

Article

Understanding Spatio-Temporal Characteristics of Urban Travel Demand Based on the Combination of GWR and GLM

Jinjun Tang ¹, Fan Gao ¹, Fang Liu ², Wenhui Zhang ^{3,*} and Yong Qi ^{4,*}

¹ School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China; jinjuntang@csu.edu.cn (J.T.); jtfangao@csu.edu.cn (F.G.)

² School of Transportation Engineering, Changsha University of Science and Technology, Changsha 410205, China; rcliufang@163.com

³ Traffic School, Northeast Forestry University, Harbin 150040, China

⁴ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

* Correspondence: zhangwenhui@nefu.edu.cn (W.Z.); qyong@njust.edu.cn (Y.Q.)

Received: 19 August 2019; Accepted: 2 October 2019; Published: 7 October 2019



Abstract: Taxis are an important part of the urban public transit system. Understanding the spatio-temporal variations of taxi travel demand is essential for exploring urban mobility and patterns. The purpose of this study is to use the taxi Global Positioning System (GPS) trajectories collected in New York City to investigate the spatio-temporal characteristic of travel demand and the underlying affecting variables. We analyze the spatial distribution of travel demand in different areas by extracting the locations of pick-ups. The geographically weighted regression (GWR) method is used to capture the spatial heterogeneity in travel demand in different zones, and the generalized linear model (GLM) is applied to further identify key factors affecting travel demand. The results suggest that most taxi trips are concentrated in a fraction of the geographical area. Variables including road density, subway accessibility, Uber vehicle, point of interests (POIs), commercial area, taxi-related accident and commuting time have significant effects on travel demand, but the effects vary from positive to negative across the different zones of the city on weekdays and the weekend. The findings will be helpful to analyze the patterns of urban travel demand, improve efficiency of taxi companies and provide valuable strategies for related polices and managements.

Keywords: travel demand; geographically weighted regression; generalized linear model; spatial heterogeneity

1. Introduction

As one of major components of the urban transportation system, taxis serve a wide range of geographical coverage while meeting travel needs. On one hand, they provide uninterrupted door-to-door services for passengers; on the other hand, they are also an important connection for other transportation modes such as buses, subways and light rail [1,2]. Recently, the development of the intelligent transportation system (ITS) and mobile devices makes it possible to easily and comprehensively obtain large-scale data to analyze the spatial-temporal variations of travel demand [3]. In addition, the analysis of travel demand based on taxi Global Positioning System (GPS) trajectories may help to investigate urban travel patterns and regularity [4,5]. It not only reflects the city function and structure, but provides reliable evidence for urban transportation planning, management and control of transit, and the development of the taxi market. Therefore, many scholars and engineers have focused research on the analysis of travel demand. These studies have mainly discussed the following

aspects: (1) travel demand management [6–9], (2) travel demand forecasting [10–13], and (3) travel demand modeling [14,15]. Spatio-temporal analysis is one of the most attractive directions amongst these various studies. For example, a spatial econometric model is widely used to explore the relationships between travel demand and a set of factors [16–21]. From the temporal perspective, applying time series models or dividing time into several periods is the most commonly used approach to understand the variations of travel demand at different times, such as weekdays and weekends, the morning peak, the evening peak and late at night [22,23].

The geographically weighted regression (GWR) is one of the most widely used methods to understand the variations of urban travel demand, and is also applied in transportation and economic areas [3,24]. In addition, many spatio-temporal data can be expressed as time series data. Therefore, the GWR can also be used to clarify the spatio-temporal variations by combining the technique of time series mining [22]. As a result, certain patterns and regularities of travel demand in different zones at different time periods can be explored. However, the GWR also has several limitations, such as inappropriate variables, inappropriateness for multiple hypotheses testing [25], and failing to decompose a global estimate into a local estimate [26]. Previous researches have generally ignored these issues, and only take into account the relationship between taxi demand and one or more categories, for example, socioeconomic and demographic factors, land use types, traffic environments and social media variables. Fortunately, a series of solutions and alternatives, mainly including Bayesian models and spatial filtering, have been proposed to solve these problems [27,28]. Therefore, the GWR can be deemed an available approach to understand the spatio-temporal variations of travel demand. Nevertheless, most of the variables utilized tend to be discrete, and it is difficult for the GWR to identify the key factors influencing travel demand. The generalized linear model can simultaneously consider discrete and continuous variables, and is a feasible approach to address the inability of the GWR to deal with discrete variables. In other words, the combination of GWR and GLM not only visualizes the spatial heterogeneity in different zones but also accurately identifies the key factors (discrete variables and continuous variables) affecting the travel demand.

The objective of this study is to investigate the spatial heterogeneity between travel demand and various variables using the taxi GPS trajectory data obtained in New York City (NYC). The main contributions of this study are: (1) We consider variation of travel demand under the influence of comprehensive factors including socio-demographic features, land use types, traffic environment and social media. (2) Construction of GWR models to capture the spatial heterogeneity of travel demand in different zones on weekdays and the weekend. (3) According to predictions of travel demand from a GWR, a GLM is estimated to further identify the key variables affecting the travel demand. The findings will be helpful to analyze the patterns and regularity of urban travel demand, and in improving the efficiency of the taxi industry and allocation of existing taxi resources.

2. Literature Review

In the last decade, several studies have been conducted to investigate how large-scale GPS trajectory data can be used to understand travel demand. Among these, travel demand forecasting is the most attractive. Based on a functional form, forecasts are modeled using the relationship between variables representing travel demand and a set of factors, studying the spatio-temporal characteristics of travel demand to understand residents' specific travel patterns and regularities. From the perspective of travel demand management, research has mainly focused on two aspects: (1) building activity-based models to analyze travel demand and (2) micro-simulation of the relationship between travel demand and traffic parameters. The previous literature has also captured the spatio-temporal variations of travel demand by exploring the difference of demand in different zones and time periods. In addition, the improvement or estimation of the travel demand model is one of the most popular topics. For example, the combined distribution and assignment (CDA) model was applied to forecast travel demand with a specific function in which travelers can choose a destination path simultaneously [9].

A network-theoretic methodology was also used to evaluate and validate travel demand models by capturing the network structure and connectivity of the modeled trips [14].

Additionally, travel demand is affected by a large quantity of factors according to the previous literature; for example, the population in the service area [15,16], employment level [17,29], income conditions [15,18], accessibility to public transit (include metro and bus) station [2,12], road density [17,30] and the land use types [4,18]. On one hand, these static environmental variables affect travel demand. On the other hand, a set of dynamic variables are also proved to have an impact on the demand, such as points of interest (POIs) [31–33], for-hire vehicles [19,34] and taxi-related accidents. According to these studies, the influence of factors on travel demand varies from positive to negative. In other words, some factors help increase travel demand, while others lead to a decrease in the demand.

A large number of models, including the traditional four-step method [35], multiple regression models [36] and the negative binomial regression model [1], are applied to analyze the spatio-temporal characteristics of travel demand. The variables utilized are spatial non-stationary, and such non-spatial methods mentioned above assume that the regression parameters are independent of the geographic location, and fail to account for local variations or the spatial heterogeneity. Therefore, the GWR was proposed to capture the spatial heterogeneity of the relationships between travel demand and various factors. The GWR is an extension of the global model by embedding the geographic locations of the sample data into the regression parameters [37]. In other words, the parameters from the GWR are local and vary from one location to another. As discussed in the introduction section, the GWR provides a feasible way to understand the spatio-temporal variations of travel demand. On one hand, the GWR can overcome the shortcoming of spatial heterogeneity; on the other hand, it is a good way to interpret the potential influencing factors on travel demand. According to a case study of San Francisco [38], a GLM can better understand how the various factors affect the relationships in different areas using the outputs from the GWR.

The reminder of paper is organized as follows. Section 3 introduces the methodology used in the study, including models and algorithms. In Section 4, we describe the study area and datasets. Section 5 reports the experiment results and analyzes them. The final section provides the conclusion of this paper.

3. Methodology

Figure 1 shows the structure of the proposed methods. First, a variety of factors or explanatory variables affecting taxi travel demand are selected from the previous literature. Second, a GWR is applied to capture the spatial heterogeneity in the relationships between travel demand and various factors. According to the results from the GWR, a GLM is used to analyze the key factors affecting travel demand and deepen understanding of how the factors affect travel demand in different areas.

