Introduction.....	8
2.1.1. Mapping Vegetation-Impervious-Soil in Urban Environments	9
2.1.2. Urbanization Identification with Time Series Landsat.....	10
2.1.3. Purposes of Study	12
2.2. Methods	13
2.2.1. Data.....	13
2.2.2. Image Preprocessing.....	14
2.2.3. V-I-S Maps Time Series Generation	14
2.2.4. Identification of Urbanization.....	16
2.3. Results.....	21
2.3.1. Endmember Selection.....	21
2.3.2. Accuracies of Individual V-I-S Maps.....	25
2.3.3. Parameters for RMSE Thresholding and Savitzky-Golay Filtering	27
2.3.4. Accuracy of Urbanization Identification	27
2.4. Discussion and Conclusion.....	29
2.4.1. Automated V-I-S Map Time Series Generation	29

2.4.2. Urbanization Identification with V-I-S Map Time Series	30
2.4.3. V-I-S Trajectories of LCLUC.....	30
2.4.4. Spatial-temporal Trends of Urbanization	31
2.4.5. Conclusion.....	32
 Chapter 3. From Land Cover to Land use: Applying Random Forest on Landsat Imagery for Detailed Urban Land-Use Mapping.....	 36
3.1. Introduction.....	36
3.2. Methods	39
3.2.1. Data.....	39
3.2.2. Spatial and Temporal Feature Generation	42
3.2.3. Feature Ranking and Selection	43
3.2.4. Land-Use Classification and Change.....	43
3.2.5. Accuracy Assessment	46
3.3. Results.....	47
3.3.1. Input Features for Land-Use Classification.....	47
3.3.2. Random Forest Feature Ranking	47
3.3.3. Accuracy of Land-Use Maps	50
3.3.4. Accuracy of Land-Use Change Maps.....	52
3.3.5. Refinement of Land-Use and Land-Use Change Map	52
3.4. Discussion and Conclusion.....	55
3.4.1. Features for Land-Use Extraction.....	55
3.4.2. Land-Use Change Map Comparison	56
3.4.3. Challenge of Inferring Land Use from Land Cover	57

3.4.4. Spatial-temporal Trend of Land-Use Change.....	59
3.4.5. Conclusion.....	60
Chapter 4. The Relative Timing of Population Growth and Land Use Change – A Case	
Study of North Taiwan from 1990 to 2015	63
4.1. Introduction.....	63
4.1.1. Urbanization	63
4.1.2. Urban Population Growth.....	64
4.1.3. Impacts on Land from Urbanization.....	65
4.1.4. Hypothetical Processes of Urbanization.....	66
4.1.5. Purposes of Study	67
4.2. Data and Methods.....	68
4.2.1. Data.....	68
4.2.2. Relationship between Population and Land Use	70
4.2.3. Relative Timing between Population Growth and Land-Use Expansion	
.....	70
4.2.4. Hypothetical Processes of Urban Growth.....	71
4.3. Results.....	72
4.3.1. Spatial-temporal Trends in Population and Land Use.....	72
4.3.2. Relationship Between Land Use and Population and Their Change...	77
4.3.3. Model Population Change based on Land-Use Expansion.....	77
4.3.4. Hypothesis Testing	84
4.4. Discussion and Conclusion.....	86

4.4.1. Relative Timing and Relationships Between Population Change and Land-Use Change	86
4.4.2. Urbanization Statuses of Metropolitan Areas	88
4.4.3. Limitation, Challenges, and Follow-on Research.....	90
4.4.4. Conclusion	91
Chapter 5. Conclusion	93
5.1. Key Contributions and Findings.....	93
5.2. Summary.....	96
5.3. Future Research	99
References.....	102

Chapter 1. Introduction

1.1. Overview

Land-cover and land-use change (LCLUC) is a global phenomenon that justifiably receives a lot of research attention (Bhatta, 2010; Lambin et al., 2001), as it does in this doctoral dissertation. Land cover is the material (e.g. vegetation, impervious surface, and soil) of the Earth surface, and land-cover change involves the conversion from one material type to another. Land use is how a parcel of land is used by human beings, so the type of land use is primarily determined by human activities (e.g. residential, industrial, and commercial). Multiple forms of LCLUC occur in urban areas and their environs, and urban expansion (e.g. urban sprawl) is a form of LCLUC that occurs globally (Seto et al., 2011). When urban population grows due to rural-urban migration and natural increase (i.e. more people are born instead of dying), the demand for housing and jobs within the urban areas increases, particularly for developing countries. Thus, new residential, industrial, and infrastructural developments appear at rural-urban fringes (e.g. suburb and peri-urban area) to relieve the growing urban population, and cities become larger over time. In addition, buildings in old town areas age and become too small or dilapidated. These buildings are torn down in exchange for new and often larger buildings. Therefore, the variety of LCLUC needs to be documented and studied for comprehensive urban planning and unraveling complicated impacts between human and their living environments.

LCLUC in urban areas has received great attention because such change has direct and obvious impacts on humans (Foley et al., 2005). Urban areas are mainly covered by three general types of land-cover materials, vegetation, impervious surface, and soil (V-I-S)

(Ridd, 1995). Human activities tend to occur in facilities and on infrastructure composed of impervious surfaces, so mapping and monitoring impervious surfaces associated with the built environment and their changes can provide useful information about human activities.

Land uses are associated with characteristic assemblages of land-cover types, so land-use types may be inferred based on land-cover composition and distributions (Cihlar & Jansen, 2001). For example, fractions of Vegetation-Impervious-Soil (V-I-S) cover are typically different between residential and commercial parcels. However, Cihlar & Jensen (2001) and Jensen (2009) pointed out the land-cover compositions can vary even for the same type of land use. In some major cities of developing countries, low socioeconomic status (SES) residential areas may have less urban vegetation cover compared to high SES residential areas based on the analysis on VHR image (Stoler et al., 2012; Stow et al., 2013). Depending on the spatial resolution of remotely sensed data, identifying the type of land use from land-cover proportions is worthy of testing and analysis.

The reasons for why LCLUC occurs varies for different places and over time (Meyer and Turner, 1992), especially for urban LCLUC. For my study, the focus is on urban growth, which includes land conversion for residential, commercial, industrial, and other activities instead of agriculture and other prime sector use. Bhatta (2010) summarizes most of the drivers of urban growth, such as population growth, economic growth, industrialization, and transportation development. Population growth stimulates new housing developments, so new residential buildings are built to accommodate the increasing demand for housing. Economic growth and industrialization prompt new factories and shopping malls, so more jobs and products are created for people's living and trading. Transportation developments enable people to move faster, so more people who live in the

suburbs can commute to cities and enjoy the service provided by cities while living in a less crowded environment. Natural drivers (e.g. hurricanes and earthquakes) sometimes massively modify the urban landscape, although usually in a temporary manner.

Population growth (either caused by natural increase or in-migration) is considered to be the most connected phenomenon to urban land-use change, which can be a driver of and be driven by LCLUC. For example, Powell et al. (2008a) monitored and analyzed the size of 10 cities in Rondonia, Brazil, relative to their human populations. A strong log-log relationship was found between city size and population, which implied that the size of city highly reflects the size of human population. Stow et al. (2016) mapped and monitored built areas at the district level in southeastern Ghana, and a strong relationship was quantified between built area and population density. Thus, the area of built land cover is a plausible demographic indicator, especially for the human population size. However, some researchers point out that the relationship of LCLUC and population is not unidirectional, but rather, reciprocal (Lambin et al., 2001; Sherbinin et al., 2007), so the processes and impacts between urban expansion and population change should be analyzed at relatively fine spatial and temporal scales.

In this dissertation, the relative timing of urban land-use change and population change was explored based on a case study for cities of northern Taiwan. Data on land-cover and land-use change were derived from dense time series of Landsat satellite multispectral image data, and annual population at fine spatial and temporal scale was acquired from local governmental agencies. A hypothetical processes of land-use change and population growth is explored based on the time series of land-use and population data. This study informs how

urban land-use change occurs in accordance with rural-urban migration across space and time.

The reasons why north Taiwan was chosen as study area are as follows. First, population data in Taiwan are available to the public at relatively fine spatial (district level) and temporal (annual) scales. To understand the relative timing of urban LCLUC and population change, such fine-scale data are necessary. Second, processes of urban expansion and urban renewal occur in Taiwan, which enabled me to study the capability of identifying and discerning both processes with remotely sensed data. Third, the urban population continued to grow under the condition of low fertility rates (from a total fertility rate of 1.72 children per woman in 1990 to 1.18 in 2015), which enabled me to study the effects on LCLUC mainly from rural-urban or urban-urban migration. Finally, Taiwan is governed by a democratic system, so land tenure and associated development are less interfered with by governmental agencies, except for land-use zoning. These reasons make Taiwan a suitable study area for exploring the relative timing of urban land-use change and human migration.

This dissertation contains three research components, and the components are closely tied together. The goal of the first component is to develop and test a semi-automatic approach to determining the type and onset time of urbanization based on dense time series of V-I-S proportion maps derived from Landsat imagery in Chapter 2. Spectral mixture analysis was implemented to estimate V-I-S proportions for each pixel in each date of Landsat imagery, and then a novel algorithm was developed and tested to identify urbanization based on V-I-S time series at the pixel level. For the second component, I tested the ability of detailed urban land-use classification based on land-cover data derived from the first component as input features by implementing machine learning techniques in

Chapter 3. Machine learning techniques were applied to understand the importance of Landsat reflectance and their derivatives for detailed urban land-use classification, and the capability of machine learning classifiers for mapping land-use and land-use change in the urban environment was also assessed. The purpose of the last component was to explore the relative timing of urban land-use change and population change at the district level based on the land-use data derived from the previous components and the fine-scale population data in Chapter 4. The relationships between population and land use were assessed with linear regression for various urban land uses, and the relative timing between population change and land-use change were determined by lagged correlation analysis. The key findings and contribution of the three research components are summarized in Chapter 5 along with future research.

1.2. Study Area and Study Period

Taiwan is an island located in East Asia on the Pacific Rim. It is separated by the Taiwan Strait from mainland China in the west. Japan is located to the north of Taiwan, separated by East China Sea. Philippines is located south of Taiwan, separated by the Bushi Channel, and the east coast of Taiwan faces the Pacific Ocean. The location of Taiwan island is strategic in the aspects of military deployment, international trade, and logistics, so the Dutch, Spanish, and Japanese have historically occupied the island as a colony. Currently, Taiwan is a territory of the Republic of China (ROC) which lost mainland China in the Chinese civil war, at the end of which the ROC government migrated to Taiwan in 1949. Mainland China is now ruled by the People's Republic of China (PRC), while the ROC is a collection of islands, the largest of which is Taiwan.

The study area for this dissertation research is located within the northern region of Taiwan (Figure 1), coinciding with the World Reference System -2 (WRS-2) coordinate is path 117, row 43 for Landsat data. The study period ranges from 1990 to 2015, which corresponds to the fastest growth of urban population for East Asia and Pacific region in Figure 1. According to the Köppen climate classification, this area has a Humid subtropical climate (Cfa) with warm temperatures and no dry season. The general physiography in the western part of the study area is generally low and flat, while it is steep and hilly in the middle and eastern areas. Forests and crops are the predominant type of land cover in the middle and the eastern part of the island, and manmade impervious surfaces (i.e. settlement) and moist soil (mostly from fallow) cover the flat areas along the west coast.

The study area contains five cities and three counties that share the same administrative level, which includes Taipei, New Taipei, Keelung, Taoyuan, Hsinchu cities, Hsinchu, Miaoli, and Yilan counties (Table 1). Taipei City is the capital of ROC, and it is also the political, financial, and transportation centers of the island. Multiple science industrial parks are scattered across the study area, especially in Hsinchu City, Taipei City, and New Taipei City, which provide high-tech products for the world. Most of the built environment is concentrated in the cores of three metropolitan areas and their vicinities, while crops and forest cover most areas of Miaoli and Yilan Counties. More jobs are created in the three metropolitan areas, which attracts the residents of Miaoli, Yilan, and other counties. According to the 2010 census report, over 11 million people (about half of the total population on the island) live in the study area, and the most populous county is New Taipei City (4.0 million), followed by Taipei (2.6 million) and Taoyuan Cities (2.1 million).

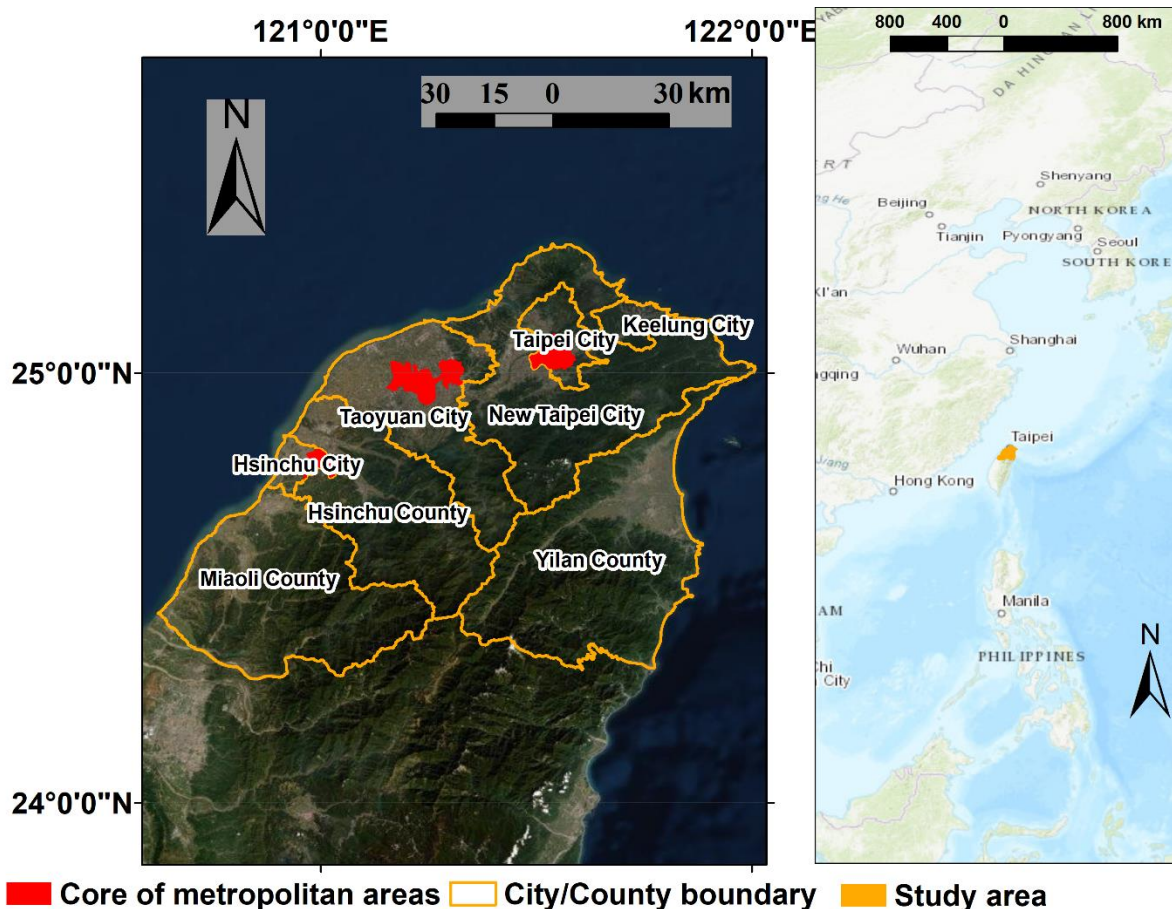


Figure 1. Study area for north region of Taiwan along with the cores of metropolitan areas within the study area.

Table 1. Metropolitan and rural areas with correspondent cities and counties along with 2010 census population

Metropolitan / rural area	City and counties	2010 census population	Metropolitan core
Taipei metropolitan area	Taipei City	2,655,515	Datong, Zhongshan, Songshan, Xinyi, Daan, Zhongzheng, and Wanhua districts
	New Taipei City	4,054,467	
	Keelung City	381,809	
Taoyuan metropolitan area	Taoyuan City	2,190,342	Taoyuan and Zhongli districts
Hsinchu metropolitan area	Hsinchu City	476,273	East district
	Hsinchu County	522,163	
Miaoli rural area	Miaoli County	530,339	---
Yilan rural area	Yilan County	426,975	---

Chapter 2. Estimating the Starting Time and Identifying the Type of Urbanization based on Dense Time Series of Landsat-Derived Vegetation-Impervious-Soil (V-I-S) Maps¹

2.1. Introduction

Land-cover and land-use change (LCLUC) in urban areas has received great attention because such change has direct and obvious impacts on humans (Bhatta, 2010; Lambin et al., 2001). Multiple forms of LCLUC occur in urban areas and their environs, and Urban Expansion (e.g. urban sprawl) is a form of LCLUC that occurs globally (Seto et al., 2011). Buildings in older urbanized areas age and become dilapidated or are too small to meet growing population pressures. Such buildings may be demolished and replaced by new buildings for higher density residential and other urban land uses. Urban areas are mainly covered by three general types of land-cover materials, Vegetation (V), Impervious surface (I), and Soil (S), according to the V-I-S model of Ridd (1995). Monitoring impervious surfaces in the built environment can provide useful information about human activities, such as the distribution of human populations and activities (Olorunfemi, 1984; Stow et al., 2016).

The ultimate goal is to understand the relative timing and relationship between LCLUC and human migration, based on LCLU data from Landsat time series and population data produced by national government organizations. This study is the first step towards that

¹ This chapter was previously published in International Journal of Applied Earth Observation and Geoinformation: Shih, H., Stow, D. A., Tsai, Y. M., & Roberts, D. A. (2020). Estimating the starting time and identifying the type of urbanization based on dense time series of Landsat-derived Vegetation-Impervious-Soil (VIS) maps—A case study of North Taiwan from 1990 to 2015. International Journal of Applied Earth Observation and Geoinformation, 85, 101987.

objective. In this chapter, I develop and test procedures for generating moderate spatial and temporal resolution land-cover data that will enable estimation of the time of onset of land-use change and provide input data for land-use classification and change analysis in subsequent Chapter 3. The relative timing and relationship between population change and LCLUC will also be explored in Chapter 4.

2.1.1. Mapping Vegetation-Impervious-Soil in Urban Environments

Remote sensing has been widely used for mapping and monitoring of land-use types, and Landsat data are the most readily and historically available and economical moderate spatial resolution imagery that supports such mapping and monitoring. The reasonable spatial resolution and moderate revisit frequency of Landsat image data are the reasons why such imagery has been widely used for mapping impervious surfaces and monitoring urbanization (Lu & Weng, 2005).

Urban areas are complex environments and mapping impervious fraction accurately can be challenging, especially with moderate spatial resolution imagery (e.g. Landsat imagery). Urban environments are more complicated than the natural environment, due to material complexity, inter-material spectral similarity (e.g. similar spectra between tile roof and soil), intra-material variability (e.g. different spectra between asphalt and concrete), illumination effects (resulting from three-dimensional effect and non-Lambertian surface from buildings), and aging of impervious materials and surfaces (Herold et al., 2003a). Also, atmospheric effects (from gases and aerosols) are more of an issue in urban environments due to air pollution from factories, power plants and automobiles (Saraswat et al., 2017).

Spectral Mixture Analysis (SMA) has been employed with moderate spatial resolution satellite data to monitor impervious cover for urban environments (Wu and Murray, 2003;

Powell et al., 2007; Franke et al., 2009). Ridd (1995) originally proposed the V-I-S model to decompose urban environment, and the V-I-S model was tested of SMA on SPOT imagery to quantify urban composition. To account for the variability of fractional impervious cover, Wu and Murray (2003) used both high-albedo and low-albedo impervious endmembers along with vegetation and soil endmembers to model land-cover composition. Wu (2004) further proposed a normalized SMA (NSMA) approach to reducing the spectral variability between high-albedo and low-albedo impervious surfaces. Multiple Endmember Spectral Mixture Analysis (MESMA: Roberts et al., 1998) has been proven for successfully mapping impervious cover within urban environments that have high spectral-radiometric variability (Powell et al., 2007). With MESMA, multiple endmembers are chosen to model the material abundance V-I-S along with amount of shadow for each pixel. The endmember model having the lowest RMSE is chosen for estimating the proportion of V-I-S.

2.1.2. Urbanization Identification with Time Series Landsat

Both image classification and SMA have been applied to multiple remotely sensed datasets to generate time series of maps depicting built land use and/or impervious surface covers. Hung and Weng (2018) applied a decision tree classifier to dense time series of Landsat imagery to map annual impervious in Hanoi, Vietnam from 1988 to 2015, but did not assess the accuracy of estimating the time of land-use change. Shih et al. (2016) mapped built expansion and the timing of expansion for Southeastern Ghana by utilizing 41 dates of Landsat imagery. A maximum likelihood classifier was applied to each image to map LCLU, and then a post-classification comparison of LCLU maps was applied to identify types of LCLUC. They found that 80% of the built expansion was accurately detected, and

the average start of built expansion was estimated within 2 years from the actual start of change.

Powell et al. (2008b) incorporated SMA to map impervious change from 1972 to 2006 based on 13 dates of Landsat images (including those from the Multi-spectral Scanner [MSS], Thematic Mapper [TM] and Enhanced Thematic Mapper [ETM+] sensors). They then applied a set of spatial and temporal rules to minimize the confusion between impervious and soil. However, the analysis was limited to the total amount of impervious change from year to year, computed as the percentage of impervious cover between the beginning and end dates of the 34-year study period.

Zhu and Wookcock (2014) developed the Continuous Change Detection and Classification (CCDC) approach to identify land-cover change with Landsat surface reflectance images by incorporating Fourier transform and random forest classifier. Liu et al. (2019) integrated temporal change rules into the CCDC to improve the accuracy of impervious change in spatial and temporal domains and tested the improved approach for Nanchang, China. Deng and Zhu (2018) incorporated a modified CCDC and random forest regressor for continuous subpixel monitoring (CSM) of urban LCLUC with dense time series of Landsat from 2000 to 2014. Instead of using the original surface reflectance of Landsat wavebands, they used coefficients from Fourier transforms of the time series of all reflectance wavebands as input features to the random forest regressor to estimate impervious cover proportions. However, only impervious trajectories were used, while the potential of Vegetation and Soil was not explored.

2.1.3. Purpose of Study

The purpose of this chapter is to: (1) develop and test an approach to semi-automatically mapping urban V-I-S cover based on SMA applied to a Landsat image time series and monitor urbanization, and (2) assess how accurately the time when transition to urban land use began to occur for areas undergoing urbanization. The specific study area is the northern part of Taiwan and the study period is from 1990 to 2015, which corresponds to a period of tremendous urban population growth in this area. A major challenge is to identify and minimize noise sources (cloud, cloud shadow, and SLC-off) existing in Landsat image series, which creates “No Data” gaps in the resultant maps. I focus on how well V-I-S trajectories derived from dense Landsat image stacks can be used to accurately monitor dynamics within and at the periphery of an urban environment. The novelty of this study is to develop, apply, and validate an urbanization identification algorithm based on observed V-I-S trajectories. The primary objective is to accurately estimate the timing and map the distribution of urban-related land-use change associated with Urban Expansion (new urban development at the periphery of urban areas) and Urban Renewal (replacement or densification of extant urban land use) based on a dense Landsat V-I-S time series derived by semi-automated SMA. The accuracy of V-I-S proportion maps derived from a semi-automatic SMA procedure was assessed as well as the accuracy of estimating start time of Urban Renewal and Expansion derived from the dense time series V-I-S proportion maps. Temporal trajectories of various urban LCLUC types were analyzed.

2.2. Methods

2.2.1. Data

Landsat series surface reflectance data (i.e. TM, ETM+, and Operational Land Imager [OLI] imagery) are the main source of imagery used in this study. The study area is located within the northern region of Taiwan, coinciding with the World Reference System-2 (WRS-2) coordinate path 117, row 43 for Landsat data. All Landsat images between 1987 to 2017 were searched on EarthExplorer, and images with less than 30% cloud cover over land were downloaded. Under this cloud cover criterion, 298 dates of Landsat surface reflectance images were accessed. These surface reflectance data sets were generated by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS; Masek et al., 2006) and Landsat Surface Reflectance Code (LaSRC) algorithms (Table 2). Landsat products dated before 1990 and after 2015 were only used to construct a complete V-I-S time series without extrapolating V-I-S beyond the two end points of time (i.e. 1990 and 2015) for accurately identifying urban land-use change close to the two end points when only using imagery between 1990 and 2015.

Reference data were acquired from two sources, Google Earth Very High Resolution (VHR) imagery and historical aerial photos. All available Google VHR images (e.g. Quickbird, Ikonos, Worldview series) from 1999 to 2015 were used as reference data. Two image frames of metric aerial photos captured between 1990 and 1999 were acquired from the Taiwanese Aerial Survey Office of Forestry Bureau, and used as a source for reference data generation. A shapefile containing 107 district polygons was downloaded from a Taiwanese governmental website (<https://data.gov.tw/>) and used as a mask for subsequent image processing steps.

Table 2. Summary of utilized Landsat scenes by satellite system

Sensor	Number of scenes	First date of imagery	Last date of imagery
Landsat 4 TM	1	4-Aug-1992	-
Landsat 5 TM	150	9-Apr-1987	5-Nov-2011
Landsat 7 ETM+	120	8-Aug-1999	29-Nov-2017
Landsat 8 OLI	27	16-Apr-2013	23-Dec-2017

2.2.2. Image Preprocessing

Image preprocessing was conducted on the downloaded Landsat reflectance images before estimating V-I-S at the sub-pixel level. The reflectance data contain an image layer called pixel QA, which was utilized to screen out cloud, cloud shadow, and SLC-off (for ETM+) pixels. For each date of Landsat, pixels marked as noise by the pixel QA were deemed as No Data pixels, and the district-level polygons were used to delineate the boundary of study area for subsequent V-I-S estimation.

2.2.3. V-I-S Maps Time Series Generation

Since shadows, particularly those from buildings in urban areas, vary over time, shadow effects on V-I-S fraction estimates need to be minimized. This was achieved through normalization based on the NSMA proposed by Wu (2004). Each waveband of normalized reflectance (\overline{R}_b) from the base image was generated with the following equation, $\overline{R}_b =$

$$\frac{R_b}{\frac{1}{N} \sum_{b=1}^N R_b} \times 100, \text{ where } R_b \text{ is the original reflectance and } N \text{ is the number of total reflectance}$$

wavebands. The normalized spectra were modeled with a constrained linear SMA to estimate V-I-S proportion for each pixel: $\overline{R}_b = \sum_{i=1}^N \overline{f}_i \overline{R}_{i,b} + e_b$ under the conditions of $\sum_{i=1}^N \overline{f}_i = 1$ and $\overline{f}_i \geq 0$, where $\overline{R}_{i,b}$ is the normalized reflectance of endmember i in waveband b ; \overline{f}_i is the fraction of the endmember i ; e_b is the residual.

Google Earth VHR imagery and Landsat imagery for 20-Jun-1990 and 13-Dec-2017 aided in the selection of endmembers for the 298 dates of Landsat surface reflectance images. Based on visual image interpretation, 600 sample pixels (i.e. 200 pixels for each land cover) were manually digitized for different manifestations of vegetation, impervious and soil cover. Only samples that were deemed stable over the study period were selected. The sample pixels were distributed over the entire study area to reduce the chance that clouds cover the sample pixels that were only collected in a small area. Specifically, samples were selected for the following cover/material types: forest, mangrove, and riparian herbaceous zones for Vegetation; clay tile, cement, steel roof, and asphalt pavement for Impervious; and baseball field, moist soil, and exposed landslide areas for Soil.

Visual inspection of endmember spectra and K-means clustering were applied to a single date of Landsat image to finalize endmember selection. A cloud-free Landsat 5 TM image acquired at 6-Feb-2011 was selected as a base image for endmember finalization. To determine the parameter K for each land cover, potential endmembers were randomly separated into two subgroups for each land-cover type: a subgroup containing 150 samples, and the other containing the remaining 50 samples. The spectral means (centers in feature space) of the clusters were derived from the K-means clustering algorithm out of the 150 subgroups, and then the smaller subgroup was used to assess the spectral means in RMSE. The RMSE was used to determine the appropriate K for each land cover. The K for each land cover was tested and allowed to range from one to three, with the sum of Ks from vegetation, impervious, and soil less than and equal to six. The processing of endmember separation, K-means clustering, and RMSE assessment was repeated 30 times. All possible

combinations of Ks for vegetation, impervious, and soil were tested, and the most accurate model with was selected to unmix the remaining 297 dates of Landsat imagery.

