Chapter 1

Application of Spatial Statistics in Transportation Engineering

Uday R.R. Manepalli and Ghulam H. Bham

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/65051

Abstract

"Everything is related to everything else, but near things are more related than distant things" is the first law of geography. It can be hypothesized that spatially, occurrence of a crash can exhibit similarities. To identify spatial patterns of crashes, this chapter presents spatial autocorrelation techniques such as Moran's I and the Getis-Ord G_i*statistics; spatial interpolation such as kriging; and nonparametric probability density function and kernel density (K). The aim of this chapter is to provide application of spatial statistics in transportation engineering specifically to identify crash concentrations and patterns of clusters in a study area.

Keywords: The Getis-Ord Gi* statistics, Kernel-Density function, kriging, Moran's I, spatial autocorrelation, highway safety, crash

1. Introduction

In this chapter, spatial data analysis and its application in the field of transportation engineering specifically for crash data analysis is presented. Analysis of spatial data extends the representation of geographic space from discrete sets of points, lines and polygonal features to mapping surfaces characterizing a continuous space. Statistics using spatial relationships for the data mapped investigates the similarities among them. The first law of geography states "Everything is related to everything else, but near things are more related than distant things" [1]. This principle has been used in various fields such as criminology, economics, transportation, etc. to identify relationships within a geographic space. In order to perform spatial data analysis, geographical locations and attributes of an object (point, line or polygon, area) are required. Spatial data analysis can answer questions such as how spatial data

distributions can be compared, and how future distributions based on current spatial data can be forecasted. In the past, different statistical techniques have been used with spatial data and they can be broadly classified as:

Spatial autocorrelation (SA): The basic principle of SA is similar to the first law of geography. SA is defined as the correlation of a variable with itself in space. SA measures the strength of autocorrelation and the assumption of independence. A variable is said to be spatially autocorrelated if there are systematic patterns in its spatial distribution. SA is positive if nearby areas (regions) are alike. Negative autocorrelation applies to neighboring areas that are unlike, and SA is not exhibited by random patterns.

SA is measured using spatial autocorrelation indices. Some of the commonly used indices are Moran's *I* and Geary's *C*. These indices are often referred to as global indices. They measure overall degree of spatial autocorrelation in a data set. For specific disaggregated estimates, local indices are used. Some of the local indices are local Moran's I [2], local Geary's C [3], and the Getis-Ord G_i^* statistics [4, 5].

Spatial interpolation: It is defined as the process of using data for locations to predict ones that are not sampled. Inverse distance weighting and kriging [6] are commonly used in spatial interpolation techniques. The latter considers a spatial lag relationship that has both systematic and random components.

Spatial regression: Due to spatial autocorrelation, ordinary regression models cannot be used. To identify the underlying effects between the dependent variable and a spatial lag of itself, geographically weighted regression (GWR) [7, 8] is used.

Additional analysis techniques include nonparametric analysis such as kernel density estimation [9], as used in point pattern analysis to identify the first-order effects, i.e., measure the variation in mean value.

This chapter is organized as follows: first, the fundamental concepts for several spatial statistics measures are explained, and it is followed by case studies related to the fundamental concepts. The chapter ends with conclusions and recommendations.

2. Fundamental concepts

This section presents the concepts related to spatial autocorrelation, i.e., Moran's I and the Getis-Ord G_i^* statistics; spatial interpolation, i.e., kriging; and nonparametric analysis, i.e., kernel density estimation. They are presented to show their use in transportation safety.

2.1. Moran's I

It is one of the oldest indicators of SA [2]. SA compares the value of a variable in one location with its value at other locations. Similar to a correlation coefficient, SA varies between −1.0 and +1.0. A positive correlation indicates clustering (i.e., higher crash concentrations in highway

http://dx.doi.org/10.5772/65051

safety), whereas negative correlation indicates dispersion or low crash concentration. Moran's I is expressed as

$$Moran's\, I = \frac{n\sum_i \sum_j w_{ij}\left(Y_i - \bar{Y}\right)\left(Y_j - \bar{Y}\right)}{\left(\sum_{i \neq j} w_{ij}\right)\sum_i \left(Y_i - \bar{Y}\right)} \tag{1}$$

The term w_{ij} represents a contiguity matrix. If location j is adjacent to location i, the interaction receives a weight of 1; otherwise, zero. The term w_{ij} compares the sum of the cross products of values at different locations weighted by the inverse of the distance between the locations.

The significance of Moran's I can be evaluated by a Z value as

$$Z(I) = \frac{I - E(I)}{S(I)} \tag{2}$$

where $E(I)$, the expected value of Moran's I, can be computed as

$$E(I) = \frac{-1}{n-1} \tag{3}$$

$S(I)$, the standard deviation, is computed as

$$S(I) = \sqrt{\frac{n^2(n-1)s_1 - n(n-1)s_2 - 2s_0^2}{(n+1)(n-1)s_0^2}} \tag{4}$$

where

$$s_0 = \sum_{i \neq j} w_{ij} \tag{5}$$

$$s_1 = \frac{1}{2}\sum_{i \neq j}\left(w_{ij} + w_{ji}\right)^2 \tag{6}$$

$$s_2 = \sum_k \left(\sum_j w_{jk} + \sum_i w_{ik}\right)^2 \tag{7}$$

In the foregoing formula, *i*, *j*, and *k* represent the location of crashes. At a level of 5%, values of Z greater than +1.96 and less than −1.96 indicate significant positive and negative SA, respectively.

