sustainability

MDPI

Article

Optimization Model of Taxi Fleet Size Based on GPS Tracking Data

Yang Yang [1,2,3,*], Zhenzhou Yuan [1,3], Xin Fu [4,*], Yinhai Wang [2] and Dongye Sun [3]

1 MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China; zzyuan@bjtu.edu.cn
2 Department of Civil and Environmental Engineering, University of Washington, More Hall, University of Washington, Seattle, WA 98195, USA; yinhai@uw.edu
3 School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; 14114218@bjtu.edu.cn
4 School of Economics and Management, Chang'an University, Middle section of South Second Ring Road, Xi'an 710064, China
* Correspondence: 14114222@bjtu.edu.cn (Y.Y.); fuxin@chd.edu.cn (X.F.)

Received: 11 December 2018; Accepted: 28 January 2019; Published: 30 January 2019

check for updates

Abstract: A reasonable taxi fleet size has a significant impact on the satisfaction of urban traffic demand, the alleviation of urban traffic congestion, and the stability of taxi business groups. Most existing studies measure the overall scale by using macro indices, and few studies are from the micro level. To meet the transportation demand for taxis, mitigating the mismatch between taxi supply and demand, this research proposes an urban taxi fleet size calculating model based on GPS tracking data. Firstly, on the basis of road network segmentation, the probability model of a passenger taxi-taking a road section as a unit is built to evaluate the difficulty of taxi-taking on a road section. Furthermore, a user queuing model is built for the "difficult to take a taxi" road section in the peak period, and the service mileage required by potential taxi users is calculated. After that, a transportation capacity measurement model is built to estimate the number of taxis required in different time periods, Finally, the income constraint model is used to explain the impact of different vehicle fleet sizes on the income of taxi groups, so as to provide a reference for the determination of the final fleet size. The model is applied to data from Xi'an. The calculation results are based on data from May 2014, and show that the scale of taxi demand is about 654–2237, and after considering the impact of different fleet size increases on income, when the income variation index is limited to 0.10, i.e., the decrease of drivers' income will not exceed 10%, an increase of 1286 taxis will be able to meet 66% of the unmet demand in the peak period. The conclusion indicates that the model can effectively calculate the required fleet size and formulate the constraint solutions. This method provided can be considered as a support for formulating the regulation strategy of an urban taxi fleet size.

Keywords: taxi; GPS tracking data; optimization fleet size model

1. Introduction

Taxis are one of the main modes of travel in cities. Although there are great differences between taxis and public transportation in terms of transport capacity, its flexibility and personalized service are important characteristics to meet the needs of urban transportation [1]. In the process of taxi operation, how to regulate the reasonable fleet size of taxis to meet residents' travel needs without causing too much traffic congestion and affecting the operating income of traditional taxi operators is an important premise for strategy formulation. On the one hand, if the amount of fleet investment is too small, where supply is less than demand, it will inevitably cause travel inconvenience and difficulties in taking a taxi.

When the amount of fleet investment is too much, the income of existing groups will decline, and lead to the aggravation of urban traffic congestion, the increase of citizens' travel costs, and even negative social impact [2]. Therefore, a proper taxi fleet size management strategy that balances passengers' benefits (measured by taxi fare and average waiting time) and operation efficiency (measured by taxi operation cost and drivers' income) is essential to mitigate the supply-demand mismatch, improve the transportation environment, and increase the transportation satisfaction of residents.

Many studies have focused on the relationship between the reasonable operation fleet scale of the taxi and its travel rules: Douglass C. North [3] introduced an aggregation model that treated unit time taxi operation cost as a constant. The model revealed that the passengers' demand for a taxi is a decreasing function of average taxi fare and expected wait time, which is in reverse correlation to the deadhead time. Although this model did not take the spatial structure of a taxi market into account, it was adopted by several economists since it succinctly summarized the characteristics of this market, leading to continuous improvement of research in this field. The research on taxi fleet size allocation and demand forecasting has been a hot spot in this field: Khaled Abbas, based on summarizing the existing studies on taxi fleet size allocation, takes a hypothetical city as an example. Three models, "future taxi fleet size based on taxi availability index", "taxi fleet requirements based on hara taxi demand model", and "generic algorithm for estimation of taxi fleet size", are applied to calculate the fleet size of future taxi allocation and demand prediction [4]. Specifically, for the optimization of the relationship between taxi fleet size and revenue, Baozhen Yao et al. [5] established a bilevel programming model to solve this problem. The lower model is the demand function model, which calculates the fleet size demand according to the given capacity configuration and ticket price. The results show that the increase of vehicle configuration size can attract potential demand, but the size of attraction mainly depends on the waiting time of passengers and taxi fares, indicating that the three are interrelated and have an impact, which should be taken into account in the study. While traditional taxis are impacted by emerging rental services like Uber and Lyft, Satish v. Ukkusuri and Zhang [6] developed a model framework to study market research based on decentralized equilibrium in order to obtain an optimal pricing and scale of taxi allocation in a given urban area. Its case study, based on data from New York City, shows that the taxi market at the time may have been oversupplied and underpriced, echoing other match research and the fact that New York taxi prices rose in 2012. Based on two Stackelberg games, two different development strategies are proposed to cope with the anticipated changes in a taxi system, such as the price and quantity elasticity of taxi demand, demand difference level, average taxi speed, passenger waiting time value and taxi service coverage. When the problem focuses on solving urban congestion, complying with the policy of energy saving and emission reduction, and considering the taxi pooling efficiency of passengers and taxi driver income, taxi pooling, and ride share are important measures to solve the problem. Some scholars have also conducted in-depth studies on taxi pooling and ride share: Wang, etc. [7] built an analytical model for taxi volume quantification based on urban taxi demand-supply balance theory and utility theory. All parameters, such as taxi effective mileage, deadhead rate, average daily operation time and average operation speed, was weighted in calculation and modeling. Yang and Sun [8,9] used entropy weight and the weight optimization method in their traffic transportation models respectively.

These studies rely on the statistics of distance and time gathered from residents' travel surveys, which oftentimes provides data that is less accurate and results in a model that is inapplicable in practical applications in large cities. In addition, there is no complete travel data for small and medium-sized cities to calculate the total scale of travel. When calculating the total demand of taxi in the research method, the geographic information attributes of taxi operation trajectory and urban residents' travel demand are not considered. As for the use of theories and methods, as well as the processing and mastering of data, there is no application of a GPS starting and ending point, trajectory and other information. Additionally, the demand and total fleet size of taxis, as well as their influencing factors, have certain differences in different times and regions. In the process of calculation and modeling, the GPS attributes of input data and geographical information factors in the results should be fully considered, and the randomness and difference of the distribution of taxi transportation demand are not considered. In the calculation process, the uniform distribution of transport demand in the road network is usually taken as the starting point, which is prone to errors. Additionally, those methods could not be extrapolated to larger areas, let alone nationwide application. The complexity of the taxi industry and original data add more challenges to this problem. In addition, these methods use historical taxi volume as input data, assuming they represent proper traffic capacity when the legitimacy of this assumption is questioned.