Accuracy of V-I-S fractions was assessed for all V-I-S maps derived from the endmember auto-selection processes based on VHR imagery as the source of reference data. Similar accuracy assessment approaches were used in past SMA studies (e.g. Wu and Murray [2003] and Deng and Wu [2013]). The district polygon-based shapefile was used as a stratum to distribute reference sample locations, and 107 sample units were delineated for stratified sampling. Five of the 107 sample units were removed because they underwent land-cover change. The sample size was determined based on a trade-off between statistical rigor and practical limitations (Wu and Murray, 2003). A 90 x 90 m square sampling unit was used (corresponding to 3 x 3 Landsat pixels to minimize misregistration effects), and polygons of V-I-S within each unit were manually digitized and areal proportions summarized in percentages. Systematic error (SE), mean absolute error (MAE), and root-mean-square error (RMSE) metrics were generated using these reference data.

2.2.4. Identification of Urbanization

The V-I-S time series maps were used to identify land-use changes related to urbanization on a per-pixel basis, namely Urban Expansion and Urban Renewal. V-I-S maps with high RMSE from SMA were removed before constructing the time series. No Data pixels (i.e. due to cloud, cloud shadow, and SLC-off) were filled by linear interpolation of modeled V-I-S estimates of the nearest before and after dates. Then, the interpolated V-I-S time series was smoothed with a Savitzky-Golay filter. Multiple forms of V-I-S trajectories are possible depending on of RMSE and Savitzky-Golay filter parameters, and RMSEs ranging from 10 to 35%, temporal windows from a month to two years and polynomial

orders from 1 to 5 were tested. Based on the smoothed V-I-S trajectories derived from No Change samples, I applied a rule that trajectories should be constructed with at least 100 dates of V-I-S maps and processed with narrow smoothing window to reduce most noise and preserve change signals at the same time.

The algorithm for identifying the onset of urbanization by primarily using Impervious fraction changes is depicted in Figure 2. A pixel is deemed as ‘No Change’ if the dynamic range of an Impervious time series (i.e. $I_{\max} - I_{\min}$) does not exceed the maximal dynamic range of V-I-S time series derived from all No Change samples, where I_{\max} and I_{\min} are the maximum and the minimum of Impervious time series from 1987 to 2017, respectively. Otherwise, multiple tests are applied to the time series of each pixel based on the average Impervious fraction for the first (1987 to 1992) and last (2013 to 2017) five years. The average Impervious fractions for the first and the last five years were designed to capture the Impervious cover at the beginning and the end of the study period, which were used for rapidly applying appropriate rules for per-pixel urbanization identification. The Urban Renewal test is applied when the average Impervious fraction is higher than 50% in both periods (shown in the lower left part of Figure 2), while the Urban Expansion test is applied when Impervious fraction is over 50% in the later period and lower in the earlier period (shown in the lower right part of Figure 2). The entire time series of Vegetation, Impervious, and Soil from 1987 to 2017 was used for the two urbanization tests, and the identified urbanization pixels between 1990 to 2015 were preserved.

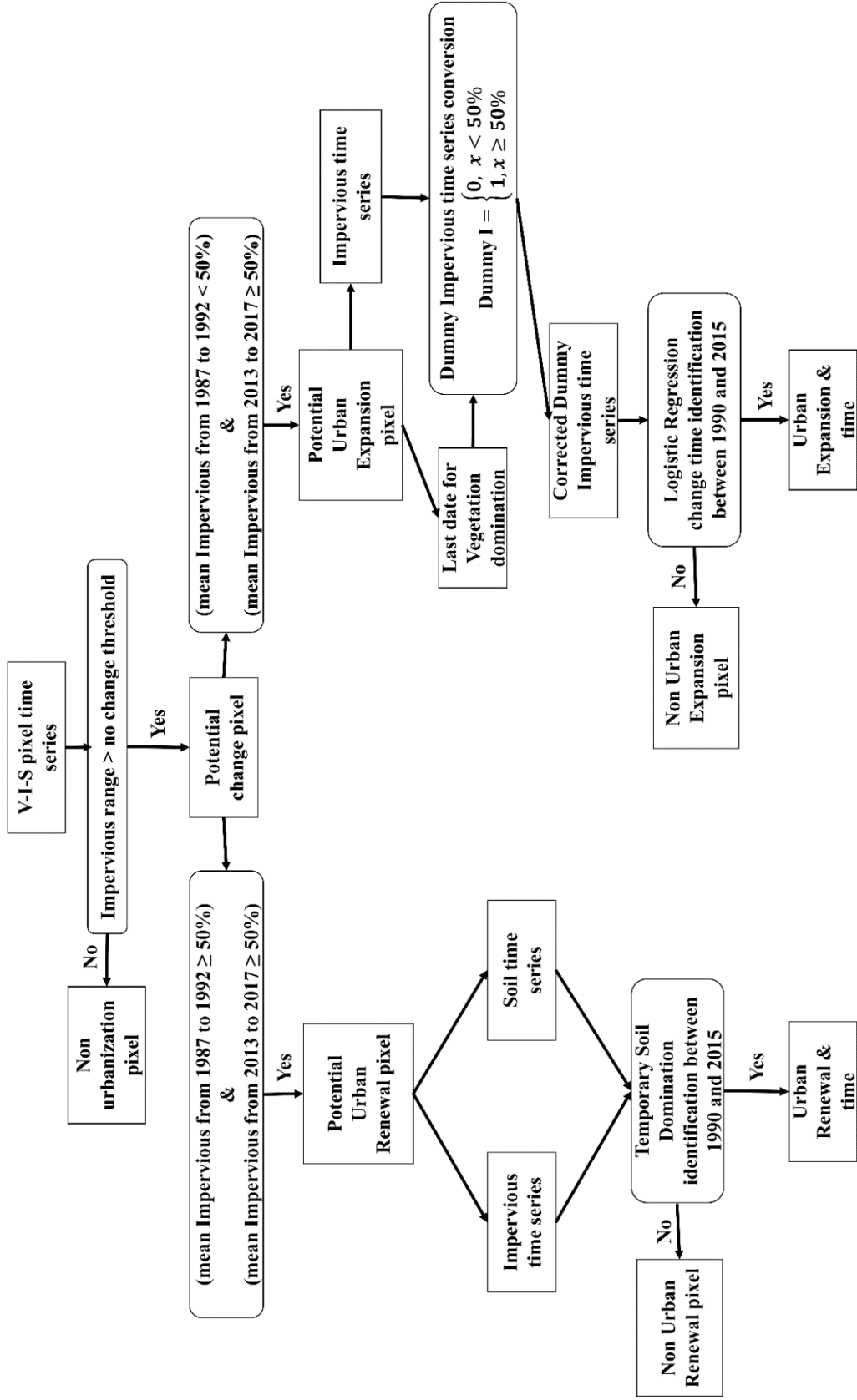


Figure 2. Flow chart depicting pixel-level urbanization identification algorithm.

Two hypothetical V-I-S trajectories are shown in Figure 3 for Urban Expansion and Renewal, respectively; these trajectories are based on theoretical considerations of urban development processes and guided by trajectories for urban land-use change samples identified by visual interpretation of the VHR imagery. For Urban Expansion, Impervious cover rapidly increases (e.g. from 0 to at least 50%) followed by a rapid decrease of Vegetation within a year (shown in Figure 3a). The hypothetical trajectory of Impervious resembles the sigmoidal shape of logistic regression, so a logistic regression model is used to estimate the time of Urban Expansion. For each pixel, the smoothed Impervious trajectory is converted into dummy variable: 1 for Impervious > 50%, and 0 for others. However, confusion between impervious and soil cover types can occur in V-I-S modeling for agricultural areas, and Vegetation or Impervious can become the dominant land cover in turns over time. To overcome this confusion, the last date when Vegetation is the dominant is recorded, and any date when Impervious is dominant before the date is assigned 0. Time (t) is used as the only one independent variable to model the dummy variable Impervious, and then a fitted logit model is used to identify whether the Impervious change at each point of time. The time of Urban Expansion onset is estimated when the modeled Impervious exceeds 50%, while the modeled Impervious trajectory is No Change when the modeled Impervious does not increase over 50%.

For a pixel trajectory to be classified as Urban Renewal (e.g. gentrification) change, the proportion of Impervious is 100% throughout most of the study period. However, a sudden reduction in Impervious fraction and concomitant increase of soil occurs when buildings and other structures are demolished, followed by a rapid increase in Impervious and decrease of Soil (shown in Figure 3b). Also, the proportion of Vegetation may increase after the

construction due to urban planning laws for requiring landscaping. The time of change is estimated as when the Impervious time series reaches its lowest point and the Soil reaches its highest point concurrently. In addition, the lowest Impervious proportion must be lower than the highest Soil proportion.

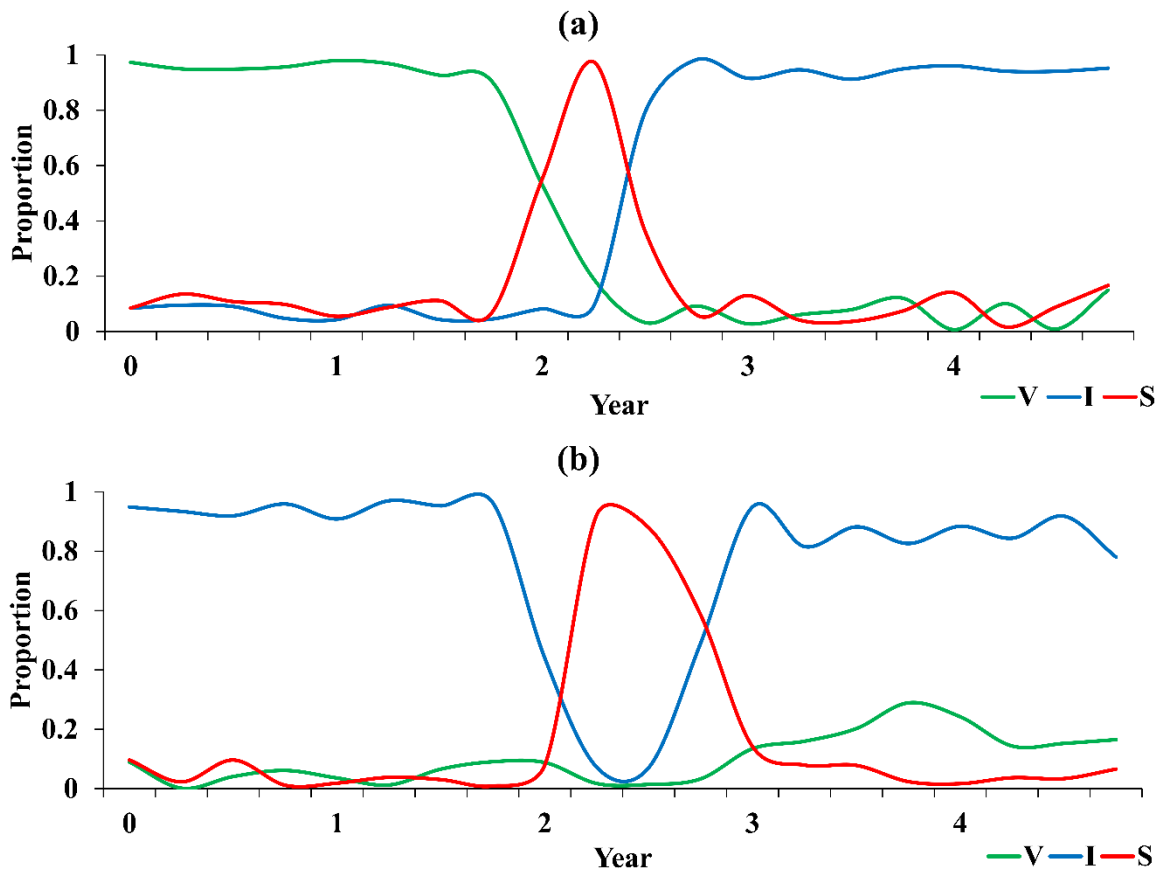


Figure 3. Hypothetical pixel-based V-I-S trajectories of (a) Urban Expansion and (b) Urban Renewal.

Accuracy of identified urbanization was assessed from an independent set of samples. With the guidance of Google Earth and historical aerial VHR imagery, 60 reference samples of urbanized areas were manually selected because the change areas are smaller than the no-change areas. Thirty of the change samples belongs to Urban Expansion examples, and the remainder 30 belongs to Urban Renewal examples. As with the No Change reference

samples, the sampling unit is a 90 by 90 m square, and V-I-S proportions over time were interpreted and digitized based on two to five dates of VHR imagery depending on its availability and coverage. Additionally, the initial date of change to Impervious was recorded according to the change date observed on the VHR images for each change sample. Change samples were determined based on whether Impervious increased or decreased at least 30% within the sampling units, so portions of the units may have experienced no change in land cover. SE, MAE, and RMSE for time intervals between the reference change time and the Landsat-derived change time were used to estimate the time effectiveness of identifying time of change.

2.3. Results

2.3.1. Endmember Selection

Spectral reflectance signature plots for all potential endmembers are shown in Figure 4. The difference between maximum and the minimum reflectance from the potential endmembers for V-I-S is much greater in all wavebands for the raw reflectance data (Figure 4a) than for the normalized reflectance (Figure 4b), which implies that the normalization is useful for reducing the intra-class spectral variation. Although Vegetation is separable from Impervious and Soil, Impervious and Soil reflectance overlap substantially, indicating that unmixing pixels with mixtures of Impervious and Soil is challenging. This confusion is common and has been observed in other studies of LCLUC based on SMA for urban and urbanizing regions (Wu, 2004; Powell et al., 2008a).

To finalize the endmembers used in NSMA, K-means clustering and resampling were repeatedly tested with various settings of the cluster parameter K, and the most and least

accurate results (assessed by RMSE) are shown in Table 3. For the three-endmember NSMA model, a single Vegetation, Impervious, and Soil endmember was used, whereas for the six-endmember model one Vegetation, two Soil, and three Impervious endmembers were used. RMSEs are lowest for the six-endmember NSMA model, especially for Impervious. For the six-endmember model, confusion between Impervious and Soil is lower than the three-endmember model.

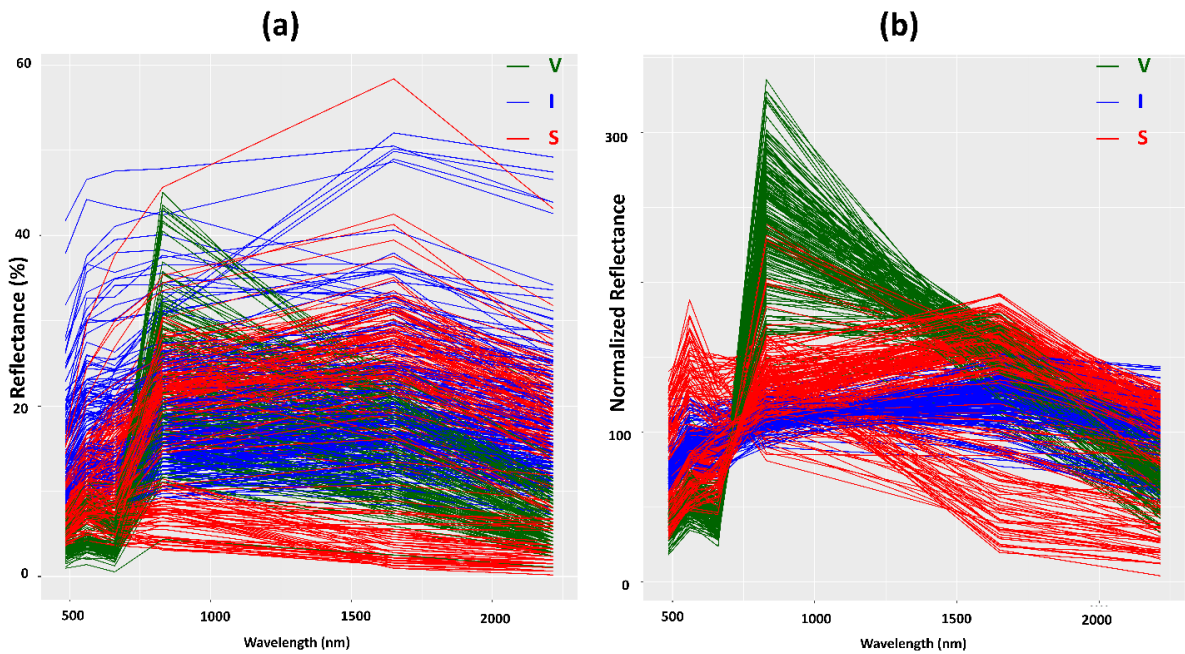


Figure 4. Landsat spectral reflectance plots of potential V-I-S endmembers. (a) Raw spectral reflectance. (b) Normalized spectral reflectance.

Table 3. Average RMSE of repetitive resampling for various settings of K-means clustering and normalized SMA

Total used Endmembers for NSMA	Endmember clusters details	Vegetation	Impervious	Soil
3	Number of clusters for Vegetation = 1 Number of clusters for Impervious = 1 Number of clusters for Soil = 1	18.78%	45.81%	54.83%
6	Number of clusters for Vegetation = 1 Number of clusters for Impervious = 3 Number of clusters for Soil = 2	18.38%	28.57%	42.66%

To understand the association between endmembers from the six-endmember model and the land-cover materials, the 600 sample pixels and their associated clusters (i.e. endmember) were visually examined on the VHR imagery. The spectral center of each cluster is shown in Figure 5. The endmembers for Vegetation, Impervious, and Soil are significantly different from each other in normalized reflectance. Dark, moist soil is separable from bright, dry soil. For Impervious, one endmember is the spectral center of white metal buildings and infrastructure (e.g. oil tanks), while the other two impervious endmembers are the spectral centers of multiple impervious materials that differ in normalized reflectance including concrete, metal, and asphalt. Compared to the multiple material impervious 2, the multiple material impervious 1 has high reflectance from 500 to 1000 nm and lower reflectance from 1000 to 2300 nm. The two complex impervious endmembers derived from the K-means clustering account for the greater variation in impervious spectra, substantially boosting the accuracy of estimated impervious fractions. In addition, more than three materials for Impervious within a pixel (e.g. metal roof, concrete, and asphalt) were observed and confirmed during the field trip and on the VHR imagery. Thus, a six-endmember model was used to unmix the 298 dates of Landsat imagery. However, three dates of imagery were found to have insufficient endmembers, so 295 dates of V-I-S maps were used for change identification procedures.

The accuracies of individual V-I-S maps for each date of Landsat imagery are shown in Figure 6 and Table 4. RMSEs for most dates of V-I-S maps range from 20 to 30%, and the maxima of RMSE for V-I-S are all over 50% respectively, which implies that the estimated V-I-S proportions are highly inaccurate for certain dates. The minima of RMSE for V-I-S are about 10%. On average, RMSE for V-I-S are 27.21%, 23.68%, and 27.94%,

respectively. The standard deviation of RMSE is lowest for Impervious fraction, suggesting that modeled Impervious is most stable and accurate. RMSE for V-I-S increases after 2003, likely related to the Scan Line Corrector failure (SLC-off) of the Landsat 7 ETM+ sensor that started that year. The total reference samples used in the assessment are also shown in Figure 6, and many of the less accurate V-I-S maps are associated with low numbers of reference samples.

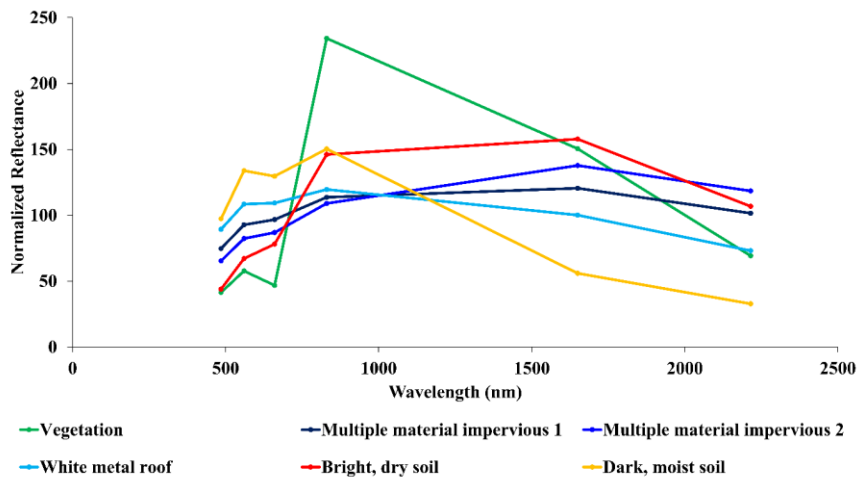


Figure 5. Six endmembers derived from K-means clustering, including a vegetation endmember, three impervious endmembers (i.e. multiple material impervious 1, 2, and white metal roof), and two soil endmembers (i.e. bright, dry soil and dark moist soil). The two multiple material impervious endmembers have different spectra though the two impervious endmembers are the spectral centers of similar impervious materials (e.g. asphalt, metal roof, and concrete).

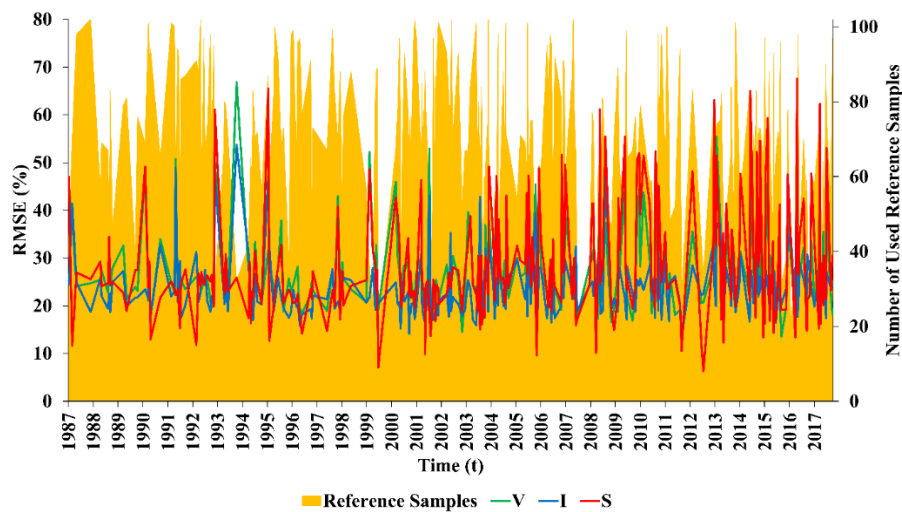


Figure 6. RMSE and used number of reference samples used for each V-I-S map derived from Landsat imagery.

Table 4. RMSE statistics over all dates of V-I-S maps

Statistics	Vegetation	Impervious	Soil
Minimum	12.41%	12.85%	6.31%
Median	25.12%	22.57%	24.62%
Average	27.21%	23.68%	27.94%
Maximum	66.81%	53.68%	67.56%
Standard deviation	8.35%	5.87%	11.53%

2.3.2. Accuracies of Individual V-I-S Maps

SMA model statistics for eight dates of V-I-S maps were selected for further analysis because 1) more than 70 reference samples were available for each map, and 2) these V-I-S maps were least affected by clouds (shown in Table 5). RMSE for Impervious in seven dates of V-I-S maps (except for ETM+ 7 28-Dec-2010) are between 19 to 25%. In terms of SE, Vegetation and Impervious fractions are underestimated in the eight V-I-S maps, while Soil was overestimated.

Table 5. Accuracy of eight selected-dates of V-I-S maps

	RMSE (%)			MAE (%)			SE (%)		
	V	I	S	V	I	S	V	I	S
TM 5 20-Jun-1990	21.4	21.3	24.6	17.0	15.9	18.5	-10.7	-6.1	16.9
TM 5 4-Feb-1993	24.3	19.1	19.7	18.6	14.6	15.2	-14.6	-0.0	14.6
ETM+ 7 6-Mar-2001	22.2	16.0	19.3	15.7	11.8	14.8	-10.8	-2.6	13.5
TM 5 19-Apr-2004	21.2	18.3	22.6	15.8	14.8	18.3	-10.8	-5.6	16.5
ETM+ 7 28-Dec-2010	34.0	28.9	45.4	29.3	21.1	40.4	-29.0	-8.6	37.7
TM 5 6-Feb-2011	21.5	15.8	24.5	15.6	11.3	18.1	-12.2	-4.8	17.1
OLI 8 29-Jan-2014	21.4	18.7	20.4	16.2	14.3	15.3	-10.9	-2.4	13.4
OLI 8 23-Dec-2017	19.0	24.8	31.2	13.4	18.9	25.6	-8.5	-16.7	25.3

Table 6. Average RMSE of V-I-S over four Landsat sensors

Sensor	V (%)	I (%)	S (%)
Landsat 4 TM	23.2	26.7	27.0
Landsat 5 TM	26.0	22.6	24.6
Landsat 7 ETM+	29.7	25.2	33.1
Landsat 8 OLI	22.8	22.4	23.3

Table 7. Average RMSE of all V-I-S maps over four seasons

Season	V (%)	I (%)	S (%)
Spring (Mar., Apr., May)	31.5	25.5	33.7
Summer (Jun., Jul., Aug.)	26.3	23.3	25.9
Fall (Sep., Oct., Nov.)	26.4	23.3	27.0
Winter (Dec., Jan., Feb.)	25.8	23.0	27.0

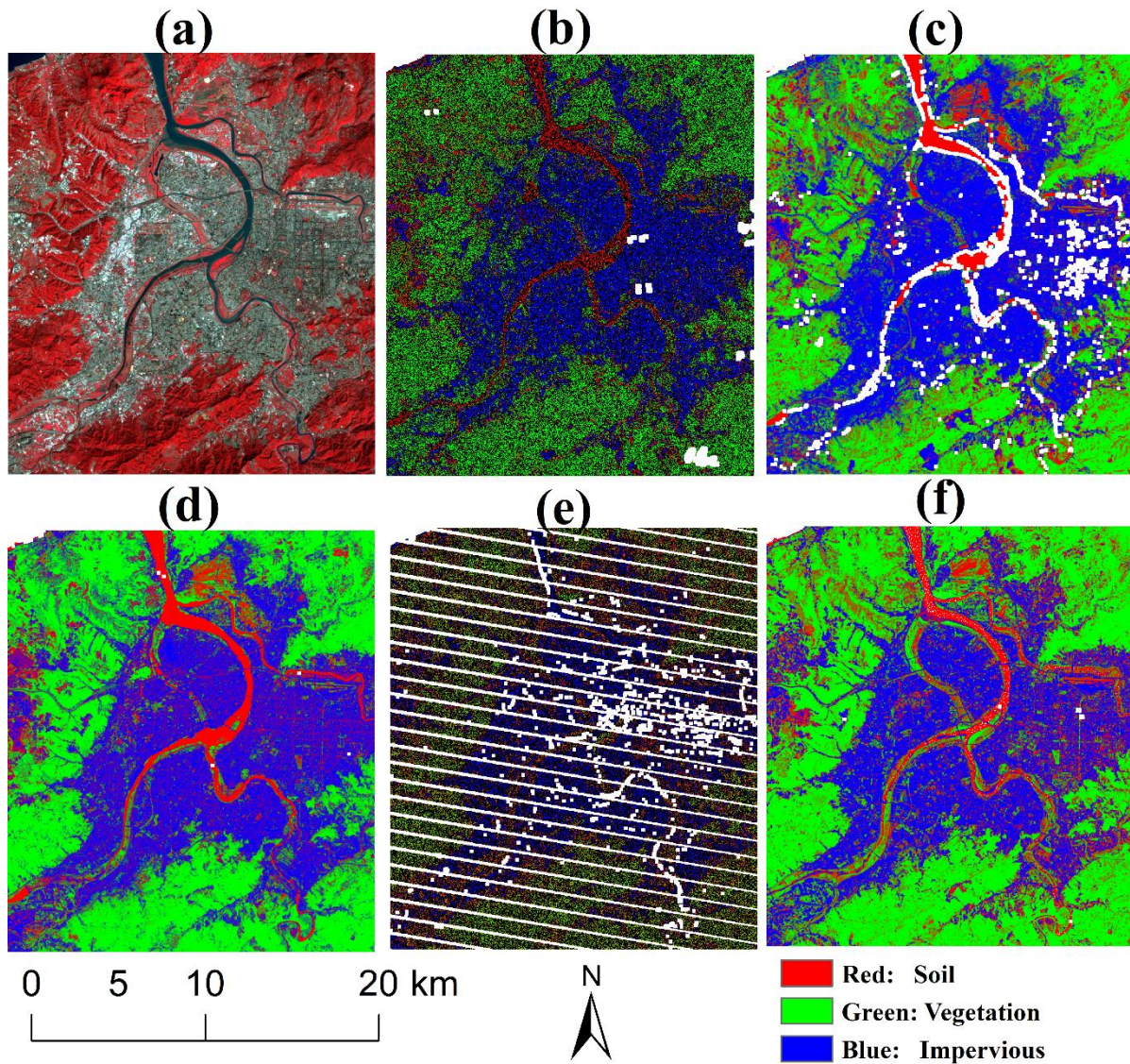


Figure 7. Selected dates of V-I-S maps along with a date of Landsat image. (a) Landsat 5 image acquired on 6-Feb-2011 in color inferred. (b) V-I-S map derived from Landsat 5 image acquired on 20-Jun-1990. (c) V-I-S map derived from Landsat 5 image acquired on 4-Feb-1993. (d) V-I-S map derived from Landsat 7 acquired on 6-Mar-2001. (e) V-I-S map derived from Landsat 7 acquired on 28-Dec-2010 with obvious SLC-off effects. (f) V-I-S map derived from Landsat 8 acquired on 23-Dec-2017.

