

# A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord $G_i^*$ Statistic

Azad Abdulhafedh\*

Department of Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA

Email: asa8cd@mail.missouri.edu

**How to cite this paper:** Abdulhafedh, A. (2017) A Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord  $G_i^*$  Statistic. *Open Journal of Civil Engineering*, 7, 208-221. <https://doi.org/10.4236/ojce.2017.72013>

**Received:** February 13, 2017

**Accepted:** June 3, 2017

**Published:** June 6, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Spatial autocorrelation is a measure of the correlation of an observation with other observations through space. Most statistical analyses are based on the assumption that the values of observations are independent of one another. Spatial autocorrelation violates this assumption, because observations at nearby locations are related to each other, and hence, the consideration of spatial autocorrelations has been gaining attention in crash data modeling in recent years, and research have shown that ignoring this factor may lead to a biased estimation of the modeling parameters. This paper examines two spatial autocorrelation indices: Moran's Index; and Getis-Ord  $G_i^*$  statistic to measure the spatial autocorrelation of vehicle crashes occurred in Boone County roads in the state of Missouri, USA for the years 2013-2015. Since each index can identify different clustering patterns of crashes, therefore this paper introduces a new hybrid method to identify the crash clustering patterns by combining both Moran's Index and  $G_i^*$  statistic. Results show that the new method can effectively improve the number, extent, and type of crash clustering along roadways.

## Keywords

Spatial Autocorrelation, Moran's Index, Getis-Ord  $G_i^*$  Statistic, Vehicle Crashes

## 1. Introduction

In many vehicle crash data, geographic relationships among crashes can exist, and this phenomenon is termed spatial autocorrelation, which is a measure of

\*PhD in Civil Engineering.

the correlation of a crash with other crashes through space. Most statistical analyses are based on the assumption that the values of observations in each sample are independent of one another. Spatial autocorrelation violates this assumption, because samples taken from nearby locations are related to each other, and hence, they are statistically not independent of one another [1] [2]. Therefore, the consideration of spatial autocorrelations has been gaining attention in crash data modeling in recent years, and researchers have shown that ignoring this factor may lead to a biased estimation of the model parameters [3]-[12]. Taking the spatial autocorrelation into account in crash modeling can improve model parameter estimation, and the overall model fit [8] [13]. The spatial autocorrelation phenomenon can be best summarized by the Tobler's first law of Geography that everything is related to everything else but those which are near to each other are more related when compared to those that are further away [14].

Spatial autocorrelation can be positive or negative among observations. Positive spatial autocorrelation occurs when observations having similar values are closer (*i.e.* clustered) to one another, and negative spatial autocorrelation occurs when observations having dissimilar values occur near one another [2] [15]. Two problems may be faced when sample data has a locational dimension: 1) the existence of spatial autocorrelation between the observations, and 2) the variation of this relationship over the space that could be described as spatial heterogeneity [16] or spatial non-stationarity [17]. Hence, spatial autocorrelation must be incorporated in modeling crash data to properly account for the effect of spatial correlation and any unobserved spatial heterogeneity that may exist in the crash data. To assess spatial autocorrelation, a distance measure must be specified in order to define what is meant by two observations being close together. These distances are usually presented in the form of a weight matrix, which defines the relationships between locations at which the observations occur [18]. If data are collected at  $n$  locations, then the weight matrix will be  $n \times n$  with zeroes on the diagonal. The weight matrix is often row-standardized, (*i.e.* all the weights in a row sum to one), and can be constructed given a variety of assumptions [2], such as, a constant distance that represents the weight for any two different locations; a fixed weight for all observations within a specified distance; or  $k$  nearest neighbors.

## 2. Background Literature

The differences between the network autocorrelation and spatial autocorrelation were examined [1]. In this study, it was shown that the network autocorrelation could influence the values associated with a network link given its relationship to another link in the network. To account for these relationships, spatial autocorrelation was only modeled between neighboring (adjacent) network links. The effect of spatial autocorrelation on traffic crashes was examined by geo-coding them to the nearest intersection or ramp, and then calculating different spatial statistics such as, mean, standard deviation, and standard deviational ellipse [19]. In another study the spatial autocorrelation of road segments was examined

by using the Moran's Index [20], and it was found that a significant level of positive spatial autocorrelation existed in the data. When investigating spatial autocorrelation among traffic crashes, [21] estimated a series of crash frequency models aggregated at the county level for the state of Texas. The rear-end crashes at signalized intersections were analyzed to model the spatial correlation between intersections [22]. In this study, three different correlation structures were considered: independent correlation, exchangeable correlation, and autoregressive correlation, where the correlation decreases as the gap between intersections increases. The models proved that high spatial correlations exist between intersections for rear-end crashes. A multivariate spatial modeling approach for excess crash frequency and severity was developed in cantons (counties) for Costa Rica [23], and results showed that the multivariate spatial model performed better than univariate spatial models. The study also reported that the effects of spatial smoothing due to multivariate spatial random effects were evident in the estimation of no-injury collisions. Generalized Poisson models were utilized [24] to explore the spatial autocorrelation of crashes, and found that spatial correlation sharply decreases at distances exceeding 7 km, and shorter road segments with high crash frequency tend to have higher spatial dependency.

