Study on the minimum fleet size based on deadheading time threshold

ZHAI hui\textsuperscript{a}, ZHANG yong\textsuperscript{b}

\textsuperscript{a}School of Rail Transportation, Soochow University, Suzhou, China
\textsuperscript{b}School of Rail Transportation, Soochow University, Suzhou, China

\textsuperscript{a}m15150686277@163.com
\textsuperscript{b}sinkey@162.com

Abstract. By integrating travel demand and car-hailing supply information, the car-hailing platforms have realized a more targeted non-cruise mobile order receiving mode. That reduces the deadheading (DH) time and the number of online car-hailing vehicles. This result will lead to a significant change in the number of online car-hailing vehicles. This paper studies the minimum fleet size based on deadheading time threshold. Then we use Suzhou taxi data to solve the minimum number of vehicles. Firstly, the space-time description of online car-hailing operation was established to study the influence of the DH time threshold on the total number of vehicles. The results show that the total number of vehicles decreases with the increase of the DH time threshold, and the trend tends to be gentle. Secondly, this paper discusses the relationship between the time cost of passenger and driver and the DH time threshold. Thirdly, this paper discussed the influence of the number of car-hailing platforms on the total number of vehicles, and it proves that the total number of vehicles in the multi-platform market is larger than that monopolized by a single platform. Finally, the feasibility of the research is verified by using Suzhou taxi data. Meanwhile, it proves that the minimum total number of vehicles has a certain robustness.

1. Introduction

Car-hailing is a business model with mobile APP as the main service platform. Through the integration of supply and demand information, it achieves a reasonable match between passengers and drivers, and provides a guaranteed ride service\cite{1}. On the one hand, too many online car-hailing makes it difficult to take orders and aggravates traffic congestion\cite{2}. On the other hand, too few online car-hailing results in difficult taxi-hailing, which significantly increases the waiting time for passengers. Therefore, optimizing the scale of the online car-hailing market is important for ensuring the balance of supply and demand in the taxi market\cite{3}. It is of great practical significance to improve the service quality of the industry, improve residents’ public travel conditions, and promote the healthy development of the online car-hailing industry.

There are three methods to study the number of vehicles, including the number of vehicle prediction method based on the minimum operating cost, the number of vehicles prediction method based on taxi and online car-hailing trajectory data, and the number of vehicles prediction method based on the road network. The research idea of the first method is to construct a car-sharing scheme, and establish a mathematical model based on the minimum operating cost to determine the minimum number of vehicles. Chen\cite{4} established a mixed-integer linear programming (MILP) model to determine the
minimum operating cost of vehicle scheduling and ride-sharing schemes. And it pointed out that car-sharing can effectively reduce the total number of vehicles. However, the study did not consider that the ride system is dynamic. In this regard, Dandl[5] considered the dynamic change rule of passenger travel demand in a shared network and established an optimization model for the total number of vehicles that minimize operating costs. Bianchessi[6] proposed dynamic adjustment of prices to change the total number of vehicles basing on the vehicle sharing system, proposed dynamic adjustment of price to change the total number of vehicles. And MP Fanti[7] established a two-stage optimization model for the total number of electric car-hailing vehicles that maximize network profits, which was based on Petri net and closed queuing network theory. Georg[8] established a closed queuing network model of the vehicle sharing system. Then they developed a profit-maximizing optimization problem for determining the optimal fleet size and obtained an approximate formulation as a reference. The above studies all aim at maximizing platform profit and solve the minimum total number of vehicles. The research idea of the second method is to study the trajectory data of taxis and shared cars, predict the development trend of passenger travel demand, and determine the minimum total number of vehicles. Alonso-Mora J[9] and Baker[10] proposed a heuristic algorithm to solve the total number of vehicles, and used New York City taxi data for verification. Wang Shuofeng[11] categorized related studies into three parts (the demand part, the supply part, and the mixed part) and solved the problem of the total number of vehicles served on demand. However, the above model did not consider the impact of mileage utilization on the total number of vehicles. Xiaofei Ye[12] proposed a dynamic adjustment mechanism of taxi fleet sizes based on a regression tree model and used the taxi operating datasets from Ningbo City to verify that the mileage utilization rate has the strongest effect on the fleet size of taxis. The above-mentioned research uses travel big data to solve the total number of vehicles. However, the larger the data scale, the lower the solution efficiency. So, to improve the efficiency of the solution, the method of combining the road network and graph theory came into being. That is, the third method, the research idea is to transform the road network into a graph network, predict the demand of taxis, and determine the minimum total number of vehicles. Xu Ying[13] proposed a novel graph and time-series learning model for city-wide taxi demand prediction and used a neural network to process sequence data. Guo Feng[14] developed a dataset, containing the travel speeds on each road link and in different time periods together with the real road network map. And they predicted taxi demand through the dataset. Wang guan[15] transformed the problem of solving the minimum SAV fleet size into the shortest path problem of Multiple Travelling Salesman Problem (MTSP) basing on the trajectory data of 50 new energy private cars in Shanghai for one year. A genetic algorithm is adopted to solve MTSP. And it was concluded that an SAV can replace 2.4 traditional private cars on average. However, the scale of data studied by this model is small, and it is difficult to apply to large-scale systems. So Vazifeh[16] and Yao Xiaoru[17] establish a graph theory model based on a vehicle-sharing network and transformed the problem of the minimum total number of vehicles into the minimum path coverage problem of a directed acyclic graph. It must be pointed out that the above research has two problems in solving the total number of vehicles: (1) Although different methods are used to research the total number of vehicles, the impact of the DH time threshold on the total number of vehicles is rarely discussed in the research; (2) Only one platform was considered in past studies on the total number of vehicles. However, in reality, there are often multiple platforms. The total number of vehicles on multiple platforms and the total number of vehicles on a single platform are rarely discussed.

