Journal of South Asian Logistics and Transport
Volume 5 Issue 1 March 2025
JSALT 5.1.1: pp. 1-23
DOI: https://doi.org/10.4038/jsalt.v5i1.92

IDENTIFICATION OF TRAFFIC ACCIDENT HOTSPOTS USING NETWORK KERNEL DENSITY: APPLICATION TO GALLE DISTRICT IN SRI LANKA

T M Rengarasu* * *and H K I Arunodi

Department of Civil & Environmental Engineering, University of Ruhuna, Sri Lanka.

* Correspondence should be addressed to *rengarasu@cee.ruh.ac.lk*

ABSTRACT

This study aims to identify traffic accident hotspots and establish the correlation between road features and accidents. Network-based Kernel Density Estimators (NKDE) were computed using traffic accident data from 2009 to 2013 for each year for road segments termed lixels. These yearly NKDEs were aggregated to formulate a traffic accident intensity index (AII), serving as the foundation for hotspot identification. A Gamma regression model with a logarithmic link was constructed to relate the AII values of lixels in the A2 highway with macro-level parameters. Population, Junction density (number of junctions in a section), Bendiness, and Urbanisation. The study determined that a cell size of 250 m with a 500 m bandwidth is optimal for NKDE calculations for the road network of the Galle district. Lixels having extreme AII values were defined as the traffic accident hotspots and subdivided into yellow, orange, and red zones. Additionally, the Gamma regression model highlighted significant correlations between road attributes and accident intensity, indicating a 131% increase in AII values in urban areas compared to rural ones. Areas with junctions and bends had a decreasing effect on the AII values. With the increase of a junction, AII decreased significantly by 29 %, and if a bend is present in the lixel, AII decreases significantly by 12%.

Keywords: *Traffic accident hotspots, GIS, network kernel density estimation, gamma regression model*

1. INTRODUCTION

Traffic accidents remain a major global health challenge, causing 1.19 million deaths in 2021, with low-income nations facing triple the fatality risk of high-income countries and 66 middle-income nations seeing rising fatalities despite overall declines.[1]. Research indicates a notable escalation in reported traffic accidents within Sri Lanka between 1989 and 2005; the number of reported traffic accidents surged from 26,196 to 52,444. Correspondingly, fatal traffic accidents also exhibited a concerning upward trajectory, with reported fatalities climbing from 1,454 in 1989 to 2,141 in 2005; this is an average growth rate of 2.4% per annum. These statistics most likely underestimate the true scope of the issue, as minor traffic accidents often go unreported to law enforcement agencies [2]. Sri Lanka has reduced its traffic accident rate and fatalities per 100,000 population, as shown in Figure 1. However, changes in these statistics have been rolling in nature and show an alternative upward and downward trend. The last recorded upward trend in the rate of fatalities due to traffic accidents was observed between 2012 and 2016; from 2017 to 2023, Sri Lanka reported a continuous decrease in fatality rates. In 2023, the fatality rate from traffic accidents in Sri Lanka stood at 10.4 deaths per 100,000 population [2]-[5]. During the United Nations' first decade of action for road safety (2011–2020), Sri Lanka reported an 8% reduction relative to the 2010 fatalities (baseline set by the UN). To achieve the target for the United Nations' second decade of action for road safety (2021-2031), Sri Lanka must undertake more targeted interventions to reduce the incidence of traffic accidents and to enhance road safety measures within the country [6].

Strategic deployment of robust traffic control devices, efficient roadway design practices, and rigorous law enforcement protocols assumes paramount importance in reducing fatalities resulting from traffic accidents. This is especially true in high-risk traffic accident zones, called "hotspots" [7]-[11]. These devices, practices and protocols equip law enforcers, transportation planners, and engineers with the ability to recognise where, when and why traffic accidents occur.

The first step in identifying hotspots is developing an accurate spatial pattern map of road traffic accidents. Numerous studies have leveraged geostatistical methods available in Geographic Information Systems (GIS) to identify hotspots associated with traffic accidents. Frequently used methods for determining traffic accident hotspots are Kernel Density Estimation (KDE) [12]-[14], Nearest-Neighbour Distance (NND) [15]-[17], Network-Based K Function,[18] Inverse Distance Weighting (IDW) [11] Moran's I method [19], Getis-Ord-Gi* statistic [19], [20].

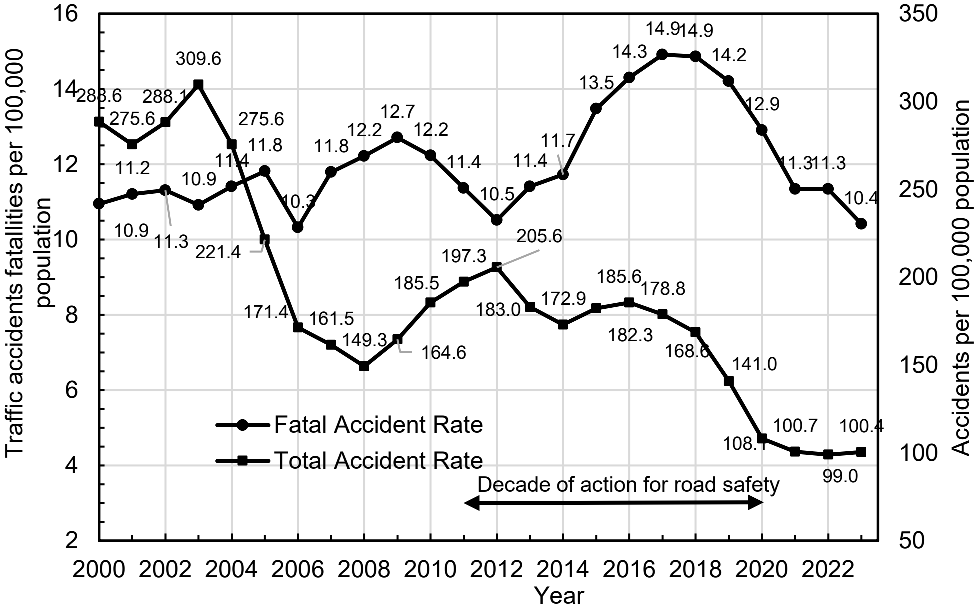


Figure 1: Variation of traffic accident and fatality rate with years

KDE generates smooth density surfaces over point event data, allowing for the computation of intensity and the identification of hotspots. The widespread adoption of KDE underscores its efficacy in resolving the spatial intricacies of point events. KDE computes the density of a given set of points in a Euclidian two-dimensional space. Although used to determine traffic accident density by previous studies, it is more suitable for two-dimensional phenomena such as rainfall, atmospheric temperature, etc. KDE used to model traffic accidents often yielded misleading conclusions and tended to overestimate cluster patterns since traffic accidents are one-dimensional events occurring along road networks. The Network Kernel Density Estimation (NKDE) method was developed in response to this limitation. Unlike conventional KDE approaches, NKDE restricts density estimation solely to the given road network. By confining the analysis to the road network, NKDE provides a more accurate depiction of traffic accident density patterns, aligning closely with the real-world distribution of incidents [13], [21].

