IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 1, 2020, accepted July 18, 2020, date of publication July 29, 2020, date of current version August 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3012689

Taxi High-Income Region Recommendation and Spatial Correlation Analysis

CHANGWEI YUAN[1,2,3], XINRUI GENG[1], AND XINHUA MAO[1,2,3]
[1]College of Transportation Engineering, Chang'an University, Xi'an 710064, China
[2]Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an 710064, China
[3]Xi'an Key Laboratory of Digitalization of Transportation Infrastructure Construction and Management, Xi'an 710064, China

Corresponding author: Xinrui Geng (1245520207@qq.com)

This work was supported by the Key Research Base Project of Philosophy and Social Science of Education Department of Shaanxi Province (Project name: Research on Taxi Governance under the Background of the Internet) under Project 19JZ008.

ABSTRACT Taxis provide essential transport services in urban areas. In the taxi industry, the income level remains a cause of concern for taxi drivers as well as regulators. Analyzing the variation trend of taxi operation efficiency indicators throughout the day, mining high-income orders hot-spots and high-income regions at different periods, will effectively improve the average hourly incomes (AHI) of drivers. This paper selects the order data for each day of holidays, working days, and non-working days through the taxi order dataset of October 2019 in Xi'an. Firstly, we analyze the variation trend of taxi operation efficiency indicators in the three days. We next divide the orders into four income levels based on the Natural Breaks accordingly. Then, we use Tyson polygon and mash map matching methods to visualize the high-income orders hot-spots and high-income regions. It is significantly to analyze and summarize the visualization results. Finally, we compute the Moran'I index to measure the spatial correlation between high-income orders regions and high-income regions. The results show that (1) the number and the spatial distribution of high-income orders hot-spots and high-income regions at different periods are different. (2) Some places are hot-spots, but neither high-income orders hot-spots nor high-income regions. (3) The high-income orders regions and high-income regions have a strong correlation in spatial distribution. This study provides suggestions and insights to taxi companies and taxi drivers to increase their average hourly income (AHI) and enhance the efficiency of the taxi industry.

INDEX TERMS High-income region, hot-spots, Moran'I, taxi operation efficiency.

I. INTRODUCTION

Taxi is an important mode of transportation to meet people's travel demands and an essential part of urban public transport system. Service refusal is a significant problem in the taxicab market, especially in developing countries where policies and regulations have not been well developed against this unpleasant phenomenon. Since 2004, there have been nearly 200 taxi strikes in various places, involving more than 100 cities in China. Especially since 2015, the trend of taxi strikes spreading to large cities is obvious. Taxi strikes in many provincial capitals including Shenyang, Changchun, Jinan, Chengdu, etc. are ultimately due to the low earnings of drivers. There are many reasons why drivers escape from serving during peak hours, i.e., drivers get a lower-than-expected income and they do not know where the high-income regions are. As a result, the drivers would rather suspend the service and take rests. The level of taxi drivers' income has attracted much attention from both taxi drivers and governors. In reality, affected by the taxi cruise service method and profit model, taxi drivers always reduce the empty-loading ratio by increasing the number of passengers to maximize profits [1]–[4].

For some experienced taxi drivers, they can roughly grasp the travel demands in different regions of the city at different periods and the location of high-income passengers, thus they can reasonably select their no-load passenger search route according to their experience in different regions at different periods. Taxi drivers with a higher income usually spend less time in finding new passengers and know some high-income regions [5]. However, the number of order and the income vary in different periods and different regions. For most drivers, they do not have enough experience to determine their range of movement maximizing the average hourly

The associate editor coordinating the review of this manuscript and approving it for publication was Daxin Tian.

income (AHI). As a result, there is an interesting phenomenon that passengers cannot find the taxis, while the drivers cannot find the passengers. The reason for this phenomenon is that the driver is blind in searching passengers, and the operation efficiency is low [6]–[9]. A recent survey in Beijing found that as many as nearly 10,000 taxi vehicles were at rest during the afternoon peak hours, evading service in congested spots and hours. It can be seen that drivers' income is closely related to the traffic situation. Often the better the traffic conditions, the higher the driver's AHI. More passenger traffic means more congestion. Under the traffic congestion environment, the waiting time fee designed according to the existing non-congested traffic conditions cannot compensate for the extra cost of taxis caused by long-term congestion, nor can it effectively regulate the taxi travel demand. This is likely to cause taxis to refuse to load because the traffic congestion takes too long to affect revenue [10].

The previous literature studied on the hot-spots. Taxi hot-spots are not high-income regions to some extent at some moments. However, there are a few researchers focused on the taxi operation efficiency and the distribution of high-income orders as well as high-income areas. The distribution of high-income orders and the high-income regions can be provided to taxi drivers and the taxi management agency, which could reduce drivers' blind search and prevent drivers from refusing services. To improve the operation efficiency of taxis, it is necessary to understand the distribution of passenger sources that bring high profits to taxi drivers. Especially during peak hours, since being trapped in congested traffic for a long time usually reduces the driver's income substantially and congestion premium cannot effectively offset the loss of the driver's income. This is because a hot-spot may not be a high-income region, which may be due to the traffic conditions or drivers habits [11].

Our study aims to identify the distribution of high-income orders and high-income regions at different periods, as well as their spatial correlation to increase the AHI and improve the operation efficiency of the entire taxi market. Taxi operation efficiency will be evaluated to increase taxi drivers' income. Most importantly, this research provides a reference for the taxi management agencies in developing charging methods as well as provides suggestions and insights to taxi companies, taxi drivers to increase their daily income and to enhance the efficiency of the taxi industry.

The potential academic contributions of this paper are as follows: (1) We not only use the Natural Breaks to divide orders into four types, but also use it to get high-income orders in hot-spots and high-income regions. (2)This study analyses the changing in holidays, working days, and non-working days about the taxi operation efficiency indicators. (3)This paper uses *Moran'I* to analyze the spatial correlation between the distribution of high-income orders and high-income regions.

The remainder of this study is as follows. In section II, relevant previous literature is reviewed. Section III presents the methodology framework and data utilized in the study. Some analysis results about taxi operation efficiency indicators and distribution of high-income orders, distribution of high-income regions are shown in section IV. Section V proposes appropriate conclusions and policy recommendations.

II. LITERATURE REVIEW

Over the past two decades, researchers from different fields have studied the problem with taxi drivers' income using different methods. The research on taxi driver income is divided into two categories: one is research that does not rely on GPS data, and the other is research based on massive taxi GPS data. In research field that does not rely on GPS data, in 1997, the authors began to explore the relationship between taxi drivers' working hours and work efficiency and proposed a daily targeting theory to explain the negative wage elasticities using the taxis drivers' average daily income from trip sheets data of New York City [12]. The new theory has sparked a lot of discussions [13], [14]. In the taxi charging field, In 1972, the authors constructed the first aggregated model of taxi price and believed that the free market competition cannot form an effective taxi price, and hence the taxi price should be regulated. All the later authors in this group used the model proposed by the first aggregated model for developing their models and examined them with different market configurations. Each model has added its value to the evolution of the taxicab market modeling. Considering the aggravation of traffic jam, the authors divided the taxi drivers' income into three groups, i.e., starting fee, mileage fee and waiting fee based on the economic demand theory, and proposed an optimization model of taxi waiting fee to analyze the influence of waiting fee on taxi demand and income. The author pointed out that the waiting time becomes longer as the expected profit arises [11], [15]. These studies are basically based on the theoretical level of research on charging methods.

Different from previous studies, more and more researchers have shifted from theoretical study to practical study. In the GPS tracking period, massive amounts of GPS data have changed the drivers' income. Relevant studies have focused on the affecting factor analysis of drivers' benefits, including the personal travel mode, travel time estimation, driving speed change, route choice, and taxi service modeling [13], [16], [17]. In [19], the author is the earliest researcher who calculated the average earnings of drivers through massive GPS data. Different from the [18], in [19], this paper divided drivers into two levels includes top drivers and ordinary drivers. They concluded that the passenger mileage and the passenger time of top drivers are significantly higher than that of ordinary drivers, while the empty driving mileage and time were lower than that of ordinary drivers [19]. In [18], the authors divided the drivers into three groups according to their daily income levels: high income, medium income, and low income [18]. In [13] and [20], the authors divided drivers into two levels according to daily income as well. They developed a visualization system to analyze spatial-temporal trajectory data and identify factors that differentiate the top drivers from ordinary drivers by incomes. They first assumed that the high-income drivers are more intelligent, serve more passengers, search actively, and change strategies less often when they have no passengers to serve. Then, they visualized trajectory data of taxi drivers in Shanghai using two novel visual encoding schemes and validated their assumptions [13], [20]. In [21], the authors proposed an optimal passenger search strategy and optimal route selection during peak hours to improve taxi drivers' earnings based on massive GPS data. By inspecting the delivery process, they then found that income during peak hours is more likely to increase while driving along the routes with good traffic conditions [21]. The authors used a generalized multi-level ordered logit model to study the influence of factors such as taxi market supply and demand ratio, passenger seeking distance, ticket price, and passenger-carrying speed on taxi drivers' earnings [13], [22]. In [16], the authors classified taxi drivers according to income. Then they used the Mash map matching method and DBSCAN(Density-Based Clustering of Applications with Noise) clustering method to compare the temporal and spatial patterns of drivers' cruise trips and parking spots at different income levels [16]. In [23], the autors formulated the passengers' searching process as a Markov Decision Process (MDP) and optimized the taxi drivers' income under uncertainties [23] These studies use massive GPS data to study taxi drivers' income from various aspects. By studying and analyzing the influencing factors of taxi service profits, including the space-time distribution of passenger sources [24],weather factors [25], hotspot regions [26]–[30], mileage or time utilization ratio [31], working time and region selection [32], travelling patterns selection [17], etc.

These previous studies have analyzed, from different perspectives, the difference between the high-income drivers and the low-income drivers in driving mode, weather affected, driving speed, etc. However, the research dimension only focuses on the changes in various indicators in the time dimension.

In contrast, our research focuses on the distribution of high-income orders and high-income regions taking the spatial structure of the taxi high-income regions into account. Tracking the high-income orders and high-income regions can provide driving guidance to improve taxi service quality and drivers' income levels.

III. METHODOLOGY AND DATA

A. DATA DESCRIPTION AND PROCESSING

The dataset used in this paper is the GPS order dataset of taxis, in Xi'an city, China. The data includes 24-hour operational data on October 5, 2019 (holiday), October 9, 2019 (working day), October 13, 2019 (non-working day). During these three days, there were no traffic accidents occurred and no abnormal events that interfered with residents' travel. Of course, the temperature in Xi'an during these three days was around 26 ° C and they were sunny.

TABLE 1. Description of the fields of taxi GPS order data.

Field	Definition	Description
car	*AD95**1	License plate number
log_time	2019-10-03 00:00:00	Order write time
get_on_time	2019-10-03 00:00:00	Get on time
get_off_time	2019-10-03 00:00:00	Get off time
on_lon	108.923742	The longitude of pickups point
on_lat	34.208508	The latitude of pickups point
off_lon	108.90318	The longitude of drop-off point
off_lat	34.237745	The latitude of the drop-off point
money	14.9	The cost of this order
mileage	5.1	Passenger miles
free_mileage	4.4	Empty miles

Therefore, these data discarded the impact of the weather on taxi operations.

