

**GUOHUA LIANG**, Ph.D.<sup>1</sup>

E-mail: lgh@chd.edu.cn

**XUJIAO SUN**, M.Sc.<sup>1</sup>

(Corresponding author)

E-mail: 2018321046@chd.edu.cn

**YIDAN ZHANG**, M.Sc.<sup>1</sup>

E-mail: 2018121259@chd.edu.cn

**MINGLI CHEN**, M.Sc.<sup>1</sup>

E-mail: 2018221235@chd.edu.cn

**WANTING ZHANG**, M.Sc.<sup>1</sup>

E-mail: 2018121244@chd.edu.cn

<sup>1</sup> College of Transportation Engineering

Chang'an University

Middle Section of South 2 Ring Rd., Xian, 710064,

Shaanxi, China

Traffic on Motorways

Original Scientific Paper

Submitted: 19 Aug. 2020

Accepted: 12 Feb. 2021

DOI: 10.7307/ptt.v33i5.3680

# IDENTIFYING EXPRESSWAY ACCIDENT BLACK SPOTS BASED ON THE SECONDARY DIVISION OF ROAD UNITS

## ABSTRACT

*For the purpose of reducing the harm of expressway traffic accidents and improving the accuracy of traffic accident black spots identification, this paper proposes a method for black spots identification of expressway accidents based on road unit secondary division and empirical Bayes method. Based on the modelling ideas of expressway accident prediction models in HSM (Highway Safety Manual), an expressway accident prediction model is established as a prior distribution and combined with empirical Bayes method safety estimation to obtain a Bayes posterior estimate. The posterior estimated value is substituted into the quality control method to obtain the black spots identification threshold. Finally, combining the Xi'an-Baoji expressway related data and using the method proposed in this paper, a case study of Xibao Expressway is carried out, and sections 9, 19, and 25 of Xibao Expressway are identified as black spots. The results show that the method of secondary segmentation based on dynamic clustering can objectively describe the concentration and dispersion of accident spots on the expressway, and the proposed black point recognition method based on empirical Bayes method can accurately identify accident black spots. The research results of this paper can provide a basis for decision-making of expressway management departments, take targeted safety improvement measures.*

## KEYWORDS

*traffic safety; accident black spots identification; expressway; division of road units; road safety index; empirical Bayes method.*

## 1. INTRODUCTION

According to the data from the China Statistical Yearbook [1], there were 203,000 traffic accidents in China in 2017, resulting in more than 60,000 deaths and more than 200,000 injuries, resulting in direct economic losses of about 1.2 billion yuan. Road sections with dense accident distribution and intersections with more accidents are called accident black spots [2]. The existing investigation results show that 25% of the total number of traffic accidents occur on the road which accounts for 0.25% of the total length of the road network [3], and the black spots of a larger proportion of traffic accidents occupy less road mileage. Therefore, accurate identification of accident black spots will be beneficial to the analysis of the characteristics of road traffic accidents and the understanding of the law of road traffic accidents, so as to reduce traffic accidents and improve driving safety.

In the identification of multiple accident points, the classical methods include accident frequency method [4], accident rate method [5, 6], matrix method [7], quality control method [8], etc. Although these methods have been widely applied in engineering practice, there are also problems such as inconsistent evaluation indicators and large deviation of identification results of different methods [9]. Since the statistical accident index cannot fully reflect the road safety condition, some scholars use Bayesian method [10-12], cluster analysis

method [13, 14], and fuzzy comprehensive evaluation method [15] to evaluate road safety and establish black spot identification method. Then Ghadi, M et al. evaluated empirical Bayesian method, excess Bayesian method, accident frequency method and accident rate method. They concluded that empirical Bayesian method had the best effect in accident black spots identification [16], and defined accident black spot [17] in combination with empirical Bayesian method and k-means clustering method. In addition, many scholars have studied the relationship between road alignment, road facilities, location speed, number of traffic conflicts, and traffic accident rate, and improved the accident black spot identification method [18-20].

With the development of computer technology, scholars have explored a series of new methods and ideas, such as dynamic traffic distribution simulator, visual interactive system, GIS, quantile regression technology, etc. Dynamic traffic distribution simulator and visual interactive system are the basic data set for estimating the accident risk by combining the traffic flow estimation of the road section with the accident data in the historical record, predicting the accident risk of each location in the future, and quantifying the accident risk index [21]. GIS is used to determine the spatial location of the accident, estimate the nuclear density, and calculate the accident risk index to identify the accident black spot [22, 23]. Quantile regression technology is based on the equivalent calculation of accident property loss to consider the accident severity and determine the traffic network accident black spot [24].

Scholars have noticed that the length and starting and ending positions of accident statistical sections will directly affect the calculation results of accident rate and the final determination of accident black spots [4]. However, the existing research on the identification of highway accident black spots is mainly based on the division of fixed length road unit [25], and substitute the data of expressway accidents, traffic conditions and linear conditions into the relevant black spots identification model to determine the black spots of accidents. On the one hand, the rationality of road unit division is rarely considered, and the accident distribution and road section division are not well combined; on the other hand, the regression effect of statistical accident number is not considered (It means that if the number of accidents beyond the normal state occurs in the road unit in a certain period of time, even if no

corresponding improvement measures are taken, the number of accidents occurring in the subsequent period of time will be relatively reduced. On the contrary, if the number of accidents in a certain period of time is significantly less than the normal state, more accidents will certainly occur in the following period of time), so the accident data cannot fully reflect the safety situation of the road section. Therefore, it is necessary to study a more accurate and reasonable road unit division method.

This paper starts from the establishment of a reasonable division method of road units, analyses the shortcomings of the traditional division method, discusses the combination of road alignment and accident distribution characteristics to divide road sections, and innovatively proposes a method of secondary division of road units. Considering the influence of the longitudinal profile of the road and the regression effect of the number of statistical accidents, the accident prediction model is established by using HSM, and the accident prediction model is used as the basis of the prior distribution to establish an empirical Bayesian method combining historical statistics and accident prediction values. Safety evaluation on each road unit of the expressway is conducted and a black point threshold determination method is established. Finally, taking the two-way eight-lane expressway from Xi'an to Baoji as an example, this paper uses the method proposed in this paper to divide the road units, establish and demarcate the accident prediction model, and identify the accident black spots.

## 2. METHODS

Combined with the distribution characteristics of road linear accidents, a reasonable road unit division method is established. Based on the HSM method, the accident prediction model is established as the prior distribution. Based on the empirical Bayesian posterior estimate, the safety evaluation is carried out on each road unit of the expressway and the determination method of the black point threshold is established. The research methodology flowchart is shown in *Figure 1*.

