Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 7621576, 20 pages
https://doi.org/10.1155/2020/7621576

Research Article

Spatial Variation of Taxi Demand Using GPS Trajectories and POI Data

Xinmin Liu,[1,2] Lu Sun ,[1] Qiuxia Sun ,[3] and Ge Gao [1,4]

[1] *College of Economics and Management, Shandong University of Science and Technology, Qingdao, China*
[2] *Qingdao Agricultural University, Qingdao, China*
[3] *College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, China*
[4] *College of Transportation, Shandong University of Science and Technology, Qingdao, China*

Correspondence should be addressed to Lu Sun; sunlu@sdust.edu.cn

Received 3 June 2019; Revised 21 August 2019; Accepted 3 September 2019; Published 13 January 2020

Academic Editor: Gonçalo Homem de Almeida Correia

Taxi as a door-to-door, all-weather way of travel is an important part of the urban transportation system. A fundamental understanding of temporal-spatial variation and its related influential factors are essential for taxi regulation and urban planning. In this paper, we explore the correlation between taxi demand and socio-economic, transport system and land use patterns based on taxi GPS trajectory and POI (point of interest) data of Qingdao City. The geographically weighted regression (GWR) model is used to describe the influence factors of spatial heterogeneity of the taxi demand and visualize the spatial distributions of parameter estimations. Results indicate that during the peak hours, there are some differences in taxi demand between workdays and weekends. Residential density and housing prices increase the number of taxi trips. Road density, parking lot density and bus station density are positively associated with the taxi demand. It is also found that the higher of the proportion of commercial area and public service area, the greater of the taxi demand, while the proportion of residential area and the land use mix have a negative impact on taxi demand. This paper provides some references for understanding the internal urban environmental factors generating from the taxi travel demand, and provides insights for reducing the taxi vacancy rate, forecasting taxi temporal-spatial demand and urban public transportation system planning.

1. Introduction

Statistics show that by the end of 2017, there were 584,400 public transport vehicles (tram, rail transit) in China and the number of taxis was 1,395,800, which means that taxi travel accounted for a large proportion of public transport. However, the spatial distribution of residents is imbalanced. Therefore, if drivers do not understand the needs of residents, oversupply will lead to unnecessary vacant travel and may cause some urban problems, such as large amounts of carbon emissions and traffic congestion [1]. Conversely, insufficient supply can lead to long waits for taxi passengers [2]. Both cases seriously downgrade the level of service for the taxi industry, such like causing inconvenience to residents' travel, weakening the service level of taxis, and increasing the cost of daily scheduling. All these bad effects result in the idleness and waste of urban public resources. Therefore, it is necessary to dig deep into the needs of residents' taxi ridership. However, exploring the demand of urban taxi ridership is an important and complex task. Firstly, considering that urban taxi travel is driven by various purposes, and urban functions vary from place to place, it is difficult to clearly articulate the determinants associated with taxi passenger travel [3, 4]. Secondly, unlike other buses and light rails modes, taxis do not have fixed routes and lines, so it is difficult to study taxi travel demand. Under such circumstances, reducing the vacant rate and waiting time, making rational use of public resources and exploring the main factors affecting residents' travel demand have become one of the most important problems we need to solve urgently. In order to solve these issues, studies begin to focus on taxi travel forecast, temporal and spatial pattern of taxi travel. Some researches also focus on the model development and optimization on taxi scheduling system, and the influence mechanism of taxis after the advent of taxi-haling applications [5–7]. Many studies concentrated on the mismatch between the potential passengers and empty taxi in spatial, then put forward the route choice model [8–10]. In addition, some scholars combine the distribution characteristics of travel demand with the urban spatial structure to consider the factors that affect residents' taxi travel [6, 11, 12]. The concept of built environment was proposed by scholars to discuss the influencing factors of residents' travel [13–15]. Although the urban built environment factors are complex and diverse, they can be summarized into three aspects: Density, Diversity, and Design [16], such as population density, employment density, land use mix and so on. The above factors are always considered negative or positive impact on travel demand [13–17]. However, most of the research objects are mainly on private cars, public bicycles and rail transit, etc., to explore the impact of built environmental factors such as population density and employment density on travel demand [18–20]. Moreover, due to limited dataset and research methods, studies do not show spatial distribution differences. For this reason, scholars tend to ignore the role of taxis in public transportation and the impact of socio-economic factors on travel demand. In recent years, scholars have begun to pay attention to the status of taxis in residents' travel, and have studied the influencing factors of taxi trips [4, 21, 22]. In this paper, "Density" factors such as road density and residential density are used to analyze the status of taxis in residents' travel.

In general, the research methods of the balance of supply and demand for using taxi are dominated by two methods, i.e., the four-step method [23] and the ordinary least squares (OLS) multiple regression model [24]. Compared with the four-step method, the regression model is relatively fast and cheaper, and is more suitable for more detailed analysis. But the assumption of the model is that all variables are static throughout the study area. For example, Yang et al. [22] used the OLS global regression model to study the related factors affecting taxi passenger travel, and found land use have a strong correlation with taxi demand. However, the above two methods ignore the spatial heterogeneity (e.g. land use types). Actually, the spatial heterogeneity has an important on residents' travel. This is because trips arise from the spatial separation of people's desired purposes, and the purposes are determined by spatial heterogeneity (the different urban functional areas). Therefore, some scholars proposed to use geographical weighted regression (GWR) model to overcome this problem [25]. This method considers the spatial nonfixed effect of independent variables due to different research areas, and it introduces the positional features of each sample point in space and uses the distance between these sample points as an important factor in defining regression weights [12, 26].

With the development of Location-Based Services (LBS), some map service providers (e.g. Google map, Baidu map, Tencent map, etc.) have gradually opened up map service application interfaces. It reduces the cost of acquisition of map data. Point of Interest (POI) data can be understood as point of information or point of interest. It is a kind of point data representing real geographical entities, which mainly contains geographical coordinates (latitude and longitude), names, categories and addresses. Such network open data can be used as the main data source for studying the functional structure of urban space [27, 28]. GPS (Global Positioning System) data has a high geographical resolution and can obtain city-wide travel trajectories. Compared with traditional survey data, it thus promoting the investigation of urban taxi passengers. Taxi GPS trajectory data is used as a type of geospatial activity record, which contains the information reflecting the trajectory, such as the time, pick-up and drop-off location of the passenger during the trip. The data has been used in many studies, such as spatial-temporal distribution of pick-up and drop-off, hotspots identification and predicting waiting time [29–31]. However, few researches focus on urban residents' travel behavior using taxi GPS trajectory data. Actually, taxi is a vital travel mode for urban residents, and it plays an important role in reflecting residents' travel. This paper analyzes the spatial-temporal variations of taxi travel demand, and considers the difference in travel demand between workdays and weekends during peak hours by using the GPS trajectory data and POI data. In order to better describe the influence factors of spatial heterogeneity on the taxi demand, the geographically weighted regression model with socio-economic attributes are used in this paper.

In summary, although existing researches on the travel demand of public transportation have been studied, there are certain restrictions on the empirical research subjects and research methods of the influencing factors. However, data used in previous research was survey data, which missed location information. Meanwhile, traditional analysis model ignores the instability of the demand spatial distribution. Considering the different travel time and purpose of residents, and the distribution of workplace and place of residence are different in spatial. Temporal-spatial variation of taxi demand distribution should be taken into account in this paper. In view of the above, this paper aims to fill this gap by having an empirical analysis on the taxi demand based on taxi trajectory and POI data. The rest of this paper is organized as follows. Section 2 describes the data and methodology, including data preprocessing, the detail of dependent variable and independent variables, kernel density estimation and geographically weight regression model. Section 3 presents the temporal-spatial distribution of travel demand. Section 4 visualized the influencing factors estimated coefficients. Finally, Section 5 we summarize findings and propose a future research agenda.

2. Data and Methodology

2.1. Data. This paper takes the taxi GPS data of Qingdao City from August 2nd to August 8th, 2017 as research object (no special events and national statutory holidays). The devices record the positions of the taxis, every 30 s during the day, generating approximately 140 million records belongs to 8,700 taxis, which covering more than 80% of the whole city. The format of the data structure like <ID, license plate number, longitude and latitude, status, speed, direction, writing time >. There are two status: 0 and 1, 1 for loaded and 0 for empty (no-load). Taxi's trajectory is composed of several GPS points, the data are arranged in chronological order to form the vehicle's trajectory and reflect the information of pick-up and drop-off location. The identification of the pick-up and drop-off points is based on the continuous change of different status, and consists of a continuous, directional linear point distribution

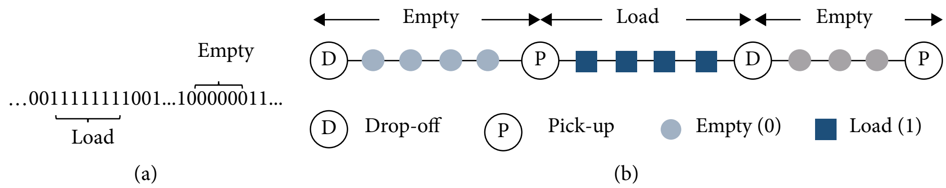


Figure 1: Taxi GPS trajectory (a) sequence of GPS trajectory (b) identification of pick-up and drop-off location.

of 0 and 1. The status form a sequence of 0 change to 1, 1 is the pick-up point, inverse, a sequence of 1 change to 0, 0 is the drop-off point (Figure 1). In the course of taxi driving, GPS equipment failure and building occlusion will occur, resulting in data loss or redundancy. It is necessary to detect and delete the wrong records to keep the good quality of the data and ensure the accuracy of the results. Three types of data should be removed: (i) records with longitude or latitude as zero; (ii) data outside the scope of study area. The main research area of this study is Qingdao urban area (Shinan District, Shibei District, Licang District, Laoshan District, Huangdao District, Chengyang District, shown in Figure 2); (iii) records with speed higher than 100 km/h. On average, 250,000 records are obtained per day.