2.3.3. Parameters for RMSE Thresholding and Savitzky-Golay Filtering

After removing V-I-S maps having high RMSE, multiple parameters for Savitzky-Golay filtering were tested, and the results are shown in Table 8. To preserve true Impervious change signals while reducing noise from SMA modeling errors, we selected 27% as the RMSE threshold for the over 100 V-I-S maps used for LCLUC analysis, one year for the filtering window, and a first-order polynomial to construct noise-reduced V-I-S time series. Based on these parameters, the maximum dynamic ranges of V-I-S time series is 22.3%, 18.7%, and 13.9%, respectively, so a pixel time series was tested for identifying urbanization (i.e. the diagram of Figure 1) if the maximum dynamic range of Impervious time series is larger than 23%.

Table 8. Maximum dynamic ranges of V-I-S time series for No Change samples

Parameters settings	Number of V-I-S map dates used	V (%)	I (%)	S (%)
Original time series	295	100.0	100.0	100.0
Single date, low RMSE filtered series (RMSE < 27%)	132	100.0	100.0	100.0
Savitzky-Golay smoothed (window = 0.5 year, polynomial = 1) series	295	38.0	34.8	25.0
Single date, low RMSE filtered (RMSE < 27%) + Savitzky-Golay smoothed (window = 1 year, polynomial = 1) series	132	22.3	18.7	13.9
Single date, low RMSE filtered (RMSE < 27%) + Savitzky-Golay smoothed (window = 1.2 year, polynomial = 1) series	132	19.4	16.1	12.2

2.3.4. Accuracy of Urbanization Identification

An urban land-use change map derived from per-pixel urbanization identification was assessed with the reference samples. The number of pixels that were identified as urban change pixels in the change and No Change sample units is shown in Figure 8. Twenty-eight of the 30 Urban Expansion samples (93%) involved two or more identified pixels and 24

(80%) involved three or more contiguous identified pixels. Nine Urban Renewal samples were identified: nine (30%) by one or more pixel, and four (13%) with two or more. For the nine detected samples, only four samples (11%) were correctly labeled as Urban Renewal, while five samples (16%) were mis-labeled as Urban Expansion. The reason why fewer than three pixels were identified for some samples is because only a portion of the 90 x 90 m sample areas experience urban densification. To avoid false positive identification, urban change samples with three or more contiguous pixels identifying urban change were preserved for later spatio-temporal analysis of urbanization.

The accuracy of the estimated time of urban change from V-I-S time series relative to reference data is shown in Table 9. The MAE and RMSE are 2.5 and 4.3 years for all identified samples. The sign of SE indicates whether the estimated time of change is later or earlier than the reference change time. The Landsat-estimated time of change is 0.6 year earlier than the reference change time (SE = -0.6). MAE and RMSE are 3.3 and 3.8 years for the Urban Renewal samples, and 2.4 and 3.3 years for the Urban Expansion samples. The estimated time of change Urban Renewal samples is two years later than that from the reference data (SE = 2.0), while the estimated Urban Expansion samples is 0.8 year earlier than the reference change time (SE = -0.8). The estimated time of Urban Expansion is more accurate than the estimated time of Urban Renewal.

Table 9. Time difference (in years) between reference and estimated urbanization onset.

	All identified change samples (yr)	Urban Renewal (yr)	Urban Expansion (yr)
SE	-0.6	2.0	-0.8
MAE	2.5	3.3	2.4
RMSE	4.3	3.8	3.3

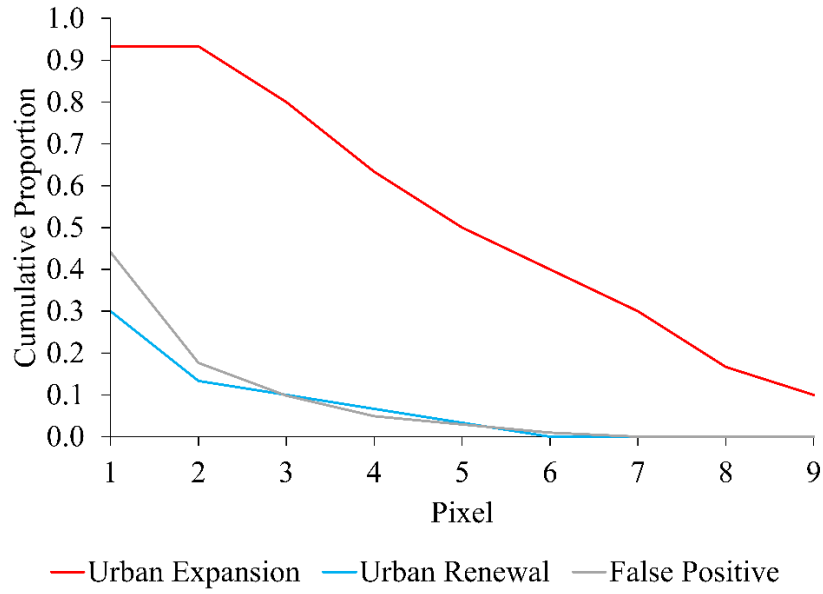


Figure 8. Cumulative proportion of successfully identified urbanized pixels within the 3 x 3 pixel reference sample units.

2.4. Discussion and Conclusion

2.4.1. Automated V-I-S Map Time Series Generation

In this chapter, V-I-S maps are derived automatically with the combined K-means clustering and NSMA modeling and used to identify two types of urbanization, which is different from other SMA studies of urban change that derive V-I-S maps for one or a few satellite images. Thus, the estimated V-I-S proportions derived from the automated processing is potentially less accurate than other estimated V-I-S proportions in other studies in nature. RMSE of Impervious is 22.2% for this study while 10.1% for Wu (2004) and 10.6% for Wu and Murray (2003). MAE of V-I-S are 25% while 12.5% for Powell et al. (2007). Three potential reasons may explain why vegetation and soil have high RMSE. These include: (1) vegetation seasonality variation, (2) VHR imagery misregistration caused by elevation variation, and (3) insufficient endmembers caused by cloud cover in some dates. For the no change reference samples, both sparse and dense herbaceous areas with

soil background were digitized as vegetation, and street trees trimming and mud overflow due to heavy rain often occurred and were observed in the study area. The inevitable error in the reference data caused by background soil and mud indicates that the estimated vegetation fraction may be more accurate than the reference data. Thus, the V-I-S distribution looks reasonable (Figure 7), but the RMSE for vegetation and soil are slightly high. However, the less accurate V-I-S maps were dropped, and the pixel V-I-S time series were smoothed for urbanization identification.

2.4.2. Urbanization Identification with V-I-S Map Time Series

Although Urban Expansion identification is accurate in this chapter, Urban Renewal identification is not. Identifying Urban Renewal is challenging because of the relatively coarse spatial and temporal resolutions of Landsat imagery, in relation to the space and time scales of Urban Renewal processes. The period of temporary soil exposure is about five days to a month during reconstruction period for some renewed buildings and skyscrapers based on the observation during field trips and VHR imagery, which was smoothed by the Savitzky-Golay filter with an-year window. Cloud contamination and mis-modeled V-I-S proportions add noise in the V-I-S time series. Additionally, Urban Renewal often occurs at a small scale (about three Landsat pixels in the study areas), which is only a small area of change sample.

2.4.3. V-I-S Trajectories of LCLUC

V-I-S trajectories of several representative urban LCLUC samples are shown in Figure 9. Samples for areas that did not undergo LCLUC tend to exhibit relatively stable V-I-S time trajectories (Figure 9a). An Urban Expansion example shown in Figure 9b, illustrates that

Vegetation was predominant during the early period, followed by temporary Soil predominance, and then conversion to Impervious, identical to our hypothesized V-I-S trajectories, which was also found in Powell et al. (2007). For the Urban Renewal sample (Figure 9c), the reference change date is in 2010, but the renewal is not shown in the figure. The signal of Impervious cover decline was removed in the processes of removal of high RMSE V-I-S maps and Savitzky-Golay filtering. Instead, Impervious cover increased around 1996 as observed in Figure 8c. However, most VHR imagery for visually inspecting Impervious are after 2000, so it is hard to ensure that the 1996 Impervious change is true change. A unique sample of park conversion is shown in Figure 9d, and the sample is in the location of an urban forest park (i.e. Daan Forest Park) based on the VHR imagery. The park was originally a ghetto with many illegal buildings, and then the buildings were torn down for improving the recreational life of urban residents in 1994. Hence, future research can focus on applying logistic regression on park conversion and assessing the results.

2.4.4. Spatial-Temporal Trends of Urbanization

An image portraying the location and time of onset of Urban Expansion is shown in Figure 10. The image extent includes the outer ring of Taipei and New Taipei city, the estuary of Danshui river, and the suburban areas of Taoyuan City. In Taipei City, substantial Urban Expansion occurred around 1996 in Neihu district close to the Danshui river, and within the Taishan and Xinzhuang districts. Urban Expansion in Neihu district occurred after the cutoff of Danshui river, while in Taishan and Xinzhuang districts Urban Expansion occurred as new factory development. Around 2007, a new scientific industrial park was established in Neihu district along with residential areas, and more industrial areas were developed in Wugu district. For the estuary of Danshui river, a new port was first built in the

coast of Bali district around 1996, and then a new residential town was planned and built in Danshui district after 2012. For Taoyuan City, Urban Expansion first occurred in the outer ring of existing urban areas of Taoyuan, Luzhu, and Guishan districts around 1996, and then Urban Expansion occurred in Linkou and Dayuan districts after 2007. New residential areas appeared in Linkou and Taoyuan districts in 1996 and 2012, respectively. while many factories appeared in Guashan, Luzhu, and Dayuan districts. Notice that the water mask in the figure was derived from the pixel QA layer with Landsat imagery dated at 6-Feb-2011 due to no cloud cover. The water mask is accurate for Danshui river, ponds in Taoyuan City, and sea, but some pixels are mislabeled as water (see the blue lines inside Taipei City). These mislabeled pixels correspond to streets affected by building cast shadows, so water should not be excluded during the processes of SMA and urbanization identification.

2.4.5. Conclusion

In this chapter, a novel approach to identifying the type and estimating the time of urbanization based on dense time series of Landsat imagery was developed and tested. The approach was tested for a study area in the north Taiwan region, a region that underwent rapid urbanization from 1990 to 2015. Accuracy of identified urbanization was assessed with Google Earth VHR imagery. To cover the study area and study period, 295 dates of Landsat imagery were used. The endmembers for 295 dates of Landsat imagery were automatically derived from K-means clustering, and then the derived endmembers were used in SMA modeling to generate V-I-S proportion maps. An algorithm was developed to identify locations and estimate dates of Urban Expansion and Urban Renewal. The time of change for Urban Expansion is estimated with logistic regression, while we attempt to

estimate the time of Urban Renewal for brief periods of soil exposure during the period between building demolition and construction.

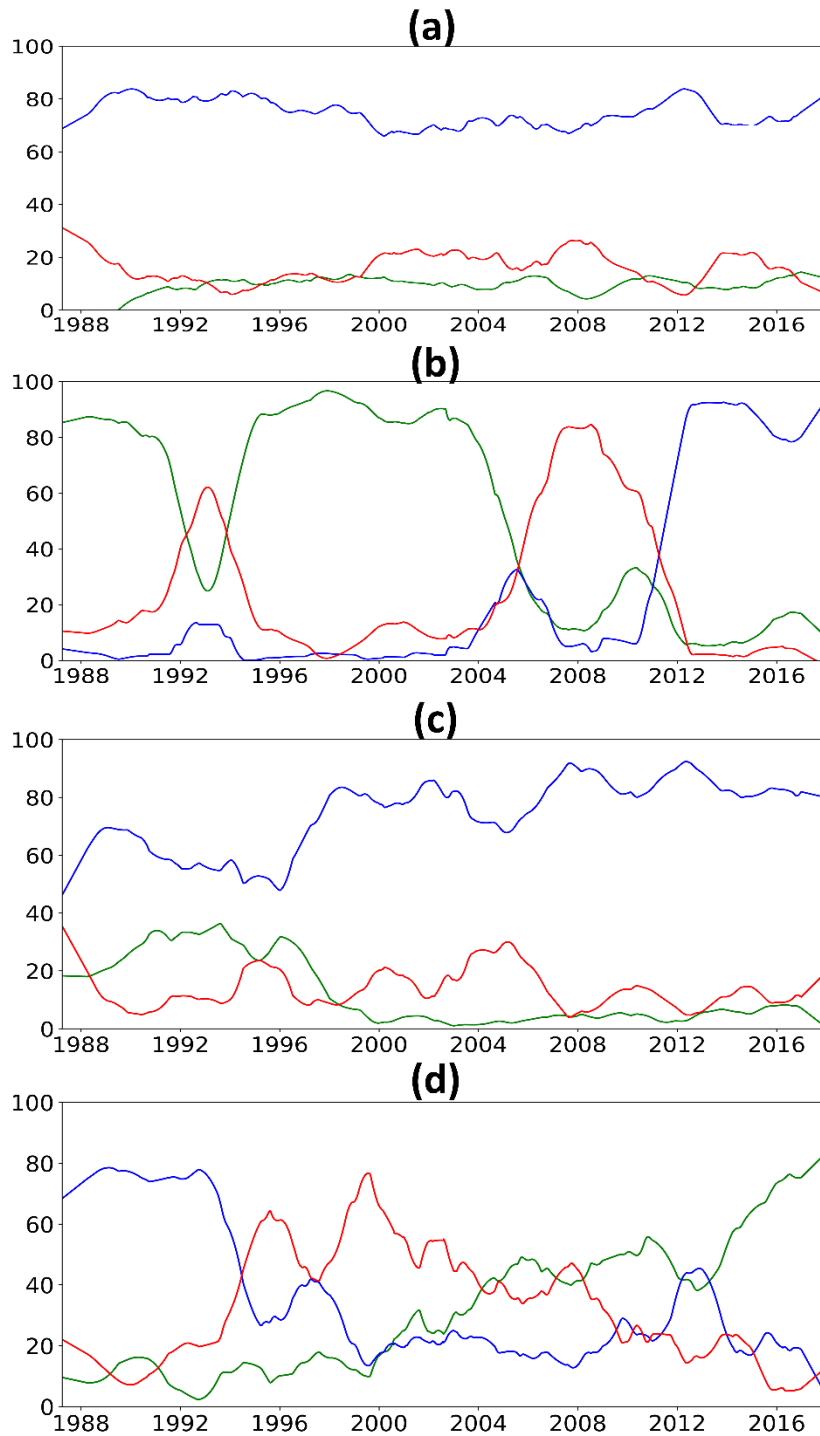


Figure 9. V-I-S time series trajectories of four urban LCLUC samples. (a) no significant land-cover change. (b) Urban Expansion in 2012. (c) Urban Renewal in 2010. (d) park conversion in 1994. Red is Soil; green is Vegetation; blue is Impervious.

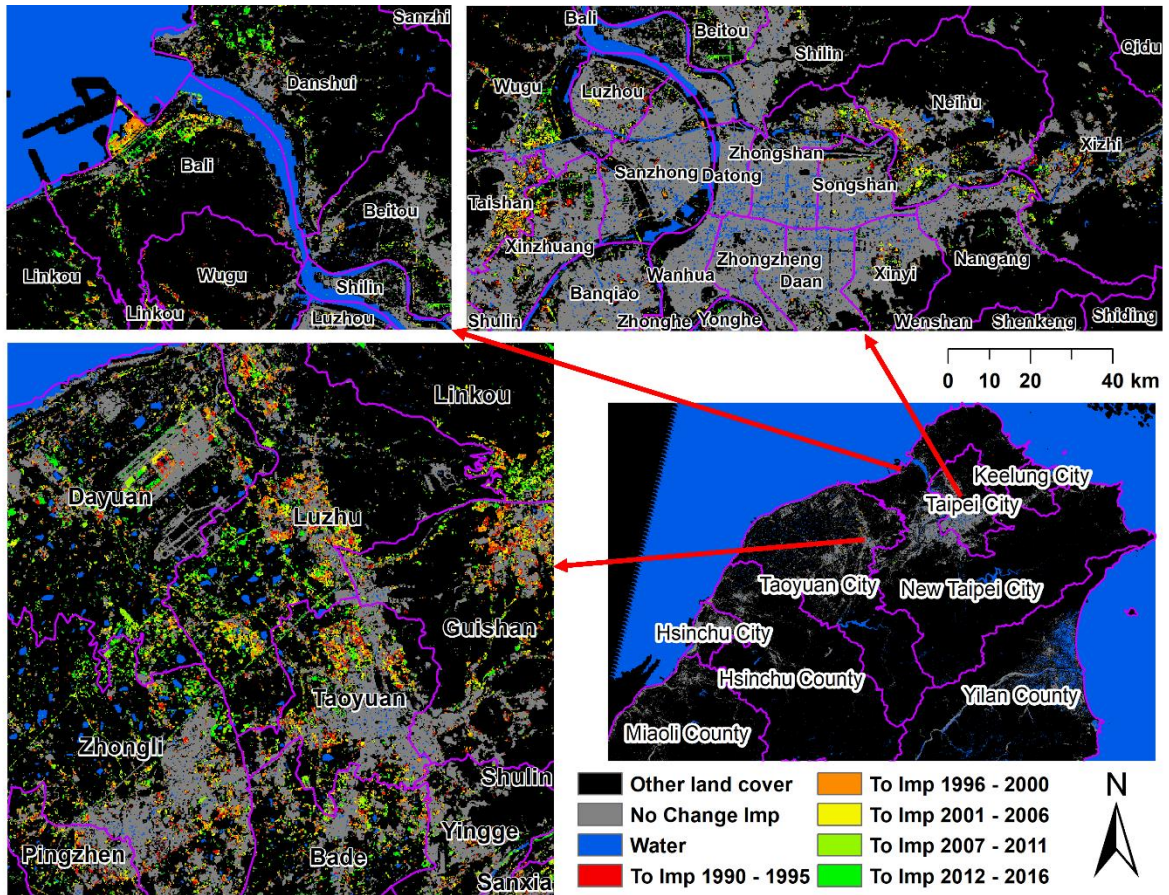


Figure 10. Time of urban expansion in the North Taiwan region.

On average, the RMSEs for the modeled V-I-S are 25%, 22%, 24% respectively. Of 295 Landsat images, 163 derived V-I-S maps with high error (i.e. high RMSE) were dropped, and the remainder of maps were used to construct a time series primarily for identifying urbanization. Urban Expansion locations and time of change onset are accurate, and 80% of Urban Expansion samples were detected with an accuracy of ± 2.4 years. However, the accuracy of identified Urban Renewal is low. SLC-off imagery acquired from Landsat 7 ETM+ sensor was found inappropriate for constructing V-I-S time series for identifying urbanization because modeled V-I-S maps derived from ETM+ imagery were found to have high errors, which adds noise into V-I-S time series. The V-I-S time series of No Change, Urban Expansion and Urban Renewal were shown and visually analyzed along with the

Urban Expansion change date map. With this automated SMA modeling and change identification approach, Urban Expansion identification with dense time series of Landsat imagery is shown to be achievable, but Urban Renewal identification was not.

Chapter 3. From Land Cover to Land Use: Applying Random Forest on Landsat Imagery for Detailed Urban Land-Use Mapping

3.1. Introduction

Land-cover and land-use change (LCLUC) is a global phenomenon that justifiably receives a lot of research attention (Bhatta, 2010; Lambin et al., 2001). Land-cover is the material (e.g. vegetation, impervious, and soil) covering the Earth's surface, and land-cover change involves the conversion from one material type to another. Land-use is how a parcel of land is used by human beings, so the type of land-use is primarily determined by human activities (e.g. residential, industrial, and commercial) (Cihlar & Jansen, 2001). Multiple forms of LCLUC occur in urban areas and their environs, which needs to be documented and studied.

Identifying, mapping, and monitoring land-use over extensive areas requires interpretation of remotely sensed data, commonly supported by field surveys. Land units associated with residential, industrial, commercial, and other types of urban land-use can be automatically classified or manually digitized with very high resolution (VHR) imagery. To generate an accurate land-use map normally requires tremendous time and effort (Jensen et al., 2001).

The type of land use is related to assemblages of land-cover types, so land-use types may be inferred based on land-cover composition and distributions (Cihlar & Jansen, 2001). For example, the fraction of Vegetation-Impervious-Soil (V-I-S) is typically different between residential and commercial parcels. However, Cihlar & Jensen (2001) and Jensen (2009) pointed out the land-cover compositions vary for the same type of land use. For example, in

some major cities of developing countries, low socioeconomic status (SES) residential areas may have less urban vegetation cover compared to high SES residential areas based on the analysis on VHR image (Stow et al., 2013; Stoler et al., 2012). Depending on the spatial resolution of remotely sensed data, identifying the type of land use from land-cover proportions is worthy of testing and analysis.

Machine learning classifiers are robust tools for classifying land-cover and land-use types based on remotely sensed data inputs (Schneider, 2012; Pal & Mather, 2005). Artificial neural network (ANN), decision tree (DT), support vector machine (SVM), and random forest (RF) are four types of machine learning classifiers widely used for LCLU classification in remote sensing, and the accuracies of LCLU maps derived from the different classifiers are often compared (Schneider, 2012; Duro et al, 2012; Dixon & Candade, 2008). Duro et al. (2012) tested DT, SVM, and RF classifiers to map a landscape in southern Saskatchewan, Canada, using SPOT-5 imagery, reporting overall accuracies over 85%. Dixon & Candade (2008) implemented maximum likelihood (ML), ANN, and SVM classifiers with Landsat imagery to map LCLU, finding that SVM and ANN outperformed the ML classifier in terms of overall accuracies. In both Duro et al. (2012) and Dixon and Candade (2008), their study areas were predominantly composed of agricultural land cover, which is relatively homogenous in the spatial domain compared to urban areas. For urban land-use classification, Jensen et al. (2001) developed an ANN classifier toolkit for image interpreters to help expedite the processes of identifying the types of land use with remotely sensed data. The study areas they tested were primarily in rural areas. Few studies have tested machine learning classifiers on images of densely developed cities for intra-urban land-use mapping.

Spectral, spatial, and texture image features have been explored with feature selection methods to analyze the importance of such image features for separating different types of land-use. Herold et al. (2003b) explored the importance of spatial metrics and image texture derived from IKONOS images in order to separate the types of land-use (e.g. residential, commercial, and forest) with Bhattacharyya Distance. Features of land-use are assumed to be normally distributed in Bhattacharyya Distance measures along with some separability-based metrics (e.g. Chernoff, Jeffreys-Matusita, and Kullback-Leibler divergence) (Guo et al., 2008). Machine learning technique can be applied to feature selection as well, and the advantage of machine learning is that it is not based on the assumption of normally distributed samples. For example, Schneider (2012) used RF to select features by ranking the importance of dense time series of Landsat wavebands, Normalized Difference Vegetation Index (NDVI), and spatial and textural metrics for land change detection. Zhu et al. (2012) applied the RF classifier on Landsat Enhanced Thematic Mapper Plus (ETM+) imagery along with Phased Array type L-band Synthetic Aperture Radar (PALSAR) imagery for urban land-use mapping, and they found that gray level co-occurrence matrix (GLCM) texture mean of near infrared (NIR) and Red wavebands were important for land-use classification in Massachusetts, USA.

The purpose of this research component is to: (1) determine the utility of land-cover features derived from spectral unmixing of V-I-S for classifying urban land-use types, (2) assess the accuracy of land-use maps derived from various spectral and land cover (V-I-S) input features to machine learning classifiers, and (3) analyze spatial-temporal trends of urban land-use change from Landsat using post-classification land-use change. The study area is the northern Taiwan region, and the study period is from 1990 to 2015, which

correspond to the area and period when most urbanization occurred in Taiwan. A novel contribution is an analysis of the importance of land-cover-related features (i.e. areal proportion, spatial arrangement, and the temporal deviations of V-I-S) for relatively detailed urban land-use mapping within the geographical and temporal contexts. At the 30 m pixel level of Landsat data, we expect that machine learning techniques that incorporate land-cover-related features might improve the accuracy of land-use classification compared to maps derived from Landsat spectral reflectance bands only. Feature selection was performed on Landsat imagery and its derivative land cover to explore the importance of spectral, spatial, and temporal features. The capability of semi-automatic machine learning classifiers for mapping land-use and land-use change in the urban environment was also assessed.

3.2. Methods

3.2.1. Data

The primary image source is from Landsat satellite systems (Table 10). Six dates of Landsat radiance images that had been converted to surface reflectance by Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS; Masek et al., 2006) and Landsat Surface Reflectance Code (LaSRC; Vermote et al., 2016) algorithms were downloaded. The six dates were chosen because of limited cloud cover near the two end points of the study period. Land cover (V-I-S proportions) was derived using spectral mixture analysis from the six dates of Landsat imagery and used as inputs to land-use classification routines. The detailed image processing steps for sub-pixel V-I-S estimation are described in Chapter 2. For each dataset, the middle date of imagery (i.e. 1992-Jul-27 and 2015-Nov-16) was the source for primary input features, while the other two dates of

Landsat imagery were used to derive temporal deviations of land-cover, described later in greater detail. The pixel QA layer for each date of surface reflectance was used to screen out cloud and cloud shadow pixels before the classification routine.

Table 10. Landsat imagery used for land-use mapping

Dataset	Sensor	Acquisition date	Land cloud cover
1990 dataset	Landsat 5 TM	1988-Feb-23	1%
	Landsat 5 TM	1992-Jul-27	2%
	Landsat 5 TM	1993-Feb-4	0%
2015 dataset	Landsat 8 OLI	2014-Jan-29	2%
	Landsat 8 OLI	2015-Nov-16	3%
	Landsat 8 OLI	2016-Jul-29	8%

Polygon-based land-use survey data collected for 2016 were primarily used to guide selection of reference data used for training and testing image classifiers. The survey data were provided through a portal of web map service (WMS) from the National Land Surveying and Mapping Center, Department of Interior, Taiwan (R.O.C.) (<http://maps.nlsc.gov.tw>). The data were generated by visual interpretation and on-screen digitization of VHR imagery and then refined by field surveys. The land-use classification scheme of the land-use survey has three levels, where Level 1 is the most general and Level 3 is the most detailed (Figure 11). However, the land-use survey data cannot be directly requested and used from the National Land Surveying and Mapping Center due to strict law limitations of Taiwanese government, which forbids data sharing to other countries. To test the reliability of relatively detailed urban land-use mapping with Landsat imagery, a modified version of the land-use survey was generated by categorical aggregation, and then I digitized land-use polygons based on the WMS land-use map for the modified land-use scheme. In my version of land-use scheme, mostly Level 2 urban land-use categories were included, while Level 1 categories were primarily used for non-urban categories (right two

columns Table 11), including six types of land-use for Level A and 16 types for Level B. The Level 3 land-use classes were not used since the information is more detailed than is needed. In total, 4660 land-use polygons were digitized for the 16 types of Level B land-use. The size of land-use polygons ranged from 9 to 48096 pixels. The number of pixels used in later training processes is shown in Figure 11.

