Cities 101 (2020) 102689

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Space-time dynamics of cab drivers' stay behaviors and their relationships with built environment characteristics

Pengxiang Zhao[a], Yang Xu[b,c,*], Xintao Liu[b], Mei-Po Kwan[d,e,f]

[a] Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland
[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[c] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China
[d] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong
[e] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong
[f] Department of Human Geography and Spatial Planning, Utrecht University, 3584 CB Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Taxi GPS trajectories
Stay behaviors
Spatiotemporal analysis
Built environment
Geographically weighted regression

ABSTRACT

Understanding cab drivers' stay activities is essential for planning and managing certain urban facilities. This study analyzes cab drivers' stay behaviors using a taxi GPS trajectory dataset collected in Wuhan, China. By extracting cab drivers' stay activities from the dataset, we measure the activity frequency at the level of traffic analysis zones (TAZs) and examine their spatiotemporal dynamics. We then derive several built environment indicators and assess their associations with these activities using ordinary least squares regression (OLS) and geographically weighted regression (GWR) models. According to the results, the stay frequency decays dramatically over the TAZs, indicating that these activities tend to be concentrated in particular areas of the city. The rates of decay, as reflected by the rank-size and power-law distributions, are similar on weekdays and weekends. Cab drivers' stay activities exhibit similar spatial patterns during the same period on weekdays and weekends. The adjusted R-squared of OLS is 0.742 for weekdays and 0.676 for weekends, which suggests a close relationship between stay activities and built environment characteristics. The GWR models further reveal the spatial variations of the activity-environment linkage across the study area. The study provides useful insights that support future urban design and transport planning.

1. Introduction

As one of the key components of urban transportation systems, taxis provide an all-weather, convenient, and personalized travel service for urban residents. Improving the efficiency of taxi services based on the knowledge of cab drivers' behavioral patterns is essential for transport planning and management of urban facilities (Kang & Qin, 2016). When and where taxis tend to stay are important but often understudied aspects of drivers' behaviors. It is necessary to investigate cab drivers' stay behaviors for several reasons. First, cab drivers' stay behaviors reflect specific types of activities that are linked to the spatial configuration of urban facilities. For instance, many vacant taxis queue up in specific areas for purposes including refueling or dining. Improving our knowledge of these behaviors could help cities better satisfy the needs of taxi drivers. Second, cab drivers' stay behaviors also reflect their operation behaviors (e.g., waiting for passengers), which in turn convey useful information about taxi supply and the areas where potential demand might be high.

Previous studies on cab drivers' activities mainly rely on data collected through travel surveys, questionnaires or interviews. Most of these studies focus on analyzing driving behavior without paying enough attention to drivers' stay activities (Kalhori, Foroughinia, & Ziapour, 2017; Ma, Yan, Huang, & Abdel-Aty, 2010; Newnam, Mamo, & Tulu, 2014; Shi, Tao, Li, Xiao, & Atchley, 2014; Zhang et al., 2018). The few studies that analyze cab drivers' dining behavior are based on questionnaire and survey data (Song et al., 2012; Zhang, Zheng, Lu, & Chai, 2009). This has led to a scarcity of knowledge on cab drivers' stay behavior, an important dimension of their activities beyond movements along road networks.

Overall, studies on cab drivers' stay behavior are sparse. As data collection of surveys and questionnaires are costly and time consuming, they usually cover small sample sizes, which may not be able to delineate cab drivers' behavior at a large scale. In recent years, the explosive growth of GPS-based taxi trajectories has provided massive opportunities for analyzing cab drivers' behaviors. Compared to traditional data based on travel surveys or questionnaires, GPS-based taxi

* Corresponding author.
E-mail address: yang.ls.xu@polyu.edu.hk (Y. Xu).

https://doi.org/10.1016/j.cities.2020.102689
Received 3 June 2019; Received in revised form 5 March 2020; Accepted 8 March 2020

trajectory data are collected by tens of thousands of taxis simultaneously at low costs and high efficiency. A strand of studies on quantifying cab driver's mobility patterns based on GPS taxi trajectory data is continuously emerging. For instance, cab drivers' operational behavior (Gao, Jiang, & Xu, 2018; Liu, Andris, & Ratti, 2010; Manley, Addison, & Cheng, 2015; Tang et al., 2016), refueling behavior (Niu, Liu, Fu, Liu, & Lang, 2016; Zhang, Yuan, Wilkie, Zheng, & Xie, 2015), and shift handover behavior (Sun & Yu, 2014; Zhang et al., 2015) have been investigated based on GPS taxi trajectory data. Despite the various types of cab drivers' behaviors examined in these studies, little effort has been devoted to systematically investigating cab drivers' stay behaviors.

In this paper, we use GPS-based taxi trajectory data collected in Wuhan, China to investigate three aspects of cab drivers' stay behaviors, namely, the regularities of cab drivers' stay behaviors across space, the spatiotemporal variations of stay frequency, and the relationship between stay frequency and certain built environment characteristics. Our specific research questions are as follows: (1) What are the key characteristics of the spatial distributions of these stay activities? (2) How do the spatial patterns of such activities change over time? and (3) How are cab drivers' collective stay patterns related to the underlying built environment? To answer these questions, we apply a trajectory prepossessing method to identify taxis' stay points, from which the intensity and evolution of cab drivers' stay activities are mapped and analyzed. Two models are utilized to measure stay activity frequency in this study, namely the rank-size distribution and the frequency probability distribution. Second, the spatial distributions of stay activities during different periods are analyzed. Finally, we apply ordinary least squares regression (OLS) and geographically weighted regression (GWR) models to investigate the relationships between stay activities and certain built environment characteristics.

2. Related work

2.1. Quantifying cab drivers' behavioral patterns

The increasing availability of GPS-based taxi trajectory data has led to an enormous amount of studies on human behavioral patterns, including both passengers and cab drivers (Liu, Kang, Gao, Xiao, & Tian, 2012; Tang, Liu, Wang, & Wang, 2015; Zhang, Xu, Tu, & Ratti, 2018; Zhao, Qin, Ye, Wang, & Chen, 2017). In this section, we mainly focus on studies on cab drivers' behavioral patterns based on GPS taxi trajectory data. The related studies are discussed as follows. For instance, Liu et al. (2010) explored cab drivers' operational patterns by analyzing their GPS digital trajectories. The results help us better understand cab drivers' behavior and motivational intelligence. Kang and Qin (2016) systematically studied the patterns of taxi demand and supply in cities using a nonnegative matrix factorization method based on a monthly collection of GPS taxi trajectory data in Wuhan, China. Zhang and Wang (2016) investigated a method to infer passenger denial behavior of taxi drivers from massive taxi trajectory data in Beijing. The results indicate that high-income taxi drivers tend to deny passengers in some situations in order to choose the destinations they prefer and make more profit. Wang, Qin, Chen, and Zhao (2018) developed a trajectory clustering method based on the edit distance and hierarchical clustering to detect anomalous trajectories, and then identify cab drivers' fraud behavior. Tang, Liang, Zhang, Huang, and Liu (2018) proposed a probabilistic model to infer driving trajectories from taxi GPS data while considering driving behavior patterns (e.g. stays). Zong, Wu, and Jia (2018) examined how the external factors (land use, traffic conditions and road grade) and internal factors (previous pick-up experience) influence taxi drivers' cruising location choice patterns.

With respect to the optimization process, cab drivers' operational choices and driving behaviors are considered. Li et al. (2011) used historical taxi trajectory data to study cab drivers' passenger-finding behavior, aiming to discover both efficient and inefficient passenger-finding strategies. Yuan, Zheng, Zhang, Xie, and Sun (2011) presented a recommendation system for both cab drivers and passengers through analyzing cab drivers' pick-up behavior patterns and passengers' mobility patterns using GPS taxi trajectory data. A two-layer decision framework was proposed to model taxi drivers' passenger-finding behaviors within urban areas using GPS taxi trajectory data in Beijing. The results can help understand taxi drivers' passenger searching decisions (Tang et al., 2016). Aiming to understand how a taxi driver seeks passengers in an unknown environment, Liu, Wang, Liu, and Krishnan (2015) proposed a framework to explore taxi drivers' route choice behavior by studying how they gather and learn information in an uncertain environment using their social network. It is found that new drivers prefer to cruise. Manley et al. (2015) investigated both the shortest path and anchor-based perspectives on route choice behavior based on an empirical analysis using appropriately 700,000 minicab routes in London. Lu, Lai, Ye, Liang, and Yuan (2017) developed a visual analytic system to study route choice behavior, which is applied to a real taxi GPS trajectory dataset in Beijing. It is observed that the majority of cab drivers prefer to select the main ring road without branches, while different choices appear among a few drivers. Using a taxi trajectory dataset with 76 million points collected in Shenzhen, Yang, Kwan, Pan, Wan, and Zhou (2017) developed a method, which extracts and incorporates the driving experience of cab drivers into a path-finding algorithm to optimize the routes between trip origins and destinations. Gao et al. (2018) proposed a novel method to optimize taxi driving strategies with the goal of maximizing global profit based on reinforcement learning.

Some previous studies have examined cab drivers' stationary behaviors that include refueling and dining events. To sense individual refueling behavior and citywide gas consumption in real-time, Zhang, Yuan, et al. (2015) developed a complete data-driven system to explore drivers' refueling behavior from the perspective of taxi mobility in urban areas. Niu et al. (2016) designed a method for detecting taxi refueling events using GPS trajectory data, which jointly takes into account stay times, trace angles, location sequences and refueling cycles of the vehicles. The extracted refueling events are used to optimize gas station site selection. Zhao, Liu, Kwan, and Shi (2018) developed a method based on support vector machine (SVM) to detect cab drivers' dining events from GPS taxi trajectory data and further displayed cab drivers' dining behavioral patterns in space and time. From the above-related literature review, it can be seen that the cab drivers' stay behaviors have been paid little attention to. Although several studies investigated cab drivers' stay behaviors, they focus on few specific types of activities (e.g. refueling or dining). However, cab drivers' stay behaviors could be linked to various activity purposes. Hence a comprehensive study on cab drivers' stay behaviors is necessary.

2.2. Understanding the regularities of human mobility

The regularities of human mobility is important to the understanding of urban dynamics. Many studies have been conducted to depict the regularities of different aspects of human mobility, such as travel distance (Gonzalez, Hidalgo, & Barabasi, 2008; Liang, Zhao, Dong, & Xu, 2013), travel time (Giannotti et al., 2011; Kung, Greco, Sobolevsky, & Ratti, 2014; Wang, Pan, Yuan, Zhang, & Liu, 2015), spatial visitation frequency (Chen, Chen, Ni, & Li, 2018; Song, Koren, Wang, & Barabási, 2010; Zhang, Xu, et al., 2018; Zheng & Zhou, 2017), among others. A bunch of models have been used to measure the observed regularities. Among them, rank-size distribution and probability distribution are often applied in human mobility studies.

In human mobility studies, travel frequency is commonly used to represent size. Hence, a rank-size distribution in human mobility normally reflects the relationship between travel frequency and the ordered ranks of all the locations in the study area, which can be fitted by a power function, an exponential function or a power law/exponential function with cut-off. For instance, Song et al. (2010) investigated the scaling properties of human mobility using mobile phone data and found that the visitation frequency of users follows a power law. Hasan, Zhan, and Ukkusuri (2013) explored the relationship between the visiting frequency of a place and the rank of the place using social media check-in data. They discovered that the visiting frequency follows a truncated power law. Zheng and Zhou (2017) studied the scaling laws of visitation frequency based on the rank-frequency law using GPS-based taxi trajectory data. The results indicated that the scaling laws of visitation frequency follow exponential distribution. Chen et al. (2018) analyzed the scaling laws of visitation frequencies using the on-demand ride services data from DiDi. It is demonstrated that the visitation frequency and the associated rank follows an exponential function across different scales. Zhang, Xu, et al. (2018) compared the rank-size distributions for incoming and outgoing trips on weekdays and weekends using smart card data and GPS-based taxi trajectory data. It is reported that the distributions can be fitted by exponential functions.

