

# THE ANALYSIS OF URBAN TAXI CARPOOLING IMPACT FROM TAXI GPS DATA

Qiang XIAO<sup>1</sup>, Ruichun HE<sup>2</sup>, Changxi MA<sup>3</sup>

<sup>1, 2, 3</sup> School of Traffic and Transportation, Lanzhou Jiao tong University, Lanzhou, China

**Contact:**

1) lzjt\_xq@126.com, 2) herc@mail.lzjtu.cn, 3) mcx9815@sina.com

---

**Abstract:**

Taxi is an important part of urban passenger transportation system. The research and analysis of taxi trip behavior is the key to meet the demand of urban passenger transport and solve the traffic congestion problem. Based on the GPS data of taxis in Nanjing, the statistical method is used to analyze the taxi characteristics of the average number of passengers, the average passenger time, the no-load distance and the passenger distance. By using the double logarithmic coordinate, the trip distance and trip time of taxi passengers are analyzed, it is found that the average trip distance of taxi passengers is mainly concentrated in 3-20km, and the average trip time of taxi passengers is mainly concentrated in 10-30 minutes. Using the information entropy theory to construct the equilibrium model of taxi passenger-carrying point, and analyze the spatial distribution of taxi, it is found that the distribution of urban taxi is unbalanced. The peak clustering algorithm is used to determine the location of passenger gathering points, and the hot spot of taxi trip is analyzed, it is found that the hot spots of taxi trip are mainly concentrated in the central city of Nanjing. Combined with the results of urban taxi trip analysis, from the perspective of taxi and passenger, we found that the number of urban taxis, the passenger carrying rate of taxis, the duration period of passenger trip, the duration and distance of passenger trip and the location of passenger trip points will have an impact on the urban taxi carpooling in Nanjing. By using the probability model of urban taxi carpooling, this paper discusses and analyzes the influence of these factors on urban taxi carpooling. The research in this paper can provide a reference for the effective implementation of urban taxi carpooling policy.

**Key words:**

taxi carpooling, taxi passenger-carrying points, equilibrium degree, trip characteristic

---

**To cite this article:**

Xiao, Q., He, R., Ma, C., 2018. The analysis of urban taxi carpooling impact from taxi GPS data. Archives of Transport, 47(3), 109-120. DOI: <https://doi.org/10.5604/01.3001.0012.6514>



## 1. Introduction

With the acceleration of urbanization in China and the rapid increase of vehicles, the problem of urban road traffic congestion has become a major problem that needs to be solved in urban development (Xiao et al., 2017). It is a common measure to develop urban public transportation and advocate green trip for all major cities. The taxi plays a significant role in transportation systems in the urban areas across the world, providing flexible and demand-responsive transport service with widely spatiotemporal coverage. However, the development of taxi market is seriously restricted by the asymmetry of information between supply and demand, the difficulty of passengers to take a taxi, the high rate of taxi empty load and the large cost of rental operation. In view of this, some scholars put forward the idea of taxi carpooling in order to solve the above problems, and quickly attracted the attention of a large number of scholars. Empirical data show that taxi carpooling has policy effectiveness in alleviating urban traffic congestion and resolving taxi operation difficulties.

The widespread adoptions of location-based services (LBS) provide unprecedented opportunities to study various mobility patterns from trillions of trails and footprints (Liang et al., 2012 and Quddus et al., 2015). A careful analysis of these digital footprints from taxi GPS localizers can provide an innovative strategy to facilitating urban public transit planning and operational decision-making, and thus have attracted considerable attentions from researchers.

Liu et al. (2012) introduced a new method to explore intra-urban human mobility and land use variations based on taxi trajectory data from Shanghai city. Castro et al. (2013) proposed an overview of mechanisms for using taxi GPS data to analyze people's movements and activities, which includes three main categories: social dynamics, traffic dynamics and operational dynamics. Yoon et al. (2007) proposed a simple yet effective method using the spatial and temporal speed information in order to estimate traffic status on surface streets with GPS location data. Zhang et al. (2013) proposed a weighted approach to estimate traffic state using GPS data by increasing the weights of recent velocity information. Li et al. (2014) presented a hybrid learning framework to appropriately combine estimation results of freeway traffic density state

from multiple macroscopic traffic flow models. Liu et al. (2015) investigate subregion information and the interaction patterns of center-local places, and reveal trip patterns and city structures, which could potentially aid in developing and applying urban transportation policies.

However, despite the variety of researches and applications, taxi GPS data has so far not been explored for urban taxi carpooling analysis (He, 2014 and Huang, 2015 and Kim, 2015). Therefore, this paper makes use of the taxi GPS data to calculate the daily average number of times of taxi passengers, the taxi no-load distance, the taxi passenger carrying distance and the duration of the taxi passengers carrying, and analyzes the characteristics of the urban taxi passenger trip. By using the spatial statistical method, the models of taxi passenger point equilibrium and passenger trip point density are established, and the spatial location and time distribution of the upper and lower passenger points are discussed, and the reasons that affect the taxi ride in urban area are analyzed from the perspective of taxi and passengers.

Our on-going work is focused on the analysis of taxi-GPS traces acquired in the city of Nanjing, China; to better understand urban taxi mobility. The contribution of this work lies on the following aspects: we analyze the spatial distribution characteristics of taxis, examine the statistical distributions of trip distances and trip time, research hot spot area of taxi trip, and explore the influencing factors of Urban Taxi carpooling from spatial area perspective.