This paper studies the total number of online car-hailing vehicles based on the taxi data in Suzhou. The research content includes: (1) According to the operating mode of the online car-hailing platform, a temporal and spatial description of the operation of online car-hailing is established, and the impact of the empty driving time threshold on the total number of vehicles is discussed. (2) The impact of the number of platforms on the total number of vehicles is considered, that is, the relationship between the total number of vehicles under multi-platform market conditions and the total number of vehicles under a single platform monopoly. (3) The model is verified to find the minimum total number of vehicles using the Suzhou taxi data.
2. Materials and Methods

2.1. Description of the method of combined vehicle transportation

2.1.1. Model assumptions
This paper makes the following assumptions to clarify the modeling conditions:

A1. It is assumed that the platform grasps the location and time information of vehicles and passengers in real-time and dispatches the driver in time. The driver responds to the dispatching instructions of the platform in time;

A2. Assuming that each vehicle travels at the same speed in the road network. The impact of the section capacity and travel demand changes on the vehicle speed is not considered;

A3. This paper does not consider factors that cause the driver to find no passengers due to small probability events;

A4. According to the current location and time of the vehicles, the platform preferentially selects the neighboring vehicles to take orders.

2.1.2. Method description of vehicle combined transportation
According to the dispatch process of the online car-hailing platform, this paper establishes a space-time coordinate system, as shown in Fig.1. The x-axis, y-axis, and t-axis respectively represent the vehicle's location and time. We are given a collection $T$ of the data released by passengers. Each trip $T_i \in T$ is defined as a tuple $T_i = \left( L^p_i, L^d_i \right)$ where $L^p_i = (x^p_i, y^p_i, t^p_i)$ represents the pick-up location and time and $L^d_i = (x^d_i, y^d_i, t^d_i)$ the drop-off location and time. The connection between the two nodes represents a trip. The solid line represents the trip of passengers before optimization. And $E = \left\{ \left( L^p_i, L^d_i \right), \ldots, \left( L^p_n, L^d_n \right) \right\}$ represents the set of solid lines. $n$ represents the total number of trips. Then Figure 1 can be expressed as $G = (T, E)$.

![Figure 1 Method description of vehicle combined transportation](image)

The platform judges the distance $d_{ij}$ between the passenger and the driver and estimates the time $t_{ij}$ for the driver to pick up the passenger. By comparing $t_{ij}$ and DH time threshold $\tau$, $d_{ij}$ and DH distance threshold $\varepsilon$, if (1) is satisfied, the passenger and the driver are matched, and the driver is assigned to
pick up the passenger at the designated place. Considering two consecutive trips $T_i$ and $T_j$ served by a single vehicle, we call the time needed to connect them the (trip) connection time, formally $t_{ij} = t_j^d - t_i^p$.

If the following relationship is satisfied by any two trips, the two trips are served by a single-vehicle.

$$\begin{align*}
d_{ij} & \leq \varepsilon \\
t_{ij} & \leq \tau
\end{align*}$$

(1)

According to (1), the itinerary set of the same vehicle service is $V = \{\delta_{ij}(\tau)\}$. And it satisfies the following relationship: if the DH time is smaller than the DH time threshold set by the platform, vehicle combined transportation is allowed; if the DH time is larger than the DH time threshold set by the platform, vehicle combined transportation is not allowed. From this, formula (2) can be obtained:

$$\delta_{ij}(\tau) = \begin{cases} 1 & t_{ij} \leq \tau \\ 0 & t_{ij} > \tau \end{cases}, \quad i, j \in \{1, 2, \ldots, n\} \cap i \neq j$$

(2)

So, the minimum total number of vehicles is

$$\min \left(Q(\tau) \right) = n - \sum_{i,j, i \neq j} \delta_{ij}(\tau)$$

(3)

According to the description on the vehicle sharing network $G = (T, E)$, solving the problem of the minimum total number of vehicles is equivalent to finding the minimum route to cover all trip demands[18]. And each trip can only be covered once. Therefore, this paper uses the Hungarian algorithm proposed by the American mathematician Harold Kuhn in 1955[19]. In this paper, the Hungarian algorithm is improved. The solution idea is to add a directed edge between the nodes representing the two trips if the two trips can be combined; if the combined transportation is not possible, then no edge is added.

