International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Identification of Traffic Accident Hotspots using Geographical Information System (GIS)

S.Lakshmi, Ishwarya Srikanth, M. Arockiasamy

Abstract: ***Limiting the number andseverity of traffic accidents is one of the major goals of road traffic safety management.The alarming rate of road accidents globally emphasizes the importance of an effective traffic safety management system. Identification of accident hotspots is the first step towards implementation of efficient traffic safety management.Until the arrival of Geographical Information System (GIS),traffic accident analyses have been performed based ontraditional statistical methods alone. The advent of GIS-based techniques has led toimproved traffic accident analysis by employing spatial statistics,enabling engineers and researchers to account for variation in the spatial characteristics of hotspot locations in the analysis. This paper discusses the different spatial and statistical methods that are employedintraffic accident hotspots identification. An example application of Planar Kernel Density Estimation (PKDE)for hotspot identification is presented based on crash data for Des Moines city of Iowa state. The effect of varying bandwidths in creating density mapsis investigated and the optimum bandwidth to obtain distinct hotspots is identified as 500 m for the chosen study area.The paper also discusses the scope for future research in traffic accident hotspot analysis.***

Keywords: Accident analysis, GIS, Hotspots, Spatial methods,Statistical tools.

I. INTRODUCTION

Traffic accidents pose a severe threat to human life, majority of which happened in developing countries owing to inadequate road safety measures [1].Since traffic crashes also cause economic loss, it affects Gross Domestic Product of a country. Hence, efforts should be taken to minimize the accidents. Hotspots, also known as “black spots”, are the locations on a section of a highway havingcrash frequency remarkably greater than anticipated at some threshold level of significance [2].

Identifying hotspots help traffic authorities to alleviate crashes by finding appropriate solution. For example, a location encountering four accidents in a year in Turkey is considered as hotspot [3].

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Dr. Lakshmi Srikanth*, Professor, Department of Civil Engineering, CMR Institute of Technology, Bengaluru, India (E-mail: lakshmi.s@cmrit.ac.in)
Ishwarya Srikanth, Graduate Research andTeaching Assistant, Department of Civil, Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton. FL 33431, USA. (E-mail:isrikanth2016@fau.edu)
Dr. Madasamy Arockiasamy, P.E., P. Eng., Fellow ASCE, Professor, Department of Civil, Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton, FL 33431, USA (E-mail: arockias@fau.edu)

There is plethora of traditional statistical models that are in practice for hotspot detection. Thomas (1996)employed conventional statistical methods on aggregated data of different segment lengths, with varying distribution and inferred that conclusion derived for one scale may not be applicable to other and the length of the road segment does have a significant role in statistical distribution of accident frequency[4]. In all the conventional models, spatial characteristics of hotspot locations were modeled as constant for a given period of time which is not true. An in-depth understanding of various factors influencing traffic accidents such as accident severity,surrounding environment is required for hotspot analysis[5]. Unlike conventional methods, spatial analysis aids in identification of traffic accident patterns and suggests reasons for those pattern characteristics [6].The capabilities of GIS as a powerful toolin identifying hazardous road locations include: i) handling large volume of various types of data, ii) visualization of accident locations, and iii) potential to analyze traffic accidents spatially with ease to identify hotspots[3].

This paper reviews the different spatial and statistical techniques for traffic accident hotspot identification with an example application of Planar Kernel Density Estimation (PKDE) on the chosen study area ofDes Moines city, Iowa state, USA. This manuscriptpresents the following sections: Section II presents the conventional hotspot identification methods along with its advantages and limitations; Section III explains the database and parameters required for spatial analysis of hotspot detection; Section IV elaborates on various spatial analysis techniques including the methods that deal with global effects and local effects; Section V presents the result analysis; conclusions and recommendation for future research are presented in section VIof this manuscript.

II. CONVENTIONAL HOTSPOT IDENTIFICATION METHODS

Different conventional or traditional hotspot identification methods that are in practice are explained below:

(a) Crash frequency (CF)method: This is a measure of occurrence of total number of accidents at a specific road stretch ina day or year. CF models are predominantly based on conventional stochastic models such as Poisson and negative binomial models. Multivariate Poisson regression models are very popular among researchers as it was found that crash occurrences were well fitted with Poisson distribution.

But the test results derived from this model may be incorrect when the assumption that mean and variance must be equal is violated.When the accident data are over-dispersed (variance exceeds mean) or under-dispersed (mean exceeds variance), it will lead to erroneous inference withthe parameters which determine the crash frequency [7].To overcome thelimitation of excessvariability in data, negative binomial regression model (Poisson-Gamma model)was extensively used in crash frequency analysis.However, this model fails to tackle data of less variance and has limitations in parameter estimate of small sample sizeand low mean values [8].To address this issue,Poisson lognormal model was found to be a viable alternative, in spite of its complexity involved in parameter estimation.Lord et. al. (2008)proposed Poisson log normal distributionfor developing accident prediction model whenever accident data are represented by small sample size and low mean values [9].

(b) Crash rate model: Crash rate is definedby the total numberof accidents per unit exposure measure of traffic volume indicated in terms of Annual Average Daily Traffic (AADT), road segment length, vehicles miles travelled [10].There exist only limited research studies on modeling crash rate distribution. Crash rate is viewed as a continuous variable and it equals zero if accident hasn'toccurred in the study area. Crash rate follows a mixed distribution due to the existence of many zero counts of accidents. This problem has been addressed byAnastasopoulos et. al.(2008) using Tobit regression model with fixed parametersover roadway segment limiting unobserved heterogeneitylikely to be present due to various factors such as environmental effects, behavior characteristics of driver etc. that are not contained in crash rate data, which result in biased parameter estimate[11, 12].To overcome this constraint, Tobit model with random parameters was proposed byAnastasopoulos et. al (2012)with fixed parameters [13]. A multivariate Tobit analysis incorporated with crash severity levels further provided insight to the factors affecting crash rate and outperformed univariate Tobit model [14]. An alternative and flexible approach among researchers to model crash rate is hurdle regression model which is a two-part model that specifies one step for zero (no accident) and another for non-zero outcomes[10].

