

## **Estimation of highway section traffic flow based on toll data**

Jun-Feng Zhang, Yun-Peng Wang and Hai-Yang Yu

*Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety  
Control, Beihang University, Beijing, China  
E-mail: bothwellzhang@163.com*

Traffic parameters acquisition for highway section mainly depends on manual counts or traffic parameters detection equipment. The toll data is not fully used. Based on the toll data of Anhui province highway considering the differences in sections and vehicle types, this exploratory study uses the Dijkstra algorithm to calculate the shortest path of the network between any two points along with programming in MATLAB, a method to estimate section traffic flow which is proposed.

*Keywords:* Toll Data; Dijkstra Algorithm; Section Flow.

### **1. Introduction**

At the end of 2015, the total mileage of China is 4.57 million kilometers, and road density 47.6 kilometers per hundred square kilometers, the mileage of country's highway is 125,400 km [1]. While highway is developing at an unprecedented speed and scale, the traffic flow is getting increasingly complex, the impact between traffic flows of different sections is also increasing. In this context, only scientific and effective management conducted by highway managers can we make full economic and social benefits of the highway. To reach this goal, the economical access to accurate and comprehensive highway sectional flow is required, which allowing managers to depth and comprehensive understanding of the operation of the highway, so that the scientific and reasonable control strategy can be made. As a part of the traffic indicators, sectional traffic flow is the visual representation of the actual operating conditions of the highway, as well as important reference for traffic management, construction planning and economic analysis. Thus, it has important significance to obtain sectional traffic flow quickly and accurately in the initial and latter part of the road evaluation.

Our technical and theoretical research on collecting traffic flow data have achieved some results. [2] Actual operating conditions of the highway affected by different sections of the actual road conditions and vehicle types. The

sectional flow mainly obtained by two ways: First, the manual survey. [3] The sectional flow is obtained by manual survey within specified time; the second is to set traffic flow detecting device in the ground or on the ground to get the sectional flow [4]. Both methods currently have some maneuverability and practicality, as well as shortcomings. Manual survey not only costs a lot of manpower and financial resources, low efficiency, but also susceptible to interference from man-made and environmental factors, resulting in a lower accuracy of the parameters obtained. Flow detecting equipment requires expensive equipment. Besides, this method can only be used in limited locations, not the whole road network. Furthermore, these two methods can not reflect the composition of traffic on the highway. Therefore, it is very important to propose an economic, convenient sectional traffic flow estimation method which covers the entire road network.

Highway toll data bring together a lot of valuable traffic information that reflects the state of the vehicle running on the highway, from which we can understand the overall state of the road network, patterns and characteristics fundamentally. Complete and accurate toll data lay a foundation for the highway sectional flow estimation and analysis. Fu Xin [5] from Chang'an University proposes a method using OD Matrix from toll traffic records to estimate sectional traffic flow in the case of simple line, Zhang Hongbin [6] uses OD matrix and traffic assignment matrix from toll data to calculate sectional flow. Wang Shengji [7] proposes the use of the shortest path of road network to solve the multipath problem. In this paper, considering the differences of road section and vehicle types, based on analyzing the vehicle running information obtained from the highway toll data, combined with MATLAB programming, a way to estimate the sectional flow is proposed.

## **2. Data Preprocessing**

The toll data includes unique identifier, vehicle entrance and exit time, vehicle entrance and exit toll information, entrance and exit time, site code, the exit lane number, vehicle types and other information. But not all the data is useful to estimation process.

The toll data contains abnormal data due to system or human causes. The causes include vehicle condition problems, system errors, half-way stop, human factors etc. In order to avoid the impact of abnormal data on the highway temporal and spatial distribution characteristics, the threshold method and quartile method are used to process toll data [8], after the original data become

reliable and accurate data which is suitable for processing, data extraction step is followed.

After data preprocessing, only useful field will be extracted from data. The extraction field includes unique identifier, entry time, entrance station ID, exit time, exit station ID and vehicle type. In this paper, Anhui charging data in July 2012 as the basis, the preprocessed data sample is as follows in table 1:

Tab. 1 The preprocessed data sample

Unique ID	Entrance Time	Entrance Code	Exit Time	Exit Code	Vehicle Type
568A003B	2012/7/20 0:00	1105	2012/7/20 3:39	701	1
1994C683	2012/7/20 0:00	1105	2012/7/20 2:54	706	1
0B232CFF	2012/7/20 0:00	10301	2012/7/20 2:33	10201	1

Combining the network topology, vehicle access information reflects the physical location and adjacency relationship of vehicles. As a virtual node of traffic generation amount, highway toll stations information can be used to calculate OD matrix of a certain period of time, which is the basis for sectional flow estimation.