### 2.1 Road unit secondary division method construction

In the general black spot identification process, a certain length (2km, 1km or 500m) [25] is generally set in advance, and the route to be analysed

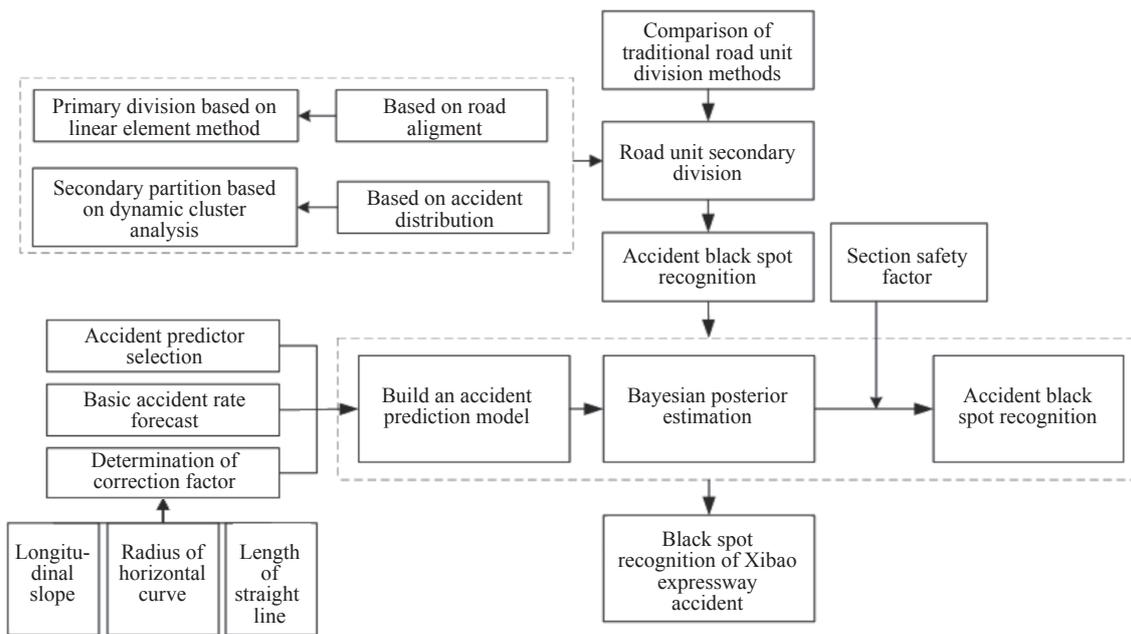


Figure 1 – Research methodology flowchart

is divided into several consecutive unit segments according to the predetermined length. At the same time, by collecting the number of traffic accidents occurring in each unit road section within a certain period of time, the black rating index of the accident is obtained, and finally, the black spot identification method is used to determine which road units are black spots of the accident. The division of roads by the fixed length division method simplifies the specific analysis of the expressway, so that it is convenient to determine the evaluation unit and use various calculation methods for further research. However, this simplified processing method inevitably produces many practical problems, such as the phenomenon of “sharp clipping”. In Figure 2, K0~K6 represents the mileage pile number, taking 2 km as the unit length of division, then K0~K6 can be divided into 3 sections, respectively labelled I~III, of which 2 accidents occurred in the K1~K2 section (indicated by black dots), two accidents occurred in the K2~K3 segment, the number of unit I accidents was 2, and the number of unit II accidents was 2. If three accidents are taken as the judgment thresh-

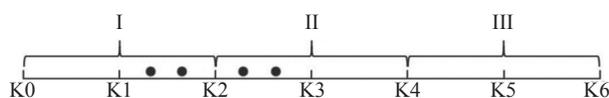


Figure 2 – Schematic diagram of isometric fixed division method

old, then the units are not included in the multiple accident points (segment). However, the length of K1~K3 is also 2 km, and 4 accidents occur, which are larger than the threshold and should be included in the multiple accident points (segment), but the judgment is missed because they are divided into other units.

The basic principle of road unit division is to fully expose the dangerous sections of the road, so the length of the divided road units cannot be too long or too short. According to domestic and foreign experience [26], the length of road units is generally in the range of 0.2 km~16 km. Of course, the specific values should be analysed according to road conditions, traffic conditions, running speed and other conditions. For the plain highway, due to the good road conditions, fast speed, ideal line shape and good viewing distance, a longer unit section should be selected, generally 2 km~16 km. For mountainous highways, due to the complex terrain, variable and complex road shape, poor line of sight and slow speed, shorter unit sections should be selected, generally 0.2 km~2 km. In addition, the statistical status of the accident data should also be considered. The longer the statistical time, the larger the accident volume and the shorter the unit road segment; the rougher the statistical accident data mileage, the worse the statistical accuracy and the longer the unit road segment. Therefore, the division of road units will inevitably affect the accuracy of accident black

spot recognition. This requires road evaluators not only to master the algorithm of accident black spot identification, but also to correctly and reasonably divide the road units, which brings some difficulties to the work of accident black spot identification.

In order to overcome the shortcomings of the fixed length division method in accident black spot recognition, this paper introduces the road linear indicators and the accumulation of traffic accidents on the road into the road unit division. Taking the road line as an indicator, the line element method is used to divide the road for the first time, and then the dynamic segmentation method is used to divide the road segment for the second time based on the first segmentation method.

The basic idea of using the linear unit division method for road division is derived from the “Guidelines for Safety Evaluation of Highway Projects” [27]. The basic idea is as follows (see Figure 3): The first unit (longitudinal slope section): the slope of the longitudinal slope is greater than 3% and the slope length is greater than the minimum design slope length;

The second unit (long straight section): the slope of the longitudinal slope is less than 3% and the length is greater than 200 m;

The third unit (short straight section): the slope of the longitudinal slope is less than 3% and the length is less than 200 m;

The fourth unit (curved section combination section): the slope of the longitudinal slope is greater than 3%, the slope length is greater than the minimum design slope length and the radius is less than 1000 meters;

The fifth unit (plane curve segment): the slope of the longitudinal slope is less than 3% and the radius is less than 1000 m;

The sixth unit (large radius curve segment): the radius is greater than 1000 m.

