electronics

MDPI

Article

Adaptive Kernel Density Estimation for Traffic Accidents Based on Improved Bandwidth Research on Black Spot Identification Model

Huimin Ge *, Lei Dong, Mingyue Huang, Wenkai Zang and Lijun Zhou

School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China
* Correspondence: hmge@ujs.edu.cn

Abstract: At present, the total length of accident blackspot accounts for 0.25% of the total length of the road network, while the total number of accidents that occurred at accident black spots accounts for 25% of the total number of accidents on the road network. This paper describes a traffic accident black spot recognition model based on the adaptive kernel density estimation method combined with the road risk index. Using the traffic accident data of national and provincial trunk lines in Shanghai and ArcGIS software, the recognition results of black spots were compared with the recognition results of the accident frequency method and the kernel density estimation method, and the clustering degree of recognition results of adaptive kernel density estimation method were analyzed. The results show that: the accident prediction accuracy index values of the accident frequency method, kernel density estimation method, and traffic accident black spot recognition model were 14.39, 16.36, and 18.25, respectively, and the lengths of the traffic accident black spot sections were 184.68, 162.45, and 145.57, respectively, which means that the accident black spot section determined by the accident black spot recognition model was the shortest and the number of traffic accidents identified was the largest. Considering the safety improvement budget of 20% of the road length, the adaptive kernel density estimation method could identify about 69% of the traffic accidents, which was 1.13 times and 1.27 times that of the kernel density estimation method and the accident frequency method, respectively.

Keywords: traffic planning and management; traffic accident; black spot identification; adaptive kernel density estimation method; road hazard index

check for updates

Citation: Ge, H.; Dong, L.; Huang, M.; Zang, W.; Zhou, L. Adaptive Kernel Density Estimation for Traffic Accidents Based on Improved Bandwidth Research on Black Spot Identification Model. *Electronics* **2022**, *11*, 3604. https://doi.org/10.3390/electronics11213604

Academic Editors: Fang Zong, Yiting Zhu, Yanyan Qin and Qiaowen Bai

Received: 6 October 2022
Accepted: 2 November 2022
Published: 4 November 2022

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

Copyright: © 2022 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

The road transportation network is a key factor in promoting the development of the world economy. It is fast, efficient, reliable, and flexible, but it is also the most dangerous means of transportation. According to statistics from the World Health Organization, the world's annual traffic accidents cause more than 1.35 million deaths and more than 50 million injuries, and traffic accidents are also the main cause of death for residents aged 15–29 years [1]. As a major public safety issue, traffic accidents not only caused serious casualties, but also brought huge economic losses to the society. In order to improve road safety, many countries have invested a lot of resources in the management of traffic accidents, but the effect has been minimal. Accident-prone sections are called black spots, in foreign countries, which refers to sections or intersections where the number of accidents is obviously prominent on the road [2–4]. Generally, it occupies a small amount of road mileage, but the concentration of traffic accidents is relatively large. Identifying accident black spots and taking corresponding measures can be used as an effective way to prevent traffic accidents [5]. The study of accident black spot identification methods can provide theoretical support for black spot identification research, which is of great significance for reducing road traffic accidents and improving road traffic safety.

Accurately identifying black spots in traffic accidents is a decisive link in improving road traffic safety. At present, many methods have been used to identify black spots in

traffic accidents, mainly including the accident number method, the accident rate method, the probability theory mathematical statistics method, the quality control method, the cumulative frequency curve method, genetic algorithm and information distribution technology, etc. [5–9]. These methods usually divide the entire highway into section units (usually 1 km) to identify black spots in traffic accidents. Thomas et al. pointed out that the use of a fixed length to identify black spots in traffic accidents may lead to deviations in the positions of the black spots, causing errors between the recognition results and the actual length of the road section [10]. At the same time, Meng Xianghai also pointed out that the use of fixed length to identify accident black spots would affect the accuracy of the results, and proposed the use of the sliding window method to identify them [11]. However, as the research continues to deepen, other researchers have found that although the sliding window method can improve the accuracy of the identification of accident black spots, it will also cause the length of the road section identified by the accident black spot to dismatch the actual road section length, thereby affecting the recognition accuracy [12,13]. Koorey et al. also discussed the impact of using variable length on the identification of traffic accident blackspots, and confirmed that choosing an appropriate segmentation method has a significant impact on reducing false positives and negatives in the results of traffic accident black spot recognition [14]. Because the clustering algorithm can not only identify accident black spots, but also analyze the causes of accident black spots, it has attracted much attention in the identification of traffic accidents [15–17]. However, the clustering algorithm also has certain limitations. The black spots of the same type of clustered traffic accidents obtained by clustering have the same degree of danger, which does not match the actual situation, which is that each accident black spot has a different degree of danger. Nuclear density estimation can reduce the severity of accident blackspots, and has gradually become a research hotspot in traffic accident black spots [18–21]. However, using a fixed bandwidth to identify black spots in an accident will also affect the accuracy of the results. At the same time, most of the current studies use the frequency of traffic accidents as the identification index and do not consider the degree of danger of the target road section.

The purpose of this paper is to solve the problem that the degree of accident danger caused by the current clustering algorithm does not match with the current situation and thereby improve the accuracy of fixed broadband accident identification results. Taking the traffic accident data of Shanghai National Highway as the research object, this paper combined the adaptive kernel density estimation method and GIS data to identify the traffic accident black spots on the basis of considering the road danger degree and compared the results with those for the accident frequency method and nuclear density estimation method, based on the crash predictive accuracy index (CPAI) and road repair and maintenance cost limit indicators (RRMCLI) to verify and evaluate the effectiveness of the adaptive kernel density estimation method. Based on the traffic data of Shanghai trunk highways, the traffic characteristics and traffic accident characteristics were analyzed to provide supportive data for the construction of the black spot identification model of urban trunk highways and the comparative analysis of the identification results. The results show that the self-adaptive kernel density estimation based on the improved bandwidth used in this paper was superior to the previous accident blackspot recognition model.

