Research on road traffic flow status based on survival analysis

Beibei Hu, Doudou Lin, Qibo Sun, Xianlei Dong*

School of Business, Shandong Normal University, Jinan, China * Corresponding author tel: +86 13280029639
E-mail:xizixinxiang@163.com, 1193535083@qq.com, 1003667231@qq.com, *sddongxianlei@163.com

Abstract: Based on the data of bayonet and floating car, this paper studies the traffic flow state of the main roads in Shenzhen, analyses the relationship between speed and flow, and establishes a speed-flow model. It studies the characteristics of traffic flow in Shenzhen and the overall distribution of congestion duration. The results demonstrate that there is a strong correlation between flow and speed, which have obvious characteristics of morning and evening peak within a day, and the flow and speed show significant differences between work days and non-work days. As for the congestion duration of each road, they are generally under 16 minutes. Specifically, when the congestion duration is less than 8 minutes, there is a high probability that the congestion state ends; while the congestion lasts more than 10 minutes, the congestion status is less likely to end in a short time. In view of the above analysis, it is recommended that relevant departments strengthen monitoring and control of real-time road traffic so that the traffic environment in Shenzhen can be improved.

1. Introduction
Along with the advancement of urbanization, traffic problems have become one of the major problems of all cities around the world. By the end of 2017, China’s motor vehicle ownership reached 310 million. The rapid growth of vehicle ownership has improved people’s travel efficiency and happiness as well as causes the traffic congestion and environmental pollution [1], which increases people's travel costs [2], and blocks the coordinated operation of urban functions. Therefore, real-time monitoring of road traffic flow and research on the distribution of road congestion are of great significance for traffic management and urban function realization.

The urban transportation systems are complex and non-linear. In addition to the traffic flow and road traffic capacity, the traffic conditions of urban roads are also affected by pedestrian crossings, traffic lights and vehicle lane changes, etc. With the development of network information technology, the intelligent transportation system makes real-time monitoring of road traffic status come true, and it is possible to obtain real-time road condition information and predict road status in real time [3]. In recent years, domestic and foreign scholars have done a lot of valuable research on road traffic flow, which mainly includes the theoretical description of road traffic flow state and model construction, as well as prediction of congestion status. For example, based on the traffic flow measurement data, they divide the free flow, steady flow and crowded flow, then establish flow-speed model, and analyze the relationship between speed, flow and density [4]; They employ vehicle speed and flow data to build based on the time-space flow-speed model and traffic flow prediction model, and study the relationship between highway traffic flow parameters [5]; they use radial basis function neural network method to collect traffic characteristics such as pedestrian characteristics, traffic flow, traffic...
facilities, traffic situation, traffic sociality, etc. Based on the urban congestion theory model of traffic 5S elements, they evaluate urban traffic congestion [6]; on the basis of the overall change law of urban road traffic congestion index, they establish a K-nearest urban road traffic congestion index prediction model to achieve short and medium term prediction of traffic congestion index [7]; a system dynamics model of urban road congestion is established to seek solutions to traffic congestion problems from the perspective of residents travel and urban logistics [8].

In summary, the previous researches on the state of road traffic flow focus more on the construction of traffic flow model and traffic congestion prediction model, while there is a small number of studies which use traffic flow big data to analyze the traffic flow state and the overall distribution of congestion law in specific urban roads. In this paper, taking specific roads as an example, we use urban road data to construct a speed-flow model. and explore the distribution law of each road congestion duration. Then we provide theoretical and methodological support for traffic management and planning.

2. Material and Methods

2.1. Data Source and Pretreatment
This research obtains the bayonet data and floating car data of 4 roads in Futian District, Shenzhen from March 25 to March 31, 2018. The bayonet data come from the four monitoring points of the Xinzhou interchange on Beihuan Road, Fulong Road’Henglongshan Entrance, Xiangmihu Road’Municipal Party School, Jingtian Road Jingmi Village Park. Floating car data are mainly collected from GPS trajectory data of vehicles such as taxis, dump truck, heavy-duty semi-trailers and coach car, etc. After the invalid data and the duplicate data are deleted, we get a total of 1,528,672 bayonet data and 1,834,340 floating car data. The floating car data includes six fields, which include the timestamp corresponding to the trajectory, the license plate number after desensitization, the longitude coordinates, the latitude coordinates, the instantaneous speed and satellite speed.