Hierarchical cluster analysis method is more commonly used as a kind of classification method. However, it is difficult to determine the number of accident classifications and it needs to calculate the inter-class distance matrix and the class diameter matrix, etc., and then carry on with the two comparisons. This results in slow and inefficient operation. Especially when the total sample size  $n$  is very large, the calculation amount of data is extremely large, which is difficult to run, and the results obtained are often not the global optimal solution, so it is not suitable for the classification of objects with a large sample size. In order to avoid the disadvantage

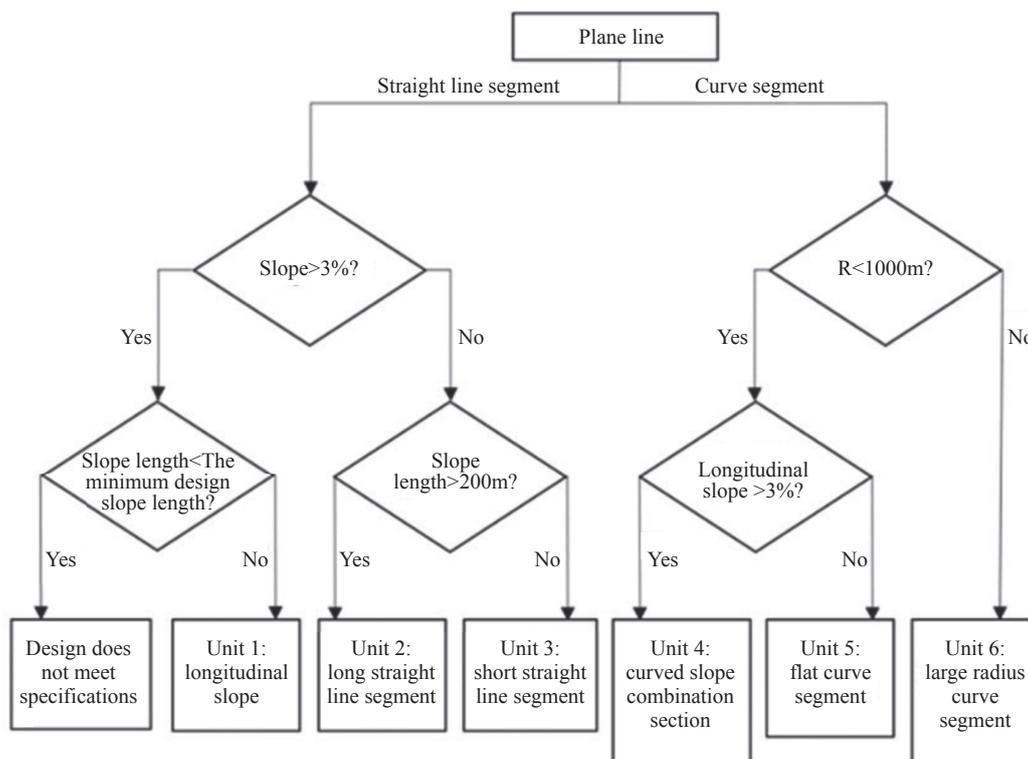


Figure 3 – The segmented process of linear unit division method

of low efficiency of hierarchical cluster analysis method, dynamic clustering method first conducts rough classification, and then modifies according to some optimal criteria to determine the new clustering centre and classification matrix. After several iterations, a more reasonable classification can be obtained. The clustering centre can well reflect the clustering status of road traffic accidents, and the interval range can objectively describe the scattered status of road traffic accidents, which can significantly improve the shortcomings of the hierarchical cluster analysis method.

The dynamic clustering analysis method considers the distribution of accidents, and can divide the road into basic evaluation units according to the safety attributes of the road. It can avoid the identification distortion caused by the division of fixed length and improve the accuracy of the identification of accident black spots.

The dynamic cluster analysis method takes each accident as a sample according to the discrete distribution of accident points on the road section, and then takes the mileage stake number of each accident as the attribute variable of the sample, and segments it according to the mileage stake number of the accident location. On the basis of specifying the number of classifications, dynamic clustering analysis method can determine the clustering centre and the number of samples (the number of accidents) of each category, and finally determine the road unit divided by the starting and ending station, as shown in Figure 4.

The algorithm divides all accident points into  $k$  categories ( $k$  is the specified number of classes, and the number of road units can be determined based on the whole mileage), takes the centre of gravity of each class as the clustering centre of this class, and establishes the initial clustering centre set. Set  $L^{(0)} = \{x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_k^{(0)}\}$ , where  $x_i^{(0)}$  is the mile-

age of the  $i$ th gathering point and the  $i$ th class after  $n$  iterations is  $G_i^{(n)}$ , thus obtaining an initial classification  $G^{(0)} = \{G_1^{(0)}, G_2^{(0)}, G_3^{(0)}, \dots, G_k^{(0)}\}$ . Taking the center of gravity of  $G_i^{(n)}$  as the new clustering centre  $x_i^{(n+1)}$ , the new cluster center set  $L^{(1)} = \{x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_k^{(1)}\}$  can be obtained. All the accident point again divided into  $k$  classes, iteratively calculate until  $G^{(m+1)} = \{G_1^{(m+1)}, G_2^{(m+1)}, G_3^{(m+1)}, \dots, G_k^{(m+1)}\}$  and  $G^{(m)} = \{G_1^{(m)}, G_2^{(m)}, G_3^{(m)}, \dots, G_k^{(m)}\}$  distance within the prescribed scope of precision (generally within  $10^{-2}$  to  $10^{-6}$ ). Then  $G^{(m)}$  is the final cluster center, and combined with the number of samples of each class, the final clustering result can be obtained and the road unit division result can be determined.

## 2.2 Establishment of black spot identification method for expressway accidents

The accident prediction model in the HSM (Highway Safety Manual) uses the same idea as the road capacity for the calculation of each accident number. First, it determines the basic number of accidents (based on the segment flow and length as the basic parameters) and then gradually corrects them to achieve the purpose of accident prediction [28]. For the correction coefficient, the model only considers the horizontal alignment and the road cross-section condition, and ignores the influence of the vertical alignment.

