WILEY | Hindawi

Research Article

Exploring the Spatial Variation of Access/Egress Distances of Subway Stations Using Mobile Phone Positioning Data in Chengdu, China

Chao Wang [ID],[1] **Zhongquan Qiu** [ID],[1] **Renbin Pan** [ID],[2] **Xiaojian Wang** [ID],[1] **and Yusong Yan**[1]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China*
[2]*School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China*

Correspondence should be addressed to Zhongquan Qiu; qiuzhongquan@swjtu.edu.cn

The distance from the origin or destination to or from the subway station is defined as the access or egress distance, which determines the service coverage of the subway station. However, little literature studies the distances at the station level, and they may vary from station to station. Therefore, this study aims to explore the influencing factors and spatial variation of the distances at the station level by using the mobile phone positioning data of more than 1.2 million anonymous users in Chengdu, China. First, this study proposes a method to extract the access and egress trips of the subway. Next, the ordinary least squares (OLS) regression models are carried out to select the significant explanatory variables. Finally, the geographically weighted regression (GWR) models are used to model the spatial variation relationship between the 85th percentile access/egress distances and the selected explanatory variables. The results show that different stations' access/egress distances vary significantly in space. Hotel, residence, life, finance, road density, and mixed land use are found to be negatively correlated with distances, while education, 36–45 years old, male, and high education are positively correlated. In addition, the GWR model reveals that the influence of explanatory variables on access/egress distance varies from space to space. The results further promote the understanding of the existing system and provide a relevant reference for planners and transportation departments to optimize land use and public transportation planning.

1. Introduction

The subway in China is experiencing rapid growth to cope with new problems related to urban expansion, traffic congestion, and air pollution [1–3]. At the same time, the service quality of the subway is constantly improving, with longer service time and higher service reliability [4]. In addition to efforts to improve the service quality of the subway itself, providing a high-quality travel environment around stations is another way to increase its attractiveness. Therefore, some cities vigorously develop transit-oriented development (TOD) projects based on compact, mixed-use, pedestrian- and bicycle-friendly urban construction concepts to realize the effective strategy of changing the mode of short-distance into walking and bicycle and long distance

from automobile to subway [3]. Naturally, the construction of subway stations must be within a reasonable distance to be convenient for residents to use. However, an important issue is how to define an acceptable and practical distance to walk, bicycle, and other modes of transportation to conveniently access/egress the station for most subway passengers and potential passengers from their homes, workplaces, schools, and other locations. The answer to this question can help provide information for planning and design decisions of TOD's scale and geographical scope.

In recent years, more and more scholars have begun to pay attention to studying the access/egress distances of subway stations. The proportion of the population served by the transportation system is a crucial indicator of system performance. Therefore, when determining the service area

around the station, most studies usually use the over-simplified method that most people walk 800 meters to the station to define it [5, 6]. However, the accuracy and appropriateness of this one-size-fits-all approach are often questioned in other studies [7]. Because people usually walk to the subway station, also by bicycle, automobile, and bus. In Sydney, Australia, only 50% walk, 34% by automobile, and 14% by bus [5]. In Beijing, China, 65.25% walk, 19.91% by bus, 14.85% by bicycle, and 5.28% by automobile [8]. In Toronto, Canada, buses and trams connected to the subway account for more than one-third of all passengers [9]. Most studies also admit that walking is a better choice for short distances, but within the acceptable travel time, using bicycles as the connection mode makes the subway more attractive and expands the subway station's service area. For example, Zuo et al. [10] indicated that the bicycle distance is 1.7 to 2.3 times walking in Shanghai. Lee, et al. [11] found that bicycles can expand urban service areas from 29.9% to 93.6% in Seoul. Other modes of transportation also increase the accessibility of the subway system, but there are apparent variations in different ways. Therefore, the planning of subway stations needs to consider the connection of various traffic modes.

The lack of appropriate data is one of the main bottlenecks in transportation research, which significantly affects the accuracy of the results. Existing research often relies on the overall inference from census data or estimates the distance at the level of individual trips based on small sample surveys and trajectory data. However, the low diversity and quantity of data are still challenging. More importantly, subway stations' access and egress distances may vary from station to station and from space to space. This has not been fully revealed in the existing research. At present, new data sources related to information and communication technology have emerged: mobile phone positioning data. These data are based on the records generated by the interaction between the mobile phone and the points of interest (POI) and the timing report every half hour under a continuous bright screen condition. Mobile phone positioning data can record users' arrival and departure times at specific places (such as subway stations). More importantly, the basic demographic information is stored in these data, which can be used to understand the travel characteristics of various user groups and reveal the observed behaviors. In this way, we can realize the complete end-to-end travel inference to detect human movement behavior's macro and micro levels.

Based on this background, this paper explores the access and egress distances of subway passengers for the first time using mobile phone positioning data of 1.2 million users in Chengdu, China, which has a larger data sample and is closer to the real travel conditions of residents than the questionnaires and other data used in previous studies. In fact, the 85th percentile access and egress distance of each subway station is an important indicator to evaluate the accessibility of subway services. Therefore, this paper proposes using the geographically weighted regression (GWR) model at the station level to reveal the influencing factors of the spatial variation of subway stations' access/egress distances. It attempts to answer three questions: (1) How do the access and egress distances of subway stations at the station-level change in space? (2) What are the influencing factors of the access/egress distances? (3) How does the influence of influencing factors on the access/egress distances change in space? These results can provide a helpful reference for planners and city managers to optimize new subway stations and transportation systems. In particular, the Chengdu rail transit group is vigorously developing the TOD project, and a large-scale analysis of urban residents' activities in Chengdu can determine the applicable distance threshold for the urban environment with a unique background. In addition, this study also shows the potential of mobile phone positioning data in exploring residents' subway access/egress distances, which can provide a certain degree of explanation for the activity intentions of different groups of people.

Next, the literature on access and egress distances of the subway is reviewed. The third section describes the study area and data. The fourth section introduces the method. The fifth section is the explanation and discussion of the results. Finally, the last section is the conclusion.

2. Literature Review

This section reviews the research related to this topic in recent years, mainly divided into two aspects: the access/egress distances of the subway for different modes of transportation and the influencing factors of access/egress distances.

2.1. Access/Egress Distances of the Subway. As the most critical factor affecting the quality of the mode exchange, researchers have perfomed much work in this area, especially for walking into subway stations. However, the distance thresholds obtained from different studies vary due to the differences in study areas and data sources. For example, on the one hand, El-Geneidy et al. [7] found that the average walking distance of the subway from home is 0.564 km, and the 85th percentile distance is 0.873 km through the Montreal OD survey in Montreal, Canada. On the other hand, Daniels and Mulley [5] determined the longer walking distance through the family travel survey in Sydney, Australia, with an average of 0.805 km and the 75th percentile of 1.018 km. Given that the distances between access and egress are not clearly distinguished in previous studies, Wang and Cao [12] analyzed the walking egress distances through the 2010 Transit Onboard Survey in the Minneapolis and St. Paul Metropolitan Area. They concluded that the average distance is 0.494 km, and the 85th percentile distance is 0.845 km. Later, Tao et al. [13] examined the distances between home and subway stations in the same city using the 2016 transit on-board survey and found a shorter average walking distance of 0.317 km. Under the Chinese background, He et al. [14]'s questionnaire survey of Nanjing Metro showed a longer distance and walking distances to the subway station range from 1.050 to 1.2 km.

The bicycles have the potential to promote the use of subways by connecting stations with origins or destinations. In recent years, the transfer between bicycles and subways

has become a research hotspot. Recent studies also show that bike-sharing has expanded the catchment area of subway stations, but the extent is various. For example, Rastogi and Krishna Rao [15] found that the access distances of subway stations are 1.8–4.05 km by investigating the operation system in Mumbai, India. However, Pan et al. [16] examined the access distances using questionnaire surveys in Shanghai and concluded that more than 70% of bicycles travel within 1.5 km, and only about 5% are more than 2.5 km. By comparing the distance variations between different cities, Hochmair [17] proposed that the service area of stations in Los Angeles is 2-3 times that of Atlanta and the Twin Cities. In addition, the median access distances observed are within the buffer radius of the proposed community hub (1.609 km) and the gateway hub (3.218 km). As an essential part of residents' mobility, commuter access can significantly reflect the service quality of transportation. According to surveys in the Seoul Metropolitan and Deajeon Metropolitan Areas, Lee et al. [11] found that the distances from home to the station and from the station to work are estimated to be 1.96 km and 2.13 km, respectively. To get a more accurate value, Zuo et al. [10] used Cincinnati GPS-based household travel survey data and examined the distance threshold of bicycles, which was more than twice that of walking (4.36 km vs. 1.30 km).

Besides walking and bicycle, the distances of other modes of transportation are also compared. For example, Wang et al. [8] compared several main modes of transportation in Beijing, indicating that the average distance of walking is the lowest, which is 0.43 km, while the others are in order: the bicycle is 1.452 km, the bus is 6.262 km, and the automobile is 9.115 km. Xi et al. [9] analyzed the Transportation Tomorrow Survey in Toronto, Canada, indicating that subways' service areas connected with buses and trams are critical because they account for more than one-third of all passengers. In space, the radius of the pedestrian service area is generally less than 1.609 km, while the bus, tram, and automobile are often many times larger.

Most of the above studies are based on the data collected from traditional travel surveys, such as questionnaire-based interviews and travel OD surveys. Access/egress distances are usually obtained from the participants' own reports, and are assumed to be capable of remembering (or willing to share) the actual activity route and movement distance. However, Weinstein Agrawal et al. [18] indicated that only half of the people's actual distances are similar to those they remember. Although GPS data can provide the most accurate spatial trajectory of personal movement patterns, they cannot be used on a larger scale [13]. Both travel surveys and personal GPS are small samples of nonpopulation data. These data limit the sample's geographic and demographic coverage, making them challenging to reflect distance patterns fully.