Due to the progress of mobile Internet, Internet, and other technologies, the vehicle-mounted activity data based on individual vehicles (mainly represented by GPS data and vehicle-mounted terminal operation data) has been widely developed and applied in the field of urban transportation and taxi research in recent years. Ghahramani and Zhou et al. [10] proposed an exploratory spatial data analysis algorithm based on cloud processing data analysis by cooperating with a telecom company. The spatial analysis method was used to detect the spatial distribution of mobile phones, and the kernel density method was used to determine these distributions. This analysis can help organizations better implement monitoring and evaluation plans at all levels and make the necessary infrastructure improvements to meet the needs of users. R. Tachet [11] used taxi travel data from cities such as New York, San Francisco, Singapore, and Vienna, calculated the shareability curve of each city in the data rather than using only some of the basic city under the premise of adjustment parameters, predicted the carpool potential of any city under the background of the rapid development in the present city, provided a planning engineer for the transportation company, and the society a sustainable path. In the research of a dynamic vehicle allocation model based on dynamic data, Javier Alonso-Mora et al. [12] studied ride-sharing services and proposed a more general real-time and large-capacity mathematical model for ride-sharing, which could be extended to a large number of passengers and trips, and dynamically generated optimal routes according to online requirements and vehicle locations. The algorithm can also be used to calculate the fleet size of self-driving cars and redirect idle vehicles to areas with high demand. Vazifeh [13] provided a web-based solution for determining the size of a vehicle for individual travel needs. Its target is the "small vehicle fleet size operation problem", which is how to determine the minimum number of vehicles required for all journeys without causing any delays to passengers. By introducing the concept of a "vehicle sharing network", an optimal solution of computing efficiency was proposed. This also means that in the process of reducing urban congestion and improving the utilization rate of rental vehicles, shared travel will be an effective measure. Paolo Santi et al. introduced the concept of a shareable network for ride-sharing taxis, modeled the shared collective interest as a function inconvenient for passengers, and efficiently calculated the optimal sharing strategy on a large number of data sets. Based on the data set of tens of thousands of taxi trips in New York City, the results showed that the cumulative total travel length can be significantly reduced under the premise of an inappropriately small increase in passengers. The results of Paolo Santi et al. [14] showed the potential and contribution of shared taxis. Further, Wenchao Xu and Haibo Zhou [15] studied the application of vehicle network big data in autonomous driving technology on the basis of analyzing and processing the relationship between a network of vehicles and big data, including vehicle geographic information. Pablo Samuel Castro et al. [16] proposed a traffic density model building method based on the GPS data of a taxi, which can be used to predict future traffic conditions and estimate the impact of emissions on urban air quality. At the same time, a new method to automatically determine the traffic capacity of each section was proposed, and its method was verified in the big data of a taxi GPS database. Zhan and Ukkusuri [17] established a probabilistic hybrid model of urban link travel time based on the GPS data of large-scale taxi travel in Manhattan, New York. The model regarded the taxi travel path as a potential path and used polynomial logit distribution to model the taxi driver's path selection behavior. Bonola et al. [18] collected data, including its GPS from small-scale taxis in Rome over six months and made experimental evaluations on delivery performance. The results showed that even with a relatively small number of cars running in parallel in Rome, a very large and irregular city, 80% urban coverage can be achieved in less than 24 h.

The emerging app-based taxi service and electrical taxis raise the turnover of existing vehicles to meet the increasing transportation demand, yet exacerbate traffic congestion. Since the operation data from an app-based taxis are unavailable, its impact on the conventional taxi market is difficult to quantify, decreasing the reliability and accuracy of a traditional traffic capacity calculation model. In previous studies based on vehicles GPS big data, most of the basic data are used to study the evaluation of urban environment, the prediction of travel time, and the service evaluation of the coverage rate of urban areas. At present, no scholars have used the basic data information of taxi GPS to calculate taxi fleet size and the income impact of existing operating groups.

Based on the above research practice and progress, this paper attempts to take the typical vehicle-mounted trajectory GPS and terminal data as the basis, under the premise of segmentation, starting from the effective operation efficiency provided by a taxi, the total demand of residents relying on taxi travel and the sensitivity of taxi driver's income change. A probabilistic model of taxi-taking based on road sections and a measurement model of the demand of unmet people are built, and a measurement model of the taxi required fleet size is formed. Then, the income change analysis on different transport fleet sizes is conducted through the income model. Finally, this paper takes the on-board GPS data of taxi in Xi'an, Shaanxi Province, China in May 2014 as an example to establish the model and verify the method.

2. Data Collection and Pre-Processing

Xi'an is the capital of Shaanxi Province, with a total area of 10,752 square kilometers and a population of 9,056,800. Xi'an is located at the junction of China's land map center and two economic regions in Central and Western China and is one of the largest node cities in the national trunk highway network.

This paper takes the main urban area of Xi'an City (within the ring expressway) as the case study and takes the GPS track record of Xi'an taxis in May 2014 as the basic data for method verification and analysis.

Taxi GPS data uses vehicle license plate and GPS system clock as the main keys to upload information such as the current longitude and latitude coordinates, speed value and speed direction of the taxi as well as whether it is carrying passengers. In May 2014, Xi'an taxi track data management system contains a total of 12,115 taxis due to vehicle maintenance, license plate replacement, suspension and other factors. According to the statistics, the average number of vehicles in operation is 11,440, and the amount of track data records generated is about 30 million/day. Among them, the morning and evening peak periods (referring to relevant studies, the morning peak period in this paper is determined as 7:00–9:00, and the evening peak period is determined as 17:00–19:00), and about 5 million pieces of track records are recorded every day. According to the vehicle operation status reflected by the on-off tags in the track data, the average passengers carried in the peak period are 85,083 times in the morning peak and 81,773 times in the evening peak respectively.

2.1. GPS Trajectory Data Semantics and Travel Event Extraction

GPS track data of a taxi is generated by vehicle terminal equipment and uploaded to data management center by instant communication. The data content mainly includes basic information such as vehicle license plate, driving time, driving mileage, active location (including the location of getting on and off the vehicle) and driving speed, etc. Taking the data collected by GPS devices of several taxis in Xi'an as an example, the trajectory data is reported at a 30-s interval, including real-time information such as longitude and latitude (including altitude), instantaneous speed, running direction (360 degrees), and passenger carrying status of vehicles operating on the internet. Based on this information, more and more accurate taxi parameters can be obtained, and based on these parameters, a more restrictive model can be established [19,20].

The trajectory data reports the real-time information of the latitude and longitude (including altitude), instantaneous speed, running direction (360 degrees) and passenger status of the operating vehicle at a certain time interval. Therefore, the daily traffic activity behavior of a taxi can be considered to be composed of multiple GPS track points, which are linearly linked in accordance with the time series to constitute the vehicle's driving track and reflect the information of the vehicle's passenger activity. Taking the data values in Table 1 as an example, the exact location of the up-down passenger event can be considered as the two continuous points between the track state of the up-down passenger (vehicle status from "0"→"1") and the track state of the down-down passenger ("1"→"0"). Therefore, the pick-up and drop-off event of a taxi can be considered as a linear event with direction and starting and ending points from "0→1, 1→0" (Figure 1). In the figure, the solid line represents the taxi's driving track in the state of carrying passengers, and the dotted line represents the taxi's driving track in the state of vacant driving.

Table 1. Research data description.

Field	Meaning	Example	Notes
Licenseplateno	License plate number	AU1234	Unique key
in_date	Record starting time	2012/8/1 0:00:00	Unique key
Longitude	Longitude	108.94380	Five decimals
Latitude	Latitude	34.22981	Five decimals
Height	Altitude	446	No decimal
Speed	Horizontal speed	26	Unit: kilometer/hour
Direction	Direction	90	Angle with north, clockwise
Eff	Current status	1	0—invalid;1—valid
car_stat1	Car status	1	"0"—vacant; "1"—occupied; "2"—turned off
car_stat3	Mileage	1325	Cumulative mileage, unit: meter
car_stat4	Cumulative time	136584540000	Cumulative time, unit: millisecond

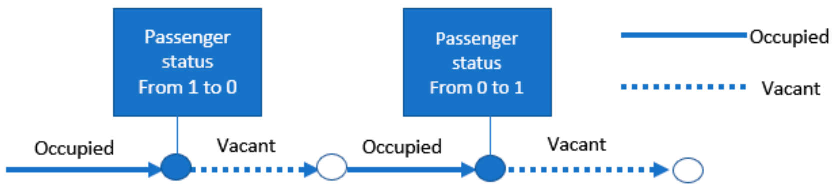


Figure 1. Trajectory data description of GPS activity events.