Many studies have investigated the statistical distributions of travel distance. Noulas, Scellato, Lambiotte, Pontil, and Mascolo (2012) investigated human mobility patterns in several metropolitan cities using check-in data of Foursquare users. The results indicated that the distribution of trip distance is well approximated by a power law. Calabrese, Diao, Di Lorenzo, Ferreira Jr, and Ratti (2013) explored individual mobility patterns from mobile phone data and found that trip length distribution is well approximated by a power law with an exponential cut-off. Wang et al. (2015) conducted a comparative analysis of intra-city human mobility using taxi trajectory data from five cities. The results suggest that the distributions of the trip distance have similar trends and follow the exponential function instead of the power law function. Riascos and Mateos (2017) studied human mobility patterns of people visiting specific locations in two big cities using Foursquare check-in data. Through analyzing the probability distributions of time and distance between two successive user check-ins, it is found that they follow power laws. The work by Zhao, Kwan, and Qin (2017) indicated that taxi is a more preferable travel mode for medium-distance trips through examining the probability distributions for human trips using taxi trajectory data.

2.3. The relationship between human movements and the built environment

The relationship between human movements and the built environment has always been an important research topic in urban studies. A growing body of research suggests that the built environment could impact on the travel behavior of people (Etminani-Ghasrodashti & Ardeshiri, 2015; Ewing & Cervero, 2010; Handy, Cao, & Mokhtarian, 2005; Xu et al., 2019; Xu, Belyi, Bojic, & Ratti, 2018; Zhang, Xu, et al., 2018; Zheng & Zhou, 2017). For example, Liu et al. (2012) investigated the association of human trips with different land use features using GPS taxi trajectory data. It is found that land use types have a great influence on daily human trips. The study by Etminani-Ghasrodashti and Ardeshiri (2015) explored how the built environment shapes human travel behaviors through examining the impact of the built environment on travel behaviors using social survey data. The work by Shen, Liu, and Chen (2017) analyzed the driving factors that affect the spatial distribution of passengers' pickup and drop-off locations based on taxi traces data. The results suggest that population density and transportation density are related to the population distribution. Zheng and Zhou (2017) examined the association between human travel and the built environment using taxi GPS traces data through analyzing the relationship between trip frequency and points of interest (POI) density. Lee and Kwan (2018) developed a hierarchical approach to accurately predict human physical activities including travel modes (e.g., biking and riding in a vehicle) based on built-environment context using machine learning techniques and GPS and accelerometer data. Using a one-year taxi ridership log data in Washington D.C., Yang et al. (2018) developed regression models to investigate the relationship between travel demand and land use characteristics as well as transit supplies. Zhang, Xu, et al. (2018) studied the association of travel demand with built environment measure as well as socio-economic factors based on smart card data and taxi trajectory data. The results enhance our understanding of the interactions between human trips and the built environment. Li, Cai, Jiang, Su, and Huang (2019) examined the relationship between explanatory variables obtained from POI and demographic data and urban taxi ridership using taxi GPS data. Tang, Gao, Liu, Zhang, and Qi (2019) investigated the spatio-temporal characteristic of travel demand with GPS-based taxi trajectory data from New York City and explored the underlying affecting variables with the geographically weighted regression (GWR) and the generalized linear models.

From the above review, it is apparent that human trips have a close association with the built environment. However, studies that focused on how the built environment influences cab drivers' stay behaviors are still limited. This study intends to fill this gap through analyzing the relationships between cab drivers' stay behaviors and the built environment using OLS regression and GWR based on GPS-based taxi trajectory data collected in Wuhan.

3. Methodology

3.1. Study area and datasets

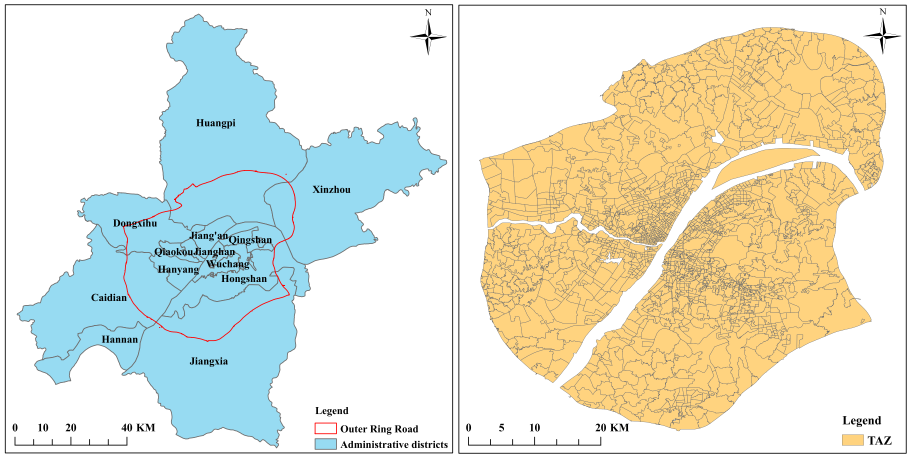
Wuhan — the capital city of Hubei Province — is a densely populated metropolitan area in central China. It has an area of 8594 km^2 and a population of about 11 million as of 2017. The city is an economic, education, and cultural hub in central China. In this study, we focus on the area within the Wuhan Outer Ring Road, where the majority of taxi trips occurred (Fig. 1a). Traffic analysis zone (TAZ) is selected as the analysis unit in this work. TAZs are normally constructed based on the socioeconomic data of census blocks, which are commonly-used units of geography in urban and transportation planning.

In this research, a trajectory dataset collected from more than 7400 taxis over 7 days (March 2nd, 2015 to March 8th, 2015) is used. The taxi trajectory data were provided by a taxi company in Wuhan. The dataset records the information of each taxi including timestamps, location, direction, speed, passenger status of the vehicles and status of taxi engine (i.e. running [ON] or not [OFF]). The sampling rate is approximately 60 s, which leads to 12.4 million GPS records per day.

3.2. Data preprocessing

Before proceeding with the analysis, data preparation is necessary. Data preparation mainly focuses on cleaning the raw data, extracting cab drivers' stay points as well as passengers' pick-up and drop-off points. Specifically, it is convenient to better understand cab drivers' stay activities by analyzing pick-up and drop-off points. Data cleaning aims to remove the false records in the raw trajectory data due to data loss or redundant fields. Pick-up and drop-off points are extracted based on the changes of taxi status (i.e., from “occupied” to “vacant” or vice versa). Stay points are extracted following the method in (Zhao, Liu, et al., 2018). According to previous studies, stay point actually corresponds to a small area a cab driver stayed over a certain period, which is denoted as a virtual location by calculating the average values of the longitudes and latitudes for the trajectory points within the stay area.

The process of identifying stay points is as follows. First, considering that a taxi's engine is not running (OFF) after the cab driver parked the taxi, the trajectory points with the engine status of OFF are extracted. Then, cab drivers' stay points are further identified from the extracted trajectory points. Specifically, if a cab driver stayed at a location over a certain period of time, a stay activity normally occurs. Hence, stay activity is determined by two spatial and temporal constraints, namely the stay time and moving distance. In this study, the thresholds of stay time and moving distance are set as 3 min and 500 m, respectively. Following this procedure, 429,201 stay points are successfully extracted



(a) Study area of Wuhan (b) Traffic analysis zones (TAZs)

Fig. 1. Study area and distribution of TAZs.

from the whole dataset for the seven days. The average number of stay points per day is 61,314. Similarly, 3,506,253 pick-up and drop-off points are detected from the trajectory data. The average number of pick-up and drop-off points per day is 500,893.

Fig. 2 illustrates the hourly variation of the cab drivers' stay activities - represented as the frequency of stay points over each one-hour time window - as well as pick-up and drop-off points extracted from the trajectory data during the seven days. As shown in Fig. 2(a), cab drivers' stay behaviors exhibit a high degree of regularity on both weekdays and weekends. The fluctuations, however, can be observed. Specifically, the period 0:00–1:00 presents as the peak of total stay frequency and with a trend of decreasing during 0:00–6:00. Additionally, there are also two local peak periods in the daytime, such as 12:00–13:00 and 16:00–17:00. The highest frequency occurring at 0:00–1:00 is reasonable since cab drivers usually go back home to have a rest during this period. The peak at 16:00–17:00 is probably attributed to taxi handovers. In order to mitigate people's difficulty in taking a taxi in the evening rush hours, Wuhan City promulgates regulations prohibiting taxis from changing shift (handovers) during evening rush hours (17:00–19:00).

We then compare the hourly variations in the frequency of pick-up and drop-off points with that of stay points in Fig. 2(b). It clearly shows that the frequency of pick-up and drop-off points exhibits regularity from Monday to Sunday. The temporal variations in the frequency of taxi trips show similar patterns between weekdays and weekends. However, there are also some differences between weekdays and weekends. For instance, the frequencies of pick-up and drop-off points at 0:00–5:00 on weekends are higher than those on weekdays. In addition, the number of trips between 16:00 and 17:00 is generally lower than that of other periods during the daytime. Interestingly, this time window corresponds to the period when stay frequency is high (Fig. 2a). Such an association may be partly attributed to the work shift of the cab drivers (e.g., taxi handovers).

3.3. Exploring distributions of stay frequency by location

According to Section 2.2, the rank-size distribution and probability distributions have been widely used to depict the characteristics of human mobility patterns. In this research, both types of distributions are used to explore the characteristics of cab drivers' stay frequency by location.

Cab drivers' stay behavior is generally regarded as a highly stochastic process in urban space. We denote F_{t1}, F_{t2}, …, F_{tn} as the sequence of the visited TAZs during period t according to the stay frequency in descending order. Stay frequency is defined as the number of stay points located in the TAZ during a certain period, which reflects the intensity of stay activities across urban locations. For instance, the stay frequency $f(r)_t$ denotes the number of stay points in the r^{th} most stayed location during period t (also known as the size). The distribution of $f(r)_t$ with respect to the rank r is referred to as the rank-size distribution.

Mathematically, a probability distribution expresses the frequency distribution of a random variable. In this research, it is used to denote the probability of stay frequency that occurred in a TAZ.

3.4. Correlating stay activities with built environment characteristics

In order to uncover the relationship between cab drivers' stay frequency and the built environment, we use OLS and GWR models to relate stay frequency with certain built environment variables at the level of TAZ.

3.4.1. Built environment variables

The built environment is normally conceptualized in terms of the three 'Ds': density, design and diversity (Cervero & Kockelman, 1997; Zhao, Kwan, & Zhou, 2018). As shown in Table 1, ten built environment indicators are selected as the variables used in this paper. These built environment indicators are computed based on the points-of-interest (POI) data and road network data of the study area. Specifically, we extract the POIs related to cab drivers' stay behaviors. We further calculate the number of each type of POIs in the TAZs, including the number of restaurants (*NoRt*), the number of shopping malls (*NoSM*), the number of residences (*NoRc*), the number of car repair and maintenance shops (*NoCS*), the number of gas stations (*NoGS*), and the number of public toilets (*NoT*).

The existence of airports or railway stations could also significantly influence the cab drivers' stay behavior (Yang et al., 2018). This is due to the large passenger volumes in these areas, which attract many cab drivers to wait for passengers. There is one airport (Wuhan Tianhe International Airport) and three railway stations (i.e. Wuchang Railway Station, Hankou Railway Station and Wuhan Railway Station) in Wuhan. The existence of an airport (*EoA*) and the existence of a railway station (*EoRS*) are formulated as dummy variables to depict whether or not a certains TAZ contains an airport or a railway station.