2.2. Methodology. In order to reflect the spatial difference of taxi travel demand, the entire urban area is divided into regions using a grid-based method. Scholars mainly use 1000∗1000 m or 500∗500 m cells in the literature. In this paper, we select 500∗500 m grid cells, which is more accuracy in spatial structure study. Secondly, according to the file "DB 3702/FW JT 010-2014" by Qingdao Municipal Transport Bureau, the distance between two neighboring stations about 500 m. It is the minimum distance of the accessibility of public service facilities. In this paper, it generated a total of 500∗500 m grid cells in ArcGIS 10.2 system. All pick-up locations of taxi and independent variables are aggregated into corresponding grid cells, the final shapefile of the study area contains more than 1300 cells. The spatial and temporal distribution characteristics of taxi travel demand are analyzed. Then hotspot areas are identified by kernel density estimation, and construct GWR model to study the relationship between taxi demand and different variables at different time.

2.2.1. Variable Description. This segment includes dependent variables and independent variables. Some research regarded socio-economic, traffic and so on as the influential factors on travel demand (independent variables) [20, 32–34]. In this paper, 10 explanatory variables are selected from 3 categories which contains socio-economic, land use and traffic factors as independent variables, and the dependent variable is the number of pick-up points of taxi. The list of the 10 independent variables is given in Table 1.

For socio-economic factors, mainly represent the impact of residents' characteristics on travel demand [19, 26, 35, 36]. For example, residential density is considered to be the most significant factor affecting travel demand, housing price can be a representative indicator of residents' economic level, and consumption level is relatively high in areas with high housing prices [37]. The data of residential density can be obtained by calling Tencent's location big data interface (https://www.geoq.cn/), and housing prices can be obtained by grabbing homepages such as Anjuke and Lianjia (https://qd.anjuke.com/?pi=PZ-360-pc-all-biaoti; https://qd.lianjia.com/?utm_source=360&utm_medium=pinzhuan&utm_term=-biaoti&utm_content=biaoti&utm_campaign=biaoti).

Traffic factors consider the field of transportation, mainly referring to the state of transportation network configuration [38, 39]. Road density, bus stop density and parking lot density are selected as three sub-variables in traffic network [34]. There are 7,106 bus stations in study area, then calculate density in each grid cell. Road network data comes from OpenStreetMap (https://www.openstreetmap.org/#map=10/35.8562/119.9185), and saved as a shapefile. The formula for calculating road density rd_i is calculated as

$$rd_i = \frac{\sum_j L_{ij}}{A_i}, \tag{1}$$

where L_{ij} is the length of road j in cell i, and A_i denotes the area of i.

Land use is also an important factor affecting residents' travel demand. Generally speaking, different land use types attract different human activities, resulting in completely different passenger patterns [40, 41]. This paper mainly considers four main types of land use, namely residential, commercial, public service and other land use. In addition, the mixing of land use types also affects residents' travel activities. Land use data, parking lot and bus station density can be obtained mainly by POI data, they can be collected by using Amap API port (http://lbs.amap.com/api/webservice/guide/api/search). We utilize the entropy of land use, which is calculated as Equation (2), to represent the land use mix [42–44]. In this formula, p_j is the proportion of one land use type (residential, commercial, public service, others) in grid cell i, and k is the number of land use types. The value of E_i ranges from 0 to 1, 0 represent that the land is exclusively dedicated to no-mixed only one type, while 1 represents well mixed.

$$E_i = -\frac{\sum_j p_j ln(p_j)}{ln(k)}. \tag{2}$$

2.2.2. Kernel Density Estimation Method. Kernel density is a nonparametric method for estimating the probability density function. In this paper, we use this method to find hotspots of travel demand in urban areas. It can visually express the change of the density distribution of the passenger's pick-up locations. The search radius is used to delineate the proximity thresholds between things. A specific spatial attenuation function is selected to describe the local spatial association of an event

Figure 2: The study area of Qingdao.

Table 1: Candidate list of explanatory variables.

	Variables	Description
Socio-economic	Residential density	The number of people per square kilometer in each grid cell
	Housing price	The average housing price per square kilometer in each grid cell
	Road density	The length of road segments per square kilometer in each grid cell
Traffic	Bus station density	The number of bus stops per square kilometer in each grid cell
	Parking lot density	The number of parking lots per square kilometer in each grid cell
	Residential area	The ratio of the total floor area used for residential purposes in each grid cell
	Commercial area	The ratio of the total floor area used for commercial purposes in each grid cell
Land use	Public service area	The ratio of the total floor area used for public service purposes in each grid cell
	Other land use	The ratio of the total floor area used for other purpose in each grid cell
	Land use mix	The degree of mixture of land use in each grid cell

with each event within the coverage of the search radius, indicating the relationship between the spatial closeness of the object and the proximity distance. The calculation process of the kernel density estimation first centers on the point feature, creates a circular surface with a given search radius. Then counts the number of point features within the circular surface and divides by the area of the circle. The density value at center is the largest, and the further away from the center point, the smaller the density value until the density value at the radius is zero [45, 46]. The format of $f(x)$ is as follows:

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{3}$$

where $x_1, \ldots, x_n$ is a sequence of independent and identically distributed random variables with F, let the probability density function be f, $K((x - x_i)/h)$ is the kernel function, n means the number of samples, $h = 500\,meters > 0$ is smoothing parameter, also called bandwidth. The larger the kernel density value, the denser the point, and the higher the probability of occurrence of a regional event.

2.2.3. Geographically Weighted Regression Method. The classical linear regression model is based on the least square method (OLS) to estimate the parameters. According to the first law of Tobler geography, all attribute values on a geographic surface are related to each other, but closer values are more strongly related than are more distant ones [47]. The distribution of taxi travel demand is spatially related to urban spatial pattern, and its influencing factors also have spatial geographic attributes. The assumption that the residual term is independent in OLS model will not be satisfied. But GWR model allows some unstable data to be simulated directly, and replaces global parameter estimation with local parameter estimation. It is more conducive to exploring the spatial variation characteristics and spatial rules of taxi travel demand [24]. It expressed as follows:

$$y_i = \beta_0(\mu_i, \theta_i) + \sum_{j=1}^{k} \beta_j(\mu_i, \theta_i) x_{ij} + \varepsilon_i, \quad (4)$$

where β_j is estimated coefficient, (μ_i, θ_i) represent the location of i, x_{ij} is independent variable, ε_i is a residual. Some methods exist that used to model the location of the estimated coefficient, such as fixed distance threshold, the Gaussian kernel function and bi-square function. The coefficients are calibrated for each grid cell and vary from locations. In our case, we use Gaussian kernel function to estimate the spatial effects:

$$\omega_{ij} = \exp\left[-\left(\frac{c_{ij}}{b}\right)^2\right] \quad (5)$$

c_{ij} represent the distance between observation j and location i, which is calculated as the orthodromic distance in this study. When one of the samples is found, the weight of the other sample point decreases as c_{ij} increases, and b is bandwidth, which refers to the nonnegative attenuation parameter of the functional relationship between weight and distance. The larger the bandwidth, the slower the weight decays with the increase of the distance.

3. Results and Analysis

3.1. Results of Temporal Distribution of Taxis. According to the status attribute of GPS data, the temporal distribution of the passengers at different time granularities is counted at intervals of 1 hour, record the change of the number of taxi passengers in one day, and comparing the changes in the number of trips on workdays and weekends. There are 8,600 taxis in Qingdao from August 2nd to August 8th, 2017 were used to record the passengers' records, and the regular pattern of the taxi passengers in different days were obtained (Shown in Figures 3 and 4).

It can be seen from Figure 3 that the distribution of passengers at each time of the week is closely related to the travel rules of residents, and consistent with the regularity of the passenger volume at each time. The number of passengers in the day varies from 1 to 2 hours, and the trips during the day is at a high level. There are certain fluctuations in different time periods. At night, as the flow of people on the street decreases, number of passengers taking taxi also decreases. In terms of time trends, the overall appearance is two peaks in the morning and evening. From 8:00 to 11:00 in the morning, as the number of residents moving to business work increases, then rises sharply to reach the first peak of the day which is the morning peak (09:00-10:00). From 16:00 to 18:00 in the evening, due to the number of vehicles on the road during the peak hours of the work, traffic congestion caused the taxi travel time become longer and the taxi trips changed significantly, showing a "*V*" trend. After 21:00, as the road gradually unblocked, the number of passengers on the taxi began to increase and reached the second peak period (21:00-22:00). After 23:00, there are fewer residents on the street, and the demand for travel has gradually decreased. Passengers have gradually declined with the night, reaching the lowest point of the day at 04:00, in line with people's daily travel rules. The maximum trips during the week is Friday, followed by the weekend.

As shown in Figure 4, there some differences between working days and weekends in terms of overall taxi demand. The overall distribution of taxi demand on weekdays and weekends is similar, but there are significant differences in local variation patterns during the afternoon (dotted line circle). There are two periods (0:00-4:00 and 14:00-10:00) the demand for weekends and much greater than weekdays, while the morning (5:00-11:00) is less than the working day. The result reflects the travel habits of many people nowadays, most people would like to have a rest on weekends morning after working days, so they delay the time outside. At the same time, people are more likely to join in leisure activities, such like social, going to the cinema and other entertainments. These activities always last until midnight. As the public transportation have been shut down, they will choose taxi back to residence places, and the taxi demand will increase.

3.2. Results of Spatial Distribution of Taxis. From the temporal distribution results, we found that the urban residents in Qingdao have some differences in time, and the morning and evening peaks are 9:00-10:00 and 21:00-22:00 respectively. From the perspective of traffic demand, the distribution of pick-up points in the peak period is visually analyzed using kernel density estimation. It can be seen that there are differences in hotspots during peak hours on weekdays and weekends.

It can be seen from Figures 5 and 6 that the location of the hotspots in the morning and evening peaks is similar, showing a trend of "middle height, low circumference", and higher kernel density values in the Shinan District and Shibei District of the city. Licang District and Huangdao District have a lower density. For the reasons why the difference in the spatial distribution of taxi demand, we need further analysis.

