

Article

A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China

Ke Nie ^{1,2}, Zhensheng Wang ^{1,2}, Qingyun Du ^{1,2,*}, Fu Ren ^{1,2} and Qin Tian ^{1,2}

¹ School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; E-Mails: nieke@whu.edu.cn (K.N.); wangzhens@whu.edu.cn (Z.W.); renfu@whu.edu.cn (F.R.); tianqin@whu.edu.cn (Q.T.)

² Key Laboratory of GIS, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

* Author to whom correspondence should be addressed; E-Mail: qydu@whu.edu.cn; Tel.: +86-27-6877-8842; Fax: +86-27-6877-8893.

Academic Editor: Marc A. Rosen

Received: 6 November 2014 / Accepted: 17 February 2015 / Published: 4 March 2015

Abstract: Research on spatial cluster detection of traffic crash (TC) at the city level plays an essential role in safety improvement and urban development. This study aimed to detect spatial cluster pattern and identify riskier road segments (RRSs) of TC constrained by network with a two-step integrated method, called NKDE-GLINCS combining density estimation and spatial autocorrelation. The first step is novel and involves in spreading TC count to a density surface using Network-constrained Kernel Density Estimation (NKDE). The second step is the process of calculating local indicators of spatial association (LISA) using Network-constrained Getis-Ord G_i^* (GLINCS). GLINCS takes the smoothed TC density as input value to identify locations of road segments with high risk. This method was tested using the TC data in 2007 in Wuhan, China. The results demonstrated that the method was valid to delineate TC cluster and identify risk road segments. Besides, it was more effective compared with traditional GLINCS using TC counting as input. Moreover, the top 20 road segments with high-high TC density at the significance level of 0.1 were listed. These results can promote a better identification of RRS, which is valuable in the pursuit of improving transit safety and sustainability in urban road network. Further research should address spatial-temporal analysis and TC factors exploration.

Keywords: network-constrained; spatial cluster pattern; traffic crash; Kernel Density Estimation; Getis-Ord Gi*; riskier road segments

1. Introduction

In recent years, the Chinese vehicle fleet has experienced rapid growth. The production and sales of vehicles in China has been ranked first in the world [1]. However, according to statistics from National Bureau of Statistics, in 2006 the death rate per million registered vehicles in China was up to 6.2, whereas the number in USA was estimated to be approximately 1.77, the number in Japan was only 0.77 [2,3]. Fatality in the transportation industry is 77.6% of the total fatality number in safety production, 15 times the number in the area coal mining industry [3]. Traffic crash (TC) has been the number one “killer” that threatens people’s lives and property in China [4]. How to keep public transit safety and sustainable has become a major concern for the Department of Road Administration. Wuhan, with thoroughfares to nine provinces in China, is one of the largest cities in the country, with a continuing expansion of urban area and highway construction. In the period of 2000–2010, the number of motor vehicle grew at 25 percent and continues to grow with confidence, however, the road mileage increased by 3 percent [5]. As the motor vehicle trip demand increases, the tense situation between traffic supply and demand emerged. Frequent TCs and traffic congestions have kept adversely affecting people’s everyday life and social development. Stringent efforts should be made to uplift the traffic safety standards and control traffic congestion for a sustainable development of transportation and urbanization. Besides, it is confirmed that TCs contribute around 10 to 15 percent of random traffic congestion, and cause the greatest amount of lost time due to congestion delays [6]. A systematic analysis of TC scenario, proper traffic control devices, suitable roadway design practices and effective traffic police activities can often help to reduce TC. Moreover, it has been proved that spatial analysis could provide an effective solution to detect the pattern and suggest reasons for the pattern characteristics [7–9]. The detection of TC pattern and identification of hot spots is the first step of TC strategies, including identification, ranking, profiling and treatment [10]. Nevertheless, city-level empirical research on spatial pattern of TC and risk road identification in China is lacking.

Two decades ago, it had been noted that “there has been little published on the geography of traffic crashes” [11]. This clearly has changed over the last two decades [12]. In a broad sense, TC is the result of interactions between human activities and geographical environment. In geographical space, TC was abstracted as discrete event in area or line. When considering factors associated with TC, there are driver factors, motor vehicle conditions, roadway conditions, traffic characteristics and environmental factors [13]. Driver factors, defined as subjective judgments, are always involved in reacting to objective conditions, such as roadway conditions. Therefore, the improvement of objective conditions can result in the decrease of TCs. In one sense, knowing the influence of road condition has on crashes would help target the maintenance effort for the road system. Nevertheless, considering the costs and resources for the improvement of objective conditions, the question becomes how to determine the road with priority to be mended, particularly with respect to which roadway segments are riskier than others. RRSs are defined as segments with more TCs in the same time interval and the equal roadway

length. Considering that TC is a kind of point event, which often occurs along roadways, RRSs can be identified by detecting the cluster pattern of TC along roadways at a the city scale. Thus, the detection of spatial clusters of TC is an essential approach to identifying RRSs for the appropriate allocation of resources for road safety improvements [14].

Research on the spatial pattern of TC has substantially progressed during the last years. Previous work has shown that the distribution of TC have apparent spatial cluster characteristics [14–17], *i.e.*, there are TC hot spots, hot road segments or hot areas that are a combined geographic unit of high-risk points, road segments or areas [10,18–20]. According to the scale of the research object and its research area, there are two main methods for analyzing the spatial pattern of geographic events: the area statistics method and the discrete event method.

As for area statistics, due to the spatial heterogeneity and spatial dependence of areas, the global spatial autocorrelation and local indicators of spatial association (LISA) are used to measure the cluster degree of attributes of the area, such as TC from different areas [7,21,22]. The global spatial autocorrelation method, including global Moran's I, global Geary's C, and Getis-Ord G_i^* , can be used to describe the distribution of TC across the study area, however the location of TC clusters and the differences among each TC clusters were ignored [23]. Afterwards, the LISA method, such as local Getis-Ord G_i^* improved from global Getis-Ord G_i^* , had been proved to be available for detecting hot areas and identifying the center of a cluster at a significant level [24]. Due to the fact that TC is spatial event in planar space but constrained in road network, the network-constrained LISA named local indicators of network-constrained clusters (LINCS) was proposed [25]. The GLINCS, based on G statistics, and ILINCS, based on I statistics, are mostly used LINCS in network space.

