Examining the spatial-temporal relationship between urban built environment and taxi ridership: Results of a semi-parametric GWPR model

Journal of Transport Geography 96 (2021) 103172

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Examining the spatial-temporal relationship between urban built environment and taxi ridership: Results of a semi-parametric GWPR model

Chao Chen [a,c], Tao Feng [b,c], Chuan Ding [d], Bin Yu [d,e], Baozhen Yao [a,*]

[a] State Key Laboratory of Structural Analysis for Industrial Equipment, School of Automotive Engineering, Dalian University of Technology, Dalian 116024, PR China
[b] Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima 739-8529, Japan
[c] Urban Planning and Transportation, Department of the Built Environment, Eindhoven University of Technology, Eindhoven 5600MB, the Netherlands
[d] School of Transportation Science and Engineering, Beihang University, Beijing 100191, PR China
[e] BDBC, Beihang University, Beijing 100191, PR China

ARTICLE INFO

Keywords:
Built environment
Points of interests (POIs)
Taxi ridership
Spatial autocorrelation
Semi-parametric geographically weighted Poisson regression (SGWPR) model

ABSTRACT

With the advance of intelligent transportation systems (ITSs) and data acquisition systems (DASs), it becomes possible in recent to explore the determinants of urban taxi ridership using multi-source heterogeneous data. This paper aims to use floating car data, points-of-interests (POIs) data and housing-price data to assess the influence of the built environment on taxi ridership. Within a scale of 0.5 km grid, critical indicators related to the economic aspect, intermodal connection, and land use factors were obtained using the multi-source data in Shanghai. To capture the spatial and temporal heterogeneity, Semi-parametric Geographically Weighted Poisson Regression (SGWPR) models are built over different time dimensions. It is found that SGWPR models result in higher goodness-of-fit than the generalized linear models. More importantly, the results show the impacts of built environment factors on taxi demand are highly heterogeneous, positive or negative in different city areas, reflected in the significant temporal variations of the effects. Overall, these findings suggest that the built environment factors have significant impacts on urban taxi demand, and the spatial context should not be ignored. Findings in this paper are expected to help better understand the relationship between urban taxi demand and built environment factors, improving the service level of the urban taxi system, and offering valuable insights into future urban and transportation planning.

1. Introduction

With the increasing improvement of living conditions, the number of private vehicles worldwide has undergone explosive growth in the past decades. Consequently, the problems of heavy traffic congestions, excessive energy consumption, and consequential severe environmental pollution have become particularly acute and cannot be ignored anymore (Zhang et al., 2017a, 2017b). On the other hand, the course of urbanization is also experiencing rapid development, especially in developing countries. The built environments in these cities have a dramatic transformation, from a high-density urban form to relatively decentralized communities in peri-urban areas (An et al., 2019). It is for these reasons that urban travel demands and patterns are being significantly influenced. Therefore, to avoid blind urban expansion and address those growing social problems, great attention has been paid to the studies on the correlation between built environment factors and travel demands. Findings in these studies are expected to provide strategic support to greener urban development because the built environments present diverse effects on travel demands (Cervero, 2013; Cordera et al., 2017; An et al., 2019; Liu et al., 2020; Zhang et al., 2020).

In recent years, taxi and online taxi-hailing services have emerged as a viable disincentive to the usage of private vehicles (Sun and Ding, 2019; Liu et al., 2020). By providing travelers with a convenient, flexible, and door-to-door service, taxis have made itself become an increasingly popular travel mode, especially in areas lacking of metro or bus services. Needless to say, taxis play a crucial role in urban mobility within the built environment. However, the interdependency between taxi demand and built environment has not been addressed adequately (Qian and Ukkusuri, 2015; Yang et al., 2018; Liu et al., 2020). This is perhaps because on the one hand taxi demand in western countries is relatively low due to the high private car ownership and high taxi cost, and on the other hand the lack of big taxi data at the city scale which

* Corresponding author.
E-mail address: yaobaozhen@dlut.edu.cn (B. Yao).

https://doi.org/10.1016/j.jtrangeo.2021.103172
Received 25 January 2021; Received in revised form 23 June 2021; Accepted 17 August 2021
Available online 31 August 2021

becomes available only in recent due to the development of sensing and intelligent transportation systems (ITSs).

Exploring urban taxi demand therefore is a non-trivial task because the determinants associated with taxi demand is difficult to articulate in terms of the varying urban functionality. Furthermore, failure to capture the local variations will undoubtedly reduce the model reliability and biased our understanding of the spatial variation of taxi ridership. Other than the spatial variation, taxi demand is also featured with a significant temporal variation, e.g. at different times of day, days of the week. Thus, the temporal factors should also be incorporated into the conceptual framework for modeling taxi demand.

In recent, the data for analyzing taxi demand and built environment are becoming increasingly available. For instance, the open taxi GPS data in many cities provides great potential to gain further understanding of taxi demand spatially and temporally. In addition, traditional surveys or official statistics are often used to explore the factors related to the built environment (Zhu et al., 2019). When combining the open taxi data with built environment data such as POI, land use data, further insight on the dependency of taxi demand on built environment may be obtained (Lim et al., 2018; Li et al., 2019).

With the facts mentioned above, this study intends to explore the spatial-temporal correlation between taxi ridership and built environment factors using multi-source data. More specifically, we study the spatial-temporal distribution of taxi demands based on floating car data and extract the built environment factors based on open POIs data of Shanghai. A series of semi-parametric GWPR models are used to explore the spatial-temporal relationship between taxi demand and built environment factors.

2. Literature review

Built environment is an innate driver of travel needs (Cervero and Kockelman, 1997; Qian and Ukkusuri, 2015; Yu and Peng, 2019; Liu et al., 2020). Significant influences of built environment factors on travel needs have been found in prior studies, present the differences among different transportation modes, such as transit, carsharing, taxi, bicycling (Cordera et al., 2017; An et al., 2019; Liu et al., 2020; Zhang et al., 2020). Since the objective of this paper is to examine the spatial-and-temporal relationship between built environment and taxi ridership, the review mainly focuses on the correlation between built environments and taxi ridership. The associated aspects are summarized as follows: (1) taxi ridership and its influential factors; (2) data sources; (3) methodologies and (4) research gaps.

2.1. Built environment factors and taxi ridership

Built environment refers to an integration of various spatial features of the environment where the taxi demands are generated. Previous studies generally found that highly mixed land use could lead to an increase in taxi demand. Different types of land use, such as residential land, commercial land, official land, and parking space, and mixed land-use indexes (e.g., mix-use entropy, land use balance, and land use dissimilarity) have been investigated (Qian and Ukkusuri, 2015; Zhang et al., 2017a, 2017b; Liu et al., 2020; Yu and Peng, 2019). In addition, the concept of the 3Ds, e.g., density, diversity, and design, is one of the popular methods to study the impacts of the diversity of land use on taxi demand in previous studies (Cervero and Kockelman, 1997).

Since the taxi ridership can be easily affected by other transportation modes like public transportation, some studies also incorporated transit accessibility, as an additional group of variables, into the concept of 3Ds to explore the relationship between alternative transportation modes. For instance, using a large set of GPS data from New York City taxis, Yang and Gonzales (2014) found that transit access time has a positive correlation with the taxi demand. ss.

Despite the different concepts representing built environment features, it should be noted that existing studies often miss representing detailed information. However, understanding the impacts of the full range of built environment factors are crucial for traffic management to formulate and implement proper policies (Yu and Peng, 2019). To better reflect the specific context of cities, in this paper, points of interests (POIs) with detailed categories are directly introduced.

2.2. Data source

Accurate data is fundamental to understand the impacts of independent variables on taxi ridership. The taxi ridership is traditionally obtained by the counting data in checkpoints (Yang et al., 2000). In recent, GPS data with the precise pick-up location information provides a better opportunity to investigate citywide taxi ridership (Qian and Ukkusuri, 2015; Sun et al., 2018; Zhang et al., 2020; Chen et al., 2020). In this study, we employed taxi GPS data to investigate urban taxi ridership. The extracted ridership is treated as the dependent variable in the model.

Regarding the data of explanatory variables, the land-use or demographic data are typically obtained through traditional surveys and official statistics (Qian and Ukkusuri, 2015; Yu and Peng, 2019; Zhu et al., 2019; Liu et al., 2020), which is not only erroneous but also time consuming. As an alternative, the data obtained via web crawler technology becomes increasingly accessible (Lim et al., 2018). As one of the representatives of such open data, POIs have become popular in land use analysis and are regarded as a practical method to precisely define and explain the land use of a specific area (Li et al., 2019; Zhang et al., 2020). Based on the data available, this study employed the POIs data together with other supplementary data, which were acquired from open Application Program Interfaces (APIs) of relevant websites.

2.3. Methodologies

The most common approach in previous studies for ridership analysis is the ordinary least square (OLS) regression model (An et al., 2019; Yu and Peng, 2019). Other methods including discrete choice models, spatial clustering models and machine learning algorithms, are also applied in different contexts (Chang et al., 2010; Jun et al., 2015; Shao et al., 2015). However, these methods are based on a common assumption that all variables are spatially stationary across the study area (Brunsdon et al., 1996). In fact, taxi ridership in one area is highly related to urban forms, and the effects of explanatory variables may vary over space (Qian and Ukkusuri, 2015). To capture the local variations, geographically weighted regression (GWR) models have been introduced (Qian and Ukkusuri, 2015; Yu and Peng, 2019), whereas the spatial-varying coefficients interpret the spatial heterogeneity of the geographical data. In this paper, we employ a direct demand model, semi-parametric geographically weighted Poisson regression model (SGWPR) to investigate the impact of built environment on taxi ridership.

