Article:

Wang, Z. orcid.org/0000-0002-6402-6280, Song, J., Zhang, Y. et al. (3 more authors) (2022) Spatial heterogeneity analysis for influencing factors of outbound ridership of subway stations considering the optimal scale range of “7D” built environments. Sustainability, 14 (23). 16314. ISSN 2071-1050

https://doi.org/10.3390/su142316314

sustainability

Article

Spatial Heterogeneity Analysis for Influencing Factors of Outbound Ridership of Subway Stations Considering the Optimal Scale Range of "7D" Built Environments

Zhenbao Wang [1,*], Jiarui Song [1], Yuchen Zhang [2], Shihao Li [1], Jianlin Jia [3] and Chengcheng Song [4]

[1] School of Architecture and Art, Hebei University of Engineering, Handan 056038, China
[2] Department of Urban Studies and Planning, The University of Sheffield, Sheffield S10 2TN, UK
[3] School of Civil Engineering, Inner Mongolia University of Technology, Hohhot 010051, China
[4] Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China
* Correspondence: wangzhenbao@hebeu.edu.cn

Abstract: The accuracy of the regression model of ridership of subway stations depends on the scale range of the built environment around the subway stations. Previous studies have not considered the Modifiable Area Unit Problem (MAUP) to establish the regression model of subway station ridership. Taking Beijing as an example, this paper expanded the built environment variables from "5D" category to "7D" category, added indicators such as parking fee standard and population density factor, and proposed a Multi-Scale Geographical Weighted Regression (MGWR) model of outbound ridership of subway stations with standardized variables. The goodness of fit of regression models under 10 spatial scales or built environment around subway stations are compared, and the spatial heterogeneity of built environment factors under the optimal spatial scale of outbound ridership of subway stations during the morning peak on weekdays is discussed. The results show that: (1) the scale range overlapped by 1000 m radius circular buffer zone and Thiessen polygon has the highest explanatory power for the regression model, and is regarded as the optimal scale range of built environment; (2) the density of office facilities, sports and leisure facilities, medical service facilities, building density and floor area ratio (FAR) has a significant impact on the outbound ridership of all subway stations; (3) office facilities, catering facilities, FAR, number of parking lots, and whether subway stations are transfer stations have a positive impact on outbound ridership. The number of medical service facilities, sports and leisure facilities, bus stops and building density have a negative impact on outbound ridership; (4) the two added factors in this study: parking charge standard and population density, as the influencing factors of the built environment, have a significant impact on the outbound ridership of some subway stations; and (5) the different local coefficients of the built environment factors at different stations are discussed, which indicate the spatial heterogeneity on the outbound ridership. The results can provide an important theoretical basis for the prediction and analysis of demand of ridership at subway stations and the integration of the built environment around the stations.

Keywords: urban traffic; built environment; multi-scale geographically weighted regression (MGWR); subway station; optimal scale range

Citation: Wang, Z.; Song, J.; Zhang, Y.; Li, S.; Jia, J.; Song, C. Spatial Heterogeneity Analysis for Influencing Factors of Outbound Ridership of Subway Stations Considering the Optimal Scale Range of "7D" Built Environments. *Sustainability* **2022**, *14*, 16314. https://doi.org/10.3390/su142316314

Academic Editors: Bin Meng, Dongsheng Zhan and Ran Liu

Received: 19 October 2022
Accepted: 5 December 2022
Published: 6 December 2022

1. Introduction

For a long time, the relationship between public transport and land use has always been a research topic of interest to scholars and policy makers in the field of transport and urban planning [1,2]. Subway has the characteristics of large traffic volume, low pollution, high speed, low energy consumption, and plays an important role in the urban public transport system. Land use is an important factor affecting subway ridership [3]. Clarifying the built environmental factors that affect the ridership of subway stations and their impact degree is conducive to putting forward targeted urban renewal planning strategies and policy suggestions for stations with overloaded ridership, and stations with insufficient ridership.

In recent years, scholars have conducted a lot of research on the factors that affect rail transit ridership. Due to different research purposes and data acquisition methods, the methods of representing public transport ridership data in the research are also different. When accurate pick-up and drop-off ridership data cannot be obtained, some studies have used the monthly average accommodation rate [4] and the average weekday accommodation rate [3,5] to represent it, but more scholars prefer to choose the one-day relationship [6] and the average daily relationship [7,8]. The models used in the existing literature include the direct demand model [9,10] and the regression model, among which the regression model is mostly used. Regression models include global regression model and local regression model. Global models include the structural equation model (SEM) [11], distance decay regression model [4] and ordinary least squares method (OLS) [3,4,12–14]. However, the default parameters in the global model are stable in the global range, so the calculated coefficients have no significant difference in space [15]. In fact, in different spaces, the estimated coefficients of the influencing factors of traffic ridership may be inconsistent. Geographically Weighted Regression (GWR) [16,17] was used to study the relationship between built environment and subway station ridership [15–17]. GWR considers spatial instability, which can reveal the spatial heterogeneity of spatial parameters. In addition, GWR has been proved to have better fitting results than the global OLS model [15]. Some recent studies have proposed a hybrid GWR model [18], which allows some variables to be local variables and others to be global variables, in order to effectively separate independent variables that produce global and local effects. However, it cannot well reflect the spatial heterogeneity. Multi-scale Geographically Weighted Regression (MGWR) [19,20] has made some improvements on this, which not only improve the goodness of fit results and model interpretation ability, but also can use the optimal bandwidth to measure the scale difference in spatial heterogeneity of influencing factors [6,21]. The MGWR model is currently applied to the relationship between built environment and ridership [6], influencing factors of second-hand house price [21] and environmental science [22,23]. To the best of our knowledge, there is a lack of research on the mechanism of the influence of built environment on ridership of subway stations by using MGWR model.

Due to the availability of data, existing studies have drawn different conclusions on factors affecting ridership of subway stations. In the selection of the influencing factors of built environment, some scholars have selected from the aspects of land use, social and economic factors and traffic environment [18,24–27]. In analyzing the factors that influence transportation, Cervero and Kockelman summarized the “3D” dimensions of the built environment, namely density, diversity and design. When studying the built environment and walking and riding behavior under its influence in Bogota, Colombia, Cervero and others added “distance to transit” and “destination accessibility” on the basis of the “3D” principle, which formed the “5D” planning principle [28–30]. Since then, many scholars have studied the built environment that affects TOD effectiveness with the “5D” dimension index as a reference [6,31–33]. With the development of the “5D” theory, Chris De Gruyter et al. added “demand management” and “demographics”, and finally formed the “7D” theory [34]. However, the spatial heterogeneity of “demand management” and “demographics” as two types of built environment influencing factors on the ridership at subway stations needs to be further analyzed.

The influence range of the built environment around the subway station is also an important issue affecting the interaction mechanism between built environment and ridership of subway stations, which is usually called “pedestrian catchment area” (PCA). The selection of PCA mainly adopts the circular buffer with the subway station as the center [9,13,17,18], but considering that in the area with dense distribution of subway stations, PCA will overlap. Therefore, some scholars use the method of combining Thiessen polygons with circular buffer zones [6,10,35]. The research radius selected by different scholars varies greatly, including 500 m [25,36], 600 m [18], 800 m [6,13,15,36] and 1500 m [37], among which 800 m is the scale chosen by most people. The radius of circular buffer is mostly selected based on previous research experience [6,17] or travel distance [18], and did not adopt quantitative research methods. Guerra et al. summarized the use of a 0.25 mile catch area around transit for jobs and a 0.5 mile catch area for population, by using the direct demand model to study 1457 public transport stations and their surrounding environments in the United States [9] Although some scholars have used quantitative methods to explore the choice of PCA, the sample size involved is large and difficult to operate. Modifiable area unit problem (MAUP) refers to the change in spatial analysis results following the defined basic area unit (grid cell or particle size). Some studies show that MAUP is an essential basic problem [38–40] in travel behavior analysis. Most of the existing studies [6,13,17] use the built environment of a single subway station as the research range, or use the single range delineation method to conduct quantitative statistical analysis of the impact of built environment, without considering the MAUP problem.