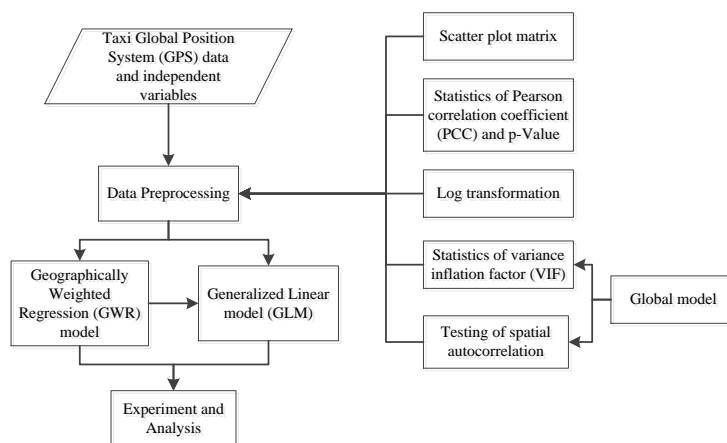


Figure 1. The structure of the proposed methods.

3.1. Terms Definition

Spatial non-stationary: Mainly represents the variable coefficients describing the spatial relationship varying over the whole space.

Corrected Akaike Information Criterion (AICc): AICc is a measure of model performance. The model with a lower value provides a better fit to the actual travel demand.

Adjusted R-squared: Also called adjusted R^2 . It is used to evaluate the prediction accuracy and the degree of model fitting, with a value ranging from 0 to 100%. For instance, an adjusted R^2 equal to 0.49 means that the model explains 49% of the variation in travel.

Residual: The residual in our study is the difference between the observations and the predictions of travel demand from the combination of the GWR and GLM.

Residual squares: The sum of the squared residuals. A smaller value means that the model fits the travel demand with higher accuracy.

3.2. Explanatory Variables

In this paper, travel demand is set as the dependent variable, and is estimated by the total number of taxi pick-ups in different blocks. Due to different patterns, destinations and times taken in every trip, we further separately divide travel demand into two parts corresponding to weekdays and the weekend [3,34].

For the influencing factors or independent variables, a large number of previous studies [15–19,32–34] have shown that taxi travel demand is affected by a great quantity of factors. According to the results shown in the above literature, we select 14 candidate independent variables divided into four categories: socioeconomic and demographic, land use, traffic environment and social media in New York City, as shown in Table 1.

For socioeconomic and demographic factors, demographic variables reflect the characteristics of urban travel and are the main factors affecting taxi travel demand; these include Hispanic Population and people with bachelor's degree or higher (BS) population [3]. In addition, employment is more significant than mean income [39]. For land use factors, it has been shown that commercial and residential districts representing the traffic volume and attraction are closely related to taxi demand [4]. In addition, taxis are an important part of the traffic system, so the impact of the traffic environment on taxi travel demand is obvious. Road density and public transit accessibility associated with infrastructure affects the aggregation and distribution of passengers [17]. Generally, higher road density is associated with a higher density of residents. The growth of road density may increase the number of trips, so road density can reflect travel demand. Public transit accessibility represents the quality of services for passengers. Uber vehicles and bicycles compete with taxis to meet travel demand [19,34], and commuting time and taxi-related accidents are also important factors affecting travel demand. Finally, the distribution of POIs represents the different attracting and distributing modes of taxi travel demand in the spatial dimension [31–33].

Following the case study in Shenzhen, China [20], road density was calculated as in Equation (1).

$$rd_i = \frac{\sum_j L_{ij} w_{ij}}{A_i} \quad (1)$$

where rd_i is the road density in the i_{th} (zip code tabulation area) ZCTA, L_{ij} and w_{ij} are the length and the width of link j in the i_{th} ZCTA, respectively, and A_i denotes the area of the i_{th} ZCTA. Note that the width of the road is obtained by the corresponding road grade. The road width and length are obtained by the following two steps: First, the length and grade of the link in each ZCTA are obtained from the graphic files. Second, according to the grade of the road, we can obtain the width of the link with different grades by referring to the Street Design Manual of New York City. Finally, we can obtain the average width of the road in each zone.

Table 1. The description of various independent variables.

| Category | Variable | Description |
|-----------------------------|------------|--------------------------------------------------------------------------------------------------------------------------------------|
| Socioeconomic & Demographic | BS Pop | The proportion of population (25 years and over) with bachelor's degree or higher to population in zip code tabulation areas (ZCTAs) |
| | His Pop | The proportion of Hispanic population to total population in ZCTAs |
| | Employment | The logarithm of the number of people in employed in ZCTAs |
| | Mean Inc | The mean income of resident in ZCTAs (\$10,000) |
| Land use | Res Area | The proportion of residential floor area to the total floor area in ZCTAs |
| | Com Area | The proportion of commercial floor area to the total floor area in ZCTAs |
| Traffic environment | Uber | The logarithm of the number of Uber vehicles trips in ZCTAs |
| | Road Den | The length of link per square meter in ZCTAs (m-1) |
| | Sub Ace | The subway accessibility in ZCTAs, the higher the value means more convenient to take the subway station |
| | Bus Ace | The bus accessibility in ZCTAs, the higher the value means more convenient to take the bus stop |
| | Com Time | The average commute time for person in ZCTAs (min) |
| | Accident | The logarithm of the number of taxi-related accidents in ZCTAs |
| Social media | Bike | The logarithm of the number of bike trips in ZCTAs |
| | POIs | The logarithm of the number of Points of interest (POIs) in ZCTAs |

It has been proved that the further away from a public transit station in a ZCTA, the fewer taxi trips will exist [4]. In view of the above rules, public transit accessibility, including bus accessibility and subway accessibility, is estimated by a linear approximation equation, as shown in Equation (2).

$$PTA_i = \frac{1}{K} \sum_k \sum_j \frac{\omega}{d_{kj}^i} \quad (2)$$

where PTA_i represents the accessibility from intersection to station, k is the road intersection or node, j denotes the public transit station in the ZCTA, and d_{kj}^i represents the geographic distance from the j_{th} road node to the k_{th} station in the ZCTA, which can be calculated by the Haversine formula. According to the latitude and longitude coordinates, the Haversine formula can be used to calculate the geographic distance between two points on earth. ω is a weighting coefficient corresponding to the characteristics of a station, such as its scale or the number of routes through the station.

3.3. GWR Model

Regression analysis is a widely and commonly used statistical method used in studying travel demand, and can evaluate relationships between two or more feature attributes. The GWR is a simple framework that extends the multiple regression model by considering local variations. In the GWR, taxi travel demand is predicted by a number of explanatory variables, and the parameters of these explanatory variables are able to vary spatially. The form of the GWR is described as Equation (3).

$$y_k(u_x, v_y) = \beta_0(u_x, v_y) + \sum_{k=1}^K \beta_k(u_x, v_y) x_k(u_x, v_y) + \varepsilon_k(u_x, v_y) \quad (3)$$

in which $y_k(u_x, v_y)$ denotes the dependent variable at location (u_x, v_y) , $x_k(u_x, v_y)$ is the independent variable, (u_x, v_y) represents the geographic location of the i_{th} ZCTA, $\varepsilon_k(u_x, v_y)$ is then random error term, and β_k is a function of location (u_x, v_y) and is allowed to vary between ZCTAs. Hence, the parameter β_k is not constant for all ZCTAs, and it is estimated by the weight in the i_{th} ZCTA.

The weight is estimated by the distance from an observation to the regression center or the geometric center of ZCTAs. As a result, the weight varies for every ZCTA (or every regression center) in the study area. As shown in Figure 2, the closer the observations to the i_{th} regression center, the greater its weight. In other words, the observations nearer the regression center have a greater influence than those farther away from the regression center for estimating the parameters β_k . Several approaches are used to estimate the weight value, such as Gaussian and bi-square kernel functions, which can be either fixed or adaptive. The Gaussian adaptive kernel function is selected to generate the weighting factors, with its value continuously and gradually decreasing from the center of the kernel, without reaching zero, as shown in Equation (4). Note that w_{ij} is the value of observations at the j_{th} location, d_{ij} is the Euclidean distance between the i_{th} regression center and the j_{th} location, and $\theta_{i(k)}$ is an adaptive bandwidth size defined as the k_{th} nearest neighbor distance. In addition, the Akaike information criterion (AICc) is used to identify an optimal distance.

$$w_{ij} = \begin{cases} \exp\left[-\left(\frac{d_{ij}}{\theta_{i(k)}}\right)^2\right] & d_{ij} < \theta_{i(k)} \\ 0 & otherwise \end{cases} \quad (4)$$

To capture the spatial heterogeneity of travel demand to explanatory variables in different blocks, we constructed GWR models responding to different times: weekdays and the weekend. The results were determined using the GWR 4.0 software [40].

