SOCIAL-SPATIAL STRUCTURE OF BEIJING: A SPATIAL-TEMPORAL ANALYSIS

by

SHANQI ZHANG

(Under the direction of Xiaobai Yao)

Abstract

Beijing experienced significant changes in its urban social spatial structure within last several decades. Economic growth, evolving urban planning policy and large-scale urban development have accelerated the pace of human displacement in Beijing. By applying a spatial-temporal analysis to identify social-spatial structural change in Beijing over time, this study aims to elucidate the driving forces of the process, draw useful findings for future urban planning in Beijing, and shed light on urban studies in other Chinese cities. Clustering analysis is performed to identify the spatial distribution pattern in Beijing in light of social-economic factors (migrant status, occupation and education level), as well as household and other demographic attributes. The GWR model is then used to examine the relationship between population change and distribution of diverse social groups during 2000 and 2009. Data on recent social phenomena are captured by the model and provide implications for planning and decision making, especially in light of new urban poverty and “social vulnerable groups.”

INDEX WORDS: social-spatial structure, Beijing, spatial-temporal, cluster analysis, GWR model, urban poverty
SOCIAL-SPATIAL STRUCTURE OF BEIJING: A SPATIAL-TEMPORAL ANALYSIS

by

SHANQI ZHANG

B.S., Nanjing Normal University, China, 2009

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2011
SOCIAL-SPATIAL STRUCTURE OF BEIJING: A SPATIAL-TEMPORAL ANALYSIS

by

SHANQI ZHANG

Major Professor: Xiaobai Yao

Committee: Steven R. Holloway

Thomas R. Jordan

Electronic Version Approved:
Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2011
ACKNOWLEDGEMENTS

I would like to acknowledge and thank Dr. Yao, who has helped me out tremendously both with this thesis and my coursework as well as in general guidance. I would also like to acknowledge and thank Dr. Holloway and Dr. Jordan for helping me out immensely both in terms of thesis research as well as my coursework. Acknowledgement and thanks are also given to my fellow students in the Geography building who were always ready to delay working for helping me out. I would like to acknowledge and thank my family and friends for supporting me through this process. Acknowledge and thanks are also given to all of the people working in the geography department office for helping me out with getting all of my paperwork completed and always helping me out with whatever I needed.
# Table of Contents

List of Figures ................................................................................................. vii
List of Tables ........................................................................................................ viii
1 Introduction ......................................................................................................... 1

2 Literature Review ................................................................................................ 4
   2.1 Urban Social-spatial Theories and Classic Models ......................................... 4
   2.2 Empirical Studies of Social-Spatial Structure ............................................... 5
   2.3 Urban Studies of Beijing ............................................................................... 7

3 Study Area and Data .......................................................................................... 9
   3.1 Study Area ..................................................................................................... 9
   3.2 Data ............................................................................................................... 11
   3.3 Data Pre-Processing ...................................................................................... 14

4 Social-Spatial Pattern of Beijing ........................................................................ 27
   4.1 Identify Social-Spatial Structure of Beijing .................................................. 27
      4.1.1 Spatial Clustering .................................................................................. 27
      4.1.2 Methodology ....................................................................................... 33
   4.2 Social -Spatial Patterns of Beijing ................................................................. 40
   4.3 Restructuring of Social Space in Beijing ...................................................... 49

5 Socioeconomic Process Related to Social Space Restructure ............................. 52
   5.1 Geographically Weighted Regression (GWR) .............................................. 52
   5.2 Empirical Results ......................................................................................... 54
   5.3 Related Social Phenomena ......................................................................... 61

6 Conclusion .......................................................................................................... 65
   6.1 Summary of Methodology and its Contribution ......................................... 65
   6.2 Major Findings and their Broader Implications .......................................... 67
   6.3 Areas for Future Study .................................................................................. 70

7 References .......................................................................................................... 73
List of Figures

Figure 1. Study Area ............................................................................................................................................ 10
Figure 2. Landsat-5 image of Beijing in 2000................................................................................................. 12
Figure 3. Landsat-5 image of Beijing in 2009................................................................................................ 13
Figure 4. WorldView-2 image of Urban Core Area, Beijing. 2009. ........................................................... 14
Figure 5. Different Building Types ................................................................................................................... 18
Figure 6. Decision Tree for Classification of Built-up Area ........................................................................... 20
Figure 7. Land-use Classification of Built-up Area .......................................................................................... 21
Figure 8. Land-cover Classification Result of Beijing in Year 2000................................................................. 22
Figure 9. Land-cover Classification Result of Beijing in Year 2009................................................................. 23
Figure 10. Estimated Population in Urban Core Area using Dasymetric Mapping ........................................ 25
Figure 11. Estimated Population of Beijing in Year 2009 using Dasymetric Mapping .................................... 26
Figure 12. U-Matrices for Interpreting SOM Result .......................................................................................... 36
Figure 13. U-Matrices Visualization of Major Components, 1990................................................................. 41
Figure 14. Social Area of Beijing in 1990 ......................................................................................................... 42
Figure 15. U-Matrices Visualization of Major Components, 2000................................................................. 44
Figure 16. Social Area of Beijing in 2000 ......................................................................................................... 45
Figure 17. U-Matrices Visualization of Major Components, 2009................................................................. 47
Figure 18. Social Area of Beijing in 2009 ......................................................................................................... 48
Figure 19. Spatial Distribution of Coefficients for the Independent Variables .............................................. 58
Figure 20. Spatial Distribution of Coefficients for the Independent Variables .............................................. 59
Figure 21. Population Density of People with College Degree or Above ....................................................... 61
Figure 22. Population Density of Temporary Residents (Living Six Month to One Year) ............................. 63
Figure 23. Population Density of Temporary Residents (Living within Six Months) ..................................... 64
List of Tables

Table 1. Distance Matrix of Land-cover Classes.......................................................... 17
Table 2. Census Variables used for Social-Area Analysis of Beijing in 1990..................... 37
Table 3. Census Variables used for Social-Area Analysis of Beijing in 2000...................... 37
Table 4. Census Variables used for Social-Area Analysis of Beijing in 2009...................... 39
Table 5. Comparison of statistical test of results from OLS, SLM, SEM and GWR model .......... 54
Table 6. Coefficients of dependent variables in OLS model and statistics of according coefficients in GWR model ........................................................................................................... 56
1 Introduction

Since the reform toward more open policies starting in 1978, China has gradually seen a transition from a centrally-planned, rural-based economy to a market-oriented, urban-based economy, leading to significant changes in spatial distributions of population and economy (Gu and Shen 2003). It was argued that following urban reforms during the mid-1980s, significant changes occurred to China’s urban structure, transforming spatial, economic, and social organization within Chinese cities (Lai and Yat-Ming 2006; Feng, Zhou et al. 2007). Market forces transformed both economic structure and the urban landscape. The emergence of private housing markets and urban redevelopment projects displaced once stable and socially homogeneous spatial structures and have gradually produced economically polarized residential patterns (Berliant and Kung 2009).

Beijing, as the capital and one of the oldest and most densely populated Chinese cities, has experienced fast economic development and an unprecedented process of urbanization since the beginning of modern reforms (He, Okada et al. 2006). Rapid economic and population growth have benefitted Beijing residents with regard to their overall living conditions, and especially in terms of infrastructure construction and improvements (Wei and Yu 2006).

After more than a decade of urban reform, differentiation of social areas was experienced in Beijing reflecting the impact of increasing differentiation of socioeconomic status. As industrialization and urbanization released large numbers of farmers from the rural land, the influx of rural migrants greatly reshaped urban structures of the area (Feng, Zhou et al. 2007). Furthermore, transformation of urban social structures has also been influenced by education and employment changes, land-use policy changes and other inner neighborhood dynamics. The
urban structure and social spatial patterns went through great changes in Beijing. After being
selected as the host city for the 2008 Olympic Games in 2001, the restructuring of Beijing’s
urban space was accelerated through a series of Olympic-related activities. Beijing has faced new
challenges, however, in the areas of increased traffic, air pollution, housing shortages and a lack
of green space, to mention several areas of challenges. Discussions and investigations started
before the event was held. Concerns about environment issues, infrastructure construction and
residential relocation and displacement are hot spots being put forward.

Many studies have examined Beijing’s rapid urban growth from a range of perspectives. Land-use changes has been tracked by use of remote sensing technologies (Chen, Gong et al.
2003; Wu, Li et al. 2006); urban growth and expansion have been modeled and assessed using
Cellular-Automated models (Chen, Gong et al. 2002; He, Okada et al. 2006; He, Okada et al.
2008); landscape and land-use patterns are investigated in consideration of ecological
knowledge and GIS techniques (Qi, Henderson et al. 2004; Wu, Hu et al. 2006); urban
transformation processes and related issues are studied and compared with those of other
Chinese cities (Chen 2008; Feng, Zhou et al. 2008; Zhao, Lu et al. 2009; Zheng, Peiser et al.
2009); statistical analysis is used to compare a range of housing issues, such as housing
inequality and housing choice, as well as institutional and economic implications of public
policy (Li 2000; Huang 2004; Huang and Jiang 2009); the influx of rural migrants and its
implication to urban forms has been investigated (Gu, Chan et al. 2006; Barabantseva 2008;
Zheng, Long et al. 2009); policy implications in transportation, urban living quality, job
accessibility and residential relocation have also been explored (Shin 2009; Zhao and Lu 2009;
Zhao, Lu et al. 2009); and most recently, impacts of the Olympic Games on urban planning and
Restructuring of Beijing’s social space has attracted considerable research attention (Smith and Himmelfarb 2007; Zhang and Zhao 2009).

No previous study has yet quantitatively examined the transition of social areas in Beijing after its dramatic change for hosting the 2008. More abundant data available nowadays provides the opportunity to explore the restructure of social space in the city, especially the most up-to-date changes. This study therefore examined the changing urban structures in Beijing and explored its underlying mechanism to provide further insight into the planning process for future development of the city. Specific objectives of this study are to: 1) to investigate the urban social-spatial structure of Beijing using clustering techniques and identify how various social areas are distributed over the city at different time periods, 2) to identify how the patterns change over time, and 3) to describe currently evolving processes impacting the urban landscape within the city from a standpoint of diverse social groups.
2 Literature Review

2.1 Urban Social-spatial Theories and Classic Models

Initialized in 1920s, three classical urban social-spatial models were developed during 1920s and 1940s. Those three models include Burgess’s concentric model, Hoyt’s sector model and Harris and Ullman’s multiple-nuclei models. They depict the urban social-spatial structure in different stages of urban development.