## 2. Analysis of taxi passenger-carrying point

### 2.1. Taxi passenger characteristics

The characteristics of taxi passenger carrying reflect the operation characteristics of taxi (Chen, 2015 and Toth, 2015 and Stephane, 2014), including the average number of times of passengers carrying per taxi, the average duration of passengers carrying per taxi, the average no-load distance per taxi and the average distance of passengers carrying per taxi. From the analysis of these characteristics, we can understand the passenger-carrying state, the duration and the daily trip time of the taxi, which provides the basis for the analysis of the influencing factors of the urban taxi carpooling.

1) The average number of times of passengers carrying per taxi

The average number of times of passengers carrying per taxi refers to the number of times the taxi carries passengers in unit time period, which reflects the degree of taxi utilization. Definition such as formula (1).

$$P_{\text{number}} = \frac{1}{n} \sum_{i=1}^n P_{\text{scary}i} \quad (1)$$

In formula (1),  $P_{\text{number}}$  is the average number of times of passengers carrying per taxi, and  $P_{\text{scary}i}$  is the number of times of passengers carried by the  $i$ th taxi per unit time and  $n$  means the number of taxis.

2) The average duration of passengers carrying per taxi

The average duration of passengers carrying per taxi is the ratio of the total taxi passenger carrying duration to the total number of times of taxi passengers carrying in unit time, which reflects the length of each taxi service. Definition such as Formula (2).

$$P_{\text{time}} = \frac{\sum_{i=1}^n P_{\text{stimer}i}}{\sum_{i=1}^n P_{\text{scary}i}} \quad (2)$$

In formula (2),  $P_{\text{time}}$  is the average duration of passengers carrying per taxi ;  $P_{\text{stimer}i}$  is the passengers carrying duration of the  $i$ th taxi in the unit time.

3) The average no-load distance per taxi

The average no-load distance per taxi is the ratio of the total mileage of no-load taxi to the numbers of no-load taxis within unit time. Definition such as Formula (3).

$$\bar{K}_{\text{empty}} = \frac{S_{\text{Nload}}}{N_{\text{Nload}}} \quad (3)$$

In formula (3),  $\bar{K}_{\text{empty}}$  is average no-load distance per taxi.

4) The average distance of passengers carrying per taxi

The average distance of passengers carrying per taxi is the ratio of the total mileage of taxi passengers carrying to the number of taxi at specified intervals, Definition such as Formula (4):

$$\bar{K}_{\text{load}} = \frac{S_{\text{sum}} - S_{\text{Nload}}}{N_{\text{sum}} - N_{\text{Nload}}} \quad (4)$$

## 2.2. Equilibrium degree model of the taxi passenger-carrying points

The concept of information entropy is introduced in this paper to explain the equilibrium degree model of the passenger points. Information entropy is a measure of uncertainty, that is, a high uncertainty of a variable corresponds to large information entropy and an even distribution of passenger-carrying points. Conversely, a low uncertainty of a variable corresponds to low information entropy and an uneven distribution of passenger-carrying points. The equilibrium degree model is established according to information entropy theory (Liu et al. 2015; Chen et al. 2013; Milev et al. 2012).

The entropy of the passenger-carrying points can be expressed according to the administrative division of the city, whose number of built districts is  $n$ , the variation  $t_i$  is the number of passenger-carrying points randomly falling into the  $i$ th district and the probability is  $P(t_i), i=1, 2, \dots, n$ :

$$H = \sum_{i=1}^n P(t_i) \log_2 P(t_i) \quad (5)$$

If all passenger-carrying points are found in the same district, their entropy value is 1, the entropy value of other districts is 0 and information entropy is the minimum. If all passenger-carrying points are evenly distributed into all of the districts, that is, the probability of all the areas is equal  $P(t_1)=P(t_2)=\dots=P(t_n)=1/n+1$ , the information entropy is the maximum ( $H_{\text{max}}$ ). The information entropy is normalised.  $H$  denotes the information entropy and  $H_{\text{max}}$  denotes the maximum information entropy expressed as follows:

$$G = H / H_{\text{max}} \quad (6)$$

$G$  is the equilibrium degree of the passenger-carrying points. Low information entropy indicates a small  $G$  and an uneven distribution of passenger-carrying points. Conversely, high information entropy indicates a large  $G$  and an even distribution of passenger points.  $H < H_{max}$ ; thus,  $0 \leq G \leq 1$ . If  $G < 0.55$ , the distribution is very even; if  $0.55 < G < 0.6$ , the distribution is uneven; if  $G > 0.6$ , the distribution is even.

**2.3. Density model of the passenger-carrying points**

The cluster centre of the passenger-carrying points' data set is determined by using the mountain clustering algorithm. The data field is divided into small grids, and the data density of each grid is estimated. The grid with a high data density denotes the peak. The point with the largest data density is the mountain top, namely, the cluster centre (Pu et al., 2007; Alexandrov et al., 2007). The detailed data are shown as follows:

- (1) Construct the grid space network according to the density distribution of the grids, the intersection set of the grid is  $V = \{v_1, v_2, \dots, v_n\}$ .
- (2) Construct the Gaussian mountain peak function of the density index,  $v_j \in V$ , the height of the mountain peak function is as follows:

$$m(v_j) = \sum_{i=1}^n \exp(-\|v_j - x_i\|^2 / 2\delta^2) \tag{7}$$

Where  $n$  is the number of the data sample,  $v_j$  is the  $j$ th intersection to be calculated;  $x_i$  is the  $i$ th intersection, the influential factor  $\delta$  is the clustering coefficient.