2. Study Area and Data Collection

In order to obtain accurate traffic accident data and determine the spatial location of traffic accidents, the Shanghai National Highway was used as the data collection location. As a provincial-level administrative region and municipality directly under the Central Government, Shanghai is located in the Yangtze River Delta, bounded between 120°52′ to 122°12′ east longitude and 30°40′ to 31°53′ north latitude. The city has 16 districts with a total area of 6340.5 square kilometers. The total length of traffic infrastructure is approximately 13,500 km, of which, approximately 972 km comprises highways, and approximately 1812 km comprises national and provincial trunk roads, as shown in Figure 1. Shanghai as one of the major cities generating Chinese financial revenue, not only because

of its unique geographical advantages, but also because of its developed road traffic conditions. The transportation of goods and import and export commodities has made the traffic volume of Shanghai trunk highways gradually increase, and the urban population and traffic flow have grown, leading to the frequent occurrence of road traffic accidents.



(**a**) (**b**)

Figure 1. Location and traffic network of Shanghai. (**a**) Shanghai location. (**b**) Shanghai traffic arterial network map.

The Shanghai Traffic Safety Integrated Service Management Platform records accident information, such as the location, time, and type of traffic accidents. Therefore, the traffic accident data studied in this article are mainly derived from the comprehensive traffic safety service management platform, a total of 3 years (2017–2019) of traffic accident data on national and provincial trunk lines in Shanghai, including the number of injured, the number of deaths, and the direct economic loss, as shown in Table 1. According to statistics in Table 1, there were a total of 2031 traffic accidents in Shanghai in the past 3 years, of which 477 were injured, 108 were killed, and the direct economic and property losses were about 2.793 million yuan.

Table 1. Statistical description of traffic accidents of different severity in Shanghai 2017–2019.

Year	Total Number of Accidents	Number of Injured	Death Toll	Direct Economic Loss [Yuan]	Injury Rate	Mortality Rate
2017	864	290	59	1,867,883	0.336	0.068
2018	618	101	29	492,473	0.163	0.047
2019	550	86	20	432,560	0.156	0.036
Total	2031	477	108	2,792,916	0.235	0.053

At the same time, this study also conducted a statistical analysis of traffic accidents in different administrative regions of Shanghai, as shown in Table 2. It can be found in Table 2 that although the maximum number of traffic accidents in Pudong New Area was 420, the injury rate and death rate were only 0.269 and 0.067, respectively. However, 44 traffic accidents occurred in Huangpu District, but the injury rate and death rate of traffic accidents there were the highest at 0.409 and 0.205, respectively.

Table 2. Descriptive statistics of traffic accidents in different administrative regions.

Administrative District	Total Number of Accidents	Number of Injured	Death Toll	Direct Economic Loss [Yuan]	Injury Rate	Mortality Rate
Baoshan District	177	47	12	54,600	0.266	0.068
Chongming District	41	4	0	23,500	0.096	0
Fengxian District	183	41	4	315,800	0.224	0.022
Hongkou Distric	25	5	0	4600	0.200	0
Huangpu District	44	18	9	110,000	0.409	0.205
Jiading District	170	38	10	192,700	0.224	0.059
Jingshan District	124	17	1	33,000	0.137	0.008
Jingan District	53	7	4	7000	0.132	0.075
Minhang District	189	54	6	347,900	0.286	0.032
Pudong New Area	420	113	28	1,218,535	0.269	0.067
Putuo District	36	13	4	8381	0.361	0.111
Qingpu District	248	35	8	164,700	0.141	0.0322
Songjiang District	154	53	19	238,700	0.344	0.123
Xuhui District	100	16	0	13,500	0.16	0
Yangpu District	43	9	0	4400	0.209	0
Changning District	24	7	3	16,000	0.292	0.125

3. Recognition Methods

3.1. Adaptive Kernel Density Estimation

The kernel density estimation method places a symmetrical surface at each traffic accident point, and by using the density method to define an arbitrary segment consistent with the entire road network, thereby identifies traffic accident black spots [20]. Its expression is:

$$\hat{f}(x) = \sum_{i=1}^{n} \frac{1}{nh} k\left(\frac{x - x_i}{h}\right) \quad (1)$$

In Equation (1): $\hat{f}(x)$ is the nuclear density value at point x; n is the number of sample points whose distance from the center x is less than or equal to h; h is the bandwidth (search radius); $k(x)$ is the kernel function, using the distance between the sample data point x_i and x to determine the effect when x_i estimating the density of point x; $x - x_i$ is the distance between event point i and event point x.

The adaptive kernel density estimation law is a natural extension of the standard kernel density estimation method, also known as the variable kernel density estimation method, which is obtained when h is no longer a global constant [14]. Its application principle relative to accident black spots is that, in the section with high accident density, relatively small bandwidth is selected to reflect the local characteristics of the accident, and in the section with low accident density, relatively large bandwidth is selected to reflect the overall characteristics of the accident. Then, the influence of outliers on the estimation value is reduced. At present, most traffic accident black spot identification methods use a fixed length of road to identify traffic accident black spots. This cannot reflect the real-time changes of traffic accidents and identify accident black spots accurately. In order to objectively select the optimal bandwidth coefficient and improve the accuracy of blackspot recognition in traffic accidents, this paper proposes to construct an adaptive kernel density estimation and improve the fixed bandwidth so that the bandwidth value changes with the change of each sample data point.

3.2. Kernel Function Selection

Commonly used kernel functions mainly include uniform, Gaussian, triangular, quadratic (Epanechnikov), and others [22]. According to the expressions of different kernel functions in Table 3, the uniform, triangular, and quadratic kernel functions intersect the x axis, and the Gaussian kernel function does not intersect the x axis, as shown in Figure 2.

Table 3. Commonly used kernel functions.

