

ORIGINAL PAPER

Open Access



GIS-based analysis of spatial–temporal correlations of urban traffic accidents

Qinglu Ma^{1,2}, Guanghao Huang^{1*}  and Xiaoyao Tang¹

Abstract

Background: Understanding the spatial–temporal distribution characteristics of urban road traffic accidents is important for urban road traffic safety management. Based on the road traffic data of Wales in 2017, the spatial–temporal distribution of accidents is formed.

Methods: The density analysis method is used to identify the areas with high accident incidence and the areas with high accident severity. Then, two types of spatial clustering analysis models, outlier analysis and hot spot analysis are used to further identify the regions with high accident severity.

Results: The results of density analysis and cluster analysis are compared. The results of density analysis show that, in terms of accident frequency and accident severity, Swansea, Neath Port Talbot, Bridgend, Merthyr Tydfil, Cardiff, Caerphilly, Newport, Denbighshire, Vale of Glamorgan, Rhondda Cynon Taff, Flintshire and Wrexham have high accident frequency and accident severity per unit area. Cluster analysis results are similar to the density analysis. Finally, the temporal distribution characteristics of traffic accidents are analyzed according to month, week, day and hour. Accidents are concentrated in July and August, frequently in the morning rush hour and at dusk, with the most accidents occurring on Saturday.

Conclusions: By comparing the two methods, it can be concluded that the density analysis is simple and easy to understand, which is conducive to understanding the spatial distribution characteristics of urban traffic accidents directly. Cluster analysis can be accurate to the accident point and obtain the clustering characteristics of road accidents.

Keywords: Traffic accidents, Density analysis, Cluster analysis, Spatial–temporal distribution

1 Introduction

Spatial–temporal distribution characteristics are important attributes of road traffic accidents. In combination with the frequency and severity of road traffic accidents, the spatial–temporal distribution characteristics of road traffic accidents in different regions of the city can be explored. It is helpful for the traffic management department to more intuitively know the distribution area and severity of traffic accidents within jurisdiction, so they can take targeted remedial and preventive measures.

The temporal distribution of road traffic accidents can be analyzed and processed with the help of spreadsheets, and the analysis of spatial distribution characteristics is more complicated.

At present, there are mainly two methods to study the spatial distribution characteristics of road traffic accidents: one is to determine the frequent area of traffic accidents by statistical analysis, which is based on the accident location field in the collected accident information [1]. Another is to visually display traffic accidents by using GIS technology, and then analyze the spatial distribution characteristics by using the spatial analysis method. Compared with the first one (determine the frequent area of traffic accidents by statistical analysis), the advantages of using GIS for analysis are these:

*Correspondence: 622190950007@mails.cqjtu.edu.cn

¹ School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China

Full list of author information is available at the end of the article

- (1) The visual characteristics of GIS can provide a more visual and intuitive understanding of the distribution of traffic accidents, to quickly form an overall grasp of the traffic safety situation in the region.
- (2) The current GIS technology system has developed a variety of spatial analysis tools, which can be used to excavate the spatial distribution characteristics of traffic accidents and the spatial relationship between different traffic accidents from multiple angles, which is difficult to be achieved by simple statistical analysis.

Some scholars have carried out different spatial analysis of traffic accidents based on GIS technology. Erdogan et al. [2] determined the accident points on the highway in the Turkish city of Afyonkarahisar by means of repetitive analysis and density analysis, and the geographical characteristics of the accident spots were analyzed. Based on pedestrian-vehicle collision data, Truong et al. [3] used spatial correlation analysis method to determine the occurrence of pedestrian vehicle accidents and to evaluate the traffic safety of urban bus stations. Colak et al. [4] proposed the hot spot analysis based on network weight and the kernel density method based on accident frequency, which carried out the spatial analysis of traffic accidents in RIZE province of Turkish. Tortum et al. [5] used Moran's I statistic and Getis-Ord $G^* i$ to identify hot spots of road traffic accidents in Turkish cities. Aslam et al. [6] used Analytic Hierarchy Process (AHP) and Point Density (PD) method to predict and verify traffic accident hot spots in Irbilof Jordan, and the distribution of hot spots in urban areas was obtained. Gholam et al. [7] researched the distribution characteristics of traffic accidents in Mashhad city of Iran by means of the nearest proximity and K-means analysis method. Temesgen et al. used the aspects of drivers, pedestrians, peaking time characteristics and other influencing factors and combined with GIS visualization technology. The road traffic accident hot spots in Ethiopia Hosanna Town were researched and corresponding countermeasures were proposed [8]. Wang et al. combined GIS technology with system clustering method, the spatial and temporal distribution characteristics were analyzed and influencing factors of traffic accidents in Guangzhou from the aspects of road characteristics, infrastructure conditions, as well as the proportion of traffic accidents during day and work days [9]. Anderson et al. used kernel density estimation, aggregated K-means clustering and spatial autocorrelation clustering models to carry out the identification of accident-prone points [10]. Fan et al. [11] carried out accident spatial distribution research based on the K-means algorithm in aspects of road sections and intersections, and excavated the black spots of

traffic accidents in Beijing. Zhang et al. [12] took accident information through a mobile App, proposed an improved K-means algorithm to effectively and quickly identify road black spots and analyze the causes of road accidents. Guo et al. used Getis-Ord $G^* i$ hot spot analysis to conduct spatial statistics of the results, the accident prone sections and boundaries were identified. By constructing a large scale Bayesian network model of traffic accidents, the probability of traffic accidents under different factors was calculated [13]. Nie et al. used the improved network kernel density method to detect traffic accident prone sections. Local Moran's I was used to test the results of kernel density analysis, which effectively and accurately located the clustering of traffic accidents in Wuhan [14]. Liu constructed a spatial-temporal network kernel density estimation model that took the severity of traffic accidents into account, which analyzed the spatial temporal hot spots of accident data. Then, the Getis-Ord $G^* i$ hot spot analysis method were used to perform spatial statistics on analysis results, which accurately identified the range and boundary of accident hot spot road sections [15]. Álvaro et al. [16] considered the spatial-temporal clustering of events on the road network, proposed a method of kernel density estimation of spatial-temporal network to detect traffic accident hot spots. Romano et al. [17] used the improved network kernel density estimation as a parameter to identify the accident occurrence points of the threshold by using the method of cumulative frequency and zero-inflated negative binomial regression model. Wang et al. proposed an improved network kernel density algorithm by optimizing the distance between events and the kernel density function of intersections. Then, a zero-inflated negative binomial regression model was used to fit the cumulative frequency distribution of the nuclear density calculation results, which greatly improved the accuracy of identification of accident-prone points [18]. Considering accidents and spatial attributes, Chen [19] constructed a genetic analysis model of hot areas based on logistic regression and spatial data mining and performed an empirical analysis in Enschede, Netherlands.