- (3) Select the maximum point of the mountain peak function value from the set  $V$  as a cluster centre, that is,

$$m(c_i) = \max(m(v_j)) \tag{8}$$

- (4) Set the new mountain peak function by subtracting the Gaussian function at point  $m(c_i)$  from the  $m(v_j)$  in the cluster centre; then, select the point from the set  $v$  with new mountain peak function as the second cluster centre.

$$m_{max}(v_j) = m(v_j) - m(c_i) \exp(-\|v_j - x_i\|^2 / 2\beta^2) \tag{9}$$

- (5) Set the number of the cluster centre according to the sample space; repeat Step (4) until all the cluster centres are found. Assume the position of the carpooler is  $P(p_x, p_y)$ ; then, work out the difference with each cluster centre; the  $P(p_x, p_y)$  is the GPS coordinates of the carpooler; the  $V(v_x, v_y)$  is the coordinates of the cluster centre. If the position of the carpooler is located in the area with high density, a closer location to the mountaintop or a closer position to the core area corresponds to a high success rate and convenient carpooling.

**3. Analyses and results**

This paper utilises the taxis' GPS data in Nanjing on 16 September 2014, a total of 7726 taxis and 18668073 records; 453019 records of 7726 taxis exist after the treatment and the calculation of the GPS trajectory data as follows:

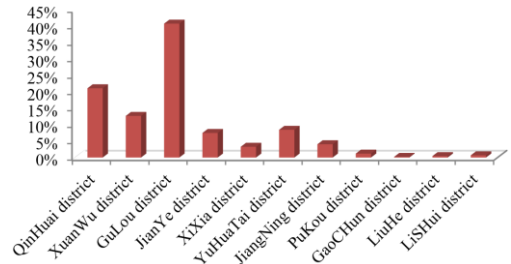


Fig. 1. The taxi passenger-carrying points distribution ratio in Nanjing City area passenger carrying points

According to the GPS trajectory data of 6503 taxis in Nanjing on 16 September 2014, the proportional distribution map of the passenger-carrying points is worked out in each administrative area according to the GPS data of each administrative district in Nanjing; from Fig.1, we can see that the QinHuai district, XuanWu district and GuLou district are the aggregation of the passenger-carrying points, which shows that the taxis in Nanjing mainly operate at the city centre. From figure.2, we can see that the number of the taxis mainly distributes from 7:00 am to 11:00 pm, the number of the taxi is less from 0:00 am to 6:00 am. The peak values of the number of times carrying the passenger per hour are mainly distributed in these following periods: 0:00–1:00, 10:00–12:00, 17:00–19:00 and 20:00–22:00, which satisfy the standards of urban city life.

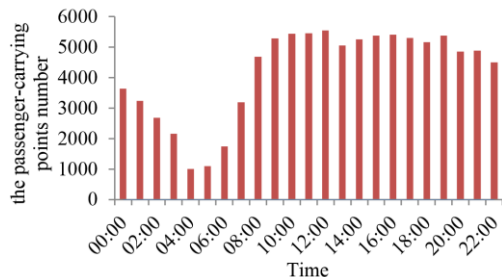


Fig. 2. Taxi passengers carrying points number per hour in Nanjing City on September 16, 2014

It can be seen from Fig.3 that the no-load rate of a single day taxi in Nanjing is the highest in the period of 3: 00-6: 00, close to 70%, while in other periods, the rate of no load basically fluctuates around 40%.Through the analysis of taxi no-load rate, we can understand the operation condition of urban taxi, the trip time distribution of the passenger and the hot spot area of the city, and help the urban taxi carpooling researchers to understand the urban taxi passenger situation more accurately.

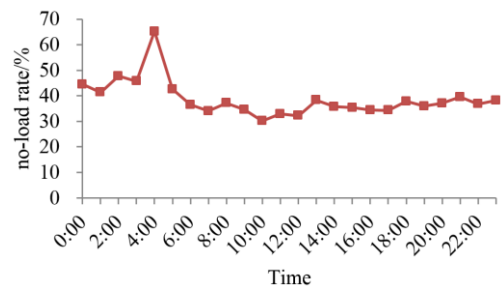


Fig. 3. The no-load rate of taxi in Nanjing City on September 16, 2014

### 3.1. Taxi passenger characteristics in Nanjing

1) The average number of times of passengers carrying per taxi

As can be seen from Fig.4, the average number of passengers carrying by a taxi in Nanjing at 7: 00-17: 00 is 2-2.5 per hour, while at 3: 00-6: 00, the average number of passengers is less than 1 per hour. Through the analysis of the number of taxi passengers carrying, it can be understood that passenger trip is mainly concentrated in the normal work time, which provides the basis for the carpooling implementation of the time period.

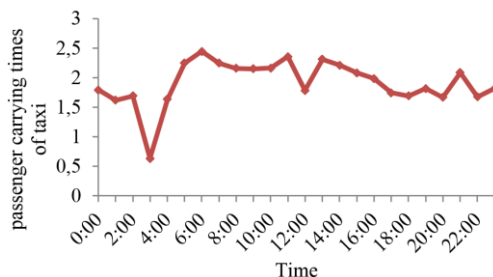


Fig. 4. The average number of times of passengers carrying per taxi in Nanjing on September 16, 2014

2) The average duration of passengers carrying per taxi

As can be seen from Fig.5, the passenger hours of taxis at each time period of a single day are curve distribution, at 7:00-9:00 and 18:00-20:00, the time of carrying passengers is longer, and the other periods are relatively short. Through the analysis of the average passenger time, we can find out the peak time of passenger trip and passenger trip duration, which provides a reference for the arrangement of carpooling service.

