


## Article

# Urban Road Accident Black Spot Identification and Classification Approach: A Novel Grey Verhul's–Empirical Bayesian Combination Method

Yan Wan <sup>1</sup>, Wenqiang He <sup>1</sup> and Jibiao Zhou <sup>1,2,\*</sup> 

<sup>1</sup> School of Civil and Transportation Engineering, Ningbo University of Technology, Ningbo 315211, China; wanyan@nbut.edu.cn (Y.W.); wenqianghe123@126.com (W.H.)

<sup>2</sup> College of Transportation Engineering, Tongji University, Shanghai 200092, China

\* Correspondence: zhoujb2014@nbut.edu.cn or zhoujibiao@tongji.edu.cn

**Abstract:** The identification and classification of accident black spots on urban roads is a key element of road safety research. To solve the problems caused by the randomness of accident occurrences and the unclear classification of accident black spots by the traditional model, we propose a method that can quickly identify and classify accident black spots on urban roads: a combined grey Verhul's–Empirical Bayesian method. The grey Verhul's model is used to obtain the predicted/expected numbers of accidents at accident hazard locations, and the empirical Bayesian approach is used to derive two accident black spot discriminators, a safety improvement space and a safety index (SI), and to classify the black spots into two, three, four and five levels according to the range of the SI. Finally, we validate this combined method on examples. High-quality and high-accuracy data are obtained from the accident collection records of the Ningbo Jiangbei District from March to December 2020, accounting for 90.55% of the actual police incidents during this period. The results show that the combined grey Verhul's–Empirical Bayesian method can identify accident black spots quickly and accurately due to the consideration of accident information from the same types of accident locations. The accident black point classification results show that the five-level rating of accident black points is most reasonable. Our study provides a new idea for accident black spot identification and a feasible method for accident black spot risk level classification.



**Citation:** Wan, Y.; He, W.; Zhou, J. Urban Road Accident Black Spot Identification and Classification Approach: A Novel Grey Verhul's–Empirical Bayesian Combination Method. *Sustainability* **2021**, *13*, 11198. <https://doi.org/10.3390/su132011198>

Academic Editors: Qiang Zeng, Pengpeng Xu, Feng Chen, Zhongxiang Feng and Marc A. Rosen

**Keywords:** traffic safety; potential for safety improvement; grey Verhul's–Empirical Bayesian method; safety index; accident black spots

Received: 16 August 2021

Accepted: 5 October 2021

Published: 11 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Road traffic safety is a world-recognized traffic problem, and according to the findings of the World Health Organization (WHO), road traffic accidents are the eighth leading cause of death worldwide. The Global Status Report on Road Safety 2018, launched by the WHO in December 2018, highlights that the number of annual road traffic deaths has reached 1.35 million. Road traffic injuries are now the leading killer of people aged 5 to 29 years. The burden is disproportionately borne by pedestrians, cyclists and motorcyclists, in particular those living in developing countries [1]. The annual report of road traffic accident statistics by the Traffic Management Bureau of the Ministry of Public Security of China shows [2] that 265,204 road traffic accidents occurred nationwide in 2019, resulting in 73,484 deaths, 304,919 injuries and USD 156.19 million in direct property losses. A reasonable assessment of road traffic safety is a necessary prerequisite and basic condition for reducing road traffic accidents and accident losses, so the study of accident black spot classification is of great significance. Accident black spots can be intersections, road sections and other areas and can refer to different contents according to the given research object. Studies have shown that accidents on road sections accounting for 0.25% of the total length of the road network account for 25% of the total number of accidents [3], and thus, black

spots compose a large proportion of accidents. Many countries [4] realize the seriousness of the problem and set up project research topics one after another, aiming at identifying the location of accident black spots on the road, proposing and determining hidden danger management measures to improve road traffic safety, of which the traffic accident black spot identification model is one of the core technologies.

The basis of accident black spot identification is accident prediction, and we refer the grey Verhul's method to the field of traffic safety to predict road traffic accidents. The reasons are as follows: First, the grey Verhul's method is one of the important models of grey system theory, which was first proposed internationally by Chinese scholar Deng Julong (1987) [5] in March 1982, and its principle is to establish a grey differential prediction model through a small amount of incomplete information to make a fuzzy long-term description of the development pattern of things, which is commonly used in the fields of population prediction, biological growth, reproduction prediction, product economic life prediction and other fields; second, in the road system as a dynamic system, the occurrence of road traffic accidents show non-smooth changes in the S-shaped development trend, and the grey Verhul's method is mainly used to describe the S-shaped process with a saturation state. Traffic accidents are random events and have discrete characteristics. The common methods for accident black spot identification in existing studies are the accident rate method [6], accident frequency method [7] and equivalent accident frequency method [8], and with the deepening of accident research, the cluster analysis method is also being applied more frequently in accident black spot identification. This type of method can carry out cause analyses for accident black spots in addition to black spot identification, and these methods more or less ignore the randomness and discrete nature of accident occurrences. Some scholars have also introduced the idea of utilizing the number of traffic conflicts [9] instead of the number of traffic accidents to identify accident black spots, but the applicability of this variable in cases with large sample sizes of accident data and the correlation between the types of traffic conflicts and the types of traffic accidents [10,11] still need to be further studied. The empirical Bayesian method is introduced to achieve the purpose of accident black spot identification in this paper. The empirical Bayesian method is a simplification of the Bayesian method, which originates from the Bayesian statistical theory proposed by the British scholar Bayes in 1763 [12]. First, empirical Bayesian methods are usually used to implement the process of estimating prior distributions and determining Bayesian decision functions using historical data when the prior distribution is unknown and are widely used in computer fields such as natural language processing, machine learning, recommender systems, image recognition and game theory. Secondly, the empirical Bayesian approach can overcome the characteristics of the randomness of accident occurrence by considering the mean value of accidents of the same type of accident objects, which also makes it gradually show its advantages in accident black spot identification. In view of the above research, in this paper, a combined grey Verhul's–Empirical Bayesian approach based on an urban road accident black spot discrimination model is proposed. The predicted/expected number of hidden accident locations is obtained by the grey Verhul's accident prediction model, which is substituted into the empirical Bayesian method to derive two accident black spot discriminatory indices, the safety improvement space and the safety index (SI), and then the advantages and disadvantages of various accident black spot evaluation methods are discussed according to the value interval of the SI, which provides a new idea for accident black spot identification and classification.

To this end, this paper aims to concentrate on the problems of accident black spot identification and classification. Our work makes the following contributions. (a) The accident identification and classification part of this paper considers highly discrete parameters, thereby effectively avoiding the randomness and discrete nature of traffic accidents and is based on a grey Verhul's–Empirical Bayesian combination model that is easy to calculate in terms of accident prediction. This part also incorporates two indices, the safety-improvement space and the SI, into the model to prevent the situation where the hazard levels of accident locations with the same number of accidents cannot be discriminated.

(b) We discuss the advantages and disadvantages of the 2-, 3-, 4- and 5-level accident black spot rating methods based on the value range of the SI and determine a reasonable accident black spot rating method to fill the current research gap with no accident black spot grading system in the current study, thus providing a sufficient theoretical basis for managers to carry out accident black spot remediation.

The remainder of this paper is organized as follows. Section 2 presents a literature review of the existing studies focusing on accident prediction models and accident black spot identification models. Section 3 formulates the proposed model algorithm and classification process. Section 4 illustrates the proposed model by a case study and analyzes the results. Conclusions and future work are discussed in Section 5.

## 2. Literature Review

Among the common accident black spot identification methods, the accident rate method [13], accident frequency method [14] and the equivalent accident frequency method [15] are widely used. For example, Danielsson (1986) [16] studied the problem of estimating in a non-experimental before-and-after investigation the effect of a countermeasure on the number of traffic accidents at road junctions. These methods are also used as the basic methods in research related to accident black spot identification, and they are improved or combined with other methods to increase the identification efficiency and accuracy. The above methods are simple, straightforward and easy to calculate, but they also have limitations. For example, two locations with the same number of accidents do not necessarily have the same hazard level, but the accident frequency method treats these locations as the same. The accident rate method is biased when applied to roads with low traffic volumes, and the method assumes a linear relationship between accident frequency and traffic volume, an assumption that may not be consistent with the actual situation. As research has progressed, cluster analysis methods such as the K-means algorithm, the density-based clustering algorithm, grey correlation analysis and the fuzzy clustering method have been widely used. Guo et al. (2018) [17] established an improved K-means clustering algorithm to eliminate the influence of isolated points in the obtained clustering results to identify and analyze the traffic accident black spots within a study area. Geng et al. (2018) [18] obtained preliminary black spot road sections by the cumulative frequency method and the road section adjacency principle and used the density-based spatial clustering of applications with noise (DBSCAN) algorithm for a clustering analysis to identify real accident black spots with short lengths and high accident concentrations. Their research results showed that the algorithm could eliminate non-black sections in preliminary road sections and identify real black spot road sections. Wang et al. (2016) [19] proposed an accident black spot identification method based on a grey clustering evaluation and combined their approach with the clustering factor method to identify accident black spot trigger factors; the results showed that the accident black spots obtained based on a clustering analysis and the actual road accident black spots basically matched. Lin et al. (2010) [20] analyzed traffic accident black spot identification methods and their adaptability based on the screening of unfavorable factors in highway traffic and found that the main factor of highway accident black spots is the environmental interference around the national highway and that the secondary factor is the failure to set road lighting facilities. Yetis et al. (2017) [21] used the fuzzy clustering method to analyze the traffic accidents data of Denizli city and reported a series of suggestions to improve traffic safety. All the above methods have certain levels of applicability and limitations. Although theoretically the accident black spots identified in the same category are considered to present the same degree of danger, in practice, the degree of danger of each accident black spot may be different.