Kernel Functions	Expressions
Uniform	$k(x) = \begin{cases} 1/2 & [-1,1] \\ 0 & other \end{cases}$
Gaussian	$k(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$
Triangular	$k(x) = \begin{cases} \frac{3(1-x^2/a^2)}{4a} & \lvert x\rvert \le a \\ 0 & \lvert x\rvert \ge a \end{cases}$
Epanechnikov	$k(x) = \begin{cases} \frac{3(1-x^2)}{4} & \lvert x\rvert \le 1 \\ 0 & \lvert x\rvert > 1 \end{cases}$

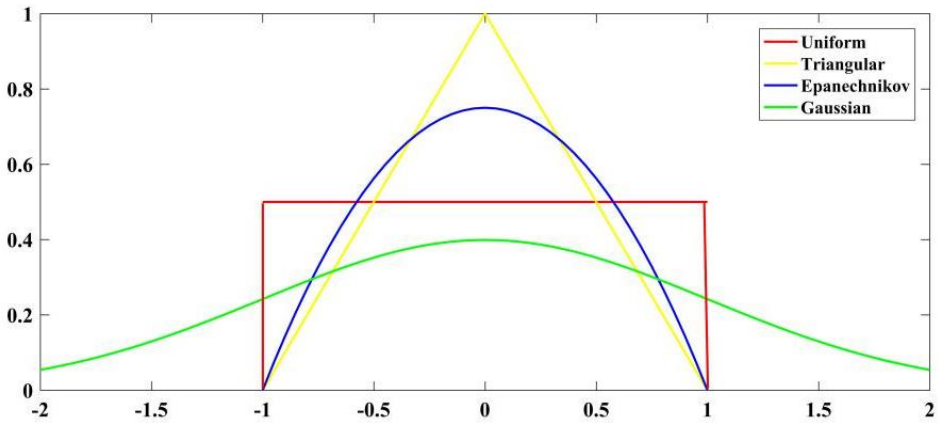


Figure 2. Comparison of different kernel functions.

In addition, this paper analyzes the influence of the kernel function on the estimated value of nuclear density by fitting the kernel density estimation curve of the above-mentioned commonly used kernel function. The curve fitting of different kernel functions is shown in Figure 3. It can be seen from Figure 3 that the kernel function has basically no effect on the kernel density estimate. Only the kernel density estimation curve obtained by using uniform kernel and quadratic kernel at the peak and boundary of the curve is different from the curves of the other two kernel functions, but the smoothness of the kernel density estimation curve in any other area is almost the same. At the same time, according to the research conclusions of relevant scholars, the determination of kernel function has little influence on the final estimation of kernel density and is not the main link between calculation and solution. In addition, under the condition that the research object has a large number of sample data, accurate estimation results of kernel density can be obtained by selecting any kernel function [23]. Although the kernel function has little influence on the kernel density estimation, the mean square error of the quadratic kernel function is the best. Therefore, the quadratic kernel function is selected as the kernel function for calculating the nuclear density of traffic accidents.

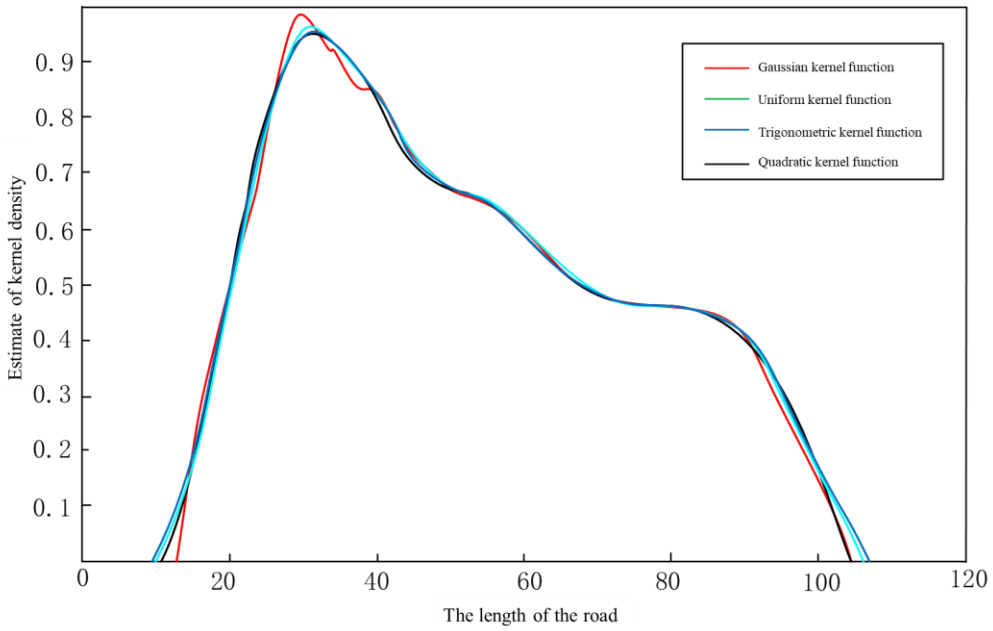


Figure 3. Fitting curves of kernel density estimation for different kernel functions.

3.3. Calculation of Adaptive Bandwidth

The key to adaptive kernel density estimation lies in the choice of bandwidth, which directly affects the accuracy of the results [24]. Among them, if the bandwidth is too large, the detailed information of the data cannot be reflected; on the contrary, the overall characteristics of the data also cannot be reflected. Therefore, in order to reflect the overall characteristics and detailed characteristics of the data as much as possible, this article adopts the interpolation method and measures the quality of the bandwidth based on the mean square error (MSE). When the mean square error is the smallest, the bandwidth is the optimal bandwidth; that is, the optimal bandwidth can be obtained from the minimum value of the mean square error (MSE) between Equation (1), the kernel density $\hat{f}(x)$ value, and the true density $f(x)$, and its expression is as follows:

$$MSE(h) = EMISE(h) = E[\hat{f}(x) - f(x)]^2 = Var(\hat{f}(x)) + Bias(\hat{f}(x))^2 \quad (2)$$

