Research Article

Understanding City-Wide Ride-Sourcing Travel Flow: A Geographically Weighted Regression Approach

Chuqiao Chen ¹, Simon Hu ², Washington Y. Ochieng ³, Na Xie ⁴,
and Xiqun (Michael) Chen ¹

¹College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

²ZJU-UIUC Institute, Zhejiang University, Haining 314400, China

³Department of Civil and Environmental Engineering, Imperial College London, South Kensington Campus, London SW72AZ, UK

⁴School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China

Correspondence should be addressed to Xiqun (Michael) Chen; chenxiqun@zju.edu.cn

Received 4 March 2021; Revised 19 April 2021; Accepted 11 June 2021; Published 24 June 2021

Academic Editor: Zhi Chun Li

Copyright © 2021 Chuqiao Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The emerging ride-sourcing service has become an important element of urban mobility. A challenging question underlying the provision of such service is how and to what extent the built environment affects origin-destination (OD) travel flows. This paper employs the geographically weighted regression (GWR) model to analyze the OD-based ride-sourcing travel flow. It makes a comparison with the existing ordinary least square (OLS) model and spatial autocorrelation model (SAM). We have collected ride-sourcing order data in Hangzhou, China, to provide an accurate source for acquiring ride-sourcing travel flow. We investigate the effects of the residential area, points of interest (POIs), and transit stations on ride-sourcing travel flow among traffic analysis zones (TAZs). The results show the following: (a) GWR has better goodness-of-fit than SAM and OLS. (b) Residential area, enterprise, and bus stations have positive correlations with ride-sourcing OD flows, but education and subway stations have negative correlations. We have further investigated the issue and found that it is not a causal relationship between the bus station and OD flow, due to collinearity between the two variables. The bus station builds on locations with high demand, but its capacity is not large enough to reduce the ride-sourcing flow to a low level, which results in a positive coefficient. (c) Based on the estimated coefficients, the prediction of ride-sourcing flows is feasible, supporting the impact analysis for urban land use and transportation planning. This paper contributes to understanding OD-based ride-sourcing travel flow distributions and provides a framework of long-term OD flow prediction for urban land use and transportation planning.

1. Introduction

The government agencies and urban planning department attach great importance to the prediction and understanding of large-scale origin-destination (OD) flow. The OD flow pattern reflects the distribution of travel demand and reveals the human mobility pattern. It helps plan traveling routes [1, 2], discover commuting regularity [3], and analyze land use properties [4]. An accurate estimation of OD matrices benefits decision-makers to better coordinate urban resources and mobility demand.

The advanced technologies of intelligent transportation systems (ITS) have offered extensive data collection methods to rectify the trip pattern identification problem. These

technologies and data, including the automatic vehicle identification (AVI) technologies [5], radio frequency identification (RFID) and the license plate-based AVI [6], taxi GPS traces data [3–5], mobile phone data [7, 8], and smart card fare data [9], have been studied extensively in the past. Apart from transit and taxi, ride-sourcing services offered by platforms like Uber and Lyft have made tremendous changes to the transportation systems and have become an essential component of city transportation mode [10]. For instance, by 2020, Uber offered ride-sourcing services for more than 900 cities in 70 countries [11]. The number of rides served by Lyft has already reached one billion in the region of the US, Toronto, and Canada by September 2018 [12].

1.1. Objective. There have been numerous studies in the literature to investigate the influences of the built environment on travel demand. However, the analysis of OD-based ride-sourcing flows is inadequate. This paper aims to fill the research gap of ride-sourcing OD flow analyses in the existing research by applying the geographically weighted regression (GWR) model and two baseline models. We analyze and explain the relationship between OD flow and independent variables (built environment) on the origin and destination levels and then illustrate the marginal effect of the residential area and subway on ride-sourcing OD flow based on urban land use and transportation planning policies.

1.2. Methodology. We adjust the GWR model to analyze the OD-based ride-sourcing travel flow and take the ordinary least square (OLS) and spatial autocorrelation model (SAM) as comparisons. The implications of model coefficients are analyzed. We collect the emerging ride-sourcing order data that can accurately reflect the OD flow distribution of on-demand ride services. We have also collected the built environment data for 70 traffic analysis zones (TAZs) in Hangzhou City, China, including the residential area, ten types of POIs (i.e., beauty, restaurant, education, enterprise, medicine, hotel, house, entertainment service, tourist spots, and shopping), and transit stations (i.e., bus and subway). We further analyze the policy influence of subway stations development in 2020 and the increase of the residential areas in 2050 on TAZs. The plausible reasons for future OD flow changes are discussed and interpreted. The results could provide insights into the importance of existing factors and a better understanding of ride-sourcing flow distribution.

1.3. Results. For the total OD flow, we find that the residential area, enterprise, and bus station variables are positively related to OD flow at both the origin and destination levels. However, education and the subway are always negatively related. When analyzing the result of different periods, we find that the enterprise attracts people in the morning rush, while in the evening rush hour, people travel from zones with dense enterprise buildings. There are some differences between weekdays and weekends. Entertainment service and tourist spots play a more important role on the weekend. We find that the residential area's marginal effect is positive in most TAZs at the inflow, outflow, and intraflow levels. In contrast, the subway station has negative effects on most TAZs.