2.2. Trajectory Data Logic Model

The linear events formed by the taxi trajectory data can be considered as the time series fed back by the vehicle GPS information, which consists of a series of point sets containing time information, space information and operation information. It can be represented as a set $\{G_1,G_2,G_3 \ldots \mathrm{G_m}\}$, where n is the total number of trajectory points formed by a taxi in a specified period of time, G_i is the locus point at a certain moment, for any G_i, we have $i \in (1, m)$. Based on the basic composition of GPS track data, according to the research needs, G_i can be designed as a one-dimensional array with six basic items of information, $G_i = \{c_i,a_i,b_i,j_i,s_i,v_i\}$, where c_i is the vehicle license plate, a_i and b_i are the longitude and latitude values of the vehicle, s_i and v_i are passenger carrying status and driving speed of the vehicle at moment j_i respectively, where $s_i = 1$ denotes occupied and $s_i = 1$ denotes vacancy.

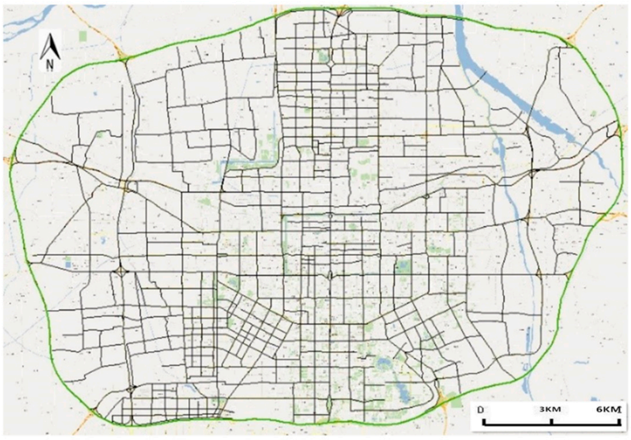
2.3. Geographic Information Matching

This part mainly matches the GPS track data of taxi with the electronic map, so as to realize the correspondence between the spatial attribute of the vehicle data and the actual geographic information. It associates points and sections in different coordinate systems, finds the road corresponding to the vehicle's driving track, and then determines the specific position (and driving direction) of the vehicle on a certain section of the map, so that the characteristics of vehicle travel activities and spatial laws can be analyzed. Taking Xi'an City as an example, with the help of the electronic map of the network (Figure 2a), based on this, the electronic map contains two parts: spatial information and attribute information. Spatial information includes geographic location information and topological attribute information. Spatial features are abstracted into the form of nodes, sections (lines) and traffic communities (planes).

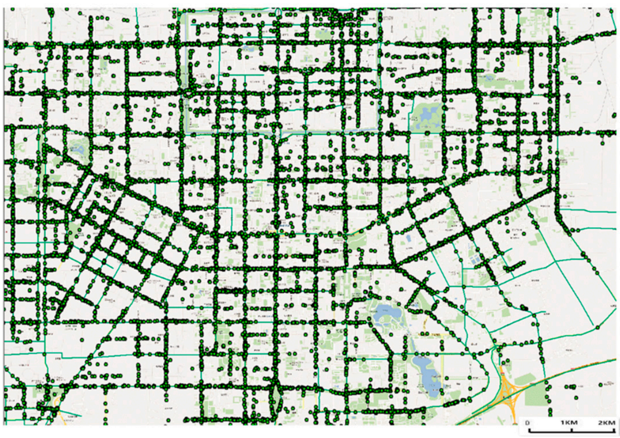
The process of geographic information matching method used in this paper is as follows:

1. Unified coordinate system which means that the coordinate system used to determine the vehicle GPS data is consistent with the map.
2. With the help of ArcGIS tool (Environmental Systems Research Institute, Inc -Esri. Redlands, CA, USA), the shortest distance method is used to determine the road section to which GPS track points belong.
3. Match the direction information of road segment with the direction information of track point, determine the direction of vehicle travel. Classify the track points and the information, such as speed, is imported into the spatial attribute database

In practical operation, this paper uses Xi'an-1980 plane coordinate and WGS-1984 earth coordinates are adopted in this paper for GPS track data. In order to ensure the consistency of the coordinate system, coordinate transformation is realized through ArcGIS. The result of map matching is shown in Figure 2b and vehicle track restoration is shown in Figure 2c.

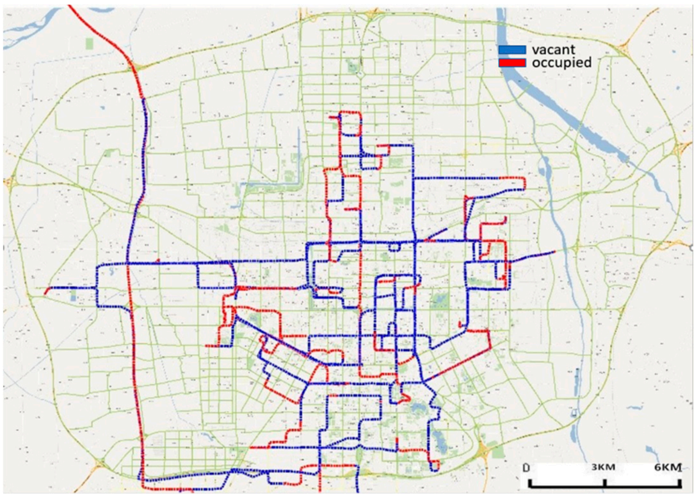


(**a**) Road network extraction.



(**b**) Trajectory matching.

Figure 2. *Cont.*



(**c**) Vehicle track restoration (one vehicle demonstration).

Figure 2. Road network extraction, trajectory matching, and vehicle track restoration.

3. Methodology

3.1. Ideas and Assumptions

As a kind of urban traffic supply, it should be the fundamental starting point to consider the reasonable fleet size from the perspective of solving the difference between supply and demand. However, it should be noted that many factors, such as the effectiveness of urban road network, the development degree of public transportation, the operating efficiency of a taxi (including time efficiency and mileage efficiency), the spatial distribution of taxi travel, and the price of a taxi will have an impact on the demand scale of a taxi [21]. In addition, different modes of taxi hiring have different efficiencies of meeting demands. An important point is that the travel demand of a taxi is essentially reflected in the time and space of one or more areas "in short supply (difficult to take a taxi)", such as the peak hours in some hot areas. However, in a flat hump period, even in some remote areas (areas that are not easily accessible or not willing to be reached by taxi), taxis are actually not in short supply; in essence, it can be understood as a scheduling problem. That is, when the fleet size is sufficient, the subproblems can be solved through reasonable operation scheduling. Therefore, this situation is not within the scope of this article. In conclusion, the reasonable fleet size of taxi should be carried out under certain assumptions. The assumptions set in this paper are as follows:

1. During a research period, residents' travel habits remain unchanged, and there is no significant change in the urban comprehensive transportation system (rail transit is newly built, etc.). The overall number of taxi trips is stable; that is, in other words, the number of taxi rides is generally stable.
2. The operating price of taxi and other operating prices are stable, which means there will be no transfer or change of the traveling population caused by economic factors.
3. The research object of this paper is the number of times to take a taxi, rather than the number of people to take a taxi, which means one person or more taking a taxi together is considered as one time.
4. The fleet size that is needed to be added in this paper's discussion, is the amount of fleet needed for unmet demand on these "not supply (hard to take a taxi)" road sections or areas during the peak hours. As explained in the previous paragraph, the problem of "difficulty in taking a taxi" in remote areas is not considered as a real lack of demand, so it is not within the scope of this paper.
5. All the car-hailing modes involved in this paper follow the principle of "first come, first served" and do not consider the taxi's no-load behavior when driving vacant; that is, if a "vacant car"

passes a road section without stopping to pick up a passenger, then the road section is deemed to have no demand.