With the development of information and communication technology, large-scale data about human spatio-temporal motion trajectories can be obtained from many sources, such as transportation network companies and social media data. Using new data, researchers have the opportunity to solve the traditional problem of distance calculation caused by limited samples. Furthermore, more

and more academic studies have shown the potential of large-scale data as an alternative source of travel behavior information, which can be used to derive the origin-destination matrix. For example, some studies have analyzed the transfer distances in bike-sharing utilizing the GPS and order data about Mobike in Shanghai. The results showed that bicycle distances have increased compared with walking distances [19, 20]. However, these current studies mainly focus on the distance of a single mode of transportation. More importantly, it is not known whether users of bike-sharing really transfer to/from the subway. Thus, the usage pattern lacks a comprehensive exploration based on sufficient data covering different modes of transportation.

2.2. Factors Influencing the Access/Egress Distances.

Many factors may affect people's use of the subway, which leads to significant differences in the access/egress distances of the subway in various environments. Therefore, the relationships between distances and critical factors such as the built environment and user characteristics must be fully understood.

2.2.1. Built Environment.

The built environment refers to the artificial environment provided for human activities, including various forms of buildings (such as residential, industrial, and commercial), infrastructure (such as transportation and parks), and urban space [21]. It is generally believed that the built environment has a significant influence on shaping the mode of human mobility and activities, which may be directly related to their accessibility to subway stations and perceived convenience [22, 23]. In particular, when there is no environment suitable for pedestrians, people's decision to drive instead of walking to the station can be affected [24]. Some studies have found that high density at intersections and roads positively correlated with walking distances. They also indicated that population density negatively correlated with walking distances [7, 13]. In addition, these factors are negatively correlated with bicycle distances [17]. Unlike previous studies that take individuals as the analysis unit, Lin et al. [20] established a regression model at the station level. Their results showed that the subway stations' catchment area is positively correlated with the distance to the city center but negatively correlated with the density of subway stations. Later, Li et al. [19] used the same level to investigate the relationships between the 85th percentile distances of different subway stations and the built environment. Many built environment factors are related to distances, but their relationships show some variations in space. Generally speaking, these studies reveal the influence of the built environment on the walking or bicycle distances of subway stations, which provide extensive enlightenment for the regional planning of the station.

2.2.2. User Characteristics.

Scholars generally believe that demographic characteristics significantly impact the distance to the subway station. However, there are apparent

differences in the direction and degree of personal characteristics (including age, gender, income, and other social factors). The representative point of view is that there are differences in walking distances in terms of age. Young people are more likely to walk to the station and have a longer walking distance [7, 14, 24, 25]. In terms of gender, males tend to walk or bicycle for a longer distance to the station than females [7, 25], but He et al. [14] thought that there is no difference between genders. Since car ownership, family income, and family size also have a negative impact on walking distances because families with more vehicles, higher income, and more members are more likely to choose to travel by car but less likely to live near the subway [7, 24, 25]. When considering the travel purpose, compared with shopping travelers, working travelers have the most extended walking distances and the highest possibility of walking to the subway station [14]. For users in bike-sharing, Ma et al. [26] found that the distances between males are higher than that of females, and urban residents are shorter than suburban residents. In addition, they also analyzed the possible influence of time on distance. Considering the impact of travel habits on distances, Lin et al. [20] proposed that the higher the proportion of a single user of the subway station, the greater the service distance.

2.3. Summary. Although more and more evidence shows that the distance between different modes of travel is different, due to the limitation of data sources, the existing literature mainly studies the distance of a single connection mode or the distance of individual travel levels. Although some existing literature has studied access and egress distances at the station level, very little literature has explored how distances vary spatially at the station level and the existence of spatial correlations of distances. In addition, there are significant differences in the direction and degree of key factors affecting distance in different studies. Therefore, the discussion on policies or plans to measure the scope of subway services is limited. In order to fill these research gaps, this paper tries to use the geographically weighted regression (GWR) model to explore the spatial change of subway access and egress distances at the station level. The variable coefficient in the GWR model is allowed to change with space, while the variable coefficient in general ordinary least squares is fixed, eliminating the spatial autocorrelation of variables [22, 27–29]. Therefore, the GWR model is more suitable for analyzing the spatial change of the distances in different subway stations.

3. Study Areas and Data

3.1. Study Areas. This study focuses on Chengdu, the capital of Sichuan Province, a high-tech industrial base, a commercial logistics center, and a comprehensive transportation hub in western China. The whole city consists of 12 municipal districts, three counties, and five county-level cities, with 14,335 km². The study areas are shown in Figure 1. The main urban areas of Chengdu are mainly composed of 12 municipal districts, namely Chenghua, Jinniu, Jinjiang,
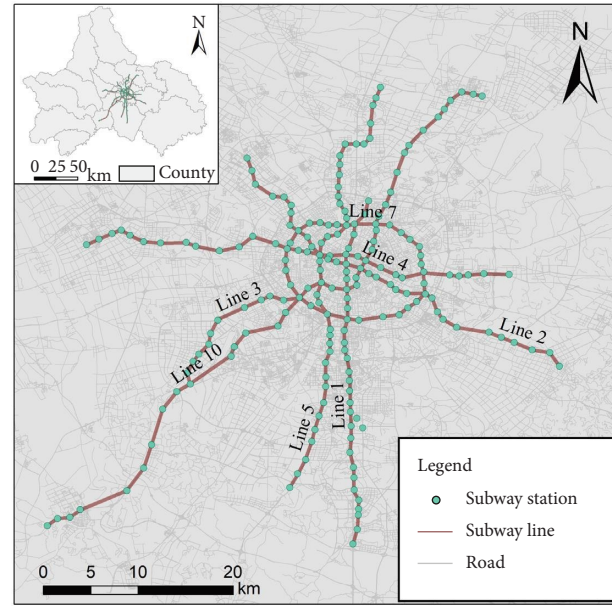

Figure 1: Study areas and the subway system in Chengdu.

Longquanyi, Pidu, Qingbaijiang, Qingyang, Shuangliu, Wenjiang, Wuhou, Xindu, and Xinjin. At the end of 2019, the resident population of Chengdu has reached 16.581 million (https://gk.chengdu.gov.cn/govInfo/detail.action?id=2576335&tn=2). With the vast population, rail transit in Chengdu has to develop rapidly to relieve the travel pressure of the metropolis. Since September 27, 2010, Chengdu metro line 1 has been put into operation. By the end of 2020, it has a relatively developed rail transit network. There are seven lines and 202 subway stations in operation, with a total length of about 518 km and an average daily passenger flow of 3.75 million. At the same time, to bring a better living and traveling environment to residents, the Chengdu rail transit group is now vigorously developing TOD projects with high density, multifunction, pedestrian-friendly environments, and high quality (https://www.chengdutod.com/#home1).

3.2. Mobile Phone Positioning Data. Mobile phone positioning data are obtained from Jike (https://www.isjike.com/). There are 4.994 million users in Chengdu, accounting for 30.12% of the 16.581 million permanent residents at the end of 2019. The average number of active users per day is 2.228 million. The users used in this paper are 1,210,252 users randomly selected from the total users. There are about 301,420,967 records, which are the continuous trajectory of sample users from October 15 to November 15, 2020. The data for this time were chosen because the weather during this period was relatively mild and more suitable for travel, and the results of the study would be more representative. Mobile phone positioning data are divided into two parts. The first part is scene data with POI records, with a total of 88,940,033 records. The main fields of scene data include user id, arrival time, longitude, latitude, departure time, scene classification, and the exact name of the POI. Table 1 is an example of scene data. The data is generated by the interaction between the

TABLE 1: Example of scene data.

User id	Arrival time	Longitude	Latitude	Departure time	Classification	POI name
3d5h###	2020/11/06 16:46:40	104. ###	30. ###	2020/11/06 16:49:16	Subway station	Dongpo road station
3d5h###	2020/11/06 16:49:12	104. ###	30. ###	2020/11/06 16:51:12	Subway station	Cultural palace station
3d5h###	2020/11/06 16:51:28	104. ###	30. ###	2020/11/06 16:54:51	Subway station	Jinsha museum station
3d5h###	2020/11/06 16:55:31	104. ###	30. ###	2020/11/06 16:59:54	Science education culture	Jinsha site museum

is used to replace more detailed information for privacy reasons.

software development kit inside the mobile phone and the POI scene through wireless fidelity, Bluetooth, and near field communication and calibrated by innovative "intelligent scene recognition" technology (https://www.cdstats.chengdu.gov.cn/htm/detail_180293.html). The identification standard is to require identification at the front gate of the subway, and the farthest 30 meters is the identification range. The shops are identified based on the entrance of the shops, and the range within 5 meters of the entrance is the effective identification range. According to the "intelligent scene recognition" technology, when a user enters or tends to enter a POI scene (such as the subway station and shop), the entry records are generated, and when a user leaves the scene, which records the time, latitude, and longitude, POI name, and other user information at this moment. Therefore, the user's stay time in a scene can be calculated. The second part is the timing report data, with a total of 212,480,934 records. The main fields include user id, time, longitude, and latitude. An example of timing report data is shown in Table 2. The generation principle of timing report data is that a report record is generated every half hour when the mobile phone continuously lights up.