Original scheme from Taiwan land survey		Modified scheme for image classification	
Level 1	Level 2	Level A	Level B
Water and water Facilities	Sea/Ocean River Reservoir Lake Pond	Hydrography	Hydrography (34654)
Forest	Coniferous Broadleaf Bamboo Mix Forest Shrubland Forestry facility	Undeveloped	Forest (222118)
Others	Wetland Grassland Barren Construction residual facility Undeveloped/vacant land		Non-forest Undeveloped (24276)
Agriculture	Wet field Dry field Orchard Aquaculture Animal husbandry Agricultural facility Cultural facility		Agriculture (41811)
Recreational	Park/Green space/open space Entertainment use/gym		Recreational (28688)
Mining	Mining and related facility Quarrying and related facility Salt industry	Nonurban Use	Mining (5320)
Water and water facilities	Flood discharge area Irrigation system Bank Water conservancy facility		Other Built (16667)
Built-up	Religious Cemetery Others	Residential	Residential (38862)
	Residential Mixed Use Commercial Industrial Storage		Mixed Use (16553) Commercial (11841) Industrial (11731)
Public Use	Governmental Agency School Medical Social welfare organization Utility infrastructure Environmental protection facility	Employment	Public Use (31016)
Transportation	Airport Port Railway		Airport (11550) Port (3602)
	High-speed railway Mass rapid transit National highway Provincial Highway Expressway General road Road facility	Transportation Corridor	Railway (1791) Road (47778)

Figure 11. Land-use classification scheme associated with the official polygon-based land-use survey and a modified scheme used for image classification. Number of pixels used in training processes shown in parentheses.

In addition to the land-use survey data, other ancillary data and available VHR images were used as a basis for generating reference data for assessing the accuracy of Landsat-derived LCLU/C products, including Google Earth VHR imagery (from 1999 to 2015) and two frames of historical aerial photos (from 1990 to 1999) acquired from the Aerial Survey Office, Forestry Bureau. Google StreetView imagery was also used as another source of reference data, by enabling visual interpretation of land-use within complex urban environments. In addition, I conducted fieldwork to confirm the current and past statuses of land-use where Google StreetView imagery was not available. Original Landsat imagery was visually interpreted to generate reference data, when the above VHR imagery and fieldwork were not available for the 1990 dataset.

3.2.2. Spatial and Temporal Feature Generation

Land use for a given place tends to be related to the land-use situation of surrounding areas, and the land use at the present is connected to the land use in the past and future. A novel contribution of this study is generating spatial and temporal features of land-cover (V-I-S) proportions and analyzing their importance in classifying urban land use. These features were derived from each date of the V-I-S proportion maps prior to conducting feature selection and land-use classification. For each pixel on two middle dates (1992-Jul-27 and 2015-Nov-16) of the V-I-S maps, GLCM texture measures were derived. Instead of applying GLCM on the original bands of surface reflectance, GLCM-based homogeneity texture was derived from the V-I-S proportion maps. The homogeneity measurements were derived with multiple kernels ranging from 3 to 9 pixels, and the appropriate kernel was evaluated later in the feature ranking and selection processes (i.e. section 3.2.3.). Temporal variation of V-I-S was derived as the temporal features with the following equation:

$$\frac{|L_{t-1} - L_t| + |L_{t+1} - L_t|}{2}$$

, where L_t is the proportion of land-cover in the middle date for each time period (1992-Jul-27 or 2015-Nov-16).

3.2.3. Feature Ranking and Selection

To understand the importance of an input feature on the resultant accuracy of land-use classification, a RF routine was run. All tested input features are shown in Table 11. The average of the input features was summarized for Level B land use to quantify the separability of land-use types. Feature selection in RF calculates the difference of Out-Of-Bag (OOB) errors before and after feature permutation, which gives an indication of relative importance of input features (Breiman, 2001). For each Level B land-use category, 1500 pixels within land-use polygons from the land-use survey data were randomly selected; extracted features for each pixel included: original surface reflectance, V-I-S proportions, GLCM Homogeneity texture from V-I-S images, and the temporal variation of V-I-S images. Coastal surface reflectance was excluded due to negative reflectance found in the training data. The pixel-based feature data were fed into the RF feature selection routine, and the top 10 important features were selected for land-use classification.

3.2.4. Land-Use Classification and Change

Because of higher overall map accuracy and shorter time for training and model selection (Shih et al., 2019), a RF classifier was used to conduct per-pixel land-use classification based on the Level B classification scheme (Figure 11). The 2015 Landsat image was selected for training and deriving land-use maps because it represents the closest date to that of the land-use survey data. All pixels were used for training, and the classifier

was trained and validated with N-folding cross-validation. The optimal RF classifier was found to result from incorporating 200 trees and six features at each split as RF hyperparameters. Three land-use maps were generated from the following input features: (1) surface reflectance data, (2) surface reflectance and V-I-S proportions, and (3) the top 10 features selected from the OOB importance analysis. The relative accuracies of the three maps reveal whether land-use maps derived with additional features tend to be more accurate.

Table 11. Input features for feature ranking and selection for Random Forest routines

Feature source	Feature
Surface reflectance	Blue waveband from 2015-Nov-16 Landsat image
	Green waveband from 2015-Nov-16 Landsat image
	Red waveband from 2015-Nov-16 Landsat image
	Near Infrared (NIR) waveband from 2015-Nov-16 Landsat image
	Shortwave Infrared 1 (SWIR 1) waveband from 2015-Nov-16 Landsat image
	Shortwave Infrared 2 (SWIR 2) waveband from 2015-Nov-16 Landsat image
Image-derived land-cover	Vegetation proportion (V) derived from 2015-Nov-16 Landsat image
	Impervious proportion (I) derived from 2015-Nov-16 Landsat image
	Soil proportion (S) derived from 2015-Nov-16 Landsat image
Spatial feature (Gray level cooccurrence matrix of land-cover)	GLCM homogeneity 3 by 3 pixel kernel of 2015 Vegetation proportion (V Homo 3)
	GLCM homogeneity 3 by 3 pixel kernel of 2015 Impervious proportion (I Homo 3)
	GLCM homogeneity 3 by 3 pixel kernel of 2015 Soil proportion (S Homo 3)
	GLCM homogeneity 5 by 5 pixel kernel of 2015 Vegetation proportion (V Homo 5)
	GLCM homogeneity 5 by 5 pixel kernel of 2015 Impervious proportion (I Homo 5)
	GLCM homogeneity 5 by 5 pixel kernel of 2015 Soil proportion (S Homo 5)
	GLCM homogeneity 7 by 7 pixel kernel of 2015 Vegetation proportion (V Homo 7)
	GLCM homogeneity 7 by 7 pixel kernel of 2015 Impervious proportion (I Homo 7)
	GLCM homogeneity 7 by 7 pixel kernel of 2015 Soil proportion (S Homo 7)
	GLCM homogeneity 9 by 9 pixel kernel of 2015 Vegetation proportion (V Homo 9)
GLCM homogeneity 9 by 9 pixel kernel of 2015 Impervious proportion (I Homo 9)	
GLCM homogeneity 9 by 9 pixel kernel of 2015 Soil proportion (S Homo 9)	
Temporal feature (temporal deviation of land-cover)	Temporal variation of Vegetation proportion (V Tmp)
	Temporal variation of Impervious proportion (I Tmp)
	Temporal variation of Soil proportion (S Tmp)

To automatically generate a land-use map using the 1990 dataset, automatic adaptive signature generalization (AASG) by Gray and Song (2013) was implemented for automatic derivation of training data for the 1990 dataset based on an existing land-use map. The most

accurate 2015 land-use map (i.e. the 2015 land-use map with the highest overall accuracy) was used as a base land-use map. Based on the base map, temporal image differencing of Impervious (I) proportion (i.e. $I_{2015}-I_{1992}$) was applied to determine stable pixels in the 1992 image. Impervious proportion maps were chosen for image differencing, instead of original surface reflectance, because I focused on the spatial-temporal trends of urban land-use change and urban areas tend to be predominantly covered by impervious land-cover types. The stable pixels were located within the range of $\mu \pm c\sigma$ for the output of image differences, where μ and σ corresponds to the mean and the standard of the output of the image difference, and the c is a user-defined parameter. A c parameter ranging from 0.25 to 1.5 (with 0.25 increment) was tested and visually examined for the 1992 and 2015 Landsat images. A parameter of 0.5 was determined as optimal for representing no change pixels and excluding most change pixels. Training pixels for the 1990 dataset were the overlapped areas of the stable pixels and the base land-use map, and the features of the training pixels (i.e. surface reflectance, V-I-S proportions, GLCM homogeneity of V-I-S, and temporal deviation of V-I-S) were extracted from the 1990 dataset. Like the 2015 dataset, three land-use maps were derived from the 1990 training pixels for comparing the accuracy of different input features.

To analyze the spatial-temporal trends of urban land-use change, two land-use change maps were generated and then compared to each other in terms of accuracy: one is from post-classification comparison of two dates of land-use maps, while the other through the incorporation of a change/no-change map derived in Chapter 2. Per-pixel post-classification comparison of the most accurate land-use maps for 1990 and 2015 was conducted to generate a land-use change map (post-classification change map hereafter). On the other

hand, the change/no change map was derived from the dense time series of Landsat V-I-S proportions for the 1990 to 2015 period, and the identified change pixels are pixels experienced urban expansion in the study period. The onset date of urban expansion was identified with logistic regression based on Impervious time series, and the identified location and estimated time of Urban Expansion is 80% accurate with the onset of expansion estimated within ± 2.4 years. Thus, the other land-use change map was derived from per-pixel overlay of the 2015 land-use and urban expansion maps. Maps were generated depicting land-use change (multi-change map hereafter) labeled with land-use types derived from the 2015 land-use map and change time, derived from the urban expansion map.

3.2.5. Accuracy Assessment

Map accuracy was assessed to (1) compare reliability of land-use maps derived with different image features, and (2) test the ability to identify and separate urban land-use types for subsequent land-use change analyses. For the first objective, 50 pixel-based samples (for each land-use type) that did not change from 1990 to 2015 were randomly selected and visually interpreted on the VHR imagery. An additional 686 no-change pixels were randomly sampled from within the entire study area for non-cloud/shadow pixels on either 1990 or 2015 datasets, which yielded 1486 no change samples for accuracy assessment. For the second objective, 300 randomly sampled pixels corresponding to urban land-use change areas were extracted, based on visual interpretation of the closest date of VHR imagery to the two original Landsat images. The 300 samples were all associated with change to urban land use (i.e. to Transportation Corridor, Residential, and Employment) because I focused on urban expansion areas. Accuracy of land-use and land-use change maps were summarized with confusion matrices generated with the 1486 and 300 samples. The samples

for accuracy assessment are mutually exclusive from the pixels that were used for training and cross-validation in the land-use classification to avoid overestimation on the accuracies of the derived land-use maps.

3.3. Results

3.3.1. Input Features for Land-Use Classification

Average value signatures for each Level B land-use class are shown in Figures 12 (spectral reflectance) and 13 (derivative features). Based on spectral reflectance only (Figure 12), 16 types of Level B land use have signatures that overlap considerably. Average reflectance of Forest and Hydrography are separable, while average reflectance of Road, Residential, and Mixed Use are inseparable. Some pixels have over 1, and the unreasonable surface reflectance stems from the misestimation of aerosol optical properties in the atmospheric correction in the Land surface reflectance procedure. The issue was pointed out in Masek et al. (2006), but whether atmospheric correction is conducted in the image preprocessing does little difference for the classification purpose (Song et al., 2001). In terms of derivative features (shown in Figure 13), Forest and Hydrography are separable in average V-I-S proportion, while urban land-use types have similar V-I-S proportions. Vegetation Homogeneity of Port is separable, and Soil Homogeneity of Forest and Airport are discernable. Vegetation temporal variation of Agriculture is separable, while Impervious and Soil temporal variations of Forest are separable.

3.3.2. Random Forest Feature Ranking

The relative importance of features derived from the RF routine is shown in Figure 14. The top 10 features are blue, green, red, NIR, SWIR1, and SWIR2 reflectance, Vegetation

Homogeneity with 7 by 7 pixel kernel, Soil Homogeneity with 9 by 9 pixel kernel, Vegetation Homogeneity with 9 by 9 pixel kernel, and Vegetation temporal variation. According to the RF ranking, green waveband reflectance is the most important, while Soil fraction is the least important. V-I-S proportions contribute less information to the RF classifier compare to the original surface reflectance data. The homogeneity texture measure based on a 9 by 9 pixel kernel was found to be most important for RF classification compared to texture from other pixel kernel sizes. Thus, GLCM homogeneity with 9 by 9 pixel kernel applied to of V-I-S images was used for subsequent land-use classification. Thus, the top 10 important features used in the later classification are six bands of surface reflectance data, V-I-S GLCM homogeneity, and V temporal variation, which were used for deriving land-use maps for the two dates of image dataset.

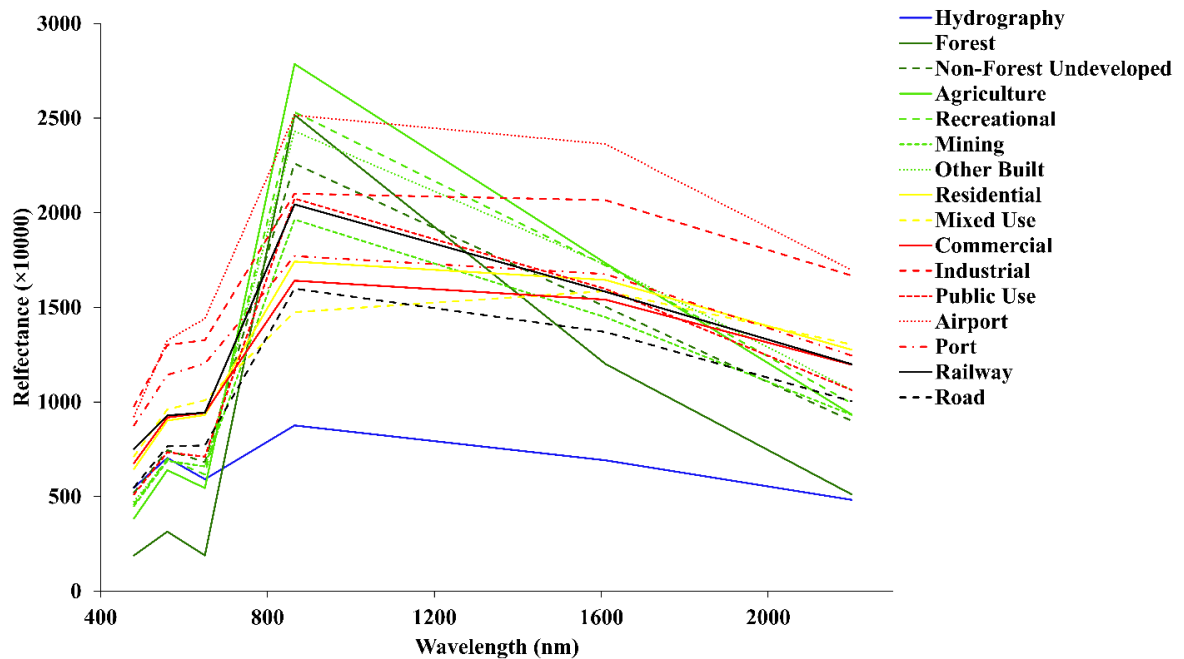


Figure 12. Average spectral reflectance for each land-use class.

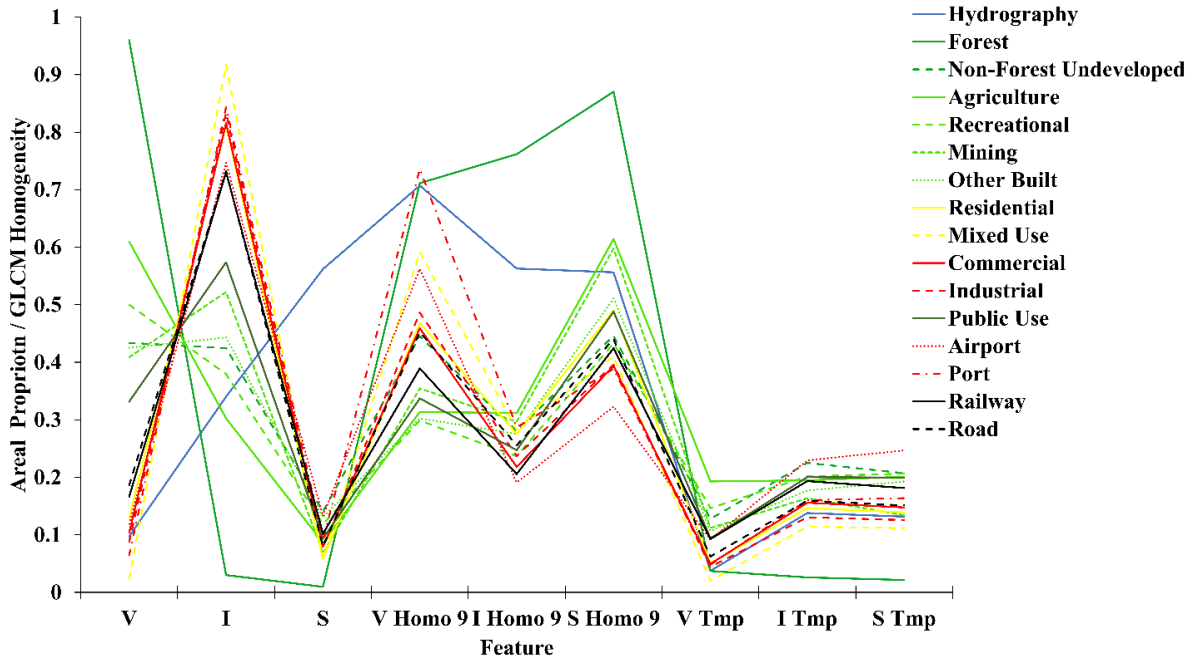


Figure 13. Average land-cover abundance, homogeneity, and land-cover temporal variation for each land use. V is vegetation, I is impervious, and S is soil. Homo is the GLCM homogeneity, and the number behind represents the number of pixel kernel. Tmp is the temporal variation for corresponding land cover.

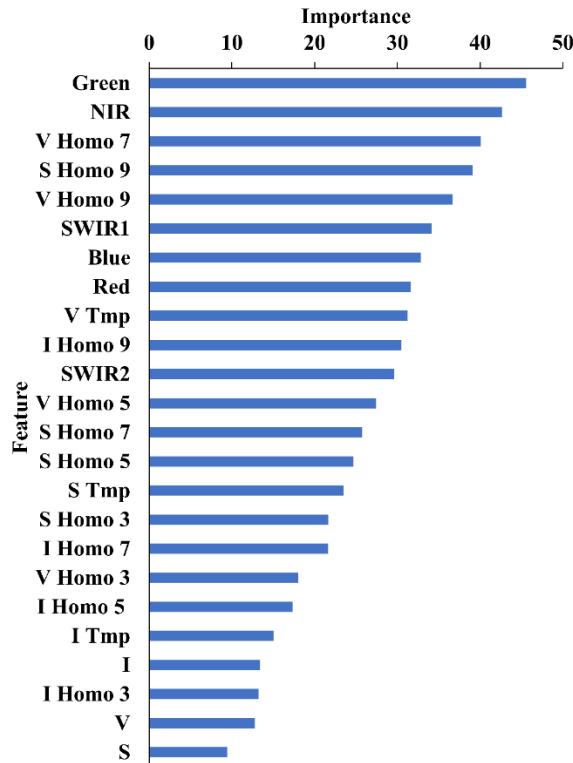


Figure 14. Out-of-bag permutation importance of features derived from random forest routines. Blue, green, red, NIR, SWIR1, and SWIR2 are the surface reflectance wavebands of Landsat 8 OLI sensor. V is vegetation, I is impervious, and S is soil. Homo is the GLCM homogeneity, and the number behind represents the number of pixel kernel. Tmp is the temporal variation for correspondent land-cover.

3.3.3. Accuracy of Land-Use Maps

Accuracy statistics for six (three feature input groups by two dates) Level B land-use maps are shown in Table 13. The six Level B maps were assessed with the 1786 independent samples (i.e. 1486 no change and 300 change samples) instead of the samples used in the training and cross-validation processes. The land-use map derived from the 2015 image dataset is more accurate than the 1990 map due to the higher quality training data for 2015 and greater uncertainty of reference data for 1990. The land-use maps derived from the top RF features are more accurate than other ones for both dates of images (52.0% for 1990 and 63.5% for 2015). Surprisingly, land-use maps derived from surface reflectance plus V-I-S proportions are the least accurate, which implies that V-I-S proportions estimated from the reflectance results in the RF classifier being overfit. The Forest class has the highest user's and producer's accuracies, while urban land-use types (e.g. Mixed Use, Commercial, Industrial) have the lower user's and producer's accuracies. Clearly, this level of land-use classification is too detailed to yield reliable maps from Landsat image data for this study area. The relative low Producer's and User's accuracies of Hydrography is because of some Hydrography samples are not covered by surface water, and these samples are on exposed riverbed covered by cobblestone or riparian vegetation.

Most categories in the Level B land-use maps derived from Top RF Features were aggregated to generate a more generalized Level A land-use map, with aggregations outlined in Table 12. The accuracies of the Level A maps are shown in Tables 13 and 14. The overall accuracy increases about 10% for both dates of land-use maps following categorical aggregation (62.0% for 1990 and 74.9% for 2015) Undeveloped has the highest producer's and user's accuracies for both dates. This is particularly encouraging when considering the

lower accuracy of the 1990 map, since for land-use change purposes most changes related to urbanization involve transitions from an initial undeveloped state.

Table 12. Land-use accuracy (%) of six Level B land-use maps

Map	Surface reflectance		1990 Surface reflectance + VIS		Top RF Features		Surface reflectance		2015 Surface reflectance + VIS		Top RF Features	
	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA
	Hydrography	47	34	50	34	40	30	62	59	65	59	59
Forest	70	86	70	84	74	90	91	91	92	91	93	94
Non-forest Undeveloped	4	3	6	6	12	10	16	17	15	15	27	26
Agriculture	67	37	66	38	69	44	74	53	72	51	71	64
Recreational	17	48	19	52	21	56	37	74	35	72	32	72
Mining	47	35	41	29	58	33	82	66	84	62	71	64
Other Built	29	25	26	23	36	37	43	53	40	51	53	50
Residential	46	45	45	46	50	51	61	57	61	56	56	55
Mixed Use	57	46	63	46	66	39	55	45	60	41	65	38
Commercial	45	30	48	34	62	34	84	43	82	43	78	41
Industrial	26	33	30	35	25	27	59	57	56	55	59	55
Public Use	7	18	8	18	9	20	36	45	29	43	26	45
Airport	50	68	46	60	70	72	50	98	51	98	65	98
Port	63	28	59	26	62	20	47	49	47	47	67	52
Railway	0	0	0	0	0	0	100	1	100	1	100	1
Road	21	44	16	38	17	42	35	66	34	66	35	52
OA	48.2		47.9		52.0		62.7		61.7		63.5	

* UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy. Top RF features are the important features derived from Random Forest routines for urban land-use classification.

Table 13. Accuracy of 2015 Level A land-use map derived from Top RF Features

Classified	Reference							UA (%)
	Hydrography	Undeveloped	Nonurban Use	Tra	Emp	Res	Total	
Hydrography	45	5	6	7	11	2	76	59.2
Undeveloped	5	439	27	12	11	4	498	88.1
Nonurban Use	9	27	313	15	37	12	413	75.7
Tra	13	6	8	75	14	25	141	53.1
Emp	7	11	22	22	312	50	424	73.5
Res	5	1	6	24	43	155	234	66.2
Total	84	489	382	155	428	248	1786	
PA (%)	53.5	89.7	81.9	48.3	72.8	62.5	OA (%)	74.9

* Tra = Transportation Corridor. Emp = Employment. Res = Residential. UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy.

Table 14. Accuracy of 1992 level A land-use map derived from Top RF Features

Classified	Reference							UA (%)
	Hydrography	Undeveloped	Nonurban Use	Tra	Emp	Res	Total	
Hydrography	26	5	11	7	14	1	64	40.6
Undeveloped	12	473	119	11	18	6	639	74.0
Nonurban use	23	49	370	24	32	14	512	72.2
Tra	9	11	29	38	56	20	163	23.3
Emp	15	13	75	23	108	27	261	41.3
Res	1	2	15	10	25	94	147	63.9
Total	86	553	619	113	253	162	1786	
PA (%)	30.2	85.5	59.7	33.6	42.6	58.0	OA (%)	62.0

* Tra = Transportation Corridor. Emp = Employment. Res = Residential. UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy.

3.3.4. Accuracy of Land-Use Change Maps

The map accuracies of two land-use change maps derived from Level A maps are shown in Tables 15 and 16. Overall change accuracy of the post-classification change map is 64.2% while 74.7% change accuracy for the multi-change map. No change Residential and all other land-use types (i.e. Res and Oth in Table 16) have higher producer's and user's accuracies, while change to Employment has the highest producer's and user's accuracy in both change maps. For the multi-change map, except change to Other (i.e. To Oth), Change to Residential has the lowest producer's accuracy (i.e. 20.0%), while Change to Transportation Corridor has the lowest user's accuracy (i.e. 19.6%).

3.3.5. Refinement of Land-Use and Land-Use Change Map

To analyze the spatial-temporal trend of urban land-use change along with population change in Chapter 4, three steps of refinement were conducted on the 2015 Level A land-use map to increase accuracy and reliability. According to the Table 13, Level A urban land uses (i.e. Transportation Corridor, Employment, and Residential) are not sufficiently high to correctly infer spatial-temporal trends of urban land-use change. In the first step, training

data for Road and Railway were excluded, and the remaining 14 types of Level B land use were trained with a RF classifier with the top 10 features. A new Level B land-use map containing 14 types of land use was derived from the trained RF classifier, and then the categorical aggregation was conducted to derive a new Level A land-use map. In the second step, manually editing processes were focused on Hydrography (e.g. exposed riverbeds), Employment, and Residential areas on the Level A land-use map. In total, 2051 polygons (ranged from 9 to 30818 pixels) were digitized to manually-correct the Level A land use, and these polygons cover 6.2% of the entire study area. Finally, motorways, primary roads, secondary roads, and rails from Open Street Map were used to reassign the pixels as Transportation Corridor in the Level A land-use map (manually-edited map hereafter). A new land-use change map was derived from the same method I applied by integrating the urban expansion map and the manually-edited map, so the new change map (manually-edited change map hereafter) were labeled with the To-land-use types along with the date of onset of land-use change.

Table 15. Accuracy of land-use change map derived from post-classification change map

		Reference (no change)				Reference (change)				Total	UA (%)
		Tra	Emp	Res	Oth	To Tra	To Emp	To Res	To Oth		
Classified (no change)	Tra	23	3	0	6	7	0	3	0	42	54.7
	Emp	3	96	5	3	0	27	6	0	140	68.5
	Res	3	7	81	0	0	0	3	0	94	86.1
	Oth	17	18	10	810	4	18	3	0	880	92.0
	To Tra	35	7	1	21	9	4	22	0	99	9.0
Classified (change)	To Emp	12	85	9	37	6	99	36	0	284	34.8
	To Res	15	28	42	12	5	8	30	0	140	21.4
	To Oth	11	18	2	66	1	7	2	0	107	0.0
Total		119	119	262	150	955	32	163	105	1786	
PA (%)		19.3	36.6	54.0	84.8	28.1	60.7	28.5	0.0	OA (%)	64.2

* UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy. Tra = Transportation Corridor. Emp = Employment. Oth = All other land-uses (i.e. Hydrography, Undeveloped, and Non-urban Use). To Tra = Change to Transportation. To Emp = Change to Employment. To Res = Change to Residential. To Oth = Change to all other land-uses (i.e. Hydrography, Undeveloped, and Non-urban Use).

Table 16. Accuracy of multi-change map

		Reference (no change)				Reference (change)				Total	UA (%)
		Tra	Emp	Res	Oth	To Tra	To Emp	To Res	To Oth		
Classified (no change)	Tra	50	9	1	20	6	0	4	0	90	55.5
	Emp	12	163	14	34	0	14	9	0	246	66.2
	Res	14	31	122	7	1	5	9	0	189	64.5
	Oth	28	33	11	854	2	4	1	0	933	91.5
Classified (change)	To Tra	8	1	0	7	10	4	21	0	51	19.6
	To Emp	3	18	0	6	6	112	33	0	178	62.9
	To Res	4	4	1	5	4	3	24	0	45	53.3
	To Oth	0	3	1	22	3	21	4	0	54	0.0
Total		119	119	262	150	955	32	163	105	1786	
PA (%)		42.0	62.2	81.3	89.4	31.2	68.7	22.8	0.0	OA (%)	74.7

* UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy. Tra = Transportation Corridor. Emp = Employment. Oth = All other land-uses (i.e. Hydrography, Undeveloped, and Non-urban Use). To Tra = Change to Transportation. To Emp = Change to Employment. To Res = Change to Residential. To Oth = Change to all other land-uses (i.e. Hydrography, Undeveloped, and Non-urban Use).