Statistical methods that have been applied to accident black spot identification include the Poisson distribution, negative binomial distribution, Poisson-log-normal distribution, random effects model considering heterogeneity, Markov transformation model, etc. Cui et al. [22] verified that the number of accidents per unit mile obeyed a Poisson distri-

bution based on a statistical analysis of traffic accident data and derived the bound value of accident spacing obeying a negative exponential distribution and abnormal accident spacing. Amoros et al. (2003) [23] compared traffic safety conditions among several counties in France by means of a negative binomial distribution for accidents with injuries and explored whether differences in safety between them could be explained by differences in the distribution of road types and differences in socioeconomic characteristics. The study found that the time trends (incidence and severity) were the same for each indicator across counties and road types. Lord (2008) et al. [24] used a Poisson-log-normal distribution model to predict motor vehicle crashes on highways and tested existing models with and without fuzzy statistical indicators. Barua et al. (2015) [25] applied a random effects model considering heterogeneity to study the inclusion of spatial autocorrelation in crashes in the cities of Richmond and Vancouver and showed that the fit did not change with the inclusion of spatial correlation and that the method had the potential to estimate safety effects more accurately under limited data conditions. Malyshkina et al. (2010) [26] proposed a two-state Markov-transformed count data model as an alternative to the zero-inflated model to predict the number of accidents occurring on roadways. However, these methods often fail to perform analyses when “poor”-quality samples are drawn.

Maen et al. (2019) [27] evaluated four common accident black spot identification methods (empirical Bayes, excess empirical Bayes, accident frequency and accident rate). In general, the empirical Bayesian method is better than other methods. The empirical Bayesian method is a simplification of the traditional Bayesian method that can overcome the influence caused by the random fluctuations of accidents because it considers the accident information of the same type of object and provides unbiased accident estimations [28]. Through a comparison of the empirical Bayesian method with several other types of accident black point discrimination methods, it was found that the consistency test results of the empirical Bayesian method were better than those of other methods [29]. In addition, empirical Bayesian methods have some advantages in terms of road accident risk identification and ranking. For example, Miaou et al. (2005) [30] statistically ranked the accident risks of urban road intersections and general rural roads by Bayesian methods. Wen et al. (2005) [31] compared the advantages and disadvantages of simple ranking, confidence interval and empirical Bayesian methods with respect to accident black spot identification, and the results showed that empirical Bayesian methods were significantly better than simple ranking and confidence interval methods. It is important to note that accident number prediction is the basis and necessary condition for accident black spot identification. Wang et al. (2006) [32] analyzed the characteristics of the Gauss-Markov (GM) (1,1) model and grey Verhulst model in road accident prediction and found that the GM (1,1) model is applicable to sequences with strong exponential laws, which can only describe monotonic change processes, while the Verhulst model is applicable to non-monotonic oscillating development sequences or S-shaped sequences with saturation states. For domestic road traffic accidents occurring in recent years, the grey Verhulst model can better predict the number of road traffic accidents and provide accurate data support for the identification of accident black spots. Based on this, this paper discriminates urban road accident black spots by a combined grey Verhulst–Empirical Bayesian model.

Despite the wide range of studies in this realm, the use of accident black spot identification methods that address the stochastic nature of accident occurrences or the ranking of accident black spot identification results is rare in the literature. To this end, this work aims to develop a state-of-the-art method framework to fill the abovementioned gaps.

### 3. Method

In this study, a model for identifying accident black spots on urban roads is proposed. The model combines the grey Verhulst method and the empirical Bayesian method, takes the numbers of accident occurrences of urban road hidden accident locations as the original data, obtains the expected numbers of accident occurrences at the hidden accident locations after the calculation of the combined model and uses the calculation results of the safety

improvement space and SI as the discriminating indices for accident black spots to discern whether the hidden accident locations are accident black spots. Finally, the levels of the detected black spots are evaluated.

### 3.1. Accident Black Spot Identification Model Construction

In practical problems, it is often seen that the original dataset itself is an oscillating series; for testing with such a data series, the grey Verhuls model can be utilized.

Suppose that  $x^{(0)}$  is the original data sequence of an accident with:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \tag{1}$$

$x^{(1)}$  is a single accumulation of  $x^{(0)}$  for generating the accumulated generation operation (1-AGO) sequence, which can be expressed as:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \tag{2}$$

After using Equation (2) to generate a sequence of immediately adjacent mean generators,  $z^{(1)}$  is the sequence of immediately adjacent mean generators of  $x^{(1)}$  and is expressed as:

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)) \tag{3}$$

$$z^{(1)}(k) = 0.5(x^{(1)}(k) + x^{(1)}(k - 1)), k = 2, 3, \dots, n \tag{4}$$

Then,

$$x^{(0)} + \alpha z^{(1)} = \beta (z^{(1)})^2 \tag{5}$$

Equation (5) is the grey Verhuls model, where  $\alpha$  and  $\beta$  are parameters known as development coefficients and are referred to as grey action quantities.

This creates a whitening differential equation for the Verhuls model, which is expressed as:

$$\frac{dx^{(1)}(t)}{dt} + \alpha x^{(1)}(t) = \beta (x^{(1)}(t))^2 \tag{6}$$

In the above equation, the meanings of  $\alpha$  and  $\beta$  are the same as those in Equation (5), and  $t$  is the time.

The grey Verhuls model parameter columns are expressed as:

$$u = (\alpha, \beta)^T = (B^T B)^{-1} B^T Y \tag{7}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & z^{(1)}(n)^2 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{8}$$

Taking  $x_0^{(1)}$  to solve the above differential Equation (6) yields the following equation for the time response function of the grey Verhuls model:

$$x^{(1)}(t) = \frac{ax_0^{(1)}}{bx_0^{(1)} + (a - bx_0^{(1)})e^{at}} \tag{9}$$

It should be noted that the time response function is a necessary process to obtain the grey prediction of the number of accidents at the location of accident hazards, with the implication that the number of accidents is assumed to vary with time.

From Equation (9), the time response sequence of the grey Verhul's model is given by the following equation:

$$\hat{x}^{(1)}(k+1) = \frac{ax^{(1)}}{bx_0^{(1)} + (a - bx_0^{(1)})e^{ak}} \quad (10)$$

Let  $x_0^{(1)} = x^{(0)}(1)$ ; then, Equation (9) becomes the following equation:

$$\hat{x}^{(1)}(k+1) = \frac{ax^{(0)}(1)}{bx^{(0)}(1) + (a - bx^{(0)}(1))e^{ak}} \quad (11)$$

Then, the cumulative reduction of the above equation yields the following equation:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (12)$$

The empirical Bayesian method is used to obtain the expected number of accidents at an urban road accident hazard location by setting the weight values and considering the average number of accidents occurring at the location and the predicted number of accidents with the same type of accident hazard location; the resulting equation has the following form:

$$Y(r|x) = Y(r) \cdot w + x \cdot (1 - w) \quad (13)$$

$$w = \left(1 + \frac{Y(r)}{k}\right)^{-1} \quad (14)$$

where  $Y(r|x)$  is the expected number of accidents at the accident hazard location with a known number of observed accidents;  $Y(r)$  is the average of the predicted number of accidents at the same type of accident hazard location;  $x$  is the observed number of accidents;  $w$  is the weight; and  $k$  is the over dispersion parameter.

It should be noted that the "same type" of potential accident location refers to the underlying spatial conditions, and the traffic flow conditions are similar for these accident potential locations. In the urban road network, the spatial conditions of the road are the main factor affecting the occurrence of accidents, such as the mismatch in the number of road lanes connecting the front and rear, the narrow lane width and so on. For two of the same types of potential accident roadways, for example, the numbers of motorized lanes and non-motorized lanes in both directions need to be the same, and the widths of the road sections are the same to ensure that the peak traffic flows of the roadways are similar; if these conditions hold, the two roadways have the same type.

*PSI* refers to the portion of the accident expectation of a hidden location that exceeds the average accident expectation across similar hidden locations [33–35]. The safety improvement space takes the average accident values of similar objects into account and is more suitable as an indicator for determining accident black spots [36–38]. For hidden accident locations, the difference between the accident expectation estimated by the empirical Bayesian method and the average accident expectation of similar hidden accident locations can be used to express the safety improvement space of the location, which is calculated as follows:

$$PSI = Y(r|x) - Y(r) \quad (15)$$

where  $Y(r|x)$  and  $Y(r)$  have the same meanings as above.

