# Analyzing the spatial-temporal influence of built environment on taxi ridership 

Que Yuhua ${ }^{\text {a }}$<br>a University of Hong Kong, Department of Civil engineering

## ARTICLE INFO

DOI: 10.31075/PIS.66.04.01
Professional paper
Received: 19/10/2020
Accepted: 01/12/2020
Corresponding author:
queyuhua@126.com
Keywords:
Taxi ridership
Built environment
Spatiotemporal non-stationarity


#### Abstract

Taxi is an important component in urban transportation system, which covers wide area and maintains 24 h available. Exploring the relationship of taxi and built environment is very important to manage taxi service and improve transportation system. The issue is addressed by capturing the influence of built environment on taxi ridership considering spatial and temporal non-stationarities. The grid cells are developed as analysis units and the global regression model is adopted for preliminary exploring. Then the GWR are implemented considering spatial heterogeneity and the GTWR is used to analyze the spatiotemporal influence of built environment on taxi ridership. An empirical study conducted in Hong Kong Island using one-week taxi's GPS data demonstrates the effectiveness of the regression models. It's verified that GWR performs better than OLS, and GTWR outperforms the rest two regression models, indicating both time and space are critical dimensions.


## 1. Introduction

A lot of studies have been conducted to explore the feature of taxi ridership and the related factors. Some studies are in micro level to model the taxi drivers' customer searching behaviors and find out the related factors. The rate of return (ROR) significantly impacts taxi drivers' searching behaviors, and taxi drivers tend to circulate around or wait at the area, where the preceding passenger drop-off. Besides, the searching behavior will change with time varying [1]Error! Reference source not found.. The relative passenger demand, cross-zonal travel distances, intra-zonal circulation distance and the rate of return would influence the zonal decision [1][2]. Besides, the surrounding routes show cumulatively impact to taxi drivers' search decisions, and their searching strategies didn't follow the random walk principle [3]. When searching for taxi station, the search districts, travel distance and congestion level will influence whether to go to nearby taxi station, and the queue lengths at taxi stands will influence taxi whether to join the queue [4]. Apart from the microscopic search, some authors also explore the macroscopic feature of taxi trips through large amounts of GPS data. When searching for customers, hunting is usually more efficient than waiting.

Besides, when dropping off in suburb area, taxis are easier to find the next passengers towards hot ones. Some areas can be more profitable in different time of a day [5]. There is one more peak hour in weekday, and the extra peak hour at night in weekday is the biggest peak hour [6]. The spatial and temporal regularities do exist in taxi travels when the trajectories are transformed into each vehicle's intrinsic reference frame [7]. The human mobility can significantly shape taxi trip patterns and the urban traffic is composed by three basic travel purposes [9]. Most taxi trips' origins and destinations are located in the center of city, and the overall distributions of demand on weekday and weekend exhibit similar but slightly different pattern [13]. The taxi GPS data is of great importance to explore spatiotemporal features of human mobility in transportation designing and planning [14]. Exploring the taxi data encourages further use of these findings in urban transportation planning and management, such as planning the night bus route 0 , identifying traffic congestion and obtain the operational states of road network [17], detecting urban network accessibility problems and finding out these isolated pockets of areas which are difficult to reach[17] [18].

Transportation is closely related with urban land use, and many studies have been conducted to explore the relationship between the fixed transit and built environment. The GWR model is widely used in these studies [20][21][24]. The improved transport accessibility can add value to land, and the relationship between transport accessibility and land value is greatly varied over space [20]. As some non-stationary land use variables exist, the GWR model outperforms tradition linear regression, which enables in-depth local analysis [22][23]. Besides, the time variable is verified to impact to the relation between land use and transit ridership [28]. The taxicabs also show some relationship between the fixed transit. The taxis are likely to gather in some certain areas, and they can either substitute or complement mass transit with time variation [11]. The newly opening subway will attract more long-distance taxi trips [12]. Many factors can influence the multimodal connection between taxi and transit, such as trip characteristics, passengers' socio-demographic features etc. [33][34].