Recent applications of NKDE have expanded significantly across various fields, particularly in criminology, transportation science, and environmental studies. NKDE has been effectively utilised to analyse crime patterns by estimating the density of crime incidents along street networks. This allows for identifying hotspots where crime is more concentrated, enabling law enforcement agencies to allocate resources more effectively [22], [23]. In transport research, NKDE is employed to assess traffic accidents, helping to visualise accident hotspots on road networks. This application aids in improving road safety measures and urban planning by highlighting areas that require intervention [23], [24]. NKDE is also applied in environmental contexts, such as analysing leaks in water distribution networks or assessing the distribution of wildlife along rivers. By estimating network event density, researchers can better understand ecological impacts and resource management [24], [22].

Recent advancements in NKDE methodologies include the development of packages like ‘spNetwork’ in R, which provides tools for performing NKDE efficiently. This package allows for analysing various types of network data, including simple, discontinuous, and continuous estimators, facilitating the study of different phenomena on networks [24], [22]. Moreover, the introduction of visualisation tools, such as ‘PyNKDV’, enhances the ability to represent network density visually, making it easier for researchers and practitioners to interpret complex data sets [23].

The versatility of NKDE makes it a valuable tool for analysing spatial data constrained to networks. Its applications in criminology, transportation, and environmental studies demonstrate its importance in addressing real-world problems through enhanced data analysis and visualisation techniques. As methodologies continue to evolve, NKDE is likely to see even broader applications and improvements in efficiency and accuracy [21], [25], [26], [27].

The study aims to delineate the density patterns of traffic accidents within the Galle district utilising the NKDE method. Through this approach, the study seeks to gain actionable insights into the spatial dynamics of traffic accidents [6].

The specific objectives of this study are twofold: firstly, to accurately plot the spatial distribution pattern of traffic accident intensity and, hence, locate traffic accident hotspots within the Galle district, and secondly, to determine the correlation between road attributes and traffic accident intensity, as revealed through NKDE techniques. Through the realisation of these objectives, valuable insights into the interplay between road characteristics and traffic accident occurrence would be revealed.

2. METHOD OF STUDY

2.1. Selection of Study Area

Narrowing the scope of study to the Galle district streamlines the analysis process while maintaining a meaningful representation of traffic accident patterns within a distinct locale. This targeted approach enables a more focused examination of road attributes and traffic accident densities, facilitating a clear understanding of the factors contributing to traffic accident occurrence within the Galle district.

2.2. Data

Traffic accident data from the Sri Lanka Traffic Police (SLTP) from 2009 to 2013 was used in this study. The traffic accident data were georeferenced according to the Kandawala projected coordinate system. The traffic accident data had numerous fields. Two fields, the east coordinate and the north coordinate of the accident locations, are needed for this study. Road network information was procured from the Survey Department of Sri Lanka (SDSL) to complement the traffic accident data. Since road network and traffic accident data use the same projected coordinate system, combining the two into a single map is easy. Figure 2 shows the locations of traffic accidents that occurred in Galle districts over the years.

2.3. Data Pre-Processing

Initially, data points found to be erroneous, and incomplete were removed from further analysis. This meticulous process aimed to minimise errors and discrepancies in the dataset, thus enhancing the validity of the subsequent findings. The valid traffic accident data were transformed into shapefiles using ArcGIS software.