2.2. The Getis-Ord G_i^* statistics

G-statistics, developed by Getis and Ord, analyzes the evidence of spatial patterns and represents a global SA index [4, 5]. The G_i^* (pronounced as G-i-star) statistics, however, is a local SA index. It is more suitable for discerning clusters of high or low concentration. A simple form of the G_i^* statistics is [10]

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j} \tag{8}$$

where G_i^* is the SA statistics of an event *i* over *n* events (e.g., crashes) [11]. The term x_j characterizes the magnitude of the variable x at event j over all n, and in highway safety, an index such as crash severity index (CSI) value determined at a particular location can be used. The G_i^* statistics can be observed from the underlying distribution of the variable *x* [11]. The threshold distance (the proximity of one crash to another) can be set to zero to indicate that all features were considered neighbors of all other features.

Further, the standardized G_i^* is essentially a Z value as well and can be associated with statistical significance

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij} x_{ij} - \bar{X} \sum_{j=1}^{n} w_{ij}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij} \right)^2 \right]}{n-1}}} \tag{9}$$

where

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X} \right)^2} \tag{10}$$

Positive and negative G_i^* statistics values correspond to clusters of crashes with high- and low-value events, respectively. A G_i^* statistics close to zero implies a random distribution of events.

http://dx.doi.org/10.5772/65051

2.3. Kriging

Kriging, a spatial prediction methodology based on spatial interpolation, was first developed by Matheron [12] based on the work of Krige [6] to predict ore reserves. Kriging has been applied widely in air quality analysis, geology, hydrology, ecology, etc. The major application of this technique is to predict values at unmeasured locations while assessing the errors of these predictions [13]. It relies on the notion that unobserved factors are autocorrelated over space, and the levels of autocorrelation decreases with distance. A trend estimate, μ(s), is determined which can be defined as [13]

$$Z_i(s) = \mu_i(s) + \varepsilon_i(s) \tag{11}$$

where $Z_i(s)$ is the variable of interest and s indicates the location of the site "i." It is composed of a deterministic trend μ_i(s) and a random error term ε_i (s). The random errors are autocorrelated over space. The expected value of Z(s) results in different types of kriging, namely simple, ordinary, universal, intrinsic kriging, and so on. However, universal kriging is preferred to other kriging methods as the trends depend on explanatory variables and (unknown) regression coefficients. The correlation between Z(s) and Z(s + h) does not depend on actual locations, but only distance "h" between the two sites. This is possible by assuming weak stationarity in all three cases. This indicates a constant variance of 2γ (h) for any s and h, where γ (h) can be expressed as

$$\gamma(h) = \frac{1}{2} var\left[Z(s+h) - Z(s)\right] \tag{12}$$

where $var[Z(s + h) - Z(s)]$ is the variance between s and s + h. When 2γ (h) is plotted versus distance, the plot is called a semivariogram. A semivariogram depicts the spatial autocorrelation of the measured sample points. One of the major steps is to select an appropriate semivariogram model that best fits the relationship between γ and h. There are three models that best explain the relationship, i.e., exponential, spherical, and Gaussian. In this chapter, only spherical model is presented, and the specifications are

$$\gamma(h) = \begin{cases} c_0 + c_1\left[1.5\frac{h}{a} 0.5\left(\frac{h}{a}\right)^3\right] if\ 0 < h < a \\ c_0 + c_1\ if\ h > a \\ 0\ otherwise \end{cases} \tag{13}$$

The different models (spherical, exponential, and Gaussian) rely on parameters that describe their shape and level of spatial autocorrelation in the data. c_0 in the above equation is called the nugget effect and reflects discontinuity in the variogram origin as caused by factors such as sampling error and short-scale variability. The origin of the term nugget originates from gold deposits, as gold commonly occurs as nuggets of pure metal that are much smaller than the size of a sample. It can result in strong variability in the sample when physically close, and therefore discontinuity of the variogram at the origin can be observed [14].

The rate of variogram reflects the degree of dissimilarity of more distant samples. At large distances, a variogram can increase indefinitely if the variability of the phenomenon has no limit. However, if the variogram stabilizes at a value, called the sill, it indicates that beyond a certain distance Z(s) and Z(s+h) are uncorrelated [14]. This distance is called the range denoted by a. It determines the threshold distance at which γ(h) stabilizes [13]. $c_0 + c_1$ is the maximum γ (h) value, called sill, and c_1 is referred to as partial sill [15]. **Figure 1** illustrates a semivariogram.