1.4. Contribution. The paper's major contributions are threefold: (a) This study is among the first attempts to utilize ride-sourcing order data to explore the influence of the built environment on OD flows. (b) We explore the influence of the built environment on ride-sourcing OD flow by applying three models (OLS, SAM, and modified GWR). The results show that GWR has better goodness-of-fit than SAM and OLS. (c) The future influences of built environment changes

due to local government policies on ride-sourcing OD flows are estimated based on marginal analysis.

The rest of the paper is organized as follows: Section 2 reviews the related work on OD flow estimation in terms of models used to analyze travel flows over time. Section 3 introduces the three models (i.e., GWR, OLS, and SAM). Section 4 introduces the data collected for this study. Section 5 presents the results of the three models. Finally, Section 6 concludes the findings and provides the future direction of research.

2. Literature Review

There has been a vast body of literature exploring the relationship between the built environment and traffic flow. However, when it comes to ride-sourcing data, only a handful of papers existed. Sabouri et al. [13] analyzed how Uber demand in 24 diverse US regions was affected by 5D variables with the multilevel modeling (MLM) method. They found that demand was negatively related to intersection density and destination accessibility variables. Yu and Peng [14] explored the relationship between ride-sourcing demand (from RideAustin company) and built environment and socioeconomic factors. The factors they chose covered income, age, education, population, and transit accessibility. By applying the geographically weighted regression model, Bao et al. [15] established the relationships between ride-sourcing usage and various environmental factors such as commercial, residential, and parking areas. Gerte et al. [16] used a linear panel model to estimate the relationship between ridesharing adoption and time, built environment, and demographic variable.

However, the studies above aggregated the ride-sourcing flow into origin-based demand flow rather than OD flow among different regions. The latter reflects travel mobility in a more detailed way. By analyzing the relationship between OD flow and built environment, we may find that the built environment factors affect inflow, outflow, and intraflow, which is a research gap worth filling.

Many studies have used spatial analysis approaches to understand the effect of the built environment on various transportation usages. The spatial autoregressive model (SAM) is one of them. LeSage and Pace [17] proposed the standard spatial autoregressive model (SAM), which considered the interaction among regions. Many researchers applied the model to predict traffic flow [18–21]. The GWR model is different from SAM by allowing the coefficients of explanatory variables to vary over space [22]. Plenty of researchers applied the GWR model on transit data. Cardozo et al. [23] explored the station-level transit ridership. The prediction of ridership at the Madrid subway stations showed that GWR outperformed traditional ordinary least squares multiple regression. Chiou et al. [24] identified the major factors for the public transit usage rate in Taiwan. The results showed that the GWR model had better accommodation of spatial autocorrelation and better prediction accuracy than the Tobit regression model. Ma et al. [25] explored the relationship between the built environment and transit ridership in Beijing using one-month transit card

data and POI data. Other studies utilized data such as taxi trajectory, walking demand, and daily activities. Li et al. [26] used taxi trajectory data within one week and POI information to estimate transportation factors such as pick-ups and drop-offs. Qian and Ukkusuri [27] modeled the spatial taxi ridership distribution through the GWR model with various sociodemographic and built-environment variables. They further investigated how the rise of TNCs infected traffic states and emissions [28]. Yang et al. [29] studied walking travel demand at intersections using walking counts over ten years in Chittenden County, Vermont, USA. Lucas et al. [30] studied the influence of travel disadvantage on travel amount with personal surveys. They found that the level of bus services, street connectivity, and neighborhood safety were all significant factors to the undertaken daily trips. Shen et al. [31] collected the car license plate recognition data and analyzed the spatial-temporal automobile travel demand with a geographically and temporally weighted regression model.

Our study estimates how different factors influence the ride-sourcing OD flow, which benefits urban land use and transportation planning. The model can predict the changes in flow once the future built environment change is determined. To the best of our knowledge, this is the first study that employs GWR to ride-sourcing OD flow analyses.

3. Models

OLS is a traditional linear regression model that maps the independent variables linearly to dependent variables. Compared with the traditional OLS model, the SAM introduces origin, destination, and OD dependence to capture the spatial autocorrelation. The GWR model is different from the above models by allowing the coefficients of explanatory variables to vary over space. The following subsections will introduce these models.

The traditional OLS has often been applied to estimate flows in population migration, transportation, and trade. However, the model assumes that observations are independent of each other, which could potentially cause inaccuracy in treating spatial problems. In this paper, we formulate the traditional OLS model as follows:

$$Y = X_o\beta_o + X_d\beta_d + \gamma L + \alpha\tau_N + \varepsilon, \quad (1)$$

where Y represents OD flows from origin to destination. It is an $n^2 \times 1$ vector in which n represents the number of TAZs in our study region. Thus, there are n^2 OD pairs in total. We stack the vector y first by origins and then by destinations. X_o and X_d represent independent variables in the origin and destination side. They are $n^2 \times k$ matrices in which k is the number of independent variables. β_o and β_d are the associated coefficients. L represents the travel distance between each OD pair. It is an $n^2 \times 1$ vector and γ is the associated scalar coefficient. $\alpha\tau_N$ is the constant term, in which τ_N is a vector of ones with the size of $n^2 \times 1$, α is the associated scalar coefficient, and ε is the random disturbance.

