Exploring spatial heterogeneity in the impact of built environment on taxi ridership using multiscale geographically weighted regression

1. Introduction

There is much evidence that the built environment has a significant and substantial impact on travel outcomes (Cervero 2002; Ewing and Cervero 2010; Ewing et al. 2015; Zhu 2012, 2013; Zhu and Mason 2014; Zhu et al. 2020, 2022a). As an imperative component of urban transportation, taxi is providing all-day and flexible service that can complement public transit system and reduce private vehicle usage. According to the National Bureau of Statistics[1], there are over one million taxis in China nationally. Especially in megacities like Beijing, they are an effective solution to the first-mile and last-mile problems (Rayle et al., 2014; Zhu et al., 2023). Understanding the factors that affect passenger demand and taxi supply during morning and evening peaks is crucial for improving the for-hire taxi industry’s service quality (Schaller, 2005). Discovering spatial patterns of taxi trips is also beneficial to mitigating traffic congestion and greenhouse gas (GHG) emissions (Çetin and Yasin Eryigit 2011). This study aims to comprehensively analyze the associations between built environment factors and taxi passengers’ travel behavior.

Some recent studies have considered the spatially heterogeneous effects of built environment variables in travel behavior analysis (eg., Blainey 2010; Blainey and Preston 2013; Feuillet et al. 2015; Qian and Ukkusuri 2015; Tu et al. 2018; Wang

[1] https://data.stats.gov.cn/easyquery.htm?cn=C01andzb=A0B07andsj=2015

and Noland 2021). An improved understanding of the associations between travel behaviors and the built environment is crucial for decision makers to make informed land use policies, and consequently influence broad sustainability objectives, such as reducing traffic congestion, energy consumption, and greenhouse gas (GHG) emissions. To better capture the spatially non-stationary process and obtain more accurate estimates, local models such as geographically weighted regression (GWR) have been introduced in transportation and urban management studies (Qian and Ukkusuri 2015; Wang and Chen 2017; Wang and Noland 2021; Yu and Peng 2019; Yu et al. 2018). However, one major assumption of GWR is that all of the associations between outcome variables and explanatory variables vary at a single spatial scale using the same bandwidth, which may not accurately characterize spatial contexts. The Multiscale Geographically Weighted Regression (MGWR) model allows multiple spatial scales to be expressed simultaneously, thus outperforming GWR. The main contribution of this paper is to employ this novel modelling approach to examine the spatially heterogeneous impacts of built environment characteristics on taxi ridership. This can provide better empirical evidence for taxi passengers' travel patterns in different spatial contexts as compared to the traditional GWR model. We visualize the local estimates, to illustrate the heterogeneity of determinants for taxi ridership's spatiotemporal distribution. The results not only expand the literature on the spatially non-stationarity associations between local built environment factors and taxi ridership distribution but also provide valuable insight into the development of transportation policies at a finer scale.

This study begins by presenting the spatial distributions of taxi ridership in two peak periods through GIS techniques. It then applies MGWR to reveal the spatial heterogeneity for the impacts of different built environment variables on taxi trips. The MGWR model is found to be a powerful tool that can provide a more accurate estimation result and nuanced analysis of the spatial variation in the impacts of different built environment characteristics on taxi ridership. This model achieves better goodness of fit and has more explanatory power. Based on the results, significant multiscale spatial heterogeneity is observed in the associations between spatial contexts and taxi trip pick-ups/drop-offs. We find that residential density is positively associated with taxi demand in areas with less public transport coverage than surrounding units. Moreover, improving bus coverage in areas with low coverage may attract more commuters to choose taxi plus bus mode for commuting instead of purchasing private vehicles. The results also indicate that the positive effects of road density on taxi ridership are spatially heterogeneous. Policymakers can benefit from the more accurate and detailed understanding of the spatial determinants of taxi ridership provided by our work. This may help to allocate available taxi resources within the urban intermodal transportation system, reducing traffic stress in the city center. In addition, since ride-hailing services like Uber and Didi have grown in popularity, our results could potentially be a valuable reference for integrating these novel mobility services into the traditional transport system effectively and efficiently.

2. Literature review

2.1 Factors that influence ridership patterns

There are a wide range of studies concerning the factors that may influence travel ridership. Apart from internal factors like cost, duration, and quality of service (Kanafani, 1983), built environment factors such as population density, development level, and public transport access have been shown to substantially affect travel behaviors. These built environment factors are often summarized as the 5D -- Density, Design, Diversity, Distance to transit and Destination accessibility (Ewing and Cervero 2010). This theory guides our study's independent variable selection. Based on the 5D principle and existing research (i.e., Currans and Muhs 2015; Qian and Ukkusuri 2015; Wang and Noland 2021; Yang et al. 2018; Yuan, Raubal and Liu 2012;), a set of standard variables are adopted in this study. For density, this paper includes residential density and employment density at the 1km*1km grid level. For design, the model includes the road area ratio for each grid, which measures the road coverage. For distance to transit, this study uses bus coverage and the subway coverage to measure public transport service level. For diversity, average housing price per m^2 and construction density (the number of buildings for each grid) are adopted.

Furthermore, for destination accessibility, the model includes the number of different categories of POIs (points of interest) to take account of the effects of different trip purposes. In particular, the existing literature has shown that transfer activities between transportation hubs have significant impacts on the travel demand for ridehailing or for-hire taxi passengers (Wang and Noland, 2021). In addition, housing prices and income have also been considered in a wide range of relevant studies as indicators of socioeconomic status (see Qian and Ukkusuri 2015; Wang and Cao 2017; Wang and Noland 2021; Yang et al. 2018; Yu and Peng 2019;).

Alongside a large body of work focusing on buses, subways and private cars (see Kain and Liu 1999; Kuby, Barranda and Upchurch 2004; Haire and Machemehl 2007; Tao et al. 2014; Chakour and Eluru 2016), ridehailing services like Uber and Didi have also received a tremendous amount of attention (Yu and Peng 2019; Zhang et al. 2020; Wang and Noland 2021). For example, Alemi et al. (2018) used online survey data in California to explore the factors affecting on-demand ride services like Uber and Lyft using binary logit models. They found that age, education level, lifestyle, land use mix and auto accessibility significantly affect the probability of using on-demand ride services. Tu et al. (2021) utilized machine learning methods with Didi ridesourcing requests and GPS data from Chengdu to examine the nonlinear effects of built environment factors on ride pooling. They argued that these built environment factors had substantial nonlinear impacts on the likelihood of ride pooling. Based on the same dataset, Zhang et al. (2020) used ordered logistic regression to investigate the correlation between the intensity of ridehailing services and POIs. Their results indicated that regions with high travel intensity showed spatial agglomeration patterns, and different amenities had different impacts on travel intensity. Nevertheless, the number of studies empirically examining the association between for-hire taxi ridership and various built environment characteristics is still limited compared to that for other transportation modes. Different travel modes may respond to the built environment differently. Drawing lessons from the classic 5D framework, this study investigates the spatial relationships between built environment context and taxi demand.

2.2 Spatial heterogeneity of built environment effects in travel behavior studies

Previous studies have examined the spatial heterogeneity of built environment effects on travel behavior and have demonstrated the importance of accounting for spatial non-stationarity in the relationship between local contexts and travel behaviors (eg., Blainey 2010; Blainey and Preston 2013; Feuillet et al. 2015; Li et al. 2021; Paez et al. 2007; Qian and Ukkusuri 2015; Tu et al. 2018; Xu et al. 2021; Wang and Noland 2021). For instance, Paez et al. (2007) adopted mixed ordered Probit models to investigate spatial and demographic variability in travel behavior of the elderly based on the travel survey data collected in the Hamilton Metropolitan Area, Canada. They pointed out that the determinants of trip-making propensity were not spatially homogeneous, and spatial models (mixed ordered Probit model) performed better than conventional non-spatial models (ordered Probit model). To predict the trip volume of local rail, Blainey (2010) compared multiple linear regression and GWR models and pointed out that the latter had a better fit and spatial parameter variation existed. Employing the geographically weighted Poisson regression (GWPR) model, Feuillet et al. (2015) investigated how residential environment characteristics influence commuting behavior. Their results indicate that the magnitude and statistical significance level of different neighborhood environment factors vary by location. Based on NYC taxi GPS data, Qian and Ukkusuri (2015) found that taxi demand was location-sensitive, meaning that the relationship between taxi ridership and its determinants, including socio-demographic and built environment characteristics, were not stationary across urban spaces. Liu et al. (2020), Tu et al. (2018), and Wang and Noland (2021) similarly argued that it is necessary to account for spatial heterogeneity when analyzing the influence of the built environment on travel behavior.

Conventional global models, such as the ordinary least squares (OLS) model and Probit model, assume that all the relationships between outcome variable and explanatory variables are stationary across spaces (Brunsdon, Fotheringham and Charlton 1996). This assumption does not hold in many real cases. Several empirical studies focusing on factors influencing travel behavior have applied local models to calibrate the spatially non-stationary process. Specifically, the GWR model is one of the typical alternative local models which is able to capture the spatial non-stationarity in the spatial process (Brunsdon, Fotheringham and Charlton 1996). It allows local regression coefficients for each independent variable to vary at different locations; this has been used to examine the spatial patterns and determinants of ridership in a few recent studies (Qian and Ukkusuri 2015, Wang and Noland 2021; Yu and Peng 2019). For example, Yu and Peng (2019) used the GWPR model to explore the spatial relationships between spatial contexts and ridesourcing demand in Austin and found significant spatial heterogeneity in the effects of socioeconomic and built environment characteristics. Wang and Noland (2021) came to similar conclusions using data on Didi Chuxing trips in Chengdu to analyze the association between the built environment and online ride-hailing ridership using the GWR model.

The GWR or GWPR models usually achieve better goodness of fit than the traditional global model (Fotheringham, Brunsdon and Charlton 2003); however, they both regard the spatial scale constant among all spatial processes. Considering the existence of multiscale spatial effects of different built environment variables on the distribution of taxi ridership, this study adopts the MGWR, a multiscale extension of the GWR model, to investigate the spatial variation in the associations between associations different built environment characters and taxi ridership. Compared with GWR, MGWR has the advantage of accurately capturing realistic spatial heterogeneity. This avoids misleading conclusions resulting from ignoring the scale variations for different spatial processes, and further helps to make more effective decisions for local governments. Meanwhile, it can reduce collinearity in the fitting process and diminish parameter estimation bias (Oshan, et al. 2019). A table summarizing the literature is provided in Appendix 1. It can be seen both spatial autocorrelation and spatial heterogeneity have been increasingly noticed in recent years. On the one hand, many studies adopt spatial econometric models to capture the spillover effects embedded in built environment effects or travel demands, namely, spatial autocorrelation. On the other hand, for spatial heterogeneity, most studies use

GWR, Geographically Weighted Poisson Regression (GWPR) and Semi-parametric Geographically Weighted Poisson Regression (SGWPR), and some studies use other methods like simultaneous equation modeling. However, none of these studies address the varying range of spatial processes for different independent variables. To the best of our knowledge, this paper is the first attempt to apply MGWR to examine the associations between spatial contexts and taxi ridership. Based on the existing theories, this study provides a new methodology for in-depth analysis of travel behavior.