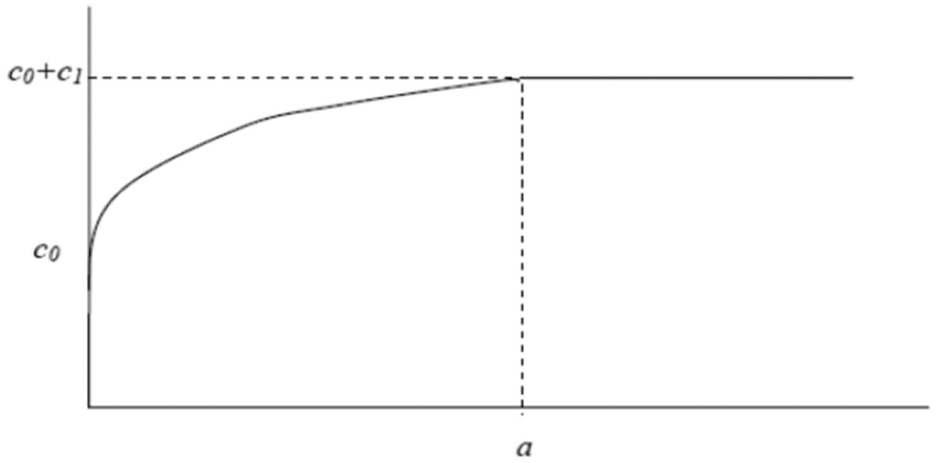

Figure 1. Illustration of a semivariogram.

2.4. Kernel density estimation

The kernel density method is a nonparametric method that uses a density estimation technique. It enables the observer to evaluate the local probability of an occurrence and degree of danger in a zone. For a given set of observations from an unknown probability density function, the kernel estimator can be defined as

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{14}$$

where h is called the smoothing parameter or bandwidth, K is called the kernel, and $\hat{f}$ is the estimator of the probability density function f. Thus, the kernel estimator depends on band-

width (h) and kernel density (K). For a given kernel, K, the kernel estimator critically depends on the choice of the smoothing parameter h. An appropriate choice of the smoothing parameter should be determined by the purpose of the estimate.

3. Case studies

The different case studies presented are related to the fields of crash data analysis, safety, and forecasting of traffic volume.

3.1. Spatial autocorrelation

A study was conducted to identify crash contributing factors on highway networks of Arkansas using a sample of crash data. In this study, spatial autocorrelation indices i.e., Moran 's I and Getis–Ord G_i^* statistics, and multinomial logistic regression were used. Autocorrelation was determined at different levels, and then multinomial logistic regression was used to identify crash-contributing factors in case a crash occurs. Based on the autocorrelation indices, the state's 75 counties were divided into zones. Further, to identify the crash contributing factors, a sample of data from the counties were compared to the statewide data.

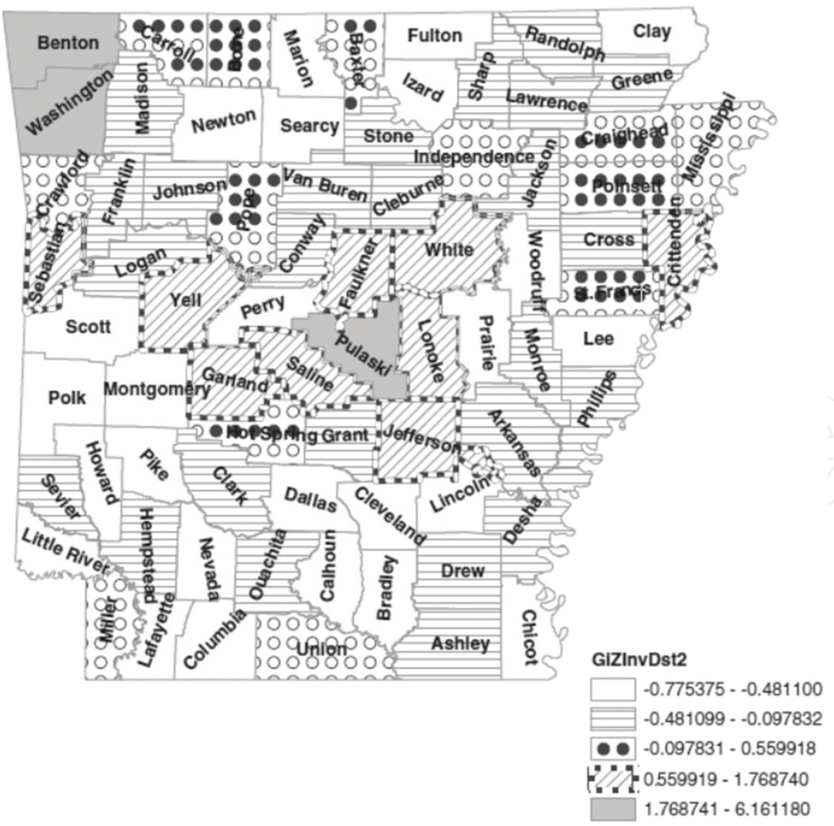

Figure 2. Counties categorized by G_i^* statistics [16].