(c) Equivalent Property Damage Only (EPDO)crash frequency method: This method unites crash severity data and crash count data, in which property damage only, that is "no injury" is considered as one unit and rest of the accidents are weighted based on their extent of severity.Oh et. al. (2010) proposed negative binomial regression modeling,in which raw crash frequency and EPDOdata are modeled separately [15]. The results wereto show that analysis of crash frequency failed toidentifysite related factors such as driving speed as they are not statistically significant that are sensitive to crash severity. A statistical approach based on lognormal hurdle model developed byMa et. al. (2016) is well suited for EPDO rate following mixed distribution where the model hurdles between zero EPDO rate and positive EPDO rates that are discretely and continuously distributed respectively [16].

All the above models discussed have certain degree of biased estimate due to fluctuations in accident count data during the span of study. This observation motivated researchers to developEmpirical Bayes (EB) methodology that can produce valid results and is less biased.EB is an effective and most reliable statistical approach to account for regression to mean effect[17].

It is very important to establish reasonable and realistic comparisons among the performance of the above models to evaluate theirsuperiority. When compared to the extensive research on methods for identification of hotspots, there is a lack of published literature on thecomparison of performance of various methods[18].Hauer et. al. (1984) made an empirical comparison on two methods: non-parametric method and EB method and concluded that EB method outperforms the other [19].Cheng et. al. (2009)applied four hot spot analyses techniques on 3-year crash data in Arizona:i) Crash frequency ranking, ii) crash rate ranking, iii) accident reduction potential and iv) EB method, and evaluated their performances by comparing them using various tests like site consistency test, total rank differences test, Poisson mean differencestest, false identification test.The results revealed that the performance of EB method is superior to that of crash rate method [20].

Elvik (2009) employed crash frequency, crash rate, combination ofcrash frequency and crash rate, EB method, and Potential for improvement method [21].Based on four yearsof accident data collected in Norwegian roads, the authors assessed the performance of the above methods in terms of sensitivity and specificity.It was concluded that the result is in good agreement with the published reference [20] and the researchers confirmed the reliability of EB technique in identifying hazardous road locations. The study carried out by many researchers unanimously agree on the superiority of EB model over others [22, 23]. The authors also suggested to replicate the study in other countries and recommended the EB method to be the standard technique for identification of black spots.

III. DATA BASE AND PARAMETERS FOR SPATIAL ANALYSIS OF HOTSPOTS

A large sample size is necessary to attain precise estimate of hotspot locations. A minimum of 3-year period of accident data is required to achieve statistically significant results [18, 24, 25].The statistical description of accident count and concentration is dependent on the choice of road section lengths [4]. Majority of the studies focus on urban areas having dense road network and major highways susceptible to high rate of crash incidence [3,6, 26, 27].When compared to theconventional methods, road segments are not clearly defined by spatial analysis methods[27].Road segments and intersections shall be analyzed separately since variables involved in each case is different [28, 29]. Intersections are susceptible to more crashes due to driving errors at partially signalized intersections [30]. Pulugurtha et. al. (2007) investigated the high pedestrian accident points at intersections[31]. Majority of the hotspot locations fall where highway connects to roads that lead to villages or small cities [3].

Due to rash driving over speeding rural roads are more prone to fatal-injury accidents compared to urban roads[32].Road intersection locationsand non-intersection locations areanalyzedseparately for bicyclist injury severity arising out of motor vehicle crashes and it isconcludedthat injury mechanism is different in each case as variables (cyclist age, influence of drugs, road curvature etc.) involved in both locations differ[29].

For each accident location, x, y coordinates, that is, longitude and latitude are specified and accidents are grouped into three categories based on the severity of accident, namely: fatal, injury, Property Damage Only (PDO).Geocodedlocations contain information (also known as attributes) like accident type, date and time of crash occurrence, road features, and weather conditions[6, 25, 33, 34]. In developing countries,pedestrians and two wheelers succumb to road accidents since they need to share thetraffic space with theheavy transport facilities[35].The vehicles involved in the accident are classified into two wheelers; light vehicles such as minibus,van; andheavy vehicles like trucks [36].

Traffic accident analysis is influenced by factors which are grouped into threecategories namely, i)human characteristics - driver's skill, performance, awareness and blood alcohol level,ii) environmental factors - weather and lighting conditions, road design, road condition, signals,iii) vehicle characteristics-braking performance, increased stopping distance[37]. Summer season encounters more accidents on highways as number of vehicles plying on highways are largercompared to other seasons[3].Shankar et. al. (1995)explained how road geometry and weather conditions influence crash occurrence in rural areas [38].Yoshiki et. al (2016) analyzed traffic accidents based on city characteristics and inferred that road length, intersections, crowded vehicles and pedestrians near public and health care facilities have significant effect on traffic accident hotspot prediction models[39].

IV. SPATIAL ANALYSIS TECHNIQUES

This section presents various spatial analysis tools integrated with ArcGIS that can identify the hotspots in the area of interest.Identification ofhotspots is the most important aspect in controlling traffic accidents because it enables effective traffic management by optimization ofroad signs and traffic police personnel, and deployment of automatic traffic monitoring systems at the hotspots[40], [41]. The accident incidentslocated on a geographical space are studied to identify any systemic pattern of occurrence of crashes to determine whether the accident locations are distributed randomly or as a clustered pattern. This process is referred to as Point Pattern Analysis(PPA). PPA methods can be categorized based on various criteria.Each method has its own advantages and limitations, and their usage is specific to a situation.PPA methods are broadly classified into methods that deal with global effects and local effects. The methods which consider the global effects refer to large scale spatial variation that describes the definite properties of local environment. In traffic accident analysis, this method refers to clustering of crashes in specific areas given the presence of built infrastructure such as bridges, tunnels, bus stops, etc. The method that deals with local effects examines the local scale variation of crashes from its mean value, i.e. interaction between accidents [42, 43, 44].

The techniques used to analyze the spatial distribution of accidents are grouped into i)Density method, ii)Distance method,and iii) Interpolation method. Methods like kernel density estimation,quadrat analysis, kriging etc. deal with first order effects of global variation. The results from first order effectshaveto be investigated further to obtain statistically significant estimate. Distance methodslike Nearest neighbor distance, and*K* function analysisexplore second order effects based on the relativedistances of individual crashes. Some of the prominent and commonly used methods in hotspot detection are presentedin the following sub-sections.