On one hand, the variance is inversely proportional to the bandwidth h, that is, the smaller the bandwidth, the greater the variance. On the contrary, the squared deviation is proportional to the sum h^2, that is, the smaller the bandwidth, the smaller the squared deviation. Therefore, the bandwidth h can be optimized by minimizing the integral mean square error. The expression is as follows:

$$MISE(h) = E\left\{\int [\hat{f}(x) - f(x)]^2 dx\right\} = \int Var\left(\hat{f}(x)\right)dx + \int Bias^2\left(\hat{f}(x)\right)dx \quad (3)$$

Due to $E(x) = \int_{-\infty}^{+\infty} xf(x)dx$ and $\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right)$,we can get

$$E[\hat{f}(x)] = E[\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right)] = \int [\frac{1}{h}k\left(\frac{x-y}{h}\right)f(y)]dy \quad (4)$$

Suppose $t = \frac{x-y}{h}$,then $y = x - ht$, substituting it into Equation (4), we can get:

$$MISE = \int Var\left(\hat{f}(x)\right)dx + \int Bias^2\left(\hat{f}(x)\right)dx \quad = \frac{1}{nh}[\int k^2(t)dt][\int f(x)dx] + \frac{1}{4}h^4[t^2k(t)dt]^2 \int [f''(x)]^2 dx \quad (5)$$

Due to $\int f(x)dx = 1$, so

$$MISE = \frac{1}{nh}[\int k^2(t)dt] + \frac{1}{4}h^4[\int t^2k(t)dt]^2 \int [f''(x)]^2dx = AMISE(h) + o(\frac{1}{nh} + h^4) \quad (6)$$

In Equation (6), $AMISE(h)$ is called the asymptotic mean square error for simplifying the equation, omitting the infinitesimal term. The final formula is as follows:

$$AMISE(h) = \frac{\int k^2(t)dt}{nh} + \frac{h^4t^4 \int [f''(x)]^2dx}{4} \quad (7)$$

By deriving Equation (7) = 0, we can get

$$h_{opt} = t^{2/5}\left\{\int k^2(t)dt\right\}^{-\frac{1}{5}}\left\{\int f''(x)^2dx\right\}^{\frac{1}{5}}n^{-\frac{1}{5}} \quad (8)$$

Meanwhile, the bandwidth h is the optimal value. The function $f(x)$ is unknown from Equation (8). For this reason, $f(x)$ can use normal distribution to estimate the value of $f(x)^2dx$. Assuming that the data of traffic accidents occurring on the road, as a whole obeys a normal distribution, the unknown function $f(x)$ can be expressed as:

$$\int f(x)^2dx = \sigma^{-5}\int \varphi(x)^2dx = \frac{3}{8}\pi^{-\frac{1}{2}}\sigma^{-5} \quad (9)$$

Meanwhile, the values of functions $\int k^2(x)dx$ and $\int k^2k(x)dx$ are directly related to the selection of specific functions. When the kernel function is selected, the values of $\int k^2(x)dx$ and $\int k^2k(x)dx$ can be determined by calculation. When the kernel function is a quadratic kernel function, the optimal bandwidth is shown in Equation (10):

$$h_{opt} = (8\sqrt{\pi})^{\frac{1}{5}}\sigma n^{-\frac{1}{5}} = 1.7\sigma n^{-\frac{1}{5}} \quad (10)$$

In the equation: σ is the sample standard deviation; n is the total number of traffic accidents.

Let the differential of Equation (3) be zero, then the adaptive bandwidth can be calculated. The mathematical formula is as follows:

$$h_{(x)} = \left\{f(x)\int k^2(x)dx\right\}^{\frac{1}{5}}\left\{\int \hat{f}''(x)^2dx\right\}^{-\frac{1}{5}}n^{-\frac{1}{5}} \quad (11)$$

In the formula: $f(x)$ is the actual nuclear density value.

Since the actual kernel density value $f(x)$ is unknown, this section uses the kernel density estimate to replace the actual kernel density value, and substituting the quadratic kernel function into Equation (11), the adaptive bandwidth can be calculated as follows:

$$\hat{h}(x) = h_{opt}{}^{\frac{9}{5}}c^{\frac{1}{5}}\frac{\hat{f}(x)^{\frac{1}{5}}}{(\sum_{i=1}^{n}\frac{1}{nh_{opt}}t^2k(t) - \hat{f}(x))^{\frac{2}{5}}} \quad (12)$$

In Equation (12): $c = \int \hat{f}''(x)dx = \frac{3}{8}\pi^{-\frac{1}{2}}h^{-5}$

By substituting Equation (12) into Equation (1), the adaptive kernel density estimation is calculated as follows:

$$\hat{f}(x)^* = \frac{1}{n\hat{h}(x)}\sum_{i=1}^{n}K\left(\frac{x - x_i}{\hat{h}(x)}\right) \quad (13)$$

In Equation (13): $\hat{f}(x)^*$ is the adaptive kernel density estimate; n is the total number of traffic accidents; $\hat{h}(x)$ is the bandwidth; K is the kernel function; x_i is the coordinates of the accident point i; x is the center point coordinates.

3.4. Road Hazard Index

For different sections, even if the number of traffic accidents is the same, the danger degree of the location is also different because of the different danger degree of specific traffic accidents. Therefore, in order to distinguish the degree of risk of traffic accidents in different locations, the Road Hazard Index (RHI) is introduced. Through weight assignment, the number of traffic accidents, the number of minor injuries, the number of serious injuries, the number of fatalities, and the direct economic loss in traffic accidents are quantified one by one to determine the risk index of the target road section. The formula for calculating the road hazard index is as follows:

$$RHI = w_1 I + w_2 S + w_3 D + w_4 PDO + TNTA \tag{14}$$

In Equation (14): I is the number of minor injuries in traffic accidents; S is the number of serious injuries in traffic accidents; D is the number of deaths in traffic accidents; PDO is the economic losses in traffic accidents; $TNTA$ is the total number of traffic accidents; w_1, w_2, w_3, w_4 are the weight coefficients of the number of minor injuries, the number of serious injuries, the number of deaths, and the economic loss, respectively.