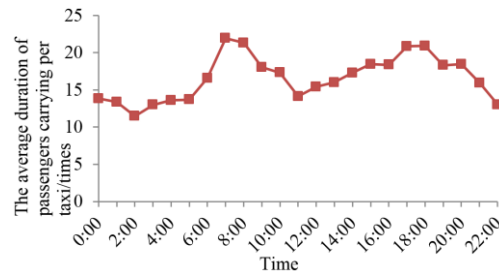


Fig. 5. The average duration of passengers carrying per taxi in Nanjing on September 16, 2014

3) The average no-load distance per taxi

As can be seen from Fig.6, the total mileage between taxis passengers carrying at all times of the day is curvilinear, ranging from 3: 00-6: 00. The average no-load distance per taxi is between 15km and 30km. The average no-load distance in other periods is relatively short. By analyzing the average no-load distance of each taxi, we can find out the driving distance of the taxi in each period, and provide a reference for the carpooling service.

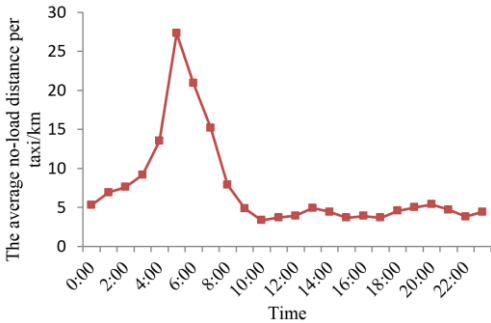


Fig. 6. The average no-load distance per taxi in Nanjing on September 16, 2014

4) The average distance of passengers carrying per taxi

As can be seen from Fig.7, the taxi passenger distance is a curve distributed at each time of the day. The average passenger distance per taxi is between 4km and 8km. It shows that taxis mainly carry passengers in short and medium distance. The analysis of the average distance of each taxi can provide a reference for the study of carpooling distance.

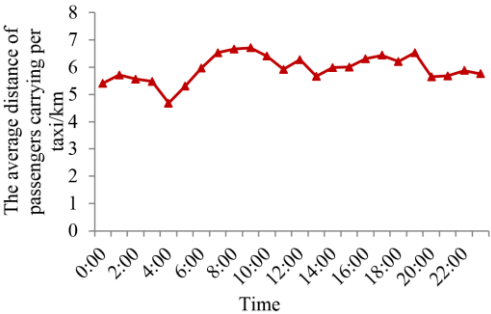


Fig. 7. The average distance of passengers carrying per taxi

5) Analysis of trip distance and duration of taxi passengers

This paper tracks the data of taxi passengers, finds out the location and time of passengers getting on and off, and makes statistics on the trip distance and trip duration of passengers in each time period. Using the double logarithmic coordinate system, the frequency of trip distance and trip duration is plotted, and the relationship between trip distance and trip duration of urban taxi passengers is analyzed.

As can be seen from Fig.8 and Table 1, on September 16, 2014, the frequency maximum value of single trip distance of taxi passengers in Nanjing is mainly in the distance between 3~10km and 10~20km, and the frequency of single trip distance distribution accounts for nearly 70% of all single trip distance distribution frequency. The frequency minimum value of single trip distance is mainly distributed in the distance above 30km. According to the statistics of the frequency of trip distance of taxi passengers on this day, it can be found that urban taxi passenger mainly trip at short and medium distances, and take taxis for a long distance less, which provides a basis for the study of carpooling distance.

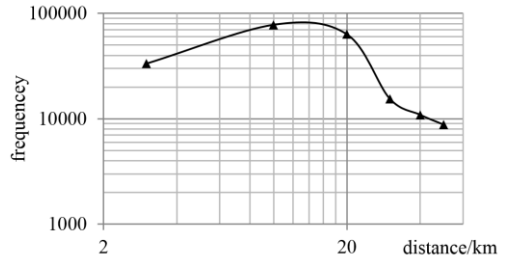


Fig. 8. Distribution of trip distance and frequency of passengers in Nanjing on September 16, 2014

Table 1. Accumulative schedule of taxi passenger trip distance in Nanjing

Average trip distance of taxi passengers [km]	frequency	Percentage [%]	Cumulative percentage [%]
0~3	33320	15.88	15.88
3~10	77906	37.13	53.01
10~20	63393	30.21	83.22
20~30	15519	7.40	90.61
30~40	10911	5.20	95.81
>40	8785	4.19	100.00

AS can be see from Fig. 9 and Table 2, on September 16, 2014, the single trip duration of taxi passengers in Nanjing is mainly in two periods: 10-20min and 20-30min, and the frequency of single trip duration distribution accounts for about 70% of the total duration distribution frequency. According

to the statistics of the trip duration of taxi passengers on this day, it is found that the passengers trip in 30min are the main service objects of taxi, which provides the basis for the study of trip duration of carpooling passengers.

