

RESEARCH

Open Access



Spatial analysis of traffic accidents based on WaveCluster and vehicle communication system data

Junhui Zhang^{1,2*} and Tuo Shi³

Abstract

The frequent occurrence of traffic accidents has always been an important problem troubling traffic safety management, so exploring the law and characteristics of case occurrence in a space area has profound significance for the prevention of traffic accidents. Starting from the space-time angle and based on the traffic accident data, this article firstly carries out the wavelet decomposition of the incident data of time series to realize the problem optimization of sparse matrix and then studies the spatial differentiation pattern of traffic accidents through the *k*-means clustering method. And under the formed differentiation pattern, the spatial and temporal laws of the incident are deeply analyzed. Finally, accident causes based on vehicle information system data are analyzed. The results show that the traffic accident space in Beijing is divided into 5 categories, among which, the hot spot space is the area with large traffic volume, diverse driver quality, or the junction of urban and rural roads, and the vehicle information system distracting the driver's attention is also the cause of accidents from a micro view through vehicle information system data.

Keywords: Wavelet analysis, Secure vehicle communication system, Spatial differentiation, *K*-means clustering

1 Introduction

In recent years, frequent traffic accidents have become an important factor threatening people's travel safety. Its high data bring unprecedented challenges to the public security organs, especially in the situation of constantly changing traffic space environment, and the case space of different types of accidents usually shows different rules and characteristics. Based on real traffic accident data and machine learning technology, analyzing traffic accident cases from the perspective of time and space can reveal the distribution law and the causes behind traffic accidents in a scientific and profound way, so as to formulate different prevention strategies according to different types of traffic accidents and make relevant departments respond to traffic accidents in a more directional and targeted way. In 2003, Kuanmin Chen and the others obtained the distribution characteristics of traffic

accidents in time and space through simple statistical methods, explored the causes, and proposed counter-measures [1] to improve road traffic safety. In 2009, T-K Anderson used kernel density estimation and *k*-means clustering method to extract road section with frequent traffic accidents [2]. In 2012, Wenhao Yu, Tinghua Ai, and others used the planar kernel density estimation method expands to the network space as the network kernel density estimation method to extract and visualize the clustering distribution area of event points under network constraints [3], providing a reference for the analysis of traffic accident spatial and temporal distribution pattern. In 2016, Jinyan Tan and others used space-time GIS technology to identify [4] urban road black spots by comprehensively considering the number, spatial location, severity, and other factors of traffic accidents. In 2016, K Z Htut and others directly used the nuclear density estimation method in ArcGIS spatial analysis tool to identify the clustered sections of expressway accidents and grade their severity [5]. The study used a variety of techniques to explore the characteristic of spatial distribution of traffic accidents, but the

* Correspondence: stshi8808@sina.com

¹Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

²Beijing Traffic Management Bureau, Beijing 100037, China

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

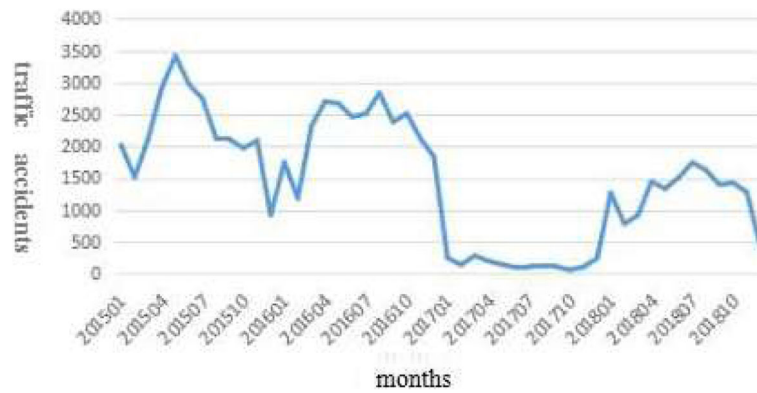


Fig. 1 Overall incidence of traffic accidents in Beijing in 2015–2018

analysis of the integration of time and space of traffic accidents is insufficient, the analysis scale is relatively macro, and most of them took a city as the research object to discuss the spatial distribution law of traffic accidents, and the application of intelligent machine learning technology is not deep enough. It also ignores the fact that multiple traffic accident spaces are similar in the time of occurrence, which leads to the weakening of the research on the spatial differentiation pattern of traffic accidents. In recent years, only a few studies have divided the traffic accident space. In 2015, Zhenhong Wang analyzed the time distribution, spatial distribution, and crowd distribution characteristics of traffic accidents [6]. In 2014, Lian Xie, Chaozhong Wu, Nengchao Lu, Yan Gao, and others proposed an improved DBSCAN clustering algorithm for the identification of road

sections with multiple traffic accidents [7] and studied the application of this algorithm in the identification of spatial distribution characteristics of traffic accidents by taking highway traffic accidents as an example. Clustering analysis is an important means of space-time analysis of traffic accidents. Based on the time series as the axis and the traffic accident law of the grassroots traffic police brigade as the research object, the integrated application of time series analysis technology and cluster technology, this study explore the space differentiation for the traffic accident, to explore the space of different types of traffic accident in time and space to carry out targeted analysis of different law of both accidents, put forward aiming at all kinds of traffic accident prevention strategies of the space, so as to greatly improve the working efficiency of the traffic police department.

