JTLU

Exploring multi-scale spatial relationship between built environment and public bicycle ridership: A case study in Nanjing

Cheng Lyu
Southeast University
cheng.lyu@seu.edu.cn

Xinhua Wu
Southeast University
xinhua.wu@seu.edu.cn

Yang Liu
Southeast University
seu_yangliu@seu.edu.cn

Xun Yang
Southeast University
yx_97@seu.edu.cn

Zhiyuan Liu (corresponding author)
Southeast University
zhiyuanl@seu.edu.cn

Abstract: A public bicycle system (PBS) is a promising countermeasure for the traffic issues induced by rapid urbanization, and it is widely acknowledged that the built environment has a significant impact on the use of a PBS. However, as the urban built-up area expands, different regions within a city can exhibit diverse characteristics. The spatial effects and differences among regions have been neglected by existing studies. To better understand how the urban built environment affects PBS ridership, this study conducts a quantitative analysis of the spatial relationship. It introduces a multi-scale geographically weighted regression (MGWR) to accomplish this task and conducts and evaluates a case study of the PBS in Nanjing, China. Six types of "D" variables (density, diversity, design, destination accessibility, distance to transit, and demand management) are involved in the analysis. The proposed method outperforms linear regression and standard geographically weighted regression (GWR) in terms of explanatory power. The modeling results demonstrate different influencing patterns between traditional downtown areas and newly built-up areas, especially for the density of population, road network, parking space, and various points of interest.

1   Introduction

With the accelerating pace of urbanization in China, the soaring population in urban areas has posed unprecedented pressure on the sustainability of cities, especially traffic issues. Public bicycle system (PBS) is an important instrument to reduce the traffic

congestion and emission in urban transportation system, by shifting travelers from using private cars to shared mobility and public transport (Liu, Jia, Xie, & Liu, 2019). A station-based public bicycle system allows users to rent a bicycle from any docking station and return it to another without worrying about the conservation and maintenance of bikes (Shaheen, Guzman, & Zhang, 2010). By the end of 2018, nearly 2,700 bike-sharing schemes have been deployed around the world (DeMaio & Meddin, 2019). PBS can serve as a crucial complement to public transport services (Yang et al., 2018) and can be synergized with strategies such as congestion pricing and park-and-ride (Liu, Wang, Zhou, & Cheng, 2017; Liu, Chen, Meng, & Kim, 2018; Shaheen & Rodier, 2018; Huang, Liu, Zhu, Kim, & An, 2019). By providing more bicycles to the public, commuters can be significantly encouraged to cycle to access public transport (de Souza, La Paix Puello, Brussel, Orrico, & van Maarseveen, 2017). As a travel mode for short-distance trips, PBS is able to solve the last-mile issue, thus enhancing the accessibility and sustainability of the entire urban transportation system.

It is widely acknowledged that built environment has a significant impact on the mode choices of travelers (Boarnet & Crane, 2001; Cervero, 2002; Ding, Wang, Liu, Zhang, & Yang, 2017; Ewing & Cervero, 2010). As defined by Handy, Boarnet, Ewing, and Killingsworth (2002), built environment includes the urban design, land use and the transportation system, in which the patterns of human activities are highly related to the corresponding physical environment. Much effort has been devoted to investigating the connection between built environment and public bicycle ridership (Bachand-Marleau, Lee, & El-Geneidy, 2012; El-Assi, Salah Mahmoud, & Nurul Habib, 2017; Faghih-Imani & Eluru, 2016; Zhang, Thomas, Brussel, & van Maarseveen, 2017), but few of them attempted to allow for the spatial heterogeneity and nonstationarity. The assumption of stationarity over space has long been realized inappropriate (Anselin, 1995; Bhat & Zhao, 2002; Ding, Wang, Yang, Liu, & Lin, 2016). In other words, it is unreasonable to expect that one factor will have exactly the same influence on the dependent variable in different locations. For example, the opening of a new shopping mall in the city center will induce much more traffic demand than that in the suburbs. Neglecting the local variations might cause reduced reliability of model and misinterpretation of the relationship between study variables (Qian & Ukkusuri, 2015). Geographically weighted regression (GWR) was proposed to deal with spatial nonstationarity (Brunsdon, Fotheringham, & Charlton 1996). However, GWR makes the assumption that all the explanatory variables share the same spatial scale. Also, it should be noted that the process of urbanization usually comes along with the expansion of the built-up area. It can be inferred that built environment elements can generate diverse effects on PBS usage in different locations of a large city. Instead of explaining the movements of urban mobility on a universal scale, the multi-scale geographically weighted regression (MGWR) allows the parameters to arise from processes of various spatial scale, so as to capture more accurate spatial relationship (Fotheringham, Yang, & Kang, 2017).

The aim of this study is to investigate the multi-scale spatial relationship between the urban built environment and the ridership of PBS. Through the introduction of the MGWR model, we manage to bring a new perspective to this topic. The common assumption holds by ordinary linear regression–built environment factors impact public bicycle usage in the same manner across the whole city–is relaxed. Moreover, the spatial variation of different factors is allowed to vary in scale compared with traditional GWR. To validate our hypothesis and evaluate the modeling results, a case study of PBS is conducted in Nanjing, China. The variations of those factors over space are verified and visualized, which can provide more reliable guidance to the development of PBS.

The following of this paper is organized as follows. In Section 2, the literature is reviewed to extract important factors of the urban built environment that might affect public bicycle usage, including the six D variables, i.e., the density, diversity, design, destination accessibility, distance to transit, and demand management. Then, Section 3 elaborates the formulation and estimation of the MGWR model.

The data and results are presented and analyzed in Section 4 and 5 respectively. Finally, Section 6 concludes the major findings and provides directions for future research.

2   Literature review

Many existing studies have attempted to investigate the influencing factors of public bicycle ridership, which can be split into three categories, namely personal preference, temporal variation, and spatial features. The studies for these three categories are reviewed as follows.

First, for personal preference, some researchers based their studies on an individual perspective to better understand the effects of personal predilections and socio-economic factors. It was found by Schoner and Levison (2013) that potential users of the public bicycle are sensitive to the walking distance to the docking station, and prefer longer cycling distance to walking distance. Raux, Zoubir, and Geyik (2017) also highlighted that the proximity to the docking stations positively affects the likelihood of an individual to use the bike based on an analysis of the bike sharing scheme Velo'v in Lyon. In a survey conducted by de Souza et al. (2017) in Brazil, the bicycle infrastructure was found to be one of the critical barriers for bicycle use as a feeder to public transit.

Second, regarding temporal variation, the usage patterns are found to be associated with weather conditions, holidays and operating time. The analysis by Gebhart and Noland (2014) suggested that high precipitation, wind speed, humidity and low temperature lead to fewer PBS ridership. The negative impacts of adverse weather conditions are also verified by Kutela and Teng (2019) as well as Frade and Ribeiro (2014). Without considering extreme temperatures, El-Assi et al. (2017) concluded a simple positive relationship between temperature and trip demand. Due to commuting demand, the ridership of weekdays is observed to be larger than weekends, and the highest usage usually shows up in evening peaks (Faghih-Imani, Hampshire, Maria, & Eluru, 2017; Faghih-Imani & Eluru, 2016; O'Brien, Cheshire, & Batty, 2014). Moreover, the study by Kutela and Teng (2019) also found that longer years in operation can contribute to increased daily ridership.