Mobile phone positioning data also contain information of each user, such as age, gender, education level, and income level, which is mainly judged by a combination of integrating real samples with Jike company's in-depth partners, users' APP online usage characteristics and offline visiting behaviors. The specific sources of user characteristics are as follows: one of them is from information such as the APP list. By analyzing the APP list installed on mobile devices and the usage as well as the reference feature labels, user characteristics are analyzed and judged. For example, in the inference of gender, the typical applications of the APP are the great aunt, male health care, and male private doctor. The second comes from the type of mobile phone information pushed, the frequency of pushing, and so on. The third comes from the location information such as the user's residential address and office address resolved based on the latitude and longitude of the user's activities to identify the income level. The fourth comes from external data sources such as UnionPay and operators, as well as public data on the Internet.

3.3. Mobile Phone Positioning Data Processing. Given that the mobile phone positioning data only records the movement trajectory of the user in continuous time, it is impossible to know the actual origin and destination of the subway trips. Therefore, the trips from the origin to the subway station and leaving the station to the destination need to extract according to some principles. In this study,

TABLE 2: Example of timing report data.

User id	Time	Longitude	Latitude
3d5h###	2020/11/06 19:56:07	104. ###	30. ###
3d5h###	2020/11/06 20:50:16	104. ###	27. ###
3d5h###	2020/11/06 22:05:57	104. ###	30. ###
3d5h###	2020/11/06 22:39:51	104. ###	30. ###

is used to replace more detailed information for privacy reasons.

a complete subway trip is defined as arriving at a subway station from the origin through a certain mode of transportation (including walking, cycling, self-driving, and so on), taking the subway through at least two subway stations, or leaving the station to reach the final destination through another mode of transportation. It is important to note that because of data quality limitations, we are not able to identify the mode of transportation by which users arrive or leave the subway station.

Next, trip extraction and screening from Steps 1 to 3 in Figure 2 are introduced. For a similar trip extraction procedure, readers can refer to Wang et al. [29] 's paper.

(i) Step 1: Clean the timing report and scene data. First of all, some scene data cannot identify detailed POI, so they are recorded as timing report data by mistake. For this reason, the timing report data whose time interval with the previous record is less than 30 minutes are deleted. Second, when the user approaches some subway stations built on the ground, scene data may be generated, but users do not actually enter the station and take the subway. Therefore, when the distance between two adjacent recorded subway stations is more than 5 km, or the time interval is more than 10 minutes, it is considered that the subway is not used to reach the adjacent stations, so these subway records are deleted.

(ii) Step 2: Trip extraction under space-time constraints. In this step, the origin and destination of the access and egress trip are judged according to the moving and staying state of the continuous track of the scene and the timing report data. Considering the data structure of this study (the principle of generating data every 30 minutes by timing report data), and referring to the related literature on travel discrimination based on mobile phone cellular network data [30, 31], this study assumes that users stay at a position for 30 minutes, and the position is regarded as a staying point (origin or destination of the trip). For scene data, when the time difference between the user's arrival and departure time in

a POI is more than 30 minutes, it is marked as a staying point. For timing report data, given that these data only record the time, longitude, and latitude of every half hour when the screen is continuously illuminated but do not know the staying time of the user in this position. Therefore, a timing report record is assumed that the user stays in the position for half an hour. For the sake of travel safety, people rarely use their mobile phones for 30 minutes continuously under a bright screen when they move continuously by walking or cycling. In the case of taking the subway, the records generated during the subway trip are mainly in the ABA and AAB forms (A is the subway station, and B is the timing report record). If the time interval between A and B is less than 30 minutes, B is regarded as a nonstay point and excluded. Then, for the records marked as staying points, the trips from the staying point (origin) to the subway station and from the subway station to the staying point (destination) are extracted as access and egress trips.

(iii) Step 3: Trip screening. After extracting subway trips in step 2, the abnormal trips must be deleted according to the following rules. First, the trips whose origin or destination does not belong to Chengdu are eliminated. Second, the trips that are not in the subway operation time (00 : 00–6 : 00) and the time difference between the origin (or destination) and the station of more than 12 hours are deleted. Third, the outliers of the distance of access and egress trips are eliminated by using the three times quantile of the box diagram.

According to the previous three steps, the number of users extracted from the original data is 166,913, and the access and egress trips are 840, 312 and 763, 086, respectively. These trips are used as follow-up analysis.

3.4. Variable Description. According to the relevant literature and available data [7, 12, 13, 19, 20], we have selected two categories of independent variables that can explain the dependent variables, namely the built environment and user characteristics. Their descriptive statistics are shown in Table 3. The built environment is the statistical value of various variables within the 1 km (approximately equal to the average distance of all the extracted trips in Section 3.4) buffer zone of the subway station, mainly including land use characteristics, traffic-related facilities, and other variables. The land-use variables are calculated using POI data. POI data are collected from Amap (also known as Gaode Map) through the application program interface (https://www.amap.com/). The total number of POI is 416,459. Each POI record usually contains the POI name, address, scene classification, longitude, and latitude of the specific location. Some scenes are too few and have been deleted. Finally, according to the scene classification, POI is mainly divided into 13 categories for research. Parking lot data are extracted from POI. Road network
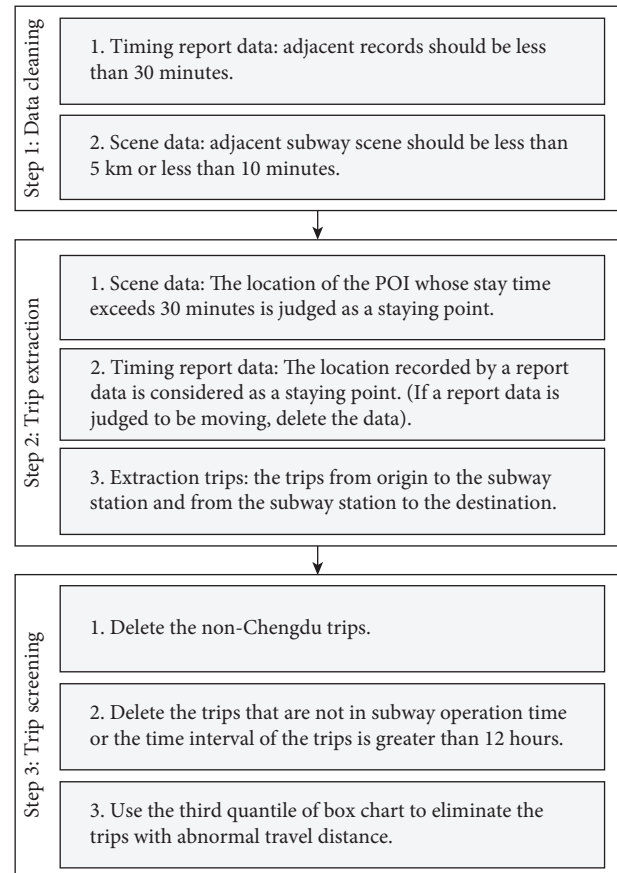

FIGURE 2: The extraction process of access and egress trips.

data are obtained from OpenStreetMap (OSM) (https://www.openstreetmap.org/). The bus stop data are obtained from the Chengdu public platform, including 10,228 bus stops. Bus stop data record each station's station name, longitude, latitude, line number, and line direction (https://www.cddata.gov.cn/oportal/index). Population data are obtained from WorldPop, counted at the grid level of 100×100 m in 2020 (https://www.worldpop.org/). This study takes Tianfu Square in Chengdu as the center and calculates the distance from each station to the city center. Finally, user characteristics are the proportion of different users in each station according to access and egress trips. In order to reduce repeated displays, the statistical values of user characteristic variables in the access trips are only presented in Table 3.

4. Method

In this study, the spatial regression model GWR was used to explore the spatial variation relationship between the 85th percentile access and egress distance and travel-related variables and the above selected built environment variables. First, the ordinary least squares (OLS) regression model was used to explore the relationship between explanatory variables and the distances. Then, to analyze the spatial change of their relationship, we calculated Moran's I to test the existence of spatial correlation of variables.

TABLE 3: Description and statistics of variables.

Variables	Description	Mean	SD
Built environment (within the 1 km buffer zone)			
Hotel	The proportion of hotel services (chain hotels, service apartments, and so on)	94.089	169.811
Sport	The proportion of sports facilities (gym center, playground, and so on)	64.470	60.775
Public	The proportion of public facilities and services (bridges, intersections, and so on)	16.436	16.698
Company	The proportion of companies (enterprise, factory, and so on)	16.436	16.698
Residence	The proportion of residences (a residence is a place used as a home or dwelling, where people reside, mainly including commercial residential buildings, and residential communities)	120.267	147.511
Life	The proportion of life services (life refers to the places or facilities that provide life services to ensure the normal operation of people's daily activities such as express delivery, laundry, mainly including post offices, travel agencies, logistics and express delivery, and laundries)	158.386	148.263
Healthcare	The proportion of medical services (general and special hospitals, clinics, and so on)	116.297	115.913
Government	The proportion of government (administrative agency, civil service, welfare institution, and so on)	77.797	97.841
Education	The proportion of science and education culture (college, university, high, middle and primary school, and so on)	118.139	127.856
Shopping	The proportion of shopping malls (retail, shopping mall, convenience store, and so on)	19.842	16.792
Finance	The proportion of financial services (banks, insurance companies, financial corporations, and so on)	58.535	68.825
Tourist	The proportion of tourist attractions (parks, museums, historical sites, and so on)	6.426	9.119
Catering	The proportion of catering services (Chinese and foreign restaurants, fast food, coffee house, and so on)	280.460	253.100
POIMix	Shannon entropy of all POI categories	3.164	0.248
Road density	Length of all roads (km)	37.065	13.775
Bus station	Number of bus stations	33.480	23.420
Parking lots	Number of parking lots	160.020	161.447
Population	The number of people	42868.607	39434.025
Distance	Euclidean distance between each station and city center (km)	10.830	7.529
User characteristics			
Age 26–35	The proportion of users aged 26–35 years old (The reference variable of age is 16–25 years old)	0.428	0.081
Age 36–45	The proportion of users aged 36–45 years old	0.053	0.017
Over 46	The proportion of users over 46 years old	0.019	0.008
Male	The proportion of male	0.552	0.067
Median education	The proportion of college and undergraduate	0.364	0.029
High education	The proportion of graduate students or above	0.125	0.030
Middle income	The proportion of users with monthly income of 7,000 to 15,000 yuan	0.319	0.042
High income	The proportion of users with a monthly income greater than 15,000 yuan	0.124	0.024

Finally, the GWR model is used to quantitatively analyze the local relationship between access/egress distances and explanatory variables. The following sections briefly introduce the principle and calculation of the model.