To measure the accessibility of the TAZs, road network density (*RND*) is also considered as an independent variable. In this study, road network density is defined as road or street length (in kilometer) divided by the area of a TAZ, which reflects the road network connectivity in the TAZ. In addition, we define a variable, named POI diversity (*POID*), based on the method for entropy computation. The

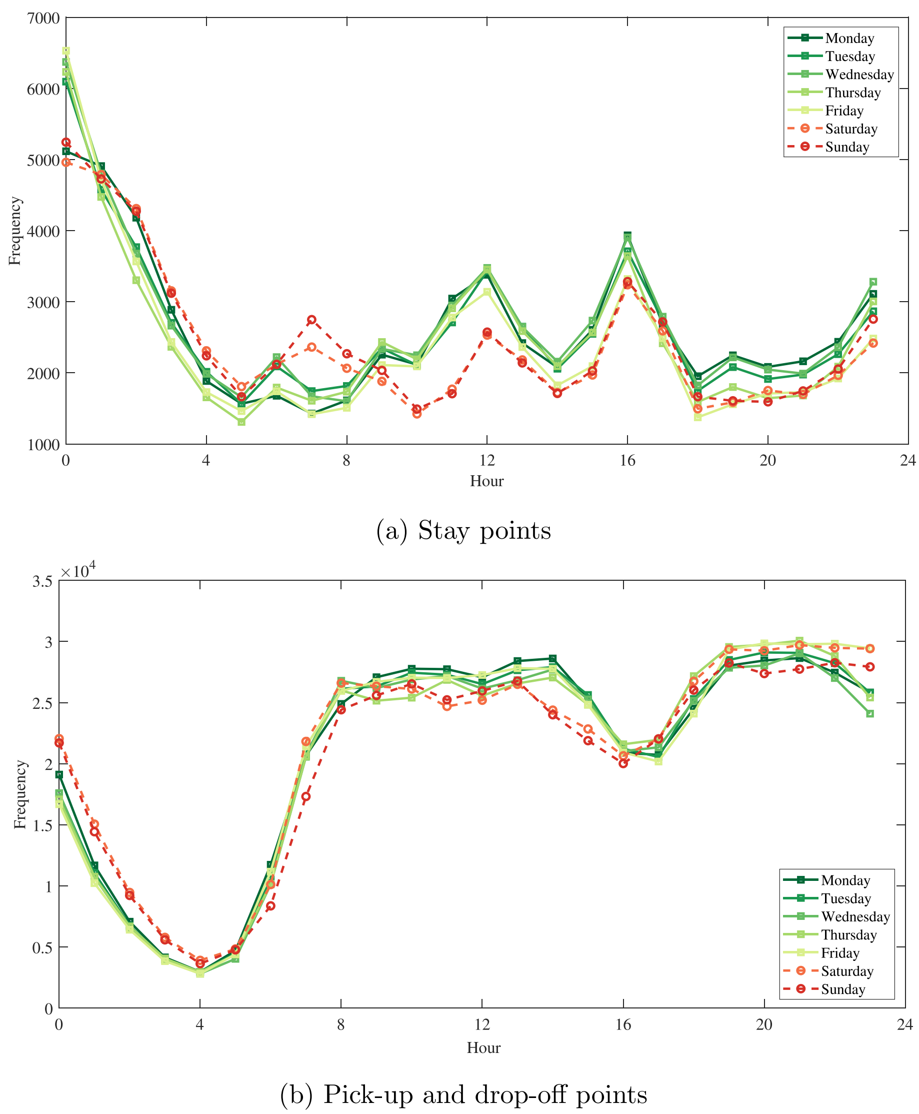


Fig. 2. Hourly variations of stay points and pick-up and drop-off points.

Table 1
Explanations of the abbreviations for the built environment variables.

Variables	Abbreviations
Number of restaurants	*NoRt*
Number of shopping malls	*NoSM*
Number of residences	*NoRc*
Number of car repair and maintenance shops	*NoCS*
Number of gas stations	*NoGS*
Existence of an airport	*EoA*
Existence of a railway station	*EoRS*
Number of public toilets	*NoT*
Road network density	*RND*
POI diversity	*POID*

proportion of the abovementioned six POI types are taken into account in the calculation process. The values of *POID* range from 0 to 1, which measure the spatial heterogeneity of the POIs in each TAZ. The values 0 and 1 correspond to a single type of POI and one with the most diverse POI, respectively. The formula for POI diversity is expressed as follows:

$$POID = -\sum_{i=1}^{n} \frac{p_i \cdot ln(p_i)}{ln(n)} \tag{1}$$

In this formula, p_i represents the proportion of the i^{th} POI type, n is the total number of POI types. On the basis of the independent variables in Table 1, OLS regression and GWR models are used to discover the relationship between stay frequency and the derived built indicators.

3.4.2. OLS regression

Ordinary least squares (OLS) regression is one of the most representative and widely used approaches among statistical methods in urban studies (Chun & Guldmann, 2018; Wang & Mu, 2018; Zhao, Deng, Song, & Zhu, 2013). In this study, the regression takes stay frequency (SF) as the dependent variable; and the independent variables are the above-mentioned built environment indicators. The relationship is examined on a TAZ basis by using OLS regression. The purpose is to analyze the effect of the built environment variables on cab drivers' stay activities. The model is formulated as follows:

$$\begin{aligned} SF = \beta_0 + \beta_1 NoRt + \beta_2 NoSM + \beta_3 NoRc + \beta_4 NoCS + \beta_5 NoGS \\ + \beta_6 EoA + \beta_7 EoRS + \beta_8 NoT + \beta_9 RND + \beta_{10} POID + \varepsilon \end{aligned} \tag{2}$$

where SF represents stay frequency; $\beta_i (0 \le i \le 10)$ denote the regression coefficients; and ε denotes the random error.

3.4.3. Geographically weighted regression

Geographically weighted regression (GWR) is specifically developed to deal with the issue that the relationship between two or more variables may vary over space, which is referred to as spatial nonstationarity (Brunsdon, Fotheringham, & Charlton, 1996; Wang, Lee, & Kwan, 2018). GWR is a localized regression model which allows the estimated parameters to vary over the spatial domain. It also takes spatial autocorrelation into account and thus can mitigate the bias in models caused by the spatial autocorrelaton effect. Hence, compared with the global model in which parameter estimation is fixed for each observation, GWR can be used to address the spatial nonstationarity and spatial autocorrelation of observations and determine whether the underlying process exhibits spatial heterogeneity. In this work, the GWR model is denoted by the following formula:

$$SF_i = \beta_{0i} + \beta_{1i}NoRt_i + \beta_{2i}NoSM_i + \beta_{3i}NoRc_i + \beta_{4i}NoCS_i + \beta_{5i}NoGS_i + \beta_{6i}EoA_i + \beta_{7i}EoRS_i + \beta_{8i}NoT_i + \beta_{9i}RND_i + \beta_{10i}POID_i + \varepsilon_i \tag{3}$$

where i represents the i^{th} TAZ, SF_i stands for the stay frequency of the ith TAZ, β_i is the estimated regression coefficients of the i^{th} TAZ, and ε_i is the error term for TAZ i.

In GWR, spatial structure is considered in the model by applying weights to the data observations. The weights are normally calculated based on a kernel function, which assigns higher weights to observations that are closer to the location (i^{th} TAZ) in space where the model is developed. Specifically, the distances between locations are input to the kernel function to calculate the weights in the setting of GWR. The Gaussian kernel function is the most commonly applied with GWR. The kernel function contains a bandwidth parameter, which determines the spatial range of the kernel. The bandwidth parameter can be set by prior knowledge or optimally estimated from the data. Two methods are widely used for estimating the bandwidth in GWR, namely cross-validation (CV) score and Akaike Information Criterion (AIC). CV score aims at searching for the kernel bandwidth in an iterative process that minimizes the prediction error, while AIC approach is compromise between the goodness of fit of the model and the simplicity of the model. In this research, the spatially adaptive band width values are obtained by using the method that minimizes AIC of regression models.

4. Results

4.1. The regularities of cab drivers' stay activities

In this subsection, we explore the regularities of cab drivers' stay frequency based on the rank-size distribution and probability distribution. Fig. 3 illustrates the rank-size distributions and probability distributions of cab drivers' stay frequency on weekday and weekend at the TAZ level. Stay frequencies on weekdays and weekends are obtained by calculating the average values of frequency on weekdays and weekends respectively. As shown in Fig. 3(a), stay frequency decays dramatically for the first few TAZs on weekdays and weekends, indicating that cab drivers' stay activities in some areas are pretty intensive. In order to compare the decay effects between weekdays and weekends, we further fit the curves in the semi-log plot using the exponential function $y = \frac{a}{b^x}$. The parameter b reflects the rate of decay. The fitted lines for the rank-size distributions on weekdays and weekends are $y = \frac{248.4}{1.004^x}$ and $y = \frac{208.1}{1.003^x}$ respectively. Note that the rates of decay are very close to each other, indicating that the decaying speed is similar between weekdays and weekends.

In Fig. 3(b), we plot the probability distributions of stay frequency for weekdays and weekends. The log-log plot shows that the probability distributions follow a linear trend, indicating that they exhibit a power law distribution. The fitted lines for the probability distributions on weekdays and weekends are $y \sim x^{-1.81}$ and $y \sim x^{-1.72}$ respectively. That means the stay frequency decays faster on weekdays than on weekends. It can also be said that the frequency difference between high-frequency TAZs and low-frequency TAZs on weekday is greater than that of weekend. The faster decay of weekday probably reflects that several TAZs with high stay frequency attracts more cab drivers' stay activities on weekdays than on weekends. For instance, a cab driver may go to the airport or railway stations to wait for the next passenger during weekdays, while he or she may go to Happy Valley (i.e. a theme park) or other places. Basically, citizens have more diverse travel choices on weekends than weekdays. The spatial heterogeneity of cab drivers' stay activities requires to be considered while configuring the urban facilities related to them.

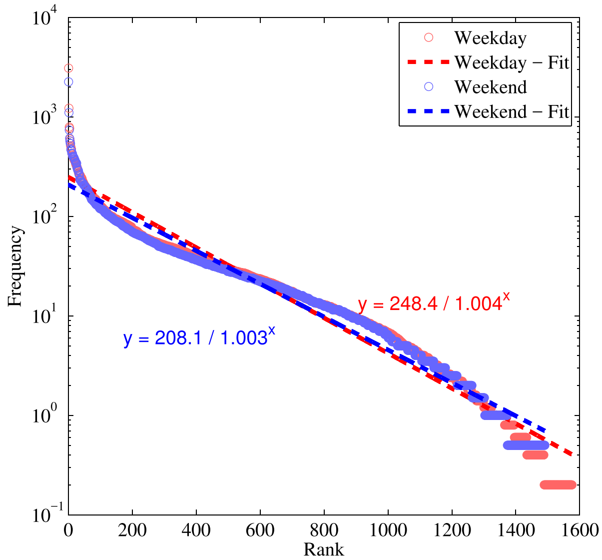
In addition, to understand the temporal dynamics, we further divide a day into six 4-h time periods and perform the same analysis for weekdays and weekends. From Fig. 2(a), we can observe that the "inflection points" of temporal variations mainly occur in the hours of 4, 8, 12, 16, 20. Hence, we select four hours as a period. The decay rate in the rank-size distribution reflects the decay difference. As shown in Fig. 4(a) and (b), it can be observed that the rank-size distributions in various periods display similar trends on weekdays and weekends. In addition, it is noteworthy that the stay frequencies during 0:00–4:00 have the highest decay rate although the temporal differences are very small. Furthermore, we quantitatively compare the decay effects across different periods by calculating the parameter b. Table 2 displays the values of parameter b during the six periods on weekdays and weekends. Note that values of b range from 1.0035 to 1.0042 on weekdays, and from 1.0033 to 1.0041 on weekends. Since the decay rates of different periods are very close, the lines are almost parallel in the semi-log plots. In addition, we can also conclude that there are no significant differences in the regularities of cab drivers' stay frequency between the different time periods on weekdays and weekends.

Furthermore, we examine the cumulative probability distributions and percentage distributions of stay frequency during various periods on weekdays and weekends. Here, the percentage refers to the proportion of the sum of stay frequency meeting a certain criterion (e.g, higher than a certain value) to the sum of all stay frequency values. Fig. 4(c) and (d) show the cumulative probability distributions and the percentage distributions changing with time on weekdays and weekends. It can be seen that some important TAZs capture a high proportion of the whole stay frequency. For instance, the percentage of TAZs with stay frequency more than 50 occupies no more than 10% in all periods. Moreover, the number of TAZs with high stay frequency (e.g, more than 50) in 0:00–4:00 is larger than those in other periods. On the contrary, the minimum quantity is observed for the number of TAZs with high stay frequency in 4:00–8:00. In addition, the cumulative probability distributions in the same period of weekdays and weekends also display no marked difference.