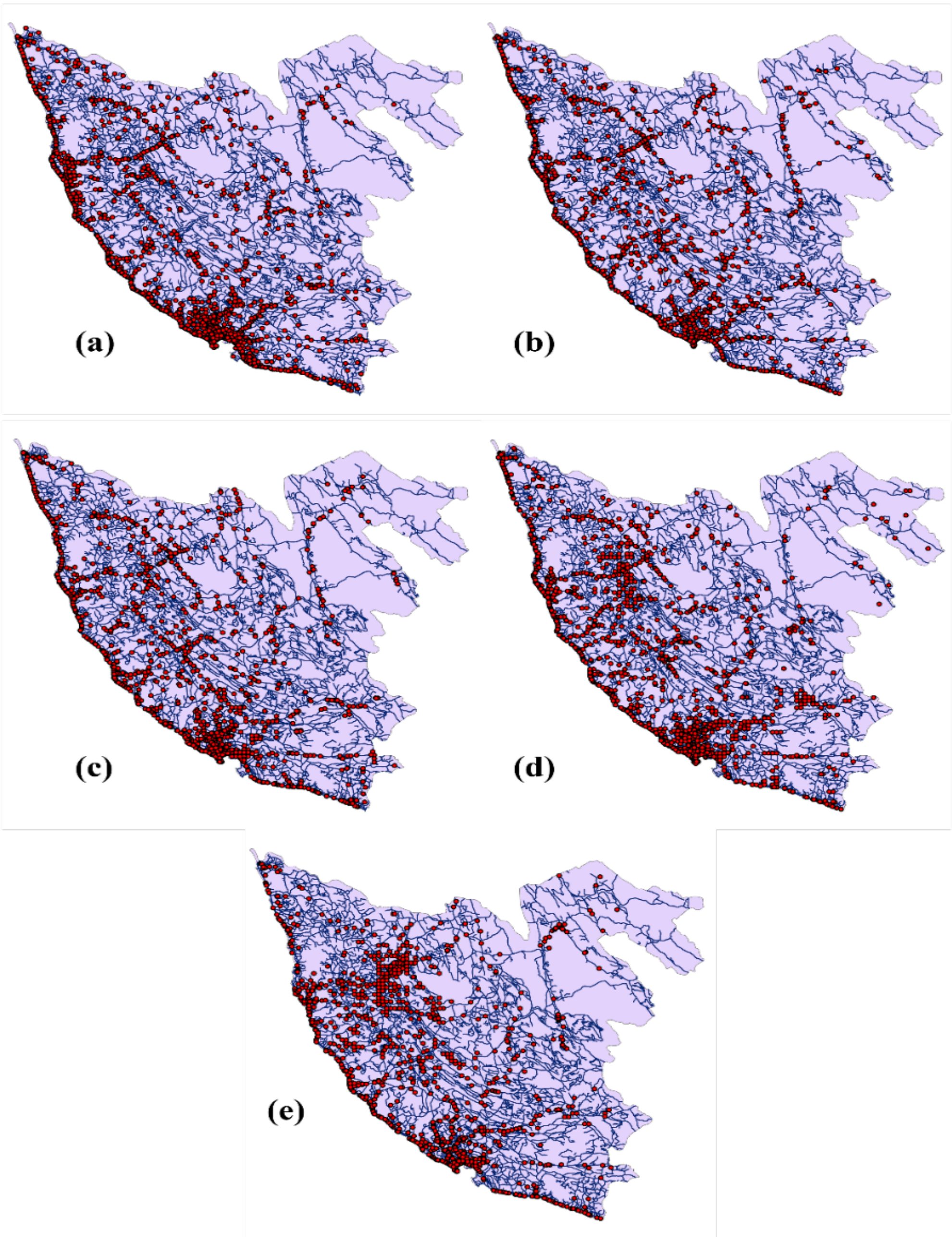


Figure 2: Traffic accident locations of Galle district in past years; (a) 2009 (b) 2010 (c) 2011 (d) 2012 (e) 2013

Source: Compiled by authors, using data from SDSL and SLTP

2.4. Hotspots Analysis using NKDE

Two KDE method variants, 'Planer Kernel Density Estimation (PKDE) and 'Network Kernel Density Estimation (NKDE), can be identified. The latter, NKDE, extending from PKDE, was introduced by Xie and Yan [13],[19], [21]. While PKDE is tailored for identifying hotspots within a 2-D homogeneous Euclidean space embedded within a region, NKDE operates within a 1-D linear space defined by a road network. A past study found that PKDE tends to overestimate density values, potentially leading to erroneous conclusions, unlike NKDE [21].

NKDE methods generate a smooth density surface over point events and facilitate the visualisation of spatial patterns and the computation of intensity, thereby enabling the identification of hotspots. NKDE plays a pivotal role in elucidating the spatial dynamics of traffic accidents, empowering researchers and policymakers with actionable insights to address road safety challenges and enhance transportation infrastructure planning and management. [12], [21], [28]-[30]. Its seamless integration within GIS platforms further enhances its appeal [21].

NKDE requires a 'kernel function' (k), 'lixel size' and 'bandwidth' (r). The 'Kernel function' used in NKDE incorporates a distance decay effect, whereby points farther from the kernel centre are assigned lower weights. NKDE assigns a proportion of the kernel density to the neighbourhood surrounding each traffic accident point, creating a continuous and smooth density surface over the network. A kernel function $k(s, z)$ is defined as a function that takes two nodes s and z from a network and returns a scalar value that represents the influence/weight of the node z on s. The kernel function is typically based on the distance between these two nodes, often utilising the shortest path distance (d_{is}). The kernel function is symmetric ($k(s, z) = k(z, s)$), non-negative ($k(s, z) \geq 0$), normalised ($\sum_{z\in(d_{is}<r)} k(s, z) = 1$) and assumes a value of 0 for all nodes located further than the bandwidth ($k(s, z)_{z\in(d_{is}\geq r)} = 0$).

During the NKDE process, the road network gets segmented; these segments are called 'lixel'. A lixel is defined as a linear equivalent of a pixel, specifically adapted for use on a network. It represents network segments from which density estimates of events such as accidents, crimes, or other occurrences are calculated. A lixel corresponds to a specific length of a line segment in the network. The centres of lixels serve as sampling points where the density of events will be estimated. KDEs for traffic accident points within the bandwidth (r) are aggregated into these lixels.

As a node z traverses along the network, the influence/weight exerted on the node s will vary with the distance following the topology of the 1-D network. The influence exerted on node s e usually diminishes with the distance. Bandwidth (r) is defined as the radial distance from the node s beyond which the influence of node z is taken to be zero. Using a small bandwidth leads to under-smoothing, where the density plot reflects the individual data points too closely, often resulting in a plot that appears jagged or noisy. This can obscure the actual underlying distribution and may lead to misinterpretation of the data. Conversely, an extensive bandwidth results in over-smoothing, where the density estimate becomes too generalised, potentially masking essential features such as multimodality. In such cases, distinct peaks in the data may be flattened into a single mode, leading to a loss of critical information about the distribution [31]-[34]. Cell size, bandwidth and Kernal functions used in the literature are summarised in Table 1.

Table 1: Summary of literature on cell sizes, bandwidths and kernel functions

No	Cell Size (m)	Bandwidth (m)	Kernel Function	Source
1	5, 10, 50, 100	20,100, 250, 500	Gaussian Quartic	[21]
2	10	100	Quartic	[13]
3	No information	200	Equal-split Equal-split continuous	[35]
4	100	1000	No information	[28]
5	40	700	Quartic	[29]
6	20	200	Quartic	[36]
7	100	200	No information	[12]
8	400	800	Normal Quartic	[30]
9	No information	100	Epanechnikov	[37]
10	250	1000	Quartic	[38]

The kernel density estimation at a network-constrained location s ($\lambda(s)$) is defined as in Equation (1).

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

Where $k\left(\frac{d_{is}}{r}\right)$ is the kernel function which indicates the weight/ influence exerted at the node s by a node i which is located at a distance of d_{is} measured along the network and within the bandwidth (r).

Okabe et al. [35] introduced a kernel method to calculate KDE values within a network, mainly focusing on density estimation at junction nodes. The study devised two unbiased functions: the equal-split discontinuous and the equal-split continuous kernel functions [39], as opposed to employing conventional biased kernel functions such as Gaussian, quartic [21], and Epanechnikov [40], among others. The equal-split continuous kernel function can be defined as in Equation 2.

$$K_p(x) \begin{cases} k(x) & \text{for } 0 \geq x \geq (2d - x) \\ k(x) - \frac{n-2}{n} K_p(2d - x) & \text{for } (2d - x) \geq x \geq d \\ \frac{2}{n} k(x) & \text{for } d \geq x \geq h \end{cases} \quad \text{-----------------------(2)}$$

Many studies have used the two unbiased kernel functions to identify network-constrained crimes [41], [36], taxi trips [42], Airbnb locations [43] and building renovation [44]. SANET software [35] was used in this study to determine the NKDE values.

2.4.1. Selection of cell size and bandwidth

The NKDE process relies on two key parameters: 'cell size' and 'bandwidth'. Defining appropriate values for these parameters is crucial for accurately representing traffic accident clusters and achieving a smoother distribution of density estimation, thereby facilitating the generation of colour-coded maps indicating hotspot locations. This selection involves a trade-off between computation time and sample size. Larger cell sizes expedite processing time, whereas smaller ones necessitate increased computation time. Furthermore, the bandwidth parameter in the NKDE method plays a critical role in determining the extent of the search area.