As is described in the Table 1, this raw GPS dataset was preprocessed to deal with some invalid data such as incomplete, noisy data, and outliers [33]. The dataset was preprocessed as follows: GPS data with incomplete values were deleted and GPS samples which are out of Xi'an city's graphical coordinates, longitude [107.40 109.49] and latitude [33.42, 34.45], were removed as well. Our study area covers most of the districts of Xi'an according to the characteristics of the frequency distribution of population activities in Xi'an, including Yanta District, Xincheng District, Beilin District, Lianhu District, Weiyang District and Baqiao District. The area (approx. area:826km^2) shown in Fig. 1 is our study area.

To better analyze the difference between high and low income taxi orders at different times, we need to clarify the charging method of taxis in Xi'an. Then, the fare of a trip is calculated as in formula (1):

$$\mathrm{F}(\mathrm{d}) = P_0 + (d-3) * P_1 + (d-12) * P_2 \tag{1}$$

where d is the distance of the occupied trip, P_0, P_1 and P_2 are standard fares according to taxis' fare system in Xian city(in 2019) as in Table 2. During the operating hours, if it is not caused by the driver, resulting in a speed of $\leq$10 km / h, the total price of one kilometer is automatically accumulated for 5 minutes from 7:00 to 9:00, and the total is 4 during 17: 00-19: 00 One kilometer freight rate per minute.

From Table 2, we can know that the taxi fare system of Xi'an is affected by multiple dimensions. On the one hand, it is affected by different moments all day and different spaces. On the other hand, it is affected by traffic conditions in the morning and evening peak hours. Therefore, for analyzing the taxi order income, it is necessary to analyze not only the

TABLE 2. Taxi fare system of Xi'an city in 2019.

Parameter	Item	Price	
		From 6am to 23pm	From 23pm to 6am
P_0	with 3 km	8.5 CNY	9.5 CNY
P_1	with 3-12 km	2 CNY/km	2.3 CNY/km
P_2	Exceed 12 km	3 CNY/km	3.45 CNY/km

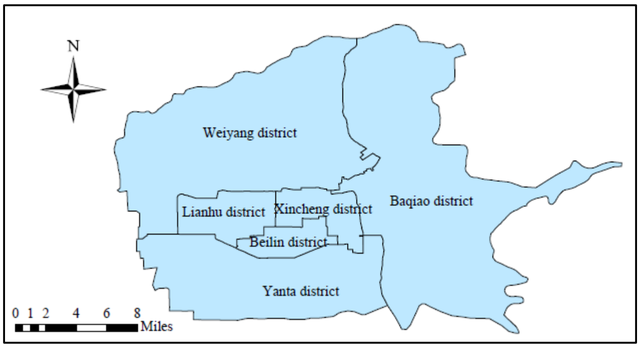


FIGURE 1. The administrative district in study area.

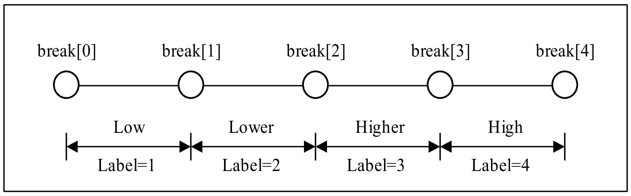


FIGURE 2. The order classification diagram.

time but also space. The income of the taxi driver in this paper is directly obtained from the money field of the taxi GPS Order Data.

B. DEFINITION FOR OPERATION EFFICIENCY INDICATORS

Generally, when studying taxi travel characteristics, taxi operation efficiency indicators will show periodic changes. In order to explore and analyze the temporal and spatial travel characteristics of urban taxi passengers, in this paper, order quantity, mileage utilization rate, average hourly incomes, average passenger time, average passenger mileage, and average empty mileage are selected for analysis. The time-varying feature analysis uses time as the horizontal axis and the time interval is 1 hour. We analyze the changes in various indicators over time [34]. The definition of each taxi operation efficiency indicator is as follows.

Denifition 1: The order quantity(N) refers to the total number of passengers carried by the operating-taxis at that moment.

$$\mathrm{N} = \sum_{i=1}^{k} j \tag{2}$$

where **N** denotes the order quantity of operating-taxis in this period; **i** = 1,2 …k denotes the number of rental cars operating in this period; **j** = 1,2…n denotes the order amount completed by each car on that day.

Denifition 2: Mileage utilization rate(MUR) refers to the ratio of the total passenger mileage of operating-taxis during this period to the total mileage of operating-taxis on that day. This indicator reflects the passenger-carrying efficiency of the vehicle. A higher MUR means that the passenger-carrying ratio is higher and the empty driving ratio is lower. For passengers, if there are not enough vehicles available for hire, passengers' waiting time will increase, which will cause a tight supply-demand relationship. If the MUR is low, the proportion of empty vehicles is high, and it is more convenient for passengers to take a taxi, but the economic efficiency of the operator will decline [35].

$$\mathrm{MUR} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} m}{(\sum_{i=1}^{k}\sum_{j=1}^{n} m) + (\sum_{i=1}^{k}\sum_{j=1}^{n} e)} \tag{3}$$

where **m** denotes passenger mileage per order and **e** is the empty travel passenger mileage per order.

Denifition 3: The average hourly income(AHI) is the ratio of the total incomes of the operating-taxis during the period to the total of the passenger-carrying time of the operating-taxis during the period.

$$\mathrm{AHI} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} s}{\sum_{i=1}^{k}\sum_{j=1}^{n} p} \tag{4}$$

where **s** represents the revenue of each order; **p** represents the length of passenger time per order.

Denifition 4: The average passenger time(APT) refers to the ratio of the total passenger-carrying time of the operating-taxis during the period to the order quantity completed by the operating-taxis during the period.

$$\mathrm{APT} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} p}{N} \tag{5}$$

Denifition 5: The average passenger mileage(APM) refers to the ratio of the total passenger mileage of operating-taxis during this period to the order quantity completed by the operating-taxis during this period. It is worth mentioning that the total passenger mileage is obtained by adding up the passenger mileage of all orders at this period.

$$\mathrm{APM} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} m}{N} \tag{6}$$

Denifition 6: The average empty mileage(AEM) refers to the ratio of the total empty mileage of operating-taxis during this period to the order quantity completed by the operating-taxis during this period.

$$\mathrm{AEM} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} e}{N} \tag{7}$$

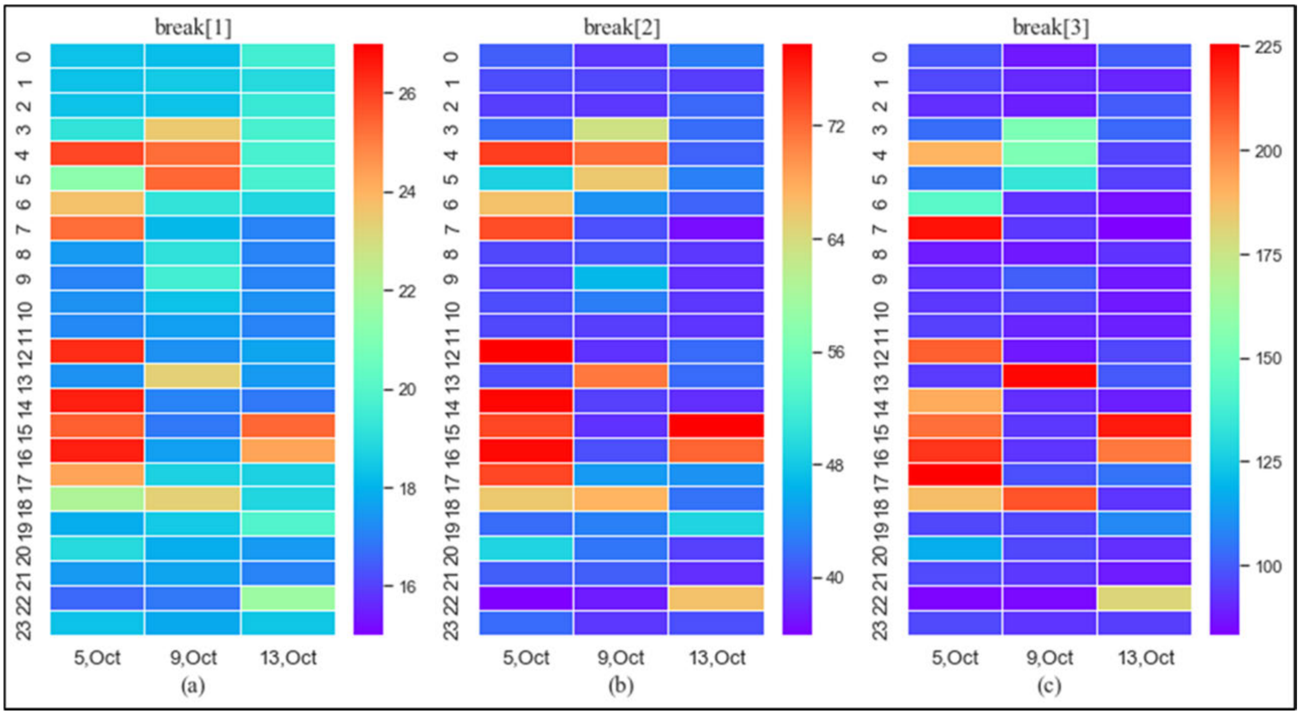


FIGURE 3. Heat map distribution of natural breakpoint values.

C. METHODOLOGY

1) NATURAL BREAKS FOR CLASSIFICATION

Natural Breaks uses the concept of clustering, namely, maximizing the internal similarity of each group, and the biggest difference between external groups. Additionally, we also consider the range and number of elements between each group as close as possible. Iteratively compares each group and the sum of the squared difference between the mean of the elements in the group and the observed value to determine the best arrangement of the values in the group. The calculated best classification can determine the breakpoint of the value in the ordered distribution to minimize the sum of the squared differences (SDCM) within the group.

Denote a number of columns $\{x1, x2, \ldots xn\}$. If the number of columns is divided into 2 types according to the natural discontinuity method, the method is as follows:

Step 1: Calculate the "Sum of Squared Deviations of Mean" (SDAM).

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{8}$$

$$\text{SDAM} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{9}$$

Step 2: Iterate through each range combination to find the smallest "sum of square deviations of category means" (SDCM_ALL).

$$SCDM_j = \sum_{i=1}^{j} (x_i - \frac{\sum_{i=1}^{j} x_i}{j})^2 + \sum_{i=j+1}^{n} (x_i - \frac{\sum_{i=j+1}^{n} x_i}{n-j})^2 \tag{10}$$

Step 3: Calculate "Variance Goodness of Fit" (GVF) GVF = (SDAM-SCDM) / SDAM, the value of GVF is between 0-1, 1 means excellent fit, 0 means extremely poor fit.

$$\text{GVF} = max\{GVF_j = (\text{SDAM}-SCDM_j)\big/SDAM\} \tag{11}$$

where i = 1,2, …n; j = 1,2,…n.

Most scholars in the past divided orders into three, four, or five levels according to the money field. This paper uses the Natural Breaks to divide the money fields of all orders in each period. When divided into three levels, the GVF value is too small and the fitting effect is poor, which is not conducive to research; when divided into five levels, the GVF value is closer to 1 and the fitting effect is good. However, due to too many levels, the number of high-income orders in this period is particularly small, which is not conducive to subsequent research on the distribution of high-income orders. Therefore, this paper chooses to divide the orders at each period into four levels based on previous research experience and the division basis of the Natural Breaks, which is more conducive to the subsequent work of this paper and better achieve the research purpose of this paper. Table 3 shows the results of the breakpoint values of orders divided at different periods; Fig. 3 describes the comparative heat map of breakpoint values, which can clearly see the difference in order classification at different times.

