

Article

Analysis of Beijing's Working Population Based on Geographically Weighted Regression Model

Yanyan Chen ^{*}, Hanqiang Qian and Yang Wang

Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China;
Qianhq@mails.bjut.edu.cn (H.Q.); wang_yang@bjut.edu.cn (Y.W.)

* Correspondence: cdyan@bjut.edu.cn

Received: 26 April 2020; Accepted: 17 June 2020; Published: 19 June 2020



Abstract: Evaluation of urban planning and development is becoming more and more important due to the large-scale urbanization of the world. With the application of mobile phone data, people can analyze the development status of cities from more perspectives. By using the mobile phone data of Beijing, the working population density in different regions was identified. Taking the working population density in Beijing as the research object and combining the land use of the city, the development status of Beijing was evaluated. A geographically weighted regression model (GWR) was used to analyze the difference in the impact of land use on the working population between different regions. By establishing a correlation model between the working population and land use, not only can the city's development status be evaluated, but it can also help city managers and planners to make decisions to promote better development of Beijing.

Keywords: working population; geographically weighted regression; mobile phone data; city evaluation

1. Introduction

Since the Industrial Revolution, many countries have begun large-scale urbanization. Over the past 40 years, China has experienced rapid economic growth and urbanization. The urbanization rate increased from 19.39% in 1980 to 58.52% in 2017. In these 40 years, the urban population has grown at a rate of 16 million people per year [1,2]. China's rural population fell by 17% between 2000 and 2010 [3]. This means that in many countries around the world, including China, cities are playing an increasingly important role in economic development [4]. A good urban development evaluation model can evaluate the current status of urban development and guide the city's future planning to help cities achieve healthy and sustainable development [5–7].

The great popularity of mobile phones and the application of mobile phone data provide more possibilities for city evaluation because mobile phone data has many advantages. First, the sample size of the mobile phone data is large, secondly, the acquisition cost of the mobile phone data is low, and finally, the mobile phone data has a wide and complete spatial coverage. Therefore, the mobile phone data can reflect the characteristics of residents in a certain area or the connections between residents in some areas, and can well reflect the projection of residents' activities in space [8]. Therefore, a large number of scholars use mobile phone data to evaluate the development of the city from different angles. For example, Tang and Lin, based on Shenzhen's mobile phone data and point-of-interest (POI) data, used geographic and time-weighted regression methods to connect residents' travel information with different types of infrastructure and evaluate the vitality of the city [9]. Liu et al. used Nanjing's mobile phone data and POI data to extract the vitality of the city through hierarchical clustering [10]. Montero combined Catalan mobile phone data with survey data through principal component analysis to analyze travel characteristics in Catalonia [11]. Zhang et al. used the K-means clustering method to analyze the relationship between Beijing's travel characteristics and land use, using Beijing's mobile

phone data and POI data. Zhang classified Beijing's land into four different categories according to travel characteristics [12].

Existing studies use mobile phone data that is often converted into resident travel data [13–15]. However, travel characteristics cannot directly show the development status of the city. In this article, mobile phone data is further processed into the residential and working places of citizens. The output value of the service industry in Beijing accounted for more than 79.65% of the total annual output value in 2015 [16]. Because jobs are mainly concentrated in the tertiary industry, the contributions of different jobs to urban development and the economy are similar. The working population density of Beijing can reflect the economic status and development status of a certain area in the city to a certain extent. Therefore, this article attempts to reflect the economic vitality of this area through the working population density estimated in different areas. The working population can promote urban development, create fiscal revenue, and maintain social stability. For the sustainable development of cities, it is very important to maintain a reasonable and stable working population [17–19]. The growth of the working population is also a necessary condition for rapid urban development. The analysis of the working population by using comprehensive land use indicators can not only analyze the factors restricting the development of the working population from a new perspective, but also evaluate the current status of urban development and, therefore, assist people to rationally plan and manage the city [20].

This study uses Beijing as the research area. Through mobile phone data, the density of working and living populations in different areas of Beijing was identified. A model of the working population and land use characteristics were established using the method of geographically weighted regression. The geographically weighted regression model takes into account spatial heterogeneity, so it can reflect the difference in the correlation of indicators between different regions. This model can analyze the development status of Beijing and predict the impact of infrastructure changes on the working population. Therefore, this study has the following characteristics: (1) estimating the population density is taken as the research objective, and the current situation of urban development is analyzed more intuitively; (2) considering the spatial heterogeneity of the impact of facilities on the working population, it is possible to quantitatively analyze the impact of infrastructure in different regions on the working population.

2. Study Area and Datasets

This article takes Beijing as the research objective. Beijing has 16 urban districts under its jurisdiction, with a total area of 16,410.54 square kilometers, of which the built-up area is 1485 square kilometers (Figure 1a). According to data from the National Bureau of Statistics of China in 2015, Beijing's permanent population at the end of the year was 21.71 million, of which the urban population was 18.77 million and the rural population was 2.93 million [16]. Beijing consists of a main urban area with the five ring roads as the boundary and several satellite cities. There are more jobs in the central part of the city. On the other hand, most citizens live in satellite cities of Beijing. The main urban area and satellite cities are connected by rail transit and urban expressways.

The citizens of Beijing have a high penetration rate of mobile phones. In the second quarter of 2015, there were 193.7 mobile phones per 100 residents in Beijing [21]. Through the mobile phone data, the life trajectory of Beijing residents can be recorded in a straightforward way. This article uses Beijing mobile phone data from 1 June to 7 June 2015. The data was provided by China Mobile. The company is Beijing's largest telecommunications provider, and its users account for 67% of all mobile phone users. Each record contains an anonymous user ID, location information and timestamp, etc.

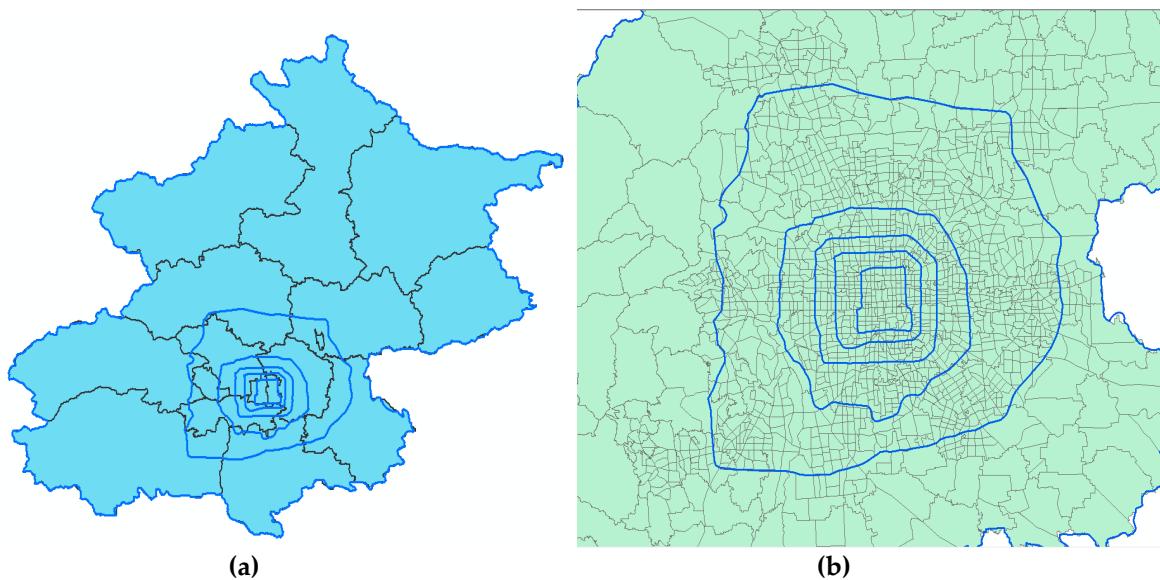


Figure 1. Study area: (a) administrative divisions of Beijing; (b) division of Beijing traffic analysis zone (TAZ).

