Exploring the influence of built environment on demand of online car-hailing travel using multi-scale geographically temporal weighted regression model

Rongjun Cheng (chengrongjun76@126.com)  
Ningbo University

Wenbao Zeng  
Ningbo University

Yuchen Zheng  
Ningbo University

Research Article

Keywords: Online car-hailing, MGTWR, POI, Travel demand, Built environment, Spatio-temporal heterogeneity

Posted Date: June 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3014459/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Exploring the influence of built environment on demand of online car-hailing travel using multi-scale geographically temporal weighted regression model

Rongjun Cheng † Wenbao Zeng Yuchen Zheng

Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China;

Abstract

The demand for online car-hailing travel is influenced by the built environment, which exhibits spatio-temporal heterogeneity in its impact. Previous studies have commonly employed geographically weighted regression (GWR) model and geographically temporal weighted regression (GTWR) model to examine the relationship between demand for online car-hailing trips and built environment. However, these studies have overlooked the scales of influence different built environment variables. This study addressed this issue by considering scale effects based on GTWR to form the multi-scale geographically temporal weighted regression (MGTWR) to explore the spatio-temporal impact of the urban built environment on the demand for online car-hailing trips. An empirical study was conducted to assess the effectiveness of MGTWR model using Point of Interest (POI) data and online car-hailing orders data in Haikou. The evaluation indicators showed that the MGTWR model has higher accuracy in fitting than the GTWR model. Moreover, the impact of each type of POI on demand of online car-hailing travel was analyzed by examining the temporal and spatial distribution of the regression coefficients.

Keywords: Online car-hailing; MGTWR; POI; Travel demand; Built environment; Spatio-temporal heterogeneity.

1. Introduction

The development of internet technology has led to a huge change in the transport...
industry, one of the products of which is the online car-hailing [1-3]. In 2012, online car-hailing emerged as a new mode of travel and rapidly swept through most Chinese cities. As of the first half of 2022, the number of active users and drivers of online car-hailing in China had reached 360 million and 4.53 million, respectively. However, the rapid expansion of online car-hailing has resulted in various challenges, including the imbalance between supply and demand. To address this issue, gaining insights into the underlying causes of passenger flow changes is crucial for effective resource scheduling. Hence, there is a need to examine the reasons for the change in demand for online car-hailing travel.

Previous studies on online car-hailing have primarily focused on predicting demand for trips [4] and analyzing factors influencing passenger flow [5,6]. Such studies have highlighted the significance of online car-hailing as a mode of transportation. Nevertheless, there has been limited investigation on the link between built environment and passenger flow.

In mainstream academic journals, studies often use terms that start with a “D” to reflect built environment features. The urban built environment is composed of “sevenDs”: density, diversity, design, destination accessibility, distance to transit, and demographics, as well as demand management [7]. Existing studies have demonstrated a strong correlation between various travel behaviors and the built environment [8]. It is evident that built environment features have varying impacts on various travel patterns. For bike-sharing, Ma et al. [9] used the GTWR model to compare the differences in the impact of built environment on between docked and dockless bike-sharing systems. Chakour et al. [10] used the Composite Marginal Likelihood (CML) based ordered response probit (ORP) model to conclude that an effective way to increase bus ridership is to increase public transport service and accessibility. Gan et al. [11] applies a gradient lift regression tree model to investigate the non-linear relationship between built environment characteristics and station-to-station ridership. For online car-hailing, Li et al. [5] employed the GWR model to examine the influence of the built environment on online car-hailing usage. They concluded that online car-
hailing services play a significant role as a complementary mode of transportation to public transport services. For walking, Tao et al. [12] adopted the gradient boosting decision trees method to examine the relationships between walking distance and spatial attributes. The studies listed above demonstrate that there is a strong link between different modes of travel and the urban built environment. While studies have been conducted to examine the relationship between the urban built environment and the demand for online car-hailing, further research is warranted on this aspect of the relationship.

Statistical modeling is commonly used in research to examine the relationship between travel behavior and the urban built environment. Among these models, the Ordinary Least Squares (OLS) model is the most representative and widely used. However, the OLS model assumes that spatial variable relationships are fixed and do not change with spatial location, which is unrealistic [13]. To address the issue of spatial heterogeneity, several models have been proposed, including distance-decay weighted regression, two-stage least squares regression, the passion model, and the GWR model [14]. The GWR model introduced by Brunsdon in 1998, examines the impact of different factors on a localized scale by constructing regression equations for each sub-region within the study area [15]. The GWR model is not only used in the analysis of passenger flow influencing factors [16-18], but also in other fields [19,20]. It is also used in the context of the demand for online car-hailing travel. Bi et al. [21] clustered passenger flows into three patterns based on the time-varying characteristics of passenger flows. And establish passenger flow spatial model through GWR model respectively. The results verified that the built environment has different degrees of influence on the passenger flow of online car-hailing travels in spatial and temporal dimensions. Zhao et al. [22] employed the same GWR model while incorporating various grid sizes during the gridding process to examine the variations in the influence of different factors on the demand for online car-hailing trips.