In Burgess’s concentric zone model, the spatial pattern of cities is produced by processes of invasion and succession, which is derived from early ecological approach forwarded by the Chicago School of Sociology. Central to the approach is the notion of social competition between various social groups that revolves around the control of urban spaces in order to maintain social distance between groups. From the perspective of a neighborhood, the Chicago scholars viewed the invasion and succession process as ongoing: any given neighborhood will be occupied by a series of immigrant groups over time. Burgess’s model depicts the social ordering of industrial cities in the early twentieth century with regard to notions of human behavior; yet ignores politics, power, transportation and other forces that shape society.

Hoyt’s sector model was forwarded firstly in 1930s. The theory was based on the observation of a top-down process of urban growth where lower class households subsequently occupied the housing left behind by the outwardly and upwardly mobile upper class. It was also recognized that cities are usually expanded along the major transportation lines. Therefore, it represents a radial instead of zonal expansion and leads to a sectoral rather than a concentric pattern.

Though the notion of social dynamic in Hoyt’s theory is quite different from that of Burgess’s, both of them consider cities as having a single major commercial core, which
characterized most urban places prior to 1950. After world war, Chauncy Harris and Edward Ullman recognized that the cities were developing multiple centers and offered their multi-nuclei model. Multiple centers of the city defined in the model might be of various functions yet all significantly impact its surroundings and together fundamentally altered the spatial structure of the city. Harris and Ullman’s model provided much more accurate descriptions of social structure in the late twentieth century and early twenty-first century.

Though these three models have been criticized, tested and modified almost ceaselessly since their introduction (Feng, Zhou et al. 2007), they provided a powerful theoretical foundation for the understanding of urban social geography that has spurred a large body of urban social geographic research.

2.2 Empirical Studies of Social-Spatial Structure

Social area analysis, based on the premise of economics status, family status as well as ethnic status, became popular in the study of urban society. Early studies on urban social structure were marked by studies in Los Angeles and San Francisco in the 1940s and 1950s, which differentiate social areas taking account of aforementioned three constructs.

Introduction of quantitative data and new analytical techniques into the study of urban society provided new perspectives looking into the social-spatial structure of cities. Complex models are developed in the latter half of twentieth century reflecting the increasingly spatially complex urban society. The factorial ecology method is one of the most popular approaches in the field of urban social-spatial studies (Gu, Wang et al. 2005) and is regarded as a very useful tool for studying intra-urban social spatial structure (Cadwallader 1996). Among a series of studies using factorial approach in western cities, Murdie’s study in Toronto (Murdie 1969), which superimposed three basic constructs-economic, family and ethnic-onto the physical space
of the city, is one of the studies that attract more attentions than other since it established a aggregated mosaic combining social-area analysis result and real space.

However, though there is rich literature on the urban social-spatial study in western cities, especially those of capitalist economy, studies on cities under socialist economy, especially socialist cities of transitional economy, is not enough (Sykora 1999; Wu 2005). As discussed in several articles, it is expected there are dramatic distinctions between socialist and capitalist cities in the character and magnitude of social-spatial difference in that socialist cities would be characterized by less segregation (Sykora 1999). The social-spatial structure of socialist cities in Eastern Europe was analyzed and modeled in previous studies (Hamilton 1979; Szelenyi 1983; Smith 1996), pointing out the existence of social-spatial difference in socialist cities. Also, the social-spatial patterns of socialist cities are compared to that of capitalist cities taking three main groups of population characteristics into account. That is, socio-economic, demographic and ethnic. More recent studies shed light on the transformation of the social-spatial structure in socialist cities in the process of post-socialist transition. The emergence of social stratification and a new middle class is addressed (Andrusz, Harloe et al. 1996). Social-spatial differentiation, which is mainly driven by uneven income caused by market mechanism, is regarded as major dimension of urban change in the post-socialist transition as well.

While continuing to adhere to socialist principles, Chinese economic reforms have fostered a great reliance on the free market, which not only accelerates economic prosperity but also leads to growing socioeconomic divisions between residents (Hu and Kaplan 2001). Previous studies of social-spatial differentiations in post-communist Eastern European cities, which have gone through the transition from socialist cities to capitalist cities, may to some degree shed light on the study of Chinese cities given that they are all in transitional economies. Differing from
many western cities, most Chinese cities are more densely populated. Spatial differentiation of social areas is studied in various cities such as Guangzhou, Shanghai, Hongkong and others (Lo 1994; Lo 2005; Wu 2005). These studies provided implications to the urban planning and cities’ future development. But the characteristic of urban social landscape in Chinese cities is still not fully addressed, especially in terms of its spatial-temporal change and comparison with western cities where there are relatively mature theories and methodology in the urban studies.

2.3 Urban Studies of Beijing

Since the reform and open policy being implemented in 1978, China has begun its transition from centrally-planned, rural-based economy to a market-oriented, urban-based economy gradually (Gu and Shen 2003). Beijing, as the capital city of China and one of the oldest and most densely populated, has been experiencing fast economic development and unprecedented process of urbanization since then (He, Okada et al. 2006). Rapid economic growth and population increasing on the one hand benefit Beijing especially with regard to the overall living conditions of the residents and infrastructure constructions (Wei and Yu 2006). On the other hand, these forces pose to Beijing many challenges and pressures such as traffic congestions, air pollution, housing shortages, lack of green space, etc. Many studies thus have been conducted focusing on Beijing’s rapid urban growth from different perspectives. Land-use change is detected employing remote sensing technologies (Chen, Gong et al. 2003; Wu, Li et al. 2006); urban growth and expansion are modeled and assessed using CA-based models (Chen, Gong et al. 2002; He, Okada et al. 2006; He, Okada et al. 2008); landscape of land-use pattern is investigated incorporating ecological knowledge and GIS techniques (Qi, Henderson et al. 2004; Wu, Hu et al. 2006); urban transformation process and related issues are studied and compared with other Chinese cities (Chen 2008; Feng, Zhou et al. 2008; Zhao, Lu et al. 2009; Zheng,
Peiser et al. 2009); housing issues such as housing inequality, housing choice and its driving forces with regard to policy, institutional and economics are studied based on statistical and comparison analysis (Li 2000; Huang 2004; Huang and Jiang 2009); influx of rural migrants and its implications to urban form is investigated (Gu, Chan et al. 2006; Barabantseva 2008; Zheng, Long et al. 2009); policy implications in transportation, urban living quality, job accessibility, residential relocation and etc are explored (Shin 2009; Zhao and Lu 2009; Zhao, Lu et al. 2009); and most recently, impacts of Olympic Games on urban planning and restructuring Beijing’s social space attract the considerable research attentions (Smith and Himmelfarb 2007; Zhang and Zhao 2009).

Despite those large volumes of studies, fewer studies have paid attention on Beijing’s urban social structure. The change in Beijing since open and reform policies were implemented has dramatically transformed the city’s overall appearance and social spaces (Feng, Zhou et al. 2007), which provide a further insight on the planning and future development of the city.

Sit (1996) was the first to study social areas in Beijing, highlight the change of housing ownership, central-planned land-use and relationship between residence and work-units as the main process within the context of social-spatial structure transformation. Gu et.al (2005) analyzed the social-spatial structure of Beijing’s built-up area and concluded that social differentiation is occurring after a decade of urban reforms, especially with regard to social-economic status. Feng et.al (2007) investigated Beijing’s social landscape during 1982 and 2000 and relative strength of processes of social differentiation by adopting a factor ecology approach. However, among existed studies, spatial factors are not fully incorporated in the evaluation of social spatial differentiation. Furthermore, few studies have focused on the comparisons of social spatial structure in cities under different economies and political environment.
3 Study Area and Data

3.1 Study Area

As the capital city of China, Beijing is currently one of the fastest-growing Chinese cities (Feng, Wang et al. 2009). The administrative organization of Chinese cities is intended to integrate urban and rural areas in order to maintain urban food resource self-sufficiency (Gu, Wang et al. 2005). The region of Beijing is a mix of urban and rural territory. Beijing consists of 18 districts (qu), which can be divided into three zones:

1) There are four central-city districts (Dongcheng, Xicheng, Chongwen and Xuanwu). They are regarded as “old urban districts”, roughly comparable to the central city area of large western cities.

2) There are four inner suburban districts (Chaoyang, Fengtai, Haidian and Shijingshan), surrounding the old urban districts as a sub-urban belt (Feng, Zhou et al. 2007).

3) There are ten outer suburban districts (Mentougou, Fangshan, Tongzhou, Daxing, Changping, Shunyi, Yanqing, Miyun, Pinggu and Huairou)
Figure 1. Study Area
3.2 Data

3.2.1 Census Data

In this study census variables are primarily used to examine the social-spatial structure of Beijing. Two types of datasets are used for 1990, 2000 and 2009.

1) National Population Census\(^1\). Included in the census are data on the demographic structure, household structure, *hukou* status\(^2\), education level, and occupation of the inhabitants. All data is collected at the sub-district level (*jiedao*), the size of which is comparable to census tract. The fourth and fifth national censuses are used to study social-areas in the years 1990 and 2000\(^3\).

2) Beijing Statistical Year Book. Since the sixth national census, taken in 2010, has not yet been released, the statistical year book is used within this study as an alternative resource for year 2009. The statistical yearbook includes similar information as that of the national population census, except that the available data is aggregated at the district level.

---

\(^1\) The first Chinese population census was conducted in 1953, followed by the second and third in year 1964 and 1982. Since the fourth population census in year 1990, the survey was consistently taken every ten years.

\(^2\) Hukou is a unique Chinese household registration system, implemented initially in Chinese cities in 1951 to maintain social peace and order and protect their freedom of residence and movement. The policy then expanded to rural areas as well, shifting the original registration purposes to prevent unplanned migration as well as to introduce formal control over the rural influx to the cities and over intra-rural and intra-urban population movement Liu, Z. Q. (2005). "Institution and inequality: the hukou system in China." *Journal of Comparative Economics* **33**(1): 133-157.. The two types of hukou, namely, urban and rural, pertain to urban and rural population, respectively. Each individual is required to register in one and only one place of residence. In urban cities, the unit of registration is a household. In rural areas, the unit of registration is a commune or village or state farm.

\(^3\) More information such as house renting, average living space has been added in the census since 2000. Therefore, more variables are available for the study in year 2000.
3.2.2 Remote Sensing Images

Landsat images of year 2000 and 2009, with the 30- meter resolution, are available for the study area. They cover the whole great Beijing area, including all 18 aforementioned districts. Following figures show the images.

