ICIC Express Letters
Part B: Applications
Volume **10**, Number **8**, August **2019**

ICIC International ©2019 ISSN 2185-2766
pp. 725–736

URBAN TRAFFIC ACCIDENT-PRONE SECTION IDENTIFICATION & ANALYSIS SYSTEM BASED ON GIS SPACE CLUSTERING

Xingpei Yan[1,2], Lang Wei[1] and Wangfang Yuan[1]

[1]School of Automobile
Chang'an University
Middle-Section of Nan'er Huan Road, Xi'an 710064, P. R. China
yanxp1989@126.com; qch_1@chd.edu.cn; autoyuanzi@163.com

[2]Department of Transport Policy & Planning Research
Road Traffic Safety Research Center of the Ministry of Public Security
B#14, New World Center Office Building, Chongwenmenwai Avenue, Beijing 100062, P. R. China

Received January 2019; accepted April 2019

Abstract. *With the increasing quantity of motor vehicles in Chinese cities, the issue of urban traffic safety is highlighted increasingly, while urban traffic accident black-spots and accident-prone sections are seen as intractable problems. However, we still lack unified conclusions on accurate identification and locating of such accident-prone sections based on existing data at present. In this paper, Shenzhen traffic has been taken as an example: based on the three-year (2014-2016) traffic accident data acquired in Shenzhen, spatial ambiguity search has been firstly adopted to call the digital map API for determining accident spots, after which the cumulative frequency analysis has been performed for analyzing accident black-spots, and accurate locating of accident spots has been finally achieved by virtue of GIS linear reference technology; after system in package and secondary development combining GIS platform, we have set up a set of urban traffic accident-prone section identification & analysis system based on GIS space clustering, which enables the accurate locating and analysis of urban traffic accident-prone sections.*
Keywords: Traffic safety, Accident black-spots, Accident-prone section, Big data, GIS

1. **Introduction.** By the end of 2016, 64 cities in China have been occupied by more than 1 million motor vehicles, and 49 cities are holding more than 1 million automobiles. What accompanies with the rapid growth of urban motor vehicles' quantity is the increasingly highlighted urban traffic safety issue: Chinese urban traffic casualties have reached over 20,000 in 2016, accounting for 31.68% of all traffic casualties [1]. The normal operation of urban social functions is being heavily jeopardized by urban traffic accidents. More than challenges to the urban operation management, both highlighting and ensuring urban traffic safety play important roles in protecting the masses and guaranteeing the vulnerable group's rights to travel [2]. Thus, the way to accurately locate, identify and analyze urban traffic accident-prone sections has become one of the most urgent issues in current road traffic safety management.

Other country's research on traffic accidents has experienced a long period of development. In the 1940s, the Florida Traffic Management Department began to use the accident number method to identify road traffic accidents. This is the earliest proposed accident point identification method [3,4]. In the 1950s, Denmark, Italy, Thailand and other countries established an accident rate evaluation index system by using traffic accident survey data and traffic volume data [5]. Subsequently, the United States successively proposed the quality control method [6] and the traffic conflict technical law [7], and the

DOI: 10.24507/icicelb.10.08.725

Soviet Union proposed the safety factor method and other accident-prone road segment analysis methods [8]. In the 1990s, the United States began to apply GIS and other technologies to traffic accident analysis. Wisconsin established a GIS-based road traffic accident information system. The Iowa Transportation Bureau Transportation Safety Office built a GIS-based accident point and GIS-based Accident Location and Analysis System (GIS-ALAS) [9].

Although China started late in the research of traffic accidents identification, the combination of advanced technologies such as big data analysis, which has been booming in recent years, has given China a corresponding advantage. However, in the analysis of urban road traffic accidents in China at present, most of them still use statistical description methods to evaluate or predict road traffic safety using indicators such as the number of accidents and the number of deaths. For example, [10] used road traffic. The random nature of the accident was analyzed by multivariate statistical methods to analyze the impact of various risk factors on traffic accidents. The study found that in urban road traffic, the traffic safety problems of the main secondary roads, intersections, and non-signal control entrances and exits are more prominent. Based on 5,190 traffic accidents, [11] uses expert opinions and data fusion methods to establish a Bayesian network structure for urban road traffic accident analysis. Through the training of this network structure, urban road traffic accidents predictive analysis can be accurately performed. According to the statistical data of traffic accidents in Urumqi for the past five years, [12] analyzed the relationship between the severity of the accident and the influencing factors of the accident. In addition, [12] also uses Logistic regression model to analyze the severity of traffic accidents, and to estimate the parameters of the model and reliability analysis. In terms of the spatial factors of comprehensive traffic accidents, China has also carried out corresponding studies [15]. For example, [9] used spatial data management tools and the number of accidents to establish an accident identification system, and achieved multiple occurrences on automatic identification of accidents point. [13] proposed a method for identifying urban road accidents based on spatial autocorrelation. In detail, using the two-dimensional distribution of traffic accident points, and urban road network topology model, this method combined GIS with spatial data mining, and constructed spatial statistical units. Then, using nuclear density space, and cluster analysis, this method generated an accidental multiple point distribution map. However, most of these studies are based on foreign traffic GIS analysis models, and China's road traffic management and accident information processing is significantly different from foreign countries. The location of accident points is very different from that of foreign countries, so it is difficult to use these foreign models directly to accident analysis in China. The results are directly applied to the field of road traffic safety in China. It is particularly important to develop a road segment accident identification and analysis system suitable for China's national conditions.