4.1. Spatial Autocorrelation Test.

Before using the spatial regression model, the spatial autocorrelation of variables should be tested. Moran's I is a widely used global spatial autocorrelation measure. Moran's I can be expressed as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(y_i - \overline{y}\right)\left(y_j - \overline{y}\right)}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}, \quad (1)$$

where n is the number of subway stations, $\overline{y}$ is the average value of y, and w_{ij} is the spatial weight between station i and station j. The value of global Moran's I is usually between −1 and 1. When Moran's I is positive, the variables have positive spatial autocorrelation; if Moran's I is negative, the variable has negative spatial autocorrelation; if Moran's I is 0, it means that the variable is random to some extent.

The Z-value of Moran's I can be calculated by the following equation:

$$Z = \frac{I - E[I]}{SD[I]}, \quad (2)$$

where $E[I]$ and $SD[I]$ are the expectation and standard deviation of the global Moran's I, respectively. A positive Z-value indicates that the variable has more spatial aggregation, while a negative Z-value indicates that the variable has more spatial dispersion. Generally, the significance of Moran's I is estimated by pseudo P value. If the pseudo P value is less than 0.05, the global Moran's I is statistically significant at the confidence level of 95%, which means that

the variable is spatially correlated. On the other hand, if the pseudo P value is greater than or equal to 0.05, it means that the variable is likely to be randomly and independently distributed in space.

4.2. Geographically Weighted Regression Model. Geographically weighted regression (GWR) model is an extended form of the OLS model, which is used to model spatial variation. Compared with the general OLS and GWR allows the coefficients of explanatory variables to change in space. In order to better understand GWR, this paper first explains the OLS model.

Assuming that the space surface is uniform, the traditional global OLS model is often used to explore the relationship between dependent and independent variables. The model formula is as follows:

$$y_i = \beta_0 + \sum_{k=1}^{m} \beta_k x_{ik} + \varepsilon_i, \tag{3}$$

where y_i is the distance of access or egress trips at the station i, β_0 is the intercept term, β_k is the estimation coefficient of the k th independent variable, x_{ik} is the environmental variable, and ε_i is the model error at station i.

Considering the global nature of the OLS model, the estimated regression coefficients are the same and constant in the whole study area. However, since spatial data are usually heterogeneous and highly dependent on local regional characteristics, the region cannot be completely homogeneous. As a basic extension of the OLS model, the geographical location factor is added to the regression parameters to quantify the spatial effect, and the neighborhood relationship is simulated by calibrating the model with local coefficients. The formula is

$$y_i = \beta_{i0}(u_i, v_i) + \sum_{k=1}^{m} \beta_{ik}(u_i, v_i) x_{ik} + \varepsilon_i, \tag{4}$$

where for the station i, (u_i, v_i) represents the geographical coordinates of the subway station, $\beta_{i0}(u_i, v_i)$ is the intercept term, $\beta_{ik}(u_i, v_i)$ is the regression coefficient associated with the k th environmental variable, x_{ik} is the k th explanatory variable.

According to the first law of geography [32], the interaction between adjacent stations is more significant than that between distant stations. This location uses the latitude and longitude location of each subway station. Therefore, constructing a spatial weight matrix is necessary to estimate the value of $\beta_{ik}(u_i, v_i)$, which can be calculated as follows:

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

where for the station i, $\widehat{\beta}_i(u_i, v_i) = (\widehat{\beta}_{i0}, \widehat{\beta}_{i1}, \cdots, \widehat{\beta}_{ik}, \cdots \widehat{\beta}_{im})^T$ is the estimated coefficient of the independent variable k, X, and Y are the vector-matrix of the independent and dependent variables, respectively, and W_i is the spatial weighting matrix, which can be expressed as follows:

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix},$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \tag{6}$$

$$W_i = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix},$$

where w_{in} represents the spatial weight value between the station i and others. In this study, the commonly used adaptive bi-square kernel is used to calculate the spatial weighting matrix, and the adaptive distance decay simulates the spatial effect of the surrounding station in the bandwidth range. It is worth noting that the bandwidth selection is also important, because it will greatly affect the coefficient estimation. This paper chooses the bandwidth selection method for the golden section search. The corrected Akaike information criterion (AIC_C) is used to evaluate the fitness to obtain the best bandwidth.

The coefficients in the OLS model are constant, and the difference between the OLS model and the GWR model is that its coefficients vary with geographical location. Therefore, we use four regression models to discuss the relationship between subway access/egress distances and explanatory variables. These four models are represented as OLS_Access, OLS_Egress, GWR_Access, and GWR_Egress. In order to further evaluate the spatial nonstationarity of the coefficients, we use AIC_C and adjusted R^2 to measure the model performance of OLS and GWR. Lower AIC_C and higher adjusted R^2 values show better model fitting.

5. Results and Discussions

5.1. Descriptive Statistics and Analysis

5.1.1. Origin and Destination of Trips. Figure 3 shows the kernel density distribution of the origin and destination of subway trips, respectively. The display rule of the distribution is quantile. The deeper the red color, the higher the number of trip generations or destinations. This figure indicates that the trip generation or attraction is mainly distributed within the loop line (Line 7), and the intensity outside the loop line is relatively low. The origin and destination of trips mainly fall near the subway station. That is, the farther away from the subway station, the less the trip. In addition, the figure also shows that there are different intensities around the same subway station, which may be related to land use.
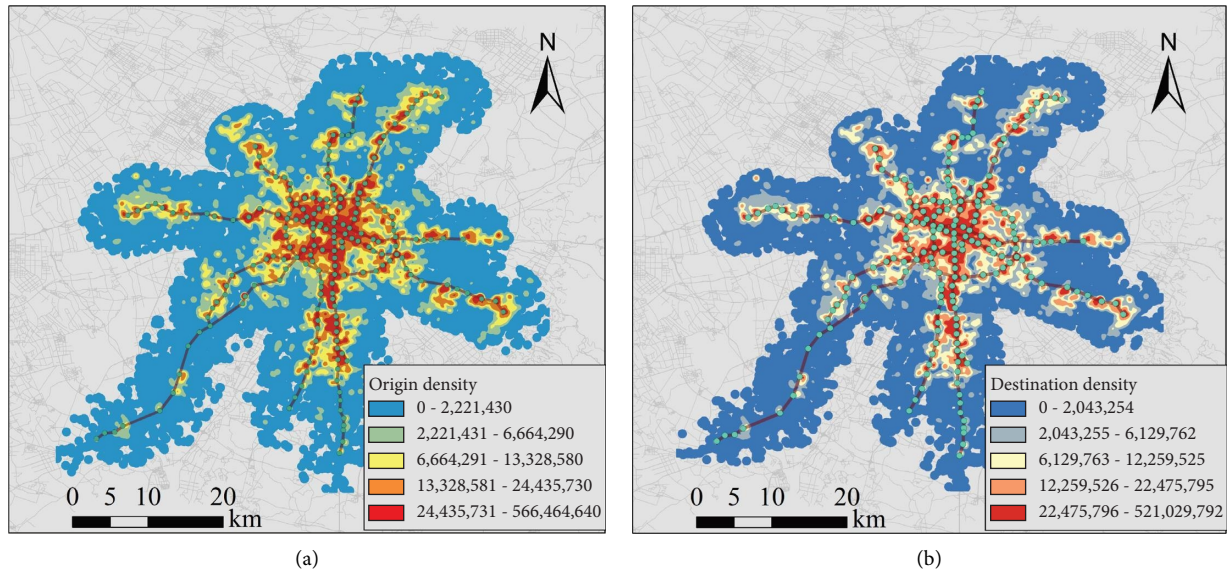
FIGURE 3: Spatial distribution of the origin and destination. (a) Kernal density of access trip and (b) kernal density of egress trip.

5.1.2. Analysis of the Access and Egress Distances.

Table 4 shows the descriptive statistics of the access and egress distances. The distances are calculated as the Euclidean distances between the origin or destination with the subway stations. The table indicates the average access distance is 1.059 km, and the 85th percentile distance is 2.031 km. However, the egress distance is lower than the access distance, with an average distance of 0.998 km and the 85th percentile distance of 1.930 km. In order to obtain the distance of the total trips, the access and egress trips of each station are added. Results show the average distance of the total trips is 1.030 km, higher than the walking distance (0.8 km) often used in practice [5, 6], and lower than the bicycle distance (2 km) calculated based on GPS trajectory in Shanghai [19]. The 85th percentile distance is often used as the threshold for people willing to walk or bicycle to reach the subway service [7, 10, 19]. The 85th percentile distance of the total trips is 1.983 km, which indicates that the subway in Chengdu provides services for most people within this distance without considering the feeder mode.

Figure 4 is a histogram of access and egress distances. The left side of the vertical axis is the frequency of trips, and the right side is the cumulative proportion of trips corresponding to the solid line. The figure shows that both the access and the egress distances are in the form of decay, and the farther the distances are, the fewer the trips are. It can also be found that the commonly used 0.8 km only accounts for 59% of trips, as shown by the blue dotted line, which means that a large proportion of trips beyond 0.8 km are still not covered, which may lead to the underestimation of the service coverage of subway stations.