The spatial autocorrelation model extends the traditional OLS by using three spatial weight matrices for origin,

destination, and origin-to-destination dependence. It can be formulated by adding three correlated terms [17]:

$$Y = \rho_o W_o Y + \rho_d W_d Y + \rho_w W_w Y + X_o \beta_o + X_d \beta_d + \gamma L + \alpha\tau_N + \varepsilon, \quad (2)$$

where W_o , W_d , and W_w represent the origin, destination, and OD dependence. They are $n^2 \times n^2$ spatial weight matrices whose elements are relevant to the distance between regions. $W = [w_{ij}]_{n \times n}$ in which w_{ij} is the distance between origin i and destination j . I_n is an identity matrix with the size of $n \times n$, and $W_o = W \otimes I_n$, $W_d = I_n \otimes W$, $W_w = W_d \cdot W_o$, and ρ_o , ρ_d , and ρ_w are the associated scalar coefficients representing the effect strength of W_o , W_d , and W_w , respectively.

The GWR model is different from the above models by allowing the coefficients of explanatory variables to vary over space. We change the model to estimate the OD flow. The model can be formulated as follows:

$$y_{ij} = \beta_{0ij} + \sum_{k=1}^K \beta_{kij} x_{ki} + \sum_{k=1}^K \beta_{(k+K)ij} x_{kj} + \varepsilon_{ij}, \quad \forall i, j = 1, \dots, n, \quad (3)$$

where (u_i, v_i) represents the location of the centroid of the i th TAZ. y_{ij} is the OD flow from origin i to destination j . x_{ki} and x_{kj} represent the k th independent variables on origin i and destination j . β_{kij} and $\beta_{(k+K)ij}$ are the associated coefficients using the same bandwidth. β_{0ij} is the constant term. ε_{ij} is the random disturbance.

Algebraically, the GWR estimates can be expressed as follows:

$$\beta_{ij} = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y, \quad (4)$$

where $\beta_{ij} = [\beta_{0ij}, \beta_{1ij}, \dots, \beta_{2Kij}]^T$ is the associated coefficient vector. $X = [X_o, X_d]$ represent independent variables on the origin and destination side. X_o and X_d are $n^2 \times K$ matrices in which K is the number of independent variables. $W(u_i, v_i) = \text{diag}[w_{i1}, w_{i2}, \dots, w_{in}]$ is a diagonal matrix, in which $w_{ij} = w_{ij} I_n$. w_{ij} denotes the allocated weight for neighboring TAZ i and TAZ j . It is determined by the adaptive Gaussian kernel function $w_{ij} = e^{-(d_{ij}/b_i)^2}$, where d_{ij} refers to Euclidean distance between TAZ i and TAZ j . b_i is an adaptive bandwidth. We use the Akaike information criterion (AIC) to choose the best specification of $W(u_i, v_i)$.

4. Data

4.1. Ride-Sourcing Passenger Flow Pattern. The city-wide ride-sourcing order data are collected during March 6–12, 2017, from Didi company in Hangzhou, China, as the ride-sourcing passenger flow input to our models. Figure 1 illustrates the hourly orders during the period. There are two peaks on weekdays in Figure 1(a) (7:00–9:00 and 17:00–19:00), while there is only one peak from 16:00 to 18:00 on weekends. As shown in Figures 1(b) and 1(c), we analyze

FIGURE 1: Temporal and spatial distributions of ride-sourcing trips. (a) Ride-sourcing flow. (b) Weekdays. (c) Weekends. (d) Origin distribution in the morning. (e) Origin distribution in the evening.

the relationship between travel time and the number of orders. There is no significant difference between the two distributions. The spatial distributions of the trip origin in the morning and evening peak hours (7:00–9:00 and 17:00–19:00) are shown in Figures 1(d) and 1(e). Most trips originate from the city center. The distribution of ride-sourcing trips reflects the urban mobility pattern to some extent.

4.2. Data Statistics in TAZs. The statistics (i.e., minimum, mean maximum, and standard deviation) of the data on the level of TAZs are shown in Table 1. The TAZ and residential area information is offered by the Hangzhou Planning Bureau, which is the government agency in charge of urban planning. We have collected the information on POIs from AMAP, one of the largest map service companies in China. There are ten types of POIs in the dataset: beauty

TABLE 1: Statistics of variables in TAZs.

Variables	Std.*	Min	Mean	Max	Number of observations
Residential area (m ²)	1,496,802.67	0	685,900	6,439,900	70
Beauty	82.38	0	87.04	390	70
Restaurant	295.95	5	341.53	1,601	70
Education	22.46	0	22.89	99	70
Enterprise	637.43	65	769.69	2,938	70
Medicine	73.39	0	69.39	463	70
Hotel	70.35	1	59.80	394	70
House	1,079.30	13	1,216.43	6,069	70
Entertainment	989.75	33	961.39	5,780	70
Tourist spots	118.46	1	49.56	897	70
Shopping	704.54	5	681.47	4,068	70
Bus	29.42	6	60.39	144	70
Subway	1.59	0	1.09	6	70
Flow (daily)	7,937.51	0	24.32	71722	4,900

*Std.: standard deviation.

(barbershop and beauty salon), restaurant, education, enterprise, medicine, hotel, house, entertainment service (government bodies, gym, and places for recreation and entertainment), tourist spots, and shopping (grocery store, supermarket, furniture, and computer market). The subway stations and bus stations are also collected. There were only three subway lines in 2019, and the largest number of subway stations in one TAZ was only six. As shown in Figure 2, there are 70 TAZs in Hangzhou. The study region covers from 119.89 to 120.57 degrees longitude and from 30.07 to 30.5 degrees latitude, which contains the urban center.