2.2. Characteristic analysis of the total number of online car-hailing vehicles

2.2.1. Impact of the DH time threshold on the total number of vehicles

When $\tau$ is decreased to 0, we approach a situation in which each trip is served by a dedicated vehicle: a solution with maximal vehicle utilization that is also optimal for traffic but incurring prohibitive costs for the mobile operator. On the other hand, when $\tau$ grows excessively, the number of vehicles is reduced, but the operational and traffic efficiency problems described previously occur. Therefore, the setting of a reasonable $\tau$ is an important design choice.

Theorem 1: The minimum total number of vehicles decreases with the increase of the DH time threshold.

Proof: The DH time threshold $\tau$ is a continuous variable and the total number of vehicles $Q(\tau)$ is a positive integer. So, the graph of $Q(\tau)$ varying with $\tau$ is a scatter graph. The proof idea is to prove that if $\tau_1, \tau_2$ meet the relationship $0 < \tau_1 < \tau_2 < +\infty$, then $Q(\tau_1) - Q(\tau_2) \geq 1$ is true.

$$R = \left\{t_{ij} | t_{ij} = t_j^d - t_i^p, i, j \in \{1, 2, \ldots, n\}, i \neq j \right\}$$ represents the DH time between the drop-off and pick-up time for any two trips, which can be calculated based on historical travel demand data. $\tau_{\min} = \min R$ represents the minimum value and $\tau_{\min}^1$ represents a small value. Let $\tau_1, \tau_2$ be the arbitrary DH time threshold, which satisfies $0 < \tau_1 < \tau_2 < +\infty$. 

Let \( C = \{ p_1, p_2, \ldots, p_k \} \) be a path cover of the vehicle-shareability network \( G = (T, E) \). Then \( p_j = \{ \delta_{12} = (L_{12}^d, L_{12}^p), \delta_{23} = (L_{23}^d, L_{23}^p), \ldots, \delta_{y} = (L_{y}^d, L_{y}^p) \} \) represents a collection of trip requirements served by a single-vehicle. A set \( S = \{ T_1, T_2, \ldots, T_m \} \) represents the trips that cannot be shared with other trips, namely \( p_j = \emptyset \). So the minimum total number of vehicles is defined as \( Q(\tau) = k + m \). Next, we talk about \( \tau_1, \tau_2 \).

(1) \( 0 < \tau_1 < \tau_2 < \tau_{\text{min}} \): If \( \tau < \tau_{\text{min}} \), this situation is equivalent to that all trips require separate vehicle services, and no trip can share a car. And the total number of vehicles required is equal to the total number of trips. But this situation does not match the actual situation.

(2) \( \tau_{\text{min}} \leq \tau_1 < \tau_2 < +\infty \cap (\tau_2 - \tau_1) \geq \tau_{\text{min}}^1 \): The DH time between two trips must be less than \( \tau_1 \) since \( \tau_1 > \tau_{\text{min}} \). Let's assume that these two trips are the two trips corresponding to the minimum value \( \tau_1 \). Then these two trips can be served by a single vehicle. That is means \( p_j \neq \emptyset \) is satisfied, where \( k \geq 1 \) is established in the set \( C \). So we can get a formula \( Q(\tau_1) = k_{\tau_1} + m_{\tau_1} \leq n - 1 \). Since \( \tau_2 \) is satisfied \( \tau_2 - \tau_1 \geq \tau_{\text{min}}^1 \), the DH time between two trips must be less than \( \tau_2 \) in the network. Suppose \( \tau_{\text{min}} \) and \( \tau_{\text{min}}^1 \) are the trips respectively. There has the relationship that \( p_j = \{ \delta_{12} = (L_{12}^d, L_{12}^p), \delta_{23} = (L_{23}^d, L_{23}^p), \ldots, \delta_{y} = (L_{y}^d, L_{y}^p) \} \neq \emptyset \) and \( p_{h} = \{ \delta_{12} = (L_{12}^d, L_{12}^p), \ldots, \delta_{y} = (L_{y}^d, L_{y}^p) \} \neq \emptyset \). That means \( k \geq 2 \) established in the set \( C \). So, we can get a formula \( Q(\tau_2) = k_{\tau_2} + m_{\tau_2} \leq n - 2 \). According to the above analysis, if \( p_j \neq \emptyset \) is satisfied, then at least two trips can be served by a single vehicle. So, when \( \tau = \tau_2 \), there are two situations:

Situation 1: The DH time threshold is increased, the number of trips of a certain vehicle service has increased based on the set \( C \) that corresponds to \( \tau = \tau_1 \), which is that the independent travel demand in the original set \( S \) can be served by some vehicles in the set \( C \). However, we do not have the new element \( p_j \) to the set \( C \), which is \( k_{\tau_1} = k_{\tau_2} \) and \( m_{\tau_1} > m_{\tau_2} \). So, the conclusion \( Q(\tau_1) > Q(\tau_2) \) is proved.