Third, spatial features are usually considered in the form of built environment. Research on the spatial relationship of the built environment and PBS ridership are classified and listed in Table 1. As mentioned in Section 1, the built environment is well acknowledged to have a tangible impact on PBS usage. The measures of the built environment were summarized by Cervero and Kockelman (1997) as "three Ds," i.e., density, diversity, and design. Beyond the classical "three Ds," three more Ds were later introduced, including destination accessibility, distance to transit, and demand management (Ewing & Cervero, 2010). As Ewing and Cervero (2010) pointed out, the "Ds" are merely rough categorization without determined boundary. Hence it is possible that one factor can be placed under multiple classes. It should also be noted that only factors relating to the built environment are included in the table.

Among all built environment factors, density is the most widely recognized. Almost every study took population density into consideration, though it was found to be insignificant in the research of Mateo-Babiano, Bean, Corcoran, and Pojani (2016). Points of interest (POI) is another class of representative factors, including commercial area (e.g., stores, shopping malls), employment (e.g., companies), residential area, as well as recreational area (e.g., parks, cinemas).

Road network density also plays an essential role in generating PBS usage. Both Faghih-Imani, Eluru, El-Geneidy, Rabbat, and Haq (2014) and Zhang et al. (2017) split the roads into major roads and branch roads. The results are similar, that branch roads positively impact the ridership.

Furthermore, competition among adjacent docking stations was recognized (El-Assi et al., 2017; Zhang et al., 2017; Zhao, Deng, & Song, 2014). Diversity was only explicitly investigated in one literature, and was found to be positively correlated to ridership (Zhang et al., 2017). Oddly, they labeled

each docking station with a unique land-use type in another research (Zhang, Brussel, Thomas, & van Maarseveen, 2018). The reason why this variable is not included in most studies might be the overlap with POI density.

Research pertaining to road design mainly focused on cycling infrastructure. Many evidence has shown the effect of bike lanes in attracting travelers to PBS (Buck & Buehler, 2011; Mateo-Babiano et al., 2016; Wang et al., 2018; Zhang et al., 2017). It was also observed that high elevation, which usually comes with steep slopes, is a hindrance to cycling (Faghih-Imani et al., 2017; Mateo-Babiano et al., 2016).

The other three "Ds" correspond with the location of the station and neighboring transportation facilities. Typical measure in prior works for destination accessibility is the distance to the city center. According to the study on BIXI in Montreal by Faghih-Imani et al. (2014), stations locating at the periphery of the city are less active compared to those near the central business district (CBD). This is consistent with the results of Zhang et al. (2017).

Also, a reciprocal relationship exists between cycling and public transit, hence the reason for examining the impacts of public transit on PBS (Kager, Bertolini, & Brömmelstroet, 2016). People tend to be attracted to combine public bicycles and transit. In the surveys conducted by Lin et al. (2018), shorter distance to transit stations will decline the use of public bikes. However, this conclusion was challenged by Tran, Ovtracht, and d'Arcier (2015) and Zhang et al. (2017), and they attributed the insignificance of the transit system to small city scale and slightly higher bus fare. In contrast, in the research of Tran et al. (2015), railway stations showed a significant relationship with PBS ridership for its role in commuting. Finally, no existing study managed to investigate the potential connection between public bicycle and parking lots.

As to the modeling tool, most existing studies drew their conclusions using linear regression. The major disadvantage of linear regression in this task is that it fails to give consideration to the effect of spatial dependency. The planning and management of bike-sharing systems, including siting of docking stations and repositioning of bicycles, usually requires understanding the heterogeneity of cycling demand over space. A few studies attempted to take spatial effects into consideration. Zhang et al. (2017) included a spatially lagged dependent variable in their model. This practice, however, merely takes the PBS demand in the vicinity of a station as an additional input. Similarly, the spatial lag model and spatial error model employed by Faghih-Imani and Eluru (2016) are still incapable of reflecting the changes in spatial relationship over space. Geographically weighted regression (GWR) is precisely designed for this purpose (Brunsdon et al., 1996), which is able to incorporate the spatial autocorrelation in a given area, and has been employed to study the impact of built environment on transit ridership (Qian & Ukkusuri, 2015). Further, the standard GWR assumes the spatial relationships are on the same scale. However, in reality, it is common that some of the parameters are homogeneous across space whilst the others are heterogeneous. To overcome this defect, in this paper, a multi-scale GWR (MGWR), which allows the scale for the coefficients of each covariate to vary, is utilized to model the relationship between built environment and public bicycle ridership.

Table 1. Literature on the relationship between built environment and PBS ridership

Ds	Built Environment Factor	Buck & Buehler, 2011	Bachand-Marleau et al., 2012	Faghih-Imani et al., 2014	Zhao et al., 2014	Tran et al., 2015-	Faghih-Imani & Eluru, 2016	Mateo-Babiano et al., 2016	El-Assi et al., 2017	Faghih-Imani et al., 2017	Zhang et al., 2017	Total
Density	population	✓	–	✓	✓	✓	✓	✓	✓	✓	✓	9
	commercial	✓	–	✓	–	✓	✓	–	–	✓	✓	6
	employment	–	–	✓	✓	✓	✓	–	✓	✓	–	6
	residential	–	–	–	–	✓	–	–	–	–	✓	2
	education	–	–	✓	–	✓	–	–	✓	✓	–	4
	recreational park	–	–	✓	–	✓	✓	–	–	✓	✓	5
	hotel	–	–	–	–	–	–	–	–	✓	–	1
	road	–	–	✓	–	✓	–	–	–	–	✓	3
	other bike stations	–	–	–	✓	–	–	–	✓	–	✓	3
Diversity	land use	–	–	–	–	–	–	–	–	–	✓	1
Design	cycling infrastructure	✓	–	–	–	–	✓	✓	–	–	✓	4
	elevation	–	–	–	–	–	–	✓	–	✓	–	2
	intersection	✓	–	–	–	–	–	–	✓	–	–	2
Destination Accessibility	location of the station	✓	✓	✓	–	–	–	–	–	–	✓	4
Distance to Transit	distance to transit stations	–	–	–	–	–	–	–	–	–	✓	1
	number of transit stations	–	✓	✓	–	✓	✓	–	✓	✓	✓	7
	number of rail stations	–	–	–	–	✓	–	–	–	–	–	1
Demand Management	parking space	–	–	–	–	–	–	–	–	–	–	0

3   Methodology for modeling spatial relationship

3.1   Multi-scale geographically weighted regression

In order to incorporate spatial variations of the relationship between variables, an intuitive solution is to allow the coefficients to vary across the study area (Brunsdon et al., 1996). Geographically weighted regression (GWR) is an extension of the traditional linear regression. Spatial weights are introduced into the model to describe the spatial relationship. Denote the location of observation point as (p_i, q_i). The coefficients β_k, which are fixed values for all the observations, are assumed to be functions of geographical locations, denoted by β_k (p_i, q_i). The general formulation of GWR is provided in Equation (1):

$$y_i = \sum_{k=0}^{m} \beta_k(p_i, q_i)\, x_{i,k} + \varepsilon_i, \quad i = 1,2,\dots,n \tag{1}$$

where m denotes the number of explanatory variables, n denotes the number of observations, and ε_i is the error term.