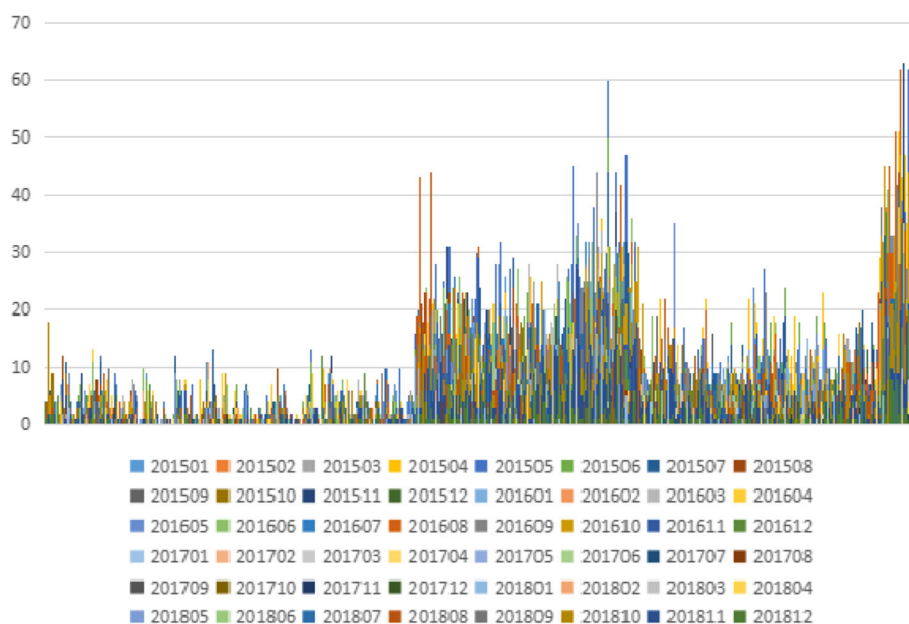


Fig. 2 Monthly cases of each traffic police brigade from 2015 to 2018

Table 1 Examples of traffic accident observation samples

Traffic police brigade	201,501	201,502	201,503
Xidan traffic police brigade	0	2	2
East traffic police brigade	0	0	1

In Section 2, the data sources are introduced and simply analyzed, and the analysis method based on WaveCluster is explained in Section 3. Spatial differentiation of traffic accidents and empirical analysis is proposed in Section 4. Finally, a conclusion is obtained in section 5.

2 Data sources and overall analysis

2.1 Data sources

The data sources (the data is protected) for this article is the number of traffic accident cases in each jurisdiction from 2014 to 2018 provided by the Beijing Traffic Management Bureau. The data includes the location of the crime, the territorial traffic police brigade, the time of the crime, the nature of the case, and other characteristics. Since this article focuses on the analysis and research of traffic accidents from the perspective of time and space, it mainly focuses on the law of the number of cases with time as the axis.

2.2 Overall analysis

Taking Beijing as the analysis object, grasping the law of traffic accident occurrence in various regions from a macro perspective is helpful for the decision-making level to understand the situation of traffic accident occurrence as a whole. According to the data, the trend of traffic accidents in Beijing is shown in Fig. 1. Figure 1 shows that the number of traffic accidents in Beijing reached the peak in April and May 2015, and with the linkage of various police services, the number of traffic accidents was reduced to the low point at the end of

Table 2 Parameter selection process

k	3	4	5	6	7	8
s	0.691	0.732	0.864	0.853	0.847	0.793

2015. In 2016, traffic accidents began to rise again. The traffic police department once again strongly rectified the traffic violations, carried out targeted wave warfare to reduce traffic accident cases, and established an effective accident prevention mechanism, which was effectively contained in 2017. The number of cases throughout the year reached the lowest level in the history of the same period, and it was relatively stable, showing no trend of sharp rise and fall. Although the number increased in 2018, the overall level was also relatively low compared with the previous 2 years. This shows that the occurrence of traffic accidents usually changes in stages due to the increase of fighting and prevention efforts, but it also reflects from the side that the response to traffic accidents at that time tends to be lagging and passive. Therefore, the inevitable choice to eradicate this stubborn social disease is to control the law of occurrence of the cases, explore the mechanism of the occurrence of cases, improve the ability to prevent cases, and make the gateway in front to curb the occurrence of cases.

Further, they master the situation of traffic accidents from the grassroots traffic police brigade and draw a time-indexed map of the situation as shown in Fig. 2. Figure 2 shows the monthly incidence of traffic police in each district.

3 The analysis method based on WaveCluster

3.1 Wavelet analysis of time series

Time domain is the most commonly used research in the study of traffic accidents. Time domain analysis has