The demand for online car-hailing travel is unstable in space and similarly unstable in the time dimension. The GWR is only able to analyses space instability, which makes
the spatio-temporal analysis incomplete. Therefore, the GTWR model was developed by improving the GWR model [23]. The model was first applied in the field of house price research and has since been used extensively in other fields as well. Shen et al. [24] analyzed the spatial and temporal distribution pattern of GTWR coefficients and compared the demand of different travel modes. Their results showed that there is strong spatial and temporal heterogeneity in the demand for different travel modes.

The GTWR model and the GWR model assume that all independent variables have the same bandwidth, meaning that the range of influence of each factor is the same. However, the range of influence varies between variables, as reflected in the bandwidth of each variable. Therefore, the Multi-scale Geographically Weighted Regression (MGWR) model was proposed. Cao et al. [25] constructs MGWR models for different time periods to derive significant spatial and temporal heterogeneity in the impact of different factors on the demand for online car-hailing trips.

In these studies, the scope and time scale of the variables are not considered simultaneously. This study incorporates both into the GWR model to form the MGTWR to analysis the relationship between the built environment and the online car-hailing.

Based on the above mentioned, our research has the following contributions:

1. In order to fully consider the range of influence of different factors and time scale, MGTWR will be used to study the influencing factors of demand for online car-hailing trips.

2. There may be differences in travel demand between weekdays and holidays, which will be investigated separately.

The rest of the paper is organized as follows. Section 2 presents the data sources and data processing. Section 3 describes the methodology used for the study. Section 4 analyzes the results of the models. Section 5 draws the conclusions of the paper and the outlook.

2. Study area and data collection

2.1 Study area
Haikou, the capital and primary transportation hub of Hainan Province, boasts a household population of 2.89 million. In 2022, the city witnessed a substantial influx of visitors, with a staggering 22.58 million recorded. Additionally, Haikou's Gross Domestic Product (GDP) reached RMB 213.477 billion, reflecting its significant economic stature.

Haikou has a well-established online car-hailing service, making it a suitable location for conducting research related to online car-hailing travel. The study area was selected based on the obtained data of online car-hailing orders and POI data, within the longitude range of 110.282° to 110.362° and latitude range of 19.972° to 20.050°, as shown in the red box in Fig.1. To facilitate the study, the area was divided into 20*20 grids, and the dates were categorized into holidays and working days, with each day divided into 12 time periods from a temporal perspective.

Fig.1 Study area

2.2 Data description and processing

The experimental dataset used in this study comprises a real-world open online car-hailing order dataset, supplied by the Drip Travel Gaia Project. The dataset consists of online car-hailing orders in Haikou, Hainan Province, China, recorded between May 1, 2017, and October 31, 2017. To ensure data privacy, all personal information has been desensitized. Table 1 provides an overview of some relevant information for the
dataset, while Table 2 presents a description of the fields.

Table 1. Online car-hailing order data sample table

<table>
<thead>
<tr>
<th>Order_id</th>
<th>Departure_time</th>
<th>Arrive_time</th>
<th>Starting_lng</th>
<th>Starting_lat</th>
<th>Dest_lng</th>
<th>dest_lat</th>
</tr>
</thead>
<tbody>
<tr>
<td>35184**</td>
<td>2017-10-10</td>
<td>2017-10-10</td>
<td>110.348</td>
<td>20.039</td>
<td>110.327</td>
<td>20.026</td>
</tr>
<tr>
<td>****</td>
<td>22:06:11</td>
<td>22:14:11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Dataset fields description.

<table>
<thead>
<tr>
<th>Field name</th>
<th>Field description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order_id</td>
<td>The order identifier</td>
</tr>
<tr>
<td>Departure_time</td>
<td>The time when the driver clicks ‘start billing’.</td>
</tr>
<tr>
<td>Arrive_time</td>
<td>The time when the driver clicks ‘arrive’.</td>
</tr>
<tr>
<td>Starting_lng</td>
<td>Longitude corresponding to the starting point filled in by passenger.</td>
</tr>
<tr>
<td>Starting_lat</td>
<td>Latitude corresponding to the starting point filled in by passenger.</td>
</tr>
<tr>
<td>Dest_lng</td>
<td>Longitude corresponding to the destination filled in by passenger.</td>
</tr>
<tr>
<td>Dest_lat</td>
<td>Latitude corresponding to the destination filled in by passenger.</td>
</tr>
</tbody>
</table>

The built environment data utilized in this study is the 2017 Haikou POI data, which was obtained by employing a Python-based web crawler to access the open platform APIs of Gaode Map and Baidu Map. This data was then imported into GIS for classification and sorting. Table 3 shows the specific form of the processed data, where the ‘Type’ column represents the type of POI. After deleting low-frequency POI and merging similar POI, there were 14 types of POI remaining, as shown in Table 4. The ‘Gpsx’ and ‘Gpsy’ columns indicate the latitude and longitude of each POI, respectively.