This paper studies the identification analysis of urban road traffic accidents in Shenzhen, and describes the research process from the acquisition of road traffic accident data to the final formation of a spatial clustering-based urban road traffic accident multi-segment identification analysis system, which is suitable for China's national conditions. The main contributions of this paper are as follows. Three different methods were used to identify and analyze the road traffic accidents in large cities in China, and the advantages of the three methods were combined. Finally, a traffic accident statistics, road element query and accident point elements were formed. Query, attribute query, generate accident point thermal map and other functions of GIS road traffic accident multi-road segment identification analysis system, the system can better match the current traffic accident field form of China traffic police. The work done in this article will be described in detail below.

2. **Data Processing and Analysis of Urban Road Traffic Accidents.** By participating in the Shenzhen Traffic Accident Data Opening Analysis Contest jointly held by the Road Traffic Safety Research Information Sharing Platform of the Road Traffic Safety Key Laboratory (subordinated to the Ministry of Public Security) and the Traffic Police Station of Shenzhen Public Security Bureau, we have acquired all traffic accident data in Shenzhen from 2014 to 2016 (over 230,000 pieces in total) as generally shown in Table 1; multiple indexes have been contained in the field types, including the accident number, accident occurrence time, road number, road name, mileage, casualty, accident cause, etc. Based on such traffic accident data acquired, the analysis on urban traffic accident-prone sections has been carried out in this paper.

Table 1. Overall situation of Shenzhen traffic accident data

Year	Quantity of accidents	Quantity of fields
2014	76300	68
2015	77028	69
2016	83927	69
In total	237255	68

In order to facilitate data analysis, the data needs to be preprocessed first. The specific process is as follows.

1) Clean the null values of each field.

Since each field contains a large number of null values, and the null value will affect the calculation and analysis of the computer program, it is necessary to write a program to delete or replace all the null values in the original data, so as to use the program to carry out large data volume calculation. At the same time, the data missing ratio of each field needs to be calculated, and the data applicability of each field is comprehensively evaluated.

2) Make a dictionary table.

Since the data provided in this competition does not contain a dictionary table, it is necessary to query the database specification, the accident scene information collection table, and the relevant information of the code filling instructions, and collate the dictionary table of the relevant fields to support the subsequent analysis. At the same time, for some cases, only the Chinese characters are not queried to the corresponding standard code, and the self-made code is used in the dictionary table to support the subsequent computer to carry out the computing application.

3) Data encoding processing.

In the original data, there are a large number of mixed Chinese characters and codes, which is not only difficult to directly carry out statistical analysis, but also not conducive to the use of computers for large-volume program calculations. Therefore, it is necessary to write a data encoding program, and use the created dictionary table to replace all the Chinese characters involved in the original data with the encoding for subsequent analysis.

3. **Discrimination Analysis Method for Urban Road Traffic Accidents Based on Spatial Clustering.** To accurately identify accident-prone sections, it is necessary to firstly accurately determine geographical locations of traffic accidents, which is generally performed during accident handling for traffic policemen in the developed European/American countries. Upon their arrival at the accident site, polices can directly acquire the accident spot's GPS longitude & latitude coordinates by contrast with GIS map based on the on-board computer, with no necessary recording of written messages, and it is also feasible to achieve modification or confirmation on the GIS map simultaneously [14]. Regardless of the gradual promotion of China's Police Geographic Information

System (P-GIS) [16], acquisition of data on traffic accident locations by traffic policemen still resort to the traditional forms of "road name + milepost No." and "accident spot information description" for locating accidents at present, which is rather outdated in the contemporary atmosphere of a digital society. An example is the traffic in Shenzhen, where diverse road types are covered, such as the expressways, ordinary highways, urban roads, etc., Given the integrity of information on national-level expressways, it is relatively accurate to adopt "road name + milepost No." field in locating accidents occurring on such expressways; however, things are different for ordinary highways and urban roads where few entities like milepost No. are available, and it even shows a world of difference in descriptions of accident spots due to the absence of a fixed format for such descriptions. Considering such a complexity of accident data, the accurate locating of accidents shall take priority. In this regard, the following approaches have been adopted in locating and identifying accident-prone sections in this paper.

3.1. **Accident coordinates locating analysis based on the digital map API.** In current traffic accident information/data acquisition domestically, traffic policemen tend to determine urban traffic accident locations via the road mileage and conspicuous characteristic locations/features, instead of longitude & latitude information, who also generally reference adjacent feature points and supplement with directions/deviations of relative reference points for such traffic accident locations. Despite its facilitation for traffic policemen in their recordings, this approach makes it extremely difficult to reacquire the accurate locating information of such accident later.

Therefore, the form of big data has been introduced in this paper to achieve search and locating via making use of existing information on accident sections and combining relative digital map API (Application Programming Interface). API refers to some predefined function interfaces aiming to provide applications and developers with the ability to access to a set of routines based on certain software/hardware, without necessary access to the source code or comprehension of the internal working mechanism in detail.