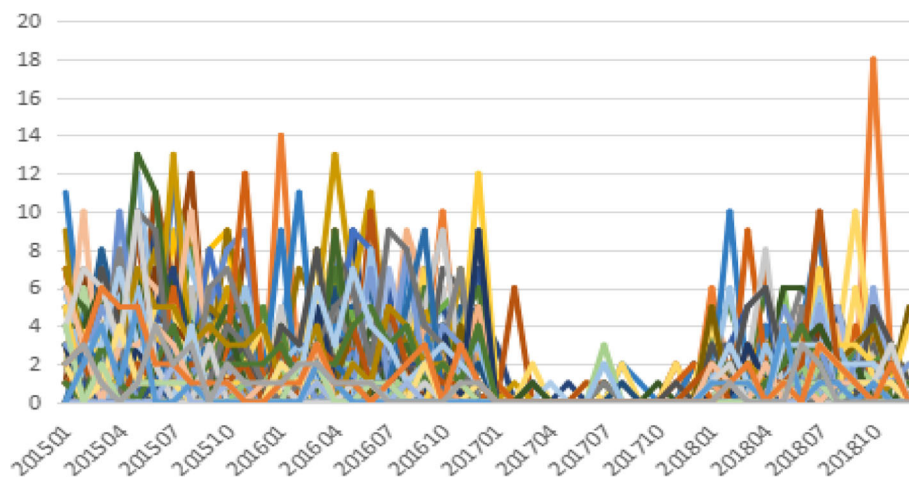
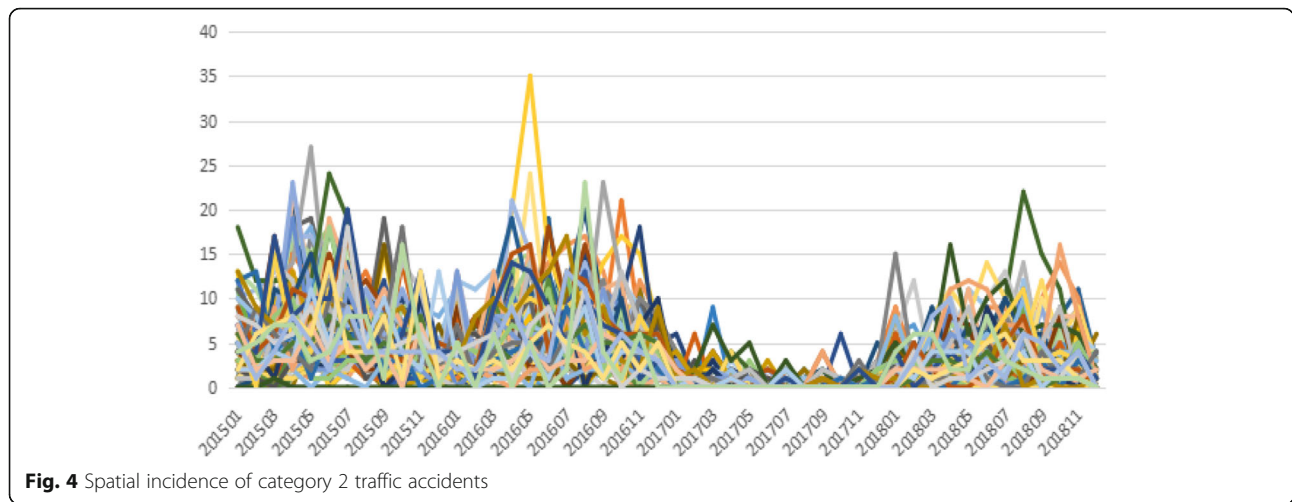


Fig. 3 Spatial incidence of category 1 traffic accidents



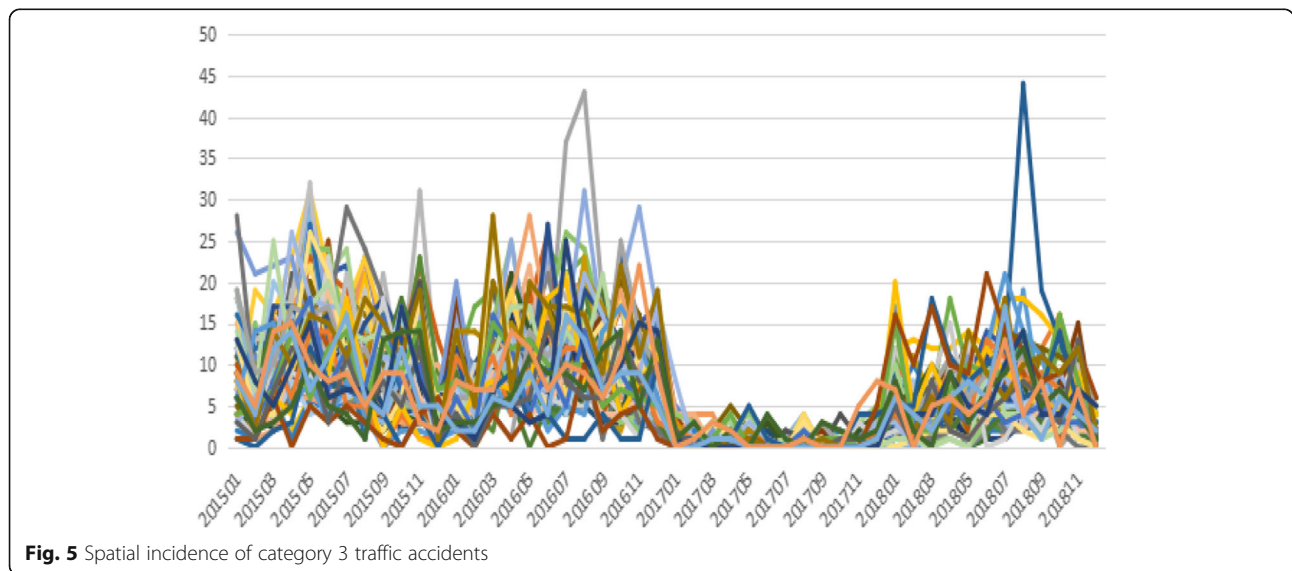
the advantages of strong time localization ability and strong interpretability, but it is unable to obtain more information about the change of time series, so some researches attempt to explore in frequency domain analysis. Although frequency domain has the function of accurate frequency positioning, it is only suitable for stationary time series analysis. However, with the change of time, traffic accidents are often subject to the comprehensive influence of a variety of factors, most of which are non-stationary sequences. Generally, they not only have the characteristics of trend and periodicity, but also have the problems of mutability, randomness, and “multi-time scale” structure, showing a multi-level evolution rule. For the study of such non-stationary time series, the corresponding time information of a certain frequency band or the frequency domain information of a certain period are usually required. Obviously, simple time domain analysis and frequency domain analysis are

