IATSS Research 47 (2023) 25–34

Contents lists available at ScienceDirect

IATSS Research

Research Article

Hotspot analysis of single-vehicle lane departure crashes in North Dakota

Ihsan Ullah Khan [a,*], Kimberly Vachal [a], Sajad Ebrahimi [c], Satpal Singh Wadhwa [b]

[a] North Dakota State University, United States of America
[b] Upper Great Plains Transportation Institute, North Dakota State University, United States of America
[c] Nicolais School of Business, Wagner College, Staten Island, New York, United States of America

ARTICLE INFO

Article history:
Received 4 May 2022
Received in revised form 20 September 2022
Accepted 18 December 2022
Available online 22 December 2022

Keywords:
Crashes
Hotspots
Moran's I
Space-Time cube
NetKDE

ABSTRACT

According to the North Dakota Department of Transportation (NDDOT), 90% of the state's fatal lane departure crashes between 2015 and 2019 occurred on rural roads. Of these, 77% were single-vehicle events. The objective here was to identify relatively high-risk areas on the rural road system. Spatial analysis techniques were explored as a beneficial tool in resource allocations aimed at single-vehicle crash prevention. Hotspot identification techniques, including Global Moran's I, local Moran's I, network kernel density estimation (NetKDE), and emerging hotspot analysis were employed. While the Global Moran's I index indicated the existence of crash clustering, the local Moran's I statistic revealed hot and cold spots in the state. The NetKDE approach was used to quantify crash clusters and prioritize locations. Results from NetKDE defined boundaries for each cluster in terms of density values embedded in the roadway. Emerging hotspot analysis evaluated the hot and cold spots with respect to time. This study will provide valuable insight and help decision makers to make more informed decisions with respect to education, enforcement and infrastructure strategies aimed at preventing single-vehicle lane departure crashes. Although limited to a narrow crash type in one state, this approach can inform other jurisdictions seeking to empirically visualize hotspots and more effectively deploy traffic safety strategies.

1. Introduction

Lane departure crashes are a primary concern in the United States. The Federal Highway Administration [1], defines a lane departure crash a one which occurs after a motor vehicle crosses a center line or an edge line, or otherwise leaves a roadway. From the driver's perspective, four components factor into these complex crashes: human error, vehicle component failure, roadway condition, and collision avoidance [2]. About 51% of all U.S. traffic fatalities were attributed to lane departure crashes from 2016 to 2018 [1]. These crashes tend to be grievous, mostly occurring at relatively high speeds.

The rural lane departure crash is the most common type among fatal traffic crashes in North Dakota. Nearly 90% of fatal lane departure crashes between 2015 and 2019 occurred on rural roads [3]. In addition, 77% of these crashes involved a single-vehicle. The two most common harmful events in single-vehicle fatal lane departure crashes were rollover/overturning (77%) and collision with fixed objects (11%). Fig. 1 shows single-vehicle lane departure crashes in North Dakota remain prevalent. The extensive rural road network and episodic crash event nature are an impediment to substantial improvement for rural roads. Moreover, these roads are often managed by local agencies that have limited resources and inadequate technical skills for safety analysis. This lack of analysis presents a challenge in determining where to apply lane departure countermeasures to effectively reduce fatalities and injuries.

Thus, tools to detect high-risk locations and ultimately deploy appropriate strategies is essential in moving traffic crash trends downward. Visual identification tools are evolving as a crucial technique. Hotspots, or crash-prone locations, are areas having clusters of relatively high-crash concentrations. Detecting these hotspots provides an opportunity to be more granular in studying and addressing crash causes. Refined knowledge is used to allocate limited resources more efficiently by prioritizing high-risk locations. The objective here was to unveil single-vehicle lane departure hotspot areas that are at a higher-risk for crashes across the state's rural road system.

2. Literature review

It is important to examine crash location and time in dispersal. One can then prioritize locations based on attribute clusters. Inherent geographical qualities increase the chance of a crash occurrence at certain locations. In other words, traffic crashes do not take place randomly. The tendency of crashes to be concentrated at given locations can be

* Corresponding author.
E-mail addresses: ihsan.khan@ndsu.edu (I.U. Khan), kimberly.vachal@ndsu.edu (K. Vachal), sajad.ebrahimi@wagner.edu (S. Ebrahimi), satpalsingh.wadhwa@ndsu.edu (S.S. Wadhwa).

https://doi.org/10.1016/j.iatssr.2022.12.003

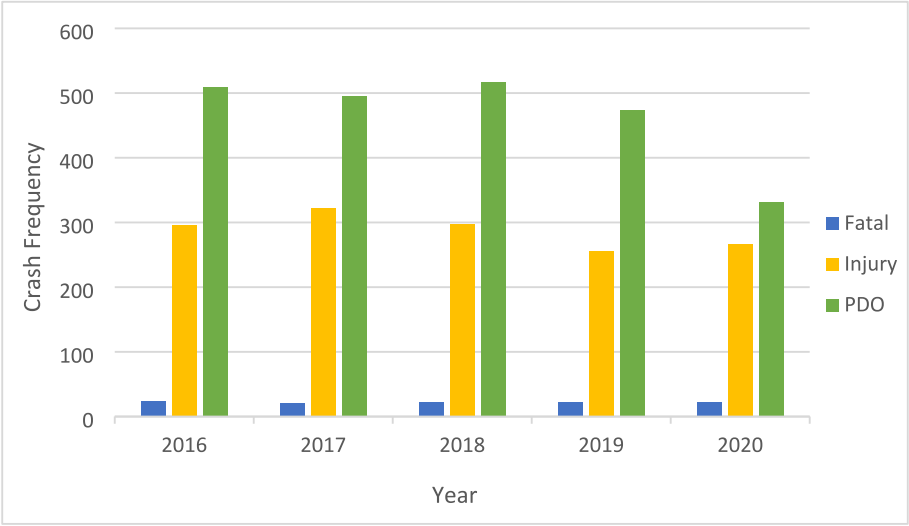

Fig. 1. Crash statistics of single-vehicle lane departure crashes in ND from 2016 to 2020.

explained by factors such as land use, socio-economic parameters, and geometric design. Hotspots or black spots refer to high-concentration crash location clusters. The systematic strategy of crash hotspots is commonly used to identify crash-prone locations, establish ranking, reveal causation, and determine countermeasure on identified locations [4,5]. The primary purpose of hotspot analysis is to give decisionmakers network insight into the select and deployment of safety measure(s) for crash prevention [6–9].

One way to conduct hotspot analysis is to identify locations by simply observing crash location maps. However, the results are subjective, providing weak support for decisions. Two empirical techniques have emerged in robust hotspot analysis for crash prevention. The first uses traditional statistical methods such as regression models [10,11], empirical Bayesian [12,13], and full Bayesian [14,15] for model based analysis. The second technique is based on geostatistical analysis such as spatial autocorrelation method [16–19] or density estimation method [20–23].

Generally, hotspot analysis based on a geostatistical approach is preferred. This approach requires less complex data and uses straightforward computation, unlike traditional statistical methods [24]. In addition, the geostatistical method provides a visual representation of results, incorporating spatial factors to detect location-specific influences on crash occurrence [25]. The Density estimation method is most commonly used in crash pattern detection, but it analyzes the crash location with no consideration for attributes. On the other hand, advanced approaches, like the spatial autocorrelation method, account for location as well as crash attributes. Hence, the underlying concept of spatial autocorrelation is associated with interdependence of a particular attribute over space.

Two categories of spatial autocorrelation are 1) global spatial analysis and 2) local spatial analysis. Global spatial analysis measures and tests the overall spatial phenomenon and identifies if the feature pattern is clustered, dispersed or randomly distributed in space with respect to their attribute values [16,18]. Local spatial analysis measures the level of spatial association at the local scale [17,19,26]. It is preferred over global spatial analysis for crash hotspot identification because local spatial patterns are undetectable when generalized for a large area [27]. Hence, local spatial analysis is used mostly for analyzing crash hotspots to capture regional heterogeneity across the study area.

The two most common approaches in local spatial analysis are local Moran's I [19,26,28] and Getis-Ord Gi* [17,29,30]. The local Moran's I has the ability to identify local clusters but cannot differentiate the high-valued from low-valued clusters. This limitation was overcome by the Gi* spatial statistic approach that distinguishes low- and high-valued clusters [16,17,31]. Another approach for hotspot analysis is the space-time cube (STC) analysis [32,33]. STC is a 3D visual representation of a geographical phenomenon such that the horizontal plane of the cube (x and y) represents space and the vertical axis represents time [33]. This approach can show spatial patterns, spatial relationships, and changes over time. Various studies that have used 3D visualizations for analysis of a wide array of spatial datasets such as atmospheric pollution, crime, crashes, and diseases [34–40].