5.1.3. Analysis of Different Users' Access and Egress Distances.

Table 5 shows the descriptive statistics of trips for different users, including age, gender, educational background, and income. Both access and egress distances show that the average distances of different types of users are various. The average distance between males (1.126 km) is higher than that of females (0.983 km), consistent with most previous studies. Whether walking or bicycling, the distance between males is higher than that of females [26, 33]. In terms of age, the distance of people aged 16–25 is the shortest. With the increase in age, the distance is longer. The higher the educational background, the farther the distance. It is an interesting discovery, which may be related to their travel purpose. Compared with other income groups, middle-income people have the longest distance. Because the subway provides an affordable travel choice for people, middle-income people are more likely to choose the subway. The egress distances are smaller, but different users show a similar pattern compared with the access distances.

The Kruskal–Wallis test is used to compare whether there are statistical differences in the travel distance of different users. The Kruskal–Wallis test is suitable for comparing grouping variables with two or more levels (if there are two levels, equivalent to the Mann–Whitney U test). In addition, it is possible to judge whether the mean values of several populations are equal or not without making any assumptions. The results show that the null hypothesis that there is no difference between samples can be rejected ($P < 0.001$), which means that the distances of different user types are significantly various.

5.1.4. Spatial Distribution of Access and Egress Distance.

Different subway stations may have different access and egress distance thresholds, considering the spatial heterogeneity. We use the standardized circle size of the subway station to represent the threshold of the 85th percentile distance shown in Figure 5. It can be found that the access and egress distances between stations vary obviously. Generally speaking, the stations with smaller distances are mainly distributed in the central city, while the stations far

TABLE 4: Descriptive statistics of distance (unit: km).

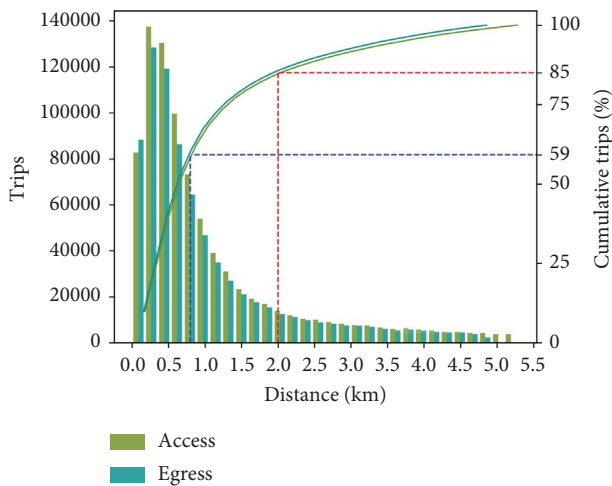| Type	Mean	SD	25th	50th	75th	85th	Max
Access distance	1.059	1.101	0.341	0.645	1.305	2.031	5.282
Egress distance	0.998	1.027	0.318	0.616	1.256	1.930	4.858
Distance of the total trips	1.030	1.067	0.330	0.631	1.282	1.983	5.282

FIGURE 4: Histogram of access and egress distances.

away from the central city have a longer distance. This result is logical because the density of subway stations is high in the central city, and people can reach the nearest subway station by traveling a short distance. However, in the outlying area of the central city, the density of the subway and other transportation facilities is low, and even if it is far from the subway station, it has to travel long distances to reach the subway station.

5.2. Model Results

5.2.1. Results of Ordinary Least Squares Regression Model. In this study, the OLS model is used to model the 85th percentile access distance to determine which factors may affect the catchment area of the subway. The Pearson correlation coefficient and variance inflation factor (VIF) are used to eliminate the collinearity between variables.

First, if the correlation coefficient between the variables is greater than 0.7, it is considered that there is a high correlation between the two variables, and it is deleted. Then, we calculate the VIF of other variables, which shows that there is no significant collinearity among variables (VIF < 5). In order to select variables better, the backward stepwise regression method is used, which allows a series of regression models to be established by deleting and adding independent variables and evaluating which variables should be kept. The results of OLS model are shown in Table 6. The adjusted R^2 of the model is 0.510 and 0.505, respectively, indicating that the independent variables in this study explain at least 50.5% of the distance variation. If the P value of a variable is less than 0.05, the null hypothesis that there is no relationship between variables can be rejected. It can be

found that hotel, residence, life, finance, road density, and mixed land are negatively correlated with distance, while other variables are positively correlated with distance.

5.2.2. Analysis Results of Global Moran's I. In order to test whether the GWR model is suitable for exploring the relationship between subway access/egress distance and explanatory variables, this study first makes a global Moran's I through ArcGIS to check whether the selected variables have spatial autocorrelation. Table 7 shows the global Moran's I result. This test measures the spatial autocorrelation of a specific element according to its position and numerical value. The null hypothesis is that there is no spatial correlation. According to the result, this hypothesis is rejected. P values of all variables are significant, showing a strong spatial correlation. The Z-value is greater than 0, indicating that each variable presents a spatial aggregation pattern. The above evidence shows that the global OLS model cannot effectively analyze the relationship between subway travel distance and interpretation. Therefore, it is advisable to use the GWR model to explore the spatial heterogeneity of data.

5.2.3. Results Analysis of the GWR Model. In order to compare the results of the global regression model, we use all variables in the OLS model to discuss the influence of spatial variation of explanatory variables. GWR 4.0 software is used to model GWR. Tables 8 and 9 show the regression results of access and egress distance in GWR, respectively. These two tables show descriptive statistics of regression coefficients: mean, standard deviation, minimum, lower quartile, upper quartile, maximum, and range. The R^2, adjusted R^2, and AICc in statistics are widely used indicators to evaluate the applicability and performance of the model. Therefore, these indexes of the GWR model in Tables 8 and 9 are compared with those of the OLS model in Table 6. As the R^2 value is larger and the AICc value in GWR is smaller, the model is more suitable for observation data, which indicates that the GWR model is superior to the traditional OLS model in this case study. In addition, descriptive statistical indicators provide an overall understanding of the distribution characteristics of regression coefficients. For example, the residence has a negative impact on the distance of subway travel (average = −5.460). However, its standard deviation (SD = 2.499) shows that the regression coefficient distribution of residence is more dispersed than other variables, and more than 75% of them have a negative impact on distance. These results will help decision-makers understand the range of local coefficients between explanatory variables and distances and then help to implement targeted planning measures at different stations.

TABLE 5: Descriptive statistics of different users (unit: km).

Type	Percentage (%)	Access distance			Egress distance		
		Mean	SD	P value	Mean	SD	P value
Gender				<0.001			<0.001
Female	39.9	0.983	1.039		0.925	0.967	
Male	60.1	1.126	1.148		1.062	1.071	
Age				<0.001			<0.001
16–25	42.7	0.967	1.018		0.912	0.948	
26–35	47.3	1.153	1.171		1.083	1.093	
36–45	7.0	1.174	1.186		1.115	1.106	
Over 46	3.0	1.231	1.222		1.165	1.152	
Education[a]				<0.001			<0.001
Low	48.6	1.048	1.092		0.993	1.022	
Median	37.2	1.060	1.102		0.997	1.027	
High	14.2	1.104	1.129		1.023	1.045	
Income[b]				<0.001			<0.001
Low	54.9	1.036	1.080		0.974	1.005	
Middle	33.2	1.107	1.137		1.045	1.064	
High	11.8	1.044	1.096		0.985	1.019	

a: low education: high school and below, medium education: college and undergraduate, high education: graduate and above. b: low income: monthly income below 7,000 yuan, middle income between 7,000 and 15,000 yuan, and high income above 15,000 yuan.
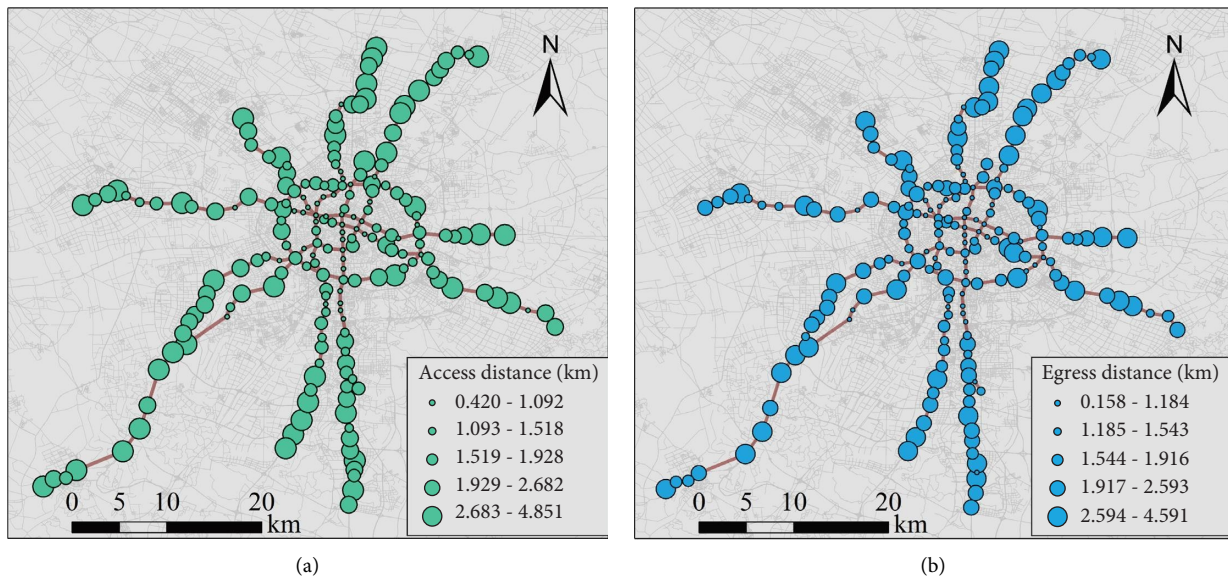

FIGURE 5: Spatial distribution of the 85th quantile distances. (a) Spatial distribution of access distance and (b) spatial distribution of egress distance.