In regard to the spatial cluster of discrete events, there are approaches, including descriptive analysis, quadrat analysis and distance analysis. Furthermore, the most typical methods based on discrete events may be the nearest neighbor distance method, Ripley's K function methods [26], Kernel Density Estimation (KDE) methods [27,28] and others. Traditionally, the KDE methods have been widely used in point-pattern analyses for discrete events, especially in TC analyses [14,29,30]. Although no single technique has emerged as the "best" for detecting and predicting TC clusters, recent research suggests that KDE outperforms other approaches due to its simplicity and easy implementation [31]. Also, the KDE method may outperform the empirical Bayesian method in the identification of hazardous road segments when only the location of the crash can be used for the analysis [32]. However, although the KDE has shown acceptable properties using density values, its homogeneous 2D assumption for events distributed in 1.5D space, such as TC on a road network, seems to be irrelevant [33–38]. To overcome this limitation, Okabe proposed the idea of the spatial analysis based on a network, Network-Constrained Kernel Density Estimation (NKDE), which can overcome the shortcomings of the KDE method and reduce the deviation of its results [39–41]. Furthermore, research has demonstrated the validity of NKDE to analyze network-based phenomena, such as TC [35,42–47].

Although KDE and NKDE are useful methods for the spatial cluster analysis in TC research, they had some limitations. One inevitable problem is the local maximums and boundary effects due to the derivation of the kernel function. Therefore, deciding which clusters are statistically significant is necessary. Nevertheless, there is not enough attention paid to the statistical significance of KDE in the current literature [48]. Meanwhile, the same question has been proposed by some researchers, such as

Xie and Anderson [14,42]. Xie noted that NKDE has one of the same fundamental drawbacks as planar KDE. Moreover, Plug said that KDE is better for visualization purposes than for identification of black spots [49,50].

Hence, in this paper, firstly KDE and NKDE are compared to portray the spatial cluster characteristic of TC in area scale and network scale, respectively. Still, as to each road network polyline, the NKDE generates a smoothing density surface with reduction of data noise and statistical bias. Secondly, considering the statistical significance of NKDE, GLINCS method is used to identify high-risk road segments by using the density value as input attributes. Next, the result of NKDE-GLINCS is compared with the GLINCS.

This paper aims to evaluate and represent the TC pattern to contribute to the traffic safety in Wuhan City. The detection can help to identify vulnerable locations and road segments that require remedial measures. A spatial method for visualization of TC spatial cluster and identification of risky road segments is expounded. The remainder of this manuscript is organized in the following manner: first, descriptions of data used in the current study; second, explanation of methods; third, results of our analyses; and finally, discussion of implications and limitations of the methods, and suggestions for future research.

2. Data and Methods

2.1. Data Source

TC in this research is defined as a motor vehicle crash with death or property loss that occurs on, and is restrained to, the roadway network in central urban areas. Simultaneously, it is a type of geographical phenomenon in the subset of 2-D geo-space. The main data sources of the present study collected from Wuhan in 2007 are as follows: (1) TC data from the Traffic Management Department; (2) roadway network; and (3) administrative map. The data from 2007 was chosen because the main form of transportation had been generated within the roadway network when neither metro nor the “three urban highway” had been employed. All the data were stored as shapefiles and shown as a TC map in an ArcGIS platform (Figure 1).

The TC data is stored in a geo-database with the basic information (such as time of day, fatality or serious injury, driver’s characteristics, environmental conditions, vehicle’s characteristics, *etc.*), roadway information (type of traffic control device, light condition, type of road, *etc.*) and handling information (insurance, compensation, cleaning up accident scene, *etc.*). According to the basic information, the total number of crashes in 2007 was 3113, consisting of 301 with fatalities, 2773 with injuries, and 39 with property losses. Only the information including address, time, and injury or fatality is extracted to enhance the computational speed and reduce the complication degree of the methods below. Since the position information is stored as addresses in a traditional way, such as “Jiefang Road to Medicine Verified Agency”, the validation of each TC according to its address on a map should be the first step of data pre-processing. In this research, the Geocoding API in “Baidu Map” is used to match the address to a location on administrative map and roadway map. The roadway network data are extracted and abstracted from the Traffic Map of Wuhan (2007). In the TC map, each roadway is shown as a polyline with attribute of its road grade. In addition, TC abstracted as point is

placed on the polyline of its matching road. Here, the urban roadway network includes main roadways, secondary roadways and branch roads and excludes metro-ways, highways, railways and ferry-ways. According to statistics at the city level, 78.3% of TC occurred on main roadways, 8.6% on secondary roadways, 2.4% on branch roads and 10.6% on other types of roadways.

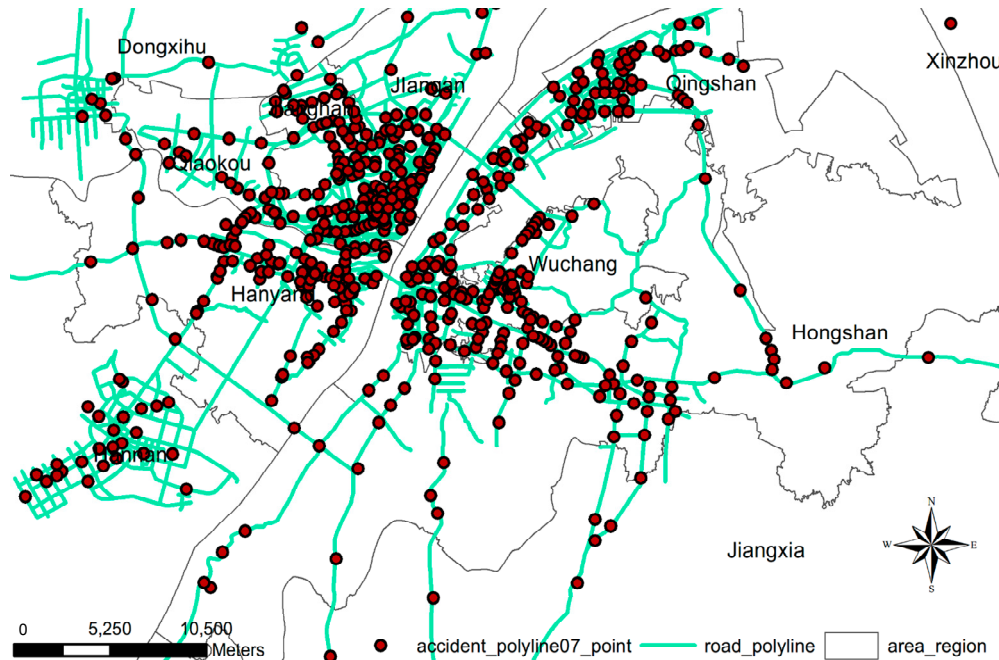


Figure 1. Locations of traffic crash in Wuhan in 2007.