According to Kveladze et al. [41], understanding the underlying meaning of complex spatiotemporal data is difficult for users. However, spatial aggregation techniques, especially point-based data, can provide a solution. The integration of hotspots with STC is referred to as emerging hotspot analysis. Both techniques are integral in deploying crash cluster decision tools. Ample literature on the use of all these methods for spatial analysis with crash detection is available but few studies have considered crash severity when identifying crash hotspots. The objective here was to apply hotspot techniques to distinguish crash clusters across the road network based on injury severity. Stakeholders can use results in data-driven decisions to prioritize locations and strategies for preventing lane departure crashes.

3. Scope and data

A total of 19,162 records for motor vehicle crashes that occurred on rural roads from 2016 to 2020 were collected from the state. At least three years of crash data is needed to effectively conduct spatial analysis [13]. Information regarding spatial location, crash, driver, vehicle, and geometrical characteristics were collected. Records were parsed to capture single-vehicle lane departure crashes on rural roads. Subsequently, rural road functional class as minor arterial, major collector, minor collector, and local roads were retained. This data deduction process resulted in a total of 3878 single-vehicle lane departure crashes that occurred on local roads. Table 1 shows the crash characteristics of the data.

Four categories were differentiated, property damage only, minor injury, serious injury, and fatal injury, for a spectrum of less- to more-severe crash outcomes, respectively. The crash frequency with serious and fatal injury categories was limited, so these categories were merged. The transformed data contained three severity levels: 2418 property damage crashes, 1075 minor injury crashes, and 385 serious or fatal injury crashes. A roadway network layer was created based on the geographic coordinate system and was projected in metric linear units. All the crashes with latitude and longitude were projected on the roadway network. SAS software 9.4 was used to clean and prepare

Table 1
Crash characteristics of the data.

Variable name		Crash severity		
		PDO	Minor injury	Serious and fatal injury
Road functional classification	Minor arterial	374	220	18
	Major collector	679	533	48
	Local	1274	686	46
Seatbelt use	Yes	1539	773	16
	No	129	423	87
	Unknown	659	243	9
Driver gender	Male	1587	1079	89
	Female	555	354	23
	Unknown	185	6	0
Speeding/too fast for conditions	Yes	846	644	38
	No	1481	795	74
Driver average age		34	35	42
Total crashes			3878	

the data. ArcMap 10.4, SANET, and ArcScene 10.4 were used for analysis and visualization.

4. Methodology

Spatial statistical methods were applied to analyze single-vehicle lane departure crashes to reveal black zone areas. Both global and local spatial analysis were used to determine significant clusters. For each crash location and the surrounding crashes, the spatial autocorrelation was performed using the linear spatial weight matrix. Approaches are described in the following sub-sections.

4.1. Global spatial autocorrelation

The spatial autocorrelation known as Global Moran's I detects general spatial patterns by considering both the location and feature values. Accounting for a set of features and related attributes, this approach examines if the pattern is clustered, random, or dispersed. Mathematically, Global Moran's I can be expressed as follows:

$$\mathrm{I} = \frac{n\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_{ij}(x_i-\overline{x})(x_j-\overline{x})}{S_0\sum_{i=1}^{n}(x_i-\overline{x})^2} \;\forall_i = 1,\ldots,n \wedge \forall_j = 1,\ldots,n \quad (1)$$

where x_i is the value of feature on location i, $\overline{x}$ is the feature mean, n is equal to the total number of locations, w_{ij} is the spatial weight representing connectivity relationships between feature i and j, and S_o is the sum of all spatial weights.

$$S_o = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \quad (2)$$

The results from the Global Moran's I statistics are interpreted with respect to its null hypothesis which states that among the features, an attribute is randomly distributed in the study area. The statistical significance of this test is calculated from the Z score assuming normal distribution with mean and variance equal to zero and one respectively. A positive Z score implies that the feature is surrounded by similar values. A negative Z score shows that the neighboring features have different values [16,17].

4.2. Local spatial autocorrelation

The local Moran's I index is used to identify local clusters and local spatial outliers of crashes in the study area. It can be written as follows:

$$I_i = \frac{z_i-\overline{z}}{\sigma^2}\sum_{j=1,j\neq i}^{n}\left[w_{ij}\left(z_j-\overline{z}\right)\right] \quad (3)$$

where I_i is the local Moran's I coefficient, z_i is the value of feature on location i, $\overline{z}$ is the feature mean, z_j is the value at all other locations such that $j \neq i$, and σ^2 is the variance of z. A high positive value of local Moran's I index indicates that the location under study is surrounded by features with similar values i.e. presence of spatial cluster. On the other hand, a high negative value of local Moran's Index shows the presence of spatial outliers i.e. the site under study is surrounded by features with different values. Local Moran's index generates four types of results: 1) high-high clusters – high values surrounded by a high-value neighbored, 2) low-low clusters - low values surrounded by a low-value neighbored, 3) high-low outlier – a high value surrounded by low-value neighbored, and 4) low-high outlier – a low value surrounded by high-value neighbored. Note that high-high clusters are considered to be crash hotspots while high-low outliers are considered to be isolated individual hotspots.

Another approach for local spatial autocorrelation is based on Getis-Ord Gi* statistics. This method distinguishes between the areas of high and low value concentration on the local scale. It determines high or low corresponding values by comparing each feature with its neighboring features. For a location to be considered as a hotspot, both the feature itself as well as the features surrounding it should have high attribute values. In ArcMap, the hotspot analysis tool uses this statistic to determine a significant red/blue spot based on neighbors' feature values. The following equation shows the general form of G_i^*,

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij}x_j - \left(\frac{\sum_{j=1}^{n} x_j}{n}\right)\times\sum_{j=1}^{n} w_{ij}}{\sqrt{\frac{\sum_{j=1}^{n} x_j^2 - \sum_{j=1}^{n} x_j}{n}}\times\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2}{n-1}}} \quad (4)$$

where G_i^* is statistic which describes the spatial dependency of feature i and x_j is an attribute value of j feature in the neighborhood. The G_i^* statistic represents the value of a target feature in the form of z-score. Features with a significant positive z-score indicate that neighboring values are similar while those with negative z-score implies that nearby values are dissimilar. The size of z-score in either direction of association shows the magnitude of clustering. In other words, a higher positive or a smaller negative z-score represents a cluster concentration for each corresponding feature. A suitable calculation of the spatial relationship between features is integral for local spatial autocorrelation analysis [42]. Some different approaches available for spatial relationship include fixed distance band, inverse distance, K nearest neighbors, and space time window. Previous research shows that the fixed distance band approach is preferred for analyzing point datasets [43]. This study used fixed distance band because we are analyzing crashes.

The degree of spatial autocorrelation changes according to distance bands because of unique groups of neighbors. Knowing the precise nature of a spatial crash patterns at a specific location is a complex phenomenon. Alternatively, previous research suggests using the optimum distance value that establishes the maximized spatial autocorrelation [43,44]. Clustering patterns of crashes will be more evident when the distance used in calculation provides an enlarged degree of spatial autocorrelation across the area of interest. For this purpose, a tool known as incremental spatial autocorrelation in ArcMap measures the variation in spatial autocorrelation as the distance bandwidths change. For a set of increasing distances, it produces several z-scores related to corresponding Moran's I. At a critical distance bandwidth, a peaked z-score shows spatial autocorrelation which forms the most prominent clusters [43,44]. But before using incremental spatial autocorrelation, the question of what starting distance should be used arises? A good starting distance would be the distance at which any given point has at least one neighbor. For this, a tool in ArcMap

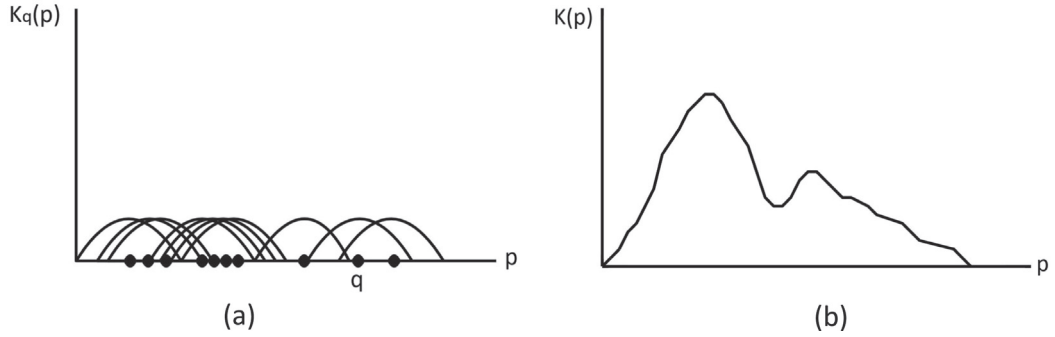

Fig. 2. Kernel density estimation procedure on a line segment: (a) mole hills or kernels (b) a big mountain or kernel density estimate (source: [45]).

known as calculate distance band from neighbor count was used. Crashes were used as input features with neighbors equal to one in order to calculate the average distance.