The coefficients of the GWR can be estimated through the weighted least square method. Denote the dependent variables as $Y=[y_1, y_2, \dots, y_n]'$ and the explanatory variables as $X=[X_1', X_2', \dots, X_n']'$, where $X_i=[x_{i,1}, x_{i,2}, \dots, x_{i,m}]$. The local coefficients can be given by Equation (2):

$$\hat{\beta}(p_i, q_i) = (X'W_iX)^{-1}X'W_iY \tag{2}$$

where $W_i = diag(w_{i1}, w_{i2}, \dots, w_{in})$ is a diagonal weighting matrix.

However, the regression coefficients do not necessarily change according to their geographical locations. In other words, some of them might have constant effects instead of being relatively non-stationary over space. By allowing coefficients to be either global or local, semi-parametric geographically weighted regression (SGWR) model, also known as mixed GWR model, can be formulated as Equation (3):

$$y_i = \sum_{k} \beta_k(p_i, q_i)x_{i,k} + \sum_{l} \gamma_l x_{i,l} + \varepsilon_i, \quad i = 1,2,\dots,n \tag{3}$$

where β_k denotes local coefficient, and γ_l denotes global coefficient.

Multi-scale geographically weighted regression (MGWR) further extends the GWR by relaxing the assumption of all the model coefficients vary at the same spatial scale (Fotheringham et al., 2017). A unique bandwidth bw_k will be assigned to each dependent variable. The formulation of MGWR is given in Equation (4).

$$y_i = \sum_{k=0}^{m} \beta_k(p_i, q_i; bw_k)\, x_{i,k} + \varepsilon_i, \quad i = 1,2,\dots,n \tag{4}$$

3.2   Scale in estimation

To allow for the multiple spatial scales of coefficients, MGWR selects a unique bandwidth for each explanatory variable. In this paper, the Gaussian weighting kernel is utilized to construct the spatial weight matrix. This kernel is set to adaptively interact with a given number of nearest neighbors. An adaptive Gaussian kernel is given by Equation (5):

$$w_{ij} = \begin{cases} \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right), & \text{if } d_{ij} < h_i \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where d_{ij} measures the distance between two points (p_i, q_i) and (p_j, q_j), and h_i is the longest distance from current point (p_i, q_i) to its neighbors. The optimal number of neighbors (i.e., the optimal bandwidth) for the kernel is determined through the golden section search. The criterion of the search is the corrected Akaike Information Criterion (AICc). Denote the estimation for the standard deviation of the error term as $\hat{\sigma}$, defined as Equation (6). The bandwidths that generate the lowest AICc are chosen as the optimal bandwidths:

$$\text{AICc} = 2n \ln \hat{\sigma} + n \ln 2\pi + n \frac{n + \text{tr}(S)}{n - \text{tr}(S) - 1} \tag{6}$$

where S denotes the hat matrix, and $\text{tr}(\cdot)$ represents the trace of a matrix.

3.3   Model calibration

The calibration method for GWR is not applicable in the MGWR case, in that the spatial weighting matrices for the explanatory variables at one location are different. Instead, the MGWR can be viewed as a generalized additive model (GAM), where the dependent variable can be written as the sum of multiple smooth functions of the explanatory variables. That is, $y_i = \sum_{k=0}^{m} f_k(x_{i,k}) + \varepsilon_i$, where $f_k(x_{i,k}) = \beta_k(p_i, q_i; bw_k) x_{i,k}$ is the additive term. One solution to GAM is the back-fitting algorithm (BF). Fotheringham et al. (2017) managed to adapt the BF algorithm to solving MGWR.

Assuming all of other additive terms are already known, the BF algorithm recursively calibrates each additive term until convergence. To boost the calibration speed, a standard GWR is pre-fitted to obtain a series of initial additive terms. Based on the initialization, the residual of the model can be calculated as $\varepsilon = y - \sum_k f_k$. Then, for a specific dependent variable x_j, GWR can be formulated as Equation (7).

$$\varepsilon_i + f_{i,j} = \beta_j(p_i, q_i; bw_j) x_{i,j} + \varepsilon_i', \quad i = 1, 2, \dots, n \tag{7}$$

By solving the GWR, a new bandwidth b_{wj} and a set of corresponding parameters can be estimated, whereby the f_j can be updated. Following this process, the bandwidths and the coefficients of all the covariates will be updated in each iteration.

4   Data description

4.1   Data source

A case study was conducted based on the cycling data of PBS in Nanjing, China, provided by Nanjing Public Bicycle Company. Nanjing PBS (NPBS) was started in 2013 and has developed to cover all the municipal districts in Nanjing. Five operating sub-regions, namely Core Urban District, Pukou District, Jiangning District, Gaochun District, and Lishui District, run independently from each other. The analysis of our research is based on the data of Core Urban District.
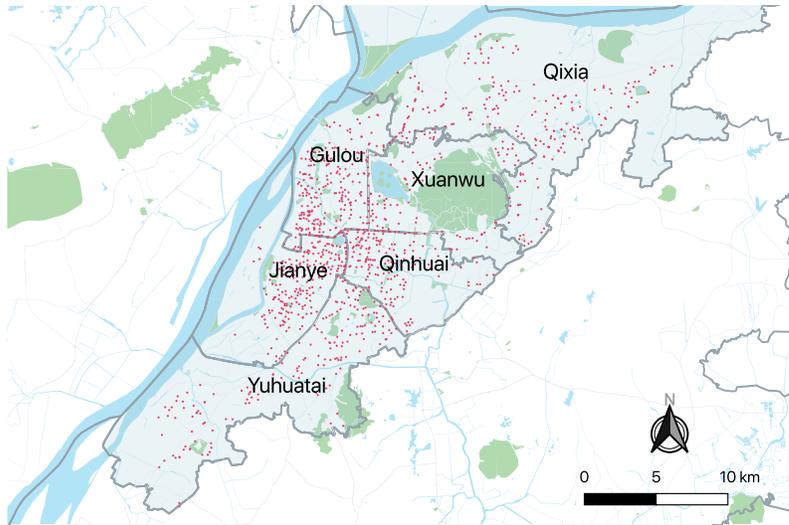
Figure 1. Docking stations of NPBS in the core urban district

The extracted dataset was collected from July to December in 2016, with 179 days (5 days are missing due to system failure), 1,064 docking stations, and more than 22 million trip records in total. The dataset contains the temporal information (start time and end time) as well as the spatial information (latitude, longitude and name of start and end stations) of each trip. Figure 1 demonstrates the location of PBS docking stations in the Core Urban District, where each station is illustrated as a red dot. The basic geometric data, including the road network and the POIs, were provided by Nanjing Institute of City & Transport Planning.

The raw dataset was cleansed to remove the abnormal trips. Firstly, only trips with their cycling duration between 1 minute to 24 hours were reserved, in order to exclude the influence of system failure. Secondly, docking stations with no trip records were removed. Thirdly, to alleviate the impacts of unpleasant weather, trips made on days with heavy rain or snow were removed. The data size after validation and cleansing is 19,708,241.