Table 3. POI dataset sample table

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Gpsx</th>
<th>Gpsy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Sun Music</td>
<td>Educational services,</td>
<td>110.323693</td>
<td>20.028170</td>
</tr>
<tr>
<td>Training Center</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhongtong Express</td>
<td>Domestic services</td>
<td>110.323693</td>
<td>20.028170</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Shopping services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Catering services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Domestic services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Address information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Corporate enterprises</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Educational services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Business residences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Government agency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Healthcare services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Accommodation services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Sports leisure services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Financial insurance services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Transport Facilities services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Car services</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The distribution of POI in the study area is presented in Fig. 2. It is observed that the left area has a significantly lower density of POI compared to the right area. Further analysis of available data suggests that the city center of Haikou lies in the upper right corner of the study area, where there is higher land use and more POIs present.
2.3 Multicollinearity test

Multicollinearity refers to the situation in which the explanatory variables in a linear regression model exhibit exact or high correlation. This means that multiple variables with strong multicollinearity can be interpreted in some way. The presence of multicollinearity can lead to large errors when explaining the effect of a particular independent variable, even though it does not affect the accuracy of model estimation. Therefore, it is important to test for multicollinearity among the alternative independent variables before performing regression analysis, to ensure the explanatory accuracy of the independent variables. To test for multicollinearity in this study, Pearson product-moment correlation coefficients (PCCs) and variance inflation factor (VIF) were used.

2.3.1 Pearson product-moment correlation coefficients test

PCCs is used in statistics to measure the degree of linear correlation between two variables and is calculated as follows [26]:

Fig. 2 Spatial distribution map of partial POIs; (a) Shopping services; (b) Catering services; (c) Corporate enterprise; (d) Government agency; (e) Transport facilities services; (f) Healthcare services; (g) Accommodation services; (h) Car services.
\[ r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \]

where \( r \) represents the PCCs, \( X_i \) and \( Y_i \) are the \( i \)-th value of variable \( X \) and the \( i \)-th value of variable \( Y \), respectively, \( \bar{X} \) and \( \bar{Y} \) denote the mean of the variable.

When the \( |r| \) value is greater than 0.7, the two variables are considered highly correlated and the choice is made to delete one of them.

Fig.3 shows that catering services POI and domestic services POI are highly correlated with a correlation coefficient of 0.85, indicating that there are likely numerous catering POIs within the domestic services POI category. Therefore, domestic services POI were removed. Similarly, educational services POI, sports leisure services POI, and financial insurance services POI were removed, whereas business residences POI was eliminated using transport facilities services. Following the removal of these POI categories, Fig.4 depicts the correlation among the different types of POI, with all correlations being less than 0.7, making them suitable for subsequent studies.

Fig.3 Correlation coefficients between different variables
2.3.2 Variance inflation factor test

The VIF is a measure of the extent of multicollinearity among the variables in a multiple linear regression model. The value of VIF ranges from 1 to positive infinity. A VIF value close to 1 indicates low multicollinearity, while a higher value indicates severe multicollinearity. If the VIF exceeds 10, the variables in the regression model are considered to have severe multicollinearity [27]. The VIF for the $k$-th independent variable can be calculated using the following formula:

$$VIF_k = \frac{1}{1 - R_k^2}$$  \hspace{1cm} (2)

$R_k^2$ is the coefficient of determination of the independent variable $x_k$ to the remaining independent variables in a regression analysis.

Table 5 presents the VIF calculations for the 9 remaining variables. The results indicate that the maximum VIF for any of these variables is 3.202, which suggests that there is no severe multicollinearity among them. Therefore, it can be assumed that these variables do not affect each other in a way that would hinder the accuracy of the regression analysis.
Table 5. Results of the VIF calculation

<table>
<thead>
<tr>
<th>Variables</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shopping services</td>
<td>1.807</td>
</tr>
<tr>
<td>Catering services</td>
<td>3.202</td>
</tr>
<tr>
<td>Address information</td>
<td>1.201</td>
</tr>
<tr>
<td>Corporate enterprises</td>
<td>2.083</td>
</tr>
<tr>
<td>Transport facilities services</td>
<td>2.976</td>
</tr>
<tr>
<td>Government agency</td>
<td>1.491</td>
</tr>
<tr>
<td>Healthcare services</td>
<td>2.295</td>
</tr>
<tr>
<td>Accommodation services</td>
<td>1.841</td>
</tr>
<tr>
<td>Car services</td>
<td>1.086</td>
</tr>
</tbody>
</table>

2.4 Spatial autocorrelation test

Spatial autocorrelation analysis is a statistical method used to examine the correlation between observations of a point in space and its neighboring points. It is useful in identifying spatial heterogeneity and spatial clustering of the study object. The most commonly used test for spatial variability is Moran's I test, which measures the degree of spatial autocorrelation of each explanatory variable. The formula for Moran's-I test is given below [27]:

\[
I = \frac{n \sum \sum w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum \sum w_{ij} \sum (y_i - \bar{y})^2}
\]  

where \( n \) is the number of spatial units; \( w_{ij} \) is the weight between positions \( i \) and \( j \); \( y_i \) and \( y_j \) are the observations at positions \( i \) and \( j \) respectively; and \( \bar{y} \) is the average of all observations.