4.3. Network kernel density estimation

There is another non-parametric approach for hotspot analysis known as kernel density estimation (KDE). This method provides a high quality visual representation of density estimates from observed data. It is known as KDE because a circular area (the kernel) of defined bandwidth is created around each point at which the indicator is observed. Fig. 2 shows 2D visualization of kernel density estimation method. For each point, individual kernel or molehill is created in the form of smooth and continuous density surface. Then, for a given cell, all overlapping density surfaces are added together to obtain a big mountain or kernel density estimate. In Fig. 2, q represents individual points (crashes), $K_q(p)$ is a network kernel density function at q, and $K(p)$ is network kernel density estimator.

Conceptually, a road crash does not occur at an exact point location but occupies and affects some length along the roadway. The use of points ignores the impact on the nearby roadway by simplifying the dimension. The uncertainty created by the crash logging process is not considered. On the contrary, characterizing a crash as a spatial event that possesses a spread of risk is more appropriate. The spread of risk is defined as the neighboring area around a crash location where the likelihood for a crash to occur differs depending on how close areas are to the crash [7]. Previous research shows that the surface based density estimation may not be suitable for characterizing crash locations occurring on the roadway network [45–47]. Network kernel density estimation (NetKDE) is an improved approach to avoid under- or over-estimation issues of the planar KDE technique for the road crashes data. The NetKDE can be expressed as follows:

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \tag{5}$$

where $\lambda(s)$ is the NetKDE at location s, r is the search radius of KDE, and $k\left(\frac{d_{is}}{r}\right)$ is weight of point i at location d_{is} to location s. The kernel function, k provides information on the distance decay effect between two points as they get farther from each other. As the distance of a point increases from location s, the effect of the point will also decrease for overall density estimation [48]. There are various types of kernel functions but their selection has little effect on both local and global density formations [42,48,49]. However, the search radius r has major impact on $\lambda(s)$ estimation because the points outside the search radius on the given road network are not considered in the estimation. Although the process of deciding the bandwidth is somewhat subjective, past studies suggest that 50 to 300 m and 1000 m bandwidth should be used for urban and rural areas, respectively [46,47]. This study used a bandwidth of 300 m for density estimation. We used a software called SANET Standalone (Ver. 4.1) for analyzing crashes that occur on the road network.

4.4. Space time cube analysis

Hotspot analysis and kernel density analysis show patterns of lane departure crash density but lack the temporal characteristics of these crashes. Hence, space-time cube (STC) analysis was used for spatiotemporal analysis of lane departure crashes. STC analysis refers to a 3D geovisualization approach that maps spatiotemporal data in the form of a cube. The x and y dimensions of a 3D space-time cube represents space while t dimension represents time (Fig. 3). In this study, the value of grid cells was set at 8000 m × 8000 m with a time duration of 6 months. This generated a total of 31,680 grid cells for the study area.

This study used a time approach to analyze the trend of lane departure crashes. The input feature points (lane departure crashes, in our case) were aggregated into NetCDF (Network Common Data Form) data structures in the form of space-time bins. The points were counted within each bin and specific attributes were aggregated. The calculated value of each bin represents the crash frequency at specific location during a given time interval. The STC were then analyzed using emerging hotspot analysis (EHSA). The goal of EHSA is to evaluate changes in hot and cold spots with respect to time. The NetCDF data structure was analyzed using a tool known as Create Space Time Cube by Aggregating Points. It then calculates the Getis-Ord Gi* statistic for each bin and time period. After the space-time hotspot analysis completion, each bin is provided with a specific z-score, p-value, and hotspot bin classification. In the next step, the Mann-Kendall trend test [51,52] is used to evaluate all the hot and cold spot trends. To help better understand the change in locations over time, 17 unique categories (Table 2) are created comparing the Getis-Ord Gi* and Mann-Kendall trend test.

5. Results and discussion

This section provides findings for the various hotspot techniques used in analyzing the single-vehicle lane departure crashes. To measure the tendency of crash events to cluster and estimate the overall degree of spatial autocorrelation, we used the Global Moran's I. As shown in Fig. 4, the value of the Moran's I index was positive, which shows significant clustering of crashes in the study area. The z-score for crashes was 6.96 and the p-value was 0.00. The assumption of random distribution for crashes was rejected because the z-score of 6.96 was significantly greater than the threshold for rejection. It means there is less than 1% likelihood that this clustered pattern could be the result of random chance. The high value of absolute z-score indicates significant spatial autocorrelation of lane departure crashes in the study area.

Local spatial autocorrelation techniques were applied to assess clustering at the local level. First of all, average distance at which any given point that has at least one neighbor was calculated to be 2180 m (=

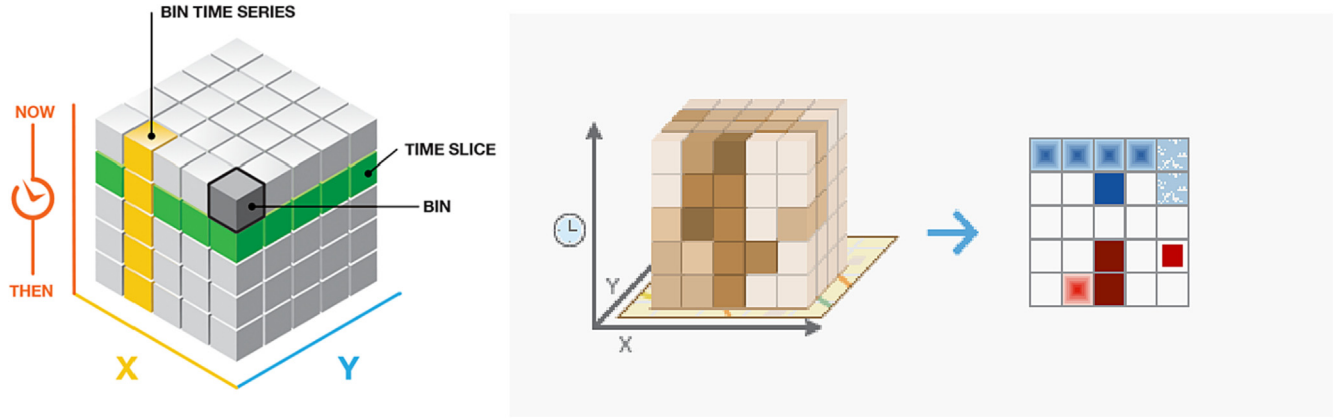

Fig. 3. (left) Structure of STC in 3D (right) generated bins in 2D by running emerging hotspot analysis. Source: [50].

2.18 km or 1.35 miles). This distance was set as a starting distance in the incremental spatial autocorrelation technique. Several trials to minimize the distance range of peaked z-scores were conducted. We extended the starting distance by increments of 500 to 2000 m and observed a decreasing trend of z-score. The concentration of spatial clustering is reflected by z-scores. The intensity of spatial clustering of crashes for the z-scores at each distance for a set number of segments is shown in Table 3. The first peak was at 3500 m with z-score of 4.27. The table also shows that, after reaching the peak, the intensity of spatial clustering decreases with the increase in distance. Where multiple peaks are present, the use of first peak is recommended because it best describes the spatial variation of the analysis [43]. In our case, the first and maximum peak was at 3500 m which would reflect more variations at the local level for crash hotspots.