In this paper, we define the concept of the *SI* to establish a discriminatory index for accident black spots. The *SI* is the ratio of the degree of the safety improvement space of the accident hazard location to the open square of the sum of the variances and has the following form:

$$SI = \frac{PSI}{\sqrt{\sigma_{r|x}^2 + \sigma_r^2}} \quad (16)$$

$$\sigma_{r|x}^2 = (1 - w) \cdot Y(r|x) \tag{17}$$

$$\sigma_r^2 = \frac{Y^2(r)}{kn_0} \tag{18}$$

where  $\sigma_{r|x}^2$  is the variance of  $Y(r|x)$ ,  $\sigma_r^2$  is the variance of  $Y(r)$ , and  $n_0$  is the observed sample size of the model.

### 3.2. Accuracy Test

According to the above derivation process, in the original series and grey model prediction series,  $x^{(0)}$  and  $\hat{x}^{(0)}$ , respectively, and the residual series can be expressed as the following equation:

$$\varepsilon^{(0)} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = (x^{(0)}(1) - \hat{x}^{(0)}(1), \dots, x^{(0)}(n) - \hat{x}^{(0)}(n)) \tag{19}$$

The corresponding relative error series can be expressed by the following equation:

$$\Delta = (\Delta_1, \Delta_2, \dots, \Delta_n) = \left( \left| \frac{\varepsilon_1}{x^{(0)}(1)} \right|, \left| \frac{\varepsilon_2}{x^{(0)}(2)} \right|, \dots, \left| \frac{\varepsilon_n}{x^{(0)}(n)} \right| \right) \tag{20}$$

Then, for  $k \leq n$ , the simulated relative error and the average relative error at point  $k$  are expressed as the following equations:

$$\Delta_k = \left| \frac{\varepsilon_k}{x^{(0)}(k)} \right| \tag{21}$$

$$\bar{\Delta} = \frac{1}{n} \sum_{k=1}^n \Delta_k \tag{22}$$

Additionally, it is assumed that the system behaves as a sequence  $x_i = (x_i(1), x_i(2), \dots, x_i(n))$ ,  $D$  is a sequence operator and  $x_i D = (x_i(1)d, x_i(2)d, \dots, x_i(n)d)$ , where  $x_i(k)d = x_i(k) - x_i(1)$ ; then,  $D$  is called the initial zeroing operator, and  $x_i D$  is called the initial zeroing image of  $x_i$ :

$$x_i D = x_i^0 = (x_i^0(1)d, x_i^0(2), \dots, x_i^0(n)) \tag{23}$$

Assuming again that the sequences  $x_0$  and  $x_i$  are of the same length, the absolute correlation between them is expressed as:

$$\varepsilon_{0i} = \frac{1 + |s_0| + |s_i|}{1 + |s_0| + |s_i| + |s_i - s_0|} \tag{24}$$

where 1 and 2 are derived from the following three equations:

$$|s_0| = \left| \sum_{k=2}^{n-1} x_0^0(k) + \frac{1}{2} x_0^0(n) \right| \tag{25}$$

$$|s_i| = \left| \sum_{k=2}^{n-1} x_i^0(k) + \frac{1}{2} x_i^0(n) \right| \tag{26}$$

$$|s_i - s_0| = \left| \sum_{k=2}^{n-1} (x_i^0(k) - x_0^0(k)) + \frac{1}{2} (x_i^0(n) - x_0^0(n)) \right| \tag{27}$$

Suppose that  $x^{(0)}$  and  $\hat{x}^{(0)}$  denote the original sequence and the predicted sequence of the grey Verhul's model for the response, respectively, and that  $\varepsilon^{(0)}$  is the residual sequence. Then, the mean and variance of  $x^{(0)}$  are expressed as the following two equations:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k) \quad (28)$$

$$S_1^2 = \frac{1}{n} \sum_{k=1}^n (x^{(0)}(k) - \bar{x})^2 \quad (29)$$

The mean and variance of the corresponding 1 are expressed as the following two equations:

$$\bar{\varepsilon} = \frac{1}{n} \sum_{k=1}^n \varepsilon_k \quad (30)$$

$$S_2^2 = \frac{1}{n} \sum_{k=1}^n (\varepsilon_k - \bar{\varepsilon})^2 \quad (31)$$

The mean–variance ratio is solved by the following equation:

$$C = \frac{S_2}{S_1} \quad (32)$$

In this paper, the accuracy of the prediction results of the grey Verhul's model determines the accuracy of the accident black spot identification model. An accuracy check of the prediction results is carried out with reference to Table 1.

**Table 1.** Accuracy test level of the grey Verhul's model.

Level	Index		
	Average Relative Error	Absolute Correlation Degree	Mean Squared Error Ratio
I	0.01	0.90	0.35
II	0.05	0.80	0.50
III	0.10	0.70	0.65
IV	0.20	0.60	0.80

### 3.3. Accident Black Spot Level Classification

From Equations (15) and (16), if the SI is positive, this means that the expected number of accidents at the hidden location is greater than the average number of accidents at the same type of location and that the examined location is relatively dangerous. The greater the SI is, the higher the degree of danger at the hidden location. If the SI is negative, the expected number of accidents at the location is less than the average number of accidents at the same type of location, and the location can be considered relatively safe. Therefore, in this paper, the principle of determining whether a hidden accident location is an accident black spot is as follows: when both the PSI and SI are greater than 0, the hidden accident location is determined as an accident black spot.

According to Equation (18), if the weight is less than 1, then the positive and negative values are determined by the corresponding size relationships. According to Equations (16)–(18), it is known that the numerator term of the value formula is the difference between values, and the denominator term is the open square of the sum of the two variances, which take values within the closed interval  $[-1, 1]$ . According to the principal accident black point determination, when  $SI \leq 0$ , the hidden accident position is not an accident black point; when  $SI > 0$ , the hidden accident position is an accident black point. To facilitate the classification of the degrees of risk of accident black spots when  $SI > 0$ , this paper classifies the SI intervals according to the secondary, tertiary, quaternary and quintuple division methods to characterize the accident black spots. The specific classification results are



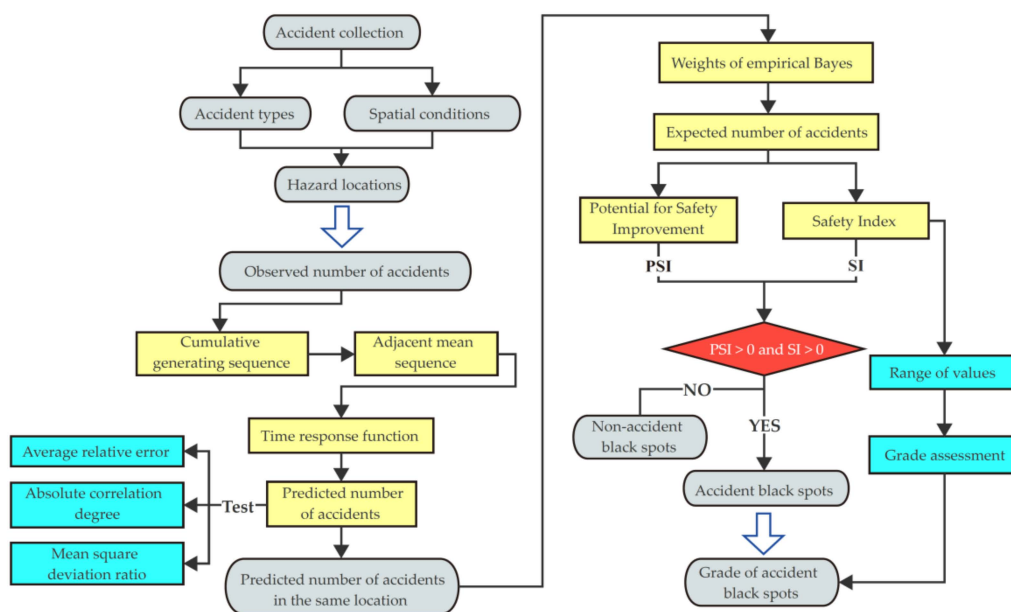
shown in Table 2, and the advantages and disadvantages of the four classification methods are discussed in Section 4.3.

