https://doi.org/10.65646/3rc20dms1s1199

IDENTIFYING ROAD ACCIDENT HOTSPOTS USING LINEAR NETWORK POINT PROCESSES: A CASE STUDY IN THE KANDY POLICE DIVISION, SRI LANKA

D. Sashinka**, R. D. Yapa and R. Punchi-Manage
Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

Road accidents are spatially constrained events that require analytical approaches that account for the underlying transport network. Traditional area-based methods often underestimate accidents along roads, as they ignore road geometry. Modeling crashes as points on a road network improves intensity estimation. Although conventional spatial point pattern techniques are widely applied in accident analysis, approaches that incorporate the linear structure of roadways are more suitable, as accidents are constrained to occur along these networks. This study aims to identify road accident hotspots using a Linear Network Point Processes (LNPP) approach, which is more suitable than traditional area-based methods for linear networks. By analyzing accident data from the Kandy Police Division in Sri Lanka, it seeks to enable precise hotspot detection and support targeted strategies for improving road safety. Accidents occur along roads; therefore, LNPP on a metric graph G = (V, E) was used, where V represents intersections and E represents road segments. The linear network kernel density estimation was used at a location x on a road segment to estimate event intensity per unit length along the network, based on the shortest path distances between observed points. Analyses used Gaussian and Epanechnikov kernels. The optimal bandwidth was selected using the Likelihood Cross-Validation (LCV) method and Scott's Rule. Diggle's edge correction method was applied to reduce boundary bias. Hence, four models were fitted. A total of 2,043 crash events were represented as a point pattern on a road network comprising 108,584-line segments and 110,361 vertices. Both Gaussian and Epanechnikov kernels produced similar density estimates on the road network for a given bandwidth. However, the LCV method, which estimates kernel density on linear networks, provided a more accurate bandwidth than Scott's rule. The smoothing bandwidths were 1.98 km and 5.07 km for the Scott and LCV methods, respectively. Residuals range from 0–45 for Scott's rule and only 0–13 with LCV. Several high-density road accident segments were found, highlighting spatial heterogeneity in accident risk along the network. LCV outperformed Scott's rule in bandwidth selection, reducing residuals and enhancing hotspot detection, demonstrating the effectiveness of LNPP for network-constrained traffic safety analysis.

Keywords: linear network point processes, point pattern analysis, road accidents, spatial processes, spatial statistic

**Corresponding Author: dilmini.sashinka@sci.pdn.ac.lk*

IDENTIFYING ROAD ACCIDENT HOTSPOTS USING LINEAR NETWORK POINT PROCESSES: A CASE STUDY IN THE KANDY POLICE DIVISION, SRI LANKA

D. Sashinka*, R. D. Yapa and R. Punchi-Manage
Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.

INTRODUCTION

Road traffic accidents represent a critical public safety challenge worldwide, with substantial impacts on human life and economic growth. The use of a linear network approach in road traffic accident analysis is motivated by several factors, such as the fact that crashes occur only on roads rather than in open fields or lakes. Classical spatial methods such as Ripley's K-function and kernel density estimation, assume that events can occur anywhere in 2D space, whereas linear network methods restrict the analysis to the lines and nodes (i.e., roads). This approach reduces bias and prevents overestimation of risk area. Moreover, classical methods calculate distance between points using Euclidean measures, while network-based analysis computes distance along the shortest path of the road, which provides a more realistic representation. Road accident risk often depends on the attributes of the road (e.g., curvature, intersections, speed limit). Therefore, the spatial distribution of these accidents, inherently constrained by the underlying road network structure, makes traditional planar spatial analysis methods insufficient for precise hotspot identification (Okabe & Sugihara, 2012). Linear Network Point Process (LNPP) models provide a more suitable framework for analyzing such events (Baddeley et al., 2015). These models enable precise estimation of event intensity along the network and improve the detection of high-risk locations. This study focuses on the Kandy Police Division in Sri Lanka, and aims to detect accident hotspots through the application of LNPP methods combined with Kernel Density Estimation (KDE). The integration of appropriate bandwidth selection techniques and edge correction methods facilitates improved spatial characterization of crash patterns. The findings intend to inform targeted road safety interventions by identifying critical segments within the road network exhibiting elevated accident risk.

METHODOLOGY

This study used 2,099 road traffic crash records reported between January 2022 and March 2024, obtained from the Kandy Traffic Police Office, Sri Lanka. Road network data for Sri Lanka were sourced from the Humanitarian Data Exchange (HDX) platform. The study area covers the Kandy Police Division, defined by a bounding box that includes the relevant spatial region for this analysis. Crash points were then projected onto the road network using the 'spatstat.linnet' package (Baddeley et al., 2015) in R. The point pattern on the network, denoted by $\{x_1, x_2, \dots, x_i, \dots, x_n\} \subset L$, where L is the linear network, was analyzed for spatial intensity. Kernel density estimation adapted for linear networks was employed to estimate the intensity function $\hat{\lambda}_L(x)$ at location x ∈ L along the network, defined by $\hat{\lambda}_L(x) = \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{d_G(x,x_i)}{h}\right)$, where $d_G(x, x_i)$ is the shortest-path distance on the network between location x and x_i , h is the smoothing bandwidth, and K is the kernel function. Both Gaussian and Epanechnikov kernels were applied to determine the most appropriate kernel for density estimation. The Gaussian kernel, defined by $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$, and the Epanechnikov kernel, given by $K(u) = \frac{3}{4}(1 - u^2), |u| \leq 1$, were both implemented. The choice between these kernels was guided by comparison of density results and standard error behavior within the linear network context.