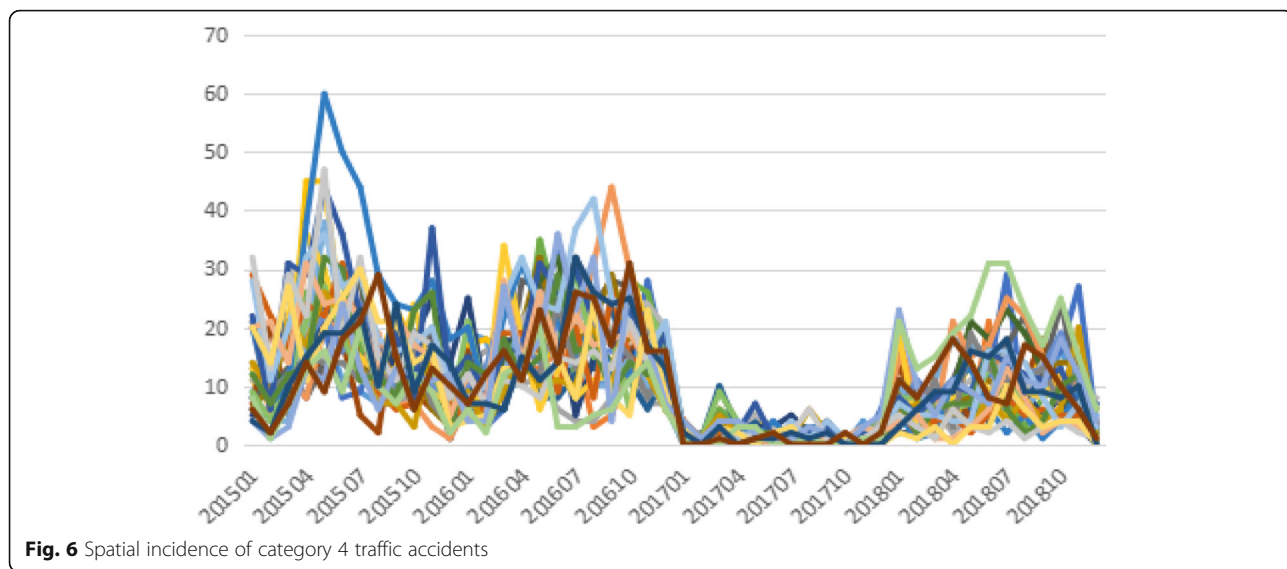
obviously weak. The proposed Morlet wavelet analysis method with time-frequency multi-resolution feature provides a way to better study time series problem, which can clearly reveal the hidden in the time series of a variety of change cycle and fully reflect the system change trend in different time scales and can make a qualitative estimate of the future development trend of the system.

In the space-time study of traffic accidents in this article, the time series of traffic accident cases observed are discrete, so the form of discrete wavelet transform is adopted:

$$W_f(a, b) = |a|^{-1/2} \Delta t \sum_{k=1}^N f(k\Delta t) \bar{\psi}\left(\frac{k\Delta t - b}{a}\right) \quad (1)$$

In this formula, $W_f(a, b)$ is the wavelet transform coefficient, $f(k\Delta t)$ is a discrete signal, a is the scaling scale, b is the translation parameter, $\psi(\frac{k\Delta t}{a})$ is the basis wavelet func-





tion, and $\overline{\psi}(\frac{kA_t-b}{a})$ is the complex conjugate function. In this study, the most important thing is to obtain the wavelet coefficients from the wavelet transform equation, use these coefficients to analyze the time-frequency variation characteristics of the time series, and conduct clustering analysis based on these coefficients to obtain the spatial differentiation pattern of traffic accidents.

3.2 K-means clustering of time series wavelet decomposition

(1) Introduction to the observation sample data set

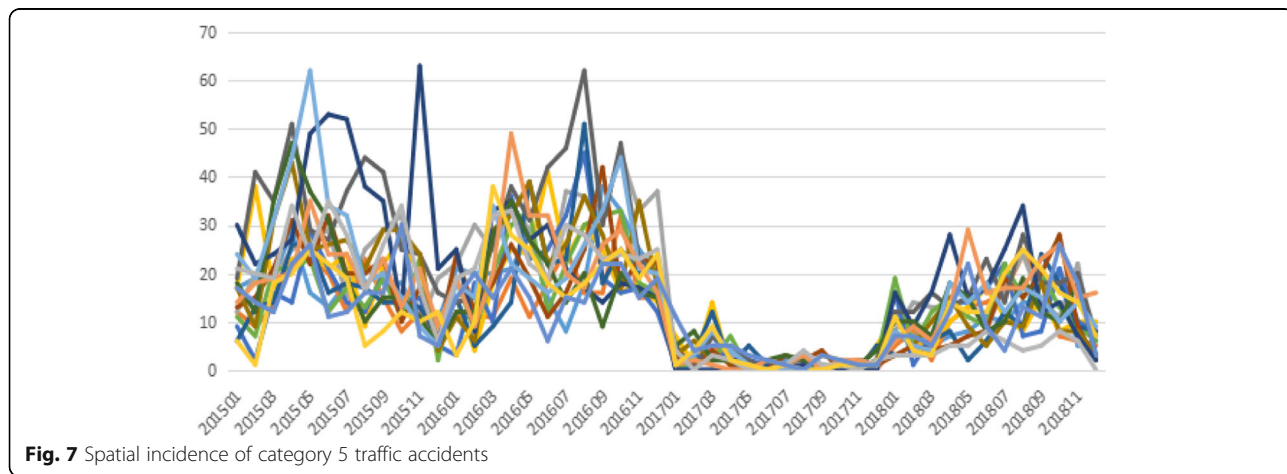
The traffic accident data set used in this paper takes the traffic police brigades in all districts of Beijing as the spatial object, the time of case occurrence as the data index (month as the basic research unit), and the number

of cases occurrence as the research object. The corresponding data set format is shown in Table 1.