Via POI (Point of Interest) retrieval interface of the digital map API, we can directly call the digital map's POI search function to search the "accident spot" field among Shenzhen traffic accident data, and determine the spot information (in such map) matching with the "accident spot" field information to the highest degree, thus acquiring coordinates corresponding to such accident spot information.

POI in this paper is mainly applied in solving the locating of urban traffic accidents. Multi-field correlated API search is adopted, covering descriptive words of the accident spot, descriptions of the road vs. certain road crossing, and fields of administrative divisions (e.g., sub-district office of the jurisdiction to which it belongs or town), thus avoiding any remarkable deviations; in addition, the uniqueness of the accident No. also guarantees the uniqueness of each accident's POI, which means that analysis and identification are achievable even in case of multiple accidents at the same spot.

Once acquired, the information on an accident spot's coordinates can be correspondingly converted into Mars coordinates, or in another word, the accident spot data can be accurately imported into the digital map. To better analyze the characteristics of major accident-prone sections, non-simple accidents are selected in this paper for locating accident coordinates and also imported into the digital map, with relevant results shown in Figure 1.

By virtue of accident coordinates locating based on the digital map API, we can intuitively see those coordinate locations of non-simple accidents occurring in Shenzhen during 2014-2016, and can also inquire/analyze each accident's location attribute to confirm accident-prone sections based on accident coordinates in the spatial analysis mode. This mode can meet demands for analyzing general accident-prone sections, acting as an

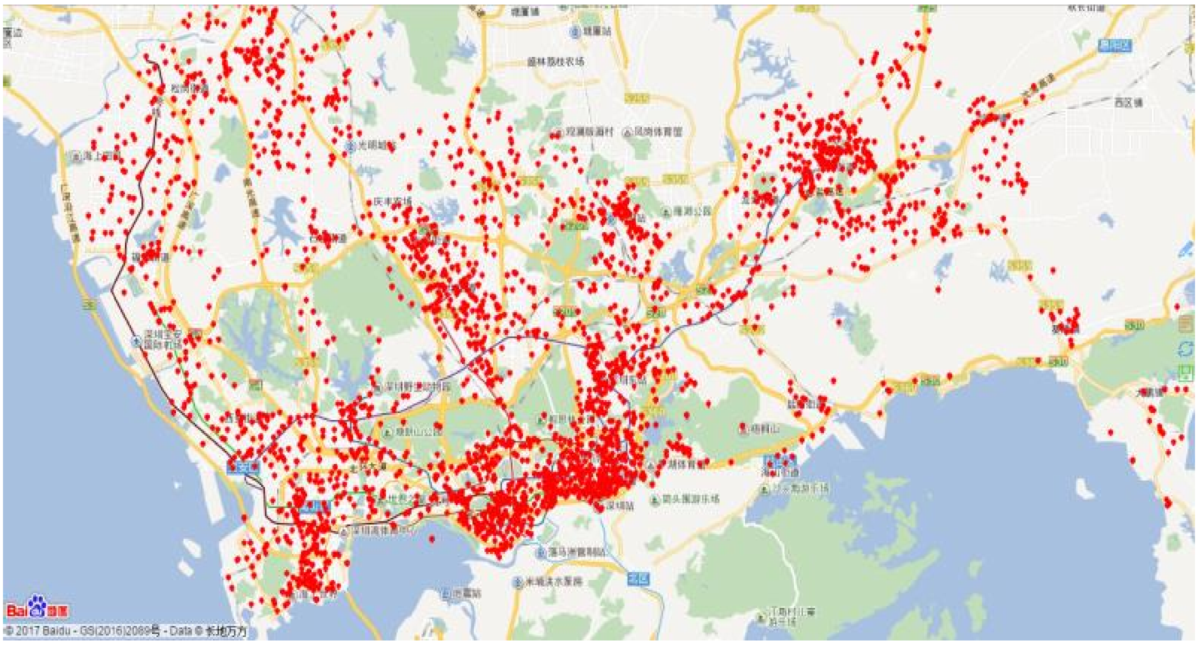

Figure 1. Shenzhen non-simple accident spot map during 2014-2016

accident-prone section identification approach (combining GIS) widely applied in China at present.

Nonetheless, it is also defective intrinsically. First of all, it acquires merely information on the digital map layer and POI spot map layer, and still lacks information on those linear elements of a road where a traffic accident generally occurs. In another word, although we can see that the accident spots lie in multiple corresponding arterials, some of such spots determined fail to be completely and accurately located within such arterials. Secondly, since the search engine relies on descriptive fields of an accident spot to search accident location, given the complexity of textual descriptions, it is still possible to result in a deviation of location identification, regardless of the selection of a spot matching with the "accident spot" field information to the highest degree as an accident spot.

3.2. **Road accident black-spot analysis based on the cumulative frequency method.** To better identify and analyze the linear layer (i.e., accident black-spots and accident-prone sections of a linear road), those arterials (i.e., expressways and some ordinary highways) with available road mileage data have been subjected to the normalized analysis on accident black-spot data in this paper. Firstly, 10 expressways, 2 national routes and 2 high-speed roads in Shenzhen (all suffering from accidents frequently) have been selected for accident black-spot identification and analyzed based on the cumulative frequency method. The so-called cumulative frequency method refers to that: the quantities (or rates) of accidents are sorted according to frequencies to result in some locations with the smaller cumulative frequencies and larger quantities of accidents, and such locations are taken as the possible locations of accident-prone sections. Examples are Boluo-Shenzhen Expressway and Huizhou-Shenzhen Coastal Expressway as shown below.