The local Moran's I approach was applied to reveal the spatial pattern of local differences. This technique evaluated the significance of the spatial difference in crashes by comparing each cell to its surrounding cells. Fig. 5 presents the results obtained for local spatial clustering through local Moran's I index. The figure shows high and low clusters areas together with high and low spatial outliers at the 95% confidence level. The red dots represent locations with high clusters, which means that serious injury crashes were surrounded by crashes with serious injuries. Similarly, locations with low clusters show the areas where PDO crashes were mostly surrounded by similar crashes i.e. PDO crashes. The high spatial outliers show locations where the serious injury crashes were surrounded by crashes with dissimilar severity, i.e. PDO and injury crashes and vice versa for low spatial outliers. The majority of the crash clusters with higher values, i.e. serious injury crashes, were located in Barnes, McKenzie and Williams counties.

The local spatial autocorrelation analysis based on Getis-Ord Gi* statistics was also conducted to reveal crash severity patterns. The z-score and *p*-value for each crash point determine a statistically significant red, blue, or gray spots while accounting for severity values of neighboring crashes. Red spots indicate statistically significant serious injury crashes surrounded by similar values, i.e. serious crashes. Blue spots indicate statistically significant minor injury crashes surrounded by minor injury crashes, and gray spots indicate points which are not statistically significant and surrounded by randomly distributed severity values. Three different lightness categories for red and blue hotspots indicate statistical confidence levels of 90%, 95%, and 99%. The darkness and intensity of hotspots color represent the statistical significance level. Fig. 6 shows serious and minor injury crash clusters in the study area.

The clusters in Fig. 6 were distinguished from general crash-prone locations because their composition was based on points with similar crash severity value. Most of serious injury crash clusters can be seen in Barnes, Emmons, Grand Forks, and Walsh counties. The south boundary of Ward County also showed some clustering of serious injury crashes. Similarly, the clustering of crashes with low severity, i.e.

Table 2
Names and definitions of space-time cube patterns.

Patterns	Pattern Name	Definition
	New Hot/Cold Spot	A location that is a statistically significant hot/cold spot for the final time step and has never been a previously statistically significant hot/cold spot.
	Consecutive Hot/Cold Spot	A location with a single uninterrupted run of statistically significant hot/cold spot bins in the final time-step intervals. The location has never been a statistically significant hot/cold spot before the final hot/cold spot run and fewer than 90% of all bins are statistically significant hot/cold spots.
	Intensifying Hot/Cold Spot	A location in which the intensity of clustering of high counts in each time step is increasing overall and for which the increase is statistically significant (90%).
	Persistent Hot/Cold Spot	A location that has been a statistically significant hot/cold spot for 90% of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time.
	Diminishing Hot/Cold Spot	A location that has been a statistically significant hot/cold spot for 90% of the time-step intervals, including the final time step. Additionally, the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant.
	Sporadic Hot/Cold Spot	A location that is an on-again off-again hot/cold spot. Fewer than 90% of the time-step intervals have been statistically significant hot/cold spots and none of the time-step intervals have been statistically significant cold/hot spots.
	Oscillating Hot/Cold Spot	A statistically significant hot/cold spot for the final time-step interval that has a history of being a statistically significant cold/hot spot during a prior time step. Fewer than 90% percent of the time-step intervals have been statistically significant hot/cold spots.
	Historical Hot/Cold Spot	The most recent time period is not hot/cold, but at least 90% of the time-step intervals have been statistically significant hot/cold spots.
	No Pattern Detected	Does not fall into any of the defined hot or cold spot patterns.

Source: [53].

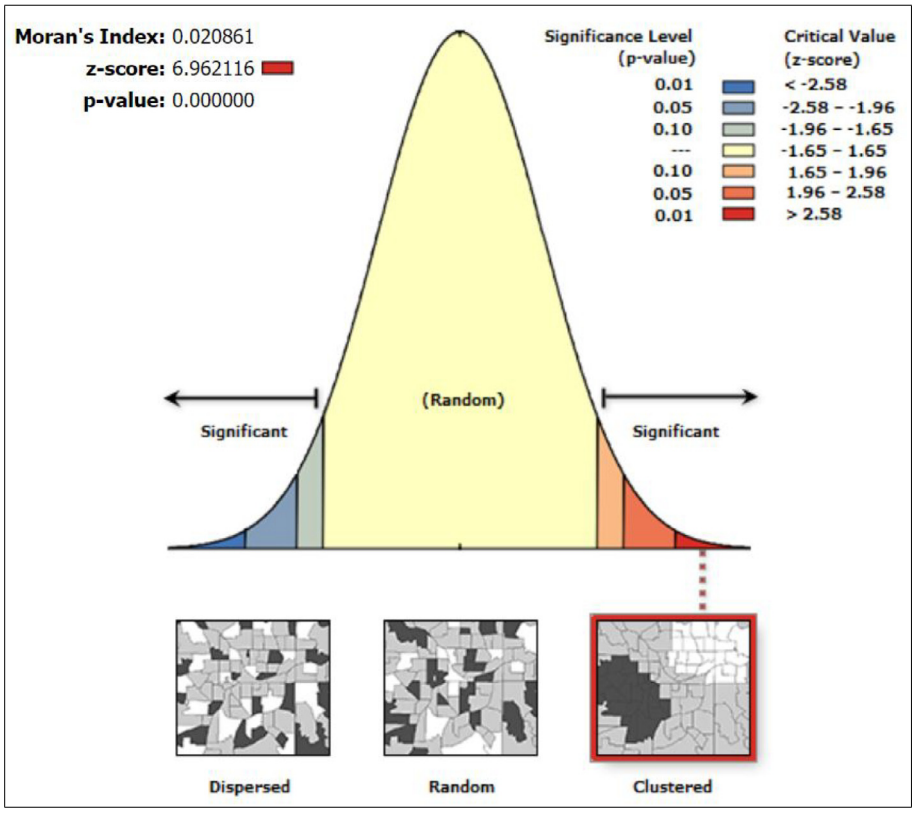

Fig. 4. Global Moran's I value for single-vehicle lane departure crashes.

minor injury, was present throughout the study area. However, most of the minor injury crash clusters were present in Barnes, Cass, McKenzie, Stark, Walsh, and Williams counties. The frequency of minor injury crash clusters in Williams, McKenzie, and Barnes counties was 32, 27, and 19 respectively.

It was observed that blue spots (minor injury crash clusters) tend to be less spread out in comparison to red spots (serious injury crash clusters). The randomly distributed crash spots (gray colored points) made the two different cluster types (i.e. red and blue) distinguishable. Moreover, the two different cluster types were not adjacent to each other. This shows the presence of obvious spatial relationship between geographic locations and the occurrence of crash severity clusters. One limitation of this approach is that it is difficult to quantify clusters along a road segment and distinguish its boundaries because they are created by points. One way to overcome this issue is to use density estimation which shows the concentration of points along the road network.

Hence, the third technique was applied. NetKDE is a successful spatial clustering method for investigating crashes on roadways. This method has a higher processing demand, so a county-level approach was needed. Williams was selected as the test county because it has the highest number of crash clusters. Fig. 7 shows a 3-D density map of crash density patterns in Williams County with subsection magnified for better visualization. The NetKDE results delivered comprehensive clustering patterns with clear boundaries. It can be observed that the majority of the crash clusters are located on curves, junctions, and intersections. Persistent clusters on horizontal curves can be seen throughout the county. The clusters emerging along the straight road segments might be related to speed. Investigation beyond this study's scope is needed to discern value with a granular understanding from this resource-intensive approach.

Table 3
The z-scores and *p*-values for a series of distance.

Distance (m)	Moran's Index	Expected Index	Variance	z-score	p-value
2000	0.05	−0.0004	0.0002	3.20	0.001
2500	0.05	−0.0004	0.0002	3.65	0.000
3000	0.05	−0.0003	0.0001	3.99	0.000
3500	0.04	−0.0003	0.0001	4.27	0.000
4000	0.04	−0.0003	0.0001	4.09	0.000
4500	0.03	−0.0003	0.0001	3.43	0.000
5000	0.03	−0.0003	0.0001	3.39	0.000
5500	0.02	−0.0003	0.0001	3.03	0.002
6000	0.02	−0.0003	0.0000	3.32	0.000
6500	0.02	−0.0003	0.0000	3.31	0.000

This study conducted a space-time cube analysis of 3878 single-vehicle lane departure crashes. The results of this analysis included six types of patterns as can be seen in Fig. 8. Persistent hotspots were distributed as spatial clusters in Burleigh, McKenzie, and Williams counties. This means that these locations have been statistically significant hotspots for 90% of the time-step intervals. There has been no visible trend with regard to a decrease or increase in the intensity of clusters over the time duration of the study. Two intensifying hotspots can also be seen in Cass and Ward counties. This refers to locations with statistically significant hotspots for 90% of the time-step intervals. Also, there has been a statistically significant increase in the intensity of clustering of high counts. Moreover, widely distributed new hotspots were generated in the western part of the state. These hotspots appeared with the passage of time. Finally, sporadic hotspots that repeatedly appeared and disappeared can be seen in both the East and West part of the state.