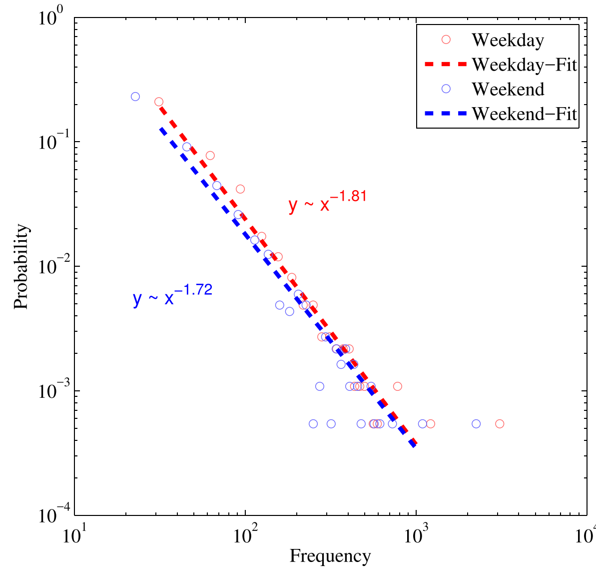
4.2. Spatiotemporal characteristics of cab drivers' stay behavior

In this section, we examine the spatiotemporal characteristics of cab drivers' stay behavior on weekdays (2 March to 6 March in 2018) and weekends (7 March to 8 March in 2018). Through the rank-size distribution analysis, it is found that the top 20% TAZs occupy 76.8% and 76.9% of all cab drivers' stay activities respectively. Hence, we mainly focus on the TAZs within top 20% according to their ranks of stay frequency in the spatiotemporal analysis. Specifically, we first calculate the average values of stay frequency during each 4-h time period (e.g. 0:00–4:00) on weekdays and weekends, respectively. Then, the TAZs within top 20% ranks are visualized in different time periods on weekdays and weekends, as shown in Figs. 5 and 6. Note that different TAZs are depicted by different colors and with uniform rendering strategy across different periods according to the value of the rank.

The average stay frequencies on weekdays and weekends display similar spatial distributions during the same period. Most of the TAZs with high stay frequency during one period on weekdays have high stay frequency in the same period on weekends, although there are small fluctuations in few TAZs. However, the spatial patterns of stay behaviors in different periods display differences. Specifically, the TAZs with high stay frequency during 0:00–4:00 are mainly distributed within the



(a) Rank-size distribution



(b) Probability density distribution

Fig. 3. Distributions of stay frequency on weekdays and weekends.

third ring road, which is on the periphery of the main city of Wuhan as a ring road expressway. By comparison, the top 20% TAZs are more dispersed in space and display expansion towards the TAZs outside the third ring road. We can conclude that cab drivers' stay activities are mainly concentrated in downtown areas at midnight.

We further analyze the dynamics of spatial patterns across different time periods on both weekdays and weekends. As illustrated in Figs. 5 and 6, the top 20% TAZs overall experience few changes over different periods while the TAZs within top 1% and 1%–5% could change from time to time. Additionally, some detailed spatiotemporal distribution characteristics can be revealed from Figs. 5 and 6. For instance, note that the airport displays as a constant area in top 1% with high stay frequency over all periods. It is probably due to the high passenger volume in the airport, which leads to cab drivers' daily activities being mainly focused on these areas.

4.3. Relationship between collective stay patterns and built environment characteristics

In this section, we discuss the relationships between the intensities of cabdrivers' stay activities and the built environment variables. First, the descriptive statistics for all the built environment variables are reported in Table 3. These built environment indicators are used as independent variables in the subsequent regression analysis, and stay frequency is used as the dependent variable. Here, stay frequencies for both weekdays and weekends are examined. Then, a series of OLS and GWR models are used to examine the relationships between stay frequency and built environment characteristics.

4.3.1. OLS regression results

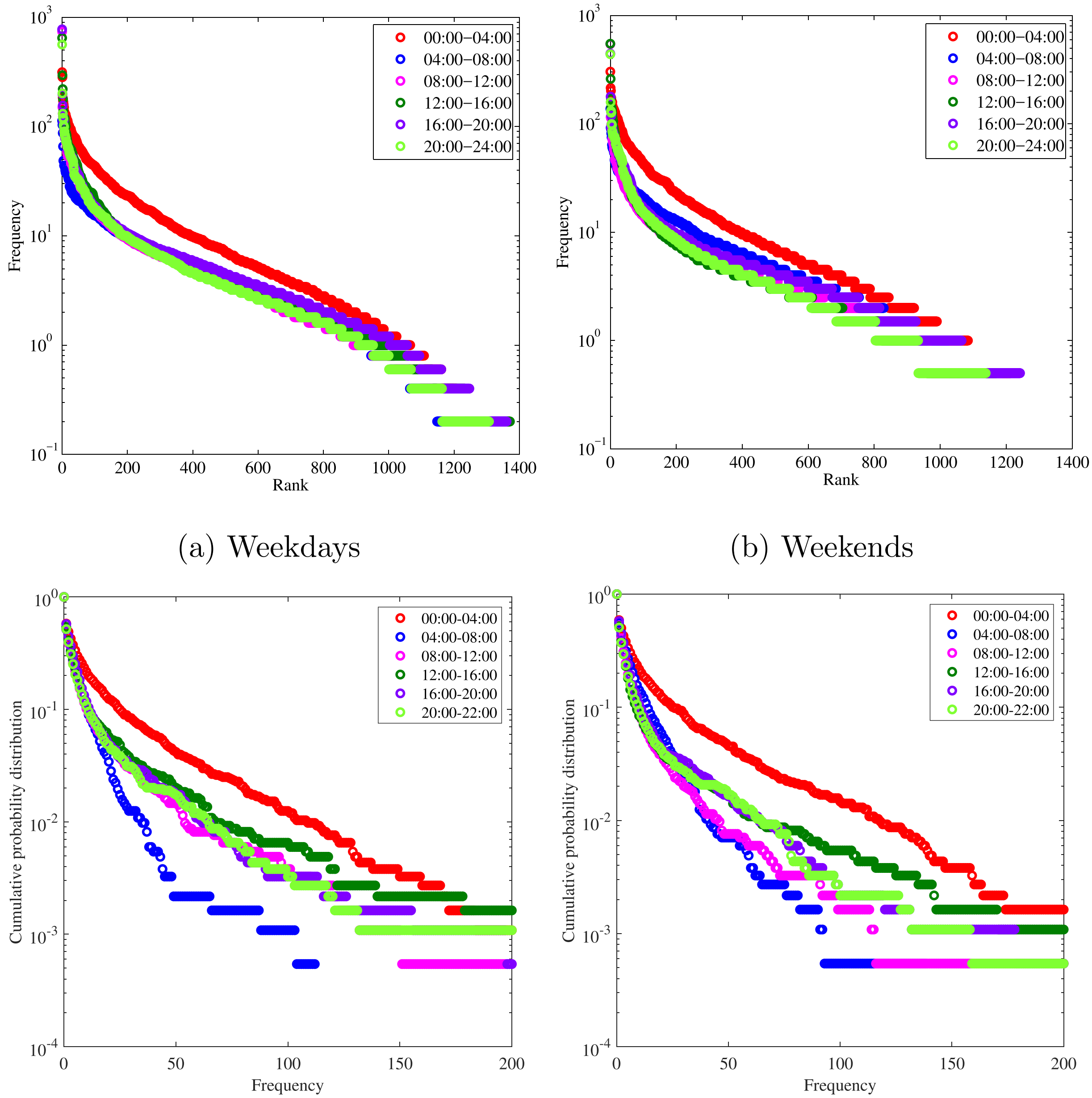
Two OLS regression models examine the global relationships between the stay frequencies on weekdays and weekends and the built environment variables. The regression results are summarized in Tables 4 and 5. As the results indicate, the number of residences, the number of car repair and maintenance shops, the number of gas stations, the existence of an airport, the existence of a railway station, road density and POI diversity are significantly associated with stay frequency. However, the number of restaurants, the number of shopping malls and the number of toilets have no significant association with stay frequency for both weekdays and weekends. Taking the number of restaurants as an example, the study by Zhang et al. (2009) indicated that the constraints such as economic affordability, passengers need and the viability to park have influence on cab drivers' selection of dining spots. It is reported that the restaurants that are suitable for taxi drivers are relatively few, especially in downtown area. This conclusion can be helpful for the interpretation of the insignificance of the number of restaurants in the regression models. The adjusted R square is 0.742 for weekdays and 0.676 for weekends, which reveals that the selected independent variables can explain 74.2% and 67.6% of the variance in cab drivers' stay frequency on weekdays and weekends.

Moreover, the coefficients show that the number of car repair and maintenance shops, the number of gas stations, the existence of an airport, the existence of a railway station and POI diversity have positive associations with stay frequency. An increase in the numbers of car repair and maintenance shops and gas stations in a TAZ will attract more taxi repairing and refueling activities. Since cab drivers live or wait for passengers in residential areas, increasing the number of residences in a TAZ can also lead to the high stay frequency of them. Because airport and railway stations are the main areas for cab drivers to wait for the next passenger, it is not surprising that the existence of an airport or railway station significantly contributes to stay frequency. Road network density and POI diversity are also positively associated with stay frequency, which suggests that increased road density and POI diversity generates more cab drivers' stay activities. In addition, we also calculate the variance inflation factor (VIF) values of each variable for the models, which are all far less than 10, indicating that there is no multi-collinearity among the independent variables.

Although the OLS regression models are able to offer a global understanding of the relationship between stay frequency and built environment characteristics, they ignore spatial nonstationary inspection. As a type of regression model that controls local variations within a range instead of estimating coefficients globally, GWR has been indicated as an effective method in human movement studies (Huang, Liu, Zhao, Zhang, & Kwan, 2019; Yang et al., 2019). Therefore, we further explored their relationship with GWR model.

4.3.2. GWR results

In this section we discuss the relationships between stay frequency and the built environment based on the results of the GWR models. According to what we find, the adjusted R^2 of the GWR models (i.e.,



(a) Weekdays (b) Weekends

(c) Weekdays (d) Weekends

Fig. 4. Rank-size distributions (a–b) and cumulative probability distributions (c–d) of stay frequency during different time periods.

Table 2
Decay parameters during different time periods on weekdays and weekends.

Time period	b	
	Weekdays	Weekends
0:00–4:00	1.0042	1.0041
4:00–8:00	1.0037	1.0036
8:00–12:00	1.0037	1.0034
12:00–16:00	1.0037	1.0035
16:00–20:00	1.0035	1.0033
20:00–24:00	1.0038	1.0037

0.889 for weekdays and 0.854 for weekends) are higher than those of the OLS models (i.e., 0.742 for weekdays and 0.676 for weekends), respectively. In addition, the Akaike Information Criterion (AIC) of GWR regression models (i.e., 18,401.85 and 18,302.01) are remarkably lower than those of OLS regression models (i.e., 19,734.58 and 19,553.79). The improvement of R^2 and the reduction in AIC values indicate that GWR regression models have better explanatory power than the OLS models for understanding the influence of the built environment on cab drivers' stay behavior.

Fig. 7 shows that the spatial distributions of local R^2 of the GWR model for weekdays and weekends. The results show that two GWR models display similar spatial distributions of local R^2 on weekdays and weekends. It can be observed that the GWR models display relatively higher explanatory capacity in the northwestern and eastern parts of the study area. It is noteworthy that the TAZs of and around the airport and railway stations have higher R^2 values than that of other TAZs. In addition, we also examine the spatial distributions of the standard residuals from the two GWR models on weekdays and weekends, as shown in Fig. 8. The ranges of standard residuals from the two models are [− 5.34,11.92] and [− 5.27,12.05] respectively, in which about