There are some studies analyzing the relationship between taxi activities and built environment. There is an inherent relationship between land-use classes and the temporal pattern of taxi pick-up/drop-off dynamics [9]. The urban sub-regions have broader internal spatial interactions, while suburban centers are more influential in local traffic, and the revealed city structure shows strong relationships with transportation facilities, which provides new view for transportation optimization and urban management [10]. Except for spatial heterogeneity, the spatiotemporal characteristic is also verified in taxi demand [26]. The taxi intends to search for passengers nearby the area where the last passenger drop-off. In addition, the effectiveness of facilities may be different in different time [25].

However, most of the research efforts to analyze taxi ridership has been exerted in a geographical perspective, and few studies have combined spatial and temporal influence quantitatively. Based on traditional GWR, the GTWR add temporal non-stationarity into geographical regression, so that it can take both temporal and spatial variation into consideration. The GTWR is adopted in this study to analyze the relationship between taxi ridership and built environment considering spatiotemporal influence.

The structure of this paper is shown as follows. Chapter 2 provides the process of data preparation. Chapter 3 presents the methodology and basic framework of GTWR. Chapter 4 describes and compares model results, including OLS, GWR and GTWR models. The regression coefficients of GTWR have been analyzed temporally, and the results of GWR have been analyzed spatially. Finally, chapter 5 conclude the findings of the paper and recommend future research directions.

## 2. Data Preparation

Hong Kong Island is chosen as study area, with several economic centers. The taxi records including driver ID, Date, Time, Latitude, Longitude, Flag Down status etc., are collected by GPS devices equipped in taxicab, at 30 -s intervals. The GPS survey data was collected in April 10~16, 2009 inclusively.

### 2.1. Data Processing

Before the work, there are a few steps to process raw GPS records and eliminate noise data: (1)The study area is Hong Kong Island, so the records with position exceeds the latitude and longitude coordinate range should be regarded as the outlier and eliminated. As a result, there are $2,075,792$ taxi records within Hong Kong Island are extracted. (2) The GPS device equipped in taxicab has a data-collected rate as 30 s, so the maximum records a driver can generate in a day is about 2280 ( 86400 s per day/ $30 \mathrm{~s}=2280$ ). If a taxi has number of daily records higher than 2280, these records in that day should be eliminated. (3) The local datum HK80 will be adopted in the study, so we need to transfer the latitude and longitude of WGS84 into HK80 coordinate system by map projection and coordinate system conversion.

### 2.2. Extracting taxi trips

The taxi trajectory can be divided into the load segment and the non-load segment as in figure 1. The trips without complete origin and destination points in the period will be filtered from valid trips. The trip origins can be extracted as the pick-up ones where the Flag Down status changes from 0 to 1 , i.e. from "unloaded segment" to "loaded segment" and vice versa for the destinations [30]. As a result, there are 5,4537 trips in total, with trip time in $0 \sim 90$ minute and trip length in $0 \sim 30 \mathrm{~km}$, the characteristics of trip length and trip time are shown in figure 2.


Figure 1. Taxi status transition diagram


Figure 2. Frequency distribution diagram of trip length and trip time

### 2.3. Grid decomposition

There have been considerable efforts to analyze taxi ridership using GPS data, and most of them focused on large and populous spatial units such as the census tract and hot clusters. However, taxis' travel pattern is highly flexible, and these presumed ridership feature within a big unit may not reflect the actual circumstances of taxi activities [29]. Hence, the smaller and homogenous spatial grid cell is preferred in this study. We intend to divide Hong Kong Island into units with common size.

The size of grid cell is determined carefully. The trip length more than 200 m accounting for $98 \%$, and the trip length more than 500 accounting for $95 \%$. Even though smaller size help to provide detailed spatial pattern of taxis, it would also cause much tedious work. Besides, too small units contain limited geographical information. To get more accurate taxi distribution pattern and to accommodate enough geographical information, the grid size of 500 m is adopted.

### 2.4 Geographical information

There are two routes of MTR in Hong Kong Island: Hong Kong Island line from Kennedy Town to Chai Wan, as well as the South Island line from Admiralty to South Horizons, located in the north side of island and north to south trend separately. The MTR stations' data is collected through web map, and the POls dataset downloaded from Opens Street Map will also be used as built environment information. After category and selection, twelve categories of POIs in Hong Kong Island have been determined. The category description of POIs is shown in table 1.