As the estimation coefficient of each independent variable varies from station to station, Figures 6–10 show that subway stations are marked with different colors in the figure based on the value of their estimation coefficient to understand better the influence of the spatial change of independent variables. Because of the layout limitation, this paper only shows and discusses that the two GWR models have common variables. On the whole, the influence of most variables on the access and egress distances, respectively, showed similar spatial variation, with only Figure 10 showing some more significant differences.

Figure 6 shows the spatially varying effects of hotel on subway stations' access/egress distance. The figure shows that the relationship between hotel and access/egress distance is negative, but it varies from space to space. From the spatial point of view, the hotel has a smaller negative impact on the distance between the city center and the north. On the contrary, in areas far away from the city center, the increase in hotels has a greater negative impact on the service scope of subway stations. These results indicate that the increase in the proportion of hotels beside suburban subway stations shortens the distance between subway stations.

Figure 7 shows the spatially varying effects of residence on subway stations' access/egress distance. The figure shows that the proportion of residence is negatively correlated with the access/egress distance. This may be because people usually like to live around subway stations, so the increase in residential proportion reduces the distance. From the spatial

TABLE 6: Results of the ordinary least squares regression model.

Variables	OLS_Access				OLS_Egress			
	Coefficient	T-value	P value	VIF	Coefficient	T-value	P value	VIF
(Intercept)	1.463	1.280	0.201	—	0.170	0.252	0.800	—
Hotel	−2.441	−2.678	0.008**	1.190	−3.050	−3.725	<0.001***	1.202
Residence	−4.240	−3.021	0.002**	1.057	−2.921	−2.269	0.024*	1.113
Life	—	—	—	—	−2.432	−2.191	0.029*	1.306
Education	2.444	1.856	0.064*	1.153	—	—	—	—
Finance	—	—	—	—	−5.634	−2.152	0.032*	1.292
Road density	−0.020	−4.753	<0.001***	1.497	−0.017	−4.113	<0.001***	1.867
Mixed land	−0.660	−2.858	0.004**	1.352	—	—	—	—
Age 36–45	—	—	—	—	8.553	2.704	0.007**	1.572
Male	5.648	6.087	<0.001***	1.585	3.528	3.999	<0.001***	1.867
High education	3.719	2.197	0.029*	1.144	6.154	4.001	<0.001***	1.058
R^2	0.529				0.527			
Adjusted R^2	0.510				0.505			
AICC	439				395			

***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

TABLE 7: Global Moran's I result of variables.

Variables	Moran's I	Expected I	Variance	Z-value	P value
Access distance	0.474	−0.004	0.001	12.247	<0.001
Egress distance	0.409	−0.004	0.001	10.574	<0.001
Hotel	0.445	−0.004	0.001	13.309	<0.001
Residence	0.364	−0.004	0.001	9.735	<0.001
Life	0.381	−0.004	0.001	9.850	<0.001
Education	0.223	−0.004	0.001	5.951	<0.001
Finance	0.323	−0.004	0.001	8.406	<0.001
Road density	0.684	−0.004	0.001	17.601	<0.001
Mixed land	0.430	−0.004	0.001	11.318	<0.001
Age 36–45 [egress]	0.277	−0.004	0.001	7.223	<0.001
Male[access]	0.399	−0.004	0.001	10.361	<0.001
Male[egress]	0.431	−0.004	0.001	11.175	<0.001
High education[access]	0.077	−0.004	0.001	2.161	0.030
High education[egress]	0.119	−0.004	0.001	3.286	0.001

Access represents the access distance. Egress represents the egress distance.

TABLE 8: Regression results of GWR_Access.

Variables	Mean	SD	Minimum	Lower quartile	Upper quartile	Maximum
(Intercept)	2.315	2.390	−2.131	0.539	3.870	6.537
Hotel	−2.379	0.845	−5.148	−2.805	−1.705	−1.442
Residence	−5.460	2.499	−10.709	−7.243	−4.180	1.991
Education	3.378	1.678	0.250	2.195	4.642	6.283
Road density	−0.019	0.003	−0.030	−0.020	−0.016	−0.014
Mixed land	−0.695	0.414	−1.424	−0.964	−0.392	0.040
Male	4.357	1.804	1.308	2.819	5.762	7.685
High education	2.430	1.357	−0.252	1.391	3.208	6.987
R^2	0.609					
Adjusted R^2	0.548					
AIC_C	434					

point of view, the impact of residential on distance is usually greater in the northeast of the city. These results indicate that people who live in the city's northeast are usually more sensitive to the access and egress distance.

Figure 8 shows the spatially varying effects of road density on subway stations' access/egress distance. The figure shows that the road density is negatively correlated with the distance. This may be because in areas with high road density, the traffic accessibility around the subway is greater, thus shortening the access/egress distance of the subway station. The access/egress distances have shown similar results, and the influence of road density on the distance is usually greater in urban suburbs. This indicates that the increase in road density around suburban subway

TABLE 9: Regression results of GWR_Egress.

Variables	Mean	SD	Minimum	Lower quartile	Upper quartile	Maximum
(Intercept)	0.198	0.918	−1.370	−0.525	0.696	2.208
Hotel	−3.058	1.368	−8.748	−3.528	−2.181	−1.513
Residence	−4.113	1.971	−7.286	−5.436	−2.871	2.923
Life	−1.143	1.477	−4.074	−2.373	0.127	1.422
Finance	−7.370	1.492	−11.680	−8.286	−6.290	−2.808
Road	−0.012	0.005	−0.025	−0.015	−0.008	−0.006
Age 36–45	9.091	2.835	3.486	7.485	10.742	17.309
Male	3.010	1.543	−0.483	1.998	4.184	5.934
High education	5.892	1.119	2.694	5.178	6.651	8.906
R^2	0.625					
Adjusted R^2	0.547					
AIC_C	393					
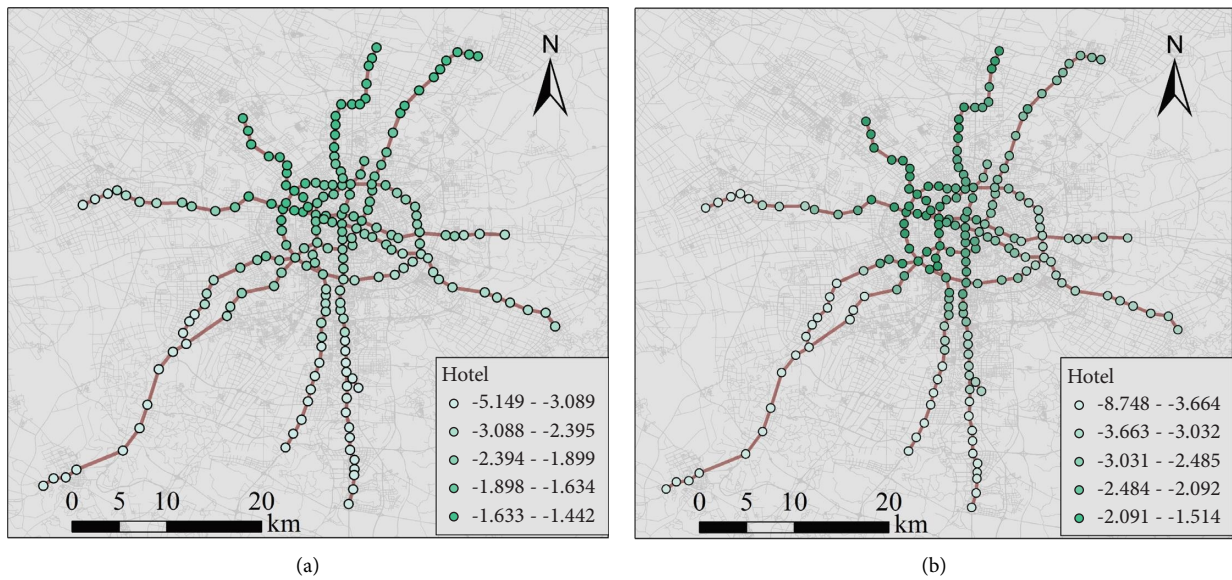
FIGURE 6: Spatially varying effects of a hotel on the access/egress distance. (a) Effects of the hotel on the access distance and (b) effects of the hotel on the egress distance.
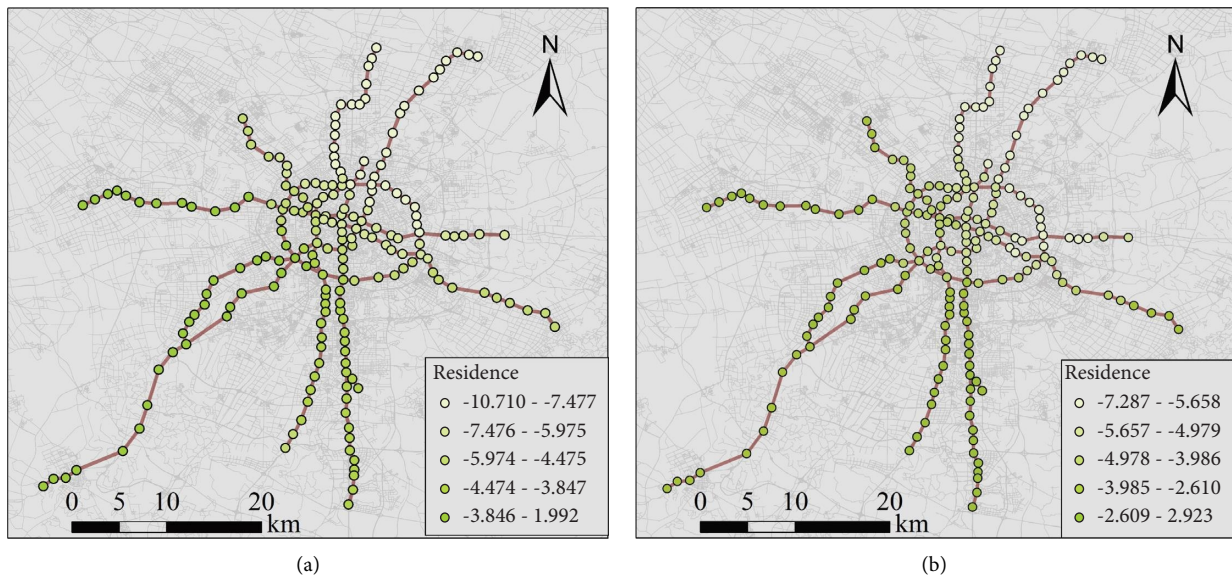