The related studies are summarized in this paper, as shown in Table 1. These studies investigated the impact of the built environment on the ridership of public transport from the aspects of analysis methods, explanatory variables used in the model, the division of the impact range of the built environment and the case study area. To sum up, the literature review shows that the methods and built environment factors used by Chinese scholars in the research are similar to those in other countries, but the methods adopted in the delineation of pedestrian catchment area (PCA) for subway stations are slightly different. Some scholars [6,10,17,35] use Thiessen polygon and circular buffer overlay method in their research, while other scholars use zones of neighborhoods [16,26] or buffer method [9,13,15,18,25]. In terms of research on the ridership and built environment, more of them choose a single scale range and less of them analyze the impact of built environment factors on the subway ridership considering the MAUP's scale and zoning effects. Compared with GWR and OLS models, MGWR has better goodness of fit and can better reflect the spatial and scale differences in the impact of the built environment on ridership. Therefore, MGWR has certain advantages compared with other models. However, the spatial heterogeneity of the variables of the built environment reflects that the impact of the explanatory variables is different according to different locations of the city, so the impact degree of the explanatory variables of the same built environment may not be comparable among different cities. Therefore, when applying the MGWR model, we should carry out targeted analysis on specific cities.

Table 1. Summary of literature review.

Author	**Main Independent Variables**	**Analysis Method**	**PCA**	**Radius of PCA**	**Study Area**	**Main Conclusion**
Guerra et al. [9]	Occupation and population	Direct demand models (DRMs)	Circle buffer	0.25 mile and 0.5 mile	21 cities of United States	The optimal circle buffer is 0.25-mile catchment area for jobs, 0.5-mile catchment area for population
Cardozo et al. [15]	Occupation, workers, land mixed use	OLS, GWR	Circle buffer	200 m and 800 m	Madrid, Spain	GWR model has better fitting than the OLS model
Jun et al. [18]	Population density, employment density, proportion of apartment units	Stepwise regression models, Mixed geographically weighted regression	Circle buffer	300 m, 600 m and 900 m	Seoul, Korea	Significant influencing factors: residential, commercial development patterns, mixed land use within 600 m circle buffer, and spatial heterogeneity of influencing factors is discussed

Table 1. *Cont.*

Author	Main Independent Variables	Analysis Method	PCA	Radius of PCA	Study Area	Main Conclusion
Calvo et al. [16]	Socioeconomic, land use, accessibility and transportation system variables	GWR	Neighborhoods		Madrid, Spain	The significant influencing factors are different in different regions
AlKhereibi et al. [26]	Land use density	Regression model	Neighborhoods		Qatar	Significant influencing factors: land use type
Zhao et al. [13]	Land use, external connectivity and intermodal connection	OLS	Circle buffer	800 m	Nanjing, China	Significant influencing factors: CBD dummy variable, the number of education buildings, entertainment venues and shop centers, bicycle Park and Ride spaces
Chen et al. [25]	Land use, intermodal connection and station characteristics	Minkowski distance-GWR	Circle buffer	500 m	Nanjing, China	Minkowski distance-GWR achieves better goodness-of-fit than the global OLS and GWR
Li et al. [10]	Population, employment, density, land use	DRMs	Overlay of circular buffer and Thiessen polygon	800 m	Guangzhou, China	Significant influencing factors: Residential buildings, business buildings, workplaces, shopping, sport, leisure services
Li et al. [35]	Population, employment, land use density, diversity and station characteristics	GWR	Overlay of circular buffer and Thiessen polygon	800 m	Guangzhou, China	Influence of built environment on ridership flow has spatial variation
Cong et al. [17]	Land use, network topology and intermodal connection	GWR	Overlay of circular buffer and Thiessen polygon	690 m	Xi'an, China	The influence of land use types on passenger flow is heterogeneous in time and space
Gao et al. [6]	"5D" built environment characteristics	MGWR	Overlay of circular buffer and Thiessen polygon	800 m	Beijing, China	Built environment features has significant spatial heterogeneity, MGWR achieves better goodness-of-fit than the global OLS and GWR
This research	"7D" built environment characteristics	MGWR	Circular buffer, and overlay range of circular buffer and Thiessen polygon	500 m, 800 m, 1000 m, 1200 m and 1500 m, respectively	Beijing, China	The area overlapped by 1000 m radius circular buffer zone and Thiessen polygon has the highest explanatory power for the MGWR model. the spatial heterogeneity of built environment factors is discussed

The objective of this research is to find a better scale range of the built environment around subway stations and improve the goodness of fit of the MGWR model between the subway stations ridership and the built environment factors. We construct multiple MGWR models under different catchment areas of the built environment around subway stations based on the "7D" theory, considering the parking charge standard and population density among the factors that affect the built environment. By comparing the goodness of fit results, the optimal scale range of the built environment is selected to evaluate the built environment factors that have significant impacts on ridership of the subway station, and the spatial heterogeneity and scale difference in the built environment factors are also explored. Discussion and suggestions for planners and policy makers are presented.

2. Methods

2.1. Environmental Influencing Factors of Subway Stations

Based on the "7D" theory of the built environment, and according to the principles of data availability and experimental operability, the built environment quantitative indicator system (as shown in Table 2) is established. This indicator system is divided into eight categories: density, diversity, design, destination accessibility, travel distance, demand management, demography and other indicators, with a total of 19 influencing factors.

Table 2. Influencing factors of the built environment and the calculation methods.

Built Environment Category	Interfering Factor	Data Sources	Computing Method	Unit
Density	Density of catering facilities Density of scenic facilities Density of commercial facilities Density of science, education and cultural facilities Density of residential facilities Density of office facilities Density of sports and leisure facilities Density of medical service facilities	Amap API(Application Program Interface) (https://lbs.amap.com/ (accessed on 10 October 2021))	$D_i = N_i/S_i$ In the formula, D_i is the density of certain type of POI facilities at subway station i; N_i is the number of POI facility points of a certain type within the analysis range of subway station i; S_i is the total area of the analysis range.	quantity/km^2
	Building density	Open street map (https://lbs.amap.com/ (accessed on 10 October 2021))	$B_i = A_i/S_i$ In the formula, B_i is the building density within the analysis range of subway station i; A_i is the total area of building base within the analysis range; S_i is the total area of the analysis range.	m^2/km^2
Diversity	Mixed utilization of land	Amap API (https://lbs.amap.com/ (accessed on 10 October 2021))	$H_i = -\sum_{i=1}^{n} (K_i) \cdot \ln(K_i)$ In the formula, K_i is the ratio of the number of POI facility points of a certain type within the analysis range of subway station i to the number of all POI facility points; H_i is the degree of mixed use of land at subway station i. The higher the H_i index, the higher the degree of mixed use of land around the subway station.	

Table 2. *Cont.*

Built Environment Category	Interfering Factor	Data Sources	Computing Method	Unit
Design	Road density	Open street map (https://www.openstreetmap.org/ (accessed on 10 October 2021))	$L_i = R_i \times 1000/S_i$ In the formula, L_i is the road density within the analysis range of subway station i; R_i is the sum of road lengths within the analysis range; S_i is the total area of the analysis range.	km/km^2
	Floor area ratio		$E_i = O_i/S_i$ In the formula, E_i is the floor area ratio of the analysis range of subway station i; O_i is the total building area within the analysis range; S_i is the total area of the analysis range.	
	Density of subway lines		$M_i = C_i \times 1000/S_i$ In the formula, M_i is the density of rail lines within the analysis range of subway station i; C_i is the sum of subway line lengths within the analysis range; S_i is the total area of the analysis range.	km/km^2
Destination accessibility	Number of parking lots	Amap API (https://lbs.amap.com/ (accessed on 10 October 2021))	Number of POI facilities	quantity
	Number of bus stops			
Distance to transit	Distance to the nearest bus stop	Amap API (https://lbs.amap.com/ (accessed on 10 October 2021))	Shortest path analysis	m
Demand management	Parking charge standard in the area where the subway station is located	Amap(https://www.amap.com/ (accessed on 10 October 2021)) and Baidu(https://map.baidu.com/ (accessed on 10 October 2021)) Map Parking service	Take the average of charging standards of multiple parking lots	Chinese yuan (CNY)/hour
Demographics	Population density	Data from WorldPop [41,42]	$P_i = G_i/S_i$ In the formula, P_i is the population density within the analysis range of subway station i; G_i is the population within the analysis range; S_i is the total area of the analysis range.	persons/km^2
Others	Attribute s of subway stations	According to the map of Beijing subway	To determine whether it is a transfer station, yes is 1, no is 0	