1) Boluo-Shenzhen Expressway

Accident frequency and cumulative frequency of corresponding section No. are shown in Table 2.

Cumulative frequency curve fitting and turning point location are shown in Figure 2.

TABLE 2. Accident frequency and cumulative frequency of corresponding section No. on Boluo-Shenzhen Expressway

Accident frequency	Section No.	Normalized accident frequency	Cumulative frequency of section No.
0	46	0	0.71875
1	6	0.1	0.8125
2	2	0.2	0.84375
3	4	0.3	0.90625
4	3	0.4	0.953125
5	0	0.5	0.953125
6	2	0.6	0.984375
7	0	0.7	0.984375
8	0	0.8	0.984375
9	0	0.9	0.984375
10	1	1	1

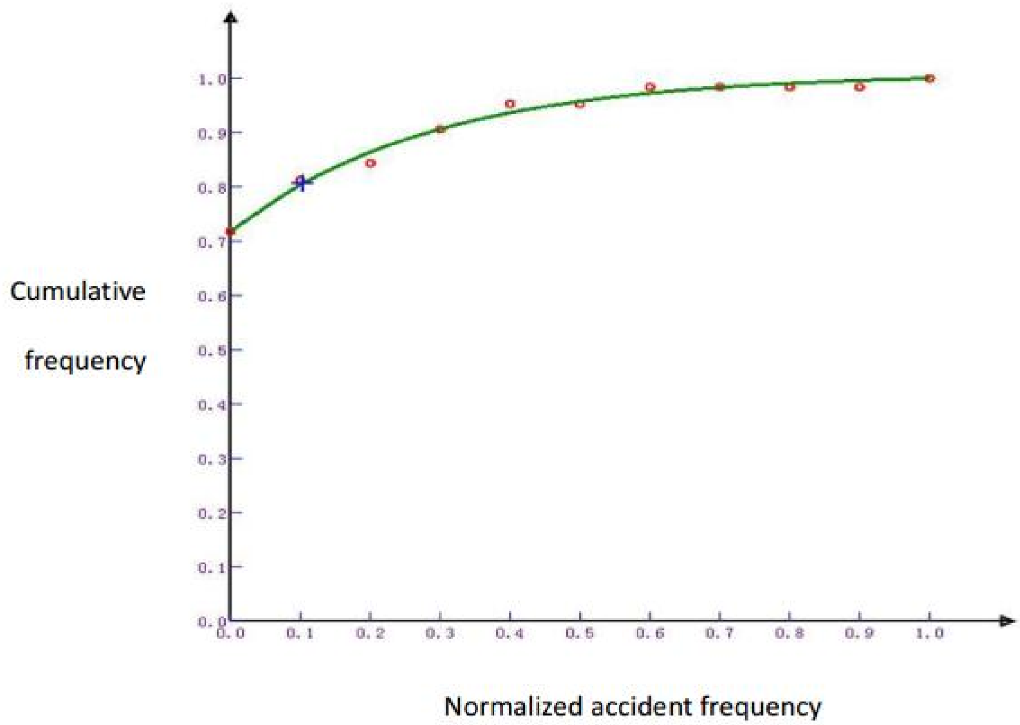

FIGURE 2. Map of cumulative frequency curve fitting and turning point location on Boluo-Shenzhen Expressway

The fitting formula for Boluo-Shenzhen Expressway obtained on the basis of its cumulative frequency curve is:

$$y = 1 - a \times \exp(-b \times x) + a \times \exp(-b)$$
$$a = 0.2905, \quad b = 3.4873$$
$$\text{SSE} = 0.001$$
$$\text{R-Square} = 0.9867$$

The turning point occurs where the normalized accident frequency is 0.103143349449696. Accident black-spot locations and accident frequencies are shown in Table 3.

Table 3. Accident black-spot locations and accident frequencies on Boluo-Shenzhen Expressway