The results of different hotspot analysis methods are summarized in Fig. 9. The high and low crash incidence locations are represented by hotspots and cold spots respectively. Using all the methods together to identify and locate crash hotspots is a promising approach. This study is focused more on the spatial analysis whereby we are trying to locate the spatial clusters for the crash severity. Relevant countermeasures can be implemented according to the specific spatial and temporal trends of these hotspots. The study results show that some of the

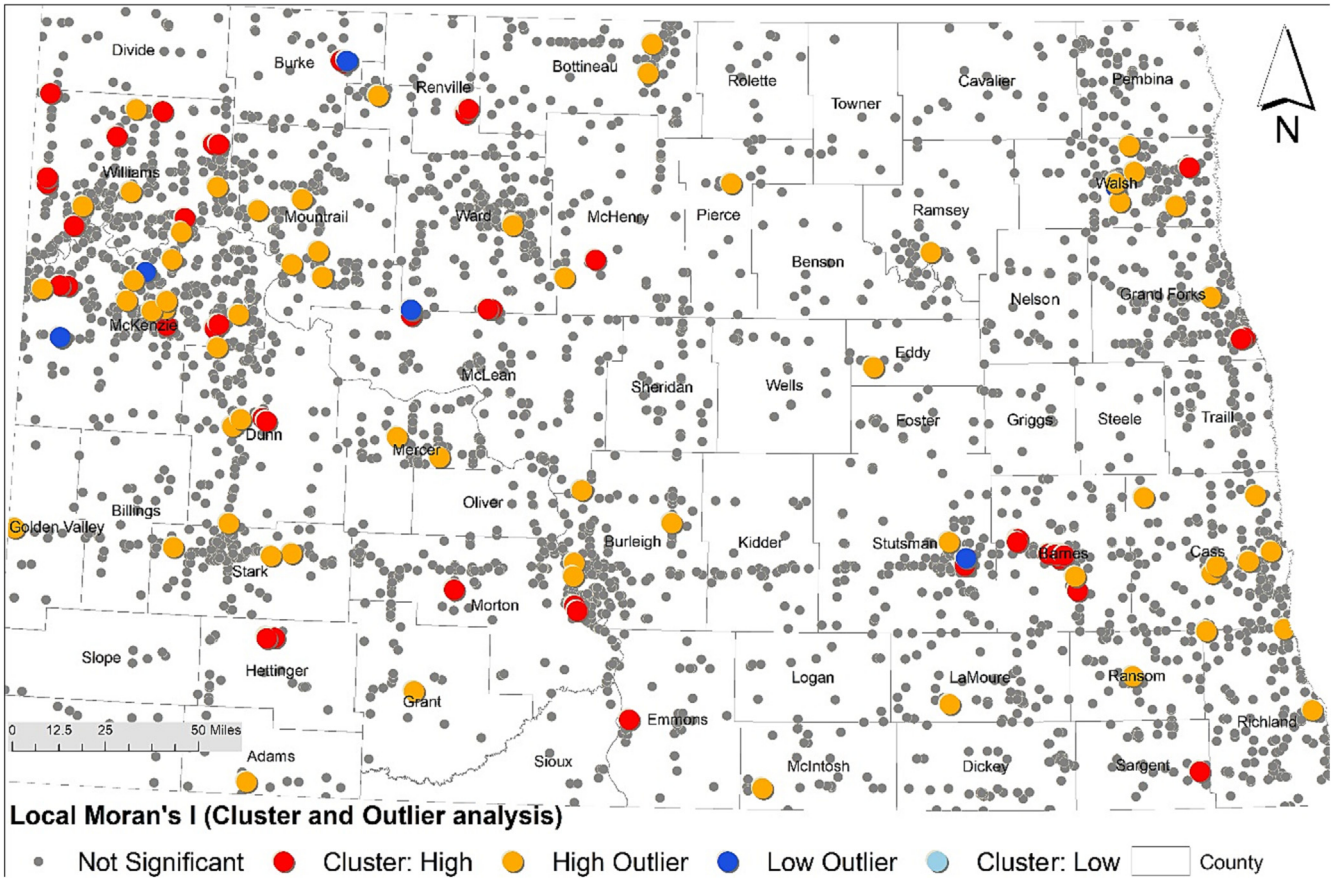

Fig. 5. Hotspot classification using local Moran's I method.

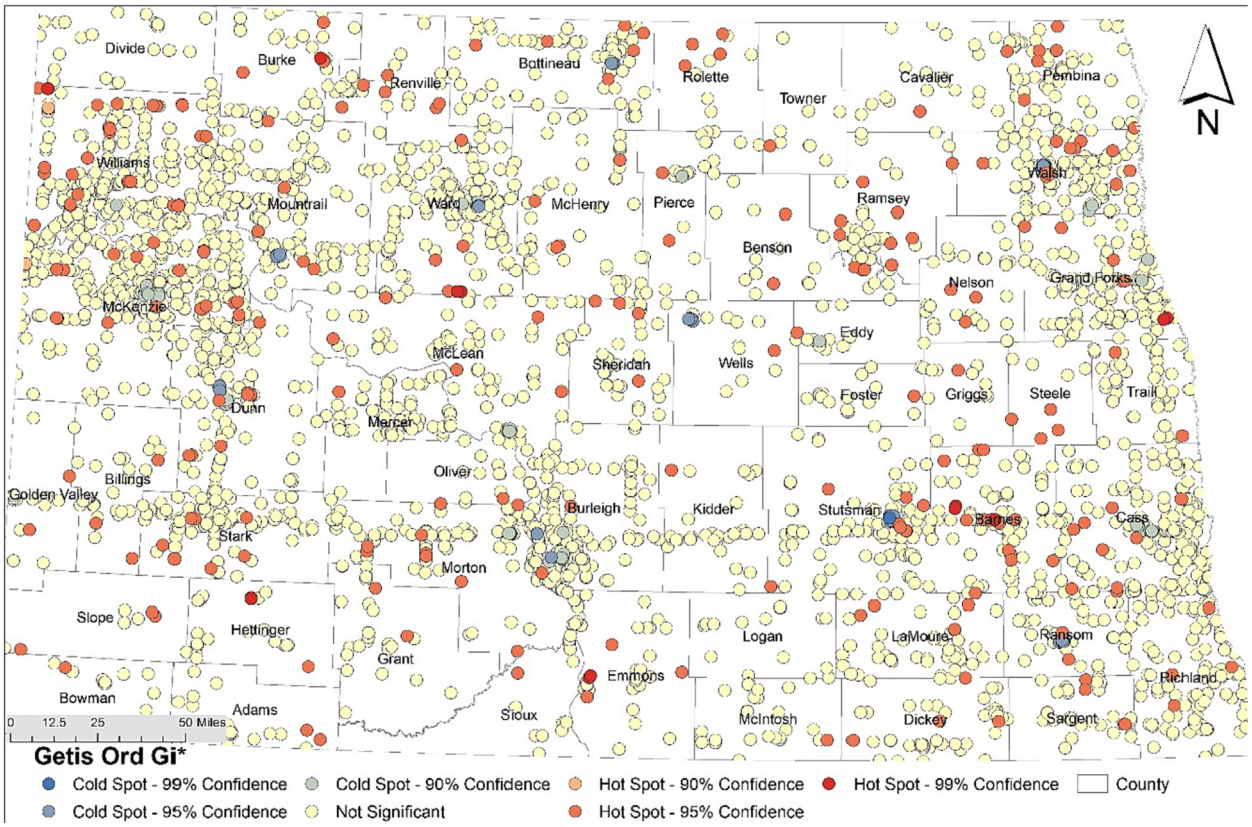

Fig. 6. Hotspot analysis using local spatial autocorrelation (Getis-Ord Gi*) approach.