2.2. Delineation of the Built Environment Range around Rail Transit Stations

Some studies use a fixed circular area within 500 m [25,36], 800 m [6,13,15,37] or 1500 m [37] radius centered on a subway station as PCA for subway ridership analysis. Due to the differences in the reality of different cities, it cannot be confirmed whether this scale is applicable to all cities. The circular buffer zone will result in the overlap of the research range of different subway stations, especially the subway stations located in the city center. The distance between them is relatively close, which will easily lead to the double calculation of the built environment. At the same time, despite being within the service radius of multiple subway stations, travelers will still choose the nearest subway station. Therefore, to avoid this phenomenon, the concept of Thiessen polygon is introduced, and the intersection of Thiessen polygon and circular buffer zone is used to solve the problem of overlapping research range [6,10,35]. By default, travelers in different regions are assigned to the designated subway station, that is, travelers will only take the subway at that subway station. However, this method also has some limitations. When the distance between travelers and two stations on different subway lines is close, travelers may choose a farther station for convenience. Therefore, in combination with the existing research, two zoning methods were selected to delimit the spatial range of the built environment: one method takes each subway station as the center, and selects 500 m, 800 m, 1000 m, 1200 m and 1500 m, respectively, as the radius to delimit the circular buffer zone; another method is called "Thiessen polygon overlay range" in which a buffer and a Thiessen polygon are created through ArcGIS intersecting analysis for each station (Figure 1).

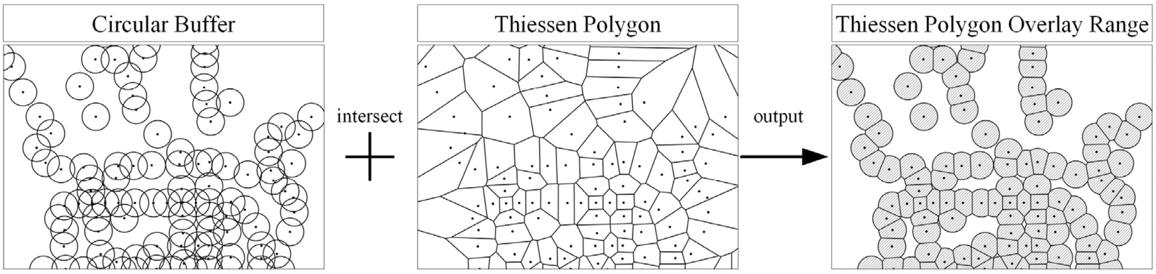


Figure 1. Schematic diagram of zoning method for built environment analysis range.

2.3. Spatial Autocorrelation Test of Explanatory Variables

The Moran's I index test was carried out on the index values (explanatory variables) of various influencing factors of the built environment around the subway station and the morning peak outbound ridership (dependent variables) of the subway station on weekdays, for which the regression model was built. The results show that the p values of all explanatory variables were less than 0.05, and the Moran's I index was between -1 and 1, indicating that the spatial autocorrelation of explanatory variables was significant, which was suitable for establishing the MGWR model.

2.4. Multiple Collinearity Test of Explanatory Variables

Variance impact factor (VIF) can reflect the severity of multicollinearity [43]. The calculation formula is:

$$\mathrm{VIF} = \frac{1}{1 - R^2} \tag{1}$$

where R^2 in the formula is the coefficient of determination in the regression equation. When VIF > 10, it is considered that there is multicollinearity, and this variable should be removed [36].

2.5. Model Construction

We took ridership of subway stations as the dependent variable, and the 19 influencing factors of the built environment as the explanatory variables; we then eliminated the explanatory variables that are unreasonable in the spatial autocorrelation and collinearity tests. Ten variable datasets were built, containing independent variables of built environment and the dependent variable. For five of them, their analysis ranges of the built environment are five circular buffers with the radius of 500 m, 800 m, 1000 m, 1200 m and 1500 m, respectively. For the other five, their analysis ranges of built environment are the overlay ranges of Thiessen polygon and the before-mentioned five circular buffers, respectively. Due to different dimensions of independent variables, in order to reflect the relative degree of influence on the dependent variable, the dependent variable and independent variable were standardized to obtain their own standardized values, with the mean value of 0 and the standard deviation of 1. The MGWR model with variables standardized is shown in Equation (2). The interpretation for the local regression coefficient β_{bwk} is that if the standardized independent variable increases by one standard deviation, on average, dependent variable increases by β_{bwk} standard deviation units [44].

$$y_i^* = \beta_{bw0}(u_i, v_i) + \sum_{k=1}^{n} \beta_{bwk}(u_i, v_i) \cdot x_{ik}^* + \varepsilon_i \tag{2}$$

$$y_i^* = \frac{y_i - \overline{y}}{s_y} \tag{3}$$

$$x_{ik}^* = \frac{x_{ik} - \overline{x_k}}{s_{xk}} \tag{4}$$

where y_i^* is the standardized value of the morning peak outbound ridership of subway station *i*; (u_i, v_i) is the longitude and latitude coordinates of subway station *i*; β_{bw0} is the regression model constant of subway station *i* with coordinate (u_i, v_i); β_{bwk} is the local regression coefficient of the *k-th* built environmental influencing factor of subway station *i*; *bwk* is the optimal bandwidth of the *k-th* built environmental influencing factor; x_{ik}^* is the standardized value of the *k-th* built environmental influencing factor indices of subway station *i*; ε_i is the random error of subway station *i*; *n* is the number of factors affecting the built environment; $\overline{y}$ is the average value of outbound ridership at the morning peak of subway station; s_y is the standard deviation of the morning peak outbound ridership of the subway station; x_{ik} is the indices of the k-th built environmental influencing factor of subway station i; $\overline{x_k}$ is the average value of the *k-th* influencing factor indices; s_{xk} is the standard deviation value of the *k-th* influencing factor indices.

According to the analysis range of different built environments and the model results obtained by the zoning method, we regard the optimal analysis range of the model as the optimal built environment scale range applicable to cities.

3. Data, Results and Discussion

3.1. Research Objects

Beijing is a mega-city with a population of 25 million, and it was also the first city in China to open a subway operation. Since the first line was put into operation in 1969, 19 subway lines had been built and operated by 2017. We selected the subway stations that have been put into use in these 19 lines as the research object, totaling 287 stations.

3.2. Passenger Flow Data

The existing research on the relationship between the built environment and ridership of subway stations has only captured the one-day ridership data for analysis, which has a certain degree of occasionality, and does not represent the real demand for the ridership on weekdays. Therefore, this paper obtained the hourly outbound ridership data of Beijing's subway operation lines within five working days from April 17 to 21, 2017. According to the trend analysis of the total outbound ridership (Figure 2), it was determined that the morning peak on weekdays is from 7:00 to 9:00 a.m. In order to reduce the impact of daily ridership fluctuation, we take the average value of a five-day passenger flow as the dependent variable. Figure 3 shows the spatial distribution of outbound ridership at 287 stations during the morning peak.

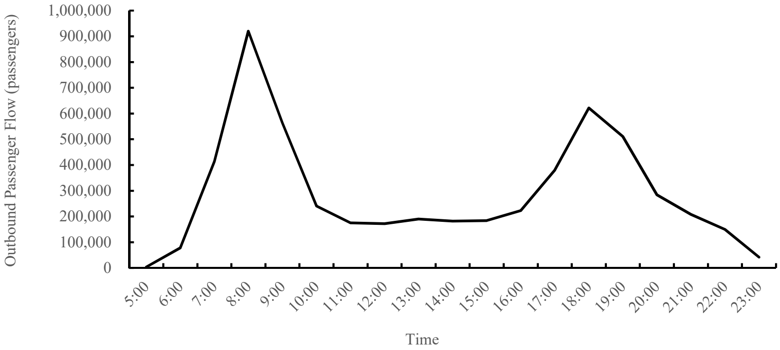


Figure 2. Outbound ridership of Beijing subway stations on weekdays.