The above accident research based on GIS spatial analysis has led to a beneficial exploration of the analysis ideas and methods of accident data, but there are still some shortcomings. Firstly, the most intuitive way to measure traffic safety is by the frequency of traffic accidents. Most of the existing literature is based on this idea and focuses on the identification of accident spots. The traffic management department, however, pays more attention to the accidents that cause serious casualties in actual traffic management. Thus, it is important to study the spatial distribution characteristics of regions with high accident severity. Secondly,

the impact of road network density on accident density is not considered in the density analysis. In the cluster analysis, there is a lack of analysis on the clustering mode of the non-aggregate accident points.

Considering the road network density, areas with frequent road traffic accidents and areas with higher severity are identified. Without considering the road network density, areas with frequent road traffic accidents and areas with higher severity are identified. Non-aggregate outlier analysis and hot spot analysis are used to analyze the severity of accidents. The spatial-temporal characteristics of accidents are analyzed. Finally, by comparing the results obtained through the two methods of density analysis and cluster analysis, the applicability of the two methods is analyzed in different scenarios.

2 Research methodology

2.1 Density analysis

In this paper, both point density and line density analysis are used to calculate the density of traffic accident points and the density of road network, respectively. The principle of point density analysis is to calculate the number of data points in a unit area, and the principle of line density analysis is to calculate the length of a line segment in a unit area [20]. The density calculation method usually adopted by GIS software is the neighbourhood method. Taking the calculation of the accident point density as an example, the calculation principle is to divide a city into several small square cells with side length d (corresponding to the pixel unit on the final GIS map). The regional accident density represented by cell k is D_k , road network density is D_r , set neighbourhood radius to ρ , $N_k(\rho)$ is the number of accidents in the neighbourhood with the center of cell k as the dot and ρ as the radius. $L_k(\rho)$ is the length of the road in the same neighbourhood, then D_k and D_r are calculated by Eq. 1.

$$\begin{cases} [D_1^a D_2^a \dots D_k^a] = \frac{1}{\pi\rho^2} [N_1(\rho) N_2(\rho) \dots N_k(\rho)] \\ [D_1^r D_2^r \dots D_k^r] = \frac{1}{\pi\rho^2} [L_1(\rho) L_2(\rho) \dots L_k(\rho)] \end{cases} \quad (1)$$

Calculating the density for each cell by using Eq. 1, and finally get the distribution map of accident density. The calculation of road network density only needs to replace the research object from the accident point with the road section. Most of the previous traffic accident studies focused on the frequency of accident points. In fact, by assigning different weights to different accident points, it is possible to study richer density information. In this paper, the severity of accidents is taken as the weight of each accident point, and then the density analysis is performed to obtain the density distribution of the severity of the traffic accident. Let the severity of the l th accident

in the neighbourhood of cell k be x_l , $l=1, 2, \dots, N_k(\rho)$. Then the accident severity density value D_s of cell k is Eq. 2.

$$[D_1^s D_2^s \dots D_k^s] = \frac{1}{\pi\rho^2} \left[\sum_{l=1}^{N_1(\rho)} x_l \quad \sum_{l=1}^{N_2(\rho)} x_l \dots \sum_{l=1}^{N_k(\rho)} x_l \right] \quad (2)$$

2.2 Cluster analysis

Cluster analysis refers to an analysis process that divides a set of physical or abstract objects into different categories composed of similar objects through certain rules. Spatial clustering analysis is based on the classification rules based on a certain spatial relationship, so as to obtain the spatial distribution characteristics of related objects [21].

Two clustering methods of outlier analysis and hot spot analysis are used to study the spatial distribution of accident severity. The meaning of non-aggregate calculation is that all calculations are based on the attributes of a single accident sample point, not the overall attributes after spatial aggregation [22]. Compared with the aggregate method, the non-aggregate method can retain the original data attributes to the maximum extent, which is conducive to the in-depth study of accident data, but it also requires more computing resources. In addition, unlike traditional methods based on hierarchical or divided clusters, which can only judge whether a sample belongs to a certain category, outlier analysis and hot spot analysis methods can identify samples that do not belong to any cluster or give confidence that the samples belong to a certain category. It is more comprehensive to describe the spatial distribution characteristics of accident points [23, 24].

Outlier analysis determines the correlation between a point and a neighbouring point in the space by calculating the local Moran index I (Local Moran's I) statistic of the data point [22]. The calculation is Eq. 3:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n \omega_{i,j} (x_j - \bar{X}) \quad (3)$$

In the Eq. 3: I_i represents the local Moran index I statistic of data point i , n is the total number of data points, x_i and x_j are the attributes of data point i j (this paper refers to the severity of the accident), and \bar{X} is the attribute global average, $\omega_{i,j}$ is the spatial weight between data point i and other data points j , usually taken as the inverse of the distance between the two points, and s is the second-order sample moment of all data point attributes except the data points. Equation 4 for the S_2 i :

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{x})^2}{n - 1} \quad (4)$$

The z score Z_{I_i} of the data point i can be calculated by Eq. 5

$$Z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \tag{5}$$

Among of which

$$E[I_i] = - \left(\sum_{j=1, j \neq i}^n \omega_{i,j} \right) / (n - 1) \tag{6}$$

$$V[I_i] = E[I_i^2] - E^2[I_i] \tag{7}$$

The generally used statistical significance confidence is 95%, when the p value is less than 0.05, it can be considered statistically significant. According to the normal distribution, the corresponding threshold value of z is ± 1.96 . Under statistically significant conditions, if the value of I is positive, it indicates that the data point has the same high or low attribute value as the neighboring point, and that point is a part of a high–high value cluster or a low–low value cluster. Whether it belongs to high–high clustering or low–low clustering depends on the relationship between the attribute value of this point and the average value of the attributes of all data points. If the value of I is negative, it means that there is a significant difference between the attribute value of the data point and the adjacent point, that is the point is an outlier.