![Figure 2. Landsat-5 image of Beijing in 2000](image)
WorldView-2 image is available for year 2009. The image covers old urban districts (Dongcheng, Xicheng, Chongwen and Xuanwu), with the area of 100 km². Launched October 2009, WorldView-2 is the first high-resolution 8-band multispectral commercial satellite, including four standard colors: blue, green, red, near-InfraRed1 and four new colors: coastal, yellow, red edge, and near-InfraRed2. WorldView-2 provides 0.46 meters panchromatic resolution and 1.84 meter multispectral resolution, which takes advantage in precise map creation, change detection and in-depth image analysis.
### 3.3 Data Pre-Processing

#### 3.3.1 Image Classification

Classification of WorldView-2 Image

Object-based image analysis (OBIA) or Geospatial object-based image analysis (GEOBIA) refers to techniques that analyzing image meaningful image objects rather than focusing on uniform pixels. Comparing to traditional per-pixel approaches for image analysis, four not commonly-used components are found in OBIA as claimed by Platt and Rapoza: 1)
segmentation, 2) the nearest neighborhood classifier, 3) incorporation of expert knowledge and 4) optimization of feature space (Platt and Rapoza 2008; Blaschke 2010).

Object-based image analysis (OBIA) techniques have been widely adopted in the last ten years given its capability of integrating additional spectral and spatial information and superiority to traditional per-pixel method in terms of accuracy and richness of information. There are rich bodies of literature focusing on the applications of OBIA in various fields such as biomass, environment, urban study and etc. (Burnett and Blaschke 2003; Dorren, Maier et al. 2003; Chubey, Franklin et al. 2006; Pascual, Garcia-Abril et al. 2008; Zhang, Pavlic et al. 2005; Kong, Xu et al. 2006; Chen, Shi et al. 2007; Lackner and Conway 2008). Recent studies have integrated ancillary socioeconomic and geospatial information to help identify urban features (Hofmann, Strobl et al. 2008; Aubrecht, Steinhocher et al. 2009; Ebert, Kerle et al. 2009) and track dynamic change in urban area, especially urban sprawl (Ivits, Koch et al. 2005; An, Zhang et al. 2007; Durieux, Lagabrielle et al. 2008).

Object-based classification methods are applied in the study. This approach involves two steps: segmentation and classification. The basic task of segmentation algorithm is to merge homogenous pixels into image elements to differentiate from heterogeneous neighboring regions and represent necessary information for interpreting image structures and understanding relationships among image components (Taubenbock, Esch et al. 2010). Multi-resolution segmentation is carried out for this study. Multi-resolution is a bottom-up region growing technique. Small objects are merged into larger ones through an iterative process. By running the pairwise clustering process, the algorithm will minimize the weighted heterogeneity expressed as \((n, h)\) of resulting image objects, where \(n\) is the size of a segment and \(h\) a parameter of
heterogeneity (Benz, Hofmann et al. 2004). This iterative process will end when the scale parameter is approached.

To make the resulting segments closer to the spatial structure of the building, meaning it will neither be too general for building extraction nor being too-detailed so that the implementation would be more compute-intensive.

Scale parameters of 10, 25, 50, and 100 are tested in this study. The result shows that scale factor of 25 is best fitted for building extraction purpose in this study. Segmentation using scale factor of 10 splits the buildings into too many small components that are not needed and get a lot of noise for the further interpretation, while using scale factor over 50 results too much unexpected segmentation group buildings and surrounding roads, which increases the difficulty for the classification.

Supervised classification using nearest neighborhood classifier is performed to classify land-cover classes. This method adopts a set of samples of different classes to assign membership values, according to object features. The Nearest Neighbor classifier returns a membership value between zero and one, based on the image object's feature space distance to its nearest neighbor. The membership has a value of one if the image object is identical to a sample. If the image object differs from the sample, the feature space distance has a fuzzy dependency on the feature space distance to the nearest sample of a class.

Five classes are utilized in this study: built-up area, water, shadow, road, and trees. Spectral information is mainly taken into consideration for the classification. Given the unknown capabilities of the newly added bands in the image, values of eight bands, combination of bands, standard deviation of each band, and the overall brightness are used as parameters for the neural network to define best combination of variables for differentiating five land-cover types. Ten
variables identified as best set for classification: 1) NIR2-Red, 2) Red Edge (RE) -Blue, 3) NIR1-RE, 4) RE – Coastal, 5) NIR2- RE, 6) RE-Green, 7) RE-Red, 8) NIR2-Yellow, 9) Std.dev of (Costal), 10) Std.dev of (NIR1). New added bands play important roles in the classification that all selected variables are related to new added bands.

Table 1 shows distance matrix between each two of land-cover type. The higher the value is in the matrix, the more different are the two according land-cover type. It could be seen that the minimum distance between samples of two land-cover types is larger than 1, meaning that the variables are sufficient for capturing the differences among land-cover classes.

Table 1. Distance Matrix of Land-cover Classes

<table>
<thead>
<tr>
<th>Classes</th>
<th>Road</th>
<th>Building</th>
<th>Tree</th>
<th>Shadow</th>
<th>Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road</td>
<td>0.00</td>
<td>1.06</td>
<td>2.95</td>
<td>1.18</td>
<td>2.48</td>
</tr>
<tr>
<td>Building</td>
<td>1.06</td>
<td>0.00</td>
<td>3.07</td>
<td>1.99</td>
<td>2.75</td>
</tr>
<tr>
<td>Tree</td>
<td>2.95</td>
<td>3.07</td>
<td>0.00</td>
<td>3.07</td>
<td>5.15</td>
</tr>
<tr>
<td>Shadow</td>
<td>1.18</td>
<td>1.99</td>
<td>3.07</td>
<td>0.00</td>
<td>2.38</td>
</tr>
<tr>
<td>Water</td>
<td>2.48</td>
<td>2.75</td>
<td>5.15</td>
<td>2.38</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Land-cover classification can tell the general land cover in the area. However, it is not enough for further land-use classification. A rule-based method integrating spectral, structural and contextual information existing in the image is then adopted to classify land-use types.

The objective for the use of expert rules in the object-based analysis is to ensure that a particular map object receives a context-dependent treatment (Steiniger, Taillandier et al. 2010). Expert rules are first set up using expert knowledge. In the case study, such expert rules will utilize semantic concepts related to urban fabric. Those concepts are contained in the map and are intuitive for map users to understand.
Geometric properties of the buildings, structural characteristic of building groups, and contextual information are adopted for this study. Three building types are taken into consideration as shown in Figure 5.

![Figure 5. (a) low-rise buildings with mixed sizes and patterns (b) high-rise buildings with consistent pattern (c) irregular shaped buildings](image)

High-rise buildings with consistent pattern as shown in (a) are most likely to be residential area since apartments are most commonly residential districts in Chinese cities and they used be built in a relatively large neighborhoods where all the apartment buildings are designed in similar or even same ways. Low-rise buildings with various sizes and patterns as shown in (b) are mostly areas where parts of old buildings are preserved while mixed with some newly established buildings. Those areas are also of mixed land-use. For example, some administrative bureaus or companies provide dormitories for the new coming workers and those who cannot
afford houses themselves in Beijing. Those dorms are usually located close to the working place. Also, there are old Chinese traditional buildings (*siheyuan*) in Beijing. They were residential areas in old times, but nowadays lots of them are preserved for commercial/historical uses, while the number of actual residents has decreased dramatically from old days. Therefore, those areas are of mixed use in terms of residential and employment. Irregular shaped buildings with relatively large size and are geographically separated from other buildings are mostly commercial/recreational places, where most hired people are working.

For the first two building types, building groups are identified based on similarities of geometric characteristics of buildings and their spatial proximity. Area, azimuth, and the length/width ratio of minimum rectangular boundary of the single building are used to describe building’s geometric characteristic. Cluster analysis is performed to find out similar buildings. Each building has a vector of attributes. Distance of attribute vectors among buildings are calculated so that close vectors in the attribute space are grouped together and their accordingly buildings are regarded as belonging to the same group. Spatial proximity is then used to group nearby buildings of same type together.

Shadows are also used in the study to help identify high-rise buildings. Proximity measurement is weighted by the size of building since the distance among buildings are related to the size of buildings or the height of buildings (the size of shadows). Closeness to shadows is used to determine the height of buildings, which indicates the density of residential area: high-rise residential area is considered to be of high population density, while lower-rise the medium. Figure.6 shows the decision tree for land-use classification after building groups are identified.
Figure 6. Decision Tree for Classification of Built-up Area
Figure 7 shows the classification result according to aforementioned framework.

Classification of Landsat Images

A supervised classification is used to classify landsat-5 images. Two major land-cover types: urbanized area and non-urbanized area are identified and labeled. Both 2000 and 2009 images
are processed so that the change of urbanized area could be calculated for the use of further regression analysis. The classification result of 2009 image is also used for the population estimation, especially in the suburban area. Following are two maps showing the land-cover

Figure 8. Land-cover Classification Result of Beijing in Year 2000
3.3.2 Dasymetric Mapping

To adjust for the use of a different reference source for 2009 data, a population interpolation technique is used to assign a population from the district level to the sub-district level.

The dasymetric mapping is a population interpolation method widely adopted in recent studies. It was originally used for cartographic representation and was defined as methods by which source zones are sub-divided into smaller constituent regions with greater consistency in the variable being mapped (Langford 2006). The application of dasymetric mapping in areal
interpolation has been grown up given its higher accuracy comparing to simple areal weighting and simplicity conceptually and in practice.

The method utilizes additional relevant knowledge to generate better target zone estimates toward a common goal. The formular description of the method can be presented as following:

$$\hat{p}_t = \sum_{s=1}^{S} \frac{A_{tsp}P_s}{A_{sp}} = \sum_{s=1}^{S} A_{tsp}d_{sp}$$

Where \(\hat{p}_t\) is the estimated population of target zone t; \(A_{tsp}\) is the area of overlap between target zone t and source zone s, and having land cover identified as populated; \(P_s\) is the population of source zone s; \(A_{sp}\) is the area of source zone s having land cover identified as populated; \(S\) is the number of source zones, and \(d_{sp}\) is the dasymetric density of the populated class in source zone s.

Once the populated area is identified, dasymetric mapping method is applied to distribute aggregated population to populated area. In this paper, a three-class dasymetric mapping is used to evaluate the urban core area and binary dasymetric mapping is used for the rest of the city, given the data available.