2.4. Research gaps

Several limitations in existing literature could be found. First, most studies measured the built environment factors using traditional surveys or official statistics. Second, the land-use indexes (e.g., mix-use entropy and land-use dissimilarity index) used to understand taxi ridership may include systematic errors. Third, the taxi ridership has been examined from the perspective of spatial variation but not temporal variations. Therefore, to fill these research gap, this study uses free open data acquired from APIs of relevant websites to explore the relationship between the built environment and taxi ridership. Other than defining complex land-use indexes, we directly employ detailed land-use categories to measure the built environment. Moreover, we explore both the spatial and temporal heterogeneity of taxi ridership using the semi-parametric GWPR model. Results obtained in this paper could serve as a reference for taxi drivers to reduce the customer-searching time and

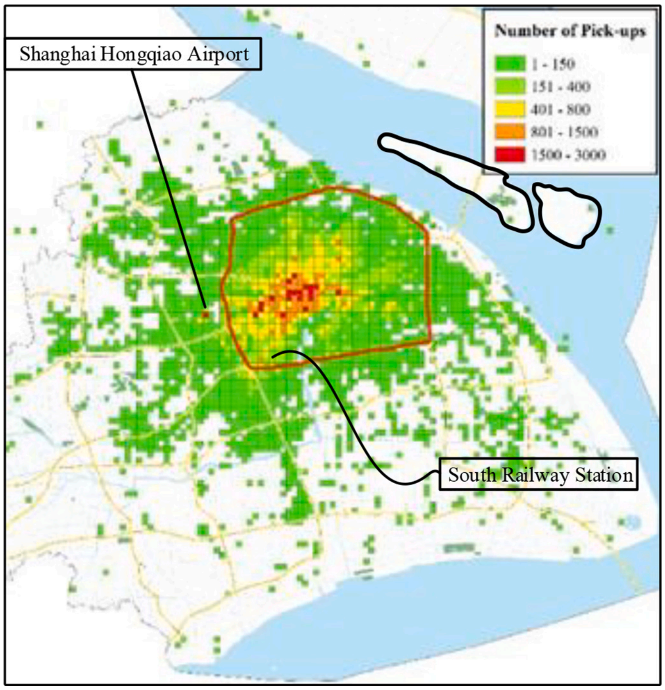


Fig. 1. The spatial distribution of pick-ups in Shanghai city during a typical weekday.

increase their income level, as well as for government planners to be more conscious of these effects and adjust their spatial planning to better improve network efficiency.

3. Data description, study area and time span

3.1. Study area, time span and dependent variables

Shanghai is the biggest and most dynamic city in China and a city where jobs and residences are mainly located in the central area. Based on the National Bureau of Statistics data, the regionalism of Shanghai city at the end of 2016 includes 16 districts, among which 15 districts, including Jing'an, Xuhui, and other 13 districts, were selected as the study areas, excluding Chongming district because it was established later than the data extraction time.

To analyze the spatial distribution of taxi ridership, previous studies commonly divided study areas into many smaller units, such as ZIP Code Tabulation Areas (ZCTAs) and Census Block Groups (CBGs) (Qian and Ukkusuri, 2015; Yu and Peng, 2019; Liu et al., 2020). In this paper, we segment the study area into a number of 500 m*500 m cells. This division is helpful to smooth out the effects of noise data and provides a reasonable scale to understand the spatial pattern of taxi demand. The selected study area finally contains 97,362 grids. After that, all pick-ups are allocated into the corresponding cell using spatial analysis. Fig. 1 shows the spatial distribution of pick-ups in Shanghai city during a typical weekday (from March 7th to 13th, 2016).

Approximately 89.85% of the daily pick-ups during a typical weekday are distributed within the Outer Ring Road of Shanghai (represented by the red circle). Affected by the urban form and functionality, taxi ridership also tends to reduce gradually from the center to peripheral areas. This is also one of the reasons why Tsungming island (denoted by the black circle) is excluded from this study. Other demand hotspots include the airports, intercity bus stations, and railway stations.

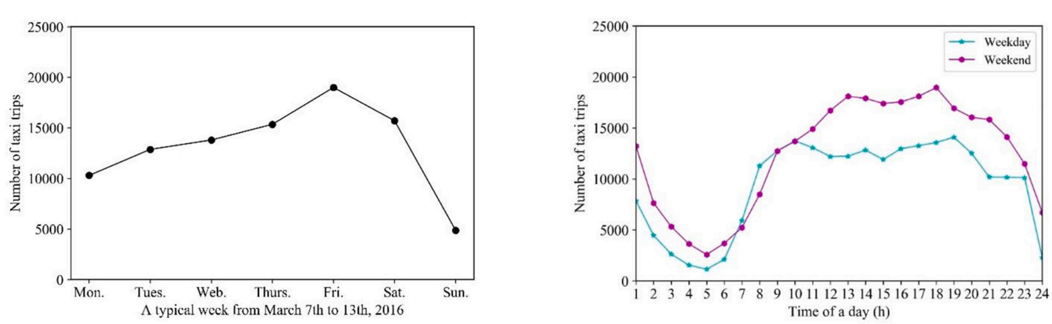
A proper study time span is essential to successfully identify the time-varying factors affecting the taxi demand. Since there is not much variation of weekly taxi trips during March 2016, models in this paper are estimated using one-week data (from March 7th to 13th, 2016). Fig. 2(a) shows the temporal distribution of taxi trips for one-week data. There were about 361 million records, and no special holidays and events are included in the selected week.

Fig. 2(b) shows the taxi trips of a typical weekday and weekend. The peak period of taxi trips happens during 10:00 and 19:00 on weekdays and during 13:00–18:00 on weekends, whereas weekends in general results in higher taxi trips than weekdays. Moreover, to disclose the varying impacts of built environment factors at different times and days, this study divides the daytime into peak hours and off-peak hours. Specifically, the continuous periods 7:00–10:00 and 16:00–19:00 are selected as morning-peak and evening-peak hours separately, while 11:00–12:00 and 20:00–22:00 are treated as off-peak hours. Thus, by aggregating the average hourly taxi trips at different time periods and days into the corresponding cells, six dependent variables (i.e., weekday morning-peak demand, weekday evening-peak demand, weekday off-peak demand, weekend morning-peak demand, weekend evening-peak demand, and weekend off-peak demand) are obtained.

3.2. Explanatory variables

3.2.1. Land use factors

POIs refer to the location of specific points in a geographic information system, and it can, to some extent, define and explain the land use type of a specific area (Li et al., 2019). In this paper, approximately half a million POIs were extracted from Amap in 2016. Based on the intended use, all the POIs could be classified into different categories. Table 1 presents the first and second-level categories provided by the Amap. However, for better application in this study, some categories need to be removed or re-classified. For instance, beauty and vehicle services fall into the living service category. A new category, namely leisure and sports service, is defined to represent both the sports and entertainment services. The education and cultural media are also combined as the culture and educational service in this paper. Concerning the categories of transport facility and facility access, the POIs of



(a) The distribution of taxi trips during a week (b) Variations of taxi trips in a typical day

Fig. 2. The temporal distribution of taxi trips in Shanghai city.

Table 1
The classification and statistics of POIs based on Amap.

No.	First-level categories	Second-level categories	Selected	Cleaned up variable	Data number
1	Catering	Chinese and foreign restaurant, fast food, snack bar, coffeehouse, teahouse, etc.	√	Catering	85,742
2	Hotel	Starred hotel, chain hotel, express hotel, apartment hotel, etc.	√	Hotel	16,631
3	Shopping	shopping center, department store, retail, convenient store, household building mall, open market etc.	√	Shopping	7845
4	Living service	Post office, logistics express, laundry, photo studio, print shop, domestic service, kiosk, pet service, etc.	√	Living Service	46,090
5	Beauty	Skincare, hairdressing, nail salon, beauty salon, etc.	√		
6	Vehicle service	Car sale, maintenance, and beauty, vehicle inspection, car renting, etc.	√		
7	Scenic spot	Park, zoological or botanical park, museum, historic site, church, bathing beach, etc.	√	Scenic Spot	4266
8	Sports	Fitness center, playground, extreme sports center, swimming pool, etc.	√	Leisure and Sports Service	29,202
9	Entertainment	Movie theatre, game center, dance hall, karaoke lounge, resort, foot bath and massage, chess room service, etc.	√		
10	Education	College, high, middle, and primary school, vocational training institution, adult education, library, etc.	√	Cultural and Educational Service	8850
11	Cultural media	Television broadcaster, art gallery, cultural palace, exhibition center, etc.	√		
12	Healthcare	General and special hospital, clinic, emergency center, the center for disease control, etc.	√	Healthcare	19,722
13	Transport facility	Airport, railway station, subway station, bus stop, parking, gas station, bridge, road-side parking space, etc.	#	Transport Hub	–
14	Finance	Bank, ATM, insurance company, financial corporation, pawnshop, etc.	√	Finance	21,005
15	Real estate	Office building, residential community, dormitory, etc.	/		–
16	Corporation	Enterprise, company, factory, industrial park, mine, etc.	√	Corporation	197,494
17	Government	Administrative agency, public security service, organizations concerning foreign affair, party's agency, welfare institution, etc.	√	Government	37,939
18	Facility Access	Access of highway, airport, railway station, parking lot, etc.	#	Transit Accessibility Parking Lot	– 39,291
19	Nature geography	Mountain, island, lake, river, etc.	/	–	–

Note: √ means selecting and using; / represents removing; # is re-classifying.

airports, railway or intercity bus stations, bus stops, metro stations, parking lots are re-classified as new variables. The category of natural geography, referring to the mountains, islands, lakes, and rivers, is removed because it cannot reflect the built environments. It is worth noting that the category of real estate is also discarded because the number of residential communities cannot truly represent the real residential information (i.e., the number of residents). Besides, the office buildings are already represented by the category of corporations. A total of 14 categories of factors are extracted based on the POIs data. The new variables and associated data numbers are presented in Table 1.

3.2.2. *Residence and housing-price factor*

As described, the POIs of real estate are discarded, leading to the loss of residential information. However, the number of residential apartments or houses is crucial in explaining taxi ridership. In this study, the residential information was extracted from the website of Lianjia, one of the largest real estate intermediary companies in Shanghai. The data collection time is, in fact, generally consistent with the collection time of POIs data. The residential data includes the number of communities, buildings, and households. Then, the number of households is employed as the category of residence.

After adding the residential information data, an example of associated data (i.e., the independent variables) around Shanghai Railway Station is shown in Fig. 3. Note that the yellow points in Fig. 3 represent the residential information extracted from Lianjia.com. To have a more explicit description of the variable, transit accessibility, this variable is presented by two different legends: bus stop and subway station. Similarly, the variable, transport hub, is depicted by the intercity bus station, airport, and railway station.

In addition, the average housing-price within an area could reflect the resident income level. It may also have positive effects on taxi ridership. Therefore, the average housing-price data is also extracted, and the distribution in Shanghai and surrounding areas is shown in Fig. 4. It can be noted that the distribution of housing-price conforms to the distribution of urban forms. As a result, the average housing-price is selected as an independent variable in this study to reflect the economic aspect.

3.3. *Descriptive statistics*

In summary, six dependent variables and three categories of explanatory variables are considered in this paper (as shown in Table 2). For the economic aspect, the average housing-price is selected. Intermodal connection variables include transit accessibility, transport hub, and parking lot. Note that transit accessibility and the existence of transport hub are calculated based on ArcGIS software version 10.5 (1 = exist and 0-not). The number of parking lots is dedicated to being extracted directly from the POIs dataset. In addition, 12 land-use related factors are extracted. Table 2 presents the descriptive statistic of all involved variables.