Accident frequency	Section milepost No.
4	50
4	54
4	63
6	51
6	53
10	61

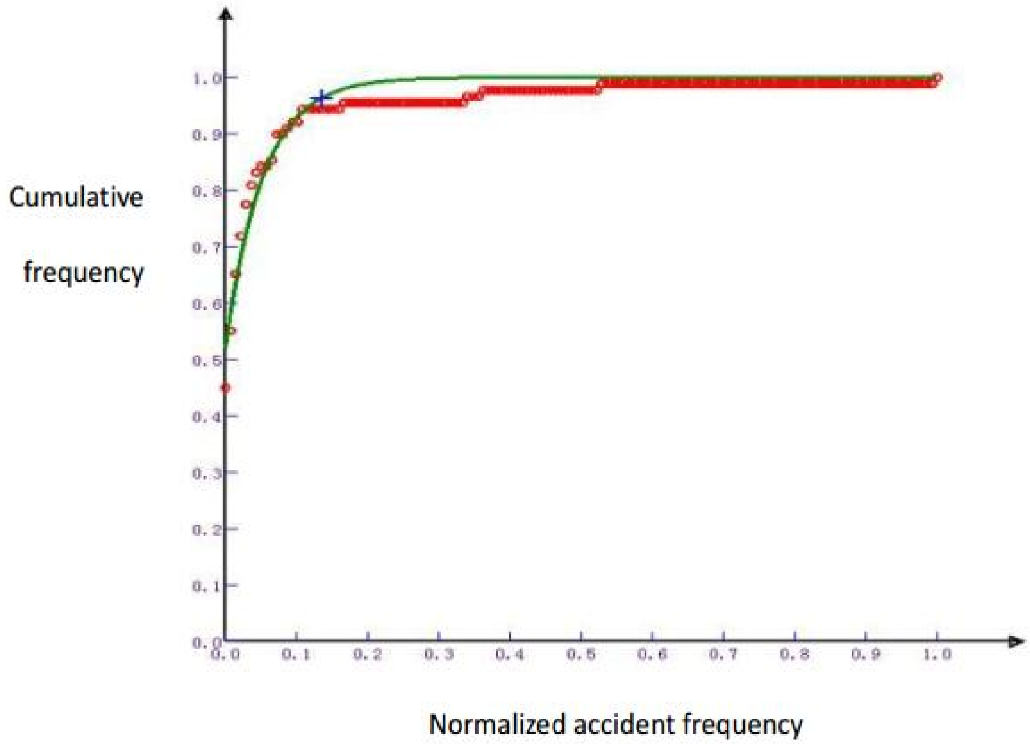

Figure 3. Map of cumulative frequency curve fitting and turning point location on Huizhou-Shenzhen Coastal Expressway

2) Huizhou-Shenzhen Coastal Expressway

Cumulative frequency curve fitting and turning point location are shown in Figure 3.

The fitting formula for Huizhou-Shenzhen Coastal Expressway obtained on the basis of its cumulative frequency curve is:

$$y = 1 - a \times \exp(-b \times x) + a \times \exp(-b)$$

$$a = 0.4794, \quad b = 19.0982$$

$$\text{SSE} = 0.0786$$

$$\text{R-Square} = 0.8979$$

The turning point occurs where the normalized accident frequency is 0.134099611316277, or in another word, where the accident frequency that is not less than 18.50575 can be regarded as an accident black-spot. Accident black-spot locations and accident frequencies are shown in Table 4.

From the above analyses on accident black-spots of Boluo-Shenzhen Expressway and Huizhou-Shenzhen Coastal Expressway, we can see an obvious advantage of identification & locating of road accident black-spots via cumulative frequency analysis: it is feasible

Table 4. Accident black-spot locations and accident frequencies on Huizhou-Shenzhen Coastal Expressway

Accident frequency	Section milepost No.
23	76
47	77
50	60
73	0
138	74

to directly obtain accurate locations of accident black-spots (based on the linear roads) via data, without resorting to any map platform, which is strongly practical and easily computed. Nevertheless, this approach is not flawless intrinsically. First of all, the cumulative frequency analysis results in textual data which are less powerful in visual display; secondly, this approach is highly demanding on data quality, since it is crucial to guarantee that every milepost No. is under the same starting point before analyzing the field of an accident spot's milepost No., otherwise a noticeable deviation may occur. In this paper, the same goes for Guangzhou-Shenzhen Riverside Expressway, National Route 107 and other roads in Shenzhen. Milepost No. ranges from 1 to 2030 on the Guangzhou-Shenzhen Riverside Expressway, and the accident spot's mileage data vary from tens to even one thousand or two thousands, which shall be attributed to respective adoptions of national mileage standard and provincial mileage standard when entering the accident data (analysis results deviate from each other obviously due to the inconsistency between both standards on the milepost No.). Besides, for Longhua-Dalingshan Expressway and Fulong High-speed Road, curve fitting fails due to the issue of data quality, leading to false calculating results.

3.3. **Accident-prone road segment analysis method based on GIS linear reference technology.** Based on the above two accident analysis approaches, to accurately determine accident spots and solve the problem of "spots and maps exist except no roads" (due to deviation of API-based characteristic extraction spot coordinates from roads) as well as "spots and roads exist except no maps" (in analyses via the cumulative frequency method), we decide to, based on linear reference technology and taking ArcGIS as a GIS platform, extract linear elements of Shenzhen roads in this paper, thus realizing the vectorization of such roads; afterwards, the identification and analysis of accident-prone sections can be performed on the basis of roads' linear elements to achieve "spots, roads and maps co-exist".

To realize the vectorization of Shenzhen roads, we need to extract such roads' linear data firstly. According to the 2014-2016 traffic accident data in Shenzhen and national road traffic grid data, linear vectorization extraction has been achieved for all roads within Shenzhen in this paper, resulting in the integration of scattered data on roads, with relative attribute information added. Specific steps are shown as follows:

The selected road has been identified, separated from a complex road network and vectorized, with its discrete vector data merged, after which a complete road has been plotted finally. After plotting, it is also necessary to add attributes to such road, thus determining its attribute integrity and identifying/discriminating attributes.