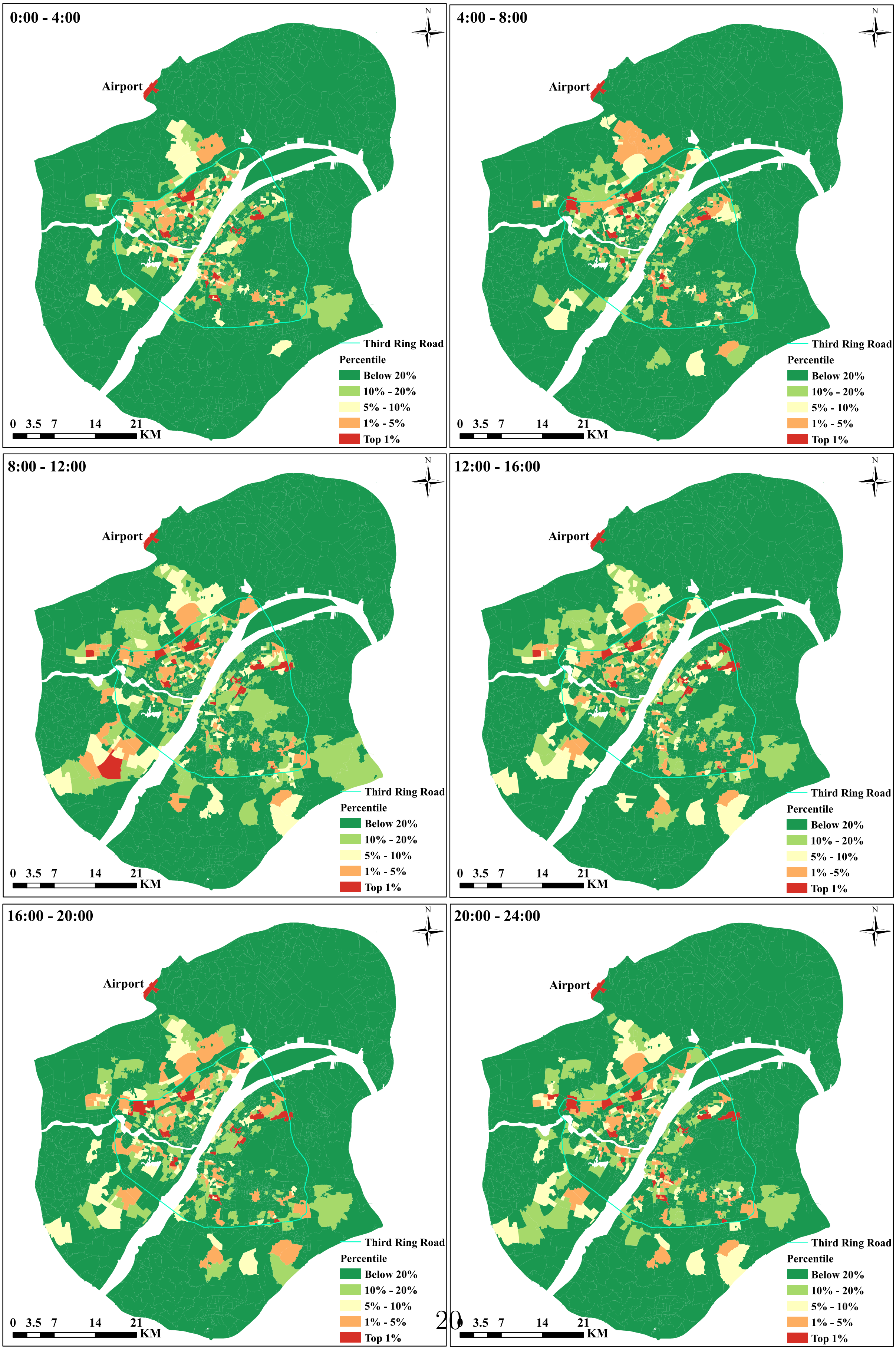


Fig. 5. Spatial distributions of cab drivers' stay activities during various periods on weekdays.

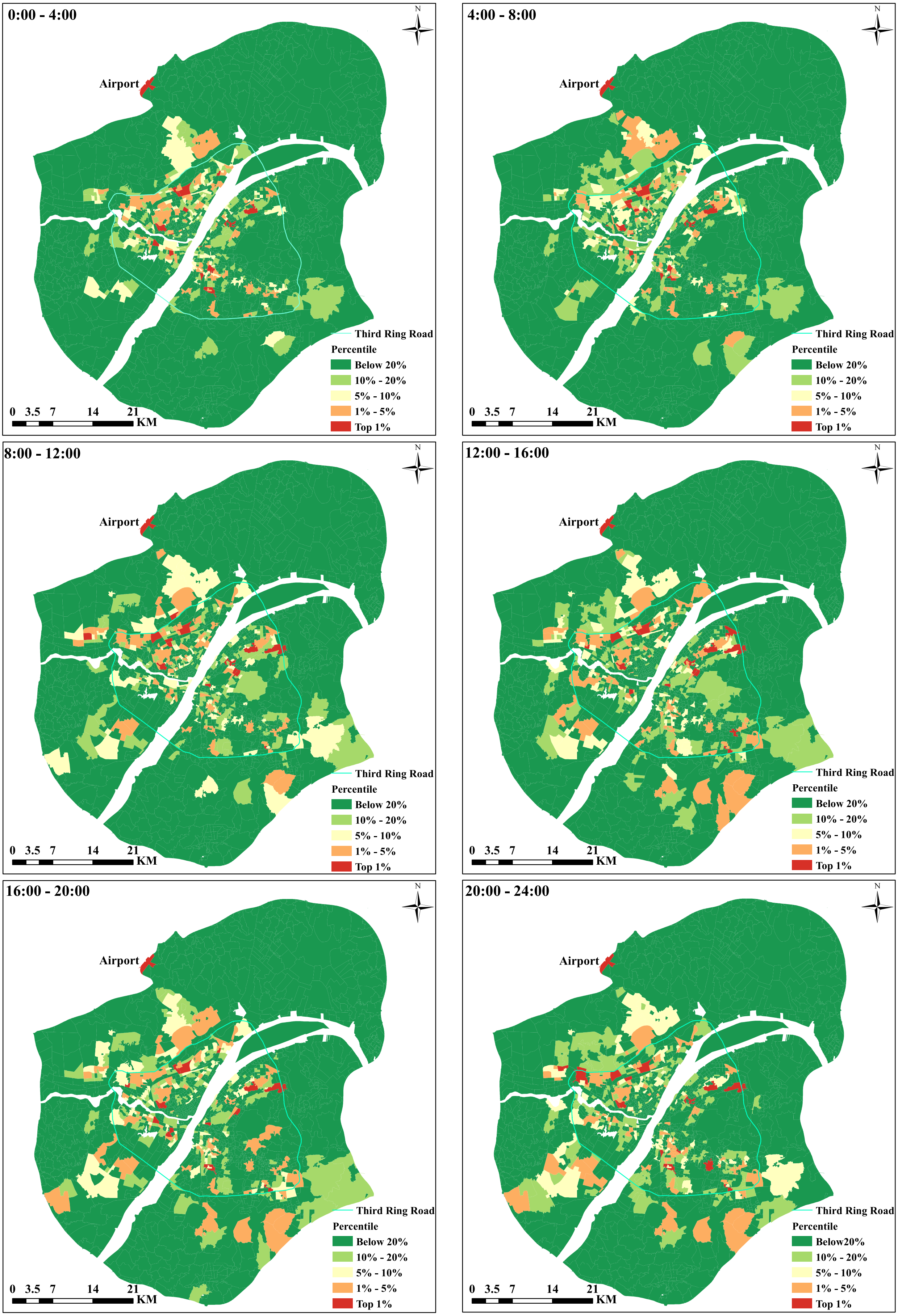


Fig. 6. Spatial distributions of cab drivers' stay activities during various periods on weekends.

Table 3
Descriptive statistics of the built environment variables.

Variables	Average	Standard deviation
NoRt (number of restaurants)	0.667	2.832
NoSM (number of shopping malls)	0.147	0.746
NoRc (number of residences)	2.195	5.110
NoCS (number of car repair and maintenance shops)	1.250	3.475
NoGS (number of gas stations)	0.028	0.180
EoA (existence of an airport)	0.0005	0.0233
EoRS (existence of a railway station)	0.0016	0.0403
NoT (number of public toilets)	0.390	0.935
RND (road network density)	8.175	6.541
POID (POI diversity)	0.214	0.250

Table 4
OLS regression results of stationary frequency on weekdays.

Variable	Coefficients	Standardized coefficients	Std	VIF
NoRt	0.297	0.008	0.462	1.197
NoSM	−1.051	−0.008	1.715	1.146
NoRc	0.770**	0.039**	0.247	1.110
NoCS	3.613**	0.124**	0.375	1.189
NoGS	196.104**	0.349**	6.932	1.089
EoA	3046.793**	0.703**	51.363	1.003
EoRS	362.745**	0.145**	30.644	1.069
NoT	0.924	0.009	1.433	1.257
RND	0.514**	0.033**	0.193	1.115
POID	54.468**	0.135**	6.117	1.634
Constant	3.682			
AIC = 19,734.58				
Adjusted R^2 = 0.742				

* Represents statistically significant at the $p < 0.05$ level.
** Represents statistically significant at the $p < 0.01$ level.

Table 5
OLS regression results of stationary frequency on weekends.

Variable	Coefficients	Standardized coefficients	Std	VIF
NoRt	0.273	0.009	0.440	1.197
NoSM	−2.077	−0.018	1.633	1.146
NoRc	0.682**	0.041**	0.235	1.110
NoCS	3.220**	0.130**	0.357	1.189
NoGS	193.226**	0.405**	6.600	1.089
EoA	2208.891**	0.60**	48.903	1.003
EoRS	335.765**	0.158**	29.176	1.069
NoT	1.056	0.012	1.365	1.257
RND	0.466*	0.036*	0.184	1.115
POID	52.632**	0.153**	5.824	1.634
Constant	3.940*			
AIC = 19,553.79				
Adjusted R^2 = 0.676				

* Represents statistically significant at the $p < 0.05$ level.
** Represents statistically significant at the $p < 0.01$ level.

96.8% and 97.1% of value belong to [− 2.58,2.58] respectively. We further test the spatial auto-correlation of standard residuals and obtain the global Moran's I index for two models, namely −0.012 and −0.010 respectively. From the spatial distributions of the standard residuals in Fig. 8, we can also see that only a few of the local regression models fail the residual tests, which correspond to the blue and red TAZs. Therefore, the distributions of the standard residuals from the GWR models are random at the 5% level of significance.

In Figs. 9–12, we display the spatial distribution of estimated coefficients of all the built environment variables. The blue colors signify that the corresponding built environment variables have a negative influence on stay frequency, while the red colors represent a positive influence. The GWR models examine the spatial varying influence of each built environment variable on stay frequency during weekdays and weekends. It can be observed that the associations fluctuate from negative (blue in Figs. 9–12) to positive (red in Figs. 9–12) for all the variables except *NoGS* and *EoA*. Additionally, the influence of each variable exhibits similar spatial distribution on weekdays and weekends. This suggests that the impact of urban built environment on cabdrivers' stay activities is relatively consistent across weekdays and weekends. This conclusion is consistent with the experimental results in Sections 4.1 and 4.2.

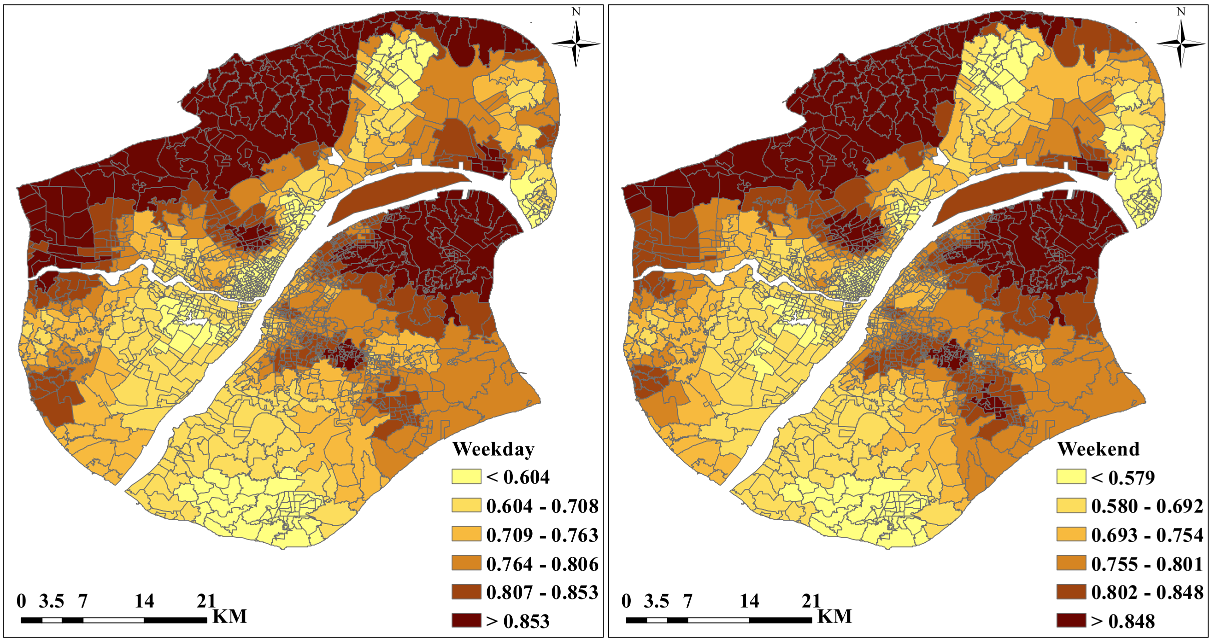
For variables such as the number of restaurants (*NoRt*), the number of shopping malls (*NoSM*) and the number of public toilets (*NoT*), the proportions of TAZs with positive and negative coefficients are comparable to each other, which suggests a complicated relationship between these variables and drivers' activities. This also explains why these variables are not significantly related to the stay frequency in the OLS models. For instance, the positive values for *NoRt* are mainly distributed in Qingshan district, Wuchang distirct, and the periphery of the study area, while the negative values occupies areas such as Jiang'an, Hanyang and Hong shan districts (Fig. 9a-1 and a-2). Since dining represents an important aspect of cabdrivers' stay activities, the result may imply that the number of restaurants in an area does not always correspond to a high probability of dining activities. Areas such as Jiang'an, Hanyang and Hongshan refer to some core parts of the inner city where the demand for taxis is usually high. It is possible that cab drivers' tended not to arrange their dining activities within these areas. The research findings are also consistent with the conclusions in the study by Zhang et al. (2009), which shows that the dining spots of taxi drivers are mainly distributed in remote areas near administrative boundaries with loose regulations. This is because these areas are more convenient for cab drivers to park their taxis while dining.