3. Data and methodology

3.1 Study area and variables

The study area was defined as the central urban area of Beijing, within the 5th Ring Road (see Figure 1), because the mobile signalling data and housing price data used in this paper are only available for the central area. The study area covers most of the areas of central Beijing, including the entire Dongcheng and Xicheng municipal districts and parts of the Haidian, Shijingshan, Chaoyang, Daxing and Fengtai municipal districts. Beijing's public management departments and social organizations are mostly distributed in central areas within the 4th Ring Road. Three traditional commercial districts (Qianmen, Xidan, Wangfujing) are all located inside the 2nd Ring Road, while many new comprehensive commercial areas such as Madian, Chaowai Street, Wudaokou and Wangjing are located between the 2nd and 4th Ring

Road in the north. In addition, more than half of the office buildings are situated in Chaoyang District and Haidian District, while lots of manufacturing is located outside the southern 4th Ring Road (Tian, Wu and Yang 2010). The chosen study area thus encompasses the core of urban life in Beijing. To address the modifiable areal unit problem (MAUP) (Openshaw and Taylor 1981), we divide this study area into 683 rectangular zones (1km × 1km) instead of using administrative units directly. This grid size is commonly used in research on Chinese metropolises such as Beijing and Shanghai (Liu et al. 2015; Kong et al. 2017; Gong et al. 2016; Zhu et al. 2022b).

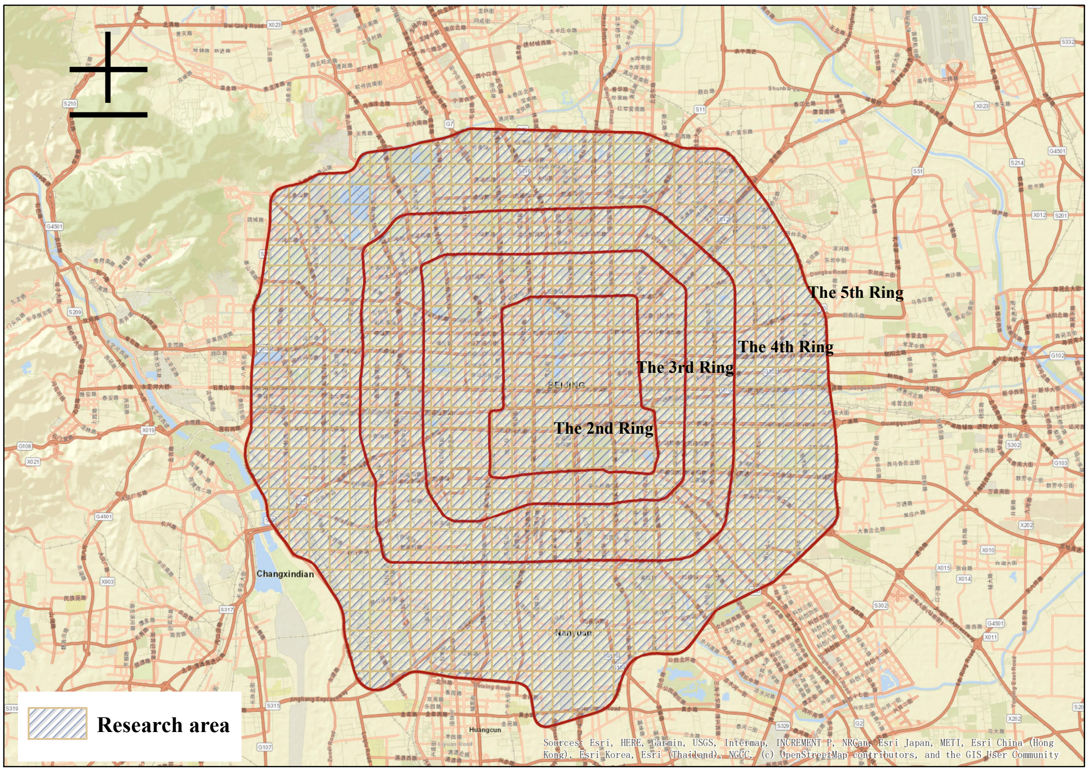

Figure 1 Research area

The dependent variables in this study are the number of taxi[2] pick-ups/drop-offs for each grid in the morning/evening peak time. The taxi trajectory data we used

[2] Note that the taxi in this paper specifically refers to traditional taxis that do not include Transportation Network Company (TNC) services.

are generated by about 18,000 taxis from several anonymous taxi companies from April 1st to 26th 2015, comprehensively reflecting the spatial characteristics of taxi passengers' travel behavior in Beijing. Given that the urban transportation system faces many challenges during rush hours, it is necessary to investigate the determinants of taxi trips' spatial distribution. Therefore, two time periods--morning peak hours (7 to 9 AM) and evening peak hours 5 to 7 PM) on weekdays are selected. We calculated the average hourly taxi trip pick-ups/drop-offs in two peak periods for each grid, respectively.

The independent variables are selected according to the 5D principle (Ewing and Cervero 2010). Specifically, we select the following built environment and socioeconomic variables as factors potentially influencing taxi ridership: residential / employment density, average housing price per m^2, bus coverage (share of 200m bus coverage in a grid), subway coverage (share of 400m subway coverage in a grid), road coverage (road area ratio for each grid), the number of buildings for each grid, the amount of floors in buildings for each grid, and the amount of certain types of POIs (Figure 2). All independent variables were collected in the same period as the dependent variable.

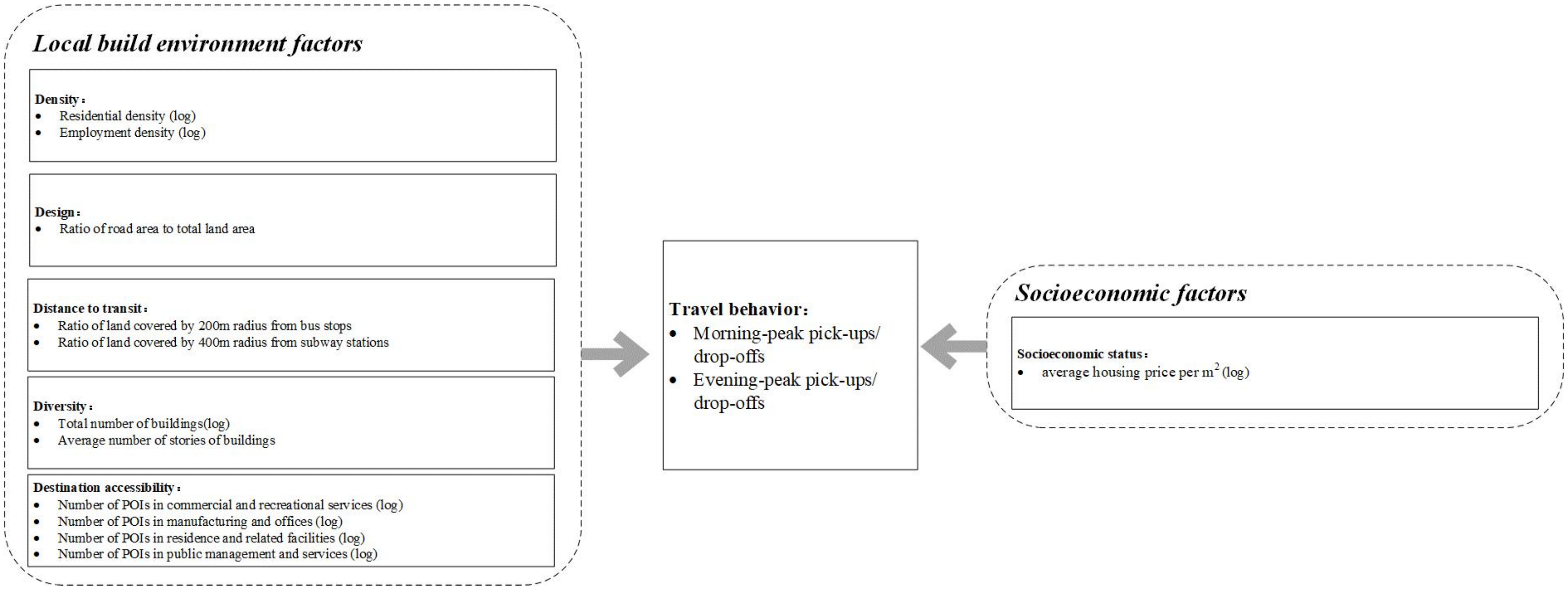

Figure 2 Research framework for major impact factors of taxi passengers' travel behavior

Since official fine-grained employment and residential density are unavailable in Beijing, we calculated these variables based on mobile signaling data. Mobile signaling data has been widely used to estimate the number of employers and residents in a specific geographical range (Ding, Niu and Song 2016; Louail et al. 2014). The mobile signalling data used in this research was collected from China Unicom, one of China's mainstream mobile carriers, provided by Smart Steps Digital Technology company. The residential location and workplace of each user were identified according to duration and frequency during weekdays, and then the number of employees and residents were aggregated by each unit.

This study uses the sum of road area per grid cell as the measure of road area density. The road network data is collected from OpenStreetMap (https://www.openstreetmap.org/). We calculated the road area of each grid by multiplying road length, lane numbers and lane width. Since each grid has a total area of $1km^2$, the road area of a grid is numerically equivalent to the road area ratio. Public transit availability must also be considered in travel behavior analysis. Though subway and bus are both major public transport modes, there are some differences regarding capacities, distribution and service ranges. We specify the geolocation of bus stops and subway stations and generate a 200-meter buffer for each bus stop and a 400-meter buffer for each subway station. The radius definition is drawn from previous studies on acceptable transfer between transport modes (Ishaque and Noland 2008; Wang and Noland 2021). To generate the variables used in the model, we calculated the ratio of the land area within a 200-meter of bus stops or a 400-meter service radius of subway stations to the total land area ($1km^2$) in each grid.

This study introduces the building density and the average amount of stories for each cell as independent variables to measure the land development intensity. The original geospatial data of buildings were collected from AMap[3]. We also include the counts of four types of POIs in each grid, which were scraped from Baidu Map[4].

Since no official micro-level socioeconomic data on actual personal income or family financial conditions is available, this paper uses housing prices to measure household income, as practiced in previous studies (Zhang, Jia and Yang 2016). We collected housing prices in 2015 from Lianjia (https://bj.lianjia.com), a popular real estate transaction platform in China. We computed the average housing price of all housing units for each grid. In addition, for some cells (around 26%) that lack housing prices, this study follows previous studies in adopting spatial interpolation methods to estimate the missing data (e.g., Chica-Olmo 2007; Hu et al. 2016; Martínez, Lorenzo and Rubio 2000; Zhang, Tan and Tang 2015). Specifically, we employ the empirical Bayesian Kriging approach for spatial interpolation (Omre 1987; Krivoruchko 2012), which has proven to be capable of achieving better global prediction with limited sample data (Montero-Lorenzo, Larraz-Iribas and Páez 2009). The spatial interpolation was implemented in ArcGIS Desktop 10.7, following the official Help documentation. We implied K-Bessel as the semivariogram model, which performs best in cross-validation.

[3] http://lbs.amap.com/
[4] http://map.baidu.com

3.2 MGWR model

As mentioned in section 1, GWR is a local regression model that can capture the spatial heterogeneity embedded in the spatial process by allowing parameter estimates for each research unit. A critical concept in GWR is the bandwidth, representing the spatial scale for each local regression equation. One shortcoming of GWR is that it assumes that the bandwidth for each independent variable is constant (Fotheringham, Yang and Kang 2017). However, it is more likely that different built environment, socioeconomic and demographic variables may have different spatial scales. That is to say, the constant bandwidth for each independent variable in GWR possibly leads to misspecification of scale for some variables (Oshan et al. 2019) and unnecessary noise (Gu et al. 2021).