4. **GIS Accident Multi-Segment Identification Analysis System Based on Multiple Identification Methods.** Once the road linear elements (with attributes done with adding) are obtained, this paper has merged the accident spot, accident black-spot information and digital map's road linear elements (acquired via the above three identification approaches) on the GIS platform, based on which the independent packaging and

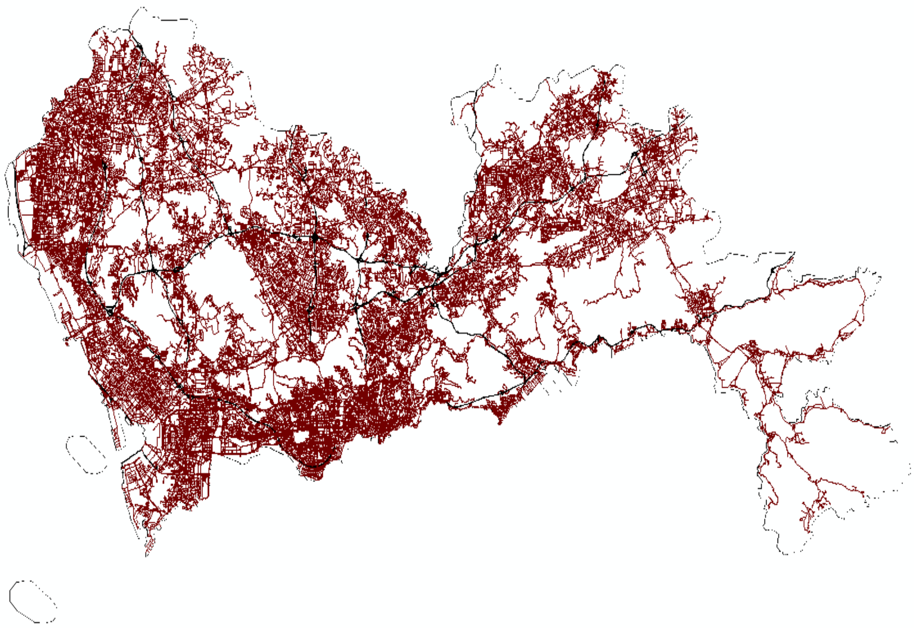

Figure 4. Scattered linear vector elements of roads within Shenzhen

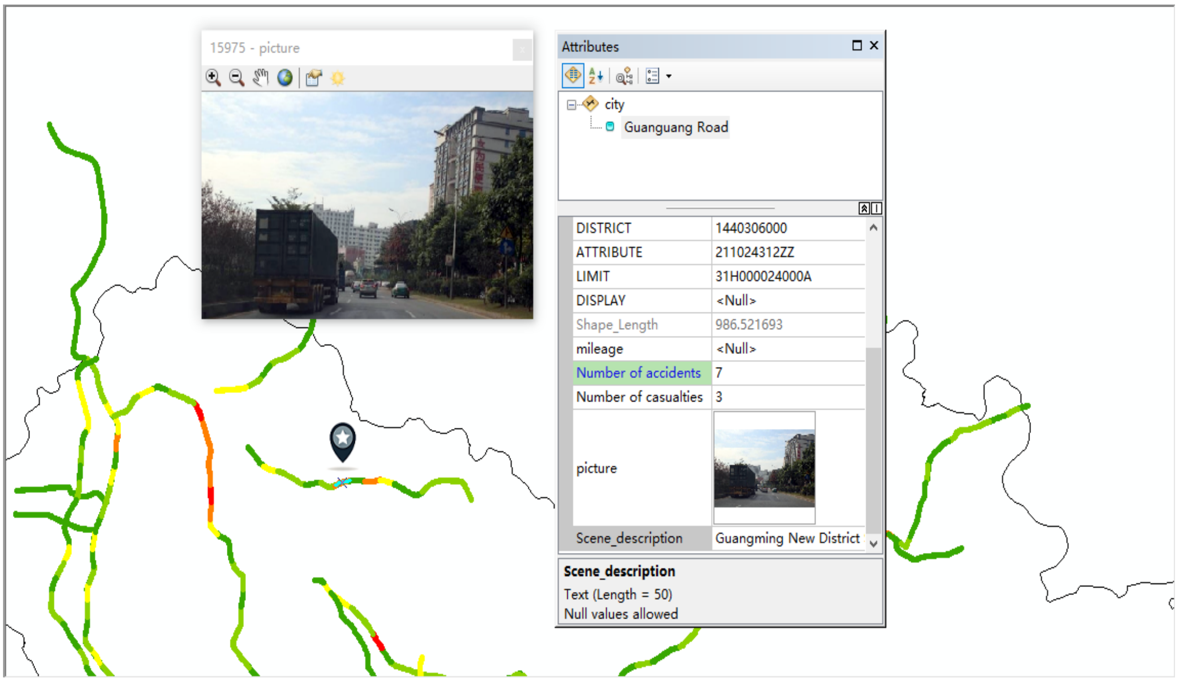

Figure 5. Addition of road element attributes

secondary development have been carried out to form an accident-prone section identification & analysis system enabling a series of functions (e.g., traffic accident statistics, inquiry of road elements, inquiry of accident spot elements, inquiry of attributes, and thermodynamic diagram analysis). The system algorithm's framework flow and the system function introduction are shown in Figure 6.