In this study, various combinations of bandwidth and cell size values were experimented with to determine optimal values. Specifically, cell sizes ranging from 10 m to 40 m, with 20 m intervals, and bandwidths ranging from 100 m to 1,000 m at 100 m intervals were experimented with.

2.4.2. Identification of hotspots

Past studies have adopted varied approaches due to the absence of a universally established threshold value for delineating hotspots. Line segments with values exceeding a predetermined threshold [14], the top 5-10% of high crash locations [45], and crash rates per mile or intersection [11] were used as hotspots. The goal is to focus limited resources on the most problematic areas. Ultimately, hotspot identification requires a combination of spatial analysis, crash frequency and severity, and engineering judgment. The KDE values for lixels were first determined year-by-year to identify hot spots. Yearly KDE estimates for lixels were combined to determine an accident intensity index (AII) for each lixel as shown in Equation (3).

$$AII = \sum_{i=1}^{n} \omega_i \times NKDE_i \quad \text{--(3)}$$

Where ω_i is the weightage given to NKDE of i^{th} year. In this study, constant weightage was used. $NKDE_i$ is the NKDE for the lixel in concern for i^{th}year

Hotspots were identified based on the distribution of I, it is expected that I would have some right-skewed distribution with a long tail. All lixels having I values less than 4.5 standard deviations (sd) away ($AII < \mu + 4.5sd$) from the mean were classified as "Cold Zones". While these cold zones have traffic accidents, they need the least attention. All lixels having AII values more than 4.5 sd away from the mean ($AII \geq \mu + 4.5sd$) were classified as "Hot Spots". Hot spots were further divided into 'Yellow Zones' having located within 4.5 to 5.5 sd from the mean ($\mu + 4.5sd \leq AII < \mu + 5.5sd$), 'Orange Zones' located within 5.5 to 6.5 sd from the mean ($\mu + 5.5sd \leq AII < \mu + 6.5sd$), and 'Red Zones' having I values more than 6.5 sd away from the mean ($AII \geq \mu + 6.5sd$).

2.5. Development of Regression Model

In several scenarios, the target variable consistently holds positive values, never reaching zero or dipping into negativity. Typically, these values concentrate towards the lower end of the spectrum, with only a tiny fraction attaining high values. Such a target variable deviates from the normality assumptions of conventional linear regression models [46], [47]. Yet, since the values are invariably positive and not strictly integers, they don't adhere to a data generation process based on Poisson or Negative Binomial distributions. Instead, they align with a process governed by a Gamma distribution, permitting estimation through linear regression techniques within the generalised linear model framework.

Lixels located on the A-2 road network within the Galle district were selected to develop the Gamma regression model. The study area encompassed a total of 3,634 lixels. Independent variables such as population, intersection density, bendiness, and urbanisation condition were designated for each lixel. The independent variables used in the development of the regression model and their definitions are shown in Table 2. This regression model, being a macro-level model, has to incorporate variables that apply to a larger area.

Table 2: The independent variables used in the model

Variable	Definition	Nature
Population	The population of the GN division in which the lixel is located	Continues
Intersection density	Intersections per unit length of lixel	Continues
Bendiness	Indicate whether there is a bend or not in the lixel	Dummy

Variable	Definition	Nature
Urbanisation condition	Indicate whether the lixel is in an urban or rural area	Dummy

The probability density function of a random variable y following a Gamma distribution with parameters ν and λ can be defined as:

$$f_{\nu,\lambda}(Y) = \frac{\lambda^{\nu}}{\Gamma(\lambda)} y^{\nu-1} e^{-\lambda y} \quad \text{---(4)}$$

Where λ (> 0) is the shape, $\nu(> 0)$ is the dispersion. The mean and the variance of the distribution are given by $\mu = E(Y) = \frac{\nu}{\lambda}$ and $V(Y) = \frac{\nu}{\lambda^2} = \frac{\mu^2}{\nu}$ respectively. The Gamma distribution is taken to have a constant shape parameter, and the regression equation is defined as shown in Equation (5).

$$g(\mu_i) = x'_i\beta \quad \text{---(5)}$$

Where, $g(.)$ is the link function, $\boldsymbol{\beta}$ is the regression parameters and $\boldsymbol{x_i}$ is the value of ith independent variable. There are three types of link functions available for the Gamma regression model they are (i) logarithmic function, $g(\mu) = log(\mu)$; (ii) the identity function, $g(\mu) = \mu$, and (iii) the inverse function $g(\mu) = 1/\mu$.

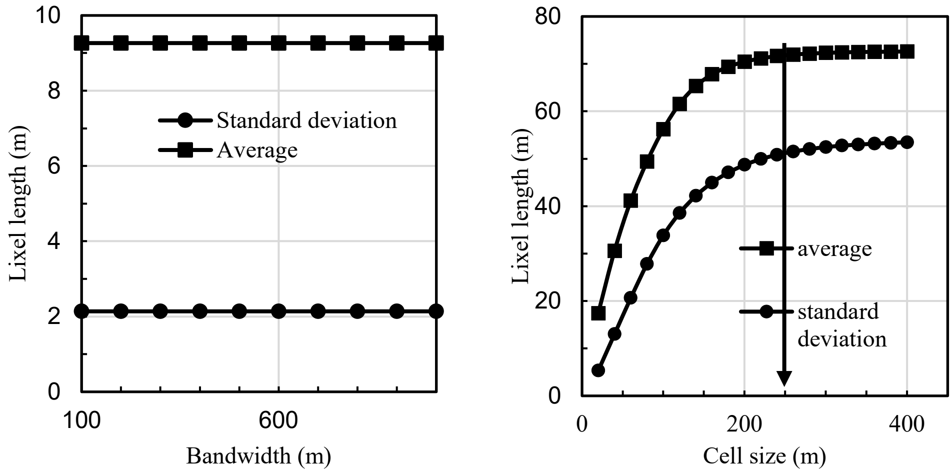
3. RESULTS AND DISCUSSION

3.1. Bandwidth and Cell Size

The optimal cell size was determined based on the mean and standard deviations of lixel lengths resulting from NKDE analysis. As shown in Figure 3 (a), the average and standard deviation of lixel lengths remained constant across bandwidths ranging from 100 m to 1000 m. This suggests that the lixel length for a given cell size is independent of bandwidth.