Next, it is about the further pretreatment of the raw data. Firstly, the timestamp field in the floating car trajectory data is converted. And then the Java program is employed to call the Aliyun Map to match the latitude and longitude of the floating car data, for obtaining the road segment where the vehicle is located. Finally, the Beihuan Road, Fulong Road, Xiangmihu Road and Jingtian Road are screened out from floating car data as the main research objects for empirical analysis.

2.2. Methodology
Based on the traffic measured data, we use the survival analysis method to study the distribution law of road congestion duration. The specific measurement steps are as follows:

(1) Computing survival time. The survival time of road congestion refers to the duration of the congestion. The floating car data are divided into two minutes as a time unit per day. According to the congestion status threshold of different types of roads, we make a judgment on the congestion status of each road in each time unit. The time unit of continuous congestion state is recorded as a congestion sample. And the sample numbers under different survival time are recorded.

(2) Computing survival function. $S(t)$ denotes the survival function, which indicates the probability that the duration of the road congestion state is longer than $t$. As $S(t)$ increases, the possibility of the road continuing to congested is greater. In this paper, we use the non-parametric method of kaplan-meier to estimate the survival function of a congestion, as shown in Equation (1):

$$S(t) = P(T > t) = \prod_{i \in S_t} \frac{n_i - d_i}{n_i}$$

(1)

Here, $T$ is the actual congestion duration of the road, $n$ denotes the number of time samples for all congestion, $d_i$ refers to the number of samples that end congestion in $t_i$. $n_i$ is the number of samples that are still in a congested state at $t_i (t_1 < t_2 < \cdots < t_k, i = 1, 2, 3 \cdots k)$. 

(3) Computing risk function. The risk function refers to the probability that the congestion state will end instantaneously within a short time $\Delta t$ after the time $t$ has elapsed. The larger the function value is, the greater possibility the congestion ends, and conversely, the congestion continues. The risk function $h(t)$ is shown in Equation (2):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

In which, $T$ denotes the road congestion duration, and $P$ is the probability that the congestion sample will end the congestion state instantly within a short time $\Delta t$ after the time $t$ has elapsed.

3. Results
The traffic flow status of roads is mainly affected by three parameters: flow, speed and density. This paper uses Greenhills’ speed-flow model \[ Q = VK \] to analyze the relationship between speed and flow, wherein $Q$ denotes the flow, $V$ denotes the speed, and $K$ refers to the density. Greenhills’ theory believes that the vehicle speed is higher when the flow of the road is not saturated, the mutual interference among vehicles is slight, and the traffic flow is in free flow. While the road is in the peak period, the traffic flow increases sharply, the mutual interference increases, and the vehicle speed decreases gradually. Finally, the free flow state will be destroyed. Even in severe cases, the vehicle speed will drop to zero.

3.1. the Research on Speed and Flow Characteristics
The daily traffic flow on work days and non-work days on the four roads, such as Beihuan Road, Fulong Road, Xiangmihu Road and Jingtian Road, is shown in Figure 1. It is noticeable that on the work days and non-work days, the daily traffic flow distribution of four roads is almost same, which demonstrates “U” curve during the period from 0:00 to 8:00, and the minimum value of traffic flow appears at around 4:00. Besides, the morning peak is mainly at 7:00-9:00, the evening peak at 17:00-19:00. Compared with the work days, the peak period on the non-work days is delayed one hour. For example, the morning peak of the work days on Beihuan Road appears at 7:00, while the non-work days appears at 8:00.

![Figure 1](image_url)

Figure 1 the Daily Traffic Flow of Each Road on Work day and Non-work day.

In Figure 1, the largest and least traffic volume is Beihuan Road and Jingtian Road respectively. The Beihuan Road, as a city expressway, undertakes the functionality of the city's large capacity, long distance and fast traffic, and the traffic is smooth and the traffic volume is relatively high.
Correspondingly, the Jingtian Road is a subordinate road, connecting with main roads and branches, and the flow is relatively low. In addition, the amount of traffic flow at the morning and evening peak of the four roads between work days and non-work days is different. For example, in the morning and evening peak on Beihuan Road, the traffic flow on non-work days is 7.1% higher than that on work days. The reasons are that the number of tourists on non-work days increases, and the limited vehicle on the work days results in relatively high traffic on non-work days.

Next, the speed of the four roads in one week is shown in Figure 2. On a whole, the speed curves of the four roads all show a "W" shape. The period of low speed is mainly in the morning and evening peaks(7:00-9:00,17:00-19:00), while the high speed is mainly from 0:00 to 6:00.