Regarding the number of residences (*NoRc*), most of the TAZs present the positive associations with stay frequency, while the TAZs with negative coefficients only cover areas such as East Lake, Wuhan University (Fig. 9c-1 and c-2). Similarly, it is found that the number of car repair and maintenance shops (*NoCS*) show positive relationships with stay frequency for most of the TAZs (see Fig. 10d-1 and d-2). It could be attributed to taxi drivers' shift handover events or parking activities for resting, since many cab drivers normally handover their taxis in the residential communities. From the Fig. 10e-1, e-2, f-1 and f-2, we can observe that the variables the number of gas stations (*NoGS*) and the existence of an airport (*EoA*) are positively related to stay frequency for all TAZs, while the positive values differ from each other. Note that the high positive values are mainly distributed in the center of the study area for *NoGS*, while the high positive values are concentrated on the north and south for *EoA*. One possible explanation is that most of the gas stations are focused on the center area. The existence of a railway station (*EoRS*), likewise, shows the a positive impact for most of the TAZs, as displayed in Fig. 11g-1 and g-2. With regard to road network density (*RND*), the TAZs with positive coefficients are mainly distributed in the center and south areas (Fig. 11i-1 and i-2). In addition, a positive influence of POI diversity (*POID*) is observed in over 80% of TAZs, covering most of the core urban areas of Wuhan (Fig. 12j-1 and j-2).

In summary, the associations of the selected built environment variables with cab drivers' stay activities present spatial variation across the study area. OLS models provide the global relationship between built environment and stay frequency for each TAZ, and GWR models discover some distinctive differences among TAZs by considering spatial autocorrelation and heterogeneity. The model results shed light on the relationship between cab drivers' stay behaviors and built environment at a more microscopic scale.

5. Discussion and conclusions

Understanding cab drivers' stay behaviors could benefit the planning and management of public facilities that are tied to taxis' daily operations. However, very limited efforts have been devoted to

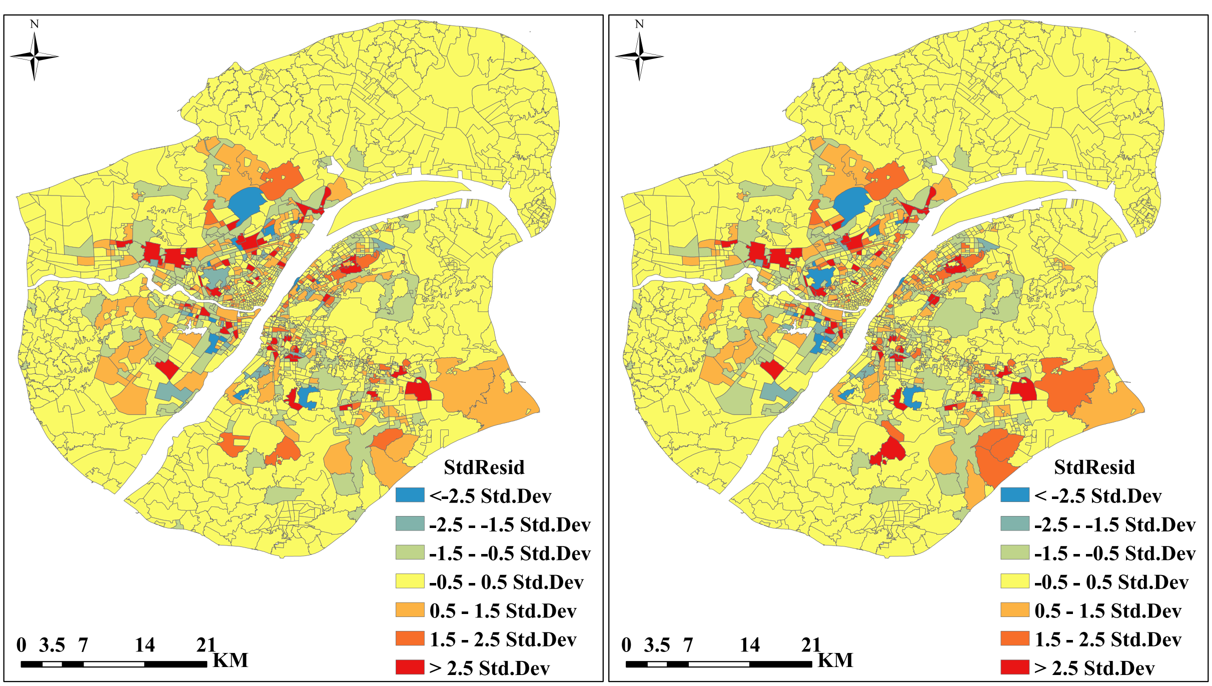


(a) Weekdays (b) Weekends

Fig. 7. Spatial distributions of local R-squared from the GWR models on weekdays and weekends.

studying this particular aspect of driver behavior. To fill the research gap, this study investigates cab drivers' stay behaviors using a GPS-based taxi trajectory dataset collected in Wuhan, China. The stay points of cab drivers are extracted from the trajectory data. These activities are then aggregated at the level of traffic analysis zones (TAZs) to describe the collective patterns of drivers' stay behaviors and the spatiotemporal dynamics. The main findings of this study are summarized as follows.

First, the rank-size and cumulative probability distributions are employed to quantify the spatial concentration of stay activities during different time periods of a day. The results suggest that stay frequency has a notable decay effect over different time periods on both weekdays and weekends. The results indicate that a substantial amount of cab drivers' stay activities tended to be concentrated in particular areas of the city. The decay effects during different periods, however, are characterized by temporal differentiation (e.g. the highest decay effect during 0:00–4:00). From another point of view, the agglomeration of



(a) Weekdays (b) Weekends

Fig. 8. Spatial distributions of standard residuals from GWR models on weekdays and weekends.

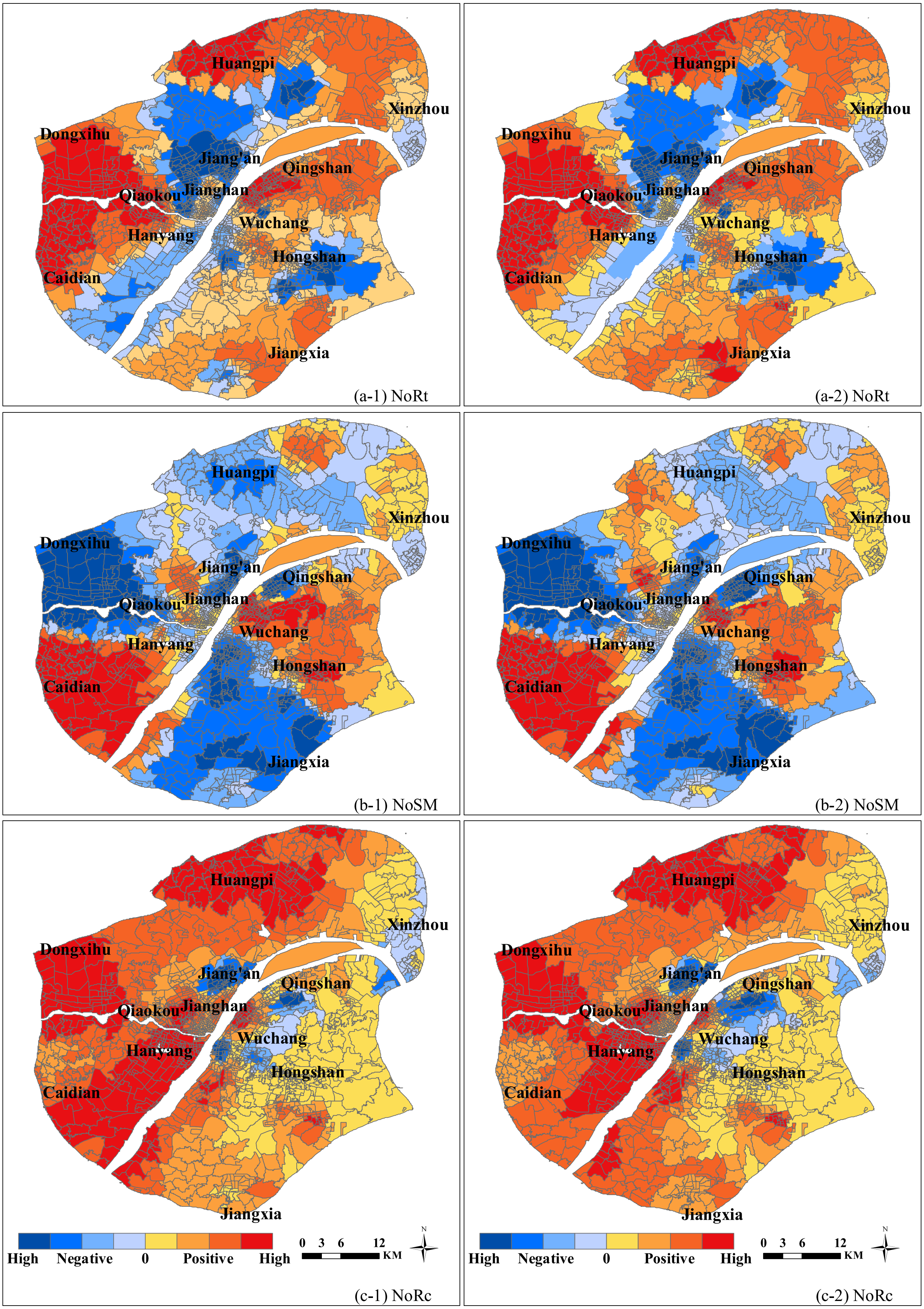


Fig. 9. Spatial distribution of regression coefficients. The left column represents weekdays, and the right column represents weekends. NoRt refers to number of restaurants, NoSM refers to number of shopping malls, NoRc refers to number of residences.

