

# Research on Prediction Method of Freeway Operation Situation Based on Short-Term Traffic Flow Multi-Parameters Regression

Xiaodan Zhang, Liande Zhong and Yongsheng Chen

*Research Institute of Highway, Ministry of Transport, 8 Xitucheng Road, Haidian District, Beijing 100088, China*

**Abstract.** In order to realize the short-term prediction of traffic flow operation security situation, so as to enhance the operation safety of freeway, based on the traffic flow detection data and traffic accident data of Beijing section in Jing-ha freeway, the security situation prediction model based on short-term multi-parameters was established in the paper. Firstly, we extracted all the traffic flow data of microwave coils and videos, as well as the traffic accidents after updating the electromechanical system of Beijing section in Jing-ha freeway, so as to establish the risk prediction database, and developed the pre-analysis software of basic data; secondly, the traffic flow data of 30 minutes prior to the time of the accident were divided into six slices at 5-minute level, and the volume, speed, occupancy as well as their statistical parameters were selected as the chosen parameters of the model; next, single-parameter Logistic regression analysis was carried out in each time slice respectively; finally, based on the results of the correlation analysis of parameters and the single-parameter modeling significance, the multi-parameters Logistic regression model was established in each time slice respectively, thus obtaining the short-term prediction model of the traffic flow operation security situation. The results indicate that, the modeling fitting effect of slice 1 is the best, that is the change of the traffic flow parameters and their statistics in 5 minutes prior to the time of the accident can effectively predict the possibility of the accident, in which the average speed has a significant impact on the risk of the accident.

## 1 Introduction

Intelligent transportation system (ITS) is widely used in the global scope, which makes the traffic managers have large real-time traffic situation data. Many researchers and practitioners have been fully aware that all the advantages of ITS will not be recognized without realizing the ability of the traffic flow short-term prediction [1] (Brian L. Smith, 2002). The traffic flow prediction model can provide such a kind of ability, and can provide forward-looking traffic management as well as comprehensive travel information service.

At present, the object of traffic forecast is mainly urban road, and it focuses on the unblocked reliability and efficiency [2]-[7], including the assessment of traffic operation situation, short-term traffic forecast, road traffic condition judgment and so on, but the researches on the traffic flow operation security situation prediction are rare.

In China, due to the lack of detailed accident data and microscopic traffic flow data as support, resulting in a serious shortage in the real-time traffic flow security analysis theory, leading to the current freeway security management lag behind the real-time traffic situation prediction in our country. The United States and Canada began to carry out researches on the traffic accident

detection algorithms and the traffic flow harbinger characteristics before the traffic accident from 1990s [8-15]. Among them, Chris adopted the speed differences between upstream and downstream, and the variances of the cross-section speed as the characterization factors of the traffic flow real-time risk discrimination [9], the results of which was referenced by Kansas state highway agency of USA in 2006. However, the main shortcomings were unknown of the risk reason and subjectivity of the risk rank assessment, and its scientificity needs to be further examined.

So, the prediction method of freeway operation situation based on short-term traffic flow multi-parameters regression is researched in this paper, thus achieving the short-term prediction of the traffic flow operation situation, and the results are helpful to reduce the risk of accident, decrease traffic accidents, and improve the operation security of freeway.

## 2 Data processing

Jing-ha freeway (Beijing section) is about 39.7 kilometers, two-direction of which has 20 microwave detectors (they can detect the traffic flow data of two directions), 16 video vehicle detectors (they can detect data of one direction), the average spacing of detectors toward Beijing direction is

1.32 kilometers, and the average spacing of detectors far away from Beijing direction is 1.53 kilometers. All the traffic flow data of microwave coils and videos, as well as the traffic accidents after updating the electromechanical system of Beijing section in Jing-ha freeway are extracted in this paper, in which, the traffic accidents include time, location, type, reason, and et. al., and the traffic flow data are speed, volume and occupancy of divided-lane at 1-minute level.