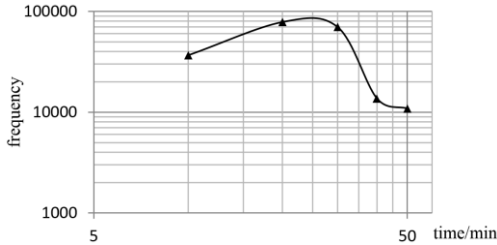


Fig. 9. Distribution of trip duration and frequency of taxi passengers in Nanjing on September 16, 2014

Table 2. Cumulative trip duration of taxi passengers in Nanjing

Average trip duration of taxi passenger [min]	frequency	Percentage [%]	Cumulative percentage [%]
0-10	36764	17.52	17.52
10-20	78735	37.52	55.04
20-30	69806	33.27	88.31
30-40	13666	6.51	94.82
>40	10863	5.18	100.00

### 3.2. The equilibrium degree of the taxi passenger points in Nanjing

Nanjing is located in the east of China, as the capital of Jiangsu Province; Nanjing consists of 11 districts, has an area of 6636.31 km<sup>2</sup> and approxi-

mately 800 million people. To analyse the equilibrium degree of the passenger-carrying points in Nanjing, the number of the passenger-carrying points of each district is calculated according to the latitude and longitude range; the average number of the passenger-carrying points of each district is calculated; meanwhile, the equilibrium degree of the passenger-carrying points is calculated according to the equilibrium degree expression, which is shown in Table 3.

In Table 3, the acreage of the LiuHe district is the largest, and the acreage of the QinHuai district is the smallest. The number of the passenger-carrying points and the average number of the passenger-carrying points in GuLou district, QinHuai district and XuanWu district are higher than the other districts. From Table 3, we know that the distribution of the passenger-carrying points in Nanjing is uneven, and the passenger-carrying points are mainly concentrated in GuLou district, QinHuai district and XuanWu district, of which the passenger density is higher than other districts. Thus, GuLou district, QinHuai district and XuanWu district are the central districts; meanwhile, the finance, culture and advertisement centre of Nanjing are the important population areas.

### 3.3. Density of the passenger-carrying points in Nanjing

To analyse the density of the passenger-carrying points in Nanjing, this paper calculates the distribution characteristics of the passenger-carrying points in each period by using the mountain clustering algorithm.

Table 3. The equilibrium degree of the passenger points in Nanjing

NO	district	Area [km <sup>2</sup> ]	passengers number	Number [km <sup>2</sup> ]	Probability	Equilibrium degree
1	QinHuai district	49.11	95350	1799.06	0.2828	0.5399
2	XuanWu district	75.46	57301	707.68	0.1106	
3	GuLou district	54.18	184362	3356.92	0.4957	
4	JianYe district	81.75	33780	417.71	0.0602	
5	XiXia district	395.38	14776	37.84	0.0054	
6	YuHuaTai district	132.39	38146	283.40	0.0420	
7	JiangNing district	1563.33	18235	11.59	0.0017	
8	PuKou district	910.51	5364	5.88	0.0009	
9	GaoCHun district	790.23	840	1.05	0.0002	
10	LiuHe district	1471	1912	1.29	0.0002	
11	LiSHui district	1063.68	2953	2.77	0.0004	
total		6587.02	453019	6865.09	-	-