6. The service form of the taxi in this paper is traditional taxi service, which does not consider the use of telephone or network for lease, because this situation will make the five-hypothesis invalid.

Based on the above assumptions, this paper takes the average effective mileage provided by a taxi and the total demand for taxi travel in the city as the core research objects, discusses taxis under the constraint of considering the change threshold of the taxi driver's income change, and looks at the relationship between the fleet size of the taxi and the total passengers transport volume when the taxi fleet size meets the transportation demand. Steps are as follows:

1. Divide the roads network into multiple section units, on which the passengers' taxi-hailing actions were abstracted into a queuing problem, based on the random process theory to build the probability model of taxi [22]. Through calculating the difficulty degree of taking a taxi for each road section, find out the road sections with "short supply (difficult to take a taxi)". These road sections are called "difficult to take a taxi areas".
2. Calculate the unmet demand for taxi according to queuing theory.
3. Based on the track data, the operating efficiency of taxis in different time ranges is calculated to seek the effective operating efficiency provided by taxis in peak hours
4. Through the unmet total demand and the taxi operating efficiency, seek the reasonable size range of taxi fleet demand.
5. Calculate the change of taxi driver's income under different fleet demand scale, or under the influence of different income, the taxi driver's fleet size.

3.2. Taxi Operation Index

3.2.1. Mileage Utilization

The mileage utilization rate of taxi refers to the percentage of passenger miles in the total mileage during the operating period of the taxi, which can reflect the operating status of taxi more intuitively. The average mileage utilization rate of the overall taxis a city can reflect, to some extent, whether taxi fleet size in a city is reasonable; the lower the mileage utilization rate is, the greater the invalid traffic flow caused by taxis in the urban road network and excess taxi fleet scale will be. Mileage utilization is calculated as follows:

The average operating mileage can be expressed as:

$$g_z = \frac{\sum g_{ij}}{N} \quad t_2 \in T \tag{1}$$

where t_2 donates unit time, T is the time range (can be taken in 24 h or taken in unit time), g_z denotes the average operating mileage in the time range of T, N denotes total amount of the taxi fleet size, i denotes the ith taxi, j denotes that the taxi is in state j, where $j = 1$ denotes carrying passengers and $j = 0$ denotes vacant driving.

The average passenger mileage can be expressed as:

$$g'_z = \frac{\sum g_{i1}}{N} \quad t \in T \tag{2}$$

where g'_z donates the average passenger mileage in the time range of T.

Mileage utilization can be expressed as:

$$K = \frac{g_z}{g'_z} \quad t \in T \tag{3}$$

where K donates the mileage utilization in the time range T.

3.2.2. Time Utilization

The time utilization rate of taxi refers to the percentage of the passenger carrying time in the total working time during the operating period of the taxi, which can reflect the relationship between the working time and efficiency of taxi drivers intuitively. The working hours of taxi drivers, on the one hand, reflects the time characteristics of urban residents who rely on taxis for travel. On the other hand, the time utilization rate can reflect whether the taxi fleet size in this city is reasonable to some extent; the lower the time utilization rate is, the less the effective working hours are and the higher the surplus of the fleet size is. Time utilization is calculated as follows:

The average working time can be expressed as:

$$t_g = \frac{\sum t_{ij}}{N} \tag{4}$$

where t_g donates the average working time.

Average passenger carrying time can be expressed as:

$$t_c = \frac{\sum t_{i1}}{N} \tag{5}$$

where t_c donates the average passenger carrying time.

Time utilization can be expressed as:

$$K = \frac{t_g}{t_c} \quad t \in T \tag{6}$$

3.3. Taxi-Hailing Probability Model

Passengers may need taxi service in any segments of the urban road network, and the demand was met randomly by any vacant taxi available on-site. With collected data, taxi utilization status and the average time utilization rate in any given segment could be calculated to represent the operation efficiency. The taxi arrival could be considered as a random "birth-and-death" process according to the random theory. Poisson process was assumed, and the taxi arrival probability was presented as:

$$p_k = \frac{(\lambda x)^k}{k!} e^{-\lambda x} \tag{7}$$

where p_k denotes the probability that k taxis would arrive in x time span, λ denotes the average interval between two consecutive taxis, and x denotes the numerical time interval.

Average taxi arrival in high demand segments could be calculated with tracking data:

$$\lambda = \frac{\sum n_i}{t} \tag{8}$$

where t denotes research time span, n_i denotes the number of taxis arriving in i status, i = 1 means the taxi was occupied, and i = 0 means the taxi was vacant.

The probability of getting a taxi could be simplified according to the Taylor formula:

$$p = 1 - e^{-\frac{n_1+n_2}{t} p(1-o)} \tag{9}$$

where p represents numerical time interval i.e., passenger waiting time, n_i denotes the number of taxis arriving in i status, t denotes the research time span, o denotes time utilization rate, i.e., the ratio between carrying time and total operation time, i = 1 meant the taxi was occupied, and i = 0 meant the taxi was vacant.

The above model reflected the probability for a passenger to get taxi service in a given waiting time. The criteria to qualify the difficulty to get taxi service was defined subjectively. In this study, for example, the probability that a passenger can get a taxi in five minutes on a certain road section is 95%, or the probability that a passenger can get a taxi in 10 min on a certain road section is 50%. When the model is calculated, the standard to measure the difficulty of taking a taxi should be set according to the actual situation. If the probability of getting taxi service in 10 min is less than 75%, that segment would be labeled as a segment with low taxi availability.

3.4. Total Taxi Service Mileage Calculation Model

In each study road section, when the arrival rate of taxi passengers is higher than that of vacant cars, based on the random theory, the taxi service process can be regarded as a random birth and death process, abstracting into a queuing problem. Many researchers have agreed on viewing the taxi service as a random "birth-and-death" process from random process [23], driver route selection experience [24], and other perspectives. Assuming passengers arrive in Poisson flow with parameter λ, when there is no available taxi they start waiting and would shift to other options if the waiting time reaches a threshold and become missed taxi customers. Assume that the queuing passengers give up following a Poisson distribution with strength $\triangle k$ and the service time of vacant taxis in the given segment obeyed negative exponential distribution with parameter μ. When $\triangle k = k\delta$, $\delta > 0$, the following equation was used to calculate the probability when the number of waiting passengers in a given segment was k. When $k = 1$:

$$\lambda p_0 = \mu p_1 \tag{10}$$

When $k = 2$:

$$p_2 = \frac{\lambda^2}{\mu(u+\delta)} p_0 = \frac{p^2}{1+\beta} p_0 (\beta + \frac{\delta}{\mu}) \tag{11}$$

where λ is the passenger arrival parameter, i.e., the passenger arrival rate, and μ was service time parameter, i.e., the vacant taxi arrival rate in unit time in a given segment. λ could be obtained through survey or calculated with equation. Assume that average waiting time is known, and all passengers are transported, then the passenger arrival rate would be:

$$\lambda = \frac{\mu^2 \tau_\varepsilon}{\mu^2 \tau_\varepsilon + 1} \tag{12}$$

where τ_ε denotes the average waiting time.

Then the probability when k passengers were waiting in a given segment could be represented as:

$$p_k = \frac{p^k p_0}{(1+\beta) + (1+2\beta)....[1+(k-1)\beta]} \tag{13}$$

where ρ denotes the system load level or intensity, $\rho = \lambda/\mu$.