Hot spot analysis is to determine whether the point belongs to the same category as neighboring points by calculating the Getis–Ord G^* statistic of each data point [22]. Getis–Ord G^* statistic G^*i can be calculated by Eq. 8

$$G_i^* = \left[\sum_{j=1}^n (\omega_{i,j}x_j) - \bar{X} \sum_{j=1}^n \omega_{i,j} \right] \cdot \left[S \sqrt{\frac{n \sum_{j=1}^n \omega_{i,j}^2 - \left(\sum_{j=1}^n \omega_{i,j} \right)^2}{n - 1}} \right]^{-1} \tag{8}$$

Among them, S can be calculated by Eq. 9

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{X}^2} \tag{9}$$

The G^*i calculated by Eq. 8 is directly the z score, so no further calculation is needed. Under statistically significant conditions (that is, the z score is greater than 1.96 or less than -1.96), the higher the z score, the closer the cluster of high values (hot spots); the lower the z score, the closer the cluster of low values (cold spots).

3 Data collection and process

In this paper, the road traffic accident data of 2017 in Wales, UK (covering 22 counties and cities) are used. Considering the impact of the accidents and the severity of the accidents, 4629 accident records with summary procedures are selected for analysis (filter out accidents that are not located in Wales). The basic work of accident analysis using GIS technology is the positioning of accident points.

Figure 1a is the accident distribution map after positioning, in which the gray lines represent the road network. To better understand the distribution of accidents in Wales’s counties, Fig. 1b shows the administrative zoning map of Wales.