2.2. Methods

2.2.1. Network Kernel Density Estimation

NKDE mainly discusses the first-order properties of spatial data in a nonparametric way to reveal cluster pattern of point events. It details not only the density value of each target but also the continuous surface map of risk targets in the study area. A symmetric and continuous surface is placed on each of the center points of the spatial units to calculate the density of the entire area considering the distances between the center point and the locations of observations within the surface. The estimators of NKDE is as followed.

$$\lambda(s) = \sum_{i=1}^n a_i \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \quad (1)$$

where $\lambda(s)$ is the density at location s ; r is the search width, which is always larger than 0 (only points within r are calculated for $\lambda(s)$); and d_{is} is the distance from the estimation point to the observation point marked as i ; k , known as the kernel function, is the function of the ratio between d_{is} and r to measure the distance decay effect; and a_i is the number of TC at location s .

The choice of the two parameters, r and k , is extremely critical. When r increases, the surface of the density becomes smoother, ignoring some details of the density. When r decreases, the surface of the density becomes uneven, enhancing the cost of the calculation. Besides, it demonstrates that the effect

of the choice of kernel function is less than the effect of the choice of the search width [47,48]. There are many kernel density functions, such as the Gaussian, Quartic, Conic, Negative exponential and Epanechnikov [51]. In this paper, the Gaussian function is used as shown by the following:

$$k\left(\frac{d_{is}}{r}\right) = \begin{cases} \sqrt{2\pi} \exp\left(-\frac{d_{is}^2}{2r^2}\right) & 0 < d_{is} < r \\ 0 & d_{is} > r \end{cases} \quad (2)$$

The NKDE has the same analysis process as the KDE, with the main discrimination being the measurement of distance. The NKDE is based on the kernel function method of the planar KDE, with an extended measurement of distance between two points from the Euclidean distance to network distance. The core point of NKDE lies in the division of the road network into a fixed length (called a lixel) and the measurement of r and using the shortest distance of the network. The procedure can be organized in six steps:

- (1) Check the topology and connectivity of road network, and merge road segments with the same road name.
- (2) Divide each road with a fixed length (marked as l) into basic road segment units (called lixels).
- (3) Calculate the number of TC in each lixel, marked as i , $i = 1, 2, \dots, n$.
- (4) Use the kernel function to determine the density distribution of each TC to each lixel inside the search width.
- (5) Determine the density value of each lixel, which is the sum of the density contributions from each TC to the lixel within the searching width.
- (6) Output the density value of each lixel.

2.2.2. Local Indicator of Network-Constrained Clusters of Getis-Ord G_i^*

Getis-Ord G_i^* , one of the most used methods for evaluating clusters, can define the actual locations where hotspots are clustered together based on a formal assessment of statistical significance [52,53]. The Getis-Ord G_i^* evaluates the degree to which each feature is surrounded by features with a similarly high or low values within a specified geographical distance (neighborhood). It can measure the concentration ratio of high or low values for the study area. Large Z -values indicate that hotspots are clustered together, whereas low Z -values indicate that cold spots are clustered together. The Getis-Ord G_i^* local statistic is given as:

$$G_i^*(d) = \frac{\sum_{j=1}^n W_{ij}(d)x_j - \bar{X} \sum_{j=1}^n W_{ij}(d)}{S \sqrt{\frac{[n \sum_{j=1}^n W_{ij}^2(d) - (\sum_{j=1}^n W_{ij}(d))^2]}{n-1}}} \quad (3)$$

where $G_i^*(d)$ is computed for a feature i within a distance d , standardized as a z -score. x_j is the attribute value for feature j within distance d of a given feature i , $W_{ij}(d)$ is a symmetric one/zero spatial weight matrix from a threshold d for the distance between features i and j , the threshold d defines the distance within which all locations are considered as neighbors (indicated by 1 in the W matrix),

and beyond which all locations are not neighbors (indicated by 0 in the W matrix) [54]. n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (4)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{X}^2} \quad (5)$$

In this case, the relationship between the lixel unit i and the other lixel units in its neighborhood distance is shown as four types: high-high, low-low, high-low and low-high. High-high and low-low mean that there is a positive correlation between the unit and its neighbors. High-low and low-high mean that there is a negative correlation between the unit and its neighbors. If the density of the lixel unit and its neighbors are all high under a level of statistical significance that is, a high-high correlation, these high-high units (H-H segments) in this paper are merged as RRSs.

GLINCS, based on Getis-Ord G_i^* , measure the autocorrelation and concentration essentially with the same equation of Equation (4). Nevertheless, the definition of weight matrix W is quite different. In GLINCS, the network is split into smaller segments to better reflect characteristic of the scale of research data and research area. W_{ij} in Equation (4) is defined as connectivity between segment i and j and can designate whether segment i and segment j share a common node [55]. The value indicates the autocorrelation and concentration value around the interest observed link i ($i = 1, \dots, n$).

The Getis-Ord G_i^* is added as a supplement to the result of the NKDE for two main reasons: one is that the smoothing density surface can decrease the noise and bias of TC point location. Traffic crash, regarded as point events with precise point location on a map, may occur and affect a certain length along a road. The use of NKDE can largely spread the counting number to the nearby road and do as a smoothing process of GLINCS. Besides, the density values can be interpreted as a risk index as input of GLINCS compared with counting index. The other is that the process of the LISA can help formally evaluate the significance of the extensiveness of locations with high-density values. Although both local Moran's I and Getis-Ord G_i^* are able to reveal hot spots, Getis-Ord G_i^* is able to differentiate autocorrelation due to the spatial associations of high-high values or low-low values. Moreover, as Xie pointed, different methods of local statistics measurements should be tested when the KDE and LISA are integrated for cluster detection [47].