A. Methods that deal with global effects

Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE) is a spatial statistical methodto identify the locations of high density of accidents in 2D Euclidean space[45].It is a commonly used method to examinetheglobal effects of point event distributionnot only due to its simplicity and easeof implementation, but also because of its ability to identify local spatial characters[25, 42, 46]. Both Planar KDE (PKDE)and Network KDE (NKDE) are useful in identifying hotspots but the latter is advantageousbecause NKDE was able to define the limits of dangerous street segments precisely, producing accurate results[33].

Planar Kernel Density Estimation (PKDE)

This method estimates the accident density by counting the number of accidents in an area. This area is known as kernel. The total study area is divided by predetermined number of cells.PKDEis evaluated by fitting a smooth function called kernel overeveryaccident point and then computingthe distance from that point to the reference location based ona mathematical function and adding the value of all the surfaces for that reference location (Fig. 1).Kernel density function weighs nearby accidents more heavily than distant ones in estimating local density[45].

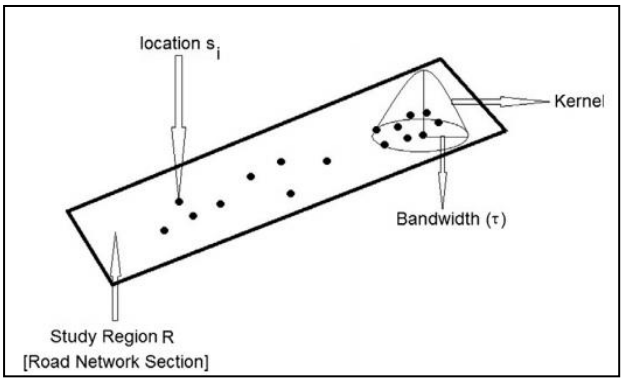

Fig 1. Illustration of PKDE for spatial point pattern analysis [47].

The general equation of PKDE is given below (Eqn.1)

$$\widehat{\lambda_\tau}(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{s-s_i}{\tau}\right) \quad (1)$$

where $\widehat{\lambda_\tau}(s)$ is density estimate at the point's location *s*,*n* is accidents count, s_i is thelocation ofi^{th} accident;τ is the bandwidth, *i.e.* search radius; *k*( ) is kernel function - a function of the distance and search radius.

Different kernel functions exist like Gaussian, Quatric, Triangular functions which are used in accident analysis[27, 48, 49].Quartic function is used in ArcGIS, which is popular in accident analysis[31].Selection of appropriate input parameters *i.e.* bandwidth and cell size,arevery crucial than the kernel function itself and this selection is also subjective in determining right density estimate[49, 50]. Narrow bandwidth yields an under-smoothed density map, with all peaks and valleys detected, apt for understanding local effects, whereas large bandwidth results in smoother density estimate, exhibiting less variability.So,one has to chooseappropriate cell size and band width based on computational time, sample size and data considered for the study. Yoshikiet. al. (2016) applied multiple regression technique, proposed by Ito et. al. (2010) [51] to arrive at suitable bandwidth for the authors' study area[39]. It is meaningless to have a bandwidth lesser than the size of grid cell. This is because the output of PKDE would be averaged over each grid[26].KDE was first used by Banoset. al. (2000) to map the spreadof school children pedestrian accidents and point source- schools [52]. Budiharto et. al. (2012) used PKDE to determine spatial distribution of hotspots and identified the fastest route to mobilize accident victim from crash site to referral hospital[53].

Table I: Cell Size and Bandwidth in Planar Kernel Density Estimation (PKDE)

S.No	Author	Cell size(m)	Bandwidth (m)	Country and area type
1	Anderson TK (2009)	100	200	United Kingdom
2	Erdoganet *al*. (2008)	500	500	Turkey
3	Erdogenet *al*. (2015)	50	700	Turkey
4	Thakalie*t al*. (2015)	400	400,800	United states
5	Hashimoto *et. al.*(2016)	250	250	Japan

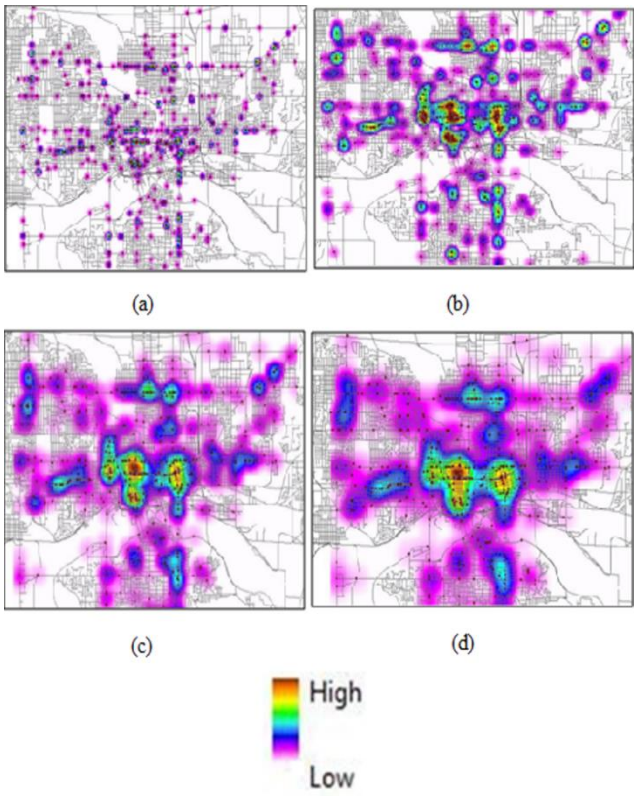

Fig. 2. PKDE Using Different Bandwidths for cell sizeof100 m: (a) Bandwidth of 250m, (b) Bandwidth of 500m, (c) Bandwidth of 750m, (d) Bandwidth of1000 m

Bíl et. al. (2013) determined the statistical significance of clusters resulting from PKDE with a dimensionless number, cluster strength, which is a function of factors such as cluster length and crash count in a cluster.Cluster strength quantifies how density estimate differs from expected value[54].