### 3. Vehicle Information Processing Model

Highway in a certain area forms road network structure with a certain size and shape, and ultimately forms a entire highway network. Highway toll data includes traffic information for the entire road network. Based on sectional flow estimation method, considering vehicle types, traffic flow of any section can be calculated. For a vehicle with same start and end points in the network can have different driving route, in order to calculate the sectional flow of specific section, the actual path of the vehicle must be determined.

#### 3.1. Network division

Based on the characteristics of highway network topology, this paper carries out two levels of highway road network division. Interchange terminal are nodes where traffic flow changes greatly, combining highway network topology and traffic flow characteristics, select the interchange terminal in highway network as the first level of division control node, the road network is divided into large sections, which is the most advanced nodes in the network. Since toll stations

are nodes where traffic flow changes, and traffic flow in the section between two adjacent toll stations changes little. Therefore, select the adjacent toll station as the second level of division control node, then large sections of first level will be divided into sub-sections. In result, highway network consists of  $m$  sections, namely Section 1, Section 2... Section  $m$ . Each section consists of  $n$  sub-sections, namely Sub-section 1, Sub-section 2... Sub-section  $n$ . Sub-section is the smallest unit for subsequent processing. The whole process is as follows:

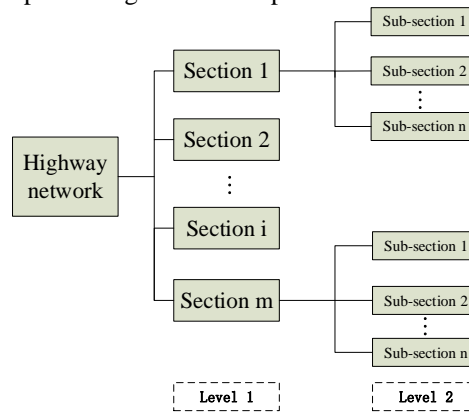


Fig. 1 Highway network division process

### 3.2. Vehicle information model

The Dutch mathematician Dijkstra proposed the shortest path algorithm<sup>[9]</sup> in 1959 for solving the shortest path problem between specified point and the other points in non-negative weights network. Based on the shortest path algorithm, the multi-path traffic flow in road network will be split in accordance with the traffic flow so that we can have the actual traveling path of vehicle.

#### 3.2.1. Shortest Path Algorithm

The basic principle of the shortest path algorithm is if the sequence  $\{v_1, v_2, \dots, v_j, \dots, v_n\}$  is the shortest path from  $v_1$  to  $v_j$ , then the sequence  $\{v_1, v_2, \dots, v_j\}$  will definitely be the shortest path from  $v_1$  to  $v_j$ . The algorithm process is as follows:

- (1) Number  $v_1$  with P, the remaining points with T, and  $T(v_i) = +\infty$  ;
- (2) If  $v_j$  is the point of the just got point P, considering the reference point

of  $T$ , namely  $v_j, (v_i, v_j) \in E$ , the right of  $(v_i, v_j)$  is  $d_{ij}$ . Make the following changes to the label of  $v_j$ :

$$T(v_j) = \min \{T(v_i), P(v_i) + d_{ij}\}. \quad (1)$$

(3) Compare all label points of  $T$ , change the smallest to the label  $P$ , namely:

$$T(v_{j0}) = \min \{T(v_j)\}. \quad (2)$$

When there are two or more smallest ones, all of them can be changed to label  $P$ . If all the points are label  $P$  then stop, otherwise replace  $v_j$  with  $v_{j0}$ , return (2). The label  $P$  of each point is the shortest length between point  $v_1$  and reference point.

### 3.2.2. The shortest path calculation

Based on the principle of Dijkstra algorithm, this article calculate the shortest path between any two nodes of 193 network nodes in Anhui highway network in the MATLAB environment. First, number the nodes in Excel, create an adjacency matrix with 193 nodes to show the shortest length of any two connected nodes. Use the dist function from MATLAB to solve the shortest path length between any two points. The output is in the form of a matrix.