First, we select the nearest detector data within 2 kilometers prior to or back to the location of accident, in order to screen data for first round, in which, the choice of accidents should avoid these caused by external factors such as weather, linear conditions, drivers and vehicles as much as possible, only in this way we can accurately excavate the rule of the traffic flow fluctuation affecting on the accidents, and in consideration that the relationship between the causes and effects of single-vehicle accidents and the traffic flow may not be strong under high service level, so the select of accident samples is more emphasis on the multi-vehicle accidents in large traffic volume (under C-classical service level); and then, we screen data for second round: check the quality of the traffic flow data, including deleting and processing the singular values such as speed with 0 km/h, diagnose the outliers by the data spatio-temporal graph method and the statistical method, and correct the abnormal data values by the simple difference method and the filtering method, in order to improve the accuracy and reliability of the analytical results, thus selecting the speed, volume and occupancy of detectors with better data quality to pair with accidents, and extract the traffic flow data of the control groups by 1:4 ratio.

The control group data meet the following requirements: the date is different with the corresponding accident; the time, week, and location are the same as the corresponding accident; the control group has no accidents at this location at the same day. Then, we select the control groups with better data quality according to the same method mentioned above, so as to establish the database required in this paper.

Because the location of traffic accident recorded by the police department is a cross-sectional stake number, therefore, it is necessary to aggregate the divided-lane traffic flow data into a cross-sectional traffic flow data, thus taking the cross-sectional data as the foundation of the research, so, the divided-lane data are transformed into

the cross-sectional data by weighting method:  $q_s = \sum_{i=1}^n q_i$ ,

$$k_s = \frac{\sum_{i=1}^n q_i \cdot k_i}{\sum_{i=1}^n q_i}, \quad v_s = \frac{\sum_{i=1}^n q_i \cdot v_i}{\sum_{i=1}^n q_i},$$

in which,  $q_i$ ,  $k_i$ ,  $v_i$  are

respectively volume, occupancy, speed of divided-lane,  $n$  is the number of lanes.

We chose a tall building on the southwest corner of the intersection and used a video-based system to record traffic conditions. Then, we selected 150 typical crossing cases, including two straight-moving vehicles from the orthogonal direction, as shown in Fig. 1(b). To avoid interference from other vehicles to drivers' crossing

behavior, we selected only the simplest crossing cases. In each case, one straight-moving vehicle encountered another straight-moving vehicle and no other object.

### 3 Mathematical model

#### 3.1 Logistic regression model

Binary Logistic regression model is commonly used to quantitatively analyze the impact of explanatory variables on binary dependent variables, also can be used to estimate the occurrence probability of a category of the dependent variables, and from the traffic flow operation results, the dependent variables just can be divided into two categories: accidents and non-accidents. The probability of the accident corresponding to one sample data is:

$$P(x_i) = \frac{1}{1 + e^{-x_i \beta}}, \quad i = 1, 2, \dots, n \quad (1)$$

The linear expression after logit transform is:

$$\ln \frac{P(x_i)}{1 - P(x_i)} = x_i \beta, \quad i = 1, 2, \dots, n \quad (2)$$

where,  $P(x_i)$  represents the probability of traffic accident;  $x_i \beta$  represents the linear combination of explanatory variables:  $x_i \beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ ,  $x_{ki}$  is the value of variable  $k$  in  $i$ -th sample;  $\beta_0$  is the regression intercept;  $\beta_1, \beta_2, \dots, \beta_k$  are the regression coefficients of explanatory variable  $x_{ki}$ ;  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  can be calculated by the maximum likelihood estimation method:

$$\ln L(\beta, x_i) = \sum_{i=1}^n [\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} - \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}})].$$

#### 3.2 Logistic model testing

In Logistic regression, likelihood ratio test, Akaike information criterion (Akaike Information Criterion, AIC) and Schwarz criterion can be used to reflect the goodness of the model fitting.

We adopts AIC to reflect the fitting effect of the final model in this paper, that is  $AIC = -2LL + 2(K + S)$ , in which,  $K$  is the number of independent variables of the model,  $S$  is the total number of response variable categories minus 1, the range of  $-2LL$  is 0 to infinity, which is the smaller the better. Under the same conditions, the smaller value of AIC indicates the better model fitting.

The prediction accuracy for classification is usually used to reflect the prediction accuracy of the model. Using Logistic model to predict the classifications needs specifying a threshold of probability, that is, when the probability calculated by Logistic model is greater than a specified threshold, it is discriminated as the traffic accident, and when the probability is less than a specified threshold, it is discriminated as the security state with non-accident. The threshold value decides the forecast accuracy of each category and the total samples, the current researches commonly use the proportion of a

category in the whole samples as the threshold value of this category prediction. Because of researching the prediction method of freeway traffic operation situation short-term in this paper, the proportion of accidents in the whole samples is adopted as the threshold value.