4. Methodology

4.1. *Multicollinearity*

Eliminating variables that have a high correlation with any other explanatory variables is an essential procedure before model estimation (Wheeler and Tiefelsdorf, 2005). In this paper, variable selection is conducted in two steps. First, a Pearson product-moment correlation coefficient (PPMCC) test is performed. Variables with coefficients larger than 0.7 are excluded from the model. Second, multicollinearity between the explanatory variables is also analyzed from the variance inflation factor (VIF), and variables with VIF greater than ten are eliminated.

4.2. *Spatial autocorrelation*

Another issue on variable selection is the non-stationary spatial effects. This effect is a critical assumption neglected in global models. Because of the functionality of urban forms, the spatial patterns of both dependent and exploratory variables have internally spatial autocorrelations. Several models in previous studies have been developed to assess spatial autocorrelation, the best known of which is the Global Moran' I test (see Moran, 1950 for more details). In this paper, this type of test method is employed to statistically validate the spatial

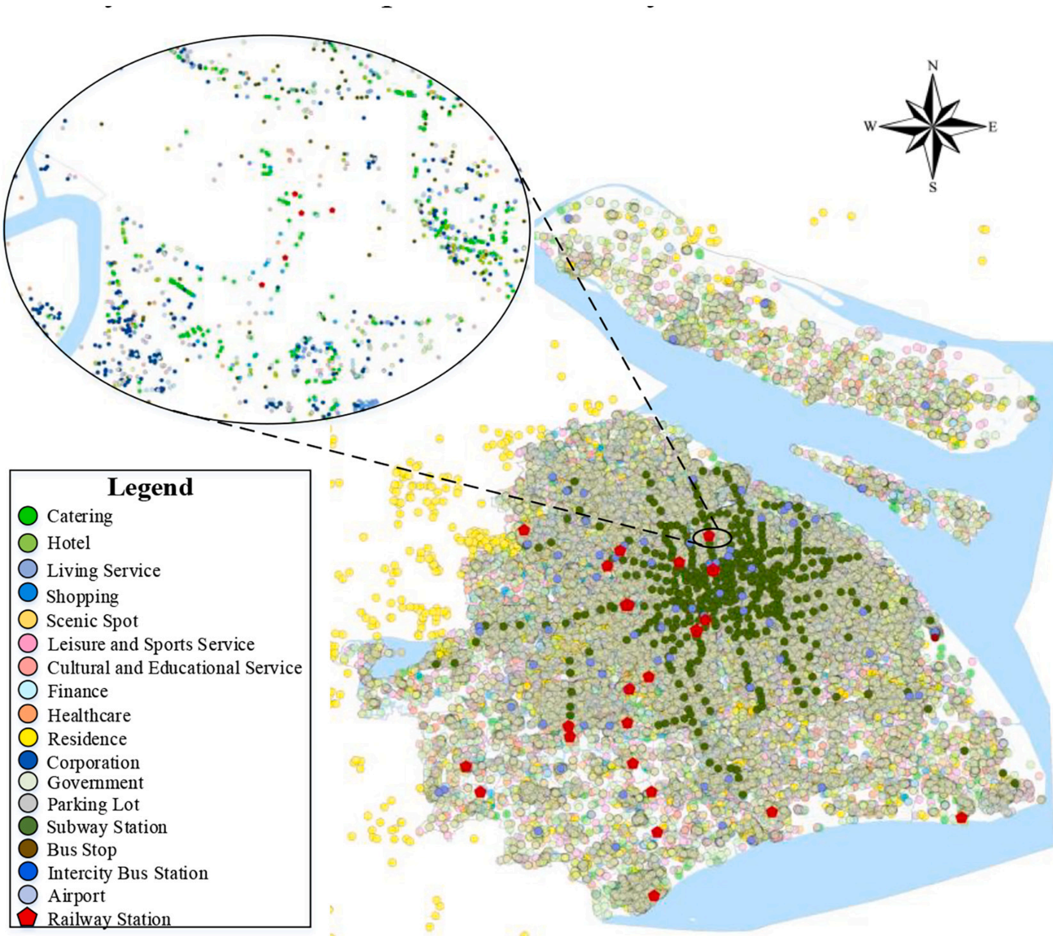


Fig. 3. An example of POIs around Shanghai Railway Station.

autocorrelation of the 22 variables. The corresponding mathematical expression is shown as follows.

$$I_k = \frac{N_k}{\sum_{i=1}^{N_k}\sum_{j=1}^{N_k} w_{k,ij}} \cdot \frac{\sum_{i=1}^{N_k}\sum_{j=1}^{N_k} w_{k,ij}\left(x_{k,i}-\overline{X_k}\right)\left(x_{k,j}-\overline{X_k}\right)}{\sum_{i=1}^{N_k}\left(x_{k,i}-\overline{X_k}\right)^2}, (i \neq j) \tag{1}$$

where $x_{k,\ i}$ and $x_{k,\ j}$ are the observations i and j of the kth variable. $\overline{X_k}$ is the mean value of the kth variable. N_k is the total number of observations (i.e., spatial cells). $w_{k,\ ij}$ is an element of the spatial weight matrix, which describes the spatial relationship between $x_{k,\ i}$ and $x_{k,\ j}$. In this paper, the spatial weight matrix is calculated based on the distance decay method (Greicius et al., 2003). The null hypothesis of Moran's I test assumes no spatial autocorrelation between explanatory variables. In other words, the Moran's I index of each variable is close to zero. *Z*-score value is also used as the indicator of the significant of the Moran's I index, and it can be calculated by the following equation.

$$Z(I_k) = \frac{I\text{-}E(I_k)}{\sqrt{Var(I_k)}} \tag{2}$$

where $E(I_k)$ and $Var(I_k)$ are the expectation and the standard deviation of the Moran's I index, respectively. The significance level in this paper is set as $p < 0.05$.

4.3. Semi-parametric geographically weighted Poisson regression (SGWPR)

As shown in Section 3.1, the taxi ridership in Shanghai is concentrated within the Inner Ring Road (denoted by the red circle), indicating that a small part of the city area generates most of the taxi demands. The distribution of taxi trips is also highly skewed (as shown in Fig. 5(a)), which is against the normality assumption of the ordinary least square (OLS) regression model. The log transformation method is a common method used in previous studies to improve the interpretability (Qian and Ukkusuri, 2015). However, in this paper, the distribution of taxi trips with log transformation still cannot satisfy the normal distribution (Fig. 5(b)).

Therefore, the model developed for analyzing the determinants of taxi ridership should address the concerns of both autocorrelation and heteroscedasticity. By assuming taxi trips follow a Poisson distribution, the SGWPR model is employed in this study, and the model can be mathematically expressed by the following equation.

$$y_i^t = \sum_k \beta_k^t\left(u_i^t, v_i^t\right)\cdot x_{k,i}^t + \sum_l \gamma_l^t \cdot z_{l,i}^t + \varepsilon_i^t \tag{3}$$

where y_i^t, $x_{k,\ i}^{\ t}$, $z_{l,\ i}^{\ t}$, and ε_i^t are, respectively, dependent variable, k th explanatory variable with a geographically varying parameter β_k^t, l th explanatory variable with a fixed parameter γ_l^t, and Poisson error of the observation i in the given time period t. (u_i^t, v_i^t) represents the geographic coordinate of the i th observation in the time period t. The

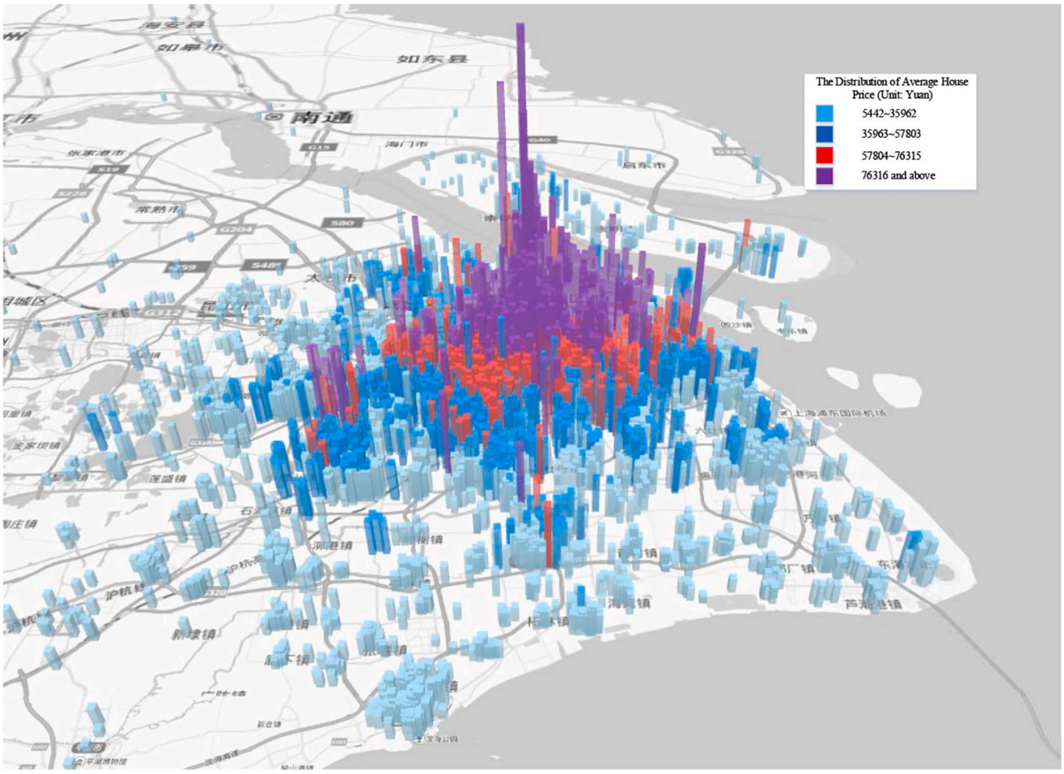
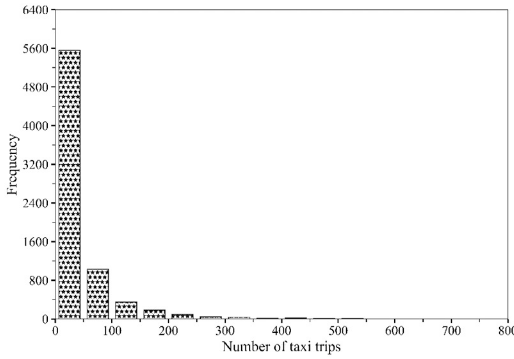


Fig. 4. The distribution of average housing-price in Shanghai and surrounding areas.