According to the preliminary observation, the matrix composed of data sets in this paper has the problem of sparsity. If the clustering analysis is carried out solely based on the original data samples, the clustering effect will be poor and the spatial differentiation result of traffic accidents will be not ideal. Therefore, this paper introduces the wavelet decomposition to eliminate the problem of sample sparsity.

(2) Calculation of wavelet coefficients.

The PyWavelets module under python was selected to conduct wavelet transform on traffic accident sequences of traffic police brigade in each jurisdiction from January 2015 to December 2018. Db4 (Daubechies limit phase wavelet) is selected as the basic wave function. The idea is to



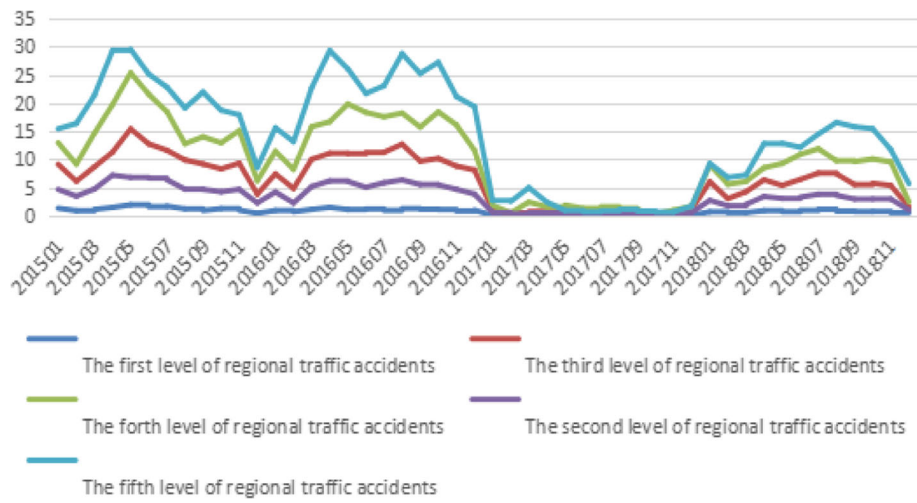


Fig. 8 Average spatial incidence level of various traffic accidents

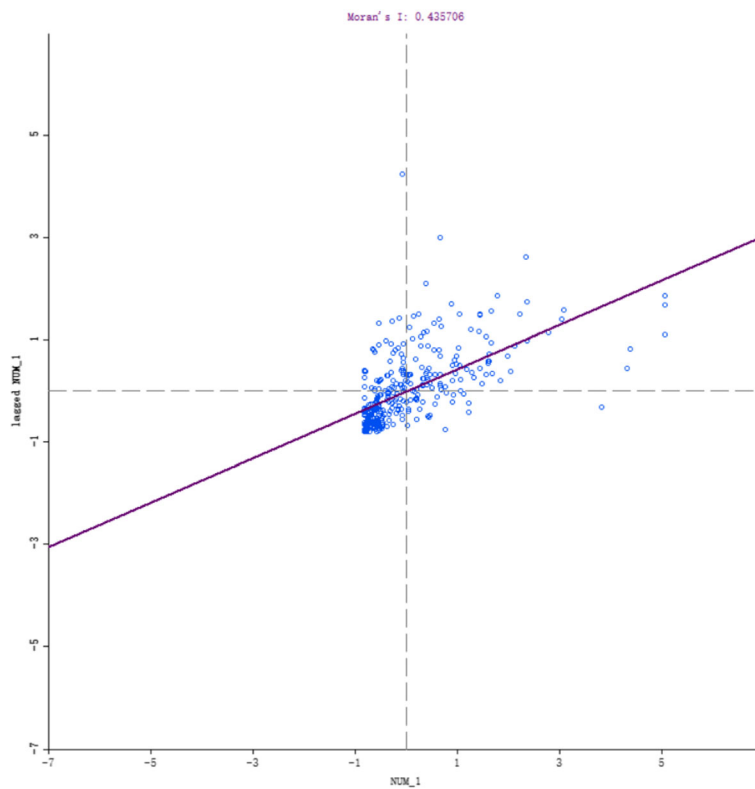


Fig. 9 Spatial correlation description of traffic accidents

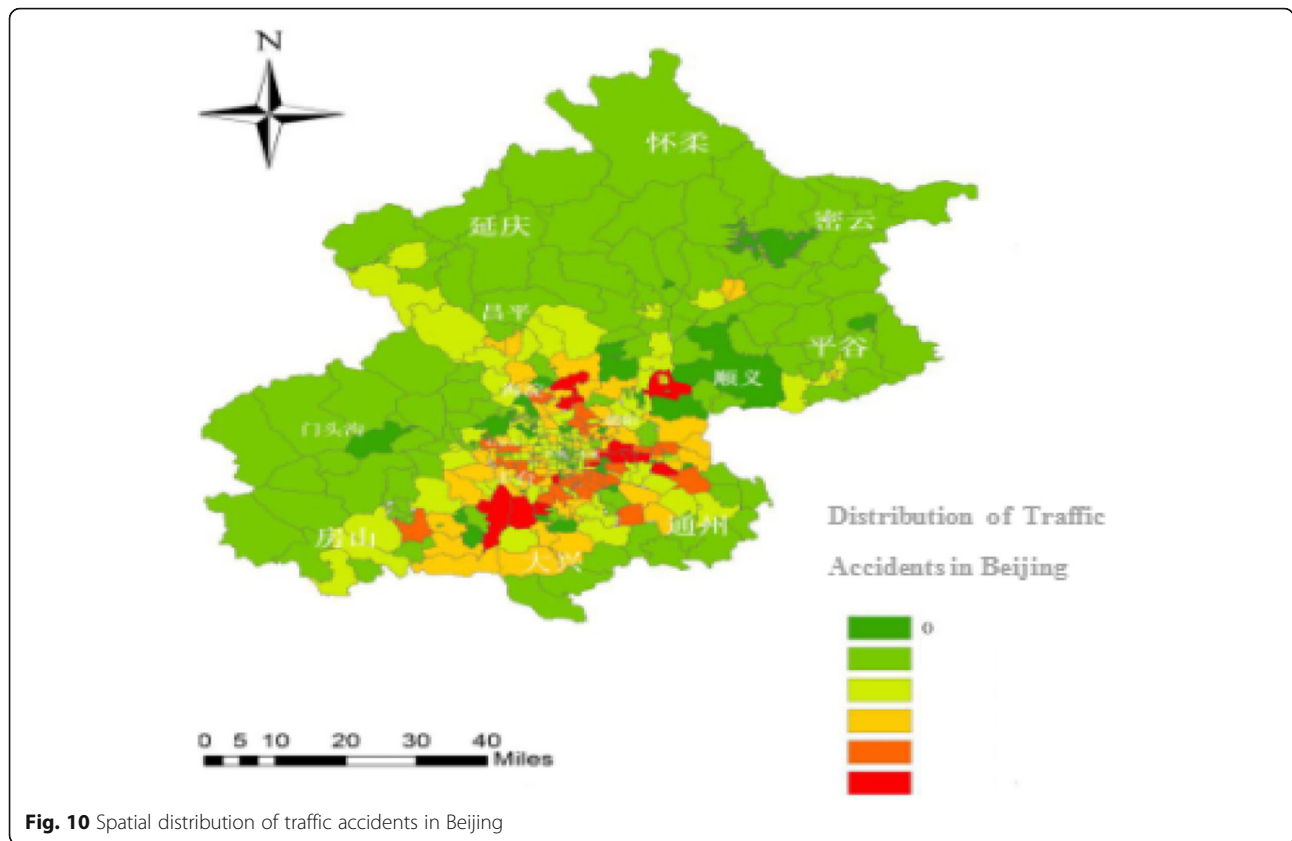


Fig. 10 Spatial distribution of traffic accidents in Beijing

decompose the data sequence into wavelet, and the result of each layer decomposition is that the low-frequency signal obtained from the previous decomposition is decomposed into low-frequency and high-frequency parts. In this way, after layer decomposition, the source signal is decomposed into:

$$X = D1 + D2 + \dots DN + AN \tag{2}$$

D1, D2... DN is the high-frequency signal obtained by the decomposition of the first layer, the second layer to the Nth layer decomposes to obtain high-frequency signals, AN is the low-frequency signal obtained by the Nth decomposition, and the wavelet coefficient is calculated accordingly.