The Z-score is often applied as an indicator of the significance of the Moran's I statistic to test the null hypothesis [9]. It can be expressed as follows:
where \( \text{Var}(I) \) and \( E(I) \) are the variance and expectation of Moran's I respectively.

The range of Moran's I is between -1 and 1. The larger the Moran's I, the more pronounced the spatial correlation; the smaller the Moran's I, the greater the spatial variation. A Moran's I value of 0 indicates a random distribution in space. To assess the significance of spatial autocorrelation and avoid random results, a P-value is used. A small P-value suggests that the observed spatial pattern is unlikely to have occurred by chance [28].

The spatial autocorrelation analysis was performed on each variable to calculate Moran's I, Z-score and P-values, which are presented in Table 6. The results demonstrate that each variable has a Moran's I greater than 0 and a P-value less than 0.000, indicating significant positive spatial correlation among the variables. Hence, all variables passed the spatial autocorrelation test at this level of significance.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Moran's I</th>
<th>Z-values</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shopping services</td>
<td>0.257</td>
<td>9.953</td>
<td>0.000</td>
</tr>
<tr>
<td>Catering services</td>
<td>0.415</td>
<td>16.02</td>
<td>0.000</td>
</tr>
<tr>
<td>Address information</td>
<td>0.249</td>
<td>9.619</td>
<td>0.000</td>
</tr>
<tr>
<td>Corporate enterprises</td>
<td>0.382</td>
<td>14.76</td>
<td>0.000</td>
</tr>
<tr>
<td>Transport facilities services</td>
<td>0.515</td>
<td>19.83</td>
<td>0.000</td>
</tr>
<tr>
<td>Government agency</td>
<td>0.410</td>
<td>15.809</td>
<td>0.000</td>
</tr>
<tr>
<td>Healthcare services</td>
<td>0.433</td>
<td>16.691</td>
<td>0.000</td>
</tr>
<tr>
<td>Accommodation services</td>
<td>0.332</td>
<td>12.825</td>
<td>0.000</td>
</tr>
<tr>
<td>Car services</td>
<td>0.319</td>
<td>12.322</td>
<td>0.000</td>
</tr>
</tbody>
</table>
3. Method

3.1 Multi-scale Geographically Temporally Weighted Regression

In spatial analysis, the relationship between variables can vary as the geographic location changes. To explore this variation, the GWR model creates a local regression equation for each region of space. The GWR model can be represented as follows [15]:

\[ y_i = \beta_{i0} \left( \mu_i, v_i \right) + \sum_{k=1}^{m} \beta_{ik} \left( \mu_i, v_i \right) x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \]  

(5)

where \( y_i \) denotes the dependent variable for the \( i \)-th sample point, \( \beta_{i0} \) is the intercept distance, \( \left( \mu_i, v_i \right) \) are the geographical coordinates of the sample points, \( m \) is number of dependent variables, \( \beta_{ik} \) indicates the regression coefficient of the \( k \)-th independent variable, \( x_{ik} \) is the \( k \)-th independent variable, \( \varepsilon_i \) indicates random error.

When time is incorporated into the model, GWR model is extended to GTWR model. It is represented by the following equation [9]:

\[ y_i = \beta_{i0} \left( \mu_i, v_i, t_i \right) + \sum_{k=1}^{m} \beta_{ik} \left( \mu_i, v_i, t_i \right) x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \]  

(6)

where \( \left( \mu_i, v_i, t_i \right) \) are the coordinates in the spatiotemporal dimension.