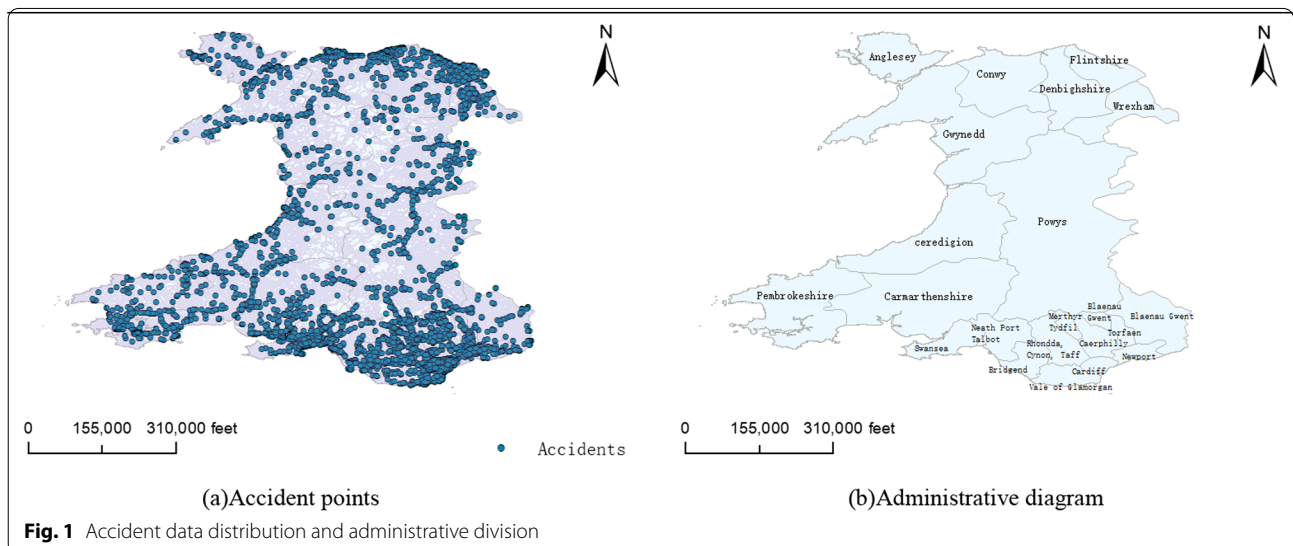


Fig. 1 Accident data distribution and administrative division

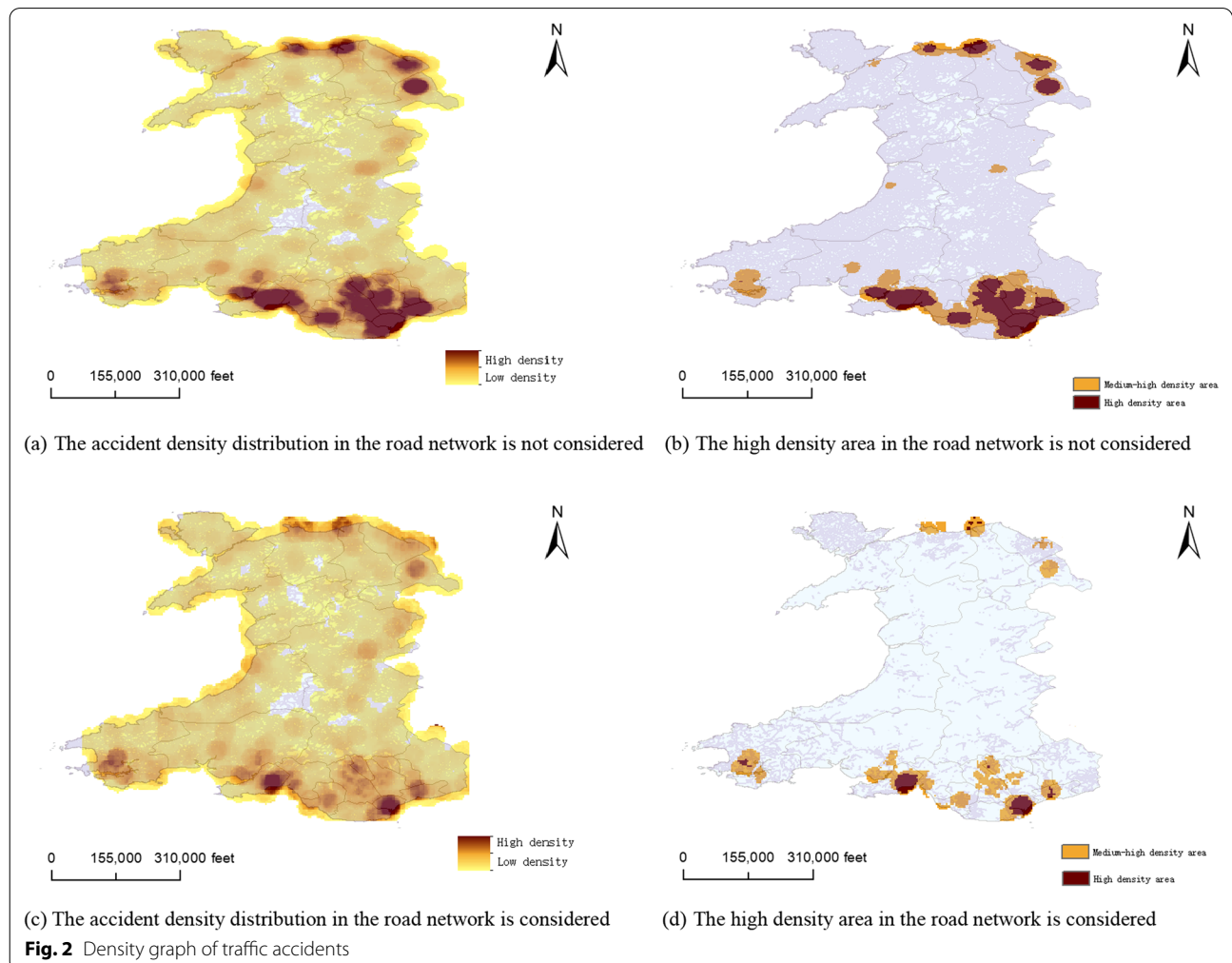
4 Data collection and process

4.1 Spatial distribution characteristics of accidents based on density analysis

An index used to measure the level of road traffic safety in an urban area is the number of traffic accidents in a unit area. According to the calculation method in Eq. 1, the cell length d and neighbourhood radius ρ need to be determined through actual calculation. Comprehensive considerations of the accuracy and efficiency of computing, the cell length and neighbourhood radius are recommended by the GIS software, which are respectively 1/30 and 1/250 of the smaller values of the height and width in the output image range neighbourhood. The latitude and longitude of accident data distribution in this study area are in the $-5.48^\circ \sim -2.65^\circ$ and $51.34^\circ \sim 53.43^\circ$, the latitude and longitude conversion to the actual distance available cell length and the neighbourhood radius after 696 m and 5797 m respectively. For the convenience of comparison, the maximum normalization method is used to normalize the density values obtained. The resulting density

calculations are shown in Fig. 2. Figure 2a, c respectively show the accident density distribution map without considering the road network density and considering the road network density. The darker the color, the higher the accident density is. In addition, on the basis of the density distribution map, an interval with a density of 0.5–0.75 and 0.75–1 is selected as the accident medium–high density area and high density area, as shown in Fig. 2b, d.

Compared with the actual administrative districts, the areas with high accident density are mainly concentrated in Swansea, Neath Port Talbot, Bridgend, Merthyr Tydfil, Cardiff, Caerphilly, Newport, Denbighshire, Vale of Glamorgan, Rhondda Cynon Taff, Flintshire and Wrexham. The density of accident points per unit area within a certain period of time does not fully reflect the frequency of accidents per unit road length. Therefore, in order to exclude the influence factor of road network density, the ratio of the density of accident points to the density of road network are calculated (the ratio of the D_{ak} and D_{rk}). Figure 2c shows the distribution of the ratio



(which can be understood as the number of traffic accidents per unit length). It can be seen from Fig. 2c that the accident frequency distribution after excluding the influence of road network density is somewhat different from the original accident frequency distribution. In Fig. 2a, the central areas of Rhondda Cynon Taff, Merthyr Tydfil, Caerphilly, Newport, Flintshire and Wrexham have a high accident density, while in Fig. 2c, these areas are relatively light in color, indicating that the accident rate per road length in these areas is not high, and the reason for the high density of accident points is that the road network is dense. Cardiff and Swansea, however, are still very dark in Fig. 2c, indicating that the area is an accident density area regardless of the number of accidents per unit area or the number of accidents per unit road length. The densities between 0.5–0.75 and 0.75–1 are also selected as the accident medium–high density area and high density area, as shown in Fig. 2d. By comparison with the actual road network, it can be determined that the area with the higher ratio in the figure is the central area of Cardiff and Swansea.

The above analysis mainly compares the frequency of accidents. The frequency of accidents is only a measure of the severity of traffic accidents, the other is the severity

of accidents. An area where there are occasional particularly serious accidents is often more important than an area where there are frequent minor accidents. Based on the classification principle of existing literature and actual data [25], this paper divides the severity of accidents into 4 grades, and the meaning of each grade is shown in Table 1.

The average accident severity distribution per unit area can be obtained by taking the accident severity as the weight, calculating the weighted density of all accident points, and then dividing by the density of accident frequency. The results for this method are shown in Fig. 3.

Figure 3a is the density distribution of accident severity. The high-density areas in Fig. 3a are selected to obtain Fig. 3b. By comparing Figs. 2b and 3b, it can be found that the density distribution centers of the two maps are quite similar. In Fig. 3b, as described above, the areas of high accident density are concentrated in Swansea, Neath Port Talbot, Bridgend, Merthyr Tydfil, Cardiff, Caerphilly, Newport, Denbighshire, Vale of Glamorgan, Rhondda Cynon Taff, Flintshire and Wrexham.

4.2 Spatial distribution characteristics of accidents based on cluster analysis

The results of traffic accident outlier analysis and hot spot analysis can be obtained by using the clustering analysis tool of GIS software and Eq. 4. The selected feature field is the severity of accidents.