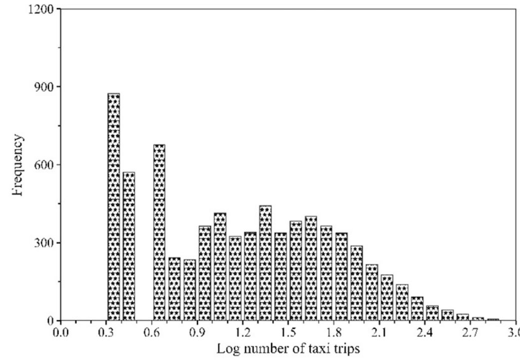
Table 2
Variable description and summary statistics.

Variables	Description	Min [b]	Mean [b]	Max [b]	IQR [a, b]	Source
Dependent Variable						
Taxi demand						
Weekday	Average demand for weekdays	[3,1,1]	[14,13,15]	[155,246,314]	[16,14,16]	SHQS [c]
Weekend	Average demand for weekends	[1,1,1]	[11,14,15]	[151,329,297]	[14,14,17]	SHQS [c]
Independent Variable						
Economic aspect						
Housing-price	Average price of houses and apartments.	9909	59,941	189,509.8	23,324.83	Lianjia.com
Intermodal connection						
Transit accessibility	Whether there exists bus stops or metro stations (1 for exist, 0 for not).	0	N/A	1	N/A	Amap
Transport hub	Whether there exists airports, railway or intercity bus stations (1 for exist, 0 for not).	0	N/A	1	N/A	Amap
Parking lot	No. of parking lots.	0	4	41	6	Amap
Land Use Factors						
Catering	No. of restaurants, bars, etc.	0	6	135	14	Amap
Corporation	No. of enterprises, companies, etc.	0	6	345	14	Amap
Finance	No. of banks, financial corporations, etc.	0	1	174	4	Amap
Shopping	No. of retails, shopping centers, etc.	0	1	13	1	Amap
Cultural and educational service	No. of cultural palaces, schools, etc.	0	0	31	1	Amap
Living service	No. of post offices, beauty salons, and car maintenances, etc.	0	5	70	9	Amap
Leisure and sports service	No. of fitness centers, movie theatres, etc.	0	2	35	3	Amap
Residence	No. of households.	0	900	12,922	1341	Lianjia.com
Healthcare	No. of hospitals, clinics, etc.	0	1	52	3	Amap
Government	No. of administrative agencies, etc.	0	2	70	6	Amap
Hotel	No. of hotels.	0	1	61	3	Amap
Scenic Spot	No. of museums, historic sites, etc.	0	2	26	4	Amap

[a] IQR is the abbreviation of Interquartile Range.
[b] Values in bracket represent the ridership in morning-peak, evening-peak, and off-peak hours, respectively.
[c] SHQS is the abbreviation of Shanghai Qiangsheng Taxi company.



(a) Taxi ridership without data transformation



(b) Taxi ridership with Log Transformation

Fig. 5. The distribution of taxi ridership.

Table 3
Moran's I test result for candidate variables.

Variables	Moran's index	Expected index	*z*-score	*p*-value
Weekday Morning-Peak demand	0.456	−0.0002	57.379	0.000
Weekday Evening-Peak demand	0.416	−0.0003	50.120	0.000
Weekday Off-Peak demand	0.470	−0.0002	57.826	0.000
Weekend Morning-Peak demand	0.383	−0.0002	48.144	0.000
Weekend Evening-Peak demand	0.385	−0.0002	48.313	0.000
Weekend Off-Peak demand	0.430	−0.0002	53.667	0.000
Transit accessibility	0.007	−0.0002	0.906	0.365
Transport hub	0.030	−0.0002	3.734	0.001
Parking lot	0.365	−0.0002	44.799	0.000
Housing-price	0.679	−0.0002	83.335	0.000
Residence	0.104	−0.0002	12.791	0.000
Catering	0.191	−0.0002	23.498	0.000
Scenic spot	0.293	−0.0002	36.543	0.000
Corporation	0.309	−0.0002	38.021	0.000
Shopping	0.061	−0.0002	7.515	0.000
Finance	0.313	−0.0002	39.135	0.000
Cultural and educational service	0.181	−0.0002	22.299	0.000
Living service	0.240	−0.0002	29.540	0.000
Leisure and sports service	0.189	−0.0002	23.190	0.000
Healthcare	0.143	−0.0002	17.597	0.000
Government	0.279	−0.0002	34.252	0.000
Hotel	0.351	−0.0002	43.376	0.000

estimated coefficient, $\beta_k^t(u_i^t, v_i^t)$, is the varying conditional of observation i.

Notice also that the dependent variable should be an integer greater than or equal to zero and follows a Poisson distribution (as shown in Eq. (3)).

$$y_i^t \sim Poission\left[M_i^t \cdot exp\left(\sum_k \beta_k^t\left(u_i^t, v_i^t\right)\cdot x_{k,i}^t + \sum_l \gamma_l^t \cdot z_{l,i}^t\right)\right] \tag{4}$$

where M_i^t is the offset variable of the observation i in the given time period t. Previous studies often use the size of population at risk or the expected size of outcome in spatial epidemiology as the term; however, in this paper, for simplicity, M_i^t is set as 1.0 for all time periods and observations.

Then, the global and local coefficients are estimated by a combination of a parametric and non-parametric method, in which global coefficients are obtained using ordinary least square while local coefficients are estimated using weighted least square (Nakaya et al., 2005; Mar'ah et al., 2017). Based on the principle that close-distance observations have higher weighting values, the local coefficients for the given time period t, $\beta_k^t(u_i^t, v_i^t)$, could be obtained by the following equation.

$$\widehat{\boldsymbol{\beta}}\left(u_i^t, v_i^t\right) = \left(X^T W\left(u_i^t, v_i^t\right) X\right)^{-1} X^T W\left(u_i^t, v_i^t\right) Y \tag{5}$$

where X and Y are, respectively, the explanatory and dependent variable matrix. X^T is the transpose of the explanatory variable matrix. $W(u_i^t, v_i^t)$ represents the spatial weighting matrix of observation i for the given time period. In this matrix, the diagonal elements are determined by the distance of observation i with other observations and off-diagonal elements are all 0. Several kernel functions have been developed in prior studies to define the spatial weighting matrix, such as Gaussian and bi-square functions. Besides, the bandwidth involved in the kernel functions is also crucial in the model estimation. In this paper, the adaptive bi-square kernel function is selected, and its mathematical expression is shown as follows.

$$w_{ij}^t = \begin{cases} \left(1 - \left(d_{ij}^t\right)^2 \Big/ \left(\theta_{i(k)}^t\right)^2\right) & d_{ij}^t \le \theta_{i(k)}^t \\ 0 & d_{ij}^t > \theta_{i(k)}^t \end{cases} \tag{6}$$

where w_{ij}^t is the weight value of observation j for the coefficient of observation i in the given time period t. d_{ij}^t is the Euclidean distance between observation i and j in the given time period t. $\theta_{i(k)}^t$ is an adaptive bandwidth size defined as the k-th nearest neighbor distance (Fotheringham et al., 2003). To find the optimal bandwidth size, a golden section search method with the objective to minimize the corrected Aikake Information Criterion (AICc) is adopted. It should be noted that the method, AICc, as well as the conventional Akaike information criterion (AIC), are also selection methods used to measure the goodness of fit of the SGWPR model. For further information on the SGWPR model, interested readers could refer to Nakaya et al. (2005).

5. Results

5.1. Results of global Moran's I test

The results of Moran's I test are shown in Table 3. It can be observed that almost all variables are statistically significant at the 0.01 level except for transit accessibility. The positive *z*-score values indicate that the data are spatially clustered. However, compared with the *z*-score values of other explanatory variables, transport hub has a relatively low value. Therefore, in this paper, transit accessibility and transport hub, are set as variables with fixed coefficients, while the remaining 14 variables are set as global variables with geographically varying parameters.

Table 4
Estimation results of the global models (i.e. GLMs).

Variable	Weekday			Weekend			VIF
	Morning-Peak	Evening-Peak	Off-Peak	Morning-Peak	Evening-Peak	Off-Peak	
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	
Intercept	2.446**	2.299**	2.407**	2.223**	2.369**	2.449**	–
Parking Lot	0.123**	0.144**	0.157**	0.091**	0.113**	0.134**	1.962
Housing-price	0.294**	0.292**	0.316**	0.263**	0.302**	0.309**	1.212
Residence	0.119**	0.030**	0.018**	0.161**	0.082**	0.084**	1.061
Catering	0.032**	0.069**	0.068**	0.042**	0.081**	0.070**	2.352
Scenic Spot	0.011**	0.033**	0.039**	−0.005*	0.034**	0.029**	1.156
Corporation	0.041**	0.114**	0.114**	0.009*	0.050**	0.056**	2.133
Shopping	0.038**	0.093**	0.087**	0.059**	0.098**	0.101**	1.501
Finance	0.027**	0.031**	0.024**	−0.006*	0.018**	0.013**	1.963
Cultural and Educational Service	0.032**	0.040**	0.035**	0.032**	0.039**	0.039**	1.079
Living Service	0.084**	0.070**	0.065**	0.099**	0.099**	0.085**	2.087
Leisure and Sports Service	0.002	0.007*	0.018**	0.017**	0.045**	0.037**	1.862
Healthcare	0.113**	0.080**	0.086**	0.094**	0.076**	0.071**	1.259
Government	0.039**	−0.018**	−0.005*	0.028**	−0.015**	−0.001	1.362
Hotel	0.027**	0.038**	0.040**	0.003	0.016**	0.028**	1.431
Transit accessibility	0.150**	0.153**	0.151**	0.136**	0.132**	0.139**	1.049
Transport Hub	0.062**	0.079**	0.070**	0.065**	0.068**	0.078**	1.006

** Significant at 0.01 level.
* Significant at 0.05 level.

Table 5
Estimation results of the SGWPR morning-peak model for local variables.