MGWR is the multiscale extension of GWR that allows different optimal bandwidths for different independent variables in the model. That enables multiscale modeling, which provides a more accurate and robust stimulation for actual spatial processes (Fotheringham, Yang and Kang 2017; Yu et al. 2018). This can be expressed as follows (Equation 1).

$$y_i = \beta_{bw0}(\alpha_i, \gamma_i) + \sum_k \beta_{bwk}(\alpha_i, \gamma_i) x_{ik} + \varepsilon_i \qquad (1)$$

Where $\beta_{bw0}(\alpha_i, \gamma_i)$ represents the intercept for observation *i* at the location (α_i, γ_i); x_{ik} represents the kth independent variable; β_{bwk} indicates the *kth* coefficient, and bwk is the bandwidth of its kernel function; and ε_i is the error term. The criterion for selecting the optimal bandwidths is the corrected Akaike information criteria

(AICc). Following Fotheringham, Yang and Kang (2017), the optimal bandwidth in this study is calculated by a bi-square kernel with a back fitting algorithm for calibration. Its main idea is to assume that all other parameters are known except the parameter that is to be calibrated currently. Proportional change in the residual sum of squares (RSS) convergence criterion is used as the convergence criterion (Equation 2).

$$SOC_{RSS} = \left|\frac{RSS_{new}-RSS_{old}}{RSS_{new}}\right| \qquad (2)$$

Where the SOC_{RSS} is the score of change according to RSS; RSS_{new} represents the RSS in the calculation this time; RSS_{old} represents the RSS in the calculation last time.

4. Results

4.1 Model comparison

As demonstrated in Table 1, the MGWR allows each independent variable to have different optimal bandwidths that depict the influencing scales for the spatial process (Lao and Gu 2020). Because both the morning pick-up model and the evening drop-off model consider the impacts of residential density on taxi trips, we align this pair together in Table 1, and so do the morning drop-off model and evening pick-up model, which both include the employment density. The influencing scales for residential density, commercial-and-recreational POI, and residence and related facilities encompass nearly the entire study area in all four models, indicating these variables have large influencing scales with global bandwidths during both peaks. In addition, bandwidth for the share of 200-meter bus coverage in the grid cell and average housing prices is also 682 grid units (out of 683) in all models except the Morning drop-off model. Compared with these variables, the influencing scales of employment density, the road area ratio, and construction density are much smaller. The bandwidth of average housing price is small in the Morning drop-off model, suggesting that the spatial variation of its influences on morning taxi drop-offs is large. Moreover, the influence scale of the construction density is quite small except in the Morning drop-off model. As for different POIs, the bandwidth of public management and service POIs is small in the two evening peak models, indicating that there is significant spatial heterogeneity during the evening peak hours. In summary, the spatial scales of influence vary for different factors. Thus, GWR, which assumes that all variables have the same bandwidth, is biased, and MGWR should be adopted to capture the spatial variation in influence factors.

Table 1 Bandwidth calculation results of MGWR

	Morning pick-up model	Evening drop-off model	Morning drop-off model	Evening pick-up model
Residential density (log)	680	682		
Employment density (log)			152	54
Share of 200-meter bus coverage in the grid cell	682	682	307	682
Share of 400-meter subway coverage in the grid cell	682	669	665	424
Road area ratio for each grid	258	287	680	133
Average housing price (log)	682	682	179	682
# of buildings for each grid (log)	223	70	682	215
# of stories in buildings for each grid (log)	680	152	664	682
# of commercial-and-recreational POI (log)	682	682	682	682
# of manufacturing-and-office POI (log)	531	682	287	682
# of residence-and-related-facility POI (log)	682	682	682	682
# of public-management-and-service POI (log)	682	49	682	56
Intercept	48	70	48	149

Note: (1): # represents the number of this variable; (2):The number in the table denotes how many nearest neighbors that contribute to the regression results for one unit.

Table 2 shows regression results of the MGWR and OLS models. The median, maximum, minimum, and standard deviation values of each independent variable's local coefficient estimate in MGWR are represented. This study's explanatory variables have passed the multicollinearity test (shown in Appendix 2) (Montgomery et al., 2012). The chosen bandwidths of the MGWR are adaptive, and the bi-square kernel function is adopted.

Table 2 Regression results of OLS and MGWR Model

	Morning Pick-up			**Morning Drop-off**			**Evening Pick-up**			**Evening Drop-off**		
	MGWR		OLS	MGWR		OLS	MGWR		OLS	MGWR		OLS
	Median	St. Dev.	Coef.	Median	St. Dev.	Coef.	Median	St. Dev.	Coef.	Median	St. Dev.	Coef.
Residential density (log)	0.107	0.010	0.109***							0.120	0.002	0.118***
Employment density (log)				0.385	0.095	0.248***	0.592	0.494	0.078**			
Share of 200-meter bus coverage in the grid cell	0.160	0.031	0.186***	0.093	0.033	0.132***	0.11	0.004	0.086**	0.122	0.005	0.139***
Share of 400-meter subway coverage in the grid cell	0.017	0.003	0.017	0.040	0.004	0.050***	0.068	0.028	0.108***	0.040	0.005	0.022
Road area ratio for each grid	0.131	0.049	0.175***	0.147	0.011	0.173***	-0.044	0.251	0.165***	0.122	0.038	0.145***
Average housing price (log)	0.017	0.002	0.072***	0.024	0.008	0.094***	-0.007	0.009	0.065**	0.044	0.003	0.102***
# of buildings for each grid (log)	-0.010	0.039	0.014	-0.018	0.059	0.014	-0.255	0.03	-0.015	-0.014	0.069	-0.006
# of stories in buildings for each grid (log)	0.169	0.006	0.228***	0.076	0.004	0.161***	0.026	0.038	0.181***	0.112	0.052	0.182***
# of commercial-and-recreational POI (log)	0.077	0.007	0.088***	-0.044	0.004	-0.048	0.116	0.142	0.106**	0.220	0.002	0.207***
# of manufacturing-and-office POI (log)	0.052	0.010	0.076***	0.047	0.057	0.143***	0.119	0.142	0.376***	0.082	0.003	0.111***
# of residence-and-related-facility POI (log)	0.090	0.006	0.038	0.021	0.005	0.007	-0.149	-0.135	-0.118	0.062	0.004	0.048**
# of public-management-and-service POI (log)	0.192	0.008	0.189***	0.190	0.006	0.203***	-0.252	0.398	0.025	0.106	0.077	0.136***
Intercept	0.036	0.197	0	0.024	0.233	0	-0.337	0.115	0	0.081	0.183	0
AICc	306.789		500.897	376.642		574.246	687.383		1130.450	102.794		334.243
R^2	0.926		0.883	0.920		0.869	0.882		0.705	0.948		0.908
Adj. R^2	0.918		0.881	0.910		0.867	0.864		0.700	0.940		0.907
Moran's *I* in residual — Contiguity matrix	0.011		0.224***	0.076*		0.276***	0.024		0.328***	0.003		0.248***
Inverse-distance matrix	0.005		0.281***	0.074*		0.285***	0.030		0.322***	0.001		0.244***

Note: (1). * p-value < 0.05; ** p-value <0.01; *** p-value <0.001; (2) Dependent variables for four models are number of morning-peak pick-ups (log), number of morning-peak drop-offs (log), number of evening-peak pick-ups (log) and number of evening-peak drop-offs (log), respectively.

Table 2 demonstrates that MGWR has a lower AICc value and a higher R^2 and adjusted R^2 than OLS, indicating relatively better goodness of fit (Charlton et al., 2009). In addition, we calculated the global Moran's I value of the residuals of each model to test whether spatial autocorrelation in the residuals exists. The results are also reported in Table 2. The residuals of the OLS model show a strong spatial autocorrelation (Moran's I value is significant and positive) regardless of whether the contiguity matrix or inverse-distance matrix is used, whereas no significant spatial autocorrelation is observed among residuals of the MGWR model. According to Finley (2011) and Gu et al. (2019), the spatially autocorrelated residuals may indicate inaccurate estimates of model parameters. These results further indicate that MGWR helps to ensure statistical validity and increase prediction precision through locally fitting the spatial variation of residuals from each spatial process. Therefore, the local MGWR model is preferred to the global OLS model in estimating the relationships between different built environment factors and taxi trips at different locations.

In addition, we further implement GWR for four models and compare the results to MGWR. The GWR results are shown in Appendix 1. The universal bandwidth for each of the four models (i.e., morning pick-up, morning drop-off, evening pick-up, and evening drop-off) is 206, 150, 150 and 166, respectively. Because Table 1 has clearly indicated that the bandwidths of most explanatory variables are different from each other and should be treated as such in analysis, we infer that adopting a universal bandwidth will be oversimplistic and lead to biased estimates. Furthermore, the differences in corrected Akaike information criterion

(AICc) and R^2 across OLS, GWR and MGWR models suggest that MGWR achieves the best model goodness of fit.

Before we proceed to discuss the MGWR model results in detail in section 4.2, we first briefly present the results of the OLS models, because they can provide intuitive interpretations on which explanatory variables are statistically significant (although their coefficient sizes are less relevant as MGWR models will provide a local coefficient estimate for each geographic unit). The OLS models report the global effects of the explanatory variables, indicating that six independent variables are significantly associated with taxi ridership in all four models, including the residential or employment densities, bus coverage, road network density, average housing price, the average amount of stories and manufacturing and office facilities. Commercial and recreational facilities shows no significant relationship with taxi drop-offs in the morning peak. That may be because most travel during this period is for commuting purposes. Likewise, residence and related facilities are only positively associated with evening-peak drop-offs. Parameter estimates for the public-management-and-service POI is statistically significant and positive, except in the Evening pick-up model, possibly because most government departments and agencies close before the evening peaks we defined (5:00-7:00 PM). Subway coverage is statistically significant only in the Morning taxi drop-off and Evening pick-up model, as people often use taxi to get from home to subway stations in the morning peak and inversely in the evening peak. Meanwhile, bus coverage has a substantially positive impact on taxi trip pick-ups and drop-offs in all four models. In other words, the importance of taxis as a solution to first-mile or last-mile problems for bus passengers is higher than that for subway passengers.

4.2 Empirical results

4.2.1 Spatial pattern of taxi trips

Figure 3 displays the spatial distribution of pick-ups and drop-offs of taxi ridership during two peak periods based on the natural breaks (Jenks) classification method.

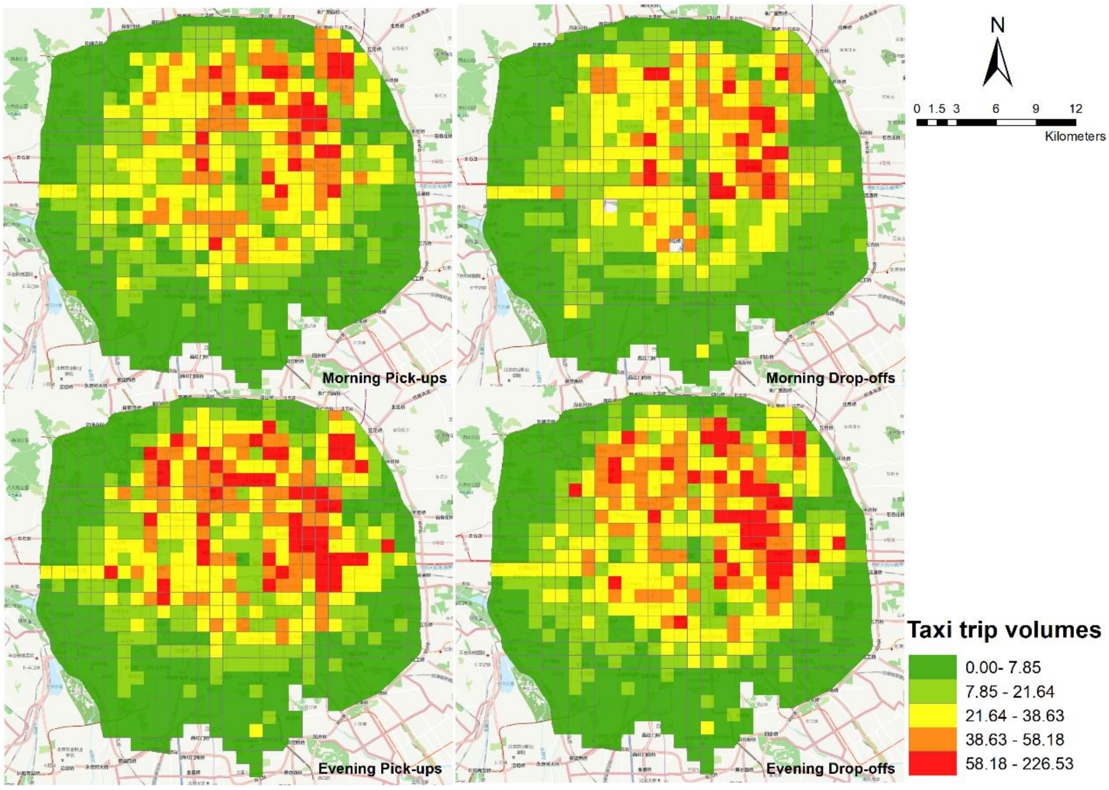

Figure 3 Spatial pattern of taxi trip volumes in grids

The amount of taxi trip pick-ups/drop-offs in each grid can also be interpreted as the density of pick-ups/drop-offs as each grid area is equal to $1km^2$. In the morning, the density of taxi trip pick-ups is generally lower than that of drop-offs, whereas the latter is less concentrated. In the evening peak hours, however, the density of taxi trip pick-ups is roughly equal to that of drop-offs. In addition, taxi drop-off density is relatively low during the evening rush hour compared to morning.