The selection of w_1, w_2, w_3, w_4 is related to the impact of traffic accidents on society. In a traffic accident, the casualties need to be considered first, followed by economic losses. So, $w_3 > w_2 > w_1 > w_4$. Drawing lessons from the long-term research of domestic and foreign scholars, the weight of the number of minor injuries is 0.5; the weight of severe injuries is 1; the weight of deaths is 3; and the weight of economic loss is 1/30,000 [25], namely:

$$RHI = 0.5I + 1S + 3D + 1/30000\ PDO + TNTA \tag{15}$$

Introducing the road hazard index using the formula of the adaptive kernel density estimation method and combining the quadratic kernel function and bandwidth, the mathematical formula of the traffic accident black spot recognition model is as follows:

$$\hat{f}(x) = \frac{1}{n\hat{h}(x)} \sum_{i=1}^{n} K\left(\frac{x - x_i}{\hat{h}(x)}\right) RHI \tag{16}$$

$$k(x) = \begin{cases} \frac{3(1-x^2)}{4} & |x| \le 1 \\ 0 & |x| > 1 \end{cases} \tag{17}$$

In Equation (17): $\hat{f}(x)$ is the estimated value of the kernel density of a certain road section; h_{opt} is the optimal bandwidth; n is the total number of traffic accidents; K is a quadratic kernel function.

4. Result Analysis

4.1. Black Spot Recognition in Traffic Accidents

Through the POI retrieval interface of the digital map API, crawler technology was used to obtain the WGS-1984 coordinate information of eac traffic accident point and convert its coordinates to the corresponding projection coordinates. At the same time, the projection coordinate information was imported into ArcGIS software for analysis, and then the traffic in Shanghai was obtained. The spatial distribution map of the traffic accident points is shown in Figure 4.

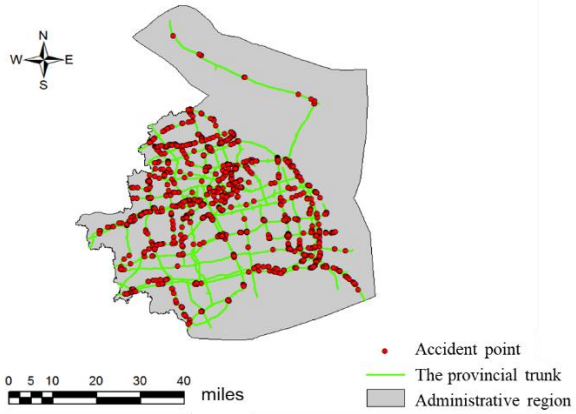
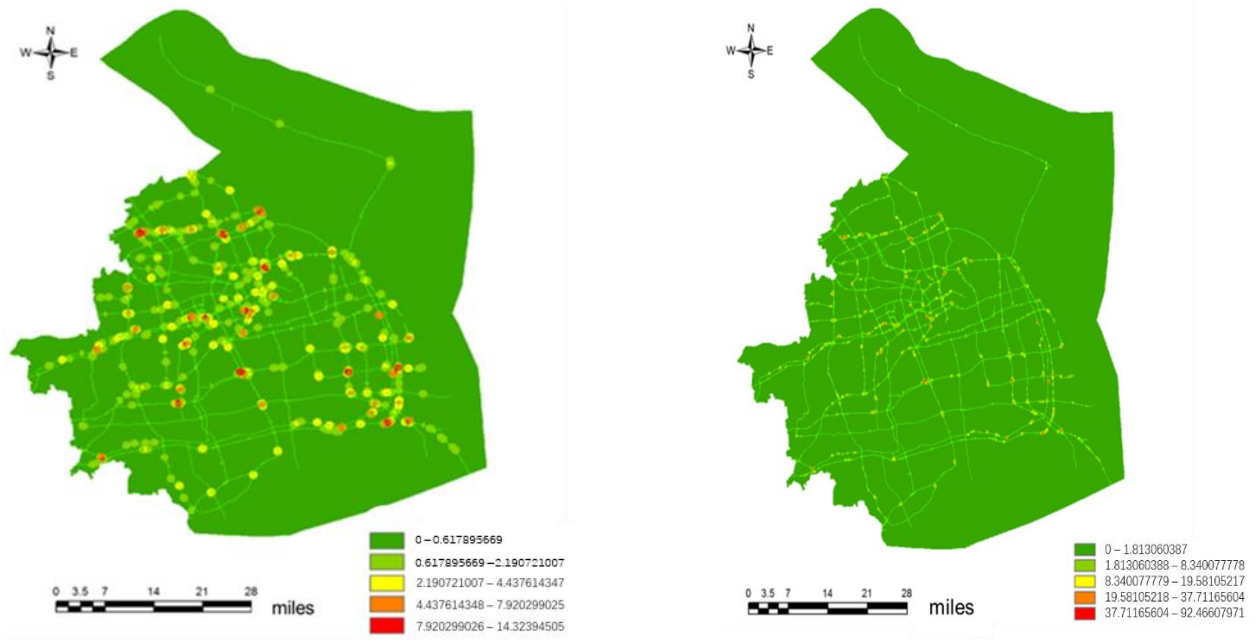


Figure 4. Distribution of Traffic Accidents in Shanghai (2017–2019).

Equation (14) was used to calculate the road hazard index. The minimum value of the road risk index was 1, the maximum value was 17, the average value was 1.58, and the standard deviation was 1.53. At the same time, the optimal bandwidth was calculated to be 458 m using Equation (10). Based on Equation (13), the adaptive kernel density estimation of the national and provincial trunk lines in Shanghai was calculated and Arc-Map 10.2 was used to identify black spots in traffic accidents, as shown in Figure 5. In Figure 5, the greener the color, the smaller the estimated value of nuclear density, that is, the lower the risk of road traffic accidents; the redder the color, the higher the estimated value of nuclear density, that is, the higher the risk of road traffic accidents. The research results show that there were 46 traffic accident black spots, and the estimated value of traffic accident nuclear density was relatively high, with a standard deviation of 1.907, and maximum adaptive kernel density estimate of 189.36.