The model involves the road form is similar to the expressway in China, and some studies show that the longitudinal profile of the road has a great influence on traffic safety, especially on the degree of the longitudinal slope of the road [29]. Therefore, this paper draws on the modelling idea of multi-lane expressway accident prediction model in HSM. According to the selection of the influence variable, it

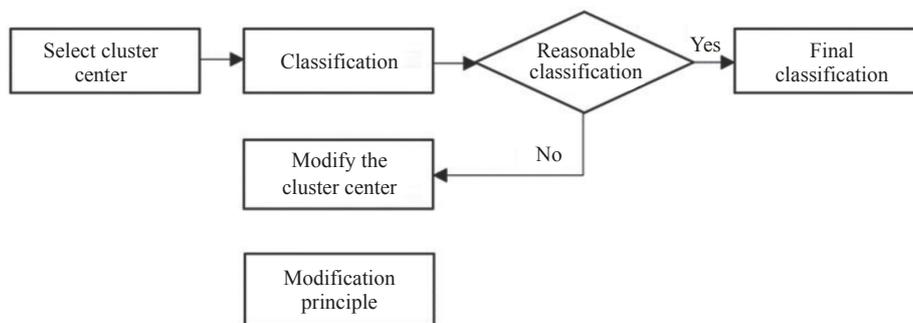


Figure 4 – Dynamic cluster analysis process

is considered that the accident rate of the road unit is determined by revising the basic accident rate under the ideal linear condition, and the correction coefficient includes the correction coefficient of the longitudinal slope of the road  $\beta_i$ , the correction coefficient of the radius of the flat curve  $\beta_{RHC}$ , and the correction coefficient of the length of the straight section  $\beta_{LT}$ . The expressway accident prediction model is shown in Equation 1.

$$N_j = N_{Bj} \cdot \beta_i \cdot \beta_{RHC} \cdot \beta_{LT} \tag{1}$$

where  $N_j$  represents the predicted accident rate of road unit  $j$ ;  $N_{Bj}$  represents the basic equivalent accident rate predicted by road unit  $i$ ;  $\beta_i$  represents the vertical slope correction coefficient of road unit slope  $i$ ;  $\beta_{RHC}$  represents the correction coefficient of the radius of the flat curve, and RHC represents the radius of the flat curve of the road element;  $\beta_{LT}$  represents the correction coefficient of straight segment length, and LT represents the length of straight segment of road element.

The accident data under the ideal road linear condition is extracted, and the relationship model between the basic accident rate  $N_{Bj}$  of the road unit and the annual average daily traffic volume ( $AADT_j$ ) is established according to the basic accident rate prediction model in HSM (see Equation 2).

$$N_{Bj} = \exp(a + b \cdot \ln(AADT_j)) \tag{2}$$

where  $a$  and  $b$  represent the parameter of the calibration.

When determining the basic accident rate of road units, it is first necessary to define the ideal linear condition of multi-lane roads to avoid the interference of the road line shape is not ideal. The ideal linear condition of the expressway refers to the geometric line shape of the road with a small slope of the longitudinal slope, a large radius of the flat curve or a moderate length of the straight section. The ideal linear condition of the expressway cited in literature is: (1) Road longitudinal slope:  $i \in [0, 2\%]$ ; (2) Flat curve radius:  $RHC \geq 4000\text{m}$ ; (3) Straight length:  $LT \in [0.5\text{km}, 1.5\text{km}]$ .

The longitudinal slope of the expressway has a significant impact on the accident. Under the same slope, the downhill section is more dangerous than the uphill section due to the braking performance of the vehicle. By analysing the relationship between the variable and the radius of the flat curve, the radius correction coefficient  $n$  of the flat curve can be determined. It is proposed to use the quadratic polynomial fitting curve of  $i$  (see Equation 3).

$$\beta_i = ci^2 - di + e \tag{3}$$

where  $\beta_i$  represents the correction factor of section slope  $i$  longitudinal slope;  $c$ ,  $d$ , and  $e$  represent the fitting calibrated coefficients respectively.

When the radius of the flat curve is less than 2500 m, the accident rate is sensitive to the change of the radius of the flat curve, and it is advisable to use  $\beta_{RHC}^{-1}$  as the dependent variable. By analysing the relationship between the variable and the radius of the flat curve, the correction coefficient of the flat curve radius can be determined. It is proposed to use RHC as the power function of the independent variable to fit the curve (see Equation 4).

$$\beta_{RHC}^{-1} = f \cdot RHC^{-g} \tag{4}$$

where  $\beta_{RHC}$  represents the flat curve correction coefficient, and RHC represents the flat curve radius of the road section;  $f$  and  $g$  represent the fitting calibrated coefficients respectively.

The linear segment length correction factor can be determined by the quadratic polynomial fitting curve of LT (see Equation 5).

$$\beta_{LT} = h \cdot LT^2 - m \cdot LT + n \tag{5}$$

where  $\beta_{LT}$  represents the correction coefficient of the long straight segment, and LT represents the long straight segment;  $h$ ,  $m$  and  $n$  represent the fitting calibrated coefficients respectively.

Empirical Bayesian posterior distribution is used to evaluate expressway safety. The basic idea is to make full use of the statistical data of expressway traffic accidents to determine the prior distribution, then transform the prior distribution into the posterior distribution according to the empirical Bayesian theory. The equation of Bayesian posterior distribution in reference [30] is shown below:

$$\begin{aligned} f_{\Lambda|Y}(\lambda|y) &= \frac{f_{Y|\Lambda}(y|\lambda) \cdot f_{\Lambda}(\lambda)}{\int_0^{\infty} f_{Y|\Lambda}(y|\lambda) \cdot f_{\Lambda}(\lambda) d\lambda} \\ &= \frac{f_{Y|\Lambda}(y|\lambda) \cdot f_{\Lambda}(\lambda)}{f_Y(y)} \\ &= \frac{[(k + u_b)/u_b]^{k+y} \lambda^{k+y-1} e^{-[(k+u_b)/u_b]\lambda}}{\Gamma(k+y)} \end{aligned} \tag{6}$$

It can be seen from Equation 6 that the posterior distribution is a gamma distribution, the shape parameter is  $\alpha = k + y$ , and the scale parameter is  $\beta = \frac{(k + u_b)}{u_b}$ , then the variance of the posterior distribution is:

$$Var(\Lambda | Y = y) = \frac{\alpha}{\beta^2} = \frac{(k+y)}{(k+u_b)^2} u_b^2 \tag{7}$$

The expected value  $\mu_a$  of the posterior distribution is:

$$\mu_a = E(\Lambda | Y=y) = \frac{\alpha}{\beta} = \frac{k+y}{k+u_b} u_b = \frac{k}{k+u_b} u + \frac{u_b}{k+u_b} y \quad (8)$$

$$= \phi u_b + (1 - \phi)$$

where

$$\phi = \frac{k}{k+u_b} = \frac{1}{1 + \frac{u_b}{k}} = \frac{1}{1 + \frac{u_b^2/k}{E(\Lambda)}} = \left(1 + \frac{Var(\Lambda)}{E(\Lambda)}\right)^{-1} \quad (9)$$

The essence of empirical Bayesian method is to treat the posterior estimate as a weighted value obtained from the prior estimate (accident model prediction value) and survey statistics. Weights are determined by the mean  $E(\Lambda)$  and variance  $Var(\Lambda)$  of the samples that determine the prior estimates (accident prediction model). When the sample mean is constant, the more the sample data deviates from the mean and the larger  $Var(\Lambda)$ , the greater the deviation between the model and similar roads, the smaller the confidence of the prior estimate and the greater the weight of the prior estimate small. The smaller the deviation of the sample data from the mean, the smaller the  $Var(\Lambda)$ , indicating that the smaller the deviation between the model and the similar road, the greater the reliability of the prior estimate and the greater the weight of the prior estimate.