Case study on crashes for Des Moines city of Iowa state

The authors of this paper carried-out a case study on crashes occurred in Des Moines city of Iowa state to show the effect of varying bandwidths in creating density maps based on Planar Kernel Density Estimation (PKDE). Crash data for the years 2008 to 2012areretrieved from https://catalog.data.gov. Prior to performing PKDE, the data are aggregated using collect event tool in ArcGIS that combines all crashes that happened in the same geographical location. This results in a new weighted point feature class called ICOUNT which is used as an input for estimating PKDE. All the spatial statistics are carried out using ArcGIS 10.6.1. PKDEis applied on major injuries to detect hotspots. Figures 2(a-d) illustrate the effect of influence of different bandwidths on density estimation for detecting hotspots. Based onpublished literatures (Table I), cell size was set as 100m and density estimate was calculated for varying bandwidths 250m, 500m, 750m and 1000m. It is quite evident that 500m bandwidth is appropriate in this case as clusters are distinct. Hotspots are hardly detected when the bandwidth is 250mand clusters get merged in higher bandwidths.

Network Kernel Density Estimation (NKDE)

Researchers started realizing the limitations of PKDE estimate in which the density of events is measured in 2D Euclidean space. However, traffic accidents occur in network space, hence leading to biased results.

Fig. 3 illustrates the idea in calculatingthe density estimate using PKDEand NKDE methods for a point *x*. To compute the crash density at a location*x*, the PKDEconsiders the complete 2D space and detects four crash locations (shown as solid dots) within a search radius *r*, whereas the NKDEfinds only two accidents within the same search radius[49]. In this case, PKDE over-estimates density values as density of road network is not taken into account.The limitation of PKDE is that some grid cells may include large number of road section and some may have few or not any. By implementing network KDE, the above drawback can be avoided[3,28, 55, 56].

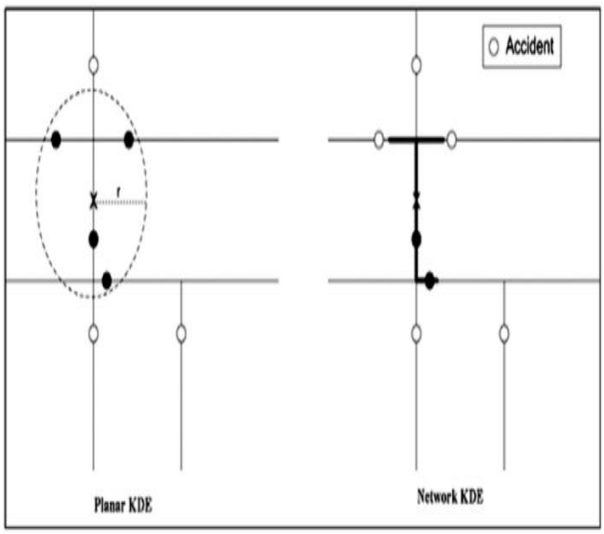

Fig.3. Difference Between PKDE and NKDE in Estimating Density Values at a point[49]

Flahaut et al. (2003) was the pioneer in developing NKDE for traffic accident analysis, but his attempt was limited to a single stretch of road, not a road network [57]. An algorithm was developed for NKDE, using Gaussian and Quartic kernel functions, where accident events are distributed in one dimensional network space[49]. In this case, grid cells are replaced by lixels and the shortest path distance from a point to crash location is defined as bandwidth. The proposed algorithm was tested on a city in Kentucky and was concluded that NKDE is superior to PKDE. The latter overestimates the density values as they are spread over the entire study.Eqn. 2 shows the density estimate at a location.

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \qquad (2)$$

where $\lambda(s)$ is the density at location *s*, *r* is the search radius,$k(\ )$ is kernel function, d_{is} is thedistance from location 's' to accident location '*i*'.

The author also found that, like *PKDE*, *NKDE* kernel functiondid not have any impact on density values but lixel length and bandwidths are subjective and influence the density patterns. Different bandwidths were proposed in literatures depending on scale of study (Table II). Okabe et. al. (2012)defined a rule of thumb that bandwidth is ten times larger than the cellsize [58].The major limitation in KDE methods is that there is no mechanism available for statistical inference that can indicate a threshold above which hotspots can be confirmed[42, 49].

The first order and second order methods were integrated to overcome this limitation [56]. The density values derived from Network KDE were used in local Moran's *I* statistical method to determine the threshold values.This method is used to analyze the second order effects on spatial pattern of accidents. This threshold value was used to identify hotspots.Application of NKDE on traffic analysis has been well explored by researchers all over the world. Kaygisiz et. al. (2015) performed a spatio-temporal analysis to determine the impact of behavior factors in driving causing traffic accidents using Network KDEand suggested changes in road structure pertaining to behavior aspects in driving [47]. The authors tested the statistical significance of the density estimate with nearest neighbor and *K* function analysis and temporal changes in hotspot was tested with chi square test.

Similar to *PKDE*, *NKDE* adopts various kernel functions [48, 55, 59]. A computational method was developed by formulating equi-split kernel function on network and showed that it can provide unbiased density estimate [60]. A preprocessing tool was also conceived to operate these methods in GIS environment, called as Spatial Analysis on a NETwork (SANET).

Kriging

This is one of the least explored method in traffic accident analysis. Kriging is an interpolation technique for estimating attribute values which are not known for any accident location in a region of accident point values. In this method, spatial variable is composed of two components: large scale trend and small-scale spatial auto correlation which is the error term [61]. Eqn. 3 shows the general equation of kriging:

$$Z_i(s) = \mu_i(s) + \varepsilon_i(s) \qquad (3)$$

$Z_i(s)$ is variable of interest (accident count);μ_i is the large-scale trend; ε_i is error component; *s* is location of accident *i*.

There are different kriging techniques such as Simple Kriging (SK),UniversalKriging (UK) and ordinary Kriging (OK). OK is the most widely used, best linear unbiased estimate, and the mean is an unknown constant across locations. SK assumes mean and distribution remain constant throughout the region. UK is used when data have strong trend. Kriging relies on semi-variogram that illustrates the spatial autocorrelation of accidents. Selection of appropriate semi-variogram model is very much essential to best fit the relationship between distance and crash locations, and variance for a given dataset. Exponential, spherical and Gaussian are commonly employed models. The shape of the models is based on parameters:i) c_0, nugget effect, which reflects the discontinuity at the origin, caused by sampling error and small-scale variability, ii) a, range-the distance at which variance stabilizes or flattens, iii) c0+c1sill-the value that the model attains at the range (Fig. 4).