Situation 2: The DH time threshold is increased, the number of trips of a certain vehicle service has increased based on the set \( C \) that corresponds to \( \tau = \tau_1 \), which is that the independent travel demand in the original set \( S \) can be served by some vehicles in the set \( C \). And we have the new element \( p_{h} \) to the set \( C \), which is \( k_{\tau_1} < k_{\tau_2} \) and \( m_{\tau_1} > m_{\tau_2} \). The trip demand of the new vehicle service \( p_{h} \) needs serves at least two independent trips.

(3) \( \tau_{\text{min}} \leq \tau_1 < \tau_2 < +\infty \cap (\tau_2 - \tau_1) < \tau_{\text{min}}^1 \): It can be seen that the total number of vehicles is equaled to \( Q(\tau_1) = k_{\tau_1} + m_{\tau_1} \) at the DH time threshold \( \tau_1 \) and the total number of vehicles is equaled to \( Q(\tau_2) = k_{\tau_2} + m_{\tau_2} \) at the DH time threshold \( \tau_2 \). For \( \tau_2 - \tau_1 < \tau_{\text{min}}^1 \), when \( \tau = \tau_2 \) is satisfied, there is no additional trip demand satisfying formula (3) that can be served by a single vehicle. Finally, the total number of vehicles corresponding to \( \tau = \tau_1 \) and \( \tau = \tau_2 \) is the same, namely \( Q(\tau_1) = Q(\tau_2) \).

In summary, when \( \tau_1 \) and \( \tau_2 \) meet the relationship \( 0 < \tau_1 < \tau_2 < +\infty \), then \( Q(\tau_1) > Q(\tau_2) \) is true. That proves the minimum total number of vehicles decreases with the increase of the DH time threshold. \( \square \)
2.2.2. The impact of DH time threshold on time cost
To determine a reasonable $\tau$, this paper establishes a mathematical model with the smallest time cost. Let the random variable $X$ represents the time between the passenger’s order and the time the driver receives the passenger. The unit is minutes, which indicates both the passenger’s waiting time and the driver’s DH time. $f(x)$ represents the probability density function of the driver and $g(x)$ represents the probability density function of the passenger.

The mathematical model is as follows:

$$W(\tau) = cn \left[ \int_0^\tau [g(x) - f(x)]dx + \int_0^\tau g(x)dx \right]$$

(4)

Restrictions: $D$: \[
g(x) - f(x) \geq 0 \quad 0 < \tau < +\infty \]

(5)

among them: $W(\tau)$ - the cost of empty driving and waiting for passengers (yuan); $c$ - empty driving cost (yuan/second).

Theorem 2: Given the probability density function $f(x)$ of empty driving time and the probability density function $g(x)$ of passenger waiting time tolerance, if both functions are of the unimodal type and meet the following conditions, then the time cost function of car-hailing has a minimum value.

Condition 1: If $f(x)$ and $g(x)$ only have one intersection, that is $g(x_0) = f(x_0)$. When the value range of $\tau$ is $\tau \in [x_0, +\infty)$, the function gets the minimum value at $\tau = x_0$.

Condition 2: If $f(x)$ and $g(x)$ only have one intersection, that is $g(x_0) = f(x_0)$ and $g(x_1) = f(x_1)$. When the value range of $\tau$ is $\tau \in [x_0, x_1]$, the function gets the minimum value at $\tau = x_0$. When the value range of $\tau$ is $\tau \in [x_1, +\infty)$, the function gets the minimum value at $\tau = x_1$.

Proof: According to the probability density function curve and constraint conditions, the graphic method is used to determine the integral area of the function, and the integrated solution is converted into the area summation to solve the minimum value of the model. The following discussion is divided into four situations.