Table 1. Description of POIs in HKI

| Variable | Description |
| :--- | :--- |
| PT facilities | Bus platform/stop, tram stop, ferry <br> terminals, interchanges |
| Automobile services | Parking/garage facilities, fuel/gas station, <br> car wash, car repair |
| Government agencies | Police station, government office |
| Companies | Office, commercial building |
| School | School, kindergarten |
| Recreation facilities | Sports center/studio, bar, café, cinema, |
| Medical facilities | Clinic, hospital, pharmacy, veterinary |
| Financial facilities | Bank, ATM, money exchange |
| Shopping facilities | Supermarket, shopping center/mall, <br> greengrocer |
| Food shops | Restaurant, fast food, bakery, deli |
| Life facilities | Post office/box, laundry/tailor, hairdresser, <br> public shower, telephone, charging station, <br> worship place etc. |
| Scenic spots | Viewpoint, picnic site, artwork, museum, <br> theater |
| Accommodations | Residential building, apartment, hotel, <br> dormitory |

## 3 Methodology

### 3.1 OLS model

The linear regression is adopted first to analyze the relationship between taxi ridership and influential factors globally. The independent variables contain POI data and Transit Access Time.

## Transit access time (TAT)

A measurement proposed by Ci Yang and Eric J. Gonzales (2014) is adopted to evaluate the accessibility of transit in an area. The measurement combines the estimated walking time a person spends to access the nearest MTR station, as well as the estimated time that person will wait for transit service. The nearest MTR station is determined by K-nearest neighbor (kNN) algorithm. Assumed a walking speed of $0.5 \mathrm{~km} / \mathrm{h}$ (3.1 mph ), the Transit Access Time (TAT) is calculated as follows Error! Reference source not found.:
$T A T=\frac{60 D}{v_{w}}+\frac{60}{f}$
Where $f$ is the frequency of subway dispatches per hour at nearest station; $D$ is distance from a location to its nearest station (mi); and $v_{w}$ is walking speed ( $\mathrm{km} / \mathrm{h}$ ).
Before the GWR or GTWR analysis, we also need to verify if the auto correlation exists in the variables. Tobler (1970) once pointed out that "the first law of geography: everything is related to other things, but the near things are more relevant than the far things", and Moran I value is used to quantify such influence. If the geographic data is positively correlated, then Moran's I value is positive, and vice versa.

### 3.2 GWR model

As taxicab's high flexibility and broad distribution, the global regression cannot reflect the geographical influence. "Brunsdon (1996) is the first to put forward the geographically weighted regression (GWR), where the regression coefficients do not remain fixed over space when cases are geographical locations" [19]. In other words, the regression model at each position is thought of a weight ordinary least squares regression, as shown in equation 4 . The space weight matrix is determined by Euclidean distance and decay-based Gaussian function.
$y_{i}=\beta_{0}\left(u_{i}, v_{i}\right)+\sum_{j} \beta_{j}\left(u_{i}, v_{i}\right) x_{i j}+\epsilon_{i}$
Where, $\left(u_{i}, v_{i}\right)$ is the coordinates of cell i , and $\beta_{0}\left(u_{i}, v_{i}\right)$, $\beta_{j}\left(u_{i}, v_{i}\right)$ is the local regression coefficient in cell i.