For the weekday morning peak and evening peak, evening peak is more significant, and the hotspots of the early peak is smaller than the range of the secondary hotspots. During the morning rush hour, the main hotspots on the workdays are similar, but the scope is different (Figure 5(a)). As can be seen from the Figure 5, the main hotspots areas are mainly distributed near the Qingdao Railway Station in the Shinan District. During the early peak period, the flow of people is relatively

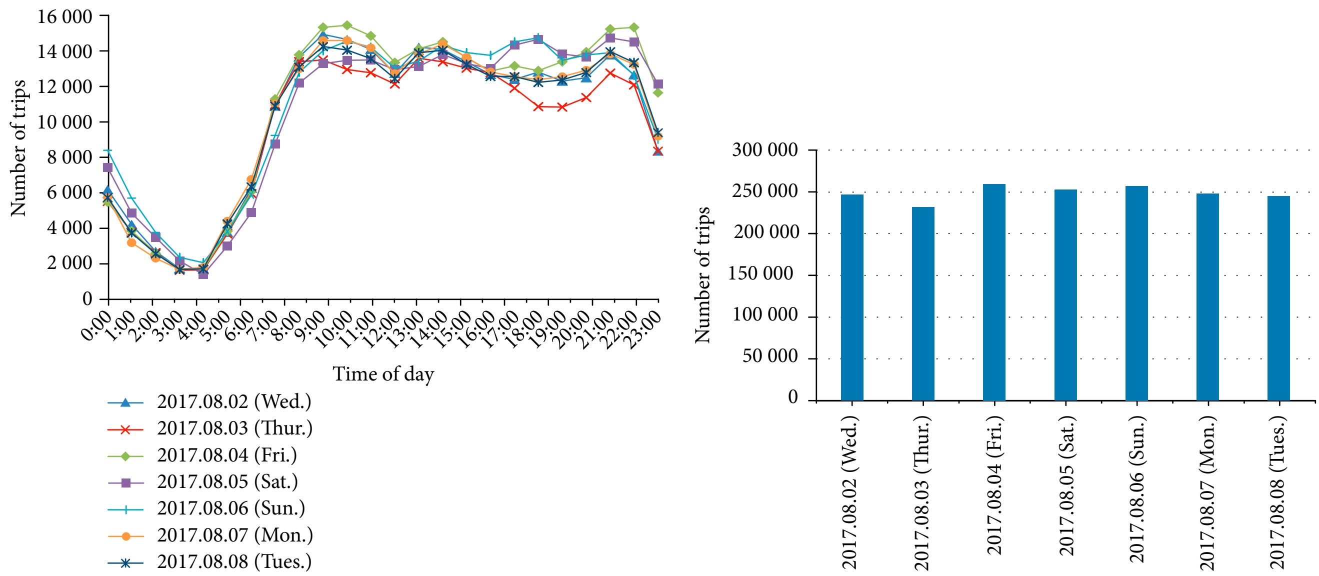


Figure 3: The distribution of number of trips for taxi in one week.

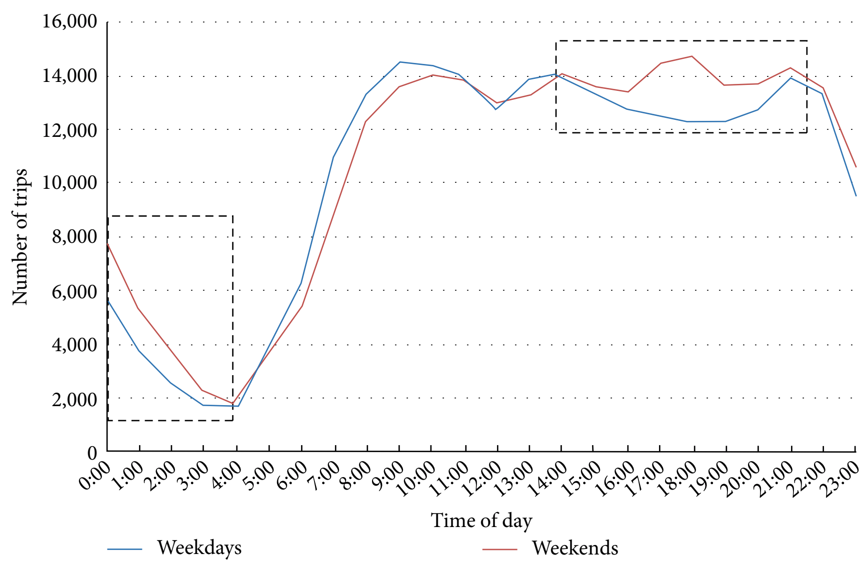
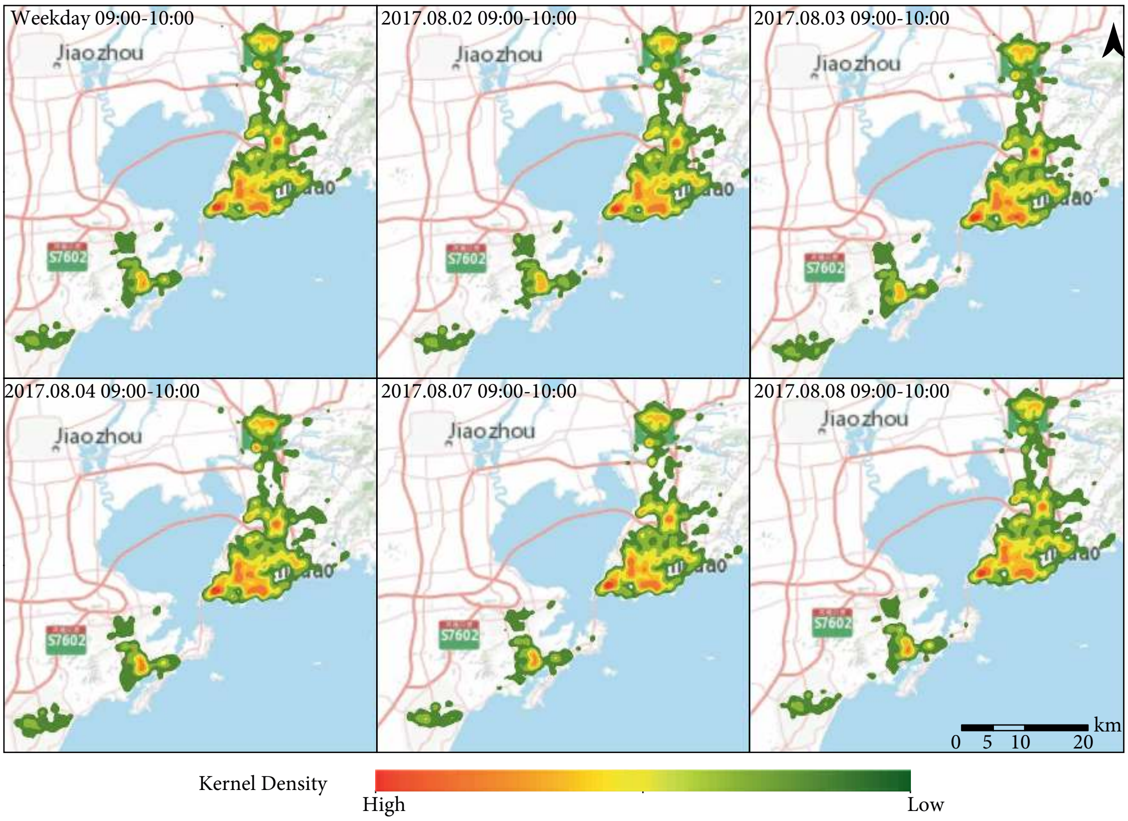


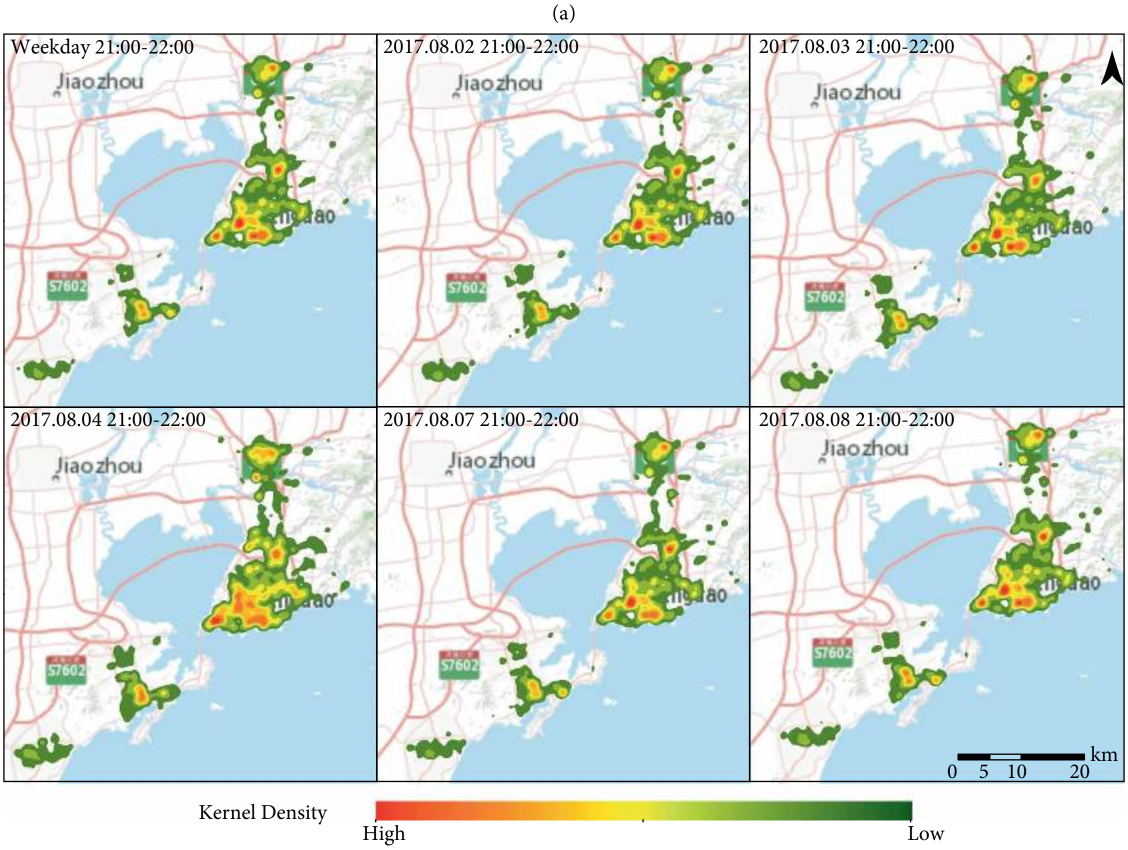
Figure 4: Variation of trips in weekdays and weekends.

dense, forming the most important hotspot area. In addition to the railway station, a small range of taxi passengers gathered around Qingdao Metro Line 3 of the Fushan institute station in the Shinan District. Secondly, there are more guests near Weihai road in Shibei District. Qingdao Taidong Commercial Pedestrian Street which is one of the five major shopping districts in Qingdao. In addition, near the Licun Metro Station in Licang District, where it is the most prosperous area in Licang District. The secondary hotspots are mainly distributed around the first-level hotspots, as well as Qingdao Liuting International Airport in Chengyang District, Jinggangshan road in Huangdao District and the area surrounded by Changjiang middle road (including Guanting, Jiajiayuan shopping mall, etc.), and Qingdao North Railway Station, etc.