To further reveal the spatial clustering pattern of taxi trips, we also implement Hot Spot Analysis. This is a spatial statistic tool for identifying significant spatial clusters of high and low values, called hot and cold spots, respectively (Ord and Getis 1995; Getis and Ord 2010). Figure 4 displays the mapping results of Hot Spot Analysis. The taxi pick-ups / drop offs during two peak periods show roughly similar results, which can be explained in two respects. First, some studies the average travel distance for taxi trips in Beijing is between 9km~20km (Jiang et al. 2018). That is, taxis are usually employed for short and medium-distance travel, which can explain the similarity of Hot Spot Analysis results of pick-ups and drop-offs during two periods. In addition, compared with auto-dependent Western cities with a comparable population and land use size, people in Beijing have relatively shorter commuting distances, which indicates a better job-housing balance within the Beijing Metropolitan Area (Zhou and Long 2014). Clusters of high-density taxi trips (hot spots) are concentrated in the northeast central urban area, between the 2nd Ring Road and the 4th Ring Road during both morning and evening peak hours. Moreover , most interchange stations like Wangjing station and the Olympic Park station are hot spots. The clusters of low-density taxi trips (cold spots) exist in two peaks, which means not only does the grid itself have low-density taxi trips, but the taxi ridership in its neighboring grids is also low, primarily distributed at the edge of the research area.

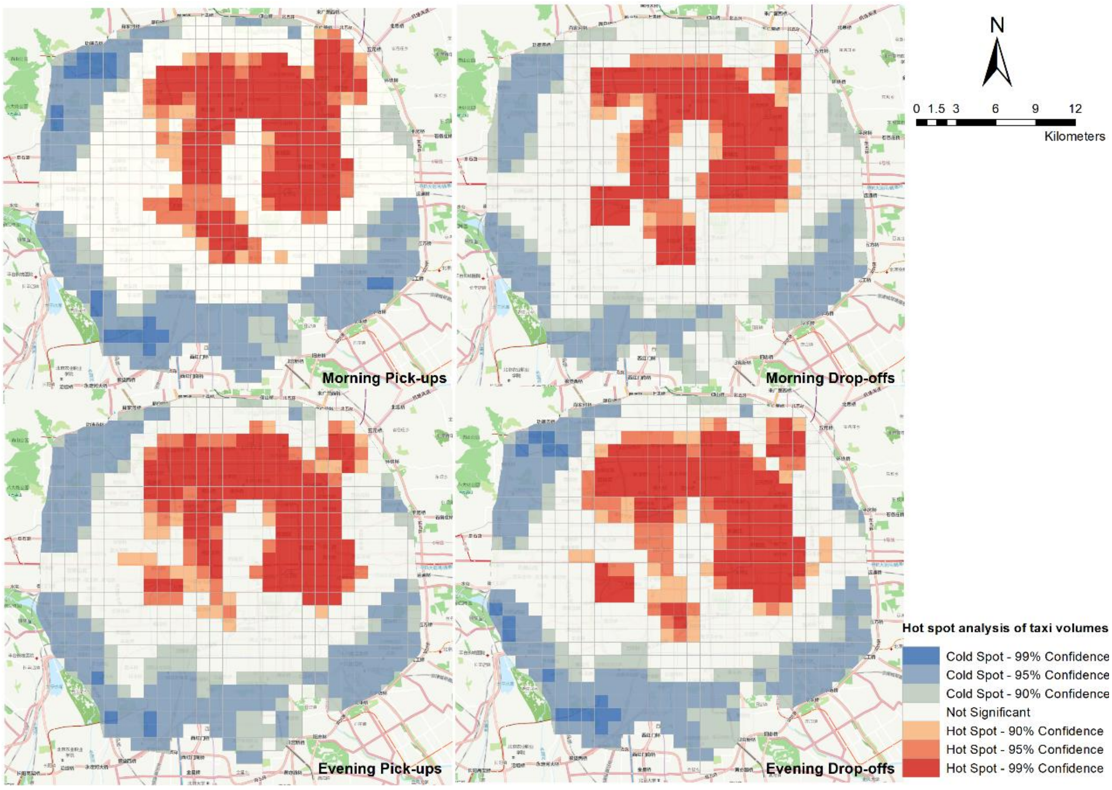

Figure 4 Hot Spot Analysis of taxi volumes in grids

4.2.2 Spatial variations of the impact on taxi trips

To reveal the spatial pattern of influence factors on taxi trips, we visualized the parameter estimates of explanatory variables using ArcGIS 10.7 software with the Jenks Natural Breaks[5] classification method, and colors from blue to red show the variation of determinants. The spatial patterns of regression coefficients for each variables are illustrated in Figure 5, note that only significant variables will be discussed in detail subsequently.

[5] It is a classification method based on variance minimization criteria, which can maximize the variance among different classes (De Smith, Goodchild, and Longley 2017). It has been widely used for local coefficient visualization (eg., Gu et al. 2021; Sha et al. 2017 and Yang et al. 2019).

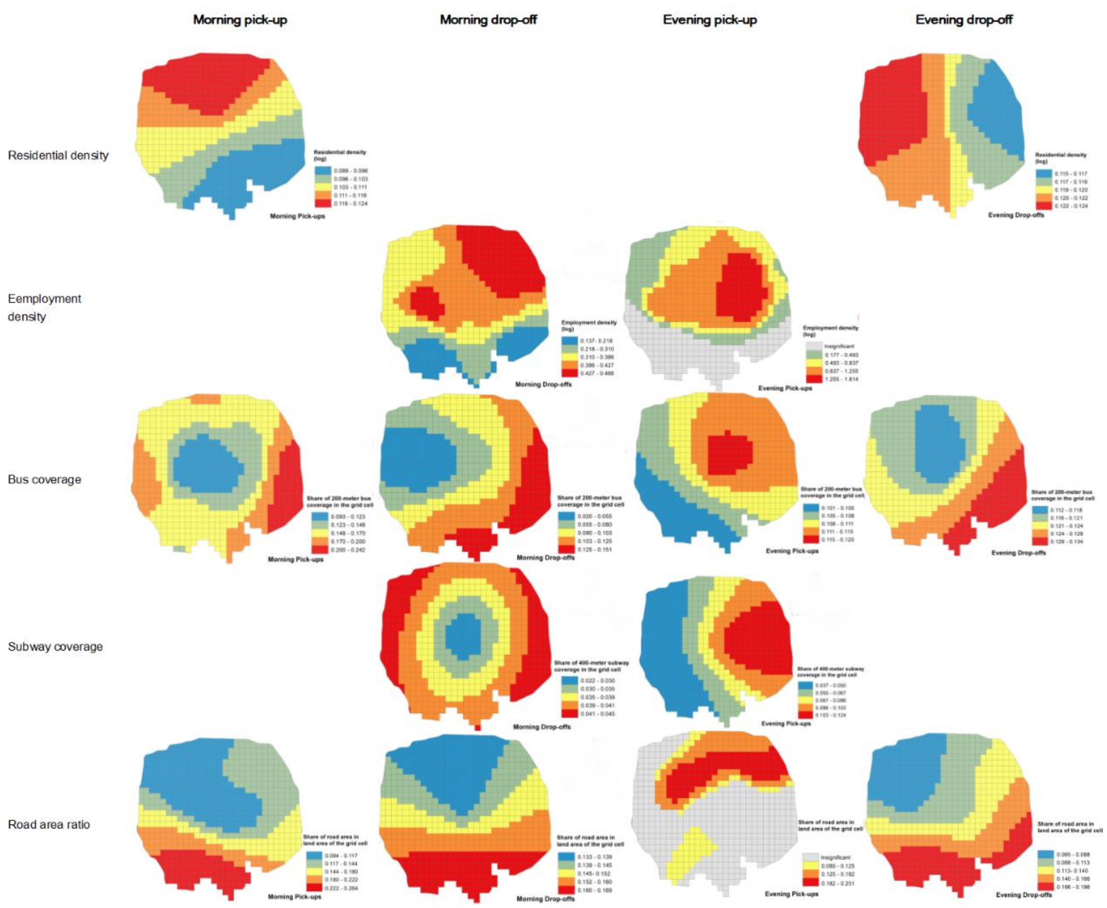

Figure 5 Local coefficients of each factor

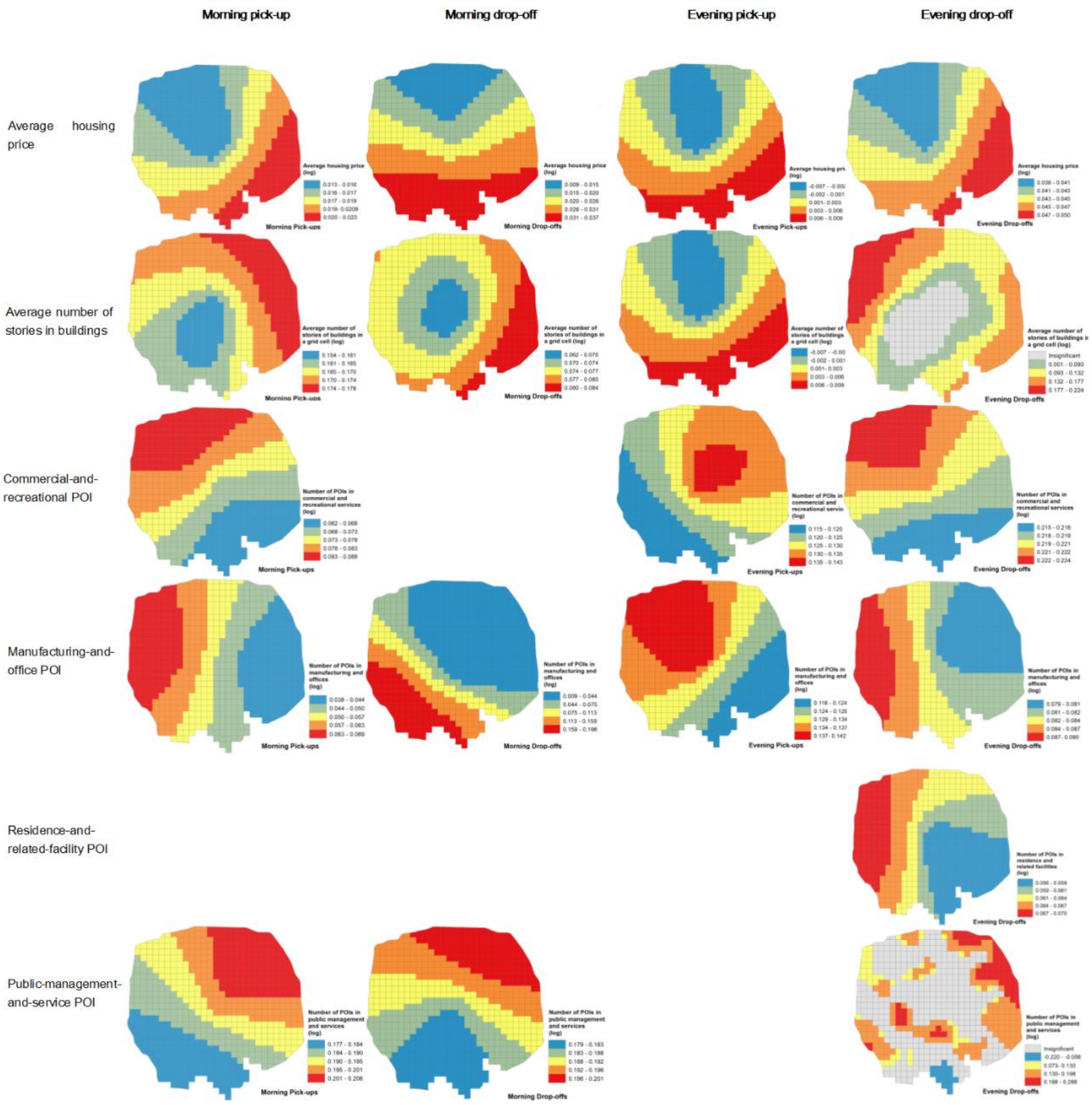

Figure 5 Local coefficients of each factor (continued)