### 3.3 GTWR model

However, the taxi ridership is also highly changeable in different time periods, but the GWR cannot consider time variation. Huang et al (2010) proposed the GTWR model, as an extension of traditional GWR model, which can take temporal and spatial non-stationarities at the same time [35]. Be similar to GWR, we construct a space-time weight matrix using space-time distance to measure geographical and temporal variation. The expression of GTWR model is expressed as follows:
$Y_{i}=\beta_{0}\left(u_{i}, v_{i}, t_{i}\right)+\sum_{j} \beta_{j}\left(u_{i}, v_{i}, t_{i}\right) X_{i j}+\varepsilon_{i}$
In the equation above, $t_{i}$ is the observing time of ridership i. Based on weighted least square regression, the parameter can be estimated by minimizing the following equation:
$\sum_{i} W_{\text {iuvt }}\left[y_{i}-\beta_{0}\left(u_{i}, v_{i}, t_{i}\right)-\sum_{j} \beta_{j}\left(u_{i}, v_{i}, t_{i}\right) X_{i j}\right]^{2}$
$W_{\text {iuvt }}$ is the space-time weight of ridership i , which can be expressed as weighting matrix, $W\left(u_{i}, v_{i}, t_{i}\right)=$ $\operatorname{diag}\left(W_{1 u v t}, W_{2 u v t}, \ldots\right)$. Thereafter, the parameter can be estimated by the equation below:
$\beta\left(u_{i}, v_{i}, t_{i}\right)=\left[X^{T} W\left(u_{i}, v_{i}, t_{i}\right) X\right]^{-1} X^{T} W\left(u_{i}, v_{i}, t_{i}\right) Y$
The space-time matrix $W\left(u_{i}, v_{i}, t_{i}\right)$ is based on spacetime distance. In this research, the calculation method of space-time distance can be developed based on the space distance. The difference is that both geographical distance and time distance are included in GTWR. In addition, for time and space should be measured in different units, Huang et al (2010) proposed an ellipsoidal coordination system to measure the "closeness" of different ridership, as shown in Fig.3.
$\mathrm{U}(\lambda)$ and $\mathrm{V}(\lambda)$ are the coordinates of the grid, and $\mathrm{T}(\mu)$ is the time period when the ridership is recorded. The spatiotemporal distance is determined by time period and geographical distance, where the $(\mathrm{U}(\lambda), \mathrm{V}(\lambda))$ plane defines space distance and $\mathrm{T}(\mu)$ defines the time distance.


Figure 1. Definition of space-time coordinate system
Source: Huang et al, 2010 [35]
When T is determined, the $(\mathrm{U}(\lambda), \mathrm{V}(\lambda))$ plane represents the taxi ridership recoded in the same time period. Based on 3D Euclidean distance calculation $d^{2}=x^{2}+y^{2}+z^{2}$, the spatiotemporal distance is calculated by integrating space distance and time distance together, which is expressed as:
$d^{S T}=f\left(d^{S}, d^{T}\right)$
Where $f\left(d^{S}, d^{T}\right)$ represents a function can be defined in different way. The space distance is denoted by $\left(d^{S}\right)^{2}=\left(u_{i}-u_{j}\right)^{2}+$ $\left(v_{i}-v_{j}\right)^{2}$, and time distance is denoted by $\left(d^{T}\right)^{2}=\left(t_{i}-t_{j}\right)^{2}$, hence the space-time distance can be expressed as:
$\left(d^{S T}\right)^{2}=\lambda\left[\left(u_{i}-u_{j}\right)^{2}+\left(v_{i}-v_{j}\right)^{2}\right]+u\left(t_{i}-t_{j}\right)^{2}$
Where $\lambda, \mu$ are parameters, $i$ and $j$ correspond to trip(i) and trip(j). Similar to the weight in GWR, space-time weight can be computed by space-time distance and bandwidth, which is a decay-based Gaussian function as:
$W_{i j}=\exp \left[-\frac{\left(d_{i j}^{S T}\right)^{2}}{h_{S T}{ }^{2}}\right]$
Where $h_{S T}$ is space-time bandwidth, which is positive. Then, the space-time weight can also be expressed as:
$W_{i j}=\exp \left\{-\left[\frac{\lambda\left[\left(u_{i}-u_{j}\right)^{2}+\left(v_{i}-v_{j}\right)^{2}\right]+\mu\left(t_{i}-t_{j}\right)^{2}}{h_{S T}{ }^{2}}\right]\right\}$
$=\exp \left\{-\left[\frac{\left[\left(u_{i}-u_{j}\right)^{2}+\left(v_{i}-v_{j}\right)^{2}\right]}{h_{S}{ }^{2}}+\frac{\left(t_{i}-t_{j}\right)^{2}}{h_{T}{ }^{2}}\right]\right\}$
$=\exp \left\{-\left[\frac{\left(d^{S}\right)^{2}}{h_{S}{ }^{2}}+\frac{\left(d^{T}\right)^{2}}{h_{T}{ }^{2}}\right]\right\}$
$=\exp \left[-\frac{\left(d^{S}\right)^{2}}{h_{S}{ }^{2}}\right] \times \exp \left[-\frac{\left(d^{T}\right)^{2}}{h_{T}{ }^{2}}\right]$
$=W_{i j}^{S} \times W_{i j}^{T}$
Where $W_{i j}^{S}$, $W_{i j}^{T}$ are space and time weight respectively, $h_{S}{ }^{2}=$ ${h_{S T}}^{2} / \lambda$ and ${h_{T}}^{2}={ }_{h_{S T}}{ }^{2} / \mu$. the corrected Akaike information criterion (AIC) is utilized to select parameter $h$.