Category	Number of counties	Counties with highest CSI[a]	CSI[b]	Total CSI[c]	CSI ratio[d]	Crash freq. ratio	G_i* statistics: range of Z values
(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
First	3	Pulaski	137,627	276,755	.50	.51	1.7678741, 6.161180
Second	9	Garland	52,189	324,668	.16	.27	0.559918, 1.768740
Third	13	Craighead	28,676	298,379	.10	.17	−0.097831, 0.559918
Fourth	25	Madison, Cleburne, Logan	45,707	273,196	.17	.16	−0.481099, −0.097832
Fifth	25	Chicot, Montgomery, Polk, Perry, Little River, Clay, Colombia	53,477	133,861	.40	.57	−0.775375, −0.481100
Total	75	13[e]	317,676	1,306,859	.24	.34	–

Note: "–" not applicable.
[a]Satisfies the condition of minimum sample size of 2000 in terms of crash frequency.
[b]CSI computed for county/counties in Column C.
[c]CSI computed for counties in Column B.
[d]Ratio of CSI values in Columns D and E.
[e]Total number of counties in Column C.

Table 1. Results presented by category, highest CSI in each category, and ratios of crash data [16].

Crash data from 2004 to 2006 were used for the study. Crashes were categorized into five levels of crash injury severity from S1 to S5, where S1 indicated fatal injury; S2, major injury; S3, minor injury; S4, complain of pain; and S5, property damage only (PDO), based on the KABCO scale. Further, crash frequency (CF), i.e., the summation of crash count at various levels of crash injury severity, and crash severity index (CSI) [16] which combines various effects of different levels of crash injury severity into an index were determined. The first step of the analysis was to determine whether spatial autocorrelation exists. Moran's I was used which identified that SA exists for the crash data used. The crash injury severity levels showed significance at various levels.

G_i* was used to discern cluster structures of high or low concentration. Z-values were also computed and the categorization of counties based on the z-values of the G_i* statistic was determined. This categorization can be based on six different classification schemes: equal interval, defined interval, quartile, natural breaks, geometric interval, and standard deviation. The natural breaks scheme was best suited for the study [17]. In the natural breaks scheme, the classes are based on inherent categorizing in the data. The classes identify the break points that best groups similar values and maximizes the differences between these classes.

http://dx.doi.org/10.5772/65051

In the study, Jenks' algorithm was used to categorize the natural breaks [17]. Jenks' algorithm is commonly used to classify the data in a choropleth map, a type of thematic map that uses shading to represent classes of a feature associated with specific areas (e.g., a population density map). Jenks' algorithm generates a series of values that best represent the actual breaks in the data as opposed to some arbitrary classification scheme. Thus, it preserves the true clustering of data values. As a result, the algorithm creates "k" classes as the variance within categories is minimized. The state of Arkansas was categorized into five categories. **Figure 2** shows these categories, and **Table 1** presents the results by category, and shows the number of counties in each category. From each category, a county or a set of counties starting with the highest CSI was selected as a data sample. The highest CSI was used as the criterion because it provided the greatest variability in the crash data.

Figure 3 presents graphically the higher and lower Z values of G_i^* for the five categories. The G_i^* Z values indicate the clustering of the attributes in the study area. The first category had higher positive Z values compared to lower Z values, indicating that the value of CSI is not random for those counties. The trend from **Figure 3** indicates that the randomness increases over the categories. This trend is similar to the trend for identification of crash casual factors identified for each category, presented next.