### 3. Global and Local Indices of Spatial Autocorrelation

There are many indices or statistics that attempt to measure spatial autocorrelation for count data, such as Moran's index (also called Moran's  $I$ ), the Geary's  $C$ , and the Getis-Ord  $G$  statistic. These indices can be computed as Global or Local measures depending on the scope of the analysis. Global spatial autocorrelation measures the overall spatial autocorrelation of the entire study area, providing a single measurement of spatial autocorrelation for an entire data. Local spatial autocorrelation measures the spatial autocorrelation of individual features and identifies the spatial patterns across the study area considering the relationship between individual features. Indices of spatial autocorrelation are based on the general index of matrix association (*i.e.* the Gamma  $\Gamma$  index). The Global Gamma index consists of the sum of the cross products of the elements  $a_{ij}$  and  $b_{ij}$  in two matrices of similarity, using spatial similarity in one matrix and value similarity in the other matrix, such that [25]:

$$\Gamma = \sum_i \sum_j a_{ij} b_{ij} \quad (1)$$

Using different value similarity would result in different indices. For example, setting  $a_{ij} = x_i x_j$  would result in Moran's  $I$  statistic, and setting  $a_{ij} = (x_i - x_j)^2$  would result in Geary's  $C$  index [25]. The Global Gamma index equals the sum of local Gamma indices within the study area. Anselin [25] outlined a general class of local indicators of spatial autocorrelation termed the Local Indicator of Spatial Autocorrelation ( $LISA$ ) statistic that satisfies two conditions, first; the  $LISA$  for each point or section in the space gives an indication of significant spatial clustering (grouping) of similar or dissimilar values around that point or section, and second; the sum of  $LISAs$  for all points or sections in a given study area is proportional to a corresponding global indicator of spatial autocorrelation for

that area, which implies that the *LISA* statistic decomposes global results into their local parts. For example, a significant global index of a study area may hide large spatial patches of no autocorrelation, and *LISA* can detect this and show the locations of these insignificant patches in space. Conversely, an insignificant global index may hide patches of strong autocorrelation, and *LISA* can detect this again.

#### 4. Moran's *I*

Moran's *I* statistic is one of the oldest indices of spatial autocorrelation and can be used to test for global and local spatial autocorrelation among continuous data. For any continuous variable,  $x$ , a mean  $\bar{x}$ , can be calculated and the deviation of any observation from that mean can be calculated based on the cross products of the deviations from the mean. The statistic then compares the value of the variable at any one location with the values at all other locations [26] [27] [28]. For  $n$  observations on a variable  $x$  at locations  $i, j$ , Global Moran's *I* can be calculated as follows [25]:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

where,

$\bar{x}$ : the mean of the variable  $x$ ;

$x_i$ : the value of variable  $x$  at location  $i$ ;

$x_j$ : the value of variable  $x$  at location  $j$ ;

$w_{ij}$ : the elements of the weight matrix;

$n$ : number of observations;

$S_0$ : is the sum of the elements of the weight matrix:  $S_0 = \sum_i \sum_j w_{ij}$ .

The local Moran's *I* for location  $i$  can be calculated as follows:

$$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x}) \quad (3)$$

$$S_i^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1} - \bar{x}^2 \quad (4)$$

Values for this index typically, range from  $-1.0$  to  $+1.0$ , where a value of  $-1.0$  indicates negative spatial autocorrelation, and a value of  $+1.0$  indicates positive spatial autocorrelation. When nearby points have similar Moran's values, their cross product is high. Conversely, when nearby points have dissimilar Moran's values, their cross-product is low. The expectation of Moran's *I* statistic is:

$$E(I) = \frac{-1}{n-1} \quad (5)$$

When a Moran's *I* value is larger than  $E(I)$ , this would indicate positive spatial autocorrelation, and if a Moran's *I* is less than  $E(I)$ , this would indicate negative spatial autocorrelation. In Moran's initial formulation, the weight variable,  $w_{ij}$ , was a contiguity matrix. Therefore, if zone  $j$  is adjacent to zone  $i$ , the product receives a weight of 1.0, otherwise, the product receives a weight of 0.0. A study