In this study we classify orders into four levels, low, lower, higher and high, by using Natural Breaks. They are break[0], break[1], break[2], break[3] and break[4]. The schematic diagram of order classification is shown in Fig. 2. From formula (1), P_0 is 8.5CNY. In this paper, we define break[0]

TABLE 3. Breakpoint value of order income based on natural breakpoint method.

time	break[1]			break[2]			break[3]			break[4]		
	5,Oct	9,Oct	13,Oct	5,Oct	9,Oct	13,Oct	5,Oct	9,Oct	13,Oct	5,Oct	9,Oct	13,Oct
00:00-00:59	18.3	18.2	19.6	41.1	38.9	42.8	98.8	87.7	100.6	497	300.7	359.4
01:00-01:59	18.3	18.5	18.9	40	39.7	39.2	96.9	90.6	90	440.8	406.8	333.4
02:00-02:59	18.3	18.3	19.4	39.2	39	41.5	91.7	88.8	99.9	334.1	427.6	411.1
03:00-03:59	19.2	23.5	19.7	41.8	63.3	41.9	103.3	153.4	102.3	315.6	434	390
04:00-04:59	25.9	25.3	19.7	74.4	71.6	41.2	189.6	153.7	95.8	442.2	356.8	493.5
05:00-05:59	21.3	25.4	19.7	48.8	65.7	43	105.4	134.4	95.4	384.4	445.2	261.6
06:00-06:59	23.7	19.2	18.8	66.3	44.2	41.4	144.1	92.4	86.5	399.7	317.7	266.1
07:00-07:59	25.3	18.1	17.1	73.6	40.3	36.7	222.8	93.7	83.2	477.9	393.7	312.4
08:00-08:59	17.5	19.1	17.1	39.7	40.6	39	88.6	88.2	92.3	401.3	429.7	475.5
09:00-09:59	17.1	19.6	17.1	39.4	46.8	38.5	92.2	100.6	88.2	397.3	375.7	244.9
10:00-10:59	17.3	18.3	17.3	40.1	42.9	39	93.6	96.1	88	418.7	304.6	360.1
11:00-11:59	17.2	17.6	17.1	39.8	39.2	38.7	94.9	90.4	89.2	450.1	491.2	354.8
12:00-12:59	26.3	17.3	17.7	77.8	38.6	41.7	208	88	96.4	465.4	232	309.6
13:00-13:59	17.3	23.3	17.5	40.1	71.2	41.7	94.3	224.1	99.6	442	485.8	425.8
14:00-14:59	26.5	17.1	16.9	77.4	39.4	38.5	192	92	89.2	419.8	425.8	360
15:00-15:59	25.5	16.9	25.4	73.9	38.6	77.8	204.7	93.7	220.9	453.5	361.4	485.2
16:00-16:59	26.5	17.6	24.3	77.3	40.3	72.3	215.8	92.9	203.2	493.8	404.8	407.1
17:00-17:59	24.3	18.7	18.7	73.9	44.7	44.1	225.7	98.2	104.4	493.2	349.8	433.9
18:00-18:59	22.1	23.3	18.8	65.7	67.2	42.2	187.6	210.5	93.2	494.5	456.7	349.3
19:00-19:59	17.9	18.5	19.9	41.9	43	49	96.4	96.4	108.9	349.6	323.4	460
20:00-20:59	18.9	17.9	17.5	49	42.3	39.5	117.4	96.3	91.8	434.8	447.4	375.7
21:00-21:59	17.5	17.7	17.1	41	41.1	38.5	97	93.4	88.7	398.5	382.9	357.6
22:00-22:59	16.6	16.9	21.7	35.9	37.5	66.2	84.1	85.3	180.1	401.8	300.6	493.5
23:00-23:59	18.3	17.8	18.4	41.7	39	40.3	96.7	93.1	94.4	476.3	452.9	482.7

is 8.5CNY. In order to avoid the impact of abnormally high values on the classified data, orders with incomes >500CNY are eliminated. The occurrence of abnormally high values is mainly due to the possible continuous operation of taxi drivers without shutting down the meter, resulting in the existence of abnormally high values. For example, at 5:00am on October 8,2019, there is an order of 110005CNY, which is impossible. Table 3 shows the hourly breakpoint values for the dates studied in this paper. We define high-income orders as dividing all orders in each period using the Natural Breaks method. The orentralders with labels equal to 4 are high-income orders.

From Table 3 and Fig. 3, we can conclude that it is reasonable to divide orders into 4 levels based on the Natural Breakpoint method, and the distribution of the four levels of orders within each period is relatively stable. From a horizontal perspective, during the period from 14:00 to 19:00 on October 5, the break[1], break[2], break[3] values are higher than October 9 and October 13. This characteristic shows that during this period, the difference between the two types of holiday orders is obvious. From a vertical perspective, during the period from 3:00 to 6:00, 13:00 to 14:00 and 18:00 to 19:00 on October 9th, from 4:00 to 5:00, 6:00 to 8:00 and 13:00 to 18:00 on October 5th, from 14:00 to 16:00 on October 13th, the breakpoint values of these periods are higher than those of other time periods in the day, indicating that the average value of order income during this period is higher, which provides a basic foundation for the average hourly income(AHI) analysis in section IV.B.

2) THIESSEN POLYGON METHOD FOR DIVIDING THE STUDY AREA

In order to more clearly describe the level of taxi order income in different regions, the study area is divided into 818 units based on the Tyson polygon method. The Thiessen polygon

TABLE 4. The number of Tyson polygons in each administrative area.

Administrative district	Area（km^2）	Quantity
Baqiao	318.34	351
Weiyang	263.24	283
Yanta	149.82	171
Lianhu	38.34	48
Xincheng	31.81	46
Beilin	23.05	36

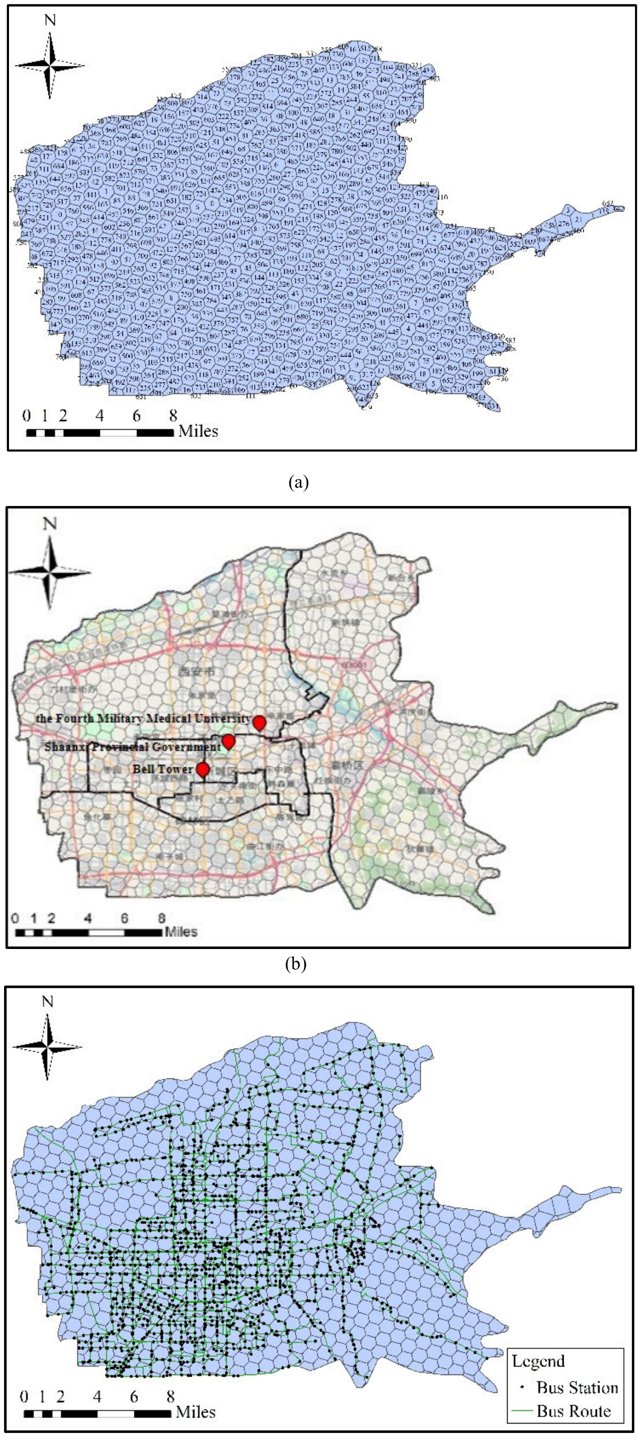
method is an interpolation analysis method proposed by the Dutch climatologist Thiessen [36]. It was originally used to calculate the average rainfall of discretely distributed meteorological stations. All neighboring points are made into vertical bisectors, and the vertical bisectors are connected and combined in sequence [37], [38]. The method overcomes the shortcomings that the division effect is not obvious due to the randomness of spatial distribution [39]. Based on the coordinates of the center point of the regular hexagon of the Didi taxi platform in Xi'an, Arcgis 10.2 was used to quickly generate a Tyson polygon, and the generated Tyson polygon was close to a regular hexagon. The results are shown in Fig. 4. Use the Voronoi subdivision to optimize urban space and use Tyson polygons for higher performance in spatial modeling [40]–[42]. Table 4 describes the number of Tyson polygons in each administrative area.

As observed in Table 4, the number of polygons in each administrative area is divided according to the size of the area. Among them, there are 935 polygons in these 6 administrative districts. Exploring the overlapping polygons of each administrative area, referring to Fig. 5(a), it can be found that the study area is divided into 818 polygons. From Fig. 5(c), it can be found that the public service, community, commercial, and other bus systems within the Second Ring Road are developed. For example, in Bell Tower, Shaanxi Provincial Government, the Fourth Military Medical University, and other places, not only are there many bus stops, but also are there many bus routes. Of course, the bus system of some scenic spots outside the Second Ring Road is also very complete, such as the Great Wild Goose Pagoda, the Small Wild Goose Pagoda, and other places.