Average system queue length could be presented as:

$$m_s = \sum_{k=0}^{n} k p_k < \sum_{k=1}^{n} \frac{k p^k}{(k-1)! \beta^{k-1}} p_0 = \rho(1 - \frac{\rho}{\beta}) p_0 e^{\rho/\beta} \tag{14}$$

In a nutshell, the average system queue length is the function of waiting time and vacant car arrival rate. Numerically, it equals to taxi trips required to satisfy the unmet demand within a given waiting time. The unmet transportation turnover could be calculated via multiplying it with average trip distance. Total taxi service mileage refers to the total carrying kilometers covered by taxis when all the transportation demand is met within a given waiting time. When the taxi-hailing probability calculated with the above model is within a receptible range in all segments, the traffic capacity is considered to be sufficient. Otherwise, unmet demand exists, and the required total mileage is represented as:

$$w_t = w_a + w_n \qquad t \in T, \tag{15}$$

where w denotes the required total taxi mileage within time span t, w_a denotes the finished taxi mileage within time span t, and w_n denotes the unmet taxi mileage within time span t.

The unmet taxi mileage within time span t could be represented as the sum of the unmet taxi trip distance in all segments with low taxi availability in the urban road network:

$$w_n = \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} m_i \bar{l}_i \tag{16}$$

where w_n denotes the unmet taxi mileage within time span t, w_i denotes the unmet taxi mileage in the i segment, m_i denotes the unmet taxi requests in the i segment, i.e., the missed trips, and $\bar{l}_i$ denotes the average trip distance in the i segment.

It is inevitable that the taxi will generate invalid mileage during the cruise, due to the difference of taxi transportation demand in different time periods and the decrease of taxi running speed caused by urban congestion. There are obvious differences in the mileage utilization rate of taxi, therefore, the operating efficiency of taxis in different time periods is affected by the speed and the online rate of vehicles (the ratio of the number of vehicles put into operation in unit time to the total number). According to the definition of mileage utilization, it can be expressed as:

$$h_t' = uh_t = uK\sum_{i=1}^{n_t} q_i \quad t \in T \tag{17}$$

where u is taxi mileage utilization (%), $h_t{}'$ is the sum of the effective mileage of all taxis operating within a unit of time (km), h_t is the total number of miles driven by all operating taxis per unit of time (km), n_t is the fleet size of taxi put into operation within the time range of t (unit), T is the time range, unit operation time, i is the ith taxi, and q_i is the mileage of the ith taxi in the time range of t.

The average effective mileage supplied by a taxi cycle in the time range t can be expressed as:

$$h_a' = \frac{u\sum_{i=1}^{n_t} q_i}{n_t} \tag{18}$$

where $h_a{}'$ donates the average effective mileage supplied by a taxi cycle in the time range t (km), and n_t is the fleet size of taxi put into operation within the time range of t (unit).

3.5. Taxi Fleet Size Adjustment Model

Above all, the incremental taxi volume required in a period could be calculated [24]:

$$\chi = \varphi \frac{w_n}{L_a'} \tag{19}$$

where φ denotes the satisfactory degree coefficient varying between 0 and 1, w_n denotes the unmet urban taxi passenger transportation turnover in an hour, and L_a' denotes the average taxi effective mileage in unit time.

Incremental taxi volume needed to satisfy unmet demand in each hour could be calculated with the volume model. The final number should lie within the range defined by the value in peak hours and that in off-peak hours, which is approximately zero. Therefore, the total incremental taxi volume was calculated as follows:

$$x^* = \left\{\partial_1 x_1, \partial_2 x_2, \cdots \partial_n x_j\right\} \tag{20}$$

where x^* denotes the final incremental capacity, ∂ denotes the satisfactory degree coefficient varying between 0 and 1, and x_j denotes the incremental capacity in each of the time unit within 24 h.

3.6. Taxi Market Supply-Demand Mechanism and Income Constraint Model

The above method was built upon the assumption that both the taxi industry and resident transportation demand would remain stable. The basic concept was to mitigate the difficulty of getting taxi service during the times where taxi demand is high. Demand for a taxi is difficult to quantify, but could be reflected by the deadhead rate, which is correlated to operation cost, operation efficiency, and average waiting time. The vacancy rate reflects the degree of passengers' demand for taxis from a certain perspective. The smaller the vacancy rate is, the greater the passengers' demand will be. [25–28]. Similarly, the operational effectiveness is also related to the vacancy rate. Under the condition of constant input quantity and freight rate, the higher the vacancy rate is, the greater the waste of resources will be and the smaller the operating profit will be. As a result of the launch of new transport capacity, passengers' waiting time will be reduced, and residents' travel choices will change. Accordingly, the number of trips completed by taxis will increase, and the transportation demand will increase. Therefore, the supply and demand mechanism of the taxi market has been in dynamic change, as illustrated in Figure 3.

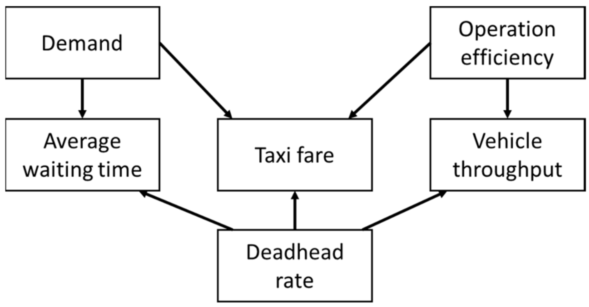


Figure 3. Supply and demand dynamic in the taxi market.

Considering the dynamics and complexity of the supply and demand balance mechanism of the taxi, after the delivery of transport capacity, it reaches a new steady state rather than an equilibrium state, which can be observed according to the calculation model. When the stability is reached again and the waiting time exceeds the threshold, the delivery can continue. The delivery of taxis also needs to comply with the laws of the market and constrain the delivery of capacity through changes in income:

$$\varepsilon = 1 - \frac{c_u}{c_o} = 1 - \frac{\sum_{i=1}^{t} c_i}{\sum_{j=1}^{t'} c_j} \tag{21}$$

where ε denotes the income variation index ranging between 0 and 1, c_o denotes the average income before volume increase, c_u denotes the average income after volume increase, c_j denotes the average income at j hour before volume increase, c_i denotes the average income at i hour after volume increase, t' denotes average working hours before volume increase, and t denotes the average working hours after volume increase.

4. Results and Discussion

4.1. Analysis on the Taxi Operation Characteristics

According to Equations (1)–(3), the average daily mileage utilization rate of taxis in Xi'an is 66%, and the mileage utilization rate in morning and evening peak hours is 76% and 75%, respectively.

The distribution of weekdays and non-weekdays is shown in Figure 4. The overall trend in mileage utilization shows two low peaks, which are 3:00–4:00 and 15:00–16:00. Compared with the first peak, the mileage utilization ratio of rest days and weekdays presents a certain degree of "dislocation" phenomenon, the first peak of the weekdays occurs at 8:00, and the first peak of the rest days occurs at 11:00, and the peak of the rest days mileage utilization lags behind the weekdays, the second peak of weekdays and rest days overlaps at 18:00 and remained stable at 22:00. Overall, the utilization efficiency still has a high mileage utilization rate even at night, which is directly related to the number of passengers and the number of operating taxis, as shown in Figure 4.

