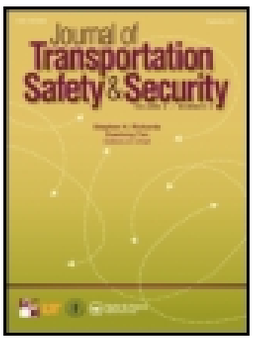

Journal of Transportation Safety & Security

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Identification and spatiotemporal evolution analysis of high-risk crash spots in urban roads at the microzone-level: Using the space-time cube method

Peijie Wu, Xianghai Meng & Li Song

To cite this article: Peijie Wu, Xianghai Meng & Li Song (2021): Identification and spatiotemporal evolution analysis of high-risk crash spots in urban roads at the microzone-level: Using the space-time cube method, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2021.1938323

To link to this article: https://doi.org/10.1080/19439962.2021.1938323

Published online: 29 Jun 2021.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2021.1938323

Identification and spatiotemporal evolution analysis of high-risk crash spots in urban roads at the microzone-level: Using the space-time cube method

Peijie Wu[a], Xianghai Meng[a], and Li Song[b]

[a]School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China; [b]Department of Civil and Environmental Engineering, USDOT Center for Advanced Multimodal Mobility Solutions and Education, University of North Carolina at Charlotte, Charlotte, NC, USA

ABSTRACT

The problem of urban crashes brings huge challenges and threats to local police and governments, especially in many cities in developing countries such as China. To reduce the frequency and severity of urban crashes, the local government in China has gradually taken interest in conducting detailed actions of traffic safety improvement at the microzone-level. Therefore, the primary goal of this study is to try a new method in spatiotemporal data mining techniques, the space-time cube method, to find high-risk crash spots at the spatiotemporal level and to obtain their spatiotemporal evolution patterns. The cumulative frequency curve method was performed to identify high-risk crash spots, and the contributory factors of forming these spots were analyzed by the latent class analysis method. The results showed that: (1) key parameters' selection is crucial in the space-time cube construction; (2) the exit ramp gore point in interchanges, intersections, and entrances of neighborhoods were prone to have many high-risk crash spots at the spatiotemporal scale; and (3) locations with consecutive, persistent, and sporadic hotspots patterns need different risk monitoring strategies and traffic safety improvement. The feasibility and advantages of the space-time cube method in hotspots identification at the microzone-level were confirmed.

KEYWORDS

urban traffic safety; space-time cube model; microzone-level safety; high-risk crash spots; spatiotemporal evolution patterns

1. Introduction

The problem of urban crashes has received an increasing amount of attention from both governments and transportation departments over the last two decades, especially in developing countries like China. In Chinese cities, crashes happen frequently in urban road networks and the number of casualties is relatively high, resulting in a large number of lives and property loss. In 2017, crashes in urban areas of China caused about 16,000

CONTACT Xianghai Meng mengxianghai100@126.com School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

deaths, 52,500 injuries and 300 million yuan in direct property loss (Ministry of Public Security of People's Republic of China, 2018). Because of the complex nature of urban traffic management, various types of vehicles, harmful driving behaviors, and other factors, the problem of urban crashes remains unsolved for most cities in China. Therefore, local governments in China gradually promote traffic safety improvement programs at the microzone-level, such as subdistricts, because regional traffic safety plans usually require a great deal of human and financial resources yet achieved unsatisfactory results. Consequently, traffic safety at the microzone-level in big cities is of great importance and huge interest of local governments in China. The microzone-level refers to one small part of the road network within a region (often urban areas). In previous studies, these networks are often called directed linear networks (Briz-Redón, Martínez-Ruiz, & Montes, 2019a; Briz-Redón, Martínez-Ruiz, & Montes, 2019b). In China, the administrative division of areas refers to these microzone-levels as sub-districts.

As a necessary and typical part of road safety programs, hotspots identification technique (namely black spots, hazard sites, high-risk sites, crash-prone sites, et al.) is widely applied in highways and urban roads. Since the first publication of the Highway Safety Manual (HSM) (AASHTO, 2010), the Empirical Bayes (EB) technique was recommended as one of the state-of-art methods in the network screening process (Elvik, 2008). Montella (2010) believed that the EB method is the most consistent and reliable method when it was compared to other hotspots identification methods with four evaluation criteria. Some limitations of the EB method have been pointed out, which has led to the Full Bayesian (FB) method development (Huang, Chin, & Haque, 2009; Lan & Persaud, 2011). The quality control method and the cumulative frequency curve method were also applied in black spots management (Yang & Li, 2012; Nguyen, Taneerananon, & Luathep, 2015). Nevertheless, conventional hotspots identification approaches sometimes meet troubles in collecting accurate traffic volume data, especially in some access roads or local streets in the urban road networks. Limited crash data, such as those of subdistrict area, also will yield unsatisfactory results of EB methods due to the poor predictive performance of the crash prediction model. To solve these deficiencies in conventional models, Getis-Ord Gi* hotspots analysis, kernel density estimation and its improved version network-based kernel density estimation became popular and were highly acknowledged by many scholars in the last few years (Xie & Yan, 2013; Benedek, Ciobanu, & Man, 2016; Jia, Khadka, & Kim, 2018; Lee & Khattak, 2019). These methods yield reasonable identification results, and they can identify hotspots microscopically. Ziakopoulos and Yannis (2020) pointed out that micro-level road safety and event analysis with spatial considerations appear to be inadequate in the last decades. Therefore, hotspots identification

methods and crash modeling still need to be further investigated at the small scale, including the microzone-level.

The majority of current literature on hotspots identification mostly focuses on spatial or temporal concentration of crash data separately (Plug, Xia, & Caulfield, 2011; Yu, Liu, Chen, & Wang, 2014; Benedek et al., 2016). Nevertheless, the identification methods of crash hotspots in space and time scale simultaneously showed huge potential in the near future (Youngok, Nahye, Serin, & Peng, 2018). Based on the spatial statistic theory, the popularity and prosperity of spatiotemporal data mining techniques were witnessed in many research fields, such as climatology, environmental health and real estate marketing (Banerjee, Carlin, & Gelfand, 2004; Cressie & Wikle, 2011). The space-time cube method, an emerging spatiotemporal data mining method, was proven to be a useful tool for us to analyze space-time closeness of events and spatiotemporal evolution patterns. It is said that the space-time cube approach offers good visual opportunities to study the relationship between time, space and additional variables (Kraak, 2003). Gatalsky, Andrienko, and Andrienko (2004) implemented the space-time cube method to explore the spatiotemporal patterns of earthquakes in the Marmara region. Cheng, Zhang, Peng, Yang, and Lu (2020) introduced the space-time cube model to analyze the spatiotemporal evolution mechanism of pollutant emissions. Because crash data essentially has spatial and temporal attributes, crash data belongs to spatiotemporal data. Therefore, the space-time cube approach can provide more insights and new understandings of dynamic spatiotemporal evolution rules about crashes.