## 4 Research results and analysis

### 4.1 Data preparation

In order to predict the traffic accident in advance, the calibrated traffic flow data within 30 minutes prior to the accident is extracted in this paper, meanwhile, the traffic flow data of the control group in the corresponding period is extracted for each accident. Through screening of data quality, we ultimately retain 112 groups of accident samples and 448 groups of non-accident samples as the control groups for modeling research, and divide them into 2 categories: accident and non-accident, that is the value of dependent variable is 1 indicating accident and 0 meaning non-accident.

Firstly, the original traffic flow data of divided-lane are aggregated into a cross-sectional traffic flow data, secondly, in order to avoid the data noise due to the short acquisition interval, the data are converged with 5-minute level in order to get averages and standard deviations, thus for 5-minute aggregation half an hour period is divided into 6 time slices: the interval between time of accident and 5 minutes prior to the accident is named as slice 1, interval between 5 to 10 minutes prior to the accident as slice 2, interval between 10 to 15 minutes prior to the accident as slice 3 and so on, interval between 25 to 30 minutes prior to the accident as slice 6.

### 4.2 Model parameters selection

We extract 9 statistical parameters as the chosen model parameters: the average values of speed  $as$ , the average values of volume  $av$ , the average values of occupancy  $ao$ , the standard deviations of speed  $ss$ , the standard deviations of volume  $sv$ , the standard deviations of occupancy  $so$ , the coefficients of variation of speed  $cvs$  ( $cvs = \frac{ss}{as}$ ), the coefficients of variation of volume  $cvv$  ( $cvv = \frac{sv}{av}$ ), the coefficients of variation of occupancy  $cvo$  ( $cvo = \frac{so}{ao}$ ) at 5-minute level.

### 4.3 Modeling steps

The binary Logistic model of the freeway traffic security operation situation deduction based on short-term traffic flow multi-parameters is established by using statistical analysis software with R programming language in this paper, the methods are as follows:

(1) We use the correlation analysis method to select parameters in each time slice, which makes the highly correlated variables be not into the Logistic model;

(2) For each parameter, we carry out a binary Logistic regression model in each time slice (single-variable Logistic regression), and select parameters which have significant impact on the traffic accident risk of freeway as the candidate variables of the subsequent steps;

(3) Combine with the analysis results of step (1) and (2) to select parameters for entering the final model, and then select the reasonable explanatory variables for modeling according to the backward selection method of the Logistic regression, thus obtaining the final models of six slices, the basic steps are as follows:

- a) First, all the variables are contained in the model;
  - b) Then, calculate the  $z$  test values of all variables, and get the corresponding  $P$  values;
  - c) Last, find the largest  $P$  value, if the  $P$  value is greater than the significance level  $\alpha_{out}$ , this variable is eliminated;
  - d) Go back to step b) for the next round of elimination.
- Among them, the significance level of the reserved values is set as: including variables is  $P \leq 0.05$ , excluding variables is  $P > 0.1$ .

### 4.4 Logistic model regression and analysis

Firstly, we use the correlation analysis method to select parameters in each time slice, the results of slice 1 are shown in Table 1.

Generally, if the correlation coefficient between two parameters is  $> 0.6$  or  $< -0.6$ , they are strongly related to each other, and they can not enter the model at the same time. Through the experimental results, it can be seen that, for slice 1:

$Cvv$  and  $cvo$ ,  $sv$  and  $av$ ,  $ao$  and  $as$ ,  $cvs$  and  $ss$  are strongly related to each other, so they can not enter the model at the same time.

The results of the other slices are obtained by the same method.

Slice 2:  $Cvv$  and  $cvo$ ,  $sv$  and  $av$ ,  $ao$  and  $as$ ,  $ao$  and  $so$ ,  $cvs$  and  $ss$  are strongly related to each other, so they can not enter the model at the same time.

Slice 3:  $Cvv$  and  $cvo$ ,  $sv$  and  $av$ ,  $ao$  and  $as$ ,  $ao$  and  $so$ ,  $cvs$  and  $ss$  are strongly related to each other, so they can not enter the model at the same time.