The traffic analysis zone (TAZ) used in this paper is a division of the TAZs by the Beijing Urban Transport Institute for Beijing transportation planning purposes (Figure 1b). According to different land use characteristics, Beijing is divided into 2006 transportation districts with an average area of 8.16 square kilometers. In this study, 18 million pieces of mobile phone data from anonymous users between 1 June and 5 June 2015, were used to obtain the living and working population of different TAZs.

According to the Athens Charter, the main functions of the city include residence, work, entertainment, and travel [22]. In this study, 12 sets of POI data including these 4 kinds of urban functions were selected from all available POI data sets to comprehensively describe the functions of the cities. The POI data sets were collected by the Google Place API [23,24]. In this study, POI is linked to TAZs and processed as density data to objectively show the current status of infrastructures in different TAZs. The table shows the POI data selected in this article (Table 1).

Table 1. Point-of-interest (POI) data.

Residential Attributes	Work Attributes	Entertainment Attributes	Travel Attributes
school density	mall density	hotel density	bus station density
pharmacy density	office building density	park density	subway station density
hospital density	government agency density		
restaurant density	bank density		

Different types of POI have different distribution characteristics in Beijing (Figure 2). For example, restaurants are most densely distributed in Beijing (Figure 2d). Office buildings are clustered in Beijing. Marked on the map are Zhongguancun Innovation Park in northwestern Beijing, Financial Street in the middle, and CBD in the east (Figure 2f). The distribution of the parks in Beijing is relatively sparse (Figure 2j). Bus stations are denser than subway stations (Figure 2k,l). The subway is the core of Beijing's public transportation system, transporting large numbers of commuters every day.

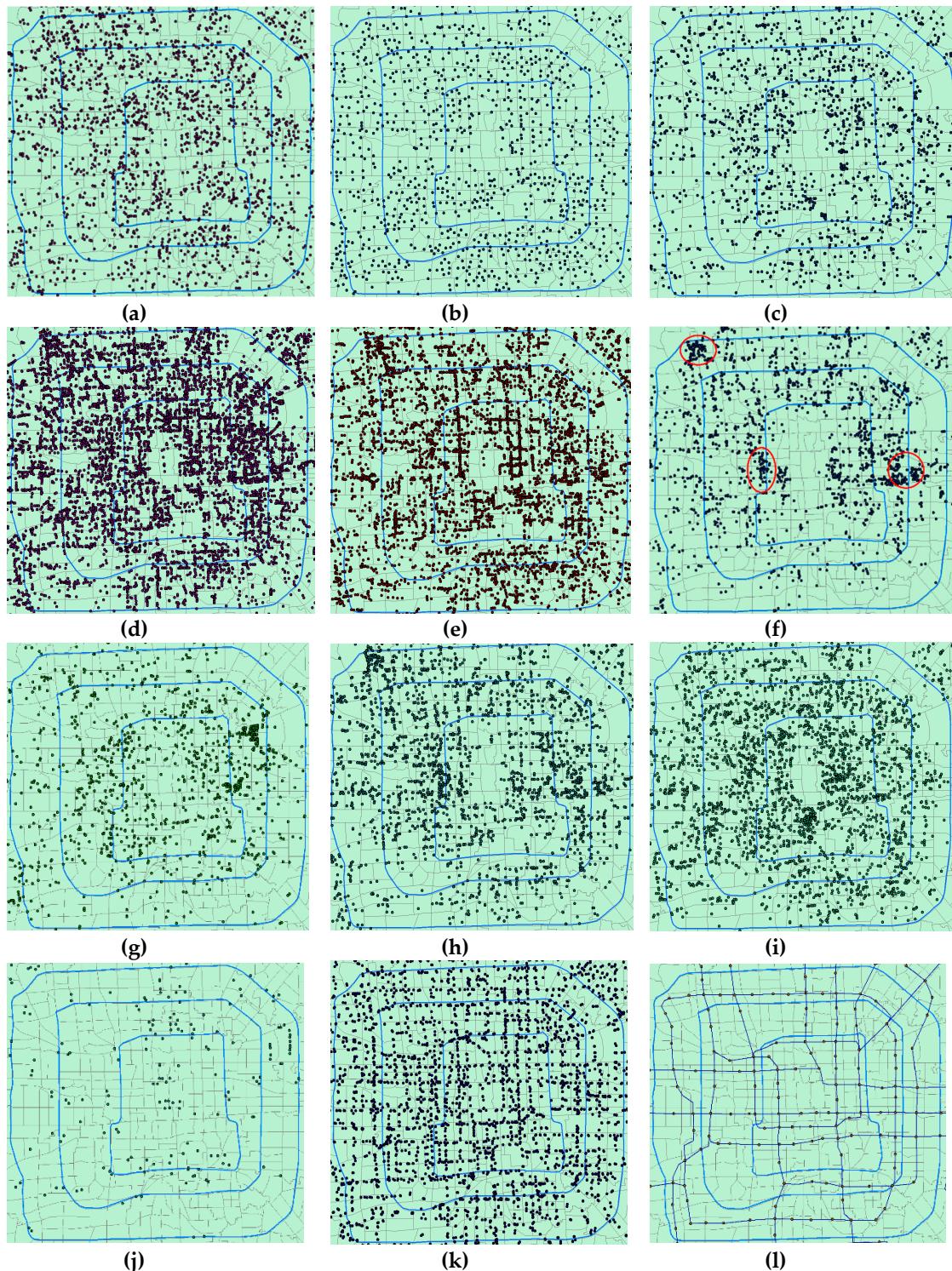


Figure 2. The distribution of POI in the center of Beijing: (a) school; (b) pharmacy; (c) hospital; (d) restaurant; (e) mall; (f) office building; (g) government agency; (h) bank; (i) hotel; (j) park; (k) bus station; (l) subway station and subway line.

3. Methodology

This study adopts the framework of a geographically weighted regression model, explores the correlation between land use characteristics and working population, and constructs a working population estimation model (Figure 3).