3. Results and Discussion

Utilizing the proposed methods, a case study was conducted with real world TC data from the city of Wuhan. Firstly, density value of TC, obtained with method of KDE and NKDE using the software of GeoDaNet, was compared to indicate the TC pattern at the city level. Secondly, the GLINCS was calculated based on the density values of NKDE and counting values with 99 iterations of Monte Carlo simulations. Finally, the cluster and riskier segments were found by using the Z-value to test the significance.

3.1. Results of the KDE and NKDE for Traffic Crash Events

To verify the most appropriate methods to detect cluster pattern, four contrast experiments were conducted, based on the methods of KDE and NKDE. As seen in Table 1, the parameters and density value were shown and compared. To contrast the clusters discrimination from scale effect of NKDE, same lixel length and different search width were set in experiments 1 and 2. To discuss the cluster discrimination from resolution effect of NKDE, different lixel length and same search width were set in experiments 2 and 3. To compare the characteristic of clusters pattern from KDE and NKDE, same search width was set in experiment 3 and experiment 4.

Table 1. Parameters and density values in experiments of KDE(kernel density estimate) and NKDE (network-constrained kernel density estimate).

Experiments		1	2	3	4
Parameters	Methods	NKDE	NKDE	NKDE	KDE
	Length of lixel (m)	10	10	40	--
	Search width (m)	40	200	200	200
	Kernel function	Gaussian	Gaussian	Gaussian	Gaussian
Value	Count	72,573	72,573	27,297	--
	Min	0	0	0	0
	Max	0.047234	0.067182	0.066892	0.001314
	Sum	20.544062	94.479292	32.121501	1.442635
	Mean	0.000283	0.001302	0.001177	0.000004
	Standard Deviation	0.001607	0.003499	0.00328	0.000027

The search width was set as 40 m and 200 m due to the minimum width of urban main road. According to the specifications of urban road planning, width of main road is defined as 45–55 m, secondary road is 40–50 m and branch road is 15–30 m. Thus, the minimum value of r should be no less than 40 m. Besides, it was suggested that the applicable value of r was 100–300 m, which was widely used in urban planning at the scale of neighborhood, block and street [56]. We used 200 m due to more obvious hot segments compared with other search widths from 100 m, 500 m and 1000 m. The lixel is like a resolution in a raster and the smaller lixel is, the higher the precision. Hence, the length of lixel in NKDE was set at 10 m and 40 m, which were also identified by Xie [47]. In addition, it may be shown that a Gaussian kernel is generally robust and a usual choice for KDE.

As shown in Table 1, higher cluster result from higher standard deviation and density sum when search width was wider and length of lixel was longer. When comparing experiments 1 and 2, sum of density value for every segment increased with search width from 40 m to 200 m. When comparing experiments 2 and 3, number of lixel segments increased approximately 1.6 times with length of lixel from 40 m to 10 m, which means that there were more segments less than the length of lixel when length was set at 40 m. Moreover, mean value of density kept stable between experiments 2 and 3, but sum value and standard deviation of experiment 2 were more than that of experiment 3. However, the density statistic value in experiment 4 decreased greatly with lower sum value and standard deviation.

A group of thematic maps was represented to reflect the density values in road network of NKDE (see Figure 2a–c) and, a density surface was displayed in the raster map to visualize the result of KDE

(see Figure 2d). In Figure 2a, the density values are displayed as a scattered distribution pattern consistent with the original location map of TC. Moreover, the road segments with higher density values were located in the main road in the center city. Together, the patterns in Figure 2b,c presented quite similar distribution characteristics with a distinct high-low cluster pattern in which the districts with higher density values were Jiang'an and Jiangnan and the roads with higher density values were Hanyang Avenue, Zhongshan Avenue, Zhongbei Road, Jiefang Avenue and others. Nonetheless, map in Figure 2b has elucidated more cluster details due to its smaller lixel length than map in Figure 2c. Due to the resolution effect of smaller lixel, more information of distribution variation is easily captured in the same length of road in experiment 2 than in experiment 3. Cluster pattern in Figure 2d portrayed approximate hot area in planar space not exactly in the road.