The MGTWR model builds on the foundation of the classical GWR model and improves on the limitations of the GTWR model by enabling variable-specific bandwidth selection. This allows for a better representation of the spatial heterogeneity among variables. The independent variable in this study is the number of POI in each area, and different bandwidths are used to determine the POI impact ranges. The MGTWR model can be expressed as follows:

\[ y_i = \beta_{00} \left( \mu_i, v_i, t_i \right) + \sum_{k=1}^{m} \beta_{ik} \left( \mu_i, v_i, t_i \right) x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \]

The regression coefficients \( \hat{\beta}_i \left( \mu_i, v_i, t_i \right) \) are calculated using the least squares method, corresponding to the following matrix expressions:

\[ \hat{\beta}_i \left( \mu_i, v_i, t_i \right) = \left[ X^T W \left( \mu_i, v_i, t_i \right) X \right]^{-1} X^T W \left( \mu_i, v_i, t_i \right) Y \]  

(7)

where \( X \) is the independent variable observation matrix; \( Y \) denotes observation vector.
of the dependent variable. \( W(\mu, v, t_i) \) is composed of weight \( w_{ij} \) space weight diagonal matrix. The expression is as follows:

\[
W(\mu, v, t_i) = \begin{bmatrix}
w_{i1} & 0 & \cdots & 0 \\
0 & w_{j2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & w_{im}
\end{bmatrix}
\] (8)

where \( w_{ij} \) represents the weight between sample point \( i \) and sample point \( j \), which can be calculated by the following equation [33]:

\[
w_{ij} = \exp \left[ -\left( \frac{d_{ij}^{ST}}{b'} \right)^2 \right]
\] (9)

where \( d_{ij}^{ST} \) is the spatio-temporal distance between sample points \( i \) and \( j \); \( b' \) indicates the spatio-temporal bandwidth.

\[
d_{ij}^{ST} = \sqrt{\lambda \left[ (u_i - u_j)^2 + (v_i - v_j)^2 + (t_i - t_j)^2 \right] + \mu (t_i - t_j)^2}
\]

where \( \lambda \) and \( \mu \) are parameters that balance spatiotemporal differences, because of the different units of space and time.

The bandwidths are calculated as shown in Fig.5. The steps are as follows [30,31]:

**Step1:** GTWR model was used to initialize the parameters.

**Step2:** Let \( f_k \) be the regression value of the \( k \)-th variable and calculate the initialized residual \( \varepsilon = y - \sum_{k=1}^{m} f_k \).

**Step3:** This residual was summed with the regression value of the 1st variable, and the summation result \( \varepsilon + f_i \) was used as the response variable to perform GTWR calculation with the independent variable to obtain the optimal bandwidth of the 1st independent variable; and the new residual was calculated.

**Step4:** Repeat the above steps until the optimal bandwidth is calculated for all variables.

**Step5:** Determine whether it converges or not, if it converges, the calculation is
finished, otherwise go back to step 3.

![Flowchart Image]

**Fig.5 Calculation of the bandwidth**

A commonly used convergence criterion is the score of change (SOC) of the Residual Sum of Squares (RSS), expressed as follows:

\[
\text{SOC}_{RSS} = \frac{\text{RSS}_{new} - \text{RSS}_{old}}{\text{RSS}_{new}}
\]  

(10)

where \( \text{RSS}_{old} \) denotes RSS from the previous cycle and \( \text{RSS}_{new} \) denotes RSS of the current cycle. When \( \text{SOC}_{RSS} \) is less than a given parameter, the algorithm determines that the iterative process satisfies the convergence condition; otherwise, it is considered as not satisfying the convergence condition. The threshold value chosen for this study is \( 10^{-5} \).

The formula for calculating RSS is as follows:

\[
\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(11)

where \( \hat{y}_i \) and \( y_i \) denote the fitted values and actual values respectively. The RSS takes values from 0 to positive infinity. The larger the value of RSS, the worse the fit.
3.2 Evaluation metrics

The Corrected Akaike Information Criterion (AICc) and the coefficient of determination (R²) were used to evaluate the fit of our model. AICc and R² are calculated as follows [32,33]:

\[
\text{AICc} = n \cdot \ln(\text{RSS}) + 2(p + 1) - n \ln(n) + \left[ 2p(p + 1) / n - p - 1 \right]
\]

(12)

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{y}_i \right)^2}
\]

(13)

where \( n \) represents the number of samples, \( p \) represents number of variables. \( \hat{y}_i \) and \( y_i \) denote the fitted values and actual values respectively.

The smaller the AICc, the better the fit of the model.

the value R² ranges from 0 to 1. R² value closer to 1 means that the prediction is better. If the R² value is 0, this means that each predicted value of the sample is equal to the mean, exactly the same as the mean model. If the R² value is less than 0, it means that the constructed model is not as good as the mean model.

4. Experiment

In this section, we will compare the model results and analyze the impacts of different POI on the demand for online taxi trips in both temporal and spatial dimensions.

4.1 Performance of the MGTWR model

Table 7 presents the fitting accuracy of the GTWR and MGTWR models for weekdays and holidays. It can be observed that for both weekdays and holidays, the AICc of the MGTWR model is smaller than that of the GWR model, with a reduction of 1540 and 2547, respectively. Additionally, the R² of the MGTWR model is higher than that of the GTWR model for both weekdays and holidays, with an increase of
0.062 and 0.057, respectively. These indicators indicate that the MGTWR model outperforms the GTWR model in fitting accuracy for both weekdays and holidays. Therefore, the subsequent analysis will use the MGTWR results to explore the impact of different types of POIs on the demand for online car-hailing trips.

