WILEY | Hindawi

Research Article

Quantifying the Impact of Rainfall on Taxi Hailing and Operation

Jian Sun (iD)**, He Dong, Guoyang Qin, and Ye Tian** (iD)

Key Laboratory of Road and Traffic Engineering of the State Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Ye Tian; tianye@tongji.edu.cn

Adverse weathers are well-known to impact the operation of transportation systems, including taxis. This paper utilizes taxi GPS waypoint data to investigate the quantitative impact of rainfall on taxi hailing and taxi operations to help improve service quality on rainy days. Through statistical analysis, the study proves that it is more difficult to hail taxis on rainy days, especially during morning peak hours. By modelling the difference value of factors for rainfall and nonrainfall conditions in a multivariate regression model and attaining the significance and elasticity of each factor, passenger demand, taxi supply, search time and velocity are proved to be the significant factors that lower the taxis' level of service on rainy days. Among them, the number of passengers and taxis are two factors that have the greatest impact. It is also shown that there is no significant difference in the total taxi supply and passenger demand between rainfall and nonrainfall conditions, but a dramatic change in the spatial distribution is discovered. The results suggest that instead of simply providing more taxis on rainy days, optimally dispatching taxicabs to high demand regions can be a more effective solution.

1. Introduction

Taxi services play a crucial role in transporting travelers in urban areas across the globe. It is an essential complement to public transit owing to its great convenience and wide availability [1]. In China, taxis carried about 35.17 billion passengers in 2018, which was 27.86% of the total urban passenger transport volume [2]. Such a huge volume highlights the importance of taxis and draws wide academic attention, and studies into factors affecting taxi operations have received considerable attention. Multiple factors have been studied in existing works, and they can be categorized into either endogenous factors (i.e., reflecting drivers' personal characteristics) such as operation region preferences, passenger search patterns, route choice behaviour, and delivery efficiency [3–6], or exogenous factors, such as varying traffic conditions, alternative transit options, online car booking, taxi prices, fuel costs, and regional land use [7, 8].

Nevertheless, the impact of adverse weather events, especially rainfall, can critically affect taxi supply and travel demand, but receives little to no attention. It is common sense in cities of high dense populations (e.g., Asia-Pacific cities, New York, etc.) that to hail a taxi become much harder in rainy days. Kamga et al. [9] investigated temporal and weather-related

variations in taxi ridership patterns in New York City from the perspective of a supply-demand equilibrium. Chen et al. [10] demonstrated that rainfall is a key influencing factor on the taxi service demand. As rain intensity increases, demands for evening rush hour and nonrush hour periods on workdays show opposite trends. However, most studies neglected the differences in spatial distribution and did not consider the causes of those variations. The other few related studies are almost all descriptive analyses under some proposed assumptions, but it lacks quantitative analyses of the performance of taxi operations [11, 12].

In contrast, tremendous attention has been placed on research into the impact of rainfall on other modes of transportation, such as the bus and subway systems [13–15]. Their findings suggest that rainfall can decrease the number of both regular and occasional travelers of subways in urban areas. Compared to the pre-defined schedules or routes, taxis have more flexibility than buses by providing a door-to door service. In adverse weather such as rainfall, some travelers would prefer taxis to avoid walking in rainfall to subway station. Thus, it is likely that the taxi demand would instead increase during rainfall. Due to distinct supply-demand structures, the analysis procedure used for those transportation modes cannot be directly transferred to studies into taxi operations. A

thorough study into rainfall's impact on taxi operations is therefore needed.

For this reason, we intend to utilize massive amounts of taxi GPS data to investigate the impact of rainfall on taxi hailing and reveal its quantitative impact on the performance of taxi operations to help improve taxi service quality on rainy days. Specifically, we first segment the study area into small cells and define a taxi level of service (TLOS) index to denote the difficulty of hailing taxis for each cell. The TLOS indices are then clustered into several levels and the spatio-temporal TLOS differences are explored. Afterwards, we identify the factors that may affect the TLOS indices and test the statistical differences of these factors between rainy and nonrainy days. Moreover, a novel method is proposed to quantify the distribution difference of supply and demand between rainy and nonrainy days. Lastly, we propose a regression model that correlates the TLOS and the impacting factors, and we obtain the significance of each factor's impact on the TLOS indices. To provide taxi regulators with quantitative advice on how to improve TLOS on rainy days, an elasticity analysis is conducted as well.

Three major contributions are made in this paper. First, we establish the TLOS indexing as a measure of the difficulty of hailing taxis and perform a specific spatio-temporal analysis of level changes between rainy and nonrainy days. Second, we thoroughly study the taxi operation factors affected by rainfall and quantitatively analyse the variation of spatial supply-demand distributions between rainy and nonrainy days. Third, we develop a multivariate regression model to find significant factors and conduct an elasticity analysis to quantify the contribution of each factor, which can provide quantitative values for regulators to take into account in order to improve taxi service quality on rainy days.

The rest of this paper is structured as follows. The next section discusses studies related to taxi services and the impact of rainfall. Section 3 describes the study area of interest and the data preparation process. Section 4 presents the methodology. Section 5 documents the results and discussion. Section 6 concludes the paper.

2. Literature Review

Due to the lack of existing research into the impact of rainfall on taxi services, this section will review literature that refers to the two most relevant study topics that are partly connected to this study. The studies of assessment of the taxi service and factors affecting taxi operations are introduced in the first part, and the impact of weather on transportation systems is reviewed in the second part.

2.1. Assessment of Taxi Service and Operations.
The taxi market, a heavily regulated industry, has received increased attention and has induced a number of studies. Studies into taxi services can be dated back to the mid-1970s [16, 17]. To evaluate the level of taxi service, scholars present multiple standards from different perspectives. Xu et al. [18] developed a neural network model to evaluate level of taxi service based on accurate endogenous variables including

passenger demand, waiting time, vacant taxi headway, average percentage of occupied taxis and taxi utilization. Yang et al. [19] used the average proportion of occupied taxis to reflect the utilization level and the average time headway between vacant taxis to reflect the availability of vacant taxis, thereby reflecting the level of taxi service. Shaaban and Kim [20] focused on passengers' satisfaction and conducted a descriptive analysis relating the demographic, accessibility, and trip purposes to taxi users with overall service satisfaction. Wong and Szeto [21] introduced a six-level TLOS (level-of-service) standard to evaluate the service quality of urban taxis based on the numerical score of a customer satisfaction survey.

With the advent of GPS tracking of taxis, a series of diverse research studies on taxi operation condition utilizing these taxi trajectories emerged. Some research focused on pattern mining of taxi operations. For example, Zhan et al. [22] explored taxi cruising patterns, travel time estimation, and travel speed variation. Li et al. [23] proposed spatio-temporal visualization analysis methods to quantify taxi operation patterns using GPS data. Other studies worked further to improve understanding of these taxi operation patterns, including taxi demand distribution predicting in order to recommend the best queuing locations for taxi drivers [24, 25]. Aided by GPS data, Gao et al. [25] and Qin et al. [26] investigated the differences in drivers' operation strategies and revealed the top drivers' strategies to help ordinary drivers increase their incomes.

To summarize, existing works to assess the level of taxi service rely on customer survey. GPS data has already become a necessary component in taxi studies. Its availability makes it possible to easily obtain cruising patterns, travel times, pick-up/drop-off locations, and the distribution of passengers, which are widely considered factors that affect the performance of taxi operation. Therefore, it is intuitive to utilize GPS data to assess level of taxi service and further facilitate analysis of taxi performance when it rains.

2.2. Impact of Rainfall on Transportation Systems.
Weather conditions are important external factors that affect transportation systems. Singhal et al. [27] utilized a linear model based on a station type of transit trip to prove the significant effects weather can have on ridership patterns. Khattak [28] found that the competitive rankings of various transportation modes often need to be adjusted in response to weather changes. A survey found that 55% of respondents would change their travel patterns after receiving weather information from secondary sources [29]. By estimating separate ordinary least square regression models for each season, Stover and McCormack [30] proved that adverse weather conditions led to lower transit ridership. Studies also suggest that a small amount of morning peak-hour rain reduces bus ridership significantly for the whole day, and that rainfall is the most influential weather event that can lead to a 3% reduction of daily bus ridership [31]. Hofmann and O'Mahony [7] analyzed the effect of rainfall on public transit in terms of service frequency, travel time and ridership variability, and identified a strong negative relationship between them.

As mentioned above, most existing studies have focused heavily on mass transit modes by using smart card databases or daily and monthly GPS data, which lack detailed

information. In the meantime, public transit systems such as buses and subways have fixed schedules and routes, while taxis are demand-responsive, which makes the findings of public transit not necessarily transferrable to the taxi industry. Therefore, an exploration on how rainfall affects taxi operations using insights obtained from GPS data is needed.