4.2   Variable selection

The average hourly pick-ups are used as the dependent variable. The trip dataset of working days was first aggregated by the morning peak (from 7:00 a.m. to 9:00 a.m.), the evening peak (from 5:00 p.m. to 7:00 p.m.) and the off-peak (other time periods beyond peak hours). The Pearson correlation coefficients (PCC) between the three subsets of data were calculated, as listed in Table 2. Coefficients larger than 0.9 are marked in bold. The pick-ups and drop-offs in peak hours are highly correlated with their opposite behaviors in another peak, indicating the vital commuting function of NPBS. It can also be

observed that the cycling demand at off-peak hours is similar to the evening peak. Therefore, only the pick-ups of the two peak periods are modeled. Furthermore, the pick-ups and drop-offs on non-working days (from 6:00 a.m. to 11:00 p.m.) have a PCC of 0.991. Hence, only the average of hourly pick-ups on non-working days is later modeled. The summary of the selected dependent variables is presented in Table 3. In Figure 2, the distribution of the raw dependent variables is plotted on the first row, while that of the log-transformed variables is plotted on the second row. Considering the long tails of the distributions, the logarithms of average pick-ups were finally taken as the dependent variables to lower the risk that the residuals of the models violate the normality assumption. As plotted in Figure 2, the dependent variables are then transformed as $\log(1+y)$, where y denotes the number of pick-ups.

Table 2. Pearson correlation coefficients of the trip data on different time periods*

	Drop-off (off)	Pick-up (off)	Drop-off (p.m.)	Pick-up (p.m.)	Drop-off (a.m.)	Pick-up (a.m.)
Pick-up (a.m.)	0.876	0.751	**0.943**	0.622	0.601	**1.000**
Drop-off (a.m.)	0.719	0.881	0.611	**0.968**	**1.000**	
Pick-up (p.m.)	0.782	**0.937**	0.671	**1.000**		
Drop-off (p.m.)	**0.953**	0.820	**1.000**			
Pick-up (off)	**0.928**	**1.000**				
Drop-off (off)	**1.000**					

* All the values are statistically significant at the 99% confidence level.

Table 3. Descriptive statistics of the average hourly pick-ups

	Pick-up (a.m.)	Pick-up (p.m.)	Pick-up (n.w.)
mean	9.545	8.759	1.571
std	10.799	11.746	1.554
min	0.000	0.000	0.002
lower quartile	2.007	1.845	0.447
median	6.457	5.227	1.182
upper quartile	13.338	11.162	2.216
max	100.855	119.224	11.452
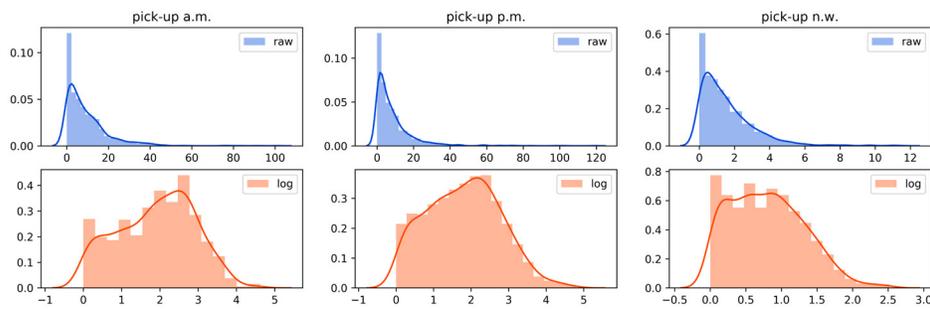
Figure 2. Distribution of the dependent variables

The built environment variable selection is based on the six D variables. Thirteen factors are finally incorporated in this study, as listed in Table 4. Based on prior research (see Table 1), the density variables considered in our analysis include the population density, POI density, and neighboring docking station density. Specifically, the POIs used in the analysis include the commercial establishment, office building, residential communities, schools, and recreational parks. Zonal population statistics were extracted from the yearbook published by Nanjing Municipal Bureau of Statistics (2017). The selected POIs are aggregated and counted using a 300-meter buffer, which is similar to prior works (Faghih-Imani et al., 2014; Tran et al., 2015). An oversized buffer might not be able to reflect the unique built environment around a specific docking station. 300m is an acceptable threshold of walking distance to a docking station, as the first quantile of cycling distance is merely 900m (Euclidean distance).

Table 4. Descriptive statistics of built environment variables

	mean	std	min	lower quartile	median	upper quartile	max
Population density	10905.58	8972.83	1815.49	3241.46	5558.65	20400.41	23529.19
Commercial density	58.20	84.03	0.00	9.00	32.00	74.00	789.00
Employment density	22.92	42.43	0.00	2.00	8.00	21.00	427.00
Residential density	5.91	4.92	0.00	2.00	5.00	9.00	29.00
School density	1.53	1.70	0.00	0.00	1.00	2.00	10.00
Park density	0.41	1.00	0.00	0.00	0.00	0.00	8.00
ke station density	1.40	1.43	0.00	0.00	1.00	2.00	8.00
Land-use diversity	2.16	0.97	0.00	2.00	2.00	3.00	4.00
Bus station density	7.91	6.27	0.00	3.00	7.00	12.00	32.00
Metro station density	0.14	0.38	0.00	0.00	0.00	0.00	2.00
Branch road density	12885.63	8845.32	12.37	6192.58	10961.32	17835.24	40047.26
Bike lane density	37507.40	14464.37	5438.13	27148.96	36866.17	44844.40	82691.39
Parking space density	3.12	3.62	0.00	0.00	2.00	4.00	23.00

Unlike the POI density, the density of the road network, including bike lanes, is calculated on the basis of a 1,500-meter buffer. The rationale behind this practice is that distant establishments do not directly affect the travel mode choice and docking station choice of travelers. However, the roads are always involved in the cycling trip of travelers, thus cannot be ignored. Zhang et al. (2017) picked 1,000 meters as the buffering threshold. Considering the median Euclidean trip distance of our dataset is 1,524m, 1,500m is appropriate as the buffer for roads in our study. Note that the main road density is not included in our analysis in that it has a PCC of 0.966 with bike lane density. In Chinese cities, especially big cities like Nanjing, separated bike lanes are usually constructed along with primary and secondary roads following the suggestions of the urban road design manual, leading to such a high correlation. Likewise, hotel density is also not included, due to the high correlation with commercial establishments.

The land-use diversity, which is often neglected in the existing literature on PBS, is measured through the count of land-use types within the 300-meter buffer. There are six types of land use that are considered, including commercial land, residential land, public service land, industrial land, transportation hub and green space. Destination accessibility used to be measured as the distance to the city center. This measure is abandoned in this research as the definition of the city center is ambiguous and Nanjing has developed into a multi-core city with distributed commercial centers. Therefore, it is unreasonable to represent the city center as a single point.

The locations of public transport stations were collected in 2016 using the API of Amap (2016). The public transport system of Nanjing is comprised of bus, metro, and tram. As the tram was still an experimental travel mode and operated in the southern part of Jianye District, where very few PBS stations were installed, it was not included in our analysis. Within the study area, there were 5623 bus stations and 67 metro stations in total. By 2016, six metro lines (i.e., Line 1, 2, 3, 10, S1 and S8) had been put into operation.