Table 7. Comparison of GWR and MGWR models

<table>
<thead>
<tr>
<th>Evaluation metrics</th>
<th>workday</th>
<th>holiday</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGTWR</td>
<td>GTWR</td>
<td>MGTWR</td>
</tr>
<tr>
<td>R²</td>
<td>0.68</td>
<td>0.618</td>
</tr>
<tr>
<td>AICc</td>
<td>72737</td>
<td>74277</td>
</tr>
</tbody>
</table>

The impact of different types of POIs on the demand for online car-hailing is different. Table 8 and Table 9 present the mean, standard deviation, minimum, and maximum values of the regression coefficients for different types of POIs in the MGTWR model. A positive regression coefficient indicates that the POI facilitates demand for online car-hailing trips, and a larger value corresponds to a greater facilitating role. Conversely, a negative coefficient suggests a disincentive effect on demand. The mean value provides an overall estimate of the impact of POIs on online car-hailing trips in the study area, while the standard deviation indicates the stability of the effect. For both weekdays and holidays, the mean regression coefficients for shopping services POI, address information POI, and car services POI are close to zero, indicating a weak impact on the demand for online car-hailing trips. In contrast, transport facilities services POI and healthcare services POI have the most significant impact, with the former contributing the most, and the latter having the largest negative effect. These two types of POIs also have the most volatile impact on the demand for online car-hailing trips, as indicated by their standard deviation. Partial regression coefficients for other types of POIs are negative, implying a disincentive effect on demand in certain regions.
Table 8. The MGTWR model coefficients on workdays

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shopping services</td>
<td>-0.017</td>
<td>0.238</td>
<td>-0.72</td>
<td>-0.559</td>
</tr>
<tr>
<td>Catering services</td>
<td>2.336</td>
<td>2.295</td>
<td>-0.794</td>
<td>8.322</td>
</tr>
<tr>
<td>Address information</td>
<td>-1.11</td>
<td>1.062</td>
<td>-3.728</td>
<td>1.541</td>
</tr>
<tr>
<td>Corporate enterprises</td>
<td>1.767</td>
<td>2.203</td>
<td>-4</td>
<td>12.153</td>
</tr>
<tr>
<td>Government agency</td>
<td>2.131</td>
<td>1.439</td>
<td>-0.429</td>
<td>5.617</td>
</tr>
<tr>
<td>Healthcare services</td>
<td>-3.201</td>
<td>10.196</td>
<td>-27.079</td>
<td>27.375</td>
</tr>
<tr>
<td>Accommodation services</td>
<td>5.603</td>
<td>5.152</td>
<td>-12.05</td>
<td>21.631</td>
</tr>
<tr>
<td>Traffic facilities services</td>
<td>21.376</td>
<td>11.87</td>
<td>1.011</td>
<td>40.752</td>
</tr>
<tr>
<td>Car services</td>
<td>0.99</td>
<td>1.958</td>
<td>-10.42</td>
<td>7.465</td>
</tr>
</tbody>
</table>

Table 9. The MGTWR model coefficients on holidays

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shopping services</td>
<td>0.227</td>
<td>0.425</td>
<td>-0.455</td>
<td>1.285</td>
</tr>
<tr>
<td>Catering services</td>
<td>3.323</td>
<td>2.710</td>
<td>-0.226</td>
<td>9.171</td>
</tr>
<tr>
<td>Address information</td>
<td>-1.056</td>
<td>1.018</td>
<td>-3.638</td>
<td>1.642</td>
</tr>
<tr>
<td>Corporate enterprises</td>
<td>1.543</td>
<td>2.078</td>
<td>-3.952</td>
<td>10.74</td>
</tr>
<tr>
<td>Government agency</td>
<td>1.142</td>
<td>0.892</td>
<td>-1.833</td>
<td>-3.741</td>
</tr>
<tr>
<td>Healthcare services</td>
<td>-8.368</td>
<td>11</td>
<td>-34.065</td>
<td>12.06</td>
</tr>
<tr>
<td>Accommodation services</td>
<td>7.01</td>
<td>6.270</td>
<td>-12.272</td>
<td>25.041</td>
</tr>
</tbody>
</table>
### Table 9 (continued)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic facilities services</td>
<td>20.61</td>
<td>11.807</td>
<td>1.149</td>
<td>38.732</td>
</tr>
<tr>
<td>Car services</td>
<td>0.602</td>
<td>1.871</td>
<td>-7.101</td>
<td>10.482</td>
</tr>
</tbody>
</table>

### 4.2 Temporal features of variable coefficients

Fig.6 to Fig.13 illustrate the fluctuations in the average coefficient of each explanatory variable over the course of a day, with the blue line representing holiday and the red line representing workday.