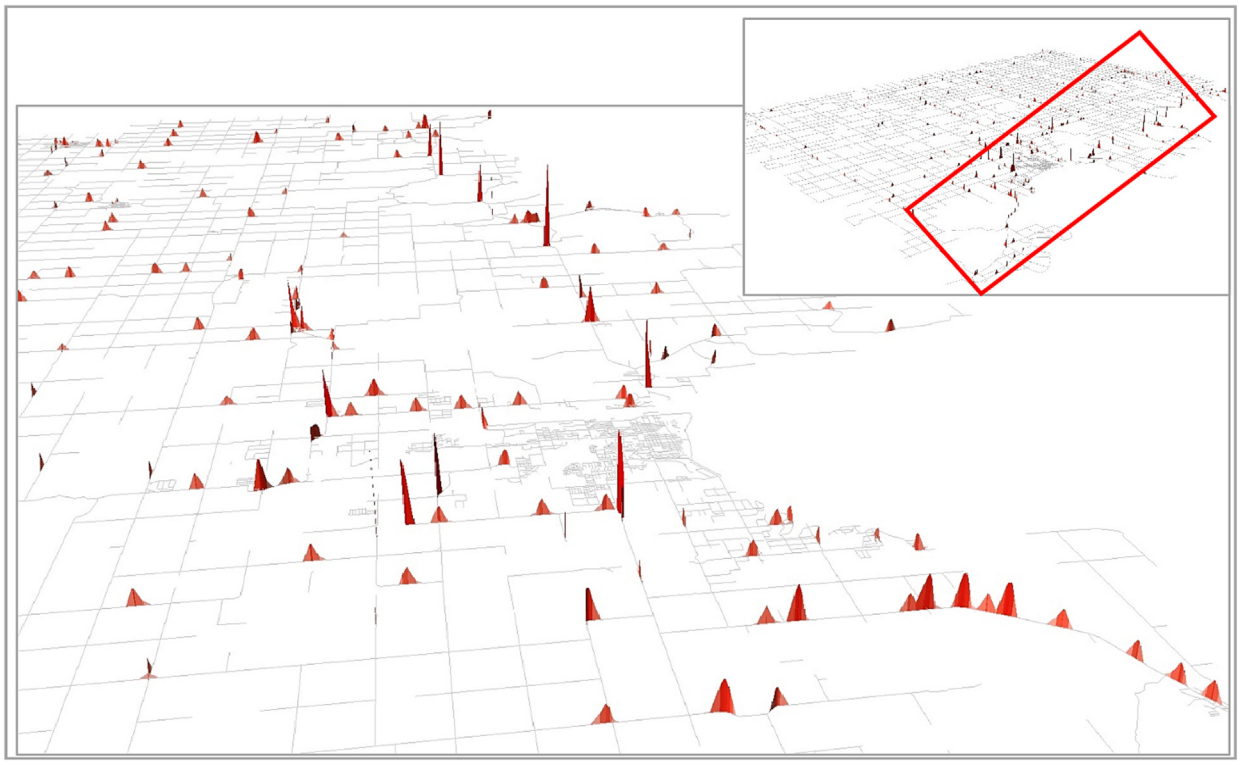

Fig. 7. Network Kernel density estimation for crashes in Williams County.

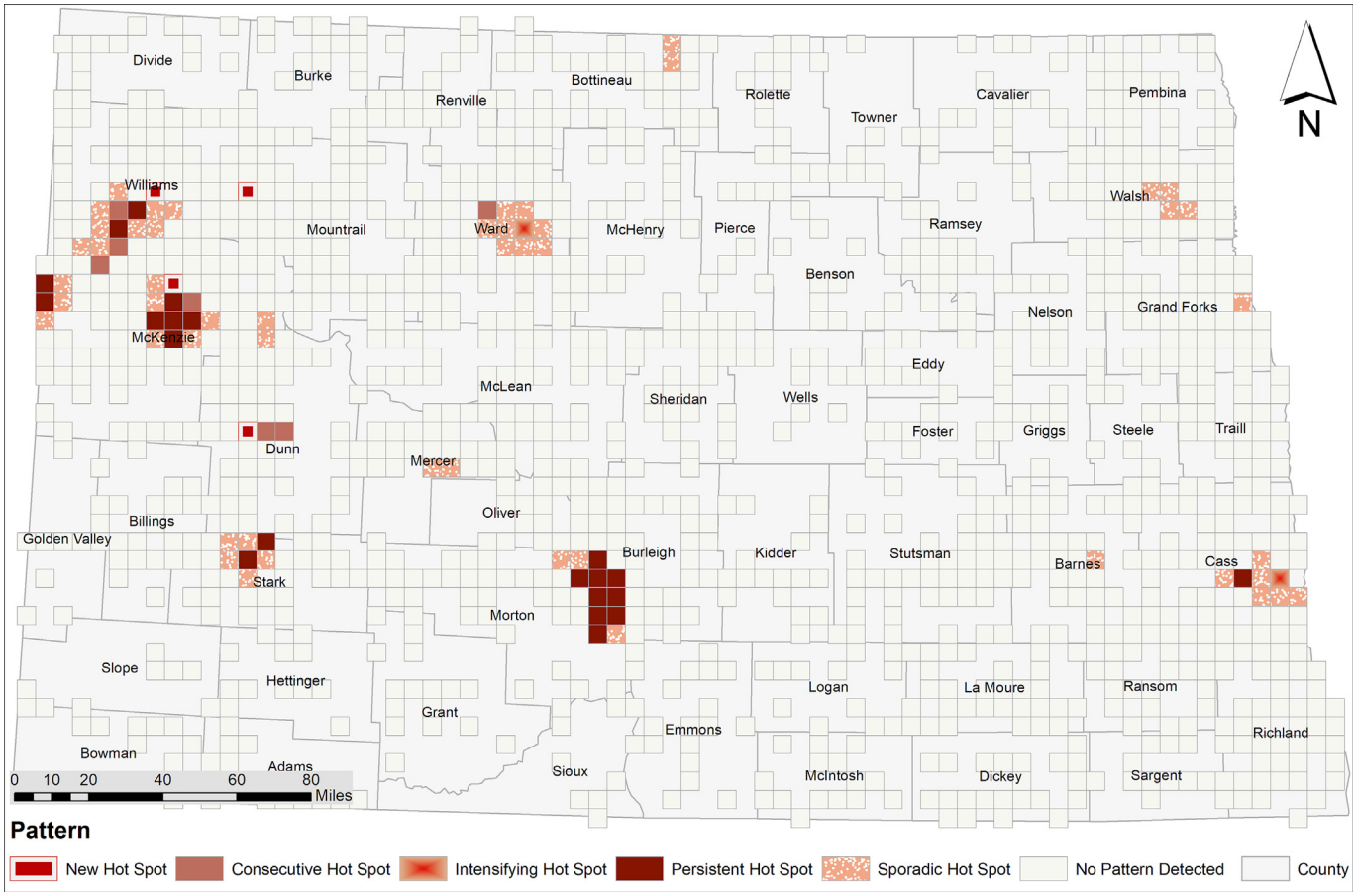

Fig. 8. Emerging hotspot analysis of single-vehicle lane departure crashes in ND.