The parking space density is another factor that was not investigated in prior works. Though PBS was first promoted to enhance the experience of public transport, it can be interesting to examine whether cars can be integrated with PBS.

5 Result analysis

5.1 Model comparison

For each dependent variable, three models, namely LR, GWR, and MGWR, are fitted on the same explanatory variables. All the explanatory variables were normalized before inputting into the models. The regressions are performed using the package mgwr in Python (Oshan, Li, Kang, Wolf, & Fotheringham, 2018), and the modeling results are demonstrated in Table 5. Note that the park density is excluded from the inputs since it is statistically insignificant in all models. It can be observed that the MGWR model significantly outperforms the other two models. For all the three dependent variables, approximately 75% of the variance can be explained by the MGWR models. In comparison, only 47.4%, 57.2% and 53.3% of the variance can be explained by global linear regression models for the three targets respectively. As to the AICc, the criterion used for bandwidth searching, the three MGWR models are much lower than the other two models. Regarding the residual sum of squares, the MGWR models reduce those of LR models by half, and are also much lower than the GWR models.

The spatial autocorrelation of residuals is also tested using the index of Moran's I. The global Moran's I can be calculated using the following formula,

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{J=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{8}$$

where w_{ij} denotes the spatial weight between two points (p_i, q_i) and (p_j, q_j), $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ is the sum of spatial weights between any two locations, z_i denotes the difference between the observed attribute of a point and the average value of the whole region. For the fitted LR models, the global Moran's I of their residuals are 0.230, 0.188 and 0.276, respectively, indicating relatively high spatial autocorrelation. As to the MGWR models, the global Moran's I of their residuals are much lower (0.041, 0.074 and 0.054 respectively), implying that the spatial heterogeneity is successfully incorporated.

The descriptive statistics of the estimated parameters of the three MGWR models are given in Appendix A. As elaborated in Section 3, a GWR/MGWR model will give a set of estimated coefficients for each location in the study area. In our case, there are in total 1,025 docking stations, indicating that there will be 1,025×13=13,325 estimated values for each model. Rather than listing all these values in a table, the results are illustrated in maps, and are further discussed in the following section.

Table 5. Descriptive statistics of all the fitted models

Dependent Variable	Pick-up (a.m.)			Pick-up (p.m.)			Pick-up (n.w.)		
Model	LR	GWR	MGWR	LR	GWR	MGWR	LR	GWR	MGWR
Residual sum	536.146	424.505	254.164	412.104	335.971	218.474	126.592	108.046	69.090
of squares	-1122.291	-1002.631	-739.747	-987.436	-882.757	-662.200	-382.529	-301.342	-72.182
Log-likelihood	2270.582	2071.105	1566.836	2000.873	1831.358	1403.764	791.057	656.301	225.668
AIC	2272.998	2073.359	1570.816	2003.288	1833.612	1407.044	793.473	657.796	229.111
AICc	0.480	0.588	0.754	0.577	0.655	0.776	0.539	0.606	0.748
R^2	0.474	0.575	0.743	0.572	0.644	0.767	0.533	0.596	0.738
Adj. R^2	0.474	0.575	0.743	0.572	0.644	0.767	0.533	0.596	0.738
Observations	1025								
Covariates	13								

5.2   Analysis

The estimated coefficients of the three MGWR models are illustrated in Figure 3 through Figure 9, where the color bar to the right of each figure shows the scale and range of estimated values. Divided by white, positive values and negative values are marked by red and blue respectively. Circles and triangles are used to assist the identification of positiveness for the color blind. Also, the estimated values that are insignificant at the 95% confidence level, i.e., p-values greater than 0.05, are labeled with grey outlines in the figures.

The discrepancy between the estimated coefficients of models in the morning peak and the evening peak has verified the commuting function of PBS. Figure 3 shows the coefficients of residential community density and employment density. It can be clearly observed that the residential community density strongly correlates with the PBS ridership in the morning peak, while in the evening peak, the relationship becomes mostly neutral. The average estimate has dropped from 0.329 in the morning to 0.004 in the afternoon. We can infer that the public bicycle is one of the preferred travel modes for home-based commuting trips. In contrast, as to the employment density, negative coefficients can be observed in the major part of the city during the morning peak. Positive relationship between the employment density and PBS ridership shows up in the outer ring of the city in the evening peak. The mean of its estimated coefficients increased from -0.110 in the morning peak to 0.046 in the evening peak, indicating that the

employment density has a larger influence on PBS usage after work. The influencing pattern on non-working days are similar to that of the morning peak, but the absolute estimates are much smaller due to less overall ridership.
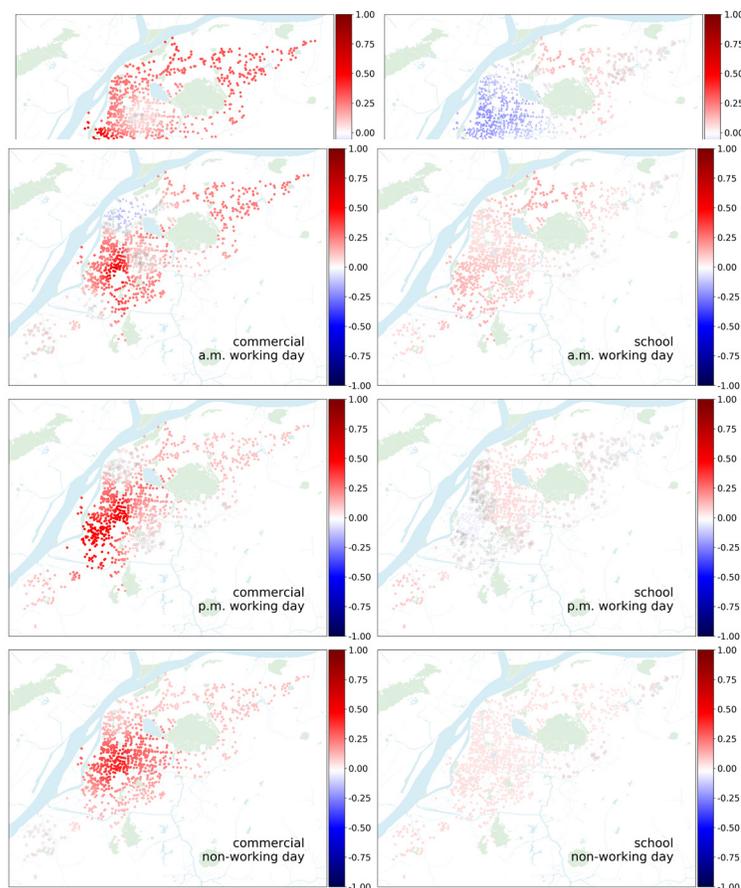
Figure 4. Estimated coefficients of the commercial density and the school density