Variable	MIN		MAX		MEN		LQ		UQ	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Intercept	−0.270	−0.577	3.937	3.519	2.326	2.157	1.959	1.818	2.833	2.649
Parking lot	−1.716	−0.785	1.825	1.743	0.037	0.069	−0.080	−0.043	0.184	0.156
Housing-price	−1.132	−1.364	1.241	0.794	0.044	0.005	−0.124	−0.139	0.189	0.150
Residence	−0.582	−0.360	0.931	0.645	0.070	0.104	−0.022	0.021	0.160	0.179
Catering	−1.322	−1.677	1.224	0.981	0.073	0.053	−0.064	−0.054	0.203	0.173
Scenic Spot	−2.138	−1.845	1.660	1.061	−0.052	−0.056	−0.214	−0.189	0.106	0.107
Corporation	−3.086	−2.588	2.136	1.128	−0.152	−0.196	−0.336	−0.365	0.048	0.011
shopping	−0.816	−0.519	0.860	0.789	0.037	0.039	−0.083	−0.061	0.145	0.124
Finance	−1.619	−1.423	2.851	2.262	0.239	0.186	0.044	−0.010	0.411	0.355
Cultural and educational Service	−0.903	−0.956	1.023	0.637	−0.024	−0.033	−0.122	−0.110	0.075	0.064
Living Service	−1.457	−0.455	1.236	1.202	0.109	0.152	−0.019	0.026	0.227	0.252
Leisure and sports Service	−0.999	−0.799	1.062	0.801	−0.024	−0.015	−0.133	−0.100	0.103	0.073
Healthcare	−1.181	−0.858	1.743	1.185	0.117	0.103	−0.008	0.012	0.235	0.188
Government	−0.992	−0.982	1.811	0.918	0.056	0.025	−0.055	−0.069	0.141	0.093
Hotel	−1.237	−1.294	1.343	1.015	0.122	0.128	−0.032	−0.007	0.276	0.261

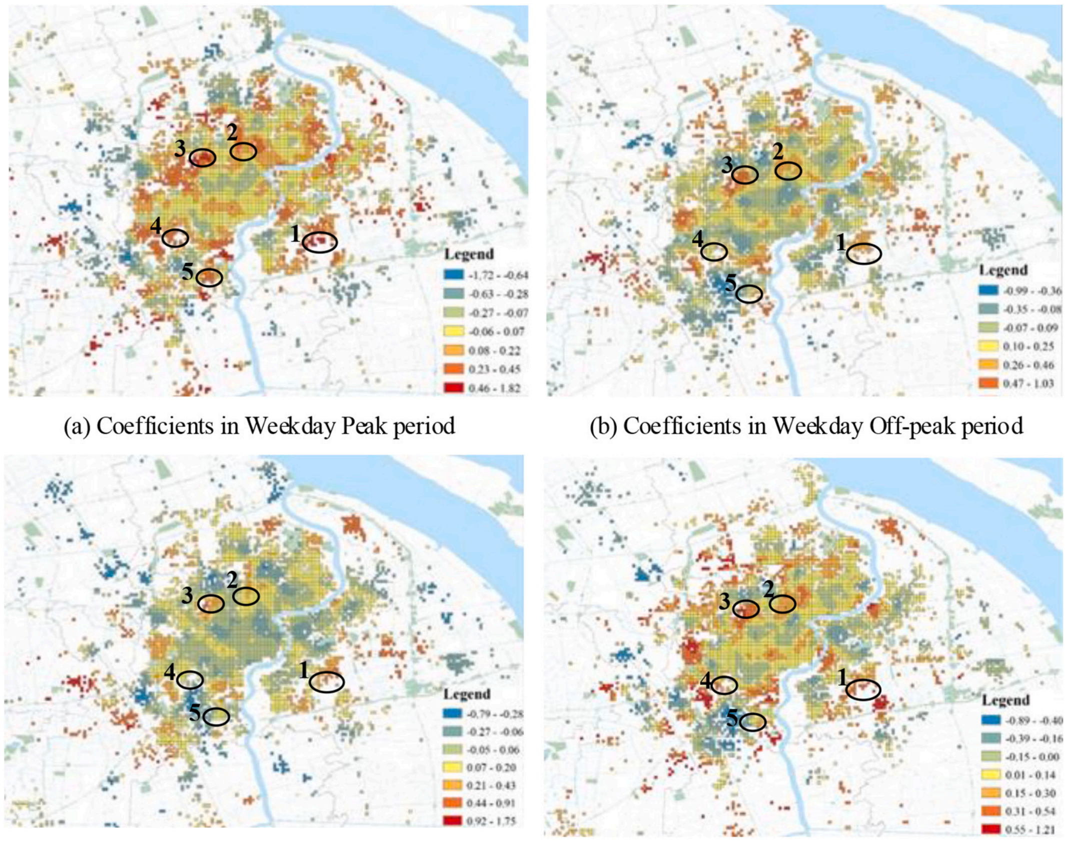
Table 6
Estimation results of the SGWPR models for global variables.

Variable	Morning-Peak				Evening-Peak				Off-Peak			
	Weekday		Weekend		Weekday		Weekend		Weekday		Weekend	
	Coeff.	*z*-Value	Coeff.	*z*-Value	Coeff.	*z*-Value	Coeff.	*z*-Value	Coeff.	*z*-Value	Coeff.	*z*-Value
TA	0.184	25.268	0.166	2.974	0.185	26.052	0.174	4.156	0.178	5.078	0.176	27.745
TH	0.035	4.320	0.040	6.445	0.049	7.883	0.045	8.735	0.035	7.341	0.041	7.340

Note: TA means Transit Accessibility; TH represents Transport Hub.

Table 7
Comparison between global and local models.

Metrics	Morning-Peak				Evening-Peak				Off-Peak			
	Weekday		Weekend		Weekday		Weekend		Weekday		Weekend	
	GLM	SGWPR	GLM	SGWPR	GLM	SGWPR	GLM	SGWPR	GLM	SGWPR	GLM	SGWPR
AIC	32,049	11,998	24,542	10,589	29,536	10,509	35,026	11,787	37,204	11,654	38,446	12,058
AICc	32,049	14,194	24,542	12,085	29,537	12,500	35,026	13,876	37,205	13,890	38,446	14,376
BIC/MDL	32,157	22,781	24,650	19,885	29,643	20,149	35,134	22,310	37,312	22,012	38,554	23,085
LL	−16,007.6	−5152.1	−12,254.3	−4564.3	−14,751.5	−4486.1	−17,496.2	−5066.2	−18,585.5	−5005.6	−19,206.3	−5163.3
R-square	0.486	0.862	0.404	0.813	0.544	0.885	0.509	0.881	0.563	0.902	0.518	0.892



(a) Coefficients in Weekday Peak period

(b) Coefficients in Weekday Off-peak period

(c) Coefficients in Weekend Peak period

(d) Coefficients in Weekend Off-peak period

Fig. 6. Spatial-and-temporal distribution for the coefficients of Parking Lot.

5.2. The overall results of the SGWPR and GLM models

To validate the consideration of spatial non-stationarity, a non-spatial model, served as a base model for the SGWPR model, is also estimated. In this paper, the generalized linear model (GLM) is selected as the global model, and its estimation is implemented in Python language on Jupyter Notebook 6.0.1 (Seabold and Perktold, 2010). Table 4 shows some intuitive results of the GLMs to understand the relationship between built environment factors and taxi ridership in different periods. The Variance Inflation Factor (VIF) of each explanatory variable is also provided. The VIF values for all 16 variables are close to 1 and much less than 10. The results give further verification that no collinearity is presented, and these candidate variables are reasonably selected.

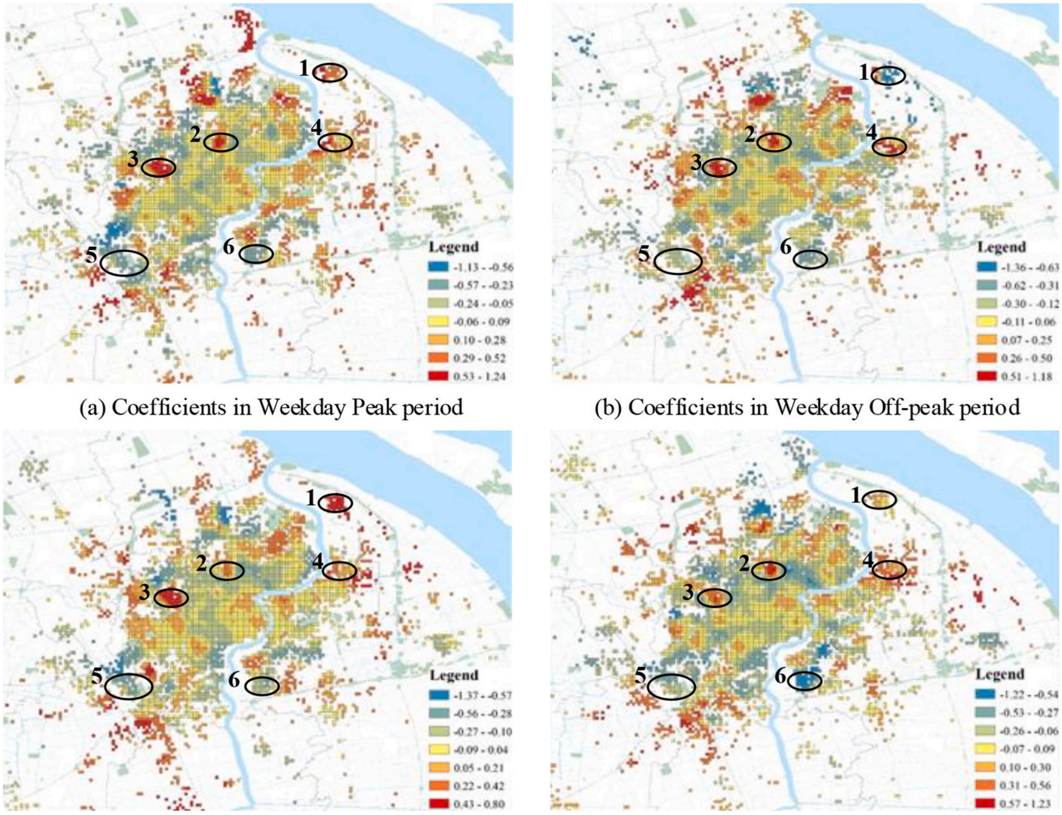
The signs and magnitudes of most coefficients are also consistent with the research hypothesis, indicating that built environment factors significantly contribute to taxi ridership. For instance, there is a positive correlation between the number of residences and taxi demand. It seems a bit surprising that the variables of the scenic spot, finance, and government in several GLMs are found to be significantly negative affecting the taxi demand. Based on shared knowledge and experience, the results of GLMs could be challenged.

With the same explanatory variables, the SGWPR models are also estimated to further analyze the impacts of built environment factors. These SGWPR models were estimated using the MGWR 1.0 software (Fotheringham et al., 2017). The results are reported in Table 5 (for explanatory variables with geographically varying parameters) and Table 6 (for explanatory variables with fixed coefficients). Note that five statistical variables, namely, minimum value (MIN), maximum value (MAX), mean (MEN), lower quartile (LQ), and upper quartile (UQ), are selected in this paper to describe the non-stationary spatial effect of each explanatory variable. The results of the SGWPR morning-peak model are provided in Table 5. Other results of SGWPR evening-peak and off-peak models could be found in Tables A2 and A3 in Appendix A.