Table 3 In-vehicle communication system types

Types	Vehicle communication system
Communication	Mobile phones, smart phones, car electronic communication stations, etc.
Navigation	GPS navigator, advanced traveler information system, etc.
Entertainment	DVD, CD, TV, etc.
Working	Car computer, PDA, etc.
Controlling	In-vehicle facility switch and adjustment control, etc.

(3) Clustering based on wavelet coefficients

In this paper, the above wavelet coefficients obtained by wavelet decomposition are utilized for *k*-means clustering. The reason for choosing *k*-means method is that the clustering effect is better. And the interpretability of the algorithm is relatively strong, and because the clustering belongs to unsupervised learning, the silhouette coefficients should be used for the determination of the number *k* of cluster. The silhouette coefficient is a way of evaluating the clustering results was first proposed by Peter J. Rousseeuw in 1986. The calculation method is as follows:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}} \tag{3}$$

$$s = \frac{1}{k} \sum_{i=1}^k s(i) \tag{4}$$

In formula (3), the smaller the average distance between sample *a(i)* and other samples in the same cluster is, the more it means that sample *i* should be clustered into this cluster; Calculation of the average distance *b_{ij}* between sample *i* and all samples of other cluster *C_j* is called the dissimilarity between sample *I* and cluster *C_j*. Defined as the sample *I*, there is no similarity