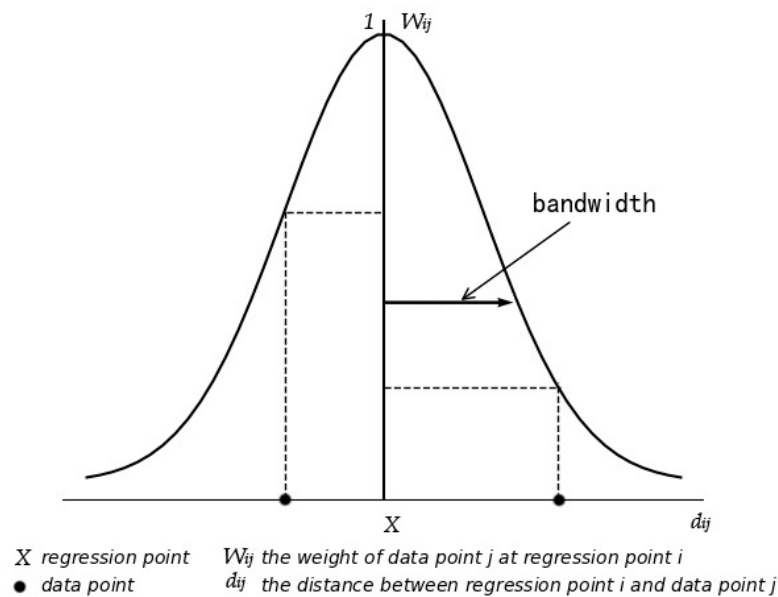


Figure 2. Gaussian weight function.

3.4. GLM

The GWR can provide a better understanding of spatial heterogeneity. However, taxi travel demand is a discrete variable. The GWR is extended by (Ordinary Least Squares) OLS, which assumes that the variable is continuous. Therefore, we propose the GLM to model travel demand. According to the predictions of travel demand provided by the GWR, we constructed the GLMs to further analyze the key variables affecting taxi travel demand. The GLMs are a broad class of model that includes linear regression, analysis of variance (ANOVA), Poisson regression and log-linear models [41]; linear regression was used in the GLM in this study. For the linear regression, the dependent variable was taxi travel demand, while the explanatory variables were the factors including socio-demographic features, land use types, traffic environment and social media. The GLM specification is given as:

$$y_j = \beta_0 + \sum_{j=1}^n \beta_j x_j + \varepsilon_j \quad (5)$$

where index j stands for each observation, y_j is the dependent variable which is estimated by GWR models, the independent variable x_j is the j_{th} feature and is linear corresponding to the parameters β_j , β_j is the j_{th} regression coefficient estimated by maximum likelihood estimation, and ε is the random error term.

4. Experiment

4.1. Study Area

The study area is selected from the core area in NYC and covers four counties (Bronx, Manhattan, Brooklyn and Queens) [17], as shown in Figure 3a. Manhattan is recognized as the city center with high density of land use. Bronx is the northernmost county and contains a large number of parks. Queens is the largest country, and the location of JFK (John F. Kennedy International) and LGA (LaGuardia) airports. Brooklyn has the largest population. All four counties have different travel demands and patterns according to different variables.

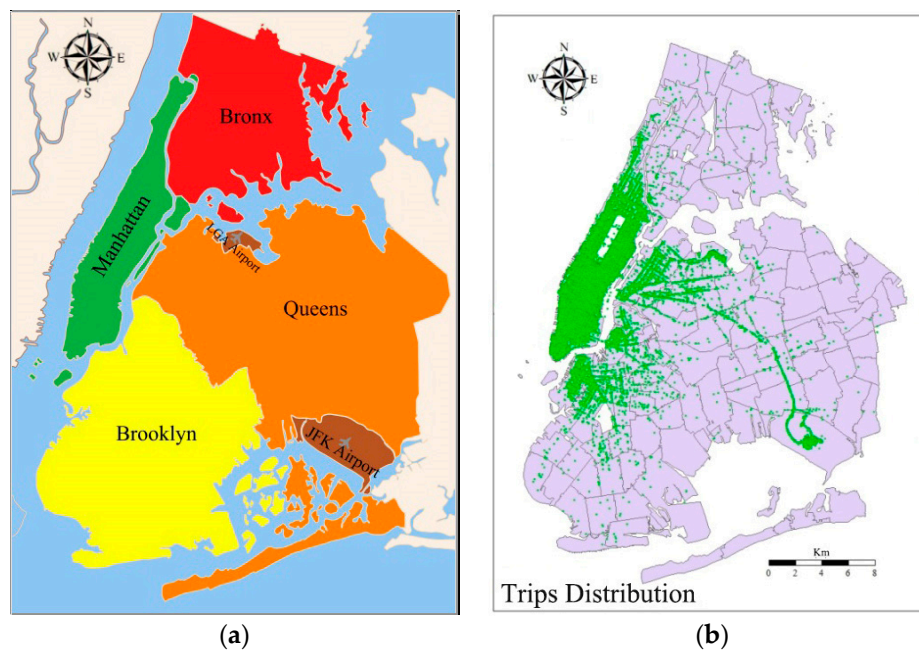


Figure 3. The study and spatial distribution of pick-ups. (a) Study area; (b) The spatial distribution of pick-ups on Oct 5, 2015.

4.2. Data Sources

The datasets used in this paper are from for kinds of sources: vehicle GPS trajectories, POIs, taxi-related accidents and other (Geographic Information System) GIS-related data. For the vehicle GPS trajectories, we extract travel trips for one week from 5 to 11 October, 2015; the total number of trips was 2,900,586. These can be freely obtained from the TLC (New York City Taxi & Limousine Commission), and the taxi datasets include information such as the times and locations of pick-ups and drop-offs, the number of passengers and travel distance, as well as tips paid. The taxi trajectory is defined as the sum of the yellow and green taxis in NYC. Note that there are no pick-up coordinates in the Uber datasets; only pick-up location ID corresponding to Taxi Zone is available. Unfortunately, the number of Taxi IDs is not as high as the number of ZCTAs. In addition, we only use the points for pick-up or pick-up location ID instead of the complete trajectories because of privacy issues. Several previous studies have revealed that the number of POIs has a positive effect on travel demand. In addition, combining the POI and the taxi GPS trajectory makes it feasible to reflect residents' travel patterns [31–33]. The POI utilized in this paper is published by Foursquare and collected by individual users from GitHub. With GitHub, all participants are free to view and download from the website. We extracted the POI data from 5 to 11 October, 2015. Taxi-related accidents are provided by an accident dataset reported by the NYPD (New York Police Department). A taxi-related accident means that the participants in an accident are more than one taxi, such as a taxi colliding with a pedestrian. Additional GIS-related data, including socio-demographic features, traffic environment, and land use types are downloaded from NYC Geodatabase. These are complex and important for analyzing the relationship between travel demand and various explanatory variables. Figure 3b shows the spatial distribution of taxi pick-ups on Monday, 5 October, 2015. It is revealed that the trips are mainly concentrated in Manhattan, northeast Brooklyn, northwest Queens and at JFK airport. As a result, these areas are more important than other areas with sparsely distributed trips for analyzing the spatio-temporal variations of travel demand.

4.3. Data Preprocessing

According to the ZCTAs, we can easily process the travel demand and explanatory variables. The original geographic shapefile of NYC contains 216 zip ZCTAs. To avoid the problem of data

invalidation, we merge the ZCTAs with small adjacent areas and delete several ZCTAs with all information are zero. The study area finally divides into 167 ZCTAs. In addition, we also carefully manipulate the variables to reduce the impact of data errors on the model results. The datasets will be processed as follows:

1. Taxi trips that do not contain the locations of pick-ups are removed.
2. According to the grid cell decomposition method, the number of occurrences of Uber vehicles in each ZCTA is counted, as shown in Equation (6).

$$\begin{cases} p_i = \frac{s_i}{S_k} \\ N_m = \sum_{n=1}^n N_{uk} \times p_i & i = (1, 2, \dots, I); k = (1, 2, \dots, K); m = (1, 2, \dots, 168) \\ \sum_{m=1}^M N_m = \sum_{k=1}^K N_{uk} \end{cases} \quad (6)$$

where s_i denotes the area of the i_{th} block, defined as the Taxi Zone divided by the ZCTAs. The Taxi Zone is released by the TLC and corresponds to each Location ID. S_k represents the area of the k_{th} Taxi Zone. p_i is the ratio of s_i to S_k . N_m refers to the number of Uber vehicle trips in the m_{th} ZCTA, while N_{uk} refers to the number of Uber vehicle trips in the k_{th} Taxi Zone.