Slice 4:  $Cvv$  and  $cvo$ ,  $ao$  and  $as$ ,  $ao$  and  $so$ ,  $cvs$  and  $ss$  are strongly related to each other, so they can not enter the model at the same time.

Slice 5:  $Cvv$  and  $cvo$ ,  $sv$  and  $av$ ,  $sv$  and  $so$ ,  $ao$  and  $as$ ,  $ao$  and  $so$ ,  $cvs$  and  $ss$  are strongly related to each other, so they can not enter the model at the same time.

Slice 6:  $Cvv$  and  $cvo$ ,  $sv$  and  $av$ ,  $ao$  and  $as$ ,  $ao$  and  $so$ ,  $cvs$  and  $ss$  are strongly related to each other, so they can not enter the model at the same time.

Secondly, for accurately analyzing the impact of a single traffic flow parameter on the probability of accident (single-factor analysis), for each parameter, we carry out a binary Logistic regression model in each time slice, for 5-minute aggregation there has  $6 \times 9 = 54$  single-variable model, due to the same modeling method, the Logistic regression results of  $as$  as the independent variable are only shown here in Table 2:

It can be seen, for parameter *as*, the significance levels of parameters in six slices are all  $P < 0.001$ , indicating that the average value of speed at 5-minute level has significant

effect on the traffic accident risk of freeway, which is selected as the candidate variable of the subsequent modeling process.

**Table 1.** Correlation coefficients of variables

	<i>cvv</i>	<i>sv</i>	<i>av</i>	<i>cvo</i>	<i>so</i>	<i>ao</i>	<i>cvs</i>	<i>ss</i>	<i>as</i>
<i>cvv</i>	1	0.123703	-0.5397	0.74723	0.115362	-0.26757	0.385528	0.406205	-0.0469
<i>sv</i>	0.123703	1	0.634693	-0.06899	0.394076	0.278207	-0.07135	-0.06005	0.081788
<i>av</i>	-0.5397	0.634693	1	-0.54284	0.249592	0.507524	-0.26532	-0.31465	-0.0451
<i>cvo</i>	0.74723	-0.06899	-0.54284	1	0.236815	-0.34532	0.364011	0.499118	0.032488
<i>so</i>	0.115362	0.394076	0.249592	0.236815	1	0.589883	0.405378	0.058607	-0.50241
<i>ao</i>	-0.26757	0.278207	0.507524	-0.34532	0.589883	1	0.191952	-0.27208	-0.69086
<i>cvs</i>	0.385528	-0.07135	-0.26532	0.364011	0.405378	0.191952	1	0.731012	-0.45599
<i>ss</i>	0.406205	-0.06005	-0.31465	0.499118	0.058607	-0.27208	0.731012	1	0.074069
<i>as</i>	-0.0469	0.081788	-0.0451	0.032488	-0.50241	-0.69086	-0.45599	0.074069	1

**Table 2.** Regression analysis results of *as*

Slice	Coefficient ( $\beta_i$ )	Standard error	z value	Significance level (Pr(> z ))	Exp( $\beta_i$ )	Exp( $\beta_i$ )-1
1	-0.05770	0.00926	-6.231	4.62e-10	0.943933	-5.6%
2	-0.058400	0.009493	-6.152	7.66e-10	0.943272	-5.7%
3	-0.05357	0.00891	-6.012	1.83e-09	0.94784	-5.2%
4	-0.046271	0.008381	-5.521	3.37e-08	0.954783	-4.5%
5	-0.043120	0.008094	-5.327	9.97e-08	0.957796	-4.2%
6	-0.046069	0.008469	-5.439	5.34e-08	0.954976	-4.5%

Odds ratio of the traffic flow parameters can be used to quantify the impacts of different traffic flow parameters on the risk of accident. The odds of an event is defined as the ratio of the probability with occurrence and the probability without occurrence, therefore, for the Logistic regression model:

$$\log\left[\frac{\text{Prob}(\text{event})}{\text{Prob}(\text{nonevent})}\right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (3)$$

$$\text{Odds} = \frac{\text{Prob}(\text{event})}{\text{Prob}(\text{nonevent})} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k} \quad (4)$$

It can be seen that when the *i*-th independent variable changes a unit, the change of odds is  $\text{Exp}(\beta_i)$ , that is:

$$\text{Odds ratio} = \frac{\text{Odds}(X_i + 1)}{\text{Odds}(X_i)} = \text{Exp}(\beta_i) \quad (5)$$

If the coefficient of an independent variable is positive, it means that odds will increase, and this value will be greater than 1; if the coefficient of an independent variable is negative, it means that odds will decrease, and this value will be less than 1; when the coefficient of an independent variable is 0, this value is equal to 1. The change percentage of odds ratio is  $\text{Exp}(\beta_i) - 1$ .