Commercial establishments and schools are also two potential motivators for cycling. From Appendix A, it can be noticed that the coefficients of these two factors are both mostly positive. This result can also be easily noticed in Figure 4, where the colors of points are red across almost the entire city. As can be expected, the impact of the commercial density is rather significant, and such impact is consistent in all three studied time periods. Although higher school density also leads to higher PBS ridership, its impact is much smaller than that of the commercial density.
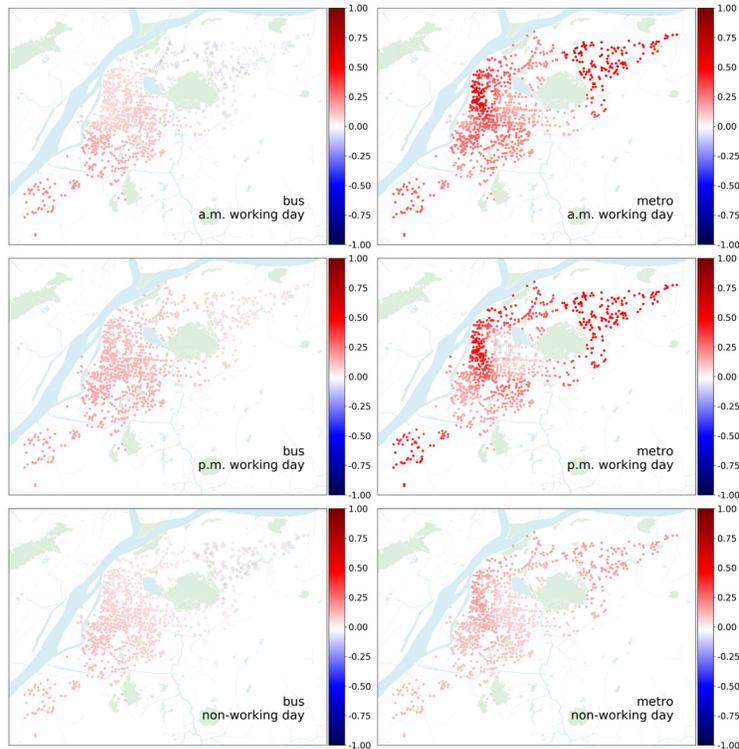
Figure 5. Estimated coefficients of the public transport density

Similarly, the public transport density shows a positive impact on bike usage regardless of the location of docking stations, as shown in Figure 5. This finding is discrepant with the results concluded by Zhang et al. (2017) but is consistent with El-Assi et al. (2017) and Faghih-Imani and Eluru (2016), which might depend on the development of the local public transport system. The figures demonstrate that the metro is a powerful booster for PBS ridership. In all the examined time periods, a stationary positive relationship can be observed in the whole city. It proves that PBS has met the expectations as a feeder to the metro system in response to the last-mile problem. The relationship with city buses is similarly homogeneous, being always positive except for Jianye District during the evening peak. However, the association of bus stations and PBS usage is not comparable with that of metro stations. This can possibly be attributed to the relatively mature planning of the bus system in Nanjing, the station density of which is adequate to cover the daily demand. Moreover, the bus stops are designed to be placed near metro stations and parking lots. In other words, the city bus can compete with PBS to some extent as a connector to fill the gap of the last mile. Therefore, if the competence of PBS is weaker than buses, like the Qixia District, where PBS station density is relatively low, it is likely that residents select buses as the traveling mode.
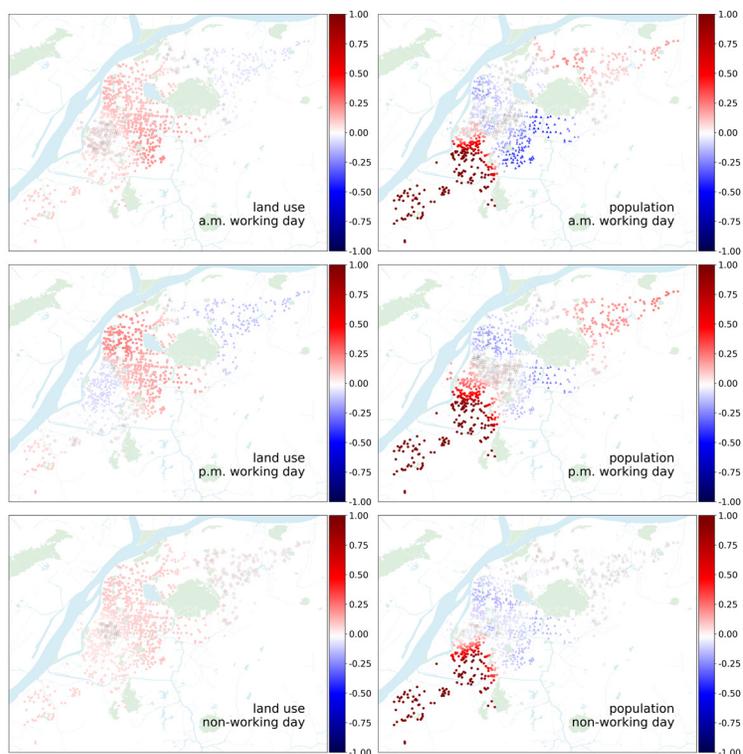
Figure 6. Estimated coefficients of land-use diversity and the population density

Land-use diversity, which is frequently discussed in research pertaining to build environment (Ding et al., 2016), was only examined in one PBS-related study (Zhang et al., 2017). Similar to the school density and the bus stop density, a complex composition of land-use types can slightly induce more PBS ridership, especially in Gulou District, Xuanwu District and Qinhuai District. However, results indicate that the population density in these three districts does not lead to larger cycling demand. For other districts, where the densities of various POIs are much lower, the population density is the major contributor to bike usage. This can be attributed to the large size of residential communities in these newly built-up areas. A docking station here may serve much more residents than another station in CBD even though their values of the residential density are the same.
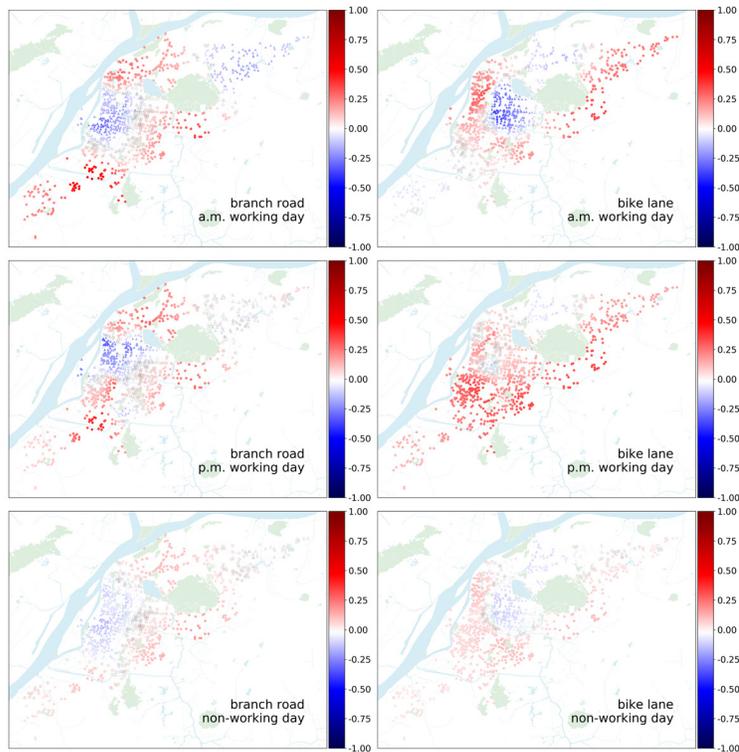
Figure 7. Estimated coefficients of the road network density