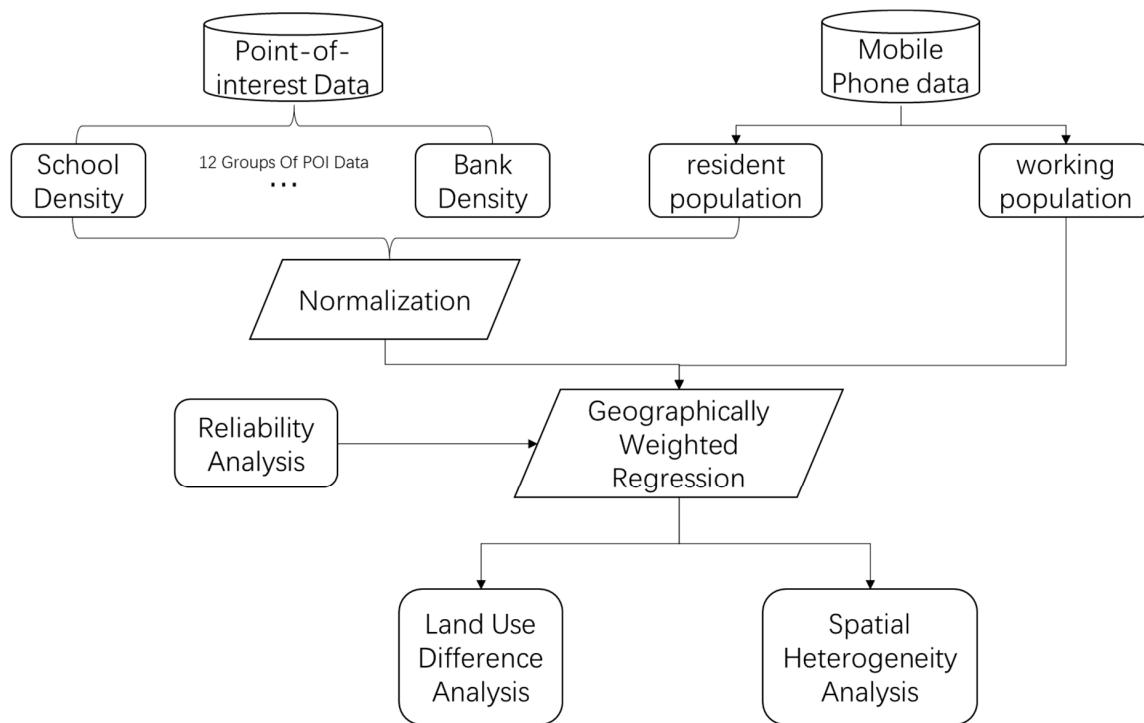


Figure 3. The workflow of this study.

First of all, this study identifies the living and working population of Beijing through mobile phone data. To prevent the inaccurate regression results caused by the magnitude difference, this paper normalizes the 12 sets of POI data and residential population density data. Next, we use the working population density as the dependent variable for geographically weighted regression. To verify the reliability of the regression analysis, this paper also performed OLS regression on this set of data to compare the goodness of fit of the two schemes. Finally, this study describes the regression results through land use difference analysis and spatial heterogeneity analysis.

3.1. Work/Residence Location Identification

In this study, the users' mobile phone data during time spent at their residence and work hours are used to determine their work location and the users' residential area. For this purpose, the fuzzy discrimination method is employed. The result of this method does not include users without a fixed workplace or users who work at night.

First of all, this paper preprocesses the original mobile phone data to remove abnormal items, then according to the distribution of work and residence time, combines it with the characteristics of mobile phone data and dwell time, feature index extraction of user work and residence data characteristics, and construction of membership function. On this basis, a standard feature vector is constructed, and a discrimination rule between the feature vector and the measured data is formed, and finally, the user's employment and residence information are obtained (Figure 4) [25]. Since most residents of Beijing work from 9:00 am to 5:00 pm, the working hours selected in this article are 9:00 to 11:00 am and 2:00 to 4:00 pm, and the residence time is 0:00 to 5:00 am.

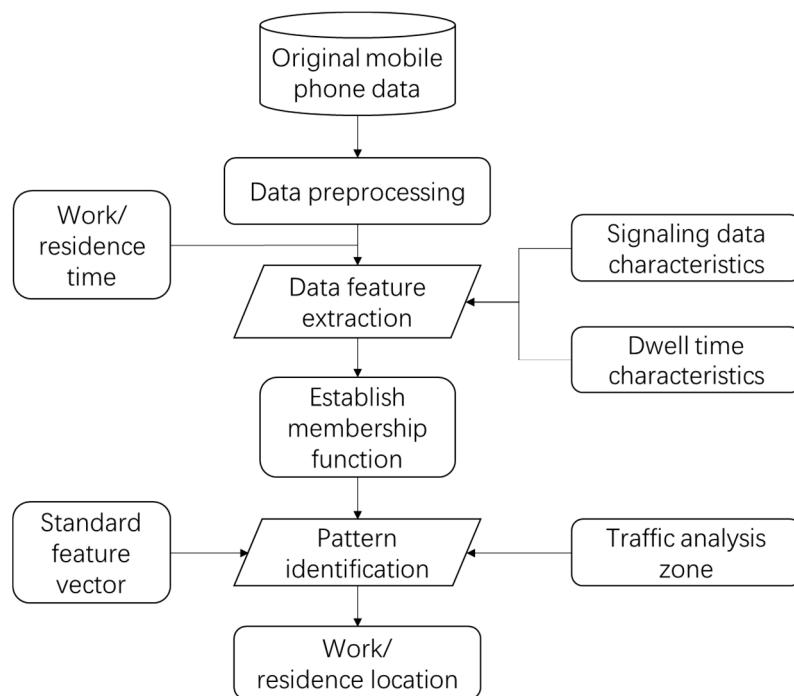


Figure 4. The workflow of work/residence location identification.

3.2. Geographically Weighted Regression

Geographically weighted regression (GWR) is a spatial analysis technique [26]. By establishing a local regression equation at each point in the spatial range, GWR explores the spatial changes and related driving factors of the research object at a certain scale and can be used to predict future results. Because it takes into account the local effects of spatial objects, its advantage is high accuracy. According to the first law of geography proposed by Tobler, everything is spatially related, and the closer the distance between the things, the greater the spatial correlation [27,28]. Therefore, unlike traditional cross-section data, the spatial correlation of spatial data will lead to the spatial heterogeneity of the regression relationship. To explore the spatial heterogeneity of spatial data, Brunsdon et al. first proposed a geographically weighted regression model in 1996, set as follows [29–31]:

$$\text{working population} = \alpha + \text{normalized population density} \times \beta + \sum_{j=1}^{12} \text{POI Density}_j \times \gamma_j + e \quad (1)$$

where α is the baseline working population in the average area, β is the association between population and working/daytime population, and γ_j is the effect of POI density $j = 1, 2, \dots, 12$ on working population.

The distance function selected in this paper is a Gaussian function because the dependent variable in the model is a continuous value with a large working population density. The Gaussian function is defined as follows:

$$\omega_{ij} = \exp\left[-(d_{ij}/b)^2\right] \quad (2)$$

where d_{ij} represents the Euclidean distance between observations i and j , and b represents the bandwidth.

Another important choice in specifying a geographically weighted regression model is the choice of fixed or variable bandwidth. There are two common bandwidth selection schemes: constant bandwidth and variable bandwidth. This paper chooses a variable bandwidth scheme. This is because if the variable bandwidth scheme is selected, the bandwidth for sparsely distributed data will be greater, while the bandwidth for densely distributed data will be reduced. The distribution of POI in the sample is uneven; this flexibility can lead to a more accurate model.

The corrected Akaike information criterion (AICc) method provides a trade-off between the degree of freedom and goodness of fit to optimize the bandwidth [26,32]. Fotheringham defined the AICc equation for GWR as:

$$AICc(b) = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right\} \quad (3)$$

where n represents the local sample size (according to bandwidth); $\hat{\sigma}$ represents the estimated standard deviation of the error term, and $\text{tr}(S)$ represents the trace of the hat matrix S . The hat matrix denotes the projection matrix from the observed y to the fitted values.

The smaller the bandwidth, the smaller the bias of the estimated regression function, but the larger the variance. Therefore, overfitting occurs. But for AICc in many cases, it can better overcome the overfitting phenomenon; that is, it tends to determine a more reasonable window width [33,34].