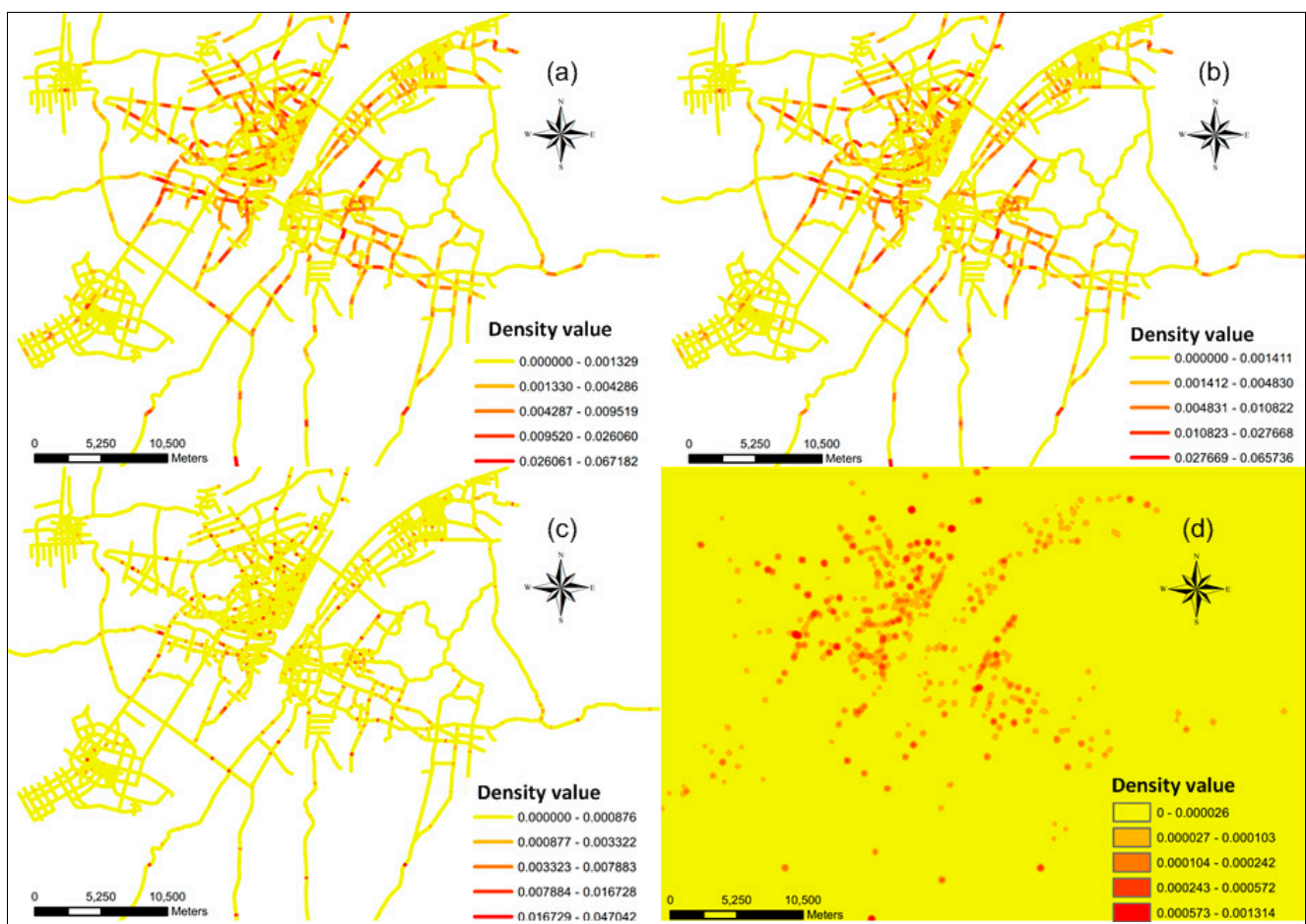


Figure 2. Density values in experiments of NKDE(network-constrained kernel density estimate) and KDE(kernel density estimate). (a) Density map of experiment 1; (b) Density map of experiment 2; (c) Density map of experiment 3; (d) Density map of experiment 4.

3.2. The GLINCS of NKDE

After the density value is calculated for each road segment using NKDE, it is then used as an attribute for computing the GLINCS to explain the method of NKDE in a quantitative way. In this part, five experiments were conducted. As seen in Table 2, input attributes in the first two experiments were the smoothing density value while in the last three experiments were crash events with a counting

approach to allocate the crash points just in its near network edges. To contrast strongly NKDE-GLINCS and GLINCS, experiment 7 and 8 were implemented as a comparative study of experiment 5 and 6. Experiment 9 was performed for an estimation of whether TC spatial cluster existed in a city level. To detect H-H road segments in different levels of statistical significance, Monte Carlo simulation was repeated 99 times via a conditional permutation process. Considering the intensive computation load in our experiments, 99 was set for a balance of random stableness and computational efficiency; the null hypothesis for Getis-Ord G_i^* is complete spatial randomness. In this hypothesis, density value represents one of many possible spatial arrangements, which are a conditional spatial permutation process by shuffling the density values among the network.

Table 2. Parameters and H-H(high-high) segment numbers in experiments of GLINCS.

Experiments		5	6	7	8	9
Parameters	Input attribute	density value of experiments 2	density value of experiments 3	crash number	crash number	crash number
	Lixel length	10m	40m	10m	40m	--
	Simulation times	99	99	99	99	99
	Simulation type	Conditional Permutation	Conditional Permutation	Conditional Permutation	Conditional Permutation	Conditional Permutation
<i>p</i> -value	0.01	1126	431	0	0	14
	0.05	1825	693	0	0	34
	0.1	2360	885	0	0	49

The results revealed that when the significance level decreases from 0.01 to 0.1, the number of H-H segments of both experiment 5 and experiment 6 increased from a relatively high speed to a relatively low speed. In addition, the number of H-H road segments in experiment 5 maintained a relatively constant time (2.6–2.8) compared with the number in experiment 6. In experiment 9, it had been proved that there is spatial cluster existing in traffic crash. Besides, in this experiment, the road network was not clip with fixed lixels and the result of Getis-Ord G_i^* had successfully revealed a global cluster pattern compared with experiments 7 and 8. Results in experiments 7 and 8 revealed that the GLINCS method failed to identify the H-H road segments. Nevertheless, results in experiments 5 and 6 demonstrated that the NKDE-GLINCS method using kernel density as input in experiments 5 and 6 could disclose and identify H-H segments successfully. The distribution of H-H segments under significance level of 0.1 in experiment 6 was shown in Figure 3a.

3.3. The Detection of Riskier Road Segments

Although experiments 5 and 6 can detect H-H segments, experiment 5 can identify the H-H segments without considering more details to keep segments in a coherent and valid length. RRSs that combines H-H road segments with the same road name under the significance level of 0.1 is detected. Here, RRSs in experiment 5 were ranked according to the GLINCS value and the density value from high to low. The first 20 RRSs were shown in the table below. Moreover, RRSs with TC can be seen in the three-dimensional visualization with ESRI ArcScene in Figure 3b. The height of red column represents mean density value of TC b in road segments. When the Z-score is greater than 1.65, which

means the road segments are significantly risky under the significance level of 0.1, the higher column denotes RRSs and the wider column denotes intense cluster of RRSs.