In addition, we also consider essential regression problems, including linear redundancy, nonlinear relationships, spatially auto-correlated residuals and normal distribution bias, to meet the assumption of the GWR and GLMs. Multicollinearity refers to redundancy of one or more explanatory variables in combination. According to tests for multicollinearity, offending variables with a Pearson correlation coefficient (PCC) greater than 0.7 and a variance inflation factor (VIF) greater than 7.5 should be removed from the model. A scatter plot matrix graph can be created to elucidate the overall relationships between travel demand and explanatory variables, such as positive correlation, negative correlation or irrelevance. Spatial autocorrelation tools such as global Moran's I are advantageous in avoiding statistically significant spatial clustering for residuals. The residuals can be defined as the difference between results obtained by observation and by computation from a formula, and can be obtained using Python. In addition, log transformation is considered as a method to correct non-normal distribution of variables, as shown in Equation (7).

$$y = \ln(x + 1), \quad (7)$$

5. Results Analysis

5.1. Statistical Features of Variables

Figure 4 shows the frequency of travel trips in different ZCTAs. As observed, the number of trips has a positively skewed distribution. It is obvious that the dependent variable cannot be used directly for modeling. To satisfy the assumptions of the GWR model and the concentration of data, we apply a log transformation to the number of travel trips, as shown in Figure 4b.

Furthermore, subway accessibility and road density reflect the quality of service of the public transit system. Therefore, the spatial distributions of these variables are expressed in Figure 5. It can be seen that the service in Manhattan is significantly more intensive than in other areas, such as Bronx.

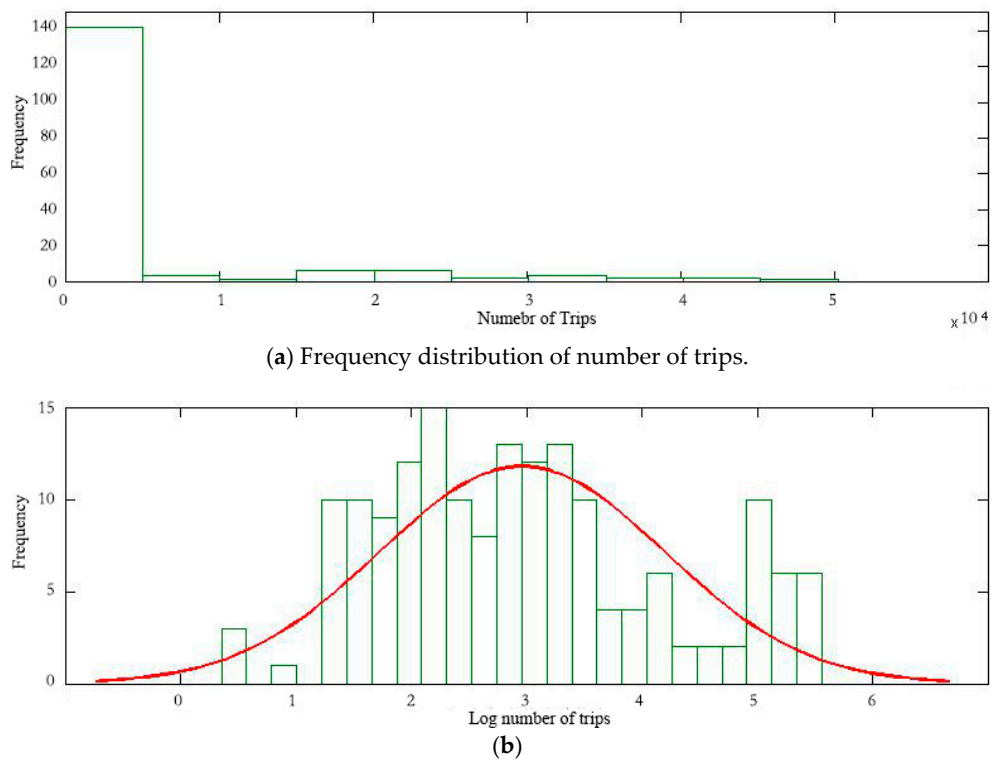


Figure 4. Frequency analysis of taxi trips. (a) Frequency distribution of number of trips. (b) Frequency distribution of log values.

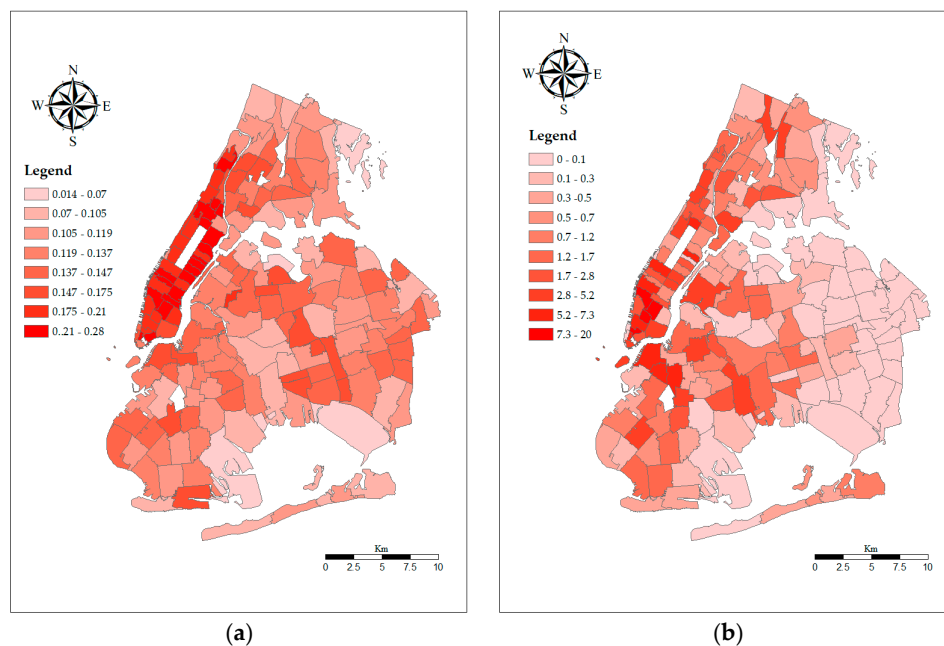


Figure 5. The spatial distributions of road density and subway accessibility. (a) Road density. (b) Subway accessibility.

5.2. Travel Demand Modeling

5.2.1. Test of Variables

The Pearson correlation coefficient (PCC) and variance inflation factor (VIF) are introduced to evaluate the effect of multicollinearity. Tables 2 and 3 show the results of the PCC values between

independent variables and the VIF values of residuals. The results show that the coefficients between any two variables are below 0.7 and the VIF of residuals are also lower than 7.5. In addition, the *p*-value used to check the statistical significance of coefficients is at the 99% significance level. As a result, all of the independent variables will avoid the effect of multicollinearity.

Table 2. The PCC and VIF value of the independent variables.

| Variable | Employment | Com Area | Uber | Road Den | Sub Ace | Com Time | POIs | Accidents | VIF |
|-----------------|------------|----------|--------|----------|---------|----------|-------|-----------|-------|
| PPC | | | | | | | | | |
| Employment | 1 | | | | | | | | 2.549 |
| Com Area | 0.012 | 1 | | | | | | | 1.356 |
| Uber | 0.211 | 0.397 | 1 | | | | | | 1.995 |
| Road Den | 0.457 | 0.229 | 0.526 | 1 | | | | | 2.351 |
| Sub Ace | 0.173 | 0.251 | 0.462 | 0.415 | 1 | | | | 1.677 |
| Com Time | -0.240 | -0.559 | -0.594 | -0.335 | -0.273 | 1 | | | 2.219 |
| POIs | 0.293 | 0.409 | 0.644 | 0.479 | 0.461 | -0.575 | 1 | | 2.249 |
| Accident | 0.204 | 0.368 | 0.463 | 0.512 | 0.523 | -0.538 | 0.558 | 1 | 2.250 |
| <i>p</i> -Value | | | | | | | | | |
| Employment | | | | | | | | | |
| Com Area | <0.01 | | | | | | | | |
| Uber | <0.01 | <0.01 | | | | | | | |
| Road Den | <0.01 | <0.01 | <0.01 | | | | | | |
| Sub Ace | <0.01 | <0.01 | <0.01 | <0.01 | | | | | |
| Com Time | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | | | | |
| POIs | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | | | |
| Accident | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | | |

Figure 6 shows the relationships between travel demand and the explanatory variables. Travel demand presents a significant positive correlation with Uber vehicles and road density, and is negatively correlated with the average commuting time.