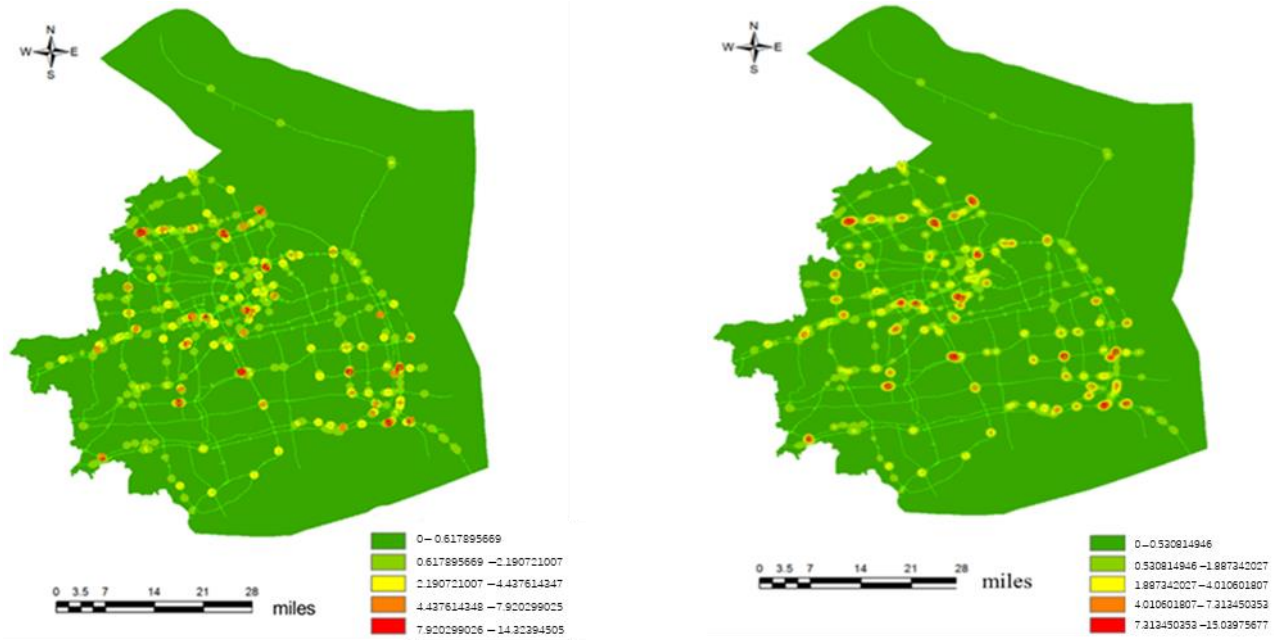
(**a**) (**b**)

Figure 5. Identification results of traffic accident black spots by adaptive kernel density estimation method. (**a**) Recognition result graph. (**b**) Black spot road network distribution.

4.2. Method Comparison

In order to verify the effectiveness of the proposed black spot identification model, the accident frequency method, kernel density estimation method, and traffic accident

identification model were compared. In the process of identifying blackspots in traffic accidents, the search bandwidth was set to 458 m and the blackspot identification results of the two methods are shown in Figure 6. According to Figure 6a, 59 traffic accident black spots were identified using the accident frequency method, with a standard deviation of 0.867 and a maximum traffic accident density of 14.32. According to Figure 6b, a total of 51 traffic accident black spots were identified using the kernel density estimation. At the black spot of traffic accidents, the standard deviation was 1.109, and the maximum traffic accident density was 15.04. Through analysis, it can be found that, compared with the accident rate and nuclear density estimation, the standard deviation of the adaptive nuclear density estimation method was relatively large, which is consistent with the results of Boroujerdian Amin, that is, the larger the variance, the better the recognition effect [26].



(**a**) (**b**)

Figure 6. Accident black spot identification results of the selected comparison method. (**a**) Recognition result of accident frequency method. (**b**) Recognition result of kernel density estimation.

In addition, the first 6 black dots identified by different traffic accident identification black dot methods were sorted, as shown in Table 4. Among the 6 black spots, the maximum kernel density estimated value recognized by adaptive kernel density estimation was 189.36, and the average length of the road section was 226 m, with a maximum road section length of 249 m, and an estimated value of kernel density of 162.43. For the nuclear density estimation method and the accident frequency method, the maximum nuclear density estimation value and the maximum density value were 176.96 and 173.62, respectively.

Table 4. Comparison of Results of Different Traffic Accident Black Spot Identification Methods.

Serial Number	Adaptive Kernel Density Estimation		Kernel Density Estimation	Accident Frequency Method
	Section Length [m]	Kernel Density Estimate	Kernel Density Estimate	Density Value
1	226	189.36	176.96	173.62
2	214	176.91	172.34	161.45
3	249	162.43	160.59	143.78
4	152	134.61	113.21	101.48
5	106	109.28	90.36	86.43
6	83	102.83	83.47	64.81

4.3. Method Testing and Evaluation

4.3.1. Accuracy Test

At present, many methods have been used to test the accuracy of accident black spot identification, such as hit rate index, recovery rate index, and prediction accuracy index. Among them, hit rate is the most commonly used validity test indicator, that is, the ratio of the total number of traffic accidents that occurred in a certain fixed period of time in the identified road section with black spots in the future to the total number of traffic accidents on the target road section. Although the usage of the hit rate index is not only simple and easy to understand, it is easily affected by the total length of the road section and the number of traffic accidents in the study area. For example, when all roads in the entire area to be studied are identified as road sections with black spots in traffic accidents, the hit rate of the effectiveness test index is 100%. However, this method and the recognition results are meaningless to road safety researchers. In view of this, prediction accuracy indicators are also used to test the quality of hot spots. Predictive accuracy index (PAI) refers to the hit rate and the percentage of the area or the percentage defined as the hot research area [27]. This index was later modified by Thakali to measure the predictive ability of collision accidents, and is called the crash predictive accuracy index (CPAI) [26]. However, traffic accidents occur on the road and are road-binding. Therefore, this study adopted the length of the road covered by black spots and the total length of the road in the study area, and introduced CPAI to test the validity of the black spot recognition model of traffic accidents. The mathematical formula expression is as follows:

$$CPAI = \frac{n/N}{l/L} \times 100\% \tag{18}$$

In the formula: n is the number of traffic accidents in the black spot section of traffic accidents; N is the total number of traffic accidents in the study area; l is the length of the black spot section of traffic accidents, and the unit is m; L is the total length of roads in the study area, and the unit is m.