Figure 2 (a) Diagram of the intersection of two functions

Figure 2 (b) Diagram of the intersection of two functions
Case a: Functions \( f(x) \) and \( g(x) \) have an intersection, and the integral region is shown in the shaded part \( D_1 \) and \( D_2 \) in Figure 2(a). The feasible region is \( x \in [x_0, \tau] \). At this point, the original function is equivalent to solving for the sum of the shaded areas in the figure, namely \( W(\tau) = cn\left[ S_{D_1} + 2S_{D_2} \right] \). \( S_{D_1} \) represents the shaded part area shown in \( D_1 \) and \( S_{D_2} \) represents the shaded part area shown in \( D_2 \), the same as below. According to Formula (5), at this time, the value range of \( \tau \) is \( \tau \in [x_0, +\infty) \). With the increase of \( \tau \), the area of the shaded part in the figure gradually increases. Therefore, when the limit of \( \tau \) is taken as \( \tau = x_0 \), the sum of the areas is the minimum, and the function gets the minimum value.

Case b: Functions \( f(x) \) and \( g(x) \) have an intersection, and the integral region is shown in the shaded part \( D_1 \) and \( D_2 \) in Figure 2(b). The feasible region is \( x \in (0, x_0] \). According to Formula (5), at this time, the value range of \( \tau \) is \( \tau \in (0, +\infty) \). When the value range of \( \tau \) is \( \tau \in (0, x_0] \), with the decrease of \( \tau \), the area of the shaded part in the figure gradually decreases. As it approaches 0, the shaded area is minimized, so the function has no minimum value. When the value range of \( \tau \) is \( \tau \in [x_0, +\infty) \), then the shadow area shown in \( D_2 \) remains unchanged and the shadow area shown in \( D_1 \) increases with the increase of \( \tau \). Therefore, when \( \tau \) is taken as \( \tau = x_0 \), the sum of the areas is the minimum.

Case c: Functions \( f(x) \) and \( g(x) \) have two intersections, and the integral region is shown in the shaded part \( D_1 \) and \( D_2 \) in Figure 2(c). The feasible region is \( x \in [x_0, x_1] \). According to Formula (5), at this time, the value range of \( \tau \) is \( \tau \in [x_0, +\infty) \). At this point, the original function is equivalent to solving for the sum of the shaded areas in the figure, namely \( W(\tau) = cn\left[ S_{D_1} + 2S_{D_2} \right] \). When the value range of \( \tau \) is \( \tau \in [x_0, x_1] \), the shadow area shown in \( D_1 \) and \( D_2 \) increases with the increase of \( \tau \), and the function gets the minimum value at \( \tau = x_0 \). When the value range of \( \tau \) is \( \tau \in [x_1, +\infty) \), then the shadow area shown in \( D_1 \) remains unchanged, while the shadow area shown in \( D_2 \) increases with the increase of \( \tau \). Therefore, when \( \tau \) is taken as \( \tau = x_0 \), the sum of the areas is the minimum.

Case d: Functions \( f(x) \) and \( g(x) \) have two intersections, and the integral region is shown in the shaded part \( D_1, D_2 \) and \( D_3 \) in Figure 2(d). The feasible region is \( x \in (0, x_0] \cap [x_1, +\infty) \). According to
Formula (5), at this time, the value range of \( \tau \) is \( \tau \in (0, +\infty) \). When the value range of \( \tau \) is \( \tau \in (0, x_0) \), the original function is equivalent to solving for the sum of the shaded areas in the figure, namely \( W(\tau) = cn\left[2S_{D_1} + S_{D_2}\right] \). With the increase of \( \tau \), the area of the shaded part increases gradually. So, there is no minimum value of the function. When the value range of \( \tau \) is \( \tau \in [x_0, +\infty) \), the original function is equivalent to \( W(\tau) = cn\left[2S_{D_1} + S_{D_2}\right] \). Then the shadow area shown in \( D_1 \) remains unchanged and the shadow area shown in \( D_2 \) increases with the increase of \( \tau \). Therefore, when \( \tau \) is taken as \( \tau = x_0 \), the function gets the minimum value. When the value range of \( \tau \) is \( \tau \in (x_0, +\infty) \), the original function is equivalent to \( W(\tau) = cn\left[2S_{D_1} + S_{D_2} + S_{D_3}\right] \). \( S_{D_3} \) represents the shaded part area shown in \( D_3 \). Then the shadow area shown in \( D_1 \) and \( D_2 \) remains unchanged, while the shadow area shown in \( D_3 \) increases with the increase of \( \tau \). Therefore, when \( \tau \) is taken as \( \tau = x_1 \), the function gets the minimum value.

Comment: In the field of transportation, there are a large number of passengers or vehicles waiting, and the waiting time is mostly random, such as the distribution of the passengers’ waiting time, the distribution of the waiting time of pedestrians at the intersection. Through the actual investigation or theoretical studies have shown that the waiting time is to obey these distributions such as logistic distribution, lognormal distribution, and Weibull distribution. These distributions have a common characteristic, that is their performance in the time distribution of the probability density function for the "single-peak" phenomenon. Therefore, this paper assumes that the DH time probability density function and the tolerance of passengers waiting time probability density function is unimodal functions.