[29] generalized these definitions to include any type of weight, and in a wider term,  $w_{ij}$  is a distance-based weight which is the inverse distance between locations  $i$  and  $j$  ( $1/d_{ij}$ ). The  $z$ -score of Moran's  $I$  can be computed as follows:

$$Z_i = \frac{I - E(I)}{\sqrt{V(I)}} \quad (6)$$

where  $E(I)$  is the expected value of  $I$ , and  $V(I)$  is the variance of  $I$ , as shown in Equation (7):

$$V(I) = E(I^2) - E^2(I) \quad (7)$$

## 5. Getis-Ord $G$ Statistic

The Getis-Ord  $G$  statistic is calculated with respect to a specified threshold distance (defined by the user) rather than to an inverse distance, as with the Moran's  $I$  [30] [31]. The Global  $G$  statistic computes a single statistic for the entire study area, while the  $G_i$  statistic is an indicator for local spatial autocorrelation for each data point. The Global  $G$  statistic can be calculated as follows [32]:

$$G = \frac{\sum_i \sum_{j \neq i} w_{ij} x_i x_j}{\sum_i \sum_{j \neq i} x_i x_j} \quad (8)$$

where,

- $x_i$ : the value of variable  $x$  at location  $i$
- $x_j$ : the value of variable  $x$  at location  $j$
- $w_{ij}$ : the elements of the weight matrix.

There are two types of local  $G_i$  statistics, although almost the two types produce identical results [31] [33]. The first one,  $G_p$  does not include the autocorrelation of a zone with itself, whereas the  $G_i^*$  includes the interaction of a zone with itself (*i.e.* the  $G_i$  statistic does not include the value of  $X_i$  itself, but only the neighborhood values, but  $G_i^*$  includes  $X_i$  as well as the neighborhood values), and both can be computed by the formulae [32]:

$$G_i(d) = \frac{\sum_{j \neq i}^n w_{ij} x_j}{\sum_{j \neq i}^n x_j} \quad (9)$$

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij} x_j}{\sum_{j=1}^n x_j} \quad (10)$$

where  $d$  is the neighborhood (threshold) distance, and  $w_{ij}$  is the weight matrix that has only 1.0 or 0.0 values, 1.0 if  $j$  is within  $d$  distance of  $i$ , and 0.0 if its beyond that distance. These formulae indicate that the cross-product of the value of  $X$  at location  $i$  and at another location  $j$  is weighted by a distance weight,  $w_{ij}$  which is defined by either a 1.0 if the two locations are equal to or closer than a threshold distance,  $d$ , or a 0.0 otherwise. The  $G$  statistic can vary between 0.0 and 1.0. The statistical significance of the local autocorrelation between each point and its neighbors is assessed by the  $z$ -score test and the  $p$ -value. ArcGIS uses the following formulae to calculate the local Getis-Ord  $G_i^*$  [34]:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{x} \sum_{j=1}^n w_{ij}}{S \sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - \left(\sum_{j=1}^n w_{ij}\right)^2}{n-1}}} \quad (11)$$

$$\bar{x} = \frac{\sum_{j=1}^n x_j}{n} \quad (12)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{x}^2} \quad (13)$$

where,

$x_i$ : the value of variable  $x$  at location  $i$ ;

$x_j$ : the value of variable  $x$  at location  $j$ ;

$w_{ij}$ : the elements of the weight matrix;

$n$ : number of observations.

The expected  $G$  value for a threshold distance,  $d$ , is defined as:

$$E[G(d)] = \frac{W}{n(n-1)} \quad (14)$$

where  $W$  is the sum of weights for all pairs of locations ( $W = \sum_i \sum_j w_{ij}$ ), and  $n$  is the number of observations. Assuming normal distribution, the variance of  $G(d)$  is defined as:

$$\text{Var}[G(d)] = E(G^2) - E^2(G) \quad (15)$$

The standard error of  $G(d)$  is the square root of the variance of  $G$ . Therefore, a  $z$ -test can be computed by:

$$S.E.[G(d)] = \sqrt{\text{Var}[G(d)]} \quad (16)$$

$$Z[G(d)] = \frac{G(d) - E[G(d)]}{S.E.[G(d)]} \quad (17)$$

## 6. Classification of Crash Clustering Patterns

The crash clustering patterns (*i.e.* type of concentration of crashes) and its statistical significance is evaluated based on the output  $z$ -scores, the correspondent  $p$ -values and the confidence level. These will determine whether a crash is classified as having a significant high spatial autocorrelation (denoted by High-High, HH), a significant low spatial autocorrelation (denoted by Low-Low, LL), a significant dispersed outlier (either a high value surrounded by low value denoted by HL, or vice versa, a low value surrounded by high value denoted by LH), or insignificant random crash. A high positive  $z$ -score for a crash point indicates a significant spatial autocorrelation (either with high values HH or with low values LL). A low negative  $z$ -score for a crash point indicates a statistically significant spatial outlier (either with high-low HL or low-high LH). A  $z$ -score of a crash point close to zero indicates that the crash is randomly and independently distributed in space. To determine if the  $z$ -score is statistically significant, it should

be compared to a range of values for a particular confidence level. For example, at a significance level of 95%, a  $z$ -score would have to be less than  $-1.96$  or greater than  $+1.96$  to be statistically significant. Typical confidence levels are 90%, 95%, or 99%. **Table 1** shows the critical  $p$ -values and  $z$ -scores for different confidence levels.