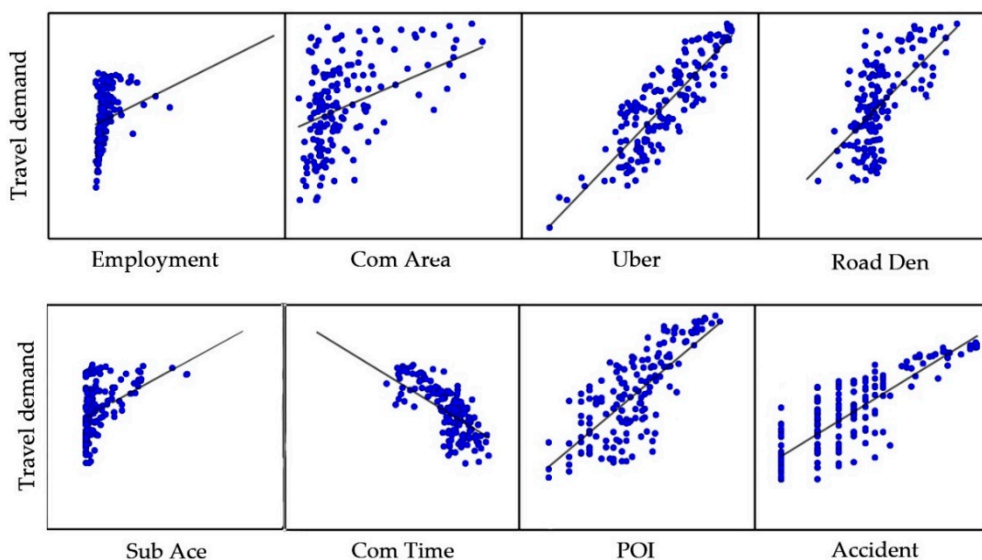


Figure 6. The relationship between travel demand and explanatory variables.

The results of spatial autocorrelation are shown in Table 3. The results indicate that the *p*-values of the residuals are very small, infinitely close to 0.000, and the minimum z-score (5.022) is also greater than 2.58. As a result, all explanatory variables are statistically significant at the 99% confidence level. In addition, the Moran’s I index values are all positive, indicating tendency toward clustering. Thus, the GWR models are capable of exploring the spatial heterogeneity of travel demand and addressing the spatial non-stationary of variables.

Table 3. Moran's I statistics for residuals.

| Variable | Moran' I | Predicting Index | z-Score | p-Value |
|------------|----------|------------------|---------|---------|
| Employment | 0.181 | −0.06 | 6.744 | 0.000 |
| Com Area | 0.310 | −0.06 | 8.875 | 0.000 |
| Uber | 0.336 | −0.06 | 9.548 | 0.000 |
| Road Den | 0.442 | −0.06 | 12.566 | 0.000 |
| Sub Ace | 0.317 | −0.06 | 9.079 | 0.000 |
| Com Time | 0.537 | −0.06 | 15.414 | 0.000 |
| POIs | 0.363 | −0.06 | 10.283 | 0.000 |
| Accident | 0.174 | −0.06 | 5.022 | 0.000 |
| residuals | 0.176 | −0.06 | 6.186 | 0.000 |

5.2.2. GWR Estimation

As discussed previously, the GWR models are applied to explore the spatial heterogeneity of travel demand to explanatory variables in different areas, and the dataset is divided into two parts: weekdays and weekends. According to variable testing results, a total number of eight explanatory variables—employment, commercial area, Uber vehicle, road density, subway accessibility, the average commuting time, POIs and taxi-related accidents—are used in the GWR models. Table 4 introduces the estimated parameters and their corresponding statistical results. The minimum value, lower quartile, median value, upper quartile and maximum value are all estimated coefficients of independent variables, and can reflect the relationship between the independent variables and travel demand in different ZCTAs. For instance, the minimum value of the average commuting time variable is negative, but the maximum value is positive, indicating that the variable is positively correlated with travel demand in some ZCTAs and negatively correlated in other ZCTAs. The Standard Deviation (STD) is an indicator of travel demand samples. A smaller value means that the GWR model can better reflect the overall travel demand.

The adjusted R-squared is 91.7% for weekdays and 90.5% for weekends. It indicates that the model or the explanatory variables can explain 91.7 percent and 90.5 percent of the variations in travel demand on weekdays and weekend. A value of more than 90% means that the GWR models provide a better understanding of travel demand. The AICc is also a measure of model performance. The value of AICc for weekdays is lower than for weekends, suggesting that the model for weekdays is more effective than the model for weekends for analyzing travel demand. In addition, the residual squares are 205.097 on weekdays and 209.283 on weekends. The spatial patterns of residuals on weekdays and weekends are plotted in Figure 7. It can be seen that the residual distributions do not present a significant spatial aggregation phenomenon over space.

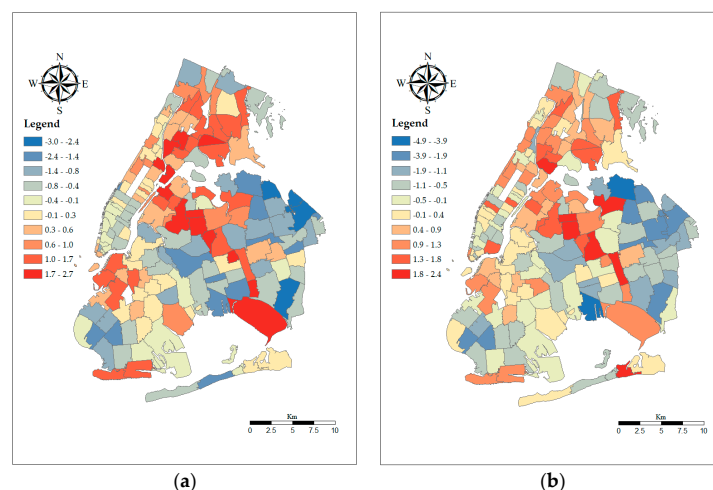
**Figure 7.** The distribution of residuals from the GWR. (a) Weekdays. (b) Weekend.

Table 4. Parameters estimations and statistical results of the GWR.

| Independent Variable | Min | Lower Quartile | Median | Upper Quartile | Max | Mean | STD |
|------------------------------------|---------|----------------|---------|----------------|--------|--------|---------|
| Weekday | | | | | | | |
| Intercept | -3.501 | -2.020 | -0.725 | 0.169 | 0.834 | 0.949 | 1.245 |
| Employment | 0.0001 | 0.0003 | 0.0006 | 0.0011 | 0.0015 | 0.0007 | 0.00004 |
| Com Area | -0.501 | -0.241 | -0.123 | 0.004 | 0.368 | -0.105 | 0.213 |
| Uber | 0.691 | 0.787 | 0.903 | 1.009 | 1.121 | 0.902 | 0.122 |
| Road Den | 0.210 | 0.278 | 0.375 | 0.481 | 0.611 | 3.854 | 1.145 |
| Sub Ace | 0.032 | 0.020 | 0.060 | 0.028 | 0.056 | 0.534 | 0.254 |
| Com Time | -0.048 | -0.039 | -0.032 | -0.024 | -0.010 | -0.031 | 0.010 |
| POIs | 0.114 | 0.227 | 0.384 | 0.459 | 0.522 | 0.344 | 0.125 |
| Accident | 0.461 | 0.569 | 0.757 | 0.889 | 1.006 | 0.737 | 0.163 |
| Akaike information criterion (AIC) | 550.946 | | | | | | |
| AICc | 555.140 | | | | | | |
| R ² | 0.947 | | | | | | |
| Adjusted R ² | 0.925 | | | | | | |
| Residual Squares | 212.811 | | | | | | |
| Weekend | | | | | | | |
| Intercept | -2.917 | -2.127 | -1.378 | -0.900 | -0.323 | -1.512 | 0.699 |
| Employment | 0.0001 | 0.00013 | 0.00017 | 0.00021 | 0.0003 | 0.0002 | 0.00004 |
| Com Area | -0.077 | 0.248 | 0.687 | 0.929 | 1.111 | 0.604 | 0.363 |
| Uber | 0.918 | 0.979 | 1.056 | 1.087 | 1.173 | 1.041 | 0.066 |
| Road Den | 0.493 | 0.540 | 0.620 | 0.747 | 0.907 | 0.650 | 0.121 |
| Sub Ace | 0.022 | 0.031 | 0.023 | 0.050 | 0.081 | 0.032 | 0.031 |
| Com Time | -0.079 | -0.074 | -0.070 | -0.061 | -0.053 | -0.067 | 0.007 |
| POIs | 0.055 | 0.138 | 0.212 | 0.278 | 0.347 | 0.207 | 0.078 |
| Accident | -0.163 | -0.110 | -0.091 | -0.025 | 0.012 | -0.076 | 0.051 |
| AIC | 559.202 | | | | | | |
| AICc | 563.234 | | | | | | |
| R ² | 0.932 | | | | | | |
| Adjusted R ² | 0.908 | | | | | | |
| Residual Squares | 224.404 | | | | | | |

5.2.3. GLM Estimation

The GLM is applied to further identify key variables that affect the spatio-temporal distribution of taxi travel demand. In GLMs, the dependent variable is provided by the GWR results, and the socio-demographic features, land use types, traffic environment and social media are recognized as the explanatory variables. Variables with a significant impact on travel demand are retained in the models, as shown in Table 5.