As shown in Table 5, the calibrated coefficients of explanatory variables vary in the study area, validating the SGWPR model's capability to address the existence of spatially non-stationary. In terms of coefficient signs, all variables have mixed estimates. Besides, as suggested by the LQ and UQ, most of the estimates of residence, living service, and healthcare during morning-peak hours on weekends are positive. Combining with the results in Tables A2 and A3, the effects of finance during different periods are also mainly positive. Furthermore, according to the MEN, LQ, and UQ, land-use factors, except for residence, demonstrate relatively stable influences during morning-peak hours on both weekdays and weekends. In contrast, the remaining three variables (e.g., parking lot, housing-price, and residence) show some moderate differences between weekdays and weekends. However, these findings cannot be found in the non-spatial model (i.e., the GLM in this paper).

The correlations between taxi ridership and two global variables (e.g., transit accessibility and transport hub) are also summarized in Table 6. In general, the variables of transit accessibility and transport hub in all models are positively correlated with taxi ridership, regardless of time period and day. The findings are in line with several previous studies (Yang and Gonzales, 2014; Yang et al., 2018; Zhang et al., 2019; Sun and Ding, 2019; Liu et al., 2020). Furthermore, compared to the estimation results in Table 6, the two variables demonstrate constant influences at different times of the day.

Table 7 provides fit statistics for all global and local models in different periods and days, in terms of AIC, AICc, Bayesian Information Criterion (BIC), Minimum Description Length (MDL), Log-Likelihood (LL), and R-square. It is shown that, for each period, the SGWPR model outperforms GLM inf terms of R-square. Taking the weekday morning-peak model as an example, the R-square values increase from



(a) Coefficients in Weekday Peak period

(b) Coefficients in Weekday Off-peak period

(c) Coefficients in Weekend Peak period

(d) Coefficients in Weekend Off-peak period

Fig. 7. Spatial-and-temporal distribution for the coefficients of housing-price.

0.486 in the GLM to 0.862 in the SGWPR model. The 0.376 improvement in the amount of variation explained is quite striking. The AICc values for all SGWPR models are also clearly smaller than those of the GLMs. The reductions in these AICc values further prove the superior of the SGWPR model over the global model. Therefore, it could be concluded that explanatory power could be significantly improved by considering the spatial information, and the detailed results of the SGWPR model will be analyzed in the next section.

5.3. Local estimates

In this section, the spatial-temporal variations of several key variables using their average values are analyzed and presented in Figs. 6 to 10. These selected variables cover the three categories of explanatory variables (see Section 3.3), and their associations with taxi ridership are not fully revealed in the existing literature. Through the pre-analysis, we found that the spatial-temporal variations in the morning-peak period models share some similarities with the evening-peak period models' results. Thus, for the analysis's brevity, the results of the evening-peak period models will not be discussed in detail in this paper. Therefore, the morning-peak period models are referred to as peak period models in later analyses for simplicity.

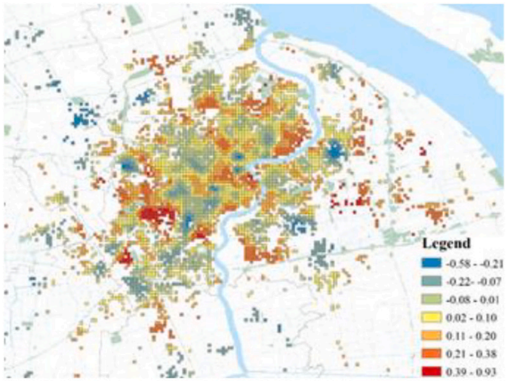
5.3.1. Intermodal connection variables

Fig. 6 presents the effects of parking on predicting taxi ridership for different time periods and days. It can be observed that most of the estimates in the weekday peak period model and weekend off-peak period model are positive except for some areas in central Shanghai and peripheral regions. The magnitudes of the relationship become smaller or even negative in the other two models. The positive impact of parking in the weekday peak period model may be attributed to that, in Shanghai, most parking lots are located in the basements of community-dwellings and office towers, indicating that taxi is one of the primary traffic modes for commuters. The results also provide indirect evidence that there is a link between the number of resident populations with taxi demands. Another reason could be that it is costly for people who live in suburban areas to park their cars at the final destinations. As a result, they are more willing to select a park-and-ride station (e.g., region 1), which usually locates around major public terminals, and take a taxi or other travel modes to satisfy their travel demands.

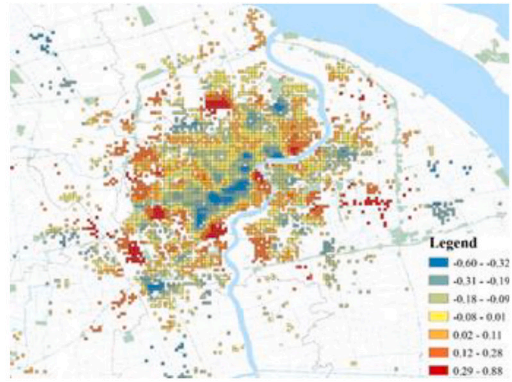
In contrast, the possible explanation of the positive impact of parking lot during weekend off-peak periods is that Shanghai is a strong tourist attraction for both local and non-local residents. Due to the restricted vehicle-entry regulations, tourists from outside Shanghai need to park their cars outside of the Inner Ring Road and take other travel modes to enter the city center. The positive estimates in regions 2–5 (e.g., the four transport hubs) confirm that assumption.

5.3.2. Economic-aspect variable

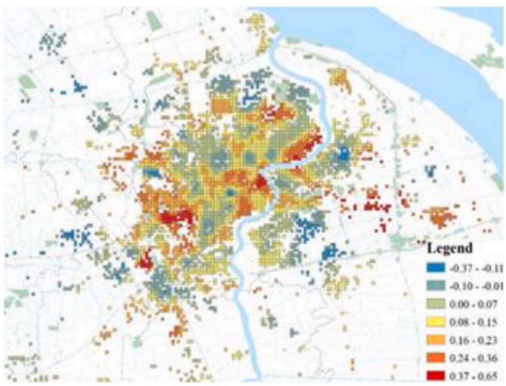
The four plots, as shown in Fig. 7, illustrate the relationship between housing-price and taxi ridership. It can be observed that coefficient estimates for all models display a similar distribution pattern. The impact of housing-price on taxi demands could be either positive or negative, depending largely on the geographical locations. As a proxy of economic indicators, housing-price could, to a certain degree, reflect the households' real spending power. Intuitively, people with higher spending power (i.e., those who live in the areas with higher housing-price) are more likely to take taxis. However, the estimates are negative for all models in some peripheral areas (as regions 5 and 6 in Fig. 7). The negative link in these areas may be attributed to the fact that suburban regions tend to have more open spaces and fewer congestions. For instance, many private villas are located in region 5, people who live here usually own their private vehicles, and thus reducing the reliance on taxi service.



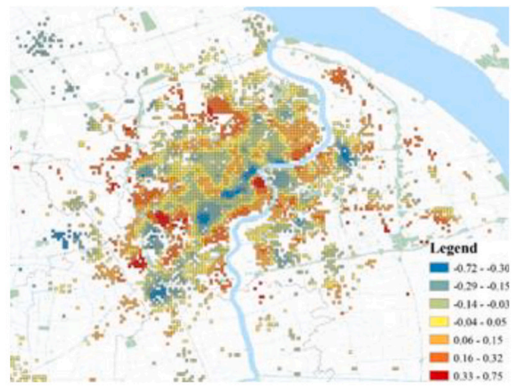
(a) Coefficients in Weekday Peak period



(b) Coefficients in Weekday Off-peak period



(c) Coefficients in Weekend Peak period



(d) Coefficients in Weekend Off-peak period

Fig. 8. Spatial-and-temporal distribution for the coefficients of Residence.

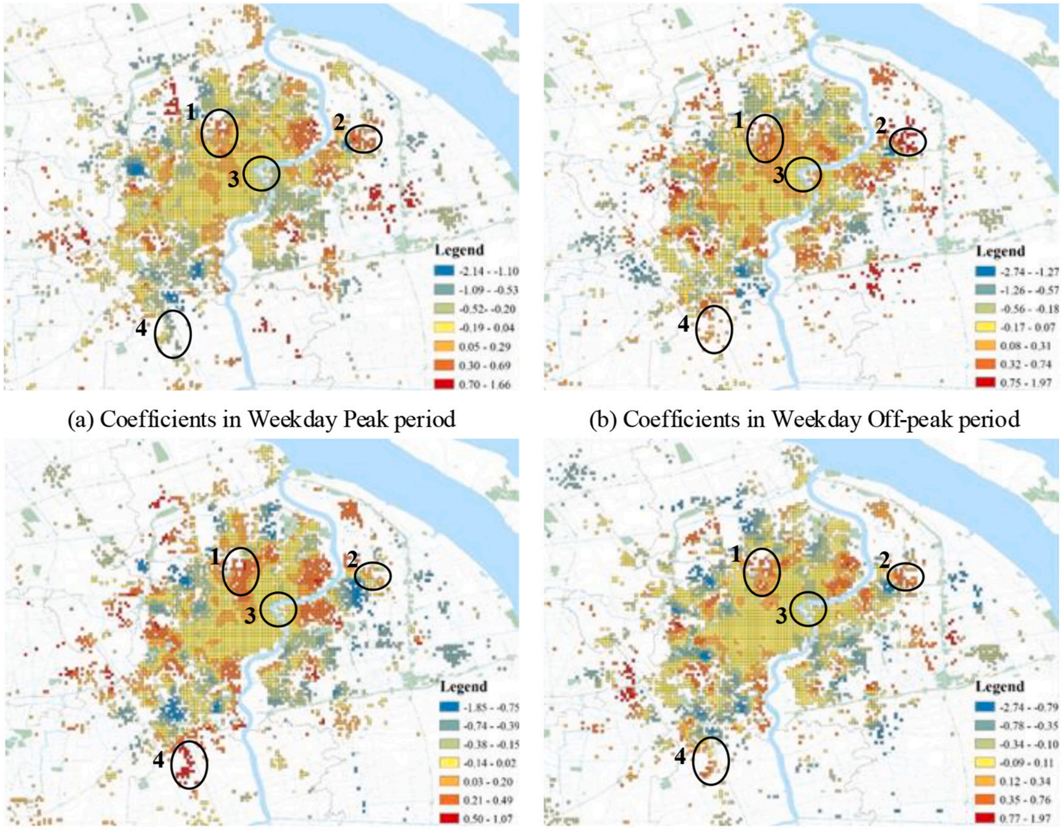
On the contrary, the coefficients are positive in some north and west suburban regions, e.g., regions 2, 3, and 4. Due to the intensive activities in these regions (e.g., a critical railway station, Shanghai West Railway Station, is situated in region 2, and both regions 3 and 4 contain a busy metro transfer station), the utility of taking taxis increases. Moreover, in some areas, the variable of housing-price illustrates a composite impact on taxi demands. For instance, the housing-price coefficients in region 1 are positive in all periods except for the off-peak hours on weekdays. The reason may be that it is not a cost-effective way of driving to work for the long distance between region 1 and the central city. Because a terminal of Metro line 6 is located close to the area, people with higher spending power are more willing to take a taxi as the first- and last-mile connection mode and then take the public transport to reach their final destinations.