2.2.3. Impact of the number of platforms on the total number of vehicles

The basic assumption proposed above is that there is only one platform. But in reality, this is more the case: there are multiple platforms in the market, and the travel demand and vehicle information data of each platform are controlled by themselves. Each platform only has the right to access its own travel demand database and dispatch its own taxis. They cannot share information with other platforms. So, if there are multiple platforms, is the total number of vehicles equal to the total number of vehicles on each platform?

In response to this problem, the research idea is to randomly divide the total travel demand data set into multiple subsets, each of which represents a database controlled by a taxi platform. For each independent platform, a vehicle intermodal network as shown in Figure 1 is established, and the minimum total number of vehicles of each platform is solved according to the method described above. The relationship between them is as follows:

Theorem 3: The total number of vehicles under the two platform market conditions is more than the total number of vehicles under the monopoly of a single platform, namely \( Q = Q_A + Q_B \). \( A \) and \( B \) represent two different taxi platforms.

Proof: We randomly divide the data set \( G = (T, E) \) into two subsets \( G_A = (T_A, E_A) \) and \( G_B = (T_B, E_B) \). The fleet sharing sets corresponding to platforms \( A \) and \( B \) are \( C_A = \{p_1^A, p_2^A, \cdots, p_n^A\} \) and \( C_B = \{p_1^B, p_2^B, \cdots, p_n^B\} \), respectively. And the independent travel demand sets corresponding to platforms \( A \) and \( B \) are \( S_A = \{T_1^A, T_2^A, \cdots, T_m^A\} \) and \( S_B = \{T_1^B, T_2^B, \cdots, T_n^B\} \). The minimum total number of vehicles of platform \( A \) is equal to \( Q_A = k_A + m_A \), and the minimum total number of vehicles on platform \( B \) is equal to \( Q_B = k_B + m_B \). So, we can get the equation \( Q_A + Q_B = k_A + m_A + k_B + m_B \).
\[ G_A \subseteq G \text{ and } G_B \subseteq G \] are independent of each other and the DH time threshold is the same in the three cases. The sum of the independent travel demand numbers in platforms \( A \) and \( B \) is equal to the independent travel demand number, namely \( m_A + m_B = m \).

We assume that \( k_A + k_B \leq k \). As can be seen from the above, 
\[ C_A \cup C_B = \{p_{iA}^1, p_{iA}^2, \ldots, p_{iA}^{k_A}, p_{iB}^1, p_{iB}^2, \ldots, p_{iB}^{k_B}\}. \]
The original set \( C \) represents the minimum total number of vehicles required to meet the vehicle intermodal transport conditions and all travel needs. It means that \( k < k_A + k_B \), which is a contradiction since \( k_A + k_B \leq k \). It can be proved that the total number of vehicles under the two platform market conditions is more than that under the monopoly of a single platform.

Corollary 1: There are \( n \) online car-hailing platforms in the market. After a new platform enters the market, only the resources of one platform will be divided up. At this time, the sum of the number of vehicles on the \( n \) platforms in the market is larger than the sum of the number of vehicles on the \((n+1)\) platforms.

According to Theorem 3, after the new platform enters the market, the number of vehicles on other platforms remains unchanged. And the number of vehicles on the two platforms is larger than the number of vehicles on the two platforms due to the addition of the new platform. Therefore, the sum of the number of vehicles on \( n \) platforms in the market is greater than the sum of the number of vehicles on \((n+1)\) platforms.

3. Results & Discussion

3.1. Data sources
To verify the practicability of the model, it is necessary to verify the model based on actual data. Therefore, this paper uses the one-week trip data of the Suzhou taxi company to verify the model. The time is from September 16th to September 22nd, 2013. Since taxis in Suzhou City still used the beckoning method in 2013, we can see how the total number of vehicles will change if the taxis adopt the online car-hailing operation model by using this data.

Part of the data is shown in Figure 3. This paper removes the original data, extracts useful fields, deletes missing data, and invalid data. We rearrange these data according to time and retains the following data: pick-up longitude, pick-up latitude, pick-up time, drop-off longitude, drop-off latitude, and drop-off time.

![Figure 3 Raw data graph](image)

According to the changing pattern of the taxi itinerary data for 24 hours a day, as shown in Figure 4. The number of passenger trips on Thursday is significantly higher than the other days. It can be seen that travel orders first show a downward trend and then show an upward trend from 22:00 to 6:00 of the next day, reaching a peak at 11:00. After that, it shows a steady trend and reaches another peak at 21:00. Based on the above rules, this paper selects travel data from 10:00-11:00 and 21:00-22:00 for analysis.
3.2. Characteristic analysis of DH time threshold