FIGURE 7: Spatially varying effects of residence on the access/egress distance. (a) Effects of residence on the access distance and (b) effects of residence on the egress distance.
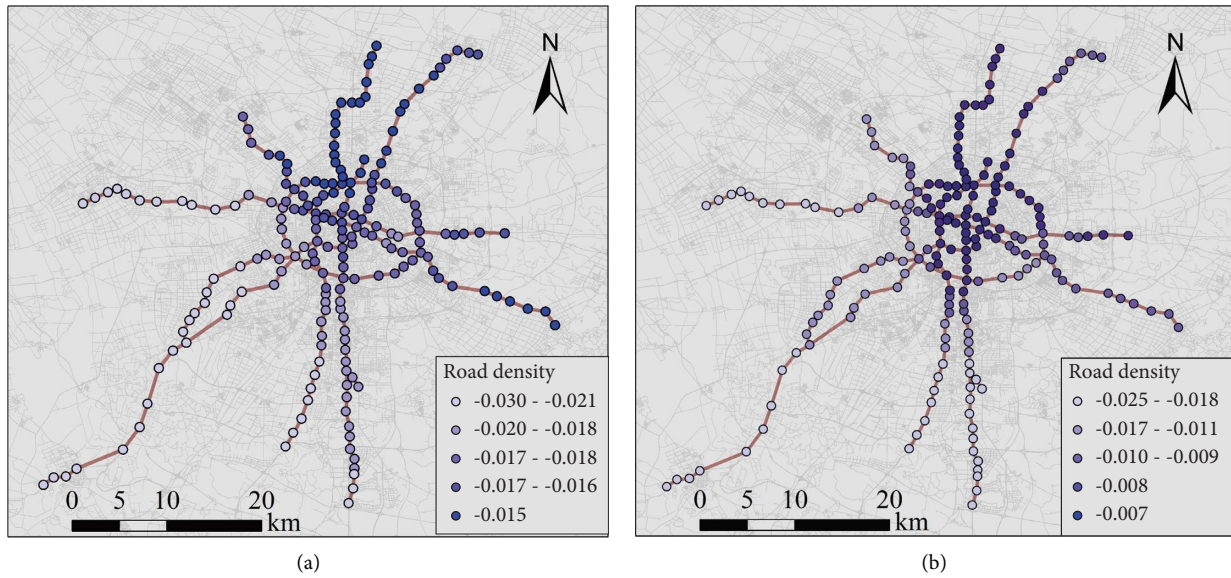
FIGURE 8: Spatially varying effects of road density on the access/egress distance. (a) Effects of road density on the access distance and (b) effects of road density on the egress distance.
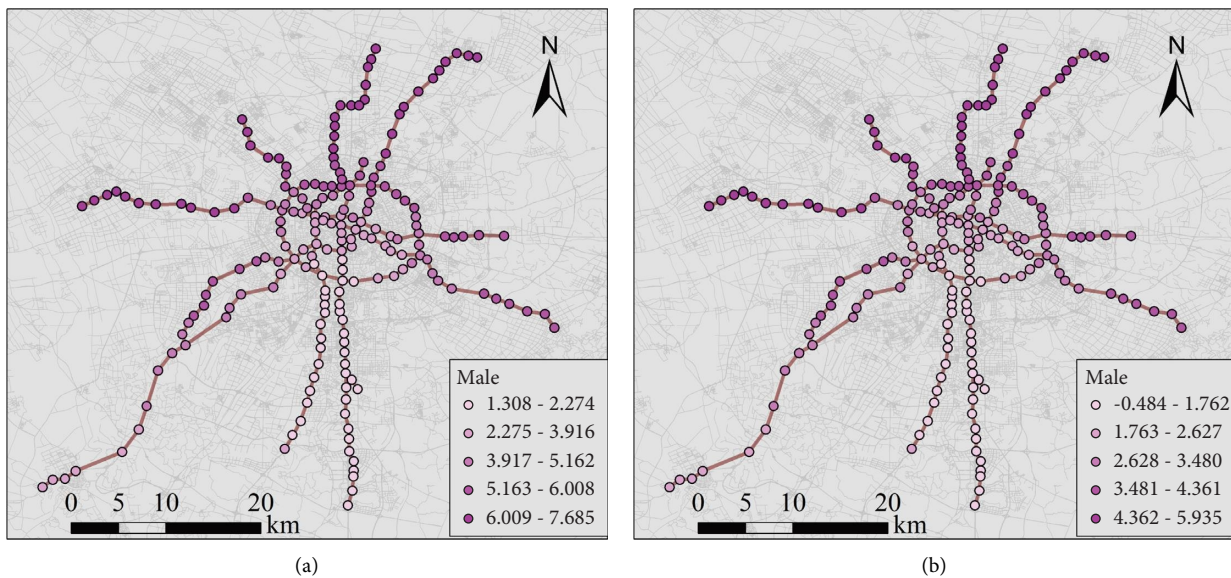

FIGURE 9: Spatially varying effects of male on the access/egress distance. (a) Effects of male on the access distance and (b) effects of male on the egress distance.

stations can affect the service coverage of subway stations more than that in downtown.

Figure 9 shows the spatially varying effects of male on subway stations' access/egress distance. The figure shows that there is a positive correlation between males and distance. In the previous descriptive statistics (Section 5.1.3), there are differences in the distance between the gender. However, their results do not reflect whether there is a statistical difference in the distance at the station level. In the GWR model, the coefficient in the north is larger, which means that in these stations in the north, males are more inclined to travel a longer distance to the subway station, thus expanding the coverage of the subway station.

Figure 10 shows the spatially varying effects of high education on subway stations' access/egress distance. The figure shows that high education is positively correlated with distance. In the previous descriptive statistics (Section 5.1.3), there are differences in the distance between education. The higher the education, the farther the distance is. The distance at the station level also shows the same result. In the GWR model, the access distance model has a larger coefficient in the north, which means that people with higher education tend to travel longer distances in these stations in the north. In addition, it can also be seen that the coefficient of egress distance is greater in the southern part compared to the access distance, perhaps because the transportation is not so
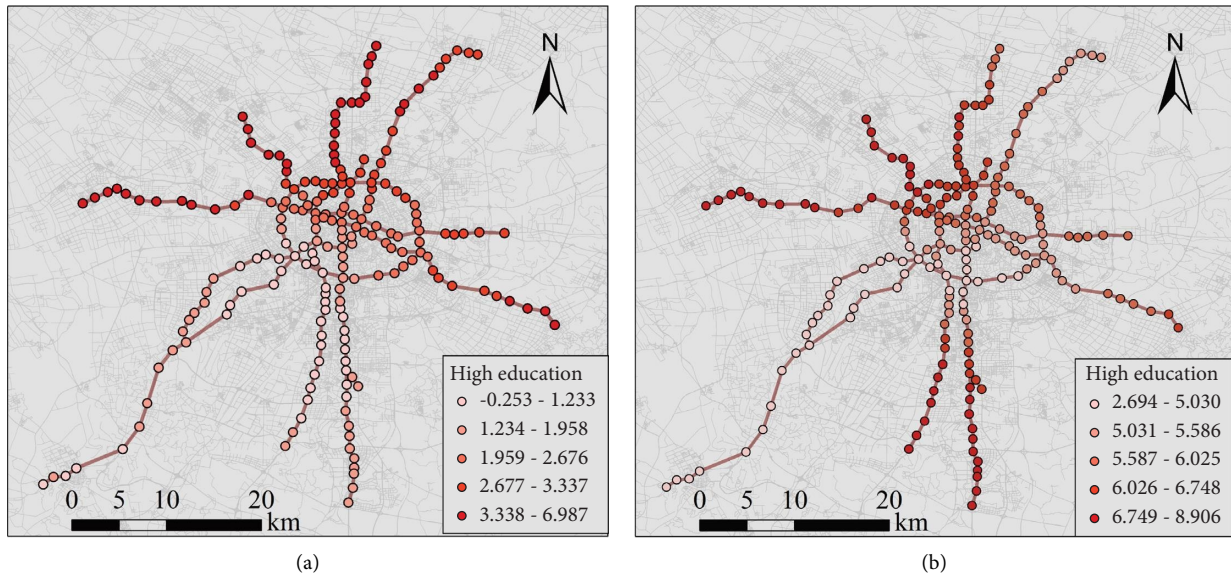
FIGURE 10: Spatially varying effects of high education on the access/egress distance. (a) Effects of high education on the access distance and (b) effects of high education on the egress distance.

convenient in the southern part of Chengdu, and the highly educated people are more likely to transfer to other transportation modes to their destinations after leaving the subway station, which is more likely to increase the egress distance.