Moreover, some studies showed that the methods used for identifying crash hotspots can be improved by embracing different perspectives, such as crash risks, collision types and vehicle types (Ferreira & Couto, 2015; Briz-Redón et al., 2019b). Crash risks can be understood as a function of crash frequency (or probability of crash occurrence) and crash severity (crash consequences) (Stipancic, Miranda-Moreno, Saunier, & Labbe, 2019; Park & Oh, 2019). Considering that the crash risk was evaluated by the sum of different severity levels of all crashes (Bao, Liu, & Ukkusuri, 2019), we define the crash risk calculated by the sum of equivalent property damage-only crashes (EPDO). Weights of different crash severity (fatal, injury, property damage-only) are determined by the quotient of the direct economic cost of each severity level crashes divided by the direct economic cost of property damage-only (PDO) crashes according to China Statistical Yearbook (National Bureau of Statistics of China, 2019). In this study, the weights of PDO, injury and fatal crashes are 1, 3 and 10, respectively. It is noted that the weights of fatal crashes and injury crashes maybe much smaller than the actual situation in China because the direct economic

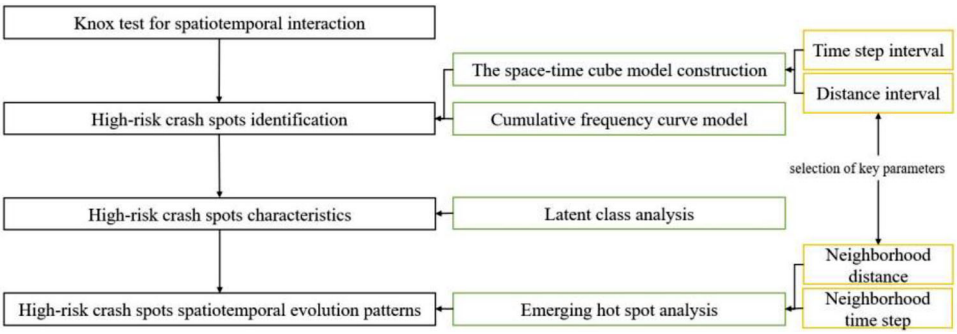

Figure 1. The overall technical route of this study (Song et al., 2020).

costs do not include the medical costs and indirect economic costs of every crash.

The main goal of this study is exploring a new approach of identifying high-risk crash spots in the urban road network and analyzing their spatiotemporal evolution patterns at the microzone-level. The overall technical route of this study is shown as Figure 1. Two important questions of this study are intended to answer: (1) if small size of grid structure applicable in urban roadways at the microzone-level? (2) how is the performance of space-time cube method in microscopic hotspots identification and spatiotemporal evolution analysis? This study contributes to the present knowledge in the following three ways: first, rarely studied methods of selecting appropriate parameters for the space-time cube method based on crash datasets were discussed and their effects on results were also investigated; second, the cumulative frequency curve method was applied to identify spatiotemporal high-risk crash spots and the latent class clustering technique was performed to find the contributory factors of these spots; and third, the results of spatiotemporal evolution patterns provide new understandings of urban crash risks, which is believed to be beneficial for conducting effective risk monitoring strategies and safety improvement actions in these locations.

2. Data

2.1. Road network structure

The road network in Chinese cities is quite different from other cities because there are many interchanges in the central urban area to alleviate traffic congestion problems. Also, the spacing between interchanges and intersections is short. For example, the smallest spacing between the interchange and the intersection is 75 meters in this study area. Therefore, the road network in this study differs from urban road networks in previous studies (Huang et al., 2016; Briz-Redón et al., 2019a; Briz-Redón et al., 2019b). There are 7 interchanges, 117 intersections, and 766 segments in the Huafu

subdistrict of Shenzhen city in China. The complex geometrical shape of interchanges and intersections was simplified by using polygons based on the satellite image. Several factors that could be associated with crashes are considered at the road segment level. These mainly include road classes, number of lanes, and cross-section forms. The road class was divided into six types, namely expressways, arterial roads, sub-arterial roads, access roads, local streets and others (e.g., auxiliary roads, ramps, under-bridge roads). There are 36 expressway segments, 121 arterial road segments, 49 sub-arterial road segments, 168 access road segments, and 216 local streets in this study area.

2.2. Crash dataset

A total of 1299 crashes recorded by the Local Traffic Police Department of Shenzhen (China) from 2014 to 2018 in the Huafu subdistrict was used as a case study. Each of these crashes was geocoded from the address information collected by the police officers. In this work, crashes have been projected into a road network structure in the Arc GIS software (Esri, 2016) and the crash points on the urban roads correspond exactly to the actual location of crashes happening. Several informative attributes were available for each crash point, such as date and time of crashes, types of crashes, types of vehicles involved, crash severity, number of people that got injured, age and gender of drivers.

2.3. Traffic flow datasets

AADT is not available for every road segment in the Huafu subdistrict, such as some local streets, but other types of road class that can be obtained from the local transportation department. Hence, a 5-level categorical covariate was defined to denote the traffic volume in every road, which has already been tested by previous authors (Hao and Daniel, 2014; Briz-Redón et al., 2019a). These five categories of traffic flow are as follows: $AADT < 7,000$ (Level 1), $7,000 \leq AADT < 16,000$ (Level 2), $16,000 \leq AADT < 25,000$ (Level 3), $25,000 \leq AADT < 55,000$ (Level 4) and $AADT \geq 55,000$ (Level 5). By field observation of traffic flows, all the traffic flow of local streets in this subdistrict belongs to level 1 of AADT, so all the traffic flow data is available in this study.

3. Methodology

3.1. Knox test for spatiotemporal interaction

Before using the space-time cube method, the Knox test is recommended for verifying whether events are close to each other within a certain time

and space range. In the Knox test, critical distances in space (δ) and time (τ) have to be specified to yield a categorization of the distances into "close" or "not close" (Knox & Bartlett, 1964). The test statistic is defined as the number of event pairs that are close both in space and time according to these distance thresholds:

$$T_{Knox} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i} \lessdot\left(d_{ij}^{S} \leq \delta\right)\mathbb{I}\left(d_{ij}^{t} \leq \tau\right) \tag{1}$$

If the point pattern exhibits clustering at the predefined spatiotemporal scales, the observed number of close pairs will be larger than the expected number under the null hypothesis of no space-time interaction (Meyer, Warnke, Rossler, & Held, 2016). By applying the Knox test for different time step intervals and distance intervals, the suitable range of time step intervals and distance intervals of space-time cubes can be found as strong spatiotemporal interactions exist.

3.2. Space-time cube model and emerging hot spot analysis

In 1970, Sweden geographer Torsten Hägertrand introduced a new three-dimensional diagram, the space-time cube, to display the life path of people from birth to death (Hägertrand, 1970). The space-time cube method is a time-space concept that describes a cube with events' geographical location (along the x- and y-axis) and time (z-axis), illustrated in Figure 2. ΔT denotes the neighborhood time step (the radius of the temporal influence when calculating the Getis-Ord Gi* index of each bin). ΔS denotes the neighborhood distance (the radius of the spatial influence when calculating the Getis-Ord Gi* index of each bin). T denotes the time step interval of each bin (the height of the square). S denotes the distance interval of each bin (the length and width of the square). This model is often seen as the beginning of time-geography studies and then was widely applied to understand many events' (earthquake, diesel emissions, et al.) movements through time and space (Gatalsky et al., 2004; Cheng et al., 2020). With the advanced computing power of computers and growing interest in spatiotemporal modeling, the space-time cube method is revived now and its improved version called the emerging hot spot analysis (Esri, 2016). This approach can effectively discover the underlying spatiotemporal evolution patterns of many geographical events.

In the emerging hot spot analysis, space-time cubes were constructed firstly as the input, then the Getis-Ord Gi* index (Getis & Ord, 2010) and the Mann-Kendall test (Mann, 1945) are used to detect spatially hot (or cold) spots and time changing trends of these spots, respectively.