**Table 2.** Assessment level classification of the SI.

Assessment Level	Non-Black Spots	I	II	III	IV	V
Two-level	$[-1, 0]$	$(0, 0.5]$	$(0.5, 1.0]$	-	-	-
Three-level	$[-1, 0]$	$(0, 0.33]$	$(0.33, 0.66]$	$(0.66, 1.0]$	-	-
Four-level	$[-1, 0]$	$(0, 0.25]$	$(0.25, 0.5]$	$(0.5, 0.75]$	$(0.75, 1.0]$	-
Five-level	$[-1, 0]$	$(0, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, 1.0]$

3.4. Accident Black Spot Identification Process

In this study, an accident black spot identification model is constructed based on the combined grey Verhul’s–Empirical Bayesian method. Step one: the accident collection records are analyzed so as to obtain the ranking of accident hazards locations based on the number of occurrences, and the number of occurrences of accidents at accident hazard locations is used as the raw data for accident prediction. Step two: the number of accidents is predicted by the grey Verhul’s method, and the time response function of the number of accidents is obtained. To verify the accuracy of the prediction model, three parameters, i.e., the mean relative error, absolute correlation and mean square error ratio, are calculated by the original data series and the predicted data series for the prediction accuracy test. Step three: the predicted data series is substituted into the empirical Bayesian model to calculate the weights of the accident hazard locations and the expected number of accidents, and the two introduced indicators (the safety improvement space and the SI) are used to discriminate whether each accident location is an accident black spot. Step four: the range of the SI is discussed; the accident black spots are classified into two, three, four and five levels, and the identified accident black spots are graded, see Figure 1. In traditional accident black spot identification approaches, accident black spots are often identified according to statistical accident quantities, but it should be noted that accident black spot identification is the final purpose of accident quantity prediction, and accident quantity prediction is equivalent to the necessary condition of accident black spot identification, which are closely related to each other.



**Figure 1.** Flow chart of the solution process.

The combined model for accident black spot identification in this paper has the following advantages. First, it can overcome the influence caused by random accident fluctuations because it considers accident information from the same type of accident location; second, the model can be analyzed even when the quality of the input accident data records is poor; third, it can calculate the hazard level, i.e., the accident black spot level, corresponding to a certain number of locations with the same number of accidents.

#### 4. Case Study

The road traffic accident data were collected from the Jiangbei District, Ningbo, China. We obtained accident data from the beginning of March until the end of December from the Jiangbei District through the cell phone application of the Jiangbei District Traffic Accident Collection Platform, which was jointly developed with the Jiangbei Traffic Police Brigade of the Ningbo City Traffic Police Department.

It should be noted that the initial data we used for the study would only have 10 months of accident collection records, so to expand the sample size, we validated the model by the number of accidents per half month. Also, our identification of each accident black spot is based on its accident data in the previous month, which is the current common way of accident black spot identification in most cities in China. In addition, this accident collection platform cannot collect other information that affects the probability of accident occurrence, such as the geometric conditions of the accident location and the daily traffic volume. We will also discuss some of these issues later in the paper.

##### 4.1. Model Calculation Process

We obtained the top 10 most accident-hazardous road sections, ranked by the number of accidents according to an analysis, see Table 3.

**Table 3.** The order of the hidden danger road sections in the Jiangbei District.

Sorting	Intersection	Number of Accidents
1	Ningci West Road of Cicheng Town (Nc)	531
2	Century Avenue of Xinming Street (Sj <sup>1</sup> )	440
3	Huancheng Road of Yongjiang Street (Hc)	422
4	North Ring West Road of ZhuangQiao Street (NR <sup>1</sup> )	382
5	Dongchang Road of Yongjiang Street (Dc)	346
6	Century Avenue of Kongpu Street (Sj <sup>2</sup> )	336
7	Qinglinwan bridge of Hongtang Street (Qlw)	326
8	North Ring West Road of Hongtang Street (NR <sup>2</sup> )	323
9	Kangzhuang South Road of Zhuangqiao Street (Kz)	295
10	G228 of Cicheng Town (G228)	291

Note, Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

Since the initial data contained only 10 months of accident collection records, the amount of accident data was limited, and to expand the sample size, the dataset was divided into half-month accident data for fitting predictions, and the number of accident data observations is shown in Table 4.

**Table 4.** Half-monthly traffic accident observations on accident-prone road sections.

Period	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
1 March	12	5	6	4	3	4	5	6	4	3
2 March	17	7	7	5	4	5	6	7	5	4
1 April	21	9	10	7	6	7	8	9	7	6
2 April	24	11	13	9	9	8	9	9	7	7
1 May	26	16	16	11	13	12	12	11	9	9
2 May	27	17	17	15	14	14	13	12	10	11

Table 4. Cont.

Period	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
1 June	29	21	19	17	18	17	15	14	13	14
2 June	28	22	22	17	19	17	15	15	14	15
1 July	30	23	21	19	18	19	18	17	16	17
2 July	30	25	24	22	21	18	19	18	17	17
1 August	28	26	24	22	21	21	19	19	17	18
2 August	29	27	25	24	22	21	20	19	18	19
1 September	29	28	27	25	21	21	21	21	20	18
2 September	28	28	27	26	23	22	20	20	19	19
1 October	29	29	28	26	23	20	19	22	20	19
2 October	28	27	27	25	22	23	21	19	18	18
1 November	30	30	27	27	22	21	21	21	21	20
2 November	28	29	26	28	21	21	23	21	20	19
1 December	28	30	29	27	24	23	20	23	21	18
2 December	30	30	27	26	22	22	22	20	19	20

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

The predicted numbers of traffic accidents per half month on the accident-hazardous roadways were obtained according to the grey Verhulst model, and the results are shown in Table 5.

Table 5. The predicted numbers of semi-monthly traffic accidents on accident-prone roads.

Period	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
1 March	12.00	5.00	6.00	4.00	3.00	4.00	5.00	6.00	4.00	3.00
2 March	16.52	6.84	7.76	5.27	4.54	5.54	6.43	7.23	5.15	4.26
1 April	20.67	9.11	9.80	6.82	6.61	7.43	8.06	8.58	6.49	5.86
2 April	23.86	11.71	12.06	8.66	9.14	9.60	9.83	10.00	8.00	7.75
1 May	25.99	14.51	14.42	10.74	11.90	11.88	11.64	11.47	9.63	9.81
2 May	27.29	17.30	16.75	12.98	14.59	14.10	13.39	12.92	11.28	11.86
1 June	28.03	19.87	18.93	15.26	16.91	16.08	14.99	14.30	12.88	13.73
2 June	28.45	22.07	20.86	17.47	18.72	17.73	16.38	15.58	14.36	15.29
1 July	28.67	23.86	22.48	19.50	20.02	19.01	17.53	16.73	15.65	16.51
2 July	28.79	25.23	23.81	21.27	20.91	19.97	18.45	17.73	16.74	17.42
1 August	28.86	26.25	24.84	22.76	21.49	20.66	19.16	18.58	17.64	18.06
2 August	28.89	26.98	25.64	23.97	21.85	21.14	19.71	19.29	18.34	18.50
1 September	28.91	27.50	26.23	24.91	22.08	21.47	20.11	19.88	18.89	18.81
2 September	28.92	27.86	26.68	25.64	22.22	21.69	20.40	20.35	19.31	19.01
1 October	28.92	28.11	27.00	26.19	22.31	21.84	20.62	20.72	19.62	19.14
2 October	28.93	28.28	27.23	26.60	22.36	21.94	20.77	21.02	19.85	19.23
1 November	28.93	28.39	27.40	26.91	22.40	22.01	20.89	21.26	20.02	19.28
2 November	28.93	28.47	27.52	27.13	22.42	22.05	20.97	21.44	20.15	19.32
1 December	28.93	28.52	27.61	27.29	22.43	22.08	21.02	21.58	20.24	19.35
2 December	28.93	28.56	27.67	27.41	22.44	22.10	21.06	21.69	20.31	19.36

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

The accident number prediction results were compared with the original accident data to obtain the average relative error between them; the absolute correlation was obtained by calculation according to Equations (23)–(27); the mean squared error ratio was calculated according to Equations (28)–(32). Finally, the prediction accuracy level test was conducted with reference to Table 1, and the results are shown in Table 6.

**Table 6.** Traffic accident prediction accuracy test for accident-prone road sections.

Accident-Prone Road	Average Relative Error		Absolute Correlation Degree		Mean Squared Error Ratio	
	Value	Level	Value	Level	Value	Level
Nc	2.19%	II	0.9984	I	0.1311	I
Sj <sup>1</sup>	2.97%	II	0.9984	I	0.0724	I
Hc	3.70%	II	0.9943	I	0.1109	I
NR <sup>1</sup>	3.44%	II	0.9914	I	0.0957	I
Dc	4.52%	II	0.9976	I	0.1187	I
Sj <sup>2</sup>	4.67%	II	0.9888	I	0.1118	I
Qlw	3.74%	II	0.9936	I	0.1055	I
NR <sup>2</sup>	4.27%	II	0.9901	I	0.1107	I
KZ	4.61%	II	0.9888	I	0.1084	I
G228	3.83%	II	0.9905	I	0.0757	I

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

Next, the accident black spot identification was performed. According to the calculation process of the empirical Bayesian method described in this paper, the expected numbers of accidents for the same types of roads were obtained first. It should be noted that accident black spot identification was performed on the basis of the monthly accident counts, and the accident prediction results for each half month listed in the previous paper needed to be summed. The expected accident counts of the same type of location for the 10 accident-hazardous road sections are shown in Table 7.