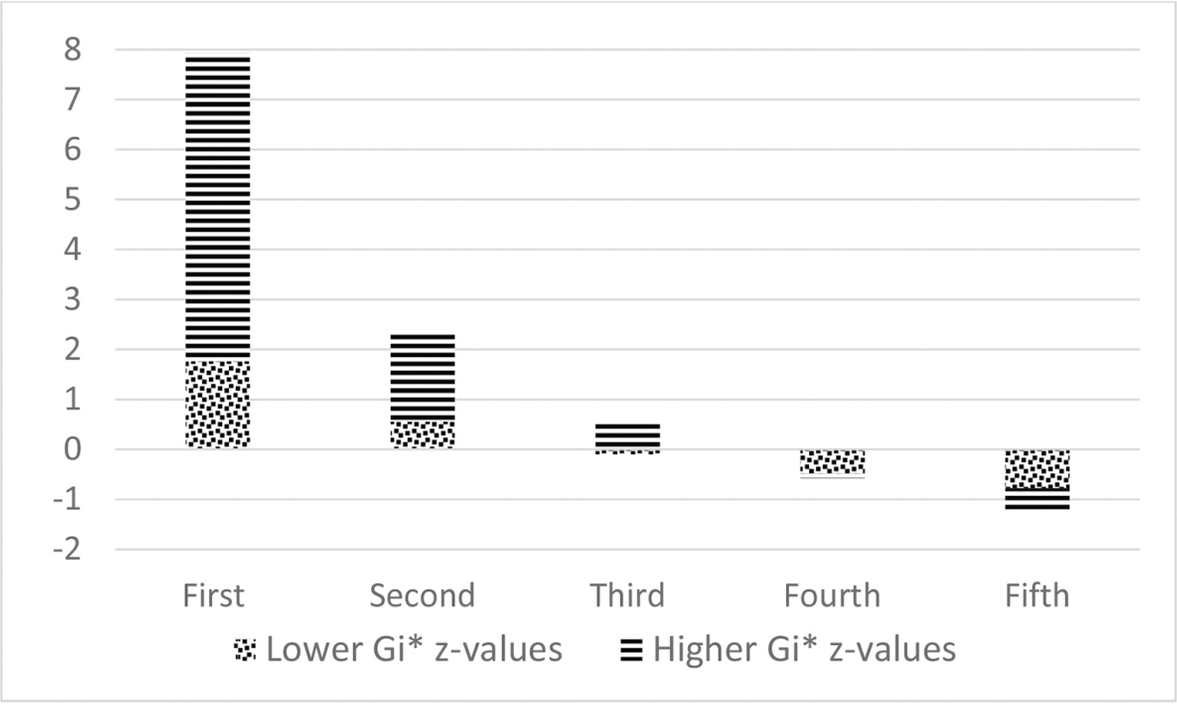

Figure 3. Comparison of G_i^* statistics values across five categories.

SA indices, however, do not explain why locations that indicate a cluster of crashes have a higher incidence of crashes compared with other locations; therefore, SA methods cannot identify crash causality factors [16]. Multinomial logistic regression (MLR) was used to identify the crash-contributing factors. The main reasons for choosing the MLR models were:

- Given that a crash has occurred, the factors that increase the chances of a fatal or a serious injury crash were considered and computed by using the odds ratio as a result of the MLR models.
- Factors that supplement the need for attainment of zero fatalities given that crashes occur because of other factors, including human factors, were identified.
- Factors for all levels of crash severity were identified, and common factors were selected as an alternate solution. However, this procedure is cumbersome when the desired results can be achieved in one model.
- A minimum sample size of 2000 is required to implement MLR models [18]. Therefore, with a decent sample size, these models can predict accurately. Details can be found elsewhere [18, 19].

Selected independent variables in the data were checked by using a variance inflation factor (VIF) to ensure that multicollinearity is not an issue. The variance inflation factor was found to be less than 10 for all of the variables; hence, multicollinearity was not observed. Variables selected for model development depended on the quality of the data. Only certain factors were retained for analysis since some factors had missing values. When more than 10% of the values were missing, that factor was not considered. For the factors presented in **Table 2**, no more than 1% of the values were missing. Mallows' Cp was used to retain the variables; a smaller value of Cp indicated a better model [19].

Abbreviations	Variables	Levels
ATM	Atmospheric conditions	Clear, rain
LGT	Light conditions	Dark, daylight
RSUR	Roadway surface	Dry, wet
RU	Roadway type	Rural, urban
RALI	Roadway alignment	Curve, straight
RPRO	Roadway profile	Grade, level
TOH	Roadway classification	Divided, undivided
TOC	Collision types	Angle, head-on, rear-end, sideswipe-same-direction (SSSD), single vehicle crashes (SVC), sideswipe–opposite direction (SWOD)
WK	Days of the week	Weekdays (M-F), weekends (Sat, Sun)
DUI	Driving under the influence	Yes, no
AADT	Annual average daily traffic	<20,000, 20,000–40,000, 40,000–60,000, 60,000–80,000, 80,000–100,000, 100,000–120,000

Table 2. List of independent variables [16].

Table 3 indicates that during darkness, fatal crashes were more likely to occur than PDO crashes, and the odds ratio increased by a factor of 1.28 if other variables remained constant. Similarly, the relative risk of fatal crashes was greater than the PDO crashes in rural areas and on curved roads.

Variables	Contributing factors	Estimate	Standard error	Chi-square value	p-Value	Odds ratio
Fatal vs property damage crashes						
LGT	Dark vs daylight	0.25	0.12	3.92	0.0476	1.28
RU	Rural vs urban	0.71	0.13	29.31	<.0001	2.04
RALI	Curve vs straight	0.34	0.13	6.25	0.0124	1.40
DUI	No vs yes	−1.17	0.13	86.71	<.0001	0.31
Major injuries vs property damage crashes						
Intercept		−2.49	0.18	185.59	<.0001	
RU	Rural vs urban	0.43	0.08	29.24	<.0001	1.54
RALI	Curve vs straight	0.29	0.08	13.05	0.0003	1.33
TOC	Angle vs SSSD	−0.39	0.17	5.36	0.0206	0.68
TOC	Head-on vs SSSD	1.86	0.23	64.43	<.0001	6.41
TOC	Rear-end vs SSSD	−0.58	0.15	15.34	<.0001	0.56
TOC	SVC vs SSSD	0.69	0.13	26.32	<.0001	2.00
TOC	SWOD vs SSSD	−1.50	0.25	36.62	<.0001	0.22
DUI	No vs yes	−0.77	0.08	91.27	<.0001	0.46

Table 3. Sample MLR results [16].