Table 5. Parameter estimation and statistical results from the GLM.

| Variable | Weekday | | | Weekend | | |
|------------------|-------------|---------------------|-----------------|-------------|---------------------|-----------------|
| | Coefficient | <i>t</i> -statistic | <i>p</i> -Value | Coefficient | <i>t</i> -statistic | <i>p</i> -Value |
| Intercept | -0.134 | -0.065 | <0.01 | -0.642 | -0.652 | <0.01 |
| Com Area | 0.073 | 3.657 | <0.01 | 0.604 | 2.905 | <0.01 |
| Uber | 0.746 | 9.067 | <0.01 | 1.215 | 11.734 | <0.01 |
| Road Den | 0.578 | 3.608 | 0.012 | 0.642 | 3.483 | <0.01 |
| Sub Ace | -0.634 | 2.159 | <0.01 | -0.684 | 3.904 | <0.01 |
| POIs | 0.293 | 3.086 | <0.01 | 0.295 | 1.960 | <0.01 |
| Accident | 0.344 | 4.098 | <0.01 | -0.771 | 5.092 | <0.01 |
| AIC | 632.327 | | | 641.675 | | |
| AICc | 621.462 | | | 638.218 | | |

The results show that the explanatory variables including commercial areas, Uber vehicle, road density, POIs and taxi-related accident are positively correlated with travel demand. However, the positive effect of the commercial area to travel demand is less notable. For the weekdays model, each percentage increase in the commercial area variable is associated with an increase in travel demand of 0.08 percent. The negative sign of the coefficient associated with subway accessibility implies that the travel demand in areas with subway stations will be suppressed because of the competition between

taxi and subway. This phenomenon is different from the results of the GWR and can be explained by two aspects. First, taxi drivers do not take kindly to entering subway-developed areas because of poor traffic conditions. Second, other modes of transportation to the subway station, such as biking and walking, have advantages over taxis in terms of price.

5.3. Discussion of the Relation between Variables and Travel Demand

As mentioned above, the GWR model can be applied to understand the variations in different ZCTAs, and further analyze the spatial heterogeneity of travel demand. The results show that the impact of explanatory variables on travel demand may be positive or negative and is highly sensitive to urban spatial structure. Figure 8a shows the spatio-temporal distribution of the coefficients for the employment variable from the GWR. It is shown that the coefficients have an obvious pattern of spatial non-stationarity. The coefficients are slightly higher on weekdays than on weekends in most ZCTAs, indicating that people in employment travel to work by taxis on weekdays. The spatial patterns of the coefficients are similar between weekdays and weekends, and the coefficients in the area near Manhattan, including northwestern Queens and northeast Brooklyn, are significantly larger than those in Manhattan. Moreover, compared with LGA airport, the relationship between employment and travel demand in JFK is weaker. One possible explanation is that the distance from LGA airport to the city center is shorter than from JFK airport to the city center.

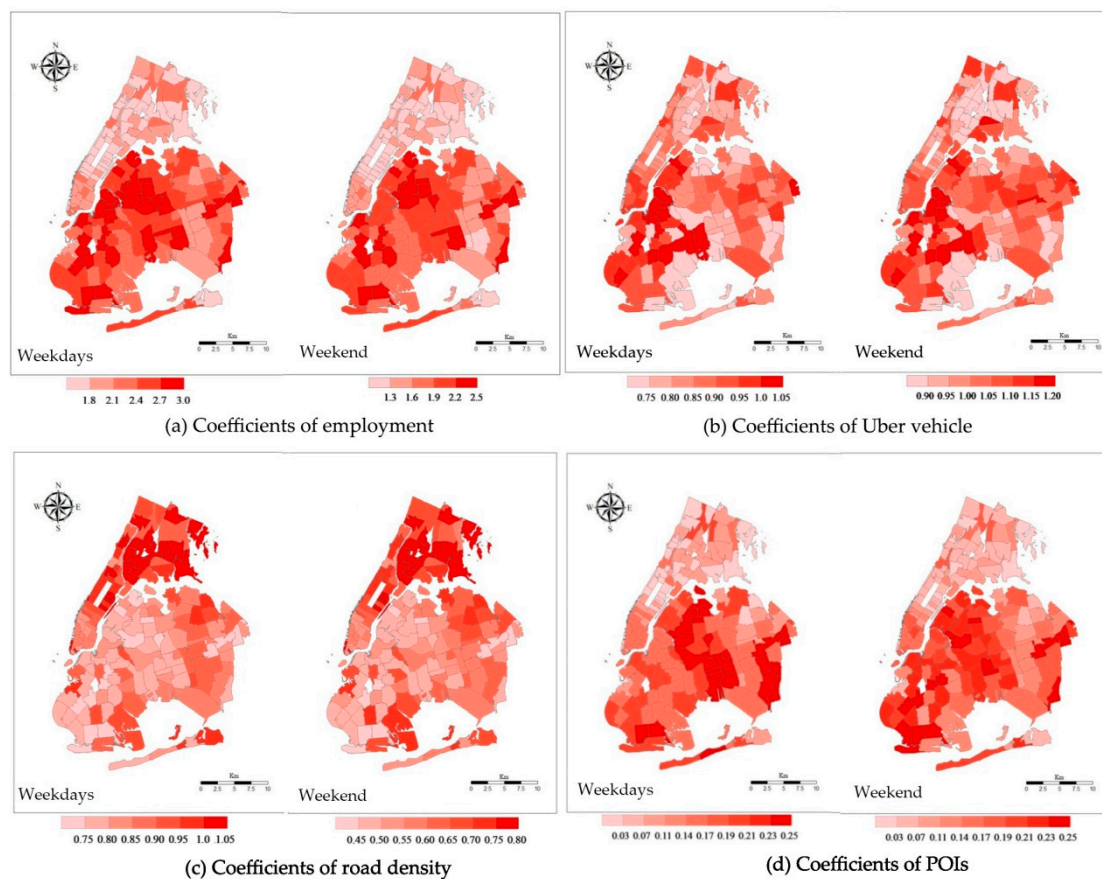


Figure 8. The spatio-temporal features of coefficients with positive values.

Similar patterns are observed in the coefficients of Uber vehicle, road density and POIs. All the coefficient signs are positive, indicating that these variables have a positive impact on travel demand in each ZCTA. Figure 8b shows the distribution of the coefficients associated with Uber vehicle on weekdays and weekends. It is revealed that the patterns are quite similar between weekdays and weekends. The coefficients in northeastern Brooklyn and northwest Queens are higher than

those in other areas, suggesting that local residents of Brooklyn and Queens are more likely to become an Uber owner, especially when they commute. Figure 8c presents the spatio-temporal distribution of the coefficients associated with road density in space. The impact of road density on travel demand is significantly different between weekdays and weekends. The travel purpose or travel time may cause this difference. In general, the increase in road density may attract more people to live in such areas, and then generate greater travel demand. Although the road density in Manhattan is high, the coefficients in the area are still lower than those in Bronx. The POIs represent the entertainment-intensive venues, and their spatial patterns can express residents' preference for certain places. Figure 8d shows that POIs always drive an increase in travel demand, whether on weekdays or weekends. As observed, the coefficients in Manhattan are less than those in southwest Queens and eastern Brooklyn. Manhattan is the city center with a large number of office buildings, while entertainment venues are mainly concentrated in southwest Queens and eastern Brooklyn. Therefore, passengers boarding in Manhattan may travel to entertainment venues at night for dinner or a performance. For subway accessibility, the coefficients are positive in the GWR and are negative in the GLMs. Therefore, subway accessibility has a significant influence on travel demand.

The spatial feature of coefficients related to average commuting time is shown in Figure 9. The negative coefficients suggest that the commuting time has a negative influence on travel demand in each ZCTA. In addition, the effects in Manhattan and southern Queens are greater than in other areas, which indicates that passengers prefer taking a taxi to using other transportation modes in Manhattan and Queens. The pattern is similar between weekdays and weekends, but the travel demand is smaller on weekdays than on weekends. The main reason for this phenomenon is that passengers are likely to take a taxi on weekdays for business trips rather than on the weekend for leisure trips.