4. Results and Discussion

4.1. Reliability Analysis

In order to verify the reliability of the GWR model, the same data in this study were subjected to global model (OLS) regression and geographically weighted regression, respectively. In this study, the working population was used as the dependent variable, the normalized population density and the 12 groups of POI density were used as independent variables for regression analysis. By comparing the parameters such as the goodness of fit of the OLS model and the GWR model, a better regression model is obtained. The results of geographically weighted regression and global regression are as follows (Tables 2 and 3):

Table 2. Global regression (OLS) model summary.

AICc	R Square	Adjusted R Square	Std. Error of the Estimate
32,854.167	0.768	0.766	1696.198

Table 3. Geographically weighted regression (GWR) model summary.

Neighbors	Residual Squares	Effective Number	AICc	R Square	Adjusted R Square
407	2,913,897,200	162.364	32,635.376	0.817	0.800

It can be seen from the above table that the R square and adjusted R square in the geographically weighted model are higher than those in the OLS model, so the geographically weighted regression model has a higher degree of goodness of fit.

Residual squares in the summary table of geographic regression models refers to the sum of squared residuals in the model (residuals are the difference between actual dependent variable value and the estimated value of the dependent variable returned by the GWR model). The smaller this measurement, the more the GWR model fits the observed data. Effective number of parameters, which is larger when bandwidths are smaller, is related to the choice of bandwidth. It is a compromise representation of the variance of the fitted values and the deviation of the coefficient estimates. AICc is a measure of model performance and helps to compare different regression models. Considering model complexity, models with lower AICc values will better fit the observed data. AICc is not an absolute measure of fitness but useful for comparing models that are applicable to the same dependent variable and have different explanatory variables. These 4 parameters are used to compare different methods of geographic regression models. After comparison, this paper chooses a variable bandwidth selection scheme.

This paper analyzes the coefficient distribution of the regression results of the least squares model. It can be seen from the bar chart (Figure 5) that the residential population has the most significant impact

on the working population, followed by office building density. However, indicators with residence attribute: school density, hospital density, pharmacy density, and restaurant density have little negative impact on the number of the working population. This is similar to the results of geographic regression analysis. Traditional global regression cannot analyze the differences of coefficients in different regions. Geographically weighted regression can better show and analyze the impact of infrastructure on the working population in different regions.

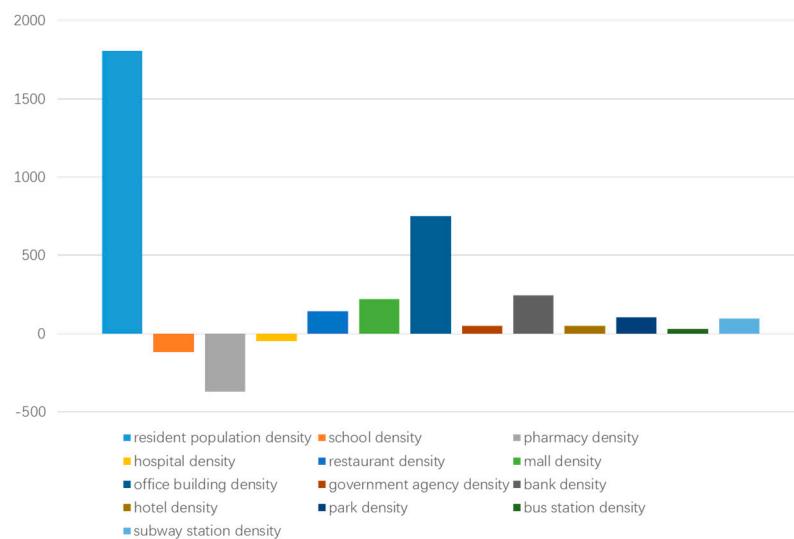


Figure 5. Coefficient distribution of the OLS model.

4.2. Land Use Difference Analysis

The box plot (Figure 6) of the coefficient distribution of the geographically weighted regression results is shown in the figure below. The coefficient distribution chart shows the factors affecting the working population of different land use characteristics. The higher the index coefficient, the more significant the impact on the working population density. If the coefficient is negative, the indicator has a negative correlation with the working population density.

The impact of the resident population density on the working population density is most significant, followed by four indicators with job attributes: mall density, office building density, government agency density, and bank density. The office building density has the highest coefficient and the highest degree of dispersion. This proves that office building density can significantly affect the working population density, and some district office buildings will have a higher degree of influence. The average coefficients of bus station density and subway station density, which are traffic indicators, are slightly higher than 0, but the density of subway stations in some areas is significantly positively related to the working population density. Indicators with entertainment attributes, hotel density and park density, have little effect on the working population. Among the indicators with residential attributes, school density, hospital density, and restaurant density have no significant effect on the working population density, but the pharmaceutical density has a negative correlation with the working population density. This may be due to the little employment created by pharmacies, and most of them are built near residential areas and far away from work areas.

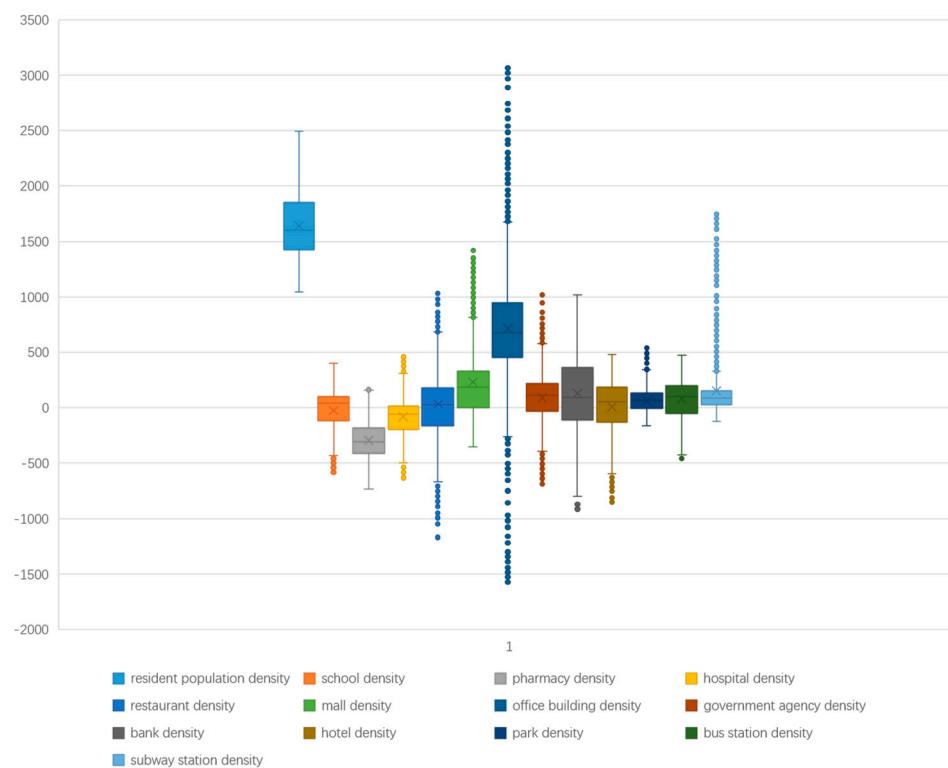


Figure 6. Coefficient distribution of the geographically weighted regression results.