For the urban core area, high-resolution WorldView-2 satellite imagery is used to classify the study area by four land-use types: non-residential, low-density residential, medium-density residential and high-density residential. Each of land-use type is assigned a density value, with the total equals to one. The values are decided going through an trial-error process, so that the sum of population counts at each district is closest to its actual value.
Figure 10. Estimated Population in Urban Core Area using Dasymetric Mapping
For the suburban area, a binary dasymetric mapping is used that all the population will be assigned to urbanized area. Figure 11 shows the population estimation of the area.

Figure 11. Estimated Population of Beijing in Year 2009 using Dasymetric Mapping
4 Social-Spatial Pattern of Beijing

4.1 Identify Social-Spatial Structure of Beijing

4.1.1 Spatial Clustering

Clustering methods are used in various research communities for grouping unlabeled data (Jain, Murty et al. 1999). In the geography field, clustering share much common ground with the long tradition of regional classification (Cliff.A.D 1975). Spatial clustering is the process of grouping a set of objects into classes or clusters based on similarity. Intuitively, objects within a valid spatially based clustering are more similar than they are to objects belonging to a different cluster (Jain, Murty et al. 1999; Han.J. 2001). As one of the new and useful data mining tools for the analysis of geographical data, spatial clustering technique has its root in mathematics and statistics, and has been applied in image processing, spatial imagery data mining, land-use classification, pattern analysis and many other areas (Steenberghen, Aerts et al.; Hallencreutz and Lundequist 2003; Wang, Luo et al. 2005; Wang, Leung et al. 2006; Baum, Haynes et al. 2007; Yee, Tung et al. 2007; Mu and Wang 2008; Steenberghen, Aerts et al. 2009). The wide application is due to its capability to reveal two types of information: 1) indicative information, which literally means the characteristic of cluster itself such as size, extent and location of the cluster, the proximity among clusters and etc, and 2) relational information, which consists of casual factors, potential explanations, cluster justification and association among themes (Estivill-Castro and Lee 2004).

In this study, we want to apply spatial clustering techniques to group people with various social status, occupations and educational backgrounds, examining is there any patterns of social spatial differentiation represented spatially.
All clustering algorithms optimize an induction principle (Cherkassky and Mulier 1998) defining a family of model and the criteria by which a model is to be regarded as the one that best fits the data (Estivill-Castro and Lee 2004). Criteria are usually defined according to a specific measure in terms of similarity or dissimilarity. That is, whether the algorithm proceeds to construct a hierarchical structure that optimizes the criteria or whether a partition strategy has been repeatedly used to modify the set of clusters.

Among large amounts of clustering techniques, they can be categorized in many ways according to different criterion, including partitional and hierarchical; distance-based or density-based, referring to the different ways to calculate the relationship between objects; agglomerative or divisive; heuristic or model-based; and etc. Given the large datasets of geographical data, spatial clustering approaches especially address the issue of classifying a large number of objects (Ng and Jiawei 2002). Partitioning method, hierarchical method, density-based method, grid-based method and artificial neuron networks method are introduced in the following.

Partitioning Methods

Partitioning methods have long been popular clustering algorithms. Given a set D of n objects in a d-dimensional space, a partitioning algorithm starts from an initial partition, and then moves observations iteratively from one group to another according to certain criterion until a convergence is met. The number of groups is specified in advance and will not change during the course of the iteration (Fraley and Raftery 1998).

K-means is an intuitive yet most commonly used algorithm employing criterion that the within-group sums of squares are minimized (MacQueen 1967; Hartigan and Wong 1979). The algorithm uses the mean value of objects in a cluster as the cluster center. According to the squared-error function criterion, K-means algorithm assigns each object to its nearest center
during every course of iteration, forming a new set of clusters. Clusters centers are then recomputed and the process is repeated until within-group sums of squares are not changed (Han.J. 2001).

Advantages of K-means algorithm are that it is relatively scalable and efficient in applications involving large data sets since its computational complexity is $O(n)$, where $n$ is the total number of observations. However, the K-means algorithm is very sensitive to the selection of the initial partition (Jain, Murty et al. 1999). If the initial partition is not properly chosen, the algorithm may converge only to a local optimum. It is also very sensitive to noise and outlier data points, since a small number of such data can substantially influence the mean value (Han.J. 2001).

For clustering via mixture models, partition methods are mostly realized on the basis of Expectation Maximization (EM) algorithm (Fraley and Raftery 1998). Work in this area is based on density estimation theory and assumes that the density distribution of each individual component can be effectively approximated by a mixture of Gaussian (Scott 1992), the parameter vectors of which are estimated using EM algorithm (Jain and Dubes 1988). Similar to the general framework of partitioning techniques, The EM algorithm procedure begins with an initial partition and parameter estimation, and iteratively rescores the patterns against the mixture density produced by parameter vector. The rescored patterns are then used to update the parameter estimates. As a criterion function, EM tries to maximize the log likelihood of the mixture model. The algorithm will terminates when the increase in the log likelihood between two successive iterations is negligible.
Hierarchical Methods

Differing from partitioning methods which only produces one partition, hierarchical methods generate a series of nested partitions. A hierarchical algorithm yields a dendrogram representing the nested grouping of patterns and similarity levels at which groupings change (Jain, Murty et al. 1999). The dendrogram can be formed in two ways: “bottom-up” and “top-down”. The “bottom-up” approach, also called as agglomerative approach, starts with each object forming a separate group. It successively merges the objects or groups according to specific algorithms until a termination condition is achieved. The “top-down” approach, also named as divisive approach, starts with all objects in the same cluster. The clusters are kept being splitted into smaller clusters according to certain measures iteratively to obtain a partition of singleton clusters (Han.J. 2001).

Most hierarchical clustering algorithms are variants of the single-link, complete-link and minimum-variance algorithms, of which single-link and complete-link algorithms are most popular (Ward 1963; King 1967; Murtagh 1983; Sneath 2005). These two approaches differ in the way they define the similarity between a pair of clusters. For the single-link algorithm, the distance between two clusters is conceptualized as the minimum of the distances among all pairs of objects drawn from two clusters. For the complete-link algorithm, the maximum of all pairwise distance between patterns in the two clusters is regarded as the distance between two clusters. The single-link approach is more versatile then complete-link one, while complete-link algorithm may produce more compact clusters. It has also been observed that the complete-link algorithm produces more useful hierarchies in many applications comparing to single-link algorithm (Jain and Dubes 1988; Jain, Murty et al. 1999).
Earlier hierarchical clustering methods like AGglomerative NESting (AGNES) and DIvisia ANAlysis (DIANA) (Kaufman 1990) have been argued to be over-simplified, which may result in erroneous clusters being found (Han.J. 2001). Recent hierarchical methods mostly adopt Clustering Using Representative (CURE) (Guha, Rastogi et al. 2001) or CHAMELEON approach (Karypis 1999), which employs more complicated principles when splitting or merging clusters, to improve the efficiency and precision of clustering.

Density-based Methods

Besides the aforementioned measurements which are mostly distance-based, other clustering methods have been developed based on the notion of density. Unlike distance-based methods, that can only find spherical-shaped clusters, density-based methods have their advantages of applying for discovering clusters of arbitrary shapes. The main idea of density-based methods is to cluster objects in the dense regions while filtering out low density regions (regarded as outliers or noise) in the data space (Han.J. 2001).

Most widely used density-based clustering algorithm includes Density-Based Spatial Clustering of Application with Noise (DBSCAN), Ordering Points to Identify the Clustering Structure (OPTICS) and DENsity-based CLUstEring (DEMCLUE). In particular, DBSCAN and OPTICS are index-based methods which are not suitable for high-dimension data. DEMCLUE algorithm is based on a set of density distribution functions and is more useful for the application involving high-dimensional datasets.

Grid-based Methods

Grid-based clustering method split the space into a grid structure given the finite number of cells, where all of the clustering procedures are performed. Those methods outperform density-
based methods due to its fast processing time especially in high-dimensional space. Typical examples of grid-based algorithms include STatistical INformation Grid (STING), WaveCluster and CLIQUE.

One advantage of grid-based method is time efficiency. The time complexity of calculating clusters using this method is not related to the number of objects, but the number of grids. Therefore, it is very useful when handling with large dataset.

Artificial Neural Networks (ANNs) methods

ANNs have been used extensively over the past few decades for cluster analysis. An artificial neural network (ANN) is a system which has been derived through models of neurophysiology. In general, it consists of a collection of simple non-linear computing elements whose inputs and outputs are tied together to form the network (Kuo, Ho et al. 2002). Objects need to be clustered are presented at the input and be grouped automatically by the network as a single unit (neuron) based on the data correlation, by which some degree of self-organization are displayed. Competitive neural networks are used to discover for itself patterns, features, and correlations in the input data and code for them in the output. Alike in classical clustering approaches, learning process iteratively re-computes the weights between input nodes and output nodes until a convergence is met (Jain, Murty et al. 1999). Examples of ANNs used for cluster analysis include self-organizing map and adaptive resonance theory (ART) models.

To apply cluster analysis approach for the analysis of spatial attributes, the notion of similarities such as Euclidean or Manhattan distances exists when the spatial attributes are in correspondence to point objects. However, the situation is more complicated for polygon objects, More specifically, the similarity (or proximity) between two polygon objects may be defined in
many ways, which raises the question that which measurement achieves the best balance in terms of results accuracy and computational efficiency when considering spatial clustering.

4.1.2 Methodology

Self-Organizing Maps

An effective data reduction method is necessary for analyzing and representing a large multidimensional spatial data set (Chen, Retherford et al. 2010). Various clustering methods has been applied in image processing, spatial imagery data mining, land-use classification, pattern analysis and many other areas (Steenberghen, Aerts et al.; Hallencreutz and Lundequist 2003; Wang, Luo et al. 2005; Wang, Leung et al. 2006; Baum, Haynes et al. 2007; Yee, Tung et al. 2007; Mu and Wang 2008; Steenberghen, Aerts et al. 2009) due to its capability to reveal two types of information: 1) indicative information, which means the characteristic of cluster itself such as size, extent and location of the cluster and etc, and 2) relational information, which consists of casual factors, potential explanations, cluster justification and association among themes (Estivill-Castro and Lee 2004).

SOM is an artificial neural network algorithm developed in the early 1980s by Kohonen to identify clusters from complex raw data. The method projects high-dimensional input data onto a regular, low-dimensional array of neurons(Kohonen 1982; Kohonen 2001). The superior features of SOM, such as unsupervised computation, handling of large data sets, and its capability to identify clusters via visualization, provide extraordinary insight into original high-dimensional data and has been demonstrated to be more effective than traditional methods (Tayman, Smith et al. 2007).
The dataset prepared for the SOM analysis is in the form of a matrix, with row representing observations and columns the attributes of interest. For each object, it has a vector of the chosen variables, denoted as \(<x_1, x_2, ... x_n>\). For each grouping unit, it has a vector with same dimensionality belonging to the same coordinate space as well. Therefore, input vector can be compared to the weight vectors of clusters.