Accuracies of the manually-edited map and the manually-edited change map are shown in Tables 17 and 18. For the manually-edited map, the producer's and user's accuracies of all Level A land-use classes substantially increased compared to the 2015 Level A map (shown in Table 13). Thus, the overall accuracy of the manually-edited change map also increased 8% because of the manually-editing processes. The manually-edited change map was used for analyses of spatial and temporal trends of urban land-use change discussed in the following section, and the land-use change map will be used in Chapter 4 for identifying the relative timing between population growth and land-use change.

Table 17. Accuracy of manually-edited map

Classified	Reference			Tra	Emp	Res	Total	UA (%)
	Hydrography	Undeveloped	Nonurban use					
Hydrography	70	5	0	3	5	0	83	84.3
Undeveloped	3	442	26	10	3	1	485	91.1
Nonurban use	4	27	328	13	11	4	387	84.7
Tra	4	6	7	48	5	1	71	67.6
Emp	3	7	19	9	237	6	281	84.3
Res	2	3	4	13	10	147	179	82.1
Total	86	490	384	96	271	159	1486	
PA (%)	81.3	90.2	85.4	50.0	87.4	92.4	OA (%)	85.5

* Tra = Transportation Corridor. Emp = Employment. Res = Residential. UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy.

Table 18. Accuracy of manually-edited change map

		Reference (no change)				Reference (change)				Total	UA (%)
		Tra	Emp	Res	Oth	To Tra	To Emp	To Res	To Oth		
Classified (no change)	Tra	74	1	2	3	5	0	0	0	85	87.0
	Emp	6	202	6	26	2	20	5	0	267	75.6
	Res	16	12	135	11	0	0	17	0	191	70.6
	Oth	8	21	5	875	2	3	1	0	915	95.6
	To Tra	11	0	0	0	10	2	1	0	24	41.6
Classified (change)	To Emp	3	23	0	5	8	119	13	0	171	69.5
	To Res	1	1	2	4	3	3	63	0	77	81.8
	To Oth	0	2	0	31	2	16	5	0	56	0.0
Total		119	262	150	955	32	163	105	0	1786	
PA (%)		62.1	77.0	90.0	91.6	31.2	73.0	60.0	0.0	OA (%)	82.7

* UA = User's accuracy. PA = Producer's accuracy. OA = Overall accuracy. Tra = Transportation Corridor. Emp = Employment. Oth = All other land-uses (i.e. Hydrography, Undeveloped, and Non-urban Use). To Tra = Change to Transportation. To Emp = Change to Employment. To Res = Change to Residential. To Oth = Change to all other land-uses (i.e. Hydrography, Undeveloped, and Non-urban Use).

3.4. Discussion and Conclusion

3.4.1. Features for Land-Use Extraction

RF feature ranking was found to be more informative than graphs of average features for determining the most appropriate feature inputs for land-use classification. This is particularly the case since many training data for Level B land-use classification have non-normal distributions. All surface reflectance wavebands were useful for urban land-use classification, especially for the higher radiometric resolution in Landsat 8 OLI imagery. This is likely one of the reasons why the land-use maps derived from Landsat 5 TM imagery are less accurate than the maps derived from Landsat 8 OLI imagery.

V-I-S land-cover proportions provided no direct contribution for boosting classification accuracy, while spatial features (i.e. GLCM Homogeneity of V-I-S) and a temporal feature (i.e. Vegetation temporal variation) derived from the V-I-S proportions slightly increased the classification accuracy. Zhu et al. (2012) input GLCM texture of original wavebands, four

seasons of Landsat ETM+ imagery, and PALSAR imagery to an RF classifier and the texture of surface reflectance was found to be important to land-use classification. In this chapter, I included GLCM texture of surface reflectance into the RF routines, and the top 10 features are Soil GLCM variance, Soil GLCM entropy, SWIR1 GLCM mean, Soil GLCM second moment, NIR GLCM contrast, Soil GLCM contrast, Soil GLCM dissimilarity, Impervious GLCM correlation, SWIR2 GLCM mean, and SWIR2 GLCM correlation. Accuracy assessment was conducted on a land-use map derived from the top 10 features, and the overall accuracy (56.8%) is lower than the land-use map derived from the surface reflectance. Although the study area and the classification scheme adopted by Zhu et al (2012) are different from this study, spatial (i.e. GLCM textural measurement of V-I-S proportion) and temporal features (i.e. Vegetation temporal variation) that I tested proved to be more useful for land-use classification of my densely urbanized study area. The same approaches should be tested and compared for other cities in East Asia (e.g. Tokyo, Seoul, Hong Kong) that are densely urbanized and expanding in future studies to confirm whether the spatial and temporal features we used are more useful than others.

3.4.2. Land-Use Change Map Comparison

Land-use change maps derived from integrating the dense-time-series-derived change map are more accurate than the one derived from post-classification comparison. Although the From (time 1) land-use types are preserved in the post-classification change map, the accuracy of mapping land-use transitions depends on the individual accuracies of two dates of land-use maps. Thus, low accuracy of the two land-use maps leads to inaccurate land-use transition in the resultant land-use change map. Conversely, the accuracy of the multi-change map is determined by the accuracies of the urban expansion map and the 2015 land-

use map. The accuracy of the urban expansion map is high due to considering land-use change with the relatively fine temporal resolution of Landsat image series, and the accuracy of the 2015 land-use map is high due to more reliable and extensive training data for the 2015 dataset.

3.4.3. Challenge of Inferring Land Use from Land Cover

To understand the reason why land-cover proportions did not contribute useful information for urban land-use classifications, histograms of pixel-based land-cover proportion shown in Figure 15 were analyzed. The histograms were derived from the 2015 training data for four exemplar Level B land-use types, and the V-I-S proportions were divided into 10 bins with equal bin interval. V-I-S proportions clustered at certain interval for non-urban area (e.g. Forest), while the proportions are more dispersed for urban land-use types (e.g. Residential, Public Use). The V-I-S proportions are similar for Residential and Public Use. V-I-S proportions of non-urban use are separable, but the proportions cannot be used to form decision boundaries among urban land use. In addition, a type of land use may contain all three types of land cover, and this situation was found and shown in Figure 5. Residential, Public Use, and Hydrography have no distinct land-cover proportions to form decision boundaries from one another at pixel level. According to the visual inspection on the V-I-S proportion maps, most surface water areas have 100% Soil proportion due to inclusion of moist soil endmember in the SMA unmixing processes in Chapter 2. During dry season, riverbeds covered by cobblestone were modeled with high Impervious proportion, while riverbeds covered by riparian vegetation were modeled with high Vegetation proportion. Thus, cobblestone riverbeds were misclassified as Commercial, Residential, or

Road, while riparian vegetation were misclassified as Non-forest Undeveloped, Agriculture, or Recreational.

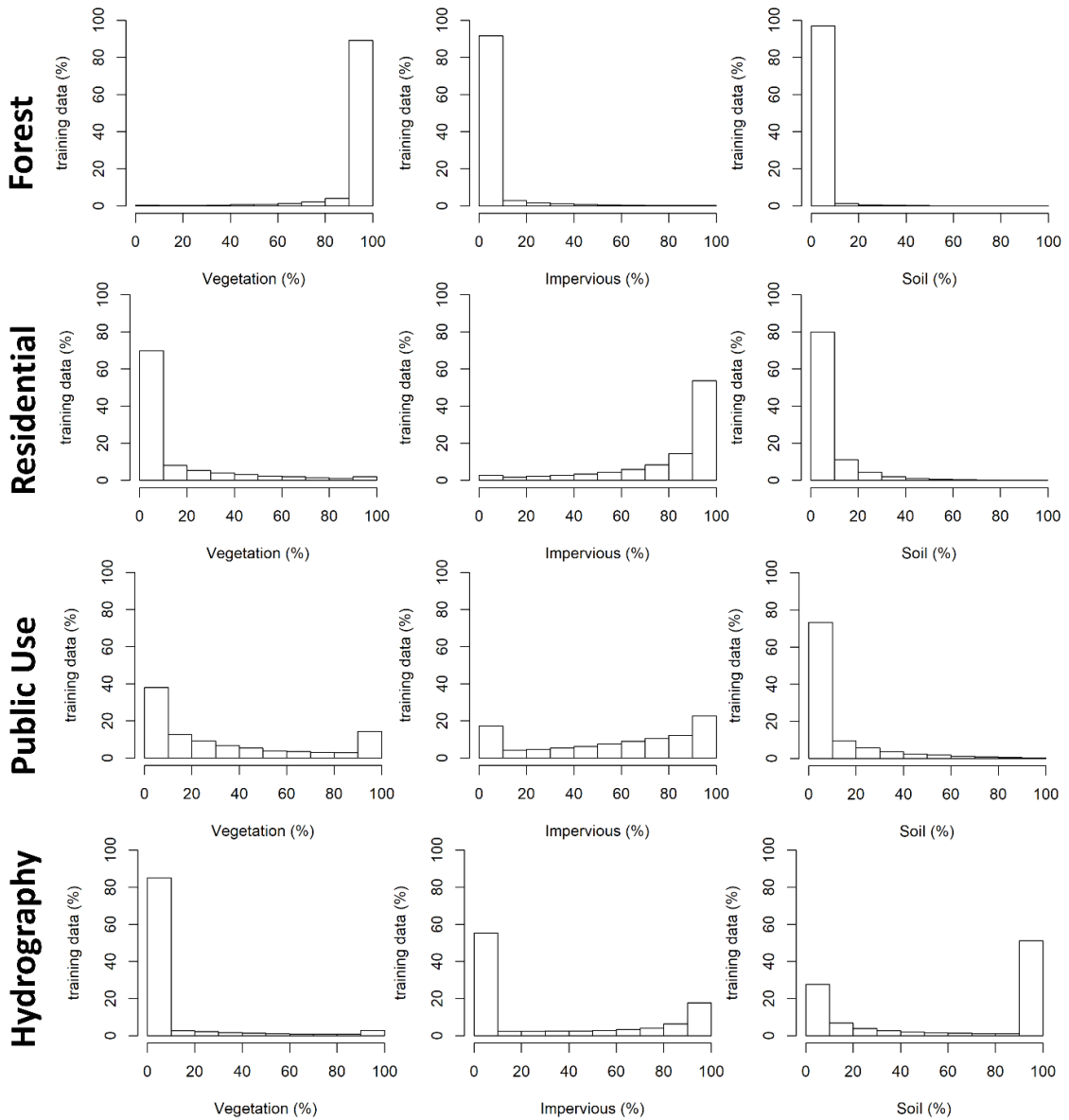


Figure 15. Histograms of pixel-based land-cover fraction derived from 2015 training data for four exemplar types of Level B land use. Forest has distinct V-I-S proportions from other land-use types. Similar V-I-S proportions are found for multiple urban land-use types, and Residential and Public Use (e.g. school, governmental agency) are two exemplar land-use types. Hydrography has complicated V-I-S proportions: surface water has high Soil proportion; riparian areas have high Vegetation proportion; dry riverbeds have high Impervious proportion.

Land-cover summarized at the V-I-S level is another reason why land-cover proportions may not be effective for inferring land-use in this case study. The land-cover features I tested are V-I-S proportions. However, forest, street trees, herbaceous, and mangrove all belong to Vegetation, while steel, cement, tile, and asphalt all belong to Impervious, but these subclasses have different spectral signatures. The V-I-S proportions are the sum of all these different materials, but some materials are more directly related to certain types of land use (e.g. asphalt to road). Estimating detailed land-cover materials proved to be accurate in past studies (e.g. Powell et al. [2007] and Okujeni et al. [2015]), and future study should focus on inferring detailed land-use from the detailed land-cover fractions.

3.4.4. Spatial-temporal Trend of Land-Use Change

Urban land-use expansion mostly occurred in Taoyuan City, New Taipei City, and Hsinchu County, which corresponds to the peri-urban areas of three core cities (Taipei City, Hsinchu City, and Taoyuan and Zhongli districts of Taoyuan City). The manually-edited change map enables visual assessment on the spatial-temporal trends of land-use expansion. Based on the assessment, newly developed area appeared at the edges of the existing built areas. Additionally, Residential and Employment land was expanded along the two sides of transportation corridor, and new Transportation Corridor tended to lead Residential and Employment development.

Five areas that experienced extensive urbanization are focused on and magnified (shown in Figure 16), including Taoyuan (Figure 16a), Linkou (Figure 16b), Taishan (Figure 16c), Hsinchu (Figure 16d), and Neihu (Figure 16f) areas. For the Taoyuan area (Figure 16a), factories for New Employment appeared first in Taoyuan district around 1990, later in Luzhu district around 1996, and then occurred in Zhongli and Dayuan districts around 2006.

New Transportation (e.g. highway) first appeared in Dayuan and Luzhu districts around 1990 and later in Guishan, Luzhu, and Zhongli districts around 2011. New Residential first occurred in Taoyuan and Bade districts around 1996 and later in Luzhu around 2011. For the Linkou area (Figure 16b), New Employment land first occurred in Guishan district around 1996 and later in Linkou district around 2011. New Residential first occurred in Guishan district from 1996 to 2005 and later in Linkou district from 2006 to 2016. For the Taishan area (Figure 16c), factories for New Employment occurred in Taishan district from 1990 to 2005 and later in Wugu district from 2006 to 2016. New Residential appeared in Xinzhuang district from 1990 to 2016. For the Hsinchu area (Figure 16d), factories and scientific labs were built in East and Zhubei districts for the entire study period, while New Residential appeared in the districts from 1996 to 2016. Most New Transportation (new highway ramp) was built in East district around 1990 while later in Zhudong and Qionglin districts from 2006 to 2016. For Neihu area (Figure 16f), New Employment for commercial buildings occurred first in Songshan district from 1990 to 1995 and later in Neihu district from 2001 to 2016. New Residential appeared in Zhongshan and Neihu districts from 1996 to 2016, while New Transportation appears in Songshan and Neihu districts around 1990.

3.4.5. Conclusion

Four objectives were addressed: (1) understanding the importance of various features derived from Landsat imagery for relatively detailed urban land-use mapping, (2) comparing the accuracy of land-use maps derived from different input features, (3) comparing the accuracy of land-use change maps derived from two different approaches, and (4) understanding the linkage between land cover to land use. RF classification routines were applied to examine the capability of mapping relatively detailed urban land-use based on

Landsat image inputs. I found that the original spectral wavebands contributed the most information to the RF classifier for urban land-use mapping, with GLCM homogeneity of V-I-S and Vegetation temporal variation also providing discrimination power. For both dates of Landsat imagery dataset (i.e. 1990 and 2015 datasets), the land-use maps derived from the top 10 features are the most accurate, while the land-use maps derived from the original spectra and V-I-S proportion were the least accurate due to overfit RF classifiers. Accuracies of Level B land-use maps are insufficient to correctly identify land-use change, so the Level B maps were categorically aggregated into Level A land-use maps. Additionally, Level A urban land uses (i.e. Transportation Corridor, Employment, and Residential) still need manually editing to correctly identify urban land-use change aside from the remotely sensed data.

Two land-use change maps were derived: one by post-classification comparison of 1992 and 2015 maps (i.e. the post-classification change map), and the other by integrating the urban expansion map (derived from chapter 2) on the 2015 land-use map (i.e. the multi-change map). The multi-change map is more accurate than the post-classification change map for identifying urban expansion areas. V-I-S proportions can be used to separate non-urban land use from urban land use (e.g. Forest versus Residential), while V-I-S proportions of multiple urban land-use types (e.g. Residential versus Public Use) are too heterogeneous to separate.

The spatial-temporal trend of urban expansion was analyzed in dense urban environments and their vicinities. Transportation Corridor land use tended to lead the development of Residential and Employment in relative undeveloped districts, and extensive

urban land-use change occurred in the peri-urban areas of New Taipei, Taoyuan, and Hsinchu cities.

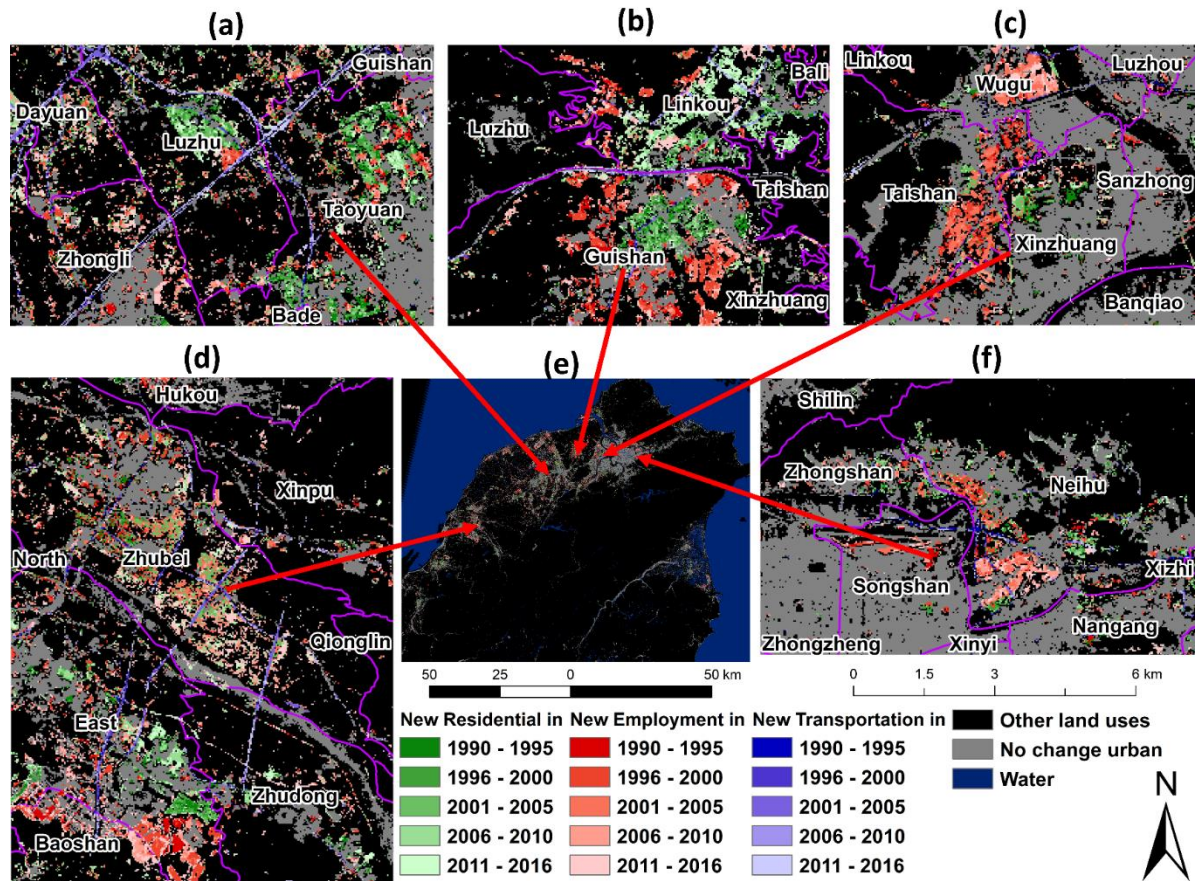


Figure 16. Spatial-temporal trend of land-use change of northern Taiwan region. (e) entire study area. (a) Taoyuan area. (b) Linkou area. (c) Taishan area. (d) Hsinchu area. (f) Neihu area. The number in the legend refers to the change time with correspondent land-use types.

Chapter 4. The Relative Timing of Population Growth and Land-Use Change – A Case Study of North Taiwan from 1990 to 2015

4.1. Introduction

Urban expansion is a form of land-cover and land-use change (LCLUC) that occurs globally (Seto et al., 2011). Population growth (caused either by natural increase or immigration) is considered to be the most connected phenomenon to urban expansion, which can be a driver of and be driven by LCLUC. For example, Powell et al. (2008a) monitored and analyzed the size of 10 cities in Rondonia, Brazil, relative to their human populations. A strong log-log relationship was found between city size and population, which implied that the size of city highly reflects the size of human population. In addition, Stow et al. (2016) mapped and monitored built areas at the district level in southeastern Ghana, and a strong relationship was quantified between built area and population density. Thus, the area of built land cover is a plausible demographic indicator, especially for the human population size. However, some researchers point out that the relationship of LCLUC and population is not unidirectional, but rather, reciprocal (Lambin et al., 2001; Sherbinin et al., 2007), so the processes and impacts between urban expansion and population change should be analyzed at relatively fine spatial and temporal scales.

4.1.1. Urbanization

Urbanization is a process of more people, and a greater fraction of people, living in urban areas over time (Weeks, 2010; Scheuer et al., 2016). Scheuer et al. (2016) synthesized and compared three urbanization models: (1) the urbanization stage model (van den Berg et al., 1982), (2) the polarization and spread model (Myrdal, 1958), and (3) the core-periphery

model (Friedmann, 1966). Van den Berg et al (1982) focused on relative population change in the core and periphery of a city over time. Myrdal (1958) saw urbanization as processes of human activities concentrating in a place and then spreading out to the nearby areas. Friedmann (1966) emphasized hierarchical growth among cities, towns, and villages, and the directions of product exchange among the human settlements.

Urbanization can be decomposed into two phenomena, urban population growth and built environment expansion. The growth of urban population results from net in-migration and natural increase. Once an urban place has been established, subsequent built environment growth usually occurs in peri-urban areas, where more land is converted to the built environment. In addition, more buildings are constructed in small villages, which transform the original villages into towns (Antrop, 2004). Sometimes, urban population growth and built environment growth are interrelated to each other, especially for the prosperous cities and towns. In such places, more people require spaces for living and working, so more land is converted into built environments for such usage. However, urbanization is a complicated process, and the nature of the transition needs to be well understood in order to best facilitate future urban planning and development.

4.1.2. Urban Population Growth

Rural-to-urban and/or urban-to-urban migration are critical sources of population growth for growing cities in addition to natural increase. Based on the push-pull theory (Lee, 1966), a reason for people to migrate is because of dissatisfaction with their current living places (i.e. push factors) and the attraction to their destinations (i.e. pull factors). Massey (1999) reviewed and synthesized multiple theories about international migration, but all such theories could also be applied and related to internal migration.

There are several reasons why people move from rural to urban areas. First, wages in cities are higher than in traditional agrarian work (a pull factor from urban areas), and neoclassical economic theory suggests that people are motivated to seek higher wages (Bencivenga & Smith, 1997). Second, some households in the rural areas would like to maximize their profit (a pull factor from urban areas) and lower the risk of conducting traditional cultivation (a push factor from rural areas), as described by the household-based economic theory (Gubhaju & De Jong, 2009). Thus, one or more members of a household will move to work in cities, and then those migrants will send remittances back to the households. Third, cities need laborers to fill in the secondary sector positions (a pull factor from urban areas) for serving people who work in the primary sector based on the segmented market labor theory (Fan, 2003). People who work in the primary sector tend to be well-educated, skilled, and earn higher salaries, while people who work in the secondary sector have low-education, are unskilled, and earn lower pay. Job vacancies in the primary sector are easy to fill, but labor in the secondary sector is still needed. Thus, these jobs in the secondary sector are gradually filled with women, teenagers, rural-to-urban migrants, and low-skilled immigrants from lower-income countries.

4.1.3. Impacts on Land from Urbanization

LCLUC occurs in both rural and urban areas due to urbanization. Massey (1999) pointed out how land use could change in rural areas. Rural households having members who work in urban areas often spend the remittances sent home on acquiring more land in home communities as prestige value or retirement income after saving enough remittances. In some cases, newly purchased land lies fallow because working in cities is more lucrative than doing agrarian work. When the households with remittance-sending family members

maintain some agricultural activities, they usually adopt new technology (e.g. machinery and fertilizer) to increase production instead of continuing traditional cultivation practices. In addition, capitalist farmers in the peripheral areas consolidate landholding, use machines, and plant cash crops to maximize profit. In urban areas, Antrop (2004) pointed out that new commercial and industrial activities may appear at the edge of large cities (e.g. Brussels, Belgium) where new peripheral roads are developed for relieving traffic congestion during the process of urbanization. Bell et al. (2010) provided examples of LCLUC due to migration, which including land abandonment due to out-migration in rural areas of Latvia, suburbanization of new luxurious housing due to in-migration of retirees to the coastal areas of Spain, and suburbanization of new illegal housing units due to in-migration of low-skilled workers in the Lisbon Metropolitan area.

4.1.4. Hypothetical Processes of Urbanization

One of the reasons why internal migration occurs is that people move to a place where more jobs and/or higher salaries exist. According to multiple migration theories, this situation is especially true for rural-urban migration or urban-to-urban migration (Massey, 1999; Zhang & Song, 2003; Lucas, 2004). The causes and the effects of human migration should reflect on land-use change as well. Therefore, I hypothesize that the dominant relationships between population change and land-use change in an urbanizing area occur as follows: (1) After new commercial and industrial establishments are built in a given place, job seekers from other places move to the vicinity of these establishments and primarily become renters of residential dwellings; (2) Increasing numbers of residents prompts the demand for housing, which leads to new residential developments in the place of work or nearby; (3) After the new residential buildings are completed, people can “officially” move

into the new housing developments, which indicates that the migrants are officially recorded in a registered population system. In other words, new industrial or commercial land use appears first in the place, residential developments follow, and finally, in-migration is officially counted in the governmental data. (4) Places in the process of urbanization will continue to densify with new small-scale businesses and transportation developments, and more migrants will move in. However, the later land-use change and human migration is not the focus of this study, as the goal is to determine whether or not most of the places (i.e. districts in this study) that were once undeveloped share the similar temporal processes of land development and human in-migration at the beginning of development.

4.1.5. Purposes of Study

The purpose of this chapter is to explore the relationship and the relative timing between population change and land-use change based on a case study of northern Taiwan from 1990 to 2015. Land-use data were generated from dense temporal satellite imagery (i.e. Landsat TM, ETM+, and OLI), and population data at the fine spatial (district) and temporal (annual) scale are based on the registered permanent address for each individual and were acquired from the local governmental website. First, the general relationship between population, land use, and their change across space and time was tested with regression analysis. Second, the relative timing between population change and land-use change was identified with lagged correlation tests. Furthermore, hypothetical processes of urbanization among population growth, residential, and employment land-use change were tested. This chapter informs how growth of the urban environment is associated with population growth across space and time.

4.2. Data and Methods

4.2.1. Data

A digital geographic information system file (shapefile) containing all district polygons was downloaded from a Taiwanese governmental website (<https://data.gov.tw/>). 107 of 368 districts located in the study area were extracted from the downloaded shapefile, and then islets that are not connected to the main island of Taiwan were manually removed for the 107 districts. The modified 107 district shapefile was used as a basis for deriving areal extent of land use.

Two main sources of data were used to represent land use and population on an annual basis. Time series data on land use from 1990 to 2015 derived from classification of dense time series of Landsat imagery were used. An accurate method for estimating areal extent of urban land-use types (Transportation Corridor, Employment, and Residential categories) in an urbanizing region over time is to generate an urban land-use change map labeled with the date that urbanization commenced, and then conduct an overlay analysis between the urban land-use change map and an accurate land-use map for the end date of study period. Thus, a semi-automatic approach to identifying the starting time of urban land change was developed and tested based on dense time series of Vegetation-Impervious-Soil (V-I-S) proportion maps derived from Landsat surface reflectance imagery (see details in Chapter 2). The location and estimated time for newly urbanized lands were generally accurate, with 80% of urban expansion estimated within ± 2.4 years. Next, random forest (RF) classification was applied to a Landsat image from 2015 to create maps of detailed urban land use (e.g. Residential, Employment, and Transportation categories) along with refinement of OpenStreetMap and manual editing, with an accuracy of 82% for land-use

categories (details provided in Chapter 3; shown in Figure 17). The areal extents of Residential, Employment, and Transportation Corridor (Transportation hereafter) land-use types were summarized for each district on an annual basis.