4.3. Spatial Heterogeneity Analysis

One of the characteristics of geographically weighted regression is that it highlights the spatial heterogeneity of the effect of the independent variable on the dependent variable. This article analyzes the correlation of different indicators on the working population density in different regions by analyzing the coefficient differences in different regions. Then, combined with the real situation in Beijing, the reasons for different regional coefficients are analyzed.

From the distribution map of the population density coefficient (Figure 7), the overall urban area coefficient is higher than that of the suburbs, and the two northern areas have the highest coefficients. The two areas are the Haidian District and the Chaoyang District of Beijing. They are the most concentrated urban areas in Beijing with the highest economic attributes. They have the perfect infrastructure and can attract a large number of people in other urban areas to work there. They have higher coefficients because of the denser working population and relatively lower residential population distribution in these two urban areas.

From the perspective of the spatial distribution of the residential index coefficients (Figure 8), the overall main urban area coefficient is lower than the suburban area, and the school density coefficient distribution is the most obvious. The greater relevance of the suburbs may be due to more backward infrastructure. However, the higher density of residential infrastructure in the main urban area means that the area is mainly responsible for residential functions rather than generating more employment. These indicators of the main urban areas and some satellite cities are negatively correlated with the working population density.

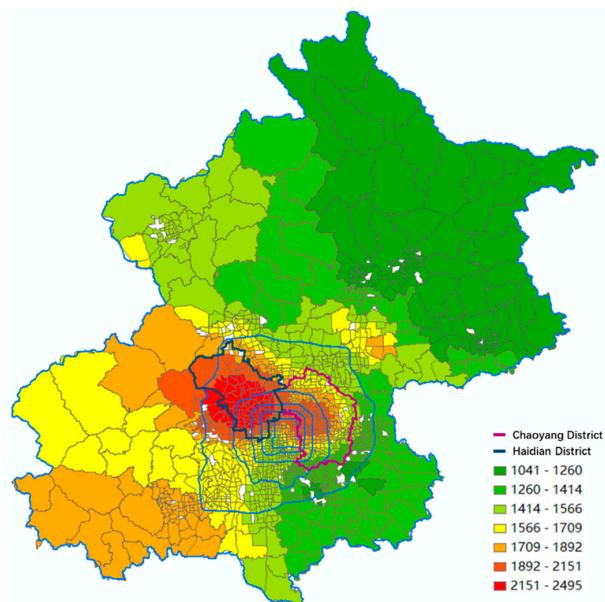


Figure 7. Distribution map of the population density coefficient.

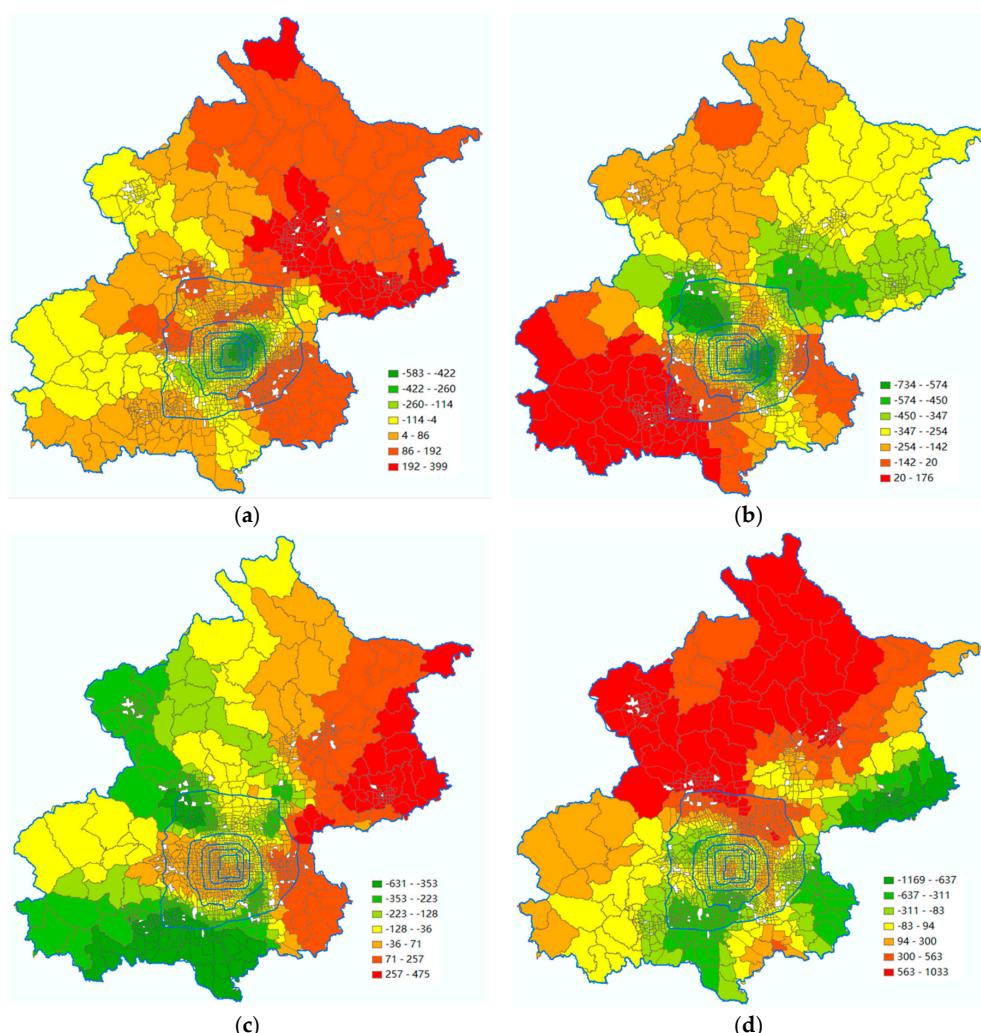


Figure 8. Distribution map of the residential index coefficients: (a) school density; (b) pharmacy density; (c) hospital density; (d) restaurant density.

The spatial distribution of the coefficients with commercial indicators varies widely, depending on the infrastructure they represent (Figure 9). The area with the highest density coefficient in the mall is Haidian District, located northwest of the main urban area, where Beijing has the most technology companies and universities. The active economy and high population have attracted a large demand for shopping. The correlation between office building and working population density is second only to residential population density. In most areas of Beijing, the coefficient of office building density is relatively high, and the coefficient of Tongzhou District, which is known as the sub-center of Beijing City, is the highest. Tongzhou District, located on the east side of Beijing's main urban area, is the area where Beijing has had the greatest potential for development in recent years, carrying many functions of Beijing's main urban area. Government agencies have less influence on the working population density. The regions with relatively higher coefficients are located on the outskirts of the north and south sides of Beijing. The bank density has the highest coefficient in Chaoyang District on the east side of Beijing. Beijing's Central Business District is located in Chaoyang District, which has the most business activities in Beijing.