3. Study Area and Data Preparation

3.1. Study Area. Taxis play an especially important role in urban transportation within large cities. As the largest city in China, Shanghai is no exception, as taxis contribute to approximately 12% of trips among public transportation modes [32]. Due to the intervention of Transportation Network Companies (TNC) such as Didi, the passenger sharing rate of cruise taxis has declined. However, the relative share of taxi hailing is still far ahead. As per 2017 Shanghai Transportation White Paper, the current share of taxis in Shanghai is 11% while it is 3% for TNC vehicles. Amongst taxi trips, roughly 70% are drawn from hailing. In the near future, it is foreseen that the share of taxi hailing will not drop since the percentage of elder passengers who are prone to TNCs is still growing. Besides, TNC service has been widely condemned due to numerous crimes or policy violations conducted by TNC drivers [33, 34]. Further, transportation authorities in Shanghai do have the right to dispatch taxis to certain regions due to taxi shortage because of extreme weather conditions. Therefore, this study takes taxi hailing as default.

In Shanghai, taxi trips are highly concentrated within the downtown area bounded by the Outer Ring Road, where about 82.25% of taxi pickups are generated, which makes it an appealing choice for our study area [26]. To facilitate the investigation of spatial variations of taxi operations, we segment the study area into small square cells with dimensions of $1 \, km \times 1 \, km$. As a result, the area of interest is divided into 744 cells in total that cover longitudes from 121.35 to 121.65, and latitudes from 31.14 to 31.37.

3.2. Data Description and Preparation. Qiangsheng, the largest taxi operator in Shanghai and even China, operates 25% of the taxis in Shanghai [35]. The share of Qiangsheng in Shanghai is larger than any other taxi companies or TNCs. A set of taxi GPS data over two months retrieved from the Shanghai Qiangsheng Taxi Company was used in this study. The data contains information of sequential trajectories and operation status from approximately 10,000 taxis. Attributes include taxi ID, date, time, longitude, latitude, speed, bearing, and operation status identifier (1 for occupied/0 for vacant). The GPS data was collected every 10 seconds, and 631 million trajectory waypoints (including passenger search and delivery data) were recorded.

March and June, identified as the months with the most rainfall by the weather data, are selected as the study period. It is intuitive that it is more difficult to hail a taxi during peak hours when demand overwhelms. Therefore, we mainly focus on the morning peak hours (7 AM–9 AM) and evening peak hours (5 PM–7 PM) on weekdays. Off-peak periods (10 AM–11 AM and 2 PM–3 PM) are also considered for comparison purpose.

Weather information was collected from the weather service *Weather Underground*. The data contains date, time, events (e.g., rain, thunderstorms, none), and weather conditions (e.g., clear, mostly cloudy, overcast, rain, showers, thunderstorms, etc.). However, it is mostly light rain if there is any precipitation in Shanghai. Rain showers or heavy rain are both rare cases (occur roughly 1% of the entire time period), which is difficult to support the investigation on the severity of rainfalls. Therefore, in order to ensure sufficient observations for each weather category, we map the weather condition for the corresponding time period into either rainfall or nonrainfall, where a condition without any precipitation (e.g., cloudy, overcast) is tagged as nonrainfall. Since the weather data was collected every 30 minutes, we split the study period into 30-minute intervals and assume the weather conditions remain unchanged within the same time interval. With data cleaning, 41 days (in total 246 hours) of data are valid for use.

4. Methodology

In this study, we intend to first confirm the impact of rain on taxi operations and hailing by meticulously investigating the relationship between the change of difficulty in hailing taxis and rainfall conditions both temporally and spatially. We then aim to propose several factors that affect taxi operations to explain how rainfall impacts taxi operations and to quantify the change in the spatio-temporal distribution of taxi supply and demand in the rain. Lastly, we combine the difficulty measure and factors affecting taxi operations into a regression model to reveal which factors affect taxi operations and how they affect the difficulty in hailing taxis in the rain, which would provide taxi regulators with a better understanding of the relationship between taxi operations and weather.

4.1. Impact of Rain on the Difficulty of Hailing Taxis. In this section, we intend to explore the taxi GPS data to find out answers to the following three questions:

(1) Is it more difficult to hail taxis when it rains? This question targets the difference of difficulty in hailing taxis between rainfall and nonrainfall conditions.

(2) Does it become more difficult to hail taxis in the rain for regions where it was already difficult to hail them? This question requires us to investigate Question 1 from a more detailed perspective.

(3) Is there a significant difference in the impact of rain on the difficulty of hailing taxis across different periods? This question calls for mining spatio-temporal patterns of the change in difficulty of hailing taxis to achieve a comprehensive understanding of the impact of rain on taxi operations and hailing.

4.1.1. Difference in Difficulty of Hailing Taxis between Rainfall and Nonrainfall Conditions. Prior to analyzing the difference, we need to first define an appropriate measure of difficulty in hailing taxis. To make the following expression concise, we use the term taxi level of service (TLOS) to reflect the

TABLE 1: Global Moran's I summary for specific periods.

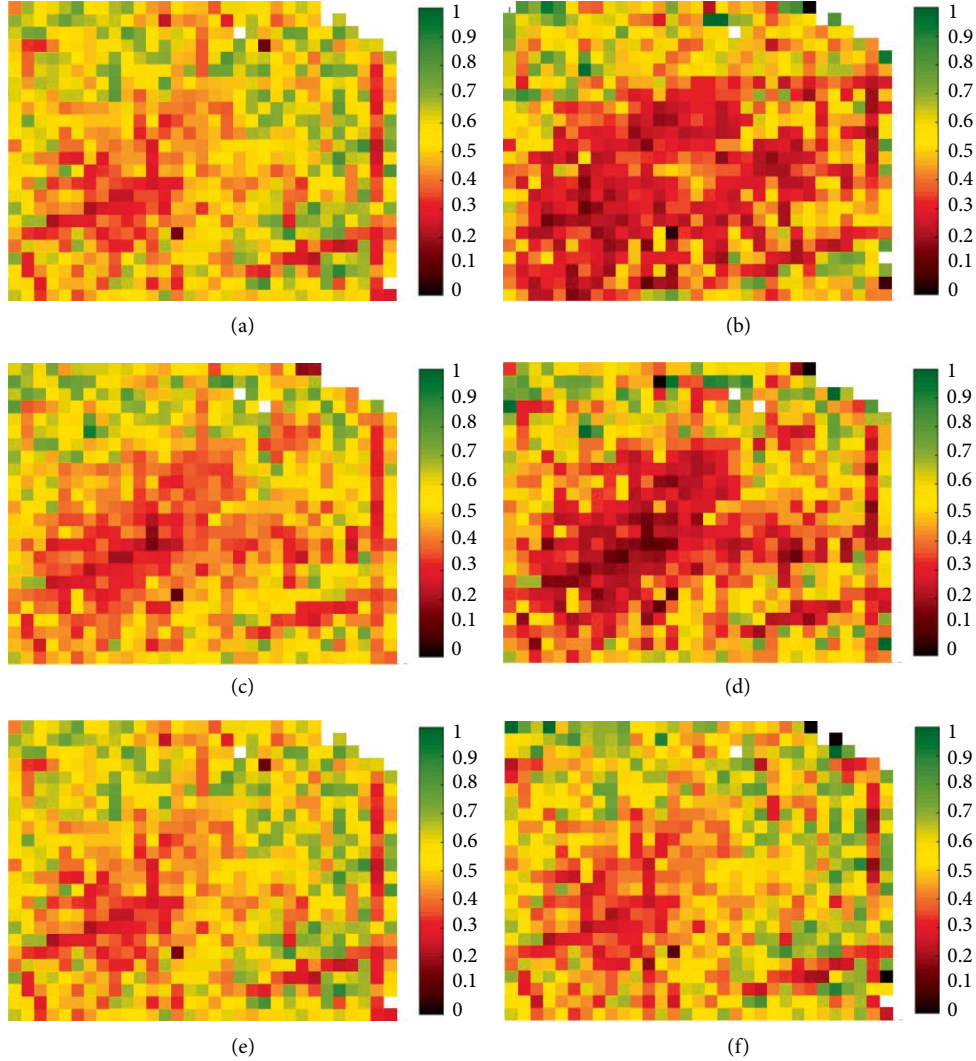| Index	Morning peak		Evening peak		Normal	
	Nonrainy	Rainy	Nonrainy	Rainy	Nonrainy	Rainy
Moran's I	0.3285	0.3818	0.3967	0.3849	0.2640	0.2642
Z Score	17.157	19.935	20.694	20.093	13.793	12.875
P-value	0.001	0.001	0.001	0.001	0.001	0.001