Similarly, the weekends peak hotspots are more pronounced, while the early peaks have smaller range (Figure 6). The hotspots in the morning rush hour on Saturday are mainly concentrated near the railway station. Qingdao Railway Station is the largest railway station in Qingdao, and is responsible for passenger and cargo transportation tasks in Qingdao and surrounding areas. In addition to the hotspots at the railway station, the morning peak on Sunday also formed a hotspot around Taidong Pedestrian Street. The pick-up point for the Saturday night peak is mainly along the metro May Fourth Square Station to Yan'er island road Station, which locate with the new CBD of Qingdao. It is a good place for residents to go shopping on weekends through the large shopping malls such as The Mixc, Hisense Plaza, Belle Plaza and MYKAL. The hotspot of Licang



(a)



(b)

Figure 5: Kernel density of weekdays peak period (a) weekdays 09:00-10:00 (b) weekdays 21:00-22:00.

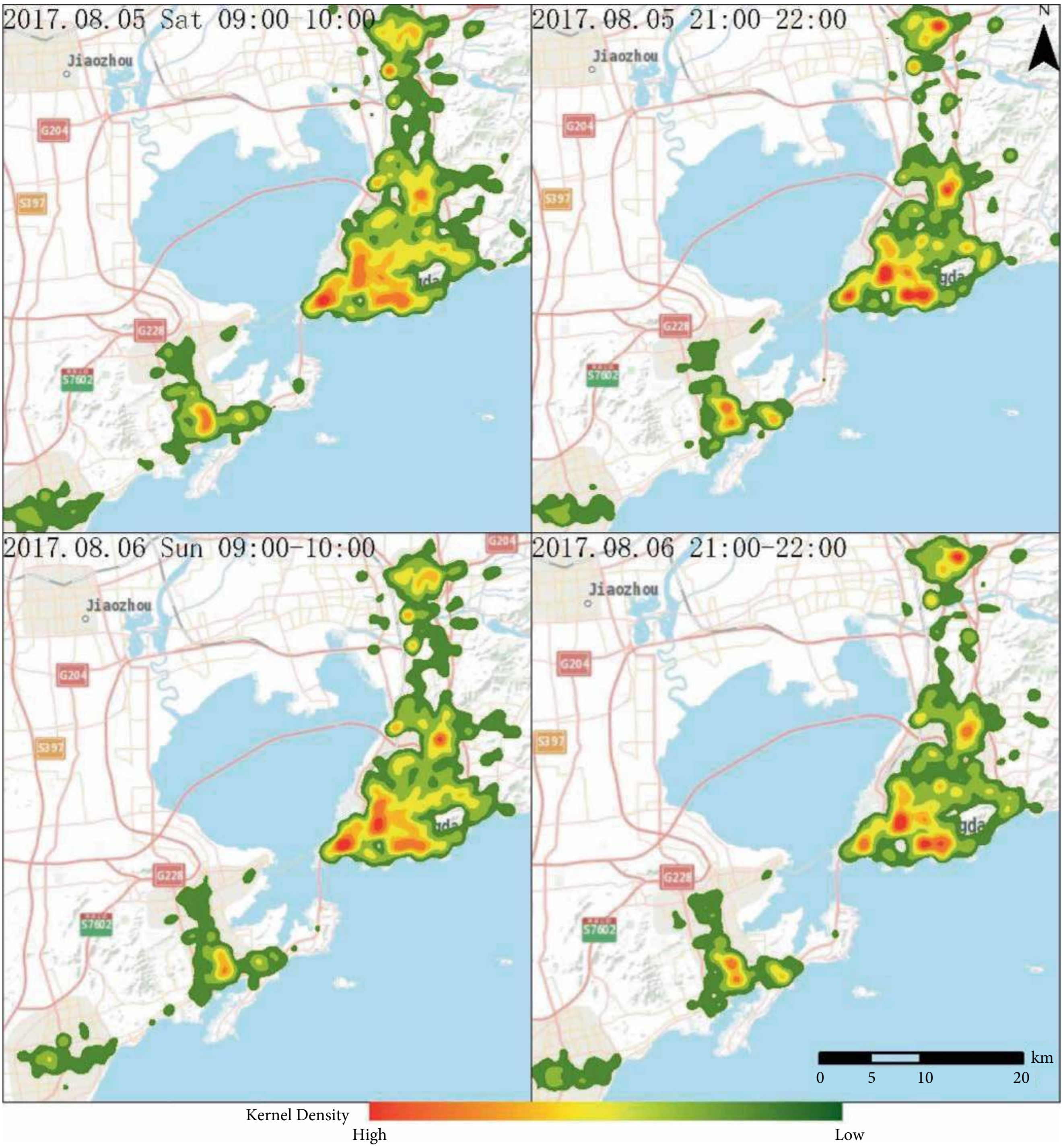


Figure 6: Kernel density of weekends peak period.

District is near the Licun metro station, which includes Suning Life Plaza, Laoshan department store, WeiKe Plaza and other shopping centers. There also locate with many colleges in Licang District, so students are always gathered here. The hotspots in Huangdao District are mainly around the Xiangjiang road traffic bureau, the Zijinshan road and the Jinggangshan road metro station. The surrounding area of Zijinshan road is a large collection of catering industry, and the intersection of Jinggangshan road and Changjiang road, there are shopping malls such as MYKAL, JUSCO, etc., these commercial places will attract more people to gather here.

3.3. Results of Geographically Weighted Regression. Before the regression analysis, in order to ensure the rationality of the model, the spatial autocorrelation and collinearity of the alternative independent variables should also be discussed.

Table 2: Correlation matrix.

	Residential density	Housing price	Road density	Bus station density	Parking lot density	Residential area	Commercial area	Public service area	Other land use	Land use mix
Residential density	1									
Housing price	0.218	1								
Road density	0.222	0.342	1							
Bus station density	0.380	0.437	0.465	1						
Parking lot density	0.408	0.479	0.497	0.442	1					
Residential area	0.047	0.090	0.033	0.164	−0.084	1				
Commercial area	0.133	0.201	0.203	0.214	0.205	−0.102	1			
Public service area	0.169	0.147	0.087	0.124	0.128	−0.126	0.038	1		
Other land use	−0.141	−0.193	−0.057	−0.196	−0.123	−0.093	0.444	0.432	1	
Land use mix	0.223	0.442	0.338	0.458	0.366	0.156	0.446	0.287	0.252	1

Table 3: Estimation results for global models (OLS).

	Variables	Coefficient		t-test		VIF	
		Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Socio-economic	Residential density	0.000	0.000	8.704	8.699	1.245	1.217
	Housing price	0.000	0.000	7.111	5.723	1.555	1.485
Traffic	Road density	0.000	0.000	4.027	4.943	1.424	1.403
	Bus station density	0.109	0.114	2.590	2.034	2.909	2.759
	Parking lot density	0.041	0.043	8.259	6.904	2.724	2.631
Land use	Residential area	3.528	4.064	−0.793	−0.292	1.431	1.512
	Commercial area	2.365	2.993	0.365	0.612	2.306	2.563
	Public service area	1.970	2.534	0.646	0.683	2.287	2.506
	Other land use	1.721	2.268	−0.676	−0.262	2.478	2.692
	Land use mix	1.460	1.816	−0.519	−1.322	2.092	2.025

Table 4: Comparison between global model and the GWR model.

	OLS		GWR	
	Weekday	Weekend	Weekday	Weekend
AICc	12166.605853	9753.084300	12035.744212	9709.16244
R^2	0.330356	0.315992	0.387968	0.360411
Adjusted R^2	0.328114	0.313254	0.387504	0.342968

Multiple independent variables are used in regression analysis, and these variables may be related to each other. Multicollinearity refers to the distortion or difficulty in estimating the model estimates due to the existence of precise correlations or high correlations between explanatory variables in linear regression models. Before proceeding to regression modeling, variables with higher correlation should be eliminates. Correlation matrix were developed to examine the correlation coefficients for each of the independent variables with other independent variables. Therefore, Pearson correlation is used to test the validity of the model of independent variable. It is observed that all correlations are less than 0.5 (Table 2).