Fig.6 specifically indicates that shopping services POI has a more pronounced effect on the promotion of online car-hailing during holidays. This can be attributed to the increase in demand for online car-hailing due to more people going out shopping during holidays. Furthermore, holidays have an even more noticeable effect on demand for online car-hailing, with the boosting effect starting to increase sharply from around 10:00. This change is due to the fact that passengers tend to leave later during holidays. During part of the weekday period, the regression coefficient is less than 0, which means that shopping services POI dampens the demand for online car-hailing trips. However, during weekday evenings, workers have time to spend on shopping, which results in a boosting effect on demand for online car-hailing trips.

![Fig.6 Temporal distribution of the MGTWR model regression coefficients](image)

(Shopping services)
Similar to shopping services POI, catering services POI have a greater impact on the demand for online car-hailing trips during holidays. Catering services POI contribute to the demand for online car-hailing trips consistently. The highest demand occurs between 20:00 and 22:00, when diners choose online car-hailing for transportation. The demand experiences a rapid growth between 12:00 and 14:00 noon due to the need for transportation after meals.

Fig.7 Temporal distribution of the MGTWR model regression coefficients
(Catering services)

There is minimal variation in the promotion of online car-hailing by POI of corporate enterprises during weekdays and holidays. This illustrates the existence of holiday work in some businesses. As commonly known, these POI dampen the demand of online car-hailing trips during late night hours. 16:00-18:00 is at the peak of the off-duty period, so the corporate enterprises POI has the greatest impact on boosting demand for online car-hailing trips.
Transport facilities services POI such as ports, airports, and train stations are the main contributors to the demand for online car-hailing trips due to their high traffic volume. However, between 2:00-6:00, the number of passengers is low and the positive effect is lowest. As the number of passengers increases, the boost in demand becomes more evident.

Fig.9 Temporal distribution of the MGTWR model regression coefficients
(Transport facilities services)

Fig.10 shows that the government agency POI consistently contributes to the demand for online car-hailing trips. However, on holidays, some government institutions are closed or operate with limited staff, resulting in less promotion of online car-hailing travel compared to weekdays. The demand for online car-hailing trips sees a greater boost between 16:00 and 20:00, when employees are likely to take a taxi to reach their destination. The promotion is also more pronounced between 10:00 and 14:00 due to the need to travel outside or go home for a lunch break. Additionally, some government agencies are in normal business on holidays and thus generate some demand.
Fig. 10 Temporal distribution of the MGTWR model regression coefficients

(Government agency)

Fig. 11 demonstrates that the impact of healthcare services POI on the demand for online car-hailing trips is highly variable. However, the changes were relatively consistent. The most significant facilitation effect is observed between 8:00 and 12:00, but this effect decreases rapidly in the afternoon and becomes negative. This indicates that healthcare services POI switch from facilitating to suppressing the demand for online car-hailing trips. There is a partial rebound during the evening peak on weekdays, after which the effect declines once again.

Fig. 11 Temporal distribution of the MGTWR model regression coefficients

(Healthcare services)

The impact of accommodation services POI on the demand for online car trips differs mainly between 8:00 and 14:00, as shown in Fig. 11. During holidays, the number of hotel occupants increases, resulting in an even greater boost. The accommodation services POI mainly refers to hotels, and the check-out times are
concentrated between 8:00 and 14:00, leading to a higher demand for online car-hailing trips during that time. However, after this period, the demand for travel decreases, and the promotion effect rapidly drops off. The demand is lowest between 20:00 and 24:00, and the accommodation services POI even has a negative impact on the demand for online car-hailing trips.

Car services POI mainly include places for car repair, car sales, car cleaning, and other car-related services. The impact of car services POI on the demand for online car-hailing trips is complex. For instance, car cleaning services are likely to have a negative effect on the demand for online car-hailing trips, whereas the impact of car repair and car sales on online car-hailing trips is uncertain. It is puzzling that most car services POI are not open at night, however, they contribute the most to the demand for online car-hailing trips during this time.
4.3 Spatial feature of variable coefficients

Fig.14 to Fig.21 illustrate the spatial variation in the regression coefficients of each explanatory variable across the study area.

According to Fig.14, the impact of shopping services POI on the demand for online car-hailing trips changes from negative to positive as move from west to east in the study area. The figure also shows that the positive impact of shopping services POI is more pronounced during holidays, especially in the upper right-hand corner of the study area. Moreover, the distribution chart of POI reveals a positive correlation between the number of shopping services POI and the demand for online car-hailing trips, as the number of shopping services POI increases from left to right. In other words, the more shopping services POI there are, the greater the demand for online car-hailing travel.