3) SPATIAL CORRELATION METHOD

The spatial relationship curve of any two points is not a simple proportional relationship with distance. To explore whether high-income orders and high-income regions are connected, this paper use *Moran's I* to measure the overall characteristics of the spatial correlation degree of the high-income orders and high-income regions. The *Moran's I*, which can reflect the spatial aggregation of the whole region, can be defined in formula (12),(13),(14) and (15) [43].

$$Moran'sI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}(y_i-y)(y_j-y)}{\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}\sum_{i=1}^{n}(y_i-y)^2} \tag{12}$$



(a)
(b)
(c)

FIGURE 4. Thiessen polygons of the study area: (a) is study area of Tyson polygon with numbers; (b) is the study area of Tyson polygon based on Open Street Map;(c) is the study area of Tyson polygon contains the bus lines and stops.

$$Z = \frac{I - E[I]}{\sqrt{V[I]}} \tag{13}$$

$$E[I_i] = -\frac{\sum_{j=1,j\neq i}^{n} w_{ij}}{n-1} \tag{14}$$

$$V[I] = E\left[I^2\right] - E[I]^2 \tag{15}$$

Moran's I is the global Moran index of high-income orders and high-income regions. Where n denotes the total number of units, n = 818; Wij is the space weight matrix, when i and j are adjacent space positions, Wij = 1; Otherwise, Wij = 0, based on Rook adjacency relation; yi and yj are the total incomes of this period of unit i and unit j respectively. y is the average income of all units; Z is the multiple of the standard deviation of the score; E[I] is the weighted mean of the total incomes in this period. In actual calculations, each parameter is regarded as a cluster with statistical significance. $|Z| > 1.65$ will be regarded as statistically significant, and the confidence level of the statistical significance is set to 90%. If the Z score passes the significance test, when the *Moran's I* is positive, it indicates a clustering trend. The clustering trend means the total incomes of orders for 818 units in Xi'an have neighboring elements that contain equally high or equally low attribute values. When *Moran's I* is negative, it indicates a discrete trend, that is, the total incomes of orders for 818 units in Xi'an have adjacent elements with different values.

IV. RESULTS AND DISCUSSIONS

A. THE VARIATION TREND IN THE DAY

The operational efficiency of the urban taxi is the essential reflection of scale and the efficiency of technology, which is the certain allocation of taxi resources (service capacity) and the satisfaction of taxi users (service level). With the fixed taxi configuration, the higher the customer's satisfaction degree of taxi travel demand is, the higher the operating efficiency will be; otherwise, the lower the efficiency will be.

From Fig. 4, the number of taxi-operated vehicles in Xi'an in October 2019 is 9,500 ~ 10,000. Fig. 5 shows the daily taxi revenue of Xi'an city in October 2019. We can conclude that the average daily income of taxis in Xi'an is 700 ~ 730 CNY, of which the average daily income of holiday taxis in Xi'an is 688.26 CNY, the average daily income of taxis on weekdays is 724.18 CNY, and the average daily income of non-working day taxis is 732.35. Fig. 5 indicates that incomes within these 31 days have a mean of 718.82 and a standard deviation of 29.38 CNY. The average income of taxis for 14 days was lower than 718.82. Among the 7-day National Day holiday, the average revenue of taxis on October 1, 2019, was 743.66, which was higher than 718.82. For another 17 days, the average income of the taxi was higher than 718.82. The average daily income of holidays is obviously lower than that of working day and non-working day, which is consistent with the [44]. On the one hand, in [44], this paper pointed out that there is less demand for rigid travel during holidays, and the demand for taxis has decreased compared to usual. On the other hand, it is related to the resting behavior of taxi drivers, more taxi drivers choose to rest on holidays. Taxi trips are different from regular public transport trips such as buses and subways [34].

Next, this paper uses the definition of taxi operational indicators in section III.B to calculate the original statistics for October 5, October 9, and October 13, as shown in Table 5 V

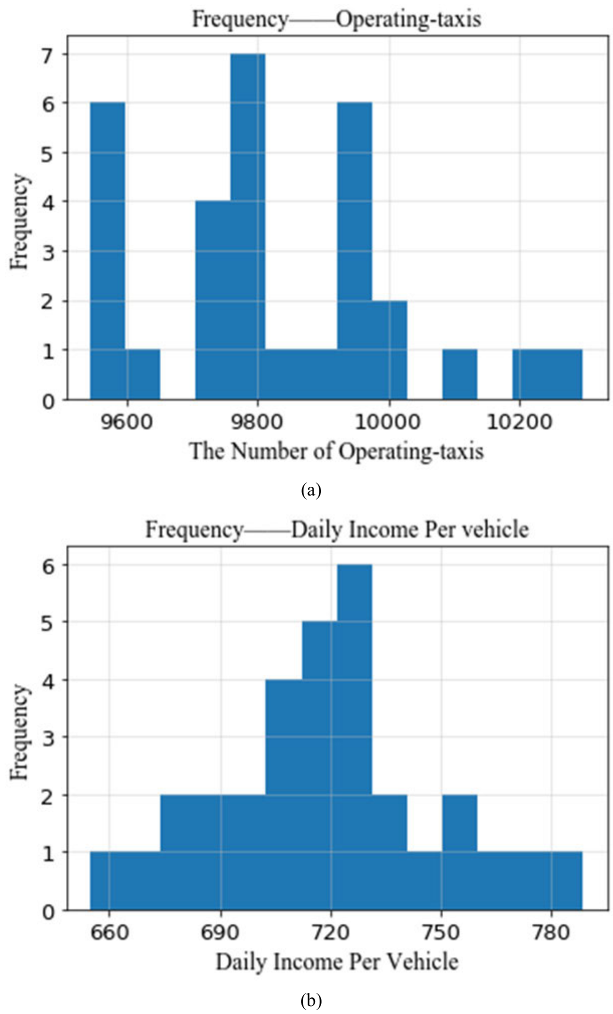
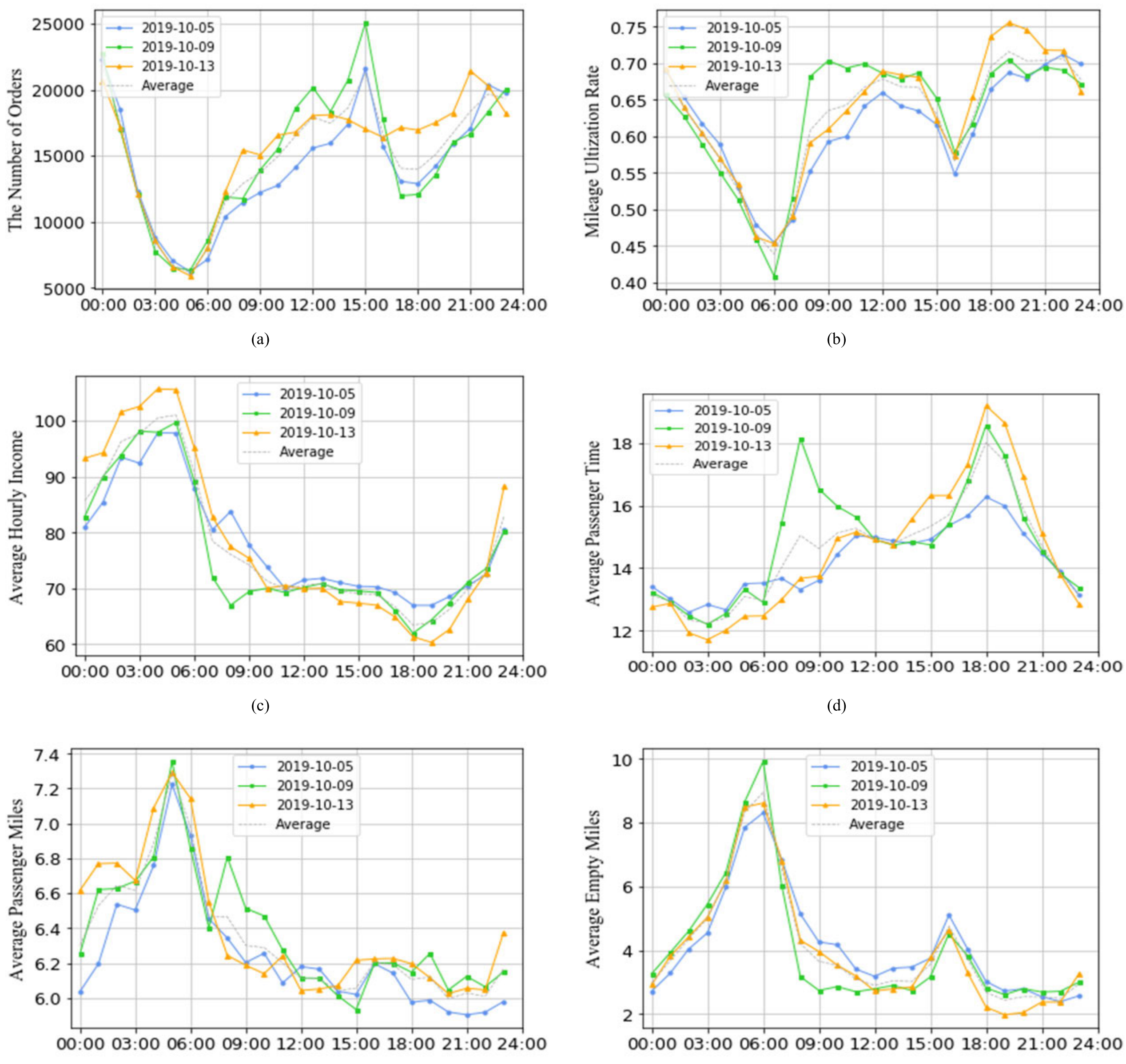


FIGURE 5. Overall operating characteristics of Xi'an taxis in October 2019: (a) is the frequency distribution histogram about the operating-taxis; (b) is the frequency distribution histogram about the daily income per vehicle.

and Fig. 6. Table 5 can reflect the fundamental situation of the three-day taxi operation. Fig. 6 can reflect the changes in these operational efficiency indicators during the day. Without considering the influence of space factors, the Travel pattern in Fig. 6 is defined based on time factors and is used to describe the temporal trend of the overall operating efficiency of taxis in Xi'an in the study area.

As Table 5 shows, the average MUR in these three days are 61.36%, 63.04%, and 63.24%, which means that the MUR of taxi drivers in Xi'an can basically reach about 60%. The maximum value is 75.53% and the minimum value is 40.86%. Some scholars have pointed out that when the mileage utilization was more than 67% and the average waiting time to catch a single taxi was more than 259s, the fleet size of cruising taxis was insufficient. Hence, the number of taxis should be increased [1].

The average AHI in these three days are 77.98CNY, 76.93CNY, and 78.94CNY respectively. The lowest value and the highest value of AHI appear on October 13, which are



(a) (b) (c) (d) (e) (f)

FIGURE 6. **The curves of the operational efficiency indicators in Oct 5, Oct 9 and Oct13.**

60.32CNY and 93.53CNY. This shows that AHI in different periods has a large gap. It is necessary that the change of AHI can provide a reference for the driver to select the time to take a break. As a result, they probably avoid missing periods with higher AHI.

The APT is maintained at about 14 minutes, and the APM of each order is about 6 kilometers, which shows that people often choose taxi to travel a short and medium distance. In [45], the authors pointed out that the average distance of highly efficient taxi trips is much longer than that of inefficient trips. The average AEM is basically maintained at about 4 kilometers.