3.2. Kriging

Kriging models were used in a study to forecast Annual Average Daily Traffic (AADT) [13]. AADT data for 27,738 sites from 1999 to 2005 were used to forecast AADT values for 2006. The initial interpolation was made for 27,738 sites and later expanded throughout the network.

The study assumed that the AADT values would be similar to values at nearby sites. Network details were obtained based on the data provided by the Texas Department of Transportation. Two functional classes were identified Class 1 (interstate) and Class 2 (other principal arterials). Each site was then matched to attributes of the closest road section using functional class. Traffic counts on segments of the same class were spatially interpolated using kriging. For each functional class, a semivariogram was estimated. For Class 1 segments, the estimated range value, a, was 1.248; nugget value, c_0, was 2.33×10^7; and partial sill, c_1, was 1.62×10^7. For Class 2 segments, a equaled 0.158, c_0 9.86×10^8, and c_1 $2.82 \times$

10^9. It was found that Class 1 scatter was higher for a given distance compared to Class 2. The larger values of sill and nugget for Class 1 indicated spatial autocorrelation for AADT that is distance dependent and sensitive. Class 1 roads had many access points which might have led to fluctuations in AADT over space. For Class 2 roads, the flow changes appeared continuous over time.

The study concluded that more data helped improve the forecast, temporal dependence was stronger than spatial dependence, and kriging methods provided reliable results in uncounted/unsampled locations.

3.3. Kernel density estimation

A study examined the spatial patterns of pedestrian crashes to identify high crash zones. The study evaluated methods to rank these zones using a Geographic Information System (GIS) [20]. To identify these high crash zones, crash concentration maps were developed. The crash concentration maps based on density values used simple and kernel density methods. Five years of crash data (1998–2002) for Las Vegas metropolitan was used in the study. For this chapter, the scope is limited to identifying the crash concentrations using the Kernel density method.

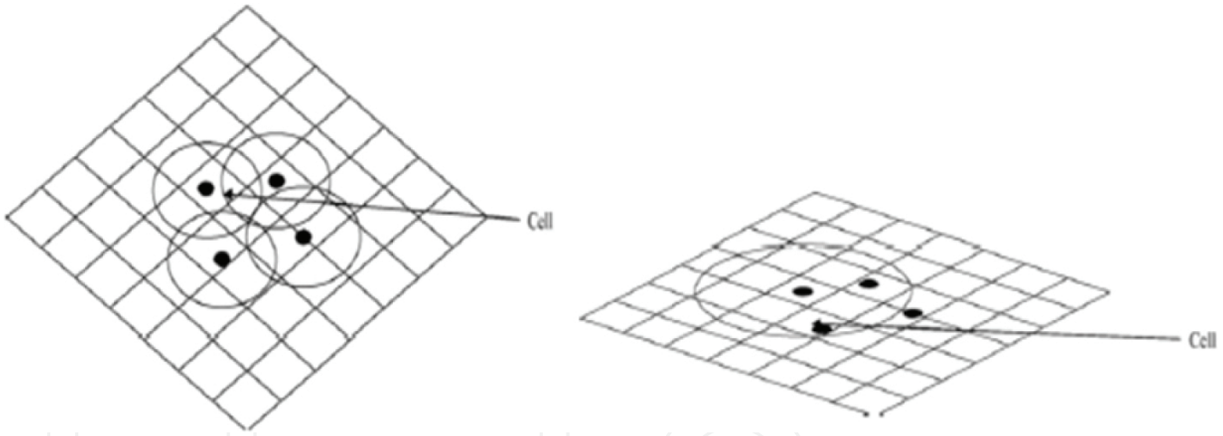

Figure 4. Illustration between kernel density (left) and simple density (right) methods [20].

The researchers identified the high crash zones using a three-step methodology: (1) geocode pedestrian crash data; (2) create crash concentration maps; and (3) identify zones, their shapes, and sizes. The geocoding of the crash data was performed using the "address match" feature. One of the major issues with point data, similar to crashes, is that when a map is plotted it may not present clusters of crash concentrations with more than just a few crashes. Developing maps with crash concentrations is therefore helpful.

Figure 4 illustrates the difference between simple density and kernel density methods, i.e., drawing a circular area of search around each crash to calculate the kernel values (K). The value of the surface is highest at the crash location and diminishes to zero at the radius of the circle. Thus, as a result, a smooth density surface is created.