OLS linear regression are first used to investigate significant factors that influence the urban taxi ridership and results are represented in Table 3. Ten independent variables are

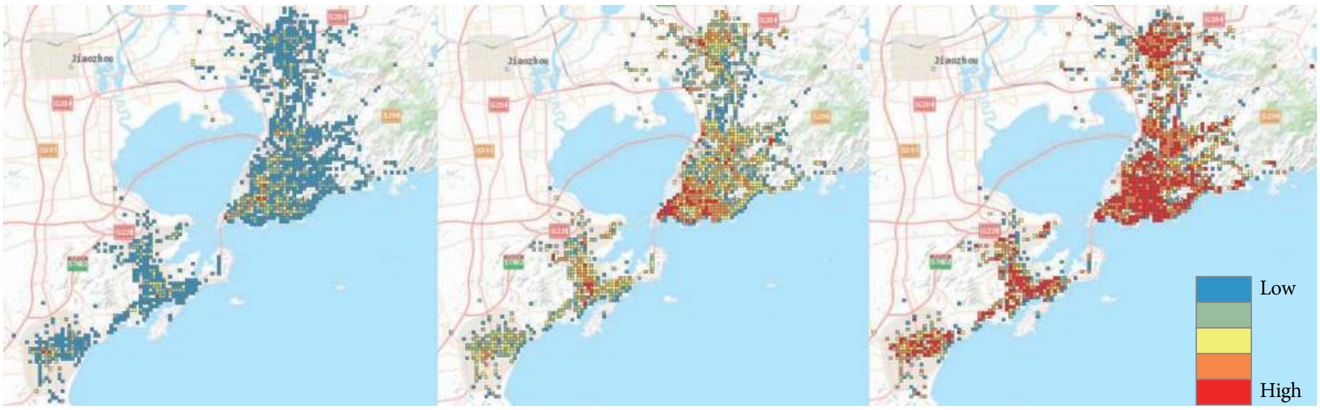


Figure 7: The distribution of main POIs (a) residential density (b) road density (c) land use mix.

reveals to be closely relate to the urban taxi ridership in both weekdays and weekends models. It was found that the variance inflation factor (VIF) was less than 7.5 and indicated that the independent variables are well selected, so that the multicollinearity issue is avoid.

As shown in Table 4, the adjusted R^2 is 0.387504 for weekday model and 0.342968 for weekend model, which corresponds to 0.05939 and 0.029714 improvement in the amount of variation explained compared to OLS linear regression. In addition, the reduction in AICc value is superior of GWR at same day.

4. Discussions

GWR model is used to analyze the spatial variation in the impact of road density, residential density, land use and other indicators on taxi travel demand, as well as the impact of different peak periods on weekdays and weekends. Here, we find that the coefficient in 5 weekdays and weekends has similar trends, so we select 4th, August as weekday and 6th, August as weekend to discuss the results. To better understand the results of each variable, the distribution of main POIs in each gird cell shown in Figure 7.

4.1. Impact of Socio-Economic Factors on Taxi Demand. Both the impact of residential density and housing prices have positive effect on residents' taxi travel, and the impact of residential density is greater than housing prices (Figure 8). For residential density, the coefficient of weekend peaks is greater than the weekday, and the impact range of weekend is wider, but the impact of housing price on taxi travel is opposite. The impact of housing price on taxi demand is relatively consistent, with changes on the weekend and the trend has gradually shifted from south to north. From a spatial perspective, the impact of residential density on taxi demand is more sensitive in parts of Shinan and Shibei District, mainly concentrated in Qingdao Railway Station and the Metro Line 2, Qingdao CBD, and Taidong Commercial Pedestrian Street, especially for weekend. The impact of housing price on the sensitive area is mainly in Huangdao. Moreover, although somewhere the housing price is high, the coefficient is relatively small. One possible explanation people with higher income may own their private cars thus reduce the choice of hailing a taxi.

4.2. Impact of Traffic Factors on Taxi Demand. In general, the impact of road density, bus station density and parking lot density also have positive impact on taxi travel. Parking lot density has the greatest impact on taxi travel demand, followed by bus station density (Figure 9). The impact of road density on taxi demand varies greatly on weekdays, with little difference between weekends. Places with higher road density may generate more taxi trips, and these places usually have a higher population density. Weekend coefficient value is lower compared to weekdays, which means that although the correlation coefficient is generally positive. The road density is lower, but more taxi trips may be generated, such as Laoshan scenic area surrounding on weekend where are fewer bus stops and bus lines. Areas with higher density of bus stations may generate more taxi trips, but the intensity on the weekend is reduced. This phenomenon can be explained in two ways. Firstly, places with bus stops are often crowded and tend to attract a large number of passengers, which in turn attracts and generates more taxi trips. Secondly, taxis may be widely used for connecting, and passengers can travel to and from the bus station's departure point and destination by taxi. The reason why the density of parking lot has a greater impact on traffic demand is related to the location of the parking lot. The spatial pattern of parking lots is generally located in hotels, shopping malls, transportation hubs and other buildings, so the crowds in these places are relatively dense and will encourage some passengers to choose taxis.

4.3. Impact of Land Use on Taxi Demand. Similar pattern is observed for the proportion of residential area, other land use and land use mix, they have negative effect on taxi demand, and other variables are positive (Figure 10). The lower the proportion of residential area, the fewer people choose to travel by taxi. From the status quo of land use in Qingdao, the area with a large living area is mainly in Shinan and Shibei District, where the housing prices are high and public transportation relatively developed. Therefore, majority people choose public transportation in these places. Residential density negatively on taxi ridership on weekday and weekend morning peak,

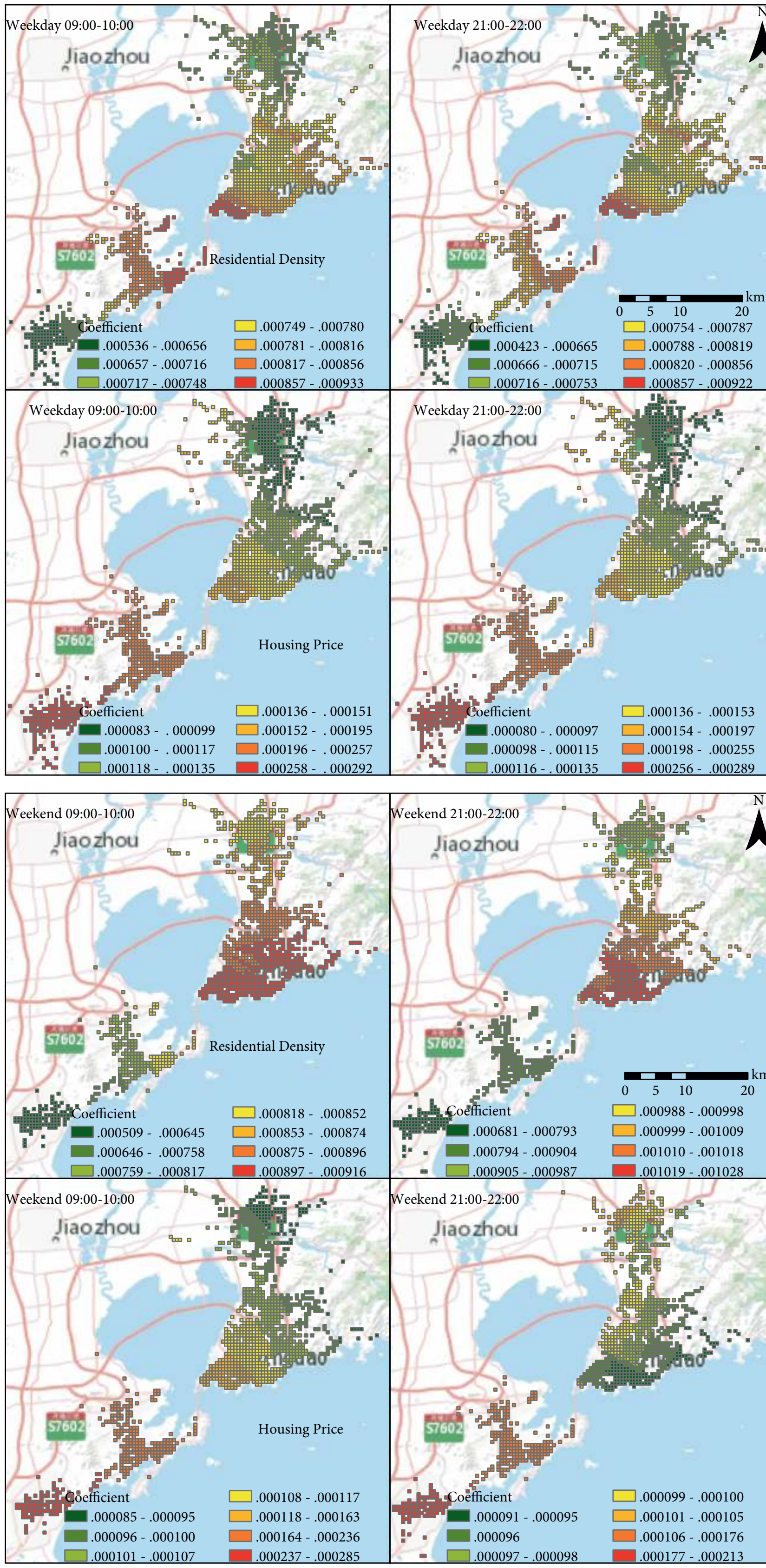
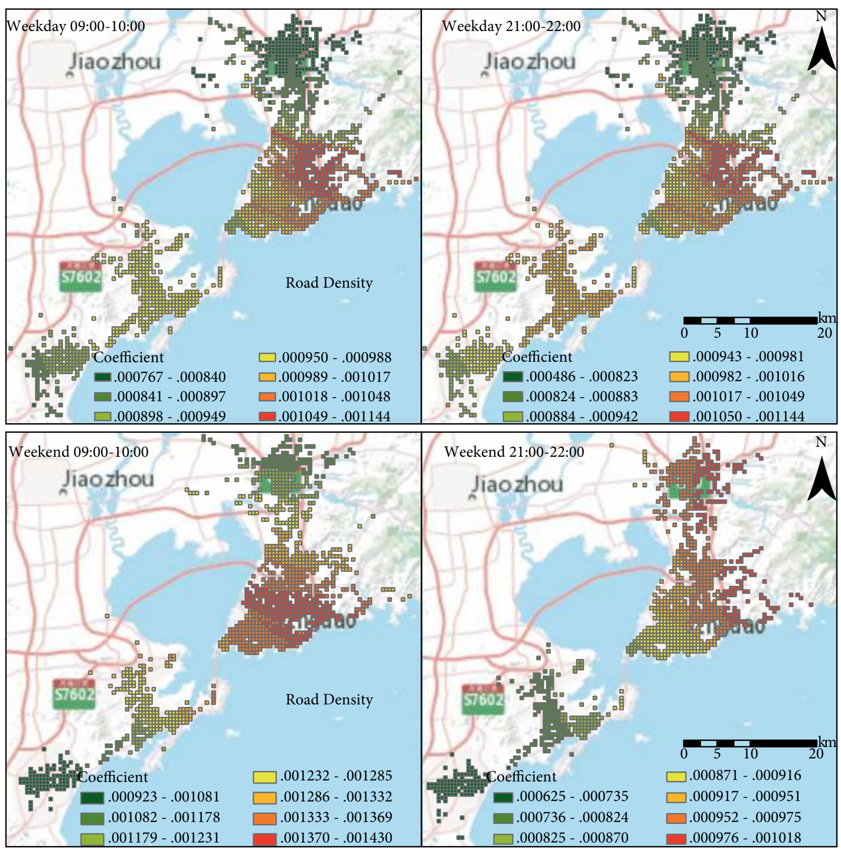
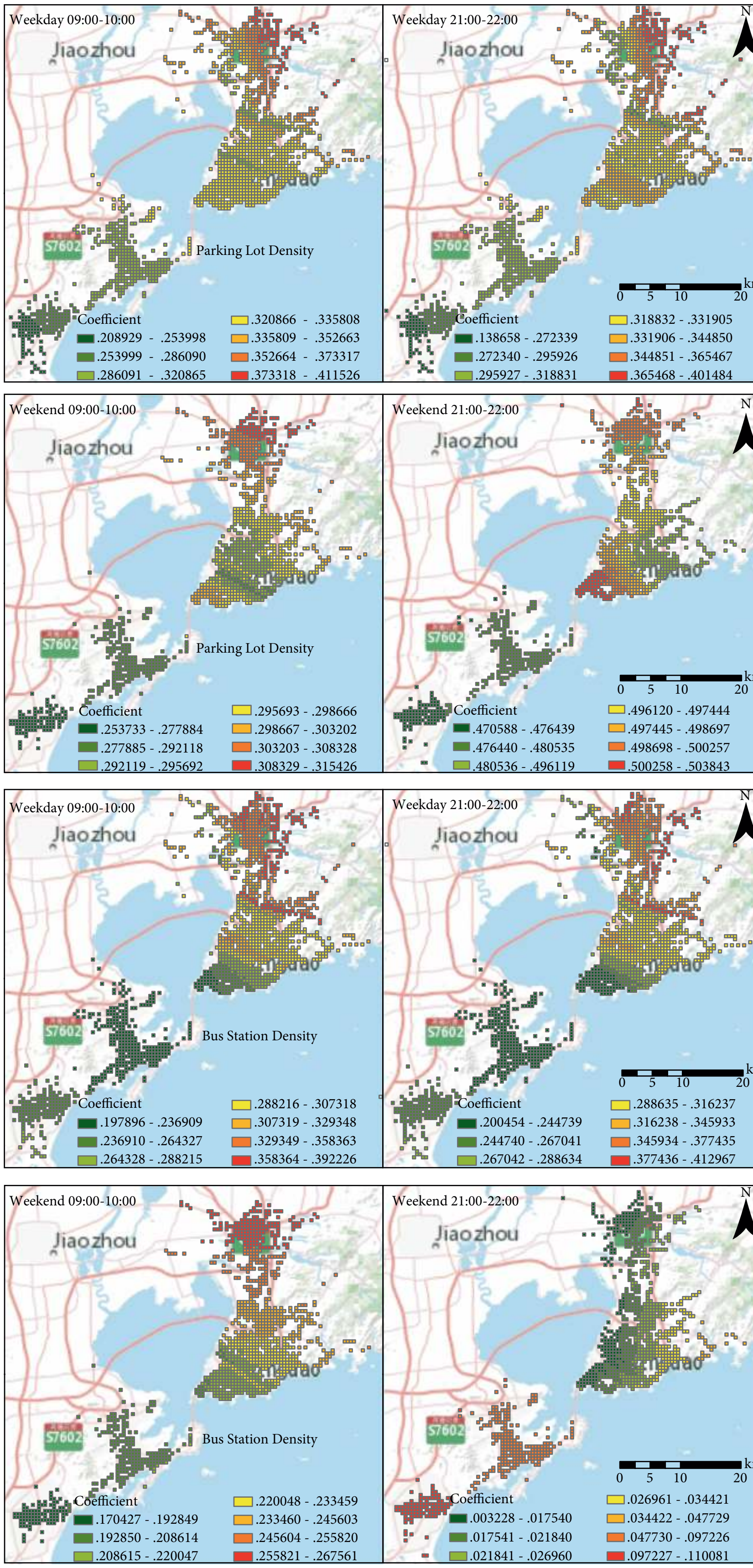