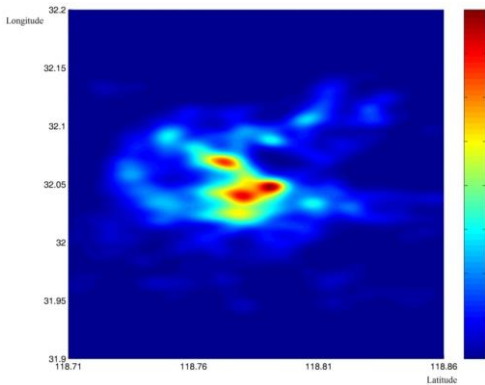


Fig. 10(a). The dense graph of the taxi passenger points between 0:00-2:59

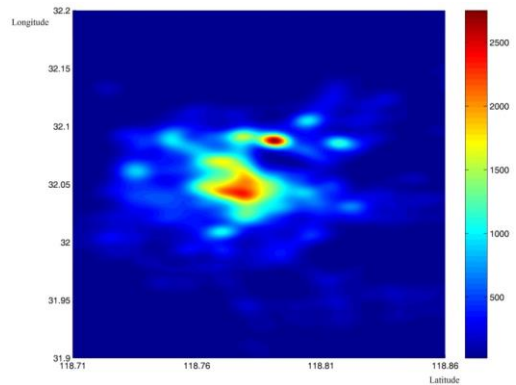


Fig. 10(d). The dense graph of the taxi passenger points between 9:00-11:59

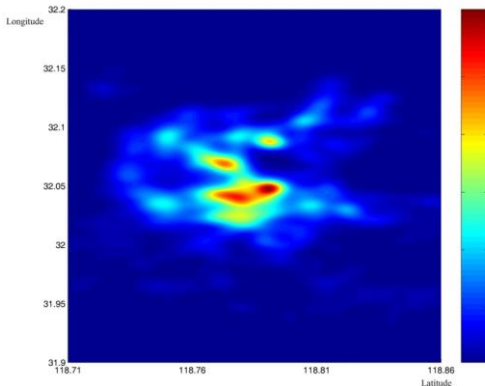


Fig. 10(b). The dense graph of the taxi passenger points between 3:00-5:59

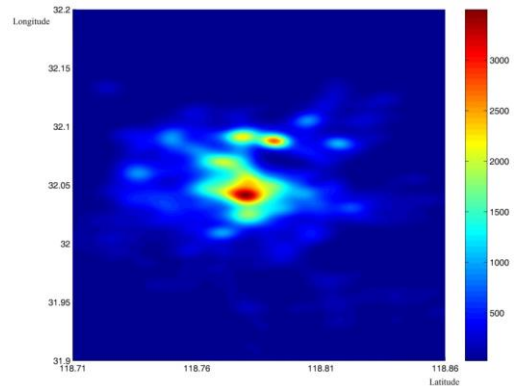


Fig. 10(e). The dense graph of the taxi passenger points between 12:00-14:59

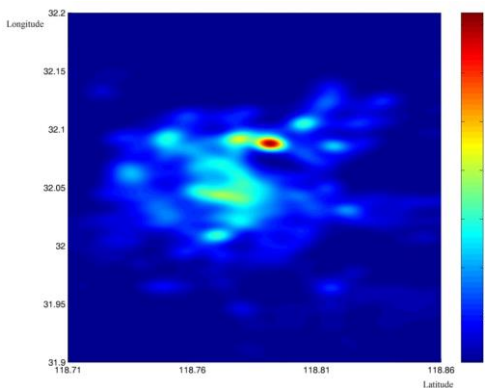


Fig. 10(c). The dense graph of the taxi passenger points between 6:00-8:59

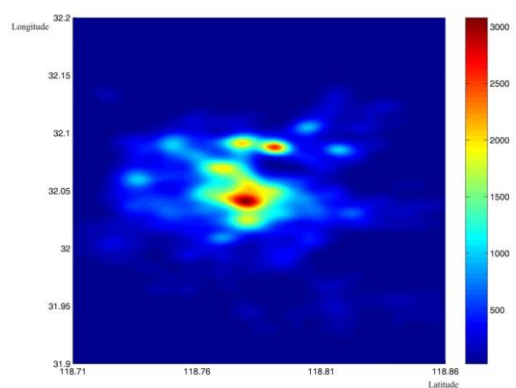


Fig. 10(f). The dense graph of the taxi passenger points between 15:00-17:59



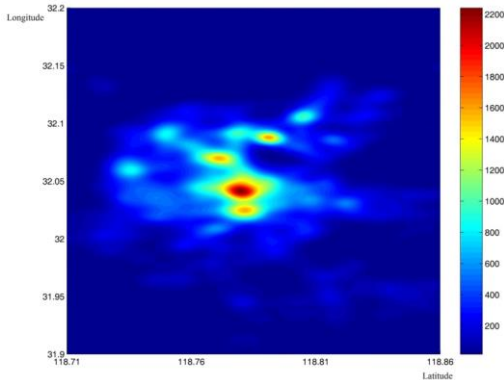


Fig. 10(g). The dense graph of the taxi passenger points between 18:00-20:59

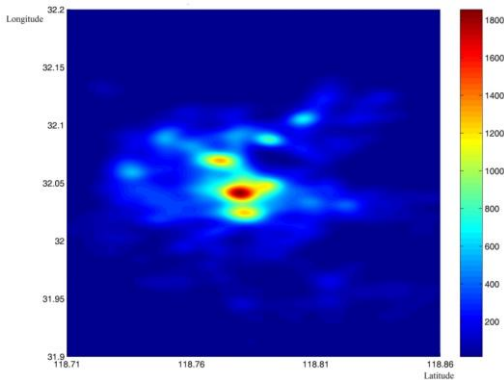


Fig. 10(h). The dense graph of the taxi passenger points between 21:00-23:59

In figure.10 (a), (b), the changes in the density of the passenger-carrying points are described with different colours. The areas in bright colour show that the density of the passenger-carrying points is high; the areas in dark colour indicate that the density of the passenger-carrying points is low. From 0:00 am to 6:00 am, the passenger-carrying points are mainly distributed in GuLou district, QinHuai district and XuanWu district, which are business districts, advertisement districts, dining districts, education districts and aggregation of the stations. Compared with the period from 0:00 am to 3:00 am and the period from 3:00 to 6:00, one more dense point exists in figure.3(a) compared with dense points in figure.10 (b), which denotes the Nanjing railway station; suggesting that in the period from

3:00 am to 6:00 am, many passengers are present in Nanjing Railway Station.

In figure.10 (c), the densest area of the passenger-carrying points from 6:00 am to 9:00 am is the Nanjing Railway Station in GuLou district, where the passenger flow volume is the largest. The passenger-carrying points in XuanWu district and in QinHuai district are low because of the function of these districts. Employees in government sectors, public institutions and markets frequently leave for work after 9:00 am; therefore, few passengers trip within this period.

In figure.10 (d)-(f), the density of the passengers in the Nanjing Railway Station increases. Thus, the railway station has a high incidence of taxi passengers all the time. The density range of the passenger-carrying points in XuanWu district and QinHuai district decreases and these points moves to the JianYe district and YuHuaTai district. Hence, the trip of people from 9:00 to 15:00 pm is distributed in business areas and education areas. From 15:00 to 18:00, most passengers trip for entertainment and mainly move to the YuHuaTai district and the Confucius Temple in QinHuai district.

In figure.10 (g), (h), the density of the passenger-carrying points from 18:00 to 24:00 am begins to move to dining and entertaining areas, such as XuanWu district, QinHuai district and GuLou district. From 21:00 to 24:00, the density of the passenger-carrying points in Nanjing Railway Station decreases with time, and this observation is consistent with regular passenger flow in the station. A later time corresponds to few trains. Trains mainly pass by, and the number of passengers in the railway station decreases. Thus, passengers in the passenger point sharply decrease.