The value range of $\tau$ is set as $\{2 \leq \tau \leq 30, \tau \in \mathbb{N}\}$. We compare the morning and evening time periods of the same day and analyze the minimum total number of vehicles with the DH time threshold. The trend of change is shown in Figure 5. It can be found that the minimum total number of vehicles at the beginning decreases with the increase of $\tau$, but as $\tau$ increases, the minimum total number of vehicles tends to be stable. When $\tau=15 \text{ min}$, although the total number of minimum vehicles gradually decreases, the downward trend tends to be flat and ultimately remains unchanged. The result shows that when $\tau=15 \text{ min}$, the minimum total number of vehicles that meet the travel demands can be determined.
Figure 5 Curve of the change of the minimum number of vehicles with the DH time threshold

Figure 6 and Figure 7 respectively show the total number of trips, the total number of current vehicles and the minimum total number of vehicles corresponding to $\tau = 3 \text{ min}$, and $\tau = 15 \text{ min}$. Although the number of sunrise trips is greatly affected by random factors, the total number of vehicles required per day remains relatively stable, and the overall trend remains relatively stable. This indicates that the minimum total number of vehicles has certain robustness. And when the travel volume suddenly increases, the travel demand can be met without adding additional vehicles.

The dotted line in Figure 6 and Figure 7 represents the ratio between the total number of trips and the total number of vehicles corresponding to $\tau = 3 \text{ min}$ and $\tau = 15 \text{ min}$. It can be seen that when $\tau = 3 \text{ min}$, a vehicle can serve two trips on average. When $\tau = 15 \text{ min}$, a vehicle can serve 4 trips on average. The utilization rate of the vehicle is significantly improved.
Figure 6 Variation trend of the minimum number of vehicles from 10:00-11:00 per day per week

In combination with Figure 5, Figure 6, and Figure 7, $\tau = 15 \text{ min}$ is selected during peak hours, and the minimum total number of vehicles that can meet the passenger travel demand is about 2,500 liang. However, the total number of vehicles in Suzhou city ranges from 3500-4000 liang. The dispatching of online car-hailing platforms can effectively reduce the empty driving rate of vehicles and the total number of vehicles.
3.3. Analysis of the total number of vehicles of multiple taxi platforms

We randomly divided the data from 10:00-11:00 and 21:00-22:00 into two groups, representing two different platforms. \( \tau = 15 \text{ min} \) was selected. The average value was taken to compare the total number of vehicles of the two platforms with the total number of vehicles of the one platform in the simultaneous segment, as shown in Figure 8 and Figure 9. The total number of vehicles of the two platforms is larger than the total number of vehicles of the one platform. The minimum total number of vehicles of the two platforms increases by about 1%-2% compared with the one platform, and the overall difference is not big.

Figure 8 Comparison of the minimum total number of vehicles between multiple platforms and one platform from 10:00-11:00

Figure 9 Comparison of the minimum total number of vehicles between multiple platforms and one platform from 21:00-22:00
4. Conclusion
Under the premise of meeting travel demand, assuming that the travel information of passengers on the road and the location of vehicles are shared, this problem is transformed into a vehicle sharing problem. The problem of the minimum total number of vehicles is solved by the idea of graph theory. We establish a mathematical model of the minimum total number of vehicles based on the DH time threshold. We establish a mathematical model based on the minimum cost of empty driving time and waiting time. Based on the data of Suzhou taxi, a reasonable threshold value of DH time is obtained, and the minimum total number of vehicles required to meet the passengers’ travel demand in the peak hour is determined through the dispatching of an online car-hailing platform.

A mathematical model is established to prove that the total number of vehicles in the multi-platform market is more than that in a single-platform monopoly. Although it is difficult for the existing market to have only one taxi platform, it can help the managers to better set the price mechanism and guide healthy competition among the platforms.

By using the taxi data of Suzhou in 2013, the above three theorems are verified. The total number of vehicles decreases with the increase of the DH time threshold. There is the value of the DH time threshold which minimizes the sum of time cost of passenger and driver. The total number of vehicles on multiple platforms is larger than the total number of vehicles on single platforms. At the same time, it is found that the total number of vehicles has certain robustness.

The solution model of the minimum total number of vehicles constructed in this paper assumes that all vehicles have the same speed. It does not consider that the intersection corner, trees, and other factors block the view, so there is still a big gap with reality. For the multi-taxi platform model, only two taxi platforms are selected for data verification. When the market is shared by three or more taxi platforms, the impact of multi-taxi platforms on the minimum total number of vehicles remains to be further studied.