![Fig.14 Spatial distribution of the MGTWR model regression coefficients (Shopping services)](image)

Catering services POI has a consistently positive impact on the demand for online car-hailing trips within the study area, which is further pronounced during holidays. The impact of catering services POI on online car-hailing trips gradually increases from the lower left corner to the upper right corner of the study area. Similar to the impact of shopping services POI, there is a positive correlation between the density of catering services POI and the demand for online car-hailing trips, where a higher density leads to a more prominent positive impact.
The demand for online car-hailing trips from corporate enterprise POI remains constant across different types of dates. The distribution map of corporate enterprise POI reveals a gradual increase in their number from the lower left corner to the upper right corner of the study area. However, the positive impact of these POI on the demand for online car-hailing is progressively decreasing in the same direction. Therefore, a negative correlation exists between the number of corporate enterprise POI and their resulting impact, which may be due to differences in travel convenience. As transportation facilities in the lower left corner are fewer and more inconvenient to travel, employees in this area may have a higher demand for online car-hailing services.
Overall, transport facility services POI contribute to the demand for online car-hailing trips. Transport facility services POI has the greatest impact on the demand for online car-hailing trips in the lower left corner of the study area. This may be due to the low number of transport facility services POI in the area, making travel more inconvenient and increasing the demand for online car-hailing.

Fig.17 Spatial distribution of the MGTWR model regression coefficients
(Transport facilities services)

The impact of the demand for online car-hailing by government agency POI is highly dependent on the type of date. The positive effect of government agency POI on the demand for online car-hailing trips during workdays is greater than during holidays, indicating that some POI staff may not work during holidays or some agencies may close. Similar to corporate enterprise POI, the bottom left region where the demand for online car-hailing trips is most influenced by government agency POI has the most significant positive impact. In the other regions, government agency POI is evenly distributed, which leads to little variation in the spatial distribution of the coefficients.
The influence of healthcare services POI on the demand for online car-hailing trips is consistently negative during the holiday, while only the top left corner has a boosting influence on weekdays. The regression coefficient in the study area increases from below to above, with sharper changes on the left side and gradual increases on the right side. In general, there is a positive correlation between the density of healthcare services POI and the demand for online car-hailing trips.

Accommodation services POI is primarily utilized by tourists, who often require online car-hailing services. Consequently, during the holiday season, the accommodation services POI has a positive impact on online car-hailing, and the
promotion is more noticeable. The facilitating effect of the accommodation services POI on the demand for online car-hailing travel increases from the lower right to the upper left corner of the study area. This effect is more prominent in areas with low POI density, and less noticeable in areas with high POI density, as indicated by the POI distribution map.

Fig. 20 Spatial distribution of the MGTWR model regression coefficients (Accommodation services)

The regression coefficient for the car services POI is positive in most regions, indicating that the car services POI contributes to the demand for online car-hailing travel. However, there is a negative impact of car services POI on the demand for online car-hailing in the bottom-right part of the region. The effect of car services POI on demand for online car-hailing trips changes from negative to positive as one moves from the lower left corner to the upper right corner, with the effect eventually decreasing to approximately zero. This change is consistent with the change in density of automotive services POIs, indicating a positive correlation between the resulting impact and density.
Fig. 21 Spatial distribution of the MGTWR model regression coefficients
(Car services)

5. Conclusion

The rapid development of internet technology has led to the emergence of online car-hailing, which has become a vital transportation mode chosen by an increasing number of individuals for travel. Addressing the issue of traffic congestion is crucial, and studying the demand for online car-hailing travel can contribute to reducing idle time and alleviating congestion. The built environment significantly influences the demand for online car-hailing, necessitating an examination of their relationship.

This paper explored the correlation between travel demand and different categories of POI by constructing the GTWR model and the MGTWR model, employing POI data and online car-hailing data. The results showed that MGTWR model has better regression effect compared with the classical GTWR model. Finally, the result of MGTWR model was used to analyze the spatio-temporal heterogeneity of the influence of different POIs on travel demand.

In the future, the following research will be explored:

First, due to the restricted access to data, this study only investigates the direct relationship between POI and travel demand. In the subsequent study, more factors related to the demand for online car-hailing trips will be investigated.

Second, the current study performed de-zoning with relatively simple meshing,
which is too idealistic and may increase errors.

Finally, the data limitations resulted in unmatched online orders not being captured, leading to a potential discrepancy between the actual demand and orders. Subsequent efforts will be made to obtain this missing data.

Acknowledgments

This research was funded by the Natural Science Foundation of Zhejiang Province, China (Grant No. LY22G010001) and Healthy & Intelligent Kitchen Engineering Research Center of Zhejiang Province and the K.C. Wong Magna Fund in Ningbo University, China.

References


[22] Zhao, G.W., Li Z.T., Shang Y.Z., & Yang, M.Z., How does the urban built environment affect online car-hailing ridership intensity among different scales?. International journal of environmental research and public health 19.9 (2022) 5325.


[33] Li, Q.Y., Ge, H.X., & Cheng, R.J., Short-term vehicle speed prediction based on BiLSTM-GRU model considering driver heterogeneity, Physica A 610 (2023) 128410.