The implementation of a SOM algorithm involves a training process during which an individual input vector is compared with weighted vectors of all neuron units, such that the best-matching vector that satisfies a minimum distance or maximum similarity criterion is included for the analysis (Pragya Agarwal 2008). Once the best-matching unit is found, weight vectors of neurons must be modified. The distance between neurons must also be adjusted accordingly:

\[ m_i(t + 1) = m_i(t) + h_{ci}[x(t) - m_i(t)] \]

In particular, \(m_i\) represents the weight vector of the \(i^{th}\) neuron; \(t\) represents the number of iterations; \(h_{ci}\) represents a neighborhood function.

Neighborhood function is regarded as one of the main influential parameters contributing to the training process. Among possible functions, linear and the Gaussian model are two most popular approaches. In the Gaussian model, the neighborhood’s size appears as kernel width, which is not a fixed parameter. The neighborhood kernel is treated as a decreasing function of the spatial distance between found best-matching unit and every other unit on the two-dimensional grid. That is, the adjustment of neurons will be maximized for the best-matching unit itself, somewhat smaller for the units in its close surroundings and even smaller for its further surroundings (Pragya Agarwal 2008).
Like other data-mining techniques, the SOM method is a “black box” that does not subscribe to any assumption with regard to inter-data relationships and is sensitive to parameters used for training (Yao 2007; Miller and Han 2009). Component planes and a unified distance matrix (U-matrix) are two most commonly used methods for summarizing a SOM. The U-Matrix, composed of a series of neurons, measures the degree of dissimilarity between clusters used (Ultsch and Siemon 1990). A component plane, by comparison, visualizes the weight of each variable in original input data sets so that the correlation between variables can be observed by comparing their corresponding component planes (Vesanto 1997; Tayman, Smith et al. 2007).

The overall U-matrix provides little insight, however, on the meaning of the observed structure. Spielman used an alternative approach to interpret SOM results by working backwards. They first selected census tracts with known characteristics and identified their locations on the SOM feature map. Locations that fell into the same bucket as known locations are then considered to share similar social characteristics. However, to pursue this approach, one must be highly familiar with the study area in order to select representative tracts. At least one tract of a pre-defined type has to be chosen so that tracts adjacent in attribute space can be labelled. Important patterns may also be neglected due to subjective choices of representatives.

To provide a more objective insight to the characteristics of social areas, we choose to work frontwards in incorporating with principle component analysis (PCA) method. It can reduce the number of variables so that a SOM feature map could be interpreted without reference to all the variables. Once the clusters were identified using SOM, the interpretation of clusters involving two steps were performed: 1) PCA was applied to extract major components in seeking to explain most variance of original data sets. 2) For each cluster identified U-Matrix was used to visualize the degree of similarity within each component. Figure.2 shows an example with four
major components found in the first step. For an identified cluster, the U-matrix of each component was displayed. The third component is shown as having the most uniform pattern comparing to others (highest proportion of low-value neurons). Therefore, the compositions in third component were regarded as representative of the cluster. A further look into component planes of variables within the third components was also needed to help label the cluster.

Figure 12. shows U-Matrix for each of the four major Figr the U-Matrix represents cluster boarder, while the low the cluster itself. High proportion of low-value neurons
Input Variables

It is noted that the 2000 census in China included additional data on indicators as household housing rent, housing cost, source of housing, employment and etc, comparing to the previous censuses. The following tables list available variables from the census data reflecting the socioeconomic and housing characteristics of Beijing residents.

**Table 2. Census Variables used for Social-Area Analysis of Beijing in 1990**

<table>
<thead>
<tr>
<th>General</th>
<th>SexRatio</th>
<th>popDensity</th>
<th>Annual growth rate</th>
<th>BRate</th>
<th>MRate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households</td>
<td>HHSize</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of Familial households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hukou Status</td>
<td>Migrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-term city Dwellers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residents currently working or studying aboard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education Level</td>
<td>Primary School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Secondary School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior College</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialized Secondary College</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduæ Coursework and Above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Census Variables used for Social-Area Analysis of Beijing in 2000**

<table>
<thead>
<tr>
<th>General</th>
<th>SexRatio</th>
<th>popDensity</th>
<th>Annual growth rate</th>
<th>BRate</th>
<th>MRate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households</td>
<td>HHSize</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of Familial households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First generation household</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nationality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minorities population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>people moving in from other provinces/cities</td>
</tr>
<tr>
<td>people moving in from other districts within same city</td>
</tr>
<tr>
<td>people moving within the same district</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hukou Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-term city Dwellers</td>
</tr>
<tr>
<td>Migrants</td>
</tr>
<tr>
<td>Residents currently working or studying aboard</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age0-14</td>
</tr>
<tr>
<td>Age15-64</td>
</tr>
<tr>
<td>Age65+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Educational Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uneducated</td>
</tr>
<tr>
<td>Having basic reading &amp; writing skills</td>
</tr>
<tr>
<td>Primary School</td>
</tr>
<tr>
<td>Junior Secondary School</td>
</tr>
<tr>
<td>Senior Secondary School</td>
</tr>
<tr>
<td>Junior College</td>
</tr>
<tr>
<td>Specialized Secondary College</td>
</tr>
<tr>
<td>Undergraduate Coursework</td>
</tr>
<tr>
<td>Graduate Coursework</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employed by types of occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managerial (government agencies, party, and enterprises)</td>
</tr>
<tr>
<td>Professional and technical</td>
</tr>
<tr>
<td>Office personnel</td>
</tr>
<tr>
<td>Commercial and Services</td>
</tr>
<tr>
<td>Farming, forestry, animal husbandary, and fisheries</td>
</tr>
<tr>
<td>Industry and transportation</td>
</tr>
<tr>
<td>others</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employed by Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farming, forestry, animal husbandary, and fisheries</td>
</tr>
<tr>
<td>Mining and quarrying</td>
</tr>
<tr>
<td>Manufacturing</td>
</tr>
<tr>
<td>Electric power, gas, and water</td>
</tr>
<tr>
<td>Construction</td>
</tr>
<tr>
<td>Geological prospecting and water conservation</td>
</tr>
<tr>
<td>Transportation, storage, postal and telecommunication services</td>
</tr>
<tr>
<td>Outside of Labor Force</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Living Space</td>
</tr>
<tr>
<td>Ownership or rental of houses by household</td>
</tr>
</tbody>
</table>

**Table 4.** Census Variables used for Social-Area Analysis of Beijing in 2009

<table>
<thead>
<tr>
<th>Employed by types of occupation</th>
<th>Managerial(government agencies, party, and enterprises)</th>
<th>Professional and technical</th>
<th>Office personnel</th>
<th>Commercial and Services</th>
<th>Farming, forestry, animal husbandary, and fisheries</th>
<th>Industry and transportation</th>
<th>others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed by Sector</td>
<td>Farming, forestry, animal husbandary, and fisheries</td>
<td>Mining and quarrying</td>
<td>Manufacturing</td>
<td>Electric power, gas, and water</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 4.2 Social –Spatial Patterns of Beijing

#### 4.2.1 Social-Spatial Pattern of Beijing in 1990

The SOM analysis was performed to derive the social-spatial pattern for Beijing in 1990, using 13 variables extracted from the fourth national census. Four social area clusters were identified. Further PCA analysis found three major components that help explain most variance.
of total variables, including: education level, *Hukou* status, and household structure. Following figures show the degree of similarity of the three major components for each class.

<table>
<thead>
<tr>
<th>Class</th>
<th>Component1</th>
<th>Component2</th>
<th>Component3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class1</td>
<td><img src="image1.jpg" alt="Image" /></td>
<td><img src="image2.jpg" alt="Image" /></td>
<td><img src="image3.jpg" alt="Image" /></td>
</tr>
<tr>
<td>Class2</td>
<td><img src="image4.jpg" alt="Image" /></td>
<td><img src="image5.jpg" alt="Image" /></td>
<td><img src="image6.jpg" alt="Image" /></td>
</tr>
<tr>
<td>Class3</td>
<td><img src="image7.jpg" alt="Image" /></td>
<td><img src="image8.jpg" alt="Image" /></td>
<td><img src="image9.jpg" alt="Image" /></td>
</tr>
<tr>
<td>Class4</td>
<td><img src="image10.jpg" alt="Image" /></td>
<td><img src="image11.jpg" alt="Image" /></td>
<td><img src="image12.jpg" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 13. U-Matrices show the degree of similarity of the six major components for each class. Class 1 to 4 represents Area of Agricultural Population; Lower Density, Inner Suburb Area; Migrants Concentrated Area and High Density, Urban Area, respectively. While component 1-3 stand for education level, *Hukou* status, and household structure.

Four social area clusters are then labelled according to comparisons of U-Matrix among each component.

Cluster 1. Area of Agricultural Population. This category includes 245 contiguous sub-districts, mainly in outer suburb areas that consist of a primarily agricultural population. This area is characterized as low population density and low educational level.
Cluster 2. Lower Density, Inner Suburb Area. This category consists of 23 sub-districts primarily located between urban and outer suburb areas. These sub-districts form a belt-like shape surrounding the urban area and are characterized as having lower population density compared to the urban core and a higher education level compared to the outer suburb area.

Figure 14. Social Area of Beijing in 1990
Cluster 3. Migrant Concentrated Area. This area consists of 13 sub-districts dispersed in the outer suburb area, characterized as having a high proportion of migrants.

Cluster 4. High Density, Urban Area. This area consists of 114 sub-districts which are clustered in old urban core areas and inner suburb areas. These sub-districts share similar characteristics, such as high population density, higher educational level and a higher proportion of residents with registered “Hukou” status.

4.2.2 Social-Spatial Pattern of Beijing in 2000

We identified six clusters after performing SOM analysis. Similar procedures as depicted above, relative to 1990 data, were implemented.