In Fig. 4a, the black dots represent high-severity accidents (high–high cluster); the blue dots represent low-severity accidents (low–low cluster); The yellow dots represent the high-low value category (high–low outlier), that is, the few high-severity accident categories contained in the space occupied by many low-severity

Table 1 Injury severity level of traffic accidents

Level	Standard
1	All other accidents
2	There were no fatalities or serious injuries but minor injuries
3	There were no fatalities but serious injuries
4	There were fatalities

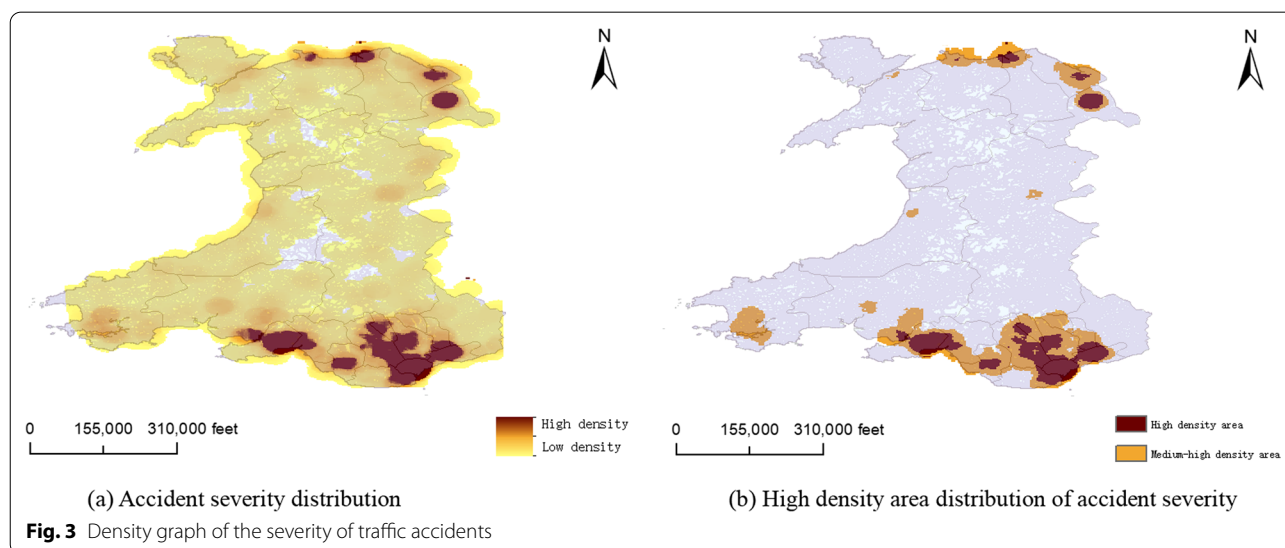
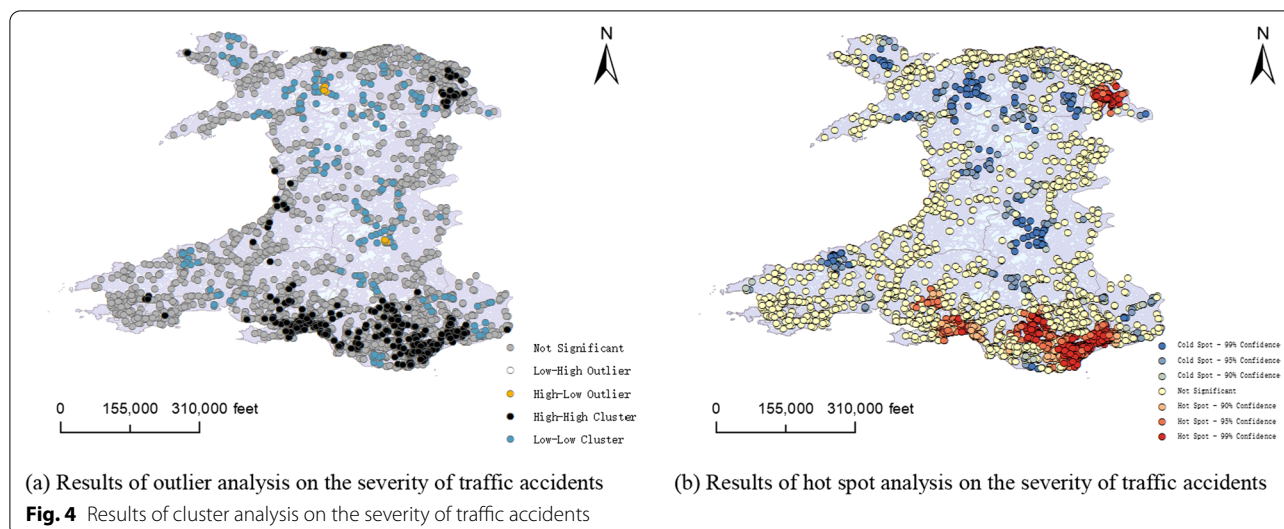


Fig. 3 Density graph of the severity of traffic accidents



accident spots; The white points represent the low–high-value category (low–high outlier), that is, the few low-severity accidents included in the space occupied by many high-severity accidents; The grey points represent the points without obvious clustering features (not significant). As can be seen from Fig. 4a, in the central areas of Swansea, Neath Port Talbot, Bridgend, Merthyr Tydfil, Cardiff, Caerphilly, Newport, Vale of Glamorgan, Rhondda Cynon Taff, Flintshire and Wrexham, traffic accident points show a clustering trend in the feature of accident severity. While in Conwy, Anglesey, Blaenau Gwent and Powys, traffic accidents show a low clustering distribution.

Figure 4b shows the clustering distribution of traffic accidents obtained through the hot spot analysis. The red dots in Fig. 4b are called “hot spots”, representing high-severity accidents; the blue dots are called “cold spots” and represent low-severity accidents; the yellow dots are the ones that are not distinctive. For “hot” and “cold” spots, the shades of color represent different confidence levels, and the darker the color, the higher the confidence level that the point belongs to the corresponding category. Hot spot analysis only focuses on the aggregation shape formed by the sample according to the level of the target eigenvalue, and does not detect outliers. Therefore, the results of hot spot analysis will reflect more features of regional distribution. By comparing Fig. 4a, b, it can be seen that outlier analysis and hot spot analysis have similar clustering results in the areas with high severity accidents and low severity accidents. The difference lies in that the outlier analysis method is to determine the values that do not conform to clustering features as insignificant or outliers, while the hot spot analysis method is to determine the values that do not conform to clustering