FIGURE 1: The distribution of mean RETTs during different time periods on nonrainy and rainy conditions respectively. The values during peak hours decrease significantly, while the difference during off-peak periods is not that significant. (a) Morning peak, nonrainfall. (b) Morning peak, rainfall. (c) Evening peak, nonrainfall. (d) Evening peak, rainfall. (e) Off peak, nonrainfall. (f) Off peak, rainfall.

difficulty of hailing taxis (the terminology of level of service was also adopted in Wong and Szeto [21]. In terms of TLOS, a straightforward index is the waiting time of passengers. However, passenger demand is difficult to retrieve via raw data or by estimation. Thus, we use vacancy rate, an effective and frequently-used index to measure the performance of taxi operation, to reflect the TLOS instead [36]. If the vacancy rate runs low, it indicates that there are less vacant taxis cruising to search passengers and that passenger wait time could potentially be long. The time vacancy rate can be calculated using the Ratio of Empty/Total Time (RETT), where RETT

not only indicates the taxi supply level from a temporal perspective, but also reflects the difficulty in hailing taxis. The lower the RETT is, the harder the potential passenger could hail a taxi. RETT is mathematically stated as follows,

$$\text{RETT}_k = \frac{\sum_{i=1}^{n} t_{ki}}{\sum_{i=1}^{n} T_{ki}} \times 100\%, \qquad (1)$$

where t_{ki} denotes the elapsed vacant time of taxi i in the kth cell, T_{ki} the total duration of taxi i in the kth cell, and n the taxi number of the kth cell.

Mean RETT is calculated for each cell for each time period with and without rainfalls respectively. RETT for specific periods are illustrated in Figure 1. It can be seen that RETTs during peak hours decrease significantly because of rainfall, while the difference during off-peak periods is not that significant. One could also tell that RETT are potentially spatially correlated, with reds concentrated in the center while greens allocated in the periphery. In order to test whether there is spatial autocorrelation and its magnitude, we refer to Global Moran's I, a spatial autocorrelation coefficient commonly used in spatial statistical analysis. The formula is stated as follows:

$$I = \frac{N\sum_p\sum_q W_{pq}\left(x_p - \overline{x}\right)\left(x_q - \overline{x}\right)}{\left(\sum_p\sum_q W_{pq}\right)\sum_p\left(x_p - \overline{x}\right)^2},\qquad(2)$$

where N denotes the number of cells, x_p and x_q is the RETT of pth cell and qth cell, $\overline{x}$ is the average of all cells, W_{pq} is the spatial weight matrix between cells, which is used to measure the relationship between regions.

The values of Global Moran's I are within $[-1, 1]$. A larger absolute value represents the more significant spatial correlation. Table 1 presents the Moran's test results for specific periods. Moran's I of all results are greater than 0 with P-value less than 0.05 and Z score more than 1.65, indicating that RETT presents a positive spatial correlation, that is, the distribution has internal correlation, which is not scattered and random distribution. Besides, the correlation magnitude varies in different periods. The larger Moran's I during peak hours indicates a higher spatial dependence. Note that the local Moran's I is not analysed here since the spatial aggregation of TLOS has been shown in Figure 1.

To verify if there is a marked change of the TLOS level of between rainy and nonrainy days, we test the changes in RETT between rainy and nonrainy days with an independent sample T-test. The test returns a p-value of 0.002, which is lower than 0.05 and indicates that there is a marked change of difficulty in hailing taxis between rainy and nonrainy days. The average RETT of nonrainy days is 47.61%, while it is 45.45% for rainy days, indicating a lower TLOS and increased difficulty in hailing taxis when it rains. The data corroborates the notion that it is generally more difficult to hail a taxi on rainy days than on nonrainy days.

4.1.2. Spatial Discrepancy of TLOS Changes in the Rain. Taxi services possess a degree of randomness as a response to random demand, which leads to taxi operations typically not being subject to a specific route or region at any given time. To reduce randomness and obtain useful and detailed patterns, we aim to cluster the cells into multiple regions where RETTs are similar to each other. To this end, we cluster cells by their TLOS in nonrainy condition into five levels (A, B, C, D, and E) using the K-Means algorithm. The K-Means algorithm is one of the most effective unsupervised learning algorithms that can solve the well-known clustering problem. In terms of difficulty in hailing taxis, these five levels correspond to very

Table 2: Outcomes of TLOS clustering.

TLOS	Counts	Average RETT		P-value
		Rainy	Nonrainy	
A (very easy)	16 (2.15%)	97.61%	97.41%	0.273
B (easy)	168 (22.58%)	66.82%	63.59%	≤0.001
C (fair)	217 (29.32%)	52.82%	50.65%	≤0.001
D (difficult)	224 (30.27%)	39.51%	36.88%	≤0.001
E (very difficult)	115 (15.54%)	27.61%	25.54%	≤0.001

Figure 2: Spatial distribution of regions with difference TLOS on nonrainy days. Regions with TLOS D and E are mainly bounded within the Middle Ring Road, and TLOS B and C are scattered around the perimeter.

easy, easy, fair, difficult, and very difficult. Table 2 presents the outcome of the cell clusters after applying the K-means algorithm. The gap between the mean of two adjacent levels is 13%. Such a difference between different levels is distinct. For example, compared with taxis in TLOS C, the operation time of taxis in TLOS E will decline by 50%. The area of regions with TLOS D and E account for about 45.82% of the study area (approx. area: 316 km^2), which indicates passengers in nearly half of the study area face considerable difficulty when hailing taxis.

The spatial distribution for different TLOS levels in nonrainy condition is visualized in Figure 2. It is worth noting that the LOS shown in the figure is an aggregated value for the entire cell it associates with. It does not imply the LOS on certain roads but the entire cell. The area of the region with TLOS A is relatively small. Regions with TLOS D and E are mainly bounded within the Middle Ring Road, and the region with TLOS D is concentrated along the Huangpu River, where there is considerable land development. Meanwhile, regions with TLOS B and C are scattered around the perimeter. To test the difference of each TLOS level between nonrainfall and rainfall conditions, an independent sample T-test is performed and the results show that all levels have significant differences between rainy and nonrainy days, except for TLOS A. This indicates that rainfall has a significant impact on regions of low TLOS on nonrainy days. As such, we choose to focus on regions with TLOS B, C, D and E as we conducted the following study.

TABLE 3: The proportion of area and level change in regions with TLOS B, C, D, and E.

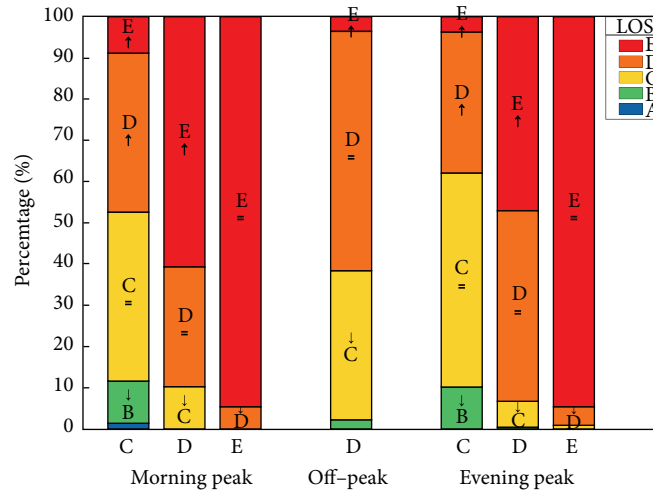| TLOS	The proportion of area harder to hail	The proportion of level change			
		Better level	Equal	One level worse	Two or more levels worse
B	53.74%	1.69%	70.07%	27.68%	0.56%
C	66.22%	7.41%	73.61%	18.98%	0.00%
D	70.82%	4.02%	76.78%	19.20%	—
E	74.67%	2.68%	97.32%	—	—

FIGURE 3: Percentage of TLOS change in each level on rainy days. The downgrades during morning peak is relatively larger compared to evening peak. The TLOS even slightly improved during off-peak when it rained.

In general, 487 square kilometers of area out of the entire study area became more difficult to hail taxis on rainy days, accounting for 70.5% of the total area (a large portion of cells may still maintain the same TLOS but drop to lower RETT when it rains). More in-depth results are presented in Table 3. A portion of the area maintains its TLOS in the rain, while the other areas have even worse levels once it rains. Moreover, taxi hailing becomes proportionately more difficult in regions that originally had worse TLOS.