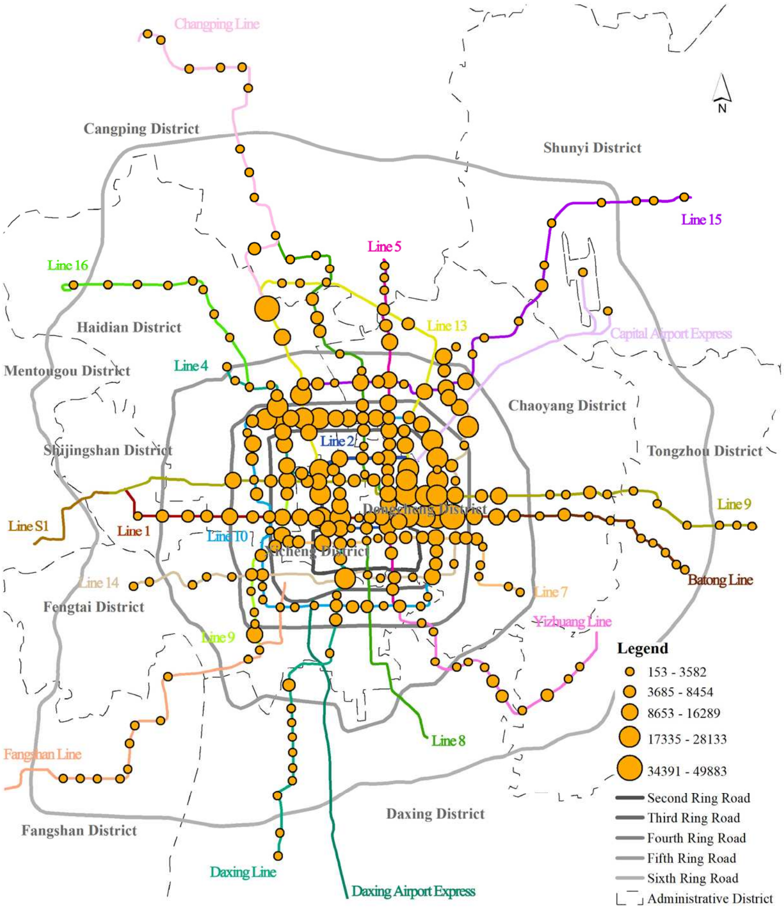


Figure 3. Spatial distribution of outbound ridership during the morning peak on weekdays in Beijing.

3.3. Model Goodness of Fit and Optimal Partition Scale Analysis

The collinearity test results show that the VIF values of the density of catering facilities and residential facilities are greater than 10 in the analysis range of a circular buffer zone with a radius of 1200 m; in the analysis range of a circular buffer zone with a radius of 1500 m, the number of parking lots, density of catering facilities, density of residential facilities, density of sports and leisure facilities, density of medical service facilities and VIF of floor area ratio (FAR) are all greater than 10, which indicates that these variables have a strong collinearity problem with one or more other variables. Therefore, these variables should be excluded from the regression analysis for their corresponding spatial range.

The coefficient of determination R^2, adjusted coefficient of determination (Adj. R^2), Akaike information criterion (AICc), and residual square sum (RSS) were used to compare the goodness of fit of regression models. The higher the coefficient of determination is, the better the fitting effect of the regression model is. The lower the AICc and RSS are, the better the fitting effect is. The goodness of fit results of MGWR models for 10 analysis ranges of the built environment are shown in Table 3. The MGWR model corresponding to the built environment scale range of 1000 m radius buffer and Thiessen polygon superposition range has the best goodness of fit. The coefficient of determination R^2 is increased to 0.932, which is higher than R^2 in other studies, such as 0.837 [6], 0.841 [17], and 0.687 [36] In addition, this optimal spatial scale range can also provide reference for the range of TOD planning and construction of subway stations. We have provided the data of the 1000 m radius buffer and Thiessen polygon overlay range for some subway stations in this scale range, as shown in Appendix A.

Table 3. Goodness of fit results of MGWR regression models under different analysis scales.

Partition Method	**Buffer Radius**	**500 m**	**800 m**	**1000 m**	**1200 m**	**1500 m**
Circular buffer	R^2	0.904	0.873	0.866	0.772	0.714
	Adj.R^2	0.853	0.827	0.815	0.726	0.659
	AICc	451.083	435.733	457.131	507.001	567.798
	RSS	27.452	36.422	38.577	65.183	82.045
Thiessen polygon overlay range	R^2	0.885	0.915	0.932	0.922	0.917
	Adj.R^2	0.835	0.868	0.889	0.881	0.872
	AICc	445.86	424.935	407.826	386.033	413.161
	RSS	32.995	24.437	19.601	22.526	23.824

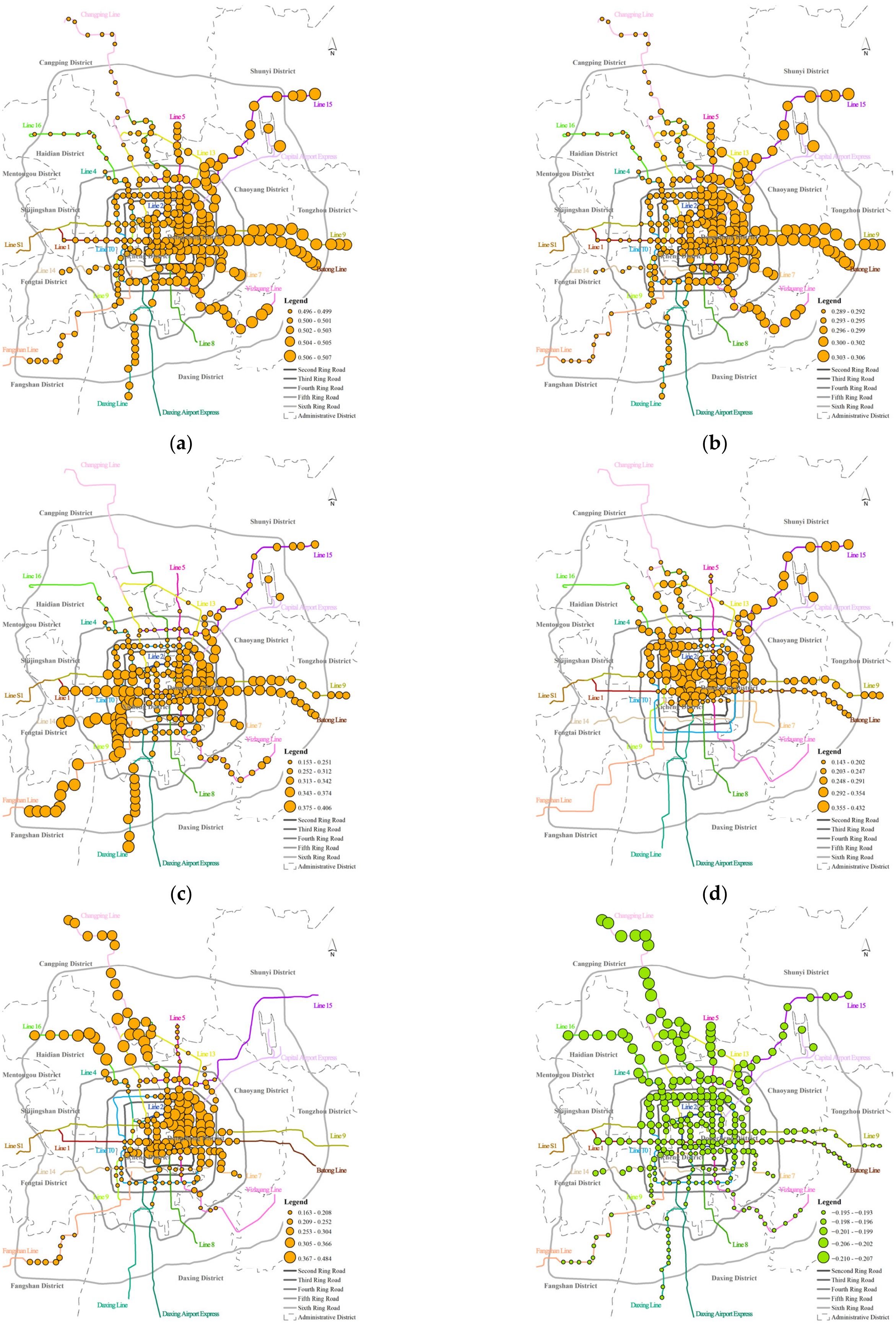
3.4. Analysis of Significant Factors

Table 4 shows the statistics of regression coefficient results of an MGWR model obtained from the overlay range of a buffer zone with radius of 1000 m and Thiessen polygon. It also includes the percentage of subway stations significantly affected based on the adjusted t-value at the 95% confidence level. It can be seen that the morning peak outbound ridership of all stations is affected by the density of office facilities, sports and leisure facilities, medical service facilities, building density and FAR. The factors that influence the subway station more than 50% include the density of catering facilities and the number of parking lots. The factors that influence the subway station less than 20% include the density of scenic spots, commercial, scientific, educational and cultural facilities, residential facilities, the density of subway lines, the density of roads, the number of bus stops, the parking charge standard of the area where the subway station is located, and the distance to the nearest bus stop. The factors that affect the subway station more than 20% but less than 50% include the degree of mixed use of land and population density.