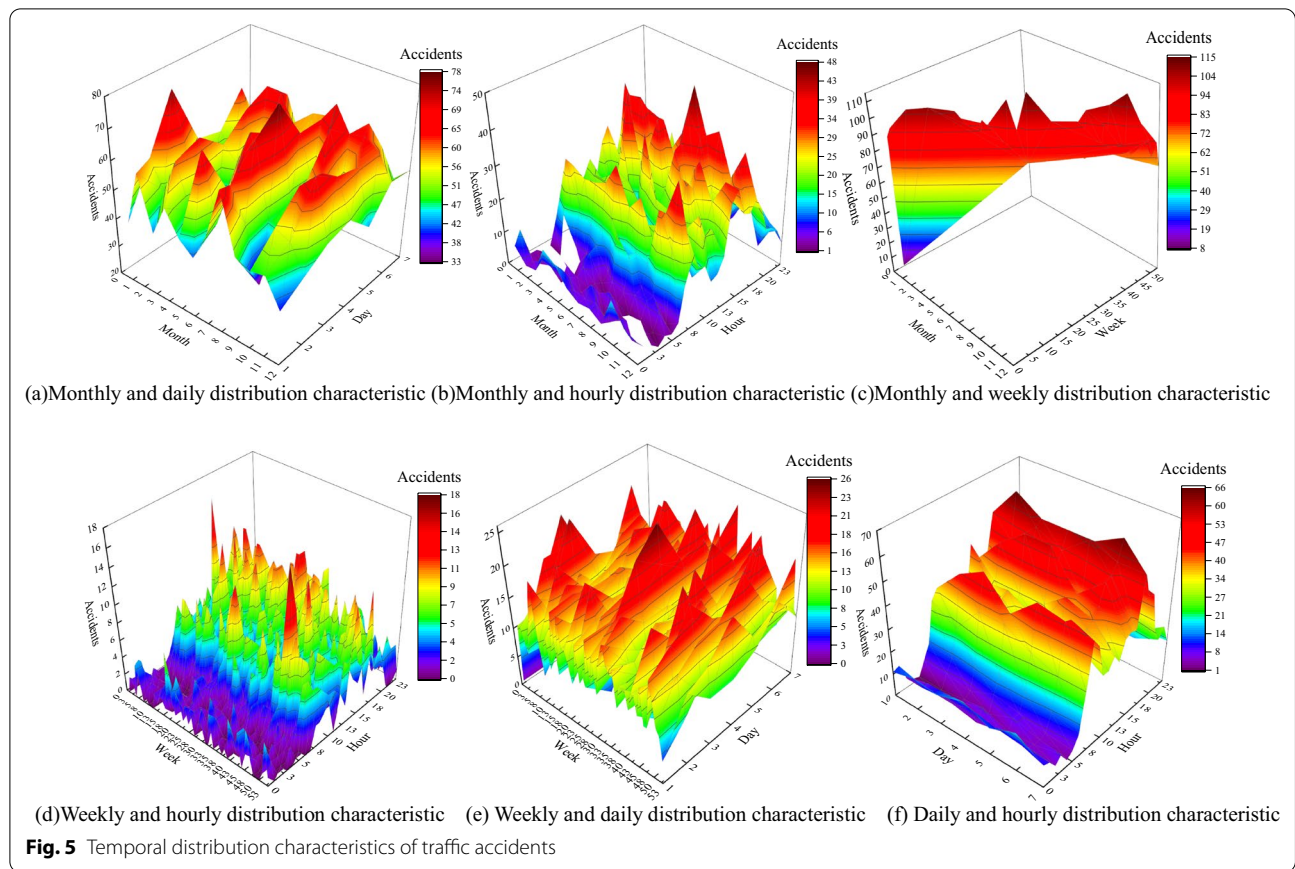
features as insignificant or express their uncertainty with a certain confidence level.

It can be seen from results of density analysis and cluster analysis, the traffic accidents in Cardiff, Swansea, Caerphilly, Newport, Vale of Glamorgan, Rhondda Cynon Taff, and Wrexham are densely distributed. Cardiff is the capital of Wales. Although the city is small, its road network is dense and prone to traffic accidents. As the home stadium of the Wales football team, the Millennium Stadium is relatively large and can accommodate a large number of spectators, which is one of the causes of traffic accidents. Swansea is the second largest city in Wales. The number of accidents is concentrated in the north, where accidents are concentrated in low-grade roads. As an important trading port city in Wales, Newport is famous for its many bridges and accidents mainly occur in urban area. Rhondda Cynon Taff’s traffic accidents are mainly concentrated on higher-grade sections, such as expressways and highways. Half of the population of Wrexham County is concentrated in the city of Wrexham. Expressway A483 is the main road to and from Wrexham and is also a concentrated section of traffic accidents.

4.3 Temporal distribution of accidents

The temporal distribution of traffic accidents has obvious aggregation characteristics, so it is helpful to understand the high incidence of traffic accidents as a whole. Based on data of traffic accidents in 2017 of Wales, the temporal characteristics are analyzed according to hours, days, weeks and months.

It can be seen from Fig. 5b, d, f that 7:00–9:00 is the first peak of the traffic accident curve, and 17:00–18:00 is the second peak of the traffic accident curve. The first



peak period coincides with the early peak period of traffic flow, indicating that the peak period of traffic flow is the period of the highest traffic accidents. The second peak lasts from 17:00 to 18:00 which coincides with the peak traffic flow at dusk.

Figure 5a, c, e also shows the daily and weekly and monthly distribution characteristics. As can be seen from the distribution of days, the respective number of accidents Saturday is more than others' days throughout the year. As can be seen from the characteristics of the weekly distribution, the numbers of accidents are more than 100 in some weeks, such as 6th, 10th, 13th, 14th, 15th, 25th, 28th, 29th, 40th, 44th and 46th weeks of the year. It can be seen from the monthly distribution characteristics that there are fewer accidents in February, April, June and December throughout the year, with the most accidents occurring in July. July and August are the time when the temperatures are the highest throughout the year.

5 Comparison and analysis of methods

Density analysis and cluster analysis have obvious differences in the recognition of traffic accident-prone areas and the recognition of high-density areas. In

order to further study these two methods, the consistency and computational efficiency in results of the density analysis and cluster analysis are calculated. In terms of consistency, the calculation equation is as Eq. 10.

$$p_q = \frac{|S_q \cap D|}{|S_q|}, q = 0, 1, 2 \tag{10}$$

In Eq. 10: S_q is the set of accident points in Wales, when $q=0$, S_0 is the original Wales accident points set; when $q=1$, S_1 is the accident points set of the high severity accident cluster obtained by outlier analysis; when $q=2$, S_2 is the accident points set of the high-severity accident cluster obtained by hot spot analysis. $|S_q|$ is the number of elements in set S_q . D is the area above high density of severity obtained by density analysis, P_q is the proportion of accident points covered by D in S_q to the total number of accidents, when $q=0$, P_0 is the proportion of accident points covered by D in S_0 to the total number of accidents; when $q=1$, P_1 is the proportion of accident points covered by D in S_1 to the total number of accidents; when $q=2$, P_2 is the proportion of accident points covered by D in S_2 to the total number of accidents.