## 4. Results and discussion

### 4.1 Geographical information

The geographical information is shown in 500 m grid system. The transit access time (TAT) is calculated as shown in figure 4, the white grids represent trip time longer than 30 minutes. The summary of POIs in the grid system is shown in table 2. It's noted that the grid system can accommodate at least one POI data.


Figure 4. Transit access time by walk (minute)
Table 1. Summary of descriptive statistics of POIs

| POIs in cell | Mean | SD | Min | Max |
| :--- | :--- | :--- | :--- | :---: |
| PT facilities | 41.9 | 49.2 | 2 | 276 |
| Automobile services | 7.4 | 6.6 | 1 | 32 |
| Government agencies | 5.9 | 5.3 | 2 | 38 |
| Companies | 10.7 | 13.5 | 1 | 70 |
| School | 10.5 | 8.7 | 1 | 46 |
| Recreation facilities | 12.9 | 25.9 | 2 | 194 |
| Medical facilities | 10.0 | 9.6 | 1 | 40 |
| Financial facilities | 13.1 | 14.7 | 1 | 78 |
| Shopping facilities | 28.9 | 57.7 | 2 | 378 |
| Life facilities | 8.6 | 9.2 | 1 | 48 |
| Scenic spots | 4.3 | 4.9 | 1 | 34 |
| Accommodations | 38.8 | 61.7 | 2 | 332 |

As shown in table 3, all the variables show positive Moran l's value, suggesting the positive autocorrelation. The variables except for company, school, medical facilities, financial facilities and accommodations, show a high level of autocorrelation.

Table 3. Moran I test results of the explanatory variables

| Variable | Moran <br> Index | Expecte <br> d Index | z-score | p-value |
| :--- | :--- | :--- | :--- | :--- |
| PT facilities | 0.3928 | -0.0046 | 8.8571 | 0.0001 |
| Automobile services | 0.3185 | -0.0046 | 7.3251 | 0.0001 |
| Gov. agencies | 0.3531 | -0.0046 | 8.1928 | 0.0001 |
| Company | 0.2494 | -0.0046 | 5.8838 | 0.0003 |
| School | 0.2064 | -0.0046 | 4.7059 | 0.0002 |
| Recreation facilities | 0.3674 | -0.0046 | 9.4445 | 0.0001 |
| Medical facilities | 0.2504 | -0.0046 | 5.8233 | 0.0001 |
| Financial facilities | 0.2486 | -0.0046 | 5.8877 | 0.0005 |
| Shopping spots | 0.4472 | -0.0046 | 11.0004 | 0.0001 |
| Life facilities | 0.3906 | -0.0046 | 8.7877 | 0.0001 |
| Scenic spots | 0.3297 | -0.0046 | 7.9507 | 0.0001 |
| Accommodations | 0.2454 | -0.0046 | 5.7008 | 0.0001 |
| TAT | 0.909 | -0.0046 | 20.2944 | 0.0001 |

### 4.2 Spatial influence

In the GWR model, the automobile services, company, recreation facilities, accommodation and transit access time (TAT) show obviously spatial non-stationary. The spatial distribution of the estimated coefficients and the t-stats for the non-stationary variables are shown in Fig.5. The automobile services in the central part of northern shore has positive influence for taxi demand, while company in the central to east part has positive influence. The recreation facilities have positive influence for taxi trips in most parts of HKI, with the west more attractive than the east. The accommodations have negative influence in taxi trips, with the central part severer than the east. This is partly because there are less accommodations located in central part of northern shore, while more taxi trips generated in the central part.

c) Recreation facilities T-test


Figure 5. Estimated coefficient for total trip in GWR

### 4.3 Model comparison

The coefficients estimation of GTWR model are shown in table 4. Based on the range and median, lower and upper quartile value of coefficients, we can speculate the general influence of variables. The mean value of most coefficients is around zero, while the minimum and maximum value show great difference, indicating a broadly possible range of influencing factors.