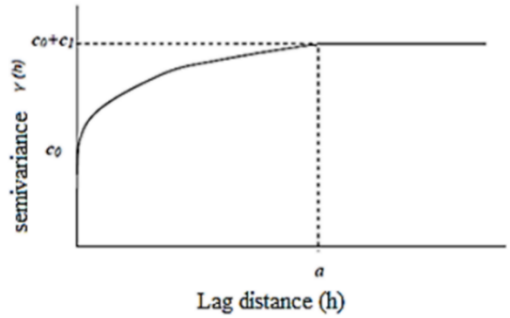

Fig.4. Illustration of Semivariogram Depicting Parameters[62]

Table II. Lixel Size and BandwidthinNetwork Kernel Density Estimation (NKDE)

S.No	Author	Lixel size (m)	Bandwidth (m)	City/Country
1	Mohaymany*et al*. (2013)	Not specified	1000	Markazi province, Iran
2	Ivan & Tesla (2015)	20	200	Ostrava, Czech Republic
3	Xie& Yan (2008, 2013)	10	100	Kentucky USA

It was observed that UK performed better than Empirical Bayes (EB) method for predicting crash frequency and crash severity[64]. UK and spherical model for semi variogram parameter estimate were used in this study. When hotspots derived from KDE were compared withthat of kriging, based on prediction accuracy index developed byChainey et. al. (2008)it was concluded that kriging outperformed KDE in hotspot identification[26, 65].

B. Methods that deal with local effects

This section deals with some of the commonly used distance methods that deal with local effects that is interaction between the accidents.

Nearest Neighbor Distance Method (NND)

Nearest Neighbor Distance (NND) method is used to identify the clustering pattern by measuring the distance between each crash location and the nearest neighboring accident. If the average of all NND is less than the expected average distance, null hypothesis, i.e., accidents are randomly distributed, is rejected, points (accident locations) are considered as clustered[42]. NND method shall be termed as planar NND method or network NND methods, that is based on where they are defined on - Euclidean or network space[58].Both planar and network NND methods shall be further classified into local NND and global NND depending on how the interaction between the accidents are considered. Shafabakhsh et. al. (2017) applied network NND analysis to show the existence of clustering pattern of crashes in a city of Iran [34].Yalcin (2015) grouped accidents based on types of vehicles involved and showed that the accidents caused by light and two wheeled vehicles were highly clustered [36].

K-function analysis

K-function is viewed as one of the popular methods to analyze traffic accident pattern spatially as this method is capable of exploring all point to point distances and examining point patters over a wide range of spatial scale unlike NNDmethod which takes into account only the distances to the closest crash locations, hence considering the smallest scales of pattern[42].Though *K*-function cannot exactly detect where the geographical location of clusters, it is a first step in identifying localized raised incidence[63]. Network K-function performs better than planar K-function and it is considered as the most reliable method. Over detecting of clustering pattern is the shortcoming in planar K-function analysis due to inappropriate distance calculations[44].

Yalcin (2015)applied Global K-function to analyze crashes involved according to vehicle type in urban areas of Osmaniye city of Turkey and concluded that two wheeled vehicles had more percentage of accidents compared to other vehicles. This was attributed to weather conditions and socio-economic environment [36]. Çela, (2013) applied Network *K*-function on smaller scale data to locate crash clusters [66].

Moran's I statistic

This method was proposed by Moran (1948) which is a measure of global spatial autocorrelation technique, based on comparison of attribute values of neighboring points. Spatial autocorrelation is a spatial arrangement technique which measures the degree of similarity of one object with its surrounding objects[67]. ArcGIS tool computes spatial autocorrelation on the basis of both feature locations and feature attributes simultaneously. For chosen accident locations and associated attributes for each accident, Moran's I index value is calculated to identify if there is any clustering pattern. p value and z score are computed to evaluate the statistical significance of *I* index. *I* index value ranges from -1 to +1. -1 implies accidents are in perfect dispersion, 0 indicates there is no autocorrelation and 1 means clustering of crashes. *I* index can be represented as (Eqn. 4):

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \quad (4)$$

x_i ,x_j represent the value of variable at crash location *i*and neighboring location *j*
n represents the crash count
w_{ij} row standardized weights, exhibiting proximity relationships between location *i* and the neighboring location *j*.
$\sigma^2$is the variance of Moran's I as shown in Eqn. 5.

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n \qquad (5)$$

The null hypothesis states that accidents are randomly distributed. When the p-value is small and the z score is significant enough that it lies outof confidence level, the null hypothesis can be rejected. Once the null hypothesis is rejected,it is necessary to examine *I* index value. If *I* index is more than 0, accidents are clustered, and dispersed if it is less than 0. The effectiveness of Moran I index method was exhibited in analyzing traffic accidents in Thiruvananthapuram, India[6].

Spatial autocorrelation may differ from one part to another part of the region. In order to capture this variability, global measures are modified to identify and compute spatial autocorrelation at local scale. Local Moran's I index, developed by authors in[5]is better suited for traffic crash analysis as they can identify clustering pattern based on the estimate of statistical significance (Eqn. 6).

$$I_i = z_i \sum_j^n w_{ij} z_j (6)$$

where,

$$z_i = \frac{x_i - \bar{x}}{\delta} (7)$$

z_i and z_j are number of standard deviations from mean for i and j locations (Eqn. 7), δ is the standard deviation of attribute value *x*.

Accidents to 100-m segments were combined and accident count of each segment was used as the attribute for computing the Moran's *I* value [68].However, Truong et. al. (2015) has computed the Moran's I index considering the distribution of Crash counts and the severity index [69]. Xie et. al. (2013) presented a novel approach for hotspot identification by integrating NKDE and local Moran's I[56].Local Moran's *I* was computed from the density values obtained from NKDE. The author carried out Monte Carlo simulation on local Moran's *I* to avoid unrealistic assumption that local Moran's *I*is normally distributed, thus making the test of significance more vigorous and flexible. Erdogan et. al. (2017) combined *PKDE* with Moran's *I* and found to it be very effective in detecting hotspots in straight roads[70]. Moran's *I* was very effective and instrumental in revealing temporal and spread of accidents in both rural and urban environments[43, 71]. The drawback with Moran's *I* is its inability to differentiate H-H vs L-L (hotspot vs. cold spot) in spatial auto-correlationenvironment. Çelaet. al. (2013) applied Network *K*-function on smaller scale data to locate crash clusters [66].