From Table 2, the odds ratios of *as* in six slices are all less than 1, and with the farther away from the time of accident, the value is increasing, which shows that the probability of accident will decrease with the increase of *as*.

The Logistic regression analysis results of the other independent variables are obtained by the same modeling method: for parameter *av*, the independent variables are

not significant in slice 1~6, which shows that the correlation between the change of volume and the accident is relatively low; for parameter *ao*, the significance levels of parameters in six slices are all  $P < 0.001$ , indicating that the average value of occupancy at 5-minute level has significant effect on the traffic accident risk of freeway, which is selected as the candidate variable of the subsequent modeling process, and the odds ratios of *ao* in six slices are all greater than 1, and with farther away from the time of accident, the value is decreasing, which shows that the probability of accident will increase with the increase of *ao*; for parameter *ss*, the independent variables are not significant in slice 1~6; for parameter *sv*, the independent variables are significant in slice 1 and slice 4, the significance levels of them are  $P < 0.05$  and  $P < 0.1$  respectively, the other slices are not significant, and the odds ratios in slice 1~6 are all less than 1, which shows that the probability of accident will decrease with the increase of *sv*; for parameter *so*, the significance levels of parameters in slice 1~6 are all  $P < 0.01$ , indicating that the standard deviation of occupancy at 5-minute level has significant effect on the traffic accident risk of freeway, which is selected as the candidate variable of the subsequent modeling process, and the odds ratios of *so* in slice 1~6 are obviously greater than 1, showing that the probability of accident will increase with the increase of *so*; for parameter *cvs*, the significance levels of parameters are more higher in slice 1, 2, 3 and 6, the significance levels are  $P < 0.01$ , the significance levels are both  $P < 0.1$  in slice 4 and 5, and the odds ratios are far greater than 1, thus the probability of accident will increase significantly with the increase of the coefficient of variation of speed; for parameter *cvv*, the independent variables are significant in slice 3 and slice 4, the significance levels of them are

$P < 0.1$  and  $P < 0.05$  respectively, the other slices are not significant, and some odds ratios are far less than 1 and some odds ratios are slightly greater than 1, which shows the impacts on the possibility of accident may be not consistent; for parameter *cvo*, the independent variables are all not significant in slice 1~6.

In summary, in slice 1, the modeling effects of parameters *sv*, *so*, *ao*, *cvs*, and *as* are the most significant; in slice 2, the modeling effects of parameters *so*, *ao*, *cvs*, and *as* are the most significant; in slice 3, the modeling effects of parameters *cvv*, *so*, *ao*, *cvs*, and *as* are the most significant; in slice 4, the modeling effects of parameters *cvv*, *sv*, *so*, *ao*, *cvs*, and *as* are the most significant; in slice 5, the modeling effects of parameters *so*, *ao*, *cvs*, and *as*

are the most significant; in slice 6, the modeling effects of parameters *so*, *ao*, *cvs*, and *as* are the most significant.

Combining with the correlation analysis results of parameters and the single-parameter modeling significance results, the variables which have significant affect on the accident risk of freeway and are uncorrelated with each other are selected for modeling, so the final variables entering the Logistic regression model of multi-parameters in each slice are respectively: for slice 1, *cvv*, *sv*, *so*, *cvs*, *as*; for slice 2, *sv*, *cvo*, *so*, *cvs*, *as*; for slice 3, *cvv*, *sv*, *so*, *cvs*, *as*; for slice 4, *cvv*, *sv*, *so*, *cvs*, *as*; for slice 5, *av*, *cvo*, *so*, *cvs*, *as*; for slice 6, *sv*, *cvo*, *so*, *cvs*, *as*.

The final regression results of model are shown in Table 3.