**Table 7.** Expected numbers of accidents at the same types of location on accident-prone road sections.

Period	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
March	21	11	14	11	12	10	12	15	10	11
April	28	18	26	13	13	11	12	17	17	14
May	29	21	27	17	16	15	15	21	22	16
June	31	28	28	19	21	18	19	21	24	18
July	31	28	24	22	20	20	17	26	21	16
August	34	37	31	23	23	23	21	29	27	23
September	39	36	34	28	21	22	26	24	25	20
October	37	39	36	36	28	28	29	28	27	26
November	40	38	33	32	25	23	31	25	24	23
December	42	41	37	31	23	25	28	26	24	24

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

The excessive dispersion parameters were obtained by the degree of dispersion of the original data from the predicted data. The excessive dispersion parameters were different for each hidden road section, and the results are shown in Table 8.

**Table 8.** Table of excessive discrete parameters for accident-prone road sections.

Road Name	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
k	1.12	1.55	1.75	1.40	1.86	2.33	2.15	2.80	1.77	2.00

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

The weights  $w$  of the empirical Bayesian formula were obtained by substituting the  $Y(r)$  and  $k$  values into Equation (14), and the results are shown in Table 9.

**Table 9.** Table of empirical Bayesian weights for accident-prone road sections.

Period	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
March	0.0506	0.1235	0.1111	0.1129	0.1342	0.1890	0.1519	0.1573	0.1504	0.1538
April	0.0385	0.0793	0.0631	0.0972	0.1252	0.1748	0.1519	0.1414	0.0943	0.1250
May	0.0372	0.0687	0.0609	0.0761	0.1041	0.1344	0.1254	0.1176	0.0745	0.1111
June	0.0349	0.0525	0.0588	0.0686	0.0814	0.1146	0.1017	0.1176	0.0687	0.1000
July	0.0349	0.0525	0.0680	0.0598	0.0851	0.1043	0.1123	0.0972	0.0777	0.1111
August	0.0319	0.0402	0.0534	0.0574	0.0748	0.0920	0.0929	0.0881	0.0615	0.0800
September	0.0279	0.0413	0.0490	0.0476	0.0814	0.0958	0.0764	0.1045	0.0661	0.0909
October	0.0294	0.0382	0.0464	0.0374	0.0623	0.0768	0.0690	0.0909	0.0615	0.0714
November	0.0272	0.0392	0.0504	0.0419	0.0692	0.0920	0.0649	0.1007	0.0687	0.0800
December	0.0260	0.0364	0.0452	0.0432	0.0748	0.0853	0.0713	0.0972	0.0687	0.0769

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

Upon employing Equation (23) to calculate the expected numbers of accidents on hazardous accident roads, the results are shown in Table 10.

**Table 10.** Expected numbers of accidents on accident-prone road sections.

Period	Nc	Sj <sup>1</sup>	Hc	NR <sup>1</sup>	Dc	Sj <sup>2</sup>	Qlw	NR <sup>2</sup>	KZ	G228
March	28.59	11.88	13.11	9.23	7.67	9.19	11.15	13.31	9.15	7.62
April	44.35	19.84	23.19	15.71	14.75	14.30	16.24	17.86	14.28	13.13
May	52.11	32.18	32.63	25.32	25.85	24.52	23.75	22.76	19.22	19.56
June	56.09	42.21	40.24	32.97	35.70	32.17	28.88	28.06	26.79	27.90
July	58.99	46.95	43.57	39.86	37.38	35.23	34.75	34.13	32.07	32.00
August	56.27	52.36	48.04	44.68	41.50	40.25	37.33	37.21	34.51	35.88
September	56.50	55.17	53.02	49.90	42.13	40.99	39.85	39.22	38.07	35.45
October	56.41	55.35	54.12	50.44	43.94	41.85	39.24	39.82	37.32	36.21
November	57.51	58.18	51.99	54.04	41.75	40.25	43.16	40.29	39.83	37.72
December	57.58	59.31	55.14	52.05	44.28	43.29	41.00	41.35	38.90	36.92

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

According to Equations (15) and (16), the PSIs and SIs of the accident-prone road sections are calculated, and the results are shown in Table 11.

**Table 11.** Accident black spot indicators for accident-prone road sections.

Road Name	Index	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Nc	PSI	7.59	16.35	23.11	25.09	27.99	22.27	17.50	19.41	17.51	15.58
	SI	0.25	0.36	0.44	0.45	0.48	0.39	0.29	0.33	0.29	0.25
Sj <sup>1</sup>	PSI	0.88	1.84	11.18	14.21	18.95	15.36	19.17	16.35	20.18	18.31
	SI	0.07	0.09	0.36	0.34	0.41	0.29	0.35	0.29	0.35	0.31
Hc	PSI	−0.89	−2.81	5.63	12.24	19.57	17.04	19.02	18.12	18.99	18.14
	SI	−0.06	−0.10	0.17	0.31	0.47	0.36	0.37	0.34	0.37	0.33
NR <sup>1</sup>	PSI	−1.77	2.71	8.32	13.97	17.86	21.68	21.90	14.44	22.04	21.05
	SI	−0.14	0.17	0.34	0.44	0.47	0.51	0.45	0.28	0.41	0.41

Table 11. Cont.

Road Name	Index	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Dc	PSI	−4.33	1.75	9.85	14.70	17.38	18.50	21.13	15.94	16.75	21.28
	SI	−0.34	0.12	0.42	0.44	0.50	0.47	0.54	0.38	0.42	0.51
Sj <sup>2</sup>	PSI	−0.81	3.30	9.52	14.17	15.23	17.25	18.99	13.85	17.25	18.29
	SI	−0.09	0.27	0.44	0.49	0.48	0.47	0.51	0.35	0.47	0.46
Qlw	PSI	−0.85	4.24	8.75	9.88	17.75	16.33	13.85	10.24	12.16	13.00
	SI	−0.08	0.30	0.41	0.37	0.57	0.48	0.37	0.27	0.29	0.33
NR <sup>2</sup>	PSI	−1.69	0.86	1.76	7.06	8.13	8.21	15.22	11.82	15.29	15.35
	SI	−0.13	0.05	0.08	0.28	0.26	0.24	0.43	0.32	0.42	0.41
KZ	PSI	−0.85	−2.72	−2.78	2.79	11.07	7.51	13.07	10.32	15.83	14.90
	SI	−0.09	−0.16	−0.12	0.10	0.36	0.22	0.36	0.28	0.42	0.40
G228	PSI	−3.38	−0.88	3.56	9.90	16.00	12.88	15.45	10.21	14.72	12.92
	SI	−0.31	−0.06	0.19	0.38	0.56	0.38	0.47	0.29	0.42	0.37

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street.

#### 4.2. Accident Black Spot Identification Results

According to the accident black spot principle in this paper, when both the PSI and SI indicators were greater than 0, the corresponding accident-prone location was judged to be an accident black spot, and the accident black spots were identified for the accident-prone road sections. The results are shown in Table 12.

Table 12. Accident black spot indicators for accident-prone road sections.

Road Name	Index	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Nc	PSI	7.59	16.35	23.11	25.09	27.99	22.27	17.50	19.41	17.51	15.58
	SI	0.25	0.36	0.44	0.45	0.48	0.39	0.29	0.33	0.29	0.25
	Result	√	√	√	√	√	√	√	√	√	√
Sj <sup>1</sup>	PSI	0.88	1.84	11.18	14.21	18.95	15.36	19.17	16.35	20.18	18.31
	SI	0.07	0.09	0.36	0.34	0.41	0.29	0.35	0.29	0.35	0.31
	Result	√	√	√	√	√	√	√	√	√	√
Hc	PSI	−0.89	−2.81	5.63	12.24	19.57	17.04	19.02	18.12	18.99	18.14
	SI	−0.06	−0.10	0.17	0.31	0.47	0.36	0.37	0.34	0.37	0.33
	Result	×	×	√	√	√	√	√	√	√	√
NR <sup>1</sup>	PSI	−1.77	2.71	8.32	13.97	17.86	21.68	21.90	14.44	22.04	21.05
	SI	−0.14	0.17	0.34	0.44	0.47	0.51	0.45	0.28	0.41	0.41
	Result	×	√	√	√	√	√	√	√	√	√
Dc	PSI	−4.33	1.75	9.85	14.70	17.38	18.50	21.13	15.94	16.75	21.28
	SI	−0.34	0.12	0.42	0.44	0.50	0.47	0.54	0.38	0.42	0.51
	Result	×	√	√	√	√	√	√	√	√	√
Sj <sup>2</sup>	PSI	−0.81	3.30	9.52	14.17	15.23	17.25	18.99	13.85	17.25	18.29
	SI	−0.09	0.27	0.44	0.49	0.48	0.47	0.51	0.35	0.47	0.46
	Result	×	√	√	√	√	√	√	√	√	√
Qlw	PSI	−0.85	4.24	8.75	9.88	17.75	16.33	13.85	10.24	12.16	13.00
	SI	−0.08	0.30	0.41	0.37	0.57	0.48	0.37	0.27	0.29	0.33
	Result	×	√	√	√	√	√	√	√	√	√
NR <sup>2</sup>	PSI	−1.69	0.86	1.76	7.06	8.13	8.21	15.22	11.82	15.29	15.35
	SI	−0.13	0.05	0.08	0.28	0.26	0.24	0.43	0.32	0.42	0.41
	Result	×	√	√	√	√	√	√	√	√	√
KZ	PSI	−0.85	−2.72	−2.78	2.79	11.07	7.51	13.07	10.32	15.83	14.90
	SI	−0.09	−0.16	−0.12	0.10	0.36	0.22	0.36	0.28	0.42	0.40
	Result	×	×	×	√	√	√	√	√	√	√
G228	PSI	−3.38	−0.88	3.56	9.90	16.00	12.88	15.45	10.21	14.72	12.92
	SI	−0.31	−0.06	0.19	0.38	0.56	0.38	0.47	0.29	0.42	0.37
	Result	×	×	√	√	√	√	√	√	√	√