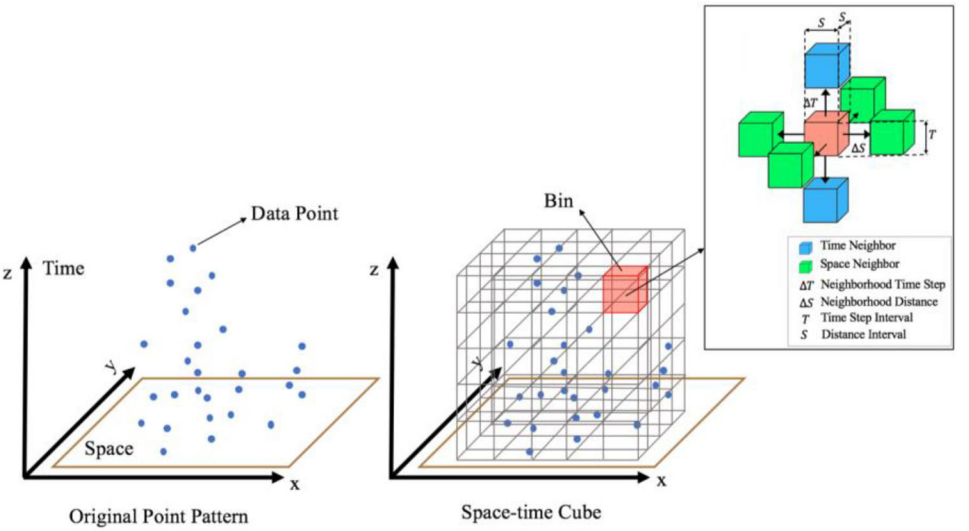

Figure 2. Diagram of space-time cubes (key parameters of the space-time cube method were illustrated, including time step intervals, distance intervals, neighborhood time steps, and neighborhood distances).

The calculation formula of Getis-Ord Gi* index is as follows:

$$G_i^* = \frac{\sum_{j=1}^{n}\omega_{i,j}x_j - \overline{X}\sum_{j=1}^{n}\omega_{i,j}}{S\sqrt{n/\left(n-1\right)\sum_{j=1}^{n}\omega_{i,j}^2 - 1/(n-1)\left(\sum_{j=1}^{n}\omega_{i,j}\right)^2}} \tag{2}$$

where G_i^* denotes the Getis-Ord Gi* index, x_j is the attribute value (crash risk in this study) of the j-th bin in space-time cubes. $\omega_{i,j} = 1$, if the j-th bin within the spatiotemporal neighborhood (including neighborhood time step and neighborhood distance) of the i-th bin. $\omega_{i,j} = 0$, if the j-th bin is out of the spatiotemporal neighborhood of the i-th bin. n represents the total number of bins in the space-time cubes. $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} x_j$ and $S = \sqrt{\frac{1}{n}\sum_{j=1}^{n} x_j^2 - (\overline{X})^2}$ denotes the average and standard deviation crash risk of all the bins in the space-time cubes, respectively. The Getis-Ord Gi* index is z-score, when the p-value has statistical significance (if $z > 0$), the more intense clustering of the higher values (hot spot) will be. When the p-value has statistical significance (if $z < 0$), the more intense clustering of the lower values (cold spot) will be (Ye, Ma, & Ha, 2018).

The Mann-Kendall test is a non-parametric statistical test method that conducts a rank correlation analysis for time-series of z-score at each location and evaluates the time changing trend of hot and cold spots identified by the Getis-Ord Gi* index. For the time series of z-score $\{x_t : t = 1, 2, \ldots, n\}$, the test statistic S is given by (Kendall & Gibbons, 1990, Section. 1.9):

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij} \tag{3}$$

$$a_{ij} = sign(x_j - x_i) = sign(R_j - R_i) = \begin{cases} 1 & x_i < x_j \\ 0 & x_i = x_j \\ -1 & x_i > x_j \end{cases} \tag{4}$$

where *sign* is a symbolic function. R_i and R_j are the ranks of observations x_i and x_j of the time series, respectively. The standard statistic Z_S can be used for the statistical significance test. Its calculation formula is:

$$Z_S = \begin{cases} \dfrac{S-1}{\sqrt{D(S)}}, & S>0 \\ 0, & S = 0 \\ \dfrac{S+1}{\sqrt{D(S)}}, & S<0 \end{cases} \tag{5}$$

When $T \geq 10$, the statistic S obeys the normal distribution approximately; its mean value is 0, and the variance $D(S) = T(T-1)(2T+5)/18$. For a given confidence level p, if $|Z_S| \geq |Z_{S,1-p/2}|$, the bin time-series has a distinct upward or downward trend. If $Z_S > 0$, it indicates an upward trend, while $Z_S < 0$, it indicates a downward trend (Cheng et al., 2020).

Finally, according to the results of the Getis-Ord Gi* index and the Mann-Kendall test, seventeen spatial-temporal evolution patterns (new hot/cold spots, consecutive hot/cold spots, intensifying hot/cold spots, persistent hot/cold spots, diminishing hot/cold spots, sporadic hot/cold spots, oscillating hot/cold spots, historical hot/cold spots, no trend detected) can be obtained (Esri, 2016).

3.3. Key parameters' selection

In the space-time cube method, four key parameters, time step interval, distance interval, neighborhood time step and neighborhood distance play an important role in detecting spatiotemporal hot spots and their evolution patterns. If the time step interval and distance interval of the constructed space-time cube is too large, the point pattern of original data may be lost while the time step interval and distance interval is too small, an excess of empty space-time cubes will be produced. To solve this problem, the Average Nearest Neighbor may provide a rational reference to the proper distance interval of space-time cubes. As for neighborhood time steps and neighborhood distances in the emerging hot spot method, their selection mostly depends on the results of control variable method because of the different roadway structure factors affecting the spatial distribution of crash data.

3.4. Latent class analysis

Latent Class Analysis (LCA), firstly proposed by Lazarsfeld and Henry (1968), is widely used to identify subgroups within a large and heterogeneous population based on observed variables (Yu, Wang, & Abdel-Aty, 2017; Sun, Sun, & Shan, 2019; Li & Fan, 2019). It is a way to discover subgroup heterogeneity and group subjects from multivariate data into "latent classes." The basic assumption of LCA is that the probability distribution of various responses to observed variables can be explained by a few mutually exclusive latent class variables, and each latent class has a specific tendency to respond to observed variables.

In this study, LCA was employed to find reasons caused high crash risk in spots. The mathematical model of LCA is as follows

$$\mathrm{P}((Y_i = yX_i = x) = \sum_{I=1}^{n_c} \gamma_I\Big(x) \prod_{m=1}^{M} \prod_{k=1}^{r_m} \rho_{mk}|I^{I(y_m=k)} \tag{6}$$

$$\gamma_I(x) = P(L_i = l|X_i = x) = \frac{\exp(\beta_{0I} + x\beta_{1I})}{\sum_{j=1}^{n_c}\exp(\beta_{oj} + \beta_{1j})} = \frac{\exp(\beta_{0l} + x\beta_{1l})}{1 + \sum_{j=1}^{n_c-1}\exp(\beta_{0j} + x\beta_{1j})} \tag{7}$$

where $\gamma_I(x)$ is the latent class membership probabilities, ρ is the item-response probabilities conditional on latent class membership, and β is the logistic regression coefficients for covariates, which predicts class membership. The LCA in this study was completed using a SAS procedure, and the parameters are estimated by maximum likelihood using the expectation-maximization procedure.