Table 4. Results statistics of local regression coefficients of Variables of MGWR model.

Categories of Built Environment	Variables	Mean	Min	Max	Percentage of Subway Stations Significantly Affected
	Constant term	−0.057	−0.397	0.307	45%
Density	Density of catering facilities	0.297	0.074	0.406	88%
	Density of scenic facilities	0.041	0.039	0.045	0%
	Density of commercial facilities	−0.057	−0.065	−0.053	0%
	Density of science, education and cultural facilities	−0.082	−0.089	−0.079	0%
	Density of residential facilities	0.070	0.064	0.074	0%
	Density of office facilities	0.502	0.496	0.507	100%
	Density of sports and leisure facilities	−0.169	−0.182	−0.164	100%
	Density of medical service facilities	−0.198	−0.210	−0.193	100%
	Building density	−0.259	−0.272	−0.243	100%
Diversity	Mixed utilization of land	−0.018	−0.878	1.041	39%
Design	Road density	0.001	−0.001	0.006	0%
	Floor area ratio	0.297	0.287	0.306	100%
	Density of subway lines	−0.114	−0.121	−0.108	0%
Destination accessibility	Number of parking lots	0.178	−0.148	0.432	62%
	Number of bus stops	−0.035	−0.131	0.070	20%
Traffic distance	Distance to the nearest bus stop	−0.002	−0.079	0.128	3%
Demand management	Parking charge standard in the area where the subway station is located	0.029	−0.236	0.274	17%
Population statistics	Population density	0.079	−0.285	0.624	32%
Other indicators	Attribute of subway station (whether it is a transfer station or not)	0.238	0.072	0.484	71%

Figure 4 shows the local coefficient distribution of some built environment factors based on MGWR model results. The percentage of outbound ridership of subway stations during the morning peak is significantly affected by these built environmental factors, which is more than 20% of all stations. The larger the circle in the figure is, the larger the absolute value of the coefficient of the independent variable is, and the more significant the influence on the ridership of the subway station is. The smaller the circle, the smaller the absolute value of the coefficient of the independent variable, and the less significant the influence on the ridership of the subway station. For the stations that do not display the circle for some independent variables, the results show that these independent variables have no significant effect on the ridership of the subway station. The orange circle means that the independent variable has a positive influence on the ridership of the subway station; that is, the increase in the independent variable will lead to the increase in the ridership of the subway station. The green circle means that the independent variable has a negative influence on the ridership of the subway station; that is, the increase in the independent variable will lead to a decrease in the ridership of the subway station. For a certain influencing factor, the difference in local coefficients reflects the spatial heterogeneity of the influence degree of this factor on the outbound ridership of subway stations. It can be seen that:

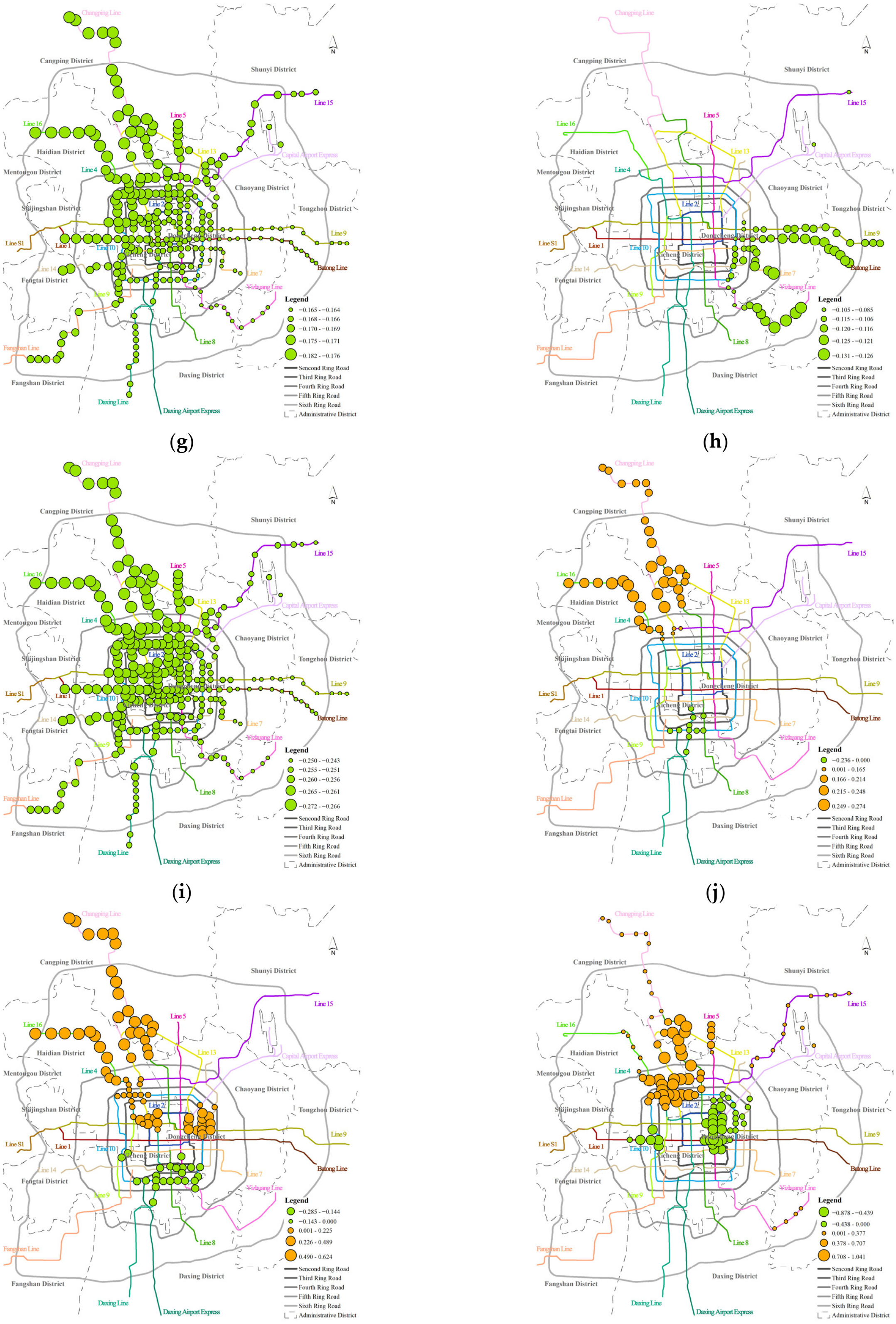


(**a**) (**b**)

(**c**) (**d**)

(**e**) (**f**)

Figure 4. *Cont.*



(**g**) (**h**)

(**i**) (**j**)

(**k**) (**l**)

Figure 4. Local regression coefficient distribution of the built environment influencing factors of outbound ridership at subway stations during the morning peak: (**a**) Office facilities; (**b**) Floor area ratio; (**c**) Catering facilities; (**d**) Number of parking lots; (**e**) Whether it is a transfer station; (**f**) Medical service facilities; (**g**) Sports and leisure facilities; (**h**) Number of bus stops; (**i**) Building density; (**j**) Parking charge standard; (**k**) Population density; (**l**) The degree of mixed land use.