4.3. Explanatory Variables. We choose the residential area, ten types of POIs, transit stations, and travel distance as the explanatory variables, all of which are defined at the TAZ level. Since we divide the study area into 70 TAZs, there are 70 observations for each variable. There are 4,900 (70 by 70) observations of ride-sourcing passenger flow and travel distance variables defined for each OD pair.

4.3.1. Residential Area. As the origin of travel demand, the residential area is a critical variable for the ride-sourcing passenger flow analysis. We have collected the land use data, including the base year of 2019 and the future year of 2050. It contains the size of residential land use as well as the commerce and residence land in each TAZ.

As shown in Figure 3(a), more than 80 thousand residents live in TAZs 1, 17, 27, 33, 38, 47, and 52. Since the residential area in these 70 TAZs varies significantly, we normalize the area size between 0 and 1. The distribution of the residential area in 2050 is consistent with that in 2019, as shown in Figure 3(b).

4.3.2. POIs. As the attraction of passengers, POIs have significant impacts on the ride-sourcing flow. We have collected the POIs data from AMAP, one of China's biggest map service companies. Among the ten types of POIs,

medicine facilities include pharmacy, clinic, and hospitals. Beauty facilities contain shops as barbershop and beauty salon where people improve their looks. Entertainment service facilities contain government bodies, gyms, and places for recreation and entertainment. Shopping facilities contain the grocery store, supermarket, furniture, and computer market. The education includes university, middle school, primary school, and kindergarten. Enterprise mainly covers office buildings. We illustrate the zonal distribution of each type of POIs in Figure 4.

4.3.3. Transit Stations. The transit station is another important factor influencing ride-sourcing passenger flows. This paper collects the public transit data, including the numbers of bus stations and subway stations of each TAZ in 2019. Generally, the ride-sourcing flow is expected to rise with the number of bus stations and subway stations. As shown in Figure 5(a), the city center owns the highest subway station density. There are 76 subway stations and three subway lines in Hangzhou in 2019. As shown in Figure 5(b), there are 4,227 bus stations in 2019, most of which are located in the central, northern, and southern parts of Hangzhou.

4.3.4. Multicollinearity Problem. The existence of multicollinearity will lead to bias in the experiment results. To solve this problem, we calculated the Pearson correlations among all the variables. The results show that some variables are strongly related to others. The beauty variable is strongly related to the restaurant, medicine, house, and entertainment variables with correlation coefficients of 0.910, 0.908, 0.910, and 0.910. The restaurant variable is closely related to education, medicine, and house variables with correlation coefficients of 0.917, 0.905, and 0.911. Medicine is closely related to the house, entertainment, and shopping variable with correlation coefficients of 0.901, 0.952, and 0.918. The house variable is related to entertainment and shopping with correlation coefficients of 0.877 and 0.869. The shopping variable is related to entertainment, with a correlation coefficient of 0.885. The remaining coefficients are below 0.7.

FIGURE 2: The layout of TAZs.

FIGURE 3: Residential area distributions of Hangzhou, China. (a) Residential area in 2019. (b) Residential area in 2050.

According to Qian et al. [28], variables with coefficients higher than 0.7 are deleted. Thus, beauty, restaurant, medicine, house, and shopping variables are removed from the explanatory variable in the experiment. The Pearson correlation coefficients among other explanatory variables are listed in Table 2. We have further calculated the variance inflation factor (VIF) for the rest of the variables to ensure no multicollinearity issues. The VIF indices of these variables are as follows: residential area: 1.093, education: 3.011,

enterprise: 5.675, hotel: 4.581, entertainment: 3.425, tourist spots: 2.930, bus: 1.885, and metro: 1.598. None of them exceed 6. Thus, there is no multicollinearity issue among the rest of the variables.

4.3.5. Spatial Autocorrelation Test. Spatial autocorrelation of an explanatory variable means its value in one zone is dependent on its value at neighborhood zones. The existence of

FIGURE 4: Continued.

FIGURE 4: Zonal distribution of POIs in Hangzhou, China. (a) Medicines. (b) Shopping. (c) Restaurant. (d) Education. (e) Enterprise. (f) Tourist spots. (g) House. (h) Hotel. (i) Beauty. (j) Entertainment.

FIGURE 5: Distributions of subway and bus stations in Hangzhou, China. (a) Subway stations. (b) Bus stations.

TABLE 2: Pearson correlation coefficients between explanatory variables.

	Residential area	Education	Enterprise	Hotel	Life	Tourist spots	Bus	Subway
Residential area	1.000	-0.150	-0.041	-0.073	-0.088	-0.118	0.002	0.011
Education	-0.150	1.000	0.653	0.670	0.601	0.281	0.672	0.526
Enterprise	-0.041	0.653	1.000	0.578	0.686	0.172	0.626	0.445
Hotel	-0.073	0.670	0.578	1.000	0.692	0.622	0.454	0.336
Life	-0.088	0.601	0.686	0.692	1.000	0.321	0.633	0.520
Tourist spots	-0.118	0.281	0.172	0.722	0.321	1.000	0.165	-0.026
Bus	0.002	0.672	0.626	0.454	0.633	0.165	1.000	0.362
Subway	0.011	0.526	0.445	0.336	0.520	-0.026	0.362	1.000

spatial autocorrelation will cause the basis of the GWR model. Thus, before conducting the GWR model, the spatial autocorrelation should be tested. We adopted Moran's I for testing our spatial autocorrelation, as it is the most commonly used index in literature. Moran's I of all explanatory variables are summarized in Table 3. The Z test value larger than 1.64 or smaller than -1.64 means the variable is statistically significant and has a strong spatial autocorrelation. As can be seen, most variables are significant except for shopping, house, and subway variables. The subway variable

is not significant, since it is very sparse. Only 76 subway stations were built in Hangzhou in 2019. However, it is a rather important variable for transportation planning. Thus, we keep the subway station and remove shopping and house variables.