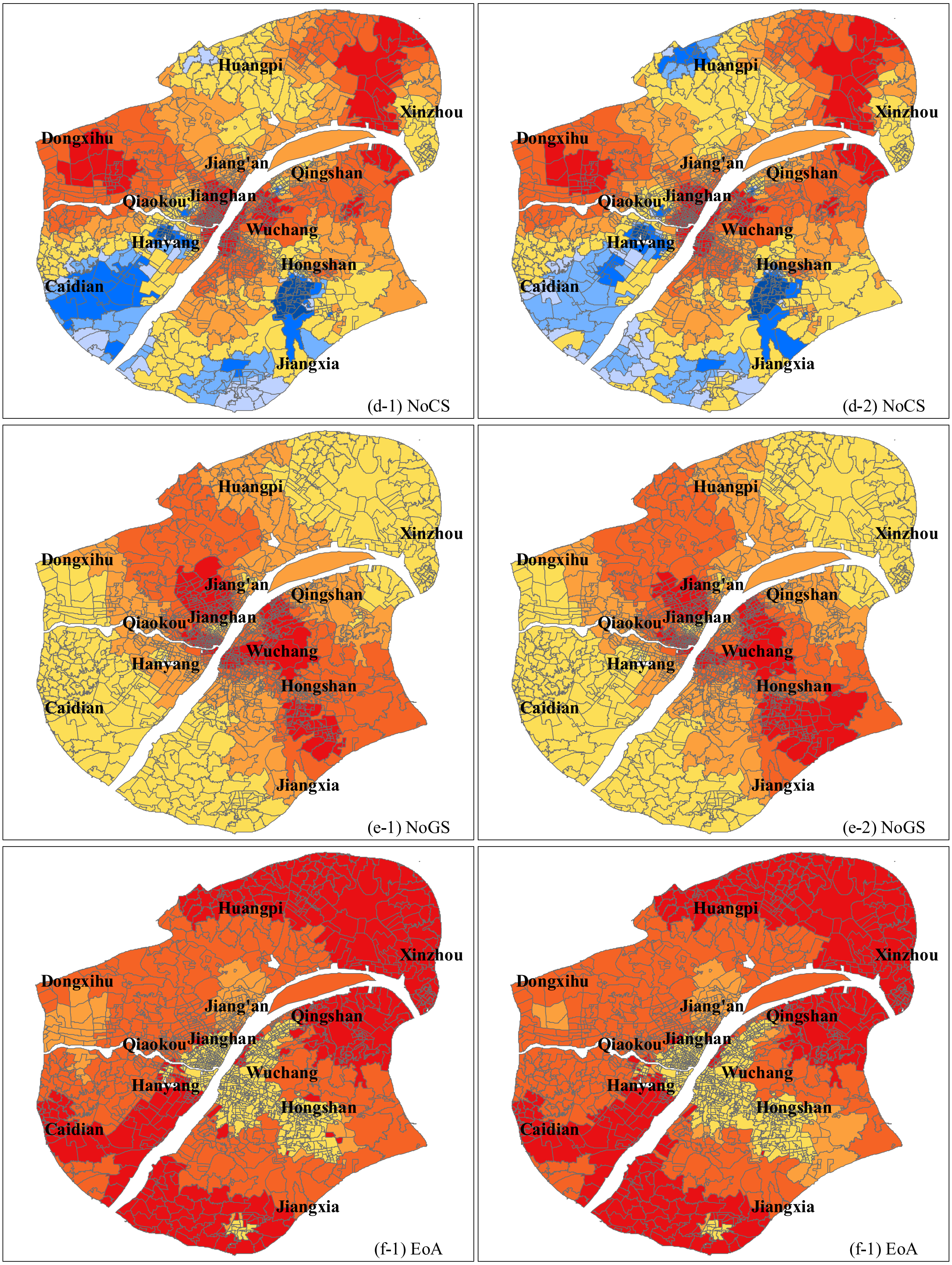


Fig. 10. Continued. NoCS refers to number of car repair and maintenance shops, NoGS refers to number of gas stations, EoA refers to existence of an airport.

cab drivers' stay activities in a small number of TAZs also indicates that cab drivers' stay behaviors are predictable at the aggregate level.

Second, the temporal dynamics of cab drivers' stay activities are explored. In this step we only focus on the top 20% of the TAZs — which are highly attractive regions that account for more than 76.8% and 76.9% of total stay activities on weekdays and weekends, respectively. It is found that the top TAZs exhibit similar spatial distributions during the same time period on weekdays and weekends. The spatial patterns of stay behaviors in different periods, however, show substantial differences. For instance, the TAZs with high stay frequencies

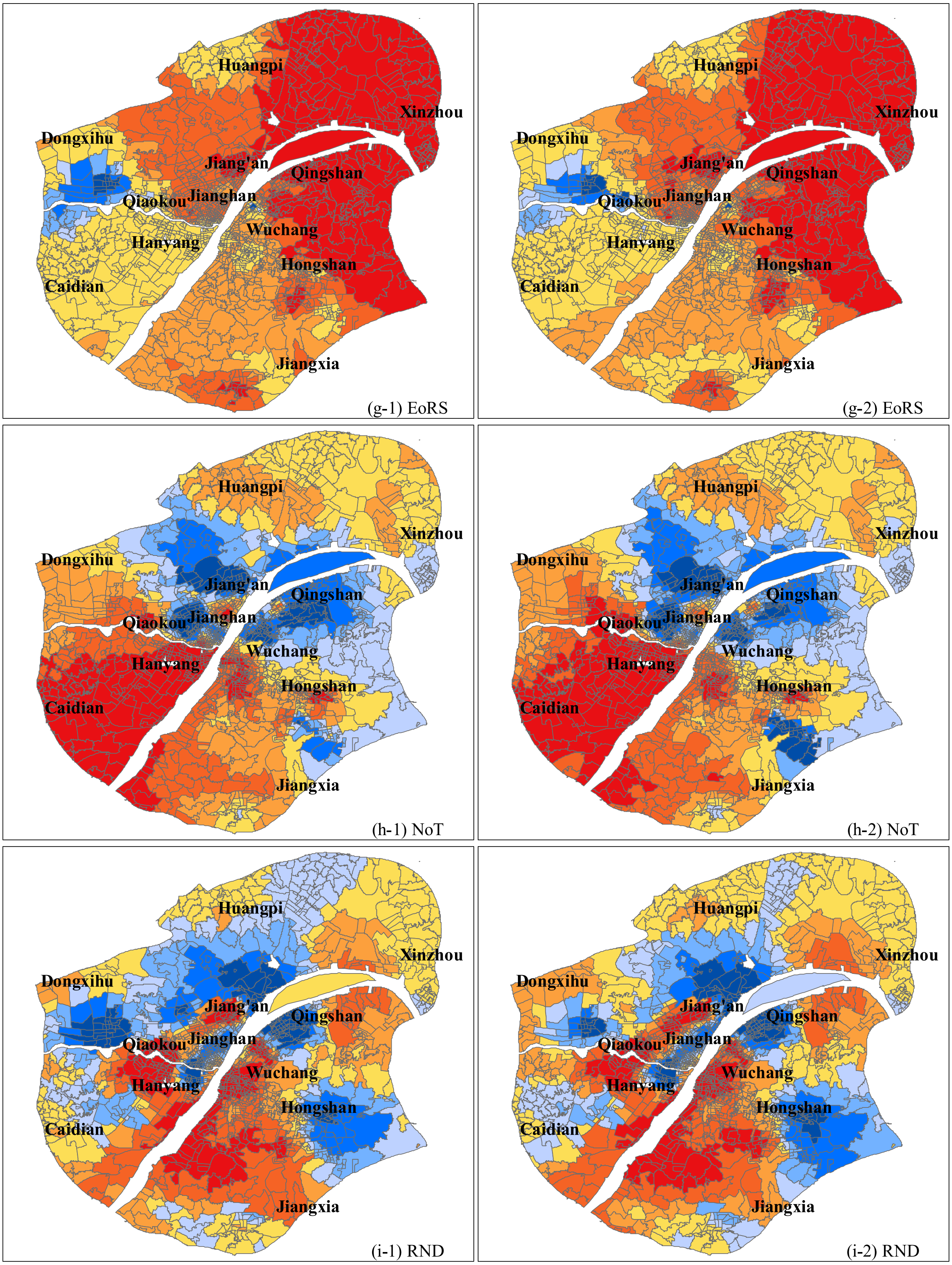


Fig. 11. Continued. EoRS refers to existence of a railway station, NoT refers to number of public toilets, RND refers to road network density.

during 0:00–4:00 are mainly distributed within the third ring road, while they display expansion towards outside the third ring road during other periods. Moreover, the top 20% TAZs overall experience few changes over different periods while the TAZs within top 1% and 1%–5% could change from time to time. The airport, however, is exceptionally displayed as a constant area within the top 1% over all periods. Summarily, we can conclude that the TAZs with high stay frequency shown in the rank-size distributions are almost constant across time and space. The urban public facilities related with cab drivers could be established in these TAZs with high stay frequency.

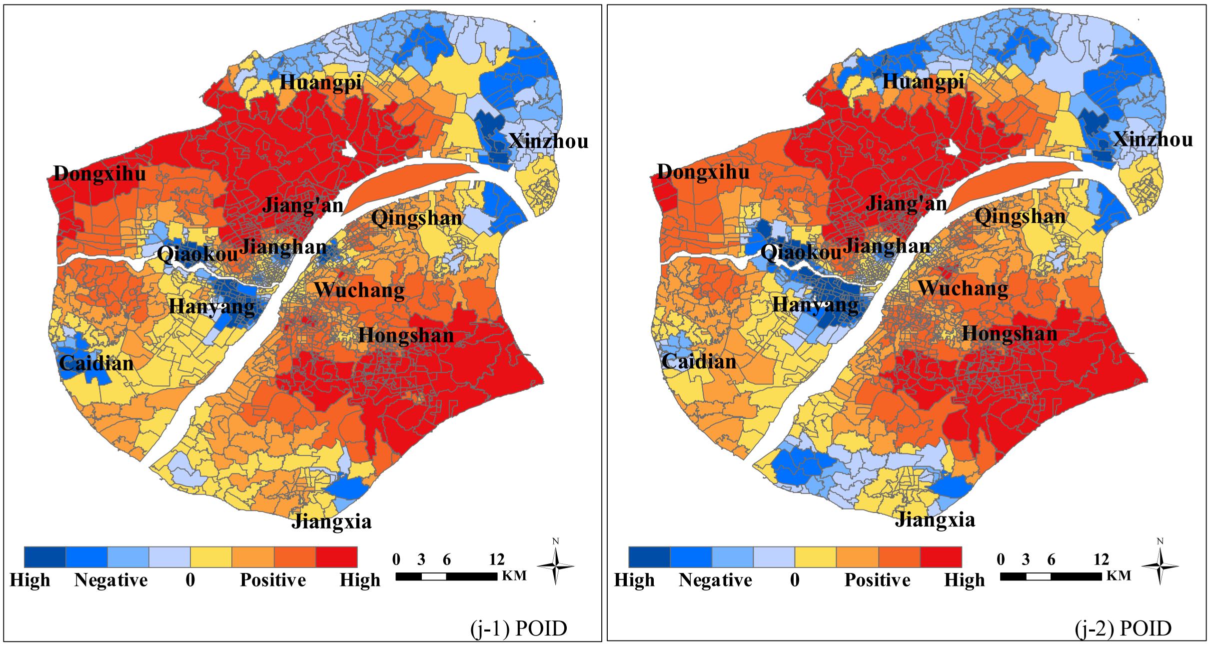


Fig. 12. Continued. POID refers to POI diversity.

Therefore, this study can be helpful for policy-makers to plan and manage urban transportation with the support of cab drivers' stay behaviors.

Third, the OLS regression and GWR models are used to examine the associations between stay frequency and built environment characteristics. The OLS regression results reveal a close relationship between the two on both weekdays and weekends. The relationships between built environment characteristics and the average stay frequencies on generic weekdays/weekends are examined. The adjusted R^2 is 0.742 for weekdays and 0.676 for weekends, indicating that 74.2% and 67.6% of the variance in cab drivers' stay frequency can be explained by the selected built environment variables on weekdays and weekends respectively. The coefficients of regression models demonstrate that the number of car repair and maintenance shops, the number of gas stations, the existence of an airport, the existence of a railway station, the number of residences, road network density and POI diversity have significantly positive associations with stay frequency. Specifically, the regression results indicate that the existence of an airport and the number of gas stations have a remarkable influence on cab drivers' stay behaviors. The number of restaurants, the number of shopping malls and the number of public toilets have no significant impact on stay frequencies. It is indicated that the current restaurants might not be able to meet cab drivers' dining requirements very well, which is supported by the study (Zhang et al., 2009).

GWR model is further employed to examine the existence of spatial nonstationarity. In this study, GWR model performs better than the OLS regression model based on the values of adjusted R^2 and AIC. The low values of Moran's I (weekday: -0.012; weekend: -0.010) in the residual maps indicate that there is no systematic error in the models. The results show that the relationships of the selected built environment variables with cab drivers' stay activities display spatial variations across the study area. Taking variables such as the number of restaurants, the number of shopping malls and the number of public toilets as examples, the comparable proportions of TAZs with positive and negative coefficients indicate a complicated relationship between these variables and stay activities. It also tends to explain why these variables are not significantly related to the stay frequency in the OLS models.

Based on the results of two sets of regression models, we find that the relationships between stay frequencies and certain variables — such as the number of car repair and maintenance shops, the number of gas stations, and the existence of an airport — are relatively consistent across OLS and GWR. While some other variables, such as the number of restaurants and the number of shopping malls, show a more complicated relationship with drivers' stay activities. Since different types of urban facilities are tied to specific types of driver activities, the comparative result suggests several hypotheses that are worth investigating. It is possible that urban facilities that link to certain aspects of drivers' activities (e.g., refueling, taxi repairing, and waiting for passengers at the airport) are very indicative of their stay activities. As these activities are important to the maintenance of the vehicles and the revenue of drivers, the corresponding urban facilities might become attractions for cabdrivers' stay activities. For other types of activities, such as dining and resting, the drivers might choose to perform them in a more flexible way, by considering to what extent they conflict with their operation strategies. For instance, the GWR model suggests a negative impact of number of restaurants on stay activities in certain areas such as Jiang'an, Hanyang and Hongshan districts (Fig. 9a-1 and a-2). These areas refer to the core parts of the city where demand for taxis is usually high. When operating in these areas, drivers may not conform to fixed schedules for dining and resting.