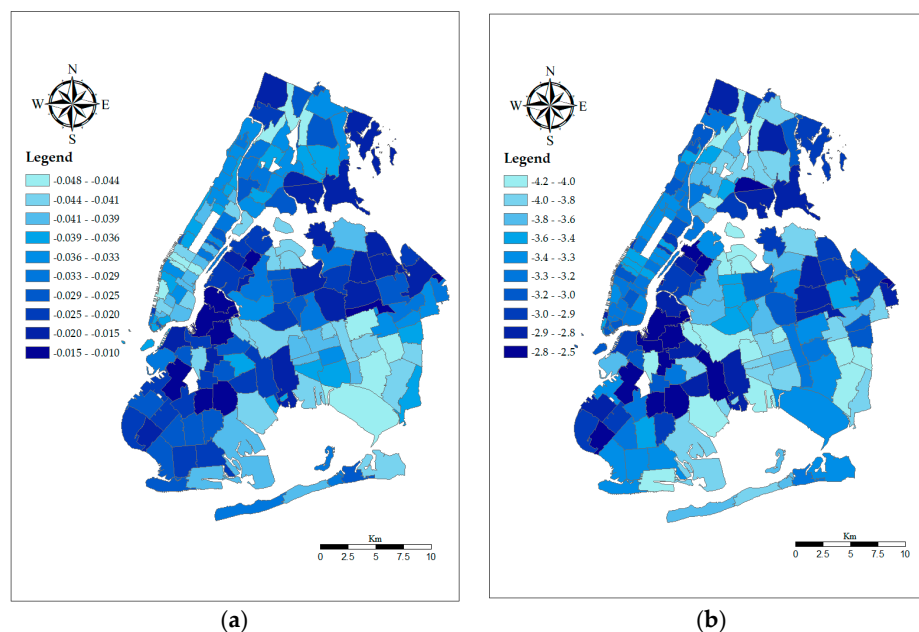


Figure 9. The spatio-temporal feature of coefficients associated with commuting time. (a) Weekdays. (b) Weekend.

According to Figure 10, the commercial area and travel demand exhibit different relationship types in different ZCTAs. In other words, the coefficients are negative in lower Manhattan, central Manhattan and southeastern Queens, but are positive in other areas. A negative coefficient means that the increase in commercial land use may decrease the number of taxi trips, while positive coefficients indicate that the increase in commercial land use may generate more trips. On one hand, the prices of merchandise in Manhattan and at JFK airport are too high to accept for passengers taking a taxi. On the other hand, Brooklyn and Queens have a significant amount of free land, so it is possible to establish more

commercial buildings such as malls or convenience stores. Moreover, taxis are more convenient than other transportation modes in the above area. As a result, people are attracted to go to these new and cheap stores by taxi. In addition, the coefficients are greater on weekends than on weekdays, suggesting that residents are likely to take a taxi to go shopping on the weekend.

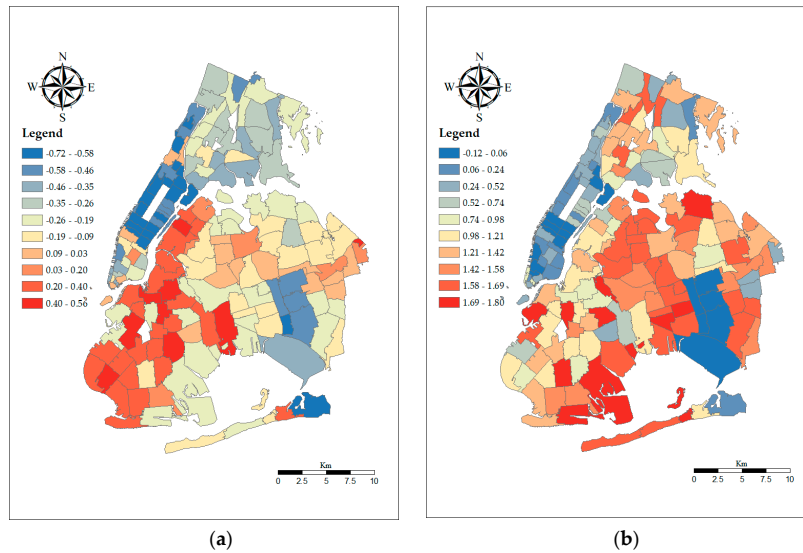


Figure 10. The spatio-temporal feature of coefficients associated with commercial area. (a) Weekdays. (b) Weekend.

The spatio-temporal distribution of coefficients associated with taxi-related accidents is shown in Figure 11. As observed, the relationship between the accidents and travel demand is positive on weekdays but negative on weekends, which indicates that the increase of accident numbers on weekdays may result in the decrease of taxi travel demand. The coefficients are greater on weekdays than on weekends, suggesting that the impact of accidents has a smaller effect on leisure trips than on business trips. It has been shown that an accident makes traffic conditions worse. Business trips are necessary on weekdays, so the impact of accidents on taxi trips is great, due to increased travel time or delays. On the other hand, accidents have a weak impact on weekend taxis because leisure trips are flexible on weekends.

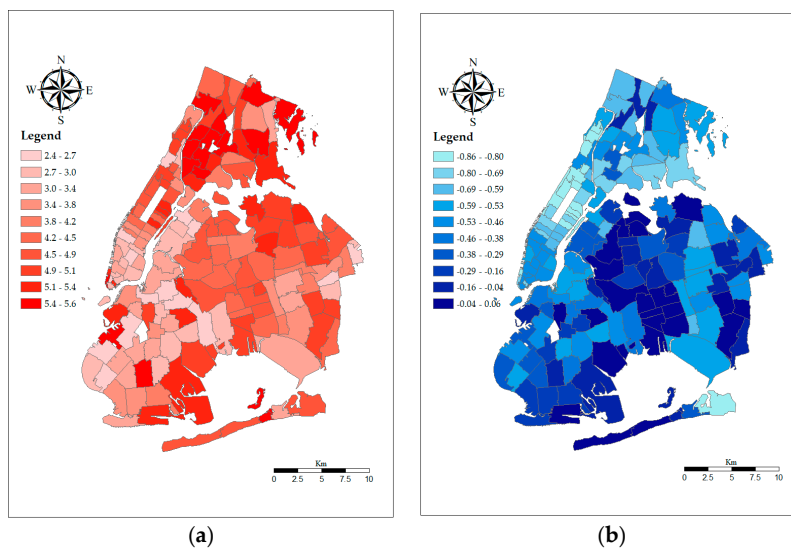


Figure 11. The spatio-temporal feature of coefficients associated with taxi-related accidents. (a) Weekdays. (b) Weekend.

According to the above discussion, we can draw a conclusion that the combination of the GWR and GLM can be used to investigate the spatial heterogeneity in the relationships between travel demand and the key variables affecting travel demand. Uber vehicles, people in employment, road density and POIs present a positive impact on travel demand. Nevertheless, the increase in commuting time will decrease the travel demand. For the proportion of commercial floor area, the coefficients signs are negative in lower Manhattan, central Manhattan and southeastern Queens, but are positive in other areas. The coefficients of POIs are positive on weekdays but negative on weekends. Finally, subway accessibility is estimated to be positively related with travel demand in the GWR, but the effect is negative in GLMs.

6. Discussion and Conclusion

The important findings and impact of this study contain the following aspects: (1) This study investigates the relationship between the spatial heterogeneity of travel demand and key variables based on the combination of a GWR and GLM using a set of taxi GPS trajectories collected in New York City. (2) Four categories, namely, socioeconomic and demographic factors, land use, traffic environment and socio media variables, are used to analyze the spatio-temporal characteristics of travel demand. The findings draw a conclusion that Uber vehicle, people in employment, road density, subway accessibility, commuting time, taxi-related accidents and POIs present an impact on travel demand, and that Uber vehicles have the strongest correlation with travel demand. (3) The relationships between travel demand and the variables vary in the spatial and temporal dimensions. For example, the impacts of POIs on travel demand in Manhattan are less than in Brooklyn, and the patterns are not similar between weekdays and weekends. (4) The GLMs are used to further explore the characteristics and patterns of travel demand using several important variables affecting travel demand. The results show that commercial areas, Uber vehicles, road density, POIs, taxi-related accidents and subway accessibility present an important impact on travel demand. In addition, the relationships of subway accessibility on travel demand are different between the GWR and GLM. The findings of this study will contribute to the sustainable and efficient development of the taxi industry. The deep exploration of local variations in travel demand aids in allocating existing taxi resources and increasing the efficiency of the transportation system.

The spatial heterogeneity in the influence of various explanatory variables on travel demand and the key factors affecting travel demand are explored through the combination of a GWR and GLM. However, this study has several shortcomings. It is focused on the analysis of taxi GPS trajectory data, and travel demand variables are simply divided into weekdays and weekends; we cannot obtain travel demand variations for other time periods, such as the morning peak, evening peak or late at night. In future work, it will be meaningful to extend this study by adding other transportation modes such as the metro, bus and shared bike. On the other hand, the GWR would be improved by taking into account time factor, the current traffic condition and the traveler's behavior. In this way, we can understand the travel demand patterns and regularity of residents from temporal and spatial perspectives in more detail. These improvements in the analysis of travel demand may assist in understanding travel patterns, improving the efficiency of the transit system, and providing a feasible way to predict and manage potential taxi resources.