4.1.3. A Temporal Analysis of TLOS Changes. In the previous sections, we have clustered cells into regions by TLOS and clarified that the rainfall has varying impacts on regions with different TLOS at aggregated level. Next, we will investigate the spatio-temporal changes in TLOS between rainy and nonrainy days. The objective is to isolate certain time periods of interests and investigate the impact of rainfall if it rains within different time periods.

The key study area where there are significant changes between rainy and nonrainy days is extracted according to the *T*-test results. It consists of the regions with TLOS C, D, and E in the morning peak, TLOS D in the off-peak, and TLOS C, D, and E in the evening peak. Focusing on these regions, we investigate the scale of each TLOS on rainy days. As illustrated in Figure 3, it becomes more difficult in hailing taxis in most regions when it rains, which is consistent with our previous findings. Furthermore, for regions with TLOS C and D, the TLOS suffers a greater decline in the morning peak hours than

in the evening peak hours, which indicates that rainfall has a greater impact on taxi operations in the morning peak hours. In contrast, it becomes easier to hail taxis in most of the regions with TLOS D during off-peak hours. As travel demand during off-peak hours can be adjusted in most cases, many optional trips may be cancelled, which leads to a better level of service.

4.2. Factors Affecting Taxi Operations. The statistical results above imply that rainfall generally deteriorates the TLOS of taxis. Such a consequence is essentially a manifestation of the negative impact of rain on taxi operations. Therefore, understanding the mechanism of how rainfall affects taxi operations can help operators develop measures to mitigate the negative impact of rain. Prior to this, we need to discuss which specific factors may affect the performance of taxi operations. In this section, we will focus on the regions with relatively poor TLOS (i.e. C, D, E) and explore factors affecting the RETT values between rainfall and nonrainfall conditions.

4.2.1. Factors Affecting Taxi Operations between Rainfall and Nonrainfall Conditions. Rainfall has a negative impact on taxi operations, which further impacts the TLOS of taxis. We aim to understand the mechanism of this impact as shown in an illustration in Figure 4. For any cell in the study area, the TLOS can be affected by either vacant (green taxis) or occupied (red taxis) taxis driving across that cell. The number of vacant taxis reflects the usable supply in those cells, while the number
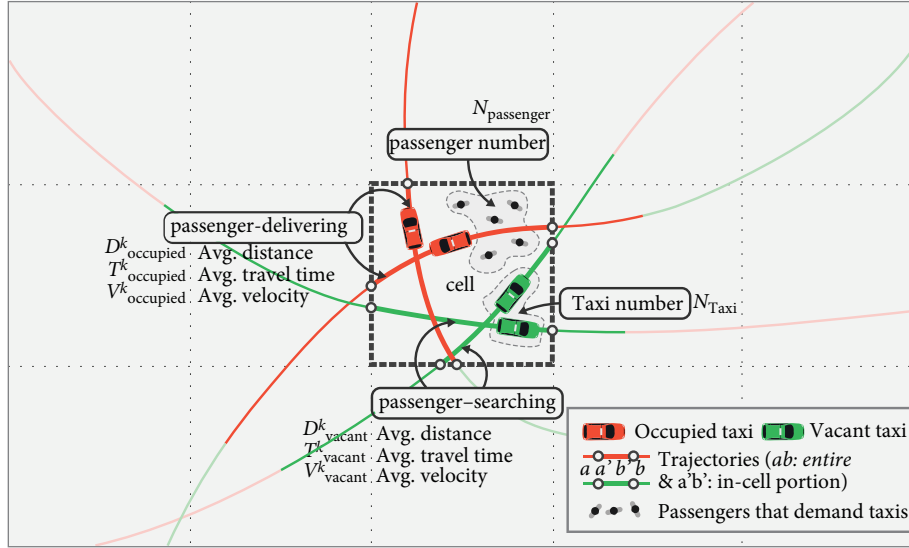
FIGURE 4: Defining factors that affect taxi operations based on search and delivery scenarios.

of occupied taxis reflects how much supply is occupied by demand in this cell or other cells. For this reason, we consider the number of taxis as well as passengers. The number of passengers here represents the demand that has been satisfied, reflecting the overall demand to a certain extent. In addition to this, it is necessary to consider the operational factors including travel time, speed, and distance for either vacant or occupied taxis, since these factors will reflect how long a vacant taxi can approach passengers in a searching process as well as how long it takes an occupied taxi to complete its current trip and return to being a vacant taxi.

In light of this illustration, we formulate the factors affecting taxi operations, including taxi number, passenger number, trip distance, travel time, and speed of vacant and occupied taxis respectively.

To state the factors concisely and clearly, we define two sets, I_k and O_k, which respectively reflect the entire occupied and vacant trajectories that start from/end in/travel through cell k during a given time interval. I'_k and O'_k are the trajectories of I_k and O_k in the cell k. For distance, we focus on the average trip length while cruising or delivering, which is trip-based. Travel time and speed are calculated for the average taxi in the cell, which is cell-based. The specific mathematical expressions are as follows:

$$D^k_{\text{occupied}} = \frac{\sum_{l\in I_k}\text{dist}(l)}{|I_k|}, \tag{3}$$

$$D^k_{\text{vacant}} = \frac{\sum_{l\in O_k}\text{dist}(l)}{|O_k|}, \tag{4}$$

where dist($\cdot$) is the distance of trajectory l.

$$T^k_{\text{occupied}} = \frac{\sum_{l\in I'_k}\text{time}(l)}{|I'_k|}, \tag{5}$$

$$T^k_{\text{vacant}} = \frac{\sum_{l\in O'_k}\text{time}(l)}{|O'_k|}, \tag{6}$$

where time($\cdot$) is the duration of trajectory l.

$$V^k_{\text{occupied}} = \frac{\sum_{l\in I'_k}\text{speed}(l)}{|I'_k|}, \tag{7}$$

$$V^k_{\text{vacant}} = \frac{\sum_{l\in O'_k}\text{speed}(l)}{|O'_k|}, \tag{8}$$

where speed($\cdot$) is the average speed of trajectory l.

Next, we analyze the characteristics of each factor. Figure 5 illustrates the values of each factor between rainfall and nonrainfall conditions during the morning and evening peaks respectively. The left-hand side (green) represents nonrainy days, while the right-hand side (yellow) represents rainy days. P_1 is the result of the independent sample T-test, which estimates whether the changes of the mean value are significant. Furthermore, we are concerned about whether the corresponding cell changes between rainfall and nonrainfall, which reflects the spatial distribution of the indicators. The paired difference T-test is then performed, and the result is recorded as P_2. In addition, the distribution characteristics of data including the mean, the spread, the asymmetry, and the outliers are also displayed.

As for passenger and taxi counts in Figure 5(a), rainfall leads to increase in demand and decrease in supply after rainfall, but the change is minor and there is no significant difference between rainy and nonrainy days in general. The average passenger count for one cell is 89, and the taxi count is 135, which is 1.52 times of passenger count. Nevertheless, the results of the paired difference T-test are significant, which means that the spatial distribution of passengers and taxis changed dramatically. In Figure 5(b), if it rains, the average cruising time decreases from 123.81 s to 112.48 s in the morning peak, while there is no significant change observed in the evening. On the other hand, the average delivery time increases dramatically by at least 48 seconds if it rains, and more long-duration trips are observed. As shown in Figure 5(c), the

$p_1 = 0.791, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 87.26, \mu_{\text{rainy}} = 88.76$

$p_1 = 0.812, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 90.52, \mu_{\text{rainy}} = 89.10$

$p_1 = 0.784, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 134.22, \mu_{\text{rainy}} = 132.04$

$p_1 = 0.530, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 135.93, \mu_{\text{rainy}} = 130.85$

(a)

$p_1 = 0.000, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 123.81, \mu_{\text{rainy}} = 112.48$

$p_1 = 0.128, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 113.79, \mu_{\text{rainy}} = 110.10$

$p_1 = 0.000, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 247.97, \mu_{\text{rainy}} = 295.99$

$p_1 = 0.000, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 227.90, \mu_{\text{rainy}} = 288.49$

(b)

FIGURE 5: Continued

$p_1 = 0.000, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 1.28, \mu_{\text{rainy}} = 1.16$

$p_1 = 0.771, p_2 = 0.589$
$\mu_{\text{nonrainy}} = 1.25, \mu_{\text{rainy}} = 1.26$

$p_1 = 0.926, p_2 = 0.897$
$\mu_{\text{nonrainy}} = 1.41, \mu_{\text{rainy}} = 1.40$

$p_1 = 0.197, p_2 = 0.002$
$\mu_{\text{nonrainy}} = 1.43, \mu_{\text{rainy}} = 1.47$

Distance (km)