Our findings have several important implications with respect to planning and policy recommendations. First, the results of rank-size and cumulative probability distributions indicate that a substantial amount of cab drivers' stay activities were highly concentrated in particular areas of the city. The top 20% of the TAZs account for more than 75% of the total activities. These areas should be primarily targeted for follow-up studies and future urban planning. For instance, surveys and interviews can be arranged at these places to better understand and validate the purposes of cab drivers' activities. Such information can be useful to guiding future deployment of infrastructures and services that would accommodate cab drivers' needs. Second, the results of the regression models suggest that certain types of facilities, such as restaurants and toilets, do not well explain the whereabouts of cab drivers' stay behavior. While there is no doubt that that dining and resting are essential activities for taxi drivers, it would be meaningful to further investigate the reasons for such a spatial mismatch. For instance, understanding how parking availability (especially short-term parking spaces) mediates the relationship between stay activities and other urban facilities (e.g., restaurants) could shed additional light on cab drivers' stay behavior and constraints of activities under complex urban contexts.

In this study, we find that the collective dynamics of cabdrivers' stay activities are tied to the spatial arrangement of urban built environment. The findings suggest that cabdrivers' stay behaviors are not random in time and space, but connected with particular activity purposes (e.g., dining, hunting, refueling). It would be meaningful to apply the methods across different cities in the world, and examine the universal characteristics of cabdrivers' stay activities, or their variations that can be caused by the varying political, institutional and social-cultural contexts of cities. This is one possible direction for future research. We want to point out a few limitations of this research. First, the study uses a one-week GPS dataset. While the regression results have captured a close relationship between driver activities and built environment characteristics, our analyses were not able to account for other factors, such as seasonality, that would mediate their inter-relationships. In the future, we plan to further investigate this when longitudinal datasets become available. Second, this study uses the frequency of stay as the sole indicator of stay behavior. Other indicators, such as the duration of stay, could be incorporated in the future. As the duration of stay is connected with the activity purpose, incorporating this variable into the analysis could provide additional insights into cabdrivers' stay behavior, and their linkage with specific types of urban infrastructure and amenities.

CRediT authorship contribution statement

Pengxiang Zhao:Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing.**Yang Xu**:Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing.**Xintao Liu**:Conceptualization, Investigation, Writing - review & editing.**Mei-Po Kwan**:Conceptualization, Investigation, Writing - review & editing.

Acknowledgement

The authors would like to thank the anonymous reviewers and the editor for their valuable comments and suggestions on earlier versions of the manuscript. This research is jointly supported by the Hong Kong Polytechnic University Start-Up Grant (no. 1-ZE6P, no. 1-BE0J), the National Natural Science Foundation of China (no. 41529101, no. 41801372), and a grant from the Research Committee on Research Sustainability of Major RGC Funding Scheme of the Chinese University of Hong Kong.

References

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis, 28*(4), 281–298.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Jr., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies, 26*, 301–313.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment, 2*(3), 199–219.

Chen, X. M., Chen, C., Ni, L., & Li, L. (2018). Spatial visitation prediction of on-demand ride services using the scaling law. *Physica A: Statistical Mechanics and its Applications, 508*, 84–94.

Chun, B., & Guldmann, J. M. (2018). Impact of greening on the urban heat island: Seasonal variations and mitigation strategies. *Computers, Environment and Urban Systems, 71*, 165–176.

Etminani-Ghasrodashti, R., & Ardeshiri, M. (2015). Modeling travel behavior by the structural relationships between lifestyle, built environment and non-working trips. *Transportation Research Part A: Policy and Practice, 78*, 506–518.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association, 76*(3), 265–294.

Gao, Y., Jiang, D., & Xu, Y. (2018). Optimize taxi driving strategies based on reinforcement learning. *International Journal of Geographical Information Science,* 1–20.

Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., & Trasarti, R. (2011). Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal, 20*(5), 695–719.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *nature, 453*(7196), 779.

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from northern California. *Transportation Research Part D: Transport and Environment, 10*(6), 427–444.

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing* (pp. 6). ACM.

Huang, J., Liu, X., Zhao, P., Zhang, J., & Kwan, M.-P. (2019). Interactions between bus, metro, and taxi use before and after the Chinese spring festival. *ISPRS International Journal of Geo-Information, 8*(10), 445.

Kalhori, R. P., Foroughinia, A., & Ziapour, A. (2017). A survey of interurban taxi drivers' driving behaviors across Kermanshah, Iran in 2015. World family medicine journal: Incorporating the Middle East. *Journal of Family Medicine, 99*(4216), 1–10.

Kang, C., & Qin, K. (2016). Understanding operation behaviors of taxicabs in cities by matrix factorization. *Computers, Environment and Urban Systems, 60*, 79–88.

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS One, 9*(6), e96180.

Lee, K., & Kwan, M.-P. (2018). Automatic physical activity and in-vehicle status classification based on GPS and accelerometer data: A hierarchical classification approach using machine learning techniques. *Transactions in GIS, 22*(6), 1522–1549.

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.

Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., & Yang, Q. (2011). Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. *Pervasive computing and communications workshops (PERCOM workshops), 2011 IEEE international conference on* (pp. 63–68). IEEE.

Liang, X., Zhao, J., Dong, L., & Xu, K. (2013). Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports, 3*(2983).

Liu, L., Andris, C., & Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems, 34*(6), 541–548.

Liu, S., Wang, S., Liu, C., & Krishnan, R. (2015). Understanding taxi drivers' routing choices from spatial and social traces. *Frontiers of Computer Science, 9*(2), 200–209.

Liu, Y., Kang, C., Gao, S., Xiao, Y., & Tian, Y. (2012). Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems, 14*(4), 463–483.

Lu, M., Lai, C., Ye, T., Liang, J., & Yuan, X. (2017). Visual analysis of multiple route choices based on general GPS trajectories. *IEEE Transactions on Big Data, 3*(2), 234–247.

Ma, M., Yan, X., Huang, H., & Abdel-Aty, M. (2010). Safety of public transportation occupational drivers: Risk perception, attitudes, and driving behavior. *Transportation Research Record, 2145*(1), 72–79.

Manley, E., Addison, J., & Cheng, T. (2015). Shortest path or anchor-based route choice: A large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography, 43*, 123–139.

Newnam, S., Mamo, W. G., & Tulu, G. S. (2014). Exploring differences in driving behaviour across age and years of education of taxi drivers in Addis Ababa, Ethiopia. *Safety Science, 68*, 1–5.

Niu, H., Liu, J., Fu, Y., Liu, Y., & Lang, B. (2016). Exploiting human mobility patterns for gas station site selection. *International conference on database systems for advanced applications* (pp. 242–257). Springer.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS One, 7*(5), e37027.

Riascos, A., & Mateos, J. L. (2017). Emergence of encounter networks due to human mobility. *PLoS One, 12*(10), e0184532.

Shen, J., Liu, X., & Chen, M. (2017). Discovering spatial and temporal patterns from taxi-based floating car data: A case study from Nanjing. *GIScience & Remote Sensing, 54*(5), 617–638.

Shi, J., Tao, L., Li, X., Xiao, Y., & Atchley, P. (2014). A survey of taxi drivers' aberrant driving behavior in Beijing. *Journal of Transportation Safety & Security, 6*(1), 34–43.

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics, 6*(10), 818.

Song, J., Lai, J., Zou, S., Huang, C., Zhu, Z., Shi, A., ... Wang, Z. (2012). Analysis on dietary behaviors among taxi drivers in Shanghai. *Chinese Journal of Health Education, 28*(2).

Sun, R., & Yu, H. (2014). Du y. time space distribution characteristics of taxi shift in Beijing. *Journal of TransportationSystems Engineering and Information Technology, 14*(1), 221–228.

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.

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.

Tang, J., Liang, J., Zhang, S., Huang, H., & Liu, F. (2018). Inferring driving trajectories based on probabilistic model from large scale taxi GPS data. *Physica A: Statistical Mechanics and its Applications, 506*, 566–577.

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.

Wang, J., Lee, K., & Kwan, M.-P. (2018). Environmental influences on leisure-time physical inactivity in the us: An exploration of spatial non-stationarity. *ISPRS International Journal of Geo-Information, 7*(4), 143.

Wang, M., & Mu, L. (2018). Spatial disparities of uber accessibility: An exploratory analysis in Atlanta, USA. *Computers, Environment and Urban Systems, 67*, 169–175.

Wang, W., Pan, L., Yuan, N., Zhang, S., & Liu, D. (2015). A comparative analysis of intra-city human mobility by taxi. *Physica A: Statistical Mechanics and its Applications, 420*, 134–147.

Wang, Y., Qin, K., Chen, Y., & Zhao, P. (2018). Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi GPS data. *ISPRS International Journal of Geo-Information, 7*(1), 25.

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems, 72*, 51–67.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., & Ratti, C. (2019). Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems, 75*, 184–203.

Yang, L., Kwan, M.-P., Pan, X., Wan, B., & Zhou, S. (2017). Scalable space-time trajectory cube for path-finding: A study using big taxi trajectory data. *Transportation Research Part B: Methodological, 101*, 1–27.

Yang, X., Fang, Z., Xu, Y., Yin, L., Li, J., & Lu, S. (2019). Spatial heterogeneity in spatial interaction of human movements—Insights from large-scale mobile positioning data. *Journal of Transport Geography, 78*, 29–40.

Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., & Zhang, L. (2018). Analysis of Washington, dc taxi demand using GPS and land-use data. *Journal of Transport Geography, 66*, 35–44.

Yuan, J., Zheng, Y., Zhang, L., Xie, X., & Sun, G. (2011). Where to find my next passenger. *Proceedings of the 13th international conference on ubiquitous computing* (pp. 109–118). ACM.

Zhang, C., Zheng, T., Lu, B., & Chai, Y. (2009). Study on urban taxi drivers' out-dining characters and planning implications in Beijing. *Progress in Geography, 3*(11).

Zhang, D., Sun, L., Li, B., Chen, C., Pan, G., Li, S., & Wu, Z. (2015). Understanding taxi service strategies from taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems, 16*(1), 123–135.

Zhang, F., Yuan, N. J., Wilkie, D., Zheng, Y., & Xie, X. (2015). Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. *ACM Transactions on Intelligent Systems and Technology, 6*(3), 37.

Zhang, S., & Wang, Z. (2016). Inferring passenger denial behavior of taxi drivers from large-scale taxi traces. *PLoS One, 11*(11), e0165597.

Zhang, W., Zhang, X., Feng, Z., Liu, J., Zhou, M., & Wang, K. (2018). The fitness-to-drive of shift-work taxi drivers with obstructive sleep apnea: An investigation of self-reported driver behavior and skill. *Transportation Research Part F: Traffic Psychology and Behaviour, 59*, 545–554.

Zhang, X., Xu, Y., Tu, W., & Ratti, C. (2018). Do different datasets tell the same story about urban mobility—A comparative study of public transit and taxi usage. *Journal of Transport Geography, 70*, 78–90.

Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2013). What influences metro station ridership in China? Insights from Nanjing. *Cities, 35*, 114–124.

Zhao, P., Kwan, M.-P., & Qin, K. (2017). Uncovering the spatiotemporal patterns of co2 emissions by taxis based on individuals' daily travel. *Journal of Transport Geography, 62*, 122–135.

Zhao, P., Kwan, M.-P., & Zhou, S. (2018). The uncertain geographic context problem in the analysis of the relationships between obesity and the built environment in Guangzhou. *International Journal of Environmental Research and Public Health, 15*(2), 308.

Zhao, P., Liu, X., Kwan, M.-P., & Shi, W. (2018). Unveiling cabdrivers' dining behavior patterns for site selection of "taxi canteen" using taxi trajectory data. *Transportmetrica A: Transport Science,* 1–24.

Zhao, P., Qin, K., Ye, X., Wang, Y., & Chen, Y. (2017). A trajectory clustering approach based on decision graph and data field for detecting hotspots. *International Journal of Geographical Information Science, 31*(6), 1101–1127.

Zheng, Z., & Zhou, S. (2017). Scaling laws of spatial visitation frequency: Applications for trip frequency prediction. *Computers, Environment and Urban Systems, 64*, 332–343.

Zong, F., Wu, T., & Jia, H. (2018). Taxi drivers' cruising patterns—Insights from taxi gps traces. *IEEE Transactions on Intelligent Transportation Systems, 20*(2), 571–582.