5. Results

5.1. Model Estimation Results. We use the overall OD data during March 6-12, 2017, and normalize the dependent variable and independent variables before processing the

TABLE 3: Moran's I test of explanatory variables.

Explanatory variable	Moran's I	Z test- N	P value
Residential area	-1.136	-1.64	0.05
Beauty	-0.035	2.14	0.02
Restaurant	-0.050	1.85	0.03
Education	-0.032	2.21	0.01
Enterprise	-0.023	2.38	0.01
Medicine	-0.017	2.50	0.01
Hotel	-0.011	2.60	<0.01
House	-0.063	1.60	0.05
Entertainment	0.002	2.86	<0.01
Tourist spots	-0.046	1.92	0.03
Shopping	-0.097	0.93	0.18
Bus	0.121	5.18	<0.01
Subway	-0.168	-0.45	0.67

model. Table 4 shows the coefficient of SAM and OLS. The -2 log-likelihood, AIC (Akaike information criterion), BIC (Bayesian information criterion), and AICc (second-order AIC) indicate that SAM is more accurate than OLS. It indicates that taking origin, destination, and OD dependence into consideration is essential. Most coefficients in SAM are statistically significant, which means that most variables have considerable influences on ride-sourcing flows. The result of the GWR is summarized in Table 5. The criterion for optimal bandwidth is AIC, and the chosen bandwidth is 342. Due to the size of the coefficients, we cannot present them in one table. Hence, the average, minimum, maximum, and standard derivation of the coefficients are presented. The -2 log-likelihood, AIC, BIC, and AICc show that the GWR model fits the data better than SAM. Hence, we will choose GWR to predict OD flow in Section 5.2.

5.1.1. Influences of Residential Area and POIs. The coefficients provide some insights into how these variables influence ride-sourcing flows. As shown in Table 4, the residential area at both origin and destination has a positive coefficient in OLS and SAM. This result is consistent with the GWR model in Table 5, which is reasonable, since the residential area is the source and the strong attraction of traffic flow.

However, in the GWR model, education plays a negative role, since most Chinese parents would prefer kindergarten, primary school, and middle school close to their home locations, and their children do not need to take ride-sourcing service. A walk or bike ride is enough to cover the distance. For high school or college/university in China, most students live on campus, indicating that this group of people does not need to take the ride-sourcing service often either. That is why the education variable does not have a positive coefficient, although it is supposed to be a strong attraction of flow. Enterprise has a positive effect on the ride-sourcing OD flow. Since office buildings are places with a large population density (hundreds of working people gather in one office building) and high access frequency (people access their workplace almost every weekday), their influences on flow are positive at both origin and destination levels.

The hotel is dense around the transportation hub, like high-speed railway stations and bus terminals. Many people would choose ride-sourcing services to travel from hotel to transportation hub and vice versa, resulting in a positive coefficient. As for entertainment services, people go to these places for entertainment. Thus, they tend to choose the most relaxed transportation mode, like taxis or ride-sourcing services. The coefficient is mainly positive. Tourist spots at destination have negative correlations with ride-sourcing flows, and the spots at origin have few effects.

5.1.2. Effects of Transit Stations. In OLS and SAM (Table 4), bus stations at origin and destination have a strong positive effect on OD flow. The results are consistent in the GWR model in Table 5. The result is surprising, since the bus, as another mode of transportation, is supposed to reduce the pressure on the roadway. Figures 1 and 5(b) show that the distribution of bus stations is consistent with trip origin/destination distribution. Thus, we infer that it is not a causal relationship between the bus station and OD flow but a correlated relationship. The bus station is designed to build on locations with high demand to reduce the traffic pressure, but its capacity is not large enough to reduce the ride-sourcing flow to a low level, which results in a large coefficient.

The effect of subway stations is consistent in GWR and SAM, where the average coefficient is negative. Some studies [25, 26, 31] find that subway stations were positively related to walking demand, transit ridership, or automobile travel demand, since subway stations do not compete with these travel modes. However, in our case, subway stations relieve traffic pressure on the roadway by attracting passengers to the subway. Thus, it reduces the ride-sourcing OD flow and obtains a negative coefficient.

5.1.3. Effects of Explanatory Variables in Different Periods. Since the GWR model outperforms SAM, we further explore the effects of explanatory variables in different periods with the GWR model. As shown in Table 6, the residential area, education, hotel, bus, and subway station coefficients on weekdays and weekends do not make any differences

TABLE 4: Coefficients of SAM and OLS.