The quality control method based on the empirical Bayesian posterior estimation value is used to determine the accident black spot recognition threshold. The quality control method considers that the probability of a traffic accident on the road follows a Poisson distribution with an accident frequency of  $\mu$ . The probability of occurrence of  $n$  traffic accidents (within statistical period  $t$ ) on the road sections studied in this paper can be expressed by Equation 10.

$$P(n | \mu, t) = \frac{e^{-\mu t} (\mu t)^n}{n!} \quad (10)$$

where  $\mu$  represents the frequency of traffic accidents;  $n$  represents the number of traffic accidents occurred during the time period  $t$ .

Using the empirical Bayesian safety estimate instead of the statistical accident indicator, the accident black spot recognition threshold determined by the quality control method is as follows:

$$R_i^+ = \lambda_i + k \sqrt{\frac{\lambda_i}{m_i}} + \frac{1}{2m_i} \quad (11)$$

$$R_i^- = \lambda_i - k \sqrt{\frac{\lambda_i}{m_i}} - \frac{1}{2m_i} \quad (12)$$

where  $R_i^+$  represents the upper limit threshold of the road unit accident rate in the similar interval  $i$ ;  $R_i^-$  represents the lower limit threshold of the road unit accident rate in the similar interval  $i$ ;  $\lambda_i$  represents the empirical Bayesian posterior estimate of the accident rate in the similar interval  $i$ ;  $m_i$  represents the accumulated number of vehicles passed by road unit  $i$ ;  $k$  represents a constant, when taking a 95% confidence interval,  $k=1.96$ .

When the posterior estimate of the accident rate of the road unit of similar interval  $i$  is greater than  $R_i^+$ , the road unit is a dangerous road section, it is considered to be an accident black spot, and it should be diagnosed and treated with certain improvement measures. When the posterior estimate of the accident rate of the road unit of similar interval  $i$  is less than  $R_i^-$ , the road unit is a safe road section. The road unit between the two is a normal road section and it should be considered whether to take some improvement measures according to the actual needs.

### 3. A CASE STUDY OF IDENTIFYING ACCIDENT BLACK SPOTS

The Xibao Expressway is from Xi'an to Baoji via Xianyang (see Figure 5). The Xibao Expressway was converted from four lanes to eight lanes and opened



Figure 5 – Location of Xibao Expressway

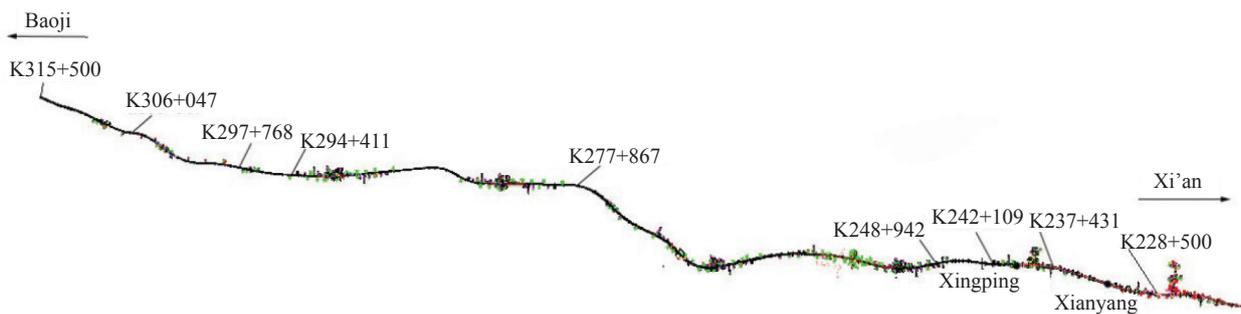


Figure 6 – Section K228+500-K315+500 of Xibao Expressway

to traffic in 2011. Since the direction from Xi'an to Baoji K1194+757-K1195+564 (length 807 meters) cannot solve the problem of insufficient horizontal clearance, the recommended speed of 100 km/h is used. The maximum speed limit of small passenger lanes on the remaining two-way eight-lane sections of the Xibao Expressway has been adjusted from 100 km/h to 120 km/h. In this thesis, an accident prediction model is established based on the entire Xibao Expressway accident data and road alignment data, and as a priori information, investigates a total of 87 km of data from K228+500-K315+500 sections of Xibao Expressway for accident black spot identification, as shown in Figure 6. As shown in the figure, the red dot area is a traffic accident-intensive spot. During the four years from October 2009 to October 2013, a total of 2091 traffic accidents occurred, of which 1938 were simple accidents and 153 were general accidents, causing 3 deaths and 50 injuries.

### 3.1 Road units division

The premise of identifying black spots in expressway accidents is to divide road sections. This paper uses the proposed road unit secondary division method to divide road units. As the

K228+500-K315+500 section of the Xibao Expressway is located in the Guanzhong Plain area, the terrain is flat, the longitudinal gradient is small and the longitudinal gradient of most sections is less than 3%. Large radius curve sections with little change in radius and slope less than 3% are merged into the same section. It can be divided into 8 segments in total, of which three segments are long straight lines and five segments are large radius curves. The results of the initial division are shown in Table 1.

In particular, according to the requirements of the quality control method to determine the threshold value of accident black spots and the initial division of road units, it can be divided into four similar intervals, which are long straight segments R1( K228+500-K237+431, K243+109-K248+942, K294+411-K297+768), large radius curve section C1( K237+431-K242+109), large radius curve section C2( K277+867-K294+411), large radius curve section C3( K248+942-K277+867, K297+768-K306+047, K306+047-K315+500).

Dynamic cluster analysis is used for these eight segments, and K-means clustering in SPSS19.0 can be used for stepwise cluster analysis. The number of accidents is converted into the number of equivalent accidents. The road unit is divided twice by

Table 1 – Linear element method division results

First division result	Longitudinal slope (average) [%]	Flat curve radius (average) [m]	Straight line length [m]
K228+500-K237+431	0.695	-	8931
K237+431-K242+109	0.488	6080	-
K243+109-K248+942	0.372	-	5833
K248+942-K277+867	0.393	4622	-
K277+867-K294+411	0.212	0.212	-
K294+411-K297+768	0.221	0.221	3357
K297+768-K306+047	0.357	0.357	-
K306+047-K315+500	0.345	0.345	-

dynamic cluster analysis, and the total number of road sections is determined by about 2 km per section. The results are shown in *Table 2*.