Table 2 Proportion of accidents covered by the areas with above high severity

Method	Density analysis	Hot spot analysis	Outlier analysis
Accident point ratio	P_0	P_1	P_2
Coverage ratio (%)	35	40	42

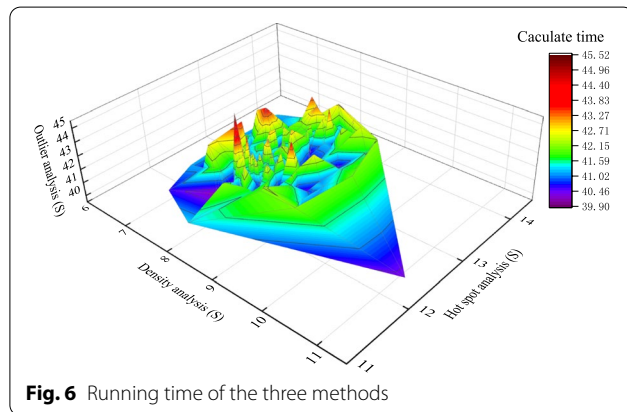


Fig. 6 Running time of the three methods

If the results of the two methods have a good consistency, then the accident points in the cluster of high-severity accident points obtained by clustering analysis should be covered by D as far as possible, because P_1 and P_2 should be greater than P_0 . The intersection tool in GIS system is used to calculate the above three proportions, as shown in Table 2.

It can be seen from Table 2 that both P_1 and P_2 are significantly greater than P_0 , indicating that the results of clustering analysis method can be well consistent with the structure of density analysis. The main reason why P_1 is less than P_2 is that the hot spot analysis cannot distinguish the abnormal point value, so some non-serious accidents may be classified as serious accidents, resulting in the increase of $|S_1|$ and the subsequent decrease of P_1 .

In terms of calculation efficiency, the calculation time required by the three methods is recorded respectively (that is, the time required to obtain Figs. 2a, 3, 4 from Fig. 1). Each method is repeated for 300 times respectively, and the single calculation time change are obtained, as shown in Fig. 6.

As shown in Fig. 6, the density analysis method takes the shortest time, and the hot spot analysis is between the density analysis and the outlier analysis. The time needed for outlier analysis is not only more than density analysis method, also significantly greater than belong to hot spot analysis method of clustering algorithm, this may be related to abnormal value analysis compared with hot spot analysis step calculation of z score more steps. The average running time, variance, standard deviation and relatively range of the three methods obtained from 300 experiments are shown in Table 3.

In general, the density analysis is simple and easy to understand, and the general information about the spatial distribution of accidents can be obtained without a complex algorithm, which can help the traffic management department to form an intuitive and quick understanding of the spatial characteristics of urban traffic accident distribution. The disadvantage is that the analysis results can only reflect the rough distribution of accident severity. The results of cluster analysis are accurate to accident points, which can identify outlier points in the severity or give the credibility of accident point clustering result, such as many high-low value points identified in outlier analysis result. It is detailed comparison information of roads or blocks, which can provide support for refined traffic safety management. However, the algorithm is difficult to understand and the actual calculation efficiency is relatively low.

6 Conclusions

Although both density analysis and cluster analysis have been practised in the spatial analysis of traffic accidents, there are still some limitations in these methods. For example, the impact of road network density on accident density is not considered in density analysis; in cluster analysis, there is a lack of analysis of non-aggregate accident point clustering mode. In addition, the comparison of the applicability of these two methods is rarely mentioned in the existing literatures. This paper proposes a new density analysis method which consider the road network density or not. Then, compares the difference in regional distribution under the two conditions and analyzes the possible reasons for the difference. At the same time, two spatial clustering models of disaggregated

Table 3 Average running time of the three methods

Analysis methods	Density analysis	Hot spot analysis	Outlier analysis
Average running time (s)	9	13	42
Variance (s)	0.24	0.25	0.72
Standard deviation (s)	0.49	0.51	0.85
Relatively range (%)	52.28	21.57	13.54

outlier analysis and hot spot analysis are proposed to further identify areas with higher accident severity. Finally, compares the results obtained by the two methods of density analysis and cluster analysis and analyzes the applicability of the two methods in different scenarios. The conclusions obtained are as follows:

- (1) In this paper, two spatial analysis methods, density analysis and cluster analysis, are used to study the spatial distribution characteristics of the frequency and severity of traffic accidents in Wales. The study shows that the frequency of road traffic accidents in Swansea, Neath Port Talbot, Bridgend, Merthyr Tydfil, Cardiff, Caerphilly, Newport, Denbighshire, Vale of Glamorgan, Rhondda Cynon Taff, Flintshire and Wrexham are high, and considering the density of the road network, the accidents in Cardiff and Swansea are still frequent. The time distribution characteristics of traffic accidents were statistically analyzed through the chart: the traffic accidents occurred frequently in the morning rush hour and at dusk in 1 day, with the most accidents occurring on Saturday of the week, and accidents occurred more frequently in July and August.
- (2) In terms of accident severity, the preliminary causative analysis indicated that the differences in road grades and traffic protection measures between counties are the main reasons for the differences.
- (3) Through the comparative analysis of the results obtained by the two methods, it can be seen that the accident spatial distribution characteristics obtained by the two methods are basically consistent, the clustering analysis method can provide more spatial feature information than the density analysis method, and the density analysis method is superior to the clustering analysis method in the simplicity of principle and computational efficiency.
- (4) Due to the limitation of the accidents data and the limitation of the spatial analysis method, the analysis of causes of accidents in different spatial characteristic areas is relatively brief. In the next step, we can combine multi-source data such as traffic flow and street view pictures to deeply explore the causes of all kinds of traffic accidents, so as to guide urban road traffic safety management more specifically.

Acknowledgements

First of all, I would like to extend my sincere gratitude to my tutor, Qinglu Ma, for his instructive advice and useful suggestions on my thesis. I am deeply grateful of his help in the completion of this thesis. I am also would like to thank the institution for providing me with the funding for my research, which can provide me with a comforting research environment. I am also deeply indebted to all the other tutors and students in translation studies for

their direct and indirect help to me. Finally, I am indebted to everyone who contributed to this article.