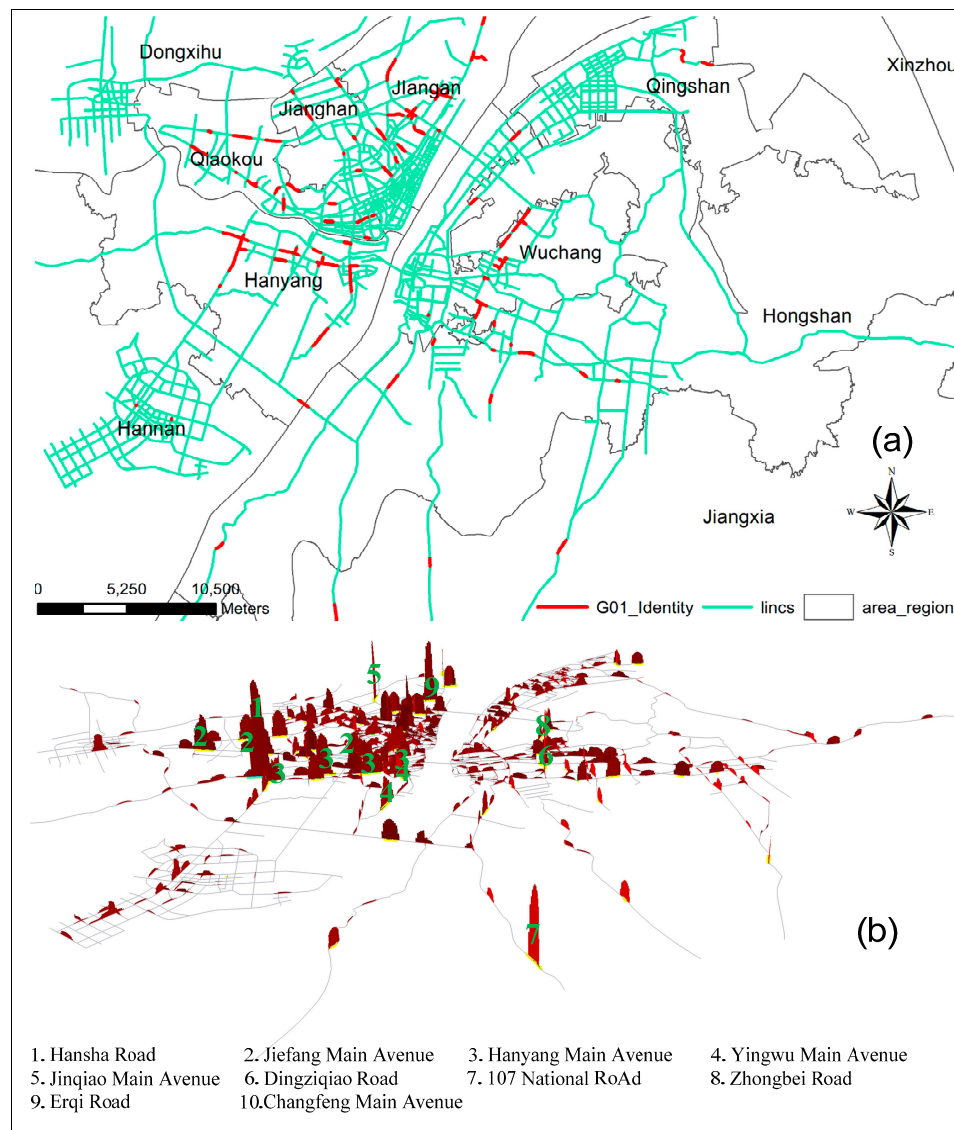


Figure 3. H-H road segments in Wuhan (0.1 level). (a) H-H segments (0.1) in experiment 6; (b) 3D map of riskier road segments in urban districts.

In Table 3 and Figure 3, the top risky road in Wuhan was Hansha Road with a considerably high-density value and the largest G^* , which was consistent with government announcements and news reports. One reason is the high speed of motor vehicles near the entrance of the Hancui Highway. Furthermore, these sections of high-risk road have been improved by the local department of transport these recent years. Moreover, the Jiefang Main Avenue was shown to be a secondary risky road with the maximum number of lixels and a steady but lower density value relative to the density values of other risk roads such as Dingziqiao Road. Till now, Jiefang Main Avenue has been verified as having been always cautioned as a hazardous road for TC due to the high traffic flow and especially the traffic jams during the morning and evening rush hours. This situation also exists on ZhongBei Road. In addition, risky roads are more likely to be located in the district of Jiangnan, considering the complexity of the road network.

Table 3. Top 20 RRSs (riskier road segments) under the significance level of 0.1.

Order	Road Name	G*	Density Value (mean)	Z (mean)	Lixels (number)	District
1	Hansha Road	0.0248	0.0409	10.6455	22	Hanyang
2	Jiefang Main Avenue	0.0217	0.0126	3.0175	63	Qiaokou
3	Hanyang Main Avenue	0.0192	0.0135	2.9929	56	Jiangnan
4	Yingwu Main Avenue	0.0181	0.0126	3.3375	48	Hanyang
5	Jinqiao Main Avenue	0.0171	0.0248	7.0348	23	Jiangan
6	Dingziqiao Road	0.0131	0.0288	7.1706	17	Hongshan
7	107 National Road	0.0149	0.0230	6.6381	21	Jiangxia
8	Zhongbei Road	0.0130	0.0110	2.6667	42	Wuchang
9	Erqi Road	0.0097	0.0228	4.6053	19	Jiangan
10	Changfeng Main Avenue	0.0117	0.0137	3.3258	31	Qiaokou
11	Jianshe Main Avenue	0.0110	0.0110	2.5378	37	Qiaokou
12	Longyang Main Avenue	0.0109	0.0089	2.6000	36	Hanyang
13	Nianyihao Road	0.0086	0.0102	2.4267	30	Qingshan
14	Handti Street	0.0085	0.0133	3.4000	22	Jiangan
15	Xiongchu Main Avenue	0.0076	0.0103	2.2714	28	Hongshan
16	Zhongshan Main Avenue	0.0060	0.0129	2.3000	22	Qiaokou
17	Sanyanqiao Road	0.0060	0.0182	3.8071	14	Jiangan
18	Yongqing Road	0.0058	0.0137	2.7889	18	Jiangan
19	Jiangda Road	0.0063	0.0127	2.9000	19	Jiangan
20	Donghu Road	0.0049	0.0096	1.9476	21	Wuchang

