Formulating an Innovative Spatial-Autocorrelation-based Method for Identifying Road Accident Hot Zones

Yiwei Feng¹, Wenjun Zhu²

¹School of Highway, Chang’an University, Xi’an, ShaanXi, 710061, China
²Central and Southern China Municipal Engineering Design & Research Institute Co., Ltd, Xi’an, ShaanXi, 710061, China

*Corresponding author’s e-mail: 892178863@qq.com

Abstract. The traditional method of identifying road accident hot zones is through the examination of accident frequency and nature, which sometimes, can be subjective and inaccurate. To overcome the limitation of the traditional method, researchers have applied Geographic Information System (GIS) approaches to identify and visualise road traffic accidences in real-time. However, these approaches still treat accidences as occasional and discrete events and can not support accurate analysis and prediction of accidences at some point. This paper takes the spatial autocorrelation nature of accidents (i.e. the interdependence of accident data and the relationship between the accident and space) into account and proposes an innovative spatial-autocorrelation-based method to identify freeway accident hot zones. Based on the spatial autocorrelation and mathematical statistics, this method constructs a point-line connectivity network to realise the space localisation and validation of accidents. Combined with GIS approaches, our approach can also automatically identify and visualise accident-prone areas. At the moment, the approach has been tentatively applied in a highway in China. The result demonstrates an algorithm behind the approach, which can effectively convert accident data into spatial data, cluster accident hot zones of any length and predict the whereabouts of likely accidents in the future. In conclusion, the robustness and accuracy of the approach innovates this study.

1. Introduction

Road traffic safety has become a critical problem nowadays. Approximately 1.35 million people die each year as a result of road traffic crashes, and more than half of these deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists. Road traffic injuries cause considerable economic losses to individuals, their families, and nations as a whole. Improved road infrastructure, lowered speed limits, safer vehicles, graduated licensing, and a range of successful programs targeting drink driving, seatbelt usage and speeding can reduce car crashes and the impact. Despite these achievements, issues around road safety are complex, myriad and changing. For example, smartphone was used while driving wasn’t an issue before. Now, it’s a growing problem. Road crashes continue to cause large numbers of deaths and serious injuries each year, and the impacts are socially and economically devastating [1].

This paper contributes to the development of road safety by improving the capacity of identifying and predicting accident-prone zones/hotspots. Accident-prone refers to a road section where a significantly higher volume of accidents or near-miss characteristics is observed within a period, compared against other road sections at the same time [2,3]. This way, a road network can be divided into
several accident-prone and non-accident-prone sections, and the investigation can be placed around examining the characteristic difference and formulating viable identification solutions.

2. A Review of the State-of-the-art Accident Identification Methods

In terms of GIS in road traffic safety, Zhu [4] used the number of equivalent accidents as a basis for discrimination, which is based on map operations in the automatic search of accident hotspot; Zhang [5] applied the principal component analysis model to identify the accident hot zone, which uses a large amount of spatial geographic data and traffic accident data as the base data. The method displays the road accident hotspots on the map in different sizes and colours, which is according to the relevant indicators of the accident location (such as the number of deaths, the number of injured, economic losses) and is made by hot zone analysis system; Erdogan et al. [6] and Gundogdu [7] converted the original accident data into tabular data, and then displayed the accident hotspot on the traffic road network map. They identified the same incident hotspot on the GIS map by using different methods: one is the Poisson statistical method and the other is the Bernoulli statistics (nuclear density analysis) method; Sando [8] proposed an accident analysis tool which is based on the cost-benefit ratio method of GIS. The accident data in the traffic collection system were combined with the spatial geographic data, and then the cost-benefit ratio method was used to analyze the accident-prone spot.

In view of the shortcomings in the traditional identification method based on accident history data, this study needs to overcome the limitations of its complicated algorithm, high application difficulty, rich requirements for accident data and insufficient visibility. Therefore, this paper tries to think that the accident data on the same accident area has potential interdependence, and combines the spatial autocorrelation theory to improve the traditional definition of the accident hot zone. Next, this paper proposes a method for identifying highway accident-prone road sections, based on the accident and spatial attributes. This method uses the spatial autocorrelation and mathematical statistics as the theoretical basis to construct the “point-line connection” system algorithm to achieve the validation of the space of the accident location. The GIS-related technology is combined to realize the automatic identification and visualization of the accident hot zone. Finally, this paper tries to apply this method to a highway in western China to verify its discriminating effect.

3. Data Preprocessing Method

3.1. Validation of the highway accident spot space

This paper proposes an innovative spatial data model to effectively identify and validate the location of traffic accidents. In this study, a point-to-line connection system (using non-buffered surface nodes) is formulated. The system can assign an accident to its nearest node in the road network and avoid accident coding information errors (namely, low-precision problems due to incorrectly assigned road traffic accidents to buffers and intersections). The improvement of the space effective accuracy of traffic accidents makes further accident analysis possible, for example, the form of accident statistics can be expanded from single-dimensional points/lines to multi-dimensional planes.