6. Conclusion

As an essential means of transportation in big cities in China, the subway significantly influences people's travel. If the transportation facilities and living environment in the proper area of the station are improved, more people may be attracted to the subway. Therefore, this study extracts the subway access and egress trips from Chengdu's mobile phone positioning data to obtain the station's service distance. Then, the access/egress distances and the environmental variables related to the trip are calculated. In order to explore the influencing factors of access/egress distance, this study applied the GWR model to test the relationship between access/egress distances and building environment variables. Finally, the spatially varying effects of these explanatory factors are analyzed. The main results are as follows:

First of all, the access and egress distance of the subway station varies with the difference in user characteristics and the spatial location of the station. Comparing gender, age, education, and income shows that the average access/egress distance of males is longer than that of women. The higher the age and educational background, the longer the distance. Compared with other income groups, middle-income groups have a longer distance. These results strengthened some previous studies, emphasizing that different social and demographic factors show significant differences in the subway access/egress distance. The major discovery in space is that the access/egress distance varies from station to station, showing significant spatial differences. Generally

speaking, the distance from the city center is shorter, and the distance in the suburbs is usually longer.

Secondly, this study takes the 85th percentile distance as the key indicator of the service coverage of the station. The OLS models are established to find out the key factors affecting the distance: hotel, residence, life, finance, road-density, and mixed land are negatively correlated with the access/egress distance of the subway station. In contrast, education, 36–45 years old, male, and high education are positively correlated.

Finally, the GWR model is used to analyze the spatial variation relationship between the distance and various factors. The goodness of fit shows that the GWR model has better performance than the OLS model for the same variable, AICc is significantly smaller, and the adjusted R^2 is higher. The influence of explanatory variables on distance also varies from space to space.

The above results in this study are obtained based on the existing stations and the historical behavior of users in Chengdu and can provide some useful information for future public transport planning or other urban constructions. In the construction of subway networks, we should carefully study the characteristics of the built environment of candidate stations in different spatial locations, and fully consider the behavioral desires of users to rationalize the planning and optimize the layout of the surroundings of different subway stations.

There are also some shortcomings in this study. First, although this study opens the relationships between the access/egress distances of the subway station and the built environment and user characteristics, travel to the subway station is a complex behavior, depending on various factors besides these factors. For example, travel purpose greatly influences access/egress distances, but these data cannot be obtained. In addition, the availability of other types of public transport (such as bike-sharing and ride-sharing) is also an

essential factor affecting the access/egress distance, as these data are not available and we are not able to quantify what effect they have on the distances, which provides a direction for future research.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Gan, M. Yang, T. Feng, and H. J. P. Timmermans, "Examining the relationship between built environment and metro ridership at station-to-station level," *Transportation Research Part D: Transport and Environment*, vol. 82, Article ID 102332, 2020.

[2] S. Li, L. Dijiang, L. Xiaoping et al., "The varying patterns of rail transit ridership and their relationships with fine-scale built environment factors: big data analytics from Guangzhou," *Cities*, vol. 99, Article ID 102580, 2020.

[3] J. Zhao, W. Deng, Y. Song, and Y. Zhu, "What influences Metro station ridership in China? Insights from Nanjing," *Cities*, vol. 35, pp. 114–124, 2013.

[4] X. Xu, Y. Lu, Y. Wang, J. Li, and H. Zhang, "Improving service quality of metro systems—a case study in the beijing metro," *IEEE Access*, vol. 8, Article ID 12573, 2020.

[5] R. Daniels and C. Mulley, "Explaining walking distance to public transport: the dominance of public transport supply," *Journal of Transport and Land Use*, vol. 6, no. 2, p. 5, 2013.

[6] B. Canepa, "Bursting the bubble: determining the transit-oriented development's walkable limits," *Transportation Research Record*, vol. 1992, no. 1, pp. 28–34, 1992.

[7] A. El-Geneidy, M. Grimsrud, R. Wasfi, P. Tétreault, and J. Surprenant-Legault, "New evidence on walking distances to transit stops: identifying redundancies and gaps using variable service areas," *Transportation*, vol. 41, no. 1, pp. 193–210, 2013.

[8] Z.-j. Wang, F. Chen, and T.-k. Xu, "Interchange between metro and other modes: access distance and catchment area," *Journal of Urban Planning and Development*, vol. 142, no. 4, 2016.

[9] Y. L. Xi, S. Saxe, and E. Miller, "Accessing the subway in Toronto, Canada: access mode and catchment areas," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2543, no. 1, pp. 52–61, 2016.

[10] T. Zuo, H. Wei, and A. Rohne, "Determining transit service coverage by non-motorized accessibility to transit: case study of applying GPS data in Cincinnati metropolitan area," *Journal of Transport Geography*, vol. 67, pp. 1–11, 2018.

[11] J. Lee, K. Choi, and Y. Leem, "Bicycle-basedtransit-oriented development as an alternative to overcome the criticisms of the conventional transit-oriented development," *International Journal of Sustainable Transportation*, vol. 10, no. 10, pp. 975–984, 2015.

[12] J. Wang and X. Cao, "Exploring built environment correlates of walking distance of transit egress in the Twin Cities," *Journal of Transport Geography*, vol. 64, pp. 132–138, 2017.

[13] T. Tao, J. Wang, and X. Cao, "Exploring the non-linear associations between spatial attributes and walking distance to

transit," *Journal of Transport Geography*, vol. 82, Article ID 102560, 2020.

[14] J. He, R. Zhang, X. Huang, and G. Xi, "Walking access distance of metro passengers and relationship with demographic characteristics: a case study of nanjing metro," *Chinese Geographical Science*, vol. 28, no. 4, pp. 612–623, 2018.

[15] R. Rastogi and K. V. Krishna Rao, "Travel characteristics of commuters accessing transit: case study," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 684–694, 2003.

[16] H. Pan, Q. Shen, and S. Xue, "Intermodal transfer between bicycles and rail transit in Shanghai, China," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2144, no. 1, pp. 181–188, 2010.

[17] H. H. Hochmair, "Assessment of bicycle service areas around transit stations," *International Journal of Sustainable Transportation*, vol. 9, no. 1, pp. 15–29, 2014.

[18] A. Weinstein Agrawal, M. Schlossberg, and K. Irvin, "How far, by which route and why? A spatial analysis of pedestrian preference," *Journal of Urban Design*, vol. 13, no. 1, pp. 81–98, 2008.

[19] W. Li, S. Chen, J. Dong, and J. Wu, "Exploring the spatial variations of transfer distances between dockless bike-sharing systems and metros," *Journal of Transport Geography*, vol. 92, Article ID 103032, 2021.

[20] D. Lin, Y. Zhang, R. Zhu, and L. Meng, "The analysis of catchment areas of metro stations using trajectory data generated by dockless shared bikes," *Sustainable Cities and Society*, vol. 49, Article ID 101598, 2019.

[21] K. Gao, Y. Yang, A. Li, and X. Qu, "Spatial heterogeneity in distance decay of using bike sharing: an empirical large-scale analysis in Shanghai," *Transportation Research Part D: Transport and Environment*, vol. 94, Article ID 102814, 2021.

[22] H. Yang, Y. Zhang, L. Zhong, X. Zhang, and Z. Ling, "Exploring spatial variation of bike sharing trip production and attraction: a study based on Chicago's Divvy system," *Applied Geography*, vol. 115, Article ID 102130, 2020.

[23] L. Yang, K. W. Chau, W. Y. Szeto, X. Cui, and X. Wang, "Accessibility to transit, by transit, and property prices: spatially varying relationships," *Transportation Research Part D: Transport and Environment*, vol. 85, Article ID 102387, 2020.

[24] J. Zhao and W. Deng, "Relationship of walk access distance to rapid rail transit stations with personal characteristics and station context," *Journal of Urban Planning and Development*, vol. 139, no. 4, pp. 311–321, 2013.

[25] B. W. Alshalalfah and A. S. Shalaby, "Case study: relationship of walk access distance to transit with service, travel, and personal characteristics," *Journal of Urban Planning and Development*, vol. 133, no. 2, pp. 114–118, 2007.

[26] X. Ma, Y. Jin, and M. He, "Measuring bikeshare access/egress transferring distance and catchment area around metro stations from smartcard data," *Information*, vol. 9, pp. 289–311, 2018.

[27] L. Yang, X. Chu, Z. Gou, H. Yang, Y. Lu, and W. Huang, "Accessibility and proximity effects of bus rapid transit on housing prices: heterogeneity across price quantiles and space," *Journal of Transport Geography*, vol. 88, Article ID 102850, 2020.

[28] R. Pan, H. Yang, K. Xie, and Y. Wen, "Exploring the equity of traditional and ride-hailing taxi services during peak hours," *Transportation Research Record*, vol. 2674, no. 9, pp. 266–278, 2020.

[29] C. Wang, X. Wang, R. Pan, and Y. Yan, "Influence of built environment on subway trip origin and destination: insights

based on mobile positioning data," *Transportation Research Record*, vol. 2676, no. 9, pp. 693–710, 2022.

[30] L. Ni, X. C. Wang, and X. M. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 510–526, 2018.

[31] X. Yang, Z. Fang, Y. Xu, L. Yin, J. Li, and S. Lu, "Spatial heterogeneity in spatial interaction of human movements—insights from large-scale mobile positioning data," *Journal of Transport Geography*, vol. 78, pp. 29–40, 2019.

[32] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic Geography*, vol. 46, pp. 234–240, 1970/06/01 1970.

[33] J. C. García-Palomares, J. Gutiérrez, and O. D. Cardozo, "Walking accessibility to public transport: an analysis based on microdata and GIS," *Environment and Planning B: Planning and Design*, vol. 40, no. 6, pp. 1087–1102, 2013.