Bike lanes have been widely acknowledged to be a contributor to generate more ridership, as an unpleasant cycling environment naturally discourages traveler from riding bicycles. The results are statistically significant in most locations, as demonstrated in Figure 7. The morning peak on working days shares a similar pattern of relationship with non-working days, with the city being split into two parts of opposite relationships. An interesting finding is that in the Jianye District, bike lanes contribute more to PBS ridership, while in other locations, the branch roads show a positive connection with cycling. It should be noted that the Jianye District is a newly developed district of Nanjing, and the traditional downtown area lies in the central area. Considering that the citizens who live in the traditional downtown are usually familiar with those branches, they prefer to take shortcuts which could save their traveling time. Nevertheless, the bike lanes become the predominant contributor in the evening peak, possibly due to the relatively dim lighting conditions in the evening.
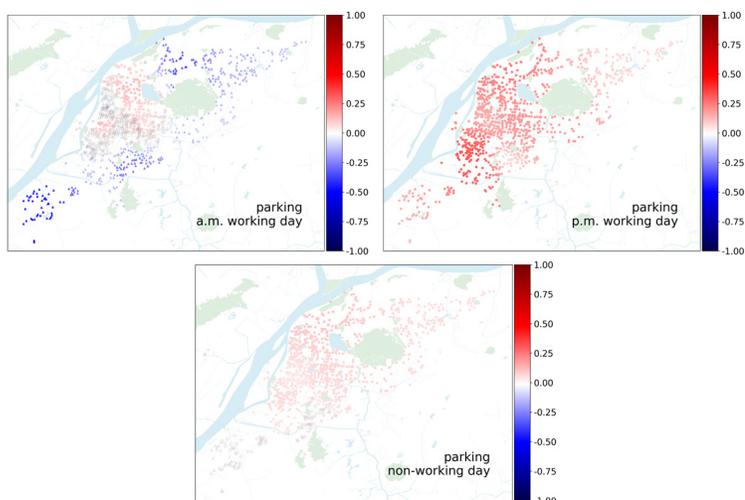
Figure 8. Estimated coefficients of the parking space density

Being neglected by most studies, parking spaces display a positive association with cycling during most of the time in central urban areas, according to Figure 8. For docking stations surrounded by adequate parking spaces in suburbs, they are more likely to fail in the competence with cars during morning peaks since most of the residents commute for a relatively long distance. In contrast, during evening peaks and on non-working days, the parking spaces consistently show a positive relationship with public bike ridership, which leads to an intuitive conjecture of Park-and-Bike (P&B). That is, travelers park their cars where there are parking spaces and then switch to PBS to reach their destinations. However, note that it remains unknown whether the driving force is truly parking spaces or some other factors that highly associate with them. It is therefore recommended that further research be conducted to investigate the existence and viability of P&B.
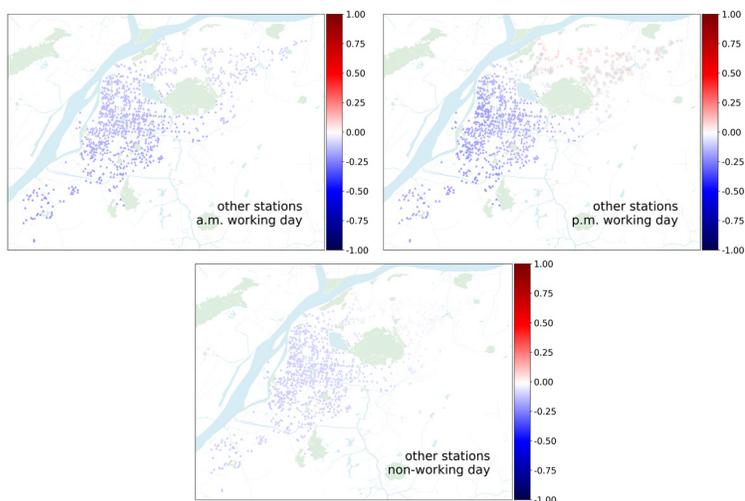
Figure 9. Estimated coefficients of the docking station density

Strong evidence reveals the existence of competition among neighboring docking stations according to Appendix A and Figure 9. Unlike PBS in western cities like BIXI (Montreal) and Citi Bike (New York), many docking stations are often installed side by side in NPBS. For example, four docking stations are arranged at the four corners of the intersection of Shanghai Road and Hanzhong Road. Intuitively, people tend to pick the closest station when faced with multiple choices. Each station will attract travelers from the others nearby, leading to lower average ridership, though their sum can be high.

6　　　Conclusion

This paper addressed the impacts of spatial heterogeneity on built environment and PBS ridership. To introduce a spatial perspective on the effects of the built environment, the MGWR was adopted and evaluated on PBS data. The results show that the MGWR outperforms the linear regression and standard GWR, which has verified our assumption that inherent non-stationarity exists in the estimation of parameters across space. The modeling results demonstrate different influencing patterns between the traditional downtown area and newly built-up areas. Disparate effects of residential communities, employment, and school, is recognized in the morning peak and evening peak, indicating the commuting function of PBS. Population density does not show a strong influence on PBS usage in central areas, but it is the major motivator of public bike usage in newly developed districts. There is also competition between adjacent PBS stations. The significant positive impact from the metro as well as the bus can also be observed across the whole city. Additionally, in traditional downtown areas, bike lanes do not have a comparable positive effect as outer areas. Instead, more trips were made where there are dense branch roads. Moreover, an unexpected high association between PBS usage and the parking lots is observed. These findings are valuable in the promotion and planning of PBS. Planning of PBS stations should consider the integration with public transport. Although bike lanes make little contribution to cycling in traditional downtown areas, renovation of chaotic aged branch roads is required to improve the cycling experience of travelers.

In conclusion, spatial effects should not be neglected when dealing with built environment, and the MGWR is robust in unveiling both the heterogeneous and homogeneous spatial relationship across the city. It should be noted that there are huge discrepancies among cities with respect to built environment and the behavior patterns of the citizen. For future work, understanding the shared functioning mechanism of built environment on PBS usage between cities would be meaningful. Also, as suggested by the results, further investigation into the feasibility of Park-and-Bike in China should be conducted.

Acknowledgements

Appendix

Appendix availabe at https://www.jtlu.org/index.php/jtlu/article/view/1568

References

Amap. (2016). Amap JavaScript API (Version 1.3). Amap. Retrieved from https://lbs.amap.com/api/javascript-api/summary

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis, 27*(2), 93–115. doi.org/10.1111/j.1538-4632.1995.tb00338.x

Bachand-Marleau, J., Lee, B. H. Y., & El-Geneidy, A. M. (2012). Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. *Transportation Research Record: Journal of the Transportation Research Board, 2314*(1), 66–71. doi.org/10.3141/2314-09

Bhat, C., & Zhao, H. (2002). The spatial analysis of activity stop generation. *Transportation Research Part B: Methodological, 36*(6), 557–575. doi.org/10.1016/S0191-2615(01)00019-4

Boarnet, M. G., & Crane, R. (2001). *Travel by design: The influence of urban form on travel.* Oxford, England: Oxford University Press.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis, 28*(4), 281–298. doi.org/10.1111/j.1538-4632.1996.tb00936.x

Buck, D., & Buehler, R. (2011, November 15). *Bike lanes and other determinants of capital bikeshare trips.* Paper presented at the Transportation Research Board 91st Annual Meeting, Washington, DC.