http://dx.doi.org/10.5772/65051

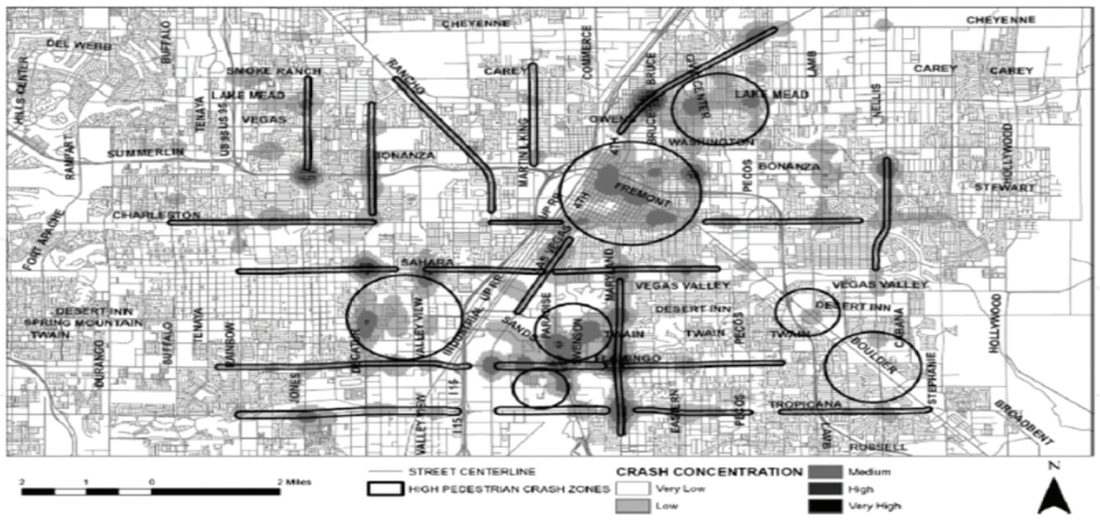

Figure 5. Las Vegas, pedestrian high crash zones [20].

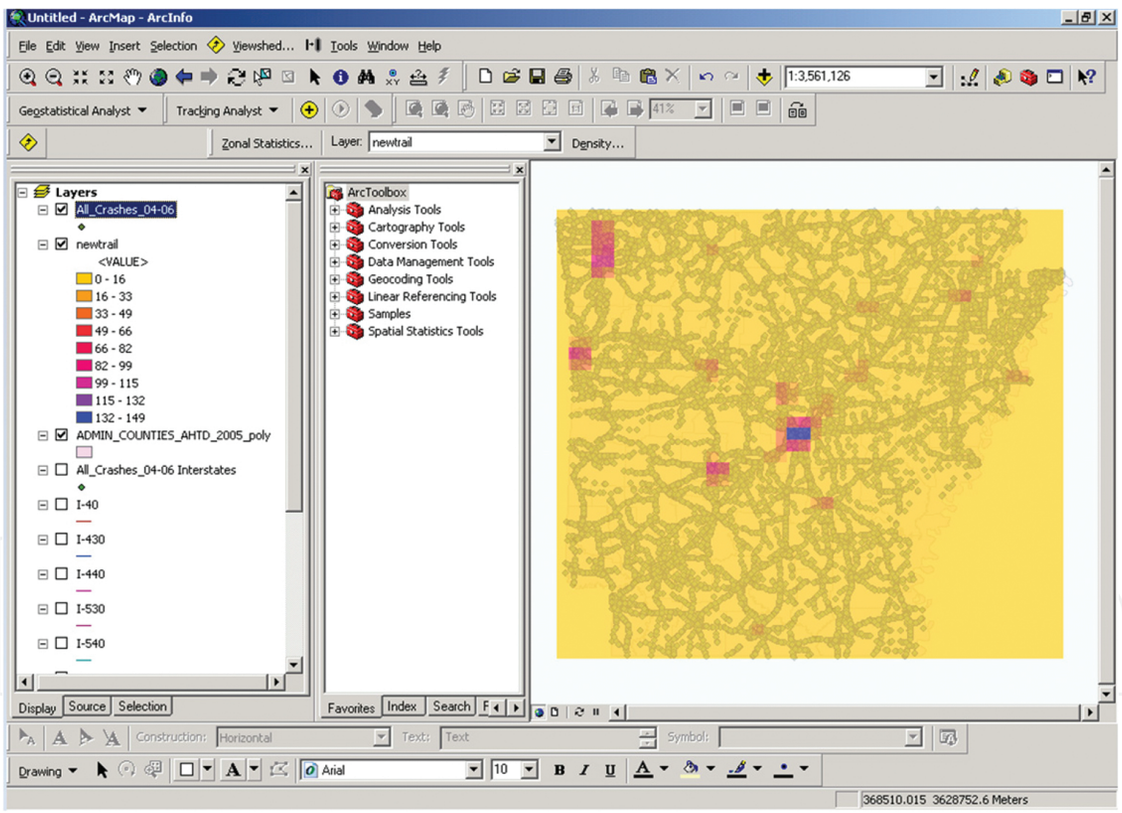

Figure 6. Identifying crash clusters using kernel density, application to Arkansas crash data.

Once the kernel density was identified, the zones of crash concentration were determined. These zones were either linear or circular. When dense clusters of crashes were observed along the route, then the zone identified was linear. When dense area was isolated at an intersection or was not linear in shape, then the zone identified was circular. When several linear zones

were closely spaced and demographic, traffic, and geometric characteristics were similar, the researchers classified it as a circular zone. The study identified 29 pedestrian high crash zones, 22 linear zones, and 7 circular zones. **Figure 5** presents the 29 different crash zones.