4. Results and discussion

4.1. Experiments of key parameters in the space-time cube

The Knox test was used to help understand the extent of spatiotemporal interaction of the crash points. The testing distance is between 20 and 1000 meters, and the testing time period is 1 day, 1 week, 1 month, 3 months and 6 months. Results of Knox test show that the distance between 20 and 100 meters has significant spatiotemporal interaction and chi-square of Knox increased with the increased time period and the decreased distance, which means strong spatiotemporal interaction exists within small space distance (e.g. $20 \sim 100$ meters) and larger time period (e.g. 6 months). In addition, the Average Nearest Neighbor method was performed to find the proper distance interval of space-time cubes. The outcomes indicated that the observed mean distance is 12.559 meters, the expected mean distance is 45.704 meters and the ratio of nearest neighbor is 0.275. Therefore, the appropriate distance

interval of space-time cubes is 20 meters, and the time step interval of space-time cubes is 6 months in this study for its significantly clustered point pattern at the spatial and temporal scale.

Next, the control variable method was applied to find the suitable neighborhood time step and neighborhood distance to get spatial-temporal evolution patterns as many as possible. The effect of different neighborhood distance and neighborhood time steps in the space-time cube is further investigated with the use of a set of values in the range [40,500] for the neighborhood distance and varying neighborhood time steps from 6 to 42 months. Tables 1 and 2 showed the testing results of the space-time cube with 20-meter distance interval and 6-month time step interval.

Table 1 showed the changes in spatial-temporal evolution patterns and the ratio of evolution patterns when the neighborhood distance increased. In the aspect of evolution pattern number and its ratio, the total number and its ratio of detected evolution patterns decreased firstly increased then with the rise of neighborhood distance in the space-time cube. This interesting outcome indicates that a different neighborhood distance may change the final evolution pattern distinctly because large neighborhood may include another high crash risk location which leads to the forming of cold spots according to the Getis-Ord Gi* index.

Table 2 illustrated that the increasing neighborhood time step and neighborhood distance will lead to the changes in identified hotspot evolution patterns, especially when the neighborhood distance larger than 60 meters. It indicates that when the neighborhood distance and the neighborhood time steps are large, the ratio of hotspot evolution patterns may be reduced. In this study, the appropriate neighborhood distance and neighborhood time of space-time cubes is 60 meters and 42 months (equals to 7 neighborhood time steps) as it can produce stable hotspot evolution patterns and many time intervals can be taken into consideration for detecting temporal trends in crash risk. However, the final selection of key parameters in the space-time cube method should remain to the decision of the researchers who take full consideration of data size and study area., because crash data in another country or city may produce different findings of spatiotemporal evolution patterns from this study.

4.2. Results of high-risk crash spots identification

High-risk crash spots in this study refer to the definition "space-time cubes with high crash risk value", which is the spatial-temporal unit with high crash risk instead of traditional spatial location with high crash risk in previous literature (Briz-Redón et al., 2019b). A total of 6,650 space-time cubes with 1,077 non-empty ones were yield in this research. Results of the cumulative frequency curve method illustrated that the critical high crash risk value of three can be determined with its cumulative frequency

Table 1. Spatiotemporal evolution results of different neighborhood distances.

Neighborhood distance (meters)	Neighborhood time step (month)	Spatial-temporal evolution patterns	Ratio of evolution patterns (%)
40	6	4 (New hot spot), 11 (Sporadic hot spot)	2.41%
60	6	10 (New hot spot), 9 (Sporadic hot spot)	2.86%
80	6	6 (New hot spot), 9 (Sporadic hot spot)	2.26%
100	6	6 (New hot spot)	0.90%
120	6	2 (New hot spot), 6 (Sporadic hot spot)	1.20%
140	6	2 (New hot spot), 7 (Sporadic hot spot)	1.35%
160	6	2 (New hot spot), 6 (Sporadic hot spot)	1.20%
180	6	2 (New hot spot), 1 (Consecutive hot spot), 6 (Sporadic hot spot)	1.35%
200	6	3 (New hot spot), 2 (Consecutive hot spot), 6 (Sporadic hot spot)	1.65%
220	6	4 (New hot spot,) 3 (Consecutive hot spot), 3 (Sporadic hot spot)	1.50%
240	6	3 (New hot spot), 4 (Consecutive hot spot), 5 (Sporadic hot spot)	1.80%
260	6	2 (New hot spot), 3 (Sporadic hot spot)	0.75%
280	6	1 (Sporadic hot spot)	0.15%
300	6	1 (New hot spot), 1 (Sporadic hot spot)	0.30%
320	6	N/S	0.00%
340	6	2 (New cold spot)	0.30%
360	6	4 (New cold spot)	0.60%
380	6	8 (New cold spot)	1.20%
400	6	8 (New cold spot)	1.20%
420	6	10 (New cold spot)	1.50%
440	6	11 (New cold spot)	1.65%
460	6	9 (New cold spot)	1.35%
500	6	13 (New cold spot)	1.95%

Note. N/S means there is no spatial-temporal pattern or trend detected.

exceeding 85%. When the crash risk of space-time cube increased by 8, the cumulative frequency reaches 99.16%. At last, 415 space-time cubes with a crash risk greater than 3 were identified as the high-risk crash spots in this crash datasets.

Spatial distribution for high-risk crash spots, namely high-risk space-time cubes, was obtained using Arc GIS data visualization tool. As shown in Figure 3, locations with only one high crash risk spot take up a great proportion (58.55%) and are scattered in many road types, such as intersections, interchanges, arterial roads, access roads and so on. There are 7 locations with 4 high-risk crash spots (1.69%), 4 locations with 5 high-risk crash spots (0.96%), and 2 locations with 6 high-risk crash spots (0.48%). Locations with more than four high-risk crash spots are usually located in the upstream of the exit ramp gore point in interchanges, intersections,

Table 2. Ratio of hotspot evolution patterns in different neighborhood time steps.

Neighborhood distance (meters)	Neighborhood time steps (months)						
	6	12	18	24	30	36	42
40	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
60	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
80	100.00%	100.00%	100.00%	100.00%	100.00%	97.44%	98.33%
100	100.00%	100.00%	100.00%	90.91%	100.00%	96.97%	96.30%
120	100.00%	100.00%	91.67%	73.33%	96.67%	97.44%	93.55%
140	100.00%	100.00%	73.33%	73.68%	89.74%	92.86%	93.44%
160	100.00%	71.43%	71.43%	48.00%	84.09%	88.64%	86.57%
180	100.00%	50.00%	38.46%	33.33%	63.49%	73.68%	81.01%
200	100.00%	45.45%	28.13%	29.73%	55.38%	72.13%	78.75%
220	100.00%	23.53%	26.19%	30.23%	54.29%	62.50%	74.68%
240	100.00%	6.90%	17.31%	20.41%	37.33%	56.41%	67.90%
260	100.00%	2.63%	9.38%	12.50%	25.33%	44.44%	52.81%
280	100.00%	1.96%	3.95%	9.86%	22.73%	32.38%	45.63%
300	100.00%	90.74%	74.12%	8.86%	19.10%	32.38%	46.79%

and entrance of neighborhoods, as these places are full of busy traffic (motorized vehicles, non-motorized vehicles and pedestrian), and complex traffic environments (Hou, Tarko, & Meng, 2018; Sun et al., 2019; Mussone, Bassani, & Masci, 2017).

For example, Figure 3g shows one special place where there is an entrance to a hospital. Based on the field observation of this area, there is busy traffic (cars, pedestrian and electric bicycle) at the entrance of the hospital during its opening hours, and traffic organization is quite complex because of the parking lot and an exclusive lane for ambulances. Besides, the intersection of local streets and arterial/sub-arterial roads also presents high crash risks (Figure 3a andf) because of the densely populated neighborhood and frequent vehicles (including cars, bicycles, motorcycles, and electric bicycle) moving in and out.