Six major components are identified according to PCA analysis, including: professional workers, education level, people occupied in farming, forestry and other agricultural related works, living conditions of households, people occupied in mining, construction and other labor-intensive works, and household structure.
<table>
<thead>
<tr>
<th>Class</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
<th>Component 4</th>
<th>Component 5</th>
<th>Component 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Class 2</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Class 3</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Class 4</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Class 5</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Class 6</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 15. U-Matrices show the degree of similarity of the six major components for each class.
Class 1 to 6 represents Inner Suburb Area; High Density, Intellectual Area; Migrants Concentrated Area; High Density, Commercial Area; Area of Agricultural Population and an Outer Suburb Area with Urban Residents, respectively.
Component 1-6 stand for professional workers, education level, people occupied in farming, forestry and other agricultural related works, living conditions of households, people occupied in mining, construction and other labor-intensive works, and household structure.
By visually interpreting U-Matrix of six major components found, identified clusters were interpreted as below.

Cluster 1. Inner Suburb Area. Consisting of 32 sub-districts, this type of social area was summarized as having a relatively high population density and employment ratio. Such sub-districts were mainly found in the margins of inner suburb and outer suburb area. A small

Figure 16. Social Area of Beijing in 2000
number of sub-districts within the urban core area were categorized as being this type, showing similar patterns as those of other suburb regions.

Cluster 2. High Density, Intellectual Area. This social area type is located within inner suburbs and partly near outer suburbs. Major characteristics of 56 sub-districts within this type are high employment ratio, high education level and high population density. Many colleges and universities are located in this area, accounting for a high concentration of intellectuals in this area.

Cluster 3. Migrants Concentrated Area. 26 sub-districts were categorized into this social area type with a high proportion of migrant population. Those sub-districts distribute primarily in the margin of the inner suburb.

Cluster 4. High Density, Commercial Area. This area features high population density, a high proportion of people engaged in commercial-related activities. This category consists of 56 sub-districts dispersed in the periphery of the urban core.

Cluster 5. Area of Agricultural Population. With a low population and a larger average living space, a high proportion of workers in agricultural related-fields and a low education level has been found in this social area, which we have labelled as area of agricultural population. Those sub-districts are concentrated in the outer suburb of the city.

Cluster 6. Outer Suburb Area with Urban Residents. Uniform patterns within a worker component identified from PCA indicates that there are workers from other than an agricultural population residing in these regions. 59 sub-districts are mainly concentrated in the southwestern outskirts of the city, while some are scattered in the northern outer suburb.

4.2.3 Social-Spatial Pattern of Beijing in 2009
Following the same process that was used for the two previous years, five clusters were found in the year 2009.

Figure 17. U-Matrices show the degree of similarity of the six major components for each class. Class 1 to 5 represents Inner Suburb Area; High Density, Intellectual Area; High Density, Migrants Concentrated Area; Lower Density, Urbanized Suburb Area; High Density, Commercial Area and an Outer Suburb Area, respectively. Component 1-5 stand for professional workers or people with decent education background, economic status, education level, household characteristic, population movement, and employment.
Five major components are identified according to PCA analysis, including: professional workers or people with decent education background, economic status, education level, household characteristic, population movement, and employment.

Interpretation of five identified clusters is stated below.

Cluster 1. High Density, Intellectual Area. Regions recognized as this social area type were characterized by rapid employment growth and high education level, indicating a sprawl of the
working area. 49 sub-districts of this type were distributed in the urban core and extend to the outskirts of the suburb area.

Cluster 2. High Density, Migrant Concentrated Area. 18 sub-districts were categorized within this type, mainly characterized as having a high proportion of migrants, high frequency of movement and rapid growth of employment. This social area includes 16 sub-districts within inner suburbs and two outer suburb sub-districts far away from the urban core.

Cluster 3. Lower Density, Urbanized Suburb Area. Comprisison of 65 sub-districts dispersed primarily over an outer suburb area, this social area was recognized as having higher population density within the suburb area and with higher income and car ownership.

Cluster 4. High Density, Commercial Area. Sub-districts within this type include an eastern inner suburb region and several regions to the north of the inner suburb edge. Inner suburb areas within type were characterized as having high a employment ratio and educational level, suggesting a concentration of industry-oriented land use type. Outer suburb areas within this type showed compartively higher frequency of movement and a high empolyment ratio. A lower education level and proportion of workers in foreign-invested enterprise, however, indicate different types of employment compared to the inner suburb area.

Cluster 5. Outer Suburb Area. Low population, low employment ratio and low education level were dominant characteristics in this social area, which are labelled as areas of agricultural population. These sub-districts are concentrated in the outer suburbs of the city.

4.3 Restructuring of Social Space in Beijing

Comparing the social-spatial structure identified in years 1990, 2000 and 2009, it is apparent that dramatic change had taken place in the city during this 19-year period.
From year 1990 to 2000, transformation of social space has mainly occurred in two ways. First, the urban area identified in 1990 had separated into different social area types, indicating a functional separation within the area. New social areas, including a new high density commercial area and a high density intellectual area have emerged in the outskirts of the urban core and inner suburb areas, resulted primarily from the expansion of markets and formation of CBD areas. Several outer suburb sub-districts, recognized as migrant concentration areas in previous years, had been incorporated into these two categories as well, while new migrants were more concentrated in the inner suburb belt in 2000. Another noticeable change is that several sub-districts in the urban core area present similar characteristics with those of the inner suburb area, which can be attributed to a suburbanization process that causes relocation of urban residents to suburb areas. Second, the predominence of a urban residents population was broken once the homogeneous outer suburb area, known for its concentration of agricultural population, had developed further. This also reflects the impact of suburbanization and the development of new satellite towns.

From the year 2000 to 2009, suburbanization had further progressed. Transformation of social space can be seen to have unfolded in the following ways. First, the intellectual area had expanded, especially into suburb areas adjacent to the intellectual area in years previous to this time period. Second, the commercial area was enlarged, represented in both the expansion of previously dispersed sub-districts in inner suburbs and the newly developed areas of outer suburb to the north of the urban core. This change reflects further suburbanization continued from the last decade and may be due to improvements in transportation networks and infrastructure. Third, in all directions more outer suburb areas were urbanized during this time period, but were primarily shown in a cluster pattern in the southeastern part of the city. In addition to the
continuing suburbanization process in the area, another noticeable change was that during this time period there was a decrease of population density in the urban core area. Contrarily, population growth and dramatically increasing housing prices in the urban core forced people to move to the more affordable periphery of the city. At the same time, the expansion of transportation networks has allowed people to commute between the periphery of the city to the urban core, which makes living and working in different areas more feasible. On the other hand, some more wealthy people moved away from the city core. They choose alternative housing in the outskirts of the city where they could have access to better living environments and avoid the crowdedness of the city. Increasing private car ownership also lessened the barrier of a long commute. A series of Olympic-related relocation projects also played an important role in this process. A lot of new cultural facilities, such as Olympic Game Park, are constructed. Lots of old industry sites have been revitalized and some noxious industrial enterprises were moved to sub-urban area for the overall city imagery. Meanwhile, multiple instances of relocation has occurred in Beijing due to various reasons such as demolition of residential areas for pre-Games construction, forcing local residents to move out of the old urban districts to planned construction projects.
5 Socioeconomic Process Related to Social Space Restructure

While SOM can help elucidate meaningful social-spatial patterns, it cannot provide direct evidence of the driving forces behind social-spatial structure change. The mechanisms underlying the restructuring of social areas are complex and involve multiple factors, such as public policy, economic markets and culture. In our study, we primarily focus on the relationship between population dynamics and residents’ social status, with the assumption that the degree of population change in one area is associated with social status, in terms of occupation, educational background, hukou status, etc., of residents living there. Once a significant relationship is found between characteristics of certain population groups exhibiting population change, related factors, such as public policy and economic markets, could be further explored.

5.1 Geographically Weighted Regression (GWR)

Spatial dependence is often found among the causes of population change, the patterns of which are mainly explained by regional economic theories, population geography theories and residential preference studies (Chi 2010). Previous studies have suggested that spatial dependence is crucial to population dynamics (Voss and Guangqing 2006). Spatial heterogeneity is also a spatial process effect, which accounts for the variation in correlations between dependent and independent variables over the space considered (LeSage 1999).

Regression models are generated to address the aforementioned assumption. The spatial heterogeneity of population change over the 2000-2009 period was also assessed using Geographically Weighted Regression (GWR). Specifically, we hypothesized that there is significant spatial variation in the relationships between different social groups and the
population dynamics in question, which are also indicators of social-spatial change during this time period considered.

GWR is a spatial data analysis technique first introduced by Fotheringham, Charlton and Brunsdon (Brunsdon, Fotheringham et al. 1998; Fotheringham, Brunsdon et al. 2002). This method is increasing being used to examine spatial variation or nonstationarity in local politics and voting patterns (Calvo and Escolar 2003), poverty related socioeconomic process (Benson, Chamberlin et al. 2005; Farrow, Larrea et al. 2005), urban growth (Partridge, Rickman et al. 2008; Luo and Wei 2009), regional industrialization and economic growth (Huang and Leung 2002; Ocal and Yildirim 2010), relationship between wealth and land cover (Ogneva-Himmelberger, Pearsall et al. 2009), environmental justice (Gilbert and Chakraborty 2011), commuting patterns (Lloyd and Shuttleworth 2005), housing prices (Bo, Bo et al. 2010) and violent crime (Cahill and Mulligan 2007).

GWR model allows parameters to be estimated locally, thus they can capture local variations in relationships between dependent and independent variables (Bo, Bo et al. 2010). The model can be expressed as

\[ Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \quad i = 1, \ldots, n \]  

(1)

Where \((u_i, v_i)\) denotes the location of observation \(i\), \(\beta_0(u_i, v_i)\) denotes the intercept, and \(\beta_k(u_i, v_i)\) represents a set of coefficients at each observation \(i\).

For each observation, GWR model intends to estimate one regression. Therefore, according to formula (1), each of the \(n\) observations has an individual set of \((k + 1)\) parameters. The parameter in GWR model is calculated by

\[ \hat{\beta}_i = (X'W_iX)^{-1}X'W_iY \quad i = 1, \ldots, n \]  

(2)
Where $W_i$ is an $n \times n$ weight matrix whose diagonal elements are the weights of each observation relative to observation $i$ (Partridge, Rickman et al. 2008).