In this study, CPAI was used to test the predictive performance of the three methods (accident frequency, nuclear density estimation, and the traffic accident black spot identification model). The comparison results of different methods are shown in Table 4. It can be seen from Table 5 that the CPAI values of the accident frequency method, the nuclear density estimation method, and the traffic accident identification model were 14.39, 16.36, and 18.25 respectively, and the lengths of the black spots of the traffic accidents were 184.68, 162.45, and 145.57 respectively. From a practical point of view, based on the results of CPAI research, it is shown that compared with the other two methods, the black spot recognition model proposed in this paper had the best recognition effect. The main reason is that the CPAI value of this method was the largest, that is, the identified black spots of traffic accidents were the shortest, accounting for 5.48% of the total length of the road network, and the number of identified traffic accidents was 514, accounting for 26.4% of the total number of traffic accidents. At the same time, the accident frequency method had the worst recognition effect (the CPAI value was the smallest), and the length of the road section with black spots in traffic accidents was 184.68, accounting for 6.95% of the total length of the road network.

Table 5. Comparison of different identification methods.

Method	Traffic Accident on the Black Spot	Total Traffic Accident	Length of Black Spots	Total Road Length	CPAI
Accident frequency method	408	1947	184.68	2657	14.39
Kernel density estimation	471	1947	162.45	2657	16.36
Black spot recognition model	514	1947	145.57	2657	18.25

4.3.2. Evaluation of Identification Methods

From the perspective of applicability, the road repair and maintenance cost limit indicators were used to evaluate the identification effects of different methods. That is, the number of traffic accidents or the largest traffic accident density identified in the shortest road section length was the best recognition effect of this method [13]. The relationship between the total percentage of traffic accidents and the percentage of road length for different traffic accident black spot identification methods is shown in Figure 7. In Figure 7, when considering the safety improvement budget of 20% of the road length, compared with the other two methods, the traffic accident black spot recognition model proposed in this paper could identify about 69% of traffic accidents, namely nuclear density estimation and accidents, at 1.13 times and 1.27 times that of the frequency method, respectively. In addition, when using this method, the length of the road that contained 70% of the traffic accidents accounted for 18.6% of the total length. Therefore, the constructed traffic accident black spot recognition model could identify black spots more effectively and accurately, and could maximize the utilization of road resources under the condition of limited road maintenance and maintenance budgets.

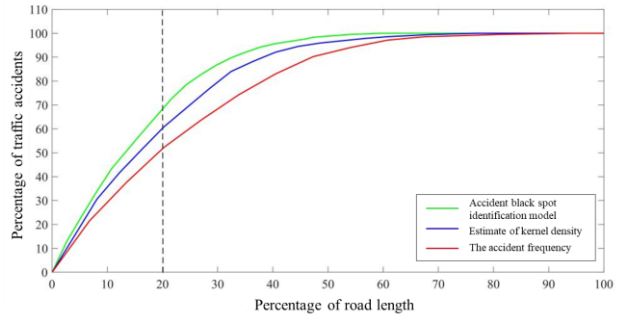


Figure 7. The relationship between the percentage of traffic accidents and the percentage of road length.

5. Conclusions

(1) General

i. Through the collection and analysis of traffic accident data on Shanghai's national and provincial trunk lines, a method for identifying black spots in traffic accidents based on improved bandwidth adaptive kernel density estimation was proposed and road hazard indexes were added as identification parameters to construct traffic accident black spot recognition model. On this basis, using ArcGIS software, the black spot recognition results were compared with the accident frequency method and the nuclear density estimation method, and it is concluded that the adaptive nuclear density estimation method had the highest degree of clustering.

ii. Using CPAI and RRMCLI to verify and evaluate the effectiveness of the three identification methods, it is concluded that the CPAI of the adaptive nuclear density estimation method was 18.25, which was higher than the accident frequency method and the nuclear density estimation. At the same time, considering the safety

improvement budget of 20% of road length, the adaptive kernel density estimation method could identify about 69% of the number of traffic accidents, which was 1.13 times and 1.27 times of the kernel density estimation method and the accident frequency method, respectively.

(2) Innovation

Based on the kernel density estimation theory, the adaptive kernel density estimation method was applied to the field of traffic accident black spot recognition. Combined with the weight assignment method, the number of traffic accidents, the number of minor injuries, the number of serious injuries, the number of deaths, and the direct economic losses were quantified one by one to establish the traffic accident black spot recognition model, which improved the recognition accuracy of the accident black spot.

(3) Deficiencies and Prospects

i. The analysis results have reference value for road safety management and control in Shanghai, and can be further applied to the design of road safety improvement schemes. Meanwhile, the proposal and application of the adaptive kernel density estimation method in traffic accident black spots is conducive to the further expansion of black spot identification methods.
ii. This study only used bandwidth and road hazard index as the main parameters when identifying traffic accident black spots and did not consider that the impact of highway spatial density may be highly related to road network structure and road parameters on accident black spots, such as lane width, road linearity, etc. In addition, the impact of different traffic flows at different sections of the road on the identification of black spots in traffic accidents needs to be further studied.

Author Contributions: Resources, W.Z.; Supervision, H.G. and L.Z.; Writing—original draft, L.D. and M.H.; Writing—review & editing, M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This article is supported by Natural Science Foundation of China under grant 51905224 and Anhui Province Communications Holding Group Co., Ltd. under grant JSKY-2019-01.