The time distribution of high-risk crash spots also needed to be taken into consideration. As can be seen from Table 3, the time period between 2015 and 2016 has the maximum number of high-risk crash spots. In the first half of 2015, the average crash risk of space-time cube is the highest among all the time step intervals, which is 3.810. The average and standard deviance of crash risk in high-risk spots are almost stable, which means there is a lot of work to do to further investigate reasons of high-risk crash spots and come up with effective traffic safety improvement countermeasures.

More importantly, characteristics of identified high-risk crash spots can provide us with more valuable information about why these space-time cubes are inclined to high crash risk. In order to unearth the underlying homogeneous clusters and common features of high-risk crash spots, the LCA method was applied in this study, which is often used in recent studies (Yu et al., 2017; Sun et al., 2019).

Considering human factors, vehicle factors, road factors, and environmental factors, a total of twelve variables were selected to describe different

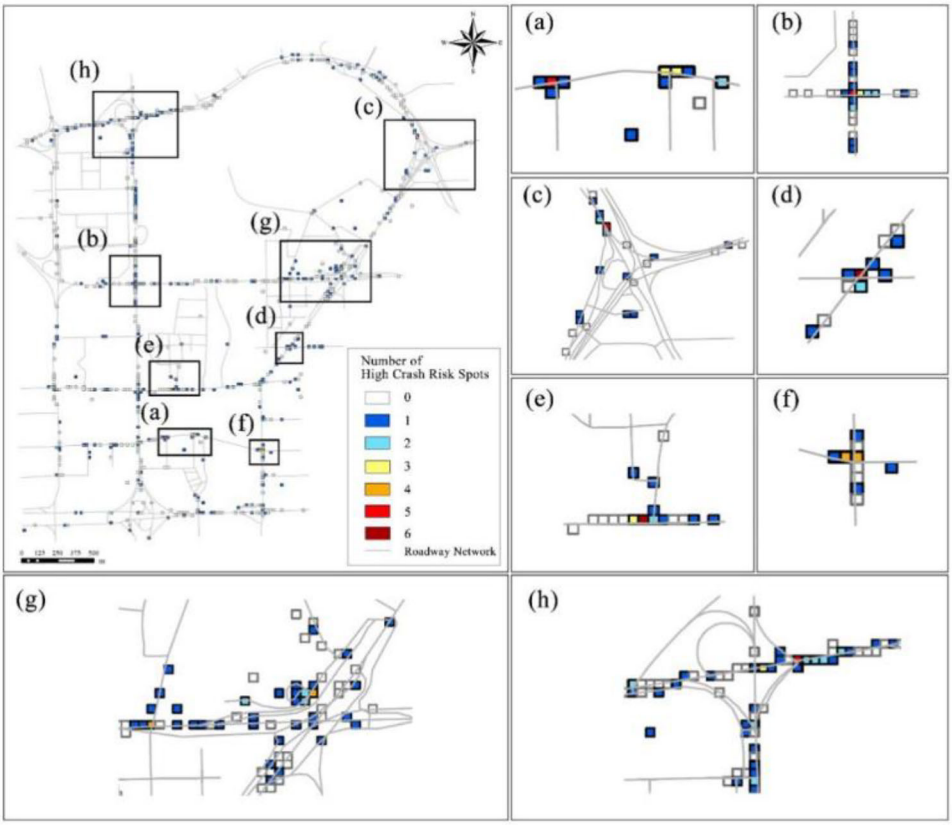

Figure 3. Distribution map for high-risk crash spots (different colors indicate different numbers of high-risk crash spots in the same location, gray lines indicate the roadway network).

characteristics of identified high-risk crash spots. To select the appropriate number of latent clusters, different numbers of clusters were tested, from one to twelve. The BIC, AIC, and CAIC criteria were used to select the final number of clusters. BIC is proved to be more reliable than other evaluation criteria, especially for large datasets (Sun et al., 2019; Li & Fan, 2019). The results of LCA show that the percentage decrease in BIC drops to less than 1% (the percentage decrease in BIC is 0.05% in this study) in the five clusters. Moreover, the quality of the clustering method was assessed by the entropy R squared criterion (McLachlan & Peel, 2000). The entropy value estimated for five clusters was 0.98, which suggests a clear separation between the latent clusters identified by LCA.

Table 4 shows the results of the final set of featured variables that are used for profiling the five classes. In class 1, 100.00% crashes happen on road segments and 68.87% on arterial roads, particularly pedestrians are injured in the 53.77% crashes and the AADT of roads where 46.23% crashes happened is between 16,000 and 25,000 pcu. Therefore, class 1 can be referred to as

Table 3. Distribution of time step interval and crash risk statistics in high-risk crash spots.

Time step interval	Number of high-risk crash spots	Crash risk of high-risk spots			
		Mean	SD	Minimum	Maximum
2013-12-30–2014-06-29	38	3.263	1.270	3	7
2014-06-30–2014-12-29	43	3.700	1.286	3	10
2014-12-30–2015-06-29	42	3.810	1.287	3	7
2015-06-30–2015-12-29	47	3.550	1.268	3	10
2015-12-30–2016-06-29	47	3.600	1.286	3	9
2016-06-30–2016-12-29	44	3.360	1.293	3	8
2016-12-30–2017-06-29	37	3.220	1.281	3	9
2017-06-30–2017-12-29	43	3.510	1.281	3	12
2017-12-30–2018-06-29	35	3.430	1.281	3	10
2018-06-30–2018-12-29	39	3.360	1.279	3	6

"Crashes on arterial road segments and pedestrians get injured frequently." Cluster 2 resembles cluster 1 for road type, road class, and traffic volume, but differs from injury type and crash form. 53.93% crashes are rear-ended collisions and people in the car get injured in the 61.80% crashes in cluster 2. For class 3, 98.96% crashes on expressway road segments with 77.08% rear-end crashes, which can be seen as "Crashes on expressway road segments and often rear-end collision." Cluster 4 has 52.98% of the crashes happening on intersections with AADT less than 7,000 pcu and the majority of crashes belong to other crash forms. Cluster 4 can be regarded as "Crashes on intersections with low traffic volume." The last class overlaps class 4 on "road type", "road class" and "traffic volume", but 52.14% crashes are other injury types (e.g. PDO) and 60.68% crashes are rear-ended collisions. By conducting the LCA method to analyze the reason of generating high crash risk locations, we can find that five crash clustering need more attention from transportation departments, and them can be considered as the main causes of high crash risks in those spatiotemporal spots.

4.3. Spatiotemporal patterns of high-risk crash spots

As it has already been explained before, 6 months for time step interval, 20 meters for distance interval, 60 meters for neighborhood distance and 42 months for neighborhood time step were set in the emerging hot spot method based on the existing data. Final spatiotemporal patterns of the space-time cube method were shown in Figure 4, including 62 locations with three spatiotemporal patterns identified (consecutive hot spot, persistent hot spot, and sporadic hot spot).

By comparing Figure 3 with Figure 4, locations with major high-risk crash spots are less likely to become the hot spots in spatiotemporal patterns (see Figures 3e and 4g; Figures 3g and 4h). Besides, locations with fewer high-risk crash spots can possibly turn into spatiotemporal pattern hot spots (see Figures 3f and 4e; Figures 3h and 4a). Therefore, hot spots as a result of spatiotemporal patterns are different from high-risk crash

Table 4. Summary of variables and their distribution in each latent cluster of high-risk crash spots.