Figure 8: Spatial distribution for the coefficients of socio-economic factors for weekday and weekend.



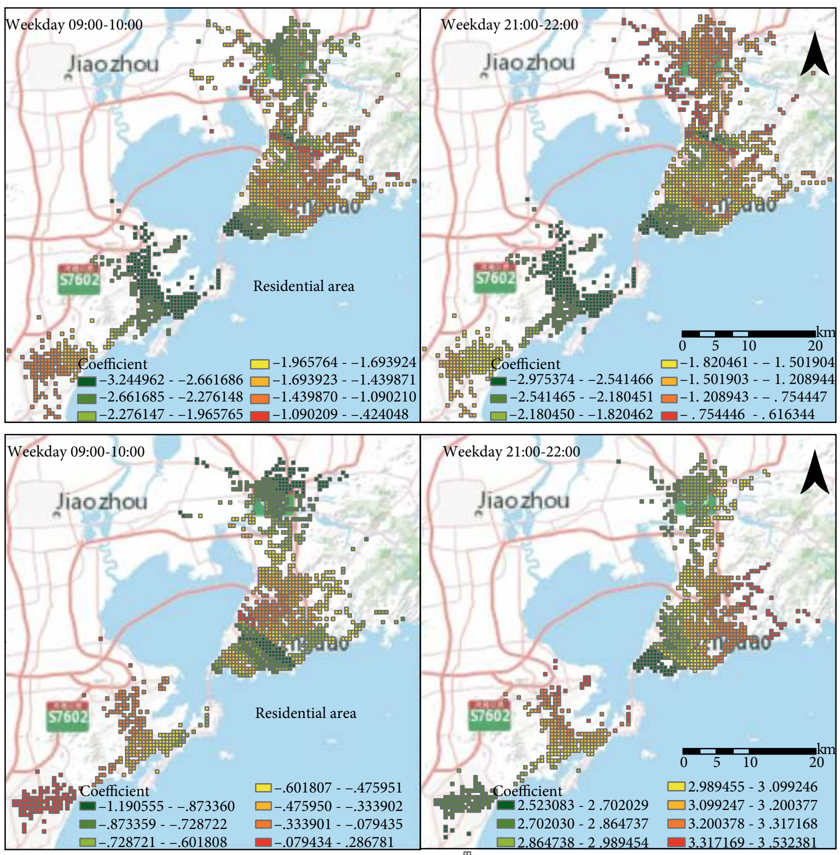
(a)

Figure 9: Continued.



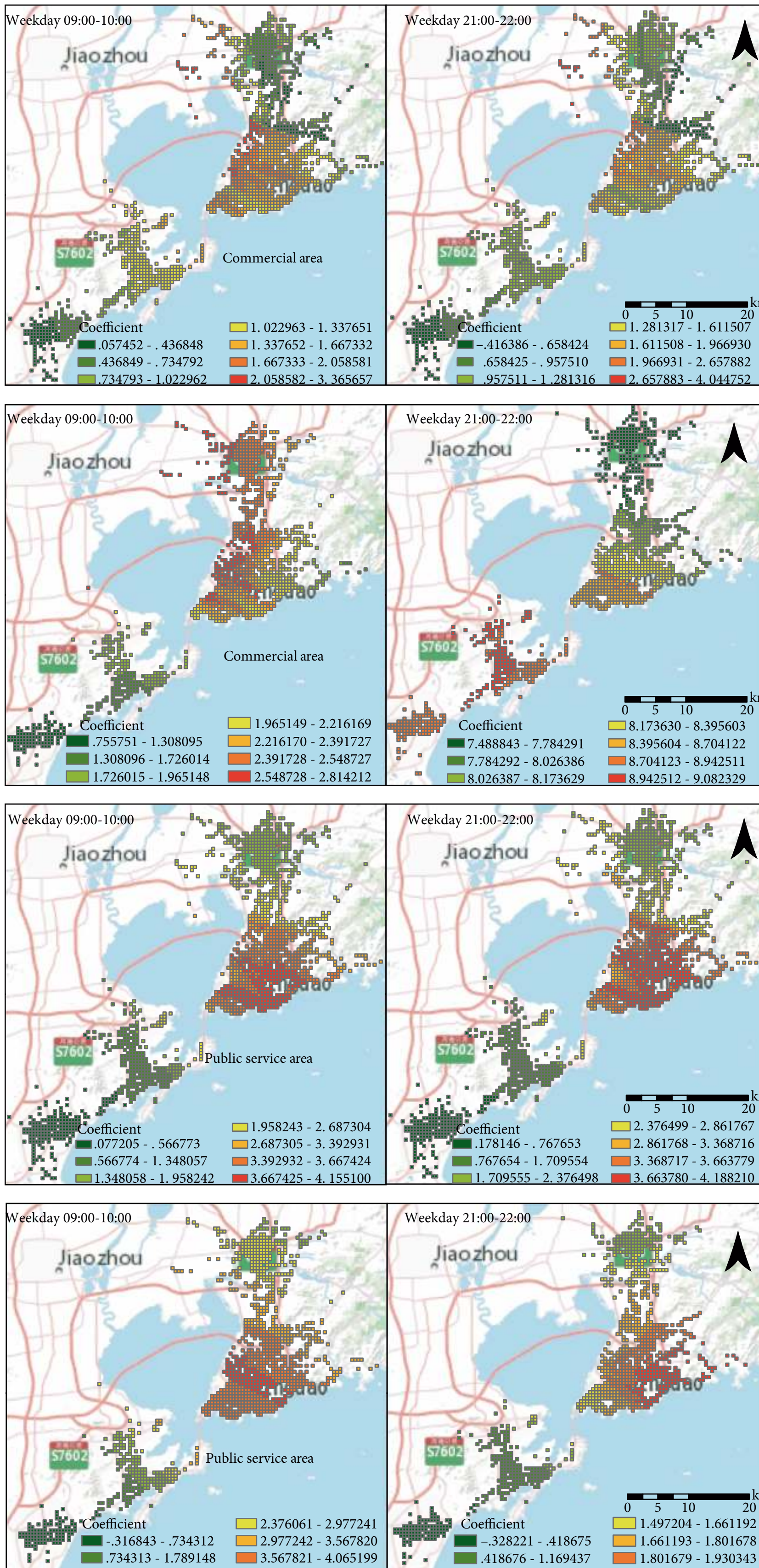
(b)

Figure 9: Spatial distribution for the coefficients of transport factors for weekday and weekend.