The proposed secondary division method was used to divide the road units of the K228+500-K315+500 section of the Xibao Expressway and division results were compared with

the actual conditions of the road. This method can divide accident-intensive points (entrances, exits, interlaced areas, etc.) and sections with roads with specific attributes (road alignment) into the same road unit, and the accidents are distributed uniformly in each road unit, which can be used as a basic evaluation unit for accident black spot recognition.

*Table 2 – Road unit secondary division results*

First division result	No.	Secondary division results	Equivalent accidents	Traffic flow [PCU/day]
K228+500-K237+431	1	K228+500-K229+511	3	25400
	2	K229+511-K231+423	68	25400
	3	K231+423-K234+543	74	25400
	4	K234+543-K237+431	87	25400
K237+431-K242+109	5	K237+753-K239+230	31	25400
	6	K239+230-K240+769	49	25400
	7	K240+769-K241+432	53	24500
	8	K241+432-K243+109	21	24500
K243+109-K248+942	9	K243+109-K243+954	24	24500
	10	K243+654-K246+331	140	24500
	11	K246+331-K248+942	71	24500
K248+942-K277+867	12	K248+942-K252+431	80	24500
	13	K252+431-K253+654	9	24500
	14	K253+654-K256+332	100	24500
	15	K256+332-K265+437	203	23700
	16	K265+437-K268+665	85	23700
	17	K268+665-K272+781	105	23700
	18	K272+781-K273+431	45	23700
	19	K273+431-K274+569	103	23700
K277+867-K294+411	20	K274+569-K277+867	87	23700
	21	K277+867-K280+331	81	23700
	22	K280+331-K282+424	49	22300
	23	K282+424-K286+306	187	22300
	24	K286+306-K289+341	52	22300
	25	K289+341-K290+432	290	22300
	26	K290+432-K292+533	35	22300
	27	K292+533-K294+411	30	22300
K294+411-K297+768	28	K294+411-K295+505	13	22300
	29	K295+505-K297+768	28	21400
K297+768-K306+047	30	K297+768-K300+511	14	21400
	31	K300+511-K301+543	42	21400
	32	K301+543-K302+525	18	21400
	33	K302+525-K306+047	97	21400
K306+047-K315+500	34	K306+047-K309+511	22	18500
	35	K309+511-K311+489	27	18500
	36	K311+489-K314+534	62	18500
	37	K314+534-K315+500	3	18500

### 3.2 Accident black spots identification

An accident prediction model is established based on the Xibao Expressway accident data and road alignment data, and used as a priori information. By selecting the basic road section under the ideal line condition of the Xibao Expressway, the data of the basic accident rate and the annual average daily traffic volume of the road section can be obtained. According to the foregoing relationship, the fitted curve can establish a model of the relationship between the basic equivalent accident rate ( $N_B$ ) of the road section and the annual average daily traffic volume of the road. Model calibration results are shown in Equation 13.

$$N_B = 3.35 \cdot 10^{-6} \cdot AADT^{1.4856} \quad (R^2 = 0.7851) \quad (13)$$

From the relationship between the basic accident rate of the road section and the traffic volume, it can be seen that if the road conditions are basically similar, the traffic volume of the road segment significantly affects the number of accidents, and the relationship between the two variables is a power function.

According to the determined correction coefficient determination method, the correction coefficient values under the conditions of longitudinal slope, flat curve radius, and straight-line length of each sample are determined separately. The fitted correction coefficients are:

$$\beta_i = 0.2014i^2 - 0.1211i + 1.0534 \quad (R^2 = 0.9263) \quad (14)$$

$$\beta_{RHC} - 1 = 96773RHC^{-1.752} \quad (R^2 = 0.8024) \quad (15)$$

$$\beta_{LT} = 0.1326LT^2 - 0.3791LT + 1.305 \quad (R^2 = 0.7885) \quad (16)$$

The accident prediction model established above was used to perform accident prediction analysis on the K228+500-K315+500 upstream of Xibao Expressway. The forecast results are shown in Figure 7.

According to the statistics of historical traffic data, the number of accidents is 45.08. If the influence of the profile is not considered, the average number of predicted accidents is 40.46, the relative error with historical accident statistics is -10.24%. However, the average number of accidents on the entire Xibao Expressway predicted by the above prediction model is 43.39. The relative error between the predicted value obtained by the accident prediction model and the historical accident statistical value is -3.69%, which improves the accuracy 6.55%. It can be concluded from this that the prediction model established in this paper has high accuracy and the model prediction results can better reflect the road safety situation.

Substituting the predicted data of each road unit into the empirical Bayesian posterior formula, the empirical Bayesian posterior estimated value of each road unit can be obtained, where  $Var(\Lambda)$  and  $E(\Lambda)$  are the mean and variance of the antecedent distribution, that is, the mean and variance of traffic accident rates for road unit samples used in modeling. It can be known from the calculation that  $E(\Lambda)$  is 3.67,  $Var(\Lambda)$  is 5.37, and  $\phi$  is 0.406, then the average Bayesian posterior estimates of equivalent accident rates for four similar intervals R1, C1, C2, and C3 are 41.82, 39.86, 49.94, and 43.54. Substituting them into Equation 11, the upper limit value of the critical ratio can be obtained, that is, the black point recognition threshold of each road unit accident. The obtained empirical Bayesian posterior estimation value is compared with the accident black spot recognition threshold, and the accident black spot is identified, as shown in Figure 8.

The results show that the empirical Bayesian posterior estimates of sections 9, 19, and 25 are 103.66, 123.92, and 109.34, respectively, which exceed their corresponding black spot recognition thresholds of 89.8, 93.29, and 104.63, so sections 9, 19, and 25 are accident black spot. Accident