Variables	Description	Whole database	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Age of driver							
Age-24	Young driver (age less than 24 years old)	7.87%	8.49%	4.49%	6.25%	9.93%	8.55%
Age 25–54	Middle-aged driver (age between 25 to 54 years old)	89.45%	89.62%	91.01%	91.67%	88.08%	88.03%
Age-55	Old driver (age greater than 55 years old)	2.68%	1.89%	4.49%	2.08%	1.99%	3.42%
Gender							
Male	Driver is male	90.34%	90.57%	91.01%	89.58%	86.75%	94.87%
Female	Driver is female	9.66%	9.43%	8.99%	10.42%	13.25%	5.13%
Injury type							
I-Ped	Pedestrian is injured in the crash	23.97%	53.77%	0.00%	0.00%	49.67%	1.71%
I-Bicy	People in the bicycle is injured in the crash	7.51%	12.26%	1.12%	0.00%	17.88%	0.85%
I-E-bicy	People in the electric bicycle is injured in the crash	11.27%	20.75%	0.00%	0.00%	27.15%	0.00%
I-M-bicy	People in the motorcycle is injured in the crash	0.18%	0.00%	0.00%	1.04%	0.00%	0.00%
I-Car	People in the car is injured in the crash	30.59%	10.38%	61.80%	45.83%	5.30%	45.30%
Other-injury		26.48%	2.83%	37.08%	53.13%	0.00%	52.14%
Crash severity							
PDO	Property damage-only in the crash	22.54%	0.00%	25.84%	45.83%	0.00%	50.43%
Injury	Someone is injured in the crash	77.28%	99.06%	74.16%	54.17%	100.00%	49.57%
Fatal	Someone is dead in the crash	0.18%	0.94%	0.00%	0.00%	0.00%	0.00%
Vehicle type (at fault)							
Car	Car at fault in the crash	81.04%	70.75%	93.26%	84.38%	76.16%	84.62%
Bus	Bus at fault in the crash	4.65%	5.66%	4.49%	6.25%	1.99%	5.98%
Van	Van at fault in the crash	2.50%	0.94%	0.00%	4.17%	3.31%	3.42%
Heavy-veh	Heavy vehicle at fault in the crash	1.61%	0.00%	2.25%	3.13%	1.32%	1.71%
M-bicy	Motorcycle at fault in the crash	0.72%	1.89%	0.00%	1.04%	0.66%	0.00%
E-bicy	Electric bicycle at fault in the crash	6.80%	16.04%	0.00%	0.00%	10.60%	4.27%
Bicy	Bicycle at fault in the crash (not electricly powered)	2.68%	4.72%	0.00%	1.04%	5.96%	0.00%
Crash form							
Side-coll	Side collision	15.56%	0.94%	39.33%	17.71%	2.65%	25.64%
Rear-end	Rear-end collision	34.53%	0.00%	53.93%	77.08%	0.00%	60.68%
Head-on	Head-on collision	1.07%	0.00%	1.12%	1.04%	0.66%	2.56%
Fix-obj	Fixed-object collision	3.76%	8.49%	4.49%	2.08%	1.99%	2.56%
Other-form		45.08%	90.57%	1.12%	2.08%	94.70%	8.55%
Road type							
Road-seg	Crash occurred in the road segment	61.18%	100.00%	100.00%	100.00%	32.45%	1.71%
Intersection	Crash occurred in the intersection	24.51%	0.00%	0.00%	0.00%	52.98%	48.72%
Interchage	Crash occurred in the interchange	10.55%	0.00%	0.00%	0.00%	2.65%	47.01%
Other-road	Crash occurred in parking lot, walkway, et al.	3.76%	0.00%	0.00%	0.00%	11.92%	2.56%
Road class							
Express-way	Expressway	22.90%	31.13%	0.00%	98.96%	0.00%	0.00%
Arterial-road	Arterial road	26.30%	68.87%	83.15%	0.00%	0.00%	0.00%

(*continued*)

Table 4. Continued.

Variables	Description	Whole database	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Subarterial-road	Sub-arterial road	6.80%	0.00%	11.24%	0.00%	18.54%	0.00%
Access-road	Access road	0.89%	0.00%	0.00%	0.00%	3.31%	0.00%
Local-street	Local street in a community	2.33%	0.00%	5.62%	0.00%	5.30%	0.00%
Other-class	Including intersection, interchange, parking lot	40.79%	0.00%	0.00%	1.04%	72.85%	100.00%
Weather							
Clear	Clear weather	84.79%	89.62%	79.78%	86.46%	84.11%	83.76%
Rainy	Rainy weather	10.91%	8.49%	11.24%	10.42%	11.26%	12.82%
Cloudy	Cloudy weather	4.29%	1.89%	8.99%	3.13%	4.64%	3.42%
Daytime/nighttime							
Daytime	Crash occurred in the daytime (6:00–20:00)	77.82%	81.13%	76.40%	78.13%	81.46%	70.94%
Nighttime	Crash occurred in the nighttime (20:00–6:00)	22.18%	18.87%	23.60%	21.88%	18.54%	29.06%
If rush-hour							
In-rushhour	Crash occurred in the rush hours (7:30–9:30 & 17:30–19:30)	20.57%	19.81%	19.10%	15.63%	26.49%	18.80%
Not-rushhour	Crash occurred not in the rush hours	79.43%	80.19%	80.90%	84.38%	73.51%	81.20%
Traffic volume							
Level_1	AADT less than 7000 pcu	41.50%	0.00%	16.85%	0.00%	97.35%	59.83%
Level_2	AADT between 7000 and 16000 pcu	9.48%	32.08%	21.35%	0.00%	0.00%	0.00%
Level_3	AADT between 16000 and 25000 pcu	21.11%	46.23%	61.80%	6.25%	1.32%	5.13%
Level_4	AADT between 25000 and 55000 pcu	3.40%	5.66%	0.00%	11.46%	1.32%	0.00%
Level_5	AADT larger than 55000 pcu	24.51%	16.04%	0.00%	82.29%	0.00%	35.04%

spots identified by the cumulative frequency curve method because Getis-Ord Gi* statistics measure how intense clustering is by taking the neighboring bins (bins within range of neighborhood time step and neighborhood distance) into consideration. The majority of these patterns are consecutive hot spots in this study, which means that there is a single uninterrupted run of statistically significant hot spot bins in the final time-step intervals and less than 90% of all bins are statistically significant hot spots in the same location. In terms of persistent hot spot patterns, like Figure 4a, b, and g, there are stable hot spot bins in at least 90% of time step intervals without the tendency of increasing nor declining. And sporadic hot spot patterns (see Figure 4b and g) indicate that there are on-again then off-again hot spot bins in the same location and can be regarded as "hot spots appeared at odd time intervals." Take one persistent hot spot (see Figure 4b) for example, the trend z-score and trend p-value of crash risk in this location are 0.716 and 0.474 respectively, which indicates a trend in time but not significant measured by the Mann-Kendall method. According to the above analysis of spatiotemporal pattern results, the temporal instability of crash risk hot spots was revealed for there is no intensifying hot spots pattern or diminishing hots spot pattern found in this