As shown in Figure 2, due to the decreasing commuting vehicles during non-work days, the speed of each road on non-work days is higher than work days, especially in the peak period on Fulong Road, Xiangmihu Road and Jingtian Road. Generally, the better the traffic conditions, the higher speed is. On Beihuan Road, the difference of speed between work days and non-work days is the smallest, and speed curves on work days and non-work days are roughly consistent.

3.2. the Speed-Flow Model

Next, we fit the traffic flow and vehicle speed data of four roads and construct the speed-flow model. The formula is shown in Equation (3):

\[ V = aQ^2 + bQ + c \]  

(3)

The fitting effect of each road is shown in Table 1.

| Road Name         | the Speed-Flow Expression | R-squared |
|-------------------|---------------------------|-----------|
| Beihuan Road      | \( -0.249Q^2 - 3.458Q + 33.42 \) | 0.4816    |
| Fulong Road       | \( -5.702Q^2 - 12.25Q + 50.21 \) | 0.4377    |
| Xiangmihu Road    | \( 1.427Q^2 - 4.24Q + 31.63 \) | 0.4262    |
| Jingtian Road     | \( -0.5829Q^2 - 5.41Q + 33.72 \) | 0.5447    |

The fitting curves of each road are shown in Figure 3.
Figure 3 the Relationship Between Speed and Flow of Each Road.

According to Figure 3, as the flow increases, the speed shows a declining trend, which indicates a negative correlation, except the speed-flow curve of Fulong Road on the early stage, which is positively correlated (as shown in Fig3(b)). On the early stage of Fulong Road, as traffic flow increases, the speed rising slowly, and there is a short period of a flat peak which indicates that the mutual interference between vehicle speed and flow is weak. The tendency of Beihuan Road, Jingtian Road and the state of the remaining part of Fulong Road are roughly same, and illustrate that at the primary step the traffic flow increases, the speed of the vehicle decreases steadily; when the flow continues to increase, the vehicle speed begins to drop sharply. By contrast, the speed-flow fitting curve of Xiangmihu Road is in a downward convex state, which illustrates as traffic flow increase gradually, the speed of the vehicle always decreases slowly and reaches a stable state. At this moment, the correlation between speed and flow is obscure.

3.3. the Distribution of Road Congestion Duration

According to the construction standard of Shenzhen Road Traffic Operation Index, congestion status is defined that the expressway speed (such as Beihuan Road, Fulong Road and Xiangmihu Road) is less than 31.5km/h and the subordinate road speed (such as Jingtian Road) is less than 25km/h. In this paper, we take two minutes as the shortest time for congestion persistence, and analyze the changes in the congestion status of the four roads in a week, and count the number of samples under different congestion durations, as shown in Table 2.

Table 2 the Number and Proportion of Samples for Congestion Duration.

| Road Name       | the Number of Samples With Congestion Duration | Sample Ratio | the Number of Samples With Congestion Duration | Ratio |
|-----------------|-----------------------------------------------|--------------|-----------------------------------------------|-------|
| Beihuan Road    | 302                                           | 87.5%        | 43                                            | 12.5% |
| Fulong Road     | 149                                           | 77.6%        | 43                                            | 22.3% |
| Xiangmihu Road  | 312                                           | 100%         | 0                                             | 0     |
| Jingtian Road   | 142                                           | 97.9%        | 2                                             | 2.1%  |

It can be seen from Table 2 that the congestion duration is mostly under 16 minutes. To make the research more universal, only the samples of congestion duration less than 16 minutes are discussed, thus the survival function and risk function of the congestion are estimated, as shown in Figure 4. Overall, when the congestion occurs, as the duration of congestion increases, the survival curve shows a downward trend. Among them, the survival curve has a large decline in the interval [0,8], and then shows a gentle trend, which indicates that the probability of ending the congestion state is relatively
large. The overall trend of the risk curve firstly rises and then falls. The specific performance is that when the survival time is within two minutes, the risk rate of congestion increases sharply, and the probability of the end of the congestion state is greater. Two minutes later, the risk curves of four roads show a downward trend, and at the point of 10 minutes gradually tend to be gentle, indicating that the probability of ending congestion is gradually decreasing.