Note that the names of the road sections in the above table are the abbreviations of the road names of the hidden sections, and the specific names are shown in Table 3. Sj<sup>1</sup> refers to the Century Avenue section of Xinming street; Sj<sup>2</sup> refers to the Century Avenue section of Kongpu street; NR<sup>1</sup> refers to the North Ring West Road section of ZhuangQiao street; NR<sup>2</sup> refers to the North Ring West Road section of Hongtang street. √ indicates the identification result is accident black spot, × indicates the identification result is non-accident black spot.

#### 4.3. Grade Classification of Accident Black Spots

According to Table 12, the numbers of accident black spots identified as non-accident black spots and accident black spots were 12 and 88, respectively. From the results, the accident black point identification results for each hidden accident location during each month were different, which means that the monthly accident black point identification calculation results for each hidden accident location could not be averaged or summed. According to the content of Section 3.3 regarding accident black spot classification, the results of Table 12 were divided into black spot classes, and the proportions of the assessment results at the various levels are shown in Figure 2.

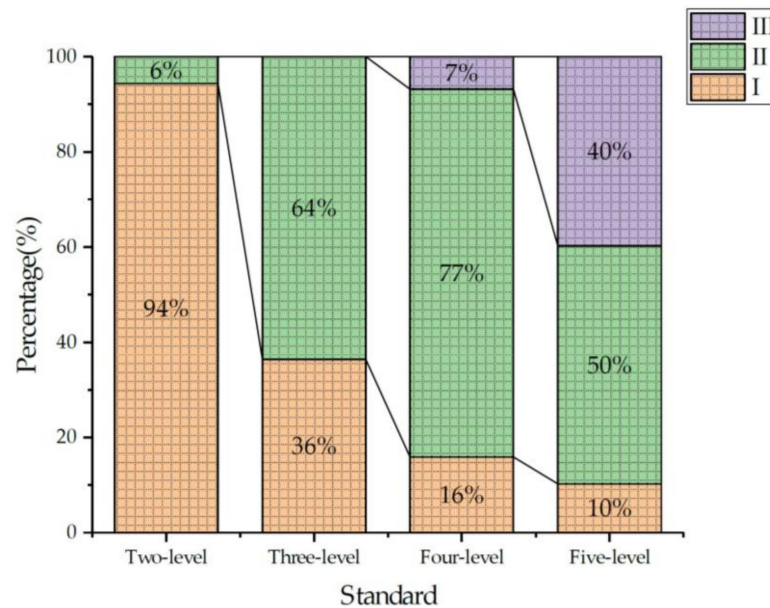
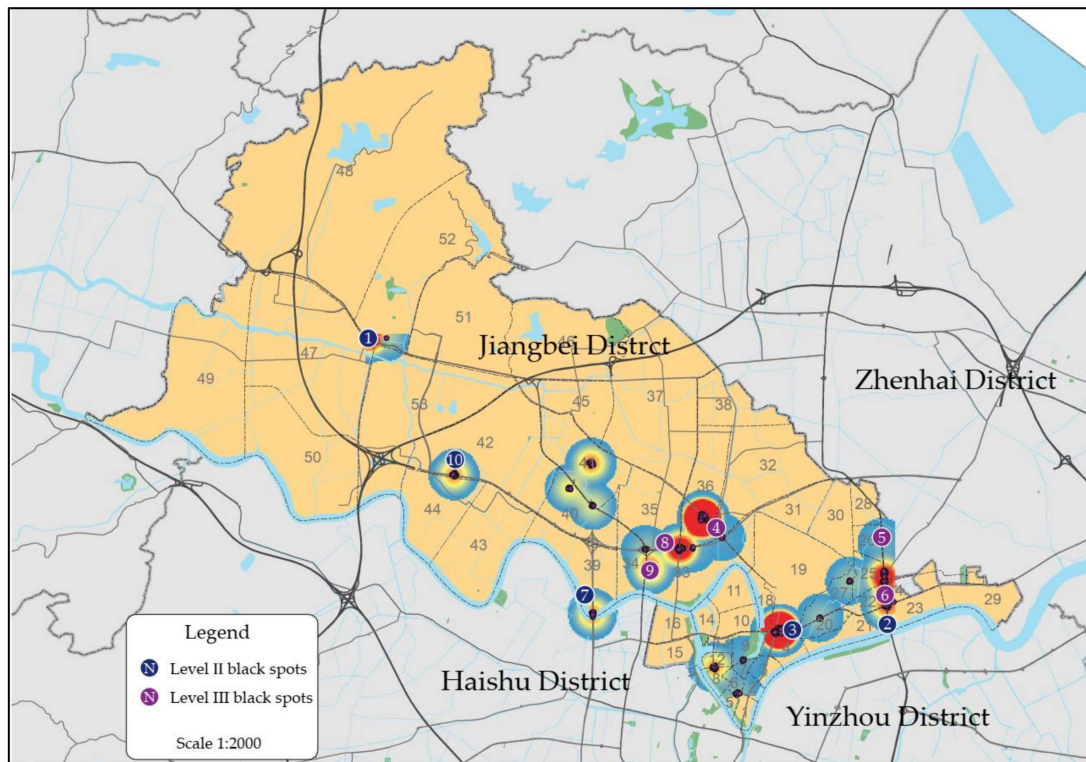


Figure 2. Proportion of evaluation results at each level.

It should be clear that due to the different thresholds for the SI division interval, the evaluation results at all levels must show different distribution trends. As seen from the above figure, according to the division results of the two-level SI evaluation table, the percentage of level I accident black spots was 94%, the percentage of level II accident black spots was 6% and the percentage of level III accident black spots was 0%, which means that all accident black spots with SI values less than or equal to 0.5 were evaluated as level I accident black spots, which is obviously unreasonable; according to the division results of the three-level SI evaluation table, the percentage of level I accident black spots was 36%, the percentage of level II accident black spots was 64% and the percentage of level III accident black spots was 0%, which is unreasonable. According to the four-level SI evaluation table, the percentage of level I accident black spots was 16%, the percentage of level II accident black spots was 77% and the percentage of level III accident black spots was 7%; from the viewpoint of rationality, the four-level evaluation > the three-level evaluation > the two-level evaluation. According to the five-level SI evaluation table, the percentage of level I accident black spots was 10%, the percentage of level II accident black spots was 50% and the percentage of level III accident black spots was 40%; from the perspective of rationality, the five-level evaluation > the four-level evaluation > the three-level evaluation > the two-level evaluation. In addition, since the maximum difference between the numbers of accidents ( $Y(r)$ ) at the same type of accident black spots was not more than double, i.e.,  $PSI \leq Y(r)$ , the SI was rarely greater than 0.6, which is actually related to the choice of the same types of accident black spots.

In summary, the five-level assessment method for accident black spots is more suitable for the qualitative study of accident black spots. We present the accident black spots of September 2020 in the form of an accident heat map according to the threshold values of the five-level assessment table, as shown in Figure 3.



**Figure 3.** Classification results for the black spots in the Jiangbei District (September 2020).

## 5. Discussion and Conclusions

### 5.1. Discussion

In this study, before performing the identification and classification of accident black spots, the hidden accident road sections were first determined based on the number of accidents occurring within a given time period. Then, model calculations were performed for the hidden accident road sections, and the specifics were listed strictly according to the model construction process. Then, the identified hidden accident road sections were screened and identified to determine accident black spots and non-accident black spots. Finally, the identified accident black spots were graded (at two, three, four and five levels) according to the proposed accident black spot assessment method.