The dependent variable used in the regression model was the sub-district-level percentage of population change in the total population for the 2000 and 2009 period. Independent variables include the percentage of increased urbanized area comprising in the total area, and the percentage of different population groups in the total population from the 2000 census. Considered variables include:

- Educational level (uneducated, primary school, middle school, high school, college, graduate or above),
- Occupations (managerial, professional and technical, commercial and service, mining and agricultural)
- *Hukou* status (permanent residents, migrants)
- Housing condition (living space, ownership of the house)
- Landuse characteristic (percentage of urban area within each sub-district)

### 5.2 Empirical Results

Given the large number of census variables, it was difficult to compare across different models, as doing so introduces the problem of multicollinearity among variables. Following the suggestion of Chi (Chi 2010), a pooled regression approach was adopted to exclude insignificant variables. A smaller number of variables that are significant and uncorrelated were selected. Final regression models include eight variables.

| Table 5. Comparison of statistical test of results from OLS, SLM, SEM and GWR model |
|---------------------------------|------|------|------|------|
|                                 | OLS  | SLM  | SEM  | GWR  |
| R-Squared                       | 0.69 | 0.70 | 0.75 | 0.92 |
An OLS regression was first performed according to claimed assumptions. The overall performance of the model (shown in Table.1) reveals the significant impact of explanatory variables on population dynamics. This validates the general hypothesis.

Examination of spatial autocorrelation of OLS residual, using Moran’s I, indicates a slight clustering tendency in the residuals. Therefore, spatial regression models, including spatial error model (SEM) and spatial lag model (SLM), were performed to taking account for the spatial homogeneity. The results of both models, however, do not show significant improvement over levels of significance indicated by the OLS model; nor did there appear to be a much effect on the coefficient values and their significance. This suggests that spatial heterogeneity was the more important factor impacting the spatial patterns of the variables.

A GWR model was then performed to examine whether heterogeneity existed in the causality between population dynamics and different social groups in comparison. According to a non-uniform distribution of sub-districts (dense distribution in the metropolitan area and sparse distribution in the remote counties), an adaptive kernel was selected as a spatially weighting scheme (Partridge, Rickman et al. 2008). This ensures consideration of a certain number of neighbors, the value of which is determined by Akaike Information Criterion (AIC) tests, as local samples, and is a better representative of spatial heterogeneity (Luo and Wei 2009). Based on the AIC test result, the optimal number of nearest neighbors was determined to be 30.

| AIC       | 577.25 | 567.52 | 522.3  | 266.07 |
Table 6. Coefficients of dependent variables in OLS model and statistics of according coefficients in GWR model

<table>
<thead>
<tr>
<th>Explanantory Variables</th>
<th>OLS</th>
<th>SEM</th>
<th>SLM</th>
<th>GWR</th>
<th>Max</th>
<th>Min</th>
<th>Median</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>% of Negative</th>
<th>% of Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>% professional and technical workers</td>
<td>-0.240***</td>
<td>-0.201***</td>
<td>-0.240***</td>
<td>0.014</td>
<td>-0.205</td>
<td>-0.095</td>
<td>-0.001</td>
<td>0.110</td>
<td>0.014</td>
<td>92.79%</td>
<td>7.21%</td>
</tr>
<tr>
<td>% age 14-65</td>
<td>0.231**</td>
<td>0.343***</td>
<td>0.231***</td>
<td>-0.070</td>
<td>-0.235</td>
<td>-0.152</td>
<td>-0.001</td>
<td>0.083</td>
<td>0.014</td>
<td>63.66%</td>
<td>36.34%</td>
</tr>
<tr>
<td>% uneducated</td>
<td>0.342***</td>
<td>0.513***</td>
<td>0.342***</td>
<td>0.637</td>
<td>0.034</td>
<td>0.335</td>
<td>0.002</td>
<td>0.301</td>
<td>0.014</td>
<td>2.10%</td>
<td>97.90%</td>
</tr>
<tr>
<td>% graduate degree</td>
<td>0.317***</td>
<td>0.312***</td>
<td>0.316***</td>
<td>3.633</td>
<td>1.978</td>
<td>2.806</td>
<td>0.017</td>
<td>0.828</td>
<td>0.014</td>
<td>1.50%</td>
<td>98.50%</td>
</tr>
<tr>
<td>% live less than 6 month</td>
<td>-0.086*</td>
<td>-0.052</td>
<td>-0.086**</td>
<td>0.034</td>
<td>-0.052</td>
<td>-0.009</td>
<td>0.000</td>
<td>0.043</td>
<td>0.014</td>
<td>61.56%</td>
<td>38.44%</td>
</tr>
<tr>
<td>% live 6 month to 1 year</td>
<td>-0.314***</td>
<td>-0.239***</td>
<td>-0.313***</td>
<td>0.034</td>
<td>-0.116</td>
<td>-0.041</td>
<td>0.000</td>
<td>0.075</td>
<td>0.014</td>
<td>79.58%</td>
<td>20.42%</td>
</tr>
<tr>
<td>% households renting house</td>
<td>0.769***</td>
<td>0.659***</td>
<td>0.768***</td>
<td>0.811</td>
<td>0.142</td>
<td>0.476</td>
<td>0.003</td>
<td>0.334</td>
<td>0.014</td>
<td>3.00%</td>
<td>97.00%</td>
</tr>
<tr>
<td>% urbanized area in past ten years</td>
<td>0.284***</td>
<td>0.051***</td>
<td>0.284***</td>
<td>0.081</td>
<td>-0.150</td>
<td>-0.034</td>
<td>0.000</td>
<td>0.116</td>
<td>0.014</td>
<td>36.04%</td>
<td>63.96%</td>
</tr>
</tbody>
</table>

Note: *** represents variables are significant at $p=0.005$ level, ** at $p=0.01$ level * at $p=0.05$ level.

Comparison among global and local models, showing great improvement of GWR specification from the OLS model and spatial regression models, according to the R square value and AIC score. The parameter estimates of GWR also show significant variation, which supports our general hypothesis of heterogeneity. For comparison, spatial regression models are performed as well.

A set of parameter estimates of explanatory variables for each sub-district, generated from the GWR model, can be pursued to analyze spatial variation of the effects of various social groups on population dynamics, as shown in Table.2. From a spatial distribution map of each variable, as shown in Figure.5, it is apparent that all parameters vary across the study area with generally
regular spatial patterns. A blank sub-district indicates that the according variable was insignificant in that area, while the sub-districts of different colors show varied coefficient values.

From the global model, the percentage of professional and technical workers in the total population had a negative effect on population change. This general pattern was confirmed by local estimates, while 92.79% of the sub-districts reflected the same trend. Several districts in the fringe of the outer suburb areas (Daxing and Shunyi) represent an opposite relationship.

Use of the global regression model indicated a positive relationship among a percentage of the age 15-64 population group, while this relationship varies in the GWR model. The impact of people in the 15-64 age group on population change was not seen to be significant in the southwestern part of the city. The urban core area, inner suburb area and outer suburb area in the northeast corner showed a consistent negative relationship under the global model, while the remaining part of the outer suburb area witnessed an opposite trend.
Figure 19. (A) the distribution of coefficients for proportion of professional workers (B) the distribution of coefficients for proportion of people aged in 15-64 (C) the distribution of uneducated population (D) the distribution of coefficients for proportion of people with College degree and above
Figure 20. (A) the distribution of coefficients for proportion of migrants living six months to one year (B) the distribution of coefficients for proportion of migrants living less than six month (C) the distribution of coefficients for proportion of households living in rented house (D) the distribution of coefficients for proportion of newly increased urbanized area
Educational background, both high education level (graduate degree) and uneducated, tended to positively effect the population change when viewed from the global model. From the local estimate, we saw that the positive impact of an uneducated population on population change was significant mainly in the urban core area and five northern and eastern outer-suburb districts. The impact of a well-educated population, on the other hand, was seen to be significant only in the inner suburb area and parts of the southeastern outer suburb area. The coefficient had an increasing tendency to be evidenced toward the outer suburb region.

Using the global model, migrants just moving into the city (having lived less than six months or living six months to one year without Beijing Hukou) were seen to be negative indicators of population change. From the GWR model, this group of people tended to have a positively effect on population change mostly in the urban core area, parts of inner suburb area, and in three outer suburb districts (Yanqing, Mentougou and Fangshan). The values of coefficients were about the same in range when the distance from the sub-district to the urban core was similarly. This pattern was especially clear for high and low values. Migrants living in the area six months to one year mainly had a negative impact on population change, as shown in local estimates by the global model, except in the northeastern part of the city. However, the impact of the global model tended to be statistically insignificant in two urban core districts (Xicheng and Xuanwu) and part of inner suburb area.

The proportion of households living in rental houses tended to have a positive effect from a global perspective. Based on local estimations, its impact was significant in three northern districts, while several sub-districts adjacent to the northern border of the city showed a negative influence.
Using the OLS model, a positive impact within the newly urbanized area during last ten years can also be observed. The GWR model suggested that the impact of urbanization was not significant in two urban core districts (Dongcheng and Xicheng), and parts of inner suburb region. It also indicated that urbanization influenced population change negatively in two other urban core districts (Chongwen and Xuanwu), and most inner suburb areas.

### 5.3 Related Social Phenomena

The result from use of the GWR model has validated the assumption that the social-spatial

![Density of Population with College Degree or Above](image)

Figure 21. Population Density of People with College Degree or Above
structure of Beijing is transforming from a homogeneous structure to heterogeneous one (Feng, Wu et al. 2008). Some critical social phenomena can also be observed from the distribution of dependent variables related to population change within the last ten years.

The proportion of graduate students has been shown to be a significant variable only in the southeastern outskirt of the city and part of inner suburb area to the north of the urban core area. “Ant Tribe,” a new term within recent years, refers to low-earning graduates from second-tier universities who can afford only ramshackle housing on the outskirts of big cities like Beijing (Ant Tribe). These college graduates are compared with ants, given the “intelligent, hardworking and strong-in-groups” characteristics of ants. Some have argued that those graduates should be considered as the fourth weakened social group, “after peasants, migrant workers and the unemployed.” (reference). It is estimated that there are around one million of the “ant tribe” population across the entire country, with about 100,000 in Beijing.

Under-employment and the gap between low entry-level salaries and high real estate prices are two major reasons for the formation of the “Ant Tribe” phenomenon. The expanding enrollment of colleges since 2003 led to increasing numbers of college graduates, while the corresponding employment opportunities have been lacking. This scenario has resulted in the formation of “settlement villages” on the periphery of Beijing, which has formed a potentially a new type of social area. Given the estimate that another 6.3 million graduates are expected to join migrant workers and other job hunters in what promises to be a fierce labour competition, the size of the “ant tribe” and its corresponding “settlement village” will continue to increase for the foreseeable future. The expansion of this social group will not only pose new challenges for the governmental policy, but is also expected to impact the social-spatial pattern in the city. On the other hand, the well-educated background of this group differentiates the “settlement villages”
from the slums in South America or Southeast Asia, where mostly poor and unemployed are concentrated, a phenomenon that poses new questions for researchers.