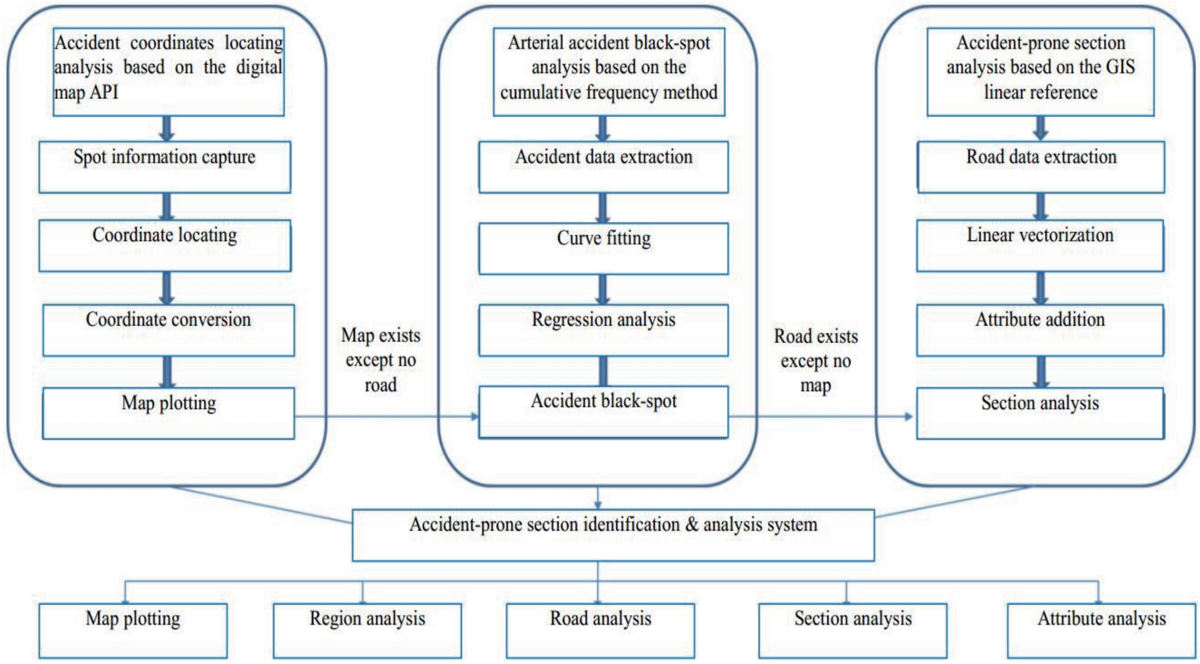

FIGURE 6. System algorithm framework flow and function display

FIGURE 7. (color online) Road sequencing distribution based on accident counting

1) Analysis of the number of road accidents.

After inputting the traffic accident data of Shenzhen city from 2014 to 2016 into the GIS accident multi-segment identification analysis system, the top 30 roads with the number of accidents are extracted and displayed on the GIS platform. The data analysis results are shown in Figure 7.

2) Analysis of road casualties.

In the GIS accident multi-discrimination analysis system, the top 30 roads in Shenzhen city from 2014 to 2016 are refined and analyzed. The analysis system is used to mark the number of casualties per kilometer in the map and the occurrence of accidents.

Figure 8. (color online) Injured and fatal accident road sections distribution map in Shenzhen city

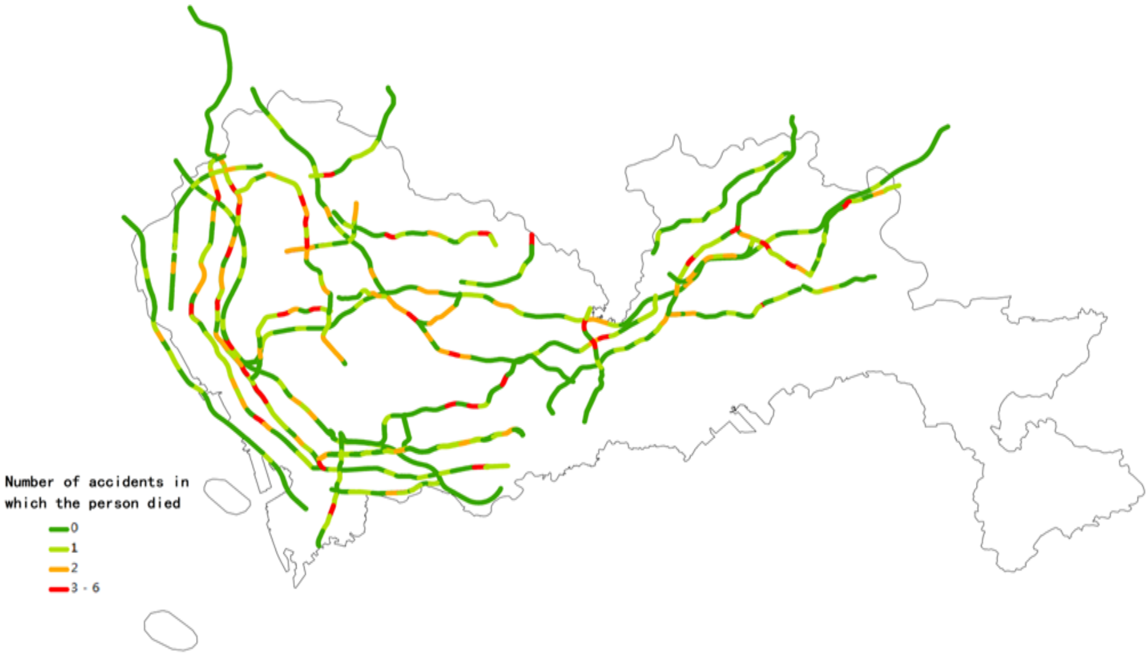

Figure 9. (color online) Fatal accident road sections distribution map in Shenzhen city

3) Analysis of road fatal accidents.