The higher the quality of the data collected, the closer the model calculation results were to the actual situation. The accident location data collected in this paper were obtained from the Jiangbei District, Ningbo, China. The accident records were obtained through the cell phone application of the Jiangbei District Traffic Accident Collection Platform, and the data statistics were collected from 0:00 on 1 March 2020 to 0:00 on 1 January 2021, a period of 10 months. A total of 29,623 accident records were collected through the cell phone application, for which the data sample of valid GPS coordinates and accident locations totaled 29,275 coordinates, as 348 invalid samples were excluded, and the accident collection efficiency of this cell phone application reached 98.83%. To further verify the accuracy and quality of the collected accident data, the effective police data of the Jiangbei Traffic Police Brigade Command Center and Accident Squadron were retrieved for comparison, and it was found that the effective police data for the same period contained 32,330 cases, and the sample volume of effective accident data collected based



on the mobile application accounted for 90.55% of the actual police cases, which satisfied the accuracy and quality requirements of the accident data analysis.

Since the accuracy of accident prediction is directly related to the accuracy of accident black spot identification, the overall accuracy of accident prediction in this paper was relatively good. From Table 8, it can be concluded that (a) the average relative errors of the prediction results of the 10 accident hazards sections were all between 0.01 and 0.05 (class II), and the overall relative error was small, indicating that the prediction results were relatively good. Among them, the maximum value of the average relative error was 4.74% (0.049) for Century Avenue 2; this was due to the small sample size regarding the number of accidents relative to other prediction objects (population, economy, etc.). (b) The absolute correlations of the predicted results for the 10 accident-hazardous road sections were all greater than 0.9 (level I), indicating a high degree of correlation between the predicted number of accidents and the actual number of observed accidents. (c) The mean–variance ratios of the prediction results for the 10 accident-hazardous road sections were all less than 0.35 (level I), indicating that the residuals between the predicted values and the original values were small, and the smaller the mean–variance ratio is, the higher the prediction accuracy. It is undeniable that the sample size of accident data in this paper was small (with a single month as the statistical time period), and the number of accidents at each accident hazard location was almost less than 100. This also leads to the fact that the expected numbers of accidents at the accident hazard locations in the prediction results must be accurate to two decimal places to facilitate the accuracy test; otherwise, the test results would produce large errors. When the number of accidents was used as the object of study in previous works, a year was used as the statistical time, which of course prevented the problem of the prediction results needing to be accurate to two decimal places.

Among the traditional accident black point identification methods, the quality control method is closer to the empirical Bayesian method in terms of black point identification ranking results compared to the accident rate method, and this phenomenon can be explained from the principle of the method [39,40]. Both the quality control method and the empirical Bayesian method consider the number of traffic accidents occurring at a point as a random variable, which is consistent with reality, while the accident rate method does not consider the random characteristics of accident occurrence and differs significantly from the other two methods. The difference between the empirical Bayesian method and the quality control method is that the empirical Bayesian method combines two aspects of information (i.e., the mean accident value of similar accident objects and site-specific historical accident data) for black spot identification, while the quality control method only refers to one aspect of information, namely, historical accident data. It should be noted that, due to the random fluctuation of accident occurrence, using historical accident data alone cannot accurately reflect the characteristics of accident distribution at a particular location, while the empirical Bayesian method can eliminate this effect and enable more accurate results to be obtained, see Table 13.

**Table 13.** Characteristic analysis of accident black spot identification methods.

Identification Methods	Was the Severity of the Accident Considered?	Was the Randomness of Accident Occurrence Considered?	Was the Accident Mean of Similar Objects Considered?
Accident number method	×	×	×
Accident rate method	×	×	√
Quality control method	×	√	×
Comprehensive index method	√√	×	×
Empirical Bayesian method	×	√√	√√

Note, × indicates that the corresponding feature is not considered, √ indicates that the corresponding feature is not fully considered, √√ indicates that the corresponding feature is fully considered.

Tian et al. (2019) [4] used the optimized value of empirical Bayesian accident number and the predicted value of accident number derived from the accident prediction model to construct the ranking index of accident black spots after accident black spot identification, and the difference and ratio between them reflected the severity of accident black spots. However, the method is still not free from the shackles of quantifying the accident severity and cannot provide managers with the specific ranking of accident black spots. We discussed the SI of accident black spots in a graded manner, and within the range  $(-1, 1)$ , the macro level can be divided into accident black spots and non-accident black spots, and the micro level can be divided into secondary, tertiary, quaternary and quintuple accident black spots according to the size of the given interval. When the SI is less than 0, according to the accident black spot discrimination rule, the hidden accident location is not an accident black spot; when the SI is greater than 0.6, it means that the PSI value of the accident black spot is extremely large, and this type of situation generally occurs less often. However, it should be noted that while the accident black spots identified by the model need more attention, accident prevention and management for non-accident black spots should also receive the same attention.

In this study, the grey Verhul's method was chosen for accident prediction based on the fact that domestic road traffic accidents have exhibited an S-shaped process with a saturation state in recent years [41]. Grey prediction aims to grasp the development law of a system and make scientific quantitative predictions regarding the future development of the system through the processing of the original data and the establishment of a grey model. The most commonly applied models in the grey prediction field are the GM (1, 1) model and grey Markov prediction model, which can be used to predict the number of traffic accidents, the number of deaths, the number of injuries, the amount of property damage and other indicators. However, the GM (1, 1) model is applicable to sequences with strong exponential laws and can only describe monotonic change processes. However, the road traffic system is a dynamic time-varying system, and traffic accidents, as a behavioral characteristic quantity [42], have a certain level of stochastic volatility and present a non-smoothly varying stochastic trend. GM (1, 1) cannot solve this situation. Furthermore, for traffic accident prediction involving S-shaped processes with saturation states, the grey Markov prediction model cannot perform state classification. Thus, in this paper, the grey Verhul's method was chosen for accident occurrence prediction, as it provides sufficient conditions for accident black spot identification and delineation [43,44].

## 5.2. Conclusions

We proposed a method that can quickly identify urban road accident black spots and classify their ranks. The model combined the grey Verhul's method and the empirical Bayesian method and took the numbers of accident occurrences at urban road hidden accident locations as the original data. After the calculation of the combined model, the proposed method obtained the expected numbers of accident occurrences at the hidden accident locations and used the calculation results of the safety improvement space and SI as the discriminating indexes of the accident black spots to discriminate whether the hidden accident locations are accident black spots or not. Finally, the accident black spots were divided into the following levels:

(a) In terms of accident black spot identification, accident black spots and non-accident black spots were identified for the accident data of 10 hidden accident sections over 10 months, and from the results, 12 non-accident black spots and 88 accident black spots were detected for a total of 100 data points, which means that the model can be used for accident black spot identification.

(b) In terms of accident black spot classification, the rationality of the proportions of the identified accident black spots with grades I, II and III in the two-grade assessment, three-grade assessment, four-grade assessment and five-grade assessment were discussed, and the analysis results showed that the five-grade assessment > the four-grade assessment > the three-grade assessment > the two-grade assessment and that the five-

grade assessment of accident black spots is most suitable for the actual situation, which provides a new way of thinking regarding accident black spot classification.

(c) The grey Verhulst–Empirical Bayesian combination method not only adapted to the current law of road traffic accidents in China but also achieved good test results for the three tested accuracy indices, the average relative error, absolute correlation, and mean square error ratio, indicating that the model was highly accurate in terms of accident black spot identification.

However, this paper has some limitations. First, the sample of this paper included only 10 months of accident data records collected by a cell phone application, and to expand the sample size, this paper fit the model according to half-month data. If accident data records for longer time periods are added appropriately as the initial data, the average relative error, absolute correlation and mean square deviation ratio of the expected number of accidents may be measured, and the test results might improve. Second, the accident black spot identification approach in this paper was mainly based on the number of accidents, and factors such as the geometric conditions of hidden accident locations and traffic volumes were not considered due to the difficulty of cell phone application-based information collection. If these factors are added appropriately, the results of accident black spot identification recognition may improve. Finally, the accident black spot identification approach in this paper was actually based on long-term cumulative accident data for classification purposes, which meant that only after one month of accidents occurred could we identify whether these hidden locations were accident black spots or not, which is not good for accident prevention. If we can consider daily traffic accident prediction or hourly traffic accident prediction, we can prevent accident occurrences at these accident black spots in a timely manner.

In future studies, we will conduct a study of accident black spot identification with a longer statistical time period and consider factors such as the geometric conditions of hidden accident locations to make the relationships between the model and the actual traffic conditions more relevant. Furthermore, the short-term prediction of traffic accidents can also be included in the study when the statistical time period of accident data is sufficiently long.

**Author Contributions:** Conceptualization, Y.W.; methodology, Y.W. and W.H.; software, W.H.; validation, W.H., and J.Z.; formal analysis, W.H.; investigation, W.H.; resources, J.Z.; data curation, J.Z.; writing—original draft preparation, Y.W.; writing—review and editing, J.Z.; visualization, W.H.; supervision, J.Z.; project administration, Y.W.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the by Philosophy and Social Science Foundation of Zhejiang Province (21NDJC163YB), and National Natural Science Foundation of China (52002282).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used to support the findings of this study are included within the article.