## 7. Data

To illustrate the analysis framework presented in this paper, Boone County, Missouri, USA crash data for the years (2013-2015) are used. Missouri crash data is reported by the Missouri State Highway Patrol (MSHP) and recorded in the Missouri Statewide Traffic Accident Records System (STARS). The total observed crashes within the three years 2013-2015 is 6886.0 along roads in Boone County. **Figure 1** shows Boone County road network in Missouri and the distribution of crashes (2013-2015).

## 8. Methodology

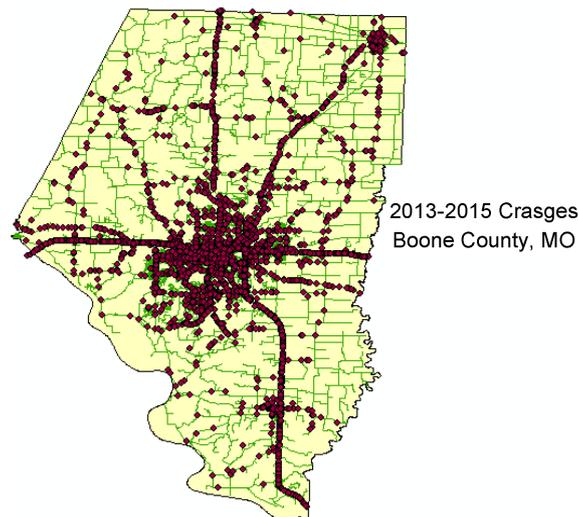
In this paper ArcGIS 10.3.1 is used to compute the Moran's  $I$ , and  $G_i^*$  statistics for crash data in Boone County, Missouri for the aggregated years of 2013-2015 using the following steps:

- Spatially join the attributes of crash incidents to road segments based on their location relationship (*i.e.* latitude/longitude) using functionalities of a GIS that try to parse roads up into consistent analysis units and matching the two features according to their relative spatial locations;
- Build a network of roads from the crash attributed road segments;
- Generate spatial weights matrix for the network arcs;
- Compute the Global Moran's  $I$  available in the ArcMap 10.3.1 Spatial Statistics toolkit;
- Compute the Global (General)  $G_i$  statistic available in the ArcMap 10.3.1 Spatial Statistics toolkit;
- Compute Anselin local Moran's  $I$  available in the ArcMap 10.3.1 Spatial Statistics toolkit;
- Compute the local Getis-Ord local  $G_i^*$  statistic available in the ArcMap 10.3.1 Spatial Statistics toolkit.

Statistically significant high spatial autocorrelation locations will have a high  $z$ -value and be surrounded by other crashes with high  $z$ -values as well (HH). Statistically significant low spatial autocorrelation locations (LL) will be found in cases where a crash point will have a low  $z$ -value and be surrounded by other

**Table 1.** Critical  $z$ -scores,  $p$ -values, and significance levels.

$z$ -score	$p$ -value	Confidence level
$z$ -score $< -1.65$ or $z$ -score $> +1.65$	$<0.10$	90%
$z$ -score $< -1.96$ or $z$ -score $> +1.96$	$<0.05$	95%
$z$ -score $< -2.58$ or $z$ -score $> +2.58$	$<0.01$	99%



**Figure 1.** Boone county, MO road network and distribution of crashes (2013-2015).

crashes with low  $z$ -values as well. If the  $z$ -value of a particular crash location is higher than the mean  $z$ -value of all crashes, then it would be considered high. If the  $z$ -value of a particular crash point is lower than the mean  $z$ -value of all crashes, then it would be considered low. The resultant  $z$ -scores and  $p$ -values indicate whether crashes with either high or low  $z$ -values are clustered. A high  $z$ -score and small  $p$ -value for a crash point indicates a spatial clustering of high values (*i.e.* HH). A low  $z$ -score and small  $p$ -value indicates a spatial clustering of low values (*i.e.* LL). The higher (or lower) the  $z$ -score, the more intense the clustering. A negative  $z$ -score for a crash point indicates an outlier (*i.e.* a dispersed crash). A  $z$ -score near zero indicates no apparent spatial clustering (*i.e.* a random crash). Both the Anselin local Moran's  $I$  and the local  $G_i^*$  statistic can be computed by the ArcMap 10.3.1 Spatial Statistics toolkit [35].