From Fig. 6(a), it can be known that the time distribution characteristics of passengers are closely related to the passenger travel pattern. The order quantity curve appears two low-value. From 3:00 to 7:00, the order quantity is a V-shaped curve. The first is at 05:00~06:00 when the order quantity is the smallest in the day. With the arrival of the morning peak, the order quantity began to increase rapidly, reaching a maximum at 15:00. This conclusion is consistent with the conclusion of the most researchers [46]–[49]. The second is at 17:00~18:00 when Xi'an began to enter the evening peak and traffic volume rose sharply. On the one hand, it is due to traffic congestion in Xi'an, the driving time is longer and the frequency of passengers getting on and off the taxis is

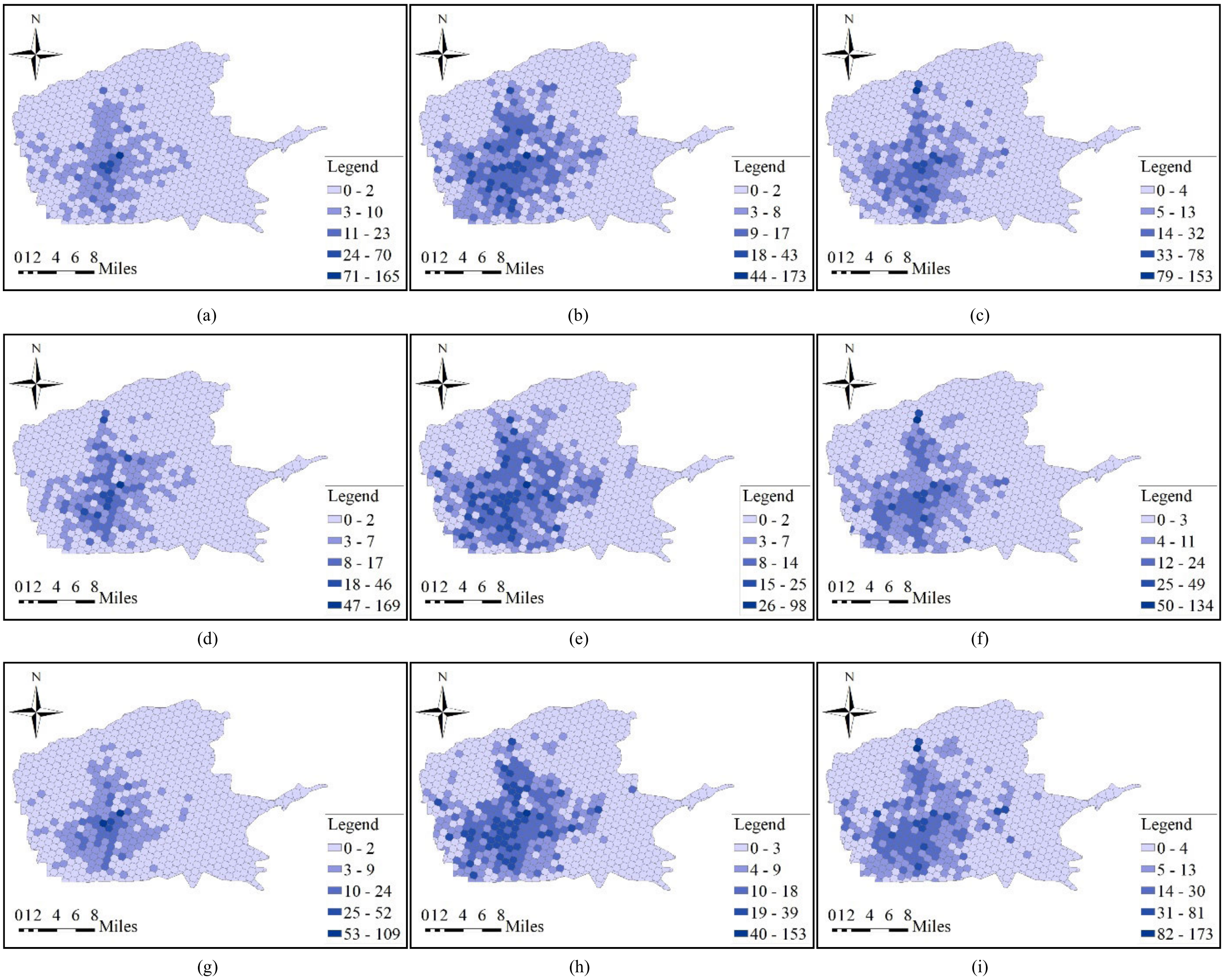


FIGURE 7. The spatial distribution of high-income orders at different time: (a), (b), (c) represent at 5 o'clock, 8 o'clock, 18 o'clock on October 5 respectively. (d), (e), (f) represent at 5 o'clock, 8 o'clock, 18 o'clock on October 9 respectively. (g), (h), (i) represent at 5 o'clock, 8 o'clock, 18 o'clock on October 13 respectively.

low. On the other hand, most taxi drivers choose to take a break at this time, so that the total order quantity goes down. According to related research, the evening peak of Xi'an is determined to be after 17:00 [47]. Unlike the order quantity curve of the non-working day(October 13, 2019), both the order quantity curve of the holiday(October 5, 2019) and working day(October 9, 2019) have similar trends. The non-working day order volume curve does not show particularly low values. From 16:00 to 22:00, the trend of orders on the non-working day is relatively flat, and the order quantity is much more than that of holidays and workdays. This phenomenon occurs because the travel demand after 16:00 on the weekend is mostly non-rigid travel demand for leisure, tourism, visiting relatives, and friends.

From Fig. 6(b), it can be seen that the curve of the mileage utilization rate has two low values as the order quantity. Their first low value occurs at the same time as the order quantity. The second appears at 16:00, which is because the order quantity decreased. The mileage utilization rate of taxis is higher at 9:00 to 15:00 in the day and 18:00 to 02:00 the next day, reaching more than 60%. The three-day mileage utilization curve trend is generally similar, showing a W-shaped fluctuation. Different from Fig. 6(f), the trend of the average empty mileage and mileage utilization rate changes in the opposite direction and the overall M-shaped fluctuation appears. Especially in the early hours of the morning, the road traffic condition is very good, the traffic demand is concentrated in a few night market areas, the taxi driver will spend more time waiting or looking for passengers, the distance which looking for passengers of all vehicles will increase.

From Fig. 6(c) and Fig. 6(e), it can be seen that the time-varying trend of the average hourly return and the average passenger load time is similar. From 18:00 to 20:00, the average hourly incomes on holidays, working days, and non-working days are less than 70 CNY, and the average passenger mileage is less than 6.2km. This is because the time period is night peak hours causing much traffic congestion. Both the

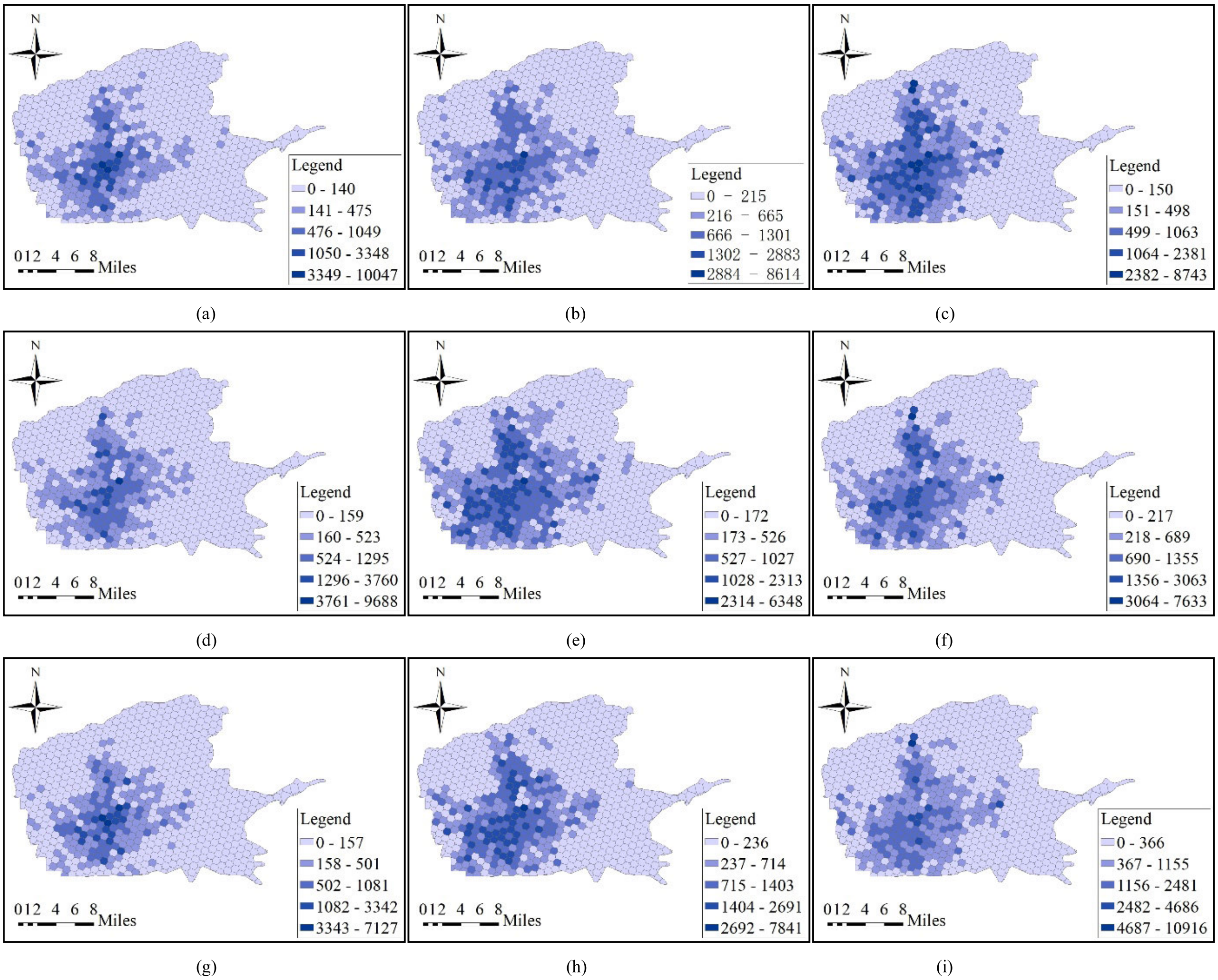


FIGURE 8. The spatial distribution of order-income regions at different time (CNY): (a), (b), (c) represent at 5 o'clock, 8 o'clock, 18 o'clock on October 5 respectively. (d), (e), (f) represent at 5 o'clock, 8 o'clock, 18 o'clock on October 9 respectively. (g), (h), (i) represent at 5 o'clock, 8 o'clock, 18 o'clock on October 13 respectively.

number of passengers served and the number of miles completed in time are reduced [17]. The taxi driver is basically in the passenger operation state, and the road traffic condition is the main factor affecting the income. The longer the stay in the congestion area, the lower the income. However, the average hourly incomes and the average passenger mileage of holidays are slightly higher than workdays and non-working days. This is because during holidays, the demand for private transportation in the city is reduced, and the traffic conditions are better, and serious traffic congestion will not occur [25].

From Fig. 6 (d), it can be seen that during 08: 00 $\sim$ 09: 00 on weekdays, the average passenger time is much longer than that on holidays and non-working days. This is due to the obvious morning peaks on working days. There is much more traffic congestion, and the passenger-carrying time increases. The average passenger time on holidays is significantly lower than that of working days and non-working days, which is also greatly related to traffic conditions and people's travel habits.

B. SPATIAL ANALYSIS

As Tobler's First Law says: Everything is related to other things, but things that are close are more closely related. Scientific selection of taxi passenger locations can reduce idle time and distance, and increase hourly income. Space analysis can reduce the operator's operating expenses and increase profitability [50]. From the analysis of section IV.A, we found that 5:00 and 18:00 in these days are the two trough moments of the day. Then at 8:00 on these days, the order quantity increased rapidly. Therefore, this paper further summarizes and analyze the distribution of high-income orders and high-income regions in these three moments. The spatial distribution of high-income orders is different from the spatial distribution of high-income regions. This paper chooses the orders that the label is equal to 4 from Fig. 2 to analyze the distribution of high-income orders. By obtaining the total income of all orders in each grid within the period, the grid with higher total income is regarded as a high-income region. The purpose is to get high-income regions in space.