In comparison to relatively confined regions impacted by college graduates, migrants have a significant effect on population change across almost the entire city. Most of the floating population does not have local Hukou status. This results in their instability and is seen in their floating nature. When compared to local residents, they lack equal access to some job opportunities (Zhu 2007). Their unstable nature causes population fluctuation as they move frequently to find

Figure 22. Population Density of Temporary Residents (Living Six Month to One Year)
better jobs and living place. A noticeable phenomenon is that since most of those migrants are occupied in service and construction, they are most likely not be well-paid and can not afford increasing rental prices in the capital city.

A group of migrants who alternate choose to live in the basements of apartment buildings, or even in former air raid shelters, named the “rat tribe” by the sociologists, has also developed, since the underground space are the only places they can afford. Unlike the 'ant tribe,' members of the 'rat tribe' live in even worse conditions and hail from a wider range of backgrounds.
6 Conclusion

Studying the differentiation of social areas at the sub-district level is crucial to the Chinese government. The basic unit of service delivery in China is the residential area, as considered from the perspective of each area’s base-level organization (Wu 2005). Though more spatial data are currently available to the public at finer resolution in urban China, the data are still confidential and thus underdeveloped when compared to such data for western cities. By combining currently available census data with remote sensing data, this study provides insight into the changing social spatial patterns in Beijing, especially with regard to the most recent changes, which occurred after Beijing was awarded host city status for 2008 Olympic Games. The underlying processes of the most recent changes are also studied by measuring associations between population change and diversity among social groups. To conclude my study, this chapter is organized as follows: the first section summarizes major methods used in the study and how they were developed to better accommodate the needs of the study. In the second section, major findings are stated, along with their broader social implications. Finally, in light of research findings within the scope of this work, several areas are suggested where further study may be fruitful.

6.1 Summary of Methodology and its Contribution

Technically this research contributes to the existing literature in the following ways. First, by adopting advanced remote sensing technologies, we interpolated spatially aggregated demographic data at the sub-district level as an initial basis for further analysis. In this effort we examined performance of spectral bands in the land-cover classifications of high-density urban areas. In our analysis four bands appeared to outperform in their respective land-cover classifications, suggesting differences among land-cover classes. Nine out of ten variables were
selected for land-cover classification using neural networks related directly to newly added bands. Red-Edge band and Near Infra Red 2 were particularly useful for the classifications shown in the study area. A rule-based land-use classification framework was then proposed for the remainder of the study. Given the difficulty in automated image interpretation of high-resolution remote sensing images, the proposed framework utilized not only spectral information, but also spatial, structural, and contextual information. This combination of data inputs allowed for the development of high-resolution imaging. Rule-based classification provided the advantage of easily incorporating expert knowledge in the process of automatically interpreting images. The proposed framework could therefore easily be extended or adjusted according to specific needs, or contribute to other applications.

Secondly, in the study of social areas of Beijing, cluster Analysis was performed to identify social-spatial patterns during three different time periods. Following the study of Spielman and Thill (2008), SOM was chosen to aid in determining the social spatial patterns of the areas in question. Combined with geographic information systems, SOM is recognized as having the advantage of retaining fundamental social fingerprints when representing general social patterns. It is capable of filtering a complex demographic reality. The method shares limitations with other clustering techniques, in that the interpretation is somewhat subjective and thus needs to be considered with caution. A two step interpretation method was used in this paper where, by integrating with the PCA method, we avoided the complication of considering the spatial pattern of each single variable. Instead, we interpreted the data by focusing on the major explanatory components. Another advantage of this method is that no presumption prior to interpretation is needed. In other words, one does not have to label the cluster with absolute precision. Instead,
clusters were defined according to the visualization of the results, which is relatively more objective.

6.2 Major Findings and their Broader Implications

In this section the significant findings from my study are used to answer the following interrelated research questions:

1) What were the urban social-spatial structures of Beijing in the years 1990, 2000 and 2009 respectively?

2) How did the social spatial patterns change over time?

3) From the standpoint of the diversity of social groups, what are the currently evolving processes impacting the urban landscape within Beijing?

In my study, four social areas were identified for the year 1990: lower density areas of agricultural population, an inner suburb area, a high density migrant-concentrated area, and a high density urban area. Six social areas were identified for the year 2000: an inner suburb area, a high density intellectual area, a high density migrant concentrated area, a high density commercial area, an area of agricultural population and an outer suburb area with urban worker residents. Five social areas were identified for the year 2009: a high density intellectual area, a high density migrant-concentrated area, a lower density urbanized suburb area, a high density commercial area, and an area of agricultural population.

The change in social-spatial patterns over time suggests the emergence of new social area types and the formation of more disaggregated social-spatial structures. Before the commercial property boom of the early 1990s, social equality within the city was considerable under the planned economy. General social-spatial patterns were closely linked to the differentiation between land-use zones, as observed in the study. From 1990 to 2000, the emergence of new
urban functions was viewed as a significant factor in the process of social-spatial pattern transformation. New social area types, namely the commercial area and the intellectual area resulting from increased economic activity, broke a once homogeneous social structure. Those two social areas represented high status zones where most residents were well-educated professional workers. These groups of people were relocated to the inner suburb areas and remain in higher status zones, the process of which is similar to that described in the concentric model in the West.

The restructuring of urban social space between the years 2000 and 2009 represents a historical continuity from the previous time period. The urbanization process expanded to outer suburban areas, leading to disaggregation in the area. High status zones, specifically the intellectual area and the commercial area, were shown to have enlarged during the time period as well. A study of the driving forces for population redistribution within this time period suggests that temporary residents were the main source contributing to population dynamics in recent years. The superior performance of the GWR model, compared to the corresponding OLS and spatial regression models, validates the assumption of spatial heterogeneity in the association between population change and the distribution of diverse social groups. The distribution of independent variables and their significance in characterizing the impact of population change suggests the contribution of various social groups to the total population change. Social groups, considered to be “social vulnerable groups”, such as migrants and new college graduates, were found to have a significant effect on the overall structure of social-spatial patterns within the city.

It is not surprising that the urbanization process has progressed rapidly under such rapid economic development in China. It may be doubtful, however, whether the restructuring of Beijing’s social space has largely diminished discrimination or alternatively worked to the
advantage of inhabitants, particularly “social vulnerable groups”. Our study shows an expansion of high status zones during the study period, indicating a possible strengthened separation between the affluent and the poor. The differentiation is rooted in socialism. For most people, the affordability of their housing has largely depended on the properties they currently owned. On the one hand, the urban housing reform provisions are mainly aimed to benefit officially registered urban residents. People previously in good quality apartments are likely to benefit more from privatization. On the other hand, urban relocation projects accompanied by rising land values forced many people, especially temporary residents living in the private-rental sector, to move to the periphery of the city, as they can no longer afford to live in their old neighborhoods. The gap between the affluent and the poor may have increased through this process. People with high social status will take advantage of redevelopment to upgrade, while lower social status people will always lag behind.

Inequalities in both the processes and outcomes of urban restructuring call for consideration of allocations for “social vulnerable groups.” In considering the major forces in shaping the social space within the city, allocations of space for these groups is crucial to the social restructuring of the city. The fast development of the city is likely to be at the expense of the rights of those social vulnerable groups, who usually experience the greatest sensitivity to changes in house markets and their related policies. They usually move from one place to the other once they cannot afford the increasing rental prices or cannot satisfy certain govermental regulations, since most of those temporary residents do not own their own houses and thus tend to move for better living conditions or lower rental prices. These social vulnerable groups are most likely to be the major forces in shaping urban social space in the next few years, given the increasing population in these groups.
To conclude, social restructuring in Beijing shows similarity to that described in the concentric model in the West. But the process is much more complex and differs from the model given China’s unique political environment, history and policies. “Social vulnerable groups” are shown to be the major forces in shaping social space within the city, and might continue to have effects on reshaping urban social areas in the city.

6.3 Areas for Future Study

Based on this framework developed in this thesis, further study in the following areas could be pursued in the future: 1) additional contextual information could be included in the study for land-use classification, such as road and transportation networks, as well as large historical and recreational sites; 2) different methods could be used to assign a proportion of each land-use class of the total population to further improve accuracy of estimation; and 3) the framework could be applied to various other types of census estimation (e.g. using industrial/ commercial area as indicators for employment estimation).

The SOM data-mining method for social area analysis is another area noteworthy for further study. Given the increased availability of additional census data, effective methods for exploring geodemographic patterns is needed. While the interpretation method used in this thesis seeks to avoid subjective judgements through analysis, there are still subjective factors, such as the determination of the numbers of the clusters. There are no well-established criteria by which one can decide whether a certain number of clusters can explain the overall pattern better than another number of clusters. Therefore, analysts may have different interpretations of the spatial pattern after applying the same analytical process. Another challenge worthy of further consideration is in the inconsistency of census variables. Though general social spatial patterns could be observed, interpreted and explained, one limitation of this study is that the change of
patterns through different time periods may be influenced by the variation of the variables. This may cause a significant difference in the results, supposing for instance, that one decides to track the changes closely or at a more precise level. More integrated data mining approaches for spatial-temporal data are worthy of discussion in the future work.

According to my study, high status zones are maintained and have enlarged through the urbanization process. Urban inequalities may present a further area of inquiry, as this present research suggests that “weakend groups” are having a significant impact within the urban social areas of the city. Will a new type of urban poor emerge in the city? How will this social group impact the overall urban social structure? And if the gap between rich and poor continues to increase, how will it impact the urban landscape and what kind of policies should be developed or imposed to address it? Allocations of social vulnerable groups calls for attention in public policy, in that the social vulnerable groups will most likely not have equal access to infrastructure, such as transportation, drinking water, etc, compared to other local urban residents. New policies for cooling down the property market have been implemented in Beijing such that millions of migrants are blocked from owning their own homes within the capital city. These policies are implemented under strict new regulations. Migrants are not allowed to buy their own houses in Beijing until satisfying stringent requirements. It is not yet clear whether the new policies will be effective in controlling the skyrocketing housing prices, or whether they will boost the rental market in the city.

Social vulnerable groups will experience the most sensitivity to changes in housing prices. How the new policies impact the market and how the social vulnerable groups respond to them seems worthy of further investigation. The possibility of an emergence of new urban poverty and the development of urban slums is an area for further consideration. While an identical or
uniform process over a city cannot be expected, the extent and trend of social differential is important for policy planning.
7 References


Miller, H. J. And J. Han (2009). Geographic Data Mining And Knowledge Discovery / Edited By Harvey J. Miller, Jiawei Han, Boca Raton, FL : CRC Press, 2009.