Next, the variation of the average and standard deviation of lixel lengths under various cell sizes was assessed while maintaining a constant bandwidth of 500 m. The analysis revealed that the average length increased with an increment in cell size up to 250 m. However, this increment gradually diminished, indicating the attainment of a stable value. Consequently, the average length exhibited a nearly constant value beyond a cell size of about 250 m (Figure 3(b)). Thus, it can be inferred that the optimum cell size for the study area is 250 m. Considering the literature, a bandwidth of twice the cell size (500 m) was chosen for this study.

3.2. Hotspot Identification



(a) **(b)**

Figure 3: Variation of lixel length with bandwidth at a cell size of 50m and cell size using traffic accidents of 2009

The AII values were found to be best represented by a Gamma distribution with a shape parameter (α) of 1.0 (i.e. an exponential distribution) and rate parameter (β)of 0.0888. A total of 644 lixels were found to have accident intensity index values exceeding ($\mu + 4.5sd =$) 138.585, hence, these lixels are classified as traffic accident hotspots. Out of which 275 lixels fall under the yellow zone, 133 lixels fall under the orange zone, and 236 lixels fall under the red zone. Figure 4 shows the histogram of the accident intensity index values overlaid with the Gamma distribution and different zones of hot spots.

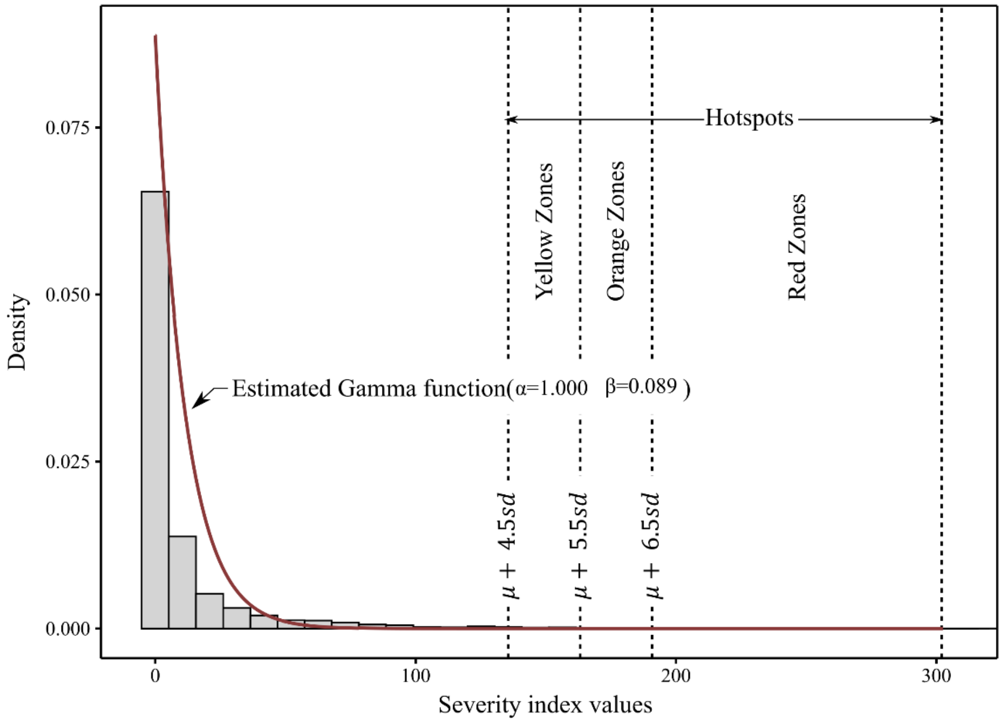


Figure 4: Histogram of accident intensity index values with gamma distribution and hotspots overlaid

Figure 5 shows the resulting traffic accident hotspots on the Galle district map. Four areas (group of lixels) were identified as red zones. They are discussed in detail below.

Area 1 is situated in the Habaraduwa locality between 129th and 130th kilometer markers on the A2 highway. This zone is approximately 310 m long, encompassing notable landmarks such as the Habaraduwa railway station, Koggala beach, Sea turtle farm, and Hatchery.

Areas 2 and 3 are situated in the Dewata locality, near Galle town, with 310 m and 410 m lengths, respectively. These areas are characterised by prominent features such as the Galle harbour, Galle fort, Galle railway station, Old Mahamodara maternity hospital, CBD of Galle, and Galle police station.

Area 4 is located in the Hikkaduwa locality between the 98th and 99th kilometre markers on the A2 highway, approximately 400 m long, and includes features such as the Hikkaduwa railway station, Hikkaduwa beach, and the Coral Sanctuary boat ticket centre.