Use vlookup function statement to match the shortest path that corresponding to the matrix to toll data by identifying the station number, then the shortest running path between any two sites can be obtained, which is the actually traveling route of vehicle. As shown in Table 2:

Tab. 2 The Shortest Path Table

In station	Out station	The shortest path
Shushan(101)	Baohedadao(103)	Shushan(101)-Jinzhaolu(102)-Xiaoxichong(18)-Baohedadao(103)
Shushan(101)	Yandian(401)	Shushan(101)-Jinzhaolu(102)-Xiaoxichong(18)-Fangxingdadao(409)-Yandian(401)
Longtang(104)	Dashu(106)	Longtang(104)-Feidong(105)-Longxi(19)-Dashu(106)
Longtang(104)	Yongzhen(206)	Longtang(104)-Feidong(105)-Longxi(19)-Wangtie(201)-Zhegao(202)-Chaohu(203)-Hanshan(204)-Wuhubei(205)-Yongzhen(206)

After getting the shortest path, based on the vehicle entrance and exit station

information, it can quickly get the actual traveling route of vehicle in the highway network, this will help us to estimate the flow of the specific sections in a certain period of time.

#### 4. Sectional Flow Estimation Model

##### 4.1. Sectional flow estimation model

When calculating the traffic section, the location of section need to be determined firstly. Due to the short length between two toll stations, there is little difference in parameters obtained at different positions in the same road. In order to facilitate the reunification process, so select the exact middle of the road as the sectional position for every road. For specific road, it is simplified as shown in Figure 2:

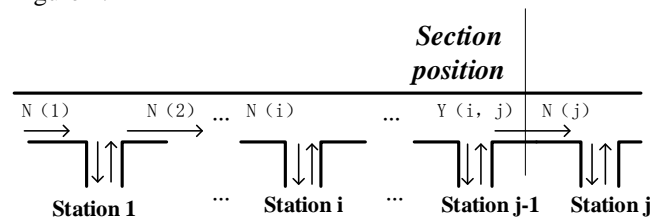


Fig. 2 The simplified schematic of highway section

Then the number of vehicles which run into highway from station  $i$  and remain on the road just before station  $j$  can be expressed as  $Y(i, j)$  :

$$Y(i, j) = Y(i, j-1) - X(i, j-1) . \quad (3)$$

Among them,  $Y(i, j-1)$  means the number of vehicles which run into highway from station  $i$  and remain on the road just before station  $j-1$ ,  $X(i, j)$  means the number of vehicles which run into highway from station  $i$  and get out of highway from station  $j$ .

Then sectional flow just before station  $j$  is  $N(j)$  :

$$N(j) = \sum Y(i, j) \quad (4)$$

##### 4.2. Vehicle type conversion

There are two kinds of vehicles on the highway namely cars and trucks. For different type of vehicles, which have very different dynamic characteristics, it will have impact on the speed of vehicle and traffic capacity. Generally, the highway has relatively stable vehicle type composition in a period of time [10].

Current vehicle information is stored according to vehicle type. Therefore, to obtain more accurate sectional flow, Vehicle type conversion is required.

According to the toll data statistics, vehicles that in the Anhui Province highway network contain a total of nine kinds of types. Wherein: 1,2,3,4, respectively, corresponding to first to fourth kind of cars in highway vehicles; 11, 12,13,14,15, respectively, corresponding to first to fifth kind of trucks in highway vehicles. Combined with "Highway Engineering Technical Standards"[11] and "Highway Charging Standards", the vehicle type conversion coefficient is as the following table [12].

Tab. 3 The Vehicle Type Conversion Coefficient Table

First Tier Category	Second Tier Category		
	Category	Vehicle code	Conversion coefficient
car	Car 1	1	1
	Car 2	2	1
	Car 3	3	1.5
	Car 4	4	1.5
truck	Truck 1	11	1
	Truck 2	12	1.5
	Truck 3	13	2
	Truck 4	14	2
	Truck 5	15	3

Based on OD information of each vehicle, through all the vehicle path, count all vehicles that pass through the selected section in a period of time. After the vehicle type conversion, the number of standard vehicles is the sectional flow of the section.