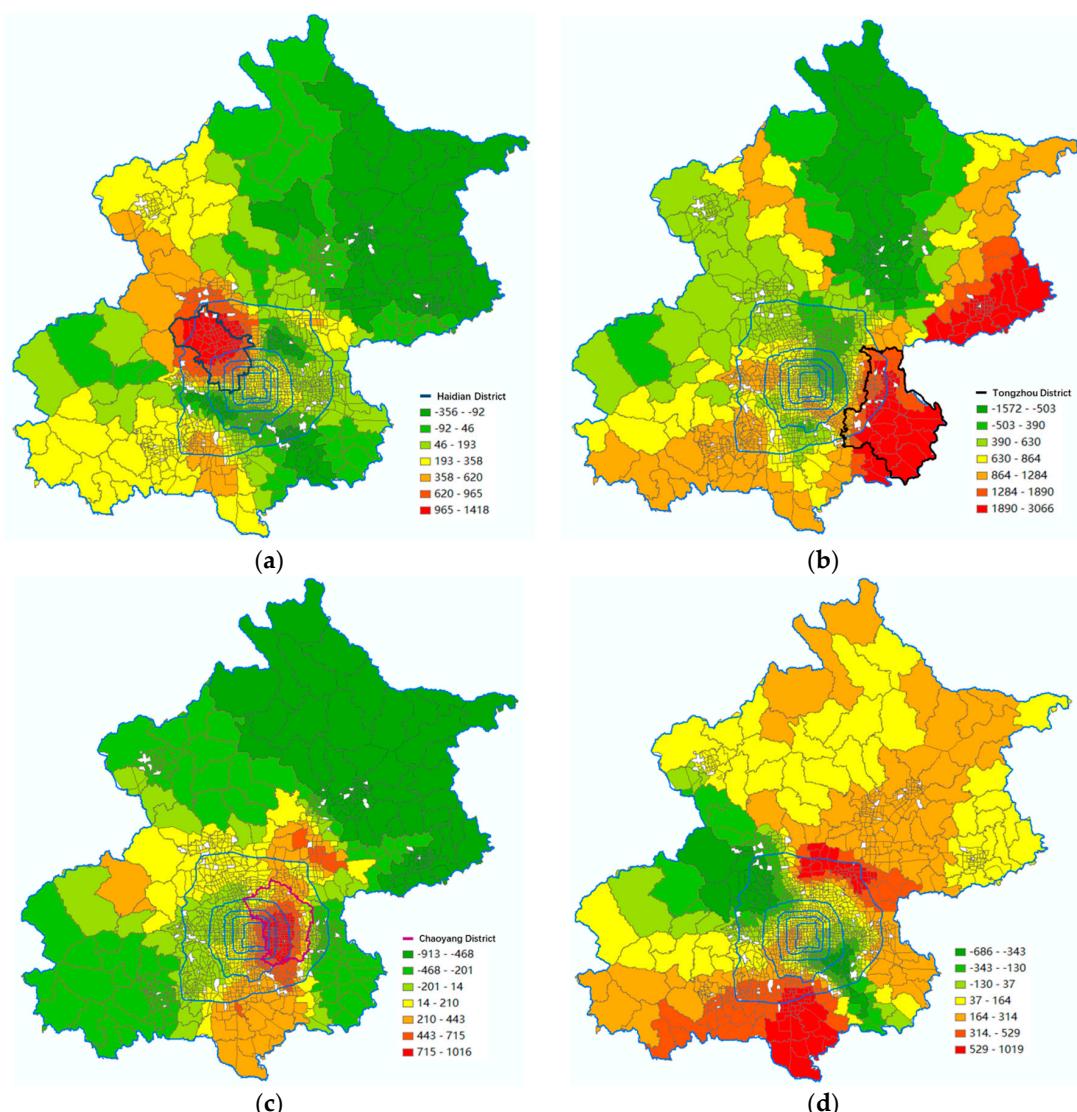


Figure 9. Distribution map of the commercial index coefficients: (a) mall density; (b) office building density; (c) bank density; (d) government agency density.

The impact of recreational indicators on the working population is not significant, but it also has a certain spatial heterogeneity. In terms of spatial distribution, the suburbs of Beijing have a higher coefficient of entertainment indicators (Figure 10).

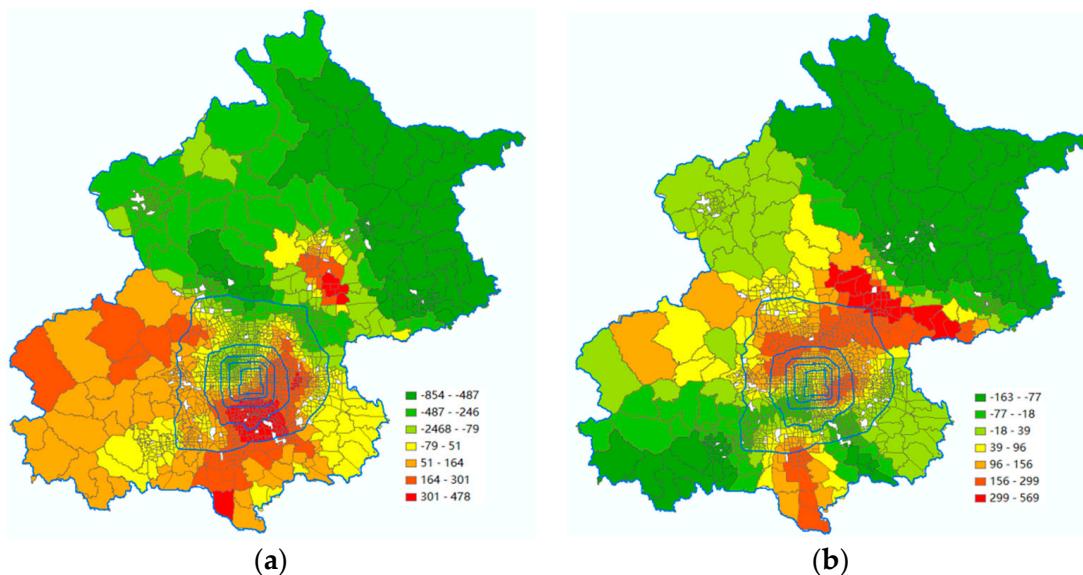


Figure 10. Distribution map of the recreational index coefficients: (a) hotel density; (b) park density.

Because Beijing's main urban areas and suburbs have large differences in infrastructure and population density, this article mainly analyzes the part of the traffic-type indicators in Beijing's main urban areas. The east and south sides of Beijing's main urban area have higher bus station density coefficients. Here are Chaoyang District and Fengtai District. The distribution map of bus stations can be seen in other areas of the city, and areas with higher coefficients have fewer bus stations (Figure 11). Compared with other urban areas, the demand for public transportation in these two urban areas is similar, and even the public transportation demand in Chaoyang District will be higher. According to the data of the smart card named Yikatong, provided by the Transportation Operations Coordination Center, in 24 hours, a total of 1.67 million passengers took the bus in Chaoyang District, while only 1.34 million passengers took the bus in Haidian District. However the bus stations are more sparse in Chaoyang District. Therefore, the reasons for the significant positive correlation between bus station density and working population density in these areas may be due to strong demand and relatively few bus stations.

Similar laws are also reflected in the correlation characteristics of subway station density and working population density. From the perspective of the distribution of subway coefficients, the regional coefficient in the northwest of the city is higher (Figure 12). There are some technology companies here, but the subway has not been opened in 2015. Now Beijing Metro Line 16 has been opened here. In areas with strong demand but relatively scarce supply, the coefficient is often higher.