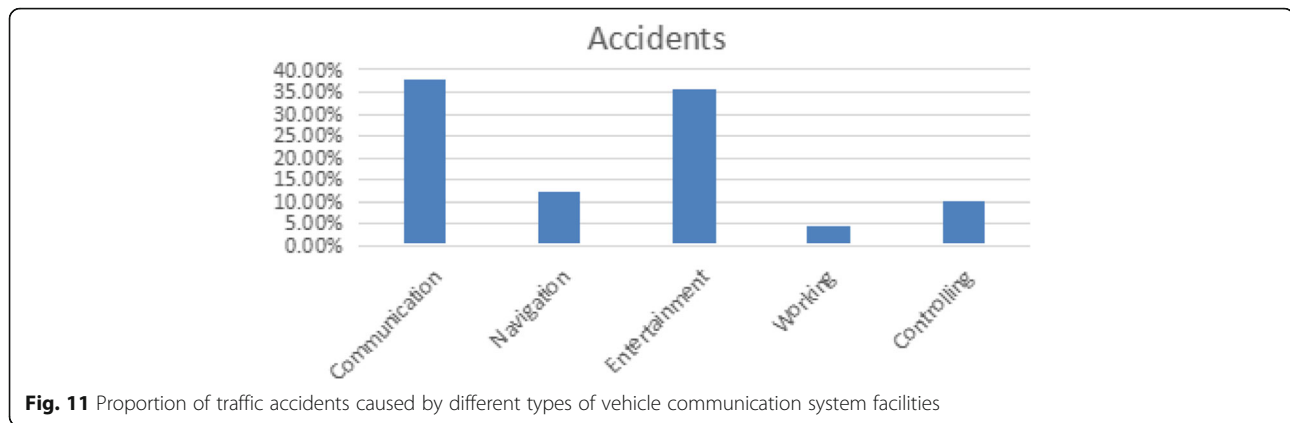


Fig. 11 Proportion of traffic accidents caused by different types of vehicle communication system facilities

between clusters, $b = \min \{b_{11}, b_{12}, \dots, b_{1k}\}$. The silhouette coefficient s is the mean value of $s(i)$ of all samples. The closer the s is to 1, the better the sample clustering effect is. Therefore, we can try to set different k values to calculate the corresponding s and select the k value closest to 1 to get a more ideal cluster number.

K -means clustering was conducted based on the wavelet coefficients of the traffic police brigade after decomposition, and finally, the conclusion of spatial differentiation pattern of traffic accidents was obtained. It means that traffic police departments with similar law of occurrence are gathered together, and the mechanism and the underlying cause of traffic accidents can be explored by analyzing the comprehensive factors of each type of traffic accident space.

4 Analysis of experimental results

4.1 Spatial differentiation of traffic accidents based on WaveCluster

The above WaveCluster method was used for the spatial differentiation experiment of traffic accidents. The spatial differentiation pattern of traffic accidents in Beijing traffic police brigade scale was determined based on the contour coefficient, which could be divided into five categories, that is, the contour coefficient was significantly higher than the others, which was 0.864, as shown in Table 2.

After analyzing the spatial differentiation results of traffic accidents of category 1 to category 5, the incidence of accidents is shown in Figs. 3, 4, 5, 6, and 7, and the average incidence level of all types of traffic accidents is shown in Fig. 8.

From Figs. 3, 4, 5, 6, 7, and 8, it can be observed that the number of traffic accidents in the space belonging to the traffic police brigade of category 1 to category 5 increases gradually. Therefore, category 1 traffic accident space is named low incidence space, category 2 traffic accident space is named medium and low incidence space, category 3 traffic accident space is named medium and high

incidence space, category 4 traffic accident space is named medium and high incidence space, and category 5 traffic accident space is named high incidence space.

In order to further explore the existence of spatial agglomeration of traffic accidents, this paper introduces the Moran index to analyze and record as shown in Fig. 9. It is not difficult to see from Fig. 9 that traffic accidents do have high spatial agglomeration, which also means that there is high spatial correlation.

4.2 Spatial distribution characteristics of traffic accidents

In order to further explore the spatial distribution characteristics of traffic accidents, the above clustering results are visually displayed on the map with ArcGIS, as shown in Fig. 10. The "0" in the figure indicates the absence of traffic accident data in the space, so no discussion will be made. In general, traffic accidents belong to the same kind of traffic accident space (traffic police brigade) are bordering, indicating that the traffic accident space with similar law of occurrence is similar in geographical location, too. This is because the social environment in the regions with similar geographical locations is relatively similar, and the nature and characteristics of the cases are quite similar, which means that the law of traffic accidents is clustered.

Figure 10 also reflects that the overall spatial distribution of traffic accidents in Beijing presents a "concentric circle attenuation law," that is, the center of city is not the traffic accident space with the highest incidence, but a quasi-circular buffer zone. On the contrary, in the second circle, Chaoyang, Haidian, and Fengtai are the concentrated areas with high incidence of traffic accidents. As the radius increases, the incidence area within the traffic accident space shows a declining trend, and the overall level of traffic accidents in the outer circle (the traffic police brigade in the outer suburbs) is relatively low. That is to say, the distribution law of cases is similar to a ring structure, with the lowest incidence in the innermost ring, the highest incidence in the middle ring,

and the lowest attenuation in the outermost ring. This conclusion is also in line with the “Concentric Circle Theory 1” of Chicago school.

4.3 Analysis of accident causes based on data of vehicle information system

Vehicular communication is defined as communication between the vehicles [8]. The communication system provide the driver with a convenient and pleasant driving experience, but also threatening driving safety at the same time [9]. In order to analyze the influence on driving safety with vehicle communication system, the data of vehicle communication system is divided into five categories based on the accident causes, as shown in Table 3.

The different types of vehicle communication facility can cause different numbers of accidents. Figure 11 shows the results. From Fig. 10, the following conclusions can be drawn: (1) Communication and entertainment causes the most traffic accidents, for example, when calling a car voice phone, it occupies hearing and cognitive language coding channels, sometimes it also takes up visual channel, but because the visual channel is significant to driving. Lots of accidents are taken place. (2) The number of gaze and the duration of gaze are the most direct evidence that the visual sub-task affects the main task of the driver. When the driver’s visual attention is concentrated on the sub-task, it is impossible to monitor the road environment well at the same time, especially for the need for visually monitoring traffic focus and thus causing traffic accidents.