1) Residential and employment densities

The spatial patterns of regression coefficients of residential or employment densities are different in the four models. As Figure 5 shows, it residential density has positive impacts on taxi demand during the morning peak in the entire research area, and the values of its coefficients decrease from north to southeast in the Morning pick-up model. For the Morning drop-off model (b), the coefficients vary significantly in space. Larger positive coefficients in the northeastern part of Beijing are observed.

Meanwhile, strong positive effects of employment density on taxi drop-offs during the morning peak hour occur in a small area west of the city center. In the evening, higher positive effects of employment density on taxi trip pick-ups can be observed in the north-eastern part, which accommodates Beijing's Central Business District (CBD). One interesting thing is that the small area west of the city center observes higher coefficients again. Referring to the hotspot analysis of bus and subway coverage (Figure 6), we find that this is a blank area for both figures, meaning it has less access to public transport compared with neighboring areas. Given that taxis are a vital complement to public transit in China (Hall et al.,2018; Wang and Noland, 2010), people who work in the vicinity of this gap in bus and subway coverage may rely heavily on taxis to connect to other transportation modes. Thus, employment density shows a stronger effect on taxi trips there. For evening-peak drop-offs, the estimates vary slightly across space, implying that the residential density is more influential in the west of the city, with the impact decreasing from west to east.

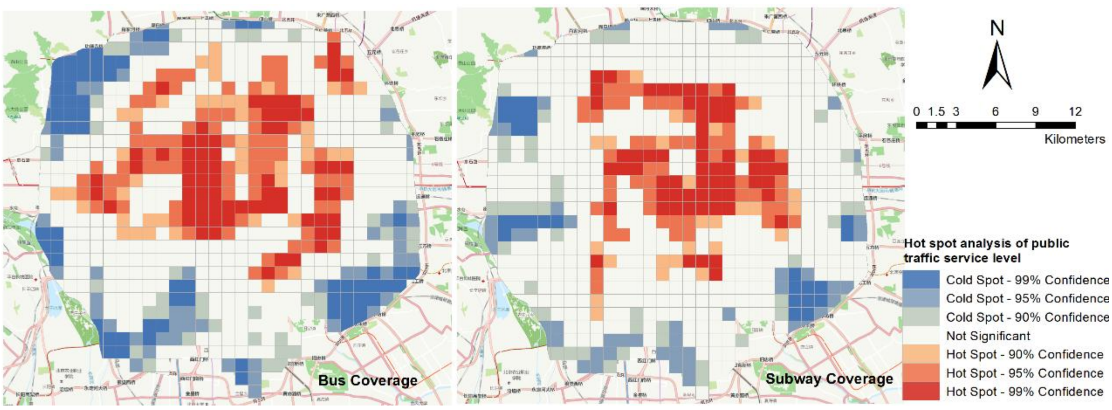

Figure 6 Hot Spot Analysis of bus/ subway coverage

2) Public transportation

- Share of 200-meter bus coverage in the grid cell

During the morning peak, we find that the local coefficients of the share of bus coverage area are positive and decrease slightly from center to periphery. That means an increase in bus coverage in areas around the city center with relatively low bus coverage may introduce more taxi trips. Increasing bus coverage in these places can encourage commuters to choose taxi plus bus mode for commuting--taking taxis to bus stops nearby and then taking buses to work. This could potentially contribute to reducing private vehicle usage, consequently relieving traffic stress and carbon emissions in the downtown area. In the Morning drop-off model, the spatial variation in the local parameter estimates for this variable is stronger, increasing from west to east. In the Evening pick-ups model, the highest local coefficient of bus coverage occurs in the northeast, which is related to the increased travel demand in this area. Bus coverage is still statistically significant for drop-offs and associated with more drop-offs during the evening peak hours, with the high coefficients concentrated in the southeast.

- Share of 400-meter subway coverage in the grid cell

According to the results, subway coverage shows a significantly positive association with taxi morning peak drop-offs and evening peak pick-ups. As shown in figure 5, the spatial pattern of its local estimated coefficients illustrates a pattern similar to the bus coverage pattern in the morning peak models. In the evening peak periods, an increase in subway coverage is likely to be associated with more taxi ridership in the eastern part of the city, while this positive effect is relatively weaker in the west.

3) Road network density

According to the estimation results, local road density is positively associated with both pick-ups and drop-offs, decreasing from south to north in the morning peak hours. Figure 7 shows that the local road network density in Beijing increases from south to north. That is, the rise of road density in southern areas with fewer roads is positively associated with the amount of taxi trips, which may suggest the current density of roads is insufficient to meet travel demand. The evening pick-up model shows a different spatial pattern of coefficients for road density compared with the morning pick-up model. There is a high clustering of area during the evening peak with high positive cumulative estimates in the north. This difference may could be explained by different travel purposes. Research points out that areas with high road density attract more people living and working, so that generates more travel demand (Tang et al. 2019). The employment density in the northern part of our study area is much higher than the southern part, therefore the positive impact of road density is quite obvious in the north, while no significant effect of it can be found in the south. In the Evening drop-off model, road density still shows a positive impact in units with a large number of taxi drop-offs. The spatial distribution of its coefficient is consistent with that in the morning.

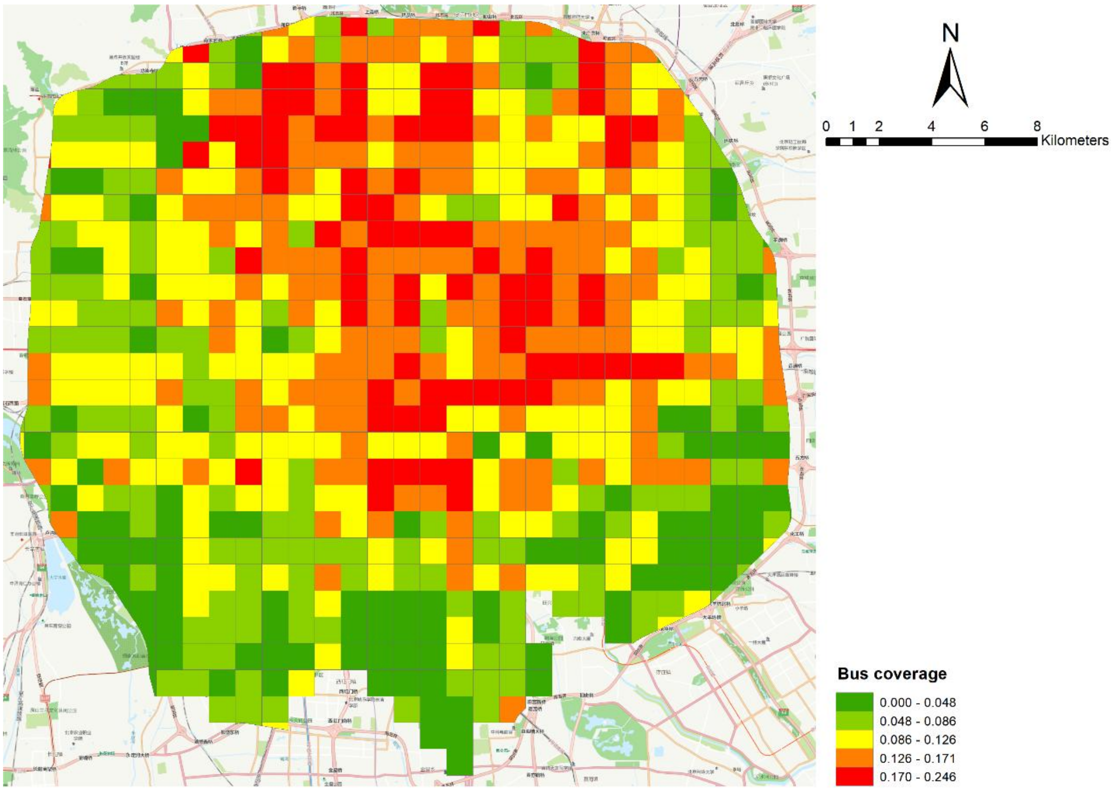

Figure 7 Spatial pattern of the road area ratio

4) Average housing price

The estimates of average housing prices in the Morning pick-up model vary slightly across the research area. According to Table 1, this factor has positive impacts on taxi demand in general. Moreover, this influence weakens from southeast to northwest, perhaps because of the higher price sensitivity for low-income groups as taxis are relatively costly compared with other travel modes (the starting price of taxis, buses, and subways in 2016 being ￥13 yuan, ￥2 yuan, and ￥3 yuan respectively). For drop-offs during the morning peak, the spatial variation in its estimates is stronger. The positive coefficient in the northern area, which concentrates lots of upscale residential neighborhoods, is smaller than that in the southern part. The spatial distributions for the coefficients of average housing price in the Evening pick-up and drop-off models are broadly similar, with the values decreasing from south to north.

5) Land development intensity

- Average number of stories of buildings

During the morning peak, the average number of stories of buildings shows a positive effect on taxi trip pick-ups, increasing slightly from southwest to northeast. For drop-offs, its local coefficients rise from the city center to the surrounding area. Meanwhile, the eastern part has the highest coefficients, indicating that high-rise buildings such as office towers and apartment blocks in the central area with intensive development have fewer effects on taxi trips during the morning peak. In contrast, such effects are more decisive in less intensively developed areas. In the Evening pick-up model, an increase in high-rising buildings is correlated with more taxi demand in the east. Nevertheless, its highest positive coefficients appear in the northwest corner, and the coefficients for central downtown are lower and even negative in the Evening drop-off model. Specifically, the eastern part of our research area has a concentration of many tall office towers and condos (see figure 8), which is more likely to generate taxi trips for commuting in the morning than areas with lower buildings. As for the different results of the Evening drop-off model, this may be related to more diverse travel purposes after work (e.g., entertainment, dining and shopping).