Acknowledgments: Thanks to all the authors for their efforts on this article. Natural Science Foundation of China and Anhui Province Communications Holding Group Co., Ltd. are gratefully funded.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sadeghi, A.; Farhad, H.; Moghaddam, A.M.; Qazizadeh, M.J. Identification of accident-prone sections in roadways with incomplete and uncertain inspection-based information: A distributed hazard index based on evidential reasoning approach. *Reliab. Eng. Syst. Saf.* **2018**, *178*, 278–289. [CrossRef]
2. Chen, W. Analysis of road sections with frequent traffic accidents. *J. Chin. People's Public Secur. Univ.* **2004**, *2*, 83–88.
3. Ali, Y.; Zheng, Z.; Haque, M.M. Modelling lane-changing execution behaviour in a connected environment: A grouped random parameters with heterogeneity-in-means approach. *Commun. Transp. Res.* **2021**, *1*, 100009. [CrossRef]
4. Nadimi, N.; Sheikh Hosseini Lori, E. Applying different analytic methods to determine black spots in two-lane highways. *J. Transp. Saf. Secur.* **2022**, *14*, 1395–1418. [CrossRef]
5. Montella, A. A comparative analysis of hotspot identification methods. *Accid. Anal. Prev.* **2010**, *42*, 571–581. [CrossRef]
6. Plug, C.; Xia, J.C.; Caulfield, C. Spatial and temporal visualisation techniques for crash analysis. *Accid. Anal. Prev.* **2011**, *43*, 1937–1946. [CrossRef]
7. Gregoriades, A.; Mouskos, K.C. Black spots identification through a Bayesian Networks quantification of accident risk index. *Transp. Res. Part C Emerg. Technol.* **2013**, *28*, 28–43. [CrossRef]
8. Ambros, J.; Havránek, P.; Valentová, V.; Křivánková, Z.; Striegler, R. Identification of Hazardous Locations in Regional Road Network—Comparison of Reactive and Proactive Approaches. *Transp. Res. Procedia* **2016**, *14*, 4209–4217. [CrossRef]
9. Prieto, F.; Gomez-Deniz, E.; Sarabia, J.M. Modelling road accident blackspots data with the discrete generalized Pareto distribution. *Accid. Anal. Prev.* **2014**, *71*, 38–49. [CrossRef]
10. Thomas, I. Spatial data aggregation: Exploratory analysis of road accidents. *Accid. Anal. Prev.* **1996**, *28*, 251–264. [CrossRef]
11. Meng, X.; Zhang, D.; Wu, P. The influence of sliding window length and sliding step length on the identification of accident-prone sections. *Transp. Syst. Eng. Inf.* **2018**, *18*, 148–155.

12. Ghadi, M.; Török, Á. Comparison Different Black Spot Identification Methods. *Transp. Res. Procedia* **2017**, *27*, 1105–1112. [CrossRef]
13. Elyasi, M.R.; Saffarzade, M.; Boroujerdian, A.M. A novel dynamic segmentation model for identification and prioritization of black spots based on the pattern of potential for safety improvement. *Transp. Res. Part A Policy Pract.* **2016**, *91*, 346–357. [CrossRef]
14. Koorey, G. Road Data Aggregation and Sectioning Considerations for Crash Analysis. *Transp. Res. Rec. J. Transp. Res. Board* **2009**, *2103*, 61–68. [CrossRef]
15. Ghadi, M.; Torok, A. A comparative analysis of black spot identification methods and road accident segmentation methods. *Accid. Anal. Prev.* **2019**, *128*, 1–7. [CrossRef]
16. Yao, S.; Loo, B.P.; Yang, B.Z. Traffic collisions in space: Four decades of advancement in applied GIS. *Ann. GIS* **2016**, *22*, 1–14. [CrossRef]
17. Nie, K.; Wang, Z.; Du, Q.; Ren, F.; Tian, Q. A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China. *Sustainability* **2015**, *7*, 2662–2677. [CrossRef]
18. Harirforoush, H.; Bellalite, L. A new integrated GIS-based analysis to detect hotspots: A case study of the city of Sherbrooke. *Accid. Anal. Prev.* **2019**, *130*, 62–74. [CrossRef]
19. Bíl, M.; Andrášik, R.; Sedoník, J. A detailed spatiotemporal analysis of traffic crash hotspots. *Appl. Geogr.* **2019**, *107*, 82–90. [CrossRef]
20. Flahaut, B.; Mouchart, M.; San Martin, E.; Thomas, I. The local spatial autocorrelation and the kernel method for identifying black zones. A comparative approach. *Accid. Anal. Prev.* **2003**, *35*, 991–1004. [CrossRef]
21. Shafabakhsh, G.A.; Famili, A.; Bahadori, M.S. GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. *J. Traffic Transp. Eng.* **2017**, *4*, 290–299. [CrossRef]
22. Kemp, C.D.; Silverman, B.W. Density Estimation for Statistics and Data Analysis. *J. R. Stat. Soc. Ser. D* **1987**, *36*, 495. [CrossRef]
23. Tang, X.; Luo, M. The optimal choice of non-parametric kernel density estimation kernel function. *J. Qiannan Teach. Coll. Natl.* **2019**, *39*, 16–19.
24. Bil, M.; Andrasik, R.; Janoska, Z. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accid. Anal. Prev.* **2013**, *55*, 265–273. [CrossRef] [PubMed]
25. Peng, Y. *Identification and Cause Analysis of Black Spots in Mountain Expressway Accidents*; Chang'an University: Xi'an, China, 2017.
26. Thakali, L.; Kwon, T.J.; Fu, L. Identification of crash hotspots using kernel density estimation and kriging methods: A comparison. *J. Mod. Transp.* **2015**, *23*, 93–106. [CrossRef]
27. Ulak, M.B.; Ozguven, E.E.; Vanli, O.A.; Horner, M.W. Exploring alternative spatial weights to detect crash hotspots. *Comput. Environ. Urban Syst.* **2019**, *78*, 101398.