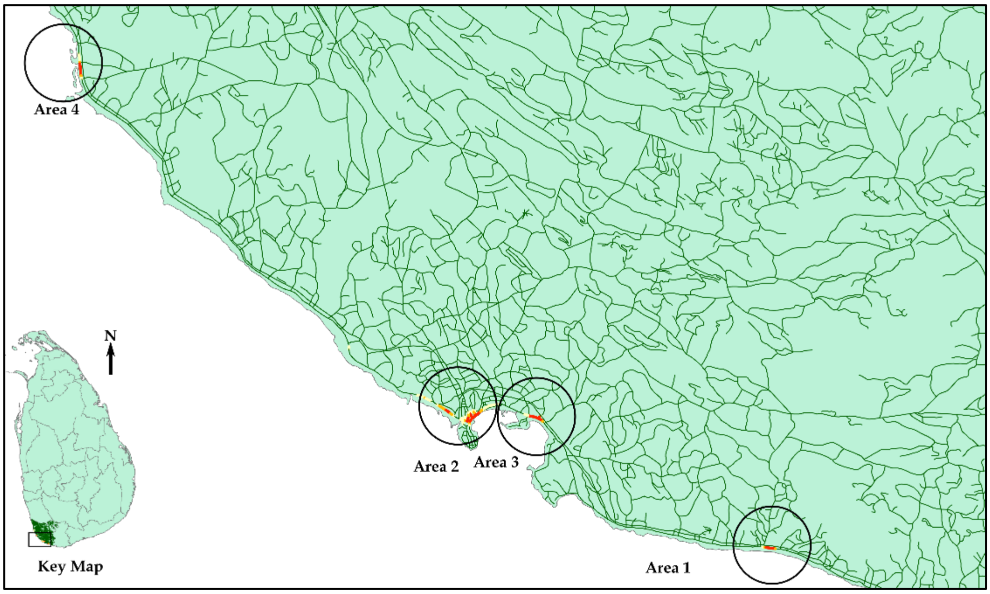


Figure 5: The traffic accident hotspots map in Galle district

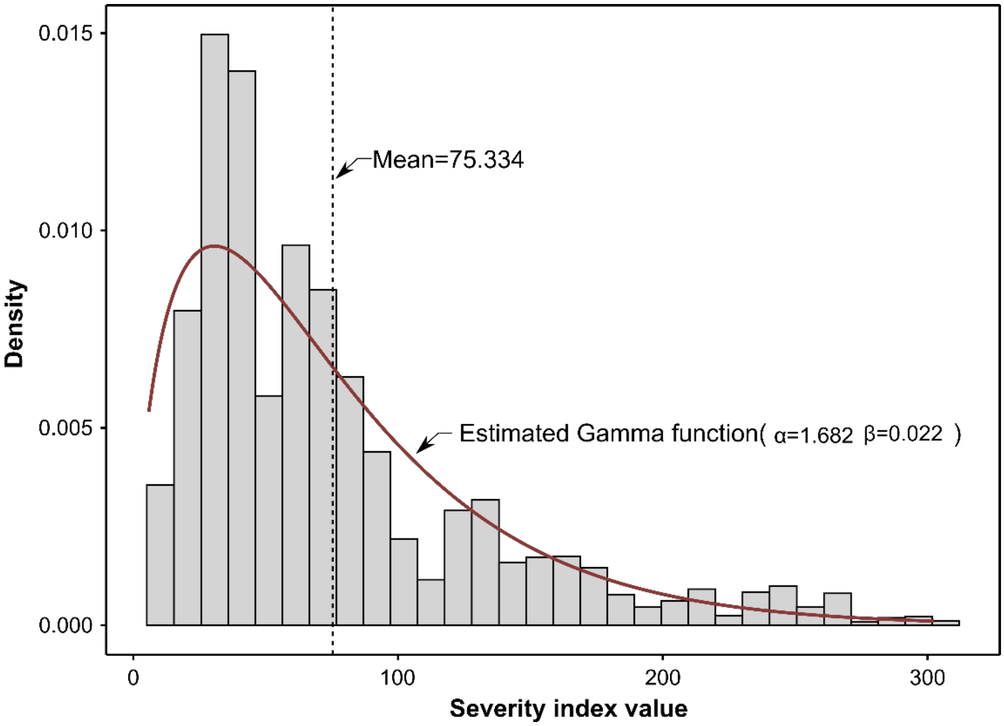


Figure 6: Histogram of all values of A2 road with Gamma distribution overlaid

3.3. Gamma Regression Model

Lixels of the A2 highway were separated from the district data used in the Gamma regression model. The analysis considered a total of 3,634 lixels. AII of A2 highway has a mean value of 75.334 and a standard deviation of 58.084. Figure 6 shows the histogram of AII values of the A2 highway with overlaid Gamma distribution. Three full Gamma regression models and three intercept-only Gamma regression models were developed. A multiple linear regression model was developed for comparison purposes.

Table 3: Model comparisons with AIC, BIC, and Bayes factor

No.	Model	AIC	BIC	Bayes Factor
1	The Liner regression model	3,7788	3,7825	0
2	**Intercept only Gamma Model**			
	with logarithmic link function	3,7818	3,7831	-2.755
	with identity link function	3,7818	3,7831	-2.755
	with inverse link function	3,7818	3,7831	-2.755
3	**Full Gamma Model**			
	with logarithmic link function	3,6200	3,6237	794.184
	with identity link function	3,6217	3,6254	785.752
	with inverse link function	3,6237	3,6274	775.648

Table 3 shows the model comparison using the AIC, BIC, and Bayes factors. For a better model, AIC and BIC should be minimal; the Bayes factor should be the most significant [48]. The full Gamma model with logarithmic link function showed the least AIC and BIC values but the most significant Bayes factor. Hence, all three parameters agree that it is the best model out of the above.

Model estimates of the Gamma regression model with logarithmic link function are shown in Table 2. Accordingly, all four independent variables showed a significant correlation ($p < 0.001$) with the dependent variable ($\boldsymbol{AII}$). With an increase of 1,000 in the population, traffic accident intensity increased significantly (p<0.001) by 25% ($EXP[0.0002 \times 1000] = 1.25$). This is as expected; increased population in an area through which the road is running would tend to increase roadside activities and AADT. With the increase of a junction, the traffic intensity decreased significantly (p<0.001) by 29% ($EXP[-0.3431] = 0.71$). This outcome, on the first look, seems counterintuitive, but the fact that most of the A2 road in Galle district runs through rural areas needs to be considered. In a rural section, the travelling speed of vehicles may be increased on an intersection-free road section, leading to higher accident intensity index. If a bend is present in the lixel traffic, accident intensity decreases significantly (p<0.001) by 12% ($EXP[-0.1325] = 0.88$). The same argument was presented for the junction density is valid for interpreting the results of the bendiness. Bend-free Road sections encourage drivers to travel faster, increasing the traffic accident intensity index. If a lixel falls within an urban area, traffic accident intensity increases significantly (p<0.001) by 131% ($EXP[0.8358] = 2.31$). Urbanisation is represented by one dummy variable that manages to catch the effect of urban built-up areas. This outcome is as expected, although the travel speeds would be lower, and increased exposure might be the main reason for this increase.

3.4. Practical Scenario

The Gamma regression model provides valuable insights that can be directly applied to enhance road safety. Several practical scenarios illustrate its potential for real-world impact:

Urban Traffic Management: City planners and traffic police can use this model to identify high-risk areas. Authorities can implement preventive measures, such as better signage, optimised traffic signal timings, or additional road safety features like speed bumps and pedestrian crossings in high-risk areas.

Targeted Law Enforcement: Traffic police can allocate resources more efficiently using the model's findings. Police patrols and traffic monitoring can be increased in this high-risk area to deter dangerous driving behaviours and ensure timely interventions.

Insurance Risk Assessment: Insurance companies could incorporate the model's predictions into their risk assessment procedures. Insurers can adjust premiums by identifying geographical areas or driver demographics associated with higher accident risk.

Infrastructure Improvements: Road Development Authority can use the model to prioritise road infrastructure upgrades. For instance, the identified hot spot is further analysed, and investments for improvements can be targeted to reduce risks in areas.

Driver Education and Awareness Campaigns: The model's insights can also inform public awareness campaigns by highlighting specific risk factors, such as speeding in rainy conditions or high accident rates in areas with poor visibility. Tailored driver education programs can be developed to address these issues, potentially reducing accident rates.