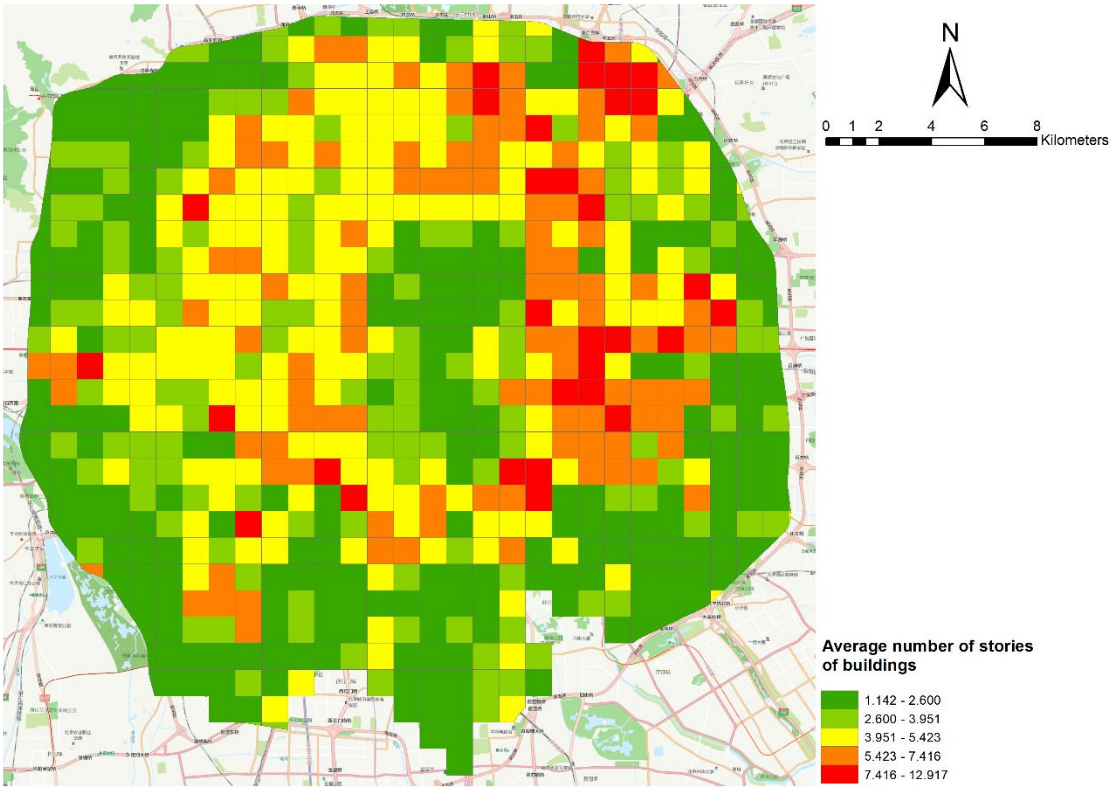

Figure 8 Spatial pattern of the average number of stories of buildings for each grid

6) Different categories of amenities

Our results show that taxi trips during the two peak periods are significantly associated with most categories of POI. The spatial patterns of various POIs' estimated coefficients are considerably different, giving us insights into the detailed association between different amenities and taxi trips in different areas within the city. For example, the manufacturing-and-office POI shows strong positive and significant coefficients in four models. The Morning pick-up model's local coefficients decline from west to east, similar to the Evening drop-off models. By comparison, the positive effects of manufacturing and office POIs exhibit an increasing trend from the northeast part to the southwest part, and the clusterings of areas with the lowest estimates cover the central area. The cause of this phenomenon may be related to the unique political position of central Beijing: a large number of people who work there are government officers with less taxi dependence, while commuters in the northern area are usually medium or high-income groups, with a higher probability of taking taxis.

5. Conclusions

Although a large body of empirical studies has been conducted on the effects of the built environment on travel behavior (Ewing and Cervero 2010), little attention has been given to the spatial heterogeneity of these effects. Only a few studies have specifically investigated the spatially heterogeneous effects of the built environment on taxi ridership (Liu et al. 2020; Qian and Ukkusuri 2015; Wang and Noland 2021). However, they have ignored the varying scales for impacts of different factors on taxi trips. This study fills this gap and examines such heterogeneity using a full sample of taxi trips during morning and evening peak hours in April 2015. We adopt the MGWR local model to reveal detailed spatial variation in the determinants of taxi ridership. The results can help policymakers better understand the spatial patterns of taxi trips and their relationship with urban built environmental characteristics. This can help policymakers to develop more contextualized policies, and may be particularly valuable for allocating taxi reception zones and transportation planning.

We first use Hot Spot Analysis to analyze the spatial- clustering pattern of taxi ridership. A spatially concentrated pattern of taxi ridership during two peak hours can be observed. Specifically, clusters of high-density taxi trips (hot spots) are concentrated in Beijing's northeast central urban area, and clusters of low-density taxi trips (cold spots) are distributed at the edge. Then the global model suggests that the residential/employment density, bus coverage, local road network density, average housing price, high-rise buildings and the manufacturing-and-office POI positively affect taxi trip pick-ups and drop-offs during both peak periods. In addition, subway coverage only has a statistically significant coefficient estimate in the Morning drop-off model and the Evening pick-up model. This provides further evidence that commuters take taxis to subway stations for commuting, implying that taxis can complement the public transit system by serving as a feeder mode.

The MGWR model provides a deeper understanding of the spatial heterogeneity in the impact of built environment characteristics on taxi ridership. First, spatial heterogeneity exists in the distributions of parameter estimation of each independent variable. For example, the positive estimated coefficients of residential density decrease from north to southeast during the morning peak, whereas bus coverage increases from the center to the periphery. Second, we find that residential density has a more profound effect on taxi demand in places with limited public transit access. Third, the positive impact of public transit on taxi demand follows the law of diminishing returns, which may suggest that improving the bus coverage in the outskirts, where there is less bus coverage, can encourage more commuters to take taxis to bus stops nearby and then take the bus to work. This may be able to reduce private vehicle usage and reduce traffic stress and carbon emissions in the downtown area. The results also reveal that an increase in road density is possibly related to more taxi demand in the southern area with less road density than in the northern parts with a dense road network. This phenomenon implies that the current road facilities cannot meet the public travel demand there. In addition, we find that the unique urban layout of Beijing, where the central part possesses crucial political function, needs to be considered in relevant research. According to our findings, we recommend that the relevant departments pay enough attention to the spatial non-stationarity in the determinants of taxi ridership' distribution. Moreover, the construction of the public transit system should be improved in suburban areas with less public transit coverage, which may effectively decrease the public demand for private cars and reduce inner-city congestion. Planning should also take into account the need for taxi services to complement these extensions of public transit, in order to serve first and last-mile needs. In addition, building-height restrictions in Beijing metropolitan area should be reconsidered, which may contribute to low-density suburbanization and increase residents' commuting costs, but provide relatively minor benefits to the service level of the urban transportation system.

Several limitations remain in this study. One is that the results of this paper are based on a 2015 Beijing dataset. Analysis using more recent data sources and data from other cities of different scales to confirm the findings of this study would be an important direction for future research. Furthermore, while this study addresses spatial heterogeneity, temporal heterogeneity may also exist. This is also an area that requires further study. There is a large scope for developing our work in the future. First, if more data are available, such as taxis trip lengths, vehicle ownership, and bikeshare ridership, we are allowed to examine the complex substitute and complementary relationships among different travel modes, which has important implications for reductions in both traffic congestion and greenhouse gas emissions. Besides, further research should investigate the causal mechanisms between the built environment characteristics and travel behavior through a more valid research design, such as combining with the qualitative research and longitudinal data collection. That will help to unravel more complexities in the impact of the built environment on travel behavior.

Reference

Alemi, F., Circella, G., Handy, S., and Mokhtarian, P. 2018. What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. *Travel Behaviour and Society* 13: 88-104.

Blainey, S. 2010. Trip end models of local rail demand in England and Wales. *Journal of Transport Geography* 18(1): 153-165.

Blainey, S. P., and Preston, J. M. 2013. A GIS-based appraisal framework for new local railway stations and services. *Transport Policy* 25: 41-51.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis* 28(4): 281-298.

Cervero, R. 2002. Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment* 7(4): 265–284.

Çetin, T., and Yasin Eryigit, K. 2011. Estimating the effects of entry regulation in the Istanbul taxicab market. *Transportation Research Part A: Policy and Practice* 45(6): 476e484.

Chakour, V., and Eluru, N. 2016. Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. *Journal of Transport Geography* 51: 205-217.

Charlton, M., Fotheringham, S., and Brunsdon, C. 2009. Geographically weighted regression. White paper. National Centre for Geocomputation. *National University of Ireland Maynooth* 2.

Chica-Olmo, J. 2007. Prediction of housing location price by a multivariate spatial method: cokriging. *Journal of Real Estate Research* 29(1): 95-114.

Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical Association* 74(368): 829-836.

Clifton, K. J., Currans, K. M., and Muhs, C. D. 2015. Adjusting ITE's trip generation handbook for urban context. *Journal of Transport and Land Use* 8(1): 5-29.

De Smith, M. J., Goodchild, M. F., and Longley, P. (2007). Geospatial analysis: a comprehensive guide to principles, techniques and software tools. Troubador publishing ltd.

Ding, L., Niu, X, and Song, X. 2016. Measuring the employment center system in Shanghai central city: A study using mobile phone signaling data. *Acta Geographica Sinica* 71(3): 484-499.

Ewing, R., and Cervero, R. 2010. Travel and the built environment: A meta-analysis. *Journal of the American planning Association* 76(3): 265-294.

Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., 2015. Varying influences of the built environment on household travel in15 diverse regions of the United States. *Urban Study* 52: 2330–2348.

Feuillet, T., Charreire, H., Menai, M., Salze, P., Simon, C., Dugas, J., ... and Oppert, J. M. 2015. Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model. International journal of health geographics 14(1): 1-14.

Finley, A. O. 2011. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. Methods in Ecology and Evolution 2(2): 143-154.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley and Sons.

Fotheringham, A. S., Yang, W., and Kang, W. 2017. Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers* 107(6): 1247-1265.

Gao, F., Li, S., Tan, Z., Wu, Z., Zhang, X., Huang, G., and Huang, Z. 2021. Understanding the modifiable areal unit problem in dockless bike sharing usage and exploring the interactive effects of built environment factors. *International Journal of Geographical Information Science* 1-21.

Getis, A., and Ord, J. K. 2010. The analysis of spatial association by use of distance statistics. In Perspectives on spatial data analysis (pp. 127-145). Springer, Berlin, Heidelberg.

Gong, L., Liu, X., Wu, L., and Liu, Y. 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science* 43(2): 103-114.

Gu, H., Meng, X., Shen, T., and Wen, L. 2019. China's highly educated talents in 2015: Patterns, determinants and spatial spillover effects. *Applied Spatial Analysis and Policy* 1-18.

Gu, H., Yu, H., Sachdeva, M., and Liu, Y. 2021. Analyzing the distribution of researchers in China: An approach using multiscale geographically weighted regression. *Growth and*

Change 52(1): 443-459.

Haire, A. R., and Machemehl, R. B. 2007. Impact of rising fuel prices on US transit ridership. *Transportation Research Record: Journal of the Transportation Research Board* 1992(1): 11e19

Hall, J. D., Palsson, C., and Price, J. 2018. Is Uber a substitute or complement for public transit?. *Journal of Urban Economics* 108: 36-50.

Hu, S., Yang, S., Li, W., Zhang, C., and Xu, F. 2016. Spatially non-stationary relationships between urban residential land price and impact factors in Wuhan city, China. *Applied Geography* 68: 48–56.

Ishaque, M. M., and Noland, R. B. 2008. Behavioural issues in pedestrian speed choice and street crossing behavior: a review. *Transport Reviews* 28(1): 61-85.

Jiang, S., Guan, W., He, Z., and Yang, L. 2018. Exploring the intermodal relationship between taxi and subway in Beijing, China. *Journal of Advanced Transportation*, 2018: 1-14.

Kain, J. F., and Liu, Z. 1999. Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transportation Research Part A: Policy and Practice* 33(7): 601e624

Kanafani, A. 1983. *Transportation demand analysis*. New York: McGraw-Hill. Kuby, M., Barranda, A., and Upchurch, C. 2004. Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice*, 38(3): 223e247

Kong, X., Liu, Y., Wang, Y., Tong, D., and Zhang, J. 2017. Investigating public facility characteristics from a spatial interaction perspective: A case study of Beijing hospitals using taxi data. *ISPRS International Journal of Geo-Information* 6(2): 38.