## 9. The New Hybrid Method

Since the Anselin Moran's  $I$  and the  $G_i^*$  can identify relatively different clustering patterns of crashes, therefore this paper introduces a new hybrid method to assess the spatial autocorrelation of crashes and identifies their clustering patterns. The new method combines both Moran's  $I$ , and the  $G_i^*$  statistic into one new hybrid index that can improve the results. Any combination maybe used by the user depending on his/her own interpretation of the results that produces the optimal outcome. For instance, a combination of 30% Moran's  $I$ , and 70%  $G_i^*$  is used in this paper to examine the new spatial clustering patterns of crashes. The hybrid method is applied using the Getis-Ord  $G_i^*$  index available in ArcMap 10.3.1 toolkits for all crashes. The results produced new statistically significant spatial clusters of high spatial autocorrelation values and low spatial autocorrelation values. Using different combination of Moran's, and  $G_i^*$ , such as 50% Moran's +50%  $G_i^*$  could result in different cluster mapping. The user can try different combination, and choose the optimal one that produces the best interpreted results. In addition, the new hybrid method could produce new clusters

if Moran's  $I$  is used in determining the hybrid results instead of the  $G_i^*$  statistic.

## 10. Results

Crashes occurred along the roads in Boone County, Missouri (2013-2015) are analyzed to assess whether they are spatially clustered, dispersed, or random. The Global Moran's  $I$  and the Global (General)  $G_i^*$  for the entire Boone County roads, were first calculated using the ArcMap 10.3.1 Spatial Statistics toolkit. The Global Moran's  $I$ , and the Global  $G_i^*$  statistic,  $z$  scores, and  $p$ -values are reported in **Table 2**.

The results of the analysis are interpreted within the context of the null hypothesis, which states that the crashes occurred in Boone County roads (2013-2015) are randomly distributed in the study area (*i.e.* there is no global spatial autocorrelation exists for the entire area). Since the  $p$ -values in **Table 2** for both Moran's  $I$  and the  $G_i^*$  are smaller than 0.05 (using a confidence level of 95%), then this indicates that the values of Global Moran's  $I$  and the Global  $G_i^*$  are significant for the entire area, and hence, we will reject the null hypothesis, and conclude that it is quite possible that the spatial distribution of the overall Boone County road crashes is the result of clustered spatial processes.

**Table 3** shows the results of the significant high spatial autocorrelation crashes, the significant low spatial autocorrelation crashes, outliers, and the non-significant random crashes of the Boone County roads by Anselin Moran's  $I$  and local  $G_i^*$  statistic respectively. The Moran's  $I$  identified (2411) significant HHs crashes for the Boone County roads, whereas the  $G_i^*$  statistic identified (2916) significant HHs crashes. The Moran's  $I$  identified (3544) significant LLs crashes, whereas the  $G_i^*$  statistic identified (3265) significant LLs crashes. The Moran's  $I$  identified (73) significant HLs and (38) significant LHs, compared to the  $G_i^*$  that identified (0) significant HLs and (0) significant LHs as the  $G_i^*$  does not identify outliers. The Moran's  $I$  identified (820) non-significant random crashes, compared to (705) non-significant random crashes by the  $G_i^*$ . So, it is clear that both Moran's  $I$  and the  $G_i^*$  statistic identify different numbers of clustering patterns.

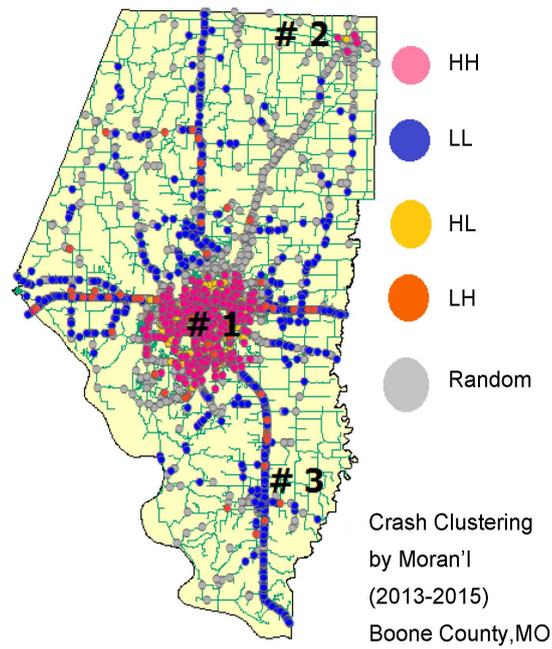
**Figure 2** shows the clustering patterns identified by Anselin local Moran's  $I$  for the Boone County roads. **Figure 3** shows the clustering patterns identified by the local  $G_i^*$  statistic for the Boone County roads. The number and extent of HHs, LLs, and random crashes differ from one method to the other. For example, cluster #1 is identified by Moran's  $I$  as mixed HHs, LLs, HLs, LHs and random crashes while it has been identified as mostly HHs, LLs and random crashes by  $G_i^*$ . Clusters #2 is identified by Moran's  $I$  as mostly random crashes, while it has been identified by  $G_i^*$  as mostly LLs. Cluster #3 is identified by Moran's  $I$  as mixed LLs, HLs, and random crashes while it has been identified as mostly LLs by  $G_i^*$ .