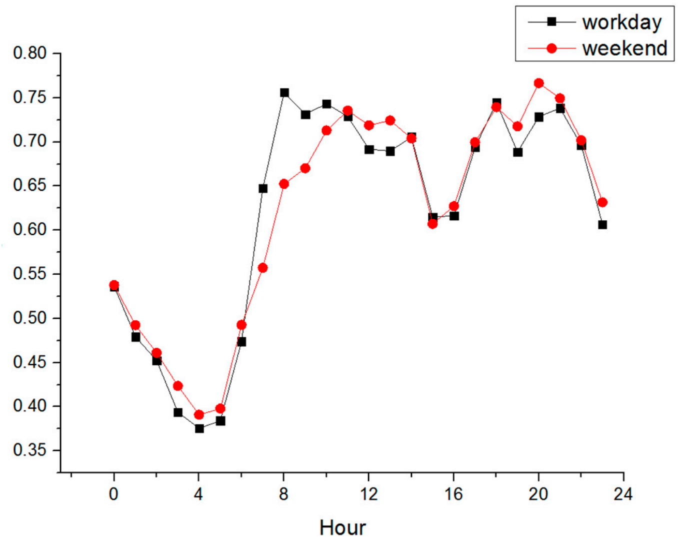


Figure 4. Interval distribution of the mileage utilization rate.

According to Equations (4)–(6), the calculation results show that the utilization rate of morning and evening peak hours in Xi'an are 76% and 75%, respectively. At the same time, there is a strong correlation between mileage utilization rate and time utilization rate, the correlation coefficients are all over 0.95, and the trend of time utilization efficiency and mileage utilization efficiency is basically consistent. According to the survey results of residents' travel in Xi'an in 2008, the trend of the time utilization rate of taxis on weekdays is basically consistent with the distribution of residents' travel time. The two peak periods are the commuting peak of urban residents respectively, and these two trips account for about 25% of the total travel volume of residents, as shown in Figure 5.

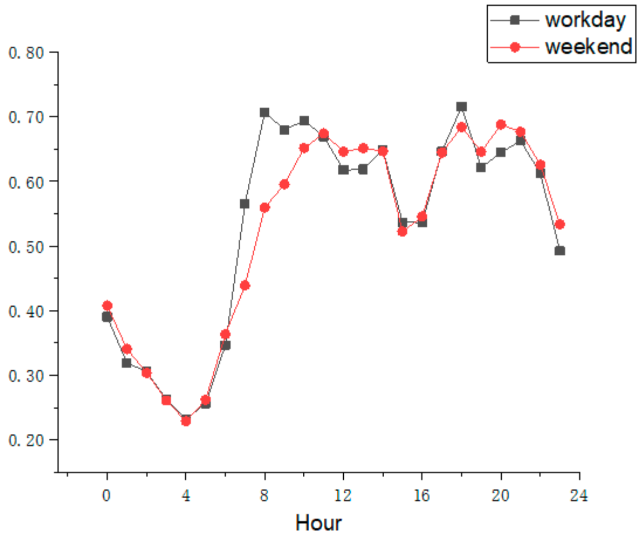
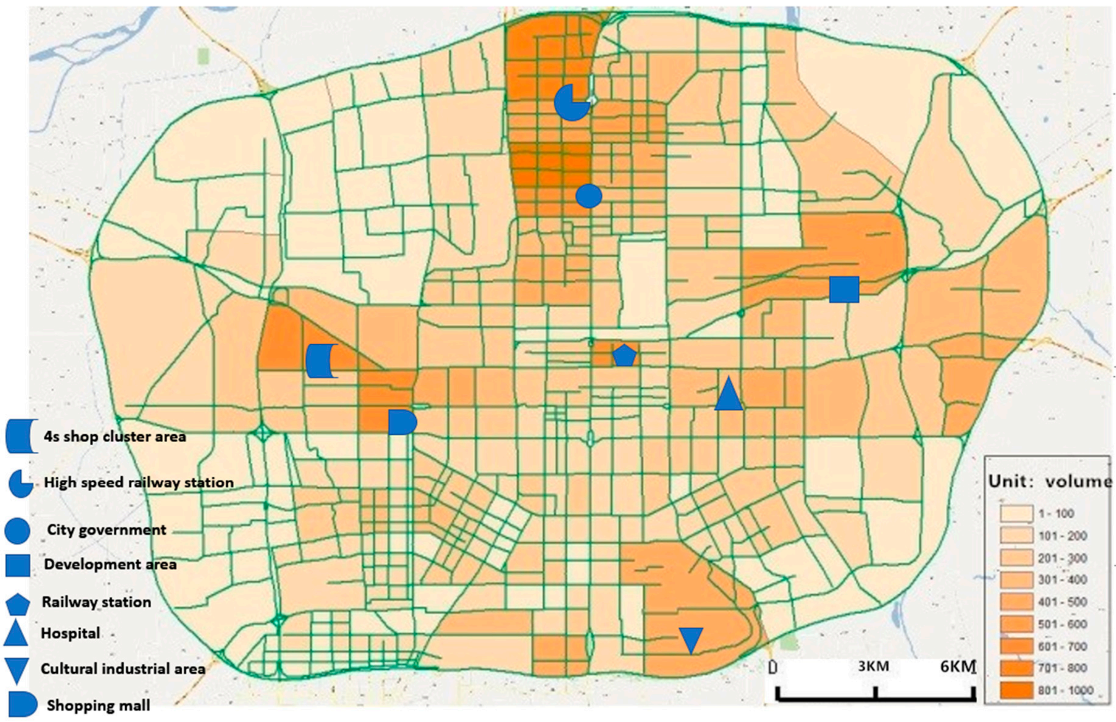


Figure 5. Interval distribution of the time utilization rate.

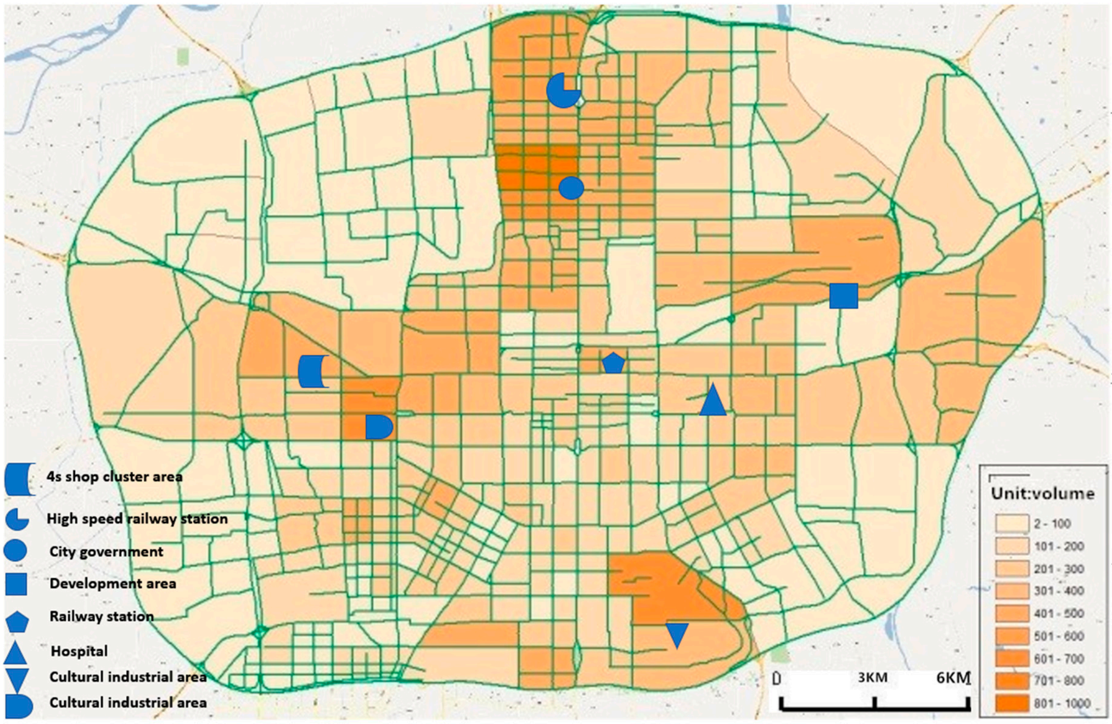
4.2. Taxi Operation Space Characteristics

In this paper, the regional taxi demand scale based on the division of the community is used to illustrate the characteristics of the taxi operation space, a total of 201 traffic communities are divided in Xi'an within three ring roads. The ArcGIS spatial analysis tool is used to study the changes of hot spots where residents use taxis to travel at any time. The comparison of getting on the taxis in morning and evening peak hours of Xi'an city is shown below as Figure 6. Volume in the maps actually means the taxi-taking times.