Kuby, M., Barranda, A., and Upchurch, C. 2004. Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice* 38(3): 223e247.

Lao, X., and Gu, H. 2020. Unveiling various spatial patterns of determinants of hukou transfer intentions in China: A multiscale geographically weighted regression approach. *Growth and Change* 51(4): 1860-1876.

Liu, X., Sun, L., Sun, Q., and Gao, G. 2020. Spatial variation of taxi demand using GPS trajectories and POI data. *Journal of Advanced Transportation* 2020

Liu, X., Gong, L., Gong, Y., and Liu, Y. 2015. Revealing travel patterns and city structure with taxi trip data. *Journal of transport Geography*, 43, 78-90.

Louail, T., Lenormand, M., Ros, O. G. C., Picornell, M., Herranz, R., Frias-Martinez, E., ... and Barthelemy, M. 2014. From mobile phone data to the spatial structure of cities. *Scientific reports* 4(1): 1-12.

Martínez, M. G., Lorenzo, J. M. M., and Rubio, N. G. 2000. Kriging methodology for regional economic analysis: Estimating the housing price in Albacete. *International Advances in Economic Research* 6(3): 438-450.

Montero-Lorenzo, J. M., B., and A. 2009. Estimating commercial property prices: an application of cokriging with housing prices as ancillary information. *Journal of Geographical Systems*, 11(4), 407-425.

Montgomery, D., Peck, E., and Vining, G. 2012. *Wiley series in probability and mathematical statistics (2nd edn.)*. New York: John Wiley and Sons, 1992. Introduction to linear regression analysis.

Openshaw, S. and Taylor, P.J. 1981. *The modifiable areal unit problem*. Wrigley, N. and Bennett, R. (Ed.) Quantitative Geography: A British View (pp. 60–69), London: Routledge and Kegan Paul.

Ord, J. K., and Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis* 27(4): 286-306.

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., and Fotheringham, A. S. 2019. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information* 8(6): 269.

Paez, A., Scott, D., Potoglou, D., Kanaroglou, P., and Newbold, K. B. 2007. Elderly mobility: demographic and spatial analysis of trip making in the Hamilton CMA, Canada. *Urban Studies* 44(1): 123-146.

Qian, X., and Ukkusuri, S. V. 2015. Spatial variation of the urban taxi ridership using GPS data. *Applied geography* 59: 31-42.

Schaller, B. 2005. A regression model of the number of taxicabs in US cities. *Journal of Public Transportation* 8(5): 63.

Sha, Z., Xie, Y., Tan, X., Bai, Y., Li, J., and Liu, X. 2017. Assessing the impacts of human activities and climate variations on grassland productivity by partial least squares structural equation modeling (PLS-SEM). *Journal of Arid Land*, 9(4), 473-488.

Tang, J., Gao, F., Liu, F., Zhang, W., and Qi, Y. 2019. Understanding spatio-temporal characteristics of urban travel demand based on the combination of GWR and GLM. *Sustainability*, 11(19), 5525.

Tao, S., Corcoran, J., Mateo-Babiano, I., and Rohde, D. 2014. Exploring Bus Rapid Transit passenger travel behaviour using big data. *Applied geography*: 53, 90-104.

Tian, G., Wu, J., and Yang, Z. 2010. Spatial pattern of urban functions in the Beijing metropolitan region. *Habitat International* 34(2): 249-255.

Tu, M., Li, W., Orfila, O., Li, Y., and Gruyer, D. 2021. Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. *Transportation Research Part D: Transport and Environment* 93: 102776.

Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., and Li, Q. 2018. Spatial variations in urban public ridership derived from GPS trajectories and smart card data. *Journal of Transport Geography* 69: 45-57.

Wang, C. H., and Chen, N. 2017. A geographically weighted regression approach to investigating the spatially varied built-environment effects on community opportunity. *Journal of transport geography* 62: 136-147.

Wang, D., and Cao, X. 2017. Impacts of the built environment on activity-travel behavior: Are there differences between public and private housing residents in Hong Kong?. *Transportation Research Part A: Policy and Practice* 103: 25-35.

Wang, S., and Noland, R. B. 2021. Variation in ride-hailing trips in Chengdu, China. *Transportation Research Part D: Transport and Environment* 90: 102596.

Xu, Y., Yan, X., Liu, X., and Zhao, X. 2021. Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. *Transportation Research Part A: Policy and Practice* 144: 170-188.

Yang, F., Ding, F., Qu, X., and Ran, B. 2019. Estimating urban shared-bike trips with location-based social networking data. *Sustainability*, 11(11), 3220.

Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., and Zhang, L. 2018. Analysis of Washington, DC taxi demand using GPS and land-use data. *Journal of Transport Geography* 66: 35-44.

Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., Mai, K., and Zhang, Y. 2017. Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science* 31(6): 1220-1244.

Yu, H., and 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. *Journal of Transport Geography* 75: 147-163.

Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., and Wolf, L. J. 2020. Inference in multiscale geographically weighted regression. Geographical Analysis, 52(1), 87–106.

Yu, H., Pang, H., and Zhang, M. 2018. Value-added effects of transit-oriented development: The impact of urban rail on commercial property values with consideration of spatial heterogeneity. *Papers in Regional Science* 97(4): 1375-1396.

Yuan, Y., Raubal, M., and Liu, Y. 2012. Correlating mobile phone usage and travel behavior–A case study of Harbin, China. Computers, Environment and Urban Systems 36(2): 118-130.

Zhang, B., Chen, S., Ma, Y., Li, T., and Tang, K. 2020. Analysis on spatiotemporal urban mobility based on online car-hailing data. Journal of Transport Geography 82: 102568.

Zhang, C., Jia, S., and Yang, R. 2016. Housing affordability and housing vacancy in China: The role of income inequality. *Journal of Housing Economics* 33: 4-14.

Zhang, Z., Tan, S., and Tang, W. 2015. A GIS-based spatial analysis of housing price and road density in proximity to urban lakes in Wuhan City, China. Chinese Geographical Science 25(6): 775–790.

Zhou, J., and Long, Y. 2014. Jobs-housing balance of bus commuters in Beijing: Exploration with large-scale synthesized smart card data. *Transportation Research Record*, 2418(1), 1-10.

Zhu, P. 2012. Are telecommuting and personal travel complements or substitutes?. The Annals of Regional Science 48(2): 619-639.

Zhu, P. 2013. Telecommuting, household commute and location choice. *Urban Studies* 50(12): 2441-2459.

Zhu, P., and Mason, S. G. 2014. The impact of telecommuting on personal vehicle usage and environmental sustainability. *International Journal of Environmental Science and Technology* 11(8): 2185-2200.

Zhu, P., Ho, S., Jiang, Y. and Tan, X. (2020). Built environment, commuting behaviour and job accessibility in a rail-based dense urban context. *Transportation Research Part D: Transport and Environment*, Volume 87, October 2020, 102438.

Zhu, P., Zhao, S. and Jiang, Y. (2022a). Residential Segregation, Built Environment and Commuting Outcomes: Experience from Contemporary China. *Transport Policy*, 116: 269-277.

Zhu, P., Wang, K., Ho, S. and Tan, X. (2023). How is Commute Mode Choice Related to Built Environment in a High-density Urban Context? *Cities*. 134, 104180.

Appendix 1 Literature summary

Spatial relationship considered	Source	Methodology	Research objectives	Transportation Mode	Research Area
NO	Schaller, B. 2005. A regression model of the number of taxicabs in US cities. *Journal of Public Transportation,* 8(5): 63.	Multiple regression modeling	Identifying the factors that generate taxi demand in the United States.	Taxi	118 U.S. cities
NO	Liu, Y., Wang, F., Xiao, Y., and Gao, S. 2012. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106(1): 73-87.	Classification tree method; Correspondence analysis	Estimating associations between land use and traffic patterns.	Taxi	Shanghai
NO	Yang, C., and Gonzales, E. J. 2014. Modeling taxi trip demand by time of day in New York City. *Transportation Research Record*, 2429(1): 110-120.	Multiple linear regression model	Estimating taxi demand for each hour of the day.	Taxi	New York City
NO	Gong, L., Liu, X., Wu, L., and Liu, Y. 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2): 103-114.	Bayes' rules	Predicting trip purposes of taxi passengers.	Taxi	Shanghai
NO	Zhang, W., Ukkusuri, S. V., and Lu, J. J. 2017. Impacts of urban built environment on empty taxi trips using limited geolocation data. *Transportation*, 44(6): 1445-1473.	Hazard-based duration model	Exploring the determinants of empty taxi trip duration.	Taxi	New York City
NO	Wu, Z. and Zhuo, J. 2018. Impact of urban built environment on urban short-distance taxi travel: the case of Shanghai. *IOP conference series: earth and environmental science,* 153: s. 062019.	Multiple regression model	Analyzing the spatial impact of the built environment on short-distance taxi riders' travel behaviour.	Taxi	Central area of Shanghai
NO	Alemi, F., Circella, G., Handy, S., and Mokhtarian, P. 2018. What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. *Travel Behaviour and Society*, 13: 88-104.	Binary logit models	Investigating influence factors for on-demand ride service utility.	On-demand ride services like Uber and Lyft	California
NO	Liu, Q., Ding, C., and Chen, P. 2020. A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand.	Generalized additive mixed model (GAMM)	Exploring factors associated with temporal and spatial distributions for taxi pick-up densities.	Taxi	Central area of the Beijing

	Travel Behaviour and Society, 18: 29-36.				
NO	Yu, H., and Peng, Z. R. 2020. The impacts of built environment on ride-sourcing demand: A neighborhood level analysis in Austin, Texas. *Urban Studies*, 57(1): 152-175.	Structural equation model	Estimating the impacts of built environment on ride-sourcing demand	Uber	Austin
NO	Zhang, B., Chen, S., Ma, Y., Li, T., and Tang, K. 2020. Analysis on spatiotemporal urban mobility based on online car-hailing data. *Journal of Transport Geography*, 82: 102568.	Ordered logistic regression	Investigating the correlations between the intensity of ride-hailing services and POIs.	Ride-hailing service through Didi platform	Chengdu
NO	Wang, J., Yamamoto, T., and Liu, K. 2021. Spatial dependence and spillover effects in customized bus demand: Empirical evidence using spatial dynamic panel models. *Transport Policy*, 105: 166-180.	Multiple regression model	Exploring the spatial impact of the built environment on short-distance taxi riders' travel behavior.	Taxi	Central area of Shanghai
NO	Dean, M. D., and Kockelman, K. M. 2021. Spatial variation in shared ride-hail trip demand and factors contributing to sharing: Lessons from Chicago. *Journal of Transport Geography*, 91: 102944.	Panel regression	Analyzing the relationship among spatial distribution of shared transportation network company ridership, demographic characters and land use conditions.	Online ride-hail service (e.g., Uber, Lyft, and Via)	Chicago
NO	Tu, M., Li, W., Orfila, O., Li, Y., and Gruyer, D. 2021. Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. *Transportation Research Part D: Transport and Environment*, 93: 102776.	Machine learning method	Examining the non-linear effects of the built environment on ride pooling.	Ride-hailing service through Didi platform	Chengdu
Spatial autocorrelation	Correa, D., Xie, K., and Ozbay, K. 2017, January. Exploring the taxi and Uber demand in New York City: An empirical analysis and spatial modeling. *In 96th Annual Meeting of the Transportation Research Board*, Washington, DC.	Linear models, spatial error models, and spatial lag models	Identifying the influence factors for taxi and Uber ridership spatial distribution.	Taxi and Uber	New York City
Spatial autocorrelation	Lavieri, P. S., Dias, F. F., Juri, N. R., Kuhr, J., and Bhat, C. R. 2018. A model of ride-sourcing demand generation and distribution. *Transportation Research Record*, 2672(46): 31-40.	Spatially lagged multivariate count model; Fractional split model	Analyzing the spatial distribution of ride-sourcing trips generated on different days of a week and identifying the influencing factors.	Ride-sourcing	Austin
Spatial autocorrelation	Pan, R., Zhang, S., Yang, H., Xie, K., and Wen, Y. 2019, October. Analysis of Spatial Equity in Taxi Services: A Case Study of New York City. *In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 2659-2664)*. IEEE.	Linear models, spatial error models, and spatial lag models	Exploring the spatial equity of taxi services and the impact of e-hailing taxis on transport equity.	Traditional taxi services and e-hailing taxis	New York City