In the GIS accident multi-discrimination analysis system, the top 30 roads in the number of fatal accidents in Shenzhen city from 2014 to 2016 were collected. In view of the fatal accident, this paper also accurately points the accident location to the mileage section per kilometer, in order to better analyze the accident occurrence point. The distribution of the death accidents in Shenzhen based on the linear factors of the road is shown in Figure 9.

5. **Conclusions.** With an opportunity of participating in the Shenzhen Traffic Accident Data Opening Analysis Contest, Shenzhen road traffic accident-prone sections have been identified and analyzed in this paper on the basis of all traffic accident data & information acquired within Shenzhen from 2014 to 2016. Three approaches have been respectively adopted in the identification and analysis of accident-prone sections, namely the accident coordinates locating analysis (based on the digital map API), road accident black-spot analysis (based on the cumulative frequency method) and accident-prone section analysis (based on the GIS linear reference). Via comprehensive comparison and analysis of advantages & disadvantages of these three approaches, the ArcGIS platform has been finally applied in the merged packaging and secondary development of the said approaches, thus generating a GIS road traffic accident-prone section identification & analysis system enabling a series of functions (e.g., traffic accident statistics, inquiry of road elements, inquiry of accident spot elements, inquiry of attributes, and generation of thermodynamic diagrams for accident spots). The system can well match the traffic accident data collected by the Chinese traffic police, so it can better adapt to the Chinese environment. In the future, the system can be transplanted to other big cities in China. Just input the road network data of other big cities into the system, and you can get better analysis results.

REFERENCES

[1] S. Dai, *Research Report of Metropolis' Road Traffic Development in China (IV)*, China Architecture & Building Press, Beijing, 2018.

[2] S. Dai, Urban traffic safety problems and countermeasures in China, *China Transportation Review*, vol.37, no.7, pp.9-21, 2015.

[3] A. Fernandes and J. Neves, An approach to accidents modeling based on compounds road environments, *Accident Analysis and Prevention*, vol.53, pp.39-45, 2013.

[4] R. A. Krammes, Interactive highway safety design model: Design consistency module, *Public Roads*, vol.77, no.5, pp.12-17, 1997.

[5] Z. A. Sawalha, *Traffic Accident Modeling: Statistical Issues and Safety Applications*, Ph.D. Thesis, University of British Columbia, 2002.

[6] J. Kononov, *Road Accident Prediction Modeling and Diagnostics of Accident Causality: A Comprehensive Methodoligy*, Ph.D. Thesis, University of Colorado at Denver, 2002.

[7] P. Hidas, Modelling vehicle interactions in microscopic simulation of merging and weaving, *Transportation Research Part C: Emerging Technologies*, vol.13, no.1, pp.37-62, 2005.

[8] G. Yannis, J. Gollas and G. Kanellaidis, A comparative analysis of the potential of international road accident data piles, *ATSS Research*, vol.124, no.2, pp.111-120, 2000.

[9] N. Li, *Identification System for Urban Traffic Accident Hazardous Locations Based on ArcGIS*, Master Thesis, Southwest Jiaotong University, 2009.

[10] M. Ma, *Research of Urban Road Crash Analysis Based on Multivariate Statistical Techniques*, Ph. D. Thesis, Wuhan University of Technology, 2010.

[11] J. Zhao, W. Deng and J. Wang, Bayesian network-based urban road traffic accidents analysis, *Journal of Southeast University (Natural Science Edition)*, vol.41, no.6, pp.1300-1306, 2011.

[12] Z. Ma, Y. Chen and L. Zhang, Influence factors of accident severity for urban road, *Journal of Chongqing Jiaotong University (Natural Science)*, vol.33, no.1, pp.111-114, 2014.

[13] G. Ren, *Black Spot Analysis and Decision Support Research of Urban Road Based on GIS*, Master Thesis, Shandong Jianzhu University, 2016.

[14] H. Jiang, A spatial autocorrelation-based approach for identifying hot spots on urban road, *Journal of Tongji University (Natural Science)*, vol.41, no.5, pp.664-669, 2013.

[15] X. Wang, S. Fang, S. Qiao et al., A comparison of crash data collecting technologies used in China and U.S., *China Safety Science Journal*, vol.22, no.10, pp.79-86, 2012.

[16] X. Wang and X. Chen, Road accident analysis system based on P-GIS, *Journal of Transport Information and Safety*, vol.31, no.5, pp.93-97, 2013.

[17] M. A. Dereli and S. Erdogan, A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods, *Transportation Research Part A: Policy and Practice*, vol.103, pp.106-117, 2017.