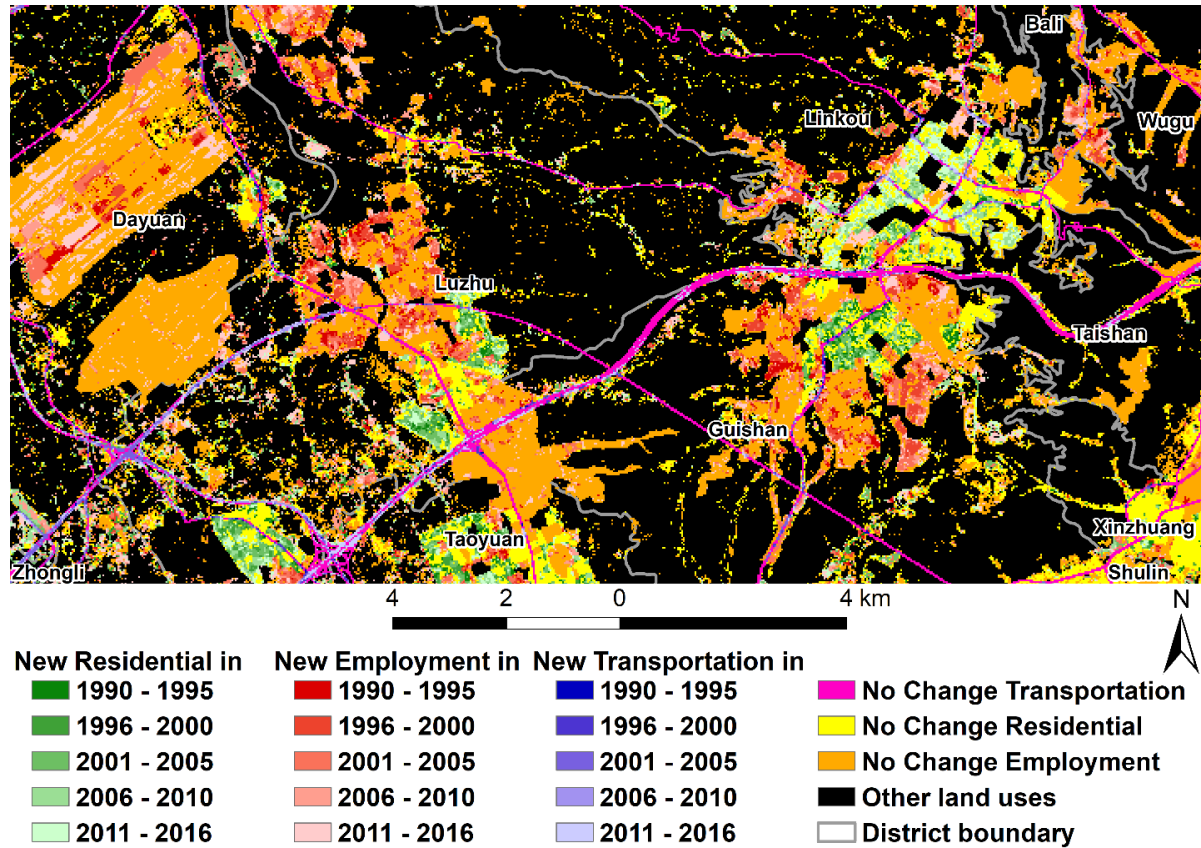


Figure 17. Subset of urban land-use change map from 1990 to 2015. The change map was derived from overlay analysis between an urban expansion map derived from Chapter 2 and a 2015 land-use map derived by random forest classification of Landsat imagery in Chapter 3.

Two sources of population data were used, including population registers and census data. The registered population data contain annual population counts at the district level. The registered population data were downloaded from Monthly Bulletin of Interior Statistics, Taiwan (R.O.C.) (https://www.moi.gov.tw/files/site_stuff/321/1/month/month.html). These data date back to 1981 on an annual basis, but district boundaries were readjusted in 1990. Thus, the population registers from 1991 to 2015 at the district level were used in the later analysis.

The numbers of births and deaths and net migrants recorded by the population register system were published at City/County level for 1992 to 2015, and the data were downloaded and used to understand the sources of population growth. Three decadal censuses for 1990, 2000, and 2010 were downloaded from the website of National Statistics (<https://www.stat.gov.tw/lp.asp?ctNode=549&CtUnit=384&BaseDSD=7&mp=4>). The census data are aggregated to the City/County level. Data from the registers and censuses were compared for consistency. Spatial-temporal trends of registered population at the district level were also analyzed in both visual and descriptive statistical manners.

4.2.2. Relationship between Population and Land Use

To understand the spatial-temporal relationships between the areal extent of urban land-use types (i.e. Residential, Employment, and Transportation) and population counts (p), linear regression models were run. The models were run for population size and areal extents of land use for the same years in a synchronized manner. In addition, regression models were run between population and Urban area extent (i.e. the sum of Residential, Employment, and Transportation land uses). Annual, five-year, and ten-year changes based on the annual land-use and population data were inputs to the linear regression models. Models were run with actual data (i.e. p and land-use area in km^2) as well as data normalized by district areas (i.e. population density in p/km^2 and land-use areal fraction in %).

4.2.3. Relative Timing between Population Growth and Land-Use Change

The relative timing between land-use expansion and population growth was assessed using lagged correlation analysis, and each type of land use was tested exhaustively against population as a series of annual data using districts as the spatial unit of analysis. Initially,

the top 30 districts with the greatest population growth or population density growth were selected for the lagged correlation analysis. Twenty-four of the top 30 population growth districts are also in the top 30 population density growth districts, so 36 districts were selected in total. Time series of population and land use are non-stationary in the temporal domain, which violates the normal distribution assumption of time series variables for any statistical test (Raffalovich, 1994). Thus, the linear growth trend of annual population and land-use data was removed before lagged correlation. Pearson's r test was applied to each pair of L and P_d , where L is the detrended areal extent of land use, P is the detrended population count, and d represents the number for lag years. d represents the relative timing between land-use expansion and population growth, and must satisfy the following conditions: $-10 \geq d \geq 10$ and $d \in \mathbf{Z}$. The relative timing between land-use expansion and population growth was determined based on the d value with the highest Pearson's r . Such a statistical test revealed whether land-use expansion tended to occur prior to population growth (i.e. $d > 0$), or vice versa (i.e. $d < 0$).

4.2.4. Hypothetical Processes of Urban Growth

Based on the hypothetical processes in section 4.1.4, the relative timing of population change and land-use change is hypothesized to be: Employment land (i.e., commercial and industrial areas) increases, followed by Residential land expansion, and finally population increases. For the selected 36 districts, the relative timing between population growth and Residential expansion, and relative timing between population growth and Employment expansion were identified by lagged correlation tests. Hence, the change sequence relationships among population growth, Employment land expansion, and Residential land expansion can be determined, enabling testing of the hypothesis.

4.3. Results

4.3.1. Spatial-temporal Trends in Population and Land Use

Population data from registers and the census are compared in Figure 18. Population growth resulting from net migration is confirmed for cities and counties having higher census-enumerated populations than the registered population. New Taipei, Taoyuan, Hsinchu Cities, and Hsinchu County exhibited positive net migration, while Keelung City, Miaoli, and Yilan Counties experienced out-migration. New Taipei City has the highest population in the study area followed by Taipei, and Taoyuan Cities (see Figure 18a) over time. According to the registered natural increase and net migration data (Figure 18b), New Taipei, Taoyuan, Hsinchu cities, and Hsinchu County were confirmed to have high net in-migration along with high natural increase. During the entire study period, population growth from net migration in percentage terms was 37%, 58%, 30%, and 47% for New Taipei, Taoyuan, and Hsinchu cities, and Hsinchu County, respectively.

The spatial-temporal trend of registered population and land use are shown in Tables 19 and 20, as well as in Figures 19 and 20. Of the 107 districts, 66 experienced population growth, while the remaining 41 districts exhibited population loss for the entire study period. The maximum of population density was lower (from 40190 to 36607 people/km²), which indicates that depopulation occurred in the core of a metropolitan area. Population originally concentrated in Taipei and Taoyuan Cities in 1991, and then increased in the southwestern districts of the 1991 populous districts (see Figure 19a to c). Districts that experienced population growth are primarily located in the periphery of three metropolitan areas and the cores of Taoyuan and Hsinchu Metropolitan areas (shown in Figure 20a). Population loss

was found in some districts within Taipei and Keelung cities along with districts in rural areas.

Table 19. Descriptive statistics of registered population, population density, and population change at the district level. Change is the difference between 1991 to 2015.

	District area (km ²)	Population count 1991	Population count 2015	Population count change	Population density 1991 (pop/km ²)	Population density 2015 (pop/km ²)	Population density change (pop/km ²)
Maximum	769.9	542942	554236	181089	40190.7	36607.7	11345.2
Mean	86.4	86723	104219	17495	4369.1	4699.6	330.5
Median	54.2	49581	53634	2725	745.1	1124.5	53.9
Minimum	4.5	3247	4606	-39049	6.9	7.8	-4884.1
Standard Deviation	125.6	100481	114591	35337	8142.7	7846.3	1724.7

Table 20. Descriptive statistics of Residential, Employment, and Transportation at the district level. Change is the difference between 1991 and 2015. % indicates the percent coverage of district area.

	Res 1991 (%)	Res 2015 (%)	Res change (%)	Emp 1991 (%)	Emp 2015 (%)	Emp change (%)	Tra 1991 (%)	Tra 2015 (%)	Tra change (%)
Maximum	53.0	55.6	8.9	34.2	37.7	10.1	15.8	15.9	1.7
Mean	9.2	10.9	1.6	8.2	10.5	2.2	3.1	3.6	0.4
Median	4.2	5.2	0.8	6.1	7.4	0.9	2.5	3.1	0.3
Minimum	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Standard Deviation	11.5	12.3	1.7	7.4	9.2	2.7	2.9	3.0	0.4

* Res = Residential; Emp = Employment; Tra = Transportation. Change is the difference between 1991 and 2015. % indicates the percent coverage of district area.

The spatial-temporal trends of the three types of land use correspond to the population dynamic trend, especially for Residential (Figure 20b). The three types of land use were in an expansion situation during the study period. Districts where local city halls are located tended to have higher population and urban land-use coverage than other districts. The distribution of population and land use corresponds to the hierarchical relationship among cities and their environs suggested by Friedmann (1966). The areal extent of Employment and Transportation increased along the west coast, especially in the northern areas of Taoyuan City (Figure 20c and d). On average, districts had 1.6 %, 2.2%, and 0.4% of

expansion for Residential, Employment, and Transportation (shown in Table 20), respectively.

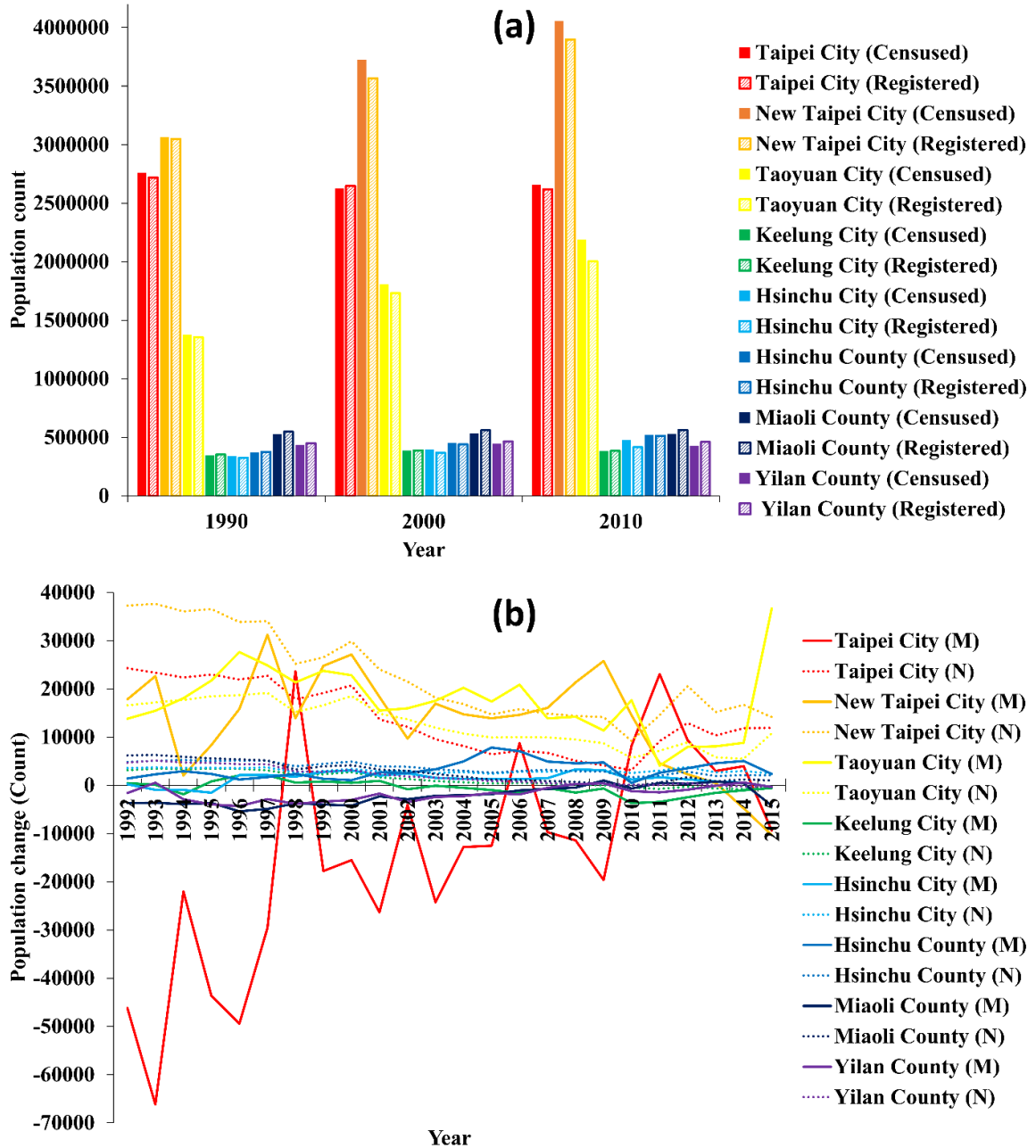


Figure 18. Temporal population dynamics for cities/counties with population data from registers and censuses. (a) shows the population count in the three census years. Cities and counties with more in-migrants had more people counted in the census than were registered as living there, while the remainder of counties experienced out-migration. (b) shows the annual natural increase (N) and net migration (M) data at the City/County level published by the registered population system.

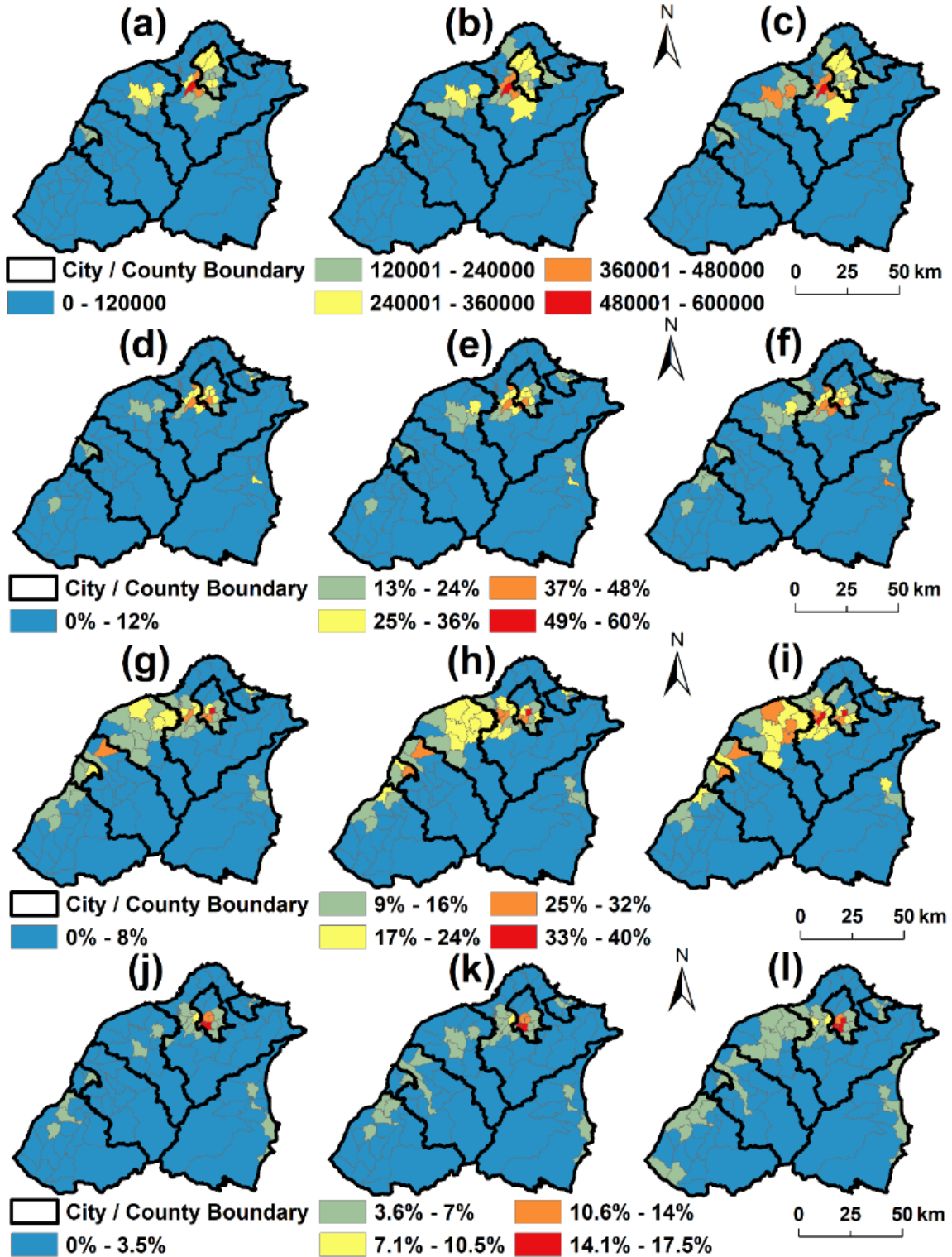


Figure 19. Equal interval choropleth maps of population and areal percent coverage of land use (% of district size) for 1991, 2003 and 2015. (a) population count 1991, (b) population 2003, (c) population 2015, (d) Residential land use (%) 1991, (e), Residential 2003, (f) Residential 2015, (g) Employment land use (%) in 1991, (h) Employment 2003, (i) Employment 2015, (j) Transportation corridor land use (%) 1991, (k), Transportation 2003, (l) Transportation 2015.

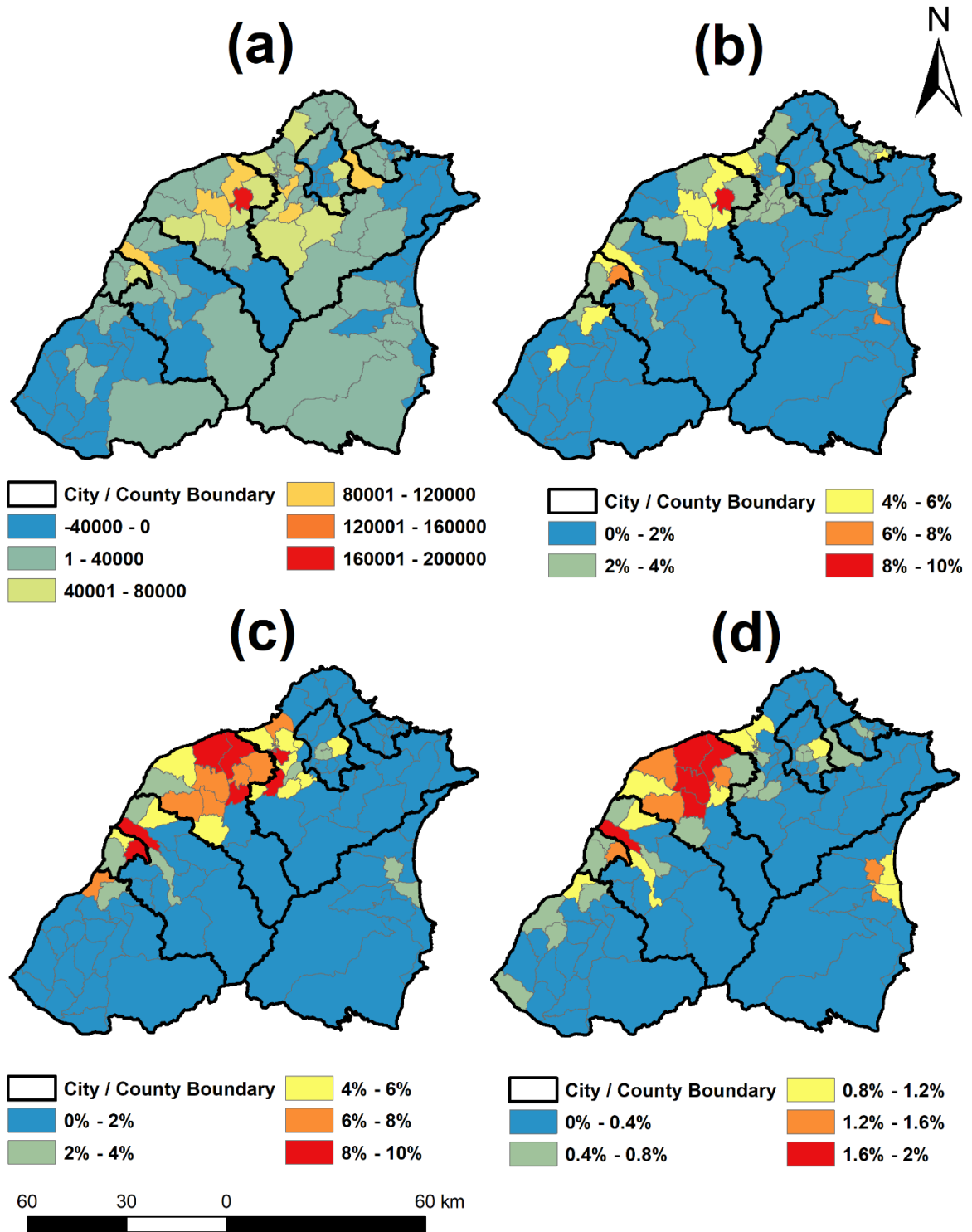


Figure 20. Equal interval choropleth maps of population count and land use (Residential, Employment, and Transportation) change (in % of district size) from 1991 to 2015. (a) population count change, (b) Residential land-use change, (c) Employment land-use change, and (d) Transportation land-use change.

4.3.2. Relationship Between Land Use and Population and Their Change

R^2 -values derived from the linear regressions are shown in Figure 21. Population was found to be positively related to the areal extent of land use at a statistically significant level for all land-use types. The relationship between population and Residential area is the strongest ($0.51 < R^2 < 0.57$ in Figure 21a). R^2 -values are higher when population and land-use are normalized by district areas, especially for Residential ($0.85 < R^2 < 0.89$). The R^2 -values for Transportation are moderate while they are lowest for Employment. However, R^2 -values were trending lower over time for normalized Transportation, while relatively stable for the remaining two types of land use.

R^2 -values of annual changes varied over time and are higher and more stable when the change data were derived from wider time intervals (shown in Figure 21b to d). Population change was found to be positively related to the areal change at a statistically significant level for all land use types. However, the relationship with the change data is weaker when the population and land-use changes are normalized by district areas. R^2 -values for Residential change (km^2) are the highest while lowest for Transportation change (km^2) (Figure 21d), which implies that Residential change moderately reflected population count change ($0.40 < R^2 < 0.56$). Thus, Residential area and its change are the land-use variables most related to population and population change, and change modeling and estimation between population and land use should be examined with a broader time interval for a more stable relationship.

4.3.3. Model Population Change based on Land-Use Expansion

Land-use expansion (i.e. increase in area of a given land-use type) was found to occur prior to population growth for most of the 36 districts except for districts in the metropolitan

cores, so land-use expansion can be inferred to be a driver of population growth, especially for Residential and Employment land expansion (shown in Figure 21d). Thus, Residential and Employment expansion were used as predictors of population change for each district of the entire study area, and four models were tested for modeling the 10-year population change due to the lowest variation in R^2 -values observed in Figure 21b to d. For one model, population change during a period was modeled by an ordinary least squares (OLS) regression with inputs of the same period of Residential and Employment change (synchronized OLS). This model was used as the standard for comparison with the other three models. Time lags for population change were incorporated into the second model, and the median of time lags was used because of the dispersed distribution of time lags, as shown in Table 22. Each 10-year population change was modeled by OLS regression with Residential change three years prior and Employment change four years prior as predictors (Time lag OLS). For example, to model population change from 1995 to 2005, the independent variables are Residential change from 1992 to 2002 and Employment change from 1991 to 2001.

Spatial effects related to population change were considered in the third model, and geographically weighted regression (GWR; Fotheringham et al. 1998) was utilized to estimate local relationships between land-use change and population change due to the possible nonstationary relationships among the three variables (i.e. spatial non-stationarity). GWR was especially useful for such situations because population varied positively and negatively over space and time. Population change for a 10-year interval was modeled with GWR, and the predictors corresponded to the same period of Residential and Employment change (Synchronized GWR). We used GWR with an adaptive bandwidth based on Akaike

information criterion correction (AICc) as the criterion to determine a suitable neighbor size at the district level. The time lags for Residential and Employment change were incorporated in the final model. GWR was implemented to model population change for a period with predictors of Residential change at a three-year earlier period and Employment change at a four-year earlier period (Time lag GWR).

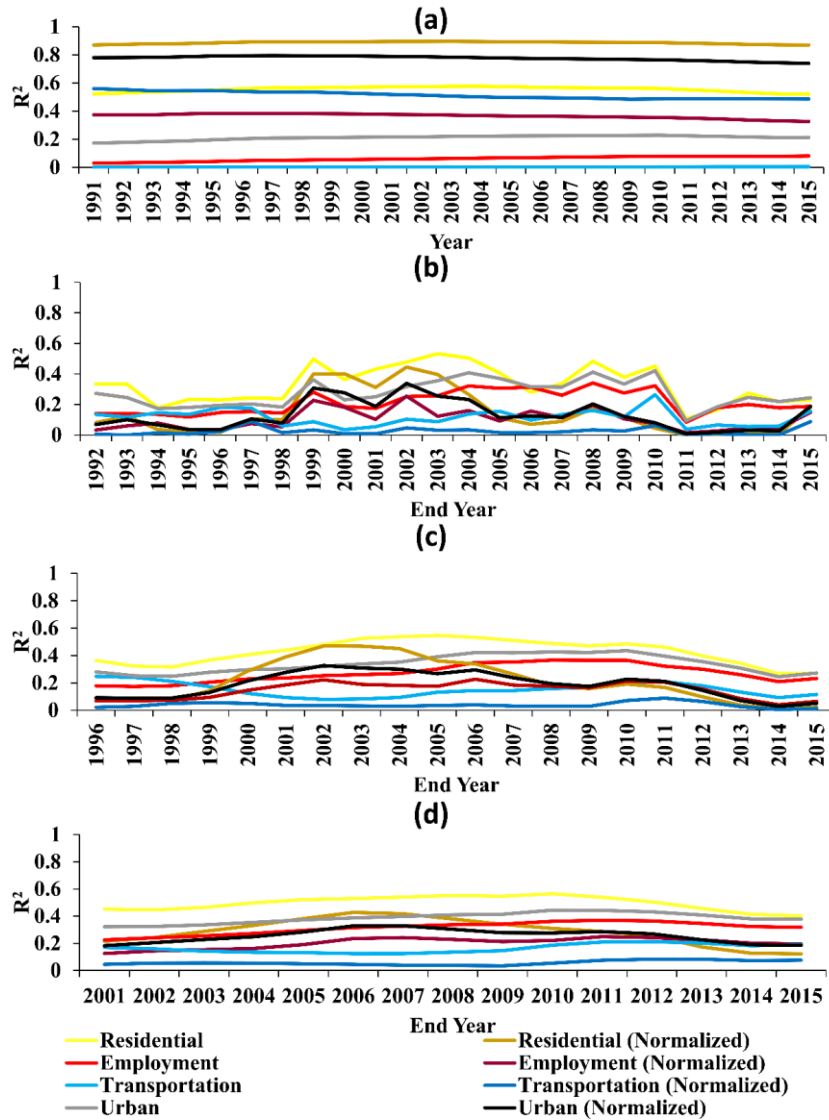


Figure 21. R^2 between population and land use and their changes. The legend only shows the corresponding land use, and population is the independent variable for all linear regression. Normalized refers to both population and the land use being normalized by the size of districts. The x-axes of (b)-(d) show the end year of the time interval. (a) annual population and land use. (b) annual change of population and land use. (c) 5-year change of population and land use. (d) 10-year change of population and land use. R^2 -values derived from population, land use area, and their changes were derived from the same year or same period without time lag.