Independent variables		SAM		OLS	
		Coefficient	<i>t</i> -statistic	Coefficient	<i>t</i> -statistic
NA	Constant	0.000	0.781	0.000	3.823
Residential area variables	O_pop	0.103	1.688**	0.147	−0.541
	D_pop	0.100	2.668**	0.145	−0.568
POIs variables	O_education	−0.243	1.696**	−0.277	0.584**
	D_education	−0.211	2.147**	−0.243	26.971**
	O_enterprise	0.182	1.782**	0.334	3.010**
	D_enterprise	0.165	2.265**	0.315	−5.423**
	O_hotel	0.131	2.025**	0.104	−9.575**
	D_hotel	0.124	2.439**	0.092	−7.252**
	O_entertainment	0.010	1.467*	−0.263	5.473**
	D_entertainment	0.012	1.247	−0.262	1.849**
	O_spots	0.005	2.336**	0.032	4.972**
	D_spots	0.007	2.041**	0.036	21.145**
Bus station variables	O_bus	0.450	1.856**	0.898	−6.016**
	D_bus	0.447	1.443*	0.900	3.183**
Subway station variables	O_subway	−0.156	0.877	0.040	−0.948
	D_subway	−0.158	1.563*	0.038	−13.232**
Spatial variables	Distance	−0.221	0.913	−1.312	2.259**
	ρ_o	0.007	1.798**	NA	NA
	ρ_d	0.007	2.013**	NA	NA
	ρ_w	−0.005	−1.923**	NA	NA
−2 log-likelihood:		−14,933.8		−13,559.3	
AIC		−14,891.8		−13,523.3	
BIC		−14,755.3		−13,406.4	
AICc		−14,891.6		−13,523.2	

**0.05 level; *0.1 level; NA: not applicable.

TABLE 5: Estimation results of the GWR model.

Explanatory variable		Average coefficient	Min coefficient	Max coefficient	STD coefficient
NA	Constant	0.000	0.000	0.001	0.000
Residential area variables	O_pop	0.102	−0.061	0.729	0.164
	D_pop	0.205	−0.210	0.690	0.281
POIs variables	O_education	−0.245	−1.298	0.497	0.402
	D_education	−0.106	−1.346	1.044	0.579
	O_enterprise	0.216	−0.857	1.012	0.402
	D_enterprise	0.223	−1.470	0.959	0.560
	O_hotel	0.165	−0.052	0.607	0.130
	D_hotel	0.271	−0.315	1.021	0.297
	O_entertainment	−0.013	−1.099	2.367	0.743
	D_entertainment	0.201	−1.913	1.531	0.814
	O_spots	0.012	−0.202	0.353	0.110
	D_spots	−0.107	−0.623	0.172	0.151
Bus station variables	O_bus	0.527	−0.866	2.015	0.657
	D_bus	0.276	−1.179	1.684	0.613
Subway station variables	O_subway	−0.202	−0.596	0.209	0.175
	D_subway	−0.222	−0.879	0.070	0.178
−2 log-likelihood:		−52,874.39			
AIC		−52,653.16			
BIC		−51,934.47			
AICc		−52,648.00			

compared with those in Table 5. The residential area, hotel, and bus variables always play a positive role in attracting and generating traffic flow, and the education and subway

stations always play a negative role. On a weekday, the enterprise has a positive effect on OD flow, while on weekends, its effect is not that strong and there even appears

TABLE 6: Average coefficient of the GWR model during various periods.

Explanatory variable		Weekday	Weekend	AM peak	PM peak	Off-peak
NA	Constant	0.000	0.000	0.000	0.000	0.000
Residential area variables	O_pop	0.098	0.108	0.097	0.094	0.107
	D_pop	0.195	0.228	0.182	0.231	0.202
POIs variables	O_education	-0.234	-0.270	-0.257	-0.276	-0.227
	D_education	-0.117	-0.085	-0.350	-0.098	-0.030
	O_enterprise	0.228	0.182	-0.002	0.191	0.296
	D_enterprise	0.216	-0.244	0.273	-0.180	-0.228
	O_hotel	0.172	0.147	0.162	0.185	0.156
	D_hotel	0.301	0.204	0.420	0.247	0.233
	O_entertainment	-0.050	0.076	0.249	0.029	-0.118
	D_entertainment	0.197	0.216	0.497	0.152	0.128
	O_spots	0.006	0.024	-0.020	0.002	0.027
	D_spots	-0.119	-0.080	-0.194	-0.081	-0.090
Bus station variables	O_bus	0.533	0.519	0.195	0.543	0.631
	D_bus	0.247	0.349	-0.053	0.289	0.380
Subway station variables	O_subway	-0.196	-0.216	-0.217	-0.206	-0.195
	D_subway	-0.225	-0.214	-0.249	-0.209	-0.218

a negative coefficient at the destination level. On the weekend, entertainment and tourist spots have stronger attractiveness for passengers, since the weekend is time for recreation and outings.

For coefficient at different times of day, the effect of the residential area, education, hotel, and the subway station is consistent with that on weekdays and weekends. Nevertheless, for enterprise, in the morning rush hour, the OD flow going to zones with more enterprise buildings would be enlarged. Meanwhile, in the evening rush hour, the OD flow from zones with more enterprise buildings would be enlarged. It is related to the activity of commuters who go to enterprise buildings in the morning rush hour and leave from enterprise buildings in the evening rush hour.

5.2. Marginal Effects of Policy Implementation

5.2.1. Marginal Effects of Residential Area Change. In 2050, the residential area is predicted to increase by 48.30%. The residential area distribution is shown in Figure 3(b). The increase in residential areas is not even across all TAZs. With the coefficients estimated in GWR, we predict the marginal effects of residential area change in 2050. The change ratio (local flow difference divided by local flow in 2019) of OD flow is illustrated in Figure 6.