(1) Office facilities, catering facilities, FAR, the number of parking lots and whether it is a transfer station have a positive impact on the morning peak outbound ridership. Office facilities (Figure 4a) and FAR (Figure 4b) have significant impacts on outbound ridership of all subway stations. Under the similar spatial distribution characteristics of the impact coefficient, the closer to the eastern region of Beijing, the greater the impact of office facility density and FAR on the morning peak outbound ridership of subway stations; catering facilities (Figure 4c), the number of parking lots (Figure 4d) and whether they are transfer stations (Figure 4e) have a significant impact on the morning peak outbound ridership of most subway stations. Catering facilities have a greater impact on the southwest and eastern region of Beijing. The number of parking lots has a greater impact on the central district of Beijing's urban area, mainly within the Third Ring Road, and it also has a greater impact on the morning peak outbound ridership of subway stations located in the east and northeast. Whether the station is a transfer station has a greater impact on the ridership of subway stations located between the East Second Ring Road and the East Third Ring Road, and it also has a greater impact on the ridership of subway stations that are located on the northwest subway lines;

(2) Medical service facilities, sports and leisure facilities, the number of bus stops and building density have a negative impact on the morning peak outbound ridership. Medical service facilities (Figure 4f) and sports and leisure facilities (Figure 4g) have significant negative impacts on all subway stations; their spatial distribution characteristics of the coefficients are similar, and the degree of influence increases from the southeast to the northwest. The main purpose of most people's travel during the morning rush hour is for work. For medical service facilities and sports and leisure facilities, the higher the density of such facilities within the determined analysis range, the lower the density of office facilities. Therefore, the higher the density of the two types of facilities, the lower the outbound ridership may result. The number of bus stops (Figure 4h) has a significant impact on the outbound ridership of subway stations, and the percentage of the number of these subway stations is 20%, mainly distributed in the southeast, which has a negative impact on the outbound ridership. The main reason is that the primary function of this area is residence, and the denser the bus stops are, the greater the residential density around the rail stops is, and the density of jobs is relatively small. Building density (Figure 4i) has a significant impact on outbound ridership for all subway stations, which has a negative impact on outbound ridership. The lower the building density is, the higher the outbound ridership is. The main reason is that the core area and the northwest area are surrounded by high-rise or super high-rise buildings, which are characterized by low building density, high FAR and more job chances. Therefore, the outbound ridership of these subway stations is relatively high; however, the buildings in the east and south are mainly residential with a high building density and relatively few job chances, resulting in a relatively low outbound ridership;

(3) Parking charge standard, population density and degree of mixed land use have different positive and negative impacts on the morning peak outbound ridership due to different locations. The parking charge standard (Figure 4j) has a positive impact on the morning peak outbound ridership of subway stations along the northern section of Changping Line and Line 16, and the increase in parking fee has a positive impact on the increase in outbound ridership of the subway stations in the region. The population density (Figure 4k) has a positive impact on the morning peak outbound ridership of subway stations, for example, stations on Changping Line which is located in northwest Beijing, stations along the northern section of Line 16, stations on Line 2, stations in the western section of Chaoyang District of Line 10, and the subway stations along the eastern section of Line 14, all of which have high working population density. However, in the east of Fengtai District, the population density around the subway stations on Line 4, Daxing Line, Line 2, Line 7 and the south section of Line 8, has a negative correlation effect on the morning peak outbound ridership, and the degree of impact is relatively small. Such stations are mainly located in the south, where the residential population density is high and the density of employment units is low, resulting in relatively small morning peak outbound

ridership. The degree of land use mix (Figure 4l) has a positive impact on the morning peak outbound ridership of the subway stations located in the southeast of Haidian District and the northern section of Line 8, and its impact is relatively large, mainly because the main function of the land around these subway stations is high-tech services, and the degree of mixed use of surrounding land is relatively low. If commercial and public service facilities are added to improve the mixed use of land, the outbound ridership will be increased; in the west of Chaoyang District, namely the subway stations along Line 2, the east section of Line 10, and the northeast section of Line 14, the mixed degree of land use has a negative impact on the outbound ridership during the morning peak, and the impact is relatively large, mainly because these subway stations are surrounded by CBD business districts and embassy districts with high density of employment units. To improve the mixed degree of land use within the limited space, increasing the scale of other land uses will reduce the land with high density of employment units, and it may reduce the outbound ridership of subway stations.

The density of office facilities, sports and leisure facilities, medical service facilities, building density and FAR have a significant impact on the morning peak outbound ridership of all subway stations. The influence of office facilities, catering facilities, FAR, the number of parking lots, whether subway stations are transfer stations and other built environmental factors on the outbound ridership during the morning peak is positive. The number of medical service facilities, sports and leisure facilities, bus stops and building density have a negative impact on the morning peak outbound ridership on weekdays. Parking charge, population density and land use mix, which are three built environment factors, have different positive and negative impacts on the morning peak outbound ridership of subway stations due to their different locations. Beijing is faced with the characteristics of overload operation of subway stations. In the process of updating and transforming the built environment for the purpose of relieving and guiding the ridership of subway stations, it is suggested that priority should be given to adjusting the built environment factors that have a significant impact on the ridership of subway stations.

Exploring the relationship between the built environment and subway station ridership can provide reference for ridership prediction and urban planning decision making. For urban policy makers, they can ease the function or guide the passenger flow according to the ridership of subway stations in different regions of the city and their significant influencing factors, so as to promote the development of urban TOD. For example, for subway stations with over-loaded ridership, on the one hand, the excessive concentration of passenger flow can be alleviated by increasing the supply of subway line facilities; on the other hand, the demand for passenger flow can be reduced by transferring or reducing those facilities with significant positive impact on the station. For subway stations with small ridership, it is recommended to add facilities that have a greater positive impact on them, and further improve the traffic connection facilities such as sharing bicycles and buses, to improve the traffic accessibility. For urban planners, the ridership of subway stations can be preliminarily predicted according to the built environment of the city, and the planning and updating strategies can be timely adjusted accordingly.

3.5. Model Bandwidth Analysis

Fotheringham et al. [19] used the back-fitting algorithm to optimize the independent variable bandwidth of MGWR, reflecting the difference in the impact scale of every independent variable, and improving the accuracy of the model. The optimal bandwidth of the subway station outbound ridership model represents the number of surrounding subway stations to be calculated when estimating the local coefficients of various influencing factors. The total number of subway stations is 287. When the optimal bandwidth is close to this number, the local coefficient solution is affected by most stations, and this influencing factor is a global scale variable. The smaller the optimal bandwidth is, the smaller the spatial influence scale is, and the local coefficients are greatly affected by nearby stations, which is a local scale variable. Table 5 lists the optimal bandwidth of each influencing factor of MGWR model.

Table 5. Optimal bandwidth of variables results of MGWR model.

Categories of Built Environment	Variables	Bandwidth
	Constant term	43
Density	Density of catering facilities	96
	Density of residential facilities	286
	Density of office facilities	286
	Density of sports and leisure facilities	286
	Density of medical service facilities	286
	Building density	286
Diversity	Mixed utilization of land	43
Design	Road density	286
	Floor area ratio	286
Destination accessibility	Number of parking lots	53
	Number of bus stops	184
Traffic distance	Distance to the nearest bus stop	157
Demand management	Parking charge standard in the area where the subway station is located	64
Population statistics	Population density	43
Other indicators	Attributes of subway stations	43

Among the categories of density and design of built environment, most of the variables have a scale of overall importance, including the density of residence, office, sports and leisure, medical service facilities, as well as building density, road density and FAR, indicating that these influencing factors have a relatively stable spatial impact on the outbound ridership of subway stations. The number of bus stops and the distance from the subway station to the nearest bus stop also have a larger scale effect. The degree of mixed use of land, the number of parking lots, the population density, and the attributes of subway stations have a small-scale effect. The subway outbound passenger flow is more sensitive to the impact of its location, and the minimum scale of land mixed use, population density and subway station attributes is 43.

4. Conclusions

Most research on subway station ridership regression models does not consider the MAUP's scale and zone effects, which may not achieve a better goodness of fit result. The objective of this paper is to increase the accuracy of the regression model between ridership of subway stations and the built environment and explore the effects of the spatial heterogeneity of built environment factors on subway station ridership. The authors expand the built environment variables from the "5D" category (density, diversity, design, distance to transit and destination accessibility) to the "7D" category (density, diversity, design, distance to transit, destination accessibility, demand management and demographics). The explanatory variables we use in this research include 16 variables used in previous research and three variables added in this study: parking fee standard, population density and whether it is a transfer station. The dependent variable is the outbound ridership of subway stations during the morning peak on weekdays. Considering the MAUP, two zoning methods were selected to delimit the spatial range of the built environment: one method takes each subway station as the center, and selects 500 m, 800 m, 1000 m, 1200 m and 1500 m, respectively, as the radius to delimit the circular buffer zone; another method is called the "Thiessen polygon overlay range", which is created through ArcGIS intersecting analysis between multiple circle buffers and multiple Thiessen polygons. This study designs

10 scale range scenarios, creates a data set with built environment as the independent variable, and subway station inbound ridership as dependent variable for each scenario. The optimal scale range of the built environment around subway stations is obtained by comparing goodness of fit of MGWR models under 10 scale range scenarios. The MGWR model corresponds to the scale range overlapped by the 1000 m radius circular buffer zone, and Thiessen polygon has the best goodness of fit result and its coefficient of determination is also higher than that in other studies. The area overlapped by a 1000 m radius circular buffer zone and Thiessen polygon is regarded as the optimal scale range of built environment around subway stations in Beijing. The MGWR local coefficients indicate the spatial heterogeneity of the built environment factors at different stations. However, the spatial heterogeneity of built environment variables is different with research results on other cities. It is very difficult to compare spatial heterogeneity among different cities. The results can provide an important theoretical basis for prediction and analysis of ridership demand at subway stations, also for the integration of built environment around the stations in Beijing.