#### 4. Analysis on the influence of Urban Taxi carpooling

According to the statistical data of GPS analysis in Nanjing city taxi, taxi passenger carrying points equilibrium degree calculation and taxi passenger density modeling and analysis, and according to the constraints of the supply and demand of urban taxi and model of urban taxi carpooling probability( Xiao et al., 2018), this model is shown in formula (10).

$$P_y = (1 - e^{-\frac{M_i VT(1-O)}{L}}) (1 - P_i)^{n-1} P_i P_c \quad (10)$$

The urban taxi carpool probability is represented by  $P_y$ ,  $L$  represents the road length,  $t_0$  represents the number of taxis,  $V$  represents the average speed of the taxis,  $T$  represents the waiting time of taxi passengers,  $O$  represents no-load rate,  $P_c$  represents the probability that passengers and taxis passenger can carpooling together,  $t_0$  represent the taxi distribution rate.

Based on the carpooling probability model of urban taxi and the statistical analysis from the perspective of urban taxi GPS data, to explore the influence of urban taxi carpooling.

#### 1) Numbers of taxis

From formula (10), when the other variables remain unchanged, the  $t_0$  gets bigger, the  $e^{-\frac{M_iVT(1-O)}{L}}$  gets the smaller, the probability of  $P_y$  will increase. On the contrary, the  $t_0$  gets smaller, the  $e^{-\frac{M_iVT(1-O)}{L}}$  gets the bigger, the probability of  $P_y$  will decrease.

According to figure.1, the number of taxis in GuLou district, QinHuai district, XuanWu district, YuHua-Tai district and JianYe district is more than that of other administrative areas. Through the analysis of formula(10), the probability of passengers to taxi carpooling in these areas is more probability, and in other administrative areas, the probability of taxi number is relatively less, and the probability of passenger to taxi carpooling in the area is relatively small.

#### 2) Taxi passenger carrying rate

From formula (10), when the other variables remain unchanged, the  $O$  gets smaller, the taxi passenger carrying rate will increase, the  $e^{-\frac{M_iVT(1-O)}{L}}$  gets the smaller, the probability of  $P_y$  will increase. On the contrary, the  $O$  gets bigger, the taxi passenger carrying rate will decrease, the  $e^{-\frac{M_iVT(1-O)}{L}}$  gets the bigger, the probability of  $P_y$  will decrease.

From figure.3, it can be seen that the taxi in Nanjing city has a higher no-load rate at 3: 00-6: 00, but less in other periods, so the probability of taxi carpooling at 3:00-6:00 is smaller, but the probability of taxi carpooling at other periods is higher.

#### 3) Taxi distribution rate

From formula (10), when the other variables remain unchanged, the taxi distribution rate  $t_0$  gets bigger, the  $(1-P_i)^{n-1}P_i$  will increase, the probability of  $P_y$  will increase. On the contrary, the taxi distribution rate  $t_0$  gets smaller, the  $(1-P_i)^{n-1}P_i$  will decrease, the probability of  $P_y$  will decrease.

From table 3, it can be seen that the distribution rate of taxis in Nanjing is extremely uneven. GuLou district, QinHuai district, XuanWu district, YuHuaTai district and JianYe district have a high distribution rate. According to the analysis of formula (10), the probability of passengers to taxi carpooling in these areas is higher than the other areas.

#### 4) Numbers of taxi passengers

From formula (10), when the other variables remain unchanged, as the number of passengers increases, the  $P_c$  gets bigger, the probability of  $P_y$  will increase. On the contrary, as the number of passengers decreases, the  $P_c$  gets smaller, the probability of  $P_y$  will decrease.

As can be seen from figure 10, the density of passengers in Nanjing urban area varies with different time periods. At 0:00-6:00, in commercial areas such as GuLou district, QinHuai district and XuanWu district, the probability of passengers carpooling is higher. at 6:00-9:00, and near the railway station in XuanWu district, the probability of passengers carpooling is higher. At 9:00-23:00, in the GuLou district, QinHuai district, XuanWu district and YuHuaTai district, the probability of passengers carpooling is obviously higher than that of other districts.

#### 5) Trip duration and distance of taxi passengers

As can be seen from figure 8-9 and table 1-2, the distance that urban passengers in Nanjing trip each time using a taxi is mainly within the range of 3-20 kilometers, and the trip duration is concentrated within the range of 10min-30min, so we can infer that, If a taxi in Nanjing wants to carry on a carpooling, the object of choice is mainly concentrated in the Medium and short distance trip, and the possibility of long-distance carpooling trip is relatively small.

#### 6) Trip points location of taxi passenger

As can be seen from table 3 and figure 10, the distribution of taxi trip points in Nanjing is very uneven. For urban passengers, the location of the trip determines whether or not to take a taxi at the right place. If the trip point is located in the no-central urban area or the non-trunk road located in the urban, the probability of taxi carpooling is smaller, and the probability of choosing the carpooling trip mode will be reduced. On the contrary, if the passengers trip point is located on the trunk road or central in the urban, the probability of taxi passing is greater, and the probability of choosing the carpooling trip mode will increase. Therefore, the location of passenger trip point has a greater impact on the passenger carpooling.