![Figure 4 Survival Curve and Risk Curve of Each Road.](image)

For the survival curve, the congestion survival rate of Fulong Road is obviously higher than the other three roads. Specifically, the survival rate of Fulong Road is the highest, and the survival rate of Xiangmihu Road is the lowest, indicating that Fulong Road has the highest probability of continuing to remain congested. Namely, Fulong Road has more serious congestion state. However, there is no significant difference in the congestion status between the Beihuan Road and Jingtian Road. For the risk curve, the risk curve of Xiangmihu Road fluctuates greatly, and the risk of congestion remains at a high state. It shows that the congestion state of Xiangmihu Road has a higher probability to end under different congestion time, and the road has better ability to dissipate congestion. In addition, the rate of risk on Beihuan Road, Fulong Road and Jingtian Road fluctuates significantly, and the difference among the three roads is unobvious, and the probability of the end of the congestion state is low.

4. Conclusions
Taking Shenzhen as an example, this paper uses the bayonet passing data and floating car data of four main roads in Futian District to establish a speed-flow model, and analyzes the state of traffic flow in Shenzhen, and uses survival analysis method to study the distribution of congestion duration. The conclusions are as follows:

- The traffic flow and speed of the four main roads in Futian District of Shenzhen City have obvious morning and evening peak characteristics, whose extent between work days and non-work days is different. Firstly, the flow of each road maintain at a high level in the morning peak, and then the downward trend occurs till the end of the evening peak, and the total daily traffic flow on non-work days is higher than work days. Then, the speed has a “W” fluctuation in one day, and the morning and evening peak characteristics are very prominent, which in non-work days is obviously higher than the work days. Next, the speed-flow model shows that there is a strong correlation between speed and flow, while due to different road type and function, as the flow increases, there is a different change in speed, and the condition of flow switch between different states.

- For the congestion phenomenon, under different congestion survival times, the probability of four roads continue to maintain and end the congestion state is different. The survival curve indicates that
when the congestion time increases, the probability of the road maintaining the congestion state is small. Fulong Road has the highest survival rate, which means the possibility of continuing to maintain congestion is the highest, and the ability to relieve congestion is poor. While the risk rate curve is hump-like, and the risk rate fluctuates significantly under different congestion time. When the congestion time is long, the probability of ending the congestion state is low.

The study of the traffic flow state is of great significance for road construction and traffic control. In order to improve the traffic situation in Shenzhen, it is recommended that relevant departments strengthen the monitoring of road conditions and enhance road reconciliation guidance and actively adjust the restrictive policies to control traffic flow during the morning and evening peak periods. It is also advised that they should improve urban intelligent transportation systems and strengthen the collection of real-time data to provide more accurate travel recommendations for residents, such as choosing the right time to go out, and from the perspective of prevention and control to alleviate road pressure and improve urban traffic, and so on.

Acknowledgments
This study was supported by the programs of National Social Science Foundation of China (Grant: 16CJY056).

References
[1] Huang W, Fan H, Qiu Y, et al. 2016 Causation mechanism analysis for haze pollution related to vehicle emission in Guangzhou, China by employing the fault tree approach. J. Chemosphere, 151:9-16.
[2] Yang Z W, Liu X M. 2012 The Effect of Urban Travel Cost on Travel Structure. J. Journal of Transportation Systems Engineering & Information Technology, 12(2):21-26.
[3] Chmiele W, Dziech A, Ernst S, et al. 2016 INSIGMA: an intelligent transportation system for urban mobility enhancement. J. Multimedia Tools & Applications, 75(17):10529-10560.
[4] Xu C, Chen X M, Coldt T T. 2014 Analysis of Characteristics of Traffic Flow Speed-density Relation Model. J. Journal of Highway & Transportation Research & Development, 31(2):114-120.
[5] Lv Y, Duan Y, Kang W, et al. 2015 Traffic Flow Prediction With Big Data: A Deep Learning Approach. J. IEEE Transactions on Intelligent Transportation Systems, 16(2):865-873.
[6] Xiong L, Yang S F, Zhang Y, et al. 2018 Research on the Urban Traffic Congestion Evaluation Model Based on 5S Theory. J. Operations Research & Management Science, (1):117-124.
[7] Wei Q B, He Z C, Zheng X S, et al. 2017 Prediction of Urban Traffic Performance Index Considering Multiple Factors. J. Journal of Transportation Systems Engineering & Information Technology, 17(1):74-81.
[8] Yang H X, Li J D, Zhang H, et al. 2014 Research on the governance of urban traffic jam based on system dynamics. J. Systems Engineering-Theory & Practice, 34(8):2135-2143.
[9] Chang Y L, Wang X T, Zhang P. 2016 Cosine Function Speed-flow Relationship Model for Road Segment Baseectiond on Upstream Arrival Traffic Flow. J. Science Technology & Engineering, 16(11):257-261.