**Table 3.** Comparison of regression analysis results of slice 1~6

Slice	Variable	Coefficient ( $\beta_i$ )	Standard error	z value	Significance level (Pr(> z ))	Exp( $\beta_i$ )
1	Constant	3.21024	0.74016	4.337	1.44e-05	24.78503
	<i>as</i>	-0.05770	0.00926	-6.231	4.62e-10	0.943933
	The fitting parameters of the model: sample size=560, -2LL = 504.15, AIC=508.15					
2	Constant	3.285394	0.761098	4.317	1.58e-05	26.71951
	<i>as</i>	-0.058400	0.009493	-6.152	7.66e-10	0.943273
	The fitting parameters of the model: sample size =560, -2LL = 506.65, AIC=510.56					
3	Constant	2.89824	0.71468	4.055	5.01e-05	18.14219
	<i>as</i>	-0.05357	0.00891	-6.012	1.83e-09	0.94784
	The fitting parameters of the model: sample size =560, -2LL = 512.45, AIC=516.45					
4	Constant	2.836702	0.735927	3.855	0.000116	17.05941
	<i>cvv</i>	-1.804372	0.893609	-2.019	0.043467	0.164578
	<i>as</i>	-0.046800	0.008598	-5.443	5.23e-08	0.954278
	The fitting parameters of the model: sample size =560, -2LL = 518.77, AIC=524.77					
5	Constant	3.033395	0.923545	3.285	0.00102	20.76762
	<i>cvo</i>	-1.535999	0.932708	-1.647	0.09960	0.215241
	<i>av</i>	-0.030455	0.014048	-2.168	0.03016	0.970004
	<i>so</i>	0.346117	0.183208	1.889	0.05886	1.413568
	<i>cvs</i>	-6.078032	3.376873	-1.800	0.07188	0.002293
	<i>as</i>	-0.041821	0.009812	-4.262	2.02e-05	0.959041
The fitting parameters of the model: sample size =560, -2LL = 520.45, AIC=532.45						
6	Constant	2.319736	0.683262	3.395	0.000686	10.17299
	<i>as</i>	-0.046069	0.008469	-5.439	5.34e-08	0.954976
	The fitting parameters of the model: sample size =560, -2LL = 524.32, AIC=528.32					

**Table 4.** Comparison of regression model prediction accuracy of slice 1~6

Model	Actual Classification	Forecast Classification		Total (rate)
		Accident (rate)	Non-accident (rate)	
1	Non-accident	135 (30.13%)	313 (69.87%)	448 (100%)
	Accident	68 (60.71%)	44 (39.29%)	112 (100%)
2	Non-accident	141 (31.47%)	307 (68.53%)	448 (100%)
	Accident	65 (58.04%)	47 (41.96%)	112 (100%)
3	Non-accident	144 (32.14%)	304 (67.86%)	448 (100%)
	Accident	64 (57.14%)	48 (42.86%)	112 (100%)
4	Non-accident	149 (33.26%)	299 (66.74%)	448 (100%)
	Accident	59 (52.68%)	53 (47.32%)	112 (100%)
5	Non-accident	150 (33.48%)	298 (66.52%)	448 (100%)
	Accident	59 (52.68%)	53 (47.32%)	112 (100%)
6	Non-accident	159 (35.49%)	289 (64.51%)	448 (100%)
	Accident	57 (50.89%)	55 (49.11%)	112 (100%)

From the results, the regression variable of the optimal final model in slice 1, 2, 3 and 4 is only *as*; the optimal model of slice 4 retains two variables: *cvv* and *as*; the final model of slice 5 retains five variables: *cvo*, *av*, *so*, *cvs* and *as*. In that case, *as* (the average value of speed at 5-minute level) is the most closely related to the probability of accident, and the significance level of each slice achieves to  $P < 0.0001$ , indicating that *as* has the remarkable

influence on the traffic accidents risk of freeway. Odds ratios of the traffic flow parameters are also given in this table to quantify the impact of different traffic flow parameters on the risk of accident, it can be seen that, the odds ratios of *as* in six slices are all less than 1, and with farther away from the time of accident, the value is increasing, which shows that the probability of accident will decrease with the increase of *as*. By comparing the

values of AIC in final model of each slice, we get that, the farther away from the time of accident, the value of AIC is increasing, illustrating that the fitting effect of the final model is the better the closer to the time of accident. So, the change of the average speed in 5 minutes prior to the accident is the most effective to predict the probability of accident, which is used as the risk characterization factor of real-time safety assessment for the traffic flow operation.