Table 4. Coefficients estimation of GTWR model

| Variable | Min | Max | Mean | Lower <br> quartile | Upper <br> quartile |
| :--- | :--- | :--- | :--- | :--- | :--- |
| PT facilities | -87.64 | 74.55 | 0.54 | -0.03 | 0.17 |
| Automobile <br> service | -38.85 | 37.99 | 0.66 | -0.10 | 0.30 |
| Governmen <br> tagency | -129.58 | 495.02 | -0.26 | -0.30 | 0.13 |
| Company | -80.42 | 721.82 | 0.76 | -0.15 | 0.32 |
| School | -45.06 | 23.89 | 0.13 | -0.13 | 0.28 |
| Recreation <br> facility | --1210.74 | 61.48 | -0.12 | -0.04 | 0.39 |
| Medical <br> facility | -604.42 | 91.86 | -0.61 | -0.20 | 0.11 |
| Finance <br> facility | -289.27 | 2436.45 | 0.50 | -0.07 | 0.30 |
| Shopping <br> spot | -834.45 | 101.60 | -0.06 | -0.05 | 0.31 |
| Life facility | -207.98 | 351.33 | 0.54 | -0.15 | 0.26 |
| Scenic spot | -770.00 | 804.94 | 0.05 | -0.19 | 0.17 |
| Accommod <br> ations | -13.25 | 6.68 | -0.16 | -0.05 | 0.12 |
| TAT | -2.50 | 15.46 | 0.68 | 0.03 | 0.20 |

As shown in table 5, the adjusted $R$ square of GTWR is improved to 0.9631 from Global regression's 0,643 and GWR's 0.787, indicating the enhanced ability of the independent variables to explain taxi ridership. In addition, the reduction in AIC value verifies that, the GTWR model is superior than the Global regression and GWR under the same dataset.

Table 5. Comparison between the three regression models

| Diagnostics | Global <br> regression | GWR | GTWR |
| :--- | :--- | :--- | :--- |
| AIC | 3100.611 | 3001.429 | 2201.939 |
| $R^{2}$ | 0.666 | 0.833 | 0.9632 |
| Adjusted $R^{2}$ | 0.643 | 0.787 | 0.9631 |

### 4.4 Temporal variation



Figure 6. Temporal variation of taxi ridership
The comparison between forecasting taxi ridership based on the GTWR model and actual ridership in Hong Kong Island is shown in Fig.6, and GTWR obtains relatively good prediction results. There are two peaks existing in a day: mid-day and night. The duration of mid-day peak is longer than night peak, while the night peak is higher.

a) PT facility

b) Transit Access Time


The temporal variation of coefficients is shown in Fig. 7 a~i. Even though the PT and Transit can be opponent to taxicabs, the area with PT facility and high Transit accessibility show positive impact to taxi ridership. The PT facility is likely to attract taxis in most time of a day except for the morning period and area with PT facilities has higher attraction in the early morning and night, when PT service is limited. The TAT also shows positive impact to taxi ridership in a day, especially in the afternoon peak. Hong Kong Island provides more jobs than living apartments, so there may be more trips come from Hong Kong Island in the afternoon. Area with transit accessibility often owns high attraction to population, leading to more travel demand. When transit capacity is saturated in the peak hours, some trips may turn to other modes like taxis. The influence of financial facility is highest in the afternoon peak, suggesting the travel demand generated in the period.


e) Life facility


## h) Recreation facility

Figure 7. The temporal variation of coefficients
The influence of company, school and automobile facility show double peaks and the two peaks of school are earlier than that of company, automobile facility. For company and school, the influence of the second peak is stronger, indicating the afternoon travel demand is larger than the morning. The lower influence of automobile facility during the two peaks, suggesting drivers are less likely to go to automobile service when there is high travel demand. The life facility, scenic spot and recreation facility show obvious double peak, with a highest peak in the noon, and the first two factors have positive impact to taxi ridership in peak hours. The recreation facility is likely to attract taxis in the night, and unlikely to attract taxis in work hours.

## 5. Conclusion

With the development of GPS technology, it's possible to explore the taxi characteristics in a large dataset. Taxicab as an indispensable part in transport system, it provides convenient point to point service with 24-hour available. The OLS, GWR, GTWR are adopted one by one, and the latter one always performs better than the former, indicating both time and space are important to taxi ridership.