As shown in Figure 6, we divide the flow into three types, that is, intraflow (trips depart from and arrive at the same TAZ), inflow (summation of trips arriving at the TAZ), and outflow (summation of trips departing from the TAZ). As shown in Figure 6(b), the outflow changes are consistent with residential area changes. TAZs with a significant residential size increase (15, 12, 24, 54, 58, and 67) will have an increase in the outflow. TAZs with a residential size decrease like 1 and 47 will have a decrease in the outflow. As shown in Figure 6(a), for most TAZs, the changes of outflow and inflow are similar. Meanwhile, for TAZ 47, a decrease in residential

areas will have an increase in inflow. The possible explanation is that TAZ 54, which has a large increase in residential areas, influences its neighbor 47, causing many people to enter TAZ 47. A similar situation happens to TAZ 33 and TAZ 52, which has no increase in the residential area but has an increase in inflow, since their neighbors TAZ 25 and TAZ 24 have a rising residential area size. The intraflow change in Figure 6(c) is much more even. The intraflow of TAZ 69 and TAZ 17 decreases. It may be caused by the increase of residential areas in TAZ 58 and TAZ 15, which attract part of the intraflow. Overall, the outflow changes are consistent with residential area changes.

5.2.2. Marginal Effects of Transit Station Change. The local government planned to construct ten new subway lines by 2022. As can be seen in Figure 7, the newly built subway stations are illustrated. There will be 260 subway stations in Hangzhou, increasing by 242.11% compared to the base year. Based on the coefficients acquired in the GWR model above, the increased ride-sourcing OD flows in the whole city can be estimated.

We have shown the OD flow changes ratio in Figure 8. Since the average coefficient of the subway station is negative, most of the TAZ's intraflow, outflow, and inflow will decrease. In Figure 8(a), TAZ 56 has the largest outflow decrease, since its subway station number dramatically increases from 0 to 11. For TAZ 38 and TAZ 66, which have no newly built subway stations, the outflow increase is caused by new stations in nearby TAZs like 42, 10, 6, and 57. People might flow out of the zone to take the subway. A similar conclusion can be drawn in the inflow case where TAZ 38 and TAZ 66 have an increase in inflow. Since TAZ 38 and TAZ 66 did not contain dense subway stations in 2019, people had to take the ride-sourcing service to enter these TAZs, causing an increase in inflow. The intraflow is consistent with the outflow.

FIGURE 6: OD flow changes due to the residential area change in 2050. (a) Outflow change ratio. (b) Inflow change ratio. (c) Intraflow change ratio.

FIGURE 7: Subway stations in Hangzhou by 2022.

FIGURE 8: OD flow changes due to subway development by 2022. (a) Outflow change ratio. (b) Inflow change ratio. (c) Intraflow change ratio.

6. Conclusions

This paper explores the influences of several built environment variables (e.g., residential area, POIs, and transit stations) on ride-sourcing OD flow. This study differs from related research by analyzing ride-sourcing OD flow rather than just origin-based demand (outflow) or destination-based inflow, which offers more detailed spatial information. The results of the OLS, SAM, and GWR models are compared. The GWR model is different from the other models by allowing the coefficients of explanatory variables to vary over space. The result shows that the GWR model outperforms both SAM and OLS models. On average, the increase in the residential area, enterprise, and bus station variables will increase OD

flow at both levels of origin and destination. The increase in education and subway will cause an opposite result, since students are not the main force of ride-sourcing passengers, and the subway competes with ride-sourcing services. For the different time of day, we find that enterprise attracts people in the morning rush hour, and people start to travel from enterprise building in the evening rush hour. It is related to the activity of commuters who go to work in the morning and go off work in the evening. Entertainment service and tourist spots play a different role on the weekends and weekdays. We also calculate and illustrate the changes in OD flow based on the residential area and the subway line construction plan. The findings and the modeling approach in this study help better understand the ride-sourcing flow and provide

planners and policymakers with scientific guidance on the design of urban land use and transportation planning.

This paper has a few limitations. All independent variables are regarded equally without considering their scales. For example, large shopping malls should have a larger weight than small ones. In the future, it is better to introduce multisource datasets such as subway transaction data to compare the impact of the built environment on different transportation modes. Several important characteristics like characteristics of the service (e.g., price, and type of car) should be considered for mode choice evaluations.

Data Availability

The POI information can be accessed by utilizing the AMAP API service by visiting <https://developer.amap.com/api/webservice/guide/api/search>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was financially supported by the National Key Research and Development Program of China (2018YFB1600900) and the National Natural Science Foundation of China (71922019, 71961137005, 71771198, and 71772195). This work was also supported in part by the Zhejiang University/University of Illinois at Urbana-Champaign Institute and was led by Principal Supervisor Simon Hu. The authors thank Didi Chuxing for offering the ride-sourcing data.