The final model of the security situation deduction is as follows:

$$P(x_i) = \frac{1}{1 + e^{-x_i \beta}} = \frac{1}{1 + e^{-(3.21024 - 0.0577as)}} \quad (6)$$

where,  $as$  is the average speed of 5 minutes prior to the accident.

After specifying a reasonable threshold value, the calibrated model can predict the risk of freeway traffic accidents in real-time. Because in the total samples, the proportion of accidents is 20%, so the threshold is set to 0.2 in this paper, namely when the probability of the model output is greater than 0.2, it is discriminated as the traffic accident; and when the probability of the model output is less than 0.2, it is discriminated as the safety state with non-accident. The comparison of the model prediction accuracy of slice 1~6 is shown in Table 4.

In slice 1, it can predict 60.71% of accidents and 69.87% of non-accidents, the total prediction accuracy is 68.04%; in slice 2, it can predict 58.04% of accidents and 68.53% of non-accidents, the total prediction accuracy is 66.43%; in slice 3, it can predict 57.14% of accidents and 67.86% of non-accidents, the total prediction accuracy is 65.71%; in slice 4, it can predict 52.68% of accidents and 66.74% of non-accidents, the total prediction accuracy is 63.93%; in slice 5, it can predict 52.68% of accidents and 66.52% of non-accidents, the total prediction accuracy is 63.75%; in slice 6, it can predict 50.89% of accidents and 64.51% of non-accidents, the total prediction accuracy is 61.79%. It can be seen that, the farther away from the time of accident, the total prediction accuracy is decreasing, and for the accident prediction accuracy, the prediction accuracy of slice 1 is significantly higher than that of the other slices, so the accident risk prediction model established in slice 1 is capable of using the real-time traffic flow data to predict the freeway traffic accident risk in real-time well.

## 5 Conclusions

The traffic flow data of the nearest detectors prior to or back to the location of the typical accidents of Beijing section in Jing-ha freeway are extracted in this paper, on

the basis, we select the binary Logistic regression method to establish the security risk prediction model and realize the real-time prediction of the probability of accident by using the average speed at 5-minute level, which providing the more scientific support for the traffic control and the traffic emergency management decisions of freeway.

## Acknowledgment

Fund: Applied Basic Research Project of Ministry of Transport (2013319223170).

## References

1. Beijing Traffic Research Center. *TMC Beijing Live-presentation Meeting Information* **12**, 15-18 Beijing, (2005).
2. G. Wang, X. Guo, Y. Jiang, J. Hou. *Journal of Highway and Transportation Research and Development* **27**(4), 155-158, (2010).
3. Y. Liang, Y. Chen, F. Ren. *Journal of Highway and Transportation Research and Development* **22**(12), 105-108, (2005).
4. K. Chen, W. Zou, X. Li. *Highways & Automotive Applications* **3**, 59-61, (2012).
5. J. Chen, W. Zhou. *Journal of Chang'an University (Natural Science Edition)* **22**(4), 52-54, (2002).
6. H. Guo, H. Shi. *Journal of Highway and Transportation Research and Development* **22**(8), 102-105, (2005).
7. Q. Xu, R. Yang. *Journal of Highway and Transportation Research and Development* **22**(12), 131-134, (2005).
8. Weil R, Wootton J, Garcia-ortiz A. *Mathematics Computer Modeling* **27**(9), 257, (1998).
9. Lee Chris Choongho. *School of Civil Engineering, The University of Waterloo, Ontario*, (2004).
10. Abdelaty M, Uddin N, Abdalla F, et al. *Transportation Research Record* **1897**, 88-95, (2004).
11. Abdelaty M, Pande A. *Journal of Safety Research* **36**(1), 97-108 (2005).
12. Abdelaty M, Uddin N, Pande A. *Transportation Research Record*, **1908**, 51-58, (2005).
13. Lee C, Hellinga B, Saccomanno F. *Transportation Research Record*, **2749**, 67-77, (2003).
14. Lee C, Saccomanno F, Hellinga B. *Transportation Research Record*, **1784**, 1-8, (2002).
15. Hossain M, Muromachi Y. *The 89<sup>th</sup> Annual Meeting of the Transportation Research Board*, 1-15 Washington, D. C., (2010).