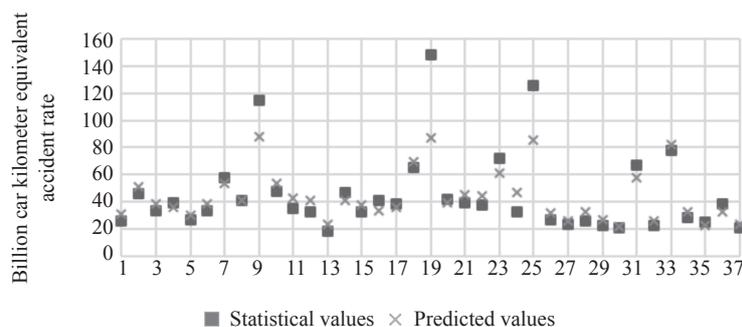


Figure 7 – Comparison of model predictions and accident statistics

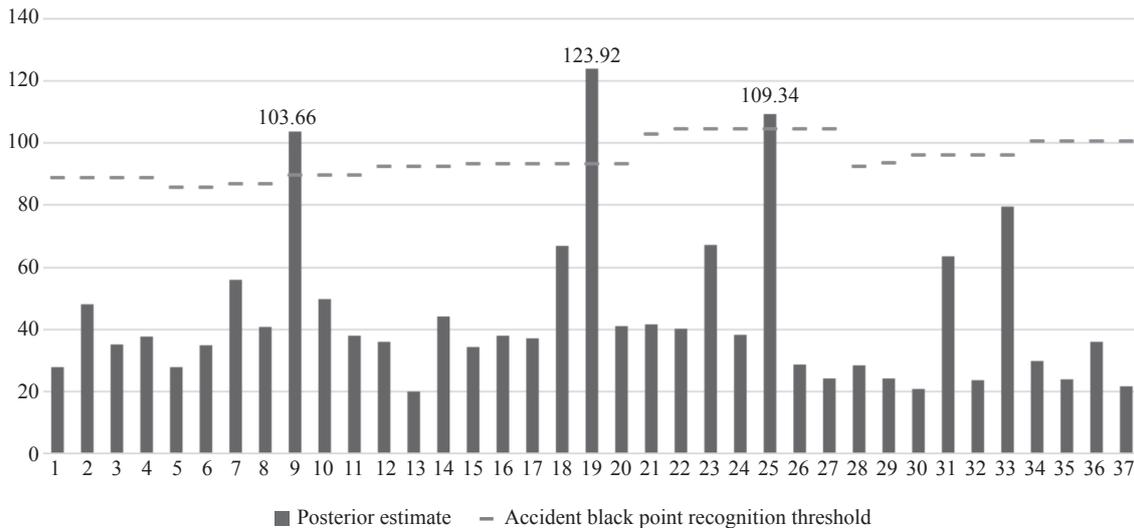


Figure 8 – Comparison of empirical Bayesian posterior estimate and accident black spot recognition threshold

black spot 19 is a large-radius curve segment with a curve radius of 3267 meters and a gradient slope of 0.428%; accident black spot 25 is a large-radius curve segment with a curve radius of 5446 meters and a longitudinal slope gradient of 0.335%; accident black spot 9 is a long straight line with a longitudinal slope of 0.438%. After identifying the black spots of highway accidents, targeted safety improvement measures can be taken for the accident black spots. Generally, a series of measures can be used to reduce the accident black spot hazards, such as setting up accident black spot warning signs, warning signs, speed limit signs, distance keeping signs, and sight guidance signs, or setting up safe lanes and adding thin layers of slip pavement, etc.

#### 4. CONCLUSIONS

In this paper, a series of research are conducted on the rational division of highway units, the establishment of accident prediction models, and the combination of empirical Bayes method and quality control method to identify black spots on highway accidents. Taking Xibao Expressway as an example, a case analysis was performed. The main research results obtained are as follows:

- 1) The road unit secondary division method can fully expose unsafe road sections, intuitively judge the safety performance of each road section, and provide a basic unit for further using the black spot recognition method to identify road sections;
- 2) Drawing on the accident forecasting ideas in HSM and increasing the consideration of the impact of longitudinal slopes on traffic accidents,

the accuracy of traffic accident prediction models is higher. This is conducive to the determination of the black point recognition threshold, identifies the black spots of expressway accidents more accurately, and the error is reduced by 6.55%.

- 3) Taking Xibao Expressway as a case, the proposed method is used to identify the accident black spots on the expressway. The results show that the empirical Bayesian posterior estimates of sections 9, 19, and 25 are 103.66, 123.92, and 109.34, respectively, which exceed their corresponding black spot recognition thresholds of 89.8, 93.29, and 104.63. Therefore, sections 9, 19, and 25 are considered accidents black spot.

The black spot identification method proposed in this paper can combine the safety evaluation of highway traffic accidents, road alignment safety evaluation, etc. to provide references for road planning, design, and operation. The scope of analysis in this article is relatively limited. The impact variables considered in the expressway accident prediction model are not comprehensive enough. Limited by the complexity of Bayesian model, human factors, road environment, weather conditions, vehicle performance, traffic composition, and traffic flow speed are not considered. At the same time, it should start from the mechanism of expressway traffic accidents and combine the safety evaluation of highway traffic accidents in advance with the road line safety evaluation, so as to study and establish the accident black spot identification method of potential expressway accidents.

## ACKNOWLEDGEMENTS

This work was supported by the Fundamental Research Funds for the Central Universities, CHD [no. 300102210204], the Shaanxi Province Natural Science Basic Research Project [no. 2020JM-222], and The National Natural Science Foundation of China [no. 52172338]. The authors would like to thank Yang Lu for his contributions to this paper.

梁国华<sup>1</sup>

Email: lgh@chd.edu.cn

孙旭姣 (通讯作者)<sup>1</sup>

E-mail: 2018321046@chd.edu.cn

张驿丹<sup>1</sup>

E-mail: 2018121259@chd.edu.cn

陈明丽<sup>1</sup>

E-mail: 2018221235@chd.edu.cn

张婉婷<sup>1</sup>

E-mail: 2018121244@chd.edu.cn

<sup>1</sup> 长安大学运输工程学院, 陕西 西安, 710064

## 基于道路单位二次划分的高速公路事故黑点识别研究

### 摘要

为了减少高速公路交通事故的危害, 提高交通事故黑点识别的准确性, 提出了一种基于道路单元二次划分和经验贝叶斯方法的高速公路事故黑点识别方法。基于HSM(高速公路安全手册)中高速公路事故预测模型的建模思想, 建立了高速公路事故预测模型, 并将其作为先验分布, 由先验分布结合经验贝叶斯法安全估计得到贝叶斯后验估计, 将后验估计值代入质量控制方法以获得黑点识别阈值。最后, 结合西安-宝鸡高速公路相关数据, 运用本文提出的方法, 对西宝高速公路进行了案例研究, 确定了西宝高速公路的第9、19、25段为黑点。结果表明, 基于动态聚类的二次分段方法可以客观地描述高速公路事故点的集中和分散, 提出的基于经验贝叶斯方法的黑点识别方法可以准确识别事故黑点。本文的研究结果可为高速公路管理部门的决策提供依据, 采取有针对性的安全改进措施。