### 4.3. Example

Based on the method in this study, we use the toll data of Anhui Province highway in July 2012 to calculate sectional flow. Select Baohedadao-Longtang, Dashu-Quanjiao, Fangxingdadao-Yandian and other 106 roads in total as examples. Given vehicles on the road are quite few in too late or too early time, select 6: 00-22: 00 time period to process the data. Use MATLAB programming to process toll data. Export flow data with one hour interval, after vehicle type conversion, Parts of the sectional flow are as follows:

Tab. 4 The Sectional Flow Table

Time	Baohedadao-Longtang	Dashu-Quanjiao	Fangxngdadao-Yandian	Feidong-Longxi	Guanji-Shuangdun
6-7	529	306	511	712	249
7-8	666	451	757	954	379
8-9	856	615	934	1211	495
9-10	1128	805	1006	1473	602
10-11	1034	898	925	1388	577
11-12	935	856	735	1124	477
12-13	850	691	713	1178	474
13-14	1082	848	913	1404	443
14-15	1105	858	913	1462	463
15-16	1139	814	915	1390	529
16-17	1126	810	1056	1544	497
17-18	1061	770	1024	1354	513
18-19	918	732	725	1122	416
19-20	869	576	647	988	408
20-21	706	604	508	812	441
21-22	637	547	465	722	361

*Note:* The section locates in the exact middle of each selected road.

Due to differences in the actual operating condition of each road, the range of flow values exhibit some degree of difference. Since the distance between adjacent toll stations is short, combined with road traffic characteristic consistency between adjacent toll stations, thus, the sectional flow does not change no matter where the section is in the same road. Therefore, the sectional flow obtained by the method in this paper has high accuracy.

Compared with the traditional way to obtain sectional flow, the method in this paper has simple process, high accuracy and high processing efficiency. The features of toll data itself make this method also has the characteristics of wide coverage. Based on this method, the sectional flow of any section in any road can be calculated. Besides, by modifying parameters, this method has the ability to obtain sectional flow per hour, sectional flow per day and even sectional flow per week, as well as other time dimensional sectional flow to meet different users' needs.

## 5. Conclusion

The method herein may save the cost of traffic detection equipment on the one



hand, on the other hand it can make full use of highway toll data. In addition, by developing corresponding sectional flow calculation software, which can be integrated into the highway toll system, the data can be generated timely and quickly, enhances the timeliness of the data. For the above characteristics of this method, this method is applicable to all highways with toll stations. It can provide a more reliable method to obtain sectional flow, which has great significance for the application of highway management, planning and so on.

### Acknowledgments

This paper is one of the initial achievement of the sub-topics “Technology and systems development based on analysis of multi-scale road network running posture large data,” which belongs to the national science and technology support program “Intelligent state highway network monitoring and security services to protect key technologies and system integration.”

### References

1. Department of Transportation. 2015 Transportation industry statistical bulletin, Beijing, Comprehensive Planning Department, 2016, pp. 1-1. (In Chinese).
2. Jin Yinli, Zhang Ying, and Han Xueting, Estimation Method of Expressway Section Traffic Flow Based on Toll Data, Traffic Information Security, 2015, vol.33, pp. 47-52. (In Chinese).
3. Li Xing, and He Zhenxing, Thinking on traffic investigation development of Guangdong province, Guangdong Highway Communications, 2002, vol.3, pp. 53-54. (In Chinese).
4. Li Baoling, and Sun Shujuan, The investigation methods of the traffic capacity, Communications Science and Technology Heilongjiang, 2004, vol.2, pp.80-82. (In Chinese).
5. Fu Xin, Wang Jiangwei, and Yuan Changwei, Algorithm for calculation of link traffic based on expressway charge record database, Communications Standardization, vol.6, pp.150-152. (In Chinese).
6. Zhang Hongbin, Ni Fujian, Yang Shunxin, and Liu Baiqiang, Freeway section traffic flow calculation based on toll data, Southeast University Journal, 2013, vol.43, pp.320-324. (In Chinese).
7. Wang Shengji, Wang Jiajie, Method of calculating the amount of the closed section of road traffic through toll data, China Information Industry Transportation, 2006, vol.12, pp.78-80. (In Chinese).

8. Yang Xiaoguang, Zhang Ruhua, Chu Hao, and Wu Sufeng, The research of basic problems of traffic information acquisition and processing based on highway toll system, *System Engineering*, 2004, vol.11, pp.4-11. (In Chinese).
9. Textbook Operations Research group, *Operational Research*, Beijing, Tsinghua University Press, 2001, pp. 320-356. (In Chinese).
10. Hao Jie, Evaluation of Highway Network Running State based on Macroscopic Fundamental Diagram, Beihang University master thesis, 2015, pp.15-15. (In Chinese).
11. JTG B01-2003, Highway engineering standard, 2003. (In Chinese).
12. Guo Xiucheng, Highway traffic running state analysis method and its application, Nanjing, Southeast University Press, 2012. (In Chinese).