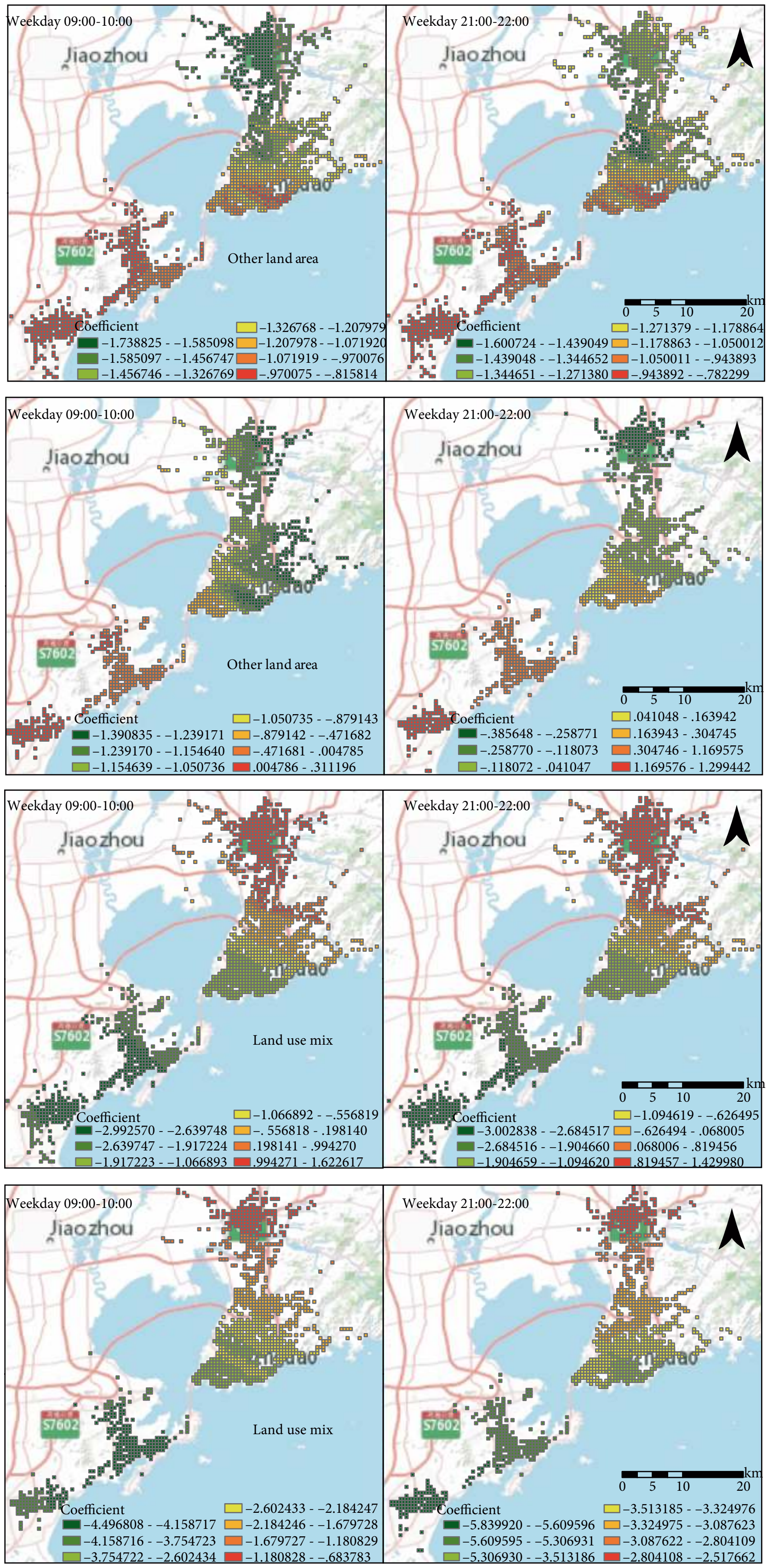
(a)

Figure 10: Continued.



(b)

Figure 10: Continued.



(c)

FIGURE 10: Spatial distribution for the coefficients of land use factors for weekday and weekend.

Table 5: Comparison of findings with existing studies.

Study	City	Ridership	Model	Significant factors
This study	Qingdao, China	Taxi	GWR	Residential density (+), housing price (+), road density (+), parking lot density (+), bus station density (+), residential area (∗), commercial area (+), public service area (∗), other land use (∗), land use mix (∗)
Tu et al. (2018)	Shenzhen, China	Taxi	GWR	Employment (∗), income (∗), land use mix (∗), road density (+), bus accessibility (∗), metro accessibility (+)
Yang et al. (2018)	Washington, DC, USA	Taxi	OLS regression model	Block (+), metro (+), bus (−), airport (+), pop (+), ResiDens (∗), RetaiDens (+), OffDens (∗), IndDens (+), OthDens (+), EmpEntropy (∗)
Qian and Ukkusuri (2015)	New York, USA	Taxi	GWR	Commuting time (−), highly educated population (+), median income (∗), road density (+), subway accessibility (+), commercial area (∗)

+positive effect, − negative effect, ∗ negative and positive effects.

but positively on weekend evening. The reason is likely to be related to the distribution of public facilities, weekend may generate more taxi trips. The larger the proportion of commercial land and public area, the more people likely to take taxis. It is more significant on weekend, the main purpose of residents' travel is work, school, entertainment, etc., and most of these destinations are concentrated in areas with public facilities and more commercial areas where tend to form a large crowd. The increase of land use mix generally induces a decrease in taxi ridership mainly due to the land use form. Reasonable land use patterns advocate compact layout, mixed use of land forms, promote high-intensity development to encourage the use of public transportation, and promote the formation of good urban structure and land use layout. These measures will effectively reduce the total demand for residents' travel, shorten the travel distance of residents, and form the traffic demand characteristics that are conducive to the development of public transportation.

To have a better understand of taxi demand results, we compared them with studies in worldwide cities such as Shenzhen [48], Washington [49] and New York [20]. Table 5 list the results of existing researches. Shenzhen and Qingdao both are coastal cities of China, but Shenzhen is the first economic special area of China, the road density, bus accessibility and income have a large influence than other transport modes. In the USA, metro accessibility and residential density in Washington and New York both have a positive impact on taxi demand. However, land use factors have different influence in cities because of the spatial heterogeneity of the economic structure and geographic location.

5. Conclusions

With the improvement of residents' living standards and the emergence of third-party platforms such as "DiDi" and "Shenzhou", the demand for urban taxis is increasing. However, the relationship between supply and demand of taxi is imbalanced no matter in temporal or spatial pattern, so it is an urgent need for us to analysis the taxi demand and its influence factors to help government make better policies. This study establishes on GWR model to include 3 categories variables to assess taxi demand pattern in Qingdao, China. Pick-up locations are derived from taxi GPS trajectory, combined with the POI data and road network data. All information then integrated at $500*500$ m grid cells in ArcGIS 10.2 system to explore the travel pattern of taxi demand and its related influence factors.

Analyzing the model and examine the temporal-spatial variations from the morning and evening peak hours of the weekend and weekdays using kernel density model. By establishing a taxi-based travel demand GWR model, which is related to socio-economic, land use and traffic factors. Results indicates that there some spatial and temporal differences in the influence of independent variables on taxi travel demand, and different factors have different impact mechanisms. The model shows strong links between demand for taxi and urban form characteristics. For example, residential density, housing price and road density are positively correlated with higher demand for taxi, but land use mix, residential area shows negatively on travel demand. The demand for taxi is higher in Shinan and Shibei District where economic level is in good condition of the whole city. These findings imply that taxi is more likely to be compete with other travel modes. Areas with higher economic levels may be more likely to choose private cars or other public transportation. This is coincided with the reality, and the areas where residents' travel demand is relatively strong are mainly concentrated in commercial areas and areas with more public service facilities. By addressing spatial nonstationary effects, the GWR model can be used to predict future variations of taxi demand. This mechanism is beneficial to further understand the spatial distribution and time evolution of urban taxi travel demand. This paper analyzes and visualizes the influence of different factors, which can better understand the spatial heterogeneity of taxi travel demand in different time periods. Firstly, the results can provide reference for urban residents' travel time planning. For example, morning peak (09:00–10:00) and evening peak (21:00–22:00) have different hotspots. Thus, taxi supply may not sufficient in these hotspot areas, residents can stagger their travel hours. Secondly, in the construction of new urban areas and the division of functional areas, the results can provide reference for urban planners making decisions for the allocation and management of transportation resources. For example, land use mix have a strong impact on taxi demand. In the construction and planning of new urban areas, improve the land use mix will relieve the pressure of traffic demand and alleviate traffic congestion. Meanwhile, the results can provide transportation managers for optimizing the taxi scheduling in daily decision-making, and smart city construction which more coincide with residents' travel demand.

In summary, the application of the GWR model can help to understand the impact mechanism of socio-economic and land use factor on taxi travel demand in different locations. At the same time, this paper examines the difference in time between weekdays and weekends, which makes up for the shortcomings of time dimension research. It can provide a reference for temporal-spatial dimensions for taxi travel demand. This study also has several limitations. Firstly, the socio-economic factors such as income and gender have not been quantitatively analyzed. In the future research framework, the dimensions of influencing factors can be further enriched. In addition, new modes of transportation and congestion are not considered, these factors can also be placed in the research framework.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper is partly supported by the National Natural Science Foundation of China (71801144, 71371111), China Postdoctoral Science Foundation Funded Project (2019M652437), National Key R&D Plan (No. 2018YFC0831002), the Key R&D Plan of Shandong Province (No. 2018GGX101045), Humanity and Social Science Fund of the Ministry of Education (No. 18YJAZH136), the Key Project of Industrial Transformation and Upgrading (Made in China 2025 No. TC170A5SW), Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents (2019RCJJ014), Shandong Postdoctoral Innovation Project (201903030) Qingdao Social Science Planning Research Project (QDSKL1901121) and Shandong University of Science and Technology Innovation Project (SDKDYC180227).