The GWR considering spatial difference, improves model precision of OLS and visualize the spatial influence of built environmental factors. Besides, the GTWR enhances model accuracy of GWR, and visualizes the temporal variation of built environmental coefficients. In fact, the GTWR can visualize both temporal and spatial variation. The Point of Interests information and Transit Access Time in Hong Kong Island are used as built environment variables. The study area is divided into grid-based system with uniform size square and two kinds of size are adopted. The 500 m grid system is developed to accommodate enough geographical information to support regression analysis.

By analyzing the relationship between built environment and taxi ridership, we find the spatial influence of automobile facility, company, recreation facility and accommodation are obvious. For time variation, the Transit Accessibility and Financial facility show single peak, indicating the surrounding area has a high demand for taxis in the afternoon, while the area with PT facility is likely to attract taxis when PT service is limited. The double peaks of school, company and automobile facility correspond to the taxi demand variation in Hong Kong Island. Life facility and scenic spots are likely to attract taxi ridership in the morning and noon, while recreation facility seems to attract taxis in the mid-night.

Based on the findings, we can find out effective traffic measurements and policies to manage taxi operation considering time and space variation and traffic authorities can make targeted measures to balance the spatial and temporal taxi ridership. However, there are some shortcomings in the study. We only include fixed factors as independent variables, the temporal variable such as human mobility is not included. Besides, the study area is Hong Kong Island, which is a small area to do research, containing limited information and taxi trips, so that the results may be not suitable to other parts of Hong Kong. In the next time, we can enlarge the study area to the whole area of Hong Kong, to make the findings more representative.

## Acknowledgements

I would like to thank Lecturer Wong, C.P of the Department of Civil Engineering at The University of Hong Kong, who offered the GPS data and instructions for me.