TABLE 5. Original statistics for taxi operational efficiency indicators calculation.

Date	Index	Mean	Std.Dev	Min	Max	Median
5 October,2019	N	14273.87	4490.15	6213	22305	14141.50
	MUR(%)	61.36	7.34	45.47	71.19	62.62
	AHI(CNY)	77.98	9.92	66.95	97.83	73.12
	APT(m)	14.22	1.09	12.59	16.28	14.17
	APM(mile)	6.25	0.34	5.90	7.22	6.17
	AEM(mile)	4.15	1.65	2.39	8.30	3.61
9 October,2019	N	15137.50	5050.47	6338	25071	15773
	MUR(%)	63.04	8.43	40.86	70.54	67.50
	AHI(CNY)	76.39	11.84	61.96	99.69	70.50
	APT(m)	14.85	1,82	12.20	18.58	14.79
	APM(mile)	6.38	0.34	5.93	7.36	6.25
	AEM(mile)	4.01	1.97	2.61	9.92	3.08
13 October,2019	N	15503.29	4388.14	5899	21405	16974.50
	MUR(%)	63.24	8.57	45.33	75.53	64.62
	AHI(CNY)	78.94	15.19	60.32	105.71	71.66
	APT(m)	14.52	2.12	11.71	19.20	14.27
	APM(mile)	6.39	0.38	6.02	7.29	6.23
	AEM(mile)	3.99	1.86	1.98	8.61	3.42

1) ANALYSIS OF SPATIAL DISTRIBUTION

The spatial distribution of high-income orders can improve the possibility of some low-income drivers accepting high-income orders and provide a reference for them. This paper is based on the high-income orders divided in section III.C.1), and then divided into 818 research units based on the latitude and longitude coordinates of the high-income orders, and divided the 818 research units into 5 levels based on the natural breakpoint method, as is shown in Fig. 7. It can be seen that in the 818 research units, the number of high-income orders can be up to 173 orders. This paper considers the regions of the 4th and 5th levels as the regional hot-spot of high-income orders. The statistical results are shown in Table 6. Table 6 counts the numbers of the high-income orders hot-spot regions located at the 4th and 5th levels so that it can be seen more clearly which are the high-income orders hot-spots.

This paper divides the total income of all orders in the study area based on the Natural Breakpoint method. As shown in Fig. 8, it describes the distribution of regions with high and low order income. In this paper, the order-income regions are divided into 5 levels, and the color of the grid unit is used to indicate the level of income of the grid. The darker the color, the higher the income. This paper defines the darkest levels as high-income regions. The statistical results are shown in Table 7. Table 7 counts the numbers of the high-income regions and the ID of high-income regions at different periods, so that we can get the high-income regions at different periods.

It can be seen from Table 6, there are 10, 28, 43, 13, 35, 44, 12, 64, and 71 high-income orders hot-spot units at 5:00, 8:00, and 18:00 on October 5, October 9, October13. From Table 7, there are 29, 66, 74, 30, 90, 74, 29, 107 and 114 high-income orders hot-spot units at 5:00, 8:00, and 18:00 on October 5, October 9, October 13. At different periods, the number and location of high-income orders hotspots and high-income regions are different. Therefore, experienced drivers can always know where to go at different periods to improve their AHI. Table 6 and Table 7 show the ID of high-income orders hot-spots and high-income regions, which can help those less experienced drivers to better grasp the travel needs of various units in the city at different periods and improve their average hourly incomes (AHI).

People, engaged in different socio-economic activities on holidays, workdays and weekends, have different travel purposes and traffic demands, which makes the distribution of high-income hot-spots and high-income regions on holidays, workdays and weekends significantly different, as shown in Fig. 7 and Fig. 8. At different periods, the high-income orders hot-spots and the high-income regions have no absolute relationship with the hot-spots in the previous study [27], [47], [51]. Besides, in regions with inconvenient transportation,

TABLE 6. The ID of high-income orders hot-spots.

Date	Time	ID	Count
Oct 5,2019	05:00-05:59	176, 237, 303, 347, 396, 461, 466, 498, 621, 750	10
	08:00-05:59	45, 107, 127, 138, 160, 176, 184, 192, 237, 255, 291, 310, 318, 326, 357, 372, 396, 419, 438, 454, 461, 471, 498, 519, 535, 537, 698, 767	28
	18:00-18:59	6, 45, 95, 97, 118, 123, 124, 127, 138, 176, 184, 192, 200, 213, 237, 241, 255, 287, 303, 310, 372, 376, 379, 396, 419, 422, 434, 438, 454, 461, 466, 498, 503, 504, 519, 522, 604, 621, 626, 746, 747, 750, 802	43
Oct 9,2019	05:00-05:59	6, 107, 176, 303, 310, 326, 396, 419, 438, 461, 466, 498, 784	13
	08:00-05:59	8, 32, 107, 134, 138, 144, 176, 184, 192, 200, 255, 266, 310, 318, 319, 326, 331, 357, 369, 372, 396, 438, 461, 466, 498, 531, 533, 535, 543, 669, 698, 706, 726, 750, 788	35
	18:00-18:59	6, 8, 32, 45, 54, 89, 107, 118, 138, 139, 176, 182, 184, 192, 200, 213, 233, 2357, 241, 247, 266, 291, 310, 326, 331, 369, 372, 376, 396, 422, 438, 454, 459, 461, 466, 498, 503, 504, 519, 522, 531, 604, 698, 706	44
Oct 13,2019	05:00-05:59	77, 107, 124, 176, 303, 326, 372, 396, 461, 466, 498, 531	12
	08:00-05:59	6, 8, 32, 45, 54, 84, 107, 138, 144, 160, 176, 184, 192, 200, 237, 241, 242, 247, 255, 266, 268, 291, 303, 310, 318, 319, 326, 331, 343, 347, 357, 369, 372, 382, 396, 422, 438, 443, 450, 454, 461, 466, 483, 491, 492, 498, 503, 519, 531, 533, 535, 536, 537, 595, 621, 626, 653, 698, 737, 746, 747, 750, 767, 788	64
	18:00-18:59	6, 8, 45, 54, 84, 95, 97, 105, 107, 118, 123, 127, 134, 138, 139, 144, 160, 176, 182, 184, 192, 200, 213, 237, 241, 247, 255, 266, 268, 287, 291, 303, 315, 318, 319, 326, 331, 335, 369, 372, 379, 396, 410, 419, 422, 434, 438, 454, 461, 466, 471, 498, 503, 504, 519, 520, 522, 531, 535, 562, 604, 621, 626, 698, 706, 708, 724, 746, 765, 767, 802	71

people are more likely to choose taxis to travel, so this region is likely to be a high-income orders hot-spot. In regions with less sparse traffic and better traffic conditions, these regions are likely to become high-income regions such as located near Zhangjiabao Street in the Weiyang District. This paper selects several main hot-spots as the main analysis objects of this paper. For example, Xi'an Railway Station(ID = 466), Bell Tower(ID = 396), Xiaozhai(ID = 718), Xi'an North Station(ID = 310), Stadium Subway Station(ID = 335), Chengnan Passenger Terminal(ID = 522), Andingmen(ID = 535), City Library Subway Station(ID = 6).

In the period 5: 00 $\sim$ 6: 00, from Fig. 7 (a) (d) (g) and Fig. 8 (a) (d) (g), there are few high-income orders hot-spots, which are scattered. In addition to format hot-spots, Xi'an Railway Station on holiday, working day, and non-working day is also a high-income orders hot-spot and a high-income region. This is because, at this time, many trains from other places to Xi'an have arrived at the station. Earlier at this period, the bus was operated at 5:30. For places where long-distance buses cannot reach, many people choose to travel by taxi, which forms an obvious high-income region. The Bell Tower is only a high-income orders hot-spot instead of a high-income region, indicating that there are many high-income orders near the Bell Tower. However, the probability of receiving high-income orders is only 23%, 19%, 24% in these three days, which is there are more low-income orders. Stadium Subway Station, Xi'an North Station, and City Library Subway Station are high-income regions. Before 5:30 in the morning, the subway and bus lines at Xi'an North Station had not yet started operation, so more people choose to travel by taxi.

In the period 8: 00 $\sim$ 9: 00, from Fig. 7 (b) (e) (h) and Fig. 8 (b) (e) (h), the high-income orders hotspot and high-income region increased significantly during this period. This is because this period is the morning peak of urban residents' travel. Most people travel at this time for commuting to work, so for places where the bus system is not perfect. For example, in the 205 communities, as high-end residential communities, residents rely more on taxis. As the number of taxi orders increases, the number of high-income orders will increase accordingly. During National Day, Xi'an Railway Station is not a high-income orders hot spot, nor is it a high-income region.. In other words, the passengers of Xi'an Railway Station at this time are mostly for travel, and most will choose to travel by bus or subway. The frequency of taxis is reduced, making it difficult to form a high-income region. The Bell Tower, Xi'an North Railway Station, Andingmen at this time

TABLE 7. The ID of high-income regions.