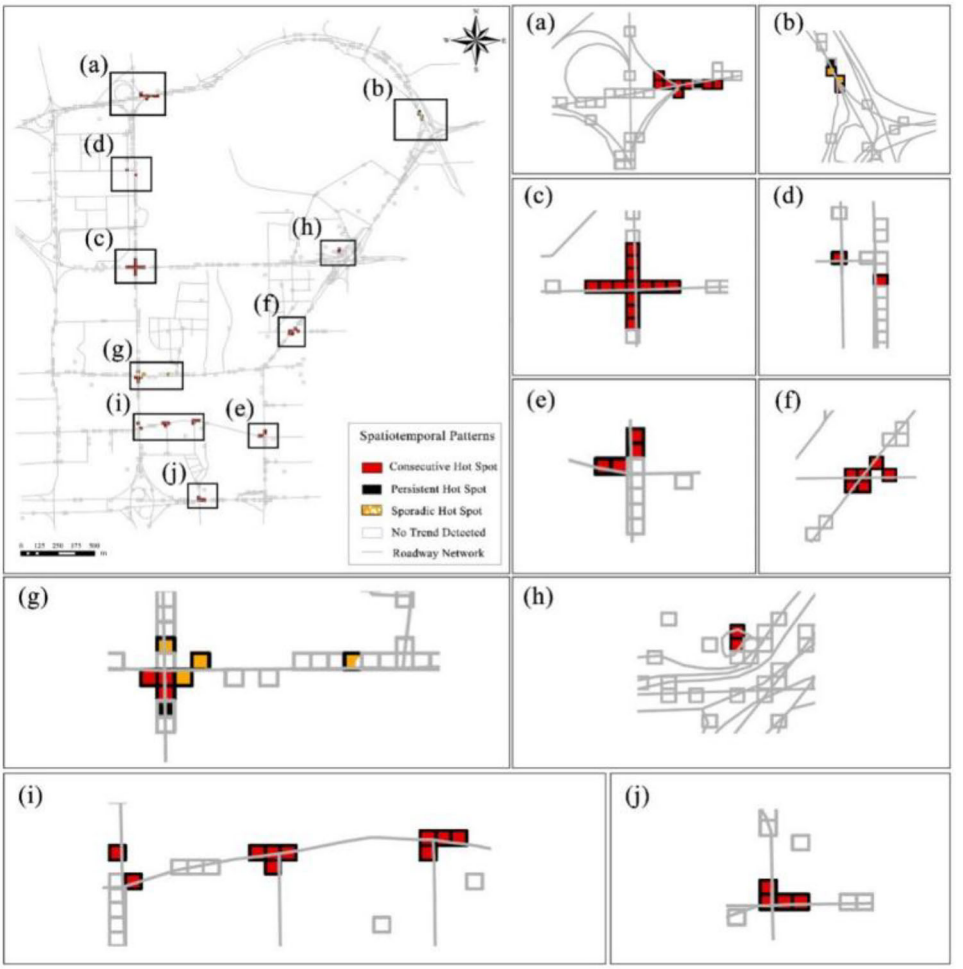

Figure 4. Spatiotemporal patterns map for different locations (different colors indicate different spatiotemporal patterns analyzed by the space-time cube method; gray lines indicate the roadway network).

study, and the temporal instability of crash rate was also found in Hou, Huo, and Leng (2020). Thus, spatiotemporal evolution patterns can help us get more insights about spatiotemporal underlying features because it takes full consideration of the dynamic time change trends of crash risk (e.g., consecutive, intensifying, persistent, diminishing, sporadic, oscillating).

5. Conclusions

Traffic safety analysis at the microzone-level has been of interest in recent years, and the quick development of spatiotemporal data mining techniques is bringing new chances and challenges for this subject. Specifically, in this

study, a subdistrict in Shenzhen city in China has been used to analyze a geocoded dataset of crashes during 2014-2018. In this regard, the identification of high-risk crash spots at the space and time scale and the spatiotemporal evolution patterns of these spots have been discussed.

In the first place, the study of high-risk crash spots along the urban road network is of great interest recently. Because crash risk taking full consideration of crash frequency and crash severity simultaneously, it can help transportation departments find more potential crash locations. However, the majority of present researches about high-risk spots identification focused on spatial distribution or temporal distribution separately instead of the spatiotemporal aspect. In this study, the cumulative frequency curve was employed to identify high-risk crash spots at the spatiotemporal scale and the underlying reasons of these spots forming are further analyzed by the LCA method. The results showed that pedestrian-injured crashes mostly happen on arterial road segments and intersections, and PDO crashes happen on expressways with the majority being rear-end collisions.

Next, the space-time cube method is a popular and emerging spatiotemporal data mining tool in other subjects, which may provide more insights about crash datasets from the perspective of spatiotemporal features. In this article, methods of selecting four key parameters in the space-time cube method based on crash datasets were proposed using average nearest neighbor and control variable method. They can be the reference methods for choosing appropriate parameters of the space-time cube method to deeply mining the spatiotemporal characteristics of crash data. The effect of different parameters on final results was also explored in this study, which indicates that spatiotemporal evolution patterns depend largely on neighborhood time steps and neighborhood distances. Due to the sample size of this study, more crash datasets can be examined by using the space-time cube method in future researches.

Spatiotemporal evolution patterns of crash risk were insufficient in the field of traffic safety. By analyzing spatiotemporal dynamic changing patterns, the complex time and space changing patterns can be discovered quickly and provide valuable information for traffic safety improvement plans for local governments. For instance, field observation and safety improvement should be emphasized on new, consecutive, intensifying and persistent hotspots because these locations may turn into high-risk spots in the future. In addition, the safety performance of different urban traffic safety actions can be evaluated by studying the spatiotemporal patterns of corresponding locations, such as the diminishing hot spots indicates the declining time trend of crash risk in these locations.

By the research of this study, two important questions mentioned in the introduction section were answered. Firstly, though grid structure seems to be

improper for certain areas, such as traffic analysis zone or regional areas, it indeed applicable at the microzone-level with small-size grid ($20 \sim 100$ meters), which perfectly divides complex road entity (e.g. interchanges) into micro analysis zones, helping find hotspots microscopically. Secondly, the actual performance of space-time cube method is a good option for traffic safety analysis, and it can reveal the spatiotemporal closeness among crash data and their dynamic spatiotemporal evolution patterns at the micro-level which rarely studied before. In conclusion, the space-time cube method was recommended to be a useful spatiotemporal data mining data technique in the traffic safety field, hopefully helping us gain more knowledge of the complex and dynamic spatiotemporal features of crash risk. However, the space-time cube method also has its drawbacks, such as containing less distances of roadways in each bin when the study road is not parallel to the edge of the square. In order to avoid this problem, we recommend that crash rates (crash number per meter) in each bin, and bins whether contain intersections should be taken account of in the future study of the space-time cube method. Besides, there is no evaluation criteria for us to compare traditional methods (e.g., quality control method, EB/FB method, and kernel density estimation) and the space-time cube method because their research objects are different (the former ones identify spatial hotspots and the latter one identifies spatiotemporal hotspots). Hence, the performance of the space-time cube method and the spatiotemporal kernel density estimation (STKDE) method to identify hotspots is needed to be further discussed in the future.

References

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press.

Bao, J., Liu, P., & Ukkusuri, S. V. (2019). A spatiotemporal deep learning approach for city-wide short-term crash risk prediction with multi-source data. *Accident Analysis and Prevention*, *122*, 239–254. doi:10.1016/j.aap.2018.10.015

Benedek, J., Ciobanu, S. M., & Man, T. C. (2016). Hotspots and social background of urban traffic crashes: a case study in cluj-napoca (romania). *Accident Analysis & Prevention*, *87*, 117–126. doi:10.1016/j.aap.2015.11.026

Briz-Redón, A., Martínez-Ruiz, F., & Montes, F. (2019a). Spatial analysis of traffic accidents near and between road intersections in a directed linear network. *Accident Analysis and Prevention.*, *132*, 105252. doi:10.1016/j.aap.2019.07.028

Briz-Redón, A., Martínez-Ruiz, F., & Montes, F. (2019b). Identification of differential risk hotspots for collision and vehicle type in a directed linear network. *Accident Analysis and Prevention.*, *132*, 105278. doi:10.1016/j.aap.2019.105278

Cheng, S., Zhang, B., Peng, P., Yang, Z., & Lu, F. (2020). Spatiotemporal evolution pattern detection for heavy-duty diesel truck emissions using trajectory mining: A case study of Tianjin. *Journal of Cleaner Production*, *244*, 118654. doi:10.1016/j.jclepro.2019.118654

Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Singapore: John Wiley.