## References

[1] Szeto, W. Y., Wong, R. C. P., Wong, S. C., \& Yang, H. (2013). A time-dependent logit-based taxi customer-search model. International Journal of Urban Sciences, 17(2), 184-198.
[2] Wong, R. C. P., Szeto, W. Y., \& Wong, S. C. (2015). A two-stage approach to modeling vacant taxi movements. Transportation Research Part C: Emerging Technologies, 59, 147-163.
[3] Tang, J., Jiang, H., Li, Z., Li, M., Liu, F., \& Wang, Y. (2016). A two-layer model for taxi customer searching behaviors using GPS trajectory data. IEEE transactions on intelligent transportation systems, 17(11), 3318-3324.
[4] Wong, R. C. P., Szeto, W. Y., \& Wong, S. C. (2014). A cellbased logit-opportunity taxi customer-search model. Transportation Research Part C: Emerging Technologies, 48, 84-96.
[5] Wong, R. C. P., Szeto, W. Y., \& Wong, S. C. (2014). Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: Modeling methodology and policy implications. Transport Policy, 33, 73-81.
[6] Zhang, D., Sun, L., Li, B., Chen, C., Pan, G., Li, S., \& Wu, Z. (2014). Understanding taxi service strategies from taxi GPS traces. IEEE Transactions on Intelligent Transportation Systems, 16(1), 123-135.
[7] Hu, X., An, S., \& Wang, J. (2014). Exploring urban taxi drivers' activity distribution based on GPS data. Mathematical Problems in Engineering, 2014.
[8] Cai, H., Zhan, X., Zhu, J., Jia, X., Chiu, A. S., \& Xu, M. (2016). Understanding taxi travel patterns. Physica A: Statistical Mechanics and its applications, 457, 590-597.
[9] Pan, G., Qi, G., Wu, Z., Zhang, D., \& Li, S. (2012). Land-use classification using taxi GPS traces. IEEE Transactions on Intelligent Transportation Systems, 14(1), 113-123.
[10]Liu, X., Gong, L., Gong, Y., \& Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43, 78-90.
[11]Austin, D., \& Zegras, P. C. (2012). Taxicabs as public transportation in Boston, Massachusetts. Transportation research record, 2277(1), 65-74.
[12]Li, M., Dong, L., Shen, Z., Lang, W., \& Ye, X. (2017). Examining the interaction of taxi and subway ridership for sustainable urbanization. Sustainability, 9(2), 242.
[13]Tang, J., Liu, F., Wang, Y., \& Wang, H. (2015). Uncovering urban human mobility from large scale taxi GPS data. Physica A: Statistical Mechanics and its Applications, 438, 140-153.
[14]Peng, C., Jin, X., Wong, K. C., Shi, M., \& Liò, P. (2012). Collective human mobility pattern from taxi trips in urban area. PloS one, 7(4), e34487.
[15]Wang, H., Liu, F., \& Tang, J. (2018). Exploring Intra-Urban Travel Mobility using Large-Scale Taxi Global Positioning System Trajectories. International Journal of Vehicle Structures \& Systems (IJVSS), 10(2).
[16]Chen, C., Zhang, D., Li, N., \& Zhou, Z. H. (2014). B-Planner: Planning bidirectional night bus routes using large-scale taxi GPS traces. IEEE Transactions on Intelligent Transportation Systems, 15(4), 1451-1465.
[17]Zhang, K., Sun, D., Shen, S., \& Zhu, Y. (2017). Analyzing spatiotemporal congestion pattern on urban roads based on taxi GPS data. Journal of Transport and Land Use, 10(1), 675-694.
[18]Cui, J., Liu, F., Janssens, D., An, S., Wets, G., \& Cools, M. (2016). Detecting urban road network accessibility problems using taxi GPS data. Journal of Transport Geography, 51, 147157.
[19]Brunsdon, C., Fotheringham, S., \& Charlton, M. (1998). Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3), 431-443.
[20]Du, H., \& Mulley, C. (2006). Relationship between transport accessibility and land value: Local model approach with geographically weighted regression. Transportation Research Record, 1977(1), 197-205.
[21]Yang, L. (2019). Evaluating the urban land use plan with transit accessibility. Sustainable cities and society, 45, 474-485.
[22]Chow, L. F., Zhao, F., Liu, X., Li, M. T., \& Ubaka, I. (2006). Transit ridership model based on geographically weighted regression. Transportation Research Record, 1972(1), 105114.
[23]Cardozo, O. D., García-Palomares, J. C., \& Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. Applied Geography, 34, 548-558.
[24]Dziauddin, M. F., Powe, N., \& Alvanides, S. (2015). Estimating the effects of light rail transit (LRT) system on residential property values using geographically weighted regression (GWR). Applied Spatial Analysis and Policy, 8(1), 1-25.
[25]Li, B., Cai, Z., Jiang, L., Su, S., \& Huang, X. (2019). Exploring urban taxi ridership and local associated factors using GPS data and geographically weighted regression. Cities, 87, 68-86.
[26]Tang, J., Gao, F., Liu, F., Zhang, W., \& Qi, Y. (2019). Understanding spatio-temporal characteristics of urban travel demand based on the combination of GWR and GLM. Sustainability, 11(19), 5525.
[27]Bao, J., Liu, P., Qin, X., \& Zhou, H. (2018). Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data. Accident Analysis \& Prevention, 120, 281-294.
[28]Ma, X., Zhang, J., Ding, C., \& Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. Computers, Environment and Urban Systems, 70, 113-124.
[29]Nam, D., Hyun, K., Kim, H., Ahn, K., \& Jayakrishnan, R. (2016). Analysis of grid cell-based taxi ridership with large-scale GPS data. Transportation Research Record, 2544(1), 131-140.
[30]Zheng, L., Xia, D., Zhao, X., Tan, L., Li, H., Chen, L., \& Liu, W. (2018). Spatial-temporal travel pattern mining using massive
taxi trajectory data. Physica A: Statistical Mechanics and its Applications, 501, 24-41.
[31]Yang, C., \& Gonzales, E. J. (2014). Modeling taxi trip demand by time of day in New York City. Transportation Research Record, 2429(1), 110-120.
[32]Qian, X., \& Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. Applied Geography, 59, 31-42.
[33]Wang, F., \& Ross, C. L. (2019). New potential for multimodal connection: exploring the relationship between taxi and transit in New York City (NYC). Transportation, 46(3), 1051-1072.
[34]Jiang, S., Guan, W., He, Z., \& Yang, L. (2018). Exploring the intermodal relationship between taxi and subway in Beijing, China. Journal of Advanced Transportation, 2018.
[35]Huang, B., Wu, B., \& Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. International Journal of Geographical Information Science, 24(3), 383-401.