D_{vacant}–Morning D_{vacant}–Evening D_{occupied}–Morning D_{occupied}–Evening

(c)

$p_1 = 0.000, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 10.54, \mu_{\text{rainy}} = 7.29$

$p_1 = 0.000, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 8.21, \mu_{\text{rainy}} = 7.44$

$p_1 = 0.357, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 22.07, \mu_{\text{rainy}} = 21.49$

$p_1 = 0.002, p_2 = 0.000$
$\mu_{\text{nonrainy}} = 22.21, \mu_{\text{rainy}} = 20.84$

Velocity (km/h)

V_{vacant}–Morning V_{vacant}–Evening V_{occupied}–Morning V_{occupied}–Evening
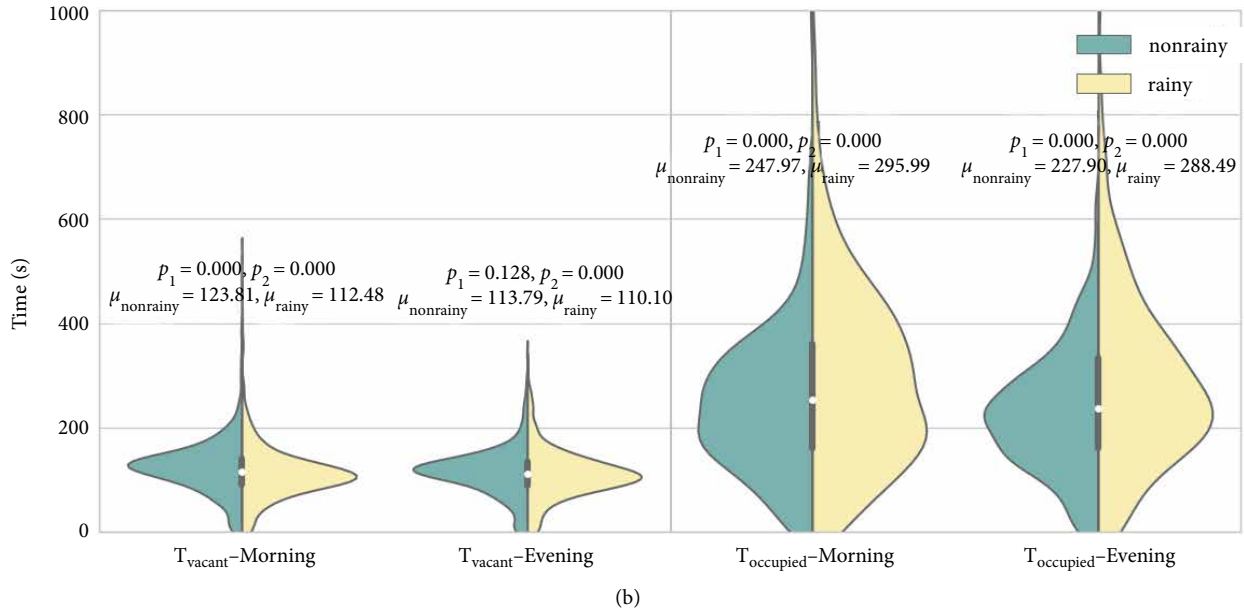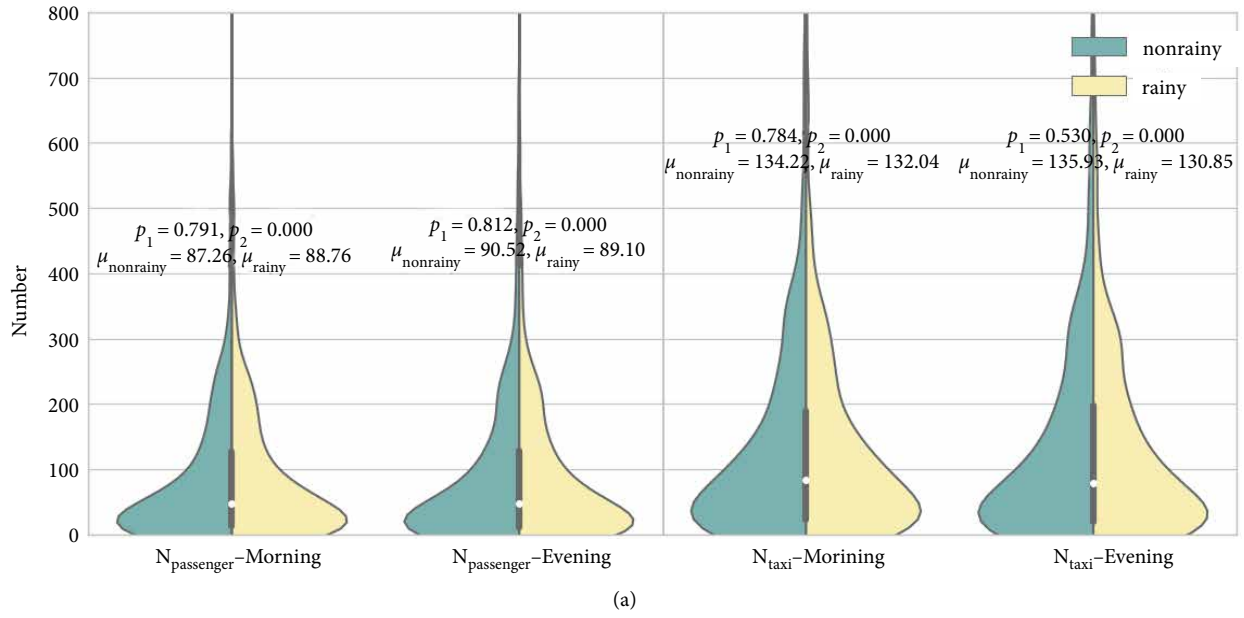
(d)

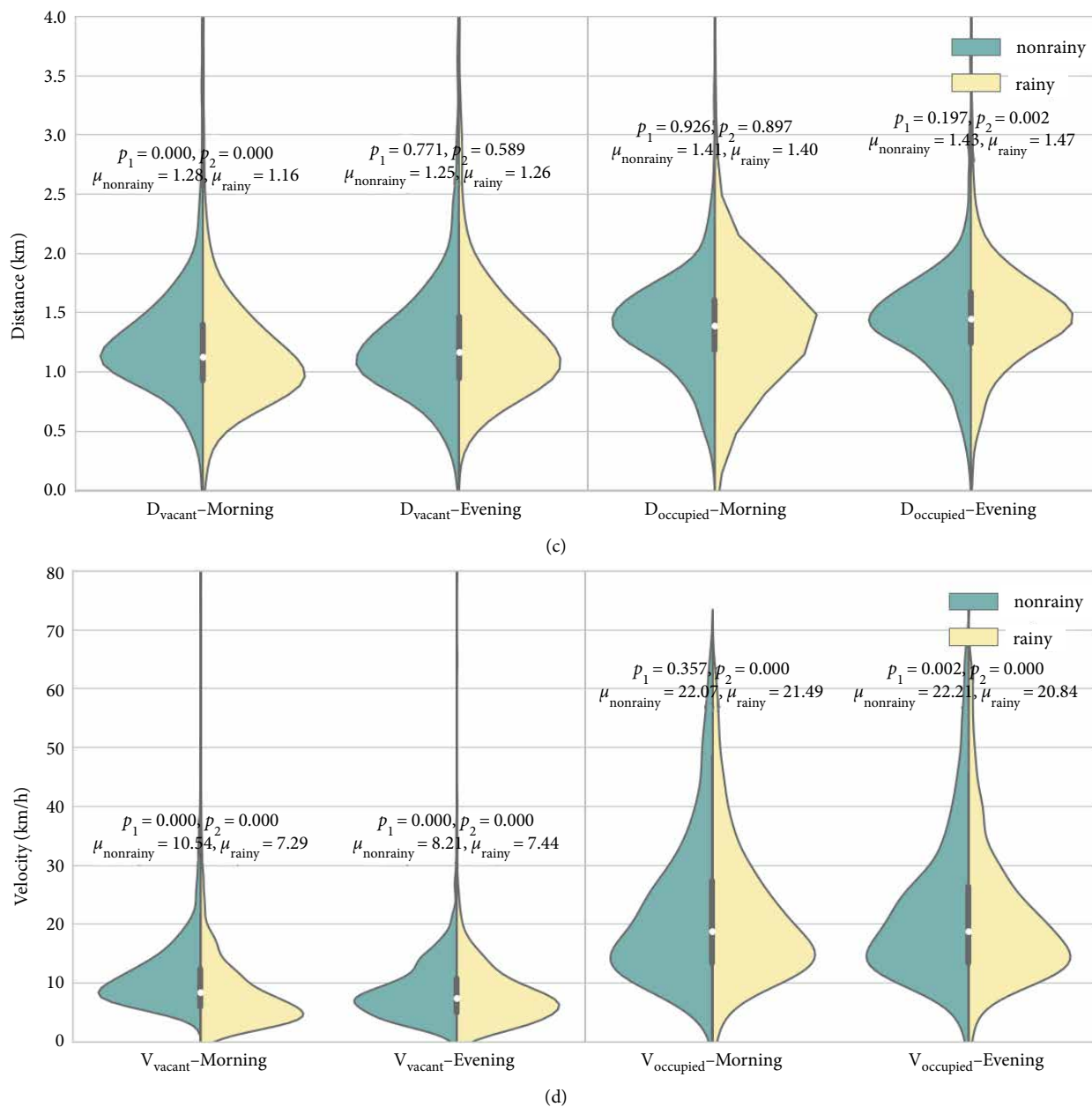FIGURE 5: Data structures and test results of each factor on rainy and nonrainy days. (a) Data distribution of $N_{\text{passenger}}$ and N_{taxi}. (b) Data distribution of T_{vacant} and T_{occupied}. (c) Data distribution of D_{vacant} and D_{occupied}. (d) Data distribution of V_{vacant} and V_{occupied}.