Authors' contributions

QM provided the general idea of the whole submission, pointed out the problems in this paper and provided guidance and solutions. GH wrote the whole article and made a simulation experiment to get the results. XT sorted out the article format, revised the basic grammar and errors. All authors read and approved the final manuscript.

Funding

The project was funded by Science and Technology Research Program of Education Commission of Chongqing Municipality, (KJZD-K202000704): Theory and practice of self-organizing optimization of interwoven zone traffic block chain under intelligent network environment; Technology Foresight and Institutional Innovation Project of Chongqing Bureau of Science and Technology (cstc2019jsyj-yzysba0058): Theory and Practice of traffic block chain organization optimization in dense interwoven areas from the perspective of big data; National Social Science Foundation of China national Emergency Management System Construction Research Project (20VYJ023): Research on ways to improve the quality and upgrade the comprehensive urban emergency management capability under major emergencies.

Availability of data and materials

We obtained the public data set from the following URL. <https://ckan.publi.shing.service.gov.uk/dataset>

Declaration

Competing interests

The authors declare that they have no competing interest.

Author details

¹School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China. ²Chongqing Key Laboratory of "Human-Vehicle-Road" Cooperation and Safety for Mountain Complex Environment, Chongqing 400074, China.

Received: 12 January 2021 Accepted: 30 August 2021

Published online: 10 September 2021

References

1. Cheng, W., & Washington, S. P. (2005). Experimental evaluation of hot spot identification methods. *Accident Analysis & Prevention*, 37(5), 870–881.
2. Erdogan, S., Yilmaz, I., Baybura, T., et al. (2008). Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis & Prevention*, 40(1), 174–181.
3. Truong, L. I., & Somenahalli, S. (2011). Using GIS to identify pedestrian vehicle crash hot spots and unsafe bus stops. *Journal of Public Transportation*, 14(1), 99–114.
4. Ebru Colak, H., Memisoglu, T., et al. (2018). Hot spot analysis based on network spatial weights to determine spatial Statistics of traffic accidents in Rize, Turkey. *Arabian Journal of Geosciences*, 11(7), 1–11.
5. Tortum, A., & Atalay, A. (2015). Spatial analysis of road mortality rates in Turkey. *Proceedings of the Institution of Civil Engineers-Transport*, 168(6), 532–542.
6. Al-Omari, A., Shatnawi, N., Khedaywi, T., & Miqdady, T. (2020). Prediction of traffic accidents hot spots using fuzzy logic and GIS. *Applied Geomatics*, 12(2), 149–161.
7. Shafabakhsh, G. A., & Famili, A. (2017). GIS-based spatial analysis of urban traffic accidents: case study in Mashhad, Iran. *Journal of traffic and transportation engineering (English Edition)*, 4(3), 290–299.
8. Hayidso, T. H., Gameda, D. O., & Abraham, A. M. (2019). Identifying road traffic accidents hotspots areas using GIS in Ethiopia: a case study of Hosanna Town. *Transport and Telecommunication*, 20(2), 123–132.
9. Li-ya, W., & Xiao-wen, Li. (2019). Urban road traffic accident black spot identification based on GIS. *Communications Science and Technology Heilongjiang*, 42(11), 212–215.

10. Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3), 359–364.
11. Peng-Cheng, F., & Xian-yuan, D. (2019). Analysis of common features of accidents prone spots in urban roads network and corresponding application based on spatial clustering. *Transportation Science and Technology*, 43(4), 85–89.
12. Jun-hui, Z., & Tuo, S. (2019). Spatial analysis of traffic accidents based on Wave Cluster and vehicle communication system data. *EURASIP Journal on Wireless Communications and Networking*, 1, 1–10.
13. Lin, G., Ji-biao, Z., et al. (2018). Analysis of urban road traffic accidents based on improved K-means algorithm. *China Journal of Highway and Transport*, 31(4), 270–279.
14. Ke, N., Zhen-sheng, W., et al. (2017). Research on spatial cluster analysis of traffic accident ignorance road network constraints. *Geomatics World*, 24(6), 50–56.
15. Yao, L. (2019). *Analysis and identification of spatial temporal hotspots and exploration of risk factors for traffic accidents*. Zhejiang University.
16. Briz-Redón, Á., Martínez-Ruiz, F., & Montes, F. (2019). Spatial analysis of traffic accidents near and between road intersections in a directed linear network. *Accident Analysis & Prevention*, 132, 1–56.
17. Romano, B., & Jiang Z. (2017). Visualizing traffic accident hotspots—based on spatial—temporal network kernel density estimation [C]//Far-noush Banaei-Kashani. In *ACM Sigspatial international conference on advances in Geographic information systems*. Redondo Beach, California, USA: Morgan; Claypool does (pp. 98–98).
18. Ying-zhi, W., & Li-jun, W. (2019). An identification method of traffic accident black point based on street-network spatial-temporal Kernel estimation. *Scientia Geographica Sinica*, 39(8), 1238–1245.
19. Jin-lin, C. (2015). *Research on identification hotspots in the urban road networks based on the network kernel density estimation method*. School of Transportation Southeast University.
20. Environmental Systems Research Institute. ArcGIS Desktop Help 10.2 Geostatistical Analyst [DB/OL]. Retrieved December 3, 2014, from <http://resources.arcgis.com/zhcn/help/main/10.2/index.html>.
21. Noland, R. B., & Quddus, M. (2004). A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention*, 36(6), 973–984.
22. Songchitruksa, P., & Zeng, X. (2010). Getis–Ord spatial statistics to identify hot spots by using incident management data. *Transportation Research Record*, 21(65), 42–51.
23. Siddiqui, C., & Choik, A.-A. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382–391.
24. Mussone, L., Bassani, M., & Masci, P. (2017). Analysis of factors affecting the severity of crashes in urban road intersections. *Accident Analysis & Prevention*, 103, 112–122.
25. Dabbour, E., Easa, S., & Haider, M. (2017). Using fixed-parameter and casual-parameter ordered regression models to identify significant factors that affect the severity of drivers' groups in vehicle-train collisions. *Journal of Accident Analysis & Prevention*, 107, 20–30.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