Date	Time	ID	Count
Oct 5,2019	05:00-05:59	6,97,107,176,200,237,255,303,310,326,335,347,357,372,396,419,422,438,454,461,471,498,531,621,706,724,746,750,784	29
	08:00-05:59	6, 8, 45, 84, 95, 107, 123, 127, 138, 144, 160, 176, 182, 184, 192, 200, 213, 234, 237, 241, 242, 247, 255, 266, 267, 268, 287, 291, 303, 310, 318, 319, 326, 335, 343, 347, 357, 372, 396, 419, 422, 438, 443, 454, 461, 471, 474, 498, 504, 519, 531, 533, 535, 537, 545, 553, 621, 626, 691, 698, 706, 718, 724, 747, 767, 784	66
	18:00-18:59	6, 8, 45, 76, 84, 95, 97, 105, 107, 118, 124, 127, 138, 139, 144, 160, 176, 182, 184, 192, 200, 213, 237, 241, 247, 255, 257, 266, 268, 287, 303, 318, 319, 326, 331, 343, 347, 357, 369, 372, 376, 379, 396, 419, 422, 434, 438, 443, 454, 461, 466, 471, 474, 485, 498, 503, 504, 519, 522, 531, 535, 553, 604, 621, 626, 698, 706, 724, 746, 747, 750, 767, 784, 802	74
Oct 9,2019	05:00-05:59	6, 95, 107, 127, 138, 176, 255, 303, 310, 326, 335, 357, 372, 396, 419, 422, 438, 443, 454, 461, 471, 498, 504, 535, 621, 706, 718, 750, 767, 784	30
	08:00-05:59	6, 8, 32, 45, 54, 84, 85, 97, 107, 118, 123, 134, 138, 144, 160, 162, 176, 182, 184, 192, 200, 202, 213, 234, 237, 241, 242, 247, 255, 261, 266, 268, 287, 291, 303, 310, 318, 319, 326, 331, 335, 343, 347, 357, 365, 369, 372, 396, 419, 422, 434, 438, 442, 443, 450, 454, 461, 466, 471, 474, 492, 498, 519, 531, 533, 535, 536, 537, 543, 562, 567, 604, 621, 626, 665, 669, 691, 698, 706, 724, 726, 728, 744, 746, 747, 750, 765, 767, 788, 813	90
	18:00-18:59	6, 8, 32, 45, 54, 84, 85, 89, 97, 107, 118, 138, 139, 160, 176, 182, 184, 192, 200, 213, 233, 234, 237, 241, 247, 254, 255, 266, 268, 291, 303, 318, 326, 331, 335, 343, 357, 369, 372, 376, 379, 396, 419, 422, 434, 438, 442, 443, 454, 459, 461, 466, 474, 498, 503, 504, 519, 522, 531, 533, 535, 537, 621, 626, 665, 669, 691, 698, 706, 724, 746, 750, 767, 802	74
Oct 13,2019	05:00-05:59	6, 77, 107, 124, 138, 176, 200, 237, 255, 303, 318, 326, 335, 372, 396, 419, 422, 438, 443, 454, 461, 471, 498, 531, 535, 746, 750, 767, 784	29
	08:00-05:59	6, 8, 32, 45, 54, 76, 84, 85, 95, 97, 105, 107, 118, 123, 138, 144, 160, 176, 182, 184, 192, 200, 208, 213, 234, 237, 241, 242, 247, 254, 255, 261, 266, 267, 268, 291, 303, 310, 318, 319, 320, 326, 331, 335, 343, 347, 357, 365, 369, 372, 376, 379, 382, 396, 419, 422, 434, 438, 442, 443, 450, 454, 461, 471, 474, 483, 489, 491, 492, 498, 519, 520, 531, 533, 535, 536, 537, 545, 553, 560, 567, 595, 604, 621, 626, 632, 641, 653, 665, 669, 691, 698, 706, 708, 724, 726, 728, 737, 746, 747, 750, 767, 784, 788, 800, 813, 817	107
	18:00-18:59	6, 8, 32, 45, 54, 76, 84, 85, 89, 95, 97, 105, 107, 118, 123, 124, 127, 134, 138, 139, 144, 160, 162, 176, 182, 184, 192, 200, 208, 213, 233, 234, 237, 241, 242, 247, 254, 255, 261, 266, 267, 268, 287, 291, 303, 315, 318, 319, 320, 326, 331, 335, 343, 347, 357, 365, 369, 372, 376, 379, 396, 410, 419, 422, 434, 438, 443, 450, 454, 461, 466, 471, 474, 491, 498, 503, 504, 519, 520, 522, 531, 535, 536, 537, 543, 545, 553, 562, 573, 595, 604, 621, 626, 659, 669, 671, 691, 698, 706, 708, 724, 726, 746, 747, 750, 765, 767, 784, 788, 798, 800, 802, 813, 817	114

are not only high-income orders hot-spots but also high-income regions. The residential area is near Andingmen, people have a high frequency of travel for commuting. Stadium Subway Station and Chengnan Passenger Terminal are high-income areas, but they are not hot areas for high-income orders. The City Library Subway Station is not a high-income orders hot-spot on working days. This is related to the type of land it uses. The nearby land is used for transportation hubs and commercial services.

In the period 18: 00 $\sim$ 19: 00, from Fig. 7 (c) (f) (i) and Fig. 8 (c) (f) (i), compared with the time period 8: 00-9: 00, the high-income orders hot-spot and high-income region

TABLE 8. Relevant indicators of taxi high-income orders space in different time.

Date	Time	*Moran's I*	Variance	Z-score
Oct 5,2019	05:00-05:59	0.2749	0.0007	10.7109
	08:00-05:59	0.2957	0.0008	10.8221
	18:00-18:59	0.4529	0.0010	14.7181
Oct 8,2019	05:00-05:59	0.1494	0.0005	6.5289
	08:00-05:59	0.4324	0.0009	13.7287
	18:00-18:59	0.4537	0.0009	14.9437
Oct 13,2019	05:00-05:59	0.3225	0.0009	10.9778
	08:00-05:59	0.4303	0.0009	14.0358
	18:00-18:59	0.4160	0.0009	13.5232

in this period have decreased. As a whole, the distribution of high-income hot-spots and high-income regions in this time period on holiday, workday, and non-working day are roughly the same. At this moment, the high-income orders hot-spots mainly include the Bell Tower, Xi'an North Station, and the City Library Subway Station. The high-income regions mainly include Bell Tower, Xi'an North Station, City Library Subway Station, Stadium Subway Station, Chengnan Passenger Station, An Dingmen.

In general, we know that the hot-spots at different times is not a high-income orders hot-spots or a high-income region. Due to the fact that Xiaozhai, as a central business area, is the hottest hot-spots in Xi'an in a day. Similarly, the Xiaozhai commercial district has a well-developed bus and subway system, but the traffic congestion index is high. Therefore, Xiaozhai is not a high-income region in the sense of our research, which is different from previous perception. The high-income orders hot-spot and the high-income region are different at different times, which is the spatially imbalanced supply-demand relationship and the relatively limited ability to regulate and control the existing urban transport system. According to the spatial distribution results, Xi'an high-income regions basically spread along the main road of Xi'an in the shape of wood; the number and distribution of high-income units in the morning and afternoon are quite different.

2) SATIAL CORRELATION ANALYSIS

In order to analyze whether the high-income orders and the total income in 818 units in the research area are spatially correlated, the Moran index and relevant statistics are calculated by using formulas (12), (13), (14) and (15), the results are shown in Table 8 and Table 9. All P-values were 0.0000 and all passed the 1% significance test.

From Table 8, we can get the following conclusion. The high-income orders of global *Moran's I* is positive in different periods, and all *Moran's I* have passed the 1% significance test, which shows that the order income of each unit in the research area shows a positive correlation in space.

TABLE 9. Relevant indicators of taxi income space in different time.

Date	Time	*Moran's I*	Variance	Z-score
Oct 5,2019	05:00-05:59	0.2564	0.0007	10.0501
	08:00-05:59	0.4818	0.0009	15.7472
	18:00-18:59	0.3647	0.0008	13.3277
Oct 8,2019	05:00-05:59	0.1587	0.0006	6.4799
	08:00-05:59	0.3814	0.0010	12.3183
	18:00-18:59	0.3939	0.0008	14.4202
Oct 13,2019	05:00-05:59	0.3549	0.0009	11.7933
	08:00-05:59	0.3849	0.0010	12.5143
	18:00-18:59	0.3452	0.0007	12.8723

It can be seen that the high-income orders of each unit in the study area is affected by the spatial correlation factors, showing a trend of agglomeration in the spatial distribution. The calculation results show that the spatial correlation between the units is significant. The minimum *Moran's I* at 5 o'clock on October 9 is 0.1587, which may be due to the less order quantity at that time, and there will not be more obvious aggregations near major passenger stations like holiday and working day. From Table 9, on the one hand, we can clearly see that as with the spatial correlation of high-income orders, the *Moran's I* of high-income regions at different times in the study area are all positive numbers, and all have passed the 1% significance test. This shows that the distribution of high-income regions also shows a tendency of agglomeration spatially. On the other hand, at 18:00, the *Moran's I* on the working day was significantly higher than that on the non-working day. This is because on the non-working day, people traveled more freely, and people often went to more dispersed places such as parks and theaters. This also validates the distribution of the high-income regions in Fig. 7 (f) and Fig. 7 (i). In the nine periods of this study, at 8 o'clock on Oct 5, the *Moran's I* which was largest is 0.4818, and the degree of aggregation was the highest at this time. Combining with the previous analysis, there were more people visiting Xi'an during the holidays, mainly concentrated in major passenger stations nearby. In summary, both the distribution of high-income orders and the distribution of high-income regions show a certain tendency of aggregation in space. Therefore, we can consider introducing this result into the online dispatch system, which provides a reference for the taxi management department in balancing the income of taxi drivers.

V. CONCLUSION

The unbalanced distribution of taxi passengers in space and time affects the taxi drivers' route choice. Taking Xi'an City, Shaanxi Province as an example, this paper applies the taxi GPS orders data to measure and analyze the taxi operation efficiency in the time dimension and high orders in space dimension in holidays, working days and non-working days. The conclusions are as follows:

First of all, from the time dimension, (1) The average daily income of taxi drivers on holidays, working days and non-working days in Xi'an are 688.26 CNY, 724.18 CNY, 732.35 CNY respectively, which shows that the average daily income of taxis drivers on holidays is significantly lower than that in working days and non-working days. (2) Taxi operation efficiency indicators vary throughout the day on Oct 5, Oct 9, and Oct 13. The order quantity, reached two low values from 5:00 to 6:00 and 18:00 to 19:00, growing fastest from 8:00 to 9:00 in these three days. (3) The average hourly incomes of drivers reach the lowest value of the day between 18:00 to 19:00, which has a great relationship with the traffic conditions.

Secondly, from the space dimension, the distribution of high-income orders and high-income regions are imbalanced in administrative and functional zones. (1) The number and the spatial distribution of high-income orders hot-spots and high-income regions vary at different periods. (2) Some places are hot-spots, but neither high-income orders hot-spots nor high-income regions, such as Xiaozhai, which is the most prosperous commercial land in Xi'an. (3) The high-income orders regions and high-income regions have a strong correlation in spatial distribution, which shows a tendency of agglomeration.

These findings provide comprehensive and quantitative insight to improve the average hourly incomes of taxi drives, as well as challenges some preconceived ideas of how to earn high incomes. This paper also provides the following two suggestions for the taxi industry management department. (1) Taxi industry managers can take traffic conditions into consideration when making taxi charging policies to avoid some drivers refusing to pick up passengers in certain places due to traffic congestion. (2) In addition, taxi industry managers can take high-income orders hotspots and high-income regions into account in the dispatch system, which provides drivers with reference suggestions on when to dispatch the orders. The two above suggestions can balance the income of taxi drivers, and improve the overall operational efficiency of taxi drivers.

This paper still has some limitations when digging into the high-income orders hot-spots and high-income regions. (1) This paper does not calculate the probability that a taxi receives a high-income orders in a unit at a certain moment. (2) The taxi operation efficiency indicators of each unit should be calculated in further research, and more objective and accurate research will be carried out by taking multiple types of influencing factors into account based on the estimation of operation efficiency.

REFERENCES

[1] X. Yu, S. Gao, X. Hu, and H. Park, "A Markov decision process approach to vacant taxi routing with e-hailing," *Transp. Res. B, Methodol.*, vol. 121, pp. 114–134, Mar. 2019.

[2] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: A recommender system for finding passengers and vacant taxis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.

[3] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger?" in *Proc. ACM Int. Conf. Ubiquitous Comput. (Ubicomp)*, 2011, pp. 109–118.

[4] H. Yang and T. Yang, "Equilibrium properties of taxi markets with search frictions," *Transp. Res. B, Methodol.*, vol. 45, no. 4, pp. 696–713, May 2011.

[5] L. Sun, D. Zhang, C. Chen, P. S. Castro, S. Li, and Z. Wang, "Real time anomalous trajectory detection and analysis," *Mobile Netw. Appl.*, vol. 18, no. 3, pp. 341–356, Jun. 2013.

[6] H. Rong, Z. Wang, H. Zheng, C. Hu, L. Peng, Z. Ai, and A. K. Sangaiah, "Mining efficient taxi operation strategies from large scale geo-location data," *IEEE Access*, vol. 5, pp. 25623–25634, 2017.