Esri. (2016). *ArcGIS 10.4 for Desktop Web Help: Create space time cube*. Retreived from http://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/create-space-time-cube.htm.

Ferreira, S., & Couto, A. (2015). A probabilistic approach towards a crash risk assessment of urban segments. *Accident Analysis and Prevention.*, *50*, 97–105. doi:10.1016/j.trc.2014.09.012

Gatalsky, P., Andrienko, N., & Andrienko, G. (2004). *Interactive analysis of event data using space-time cube*. Eighth International Conference on Information Visualisation, Proceedings [Paper presentation]. , IEEE.

Getis, A., & Ord, J. K. (2010). *The analysis of spatial association by use of distance statistics*. Berlin, Heidelberg: Springer.

Hägertrand, T. (1970). *What about people in regional science?* Ninth European Congress of the Science Association.

Hou, Q., Huo, X., & Leng, J. (2020). A correlated random parameters tobit model to analyze the safety effects and temporal instability of factors affecting crash rates. *Accident Analysis and Prevention.*, *134*, 105–326.

Hou, Q., Tarko, A. P., & Meng, X. (2018). Investigating factors of crash frequency with random effects and random parameters models: New insights from Chinese freeway study. *Accident Analysis and Prevention*, *120*, 1–12. doi:10.1016/j.aap.2018.07.010

Hao, W., & Daniel, J. (2014). Motor vehicle driver injury severity study under various traffic control at highway-rail grade crossings in the united states. *Journal of Safety Research*, *51*, 41–48. doi:10.1016/j.jsr.2014.08.002

Huang, H., Chin, H. C., & Haque, M. M. (2009). Empirical evaluation of alternative approaches in identifying crash hot spots. Naive ranking, empirical Bayes, and full Bayes methods. *Transportation Research Record: Journal of the Transportation Research Board*, *2103*(1), 32–41. doi:10.3141/2103-05

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M. (2016). Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography*, *54*, 248–256. doi:10.1016/j.jtrangeo.2016.06.012

Jia, R., Khadka, A., & Kim, I. (2018). Traffic crash analysis with point-of-interest spatial clustering. *Accident Analysis and Prevention*, *121*, 223–230. doi:10.1016/j.aap.2018.09.018

Kendall, M. G., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). London: Edward Arnold.

Knox, E. G., & Bartlett, M. S. (1964). The detection of space-time interaction. *Applied Statistics*, *13*(1), 25–99. doi:10.2307/2985220

Kraak, M. (2003 *The space-time cube revisited from a geovisualization perespective [Paper presentation]*. Proceedings. of the 21st International Cartographic Conference (ICC), 1988–1996.

Lan, B., & Persaud, B. (2011). Fully Bayesian approach to investigate and evaluate ranking criteria for blackspot identification. *Transportation Research Record: Journal of the Transportation Research Board*, *2237*(1), 117–125. doi:10.3141/2237-13

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. New York: Houghton Mifflin.

Lee, M., & Khattak, A. J. (2019). Case study of crash severity spatial pattern identification in hot spot analysis. *Transportation Research Record: Journal of the Transportation Research Board* Advanced online publication. doi:10.1177/0361198119845367

Li, Y., & Fan, W. (2019). Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: A case study of North Carolina. *Accident Analysis & Prevention*, *131*, 284–296. doi:10.1016/j.aap.2019.07.008

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245. doi:10.2307/1907187

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics, John Wiley & Sons, New York.

Meyer, S., Warnke, I., Rossler, W., & Held, L. (2016). Model-based testing for space-time interaction using point processes: An application to psychiatric hospital admissions in an urban area. *Spatial and Spatio-Temporal Epidemiology*, *17*, 15–25. doi:10.1016/j.sste.2016.03.002

Ministry of Public Security of People's Republic of China (2018). *Annual Statistical Reports of Road Traffic Accidents*.

Montella, A. (2010). A comparative analysis of hotspot identification methods. *Accident Analysis & Prevention*, *42*(2), 571–581. doi:10.1016/j.aap.2009.09.025

Mussone, L., Bassani, M., & Masci, P. (2017). Analysis of factors affecting the severity of crashes in urban road intersections. *Accident Analysis and Prevention*, *103*, 112–122. doi: 10.1016/j.aap.2017.04.007

National Bureau of Statistics of China (2019). *China statistical yearbook*. Beijing: China Statistics Press.

Nguyen, H. H., Taneerananon, P., & Luathep, P. (2015). Approach to Identifying Black Spots Based on Potential Saving in Accident Costs. *Engineering Journal*, *20*(2), 110–122.

Park, H., & Oh, C. (2019). A vehicle speed harmonization strategy for minimizing inter-vehicle crash risks. *Accident Analysis and Prevention*, *128*, 230–239. doi:10.1016/j.aap.2019.04.014

Plug, C., Xia, J. C., & Caulfield, C. (2011). Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis and Prevention*, *43*(6), 1937–1946. doi:10.1016/j.aap.2011.05.007

Song, L., Li, Y., Fan, W., & Wu, P. (2020). Modeling pedestrian-injury severities in pedestrian-vehicle crashes considering spatiotemporal patterns: insights from different hierarchical bayesian random-effects models. *Analytic Methods in Accident Research*, 100137.

Stipancic, J., Miranda-Moreno, L., Saunier, N., & Labbe, A. (2019). Network screening for large urban road networks: Using GPS data and surrogate measures to model crash frequency and severity. *Accident Analysis and Prevention*, *125*, 290–231. doi:10.1016/j.aap.2019.02.016

Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis and Prevention*, *124*, 50–57. doi:10.1016/j.aap.2018.12.016

Xie, Z., & Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography*, *31*, 64–71. doi:10.1016/j.jtrangeo.2013.05.009

Yang, J., & Li, Q. (2012). *Progress in industrial and civil engineering (Part 1)*. Switzerland: Trans Tech Publications.

Ye, W., Ma, Z., & Ha, X. (2018). Spatial-temporal patterns of PM2.5 concentrations for 338 Chinese cities. *The Science of the Total Environment*, *631-632*, 524–533. 631e632, 524e533. doi:10.1016/j.scitotenv.2018.03.057

Youngok, K., Nahye, C., Serin, S., & Peng, C. (2018). Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation. *PLOS One*, *13*(5), e0196845.

Yu, H., Liu, P., Chen, J., & Wang, H. (2014). Comparative analysis of the spatial analysis methods for hotspot identification. *Accident Analysis and Prevention*, *66*, 80–88. doi:10.1016/j.aap.2014.01.017

Yu, R., Wang, X., & Abdel-Aty, M. (2017). A hybrid latent class analysis modeling approach to analyze urban expressway crash risk. *Accident Analysis and Prevention*, *101*, 37–43. doi:10.1016/j.aap.2017.02.002

Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. *Accident Analysis and Prevention*, *135*, 105323. doi:10.1016/j.aap.2019.105323