Table 21. Districts with the greatest population change or population density change in the descendent order of population change. Population is in count. Employment, Residential, and Transportation Corridor are in km². Population density is in population count / km².

District	Cy	Pop 1991	Pop 2015	Emp 1991	Emp 2015	Res 1991	Res 2015	Tra 1991	Tra 2015	area	Pop Chg	PopDen Chg	Res Chg	Emp Chg	Tra Chg
Taoyuan	TYC	246056	427145	6.09	8.71	7.96	11.05	1.35	1.84	34.3	181089	5273	3.08	2.62	0.49
Zhongli	TYC	276878	389782	10.92	16.38	9.14	12.38	2.42	3.73	76.1	112904	1483	3.25	5.46	1.31
Xinzhuang	NTPC	308293	413243	5.77	6.80	5.90	6.72	0.90	0.97	20.6	104950	5074	0.82	1.04	0.07
Zhubei	HO	67178	170790	5.14	9.62	2.21	4.57	1.45	2.26	50.3	103612	2056	2.36	4.48	0.81
Xizhi	NTPC	97261	196028	2.85	3.36	5.00	6.17	2.06	2.41	71.7	98767	1376	1.18	0.51	0.35
Luzhu	TYC	58156	155403	9.94	17.37	3.47	7.12	1.61	2.83	74.8	97247	1299	3.65	7.42	1.22
Tucheng	NTPC	142348	238703	4.27	5.50	3.42	4.20	0.84	0.97	29.9	96355	3218	0.78	1.23	0.13
Luzhou	NTPC	112560	200095	1.50	1.82	3.27	3.64	0.31	0.32	7.7	87535	11345	0.37	0.32	0.02
Danshui	NTPC	85980	162221	4.13	5.39	5.20	7.48	1.26	1.54	73.0	76241	1043	2.29	1.26	0.28
Neihu	TPC	214750	287566	2.71	4.34	3.77	4.63	1.34	1.66	31.9	72816	2276	0.85	1.62	0.32
Shulin	NTPC	115581	184249	4.87	7.43	3.94	5.09	1.05	1.23	31.9	68668	2151	1.14	2.56	0.18
Pingzhen	TYC	150703	217887	6.30	9.56	4.79	6.83	1.60	2.37	43.4	67184	1545	2.05	3.25	0.78
Xindian	TPC	233277	300267	3.80	4.76	8.03	9.80	2.01	2.20	121.1	66990	552	1.78	0.96	0.19
Linkou	NTPC	34125	100350	4.83	7.62	3.99	6.93	1.37	1.97	54.2	66225	1219	2.94	2.80	0.61
Yangmei	TYC	98943	161098	8.17	14.34	3.91	6.61	2.24	3.35	89.6	62155	693	2.70	6.18	1.11
Bade	TYC	135897	187420	5.33	8.76	3.60	5.08	0.81	1.18	33.7	51523	1524	1.48	3.43	0.37
Sanxia	NTPC	61627	112708	2.91	4.28	3.56	4.96	1.39	1.55	189.5	51081	269	1.41	1.37	0.16
East	HC	157676	208142	7.42	10.65	5.66	7.88	1.48	1.99	34.8	50466	1448	2.23	3.23	0.51
Guishan	TYC	95756	145580	11.77	16.60	6.04	8.58	2.12	2.63	71.4	49824	697	2.55	4.83	0.51
Wenshan	TPC	225755	275367	3.28	3.48	4.50	4.91	1.40	1.46	31.2	49612	1588	0.41	0.20	0.06
Longtan	TYC	82698	118433	9.63	13.23	4.35	5.72	1.74	2.11	75.9	35735	470	1.37	3.60	0.37
Zhonghe	NTPC	379968	414304	2.44	3.25	6.46	7.19	0.87	0.94	19.4	34336	1767	0.73	0.82	0.07
Wugu	NTPC	48796	82923	5.29	7.20	1.46	1.77	1.35	1.48	34.3	34127	994	0.31	1.91	0.13
North	HC	120338	149156	3.82	4.81	4.34	5.41	0.48	0.60	23.9	28818	1203	1.08	0.99	0.13
Xiangshan	HC	50897	76762	6.86	8.54	2.81	4.15	1.63	2.00	65.6	25865	393	1.35	1.68	0.37
Taishan	NTPC	53658	78707	3.15	4.77	1.46	1.75	1.17	1.30	18.0	25049	1385	0.29	1.62	0.13
Dayuan	TYC	61219	85565	15.97	23.41	3.15	5.57	1.68	3.29	90.1	24346	270	2.42	7.45	1.61
Toufen	MO	79629	103239	4.87	6.25	4.35	6.87	1.83	2.17	52.9	23610	445	2.52	1.38	0.34
Hukou	HO	54078	77280	14.01	16.54	2.07	3.17	2.00	2.52	58.0	23202	399	1.10	2.53	0.52
Zhunan	MO	62792	84469	5.86	8.49	3.78	5.40	1.89	2.34	41.6	21677	520	1.63	2.63	0.45
Anle	KC	60333	81329	1.24	1.40	2.12	2.76	0.62	0.71	18.5	20996	1131	0.64	0.16	0.08
Yingge	NTPC	67214	87965	3.11	4.21	2.26	2.82	1.01	1.14	21.8	20751	950	0.56	1.10	0.13
Banqiao	NTPC	542942	554236	3.00	3.68	8.61	9.25	1.42	1.49	21.2	11294	530	0.64	0.69	0.06
Sanzhong	NTPC	378397	388447	3.20	3.40	5.78	5.91	1.48	1.52	17.9	10050	558	0.14	0.20	0.04
Xinyi	KC	42547	51604	0.70	0.79	1.61	2.14	0.23	0.27	10.7	9057	839	0.53	0.09	0.04
Luodong	YO	65456	72404	1.03	1.19	3.40	4.08	0.58	0.73	11.2	6948	616	0.69	0.16	0.15

*Cy=City/County, Pop=population, Emp=Employment, Res=Residential, Tra=Transportation, Chg=change in original units (count for population, and km² for land use), Den=density, TPC=Taipei City, NTPC=New Taipei City, TYC=Taoyuan City, HC=Hsinchu City, HO=Hsinchu County, KC=Keelung City, MO=Miaoli County, YO=Yilan County.

Table 22. Time lag (in years) between population and land use identified by max lag correlation. Positive sign of time lag indicates that population growth occurred later than land-use change, while the negative sign of time lag indicates that population growth occurred earlier than land-use change.

	Residential	Employment	Transportation	Urban
Mean	1.5	1.3	-0.1	1.0
Standard deviation	6.7	7.8	7.1	7.8
Minimum	-10	-10	-10	-10
25% quantile	-2	-6.5	-6	-8
median	2.5	3.5	-1.5	2.5
75% quantile	6	9	8	9
Maximum	10	10	10	10

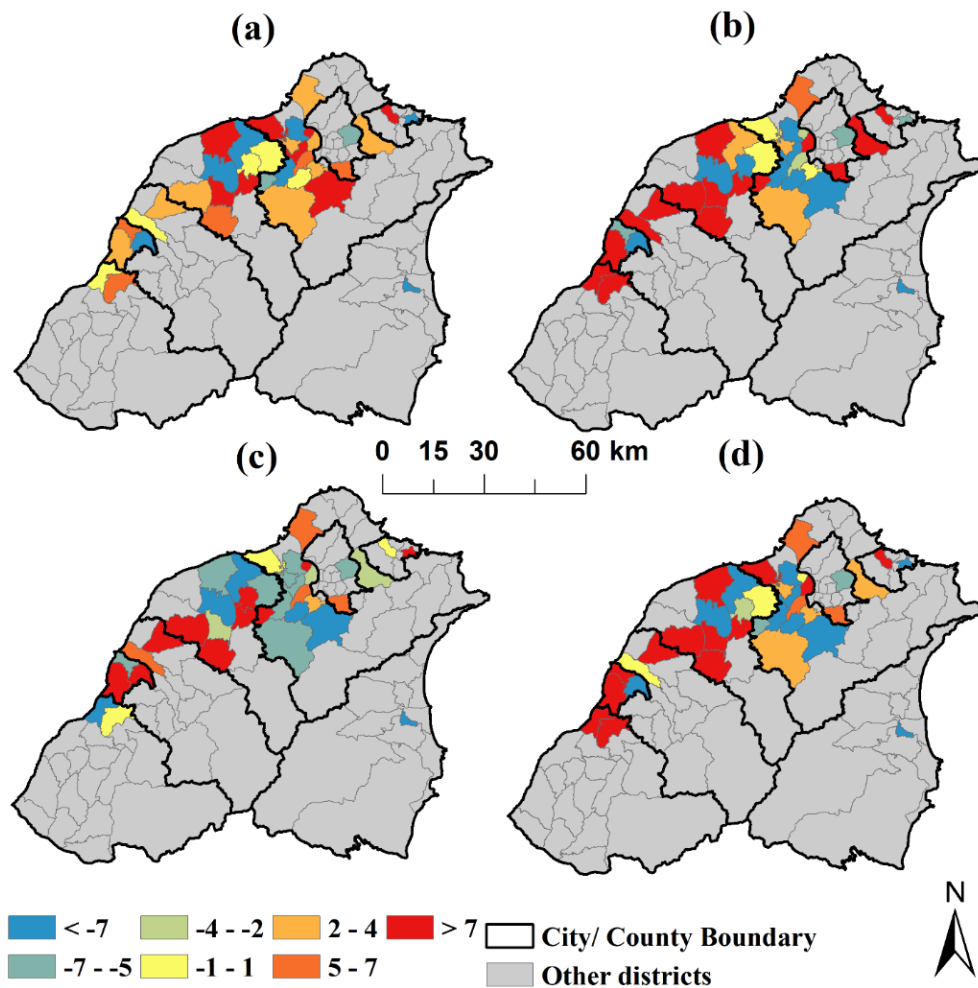


Figure 22. Maps depicting time lag (in years) between population growth and land-use expansion by district. (a) Residential land use, (b) Employment land use, (c) Transportation corridor, and (d) General Urban (the sum of three land-use types). Positive signs of time lags indicate that population growth occurred later than land-use expansion, while the negative signs indicate that population growth occurred earlier than land-use expansion.

Adjusted R^2 -values were used to evaluate the four models, as shown in Table 23. Synchronized OLS yielded the lowest adjusted R^2 -values, while Time lag GWR resulted in the highest adjusted R^2 -values. Adjusted R^2 -values slightly increased when only time lag between population change and land-use change was incorporated. On the other hand, the adjusted R^2 -values substantially increased in the synchronized GWR. The synchronized and Time lag GWR models yielded statistically significant improvement compared to the Synchronized and Time lag OLS models based on ANOVA tests (p -values < 0.0001). The reason why the adjusted R^2 -values were substantially boosted is that districts that experienced population growth or decline were modeled together with Residential and Employment changes, but the impacts from Residential and Employment changes varied over depopulating and population-growing districts. Specifically, local R^2 -values and coefficients of predictors of Time lag GWR only slightly vary over different time periods, and the metrics of an exemplar model of Time lag GWR (i.e. population change from 2005 to 2015) are shown in Figure 23. The local R^2 -values are high (local $R^2 > 0.62$) for districts in the western part of the study area and gradually decrease to the southeast direction (Figure 23a), which implies that the local models among population change, Residential change, and Employment are accurate in the western part of the study area. In terms of intercepts (Figure 23b), districts around the boundary between New Taipei and Taoyuan cities had the highest local intercept, which indicates that population growth in these districts were substantially higher than other districts. Residential change (Figure 23c) had the greatest positive impacts on population change in districts of Taoyuan city, and the impacts gradually decreased. Employment change (Figure 23d) had both positive and negative impacts on population change. Districts in Keelung and northeast part of New Taipei City had the greatest positive

impacts on population change due the expansion of science industrial parks and nuclear power plants. However, Employment change had negative impact on districts of Taoyuan City, the western part of New Taipei City, and the eastern part of Yilan County because Employment and Residential usage competed for the limited undeveloped land these districts. Thus, both spatial effects and time lags between population change and land-use change should be considered for future modeling of population change with land-use change as predictors.

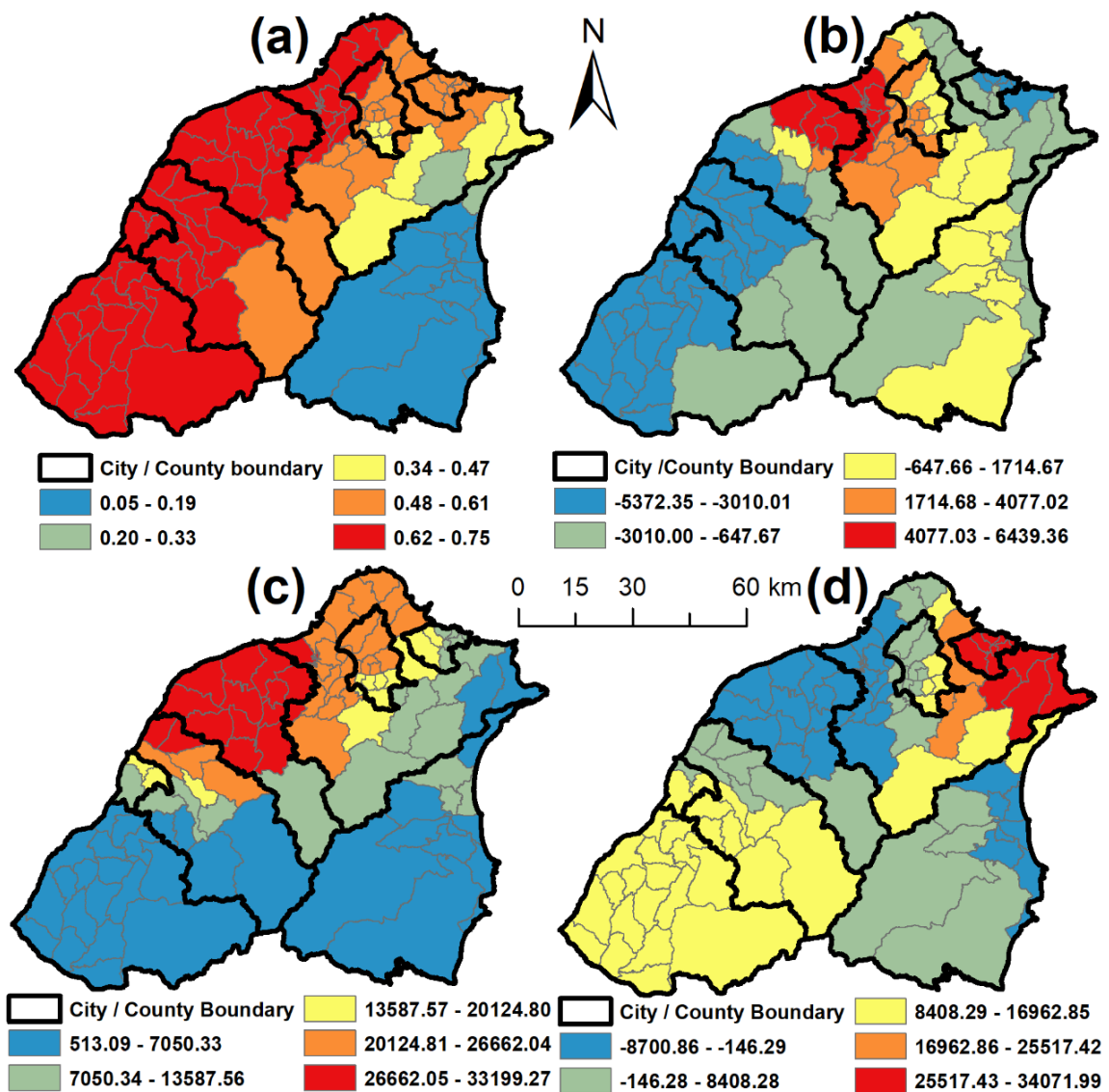


Figure 23. Equal interval choropleth maps of local R^2 and coefficients derived from geographically weighted regression for modeling population change from 2005 to 2015 with Residential change from 2002 to 2012 and Employment change from 2001 to 2011. (a) Local R^2 . (b) Coefficients of local intercept. (c) Coefficients of local Residential change. (d) coefficients of local Employment change.

Table 23. Adjusted R² of 10-year interval change in population and land use

year	Synchronized OLS	Time lag OLS	Synchronized GWR	Time lag GWR
2001	.45	-	.60	-
2002	.44	-	.59	-
2003	.46	-	.61	-
2004	.49	-	.67	-
2005	.52	.55	.69	.72
2006	.52	.57	.68	.76
2007	.53	.60	.68	.76
2008	.54	.60	.84	.79
2009	.53	.60	.73	.79
2010	.55	.61	.66	.78
2011	.53	.60	.67	.73
2012	.49	.58	.66	.72
2013	.45	.57	.65	.73
2014	.41	.54	.65	.78
2015	.41	.54	.65	.73

4.3.4. Hypothesis Testing

To test the hypothesis that Employment land use expanded, followed by Residential expansion, and then population growth (described in section 4.1.4), the relative timing among population growth, Residential, and Employment expansion was examined by the time lags of population-Residential and population-Employment. According to Table 22, the general change sequence can be determined based on the median of time lags among the three variables over the 36 districts. According to the median of time lags, the general progression was Employment expansion occurring first, followed by Residential expansion, and then population growth at the end, which matches my hypothesis.

The specific change sequences for the 36 districts are mapped in Figure 24. Twelve of the 36 districts matched the change sequence of Employment-Residential-population, while other districts had other sequences. I found that the cores of metropolitan areas (e.g. Taoyuan, Zhongli, and East districts) tended to exhibit population growth prior to land-use

change (i.e. the districts showing in reddish hues). On the other hand, districts in the periphery of the metropolitan areas exhibited Employment and Residential expansion prior to population change (i.e. the districts showing green and blue hues).

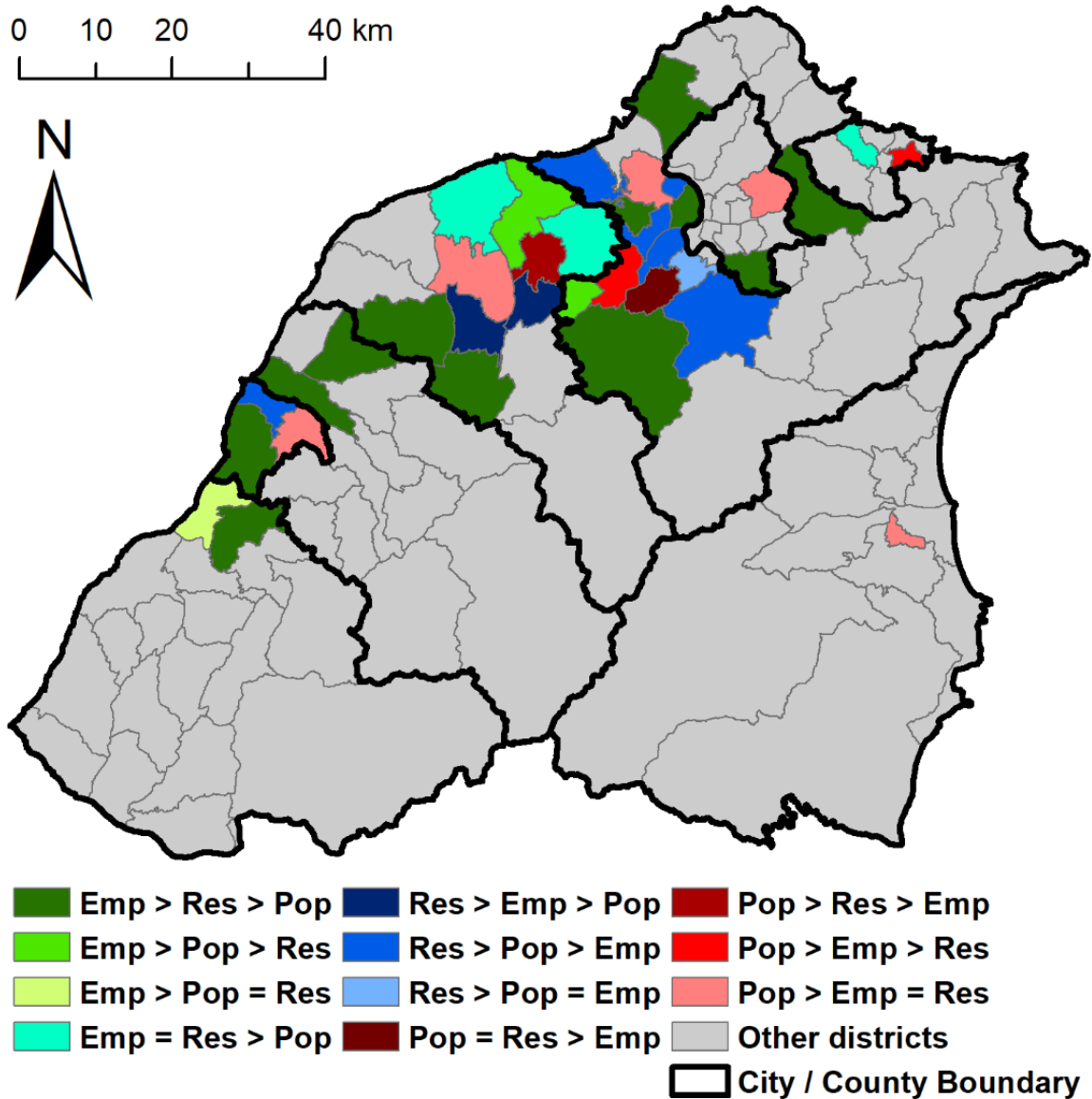


Figure 24. Map of change sequence among population (Pop), Residential (Res), and Employment (Emp) for the 36 districts with greatest population change or population density change. Emp > Res refers that employment expansion occurred prior to Residential. Emp = Res indicates Employment and Residential expansion occurred at the same time.

My explanation for why some districts with population growth experienced different change sequences than the one I proposed is that cores of metropolitan areas provide more

jobs (e.g. tertiary and quaternary sectors of economy) than the periphery and other agriculture-based districts. When people moved into the core area, landlords divided existing housing units in the core area into smaller rooms and rented to the in-migrants. People crowded into the core area as renters even though available residential spaces were less than the demand. With landlords' permission, tenants could change their permanent addresses to their renting places for their child(ren) for attending school close to the renting addresses. Residential expanded slowly in the core areas owing to high land prices and limited undeveloped space. Additionally, construction companies perceived the need for residential spaces in the core areas, and developed new residential areas in the periphery, where land prices were lower, before the added population arrived. Although people moved into the periphery after residential construction was completed, they still commuted to the core areas for work. Therefore, the spatial connection between the core and periphery of metropolitan areas should be considered along with the hypothesis I proposed.

4.4. Discussion and Conclusion

4.4.1. Relative Timing and Relationships Between Population Change and Land-Use Change

To examine differences in time lags at two administrative levels (city/county and district), the relative timing between population change and land-use change was examined for Hsinchu City and its associated districts (shown in Table 24). Population growth was found to occur nine years later than three types of land-use expansion for Hsinchu city, and Xiangshan and North districts that had similar time lag relationships. On the other hand, the time lags relative to population growth for East district were -10, -10, and 10 years for

Residential, Employment, and Transportation, respectively. The differences in time lags derived from East district were lost when the district was aggregated to Hsinchu City. Thus, the relative timing between population change and land-use change is affected by the modifiable areal unit problem (MAUP), and the differences due to MAUP should be explored in future studies. However, aggregation from district to city level is inappropriate for cities with large coverage (e.g. Taoyuan and New Taipei Cities) because population growth in a district hardly drives or is driven by land-use change in other distant districts within the same city. For example, Sanxia and Danshui districts exhibited great population growth and Residential land expansion under New Taipei City, but the Euclidean distance between the two districts is 33 km and there are no rapid transportation connections. Therefore, population growth in Sanxia district should only drive or be driven by Residential expansion in the same district. However, population growth in Sanxia district was deemed to be a driver, or be driven by Residential expansion in Danshui district after aggregating into New Taipei City. Thus, time lags derived from the aggregated level data are erroneous for these large cities.

Table 24. Time lags (in years) for Hsinchu City and its associated district

	Residential	Employment	Transportation	Urban
Hsinchu City	9	9	9	9
East district	-10	-10	10	-10
North district	6	6	-7	8
XiangShan district	3	9	8	9

The statistical relationship between population and Residential proved to be the strongest compared to relationships with other land uses. A district with extensive residential fraction indicates high population density at present, and more residential expansion refers to higher population growth in the future. In past studies (e.g. Alig and Healy 1987 and Stow et al.

2016), built area (General Urban in our study) was directly used as a dependent variable with population as a predictor. However, for this study I found that the areal extent of Residential is more relevant than directly using built area. Moreover, I found that population growth tended to occur later than Residential and Employment expansion in general, so population change seemed to have been driven by land-use change.

4.4.2. Urbanization Statuses of Metropolitan Areas

Population, urban activities, and their changes were used to represent the stages of urbanization in van den Berg et al. (1982) and Myrdal (1958). Here I utilized relative changes of population and land use (as proxies of urban activities) to examine the three metropolitan areas. Van den Berg et al. summarized four stages of urbanization based on population change, including urbanization, suburbanization, disurbanization, and reurbanization. Population increases and concentrates in the core of a city during the urbanization stage, and then population keeps increasing in the suburbs during the suburbanization stage. Later during disurbanization, the city core experiences depopulation, while the suburbs experience continued population growth. Finally, during reurbanization, the suburb experiences depopulation, and yet the city center exhibits population regrowth through the process of gentrification. Population and land-use change in three districts over Taipei metropolitan area are shown in Figure 25 as a way of showing these urbanization stages. The core of the metropolitan area (i.e. Wanhua district), experienced depopulation before 2010, while Luzhou (7 km away from the core in the northwestern direction) and Danshui districts (11 km away from the core in the same direction) experienced population growth during the same period (Figure 25a). Population kept increasing after 2010 in Danshui district, while Wanhua district had slightly population growth. Suburbanization,

disurbanization, and reurbanization are found in the study period in the three districts for Taipei metropolitan area. Population continued to grow in the core and the suburb of Taoyuan and Hsinchu metropolitan areas (shown in Figure 19 and 20, and Table 21), which corresponds to urbanization and suburbanization stages.