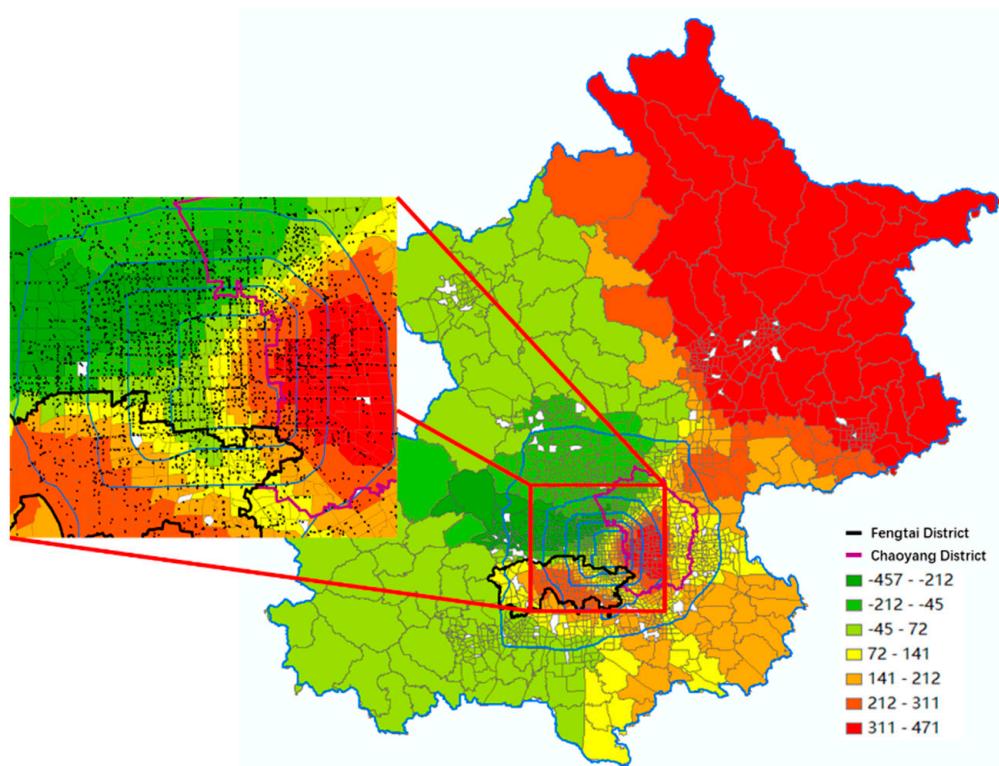


Figure 11. Distribution map of the bus station density coefficients.

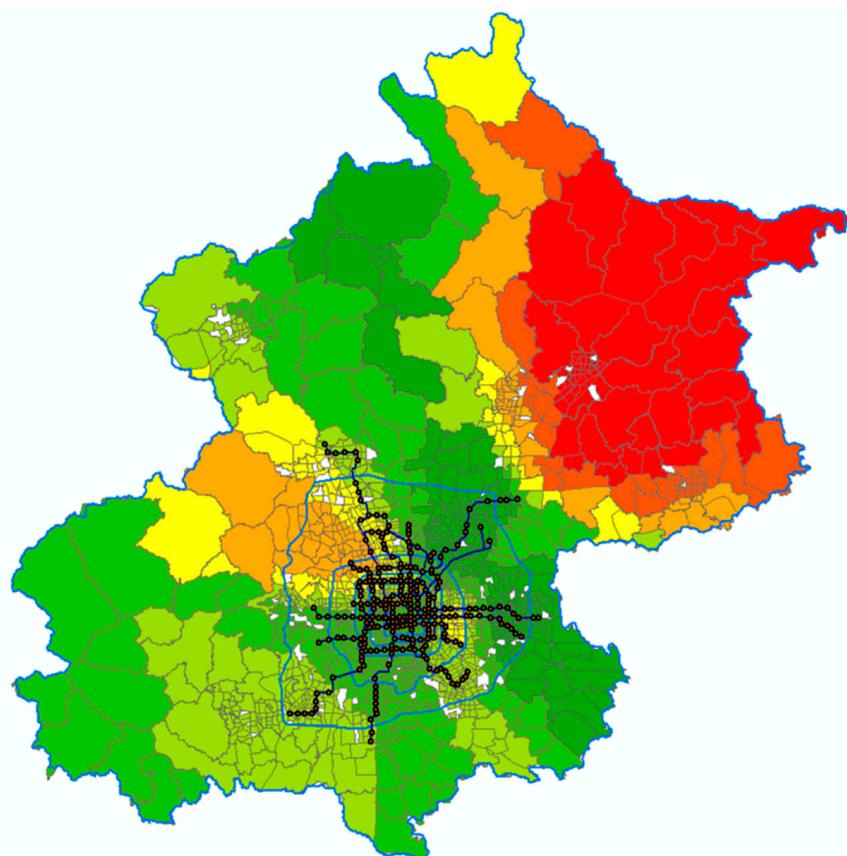


Figure 12. Distribution map of the subway station density coefficients.

5. Conclusions

With the maturity of mobile communication means, we can obtain more information through mobile phone signaling data. In this paper, the densities of working and residential population in Beijing were obtained through mobile phone signaling data. In order to fully describe the working population, we selected 12 groups of POIs according to the description of urban functions in the Athens Charter. Through regression analysis of the working population, the impact and correlation of different infrastructure density and other indicators on the working population in different regions were obtained. This paper proposes a regression model with working population density as the dependent variable, which not only explains the impact of different indicators on the working population, but also provides an estimation model of the working population. This article examines the impact of different infrastructures on working population density. Additionally, because we chose the geographic regression model, which takes into account the first law of geography, this paper studies the spatial heterogeneity of different indicators. By studying how these facilities affect the working population density in different regions, we can explore the relationship between each indicator and the working population with greater accuracy. Urban planners and managers can use this approach to explore the impact of infrastructure adjustments on the working population. Finally, it can promote the scientific development of the city.

The achievements and innovations in this paper include the following.

The working population density obtained using mobile phone data extraction is the object of model research. Compared with the travel intensity data used in other studies, the working population distribution has a more direct role in guiding urban development.

The relationship between working population and land use was analyzed based on the geographic regression model. The geographic regression model takes into account the characteristics of spatial heterogeneity and the impact of various infrastructures with space attenuation, which is more accurate than traditional regression models, and quantitatively, the impact of each index on the working population is expressed by region.

This study has several limitations that deserve further exploration: due to data constraints, the types of POI extracted are insufficient, many other infrastructures will also affect the distribution of the working population, and there is space for improvement in model accuracy. Due to the limitations of the extraction method of the working population, the model accuracy is limited by TAZs. This results in lower accuracy when studying spatial heterogeneity, and the differences between regions are not obvious. In addition, the working population identification method is to divide the user's working location by working hours, which is not accurate. The mobile phone data selected in this article is not a full sample or random sample, and the description of the population may not be accurate. If there is more data, such as data at different times, the model will become more accurate and interesting.