There are several other limitations to note. First of all, due to the limitation of data acquisition, this study only acquired the outbound ridership data of the subway station within a week. In future research, it is hoped to obtain more abundant ridership data of different periods of time, to explore the similarities and differences in the influences of the built environment on different periods of the day and on the weekend ridership. Secondly, the explanatory variables in this paper lack the authoritative data of Beijing's resident population and employment distribution, and there may be some errors in the estimation results of the model. In the future, more detailed population survey data should be used for research to improve the accuracy of the model. Third, this study only uses the subway ridership data of Beijing. In future studies we hope to obtain more passenger flow data from other cities, to explore whether there is regularity of significant factors affecting different cities. The method proposed in this paper has obtained useful goodness of fit results in the regression analysis of 287 subway stations in Beijing. However, this method has certain limitations in the application of cities or other counties with fewer subway stations, because the MGWR model requires a certain sample size to obtain effective goodness of fit results. Therefore, how many subway stations can achieve satisfactory goodness of fit of the regression model is a problem that needs further discussion.

Author Contributions: Conceptualization, Z.W. and J.J.; Data curation, J.S., S.L., J.J. and C.S.; Methodology, Z.W. and C.S.; Software, J.S. and S.L.; Supervision, Z.W.; Writing—original draft, Z.W., J.S. and Y.Z.; Writing—review and editing, Z.W. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Hebei Social Science Development Research Project, China (grant No. 20210201407).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

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

Abbreviations

Adj. R^2	Adjusted Coefficient of Determination
AICc	Akaike Information Criterion
API	Application Program Interface
DRM	Direct Demand Models
FAR	Floor Area Ratio
GWR	Geographically Weighted Regression
MAUP	Modifiable Area Unit Problem
GWR	Geographically Weighted Regression
MGWR	Multi-Scale Geographical Weighted Regression
OLS	Ordinary Least Squares
PCA	Pedestrian Catchment Area
RSS	Residual Square Sum
SEM	Structural Equation Model
TOD	Transit Oriented Development
VIF	Variance Impact Factor

Notations

A_i	The total area of building base within the analysis range
B_i	The building density within the analysis range of subway station *i*
C_i	The sum of subway line lengths within the analysis range
D_i	The density of certain type of POI facilities at subway station
E_i	The floor area ratio of the analysis range of subway station *i*
G_i	The population within the analysis range
H_i	The degree of mixed use of land at subway station *i*
K_i	The ratio of the number of POI facility points of a certain type within the analysis range of subway station *i* to the number of all POI facility points
L_i	The road density within the analysis range of subway station *i*
M_i	The density of rail lines within the analysis range of subway station i
N_i	The number of POI facility points of a certain type within the analysis range of subway station *i*
O_i	The total building area within the analysis range
P_i	The population density within the analysis range of subway station
R_i	The sum of road lengths within the analysis range
S_i	The total area of the analysis range
y_i^*	The standardized value of the morning peak outbound ridership of subway station *i*
β_{bw0}	The regression model constant of subway station *i* with coordinate (u_i, v_i)
β_{bwk}	The local regression coefficient of the *k-th* built environmental influencing factor of subway station *i*
x_{ik}^*	The standardized value of the *k-th* built environmental influencing factor indices of subway station
ε_i	The random error of subway station *i*; *n* is the number of factors affecting the built environment
$\overline{y}$	The average value of outbound ridership at the morning peak of subway station
s_y	The standard deviation of the morning peak outbound ridership of the subway station
x_{ik}	The indices of the k-th built environmental influencing factor of subway station *i*
$\overline{x_k}$	The average value of the *k-th* influencing factor indices
s_{xk}	The standard deviation value of the *k-th* influencing factor indices

Appendix A

Table A1. Some subway station data under the 1000 m radius buffer and Thiessen polygon overlay range.

Variables	**Stations**	**Chang pingxishankou**	**Gonghua cheng**	**Beianhe**	**Daoxiang hulu**	**Tundian**	**Shange zhuang**	**Cuige zhuang**	**Dongxia yuan**	**Lucheng**	**Huagong**
Morning peak ridership of subway		299	152	477	1204	184	286	875	709	602	356
Density	Density of catering facilities	0.75	0.32	1.27	0.32	14.34	11.72	20.41	0.85	1.88	13.78
	Density of scenic facilities	0.75	0.32	0.32	0.00	0.64	0.42	0.00	0.43	0.75	0.49
	Density of commercial facilities	0.00	0.00	0.32	0.00	1.91	0.42	1.51	0.00	0.75	0.98
	Density of science, education and cultural facilities	3.38	1.27	0.64	1.27	7.33	4.19	9.82	1.28	3.75	7.38
	Density of residential facilities	1.50	0.32	2.55	1.59	4.78	2.51	7.94	3.41	0.75	7.87
	Density of office facilities	3.01	1.59	7.32	5.73	18.48	22.61	12.09	8.10	18.77	74.80
	Density of sports and leisure facilities	1.13	0.32	5.09	0.00	2.87	5.44	1.89	0.00	2.63	2.95
	Density of medical service facilities	0.75	0.00	0.64	0.00	2.87	1.26	2.27	0.00	1.50	0.98
	Building density	0.01	0.01	0.02	0.03	0.08	0.06	0.04	0.01	0.00	0.23
Diversity	Mixed utilization of land	0.03	0.03	0.03	0.08	0.03	0.03	0.04	0.06	0.03	0.17
Design	Road density	7.50	4.20	7.02	9.05	6.15	6.20	2.52	8.10	5.89	9.60
	Floor area ratio	0.04	0.05	0.07	0.05	0.17	0.18	0.12	0.06	0.00	0.65
	Density of subway lines	0.64	0.70	0.88	0.64	0.63	0.30	0.75	0.54	0.33	0.57
Destination accessibility	Number of parking lots	0	1	1	0	0	6	3	2	8	19
	Number of bus stops	8	1	3	5	5	2	4	7	7	1
Distance to transit	Distance to the nearest bus stop	88.58	8.92	114.15	268.65	379.19	529.83	33.63	259.35	20.33	819.20
Demand management	Parking charge standard in the area where the subway station is located	0.5	0.4	0.5	1	0.5	2	0.5	0.3	0.5	1.5
Demographics	Population density	3458	5460	6560	3322	5554	11821	9353	2759	1930	20441
Others	Attribute of subway station	0	0	0	0	0	0	0	0	0	0

References

1. Ratner, K.A.; Goetz, A.R. The reshaping of land use and urban form in Denver through transit-oriented development. *Cities* **2013**, *30*, 31–46. [CrossRef]
2. Tan, Z.; Li, S.; Li, X.; Liu, X.; Chen, Y.; Li, W. Spatio-temporal effects of urban rail transit on complex land-use change. *Acta Geogr. Sin.* **2017**, *72*, 850–862.
3. Kuby, M.; Barranda, A.; Upchurch, C. Factors influencing light-rail station boardings in the United States. *Transp. Res. Part A Policy Pract.* **2004**, *38*, 223–247. [CrossRef]
4. Gutiérrez, J.; Cardozo, O.D.; García-Palomares, J.C. Transit ridership forecasting at station level: An approach based on distance-decay weighted regression. *J. Transp. Geogr.* **2011**, *19*, 1081–1092. [CrossRef]
5. Loo, B.P.Y.; Chen, C.; Chan, E.T.H. Rail-based transit-oriented development: Lessons from New York City and Hong Kong. *Landsc. Urban Plan.* **2010**, *97*, 202–212. [CrossRef]
6. Gao, D.; Xu, Q.; Chen, P.; Hu, J.; Zhu, Y. Spatial characteristics of urban rail transit passenger flows and fine-scale built environment. *J. Transp. Syst. Eng. Inf. Technol.* **2021**, *21*, 25–32.
7. Sung, H.; Choi, K.; Lee, S.; Cheon, S.H. Exploring the impacts of land use by service coverage and station-level accessibility on rail transit rid-ership. *J. Transp. Geogr.* **2014**, *36*, 134–140. [CrossRef]