Local G-statistic

Local *G* statistic (or Getis-Ord *Gi**) is used to identify the spatial association of the high and low values of the feature.

Local G-statistic is defined as (Eqn. 7):

$$G_i(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j} \qquad (7)$$

d is distance threshold
$x_j$is the value of the attribute at location *j*
$w_{ij}$is weight of the target neighbor pair

Truong et. al. (2015) generated pedestrian vehicle crash hotspot map using Local *G*-Statistic method to prioritize unsafe bus stops based on severity index. The authors have mentioned that while ranking unsafe bus stops, sensitivity analysis must be performed to identify an appropriate buffer size.This is because severity indices vary for different buffers and hence influences the ranking. Erdogan et. al. (2008) showed that combining PKDE with local *G* statistic could produce accurate hotspots in junction locations rather than in a road stretch [3].

V. RESULTS ANALYSIS

In this paper, various spatial and statistical methods that are employed in detecting accident hotspots are discussed and the influence of bandwidth in identifying the hotspots is investigatedfor the study area ofDes Moines city, Iowa state, USA. There are 24660 accident locations and the accidents are grouped into three categories based on the severity of the accidents, namely: fatalities, injuries and property damage only. In this study only injury-type accidents are considered for the purpose of spatial analysis. There are 7086 locations having injury-type accident. PKDE is carried out using four bandwidths –250m, 500m, 750m and 1000 m. In all cases cell size is maintained as 100 m. The following table illustrates the results obtained from PKDE analysis.

Table III. Results from PKDE analysis

S.No.	Bandwidth (in m)	No of hotspots detected	Density estimate (No.of accident counts per square km)
1	250	9	8573
2	500	12	1340
3	750	5	2297
4	1000	4	1710

It is evident that when the bandwidth is 500 m, larger number of distinct hotspots are detected,and the maximum density estimate for 500 m bandwidth is obtained as 1340accidents per square kilometer. Hotspots get merged when the search radius is greater than 500m. Hence,a 500m bandwidth is ideal to locate the hotspots for the chosen study area.

VI. CONCLUSIONS AND RECOMMENDATIONS

Accident hotspotsidentification methods have significantly improved over the past two decades and play a crucial role in the enforcement of effective strategies for traffic safety management. In this paper, various spatial and statistical methods used in detecting accident hotspots are discussed, followed by acase studyusing Planar Kernel Density Estimation (PKDE) to investigate the effect of bandwidths in hotspot identification. The study considers the injury-type accidents in the Des Moines city, Iowa state, USA,based on crash data for the years 2008 to 2012. The influence of bandwidth in PKDE is investigatedfor the cell size of 100 m and the optimum bandwidth is identified. It is concluded that a 500m bandwidth is ideal to locate hotspots, for the chosen study area. In the case of other bandwidths (i.e. 250 m, 750 m, and 1000 m), distinct hotspots were not identified.

Some of the recommendations for future research in the area of spatial analysis of traffic accidents for hotspots identification are listed below:

i) In depth analysis on different locations such as straight roads, junction, and parking lots could be carried out to investigate which type of analysis method is appropriate for these locations.
ii) Influence of characteristics of land-use, driver's behavior, road geometry and weather conditions could be incorporated in hotspot detection.
iii) Most studies consider traffic volume as constant for each road segment.A more realistic treatment could consider the as-is condition in NKDE.
iv) Sensitivity analysis could be performed using different segment lengths and bandwidths.
v) Incorporating multi-criteria ranking methods to prioritize hotspots.
vi) Effects of exponential, Gaussian and spherical functions can be studied in crash prediction based on kriging method.
vii) Hotspot identification can be developed by using single parameter based on integration of severity of individualcrash data in kriging.
viii) Identify the contribution of crash influencing factors like weather conditions and traffic exposure in hotspot detection using kriging.
All the above suggestions can improve the existing techniques to provide more robust and accurate hotspot identification and prediction, which can create an avenue for planning, implementation and decision making for traffic safety management.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge Florida Atlantic University, Boca Raton, USA and CMR Institute of Technology, Bengaluru, India for providing the facilities to carry out the study.