Table 4: Gamma regression model estimates with log link

Variable	Estimate	Std. Error	t value	P(>|t|)	Significance
Intercept	3.0250	0.0369	81.95	0.00	***
Population	0.0002	0.0000	11.986	0.00	***
Junction density	-0.3431	0.0450	-7.625	0.00	***
Bendiness	-0.1325	0.0233	-5.698	0.00	***
Rural/Urban	0.8358	0.0296	28.215	0.00	***
Dispersion parameter	0.3316				

Real-Time Navigation: Integrating identified accident hotspots into real-time navigation systems could give drivers advanced warnings about high-risk areas. This would allow drivers to take precautionary measures, such as slowing down or choosing alternative routes, reducing the likelihood of accidents in high-risk zones.

4. CONCLUSIONS AND FUTURE DIRECTIONS

The primary conclusions of this paper are as follows. This study employed the NKDE to identify the traffic accident hotspots. It was determined that a cell size of 250 m and a bandwidth of 500 m is suitable for the Galle district for developing NKDE values. Cell size and bandwidth are loosely related to the network geometry; hence, proposed values are considered suitable only for the Galle district road network. Appropriate values must be determined when NKDE is applied to other geographical areas.

NKDE values obtained for each year were combined into a traffic accident intensity index (AII) and was used to delineate the traffic accident hotspots into three zones: yellow, orange, and red. Four distinct red zones were observed along the A2 highway. This observation is justified by these areas' high population density and public facilities, resulting in heightened traffic volumes and increased traffic accident risks.

This paper further concluded that a Gamma model with a log link was the best model explaining the relationship between traffic accident intensity index (AII) and road attributes. Upon examination of the unstandardised coefficients, it was observed that both population density and urbanisation variables significantly increase the traffic accident density. However, the impact of urbanisation on traffic accident density was found to be more pronounced. Therefore, it can be concluded that urbanisation emerges as the most influential factor affecting traffic accident density among the four variables considered in this study.

The main limitation of this study is the recency of traffic accident data. This study used relatively old traffic accident data. In contrast, the potential and the validity of the procedures developed in this study remain binding. Secondly, this study used an year as the temporal unit for traffic accident aggregation to reduce the computation time for NKDE values. A smaller temporal unit, such as a quarter, may help to reveal seasonal variations in the traffic accident patterns. Any study that emulates the procedures used in this study must define suitable bandwidth values according to the specific characteristics of the network in the study area.

REFERENCES

[1] Global status report on road safety 2023. Geneva: World Health Organisation; 2023.

[2] A. Somasundaraswaran, "Accident statistics in Sri Lanka," *IATSS research*, vol. 30, Art. no. 1, 2006, doi: https://doi.org/10.1016/S0386-1112(14)60162-X.

[3] S. D. Dharmaratne, A. U. Jayatilleke, and A. C. Jayatilleke, "Road traffic crashes, injury and fatality trends in Sri Lanka: 1938-2013," *Bulletin of the World Health Organisation*, vol. 93, pp. 640–647, 2015.

[4] A. S. Kumarage, S. Wickramasingha, and M. Jayarathna, Analysis of road accidents in Sri Lanka: Analysis for 2002. Kelaniya, Sri Lanka: Department of Civil Engineering University of Moratuwa, 2003.

[5] S. Kumarage, C. Abeygoonawardena, and R. Wijesundera, "Identifying causal factors of traffic accidents in Sri Lanka," *94th Annual Proceedings of Institute of Engineers Sri Lanka*, 2000.

[6] 2020 Summit of the United Nations, GA Res 74/299, UN GAOR, 74th session, Agenda Item 12, UN Doc A/RES/74/299 (2 September 2020, adopted 31 August 2020).

[7] G. Shafabakhsh and Y. Sajed, "Identification methods of accident hotspots and providing a model for evaluating the number and severity of accidents on roadways," *International Journal of Transportation Engineering*, vol. 10, Art. no. 1, 2022.

[8] S. A. Aldala'in, N. S. Abdul Sukor, M. T. Obaidat, and T. S. B. Abd Manan, "Road Accident Hotspots on Jordan's Highway Based on Geometric Designs Using Structural Equation Modeling," *Applied Sciences*, vol. 13, Art. no. 14, 2023.

[9] E. K. Pour, K. Mohaddes, and M. Saadati, “Using road traffic accident data for hotspots identification: a case study in Marand, Iran,” *Journal of Injury and Violence Research*, vol. 11, Art. no. 4, 2019.

[10] H. Zhu, Y. Zhou, and Y. Chen, “Identification of potential traffic accident hot spots based on accident data and GIS,” *in MATEC Web of Conferences*, 2020, vol. 325, p. 1005.

[11] N. P. Kodippili, M. Akther, and G. Naveendrakumar, “Accident Hotspots in Southern Expressway of Sri Lanka: Interpolation Evaluation using GIS,” *Advances in Technology*, pp. 191–206, 2021.

[12] T. K. Anderson, “Kernel density estimation and K-means clustering to profile road accident hotspots,” *Accident Analysis & Prevention*, vol. 41, Art. no. 3, 2009.

[13] Z. Xie and J. Yan, “Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach,” *Journal of Transport Geography*, vol. 31, pp. 64–71, 2013.

[14] J. Benedek, S. M. Ciobanu, and T. C. Man, “Hotspots and social background of urban traffic crashes: A case study in Cluj-Napoca (Romania),” *Accident Analysis & Prevention*, vol. 87, pp. 117–126, 2016.

[15] W. Alkhadour, J. Zraqou, A. Al-Helali, and S. Al-Ghananeem, “Traffic Accidents Detection using Geographic Information Systems (GIS),” *Traffic*, vol. 12, Art. no. 4, 2021.

[16] M. Kamran, W. Chaudhry, B. O. Taiwo, S. Hosseini, and H. Rehman, “Decision intelligence-based predictive modelling of hard rock pillar stability using K-nearest neighbour coupled with grey wolf optimisation algorithm,” *Processes*, vol. 12, Art. no. 4, 2024.

[17] M. Haddad et al., “Visual hazardous models: a hybrid approach to investigate road hazardous events,” *Accident Analysis & Prevention*, vol. 200, p. 107556, 2024.

[18] S. Feng, A. Wang, Z. Tian, and S. Park, “Exploring the correlation between hard-braking events and traffic crashes in regional transportation networks: A geospatial perspective,” *Multimodal Transportation*, vol. 3, Art. no. 2, 2024.

[19] W.-F. Xie, J.-K. Li, K. Peng, K. Zhang, and Z. Ullah, “The Application of Local Moran’s I and Getis–Ord G i* to Identify Spatial Patterns and Critical Source Areas of Agricultural Nonpoint Source Pollution,” *Journal of Environmental Engineering*, vol. 150, Art. no. 5, 2024.