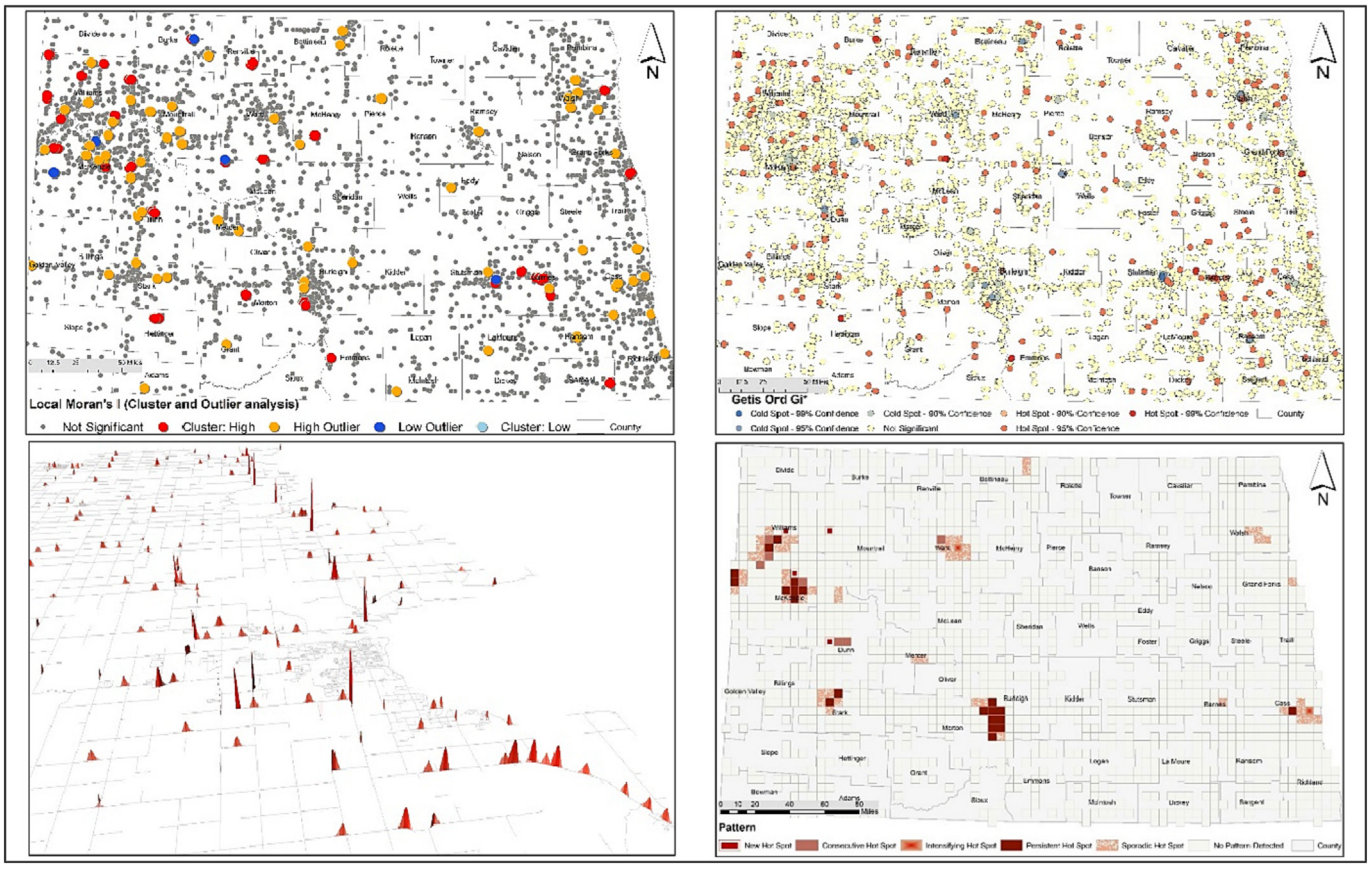

Fig. 9. Comparison of results from different hotspots methods – (top left) local Moran's I (top right) Getis Ord Gi* (bottom left) network kernel density estimation (bottom right) emerging hotspot analysis.

locations in Burleigh, McKenzie, and Williams counties have different phenomena which should attract the attention of the traffic safety agencies. The findings of the study can be used to direct the limited resources and to propose preventive strategies targeting the crash hotspots.

6. Conclusions

The benefit of classifying crash hotspots cannot be over-emphasized as an important method of discerning geospatial themes in road safety. Hence, this study employed hotspot analysis techniques to reveal spatial patterns of single-vehicle lane departure crashes. While the Global Moran's I index indicated the existence of crash clustering, use of the local Moran's I statistic enabled the identification of hot and cold spots within the study area. Furthermore, data-driven enhancements were placed side by side with locations in terms of statistically significant high and low crash concentrations. A more granular crash cluster quantification, using the NetKDE technique, was presented for Williams County. A 3D density map created in ArcScene also defined boundaries for each cluster in terms of density values embedded in the roadway. Finally, the space-time cube analysis enabled us to identify different patterns in the spatial data with respect to time.

As effective as this method was, it does have shortcomings. Particularly, it fails to give an accurate statistical significance of the resulting crash clusters, which therefore suggests the need for further investigation. This article focused on lane departure crashes that involved only single vehicles. Another limitation of the study was limiting events to a specific road type in the state network. Future studies can overcome these limitations by broadening the crash type and locations. Results are valuable as a sample for road safety stakeholders to consider for future endeavors aimed at data-driven tools to reveal crash hotspots and to distinguish patterns across the state. Scrutiny of characteristics for crashes in the areas with high frequency of clusters will further enhance the understanding of associated risk factors and appropriate prevention strategies such as heightened enforcement, public education, and infrastructure enhancements including rumble strip installations, safety edges, post-mounted delineators, oversized chevrons, and advisory speed marking in lanes on curves. The tools approach used in this study can be adopted by other jurisdictions seeking to empirically visualize hotspots and more effectively deploy traffic safety strategies.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was funded by the U.S. Department of Transportation's Office of the Assistant Secretary for Research & Technology. Grant 69A3551747108 was awarded to the Mountain Plains Consortium under the competitive University Transportation Center program. The authors greatly appreciate North Dakota Department of Transportation support in this effort to enhance insight into lane departure crashes in the state. Contents of this paper reflect the views of the authors, who are responsible for the facts and accuracy of the information presented.

References

[1] Roadway Departure Safety - FHWA, https://safety.fhwa.dot.gov/roadway_dept/ 2022 (accessed September 28, 2021).
[2] 2019 FHWA, Driving FoRRRwD-A Systemic Approach to Reducing Roadway Departures, 2019.