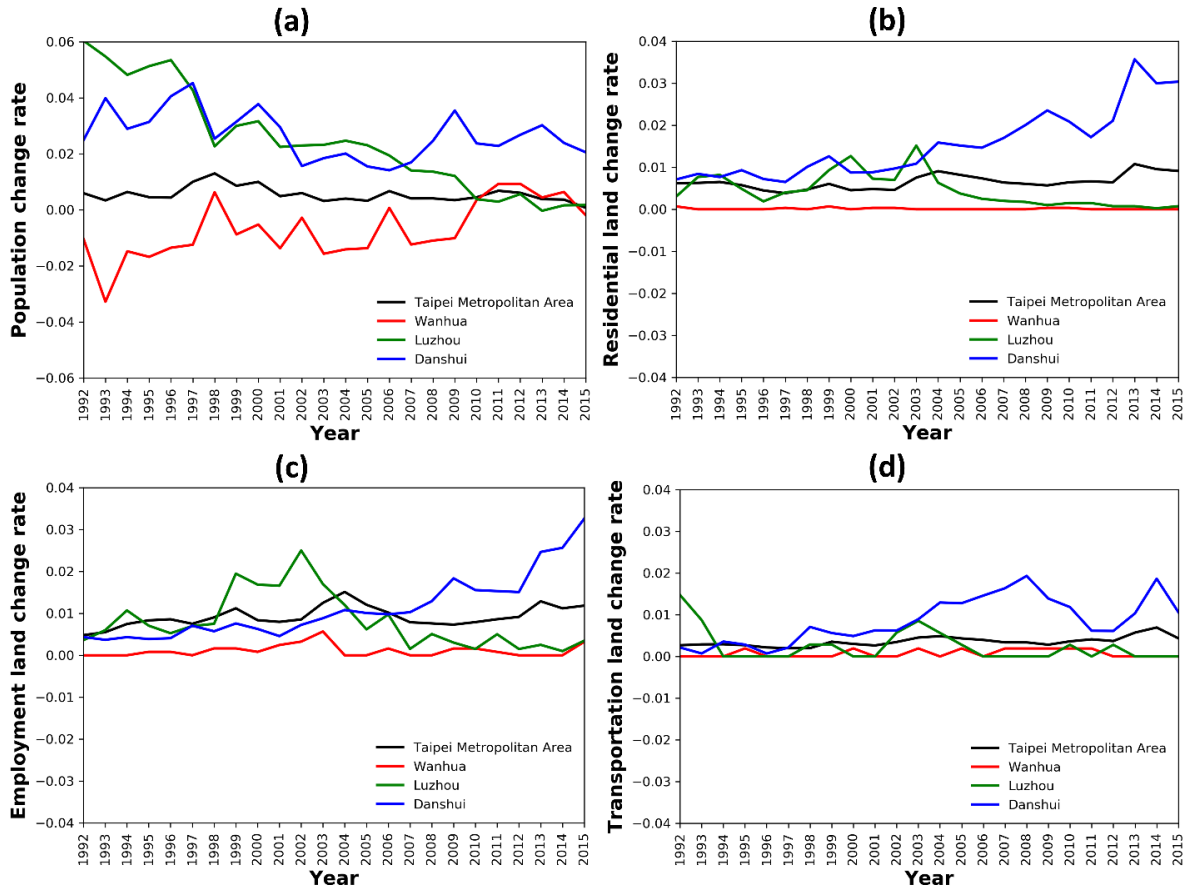


Figure 25. Time trajectory plots of annual change rate of population and land use within Taipei metropolitan area with exemplar districts. The metropolitan area includes Taipei City and New Taipei City. Wanhua district represents for the core of the metropolitan area; Luzhou district is 7 km northwestern away from the core, and Danshui is 11 km northwestern away from the core. (a) Population change, (b) Residential land change, (c) Employment land change, and (d) Transportation corridor land change.

The polarization and spread model proposed by Myrdal (1958) is found in the relative change rate of land use, as well (Figure 25b to d). Myrdal focused on urban activities instead of population change as urbanization processes. The polarization phase is defined as urban activities increasing until saturated in an urban area. And then, the urban activities spread

out to the hinterland of the urban areas after the polarization. Wanhua district had the lowest land-use change rate, Luzhou district had moderate land-use change rate, while Danshui had the highest change rate. Residential and Employment expanded substantially in Luzhou district before 2003, and then the expansion continued in Dnashui district after 2003. Employment land, including commercial and industrial land use, is a typical indicator of urban activities. More undeveloped land was used for Employment or Residential in Luzhou district over time, and then the spread of employment land occurred after the polarization in the Luzhou district.

4.4.3. Limitations, Challenges, and Follow-on Research

Several limitations should be addressed that likely influenced the results of this study. Differences between the people enumerated in the population register and the census exist, and the actual population for districts in urban areas is underestimated while overestimated for rural districts. The approaches I used to derive annual land-use data were designed within the context of urban expansion, and we assumed that no changes occurred within developed areas. However, urban renewal and park conversion were found in the study area (e.g. Taipei Bus Station, Daan Forest Park, and Songshan Cultural and Creative Park), but the occurrence of this land-use type is unknown before urban renewal and park conversion according to the mapping approaches I used.

Mobile phone signal data with geographical coordinates can be directly used for inferring land use (Toole et al., 2012) and population estimates (Deville et al., 2014). Toole et al. derived weekday-weekend human activity schedules from individual-based time series mobile phone data, and then the type of land use can be inferred based on the weekday-weekend daily activity schedule. Deville et al. estimated population density based on mobile

phone signals at fine spatial (cellular tower zone) and temporal (seasonal) level. The resultant land-use map and estimated population are accurate and precise. Thus, mobile phone signals and web-based socioeconomic data (e.g. Yelp, Google maps and StreetView) could be incorporated into the land-use mapping processes to improve the accuracy of land-use maps derived from remotely sensed imagery, as well as into population estimation at fine spatial and temporal scales.

The estimated relative timing between population change and land-use change is restricted to the study period for which I have data. The entire trajectories of population and land use were not recorded in the data for some developed districts, and the initial statuses of population and land use remain unknown for these data. Thus, time lags that I estimated through lagged correlation could be mis-specified. Therefore, more case studies should be conducted with the same approaches with a longer study period to reconfirm my findings and test my hypothesis of urbanization processes.

4.4.4. Conclusion

In this chapter, the relative timing between population growth and land-use change was assessed based on a case study of north Taiwan from 1990 to 2015. Annual land-use data (Residential, Employment, and Transportation corridor categories) were generated from the overlay analysis of the satellite image-derived urban expansion map and the land-use map at the end date of the study period. Annual registered population at district level were publicly available and acquired from a governmental website.

Three major contributions stem from this study. This is the first study to examine relationships between population and land-use changes at an annual time step. This was possible because of the unique availability of annual population data for Taiwan, and readily

accessible and free long-term archive of Landsat surface reflectance data. Linear regression models were run for each type of land use versus population to understand the general relationship between the abundance of land use and population. Additionally, the relative timing between population growth and land-use change was estimated with lagged correlation. Finally, the hypothetical change sequence of population growth and land-use expansion was validated within the context of urbanization.

The areal extent of Residential land use (and change) was most related to population count (and change), and the relationship is stronger when Residential and population were normalized by district areas. Population growth generally occurred 2.5 years later than Residential land expansion based on the median time lag, and most districts experienced population growth later than Residential land expansion. The expansions of Residential and Employment land uses were deemed to be the drivers, and population change was inferred to be the result of changes in Residential and Employment. Synchronized OLS, time lag OLS, synchronized GWR, and time lag GWR were applied to model population change with the Residential and Employment change as predictors. Population change modeled with GWR, along with time lags of Residential and Employment change, was found to have the highest adjusted R^2 -values. A third of districts with population growth had the change sequence of population growth and land-use expansion we hypothesized, while other districts had different change sequences due to the closeness of metropolitan cores that provide abundant jobs.

Chapter 5. Conclusion

5.1. Key Contributions and Findings

This appears to be the first study to examine relationships between population and land-use changes at an annual time step, made possible by the availability of annual population data for Taiwan, and the long-term archive of Landsat surface reflectance data. A novel approach to identifying the type and estimating the time of urban land change based on dense time series of V-I-S proportion maps derived from Landsat imagery was developed and tested. Additionally, land-cover proportions and their derivatives were tested as input features of RF classifiers for the ability of generating accurate detailed land-use maps. Finally, the relationship and relative timing between population growth and urban land-use change were explored, and the cause-effect relationship between population change and land-use change were determined.

Several findings resulted from the study of identifying urbanization based on dense time series of V-I-S proportion maps derived from Landsat imagery. Estimated proportion of Impervious cover was the most accurate compared to the estimated proportions of Vegetation and Soil, likely because multiple endmembers for impervious materials were used in the NSMA modeling. The estimated V-I-S proportions derived from Landsat 8 OLI imagery were found to be the most accurate compared to the V-I-S proportions derived from TM and ETM+ imagery, likely due to the improvement of sensors specification. Additionally, the estimated V-I-S proportions for Landsat imagery acquired in winter seasons are the most accurate while those from spring season images are least accurate.

The algorithm for identifying and mapping urbanization yielded different accuracies for Urban Expansion compared to Urban Renewal processes. The reasons for why the

accuracies are different lies in the different spatial and temporal scales associated with the two types of urbanization. Urban Expansion areas tended to occur as larger change parcels and the V-I-S trajectories are relative stable before and after change, which can be easily identified based on dense time series of Landsat imagery. However, Urban Renewal areas were challenging to identify on the Landsat time series because the Urban Renewal signal, mostly exposed soil, occurs for a short amount of time window (shorter than a month) and as small spatial units (less than or equal to three Landsat pixels).

For urban land-use classification and mapping, RF classification routines were applied to (1) understand feature importance of Landsat surface reflectance and its derivatives, and (2) derive accurate land-use maps in Chapter 3. The feature importance derived from RF classification routines was informative for such purpose, while directly including V-I-S proportions along with Landsat surface reflectance as input features of RF classifiers led to the overfit RF classifiers. However, spatial and temporal derivatives of V-I-S proportions (i.e. GLCM Homogeneity of V-I-S and Vegetation variation) were proved to be more contributive than the spatial-temporal derivatives directly derived from surface reflectance proposed by Zhu et al (2012). V-I-S proportions can be used to separate non-urban land use from urban land use (e.g. Forest versus Residential), while V-I-S proportions of multiple urban land-use types (e.g. Residential versus Public Use) are too heterogeneous and spectrally similar to discriminate from each other.

In Chapter 3, two land-use change maps were derived and compared in terms of accuracies, and the land-use change map derived from overlay analysis was confirmed to be more accurate than the post-classification change map. The accuracies of the two change maps depended on the accuracies of three maps which are the 1990 and 2015 land-use maps,

and urban expansion map. The accuracy of the urban expansion map is high from exploiting the relatively fine temporal resolution of Landsat image series, and the accuracy of the 2015 land-use map is high due to more reliable and extensive training data for the 2015 dataset. However, the 1990 land-use map is relatively inaccurate due to limited training data with high uncertainty. Therefore, deriving reliable time series of land-use data within the context of urbanization should rely on dense time series of Landsat imagery instead of post-classification comparison.

The general relationship between population count (and change) and areal extent of land use (and change) was explored based on annual population registers and land-use areal extent data. Linear regression models were run for population versus Residential, Employment, Transportation, and Urban (sum of all three types of land use), and the areal extent of Residential is found to be more related to population counts and change than for areal extent of Urban.

This appears to be the first study to examine the relative timing between population change and land-use change and to identify the cause-effect relationships based on annual population and land-use data. The time lags between population change and land-use change vary over the core and periphery of metropolitan areas. Population growth tended to occur prior to land-use change in the core areas due to high land prices and limited developable space. On the other hand, land-use change tended to occur prior to population change in the periphery area because construction companies perceived the need for activity spaces in the core areas and developed new employment and residential areas.

5.2. Summary

Urban expansion is a form of LCLUC that occurs globally, and population growth can be a driver and be driven by LCLUC. Determining the cause-effect relationship is challenging because the temporal resolution of population data is generally limited by decadal or even less frequent demographic censuses for most countries. In this dissertation, the relative timing between population growth and land-use change was assessed based on a case study of north Taiwan from 1990 to 2015. Data on population were acquired from governmental-based population registers on an annual basis at the district level, and land-use data were derived from dense time series of Landsat imagery.

In Chapter 2, a semi-automatic approach of identifying the type and starting time of newly urbanized areas was developed and tested based on dense time series of V-I-S proportion maps derived from Landsat imagery. To cover the study area and study period, 295 dates of Landsat imagery were used. Spectral-radiometric endmembers were automatically derived from K-means clustering, and then the derived endmembers were used in NSMA models to generate V-I-S proportion maps. An algorithm was developed to identify locations and estimate dates of Urban Expansion and Urban Renewal. The onset for Urban Expansion is estimated with logistic regression, while the time of onset Urban Renewal is estimated based on brief periods of soil exposure during the period between building demolition and construction. Urban Expansion locations and starting time of change are reasonably accurate, with 80% of Urban Expansion samples being detected with an accuracy of ± 2.4 years. However, the accuracy of estimated Urban Renewal was low because the brief period of soil exposure is often too short to be identified. Thus, only Urban

Expansion pixels were preserved in the urban change map for deriving time series of land-use data.

In Chapter 3, RF classification was applied to examine the capability of mapping relatively detailed urban land-use based on Landsat images and derived V-I-S proportions. The original surface reflectance data contributed the most information to the RF classifier for urban land-use mapping, with GLCM homogeneity of V-I-S and Vegetation temporal variation also providing discrimination power. For two dates of Landsat imagery dataset (i.e. 1990 and 2015 datasets), land-use maps derived from the top 10 most contributing features (original surface reflectance, GLCM Homogeneity of V-I-S, and Vegetation temporal variation) are the most accurate, while the land-use maps derived from the original spectra and V-I-S proportion were the least accurate due to overfitting of RF classifiers. Accuracies of Level B land-use maps are insufficient to correctly identify land-use change, so the Level B maps were categorically aggregated to produce generalized (Level A) land-use maps. Level A urban land uses (Transportation Corridor, Employment, and Residential categories) required manual editing to yield urban land-use change information suitable for comparative analyses with population change (in Chapter 4). Two land-use change maps were derived: one by post-classification comparison of 1990 and 2015 maps, and the other by integrating the urban expansion map (summarized in Chapter 2) on the 2015 land-use map. The land-use change map derived from overlay analysis was confirmed to be more accurate (overall accuracy = 82.7%) than the post-classification change map for identifying urban expansion. Therefore, the land-use change map derived from overlay analysis containing the starting time of urban expansion was used to generate a time series of land-use data.

In Chapter 4, the spatial relationship and relative timing between population change and land-use change were explored at the district level. Annual land-use data (Residential, Employment, and Transportation corridor categories) from the overlay analysis in Chapter 3 were utilized. The spatial-temporal trends of population and land use were first analyzed in both visual and descriptive statistical manners. Linear regression models were run for each type of land use versus population to understand the general relationship between the areal extent of land-use types and population count. Additionally, the relative timing between population growth and land-use change was estimated with lagged correlation test. Finally, the hypothetical change sequence of population growth and land-use expansion was tested within the context of urbanization. I hypothesized that Employment land increases, followed by Residential land expansion, and finally population increases. The areal extent of Residential land use and change covaried most with population count and change, and the relationship is stronger when Residential and population data were normalized by district areas. Population growth generally occurred 2.5 years later than Residential land expansion based on the median time lag, and most districts, except districts in metropolitan cores, experienced population growth later than Residential land expansion. The expansions of Residential and Employment were deemed to be the primary drivers of population change. Synchronized OLS, time lag OLS, synchronized GWR, and time lag GWR were applied to model population change with the Residential and Employment change as predictors. Population change modeled with GWR along with time lags of Residential and Employment change yielded the highest adjusted R^2 -values. A third of districts with population growth experienced a change sequence of population growth and land-use expansion, as

hypothesized, while the remainder two thirds of districts had different change sequences due to the closeness of metropolitan cores that provide abundant jobs.

5.3. Future Research

Multiple forms of urban LCLUC occurred in the urban areas and their vicinities, including urban expansion, urban renewal, and park conversion. Algorithms for identifying both Urban Expansion and Renewal were developed and tested in this study, while identification of park conversion was not conducted and tested. The approaches I used to derive annual land-use data were primarily designed to map and explore urban expansion, and I assumed that no changes occurred within developed areas. Urban renewal, including creation of recreational spaces, occurred in the study area and period in communities such as Taipei Bus Station, Daan Forest Park, and Songshan Cultural and Creative Park. However, the Urban Renewal places cannot be identified based on the algorithms. To improve the accuracy of land-use change map, the urban LCLUC identification algorithm should be refined, tested, and compared with the Continuous Change Detection and Classification developed by Zhu (2014). The urbanization identification approaches should also be tested with Moderate Resolution Imaging Spectroradiometer (MODIS) image time series, both independently and integrated with Landsat image time series.

RF classification routines were implemented to generate urban land-use maps, and GLCM texture of surface reflectance (Zhu et al. 2012) was utilized in the feature ranking and classification. Spatial (i.e. GLCM textural measurement of V-I-S proportion) and temporal features (i.e. Vegetation temporal variation) that I tested proved to be useful for land-use classification of my densely urbanized study area. The same approaches should be

tested and compared in future studies for other cities in East Asia (e.g. Tokyo, Seoul, Hong Kong) that are densely urbanized and expanding in future studies.

Other sources of data archives should be compiled and stored in a spatial-temporal data repository along with remotely sensed data archives, for future refinement of the land-use map, to generate time series population data, and to conduct urban environment change simulations. For example, mobile phone signal data with geographical coordinates can be directly used for inferring land use (Toole et al., 2012). Toole et al. derived weekday-weekend human activity schedules from individual-based time series mobile phone data, and then the type of land use can be inferred based on the weekday-weekend daily activity schedule. Deville et al. (2014) tested mobile phone signal data derived from cellular towers with geographical coordinates for estimating population counts over space and time, and the population estimates proved to be accurate. Thus, mobile phone signal along with web-based socioeconomic data (e.g. Yelp, Google maps, Streetview, etc.) could be utilized in a recurrent manner for simulating population growth and land-use change in future studies.

The estimated relative timing between population change and land-use change is restricted to my study period. The entire trajectories of population and land use were not recorded in the data for some developed districts, and the initial statuses of population and land use remain unknown for these data. Thus, time lags that I estimated through lagged correlation could be mis-specified. Therefore, more case studies should be conducted with the same approaches with a longer study period to reconfirm my findings and test my hypotheses of urbanization processes. The problem is that few countries, except some in East Asia (Campbell, 2015), that conduct population registers, do not necessarily make annual population data accessible. The reason why developed countries do not compile,

centralize, and summarize personal information is because of a threat of privacy and personal freedom (Weeks, 2020).

References

- Alig, R. J., & Healy, R. G. (1987). Urban and built-up land area changes in the United States: an empirical investigation of determinants. *Land Economics*, 63(3), 215-226.
- Antrop, M. (2004). Landscape change and the urbanization process in Europe. *Landscape and urban planning*, 67(1-4), 9-26.
- Bell, S., Alves, S., de Oliveira, E. S., & Zuin, A. (2010). Migration and land use change in Europe: a review. *Living Reviews in Landscape Research*, 4(2), 1-49.
- Bencivenga, V. R., & Smith, B. D. (1997). Unemployment, migration, and growth. *Journal of political Economy*, 105(3), 582-608.
- Bhatta, B. (2010). *Analysis of urban growth and sprawl from remote sensing data*: Springer Science & Business Media, Berlin. pp. 17-36. <https://doi.org/10.1007/978-3-642-05299-6>.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Campbell, C. D. (2015). Demographic Techniques: Family Reconstitution. In J. D. Wright (2nd Eds.). *International encyclopedia of the social & behavioral sciences* (pp 138-142). Elsevier.
- Cihlar, J., & Jansen, L. J. (2001). From land-cover to land-use: a methodology for efficient land-use mapping over large areas. *The Professional Geographer*, 53(2), 275-289.
- Campbell, C. D. (2015). Demographic Techniques: Family Reconstitution.
- Deng, C., & Wu, C. (2013). A spatially adaptive spectral mixture analysis for mapping subpixel urban impervious surface distribution. *Remote Sensing of Environment*, 133, 62-70. <https://doi.org/10.1016/j.rse.2013.02.005>
- Deng, C., & Zhu, Z. (2018). Continuous subpixel monitoring of urban impervious surface using Landsat time series. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2018.10.011>.
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., ... & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45), 15888-15893.
- Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: one or the other, or both?. *International Journal of Remote Sensing*, 29(4), 1185-1206.
- Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259-272.
- Fan, C. C. (2003). Rural-urban migration and gender division of labor in transitional China. *International Journal of Urban and Regional Research*, 27(1), 24-47.
- Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... & Helkowski, J. H. (2005). Global consequences of land use. *science*, 309(5734), 570-574.
- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, 30(11), 1905-1927.
- Franke, J., Roberts, D. A., Halligan, K., & Menz, G. (2009). Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban

- environments. *Remote Sensing of Environment*, 113(8), 1712-1723.
<https://doi.org/10.1016/j.rse.2009.03.018>.
- Friedmann, J. (1966). Regional development policy: a case study of Venezuela (No. HT395. V4 F7).
- Gray, J., & Song, C. (2013). Consistent classification of image time series with automatic adaptive signature generalization. *Remote Sensing of Environment*, 134, 333-341.
- Gubhaju, B., & De Jong, G. F. (2009). Individual versus household migration decision rules: Gender and marital status differences in intentions to migrate in South Africa. *International migration*, 47(1), 31-61.
- Guo, B., Damper, R. I., Gunn, S. R., & Nelson, J. D. (2008). A fast separability-based feature-selection method for high-dimensional remotely sensed image classification. *Pattern Recognition*, 41(5), 1653-1662.
- Herold, M., Gardner, M. E., & Roberts, D. A. (2003a). Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and remote sensing*, 41(9), 1907-1919. <https://doi.org/10.1109/TGRS.2003.815238>.
- Herold, M., Liu, X., & Clarke, K. C. (2003b). Spatial metrics and image texture for mapping urban land-use. *Photogrammetric Engineering & Remote Sensing*, 69(9), 991-1001.
- Hung, Q. A., & Weng, Q. (2018). Mapping Impervious Surfaces in the Greater Hanoi Area, Vietnam, from Time Series Landsat Image 1988–2015, in: Weng, Q., Quattrochi, D., & Gamba, P. E., (2 Ed.), *Urban remote sensing*. CRC press., Boca Raton, pp. 173-194.
- Jensen, J. R. (2009). *Remote sensing of the environment: An earth resource perspective 2/e*.
- Jensen, J. R., Qiu, F., & Patterson, K. (2001). A neural network image interpretation system to extract rural and urban land-use and land-cover information from remote sensor data. *Geocarto International*, 16(1), 21-30.
- Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., ... & George, P. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global environmental change*, 11(4), 261-269.
- Lee, E. S. (1966). A theory of migration. *Demography*, 3(1), 47-57.
- Liu, C., Zhang, Q., Luo, H., Qi, S., Tao, S., Xu, H., & Yao, Y. (2019). An efficient approach to capture continuous impervious surface dynamics using spatial-temporal rules and dense Landsat time series stacks. *Remote Sensing of Environment*, 229, 114-132. <https://doi.org/10.1016/j.rse.2019.04.025>
- Lu, D., & Weng, Q. (2005). Urban classification using full spectral information of Landsat ETM+ imagery in Marion County, Indiana. *Photogrammetric Engineering & Remote Sensing*, 71(11), 1275-1284. <https://doi.org/10.14358/PERS.71.11.1275>.
- Lucas, Jr, R. E. (2004). Life earnings and rural-urban migration. *Journal of political economy*, 112(S1), S29-S59.
- Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., ... & Lim, T. K. (2006). A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, 3(1), 68-72.
<https://doi.org/10.1109/LGRS.2005.857030>.

- Massey, D. S. (1999). Why does immigration occur?: a theoretical synthesis. In C. Hirschman, P. Kasinitz, & J. Dewind. (Eds.), *The handbook of international migration: The American experience* (pp. 34-52). Russell Sage Foundation.
- Meyer, W. B., & Turner, B. L. (1992). Human population growth and global land-use/cover change. *Annual review of ecology and systematics*, 23(1), 39-61.
- Myrdal, G. (1958). *Rich lands and poor: the road to world prosperity* (No. HD82 M9 1958).
- Okujeni, A., van der Linden, S., & Hostert, P. (2015). Extending the vegetation–impervious–soil model using simulated EnMAP data and machine learning. *Remote Sensing of Environment*, 158, 69-80.
- Olorunfemi, J. F. (1984). Land Use and Population: A Linking Model. *Photogrammetric Engineering & Remote Sensing*, 50, 221-227.
- Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), 1007-1011.
- Powell, R. L., & Roberts, D. A. (2008a). Characterizing variability of the urban physical environment for a suite of cities in Rondonia, Brazil. *Earth Interactions*, 12(13), 1-32.
- Powell, R. L., Roberts, D. A., Dennison, P. E., & Hess, L. L. (2007). Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, 106(2), 253-267.
<https://doi.org/10.1016/j.rse.2006.09.005>.
- Powell, S. L., Cohen, W. B., Yang, Z., Pierce, J. D., & Alberti, M. (2008b). Quantification of impervious surface in the Snohomish water resources inventory area of western Washington from 1972–2006. *Remote Sensing of Environment*, 112(4), 1895-1908. <https://doi.org/10.1016/j.rse.2007.09.010>
- Raffalovich, L. E. (1994). Detrending time series: A cautionary note. *Sociological Methods & Research*, 22(4), 492-519.
- Ridd, M. K. (1995). Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. *International journal of remote sensing*, 16(12), 2165-2185.
<https://doi.org/10.1080/01431169508954549>.
- Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. O. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote sensing of environment*, 65(3), 267-279.
[https://doi.org/10.1016/S0034-4257\(98\)00037-6](https://doi.org/10.1016/S0034-4257(98)00037-6).
- Saraswat, I., Mishra, R. K., & Kumar, A. (2017). Estimation of PM10 concentration from Landsat 8 OLI satellite imagery over Delhi, India. *Remote Sensing Applications: Society and Environment*, 8, 251-257.
<https://doi.org/10.1016/j.rsase.2017.10.006>
- Scheuer, S., Haase, D., & Volk, M. (2016). On the Nexus of the Spatial Dynamics of Global Urbanization and the Age of the City. *PloS one*, 11(8).
- Schneider, A. (2012). Monitoring land-cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach. *Remote Sensing of Environment*, 124, 689-704.

- Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A meta-analysis of global urban land expansion. *PloS one*, 6(8).
<https://doi.org/10.1371/journal.pone.0023777>.
- Sherbinin, A. D., Carr, D., Cassels, S., & Jiang, L. (2007). Population and environment. *Annu. Rev. Environ. Resour.*, 32, 345-373.
- Shih, H., Stow, D. A., & Tsai, Y. H. (2019). Guidance on and comparison of machine learning classifiers for Landsat-based land-cover and land-use mapping. *International Journal of Remote Sensing*, 40(4), 1248-1274.
- Shih, H., Stow, D. A., Weeks, J. R., & Coulter, L. L. (2016). Determining the type and starting time of land cover and land use change in southern Ghana based on discrete analysis of dense Landsat image time series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 2064-2073.
<https://doi.org/10.1109/JSTARS.2015.2504371>.
- Shih, H., Stow, D. A., Tsai, Y. M., & Roberts, D. A. (2020). Estimating the starting time and identifying the type of urbanization based on dense time series of landsat-derived Vegetation-Impervious-Soil (VIS) maps—A case study of North Taiwan from 1990 to 2015. *International Journal of Applied Earth Observation and Geoinformation*.
- Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM data: when and how to correct atmospheric effects?. *Remote sensing of Environment*, 75(2), 230-244.
- Stoler, J., Daniels, D., Weeks, J. R., Stow, D. A., Coulter, L. L., & Finch, B. K. (2012). Assessing the utility of satellite imagery with differing spatial resolutions for deriving proxy measures of slum presence in Accra, Ghana. *GIScience & Remote Sensing*, 49(1), 31-52.
- Stow, D. A., Weeks, J. R., Shih, H. C., Coulter, L. L., Johnson, H., Tsai, Y. H., ... & Mensah, F. (2016). Inter-regional pattern of urbanization in southern Ghana in the first decade of the new millennium. *Applied Geography*, 71, 32-43.
<https://doi.org/10.1016/j.apgeog.2016.04.006>
- Stow, D. A., Weeks, J. R., Toure, S., Coulter, L. L., Lippitt, C. D., & Ashcroft, E. (2013). Urban vegetation cover and vegetation change in Accra, Ghana: Connection to housing quality. *The Professional Geographer*, 65(3), 451-465.
- Toole, J. L., Ulm, M., González, M. C., & Bauer, D. (2012, August). Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing* (pp. 1-8).
- Van den Berg, L., Drewett, R., & Klaassen, L. H. (1982). *A study of growth and decline: Urban Europe*. Elsevier.
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46-56.
- Weeks, J. (2010) *Defining Urban Area*. In T. Rashed, & C. Jürgens. (Eds.). (2010). *Remote sensing of urban and suburban areas* (Vol. 10). Springer Science & Business Media.
- Weeks, J. R. (2020). *Population: An introduction to concepts and issues*. Cengage Learning.

- Wu, C., & Murray, A. T. (2003). Estimating impervious surface distribution by spectral mixture analysis. *Remote sensing of Environment*, 84(4), 493-505.
[https://doi.org/10.1016/S0034-4257\(02\)00136-0](https://doi.org/10.1016/S0034-4257(02)00136-0)
- Wu, C. (2004). Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery. *Remote Sensing of Environment*, 93(4), 480-492.
<https://doi.org/10.1016/j.rse.2004.08.003>.
- Zhang, K. H., & Song, S. (2003). Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Economic Review*, 14(4), 386-400.
- Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote sensing of Environment*, 144, 152-171. <https://doi.org/10.1016/j.rse.2014.01.011>
- Zhu, Z., Woodcock, C. E., Rogan, J., & Kellndorfer, J. (2012). Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land-cover classification using Landsat and SAR data. *Remote Sensing of Environment*, 117, 72-82.