4. Summary and Conclusions

The intense demand of roads to cater to the rapid economic development has made road traffic crashes causing traffic congestion one of the most pervasive forms of “bottle neck” in Wuhan, China. Secure and efficient transportation and mobility are key components and central to sustainable development of all-round urbanization. For an effective solution of the TC problem with a limited budget, risky road segments with higher probability of TCs should be given the priority to be maintained. The detection of spatial clusters of TC is the first vital step for the appropriate allocation of resources for safety improvement in a sustainable way. To identify riskier road segments in Wuhan, a two-step approach using NKDE, extended from KDE in planar space, and GLINCS, based on Getis-Ord G_i^* was illustrated. As presented, TC on the road network in Wuhan, with a total of 3113 crashes between motor vehicles, were selected for testing and verifying. It is confirmed that NKDE-GLINCS perform better than traditional GLINCS in identifying the cluster due to the preprocessing of NKDE smoothing. The case study also provides evidence of effectiveness and robustness of the NKDE-GLINCS method. In addition, the top 20 roads with high-high TC density at the significance level of 0.1 are listed and presented in 3-D visualization. The results of this case study should be useful in assisting transportation agencies and motorists to identify risky roads quickly and play an important role in the further analysis and prediction of TC.

Compared with conventional TC analysis methods, NKDE can be used not only for analyzing the properties of point events and measuring the variation in the mean values of spatial processes but also

for a preprocess for a smoothing density value from the origin data. The main advantage of the NKDE method is that the uncertainty about the process can be understood and implement easily. However, NKDE methods may always be used as visualization tools, due to the absence of significance testing. Herein, NKDE result was input as attribute for GLINCS to use the density indicator formally for evaluating the significant locations with high-density values.

Although the NKDE-GLINCS method for detecting the cluster pattern of TC has availability and advantage, there are still some fields to be improved. As discussed previously, in this study, only the spatial characteristics of TC was analyzed, whereas previous research has shown that the factors associated with TC may be diverse and complicated [15,17,47,57,58]. Thus, further study is needed to add other parameters to the kernel function and weight matrix, such as road density, road accessibility and land-use of the study area. Despite these improvements existing, TC distribution analysis using NKDE-GLINCS in other areas or cities in different scales are still expected. Besides, some other applications for geographical events constrained by or associated with networks are encouraged in the future.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Project No. 41371427/D0108 and 41271455/D0108), the Fundamental Research Funds for the Central Universities (Project NO. 2012205020215) and the Traffic Management Department of Wuhan.

Author Contributions

Ke Nie and Zhensheng Wang conceived and designed the study with the support of Qingyun Du and Fu Ren. Ke Nie and Zhensheng Wang analyzed the data and performed the experiments. All the co-authors drafted and revised the article together. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Wu, T.; Zhao, H.; Ou, X. Vehicle Ownership Analysis Based on GDP per Capita in China: 1963–2050. *Sustainability* **2014**, *6*, 4877–4899.
2. Evans, L. *Traffic Safety*; Science Serving Society: Bloomfield Hills, MI, USA, 2004.
3. Zhang, H.Y. Analyzing Traffic Accident Situation with Bayesian Network. Ph.D. Thesis, Jilin University, Changchun, Jilin, China, June 2013.
4. Zhang, G.; Yau, K.K.W.; Zhang, X. Analyzing fault and severity in pedestrian-motor vehicle accidents in China. *Accid. Anal. Prev.* **2014**, *73*, 141–150.
5. Chen, D. Research and Application on Evaluation Methods of Wuhan Public Transportation System. Master's Thesis, Wuhan University of Technology, Wuhan, Hubei, China, June 2010.

6. Flak, M.A. Reducing accident impacts under congestion. In *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities*; American Society of Civil Engineers: New York, NY, USA, 1997.
7. Prasannakumar, V.; Vijith, H.; Charutha, R.; Geetha, N. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia-Soc. Behav. Sci.* **2011**, *21*, 317–325.
8. Maher, M.J.; Mountain, L.J. The identification of accident blackspots: A comparison of current methods. *Accid. Anal. Prev.* **1988**, *20*, 143–151.
9. Schuurman, N.; Cinnamon, J.; Crooks, V.A.; Hameed, S.M. Pedestrian injury and the built environment: An environmental scan of hotspots. *BMC Public Health* **2009**, *9*, Article 233.
10. Moons, E.; Brijs, T.; Wets, G. Improving Moran's Index to Identify Hot Spots in Traffic Safety. In *Geocomputation and Urban Planning*; Murgante, B., Borruo, G., Lapucci, A., Eds.; Springer: Berlin, Germany, 2009; pp. 117–132.
11. Joly, M.-F.; Foggin, P.M.; Pless, I.B. Geographical and socio-ecological variations of traffic accidents among children. *Soc. Sci. Med.* **1991**, *33*, 765–769.
12. Holz-Rau, C.; Scheiner, J. Geographical Patterns in Road Safety: Literature Review and a Case Study from Germany. *EJTIR* **2013**, *13*, 99–122.
13. Li, L.; Zhu, L.; Sui, D.Z. A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *J. Transp. Geogr.* **2007**, *15*, 274–285.
14. Anderson, T.K. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accid. Anal. Prev.* **2009**, *41*, 359–364.
15. Eckley, D.C.; Curtin, K.M. Evaluating the spatiotemporal clustering of traffic incidents. *Comput. Environ. Urban Syst.* **2013**, *37*, 70–81.
16. Dai, D. Identifying clusters and risk factors of injuries in pedestrian-vehicle crashes in a GIS environment. *J. Transp. Geogr.* **2012**, *24*, 206–214.
17. Levine, N.; Kim, K.E.; Nitz, L.H. Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accid. Anal. Prev.* **1995**, *27*, 663–674.
18. Elvik, R. Evaluations of road accident blackspot treatment: A case of the iron law of evaluation studies? *Accid. Anal. Prev.* **1997**, *29*, 191–199.
19. Flahaut, B.; Mouchart, M.; Martin, E.S.; Thomas, I. The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accid. Anal. Prev.* **2003**, *35*, 991–1004.
20. Chini, F.; Farchi, S.; Ciaramella, I.; Antoniozzi, T.; Rossi, P.G.; Camilloni, L.; Valenti, M.; Borgia, P. Road traffic injuries in one local health unit in the Lazio region: Results of a surveillance system integrating police and health data. *Int. J. Health Geogr.* **2009**, *8*, 21.
21. Erdogan, S. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *J. Saf. Res.* **2009**, *40*, 341–351.
22. Loo, B.P.Y. The identification of hazardous road locations: A comparison of the blacksite and hot zone methodologies in Hong Kong. *Int. J. Sustain. Transp.* **2009**, *3*, 187–202.
23. Black, W.R.; Thomas, I. Accidents on Belgium's motorways: A network autocorrelation analysis. *J. Transp. Geogr.* **1998**, *6*, 23–31.
24. Getis, A. A history of the concept of spatial autocorrelation: A geographer's perspective. *Geogr. Anal.* **2008**, *40*, 297–309.