[3] Vision Zero NDDOT, https://visionzero.nd.gov/strategies/LaneDeparture/ 2022 (accessed September 27, 2021).
[4] E. Moons, T. Brijs, G. Wets, Improving moran's index to identify hot spots in traffic safety, Stud. Comput. Intell. 176 (2009) 117–132, https://doi.org/10.1007/978-3-540-89930-3_7.
[5] M.A. Dereli, S. Erdogan, A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods, Transp. Res. Part A Pol. Pract. 103 (2017) 106–117, https://doi.org/10.1016/J.TRA.2017.05.031.
[6] L. Li, L. Zhu, D.Z. Sui, A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes, J. Transp. Geogr. 15 (2007) 274–285, https://doi.org/10.1016/J.JTRANGEO.2006.08.005.
[7] T.K. Anderson, Kernel density estimation and K-means clustering to profile road accident hotspots, Accid. Anal. Prev. 41 (2009) 359–364, https://doi.org/10.1016/J.AAP.2008.12.014.
[8] S.S. Vemulapalli, M.B. Ulak, E.E. Ozguven, T. Sando, M.W. Horner, Y. Abdelrazig, R. Moses, GIS-based spatial and temporal analysis of aging-involved accidents: a case study of three counties in Florida, Appl. Spat. Anal. Pol. 10 (2017) 537–563, https://doi.org/10.1007/S12061-016-9192-4/FIGURES/9.
[9] Z. Xie, J. Yan, Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach, J. Transp. Geogr. 31 (2013) 64–71, https://doi.org/10.1016/J.JTRANGEO.2013.05.009.
[10] A. Vogt, J. Bared, Accident models for two-lane rural segments and intersections, Transp. Res. Rec. J. Transp. Res. Board. (1998) 18–29, https://doi.org/10.3141/1635-03.
[11] C. Zhang, J.N. Ivan, Effects of geometric characteristics on head-on crash incidence on two-lane roads in connecticut, Transp. Res. Rec. J. Transp. Res. Board. 2005 (1908) 159–164, https://doi.org/10.1177/0361198105190800119.
[12] R. Elvik, Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads, 2008 72–75, https://doi.org/10.3141/2083-08.
[13] W. Cheng, S.P. Washington, Experimental evaluation of hotspot identification methods, Accid. Anal. Prev. 37 (2005) 870–881, https://doi.org/10.1016/J.AAP.2005.04.015.
[14] E. Sacchi, T. Sayed, K. El-Basyouny, Multivariate Full Bayesian Hot Spot Identification and Ranking: New Technique, 2515, 2015 1–9, https://doi.org/10.3141/2515-01.
[15] H.L. Huang, H.C. Chin, M.M. Haque, Hotspot identification: a full bayesian hierarchical modeling approach, Transp. Traffic Theory 2009 Golden Jubil. (2009) 441–462, https://doi.org/10.1007/978-1-4419-0820-9_22.
[16] A. Getis, J.K. Ord, The analysis of spatial association by use of distance statistics, Geogr. Anal. 24 (1992) 189–206, https://doi.org/10.1111/J.1538-4632.1992.TB00261.X.
[17] J.K. Ord, A. Getis, Local spatial autocorrelation statistics: distributional issues and an application, Geogr. Anal. 27 (1995) 286–306, https://doi.org/10.1111/J.1538-4632.1995.TB00912.X.
[18] P.A.P. Moran, Notes on continuous stochastic phenomena, Biometrika. 37 (1950) 17–23, https://doi.org/10.2307/2332142.
[19] L. Anselin, Local indicators of spatial association—LISA, Geogr. Anal. 27 (1995) 93–115, https://doi.org/10.1111/J.1538-4632.1995.TB00338.X.
[20] C. Sabel, S. Kingham, A. Nicholson, P. Bartie, Road traffic accident simulation modelling - A Kernel estimation approach, Proc., 17th Annu. Colloq. Spat. Inf. Res. Cent. Univ. Otago, Dunedin, New Zeal 2005, pp. 67–75.
[21] S. Erdogan, I. Yilmaz, T. Baybura, M. Gullu, Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar, Accid. Anal. Prev. 40 (2008) 174–181, https://doi.org/10.1016/j.aap.2007.05.004.
[22] Ö. Kaygisiz, Ş. Düzgün, A. Yildiz, M. Senbil, Spatio-temporal accident analysis for accident prevention in relation to behavioral factors in driving: the case of south Anatolian motorway, Transp. Res. Part F Traffic Psychol. Behav. 33 (2015) 128–140, https://doi.org/10.1016/J.TRF.2015.07.002.
[23] P.-F. Kuo, X. Zeng, D. Lord, Guidelines for choosing hot-spot analysis tools based on data characteristics, network restrictions, and time distributions, 91st Annu. Meet. Transp. Res. Board, January 22–26, Washington, DC, 2012.
[24] L. Thakali, T.J. Kwon, L. Fu, Identification of crash hotspots using kernel density estimation and kriging methods: a comparison, J. Mod. Transp. 23 (2015) 93–106, https://doi.org/10.1007/S40534-015-0068-0/FIGURES/10.
[25] S. Mitra, Enhancing road traffic safety: a GIS based methodology to identify potential areas of improvement, Proj. Rep. Submitt. to William Barbar. Leonard Univ. Transp. Cent. 2008, pp. 1–23.
[26] E. Moons, T. Brijs, G. Wets, Improving Moran's index to identify hot spots in traffic safety, Geocomp. Urban Plan. 176 (2009) 117–132, https://doi.org/10.1007/978-3-540-89930-3_7.
[27] P. Songchitruksa, X. Zeng, Getis–Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data, 2010 42–51, https://doi.org/10.3141/2165-05.
[28] Z. Dezman, L. de Andrade, J.R. Vissoci, D. El-Gabri, A. Johnson, J.M. Hirshon, C.A. Staton, Hotspots and causes of motor vehicle crashes in Baltimore, Maryland: a geospatial analysis of five years of police crash and census data, Injury. 47 (2016) 2450–2458, https://doi.org/10.1016/J.INJURY.2016.09.002.
[29] H. Zubaidi, I. Obaid, H.A. Mohammed, S. Das, N.S.S. Al-Bdairi, Hot spot analysis of the crash locations at the roundabouts through the application of GIS, J. Phys. Conf. Ser. 1895 (2021), 012032, https://doi.org/10.1088/1742-6596/1895/1/012032.
[30] K. Hazaymeh, A. Almagbile, A.H. Alomari, K. Hazaymeh, A. Almagbile, A.H. Alomari, Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques, ISPRS Int. J. Geo-Inform. Vol. 11 (2022) 260, https://doi.org/10.3390/IJGI11040260.
[31] M. Chance Scott, S. Sen Roy, S. Prasad, Spatial patterns of off-the-system traffic crashes in Miami-Dade County, Florida, during 2005-2010, Traffic Inj. Prev. 17 (2016) 729–735, https://doi.org/10.1080/15389588.2016.1144878.
[32] Z. Cheng, Z. Zu, J. Lu, Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections, Sustain. 11 (2019) 160 , 11 (2018) 160 https://doi.org/10.3390/SU11010160.
[33] I. Kveladze, M.J. Kraak, C.P.J.M. van Elzakker, A Methodological Framework for Researching the Usability of the Space-Time Cube, 50, 2013 201–210, https://doi.org/10.1179/1743277413Y.0000000061.
[34] U. Demšar, K. Virrantaus, Space–time Density of Trajectories: Exploring Spatio-temporal Patterns in Movement Data, 24, 2010 1527–1542, https://doi.org/10.1080/13658816.2010.511223.
[35] T.B. Fang, Y. Lu, Constructing a near real-time space-time cube to depict urban ambient air pollution scenario, Trans. GIS 15 (2011) 635–649, https://doi.org/10.1111/J.1467-9671.2011.01283.X.
[36] T. Nakaya, K. Yano, Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density estimation and scan statistics, Trans. GIS 14 (2010) 223–239, https://doi.org/10.1111/J.1467-9671.2010.01194.X.
[37] R.C. Sadler, J. Pizarro, B. Turchan, S.P. Gasteyer, E.F. McGarrell, Exploring the spatial-temporal relationships between a community greening program and neighborhood rates of crime, Appl. Geogr. 83 (2017) 13–26, https://doi.org/10.1016/J.APGEOG.2017.03.017.
[38] J. Yoon, S. Lee, Spatio-temporal patterns in pedestrian crashes and their determining factors: application of a space-time cube analysis model, Accid. Anal. Prev. 161 (2021), 106291 https://doi.org/10.1016/J.AAP.2021.106291.
[39] Y. Kang, N. Cho, S. Son, Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation, PLoS One 13 (2018), e0196845 https://doi.org/10.1371/JOURNAL.PONE.0196845.
[40] P. Purwanto, S. Utaya, B. Handoyo, S. Bachri, I.S. Astuti, K. Sastro, B. Utomo, Y.E. Aldianto, Spatiotemporal analysis of COVID-19 spread with emerging hotspot analysis and space–time cube models in East Java, Indonesia, ISPRS Int. J. Geo-Inform. 10 (2021) 133 , 10 (2021) 133 https://doi.org/10.3390/IJGI10030133.
[41] I. Kveladze, M.J. Kraak, C.P.J.M. van Elzakker, Cartographic Design and the Space–Time Cube, 56, 2018 73–90, https://doi.org/10.1080/00087041.2018.1495898.
[42] D. O'Sullivan, D.J. Unwin, Geographic Information Analysis: Second Edition, Geogr. Inf. Anal. Second ed., 2010 https://doi.org/10.1002/9780470549094.
[43] A. Mitchell, The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics, second edition Esri Press, 2021 https://www.esri.com/en-us/esri-press/browse/the-esri-guide-to-gis-analysis-volume-2-spatial-measurements-and-statistics-second-edition (accessed December 8, 2021).
[44] J.K. Maingi, J.M. Mukeka, D.M. Kyale, R.M. Muasya, Spatiotemporal patterns of elephant poaching in South-Eastern Kenya, Wildl. Res. 39 (2012) 234–249, https://doi.org/10.1071/WR11017.
[45] A. Okabe, K. Sugihara, Spatial Analysis Along Networks: Statistical and Computational Methods, 2012 288 https://www.wiley.com/en-us/Spatial+Analysis+Along+Networks%3A+Statistical+and+Computational+Methods-p-9780470770818 (accessed December 10, 2021).
[46] T. Steenberghen, K. Aerts, I. Thomas, Spatial clustering of events on a network, J. Transp. Geogr. 18 (2010) 411–418, https://doi.org/10.1016/J.JTRANGEO.2009.08.005.
[47] Z. Xie, J. Yan, Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach, J. Transp. Geogr. 31 (2013) 64–71, https://doi.org/10.1016/j.jtrangeo.2013.05.009.
[48] Z. Xie, J. Yan, Kernel density estimation of traffic accidents in a network space, Comput. Environ. Urban. Syst. 32 (2008) 396–406, https://doi.org/10.1016/J.COMPENVURBSYS.2008.05.001.
[49] B.W. Silverman, Density Estimation for Statistics and Data Analysis, 1998 175.
[50] Emerging Hot Spot Analysis (Space Time Pattern Mining)—ArcGIS Pro | Documentation, https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/emerginghotspots.htm 2022 (accessed September 10, 2022).
[51] Henry B. Mann, Nonparametric Tests Against Trend, 13, 1945 245–259 https://www.jstor.org/stable/1907187 (accessed September 9, 2022).
[52] G. Maurice, Kendall, Rank Correlation Methods, 4th ed. Griffin, London, 1975 https://www.worldcat.org/title/rank-correlation-methods/oclc/3827024 (accessed September 9, 2022).
[53] ESRI, How Emerging Hot Spot Analysis works—ArcGIS Pro | Documentation, https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/learnmoreemerging.htm 2022 (accessed September 10, 2022).