**Conflicts of Interest:** The authors declare that they have no conflict of interests.

## References

1. World Health Organization. *Global Status Report on Road Safety 2018: Supporting a Decade of Action*; WHO: Geneva, Switzerland, 2018.
2. Traffic Administration Bureau of the Ministry of Public Security. *Annual Report on Road Traffic Accident Statistics of the People's Republic of China (2019)*; Ministry of Public Security Traffic: Wuxi, China, 2020.
3. Chen, W. Analysis of road traffic accident-prone road. *J. Chin. People's Secur. Univ.* **2004**, *2*, 83–88.
4. Tian, Z.; Zhang, S. Identification and ranking of accident black spots using advanced empirical Bayes method. *J. Chang. Univ.* **2019**, *39*, 115–126.
5. Deng, J. *Grey System Theory*; Huazhong University Press of Science and Technology: Wuhan, China, 1987.

6. Hu, X.; Liu, T.; Zhang, T.H.; Huang, Y.; Xie, H.W.; Yu, T. Highway black spot recognition and improvement. *J. Traffic Transp. Eng.* **2004**, *4*, 106–109.
7. Sha, H.; Chen, Y.; Chai, G. Design on cause analysis system for road black spots. *China Saf. Sci. J.* **2007**, *17*, 75–80.
8. Xiao, S.; Guo, X.; Song, J. Study on road traffic accident black spot identification method. *J. Highw. Transp. Res. Dev.* **2003**, *20*, 70–73.
9. Sun, L.; Shao, Y.; Yan, X.; Lei, X. Identification of highway traffic accident black-spots based on traffic conflict technique. *J. Chongqing Jiaotong Univ.* **2012**, *31*, 63–67.
10. Guo, Y.; Liu, P.; Wu, Y.; Li, X. Bayesian traffic conflict model accounting for heterogeneity. *China J. Highw. Transp.* **2018**, *31*, 296–303.
11. Zeng, Q.; Wen, H.; Huang, H.; Wang, J.; Lee, J. Analysis of crash frequency using a Bayesian underreporting count model with spatial correlation. *Phys. A Stat. Mech. Appl.* **2020**, *545*, 123754. [[CrossRef](#)]
12. Mao, S. *Bayes Statistics*; China Planning Press: Beijing, China, 2000.
13. Song, J.; Chen, X.; Zhang, B. Study on traffic security countermeasures of expressway. *Commun. Stand. Issue* **2006**, *7*, 194–198.
14. Guo, X.; Sheng, Y.; Pan, Z.; Pan, M.; He, M. Analysis of general characteristics for highway traffic accident black-spot. *J. Southeast Univ.* **2007**, *37*, 930–933.
15. Wen, H.; Wu, Y.; Qi, W. Models of identification and formulation cause analysis on accident black spots of mountainous highway. *J. Chongqing Jiaotong Univ.* **2016**, *35*, 137–140.
16. Danielsson, S. A comparison of two methods for estimating the effect of a countermeasure in the presence of regression effects. *Accid. Anal. Prev.* **1986**, *18*, 13–23. [[CrossRef](#)]
17. Guo, L.; Zhou, J.; Dong, S.; Zhang, S. Analysis of urban road traffic accidents based on improved k-means algorithm. *China J. Highw. Transp.* **2018**, *31*, 270–279.
18. Geng, C.; Peng, Y. Identification method of traffic accident black spots based on dynamic segmentation and DBSCAN algorithm. *J. Chang. Univ.* **2018**, *38*, 131–138.
19. Wang, J.; Lu, X. Highway black spot verification and cause analysis based on cluster analysis. *Technol. Highw. Transp.* **2016**, *32*, 114–119.
20. Lin, L.; Zhang, Y.; Gao, M. Formation cause identification and improvement of the road traffic accidents black spots. *J. Nanjing For. Univ.* **2010**, *34*, 157–160.
21. Yetis, S.M.; Cakici, Y. An integration of different computing approaches in traffic safety analysis. *Transp. Res. Procedia* **2017**, *22*, 265–274.
22. Cui, H.; Shen, X.; Zhao, S.J.; Sun, X.X. New traffic accident black spots identification method based on abnormal accidents space. *J. Wuhan Univ. Technol.* **2012**, *34*, 54–58.
23. Amoros, E.; Martin, J.L.; Laumon, B. Comparison of road crashes incidence and severity between some French countries. *Accid. Anal. Prev.* **2003**, *35*, 537–547. [[CrossRef](#)]
24. Lord, D.; Miranda-Moreno, L.F. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. *Saf. Sci.* **2008**, *46*, 751–770. [[CrossRef](#)]
25. Barua, S.; El-Basyouny, K.; Islam, M.T. Effects of spatial correlation in random parameters collision count-data models. *Anal. Methods Accid. Res.* **2015**, *5*, 28–42. [[CrossRef](#)]
26. Malyskina, N.; Mannering, F. Zero-state Markov switching count-data models: An empirical assessment. *Accid. Anal. Prev.* **2010**, *42*, 122–130. [[CrossRef](#)] [[PubMed](#)]
27. Maen, G.; Török, Á. A comparative analysis of black spot identification methods and road accident segmentation methods. *Accid. Anal. Prev.* **2019**, *128*, 1–7.
28. Hauer, E.; Kononov, J.; Allery, B.; Griffith, M.S. Screening the road network for sites with promise. *Transp. Res. Rec.* **2002**, *1784*, 27–32. [[CrossRef](#)]
29. Montella, A. A comparative analysis of hotspot identification methods. *Accid. Anal. Prev.* **2010**, *42*, 571–581. [[CrossRef](#)] [[PubMed](#)]
30. Miaou, S.; Song, J. Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accid. Anal. Prev.* **2005**, *37*, 699–720. [[CrossRef](#)] [[PubMed](#)]
31. Wen, C.; Washington, S.P. Experimental evaluation of hotspot identification methods. *Accid. Anal. Prev.* **2005**, *37*, 870–881.
32. Wang, F.; Li, T.; Yu, C. Grey Verhulst Predictive Model of Road Traffic Accidents. *J. Traffic Transp. Eng.* **2006**, *6*, 122–126.
33. Mcguigan, D. The use of relationships between road accidents and traffic flow in 'black-spot' identification. *Traffic Eng. Control* **1981**, *8*, 448–453.
34. Persaud, B.; Lyon, C.; Nguyen, T. Empirical Bayes procedure for ranking sites for safety investigation by potential for improvement. *Transp. Res. Rec.* **1999**, *1665*, 7–12. [[CrossRef](#)]
35. Debrabant, B.; Halekoh, U.; Bonat, W.H.; Hansen, D.L.; Hjelmberg, J.; Lauritsen, J. Identifying traffic accident black spots with Poisson-Tweedie models. *Accid. Anal. Prev.* **2018**, *111*, 148–157. [[CrossRef](#)]
36. Guo, Y.; Liu, P.; Wu, Y.; Yang, M.-L. Traffic conflict model based on Bayesian multivariate Poisson-lognormal normal distribution. *China J. Highw. Transp.* **2018**, *31*, 101–109.
37. Wang, X.; Xu, X. Assessing the relationship between self-reported driving behaviors and driver risk using a naturalistic driving study. *Accid. Anal. Prev.* **2019**, *128*, 8–16. [[CrossRef](#)]

38. Feng, Z.; Yang, L.; Jing, L.; Zhou, M.; Zhang, W.; Kun, W. Evolutionary game analysis of the illegal occupancy behavior of the motorway emergency lane. *Cogn. Technol. Work* **2017**, *19*, 461–467. [[CrossRef](#)]
39. Zeng, Q.; Xu, P.; Wang, X.; Wen, H.; Hao, W. Applying a Bayesian multivariate spatio-temporal interaction model based approach to rank sites with promise using severity-weighted decision parameters. *Accid. Anal. Prev.* **2021**, *157*, 106190. [[CrossRef](#)]
40. Zhou, J.; Mao, X.; Wang, Y.; Zhang, M.; Dong, S. Risk Assessment in urban large-scale public spaces using Dempster-Shafer theory: An empirical study in Ningbo, China. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2942. [[CrossRef](#)] [[PubMed](#)]
41. Ma, C.; Dai, G.; Zhou, J. Short-Term traffic flow prediction for urban road sections based on time series analysis and LSTM\_BILSTM method. *IEEE Trans. Intell. Transp. Syst.* **2021**, *99*, 1–10.
42. Wan, Y.; Zhou, J.; He, W.; Ma, C. Modeling the curb parking price in urban center district of China using TSM-RAM approach. *J. Adv. Transp.* **2020**, *2020*, 4905059. [[CrossRef](#)]
43. Uma, S.; Eswari, R. Accident prevention and safety assistance using IOT and machine learning. *J. Reliab. Intell. Environ.* **2021**, *10*, 1–25.
44. Tanprasert, T.; Siripanpornchana, C.; Surasvadi, N.; Thajchayapong, S. Recognizing traffic black spots from street view images using environment-aware image processing and neural network. *IEEE Access* **2020**, *99*, 121469–121478. [[CrossRef](#)]