Cervero, R. (2002). Induced travel demand: Research design, empirical evidence, and normative policies. *Journal of Planning Literature, 17*(1), 3–20. doi.org/10.1177/088122017001001

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment, 2*(3), 199–219. doi.org/10.1016/S1361-9209(97)00009-6

de Souza, F., La Paix Puello, L., Brussel, M., Orrico, R., & van Maarseveen, M. (2017). Modelling the potential for cycling in access trips to bus, train and metro in Rio de Janeiro. *Transportation Research Part D: Transport and Environment, 56,* 55–67. doi.org/10.1016/j.trd.2017.07.007

DeMaio, P., & Meddin, R. (2019). The bike-sharing world map. Retrieved from www.bikesharingmap.com

Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice, 100,* 65–80. doi.org/10.1016/j.tra.2017.04.008

Ding, C., Wang, Y., Yang, J., Liu, C., & Lin, Y. (2016). Spatial heterogeneous impact of built environment on household auto ownership levels: Evidence from analysis at traffic analysis zone scales. *Transportation Letters, 8*(1), 26–34. doi.org/10.1179/1942787515Y.0000000004

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. *Transportation, 44*(3), 589–613. doi.org/10.1007/s11116-015-9669-z

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association, 76*(3), 265–294. doi.org/10.1080/01944361003766766

Faghih-Imani, A., & Eluru, N. (2016). Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography, 54,* 218–227. doi.org/10.1016/j.jtrangeo.2016.06.008

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography, 41,* 306–314. doi.org/10.1016/j.jtrangeo.2014.01.013

Faghih-Imani, A., Hampshire, R., Marla, L., & Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice, 97*, 177–191. doi.org/10.1016/j.tra.2016.12.007

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers, 107*(6), 1247–1265. doi.org/10.1080/24694452.2017.1352480

Frade, I., & Ribeiro, A. (2014). Bicycle sharing systems demand. *Procedia–Social and Behavioral Sciences, 111,* 518–527. doi.org/10.1016/j.sbspro.2014.01.085

Gebhart, K., & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation, 41*(6), 1205–1225. doi.org/10.1007/s11116-014-9540-7

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity. *American Journal of Preventive Medicine, 23*(2), 64–73. doi.org/10.1016/S0749-3797(02)00475-0

Huang, K., Liu, Z., Zhu, T., Kim, Y., & An, K., (2019). Analysis of the acceptance of park-and-ride by users: A cumulative logistic regression approach. *Journal of Transport and Land Use, 12*(1) 637–647. doi.org/10.5198/jtlu.2019.1390

Kager, R., Bertolini, L., & Te Brömmelstroet, M. (2016). Characterization of and reflections on the synergy of bicycles and public transport. *Transportation Research Part A: Policy and Practice, 85*, 208–219. doi.org/10.1016/j.tra.2016.01.015

Kutela, B., & Teng, H. (2019). The influence of campus characteristics, temporal factors, and weather events on campuses-related daily bike-share trips. *Journal of Transport Geography, 78*, 160–169. doi.org/10.1016/j.jtrangeo.2019.06.002

Lin, J.-J., Zhao, P., Takada, K., Li, S., Yai, T., & Chen, C.-H. (2018). Built environment and public bike usage for metro access: A comparison of neighborhoods in Beijing, Taipei, and Tokyo. *Transportation Research Part D: Transport and Environment, 63,* 209–221. doi.org/10.1016/j.trd.2018.05.007

Liu, Z., Chen, X., Meng, Q., & Kim, I. (2018). Remote park-and-ride network equilibrium model and its applications. *Transportation Research Part B: Methodological, 117*, 37–62. doi.org/10.1016/j.trb.2018.08.004

Liu, Y., Jia, R., Xie, X., & Liu, Z. (2019). A two-stage destination prediction framework of shared bicycle based on geographical position recommendation. *IEEE Intelligent Transportation Systems Magazine, 11*(1), 42–47. doi.org/10.1109/MITS.2018.2884517

Liu, Z., Wang, S., Zhou, B., & Cheng, Q., (2017). Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics. *Transportation Research Part C, 79*, 58–72.

Mateo-Babiano, I., Bean, R., Corcoran, J., & Pojani, D. (2016). How does our natural and built environment affect the use of bicycle sharing? *Transportation Research Part A: Policy and Practice, 94,* 295–307. doi.org/10.1016/j.tra.2016.09.015

Nanjing Municipal Bureau of Statistics. (2017). *Statistical yearbook of Nanjing.*

O'Brien, O., Cheshire, J., & Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography, 34*, 262–273. doi.org/10.1016/j.jtrangeo.2013.06.007

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2018). *MGWR: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale.* OSF Preprints. doi.org/10.31219/osf.io/bphw9

Qian, X., & Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. *Applied Geography, 59*, 31–42. doi.org/10.1016/j.apgeog.2015.02.011

Raux, C., Zoubir, A., & Geyik, M. (2017). Who are bike sharing schemes members and do they travel differently? The case of Lyon's "Velo'v" scheme. *Transportation Research Part A: Policy and Practice, 106,* 350–363. doi.org/10.1016/j.tra.2017.10.010

Schoner, J., & Levinson, D. M. (2013). *Which station? Access trips and bike share route choice* (Working Papers No. 000117; Nexus Working Papers, pp. 15). Minneapolis: Center for Transportation Studies.

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board, 2143,* 159–167. doi.org/10.3141/2143-20

Shaheen, S., & Rodier, C. J. (2018). *EasyConnect: Low-speed modes linked to transit planning project* (UCD-ITS-RR-08-33). Berkeley, CA: Institute of Transportation Studies, University of California.

Tran, T. D., Ovtracht, N., & d'Arcier, B. F. (2015). Modeling bike sharing system using built environment factors. *Procedia CIRP, 30,* 293–298. doi.org/10.1016/j.procir.2015.02.156

Wang, Y., Liu, Y., Ji, S., Hou, L., Han, S. S., & Yang, L. (2018). Bicycle lane condition and distance: Case study of public bicycle system in Xi'an, China. *Journal of Urban Planning and Development, 144*(2), 05018001. doi.org/10.1061/(ASCE)UP.1943-5444.0000436

Yang, X.-H., Cheng, Z., Chen, G., Wang, L., Ruan, Z.-Y., & Zheng, Y.-J. (2018). The impact of a public bicycle-sharing system on urban public transport networks. *Transportation Research Part A: Policy and Practice, 107,* 246–256. doi.org/10.1016/j.tra.2017.10.017

Zhang, Y., Brussel, M. J. G., Thomas, T., & van Maarseveen, M. F. A. M. (2018). Mining bike-sharing travel behavior data: An investigation into trip chains and transition activities. *Computers, Environment and Urban Systems, 69*, 39–50. doi.org/10.1016/j.compenvurbsys.2017.12.004

Zhang, Y., Thomas, T., Brussel, M., & van Maarseveen, M. (2017). Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. *Journal of Transport Geography, 58*, 59–70. doi.org/10.1016/j.jtrangeo.2016.11.014

Zhao, J., Deng, W., & Song, Y. (2014). Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transport Policy, 35*, 253–264. doi.org/10.1016/j.tranpol.2014.06.008