**Table 2.** Global Moran's  $I$  and global  $G_i^*$ .

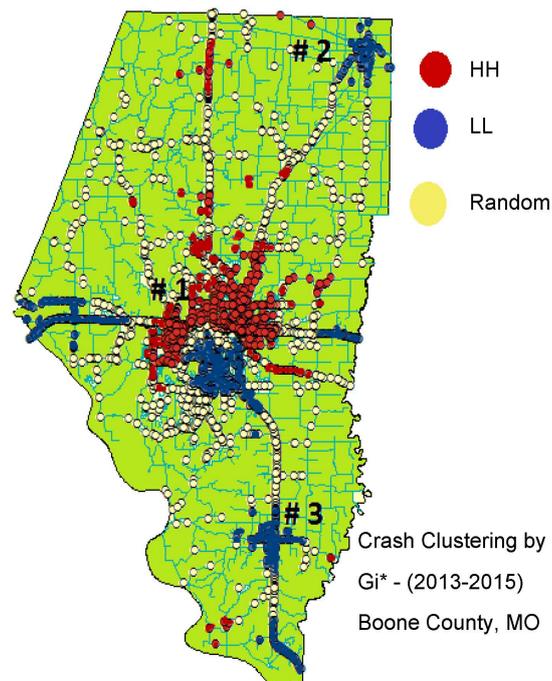
Index type	Index value	$z$ -score	$p$ -value
Global Moran's $I$	0.472	4.856	0.0000
Global $G_i^*$ statistic	0.255	2.817	0.0000

**Table 3.** Number of crash clustering patterns by Moran's  $I$  and  $G_i^*$ .

Index	High-High HH	Low-Low LL	Outliers HL	Outliers LH	Random	Total crashes
Anselin Moran's $I$	2411	3544	73	38	820	6886
$G_i^*$ statistic	2916	3265	0	0	705	6886



**Figure 2.** Crash clustering patterns by Anselin local Moran's  $I$ .



**Figure 3.** Crash clustering patterns by the local  $G_i^*$  statistic.

Using the new hybrid method by combining 30% Moran's  $I$ , and 70%  $G_i^*$  renders another measure of spatial autocorrelation as shown in **Figure 4** for the Boone County roads. The results of the new method produced new statistically significant spatial clusters of high spatial autocorrelation values and low spatial autocorrelation values. From **Figure 4**, it can be seen that cluster #1 near the city of Columbia area is mixed of HHs, LLs, and random crashes compared to mostly HHs and LLs in  $G_i^*$  and mostly outliers, HHs, and LLs in Moran's  $I$ .

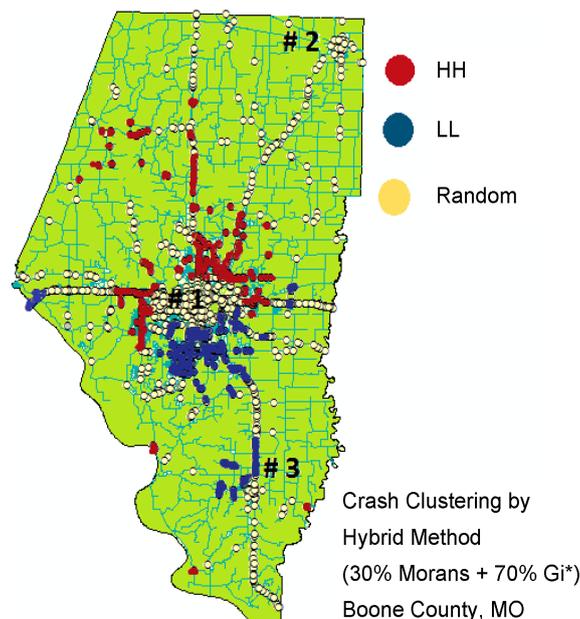
Cluster #2 now presents insignificant random crashes compared to mostly LLs in  $G_i^*$  and mostly outliers in Moran's  $I$ . Clusters #3 becomes mostly insignificant random crashes with some LLs compared to mixed LLs, and outliers in Moran's  $I$  and LLs in  $G_i^*$ . This change makes sense because clusters of LLs and random crashes are more likely happen in big cities (*i.e.* the City of Columbia), and clusters of HHs are more likely happen in the suburban areas of big cities [36]. These results obtained by the new hybrid method show an effective improvement in the clustering patterns of crashes along Boone County roads.