Spatial autocorrelation	Wang, M., Chen, Z., Mu, L. and Zhang, X. 2020. Road network structure and ride-sharing accessibility: Evidence from a network science perspective. *Computers Environment and Urban Systems*, 80: 101430.	Spatial Durbin model (SDM)	Investigating the relationship between road network structure and ride-sharing accessibility	Uber	Atlanta
Spatial autocorrelation	Ni, Y., and Chen, J. 2020. Exploring the Effects of the Built Environment on Two Transfer Modes for Metros: Dockless Bike Sharing and Taxis. *Sustainability*, 12(5): 2034.	K-means clustering; spatial lag model	Comparing the temporal-spatial distribution of dockless bike sharing and taxis as first/last-mile solutions and exploring how sociodemographic and built-environment factors influence their usage.	Dockless bike sharing (DBS) and taxis	Beijing
Spatial autocorrelation	Zhang, W., Le, T. V., Ukkusuri, S. V., and Li, R. 2020. Influencing factors and heterogeneity in ridership of traditional and app-based taxi systems. *Transportation*, 47(2): 971-996.	Mixture modeling structure of spatial lag and simultaneous equation model	Investigating the factors influencing traditional taxi and app-based taxi service demand considering spatial, temporal, and modal heterogeneity.	Traditional taxi services and app-based taxi service	New York City
Spatial autocorrelation	Wang, J., Yamamoto, T., and Liu, K. 2021. Spatial dependence and spillover effects in customized bus demand: Empirical evidence using spatial dynamic panel models. *Transport Policy*, 105: 166-180.	Spatial dynamic panel model	Modelling proposes customized bus (CB) demand by considering service supply level, demographic characteristics, land use and public service accessibility.	Customized bus	Dalian
Spatial heterogeneity	Qian, X., and Ukkusuri, S. V. 2015. Spatial variation of the urban taxi ridership using GPS data. *Applied geography*, 59: 31-42.	Geographically weighted regression (GWR)	Modeling the spatial heterogeneity of taxi ridership.	Taxi	New York City
Spatial heterogeneity	Nam, D., Hyun, K., Kim, H., Ahn, K., and Jayakrishnan, R. 2016. Analysis of grid cell–based taxi ridership with large-scale GPS data. *Transportation Research Record*, 2544(1): 131-140.	Geographically weighted regression (GWR)	Exploring spatial correlations among transit, urban density and taxi ridership.	Taxi	Seoul
Spatial heterogeneity	Li, B., Cai, Z., Jiang, L., Su, S., and Huang, X. 2019. Exploring urban taxi ridership and local associated factors using GPS data and geographically weighted regression. *Cities*, 87: 68-86.	Hierarchical clustering; Stepwise linear regression; Geographically Weighted Regression (GWR)	Exploring the spatial-temporal pattern and local associated factors of taxi trajectory.	Taxi	Majority of the metropolitan area of Beijing
Spatial heterogeneity	Yu, H., and Peng, Z. R. 2019. Exploring the spatial variation of ride-sourcing demand and its relationship to build environment and socioeconomic factors with the geographically weighted Poisson regression. *Journal of Transport Geography*, 75: 147-163.	Geographically Weighted Poisson Regression (GWPR)	Analyzing the impact of the built environment on ride-sourcing demand considering the spatial heterogeneity.	Ride-sourcing trip through Ride Austin platform	Austin
Spatial heterogeneity	Zhang, W., Le, T. V., Ukkusuri, S. V., and Li, R. 2020. Influencing factors and heterogeneity	Mixture modeling structure of spatial lag and simultaneous	Investigating the factors influencing traditional taxi and app-based taxi	Traditional taxi services and app-	New York City

	in ridership of traditional and app-based taxi systems. *Transportation*, 47(2): 971-996.	equation model	demand considering the spatial, temporal, and modal heterogeneity.	based taxi service	
Spatial heterogeneity	Yuan, C., Duan, Y., Mao, X., Ma, N. and Zhao, J. 2021. Impact of the mixed degree of urban functions on the taxi travel demand. *PLOS ONE*, 16 (3): s. e0247431.	Geographically Weighted Regression (GWR)	Investigating the relationship between the mixed degree of urban internal functions and the residents' taxi travel demand.	Taxi	Xi'an
Spatial heterogeneity	Chen, C., Feng, T., Ding, C., Yu, B., and Yao, B. 2021. Examining the spatial-temporal relationship between urban built environment and taxi ridership: Results of a semi-parametric GWPR model. *Journal of Transport Geography*, 96: 103172.	Semi-parametric Geographically Weighted Poisson Regression (SGWPR)	Exploring the determinants of urban taxi ridership.	Taxi	Shanghai
Spatial heterogeneity	Wang, S., and Noland, R. B. 2021. Variation in ride-hailing trips in Chengdu, China. *Transportation Research Part D: Transport and Environment*, 90: 102596.	Geographically Weighted Regression (GWR)	Investigating relations between spatial, social-economic factors and ride-hailing service demand.	Ride-hailing service through Didi platform	Chengdu

Appendix 2 VIF test results

	Morning peak taxi pick-ups	Morning peak taxi drop-offs	Evening peak taxi pick-ups	Evening peak taxi drop-offs
Employment density (log)		4.09	4.09	
Residential density (log)	3.27			3.27
Average housing price (log)	1.53	1.59	1.59	1.53
Share of 200-meter bus coverage in the grid cell	2.57	2.58	2.58	2.57
Share of 400-meter subway coverage in the grid cell	1.44	1.44	1.44	1.44
Road area ratio for each grid	1.96	1.94	1.94	1.93
# of buildings for each grid (log)	2.58	2.54	2.54	2.58
# of stories in buildings for each grid (log)	3.04	3.00	3.00	3.04
# of commercial-and-recreational POI (log)	5.25	5.22	5.22	5.25
# of manufacturing-and-office POI (log)	1.93	2.70	2.93	2.36
# of residence-and-related-facility POI (log)	5.43	5.25	5.25	5.43
# of public-management-and-service POI (log)	5.70	5.68	5.68	5.70
Mean VIF	3.19	3.28	3.28	3.19

Note: # represents the number of this variable.

Appendix 3 Regression results of GWR Model

(a) Morning Pick-up Model

Morning Pick-up Model				
Dependent variable: Number of morning-peak pick-ups (log)	GWR			
	Median	Min	Max	St. Dev.
Residential density (log)	0.128	-0.001	0.332	0.068
Share of 200-meter bus coverage in the grid cell	0.156	0.048	0.3	0.051
Share of 400-meter subway coverage in the grid cell	0.01	-0.048	0.116	0.033
Road area ratio for each grid	0.148	0.059	0.389	0.082
Average housing price (log)	0.038	-0.06	0.162	0.047
# of buildings for each grid (log)	0.002	-0.193	0.087	0.058
# of stories in buildings for each grid (log)	0.198	-0.001	0.297	0.071
# of commercial-and-recreational POI (log)	0.059	-0.149	0.269	0.079
# of manufacturing-and-office POI (log)	0.043	-0.038	0.217	0.065
# of residence-and-related-facility POI (log)	0.071	-0.106	0.259	0.072
# of public-management-and-service POI (log)	0.219	-0.064	0.571	0.111
Intercept	0.047	-0.154	0.475	0.144
Bandwidth	206			
AICc	396.421			
R^2	0.924			
Adj. R^2	0.912			

(b) Morning Drop-off Model

Morning Drop-off Model				
Dependent variable: Number of Evening-peak Drop-offs (log)	GWR			
	Median	Min	Max	St. Dev.
Employment density (log)	0.369	0.029	0.93	0.211
Share of 200-meter bus coverage in the grid cell	0.1	-0.083	0.207	0.059
Share of 400-meter subway coverage in the grid cell	0.047	-0.056	0.161	0.048
Road area ratio for each grid	0.152	0.009	0.346	0.072
Average housing price (log)	0.044	-0.159	0.2	0.08
# of buildings for each grid (log)	-0.033	-0.277	0.133	0.085
# of stories in buildings for each grid (log)	0.081	-0.108	0.246	0.061
# of commercial-and-recreational POI (log)	-0.046	-0.342	0.083	0.095
# of manufacturing-and-office POI (log)	0.084	-0.088	0.337	0.095
# of residence-and-related-facility POI (log)	-0.004	-0.311	0.275	0.124
# of public-management-and-service POI (log)	0.222	-0.026	0.612	0.114
Intercept	0.087	-0.203	0.627	0.161
Bandwidth	150			
AICc	475.122			
R^2	0.926			

Adj. R^2	0.909

(c) Evening Pick-up Model

Evening Pick-up Model				
Dependent variable: Number of morning-peak pick-ups (log)	GWR			
	Median	Min	Max	St. Dev.
Employment density (log)	0.237	0.035	0.902	0.159
Share of 200-meter bus coverage in the grid cell	0.141	-0.024	0.249	0.061
Share of 400-meter subway coverage in the grid cell	0.027	-0.059	0.141	0.038
Road area ratio for each grid	0.146	0.017	0.401	0.087
Average housing price (log)	0.041	-0.114	0.177	0.068
# of buildings for each grid (log)	-0.02	-0.227	0.144	0.082
# of stories in buildings for each grid (log)	0.078	-0.08	0.224	0.066
# of commercial-and-recreational POI (log)	0.135	-0.133	0.278	0.079
# of manufacturing-and-office POI (log)	0.119	-0.036	0.281	0.07
# of residence-and-related-facility POI (log)	0.001	-0.222	0.216	0.097
# of public-management-and-service POI (log)	0.126	-0.164	0.396	0.116
Intercept	0.06	-0.17	0.666	0.183
Bandwidth	150			
AICc	270.966			
R^2	0.945			
Adj. R^2	0.932			

(d) Evening Drop-off Model

Evening Drop-off Model				
Dependent variable: Number of Evening-peak Drop-offs (log)	GWR			
	Median	Min	Max	St. Dev.
Residential density (log)	0.134	-0.117	0.387	0.1
Share of 200-meter bus coverage in the grid cell	0.114	-0.025	0.198	0.041
Share of 400-meter subway coverage in the grid cell	0.033	-0.023	0.094	0.027
Road area ratio for each grid	0.121	0.037	0.286	0.067
Average housing price (log)	0.041	-0.098	0.197	0.06
# of buildings for each grid (log)	0.121	0.037	0.286	0.067
# of stories in buildings for each grid (log)	-0.014	-0.277	0.12	0.068
# of commercial-and-recreational POI (log)	0.220	0.216	0.223	0.002
# of manufacturing-and-office POI (log)	0.082	0.079	0.089	0.003
# of residence-and-related-facility POI (log)	0.062	0.056	0.070	0.004
# of public-management-and-service POI (log)	0.106	-0.220	0.287	0.077
Intercept	0.035	-0.18	0.606	0.168
Bandwith	166			
AICc	254.677			

R^2	0.944
Adj. R^2	0.932

Note: (1) Note: # represents the number of this variable; (2) * p-value < 0.05; ** p-value <0.01; *** p-value <0.001