### 关键词

交通安全; 事故黑点识别; 高速公路;  
道路单元划分; 贝叶斯经验方法

## REFERENCES

- [1] Statistics Bureau of the People's Republic of China. *China Statistical Yearbook*. Beijing: China Statistics Press; 2018.
- [2] Elvik R. A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis and Prevention*. 2008;40(6): 1830-1835. DOI: 10.1016/j.aap.2008.08.001
- [3] Zhang D. *Analysis of road traffic accidents and black spots*. Beijing: People's Communications Press; 2005.
- [4] Geng C, Peng Y. [A black spot identification method for traffic accidents based on dynamic segmentation and DBSCAN algorithm]. *长安大学学报(自然科学版)*. 2018;38(5): 131-138. Chinese.
- [5] Yakar F. Identification of accident-prone road sections by using relative frequency method. *Promet – Traffic&Transportation*. 2015;27(6): 539-547. DOI: 10.7307/ptt.v27i6.1609
- [6] Borsos A, Cafiso S, D'Agostino C, Miletics D. Comparison of Italian and Hungarian black spot ranking. *Proceedings of 6<sup>th</sup> Transportation Research Arena*; 2016.
- [7] Jordan P. ITE and Road Safety Audit – A partnership for traffic safety. *ITE Journal*. 1999;69(3): 24-27. DOI: 10.1109/25.752590
- [8] Li Q, Cheng C, Chen L. [Black point model of traffic accident based on GA-BP neural network algorithm and rough set theory]. *武汉理工大学学报(交通科学与工程版)*. 2011;35(4): 756-760. Chinese.
- [9] Guan M. [Comparative study on identification methods of traffic accident frequent point]. *公路*. 2009;2009(4): 191-195. Chinese.
- [10] Park B, Lord D, Lee C. Finite mixture modeling for vehicle crash data with application to hotspot identification. *Accident Analysis and Prevention*. 2014;71: 319-326. DOI: 10.1016/j.aap.2014.05.030
- [11] Elyasi MR, Saffarzade M, Boroujerdian AM. A novel dynamic segmentation model for identification and prioritization of black spots based on the pattern of potential for safety improvement. *Transportation Research Part A: Policy and Practice*. 2016;91: 346-357. DOI: 10.1016/j.tra.2016.06.020
- [12] Richard KR, Kim S, Ulfarsson GF. A hierarchical Bayesian logistic regression with a finite mixture for identifying higher-than-expected crash proportions at intersections. *Journal of Transportation Safety & Security*. 2019;11(1): 1-20. DOI: 10.1080/19439962.2017.1337054
- [13] Meng X, Li M, Mai Q, Guan Z. Research on identification of black spot and accident inducing factor for freeway. *Journal of Transportation Systems Engineering & Information Technology*. 2011;11(1): 114-120. DOI: 10.1186/s12889-016-2722-9
- [14] Niu Z, et al. [Identification of frequent occurrences of highway traffic accidents considering system clustering]. *中国安全科学学报*. 2018;28(11): 104-109. Chinese.
- [15] Jia H, Sang J, Yang L, Feng T. [Identification method of potential accident risk points (sections) for newly built expressways]. *北京工业大学学报*. 2016;42(8): 1233-1238. Chinese.
- [16] Ghadi M, Torok A. A comparative analysis of black spot identification methods and road accident segmentation methods. *Accident Analysis and Prevention*. 2019;128: 1-7. DOI: 10.1016/j.aap.2019.03.002
- [17] Ghadi M, Torok A, Tanczos K. Integration of probability and clustering based approaches in the field of black spot identification. *Periodica Polytechnica Civil Engineering*. 2019;63(1): 46-52. DOI: 10.3311/PPci.11753
- [18] Ahmed M, Huang H, Abdel-Aty M, Guevara B. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. *Accident Analysis and Prevention*. 2011;43(4): 1581-1589. DOI: 10.1016/j.aap.2011.03.021

- [19] Ana F, Jose N. An approach to accidents modeling based on compounds road environments. *Accident Analysis and Prevention*. 2013;53: 39-45. DOI: 10.1016/j.aap.2012.12.041
- [20] Uzundu C, Jamson S, Lai F. Investigating unsafe behaviours in traffic conflict situations: An observational study in Nigeria. *Journal of Traffic and Transportation Engineering (English Edition)*. 2019;6(5): 482-492. DOI: 10.1016/j.jtte.2018.06.002
- [21] Gregoriades A, Mouskos KC. Black spots identification through a Bayesian networks quantification of accident risk index. *Transportation Research Part C: Emerging Technologies*. 2013;28(3): 28-43. DOI: 10.1016/j.trc.2012.12.008
- [22] Sandhu HAS, Singh G, Sisodia MS, Chauhan R. Identification of black spots on highway with kernel density estimation method. *Journal of the Indian Society of Remote Sensing*. 2016;44(3): 457-464. DOI: 10.1007/s12524-015-0500-2
- [23] Colak HE, Memisoglu T, Erbas YS, Bediroglu S. Hot spot analysis based on network spatial weights to determine spatial statistics of traffic accidents in Rize, Turkey. *Arabian Journal of Geosciences*. 2018;11(151). DOI: 10.1007/s12517-018-3492-8
- [24] Washington S, Haque MM, Oh J, Lee D. Applying quantile regression for modeling equivalent property damage only crashes to identify accident black spots. *Accident Analysis & Prevention*. 2014;66(1): 136-146. DOI: 10.1016/j.aap.2014.01.007
- [25] Zhang C, Ivan JN. Effects of geometric characteristics on head-on crash incidence on two-lane roads in Connecticut. *Transportation Research Record*. 2005;1908(1): 159-164. DOI: 10.3141/1908-19
- [26] Ma Z, Zhang H, Zhang H, Wang J. [A model for predicting the number of roadside accidents on expressways]. *长安大学学报(自然科学版)*. 2017;37(4): 119-126. Chinese.
- [27] Ministry of Transport of the People's Republic of China. *Guidelines for Safety Evaluation of Highway Projects: JTG/TB05-2015*. Beijing: Ministry of Transport of the People's Republic of China; 2015.
- [28] AASHTO. *Highway Safety Manual (1<sup>st</sup> Edition)*. Washington D. C.: AASHTO; 2010.
- [29] Xie L, Wu C, Lyu N, Duan Z. Studying the effects of freeway alignment, traffic flow, and sign information on subjective driving workload and performance. *Advances in Mechanical Engineering*. 2019;11(5). DOI: 10.1177/1687814019853925
- [30] Gelman A, et al. *Bayesian Data Analysis*. 3<sup>rd</sup> ed. New York: Chapman and Hall/CRC; 2013.