References

[1] T. Çetin and Y. K. Eryigit, "Estimating the effects of entry regulation in the Istanbul taxicab market," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 6, pp. 476–484, 2011.

[2] B. Schaller, "A regression model of the number of taxicabs in US cities," *Journal of Public Transportation*, vol. 8, no. 5, pp. 6363–78, 2005.

[3] Y. Yue, T. Lan, G. O. Yeh Anthony, and Q. Li, "Zooming into individuals to understand the collective: a review of trajectory-based travel behavior studies," *Travel Behaviour and Society*, vol. 1, no. 2, pp. 719–723, 2014.

[4] W. Tu, J. Z. Cao, Y. Yue, S. L. Shaw, M. Zhou, and Z. Wang, "Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns," *International Journal of Geographical Information Science*, vol. 31, no. 12, pp. 2331–2358, 2017.

[5] C. Z. Yao and J. N. Lin, "A study of human mobility behavior dynamics: a perspective of a single vehicle with taxi," *Transportation Research Part A: Policy and Practice*, vol. 87, pp. 51–58, 2016.

[6] J. J. Tang, F. Liu, Y. H. Wang, and H. Wang, "Uncovering urban human mobility from large scale taxi GPS data," *Physica A: Statistical Mechanics and its Applications*, vol. 438, no. 15, pp. 140–153, 2015.

[7] X. Liu, L. Gong, Y. X. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *Journal of Transport Geography*, vol. 43, pp. 78–90, 2015.

[8] H. Yang and S. C. Wong, "A network model of urban taxi services," *Transportation Research Part B: Methodological*, vol. 32, no. 4, pp. 235–246, 1998.

[9] K. I. Wong, S. C. Wong, and H. Yang, "Modeling urban taxi services in congested road networks with elastic demand," *Transportation Research Part B: Methodological*, vol. 35, no. 9, pp. 819–842, 2001.

[10] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: a recommender system for finding passengers and vacant taxis," *IEEE Transaction on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2390–2403, 2013.

[11] K. Kim, "Exploring the difference between ridership patterns of subway and taxi: case study in Seoul," *Journal of Transport Geography*, vol. 66, pp. 213–223, 2018.

[12] H. T. Yang, X. Z. Lu, C. Cherry, X. H. Liu, and Y. L. Li, "Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression," *Journal of Transport Geography*, vol. 64, pp. 184–194, 2017.

[13] K. Maat and H. Timmermans, "Influence of the residential and work environment on car use in dual-earner households," *Transportation Research Part A: Policy and Practice*, vol. 43, no. 7, pp. 654–664, 2009.

[14] C. Ding, Y. Lin, and C. Liu, "Exploring the influence of built environment on tour-based commuter mode choice: a cross-classified multilevel modeling approach," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 230–238, 2014.

[15] R. Rixey, "Station-level forecasting of bike sharing ridership: station network effects in Three US Systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2387, no. 1, pp. 46–55, 2013.

[16] R. Cervero, "Built environments and mode choice: toward a normative framework," *Transportation Research Part D: Transport and Environment*, vol. 7, no. 4, pp. 265–284, 2002.

[17] C. Chen, D. Varley, and J. Chen, "What affects transit ridership? a dynamic analysis involving multiple factors, lags and asymmetric behavior," *Urban Studies*, vol. 48, pp. 1893–1908, 2011.

[18] J. Gutiérrez, O. D. Cardozo, and J. C. García-Palomares, "Transit ridership forecasting at station level: an approach based on distance-decay weighted regression," *Journal of Transport Geography*, vol. 19, pp. 1081–1092, 2011.

[19] J. Zhao, W. Deng, Y. Song, and Y. Zhu, "Analysis of metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models," *Transportation*, vol. 41, pp. 133–155, 2014.

[20] G. Gao, H. J. Sun, and J. J. Wu, "Activity-based trip chaining behavior analysis in the network under the parking fee scheme," *Transportation*, vol. 46, no. 3, pp. 647–669, 2019.

[21] X. W. Qian and S. V. Ukkusuri, "Spatial variation of the urban taxi ridership using GPS data," *Applied Geography*, vol. 59, pp. 31–42, 2015.

[22] Y. Yang, Z. He, Z. Y. Song, X. Fu, and J. W. Wang, "Investigation on structural and spatial characteristics of taxi trip trajectory network in Xi'an, China," *Physica A: Statistical Mechanics and its Applications*, vol. 506, pp. 755–766, 2018.

[23] M. G. McNally, "The four-step model. UC Irvine: center for activity systems analysis," 2008, http://escholarship.org/uc/item/0r75311t.

[24] M. H. Kutner, C. J. Nachtsheim, and J. Neter, *Applied Linear Regression Models*, McGraw-Hill/Irwin Education, 2004.

[25] C. Brunsdon, A. S. Fotheringham, and M. Charlton, "Geographically weighted regression: a method for exploring spatial non-stationarity," *Geographical Analysis*, vol. 28, no. 4, pp. 281–298, 1996.

[26] O. D. Cardozo, J. C. García-Palomares, and J. Gutiérrez, "Application of geographically weighted regression to the direct forecasting of transit ridership at station-level," *Applied Geography*, vol. 34, pp. 548–558, 2012.

[27] G. Gao, Z. Wang, X. M. Liu, Q. Li, W. Wang, and J. Zhang, "Travel behavior analysis using 2016 Qingdao's household traffic

surveys and Baidu electric map API data," *Journal of Advanced Transportation*, vol. 2019, Article ID 6383097, 18 pages, 2019.

[28] Z. Wang, G. Gao, X. M. Liu, and W. Lyu, "Verification and analysis of traffic evaluation indicators in urban transportation system planning based on multi source data a case study of Qingdao city, China," *IEEE Access*, vol. 7, no. 1, pp. 110103–110115, 2019.

[29] Z. Fang, S. Shaw, W. Tu, Q. Li, and Y. Li, "Spatio temporal analysis of critical transportation links based on time geographic concepts: a case study of critical bridges in Wuhan, China," *Journal of Transport Geography*, vol. 23, no. 1, pp. 44–59, 2012.

[30] Y. Zhou, Z. X. Fang, J. C. Thill, Q. Q. Li, and Y. G. Li, "Functionally critical locations in an urban transportation network: Identification and space-time analysis using taxi trajectories," *Computers, Environment and Urban Systems*, vol. 52, pp. 34–47, 2015.

[31] W. B. Zhang, S. V. Ukkusuri, and J. J. Lu, "Impacts of urban built environment on empty taxi trips using limited geolocation data," *Transportation*, vol. 44, no. 6, pp. 1445–1473, 2017.

[32] B. D. Taylor and N. Y. Fink Camille, *The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature*, UCLA, Los Angeles Transit Administration, 2003.

[33] X. Chu, *Ridership Models at the Stop Level*, National Center of Transit Research, University of South Florida, 2004.

[34] C. Chalkias, A. G. Papadopoulos, and K. Kalogeropoulos, "Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: empirical evidence from Athens, Greece," *Applied Geography*, vol. 37, pp. 34–43, 2013.

[35] L. E. Ramos-Santiago and J. Brown, "A comparative assessment of the factors associated with station-level streetcar versus light rail transit ridership in the United States," *Urban Studies*, vol. 53, no. 5, pp. 915–935, 2016.

[36] R. Hughes and D. MacKenzie, "Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators," *Journal of Transport Geography*, vol. 56, pp. 36–44, 2016.

[37] F. Des Rosiers, M. Lagana, M. Theriault, and M. Beaudoin, "Shopping centres and house values: an empirical investigation," *Journal of Property Valuation and Investment*, vol. 14, no. 4, pp. 41–62, 1996.

[38] L. Zhang, J. H. Hong, A. Nasri, and Q. Shen, "How built environment affects travel behavior: a comparative analysis of the connections between land use and vehicle miles traveled in US cities," *Journal of Transport and Land Use*, vol. 5, no. 3, pp. 40–52, 2012.

[39] V. Chakour and N. Eluru, "Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal," *Journal of Transport Geography*, vol. 51, pp. 205–217, 2016.

[40] Y. C. Chiou, R. C. Jou, and C. H. Yang, "Factors affecting public transportation usage rate: geographically weighted regression," *Transportation Research Part A: Policy and Practice*, vol. 78, pp. 161–177, 2015.

[41] N. Hu, E. F. Legara, K. K. Lee, G. G. Hung, and C. Monterola, "Impacts of land use and amenities on public transport use, urban planning, and design," *Land Use Policy*, vol. 57, pp. 356–367, 2016.

[42] C. Schreiber, "The economic reasons for price and entry regulation of taxicabs," *Journal of Transport Economics and Policy*, vol. 15, no. 1, pp. 81–83, 1981.

[43] B. Schaller, "Elasticities for taxicab fares and service availability," *Transportation*, vol. 26, pp. 283–297, 1999.

[44] J. P. Toner, "The welfare effects of taxicab regulation in English towns," *Economic Analysis and Policy*, vol. 40, no. 3, pp. 299–312, 2010.

[45] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, 1956.

[46] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[47] W. Tober, "A computer movie simulation urban growth in the Detroit region," *Economic Geography*, vol. 46, no. 2, pp. 234–240, 1970.

[48] T. Wei, C. Rui, Y. Yang, B. Zhou, Q. Li, and Q. Li, "Spatial variations in urban public ridership derived from GPS trajectories and smart card data," *Journal of Transport Geography*, vol. 69, pp. 45–57, 2018.

[49] Y. Zhou, M. L. Franz, S. Zhu, J. Mahmoudi, A. Nasri, and L. Zhang, "Analysis of Washington, DC taxi demand using GPS and land-use data," *Journal of Transport Geography*, vol. 66, pp. 34–44, 2018.