Author Contributions: Methodology, H.Q.; project administration, Y.C.; resources, Y.C.; visualization, H.Q.; writing—original draft preparation, H.Q.; writing—review and editing, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Airport Construction Corporation of Caac (MHRD20130107). This publication was further supported by Beijing Transport Institute (Z181100003918011)

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sumita, M. Communiqué of the national bureau of statistics of People's Republic of China on major figures of the 2010 population census(No.1). *China Pop. Today* **2011**, *2*, 21–25.
2. Zheng, M.; Liu, F.; Guo, X.; Lei, X. Assessing the distribution of commuting trips and jobs-housing balance using smart card data: A case study of Nanjing, China. *Sustainability* **2019**, *11*, 5346. [[CrossRef](#)]
3. Li, S.; Sun, Z.; Tan, M.; Li, X. Effects of rural-urban migration on vegetation greenness in fragile areas: A case study of Inner Mongolia in China. *J. Geogr. Sci.* **2016**, *26*, 313–324. [[CrossRef](#)]

4. Xiao, X.; Hu, M.; Tan, M.; Li, X.; Li, W. Changes in the ecological footprint of rural populations in the Taihang Mountains, China. *Sustainability* **2018**, *10*, 3741. [[CrossRef](#)]
5. Zhang, X.; Lin, X.; Zhu, S. Modeling urban growth by cellular automata: A case study of Xiamen City, China. In Proceedings of the 10th International Conference on Computer Science & Education (ICCSE), Cambridge, UK, 22 June 2015; pp. 645–650.
6. Peng, M.; Chen, D.; Ma, Y.; Meng, C. Comprehensive degree of land use analysis based on BGWR model: A case study of Wuhan. In Proceedings of the 2017 25th International Conference on Geoinformatics, Buffalo, NY, USA, 2–4 August 2017; pp. 1–5.
7. Zeng, C.; Yang, L.; Dong, J. Management of urban land expansion in China through intensity assessment: A big data perspective. *J. Clean. Prod.* **2017**, *153*, 637–647. [[CrossRef](#)]
8. Cai, J.X.; Huang, B.; Song, Y.M. Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sens. Environ.* **2017**, *202*, 210–221. [[CrossRef](#)]
9. Tang, L.; Lin, Y.; Li, S.; Li, S.; Li, J.; Ren, F.; Wu, C. Exploring the influence of urban form on urban vibrancy in shenzhen based on mobile phone data. *Sustainability* **2018**, *10*, 4565. [[CrossRef](#)]
10. Liu, S.; Zhang, L.; Long, L. Urban vitality area identification and pattern analysis from the perspective of time and space fusion. *Sustainability* **2019**, *11*, 4032. [[CrossRef](#)]
11. Montero, L.; Ros-Roca, X.; Herranz, R.; Barceló, J. Fusing mobile phone data with other data sources to generate input OD matrices for transport models. *Transp. Res. Procedia* **2019**. [[CrossRef](#)]
12. Chen, Y.; Zhang, Z.; Liang, T. Assessing urban travel patterns: An analysis of traffic analysis zone-based mobility patterns. *Sustainability* **2019**, *11*, 5452. [[CrossRef](#)]
13. Liu, L.; Peng, Z.; Wu, H.; Jiao, H.; Yu, Y. Exploring urban spatial feature with dasymetric mapping based on mobile phone data and LUR-2SFCAe method. *Sustainability* **2018**, *10*, 2432. [[CrossRef](#)]
14. Liu, L.; Peng, Z.; Wu, H.; Jiao, H.; Yu, Y.; Zhao, J. Fast identification of urban sprawl based on K-Means clustering with population density and local spatial entropy. *Sustainability* **2018**, *10*, 2683. [[CrossRef](#)]
15. Zhou, X.; Chen, X.; Zhang, T. Impact of megacity jobs-housing spatial mismatch on commuting behaviors: A case study on central districts of Shanghai, China. *Sustainability* **2016**, *8*, 122. [[CrossRef](#)]
16. National Bureau of Statistics of China. Available online: <http://data.stats.gov.cn> (accessed on 10 March 2020).
17. Lucas, K.; Philips, I.; Mulley, C.; Ma, L. Is transport poverty socially or environmentally driven? Comparing the travel behaviours of two low-income populations living in central and peripheral locations in the same city. *Transp. Res. Pt. A-Policy Pract.* **2018**, *116*, 622–634. [[CrossRef](#)]
18. Guo, Y.Q.; Chang, S.S.; Sha, F.; Yip, P.S.F. Poverty concentration in an affluent city: Geographic variation and correlates of neighborhood poverty rates in Hong Kong. *PLoS ONE* **2018**, *13*. [[CrossRef](#)]
19. Li, H.Q.; Cheng, S.M.; Haynes, K.E. The employment effects of new business formation: A regional perspective. *Econ. Dev. Q.* **2011**. [[CrossRef](#)]
20. Rodrigues, F.; Alves, A.; Polisciuc, E.; Jiang, S.; Ferreira, J.; Pereira, F. Estimating disaggregated employment size from points-of-interest and census data: From mining the web to model implementation and visualization. *Int. J. Adv. Intell. Syst.* **2013**, *6*, 41–52.
21. Ministry of Industry and Information Techology of the People's Republic of China. Available online: www.miit.gov.cn (accessed on 10 March 2020).
22. Corbusier, L.; Eardley, A. *The Athens Charter*; Grossman Publishers: New York, NY, USA, 1973.
23. GoogleMap. Available online: <https://developers.google.cn/maps/documentation> (accessed on 4 March 2020).
24. Yang, S.; Shen, J.; Konečný, M.; Štampach, R. Study on the spatial heterogeneity of the POI quality in OpenStreetMap. In Proceedings of the 7th International Conference on Cartography and GIS, Sozopol, Bulgaria, 18–23 June 2018.
25. Lai, J.; Chen, Y.; Zhang, W.; Chen, Z. Identification method of residence and employment locations based on cellular phone data. *J. Netw.* **2014**, *9*, 8. [[CrossRef](#)]
26. Fotheringham, A.S.; Charlton, M.; Brunsdon, C. *Measuring Spatial Variations in Relationships with Geographically Weighted Regression*; Springer: Berlin/Heidelberg, Germany, 1997.
27. Chow, L.F.; Zhao, F.; Liu, X.M.; Li, M.T.; Ubaka, I.; Natl Acad, T.R.B. Transit ridership model based on geographically weighted regression. *Travel Survey Methods Inf. Technol. Geospat. Data* **2006**, *1972*, 105–114. [[CrossRef](#)]
28. Xu, C.C.; Zhao, J.Y.; Liu, P. A geographically weighted regression approach to investigate the effects of traffic conditions and road characteristics on air pollutant emissions. *J. Clean. Prod.* **2019**, *239*. [[CrossRef](#)]

29. Tang, J.J.; Gao, F.; Liu, F.; Zhang, W.H.; Qi, Y. Understanding spatio-temporal characteristics of urban travel demand based on the combination of GWR and GLM. *Sustainability* **2019**, *11*, 5525. [[CrossRef](#)]
30. Zhang, X.X.; Huang, B.; Zhu, S.Z. Spatiotemporal influence of urban environment on taxi ridership using geographically and temporally weighted regression. *ISPRS Int. Geo-Inf.* **2019**, *8*, 23. [[CrossRef](#)]
31. Bao, J.; Shi, X.M.; Zhang, H. Spatial analysis of bikeshare ridership with smart card and POI data using geographically weighted regression method. *IEEE Access* **2018**, *6*, 76049–76059. [[CrossRef](#)]
32. Yang, F.; Ding, F.; Qu, X.; Ran, B. Estimating urban shared-bike trips with location-based social networking data. *Sustainability* **2019**, *11*, 3220. [[CrossRef](#)]
33. Ma, Y.; Sucharita, G. Geographically weighted regression models in estimating median home prices in towns of Massachusetts based on an urban sustainability framework. *Sustainability* **2018**, *10*, 1026. [[CrossRef](#)]
34. Du, H.B.; Mulley, C. Relationship between transport accessibility and land value - Local model approach with geographically weighted regression. *Transp. Res. Rec. Ser.* **2006**, *1977*, 197–205. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).