REFERENCES

1. World Health Organization,https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. Hakkert, A. S., and Mahalel. D. (1978). "Estimating the Number of Accidents at Intersections from a Knowledge of the Traffic flow on the Approaches". Accident Analysis and Prevention, Vol. 10, pp. 69-79
3. Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). "Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar". Accident Analysis and Prevention, 40(1), 174–181.
4. Thomas, I. (1996). "Spatial data aggregation: Exploratory analysis of road accidents". Accident Analysis and Prevention, 28(2), 251–264.
5. Anderson, T. K. (2009). "Kernel density estimation and K-means clustering to profile road accident hotspots". Accident Analysis and Prevention, 41(3), 359–364.
6. Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). "Spatio-temporal clustering of road accidents: GIS based analysis and assessment". Procedia - Social and Behavioral Sciences, 21, 317–325.
7. Lord, D., & Mannering, F. (2010). "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives". Transportation Research Part A: Policy and Practice, 44(5), 291–305.
8. Lord, D. (2006). "Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter". Accident Analysis and Prevention, 38(4), 751–766.
9. Lord, D., & Miranda-Moreno, L. F. (2008). "Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective". Safety Science, 46(5), 751–770.
10. Lu Ma, Xuedong Yan and JinxianWeng (2015)."Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter"J. Adv. Transp., 49:928–940
11. Anastasopoulos, P. C., Tarko, A. P., & Mannering, F. L. (2008). "Tobit analysis of vehicle accident rates on interstate highways". Accident Analysis and Prevention, 40(2), 768–775.
12. Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Econometric methods for Transportation data analysis, Chapman & Hall/CRC, Taylor & Francis Group.
13. Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N., & Haddock, J. E. (2012). "study of factors affecting highway accident rates using the random-parameters tobit model". Accident Analysis and Prevention, 45, 628–633.
14. Anastasopoulos, P. C., Shankar, V. N., Haddock, J. E., & Mannering, F. L. (2012). "A multivariate tobit analysis of highway accident-injury-severity rates". Accident Analysis and Prevention, 45, 110–119.
15. Oh, J., Washington, S., & Lee, D. (2010). "Property Damage Crash Equivalency Factors to Solve Crash Frequency-Severity Dilemma". Transportation Research Record: Journal of the Transportation Research Board, 2148, 83–92.
16. Ma, L., Yan, X., Wei, C., & Wang, J. (2016). "Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework". Analytic Methods in Accident Research, 11, 48–61.
17. Persaud, B., & Lyon, C. (2007). "Empirical Bayes before-after safety studies: Lessons learned from two decades of experience and future directions". Accident Analysis and Prevention, 39(3), 546–555.
18. Cheng, W., & Washington, S. P. (2005). "Experimental evaluation of hotspot identification methods". Accident Analysis and Prevention, 37(5), 870–881.
19. Hauer, E. and Persaud, B. N. (1984). Problem of identifying hazardous locations using accident data . Transportation Research Record, (No. HS-03(975), 49.
20. Cheng, W., & Washington, S. (2009). "New Criteria for Evaluating Methods of Identifying Hot Spots". Transportation Research Record: Journal of the Transportation Research Board, 2083, 76–85.
21. Elvik, R. (2009). "Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads". Transportation Research Record: Journal of the Transportation Research Board, 2083(0349), 72–75.
22. Montella, A. (2010). "A comparative analysis of hotspot identification methods". Accident Analysis and Prevention, 42(2), 571–581.
23. Valentova, V., Ambros, J., & Janoska, Z. (2014). "A comparative analysis of identification of hazardous locations in regional rural road network". Advances in Transportation Studies, 34, 57–66.
24. Elvik, R. (2008). "A survey of operational definitions of hazardous road locations in some European countries". Accident Analysis and Prevention, 40(6), 1830–1835.
25. Steenberghen, T., Aerts, K., & Thomas, I. (2010). "Spatial clustering of events on a network". Journal of Transport Geography, 18(3), 411–418.