Author Contributions: Conceptualization, W.Z. and Y.Q.; Data curation, F.G.; Methodology, J.T., W.Z. and Y.Q.; Resources, Y.Q.; Software, F.L.; Validation, F.G., F.L. and W.Z.; Writing – original draft, J.T. and F.G.

Funding: This research was funded in part by the National Natural Science Foundation of China (No. 71701215), Foundation of Central South University (No. 502045002), Science and Innovation Foundation of the Transportation Department in Hunan Province (No. 201725), Postdoctoral Science Foundation of China (No. 2018M630914 and 2019T120716), the National Key Research and Development Program of China: Key projects of international scientific and technological innovation cooperation between governments (Grant No. 2016YFE0108000).

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

References

1. Hochmair, H.H. Spatiotemporal pattern analysis of taxi trips in New York City. *Transp. Res. Rec.* **2016**, *2542*, 45–56. [[CrossRef](#)]
2. Kuang, L.; Yan, X. Predicting Taxi Demand Based on 3D Convolutional Neural Network and Multi-task Learning. *Remote Sens.* **2019**, *11*, 1265. [[CrossRef](#)]
3. Tang, J.; Wang, Y. A Mixed Path Size Logit-Based Taxi Customer-Search Model Considering Spatio-Temporal Factors in Route Choice. *IEEE Trans. Intell. Transp. Syst.* **2019**. [[CrossRef](#)]
4. Qian, X.; Ukkusuri, S.V. Spatial variation of the urban taxi ridership using GPS data. *Appl. Geogr.* **2015**, *59*, 31–42. [[CrossRef](#)]
5. Wang, Y.; Gu, Y. Using spatial semantics and interactions to identify urban functional regions. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 130. [[CrossRef](#)]
6. Bassolas, A.; Ramasco, J.J. Mobile phone records to feed activity-based travel demand models: MATSim for studying a cordon toll policy in Barcelona. *Transp. Res. Part A Policy Pr.* **2019**, *121*, 56–74. [[CrossRef](#)]
7. Piras, F.; Sottile, E. Modal Share Change following Implementation of Travel Demand Management Strategies. *Transp. Res. Rec. J. Transp. Res. Board* **2018**, *2672*, 731–741. [[CrossRef](#)]
8. Salon, D.; Gulyani, S. Commuting in Urban Kenya: Unpacking Travel Demand in Large and Small Kenyan Cities. *Sustainability* **2019**, *11*, 3823. [[CrossRef](#)]
9. Zhang, S.; Liu, X. Spatio-temporal modeling of destination choice behavior through the Bayesian hierarchical approach. *Phys. A Stat. Mech. Its Appl.* **2018**, *512*, 537–551. [[CrossRef](#)]
10. Tan, H.; Du, M. The Combined Distribution and Assignment Model: A New Solution Algorithm and Its Applications in Travel Demand Forecasting for Modern Urban Transportation. *Sustainability* **2019**, *11*, 2167. [[CrossRef](#)]
11. Cheng, Y.; Chen, X. Optimizing Location of Car-Sharing Stations Based on Potential Travel Demand and Present Operation Characteristics: The Case of Chengdu. *J. Adv. Transp.* **2019**. [[CrossRef](#)]
12. Tang, J.; Liang, J. Inferring driving trajectories based on probabilistic model from large scale taxi GPS data. *Phys. A Stat. Mech. Its Appl.* **2018**, *506*, 566–577. [[CrossRef](#)]
13. Tang, J.; Zhang, S. Taxi trips distribution modeling based on Entropy-Maximizing theory: A case study in Harbin city—China. *Phys. A Stat. Mech. Its Appl.* **2018**, *493*, 430–443. [[CrossRef](#)]
14. Friedrich, M.; Pestel, E. Scalable GEH: A Quality Measure for Comparing Observed and Modeled Single Values in a Travel Demand Model Validation. *Transp. Res. Rec.* **2019**. [[CrossRef](#)]
15. Saberi, M.; Rashidi, T.H. A complex network methodology for travel demand model evaluation and validation. *Netw. Spat. Econ.* **2018**, 1–23. [[CrossRef](#)]
16. Tang, J.; Zhu, Y. Identification and interpretation of spatial-temporal mismatch between taxi demand and supply using global positioning system data. *J. Intell. Transport. Sys.* **2019**, *23*, 403–415. [[CrossRef](#)]
17. Li, T.; Jing, P. Revealing the Varying Impact of Urban Built Environment on Online Car-Hailing Travel in Spatio-Temporal Dimension: An Exploratory Analysis in Chengdu, China. *Sustainability* **2019**, *11*, 1336. [[CrossRef](#)]
18. Wang, F.; Ross, C.L. New potential for multimodal connection: Exploring the relationship between taxi and transit in New York City (NYC). *Transportation* **2019**, *46*, 1051–1072. [[CrossRef](#)]
19. King, D.A.; Saldarriaga, J.F. Measuring changes in taxi trips near infill development and issues for curbside management of for-hire vehicles. *Res. Transp. Bus. Manag.* **2019**, *29*, 93–100. [[CrossRef](#)]
20. Yu, H.; Peng, Z.R. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. *J. Transp. Geogr.* **2019**, *75*, 147–163. [[CrossRef](#)]
21. Handley, J.C.; Fu, L. A case study in spatial-temporal accessibility for a transit system. *J. Transp. Geogr.* **2019**, *75*, 25–36. [[CrossRef](#)]
22. Zong, F.; Tian, Y. Trip destination prediction based on multi-day GPS data. *Phys. A Stat. Mech. Its Appl.* **2019**, *515*, 258–269. [[CrossRef](#)]
23. Wang, H.; Huang, H. Revealing Spatial-Temporal Characteristics and Patterns of Urban Travel: A Large-Scale Analysis and Visualization Study with Taxi Gps Data. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 257. [[CrossRef](#)]
24. Sultana, S.; Pourebrahim, N. Household Energy Expenditures in North Carolina: A Geographically Weighted Regression Approach. *Sustainability* **2018**, *10*, 1511. [[CrossRef](#)]

25. Griffith, D.A. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environ. Plan. A Econ. Space* **2008**, *40*, 2751–2769. [[CrossRef](#)]
26. Wheeler, D.C.; Waller, L.A. Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *J. Geogr. Syst.* **2009**, *11*, 1–22. [[CrossRef](#)]
27. Wheeler, D.C. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environ. Plan. A Econ. Space* **2007**, *39*, 2464–2481. [[CrossRef](#)]
28. Ahmed, B. The Traditional Four Steps Transportation Modeling Using a Simplified Transport Network: A Case Study of Dhaka City, Bangladesh. *IJAST* **2012**, *1*, 19–40.
29. Xu, Y.; Belyi, A. Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Comput. Environ. Urban Syst.* **2018**, *72*, 51–67. [[CrossRef](#)]
30. Tu, W.; Cao, R. Spatial variations in urban public ridership derived from GPS trajectories and smart card data. *J. Transp. Geogr.* **2018**, *69*, 45–57. [[CrossRef](#)]
31. Gao, S.; Janowicz, K. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. Gis.* **2019**, *21*, 446–467. [[CrossRef](#)]
32. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: Beijing, China, 2012; pp. 186–194.
33. Liu, Y.; Wang, F. Urban land uses and traffic ‘source-sink areas’: Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan.* **2012**, *106*, 73–87. [[CrossRef](#)]
34. Yang, Z.; Franz, M.L. Analysis of Washington, DC taxi demand using GPS and land-use data. *J. Transp. Geogr.* **2018**, *66*, 35–44. [[CrossRef](#)]
35. Hutcheson, G.D. Ordinary least-squares regression. In *The SAGE Dictionary of Quantitative Management Research*; Moutinho, L., Hutcheson, G.D., Eds.; Sage: New York, NY, USA, 2011; pp. 224–228.
36. Brunson, C.; Fotheringham, A.S. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [[CrossRef](#)]
37. Pu, Z.; Li, Z. Evaluation of spatial heterogeneity in the sensitivity of on-street parking occupancy to price change. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 67–79. [[CrossRef](#)]
38. Hu, B.; Xia, X. Analyses of the Imbalance of Urban Taxis’ High-Quality Customers Based on Didi Trajectory Data. *J. Adv. Transp.* **2019**. [[CrossRef](#)]
39. Taylor, B.D.; Fink, C.N. *The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature*; Working Paper; UCLA Institute of Transportation Studies: Los Angeles, CA, USA, 2003.
40. Nakaya, T.; Fotheringham, S. Semiparametric Geographically Weighted Generalised Linear Modelling in GWR 4.0. 2009. Available online: http://mural.maynoothuniversity.ie/4846/1/MC_Semiparametric.pdf (accessed on 1 January 2009).
41. Agresti, A. Introduction to generalized linear models. *Categ. Data Anal.* **2002**, *146*, 148.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).