5 Conclusion

According to the traffic accident incidence data of some traffic police brigade in Beijing in each month from 2015 to 2018, this paper firstly uses wavelet decomposition method to decompose the time series of traffic accident occurrence and further explores the spatial differentiation pattern of traffic accident occurrence based on the decomposed wavelet coefficient and k -means clustering. The results show that traffic accident spaces of the same category (traffic police brigade) are similar in the law of occurrence, which is due to the fact that the spaces close to the occurrence of traffic accidents are more similar in social and environmental conditions, and more adjacent in geographical space. The space of high-incidence cases usually has complex traffic demands, lack of traffic facilities, large difference in driver quality, complex traffic environment, and the security level needs to be improved. And the space of low-incidence cases mainly is distributed in suburb areas and core area. This is because the suburb areas of various kinds of roads in the outer suburbs and counties are relatively low, which obviously lead to traffic participants paying more attention to traffic safety, a small floating population, the pilot

quality is relatively stable, and the path to be familiar with the environment and economic level is not high, which make over speed behavior (easily happened traffic accident) returns relatively low. The reasons for the low incidence rate in some core areas are that the traffic security facilities in these spaces are relatively complete, the surveillance and camera equipment are ubiquitous, and the probability of seeing traffic police is relatively high, which has a strong deterrent effect on drivers and prevents the high incidence of traffic accidents. And through analyzing the in-vehicle communication system data microscopically, when the driver uses the in-vehicle information system during driving, the operations involving visual resource consumption mainly include information input and display observation, and the impact on driving safety includes the occupation of the driver’s visual attention resources and the decline in driving performance, and further threatening driving safety.

In order to effectively curb the occurrence of traffic accidents, we must adopt the strategy of combining prevention with prevention and prevention first. The following suggestions are put forward:

First, grasp the law of space crime, predict, forewarn the trend of crime, make the plan in advance, deploy the countermeasures, and reasonably allocate the police resources.

Second, strengthen the crackdown on high-incidence case, pay attention to and follow-up the trajectory of serious traffic violations that easily lead to traffic accidents, monitor at any time, detect, and control in advance.

Third, strengthen safety education. For drivers with serious traffic violations in the space, do a good job in traffic safety re-education. Especially in the urban and rural areas where the moving population is more and the drivers are more diverse, try to do the best to get about the situation. Traffic safety re-education personnel should update in time, the traffic police brigade of possession should call each other, drivers should grasp a comprehensive knowledge about dangerous driving.

Fourth, to improve the level of security, we need to sort out the weak space for physical and technical prevention, especially in the urban and rural areas prone to traffic accidents. Relevant departments should promptly promote the deployment of physical and technical defense forces, update the supporting resources of road infrastructure, and reduce the number of traffic accidents.

Fifth, in the design safety research content of the vehicle communication system, the workload of the human-computer interaction field, the input device, and its interface integration must be suitable for different groups of people [10]. Both usability researches are the focus of attention, and we need to blend management with industry associations. Designing criteria and measures will be developed to find effective solutions to the potential hazards of sub-tasks.

Authors' contributions

JZ is the main writer of this paper. He proposed the main idea, put forward the WaveCluster algorithm, and analyzed the result. TS gave some important suggestions for the related method. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China. ²Beijing Traffic Management Bureau, Beijing 100037, China. ³Beijing Police College, Beijing 102202, China.

Received: 20 February 2019 Accepted: 25 April 2019

Published online: 22 May 2019

References

1. K. Chen, Y. Wang, Distributing characteristics and preventive countermeasures of traffic accidents on urban roads. *J. Transp. Eng.* **3**(1), 84–87 (2003)
2. T.K. Anderson, Kernel density estimation and K-means clustering to profile road accident hotspots. *Accid Anal Prev* **41**(3), 359–364 (2009)
3. W. Yu, T. Ai, POI visualization and analysis on network space supported by kernel density estimation method. *Acta Geodaetica et Cartographica Sinica* **44**(1), 82–90 (2015)
4. J. Tan, X. Zhang, L. Shao, et al., The application of Spatio-temporal GIS in black spot prediction and identification of traffic accidents in Guangzhou. *Software Guide* **15**(12), 116–118 (2016)
5. K.Z. Htut, P. Piyatadsananon, V. Ratanavaraha, in *AtransSymposium: Young researchers forum 2016, transportation for A Better Life: Safe and Smart Cities*. Identifying the spatial clustering of road traffic accidents on Naypyitaw-Mandalay Expressway (2016)
6. Z. Wang, Distributing and preventive countermeasures of traffic accidents in Chinese urban roads. *Architect Eng Technol Design* **32**(35), 55–56 (2015)
7. X. Xie, C. Wu, N. Lu, Y. Gao, Research on identification method of road traffic accident-prone sections based on improved clustering algorithm. *J Wuhan Univ Technol (Edition of Transportation Science and Engineering)* **38**(4), 94–97 (2014)
8. S.K. Bhoi, P.M. Khilar, Vehicular communication: a survey. *Networks let* **3**(3), 204–217 (2013)
9. G. Calandriello, P. Papadimitratos, J.P. Hubaux, et al., On the performance of secure vehicular communication systems. *IEEE Trans. Dependable Secure Comput* **8**(6), 898–912 (2011)
10. O.A. Basir, *Vehicle immersive communication system: U.S. patent 9,930,158* (2018), pp. 3–27

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)