TABLE 4: The variation of the spatio-temporal distribution of supply and demand.

TLOS	Morning peak			Evening peak		
	VRD_{taxi}	$VRD_{passenger}$	VD	VRD_{taxi}	$VRD_{passenger}$	VD
C	0.0097	0.5094	−0.4947	−0.0497	0.2246	−0.3001
D	−0.2902	0.1759	−0.6170	−0.5962	−0.2912	−0.6150
E	−0.8381	−0.6082	−0.6657	−0.9253	−0.8320	−0.5745

effect of rainfall on distance is not as remarkable as on other variables. The distance of cruising is shortened from 1.28 km to 1.16 km in the morning peak and is maintained at around 1.25 km in the evening peak. The travel distance of the delivery fluctuates slightly between 1.40 km and 1.47 km during peak hours. In Figure 5(d), a spatial redistribution of velocity appears in peak hours. Such differences in terms of cruising velocity are greater during the morning peak. The velocity during morning peak hours decreases from 10.54 km/h to 7.29 km/h. The changes in delivery velocity are less pronounced. Velocity remains constant at around 21 km/h during the morning peak and decreases slightly in the evening peak.

In general, the data distribution of factors during cruising and delivering are quite different, while the difference between the morning and evening peaks is small. For $N_{passenger}$, N_{taxi}, and $V_{occupied}$, a spatial redistribution occurs even though the average value remains unchanged.

4.2.2. Redistribution of the Supply and Demand between Rainfall and Nonrainfall Conditions.

As the analysis above indicates, taxi supply and passenger demand should be affected by rainfall but result in no significant statistical difference. A further paired difference test shows that the spatial distributions of these two factors significantly change despite essentially unchanged total numbers. This implies that the rainfall has an impact in spatially redistributing the taxi supply and passenger demand. To have a deeper understanding of how the distributions are relocated, we apply a matrix operation to measure the spatial change with the following method.

The distributions across nonrainy and rainy days are described by matrix P and Q respectively. The value of row i and column j in the matrices (p_{ij}, q_{ij}) represents the quantity of taxis or passengers in the region of longitude i and latitude j. To obtain the change between two weather conditions, a judgment matrix F is implemented. The values in the matrix F are determined by the following rules: if $p_{ij} > q_{ij}$ then $f_{ij} = 1$, if $p_{ij} < q_{ij}$, then $f_{ij} = -1$, and otherwise $f_{ij} = 0$. The variation ratio of each region's distribution (VRD) can be defined as:

$$VRD = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \cdot f_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij} \cdot f_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}}, \quad (9)$$

where VRD denotes the variation ratio of distribution in TLOS C, D or E. m and n are the counts of longitude and latitude. More specifically, $VRD_{passenger}$ denotes the variation ratio of passengers, and VRD_{taxi} denotes the variation ratio of taxis.

The values of the elements in VRD are within the range of [−1,1]. When the quantity of the passengers or the taxis in all

cells in this level increases with rainfall, VRD = 1. On the other hand, when the quantity of all areas decrease, VRD = −1. Furthermore, VRD > 0 indicates that passengers or taxis shift to the area in general, and VRD < 0 indicates that passengers or taxis shift out of the area in general. The greater the absolute value of VRD is, the greater the variation of the distribution is.

As noted above, the quantity of taxis for one grid is about k (k = 1.52 in our study) times that of the number of passengers. Then the variation difference between the supply and demand (defined as VD) is calculated as follows:

$$VD = k \times VRD_{taxi} - VRD_{passenger}. \quad (10)$$

The variation ratio and the difference between supply and demand is summarized in Table 4. One can tell that the taxis drive out of regions with TLOS D and E in the morning peak. Moreover, the decrement during the evening peak is greater than that during morning peak. Since there is no significant change in the total amount of all level regions, taxis likely transfer from regions of TLOS C, D, and E to regions of TLOS A and B. 2.7% of taxis shift from regions with high demand to peripheral areas during the morning peak hours. Moreover, 5.5% of taxis drive out during the evening peak hours.

However, for passengers $VRD_{passenger}$, the quantity in regions of TLOS C increases during both the morning and evening peaks. In regions with TLOS D, the ridership rises in the morning peak and declines in the evening peak. In the hardest-hit areas of regions with TLOS E, ridership always decreases. Passengers may travel using underground public transport, which is smoother by contrast.

The variation difference between the supply and demand (VD) reflects the balance change of supply and demand. The difference being less than zero means that the change in passenger demand is greater than the change in supply. Due to the huge quantitative differences in taxis and passengers, there is a natural shortage in supply. Once it rains, such inherent imbalance of supply and demand is aggravated, and it becomes more difficult to hail taxis. The imbalance of the morning peak is larger than that of the evening peak, and as the difficulty increases, this imbalance is further accentuated.

4.3. Quantitative Impact of Rain on Taxi Operations.

The previous section illustrated the relationship between each proposed factor and RETT as either positive or negative, which completes a qualitative study into the individual impact of each factor on the TLOS. To inspect deeper and provide more quantitative insights for taxi regulators, a further study is performed in this section to quantify the impact and contribution of each factor on the TLOS. Multivariate

TABLE 5: Model results for peak hours.

| Variables	Morning-peak				Evening-peak			
	Coefficient	Sig.	Tolerance	VIF	Coefficient	Sig.	Tolerance	VIF
(Constant)	1.672	.000	/	/	1.049	.000	/	/
$\Delta N_{passenger}$	−0.060	.000	.821	1.217	−0.325	.000	.725	4.101
ΔN_{taxi}	0.022	.000	.776	1.289	0.240	.000	.822	4.473
ΔT_{vacant}	/	/	/	/	40.361	.000	.970	1.331
$\Delta T_{occupied}$	−1.708	.019	.833	1.200	−4.585	.003	.811	2.448
ΔD_{vacant}	14.790	.000	.410	2.441	2.115	.000	.613	3.192
$\Delta D_{occupied}$	−10.188	.000	.387	2.582	/	/	/	/
ΔV_{vacant}	1.159	.000	.803	1.246	1.319	.000	.745	2.106
$\Delta V_{occupied}$	−0.077	.017	.856	1.169	−0.398	.000	.647	2.236
Adjusted R squared	0.674				0.823			
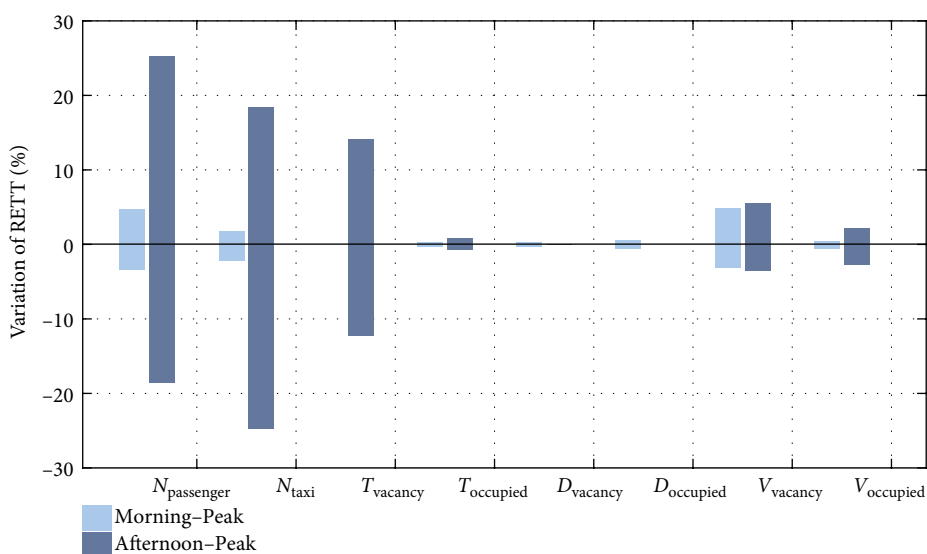
FIGURE 6: Variation of RETT caused by a single factor.