References
[1] Wang X, Agatz N A H, Erera A. (2015) Stable Matching for Dynamic Ride-sharing Systems. ERIM Report Series Research in Management.
[2] People's Daily Online. (2019) Online car-hailing aggravates congestion, but it is not the only real culprit. http://finance.people.com.cn/n1/2019/0522/c1004-31097215.html.
[3] Xi YF, Liu ZK, Yang PY. (2020) Travel demand forecast methods for Internet private hire vehicles. JOURNAL OF SHANGHAI UNIVERSITY (NATURAL SCIENCE EDITION), 26(03):328-341.
[4] Chen S, Wang H, Meng Q. (2019) Solving the first-mile ridesharing problem using autonomous vehicles. Computer Aided Civil and Infrastructure Engineering, 35:45-60.
[5] Tu M, Li Y, Li WX. (2019) Improving ridesplitting services using optimization procedures on a shareability network: A case study of Chengdu. TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE, 149.
[6] Bianchessi A G, Formentin S, Savarese S M. (2013) Active fleet balancing in vehicle sharing systems via Feedback Dynamic Pricing. In: 2013 16th International IEEE Conference on Intelligent Transportation Systems. Hague, Netherlands.
[7] Fanti M P, Mangini A M, Pedroncelli G. (2014) Fleet Sizing for Electric Car Sharing Systems in Discrete Event System Frameworks. In: 2014 IEEE Transactions on Systems, Man, and Cybernetics Systems. San Diego. Pp. 1-17.
[8] George D K, Xia C H. (2011) Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. European Journal of Operational Research, 211(1):198-207.
[9] Alonso-Mora J, Samaranayake S, Wallar A. (2017) On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc Natl Acad Sci U S A, 114(3):462-467.
[10] Baker, B M, Ayechew, M A. (2003) A genetic algorithm for the vehicle routing problem. World Congress on Intelligent Control & Automation IEEE, 30:787-800.
[11] Wang SF, Li L, Ma WJ. (2019) Trajectory analysis for on-demand services: A survey focusing on spatial-temporal demand and supply patterns. Transportation Research Part-C: Emerging
Technologies, 108:74-99.

[12] Ye XF, Li M, Yang ZZ. (2020) A Dynamic Adjustment Model of Cruising Taxicab Fleet Size Combined the Operating and Flied Survey Data. Sustainability, 12(7):1-8.

[13] Xu Y, Li DS. (2019) Incorporating Graph Attention and Recurrent Architectures for City-Wide Taxi Demand Prediction. ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION, 8(9):414.

[14] Feng G, Zhang DQ, Dong YC. (2019) Urban link travel speed dataset from a megacity road network. SCIENTIFIC DATA, 6:61-68.

[15] Wang G, Yang C, Zhang YJ. (2019) Addressing the Minimum Fleet of Shared Autonomous Vehicles Using Real-world Trajectory Data. CHINA TRANSPORTATION REVIEW, 41:48-53.

[16] Vazifeh M M, Santi P, Resta GL. (2018) Addressing the minimum fleet problem in on-demand urban mobility. Nature, 557(7706):534-538.

[17] Yao XR, Wang G, Yang C. (2019) Exploring Fleet Size of Shared Autonomous Vehicles in Future City: A Case Study in Shanghai. Journal of Transportation Systems Engineering and Information Technology, 19(06):85-91.

[18] VAZIFEH M M, SANTI P, RESTA G. (2018) Addressing the minimum fleet problem in on-demand urban mobility. Nature, 557: 534-538.

[19] J.Munkres. (1957) Algorithms for the Assignment and Transportation Problems. Journal of the Society for Industrial and Applied Mathematics, 5(1):32–38.

[20] Zhang SF. (2019) Research on Signal Control Method Based on Passenger Demand in Connected Environment. Hefei University of Technology.

[21] Ceder, A., Marguier, P. H. Passenger waiting time at transit stops. Traffic Engineering & Control, 26(6), 327–329.

[22] Turnquist M A. (1978) A model for investigating the effects of service frequency and reliability on bus passenger waiting times. Transportation Research Record Journal of the Transportation Research Board, 663(663):70-73.

[23] Larry, A, Bowman. (1981) Service frequency, schedule reliability and passenger wait times at transit stops. Transportation Research Part A General.

[24] Currie G, Csikos D R. (2007) The Impacts of Transit Reliability on Wait Time: Insights from Automated Fare Collection System Data[C]// Transportation Research Board Meeting.

[25] Lüthi, Marco, Weidmann U, Nash A. (2007) Passenger arrival rates at public transport stations. Transportation Research Board Meeting.

[26] Liu XM, Li ML, Wang HP. (2020) Research on Passengers’ Willingness to Wait at Bus Stations Based on SP Survey. Traffic Engineering, 20(04):26-31.

[27] Ahmad T, Mahmoud M, Ameneh S. (2018) Modelling passenger waiting time using large-scale automatic fare collection data: An Australian case study. Transportation Research Part F Traffic Psychology and Behaviour, 58:500-510.

[28] Zhang ZY, Hao XY, Wang D. (2015) Endurance Time of Pedestrians Crossing at Signalized Intersection of Beijing. Journal of Transportation Systems Engineering and Information Technology, 15(06):212-219.