(**a**) Morning peak hours.

Figure 6. *Cont.*



(**b**) Evening peak hours.

Figure 6. Traffic zone's pick-up volume distribution in peak hours.

The results show that the regions with high demand frequency in the Xi'an taxi network are located in the north, west, southeast, and northeast, respectively. During peak hours, the amount of taxi rides in hot spots totaled nearly 3000 times. After map matching, it can be seen that the geographic locations corresponding to the four regions are the municipal government, municipal library, large supermarket, shopping mall, office building, hospital, railway station, and other major functional sites in cities, which are basically consistent with the actual situation, where these "hot spots" are the focus of this paper.

4.3. Fleet Size Calculation and Adjustment

According to the taxi probability model, to calculate the taxi probability of each section, due to the subjective differences in the degree of difficulty in taxi taking between different regions and different groups. In this study, if the probability of a passenger getting a taxi within five minutes of a section is less than 75%, it is considered as a "taxi difficult" section. Bringing the taxi data into the above calculations indicates that there are road sections that are difficult to travel in multiple time frames, and it is necessary to increase the fleet size to meet the transportation needs.

With the available data and the above methodology, let the value of ∂ be 1, and the incremental taxi volume in Xi'an to completely satisfy transportation demand for each hour from 8:00–13:00 and from 17:00–22:00 was calculated and shown in Table 2.

The results show that the required incremental volume ranged between 654 and 2237 taxis, with the requirement for peak hours significantly higher than that for off-peak hours. In practice, an increase in accordance to the maximum demand would result in excessive capacity in off-peak hours.

Table 2. Incremental taxi volume in specific hours.

Time [1]	8:00	9:00	10:00	11:00	12:00	13:00
Current Mileage Utilization Rate (%)	0.76	0.73	0.74	0.73	0.69	0.69
Current Speed (km/h)	22.86	22.81	22.36	23.81	26.80	26.31
Calculated Incremental Volume (cars)	2150	2022	1126	824	658	784
Time	17:00	18:00	19:00	20:00	21:00	22:00
Current Mileage Utilization Rate (%)	0.69	0.74	0.69	0.73	0.74	0.70
Current Speed (km/h)	22.02	20.02	24.94	25.99	26.68	28.47
Calculated Incremental Volume (cars)	1197	2237	1525	1474	1211	654

[1] Referred to the hour starting from the indicated time.

One discussion point to note is the relationship between the results and the assumptions. Firstly, with regard to assumptions 1 and 2, the price level and overall demand of the taxi during this period are assumed to be constant, once the price changes. Then, the demand for taxis will drop. This paper does not specifically analyze the sensitivity of prices. However, if the price of taxis is lower, the demand will increase at any time. Secondly, with regard to assumption 3, this article focuses on the number of taxis, not the number of people traveling because the number of taxis essentially determines the demand for the vehicle, not the number of people. With regard to assumption 4, this paper is based on the demand gap of the "taxi-difficult" sections to measure the required capacity. However, under different levels of difficulty in taxiing, the scale of demand for taxis is also different if city residents have a low threshold for waiting for taxis and the demand for capacity is greater. With regard to assumptions 5 and 6, it is a certain constraint on the form of service of the taxi, that is, without considering the refusal of a taxi call and other means of recruiting. This is because once the situation of the refusal is considered, the existing vehicle utilization efficiency will decrease, and more vehicles will be needed. In addition, if the vehicle is called by other means, the number of vacant vehicles on a certain section of the road section will not truly reflect the utilization efficiency of the taxi. Finally, on the premise of the above assumptions, another point to note is that the value of the capacity of the taxi car, if configured according to the "gap" of the peak period, is achieved according to the peak configuration. This inevitably leads to a decrease in the utilization efficiency at the peak period. Therefore, the capacity configuration of the taxi will determine an initial increase in the case of determining the total "gap". In this paper, the actual incremental volume was set to be 70% of the maximum value, i.e., 1566 taxis. In theory, 70% of the unmet demand in peak hours could be met and basically all demand in off-peak hours could be satisfied after this increase.

After that, the tracking data is used to calculate the operating income of taxi in unit time before and after capacity adjustment. This paper takes single-shift taxis as an example. Pre- and post-adjustment taxi driver income in each hour from 6:00 to 15:00 was calculated with tracking data and summarized in Table 3 and Figure 7.

Table 3. Xi'an taxi driver income changes.

Time	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	Total
Pre-adjustment	20.12	41.09	45.42	44.07	43.30	43.51	42.75	44.64	44.40	35.65	404.96
Post-adjustment	17.97	32.66	44.43	43.22	38.19	37.44	36.30	38.26	35.11	28.33	351.92

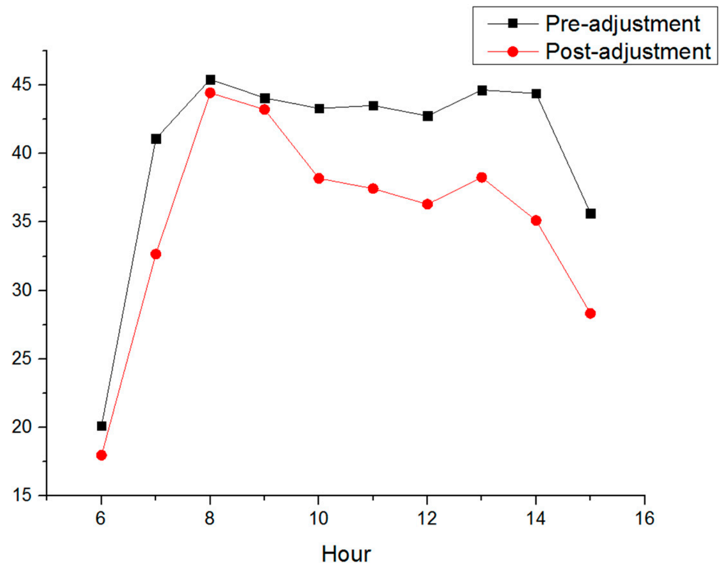


Figure 7. Driver income changes between pre-adjustment and post-adjustment.

According to calculation, taxi driver income variation index was 0.13 when the capacity was increased to fully satisfy demand. When the variation index was restrained to less than 0.10, the incremental capacity would be 1286 vehicles, which could meet 66% of the total unmet demand in peak hours without leading to a decrease larger than 10% in taxi driver income. The results show that the overall taxi capacity of Xi'an city has not yet reached the reasonable supply quantity, and the impact on taxi operation utility should be considered while increasing the capacity scale. The calculation results can be used to guide the increase of the city's capacity, as in theory the demand during peak and off-peak hours should be able to be improved and met with greater efficiency than the current market performance. The introductory model of taxi hailing can be introduced to identify taxi hailing difficulty areas to analyze the difficulty degree of taxi hailing. Based on previous studies, the distribution process of vehicle arrival is simplified, and the distribution calculation for the arrival of taxis needs to be further improved.

5. Conclusions

The study fully considers the spatial difference and transportation capacity of the taxi transportation demand and provides accurate taxi trajectory data and trajectory data identification method based on the collected taxi GPS data. From this research, an incremental taxi volume model based on queuing theory was created to address the supply and demand issue of the current taxi market to understand better how to mediate the difficulty of hailing a taxi during peak demand in certain areas by utilizing driver income as a constraint.