TABLE 6: The adjustment of each factor for a 5% increment in RETT.

	$\Delta N_{passenger}$	ΔN_{taxi}	ΔT_{vacant}	$\Delta T_{occupied}$	ΔD_{vacant}	$\Delta D_{occupied}$	ΔV_{vacant}	$\Delta V_{occupied}$
Morning-peak	83	227	/	2.927	0.338	0.491	4.314	64.935
Evening-peak	15	21	0.124	1.091	3.364	/	3.791	12.563
Maximum-adjustable	78	103	0.349	0.200	0.020	0.062	4.210	6.725

regression analysis, a statistical model for estimating the relationships between variables, is perfect for this purpose, especially as it helps us understand the quantitative impact these factors have by exploring how TLOS changes when any one of them is varied, while the others remain fixed.

Specific regression models such as linear regression [30, 31], and logistic regression [26, 37], are widely applied in studies of impact of weather on transportation system operations. Other complicated regression models, such as ridge and lasso regression are rarely, but still occasionally, used. The dependent variable RETT in this study is a continuous variable, so a logistic model that considers discrete dependent variables is not suitable. This section aims to investigate the impact of

supply-demand ratio on taxi operational characteristics but not the spatial geographic information. Therefore, a linear regression model is adopted herein instead of spatial model.

Models are developed for two peak periods (i.e., the morning evening peak hours) of low TLOS (i.e., C, D, E) regions. Also, since the model should consider the impact of the changes in each factor between rainy and nonrainy days on the RETT, it becomes straightforward to regress the differences in the RETT against the differences in each factor.

4.3.1. Regression Model Specification. The analysis performed above was all cell-based, which provides us with a specific spatial analysis of the TLOS. However, the TLOS is also

impacted by operation factors not only of one single cell but of other neighboring cells. Since we cannot figure out the accurate impact of each cell, we take an average to eliminate spatial dependence. For each variable, we combine the spatial average with the temporal data and divide it into two sets, one for rainy days (set "R") and the other for nonrainy days (set "N"). We then combine and subtract the data from the two sets to construct a new set "V," and the differences are utilized for regression modeling. The process can be expressed as follows:

$$V = \{v | v = r - n, r \in R, n \in N\}. \tag{11}$$

As regions of TLOS C, D, and E during peak hours are intensively affected by rainfall, we consider them as a whole when modeling, which can also expand the data volume. The model specification is as follows:

$$\begin{aligned}\Delta \mathrm{RETT} =&\beta_0 + \beta_1 \Delta N_{\mathrm{passenger}} + \beta_2 \Delta N_{\mathrm{taxi}} + \beta_3 \Delta T_{\mathrm{vacant}} \\ &+ \beta_4 \Delta T_{\mathrm{occupied}} + \beta_5 \Delta D_{\mathrm{vacant}} + \beta_6 \Delta D_{\mathrm{occupied}} \\ &+ \beta_7 \Delta V_{\mathrm{vacant}} + \beta_8 \Delta V_{\mathrm{occupied}},\end{aligned} \tag{12}$$

where β_i denotes the change in response to the unit change in the independent variable, and Δ is the difference value of factors for rainfall and nonrainfall conditions.

The results of the model are presented in Table 5. The adjusted R square value for the morning peak is 0.674, and it is 0.823 for the evening peak, indicating that the results fit the observation well.

5. Results and Discussion

As reflected in the model,$\Delta N_{\mathrm{passenger}}$, $\Delta T_{\mathrm{occupied}}$, $\Delta D_{\mathrm{occupied}}$, and $\Delta V_{\mathrm{occupied}}$ are negatively correlated with the dependent variable. Increasing ridership and its corresponding factors of carrying passengers reduces the RETT, which means that it becomes more difficult to hail a taxi. Meanwhile, ΔN_{taxi}, $\Delta T_{\mathrm{vacant}}$, $\Delta D_{\mathrm{vacant}}$, $\Delta V_{\mathrm{vacant}}$ have the same trend as $\Delta \mathrm{RETT}$. Increasing taxi count and its corresponding factors of cruising improves the RETT to ease the tension of hailing. For the same independent variables, the effect of changing them has on the evening peak is greater than that from the morning peak.

While cruising, travel time plays an important role in impacting RETT during the evening peak hours, while distance and velocity are influential on the RETT for both morning and evening peak hours. Similarly, distance, elapsed time, and velocity of delivering also influence taxi-hailing.

To explore the magnitude of the impact caused by individual factors, statistics are utilized to analyze the data. According to the judgment method of outliers in a box plot, we select the data of each factor in the interval from "$Q_1 - 1.5\mathrm{IQR}$" to "$Q_3 + 1.5\mathrm{IQR}$" as its adjustable range (Q_1 and Q_3 refer to the upper lower quartile of the data, and IQR equal to Q3 minus Q1). With the corresponding coefficients taken into consideration, we obtain the adjustable range of dependent variables caused by these factors. The responsiveness is presented in Figure 6. It can be seen that the influence of these factors on evening peaks is much greater than for the morning peaks, since travel in the morning on weekdays mostly consists

of commuting trips, which are hard to adjust. Compared with $N_{\mathrm{passenger}}$, N_{taxi}, T_{vacant}, and V_{vacant}, the effects of other variables are negligible.

From the perspective of market regulators, measures to ease taxi shortages during rainfall, essentially maintaining a higher TLOS, is of interest. In order to achieve a 5% increment in RETT, a single factor is tested by keeping the other variables constant. Concretely, the quantities to be manipulated and the adjustable maximums are listed in Table 6.

Unfortunately, the adjustment required for $\Delta T_{\mathrm{occupied}}$, $\Delta D_{\mathrm{vacant}}$, $\Delta D_{\mathrm{occupied}}$, and $\Delta V_{\mathrm{occupied}}$ exceeds the maximum adjustable threshold. It is difficult to achieve a 5% RETT increment by manipulating these variables. The other factors, namely $\Delta N_{\mathrm{passenger}}$, ΔN_{taxi}, $\Delta T_{\mathrm{vacant}}$, $\Delta V_{\mathrm{vacant}}$, should be adjusted in order to achieve the desirable outcomes. Through model establishment and discussion, the following conclusions can be drawn:

(1) The number of passengers and taxis are factors of great importance that affect taxi operations. Moderate regulation of the quantity of taxis alone can relieve the tension felt during evening peak hours. On this basis, the situation in the morning peak is exacerbated by rainfall, and other public transportation options should be promoted to decentralize taxi ridership.

(2) Distance, elapsed time and velocity of delivering influence taxi-hailing. Nonetheless, they have short controllable range, which can be easily exceeded and thus they are incapable of serving as effective regulation measures.

6. Conclusion

In this paper, we aim to seek quantitative understanding of the impact of rain on taxi hailing and operations, thereby providing regulators with useful advice that can be used to improve the performance of taxi operations in the rain. First, we use GPS waypoints in Shanghai to confirm that it is significantly more difficult to hail taxis when it rains for most regions. Afterwards, we cluster the cells in the study area into five levels according to their RETT and reveal the spatial and temporal differences in the difficulty of hailing taxis between rainfall and nonrainfall conditions. Next, we investigate eight factors that affect taxi operations to figure out how the impact of rain takes effect. Lastly, we develop multivariate regression model for peak hours between rainfall and nonrainfall conditions and obtain useful findings about how to improve the performance of taxi operations in the rain. The conclusions are three-folded:

(1) It's revealed in our study that a large proportion of area has a declining TLOS (accounting for 70.5% of the total area) when it rains. The TLOS declines remarkably especially in regions with low TLOS to begin with in nonrainy condition. Also, such a decline in TLOS is more considerable during morning peak than during evening peak.

(2) There is no significant change of total taxi supply and passenger demand between rainfall and nonrainfall conditions, which is inconsistent with previous study. However, a dramatic change in the spatial distribution of the supply and demand can be found. Results show that taxis normally transfer from central regions to the periphery of our study area where hailing a taxi is relatively easier in the rain. It is revealed by field data that taxi drivers are relatively prone to return back to city center in rain. That said, previous works on the taxi demand/supply patterns once it rains mainly focused on peak areas while omitting fringe areas, and conclude that the total demand may increase and the total supply may drop. Besides, since heavy rain rarely occurs in Shanghai, the commonly-seen light rains in Shanghai may not be severe enough to draw statistically significant change in total demand/supply.