8. Sun, L.S.; Wang, S.W.; Yao, L.Y.; Rong, J.; Ma, J.M. Estimation of transit ridership based on spatial analysis and precise land use data. *Transp. Lett.* **2016**, *8*, 140–147. [CrossRef]
9. Guerra, E.; Cervero, R.; Tischler, D. The half-mile circle: Does it represent transit station catchments? *Univ. Calif. Transp. Cent. Work. Pap.* **2012**, *2276*, 101–109. [CrossRef]
10. Li, S.; Lyu, D.; Liu, X.; Tan, Z.; Gao, F.; Huang, G.; Wu, Z. The varying patterns of rail transit ridership and their relationships with fine- scale built environment factors: Big data analytics from Guangzhou. *Cities* **2020**, *99*, 102580. [CrossRef]
11. Sohn, K.; Shim, H. Factors generating boardings at Metro stations in the Seoul metropolitan area. *Cities* **2010**, *27*, 358–368. [CrossRef]
12. Sung, H.; Oh, J.T. Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities* **2011**, *28*, 70–82. [CrossRef]
13. Zhao, J.; Deng, W.; Song, Y.; Zhu, Y. What influences Metro station ridership in China? Insights from Nanjing. *Cities* **2013**, *35*, 114–124. [CrossRef]
14. Zhao, J.; Deng, W.; Song, Y.; Zhu, Y. Analysis of Metro ridership at station level and station-to-station level in Nanjing: An approach based on direct demand models. *Transportation* **2014**, *41*, 133–155. [CrossRef]
15. Cardozo, O.D.; García-Palomares, J.C.; Gutiérrez, J. Application of geographically weighted regression to the di-rect forecasting of transit ridership at station-level. *Appl. Geogr.* **2012**, *34*, 548–558. [CrossRef]
16. Calvo, F.; Eboli, L.; Forciniti, C.; Mazzulla, G. Factors influencing trip generation on metro system in Madrid (Spain). *Transp. Res. Part D Transp. Environ.* **2019**, *67*, 156–172. [CrossRef]
17. Cong, Y.; Wang, Y.; Yu, L.; Li, G. Spatial-temporal effects of land use factors on metro station passenger flow. *Urban Mass Transit* **2021**, *24*, 116–121.
18. Jun, M.J.; Choi, K.; Jeong, J.E.; Kwon, K.H.; Kim, H.J. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *J. Transp. Geogr.* **2015**, *48*, 30–40. [CrossRef]
19. Fotheringham, A.S.; Yang, W.; Kang, W. Multiscale geographically weighted regression (MGWR). *Ann. Am. Assoc. Geogr.* **2017**, *107*, 1247–1265. [CrossRef]
20. Yu, H.; Fotheringham, A.S.; Li, Z.; Oshan, T.; Kang, W.; Wolf, L.J. Inference in multiscale geographically weighted regression. *Geogr. Anal.* **2019**, *52*, 87–106. [CrossRef]
21. Shen, T.; Yu, H.; Zhou, L.; Gu, H.; He, H. On hedonic price of second-hand houses in Beijing based on multi-scale geographically weighted regression: Scale law of spatial heterogeneity. *Econ. Geogr.* **2020**, *40*, 75–83.
22. Chen, Z.; Deng, S. Research on the relationship between block morphology and heat island intensity in Guangzhou's main urban area based on multi-scale geographically weighted regression. *Intell. Build. Smart City* **2021**, *10*, 13–17.
23. Zhou, Z.; Cheng, X. Spatial heterogeneity of influencing factors of PM2.5 in Chinese cities based on MGWR model. *China Environ. Sci.* **2021**, *41*, 2552–2561.
24. Wang, Y. Research on the Interaction between Land Use of Station Areas and Passenger Flow of Subway in Beijing. Ph.D. Thesis, Tsinghua University, Beijing, China, 2018.
25. Chen, E.; Ye, Z.; Wang, C.; Zhang, W. Discovering the spatio-temporal impacts of built environment on metro ridership using smart card data. *Cities* **2019**, *95*, 102359. [CrossRef]
26. Alkhereibi, A.; Alsuwaidi, M.; Al-Mohammed, R.; Pokharel, S.; Ayari, M. An integrated urban-transport smart growth model around metro stations: A case of Qatar. *Transp. Res. Interdiscip. Perspect.* **2021**, *10*, 100392. [CrossRef]
27. Yang, L.; Song, X. TOD typology based on urban renewal: A classification of metro stations for Ningbo city. *Urban Rail Transit* **2021**, *7*, 240–255. [CrossRef]
28. Ewing, R.; Cervero, R. Travel and the built environment: A synthesis. *Transp. Res. Rec.* **2001**, *1780*, 87–114. [CrossRef]
29. Cervero, R.; Sarmiento, O.L.; Jacoby, E.; Gomez, L.F.; Neiman, A. Influences of Built Environments on Walking and Cycling: Lessons from Bogotá. *Int. J. Sustain. Transp.* **2009**, *3*, 203–226. [CrossRef]
30. Ewing, R.; Cervero, R. Travel and the built environment. *J. Am. Plan. Assoc.* **2010**, *76*, 265–294. [CrossRef]
31. Chow, A.S.Y. Urban Design, Transport Sustainability and Residents' Perceived Sustainability: A Case Study of Transit-oriented Development in Hong Kong. *J. Comp. Asian Dev.* **2014**, *1*, 73–104. [CrossRef]
32. Higgins, C.D.; Kanaroglou, P.S. A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region. *J. Transp. Geogr.* **2016**, *52*, 61–72. [CrossRef]
33. Li, C.; Lin, C.; Hsieh, T. TOD District Planning Based on Residents' Perspectives. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 52. [CrossRef]
34. De Gruyter, C.; Ma, L.; Saghapour, T.; Dodson, J. How does the built environment affect transit use by train, tram and bus? *J. Transp. Land Use* **2020**, *13*, 625–650. [CrossRef]
35. Li, S.; Lyu, D.; Huang, G.; Zhang, X.; Gao, F.; Chen, Y.; Liu, X. Spatially varying impacts of built environment factors on rail transit ridership at station level: A case study in Guangzhou, China. *J. Transp. Geogr.* **2020**, *82*, 102631. [CrossRef]
36. Yang, H.; Xu, T.; Chen, D.; Yang, H.; Pu, L. Direct Modeling of Subway Ridership at Station Level: A Study Based on Mixed Geographically Weighted Regression. *Can. J. Civ. Eng.* **2020**, *47*, 534–545. [CrossRef]
37. Bai, T.; Cai, L.; Zhu, J.; Lin, X. Spatial Differentiation of Rail Transit and Urban Development Coordination: A Case Study of Beijing. *J. Transp. Syst. Eng. Inf. Technol.* **2020**, *20*, 14–19.
38. Clark, A.; Scott, D. Understanding the impact of the modifiable areal unit problem on the relationship between active travel and the built environment. *Urban Stud.* **2014**, *51*, 284–299. [CrossRef]

39. Yang, L.; Hu, L.; Wang, Z. The built environment and trip chaining behavior revisited: The joint effects of the modifiable areal unit problem and tour purpose. *Urban Stud.* **2019**, *56*, 795–817. [CrossRef]
40. Zhou, X.; Yeh, A. Understanding the modifiable areal unit problem and identifying appropriate spatial unit in jobs–housing balance and employment self-containment using big data. *Transportation* **2021**, *48*, 1267–1283. [CrossRef]
41. Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* **2015**, *10*, e0107042. [CrossRef]
42. Lloyd, C.T.; Sorichetta, A.; Tatem, A.J. High resolution global gridded data for use in population studies. *Sci. Data* **2017**, *4*, 170001. [CrossRef] [PubMed]
43. Mansour, S.; Al Kindi, A.; Al-Said, A.; Al-Said, A.; Atkinson, P. Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). *Sustain. Cities Soc.* **2020**, *65*, 102627. [CrossRef] [PubMed]
44. Gujarati, D.N.; Porter, D.C. *Basic Econometrics*, 5th ed.; McGraw-Hill/Irwin: New York, NY, USA, 2009; pp. 157–158.