The space accident spot validation system proposed in this study is based on the accident data table and the road network database. ArcGIS (version 8.2) and ArcObjectTM are used for spatial data integration, spatial visualisation, and traffic accident spot and road centerline alignment (ArcGISTM contains Microsoft® Visual BasicTM coding function modules for customising new functional modules). The operation flow is shown in Figure 1.
3.2. Basic analysis unit and accident statistics
After validating the space of the accident spot, the next step is to define the basic analysis unit and perform the corresponding accident statistics analysis. Since most of the accident location information is recorded by the nearest mileage station, this paper takes a 100-meter pile as a fixed-length section of the expressway and generates “m” numbers of basic analysis units. Then the cross-tool of the spatial statistical analysis module in ArcGIS is used to realise the statistical work of traffic accident frequency in each basic analysis unit, and the accident frequency is recorded as $a_i$, $i = 1, 2, ..., n$. If a traffic accident occurs at the intersection of the basic analysis unit, the module can assign it to one of the units according to the established principle to avoid repeated counts. When a traffic accident occurs near two basic analysis units, the average values of two basic analysis units can be compared and the basic analysis unit with a smaller average number of stations can be used for statistics.

4. High-speed Road Accident Hotspot Identification Method

4.1. Spatial analysis of accidents
At present, the classification of spatial correlation in China and foreign research mainly includes the following three types: spatial correlation, spatial negative correlation and spatial uncorrelation. The three correlations mainly depend on the relationship between the similarity of each spatial observation value and the spatial distance. When the similarity of spatial observations is stronger with the reduction of spatial distance, it is called spatial positive correlation; when the similarity of spatial observations is weaker with the reduction of spatial distance, it is called spatial negative correlation; When the similarity of spatial observations does not change significantly with the reduction or increase of spatial distance, it is called spatial irrelevance. The calculation formula of spatial correlation is as follows:

$$X_i = \rho \sum_j w_{ij} x_j + u_i$$  \hspace{1cm} (1)
In the formula, $\rho$ represents spatial autocorrelation coefficient between $x_i$, $w_{ij}$ represents the distance relationship between the research points, usually a function of distance, ie $w_{ij} = d_{ij}^{-\alpha}$, $u_i$ represents an obeying independent and identically distributed error term, the coefficient of variation is $\sigma^2$.

The analysis based on spatial autocorrelation can explain the correlation between the value at the research point $X$ and its continuous neighboring point value. Based on the Moran’s index, the accident hot zone identification method can fully analyze the difference between an accident analysis unit $i$ and its adjacent accident analysis unit. The Moran’s index\[^{[10]}\] (LISA) of the accident unit $i$ is calculated as follows:

$$I_i = (x_j - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})$$

In the formula, $x_i$ is the total number of traffic accidents at the accident analysis unit $i$, while $\bar{x}$ is the mean of the number of accidents at the accident analysis unit $i$ and its adjacent accident traffic spot; $W_{ij}$ is a position space weight matrix, which reflects the spatial distance relationship between the accident analysis unit $i$ and its adjacent accident analysis units based on the road network; The value of $I_i$ can be either positive or negative or zero. If the value of $I_i$ is positive, it indicates that there is a similarity between the accident spot $i$ and its adjacent accident locations (that is, the same value is high - high value or low - low value).

### 4.2. Spatial Weight Matrix

According to the number of accidents and the spatial weight matrix of each accident research point, the LISA value can be calculated by formula (2). The LISA value can be used to evaluate the risk degree of accidents in various sections of the road network. The LISA value is positively correlated with the concentration of accidents and the probability of occurrence. In addition, the parameters $Z$ and $P$ values of the response significance level can be used to test the significance level of each analysis unit of each study point $i$. When $Z$ is positive, it indicates that $i$ and the adjacent study points have similar high-value distribution. When $Z$ is a negative value, it indicates that there is a concentrated low-value distribution between $i$ and the adjacent research points.

In this paper, the $Z$ value and the LISA value are both positive values when the significance level is $5\%$. Meanwhile, there is a set of adjacent accident spots with high traffic accidents in each basic analysis unit, and the accident hot zone can be identified (as shown in Figure 3).

**Figure 2. LISA result classification**

### 5. Case Study and Result Analysis

#### 5.1. Data Sources

In order to verify the identification effect of the accident hot zone identification method considering the spatial attributes proposed in this paper, the accident statistics of the highway section K43+0.000 to
K71+0.000 in the western part of China from 2012 to 2016 are analyzed. During the research period, there were 1118 traffic accidents and 280 accident records.