25. Yamada, I.; Thill, J.C. Local indicators of network-constrained clusters in spatial point patterns. *Geogr. Anal.* **2007**, *39*, 268–292.
26. Okabe, A.; Yamada, I. The K-function method on a network and its computational implementation. *Geogr. Anal.* **2001**, *33*, 271–290.
27. Chaikaew, N.; Tripathi, N.K.; Souris, M. Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand. *Int. J. Health Geogr.* **2009**, *8*, 36.
28. Carlos, H.A.; Shi, X.; Sargent, J.; Tanski, S.; Berke, E.M. Density estimation and adaptive bandwidths: A primer for public health practitioners. *Int. J. Health Geogr.* **2010**, *9*, 39.
29. Pulugurtha, S.S.; Krishnakumar, V.K.; Nambisan, S.S. New methods to identify and rank high pedestrian crash zones: An illustration. *Accid. Anal. Prev.* **2007**, *39*, 800–811.
30. Anderson, T. Comparison of spatial methods for measuring road accident ‘hotspots’: A case study of London. *J. Maps.* **2007**, *3*, 55–63.
31. Chainey, S.; Tompson, L.; Uhligm S. The utility of hotspot mapping for predicting spatial patterns of crime. *Secur. J.* **2008**, *21*, 4–28.
32. Yu, H.; Liu, P.; Chen, J.; Wang, H. Comparative analysis of the spatial analysis methods for hotspot identification. *Accid. Anal. Prev.* **2014**, *66*, 80–88.
33. Yamada, I.; Thill, J.C. Comparison of planar and network K-functions in traffic accident analysis. *J. Transp. Geogr.* **2004**, *12*, 149–158.
34. Loo, B.P.Y.; Yao, S. The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. *Comput. Environ. Urban Syst.* **2013**, *41*, 249–261.
35. Mohaymany, A.S.; Shahri, M.; Mirbagheri, B. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-spat. Inf. Sci.* **2013**, *16*, 113–119.
36. Young, J.; Park, P.Y. Hotzone identification with GIS-based post-network screening analysis. *J. Transp. Geogr.* **2014**, *34*, 106–120.
37. Okabe, A.; Satoh, T.; Sugihara, K. A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 7–32.
38. Erdogan, S.; Yilmaz, I.; Baybura, T.; Gullu, M. Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accid. Anal. Prev.* **2008**, *40*, 174–181.
39. Okabe, A.; Yomono, H.; Kitamura, M. Statistical analysis of the distribution of points on a network. *Geogr. Anal.* **1995**, *27*, 152–175.
40. Okabe, A.; Okunuki, K. A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. *Trans. GIS* **2001**, *5*, 209–220.
41. Okabe, A.; Okunuki, K.I.; Shiodem S. The SANET toolbox: New methods for network spatial analysis. *Trans. GIS* **2006**, *10*, 535–550.
42. Xie, Z.; Yan, J. Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst.* **2008**, *32*, 396–406.
43. Sugihara, K.; Okabe, A.; Satoh, T. Computational method for the point cluster analysis on networks. *GeoInformatica* **2011**, *15*, 167–189.
44. Steenberghen, T.; Aerts, K.; Thomas, I. Spatial clustering of events on a network. *J. Transp. Geogr.* **2010**, *18*, 411–418.

45. Li, Q.; Zhang, T.; Wang, H.; Zeng, Z. Dynamic accessibility mapping using floating car data: A network-constrained density estimation approach. *J. Transp. Geogr.* **2011**, *19*, 379–393.
46. Shiode, S.; Shiode, N. Network-based space-time search-window technique for hotspot detection of street-level crime incidents. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 866–882.
47. Xie, Z.; Yan, J. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach. *J. Transp. Geogr.* **2013**, *31*, 64–71.
48. Bíl, M.; Andrášik, R.; Janoška, Z. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accid. Anal. Prev.* **2013**, *55*, 265–273.
49. Plug, C.; Xia, J.C.; Caulfield, C. Spatial and temporal visualization techniques for crash analysis. *Accid. Anal. Prev.* **2011**, *43*, 1937–1946.
50. Nakaya, T.; Yano, K. Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Trans. GIS* **2010**, *14*, 223–239.
51. O’Sullivan, D.; Unwin, D.J. *Geographic Information Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
52. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **1992**, *24*, 189–206.
53. Ord, J.K.; Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **1995**, *27*, 286–306.
54. Peeters, A.; Zude, M.; Käthner, J.; Kanber, R.; Hetzroni, A.; Gebbers, R.; Ben-Gal, A. Getis–Ord’s hot-and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. *Comput. Electron. Agric.* **2015**, *111*, 140–150.
55. Yamada, I.; Thill, J.C. Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. *Ann. Assoc. Am. Geogr.* **2010**, *100*, 269–285.
56. Porta S.; Latora V.; Wang F.; Strano, E.; Cardillo, A.; Scellato, S.; Iacoviello, V.; Messori, R. Street centrality and densities of retail and services in Bologna, Italy. *Environ. Plan. B Plan. Des.* **2009**, *36*, 450–465.
57. Jones, A.P.; Haynes, R.; Kennedy, V.; Harvey, I.M.; Jewell, T.; Lea, D. Geographical variations in mortality and morbidity from road traffic accidents in England and Wales. *Health Place* **2008**, *14*, 519–535.
58. Wier, M.; Weintraub, J.; Humphreys, E.H.; Seto, E.; Bhatia, R. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accid. Anal. Prev.* **2009**, *41*, 137–145.