[20] N. Manap, M. N. Borhan, M. R. M. Yazid, M. K. A. Hambali, and A. Rohan, “Identification of hotspot segments with a risk of heavy-vehicle accidents based on spatial analysis at the controlled-access highway,” *Sustainability*, vol. 13, Art. no. 3, 2021.

[21] Z. Xie and J. Yan, “Kernel density estimation of traffic accidents in a network space,” Computers, environment, and urban systems, vol. 32, Art. no. 5, 2008.

[22] J. Gelb, "spNetwork: A package for network kernel density estimation," *R Journal*, vol. 13, no. 2, 2021.

[23] T. N. Chan, R. Zang, P. L. Ip, L. H. U, and J. Xu, "PyNKDV: An efficient network kernel density visualisation library for geospatial analytic systems," *in Companion of the 2023 International Conference on Management of Data*, 2023, pp. 99-102.

[24] J. Gelb and P. Apparicio, "Temporal network kernel density estimation," *Geographical Analysis*, vol. 56, no. 1, pp. 62-78, 2024.

[25] Z. Wang, Z. Su, Y. Deng, J. Kurths, and J. Wu, "Spatial network disintegration based on kernel density estimation," *Reliability Engineering & System Safety*, vol. 245, p. 110005, 2024.

[26] A. Karatzoglou, "Applying network kernel density estimation (NKDE) and temporal network kernel estimation (TNKDE) for generating safer routes," in Proceedings of the 15th ACM SIGSPATIAL *International Workshop on Computational Transportation Science (IWCTS '22)*, 2022, pp. 1-10.

[27] S. Węglarczyk, "Kernel density estimation and its application," *ITM Web Conference*, vol. 23, 2018, p. 00037.

[28] A. Blazquez and M. S. Celis, “A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile,” *Accident Analysis & Prevention*, vol. 50, pp. 304–311, 2013.

[29] Ö. Kaygisiz, Ş. Düzgün, A. Yildiz, and M. Senbil, “Spatio-temporal accident analysis for accident prevention in relation to behavioural factors in driving: The case of South Anatolian Motorway,” *Transportation research part F: traffic psychology and behaviour*, vol. 33, pp. 128–140, 2015.

[30] L. Thakali, T. J. Kwon, and L. Fu, “Identification of crash hotspots using kernel density estimation and kriging methods: a comparison,” *Journal of Modern Transportation*, vol. 23, Art. no. 2, 2015.

[31] S. Abramson, "On bandwidth variation in kernel estimates-a square root law," *The Annals of Statistics*, pp. 1217-1223, 1982.

[32] O. Cronie and M. N. M. Van Lieshout, "A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions," *Biometrika*, vol. 105, no. 2, pp. 455-462, 2018.

[33] Gelb and P. Apparicio, "Temporal network kernel density estimation," *Geographical Analysis*, vol. 56, pp. 62-78, 2024. [Online]. Available: https://doi.org/10.1111/gean.12368

[34] J. Henderson, A. Papadopoulos, and C. F. Parmeter, "Bandwidth selection for kernel density estimation of fat-tailed and skewed distributions," *Journal of Statistical Computation and Simulation*, vol. 93, no. 12, pp. 2110-2135, 2023.

[35] A. Okabe, T. Satoh, and K. Sugihara, “A kernel density estimation method for networks, its computational method and a GIS-based tool,” *International Journal of Geographical Information Science*, vol. 23, Art. no. 1, 2009.

[36] S. Khalid et al., “Network constrained spatio-temporal hotspot mapping of crimes in Faisalabad,” *Applied Spatial Analysis and Policy*, vol. 11, Art. no. 3, 2018.

[37] M. Bíl, R. Andrášik, and Z. Janoška, "Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation," *Accident Analysis & Prevention*, vol. 55, pp. 265-273, 2013.

[38] S. Bisht and G. Tiwari, "Identification of road traffic crashes hotspots on an intercity expressway in India using geospatial techniques," *IATSS Research*, vol. 47, no. 3, pp. 349-356, 2023.

[39] Sugihara, T. Satoh, and A. Okabe, "Simple and unbiased kernel function for network analysis," *in 2010 10th International Symposium on Communications and Information Technologies*, 2010, pp. 827-832.

[40] Bíl, R. Andrášik, and Z. Janoška, “Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation,” *Accident Analysis & Prevention*, vol. 55, pp. 265–273, 2013.

[41] Milić, Z. Đurđević, S. Mijalković, and D. Erkić, “Identifying Street Hotspots Using a Network Kernel Density Estimation,” *Revija za kriminalistiko in kriminologijo/Ljubljana*, vol. 71, Art. no. 4, 2020.

[42] B. Shen, X. Xu, J. Li, A. Plaza, and Q. Huang, "Unfolding spatial-temporal patterns of taxi trip based on an improved network kernel density estimation," *ISPRS International Journal of Geo-Information*, vol. 9, no. 11, p. 683, 2020.

[43] R. M. Correia, R. Guerreiro, and F. J. da Silva Brandão, "Spatial analysis of Airbnb in Lisbon: A network kernel density estimation," *Spatial Humanities 2021*, Lisbon, Portugal, 2021.

[44] R. M. Correia, R. Guerreiro, and F. Brandão, "Data-driven spatial analysis of urban renewal. Network kernel density estimation of building renovation," *in Formal Methods in Architecture: Proceedings of the 5th International Symposium on Formal Methods in Architecture (5FMA), Lisbon 2020*, Cham: Springer International Publishing, 2021, pp. 185-195.

[45] T. H. Hayidso, D. O. Gemeda, and A. M. Abraham, “Identifying road traffic accidents hotspots areas using GIS in Ethiopia: a case study of Hosanna Town,” *Transport and Telecommunication Journal*, vol. 20, Art. no. 2, 2019.

[46] Y. Zhang, X.-S. Wei, B. Zhou, and J. Wu, “Bag of tricks for long-tailed visual recognition with deep convolutional neural networks,” *in Proceedings of the AAAI Conference on Artificial Intelligence, 2021*, vol. 35, pp. 3447–3455.

[47] Makansi, Ö. Cicek, Y. Marrakchi, and T. Brox, “On exposing the challenging long tail in future prediction of traffic actors,” *in Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13147–13157.

[48] R. Wetzels and E.-J. Wagenmakers, “A default Bayesian hypothesis test for correlations and partial correlations,” *Psychonomic Bulletin & Review*, vol. 19, pp. 1057–1064, 2012.