(3) Passenger demand, taxi supply, search time, and velocity are the significant factors that dominate contributions to declining TLOS on rainy days. Since the urban traffic administration has the right to manage and dispatch taxis, rather than the coarse management of dispatching rush-hour taxis which is a costly and low-yielding choice, we suggest regulators to invest in taxi subsidies or pricing adjustments to redistribute taxis by recalling taxis from peripheral regions to central area. Furthermore, operators should consider the improvement of public transit services on rainy days to shift excess passengers away from taxis and utilize surge pricing to give drivers incentives to head to certain regions.

In the future work, we expect to retrieve taxi hailing app data so that we can obtain passenger waiting time for taxis and have a more accurate estimation of the TLOS. Meanwhile, since passenger demand is also known from such a data source, we aim to design incentive-based management strategy to give drivers incentives and mitigate the difficulty in hailing taxis. Another worthwhile research direction is to explain the spatial and temporal correlation and random effects with consideration of spatial and temporal independent variables in the regression model.

Data Availability

The taxi GPS data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. King and J. Peters, "Taxicabs for improved urban mobility: are we missing an opportunity?" in *Transportation Research Board 91st Annual Meeting*, Transportation Research BoardWashington, DC, USA, 2012.

[2] China Highway, "Statistical bulletin of the development of the transportation industry in 2018," 2018, http://www.chinahighway.com/article/63052.html.

[3] J. Lee, I. Shin, and G. L. Park, "Analysis of the passenger pick-up pattern for taxi location recommendation," in *Fourth International Conference on Networked Computing and Advanced Information Management*, IEEEGyeongju, 2008.

[4] J. W. Powell, Y. Huang, F. Bastani, and M. Ji, "Towards reducing taxicab cruising time using spatio-temporal profitability maps," in *International Conference on Advances in Spatial and Temporal Databases*, SpringerBerlin, 2011.

[5] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, "Recommending friends and locations based on individual location history," *ACM Transactions on the Web*, vol. 5, no. 1, pp. 1–44, 2011.

[6] X. Wang, T. Fan, W. Li et al., "Speed variation during peak and off-peak hours on urban arterials in Shanghai," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 84–94, 2016.

[7] M. Hofmann and M. O'Mahony, "The impact of adverse weather conditions on urban bus performance measures," in *Proceedings of the Intelligent Transportation Systems*, IEEEVienna, Austria, 2005.

[8] C. Zeng and N. Oren, "Dynamic taxi pricing," *Frontiers in Artificial Intelligence and Applications*, vol. 263, pp. 1135–1136, 2014.

[9] C. Kamga, M. A. Yazici, and A. Singhal, "Hailing in the rain: temporal and weather-related variations in taxi ridership and taxi demand-supply equilibrium," in *Proceedings of the 92nd Annual Meeting of Transportation Research Board*, Transportation Research BoardWashington, DC, USA, 2013.

[10] D. Chen, Y. Zhang, L. Gao, N. Geng, and X. Li, "The impact of rainfall on the temporal and spatial distribution of taxi passengers," *PLoS One*, vol. 12, no. 9, p. e0183574, 2017.

[11] C. Camerer, L. Babcock, G. Loewenstein, and R. Thaler, "Labor supply of New York City cabdrivers: one day at a time," *The Quarterly Journal of Economics*, vol. 112, no. 2, pp. 407–441, 1997.

[12] H. S. Farber, *Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers*, Social Science Electronic Publishing, New York, NY, USA, 2014.

[13] Z. Guo, N. H. M. Wilson, and A. Rahbee, "Impact of Weather on Transit Ridership in Chicago, Illinois," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2034, no. 2034, pp. 3–10, 2007.

[14] P. Arana, S. Cabezudo, and M. Peñalba, "Influence of weather conditions on transit ridership: a statistical study using data from Smartcards," *Transportation Research Part A: Policy and Practice*, vol. 59, no. 1, pp. 1–12, 2014.

[15] M. Zhou, D. Wang, Q. Li, Y. Yue, W. Tu, and R. Cao, "Impacts of weather on public transport ridership: results from mining data from different sources," *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 17–29, 2017.

[16] E. Weiner, "The characteristics, uses and potentials of taxicab transportation," *Transportation*, vol. 4, no. 4, pp. 351–367, 1975.

[17] M. Wohl, "The taxi's role in urban America: today and tomorrow," *Transportation*, vol. 4, no. 2, pp. 143–158, 1975.

[18] J. Xu, S. C. Wong, H. Yang, and C.-O. Tong, "Modeling level of urban taxi services using neural network," *ASCE Journal of Transportation Engineering*, vol. 125, no. 3, pp. 216–223, 1999.

[19] H. Yang, Y. W. Lau, S. C. Wong, and H. K. Lo, "A macroscopic taxi model for passenger demand, taxi utilization and level of services," *Transportation*, vol. 27, no. 3, pp. 317–340, 2000.

[20] K. Shaaban and I. Kim, "Assessment of the taxi service in Doha," *Transportation Research Part A: Policy and Practice*, vol. 88, pp. 223–235, 2016.

[21] R. C. P. Wong and W. Y. Szeto, "An alternative methodology for evaluating the service quality of urban taxis," *Transport Policy*, vol. 69, pp. 132–140, 2018.

[22] X. Zhan, S. Hasan, S. V. Ukkusuri, and C. Kamga, "Urban link travel time estimation using large-scale taxi data with partial information," *Transportation Research Part C: Emerging Technologies*, vol. 33, no. 4, pp. 37–49, 2013.

[23] B. Li, D. Zhang, L. Sun et al., "Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset," in *IEEE International Conference on Pervasive Computing and Communications Workshops*, IEEESeattle, WA, USA, 2011.

[24] T. Takayama, K. Matsumoto, A. Kumagai, and N. Sato, "Waiting/cruising location recommendation for efficient taxi business," *International Journal of System Applications, Engineering & Development*, vol. 2, no. 5, 2011.

[25] Y. Gao, P. Xu, L. Lu, H. Liu, S. Liu, and H. Qu, "Visualization of taxi drivers' income and mobility intelligence," in *International Symposium on Visual Computing*, SpringerBerlin, 2012.

[26] G. Qin, T. Li, B. Yu, Y. Wang, Z. Huang, and J. Sun, "Mining factors affecting taxi drivers' incomes using GPS trajectories," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 103–118, 2017.

[27] A. Singhal, C. Kamga, and A. Yazici, "Impact of weather on urban transit ridership," *Transportation Research Part A: Policy and Practice*, vol. 69, no. 69, pp. 379–391, 2014.

[28] A. J. Khattak, "Driver response to unexpected travel conditions: effect of traffic information and other factors," *Journal of Medical Virology*, vol. 62, no. 1, pp. 84–90, 1991.

[29] A. D. Palma and D. Rochat, "Understanding individual travel decisions: results from a commuters survey in Geneva," *Transportation*, vol. 26, no. 3, pp. 263–281, 1999.

[30] V. W. Stover and E. D. Mccormack, "The impact of weather on bus ridership in Pierce County, Washington," *Transportation*, vol. 15, no. 1, pp. 95–110, 2010.

[31] K. Syeed, J. Bunker, and J. B. Lee, "Impact of rain on daily bus ridership: a Brisbane case study," in *Australasian Transport Research Forum 2013 Proceedings*, At Brisbane, Australia, 2013.

[32] Shanghai Municipal Transport Research Center, *Shanghai Transportation Industry Development Report*, Shanghai Municipal Transport Research Center, China, 2018.

[33] B. Chaudhry, S. El-Amine, and E. Shakshuki, "Passenger safety in ride-sharing services," *Procedia Computer Science*, vol. 130, pp. 1044–1050, 2018.

[34] P. Lahkar, *Understanding Use of Transport Network Companies (TNC) in Virginia*, Virginia Tech, Virginia, 2018.

[35] Sohu, "Qiangsheng Holdings currently owns about 12,000 taxi licenses," 2019, http://www.sohu.com/a/319723466_115124.

[36] Y. Cao and X. Luo, "Influence of usage rate of taxi-hailing apps on taxi vacancy ratio," *Computer Engineering and Applications*, 2016.

[37] R. C. P. Wong, W. Y. Szeto, and S. C. Wong, "A cell-based logit-opportunity taxi customer-search model," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 84–96, 2014.