Using the trajectory data, combined with the queuing theory model, the research started from the perspective of meeting the passenger demand in the area where taxiing is difficult. To meet the passenger demand, the passenger mileage provided by the taxi is used as the research object, and the traffic capacity is controlled under the constraint of the taxi operation utility. Incremental volume was calculated based on carrying mileage required to satisfy the unmet transportation demand. The reasonable capacity of urban taxis was studied, and the calculation of the reasonable increase in capacity was carried out. The relationship between the increase in capacity and the change coefficient of driver income was obtained. The case study in Xi'an demonstrates that the method is effective and operable, and it can provide a certain practical application basis and calculation method for the development of the taxi industry. In addition, results also suggested that the current supply in Xi'an was not yet sufficient and further increase is needed in order to meet the current met and unmet demand during peak and off-peak hours.

Author Contributions: Conceptualization: Y.Y. and X.F.; methodology: Y.Y. and X.F.; software: Y.Y.; validation: Y.Y., Z.Y., and X.F.; formal analysis: Y.Y.; investigation: D.S.; resources: X.F.; data curation: D.S.; writing—original draft preparation: Y.Y. and X.F.; writing—review and editing: Y.Y. and X.F.; visualization: D.S.; supervision: Y.W.; project administration: X.F. and Y.Y.; funding acquisition: Z.Y.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities, grant number 2017YJS113 and the Natural Science Foundation of Hebei Province, grant number E2016210131.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fu, X.; Sun, M.P.; Sun, H. Taxi Commute Recognition and Temporal-spatial Characteristics Analysis Based on GPS Data. *China J. Highw. Transp.* **2017**, *30*, 134–143.
2. Yang, Y.; He, Z.; Song, Z.; Fu, X.; Wang, J. Investigation on structural and spatial characteristics of taxi trip trajectory network in Xi'an, China. *Phys. A Stat. Mech. Its Appl.* **2018**, *506*, 755–766. [CrossRef]
3. Douglas, G.W. Price regulation and optimal service standards: The taxicab industry. *J. Transp. Econ. Policy* **1972**, *6*, 116–127.
4. Abbas, K.A. Estimation of taxi fleet size: A generic algorithm. *Sigma J. Eng. Nat. Sci.* **2016**, *7*, 21–29.
5. Yao, B.; Jin, L.; Cao, Q.; Gao, J.; Zhang, M. Fleet size and fare optimization for taxi under dynamic demand. *J. Transp. Lit.* **2016**, *10*, 45–50. [CrossRef]
6. Zhang, W.; Ukkusuri, S.V. Optimal fleet size and fare setting in emerging taxi markets with stochastic demand. *Comput. Aided Civ. Infrastruct. Eng.* **2016**, *31*, 647–660. [CrossRef]
7. Wang, Y.; Su, J.; Hou, J. Determination of reasonable ownership scale for urban taxicab system. *J. Chongqing Jiaotong Univ. Nat. Sci.* **2014**, *33*, 131–134.
8. Yang, Y.; Yuan, Z.; Chen, J.; Guo, M. Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction. *Environ. Eng. Manag. J.* **2017**, *16*, 2413–2423. [CrossRef]
9. Sun, D.; Jia, Y.; Qin, L.; Yang, Y.; Zhang, J. A Variance Maximization Based Weight Optimization Method for Railway Transportation Safety Performance Measurement. *Sustainability* **2018**, *10*, 2903. [CrossRef]
10. Ghahramani, M.; Zhou, M.C.; Hon, C.T. Mobile Phone Data Analysis: A Spatial Exploration Toward Hotspot Detection. *IEEE Trans. Autom. Sci. Eng.* **2018**, 1–12. [CrossRef]
11. Tachet, R.; Sagarra, O.; Santi, P.; Resta, G.; Szell, M.; Strogatz, S.H.; Ratti, C. Scaling law of urban ride sharing. *Sci. Rep.* **2017**, *7*, 42868. [CrossRef] [PubMed]
12. Alonso-Mora, J.; Samaranayake, S.; Wallar, A.; Frazzoli, E.; Rus, D. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 462–467. [CrossRef] [PubMed]
13. Vazifeh, M.M.; Santi, P.; Resta, G.; Strogatz, S.H.; Ratti, C. Addressing the minimum fleet problem in on-demand urban mobility. *Nature* **2018**, *557*, 534–538. [CrossRef] [PubMed]
14. Santi, P.; Resta, G.; Szell, M.; Sobolevsky, S.; Strogatz, S.H.; Ratti, C. Quantifying the benefits of vehicle pooling with shareability networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13290–13294. [CrossRef] [PubMed]
15. Xu, W.; Zhou, H.; Cheng, N.; Lyu, F.; Shi, W.; Chen, J.; Shen, X. Internet of vehicles in big data era. *IEEE/CAA J. Autom. Sin.* **2018**, *5*, 19–35. [CrossRef]
16. Castro, P.S.; Zhang, D.; Li, S. Urban traffic modelling and prediction using large scale taxi GPS traces. In *International Conference on Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 57–72.
17. Zhan, X.; Ukkusuri, S.V. Probabilistic Urban Link Travel Time Estimation Model Using Large-Scale Taxi Trip Data. In Proceedings of the Transportation Research Board 94th Annual Meeting, Washington, DC, USA, 11–15 January 2015.
18. Bonola, M.; Bracciale, L.; Loreti, P.; Amici, R.; Rabuffi, A.; Bianchi, G. Opportunistic communication in smart city: Experimental insight with small-scale taxi fleets as data carriers. *Ad Hoc Netw.* **2016**, *43*, 43–55. [CrossRef]
19. Hu, J.; Huang, Z.; Deng, J. A Hierarchical Path Planning Method Using the Experience of Taxi Drivers. *Procedia Soc. Behav. Sci.* **2013**, *96*, 1898–1909.
20. Yang, Y.; Yuan, Z.; Chen, J.; Zheng, Y.; Sun, D. A New Method to Calculate the Cost of Urban Rail Transit Operation. In Proceedings of the International Conference on Chemical, Material and Food Engineering, Kunming, China, 25–26 July 2015.

21. Yang, Y.; Yuan, Z.Z.; Li, J.Y.; Wang, Y.H.; Wang, W.C. Multi-Mode Public Transit OD Prediction and Scheduling Models-ATS-final. *Adv. Transp. Stud.* **2018**, *s3*, 133–146.
22. Wong, K.I.; Wong, S.C.; Yang, H. Modeling urban taxi services in congested road network with elastic demand. *Transp. Res. Part B* **2001**, *35*, 819–842. [CrossRef]
23. Zhang, J.; Meng, W.; Liu, Q.; Jiang, H.; Feng, Y.; Wang, G. Efficient vehicles path planning algorithm based on taxi GPS big data. *Opt. Int. J. Light Electron Opt.* **2016**, *127*, 2579–2585. [CrossRef]
24. Hu, J.; Xie, H. Floating car data based on method for determining taxi scale. *Transp. Stand.* **2011**, *18*, 43–49.
25. Flores-Gurl, D. An Economic Analysis of Regulated Taxicab Market. *Rev. Ind. Organ.* **2003**, *23*, 255–266. [CrossRef]
26. Coffman, R. The economic reasons for price and entry regulation of taxicabs. *Transp. Econ. Policy* **1977**, *11*, 3.
27. Schaller, B. Elasticities for taxicab fares and service availability. *Transportation* **1999**, *26*, 283–297. [CrossRef]
28. Kumarage, A.S.; Bandara, M.; Munasinghe, D. Analysis of the economic and social parameters of the Three-Wheeler Taxi service in Sri Lanka. *Res. Transp. Econ.* **2010**, *29*, 395–400. [CrossRef]