References

- [1] C. Chen, D. Zhang, N. Li, and Z.-H. Zhou, "B-planner: planning bidirectional night bus routes using large-scale taxi GPS traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1451–1465, 2014.
- [2] C. Chen, D. Zhang, B. Guo, X. Ma, G. Pan, and Z. Wu, "Trip planner: personalized trip planning leveraging heterogeneous crowdsourced digital footprints," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1259–1273, 2015.
- [3] C. Peng, X. Jin, K. C. Wong, M. Shi, and P. Liò, "Collective human mobility pattern from taxi trips in urban area," *PLoS One*, vol. 7, no. 4, Article ID e34487, 2012.
- [4] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi GPS traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 113–123, 2013.
- [5] M. P. Dixon and L. R. Rilett, "Population origin-destination estimation using automatic vehicle identification and volume data," *Journal of Transportation Engineering*, vol. 131, no. 2, pp. 75–82, 2005.
- [6] C. Chen, L. Zheng, C. Cui, and W. Liu, "Estimating origin-destination flows using radio frequency identification data," in *Green, Pervasive, and Cloud Computing*, S. Li, Ed., Springer International Publishing, Cham, Switzerland, pp. 215–225, 2019.
- [7] M. G. Demissie, F. Antunes, C. Bento, S. Phithakkitnukoon, and T. Sukhviul, "Inferring origin-destination flows using mobile phone data: a case study of Senegal," in *Proceedings of the 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 1–6, IEEE, Chiang Mai, Thailand, June 2016.
- [8] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin-destination trips by purpose and time of day inferred from mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 240–250, 2015.
- [9] A. Alsger, B. Assemi, M. Mesbah, and L. Ferreira, "Validating and improving public transport origin-destination estimation algorithm using smart card fare data," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 490–506, 2016.
- [10] T. Cetin and E. Deakin, "Regulation of taxis and the rise of ridesharing," *Transport Policy*, vol. 76, pp. 149–158, 2019.
- [11] Uber Technologies Inc., "Use uber in cities around the world," 2020, <https://www.uber.com/global/en/cities/>.
- [12] CNBC, "Lyft has now delivered 1 billion rides," 2018, <https://www.cnbc.com/2018/09/18/lyft-hits-1-billion-rides.html>.
- [13] S. Sabouri, K. Park, A. Smith, G. Tian, and R. Ewing, "Exploring the influence of built environment on uber demand," *Transportation Research Part D: Transport and Environment*, vol. 81, Article ID 102296, 2020.
- [14] H. Yu and Z.-R. Peng, "Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted poisson regression," *Journal of Transport Geography*, vol. 75, pp. 147–163, 2019.
- [15] J. Bao, P. Liu, H. Yu, and J. Wu, "Spatial analysis for the usage of ride-sourcing services, an application of geographically weighted regression," in *Proceedings of the 17th COTA International Conference of Transportation Professionals*, Shanghai, China, July 2017.
- [16] R. Gerte, K. C. Konduri, and N. Eluru, "Is there a limit to adoption of dynamic ridesharing systems? evidence from analysis of uber demand data from New York city," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 42, pp. 127–136, 2018.
- [17] J. P. LeSage and R. K. Pace, "Spatial econometric modeling of origin-destination flows," *Journal of Regional Science*, vol. 48, no. 5, pp. 941–967, 2008.
- [18] K. Kerkman, K. Martens, and H. Meurs, "A multilevel spatial interaction model of transit flows incorporating spatial and network autocorrelation," *Journal of Transport Geography*, vol. 60, pp. 155–166, 2017.
- [19] K. Kerkman, K. Martens, and H. Meurs, "Predicting travel flows with spatially explicit aggregate models," *Transportation Research Part A: Policy and Practice*, vol. 118, pp. 68–88, 2018.
- [20] L. Ni, X. Wang, and D. Zhang, "Impacts of information technology and urbanization on less-than-truckload freight flows in China: an analysis considering spatial effects," *Transportation Research Part A: Policy and Practice*, vol. 92, pp. 12–25, 2016.
- [21] L. Ni, X. Wang, and X. 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.
- [22] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: a method for exploring spatial nonstationarity," *Geographical Analysis*, vol. 28, pp. 281–298, 1996.

- [23] O. D. Cardozo, J. C. García-Palomares, and J. Gutiérrez, "Application of geographically weighted regression to the direct forecasting of transit ridership at station-level," *Applied Geography*, vol. 34, pp. 548–558, 2012.
- [24] Y.-C. Chiou, R.-C. Jou, and C.-H. Yang, "Factors affecting public transportation usage rate: geographically weighted regression," *Transportation Research Part A: Policy and Practice*, vol. 78, pp. 161–177, 2015.
- [25] X. Ma, J. Zhang, C. Ding, and Y. Wang, "A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership," *Computers, Environment and Urban Systems*, vol. 70, pp. 113–124, 2018.
- [26] B. Li, Z. Cai, L. Jiang, S. Su, and X. Huang, "Exploring urban taxi ridership and local associated factors using GPS data and geographically weighted regression," *Cities*, vol. 87, pp. 68–86, 2019.
- [27] X. Qian and S. V. Ukkusuri, "Spatial variation of the urban taxi ridership using GPS data," *Applied Geography*, vol. 59, pp. 31–42, 2015.
- [28] X. Qian, T. Lei, J. Xue, Z. Lei, and S. V. Ukkusuri, "Impact of transportation network companies on urban congestion: evidence from large-scale trajectory data," *Sustainable Cities and Society*, vol. 55, Article ID 102053, 2020.
- [29] H. Yang, X. Lu, C. Cherry, X. Liu, and Y. Li, "Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression," *Journal of Transport Geography*, vol. 64, pp. 184–194, 2017.
- [30] K. Lucas, I. Philips, C. Mulley, and L. Ma, "Is transport poverty socially or environmentally driven? comparing the travel behaviours of two low-income populations living in central and peripheral locations in the SAME city," *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 622–634, 2018.
- [31] X. Shen, Y. Zhou, S. Jin, and D. Wang, "Spatiotemporal influence of land use and household properties on automobile travel demand," *Transportation Research Part D: Transport and Environment*, vol. 84, Article ID 102359, 2020.