5.3.3. *Selected land-use factors*

Based on the dataset, the residential factor has a significant contribution to taxi ridership (especially for peak hours), which is reasonably consistent with most previous studies. For peak period models, the spatial effect of the residential factor on taxi ridership is similar. The number of residential apartments is positively correlated with taxi demand in the study area, except for the south and east suburban areas. This positive effect is consistent with the result of Yu and Peng (2019), who found that the residential factor plays a crucial role in taxi demand during peak periods. On the contrary, the coefficients are negative in the city center in the off-peak periods. The negative effect may be explained by the fact that people who live in the city center usually have a higher income and own their private vehicles, thus reducing the chance of taking a taxi in off-peak hours.

The partial dependence plots for the scenic spot in different time periods and days are shown in Fig. 9. Remarkably, in some peripheral regions (e.g., regions 1 and 2), the scenic spots are found to be significantly positive in explaining taxi ridership, even on weekdays. The results indicate that, apart from those common factors considered in previous studies (e.g., residential, office, and commercial land-use factors), the impact of scenic spots on taxi ridership can also not be ignored. It may because Shanghai is one of the most attractive tourist cities in China; a substantial proportion of taxi ridership is contributed by non-local travelers, regardless of whether on weekdays or weekends. In contrast, the coefficients are smaller and become even negative in region 3 (i.e., the central business district, CBD) compared with other areas, indicating that higher scenic spot density leads to less taxi demand. The results may be attributed to the dense and pedestrian-friendly street network, as well as the well-connected public transport system. Despite taxis, there are more means of transport available to people traveling around in this region. By comparing the estimates in region 4, it can be found that the effects of the scenic spot on weekends are more serious than on weekdays. One possible reason is that, according to our POIs dataset, the variable of scenic also includes spots parks, historic sites, and museums that may attract both non-local travelers and natives. Because of the involvement of natives, the coefficient of the scenic spot has an outstanding performance on weekends.

The impacts of corporation on taxi ridership are summarized in Fig. 10. Most of the coefficients estimated in weekday models are negative. The relationship is more apparent in the west bank of the Huangpu River where the subway network is extremely intensive. There are also some exceptions for the weekday models that the estimates are positive in downtown. A possible explanation for the result is that many corporations are situated in areas that are not walkable from metro stations. Taxi served as the first- and last-mile connection mode could empower commuters to easily reach their destination. Interestingly, compared to the estimations of weekday models, the coefficients in the weekend peak period model exhibit a different distribution pattern. As shown in Fig. 10(c), the impact of the number of corporations on taxi



(a) Coefficients in Weekday Peak period

(b) Coefficients in Weekday Off-peak period

(c) Coefficients in Weekend Peak period

(d) Coefficients in Weekend Off-peak period

Fig. 9. Spatial-and-temporal distribution for the coefficients of Scenic Spot.

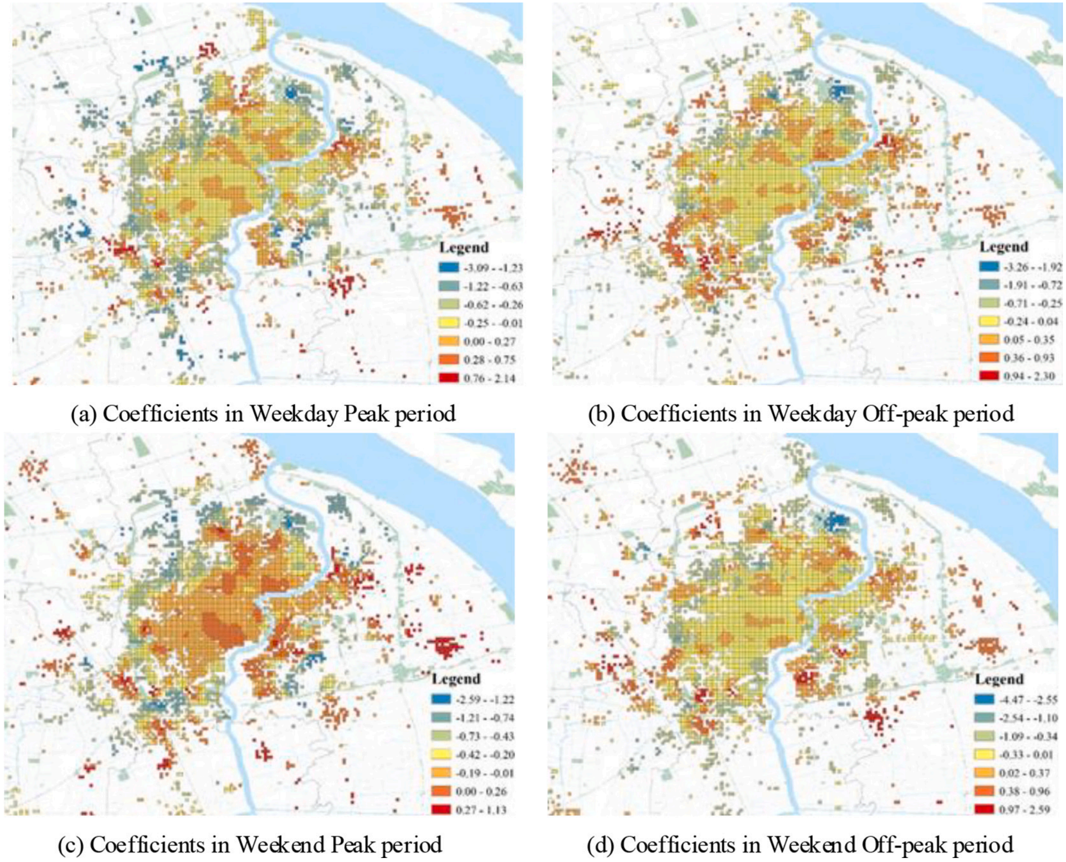
demand is remarkable and positive in the central city. This finding is inconsistent with our shared knowledge and experience that the trip demands for working should be much lower on weekends for most commuters. The reason may be that the utility of using a taxi varies for different periods. For example, in normal working weekday, driving or public transport, such as subway and conventional buses, are more reliable and preferred for commuters. However, many high-tech companies in the central city frequently require staff to work overtime on Friday night. After an 'on-the-night' working, compared to other traffic modes, commuters would like to choose a convenient, fast, and comfortable travel mode to go home and get a rest; thus, the utility of using taxis is highly increased.

6. Conclusions

As a viable option for promoting greener travel, taxis have been attracting ridership increasingly in recent years. However, the potential taxi demand in urban areas remains a cause of concern for taxi drivers and regulators. In order to assist policy makers in framing better management strategies, it is necessary to explore and comprehend the influential factors related to taxi ridership. This paper is one of the pioneers to examine the influences of built environment on taxi ridership using multi-source heterogeneous data. The results of this paper aim to provide policy and planning recommendations for both operators and regulators to improve their operations.

To this end, this paper first analyzed the spatial-temporal distribution of taxi trips by exploring taxi GPS data over a month in the city of Shanghai, China. The analysis shows that taxi ridership is unevenly distributed in space and presents variations in the time dimension. To understand the impacts of the full range of built environment factors, multi-sourced data, including POIs and housing-price, were acquired through open APIs of Amap and Lianjia.com. After data cleaning, a total of 16 categories of factors were defined. Next, a serial of SGWPR models and generalized linear models were built to examine the spatial-temporal influence of built environment on taxi ridership based on the multisource data. Results show that the SGWPR models have a better performance in terms of model fit than the generalized linear models. Besides, by visualizing the coefficients at the level of 0.5 km grid, the impacts of built environment factors are found to be highly heterogeneous in different city areas and time periods. It is found that, when considering spatial and temporal variations, all built environment factors have positive or negative effects on taxi ridership.

These findings yield several important policy implications for taxi drivers and regulators to improve the service level of the urban taxi system. First, the density of scenic spots is positively correlated with taxi usage in some suburban areas, especially on weekends. At the same time, it impedes its usage in the central urban area. Therefore, it is recommended that taxi drivers should pay more attention to their



(a) Coefficients in Weekday Peak period (b) Coefficients in Weekday Off-peak period

(c) Coefficients in Weekend Peak period (d) Coefficients in Weekend Off-peak period

Fig. 10. Spatial-and-temporal distribution for the coefficients of Corporation.

potential passengers in scenic spots. Besides, policy makers should also design and promote special-purpose tourist taxi services and consider the diverse taxi demands for non-local travelers.

Second, it is common that taxis are oversupplied in city centers and underserved in suburban areas. To avoid the spatial-temporal mismatch between taxi demand and supply, taxi operators could design space-time-dependent pricing strategies to adjust taxi drivers' behavior. For instance, through incentive policies, they can encourage taxi drivers to serve areas with low demand. Conversely, they can also push a low price as punishment to make taxi drivers circulate away from areas with heavy traffic. More meaningfully, the space-time-dependent pricing strategies could reduce the waiting time for people living in suburban areas, increase the opportunities for taxi drivers to reach passengers, and consequently improve the general service level of the urban taxi system.

Third, we found the impacts of built environment factors on taxi demand are highly heterogeneous. This is in line with that findings that urban areas with dense populations are associated with a high demand for travel (Yang and Gonzales, 2014; Liu et al., 2020). These results indicate the importance of balanced and mixed land use and suggested that government planners should strengthen the rationality of urban infrastructure allocation.

In spite that the spatial-and-temporal relationship between the physical built environment and taxi ridership was found, the taxi ridership is also influenced by some socio-demographic features (e.g., age, education, and employment) (Yu and Peng, 2019). However, because the large and countless number of floating populations are from outside of Shanghai city, the socio-demographic data are not available in the current demographic census data. Future analysis may include the effects of social demographics if the data is available. In addition, because of the limitation of available data regarding other transportation modes, the interaction between taxis and public transportation modes was not analyzed. Future studies may extend along these considerations.

Declaration of Competing Interest

None.

Acknowledgements

The work described in this paper was jointly supported by grants from the National Key R&D Program of China (2018YFB1600500) and the State Key Laboratory of Structural Analysis for Industrial Equipment (S18307). The authors also would like to thank the support from the China Scholarship Council (CSC, No.: 201906060030) to C. CHEN.

Appendix A

Table A1 shows the Pearson correlation matrix between explanatory variables. The estimation results of the SGWPR models for local variables are given in Table A2 (for the evening-peak model) and Table A3 (for the off-peak model)

Table A1
Results of the Pearson Product-Moment Correlation Coefficient Test.