[7] L. Tang, F. Sun, Z. Kan, C. Ren, and L. Cheng, "Uncovering distribution patterns of high performance taxis from big trace data," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 5, p. 134, Apr. 2017.

[8] R.-H. Hwang, Y.-L. Hsueh, and Y.-T. Chen, "An effective taxi recommender system based on a spatio-temporal factor analysis model," *Inf. Sci.*, vol. 314, pp. 28–40, Sep. 2015.

[9] S. Zhang and Z. Wang, "Inferring passenger denial behavior of taxi drivers from large-scale taxi traces," *PLoS ONE*, vol. 11, no. 11, Nov. 2016, Art. no. e0165597.

[10] C. W. Yuan, X. Y. Mi, Q. Q. Wu, and D. L. Wei, "Optimal model of taxi waiting time fee under traffic congestion condition," *J. Traffic Transp. Eng.*, vol. 14, no. 2, pp. 75–81, 2014.

[11] C. Yuan, D. Wu, D. Wei, and H. Liu, "Modeling and analyzing taxi congestion premium in congested cities," *J. Adv. Transp.*, vol. 2017, Mar. 2017, Art. no. 2619810.

[12] C. Camerer, L. Babcock, G. Loewenstein, and R. Thaler, "Labor supply of New York City cabdrivers: One day at a time," *Quart. J. Econ.*, vol. 112, no. 2, pp. 407–441, May 1997.

[13] G. Qin, T. Li, B. Yu, Y. Wang, Z. Huang, and J. Sun, "Mining factors affecting taxi drivers' incomes using GPS trajectories," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 103–118, Jun. 2017.

[14] D. Card and G. B. Dahl, "Family violence and football: The effect of unexpected emotional cues on violent behavior," *Quart. J. Econ.*, vol. 126, no. 1, pp. 103–143, Feb. 2011.

[15] D. Wei, C. Yuan, H. Liu, D. Wu, and W. Kumfer, "The impact of service refusal to the supply–demand equilibrium in the taxicab market," *Netw. Spatial Econ.*, vol. 17, no. 1, pp. 225–253, Mar. 2017.

[16] H. A. H. Naji, C. Wu, H. Zhang, and L. Li, "Towards understanding the impact of human mobility patterns on taxi drivers' income based on GPS data: A case study in Wuhan-China," in *Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Aug. 2017, pp. 1152–1160.

[17] G. Ou, Y. Wu, G. Wang, and Z. Guo, "Big-data-based analysis on the relationship between taxi travelling patterns and taxi drivers' incomes," in *Proc. 16th Int. Conf. Service Syst. Service Manage. (ICSSSM)*, Jul. 2019, pp. 1–6.

[18] Y. Dong, Z. Zhang, R. Fu, and N. Xie, "Revealing New York taxi drivers' operation patterns focusing on the revenue aspect," in *Proc. 12th World Congr. Intell. Control Automat.*, 2016, pp. 1052–1057.

[19] L. Liang, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," *Comput., Environ. Urban Syst.*, vol. 34, no. 6, pp. 541–548, Nov. 2010.

[20] Y. Gao, P. Xu, L. Lu, H. Liu, S. Liu, and H. Qu, "Visualization of taxi drivers' income and mobility intelligence," in *Proc. Int. Symp. Vis. Comput.* Berlin, Germany: Springer, 2012, pp. 275–284.

[21] D. Zhang, L. Sun, B. Li, C. Chen, G. Pan, S. Li, and Z. Wu, "Understanding taxi service strategies from taxi GPS traces," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 123–135, Feb. 2015.

[22] F. Sun, X. Zhang, L. Tang, Z. Liu, X. Yang, and K. Dong, "Temporal and spatial distribution of high efficiency passengers based on GPS trajectory big data," *J. Geo-Inf. Sci.*, vol. 17, no. 3, pp. 329–335, 2015.

[23] H. Rong, X. Zhou, C. Yang, Z. Shafiq, and A. Liu, "The rich and the poor: A Markov decision process approach to optimizing taxi driver revenue efficiency," presented at the 25th ACM Int. Conf. Inf. Knowl. Manage., Indianapolis, IN, USA, 2016, doi: 10.1145/2983323.2983689.

[24] L. Tang, W. Zheng, Z. Wang, H. Xu, J. Hong, and K. Dong, "Space time analysis on the pick-up and drop-off of taxi passengers based on GPS big data," *J. Geo-Inf. Sci.*, vol. 17, no. 10, pp. 1179–1186, 2015.

[25] S. Y. Oleyaei-Motlagh and A. E. Vela, "Inferring demand from partially observed data to address the mismatch between demand and supply of taxis in the presence of rain," Mar. 2019, pp. 7–15, *arXiv:1903.06619*. [Online]. Available: https://arxiv.org/abs/1903.06619

[26] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers Comput. Sci.*, vol. 6, no. 1, pp. 111–121, Feb. 2012.

[27] B. Zhou, L. Ma, J. Hu, S. Wu, and G. He, "Extraction of urban hotspots and analysis of spatial interaction based on trajectory data field: A case study of Shenzhen City," *Tropical Geogr.*, vol. 39, no. 1, pp. 117–124, 2019.

[28] K. Qin, Q. Zhou, Y. Xu, W. Xu, and P. Luo, "Spatial interaction network analysis of urban traffic hotspots," *Prog. Geogr.*, vol. 36, no. 9, pp. 1149–1157, 2017.

[29] P. Zhao, K. Qin, X. Ye, Y. Wang, and Y. Chen, "A trajectory clustering approach based on decision graph and data field for detecting hotspots," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 6, pp. 1101–1127, 2017.

[30] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in *Proc. 18th Int. Conf. World Wide Webs (WWW)*, 2009, pp. 791–800.

[31] J. Cramer and A. B. Krueger, "Disruptive change in the taxi business: The case of Uber," *Amer. Econ. Rev.*, vol. 106, no. 5, pp. 177–182, May 2016.

[32] R. C. P. Wong, W. Y. Szeto, S. C. Wong, and H. Yang, "Modelling multi-period customer-searching behaviour of taxi drivers," *Transportmetrica B, Transp. Dyn.*, vol. 2, no. 1, pp. 40–59, Jan. 2014.

[33] A. Duran and M. Earleywine, "GPS data filtration method for drive cycle analysis applications," in *Proc. SAE World Congr. Exhib.*, 2012, pp. 1–9.

[34] Y. Tang, C. Jiang, B. Zheng, and Q. Li, "Taxi on service trip characteristics based on multi-source data fusion: A case of Yueyang," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 18, no. 2, pp. 45–51, 2018.

[35] D. Flores-Guri, "An economic analysis of regulated taxicab markets," *Rev. Ind. Org.*, vol. 23, nos. 3–4, pp. 255–266, Dec. 2003.

[36] C. M. Gold, J. Nantel, and W. Yang, "Outside-in: An alternative approach to forest map digitizing," *Int. J. Geograph. Inf. Syst.*, vol. 10, no. 3, pp. 291–310, Apr. 1996.

[37] Z. Zhou, S. Zhang, K. Xiong, B. Li, Z. Tian, Q. Chen, L. Yan, and S. Xiao, "The spatial distribution and factors affecting karst cave development in Guizhou province," *J. Geograph. Sci.*, vol. 27, no. 8, pp. 1011–1024, Aug. 2017.

[38] Q. Yu, D. Yue, D. Yang, H. Ma, Q. Zhang, and B. Yin, "Layout optimization of ecological nodes based on BCBS model," *Trans. Chin. Soc. Agricult. Machinery*, vol. 47, no. 12, pp. 329–330, and 336, 2016.

[39] Z. Shuxu, Q. Ruitao, and L. Changrong, "An traffic trajectory data analysis method based on trajectory feature division," *Bull. Surv. Mapping*, vol. 11, p. 73, Nov. 2018.

[40] N. Davis, G. Raina, and K. Jagannathan, "Taxi demand-supply forecasting: Impact of spatial partitioning on the performance of neural networks," Dec. 2018, pp. 7–10, *arXiv:1812.03699*. [Online]. Available: https://arxiv.org/abs/1812.03699

[41] R. Gelda, K. Jagannathan, and G. Raina, "Forecasting supply in Voronoi regions for app-based taxi hailing services," in *Proc. 6th IEEE Int. Conf. Adv. Logistics Transp. (ICALT)*, Jul. 2017, pp. 47–52.

[42] N. Davis, G. Raina, and K. Jagannathan, "Taxi demand forecasting: A HEDGE-based tessellation strategy for improved accuracy," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3686–3697, Nov. 2018.

[43] G. P. Hammond and J. B. Norman, "Decomposition analysis of energy-related carbon emissions from UK manufacturing," *Energy*, vol. 41, no. 1, pp. 220–227, May 2012.

[44] Z. Lyu, J. Wu, S. Yao, and L. Zhu, "FCD-based analysis of taxi operation characteristics: A case of Shanghai," *J. East China Normal Univ. Natural Sci.*, no. 3, pp. 133–144, 2017 2017.

[45] W. Zhai, X. Bai, Z.-R. Peng, and C. Gu, "A bottom-up transportation network efficiency measuring approach: A case study of taxi efficiency in New York City," *J. Transp. Geogr.*, vol. 80, Oct. 2019, Art. no. 102502.

[46] C. Sun, X. Chen, H. M. Zhang, and S. Chen, "Big data assessment on the operating characteristics of on-line taxis and thoughts on relevant policies," *J. Southeast Univ., English Ed.*, vol. 34, no. 3, pp. 394–401, 2018.

[47] Y. Li, H. Chen, X. Sun, T. Luo, and Z. Shi, "An analysis of travel characteristics of urban residents based on hot spot detection model," *J. Transp. Inf. Saf.*, vol. 37, no. 1, pp. 128–136, 2019.

[48] Y. Yang, Z. He, Z. Song, X. Fu, and J. Wang, "Investigation on structural and spatial characteristics of taxi trip trajectory network in Xi'an, China," *Phys. A, Stat. Mech. Appl.*, vol. 506, pp. 755–766, Sep. 2018.

[49] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, "Improved deep hybrid networks for urban traffic flow prediction using trajectory data," *IEEE Access*, vol. 6, pp. 31820–31827, 2018.

[50] J. Azevedo, P. M. d'Ore, and M. Ferreira, "On the mobile intelligence of autonomous vehicles," in *Proc. IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS)*, S. O. Badonnel, M. Ulema, C. Cavdar, L. Z. Granville, and C. R. P. DosSantos, Eds., Apr. 2016, pp. 1169–1174.

[51] D. Liu, S.-F. Cheng, and Y. Yang, "Density peaks clustering approach for discovering demand hot spots in city-scale taxi fleet dataset," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst., IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2015, pp. 1831–1836.

CHANGWEI YUAN received the B.S. degree in statistics and the M.S. and Ph.D. degrees in transportation planning and management from Chang'an University, in 2003, 2005, and 2007, respectively. His research interests include transportation infrastructure network design, traffic energy, environmental management, and taxi management.

XINRUI GENG received the B.S. degree in transportation engineering from Chang'an University, in 2018, where she is currently pursuing the M.S. degree in transportation engineering. Her research interests include traffic big data, traffic flow modeling, and taxi control.

XINHUA MAO received the B.S., M.S., and Ph.D. degrees in transportation engineering from Chang'an University. He is currently a Lecturer with Chang'an University. His research interests include traffic flow modeling, transportation infrastructure network design, and logistics network optimization.