**Table 4** summarizes the HHs, LLs, HLs, LHs, and random crashes identified by Anselin local Moran's  $I$ , the local  $G_i^*$ , and the new hybrid method for Boone County roads.

From **Table 4**, it can be seen that the number of the significant (HHs) and (LLs) identified by the hybrid method have decreased compared to Moran's  $I$  and  $G_i^*$ . However, the number of insignificant random crashes identified by this method has increased compared to the other two methods, which indicates an improvement of the crash clustering patterns.

## 11. Conclusion

In many vehicle crash data, locational relationships among crashes can exist



**Figure 4.** Crash clustering patterns by the new hybrid method.

**Table 4.** Number of crash clustering patterns by Moran's  $I$ ,  $G_i^*$ , and the new hybrid method.

Index	High-High HH	Low-Low LL	Outliers HL	Outliers LH	Random	Total crashes
Anselin Moran's $I$	2411	3544	73	38	820	6886
$G_i^*$ statistic	2916	3265	0	0	705	6886
Hybrid	1841	2417	0	0	2628	6886

given that movement is confined to roadways which are traversed by many users. This phenomenon is termed spatial autocorrelation and if not appropriately accounted for, can lead to incorrect parameter estimates in the modeling process. This paper examined two spatial autocorrelation indices: Moran's  $I$ ; and Getis-Ord  $G_i^*$  statistic to differentiate between spatially clustered, dispersed, or random crash events that occurred in Boone County roads, Missouri in years (2013-2015). Since the two indices can identify relatively different numbers of clustering patterns of crashes, therefore this paper introduced a new hybrid method to assess the spatial autocorrelation of crashes and identify their clustering patterns. The new method combined both Moran's  $I$ , and the  $G_i^*$  statistic into one new hybrid index that can improve the results. Any combination maybe used by the user depending on his/her own interpretation of the results that produces the optimal outcome. A combination of 30% Moran's  $I$ , and 70%  $G_i^*$  was used in this paper to examine the new spatial clustering patterns of crashes. The hybrid method was applied using the Getis-Ord  $G_i^*$  index available in ArcMap 10.3.1 toolkits for all crashes. The results produced new statistically significant spatial clusters of high spatial autocorrelation values and low spatial autocorrelation values that effectively improved the clustering patterns of crashes.

## References

- [1] Black, W. (1992) Network Autocorrelation in Transport Network and Flow Systems. *Geographical Analysis*, **24**, 207-222. <https://doi.org/10.1111/j.1538-4632.1992.tb00262.x>
- [2] Bailey, T.C. and Gatrell, A.C. (1995) Interactive Spatial Data Analysis. Addison Wesley Longman Limited, Harlow.
- [3] Lord, D. and Persaud, N. (2000) Accident Prediction Models with and without Trend: Application of the Generalized Estimating Equations Procedure. *Transportation Research Record*, **1717**, 102-108. <https://doi.org/10.3141/1717-13>
- [4] Wood, G.R. (2002) Generalized Linear Accident Models and Goodness of Fit Testing. *Accident Analysis and Prevention*, **34**, 417-427. [https://doi.org/10.1016/S0001-4575\(01\)00037-9](https://doi.org/10.1016/S0001-4575(01)00037-9)
- [5] MacNab, Y.C. (2004) Bayesian Spatial and Ecological Models for Small-Area Accident and Injury Analysis. *Accident Analysis and Prevention*, **36**, 1028-1091. <https://doi.org/10.1016/j.aap.2002.05.001>
- [6] El-Basyouny, K. and Sayed, T. (2006) Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models. *Transportation Research Record*, **1950**, 9-16. <https://doi.org/10.3141/1950-02>
- [7] Mitra, S. and Washington, S. (2007) On the Nature of Over-Dispersion in Motor