26. Thakali, L., Kwon, T. J., & Fu, L. (2015). "Identification of crash hotspots using kernel density estimation and kriging methods: a comparison". Journal of Modern Transportation, 23(2), 93–106.
27. Yu, H., Liu, P., Chen, J., & Wang, H. (2014). "Comparative analysis of the spatial analysis methods for hotspot identification". Accident Analysis and Prevention, 66, 80–88.
28. Ivan, I., & Tesla, J. (2015). "Road and intersection accidents: Localization of black spots in Ostrava". Geograficky Casopis, 67(4), 323–340.
29. Moore, D. N., Schneider IV, W. H., Savolainen, P. T., & Farzaneh, M. (2011). "Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations". Accident Analysis and Prevention, 43(3), 621–630.
30. Young, K. L., Salmon, P. M., & Lenné, M. G. (2013). "At the cross-roads: An on-road examination of driving errors at intersections". Accident Analysis and Prevention, 58, 226–234.
31. Pulugurtha, S. S., Krishnakumar, V. K., & Nambisan, S. S. (2007). "New methods to identify and rank high pedestrian crash zones: An illustration". Accident Analysis and Prevention, 39(4), 800–811.
32. Ma, L., Yan, X., & Qiao, W. (2014). "A Quasi-Poisson Approach on Modeling Accident Hazard Index for Urban Road Segments". Discrete Dynamics in Nature and Society, 2014, 1–8.
33. Benedek, J., Ciobanu, S. M., & Man, T. C. (2016). "Hotspots and social background of urban traffic crashes: A case study in Cluj-Napoca (Romania)". Accident Analysis and Prevention, 87(February), 117–126.
34. Shafabakhsh, G. A., Famili, A., & Bahadori, M. S. (2017). "GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran". Journal of Traffic and Transportation Engineering (English Edition), 4(3), 290–299.
35. Khan, A.M. and Tehreem, A., 2012. "Causes of road accidents in Pakistan". Journal of Asian Development Studies, 1(1), pp.22-29.
36. Yalcin, G., & Duzgun, H. S. (2015). "Spatial analysis of two-wheeled vehicles traffic crashes: Osmaniye in Turkey". KSCE Journal of Civil Engineering, 19(7), 2225–2232.
37. Fell, J. C. (1976). "A Motor Vehicle Accident Causal System: The Human Element. Human Factors": The Journal of Human Factors and Ergonomics Society, 18(1), 85–94.
38. Shankar, V., Mannering, F., Woodrow, B., 1995. "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies". Accident Anal. and Prev. 27 (3), 371–389.
39. Yoshiki, S., Hashimoto, S., Mimura, Y., Saeki, R., Nanba, S., & Ando, R. (2016). "Development and application of traffic accident density estimation models using kernel density estimation". Journal of Traffic and Transportation Engineering (English Edition), 3(3), 262–270.
40. Biswas, D., Su, H., Wang, C., Blankenship, J., & Stevanovic, A. (2017). "An automatic car counting system using overfeat framework". Sensors (Switzerland), 17(7), 1–13.
41. Biswas, D., Su, H., Wang, C., Stevanovic, A., & Wang, W. (2018). "An automatic traffic density estimation using Single Shot Detection (SSD) and MobileNet-SSD". Physics and Chemistry of the Earth, (December), 0–1.
42. Bailey, T. C., &Gatrell, A. C. (1995). Interactive spatial data analysis (Vol. 413). Essex: Longman Scientific & Technical.
43. Mohaymany, A. S., Shahri, M., & Mirbagheri, B. (2013). "GIS-based method for detecting high-crash-risk road segments using network kernel density estimation". Geo-Spatial Information Science, 16(2), 113–119.
44. Yamada, I., & Thill, J. C. (2004). "Comparison of planar and network K-functions in traffic accident analysis". Journal of Transport Geography, 12(2), 149–158.
45. Silverman, B. W. (1986). Density estimation for statistics and data analysis. London: Chapman Hall.
46. Steenberghen, T., Dufays, T., Thomas, I., & Flahaut, B. (2004). "Intra-urban location and clustering of road accidents using GIS: A belgian example". International Journal of Geographical Information Science, 18(2), 169–181.
47. Kaygisiz, Ö., Düzgün, Ş., Yildiz, A., & Senbil, M. (2015). "Spatio-temporal accident analysis for accident prevention in relation to behavioral factors in driving: The case of South Anatolian Motorway". Transportation Research Part F: Traffic Psychology and Behaviour, 33, 128–140.
48. Gibin, M., Longley, P., & Atkinson, P. (2007). "Kernel density estimation and percent volume contours in general practice catchment area analysis in urban areas". The Proceedings of GISRUK, C, 11–13.
49. Yalcin, G., & Duzgun, H. S. (2015). "Spatial analysis of two-wheeled vehicles traffic crashes: Osmaniye in Turkey". KSCE Journal of Civil Engineering, 19(7), 2225–2232.
50. O'sullivan, D., & Wong, D. W. S. (2007). "A surface-based approach to measuring spatial segregation|". Geographical Analysis, 39(2), 147–168.
51. Ito, F., Itogawa, E., Umemoto, M., 2010. "Occurrence factors from a microscale environmental characteristics point of view: case study of bag snatching in Itabashi Ward, Tokyo". Journal of Social Crime Safety Science (13), 109-118.
52. Banos, A., & Huguenin-Richard, F. (2000). "Spatial Distribution of Road Accidents in the Vicinity of Point Sources Application to Child Pedestrian Accidents". Geography and Medicine, 54–64.
53. Budiharto, U., & Saido, A. P. (2012). "Traffic Accident Blackspot Identification and Ambulance Fastest Route Mobilization". Jurnal Transportasi, 12(3), 237–248.
54. Bíl, M., Andrášik, R., & Janoška, Z. (2013). "Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation". Accident Analysis and Prevention, 55, 265–273.
55. Okabe, A., Okunuki, K. I., & Shiode, S. (2006). The SANET toolbox: New methods for network spatial analysis. Transactions in GIS, 10(4), 535–550.
56. Xie, Z., & Yan, J. (2013). "Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: An integrated approach". Journal of Transport Geography, 31, 64–71.
57. Flahaut, B., Mouchart, M., Martin, E. S., & Thomas, I. (2003). "The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach". Accident Analysis and Prevention, 35(6), 991–1004.
58. Okabe, A., & Sugihara, K. (2012). Spatial analysis along networks: statistical and computational methods. John Wiley & Sons
59. Levine, N. (2004). CrimeStat III: A spatial statistics program for the analysis of crime incident locations. Houston, TX/Washington, DC: Ned Levine & Associates/The National Institute of Justice.
60. Okabe, A., Satoh, T., & Sugihara, K. (2009). "A kernel density estimation method for networks, its computational method and a GIS-based tool". International Journal of Geographical Information Science, 23(1), 7–32.
61. Wang, X., Kockelman, K. M., Murray, W. J., & Fellow, J. (2009). "Forecasting Network Data: Spatial Interpolation of Traffic Counts using Texas Data". Trasportation Research Board, Transporta.
62. Wang, X., and Kockelman, K. M. (2009). "Forecasting Network Data: Spatial Interpolation of Traffic Counts using Texas data'. Transportation Research Record:Journal of the Transportation Research Board, No. 2105, TRB, Washington, D.C., pp. 100-108.
63. Jones, A. P., Langford, I. H., & Bentham, G. (1996). "The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England". Social Science and Medicine, 42(6), 879–885.
64. Manepalli, U. R. R., Bham, G. H., & Kandada, S. (2011). "Evaluation of Hotspots Identification Using Kernel Density". 3rd International Conference on Road Safety and Simulation, 1750, 1–17.
65. Chainey S, Tompson L, Uhlig S (2008). "The utility of hotspot mapping for predicting spatial patterns of crime". Secur J 21(1):4–28
66. Çela, L., Shiode, S., & Lipovac, K. (2013). "Integrating GIS and Spatial Analytical Techniques in an Analysis of Road Traffic Accidents in Serbia". International Journal for Traffic and Transport Engineering, 3(1), 1–15.
67. Levine N., 2000. CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations, vol. 1.1. Ned Levine & Associates/National Institute of Justice, Annandale, VA/Washington, DC.
68. Moore, D. N., Schneider IV, W. H., Savolainen, P. T., & Farzaneh, M. (2011). "Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations". Accident Analysis and Prevention, 43(3), 621–630.
69. Truong, L., & Somenahalli, S. (2015). "Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops". Journal of Public Transportation, 14(1), 99–114.
70. Erdogan, S., Ilçi, V., Soysal, O. M., & Kormaz, A. (2015). "A Model Suggestion for the Determination of the Traffic Accident Hotspots on the Turkish Highway Road Network: a Pilot Study". Boletim de Ciências Geodésicas, 21(1), 169–188.
71. Soltani, A., & Askari, S. (2017). "Exploring spatial autocorrelation of traffic crashes based on severity". Injury, 48(3), 637–647.

AUTHORS PROFILE

Dr. Lakshmi Srikanth, is a Professor at the Department of Civil Engineering, CMR Institute of Technology, Bengaluru. She has more than 2 decades of industrial, research and academic experience in Structural Engineering, Remote Sensing and GIS. She is a life member of Indian Society of Geomatics and Indian Society of Remote Sensing. Her areas of research interest are GIS applications in Traffic accident analysis, Structural Engineering and Ocean and Atmospheric Research.

Ms. Ishwarya Srikanth, is a Doctoral Candidate at the Department of Civil, Environmental and Geomatics Engineering in Florida Atlantic University. She is an active student member of American Society of Civil Engineers (ASCE) and Florida Structural Engineers Association (FSEA). Her research interests include Bridge Deterioration and Maintenance models, Structural Health Monitoring, Structural Reliability, Design of Offshore Structures, Machine learning and GIS applications in Civil and Structural Engineering.

Dr. M. Arockiasamy, is a Professor at the Department of Civil, Environmental and Geomatics Engineering in Florida Atlantic University (FAU) and the Director of Center for Infrastructure and Constructed Facilities at FAU. He is a Professional Engineer in the states of Florida, Alabama, Louisiana, Wisconsin and Newfoundland in Canada, and a Fellow ASCE member. His research interests include Ocean, Wind and Wave Energy Utilization, Offshore/Coastal Structures, Advanced High Strength Composites, Fire and Blast Resistance of Structures, Sustainability and Climate Change Impact on Infrastructure.