The study concluded that the GIS-based methodology helps quantify the concentration of crashes and thus reduce the degree of subjectivity involved in identifying high crash zones. This approach is practical and easy to implement as most agencies collect crash, census, and traffic data sets in a GIS format.

In another study, undergraduate civil engineering students were exposed to the application of GIS in a mandatory course in transportation engineering [21]. The GIS tutorial was implemented in a laboratory environment developed as a self-guided activity supported by a web-based learning system. One of the tasks was to create a crash concentration map based on the data provided for a state highway network using the kernel density method. **Figure 6** presents a sample output from one of the students in the laboratory. The kernel density method is therefore easy to implement, and students in a laboratory when provided with a self-guided tutorial can implement it. The method when based in a GIS can also serve as a powerful tool to visualize crash clusters in a network.

4. Conclusions and recommendations

This chapter summarizes the fundamental concepts associated with spatial analysis of data in transportation engineering. Further, the application of these concepts is presented with interesting case studies from the literature specifically to improve highway safety and forecast of traffic volume for planning-level applications.

In various case studies presented in this chapter, a different spatial statistics model has been used. Depending on the type of problem, availability of data, expected outcomes, and ingenuity have led researchers to different techniques in spatial data analysis. These techniques help improve understanding of the phenomenon and thereby the solution to the problem. The future of spatial statistics lies in creative thinking and seeking solutions in more than one way. In terms of problem solving, solutions can be derived both objectively and subjectively. The more one experiments with the available techniques, the closer one can reach an ideal solution.

Author details

Uday R.R. Manepalli[1] and Ghulam H. Bham[2*]

*Address all correspondence to: ghbham@gmail.com

1 Agile Assets Inc., Austin, TX, USA

2 University of Alaska Anchorage, Anchorage, AK, USA

References

[1] Tobler W. A computer movie simulating urban growth in the Detroit region. Economic Geography, 1970; V. 46, 9: 234-240.

[2] Moran PAP. Notes on continuous stochastic phenomena. Biometrika, 1950; 37: 17-33.

[3] Geary R. The contiguity ratio and statistical mapping. The Incorporated Statistician, 1954; 5: 115-45.

[4] Getis A. Ord JK. The analysis of spatial association by use of distance statistics. Geographic Analysis, 1992; Vol. 24, No. 3: 189-206.

[5] Ord JK. Getis A. Local spatial autocorrelation statistics: distributional issues and an application. Geographic Analysis, 1995; Vol. 27, No. 4: 286-306.

[6] Krige DG. A Statistical Approach to Some Mine Valuations and Allied Problems at The Witwatersrand [Master's thesis]. Witwatersrand: University of Witwatersrand; 1951.

[7] Fotheringham A. S., Brunsdon C., Charlton M. E. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Chichester, Wiley; 2002.

[8] Silverman B.W. Density Estimation for Statistics and Data Analysis, London, Chapman and Hall; 1985.

[9] Diggle, P. J., A kernel method for smoothing point process data. Journal of the Royal Statistical Society C, 1985; 34, 138-147.

[10] McGuigan DRD. The use of relationships between road accidents and traffic flow in "black-spot" identification. Traffic Engineering and Control, 1981; 448–453.

[11] Songchitruksa P, Zeng X. Getis-Ord Spatial Statistics for Identifying Hot Spots Using Incident Management Data. In Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., 2010; 2165: 42-51.

[12] Matheron G. Principles of Geostatistics. Economic Geology, 1963; 58: 1246-1266.

[13] Wang X, Kockelman KM. Forecasting Network Data: Spatial Interpolation of Traffic Counts using Texas Data. In Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C., 2009; 2105: 100-108.

[14] Chiles J-P., Delfiner P. Geostatistics: Modeling Spatial Uncertainty, 2nd ed., New Jersey, Wiley; 2012.

[15] Cressie N. Statistics for Spatial Data, New York, Wiley Interscience; 1993.

[16] Manepalli URR., Bham GH. Identification of Crash-Contributing Factors: Effects of Spatial Autocorrelation and Sample Data Size, Transportation Research Record: Journal

of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., 2013; 2386: 179–188.

[17] Geography lecture notes [Internet]. Available from: http://go.owu.edu/~jbkrygie/krygier_html/geog_353/geog_353_lo/geog_353_lo07.html [Accessed: 2016-03-10].

[18] Ye F., Lord D. Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit. In Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., 2011; 2241: 51-58.

[19] Bham GH., Javvadi BS, Manepalli, URR. Multinomial logistic regression model for single-vehicle and multivehicle collisions on Urban U.S. Highways in Arkansas. Journal of Transportation Engineering, 2012; 138, No. 6: 786-797.

[20] Pulugurtha SS., Krishnakumar VK., Nambisan SS. New Methods to Identify and Rank High Pedestrian Crash Zones: An Illustration. Accident Analysis and Prevention, 2007; Vol. 39, No. 4: 800-811.

[21] Bham GH, Cernusca D, Luna R, Manepalli, URR. Longitudinal Evaluation of a GIS Laboratory in a Transportation Engineering Course, ASCE Journal of Professional Issues in Engineering Education and Practice, 2011; 137: 258-266.