	TA	TH	PL	HP	Residence	Catering	SS	Corporation	Shopping	Finance	CES	LS	LSS	Healthcare	Government	Hotel
TA	1															
TH	0.016	1														
PL	0.129**	0.034*	1													
HP	0.009	−0.008	0.264**	1												
Residence	0.069**	−0.02	0.041**	−0.009	1											
Catering	0.159**	0.029	0.413**	0.060**	0.050**	1										
SS	0.038*	0.005	0.184**	0.237**	−0.078**	0.218**	1									
Corporation	0.061**	0.021	0.582**	0.214**	−0.039*	0.327**	0.159**	1								
Shopping	0.109**	0.026	0.268**	−0.011	0.089**	0.541**	0.142**	0.146**	1							
Finance	0.126**	0.011	0.524**	0.206**	0.001	0.383**	0.163**	0.649**	0.220**	1						
CES	0.024	−0.02	0.126**	0.155**	0.071**	0.099**	0.089**	0.073**	0.056**	0.047**	1					
LS	0.151**	−0.006	0.418**	0.108**	0.164**	0.637**	0.157**	0.351**	0.454**	0.379**	0.120**	1				
LSS	0.104**	0.018	0.482**	0.152**	0.066**	0.590**	0.159**	0.395**	0.408**	0.386**	0.097**	0.563**	1			
Healthcare	0.103**	0.003	0.316**	0.148**	0.080**	0.320**	0.120**	0.201**	0.239**	0.260**	0.137**	0.373**	0.280**	1		
Government	0.083**	−0.013	0.348**	0.236**	0.115**	0.195**	0.239**	0.311**	0.126**	0.253**	0.224**	0.291**	0.220**	0.259**	1	
Hotel	0.115**	0.017	0.382**	0.223**	0.049**	0.431**	0.235**	0.294**	0.228**	0.287**	0.122**	0.433**	0.368**	0.254**	0.326**	1

Note: TA: Transit Accessibility; TH: Transport Hub; PL: Parking Lot; HP: Housing-Price; SS: Scenic Spot;
CES: Cultural and Educational Service; LS: Living Service; LSS: Leisure and Sports Service;
** Significant at 0.01 level; * Significant at 0.05 level.

Table A2
Estimation results of the SGWPR evening-peak model for local variables.

Variable	MIN		MAX		MEN		LQ		UQ	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Intercept	−1.249	−0.443	3.674	3.766	2.125	2.223	1.73	1.876	2.671	2.807
Parking lot	−1.108	−1.009	1.209	2.47	0.085	0.109	−0.034	−0.052	0.215	0.216
Housing-price	−1.517	−1.067	1.488	1.048	−0.008	−0.031	−0.163	−0.164	0.143	0.092
Residence	−0.65	−0.736	0.654	0.736	−0.014	0.001	−0.101	−0.091	0.068	0.106
Catering	−1.307	−1.495	1.147	1.185	0.126	0.105	−0.013	−0.03	0.25	0.225
Scenic Spot	−2.158	−2.039	1.821	1.666	−0.021	0.013	−0.169	−0.136	0.146	0.178
Corporation	−3.811	−4.581	1.825	1.575	−0.03	−0.152	−0.17	−0.302	0.157	0.05
shopping	−0.7	−0.645	0.847	0.846	0.026	0.031	−0.083	−0.113	0.145	0.167
Finance	−1.454	−0.931	2.028	2.249	0.274	0.297	0.047	0.064	0.466	0.486
Cultural and educational Service	−1.311	−1.319	0.988	1.101	−0.058	−0.012	−0.173	−0.124	0.071	0.093
Living Service	−1.017	−0.862	0.921	1.529	0.063	0.152	−0.053	0.011	0.196	0.275
Leisure and sports Service	−1.044	−0.977	1.02	1.385	−0.007	0.013	−0.114	−0.103	0.097	0.114
Healthcare	−0.652	−0.609	0.954	1.199	0.087	0.072	−0.034	−0.049	0.204	0.178
Government	−0.882	−0.97	1.76	0.801	−0.005	−0.025	−0.121	−0.122	0.083	0.077
Hotel	−1.305	−0.889	1.913	1.874	0.151	0.117	−0.017	−0.03	0.312	0.295

Table A3
Estimation results of the SGWPR off-peak model for local variables.

Variable	MIN		MAX		MEN		LQ		UQ	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Intercept	−0.889	−1.208	3.958	3.845	2.22	2.33	1.796	1.906	2.802	2.934
Parking lot	−0.98	−0.886	2.138	1.204	0.134	0.085	−0.008	−0.041	0.273	0.211
Housing-price	−1.356	−1.212	1.179	1.222	−0.015	−0.008	−0.199	−0.172	0.156	0.148
Residence	−0.595	−0.717	0.88	0.742	−0.02	0.019	−0.12	−0.065	0.072	0.107
Catering	−1.214	−1.294	1.088	1.015	0.105	0.092	−0.041	−0.044	0.239	0.218
Scenic Spot	−2.735	−2.738	1.964	1.967	0.003	−0.037	−0.15	−0.185	0.191	0.124
Corporation	−3.257	−4.466	2.293	2.586	0.012	−0.069	−0.153	−0.236	0.176	0.125
shopping	−0.806	−0.469	0.717	0.793	0.011	0.05	−0.112	−0.066	0.141	0.161
Finance	−1.485	−1.658	2.655	2.49	0.27	0.294	0.043	0.0514	0.447	0.496
Cultural and educational Service	−1.452	−1.246	1.021	1.121	−0.047	−0.015	−0.156	−0.116	0.068	0.082
Living Service	−2.271	−1.661	1.536	1.284	0.074	0.119	−0.064	−0.012	0.226	0.24
Leisure and sports Service	−1.238	−1.074	1.251	0.702	−0.005	−0.006	−0.12	−0.121	0.116	0.11
Healthcare	−1.198	−0.758	1.11	1.206	0.078	0.078	−0.035	−0.04	0.212	0.194
Government	−1.669	−2.079	1.071	1.331	−0.036	−0.003	−0.122	−0.102	0.078	0.092
Hotel	−1.242	−0.913	2.078	2.322	0.132	0.161	−0.021	−0.011	0.287	0.317

References

An, D., Tong, X., Liu, K., Chan, E.H., 2019. Understanding the impact of built environment on metro ridership using open source in Shanghai. Cities 93, 177–187.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28 (4), 281–298.

Cervero, R., 2013. Linking urban transport and land use in developing countries. J. Trans. Land Use 6 (1), 7–24.

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: density, diversity, and design. Transp. Res. Part D: Transp. Environ. 2 (3), 199–219.

Chang, H.W., Tai, Y.C., Hsu, J.Y.J., 2010. Context-aware taxi demand hotspots prediction. Int. J. Business Intell. Data Min. 5 (1), 3–18.

Chen, F., Yin, Z., Ye, Y., Sun, D.J., 2020. Taxi hailing choice behavior and economic benefit analysis of emission reduction based on multi-mode travel big data. Transp. Policy 97, 73–84.

Cordera, R., Coppola, P., dell'Olio, L., Ibeas, A., 2017. Is accessibility relevant in trip generation? Modelling the interaction between trip generation and accessibility taking into account spatial effects. Transportation 44 (6), 1577–1603.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons.

Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). Ann. Am. Assoc. Geograph. 107 (6), 1247–1265.

Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. 100 (1), 253–258.

Jun, M.J., Choi, K., Jeong, J.E., Kwon, K.H., Kim, H.J., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. J. Transp. Geogr. 48, 30–40.

Li, B., Cai, Z., Jiang, L., Su, S., Huang, X., 2019. Exploring urban taxi ridership and local associated factors using GPS data and geographically weighted regression. Cities 87, 68–86.

Lim, C., Kim, K.J., Maglio, P.P., 2018. Smart cities with big data: reference models, challenges, and considerations. Cities 82, 86–99.

Liu, Q., Ding, C., Chen, P., 2020. A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand. Travel Behav. Soc. 18, 29–36.

Mar'ah, Z., Djuraidah, A., Wigena, A.H., 2017. Semi-parametric geographically weighted regression modelling using linear model of coregionalization. Int. J. Sci. Basic Appl. Res. (IJSBAR). 34 (2), 178–186.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1/2), 17–23.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping. Stat. Med. 24 (17), 2695–2717.

Qian, X., Ukkusuri, S.V., 2015. Spatial variation of the urban taxi ridership using GPS data. Appl. Geogr. 59, 31–42.

Seabold, S., Perktold, J., 2010, June. Statsmodels: econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference, Vol. 57, p. 61.

Shao, D., Wu, W., Xiang, S., Lu, Y., 2015, November. Estimating taxi demand-supply level using taxi trajectory data stream. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE, pp. 407–413.

Sun, D.J., Ding, X., 2019. Spatiotemporal evolution of ridesourcing markets under the new restriction policy: a case study in Shanghai. Transp. Res. A Policy Pract. 130, 227–239.

Sun, D.J., Zhang, K., Shen, S., 2018. Analyzing spatiotemporal traffic line source emissions based on massive didi online car-hailing service data. Transp. Res. Part D: Transp. Environ. 62, 699–714.

Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7 (2), 161–187.

Yang, C., Gonzales, E.J., 2014. Modeling taxi trip demand by time of day in New York City. Transp. Res. Rec. 2429 (1), 110–120.

Yang, H., Lau, Y.W., Wong, S.C., Lo, H.K., 2000. A macroscopic taxi model for passenger demand, taxi utilization and level of services. Transportation 27 (3), 317–340.

Yang, Z., Franz, M.L., Zhu, S., Mahmoudi, J., Nasri, A., Zhang, L., 2018. Analysis of Washington, DC taxi demand using GPS and land-use data. J. Transp. Geogr. 66, 35–44.

Yu, H., Peng, Z.R., 2019. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. J. Transp. Geogr. 75, 147–163.

Zhang, K., Sun, D., Shen, S., Zhu, Y., 2017a. Analyzing spatiotemporal congestion pattern on urban roads based on taxi GPS data. J. Transp. Land Use 10 (1), 675–694.

Zhang, W., Ukkusuri, S.V., Lu, J.J., 2017b. Impacts of urban built environment on empty taxi trips using limited geolocation data. Transportation 44 (6), 1445–1473.

Zhang, X., Huang, B., Zhu, S., 2019. Spatiotemporal influence of urban environment on taxi ridership using geographically and temporally weighted regression. ISPRS Int. J. Geo Inf. 8 (1), 23.

Zhang, B., Chen, S., Ma, Y., Li, T., Tang, K., 2020. Analysis on spatiotemporal urban mobility based on online car-hailing data. J. Transp. Geogr. 82, 102568.

Zhu, Y., Chen, F., Wang, Z., Deng, J., 2019. Spatio-temporal analysis of rail station ridership determinants in the built environment. Transportation 46 (6), 2269–2289.