### 5. Conclusion

From the perspective of taxi GPS data, this paper analyzes the factors influencing urban taxi carpooling. By means of statistical method, equilibrium model and density model, this paper analyzes the trip characteristics of taxis and passengers, and finds out that the number of urban taxis, the passenger carrying rate of taxis, the duration period of passenger trip, The duration and distance of passenger trip and the location of passenger trip points will have an impact on the urban taxi carpooling in Nanjing. The results in this paper can be used as reference for the implementation of taxi carpooling policy in Nanjing.

### Acknowledgements

This paper is supported by the National Natural Science Foundation of China (Grant no. 61364026 and 51408288) and The humanities and social sciences fund of the Ministry of Education (Grant no.15YJCZH107).

### Reference:

- [1] ALEXANDROV, G. A. & YAMAGATA, Y. , 2007. A peaked function for modelling temperature dependence of plant productivity. *Ecological Modelling*. 200(1-2), 189-192.
- [2] CASTRO, P. S. , & ZHANG, D. & CHEN, C. & LI, S. , 2013. From taxi GPS traces to social and community dynamics: a survey. *ACM Comput Surv*. 46 (2), (article 17).
- [3] CHEN X. W. , 2018. A Spatial and Temporal Analysis of the Socioeconomic Factors Associated with Breast Cancer in Illinois Using Geographically Weighted Generalized Linear Regression. *Journal of Geovisualization and Spatial Analysis*. 2 (1), 1-16.
- [4] CHEN, Y. M. & WU, K. S. & LI, X. J. , 2013. A kind of outlier mining algorithm based on information entropy. *Control and Decision*. 28 (6), 867-872.
- [5] HE, W. & LI, H. W. , 2014. Intelligent carpool routing for urban ridesharing by mining GPS trajectories, *IEEE Transactions on Intelligent Transportation Systems*. 15(5), 2286-2296.
- [6] HUANG, S. C. & JIAU, M. K. & LIN C. H. . , 2015. A genetic-algorithm-based approach to solve carpool service problems in cloud computing. *IEEE Transactions on Intelligent Transportation Systems*. 16 (1), 352-364.
- [7] KIM, J. W. & MAHMASSANI, H. S. , 2015. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Part C: Emerging Technologies*. 59, 375-390.
- [8] LIANG, X. & ZHENG X. & LU, W. , 2012. The scaling of human mobility by taxis is exponential. *Phys. A*. 391, 2135-2144.
- [9] LIU, Y. & KANG, C. & GAO, S. , 2012. Understanding characteristics of intraurban trips using taxi trajectory data. *J. Geogr. Syst*. 14 (4), 463-483.
- [10] LIU, Y. & WANG, F. & XIAO, Y. & GAO, S., 2012. Urban land uses and traffic source-sink areas: evidence from GPS-enabled taxi data in Shanghai. *Landsc Urban Plan*. 106 (1), 73-87.
- [11] LI, L. & CHEN, X. & ZHANG, L. , 2014. Multimodel ensemble for freeway traffic state estimations, *IEEE Trans. Intell. Transp. Syst*. 15 (3), 1323-1336.
- [12] LIU, X. & GONG, L. & GONG, Y. X. & LIU, Y. , 2015. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*. 43, 78-90.
- [13] LIU, X. J. & LI, Q. L. & LI, Y. J, 2015. Learning with information entropy method for transportation image retrieval. *International Journal of Multimedia and Ubiquitous Engineering*. 10 (7), 317-328.

- [14] MA, Z. L. & SHAO, C. F. & HU, D. W. , 2012. Temporal-spatial analysis model of traffic accident frequency on expressway. *Journal of Traffic and Transportation Engineering*. 12, 93-99.
- [15] MILEV, M. & INVERARDI, P. N. & TAGLIANI, A. , 2012. Moment information and entropy evaluation for probability densities. *Applied Mathematics and Computation*. 218(9), 5782-5795.
- [16] PU, H. Z. & ZHEN, Z. Y. & WANG, D. B. , 2007. Improved jumping gene genetic algorithm for multi-peak function optimization. *Nanjing Hangkong Hangtian Daxue Xuebao/Journal of Nanjing University of Aeronautics and Astronautics*. 39 (6), 829-832.
- [17] QUDDUS, M. & WASHINGTON S. , 2015. Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. *Transportation Research Part C: Emerging Technologies*. 55, 328-339.
- [18] STÉPHANE, G. A. & LUK, K. B. & YASAR, A. U. H. , 2014. Multi-agent simulation of individual mobility behaviour in carpooling. *Transportation Research Part C*. 45, 83-98.
- [19] TOTH, C. , 2015. Carpooling in Hungary: Can it reduce the GHG emissions of personal transport?. *European Transport - Trasporti Europe*. 58 (4), 1-24.
- [20] XIAO, Q. & HE, R. C. , 2017. Carpooling scheme selection for taxi carpooling passengers: a multi-objective model and optimisation algorithm. *Archives of Transport* . 42(2), 85-92.
- [21] XIAO Q, HE, R. C. , YU J. N. , 2018. Method of taxi carpooling probability and wait time based on Poisson distribution. *China Journal of Highway and Transport*. 31 (5), 151-159.
- [22] YOON, J. & NOBLE, B. & LIU, M. , 2007. Surface street traffic estimation. *Proceedings of the 5th International ACM Conference on Mobile Systems: Applications and Services*. New York, USA, 220-232.
- [23] ZHANG, J. D. & XU, J. & LIAO, S. S, 2013. Aggregating and sampling methods for processing GPS data streams for traffic state estimation, *IEEE Trans. Intell. Transp. Syst.* 14 (4), 1629-1641.