5.2 Data preprocessing and hot zone identification

Because the traffic police recorded the accident information with the accident spot approaching the 100-meter station, this paper records the basic unit of the road section with a length of 100 m. Due to the inevitable fragmentation problem, 2816 are finally generated. The basic analysis unit is roughly equal to the number of routes with a total mileage of 28km divided by 100 meters, and its accuracy is 99.42%. Firstly, the space accident space is validated, and then the number of traffic accidents of each basic analysis unit is counted, and the spatial statistical analysis tool of ArcGIS is used to identify the accident hot zone. The result is shown in Fig. 4.

![Figure 3. accident hot zone](image)

From Figure 3, we can see that at the 95% significance level, 42 of 280 accident spots are identified as accident hotspots. The specific parameters of the accident hot zone are shown in Table 1.

| Table 1. Accident hot zone statistics |
|--------------------------------------|
| **Entire road basic unit** | **Record the accident spot** | **Accident hot zone** |
| Total number of units | 2816 | 280 | 42 |
| Total number of traffic accidents (from) | 1118 | 1118 | 281 |
| Average number of traffic accidents per unit (start / accident spot) | 0.40 | 3.99 | 6.69 |
Minimum number of traffic accidents per unit (from)  0  1  4
Maximum number of traffic accidents per unit (from)  13  13  13

In Table 1, the number of traffic accidents in the accident hot zone is as small as 4, up to 13, with an average of 6.69/unit, and the number of traffic accidents is significantly higher than the entire road. It can be seen from the results of the accident hot zone that the condition of an accident spot being determined as an accident hot zone is that a large number of traffic accidents have occurred and the frequency of accidents at surrounding accident spot is high. Meeting the above two conditions can be defined as the accident hot zone. Through the above analysis of the hot zone identification case, it can be seen that the hot zone identification method considers the continuity of the accident hot zone compared with the hot spot method, and continuously identifies the single accident-prone road section with spatial autocorrelation relationship, and completes the complete accident hot zone identification. Secondly, from the perspective of the traffic accident hotspot, the spatial statistical algorithm can realize the clustering of arbitrary length road segments, avoiding the influence of the fixed length unit division on the accident hot zone section, and the identification result can more objectively reflect the space distribution characteristics of the traffic accident hot zone.

6. Conclusion
This paper applies the ArcGIS software as a technical development tool and formulates a new method for accident hot zone identification based on the spatial autocorrelation algorithm. This method takes the attributes and spatial properties of an accident into account and realises the automatic identification of accident hot zones. The analysis of the accident hot zone shows that the method proposed in this paper can effectively identify the whereabouts of accidents regardless of the accident rate, which advances the existing hotspot-based identification method. Therefore, this paper recommends the use of the accident hot zone method in the identification of accident-prone road sections in the future.

However, there are still some issues worthy of further research in the future. First, this paper divides the road sections into several 100 meters long piles, which need further validation. Secondly, this study has not considered the hot zone difference which can be reflected by the number and consequence of accidents, and third, the hot zone identification method can be used to prioritise the accident hot zone according to the certain rules such as the length of the accident hot zone and the risk of the accident hot zone.

Acknowledgments
The authors appreciate Engineering Research Center of Highway Infrastructure Digitalization for providing software and CCCC First Highway Consultants Co., Ltd. for providing necessary information.

References
[1] NGUYEN H H. (2016) Multidimensional analysis of road safety case study of Amsterdam[D]. Master thesis. University Paris Diderot.
[2] Wang H, Li R. (2016) Application Research of Buffer Analysis Method in Accident Identification[J]. Journal of Highway Engineering, 41(1): 103–107.
[3] ZHOU Z et al. (2017) Identification of multiple occurrences of rural road traffic accidents based on matrix model[J]. Journal of Shandong Jiaotong University, 25(2): 33–39.
[4] Zhu L. (2006) GIS-based automatic detection of road traffic accidents and traffic accident prediction [D]. Urumqi: Xinjiang Agricultural University.
[5] Tension. (2008) Research on Black Spots of Road Traffic Accidents Based on GIS and Principal Component Analysis[J]. Geospatial Information. 6(5): 79-82.
[6] S Erdogan, I Yilmaz, T Baybura. (2008) Geographical information systems aided traffic accident analysis system case study [J]. Accident Analysis and Prevention, 40:174-181.

[7] I B Gundogdu. (2010) Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya[J]. Safety Science, 48: 763-769.

[8] U.S. Department of Transportation Federal highway Administration. (1999) Using GIS in the Analysis of Truck Crashes Research[R]. HSIS (Highway Safety Information System) SUMMARY REPORT.

[9] Yamada I, THILL J. (2010) Local indicators of network-constrained clusters in spatial patterns represented by a link attribute[J]. Annals of the Association of American Geographers ISSN, 100(2): 269–285.

[10] FLAHAUT B. (2004) Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation [J]. Accident Analysis and Prevention, 36(6): 1055–1066.