- Vehicle Crash Prediction Models. *Accident Analysis and Prevention*, **39**, 459-468. <https://doi.org/10.1016/j.aap.2006.08.002>
- [8] Aguero-Valverde, J. and Jovanis, P. (2008) Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record*, **2061**, 55-63. <https://doi.org/10.3141/2061-07>
- [9] Washington, P., Karlaftis, G. and Mannering, F. (2010) *Statistical and Econometric Methods for Transportation Data Analysis*. 2nd Edition, Chapman Hall, CRC, Boca Raton.
- [10] Lord, D. and Mannering, F. (2010) The Statistical Analysis of Crash Frequency Data: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, **44**, 291-305. <https://doi.org/10.1016/j.tra.2010.02.001>
- [11] Savolainen, P., Mannering, F., Lord, D. and Quddus, M. (2011) The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, **43**, 1666-1676.
- [12] Mohammadi, M., Samaranayake, V. and Bham, G. (2014) Crash Frequency Modeling Using Negative Binomial Models: An Application of Generalized Estimating Equation to Longitudinal Data. *Accident Analysis and Prevention*, **2**, 52-69.
- [13] El-Basyouny, K. and Sayed, T. (2009) Collision Prediction Models Using Multivariate Poisson-Lognormal Regression. *Accident Analysis and Prevention*, **41**, 820-828.
- [14] Tobler, W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234-240. <https://doi.org/10.2307/143141>
- [15] Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht. <https://doi.org/10.1007/978-94-015-7799-1>
- [16] LeSage, J.P. and Pace, R.K. (2009) *Introduction to Spatial Econometrics*. Chapman and Hall, CRC, New York. <https://doi.org/10.1201/9781420064254>
- [17] Fotheringham, A.S., Brunson, C. and Charlton, M.E. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. Wiley, Chichester.
- [18] Cliff, A.D. and Ord, K. (1981) *Spatial Processes: Models and Applications*. Pion, London.
- [19] Levine, N., Kim, K.E. and Nitz, L.H. (1995) Spatial Analysis of Honolulu Motor Vehicle Crashes: A Spatial Pattern. *Accident Analysis and Prevention*, **27**, 663-674.
- [20] Black, W.R. and Thomas, I. (1998) Accidents on Belgium's Motorways: A Network Autocorrelation Analysis. *Transport Geography*, **6**, 23-31.
- [21] Miaou, S.P., Song, J.J. and Mallick, B.K. (2003) Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. *Accident Analysis and Prevention*, **6**, 33-57.
- [22] Wang, X. and Abdel-Aty, M. (2006) Temporal and Spatial Analyses of Rear-End Crashes at Signalized Intersections. *Accident Analysis and Prevention*, **38**, 1137-1150.
- [23] Aguero-Valverde, J. (2013) Multivariate Spatial Models of Excess Crash Frequency at Area Level: Case of Costa Rica. *Accident Analysis and Prevention*, **59**, 365-373.
- [24] Chiou, Y.C., Fu, C. and Chih-Wei, H. (2014) Incorporating Spatial Dependence in Simultaneously Modeling Crash Frequency and Severity. *Analytic Methods in Accident Research*, **2**, 1-11.
- [25] Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographic Analysis*, **27**, 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- [26] Anselin, L. (1992) *Space Stat: A Program for the Statistical Analysis of Spatial Data*.

National Center for Geographic Information and Analysis, University of California, Santa Barbara.

- [27] Goodchild, M.F. (1987) Spatial Autocorrelation. *Concepts and Techniques in Modern Geography*, **48**, 56-63.
- [28] Griffith, D.A. (1987) Spatial Autocorrelation: A Primer. Resource Publications in Geography. The Association of American Geographers, Washington DC.
- [29] Cliff, A.D. and Ord, K. (1975) The Choice of a Test for Spatial Autocorrelation. Pion, London.
- [30] Getis, A. and Ord, J.K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, **24**, 189-206.  
<https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- [31] Getis, A. and Ord, K. (1996) Local Spatial Statistics: An Overview. Geo Information International, Cambridge.
- [32] Fischer, M.M. and Wang, J. (2011) Spatial Data Analysis: Models, Methods, and Techniques. Springer, New York. <https://doi.org/10.1007/978-3-642-21720-3>
- [33] Berglund, S. and Karlstrom, A. (1999) Identifying Local Spatial Association in Flow Data Geographical Systems. *Transportation Geography*, **1**, 219-236.
- [34] ESRI (2016) ArcGIS Resources Center.  
[http://resources.esri.com/help/10.1/ArcGISEngine/java/Gp\\_ToolRef/Spatial\\_Statistics\\_tools/how\\_hot\\_spot\\_analysis\\_colon\\_getis\\_ord\\_gi\\_star\\_spatial\\_statistics\\_works.htm](http://resources.esri.com/help/10.1/ArcGISEngine/java/Gp_ToolRef/Spatial_Statistics_tools/how_hot_spot_analysis_colon_getis_ord_gi_star_spatial_statistics_works.htm)
- [35] ESRI (2016a) ArcGIS Resources Center.  
<https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/an-overview-of-the-spatial-statistics-toolbox.htm>
- [36] Myers, R., Branas, C., French, C., Nance, L., Kallan, J., Wiebe, J. and Carr, G. (2013) Safety in Numbers: Are Major Cities the Safest Places in the United States? *Annals of Emergency Medicine*, **62**, 408-418.



**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.  
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)  
Providing 24-hour high-quality service  
User-friendly online submission system  
Fair and swift peer-review system  
Efficient typesetting and proofreading procedure  
Display of the result of downloads and visits, as well as the number of cited articles  
Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [ojce@scirp.org](mailto:ojce@scirp.org)