Bandwidth selection was performed using two methods. The first was Scott's Rule, which estimates bandwidth as $h_{scott} = \sigma_{scott}.n^{-1/6}$, where σ_{scott} denotes the standard deviation. The second was Likelihood Cross-Validation (LCV) method, which optimizes the bandwidth by maximizing the leave-one-out log-likelihood, $\hat{h}_{LCV} = argmax \sum_{i=1}^{n} log\lambda_i(x_i; h)$, where $\lambda_i(x_i; h)$ denotes the intensity estimated without i-th event. To reduce the bias in intensity estimation near the network boundaries, Diggle's edge correction method was applied. The analysis conducted using the 'spatstat', 'sf', and 'sp' packages in R software (R Core Team, 2023).

RESULTS AND DISCUSSION

The constructed linear road network consists of 110,361 vertices and 108,584 line segments. A total of 2,043 crash events were mapped onto this network and treated as a linear point pattern. Two bandwidth selection methods were used prior to KDE: Scott's Rule and LCV. The bandwidth estimated under Scott's Rule was approximately 1.98 km, while the LCV method produced a larger value of approximately 5.07 km.

The density maps are displayed using a color gradient, where blue indicates low accident intensity and yellow indicate segments with higher accident intensities. Gaussian and Epanechnikov kernels produced nearly identical intensity surfaces when used with the same bandwidth as shown in Figure 1.

However, the Epanechnikov kernel yielded marginally lower intensity values than the Gaussian kernel for a given bandwidth. This pattern was consistent under both bandwidth selection methods. These results indicate that the Epanechnikov kernel offers a more stable and less overestimated representation of the spatial pattern. Therefore, the Epanechnikov kernel was selected for subsequent analysis.

To select the best-fitted bandwidth, both Scott's Rule and LCV methods were compared using the Epanechnikov kernel, as shown in Figure 2. The figure reveals distinct differences between the two bandwidth selection methods.

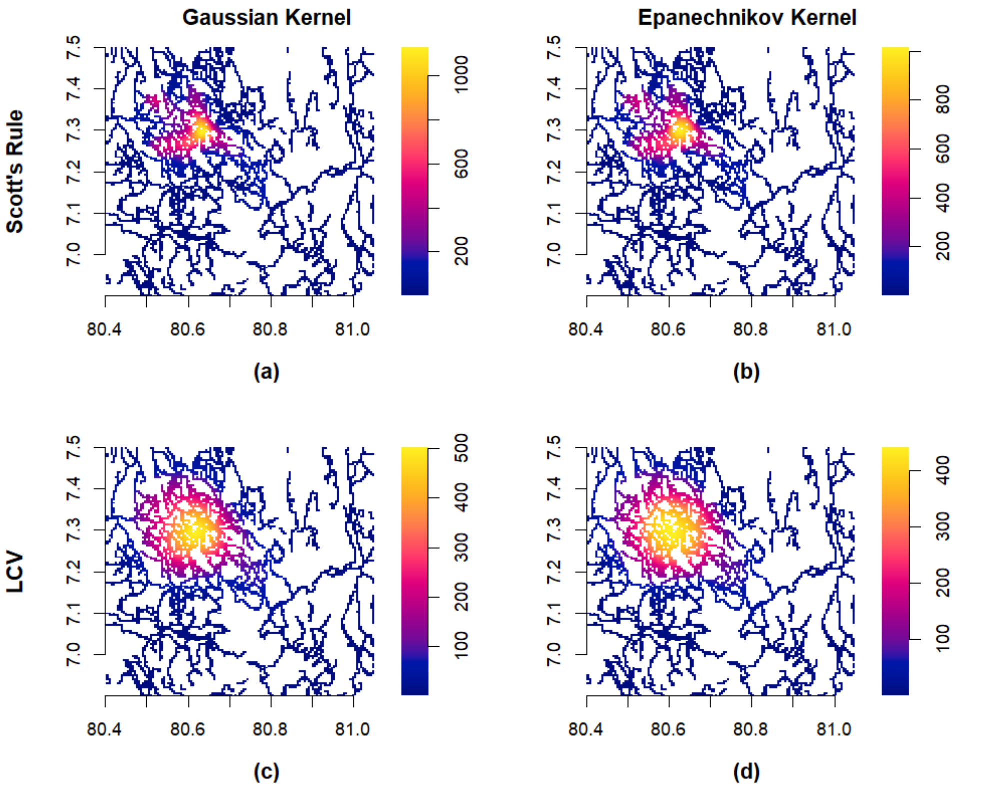

Figure 51: Kernel density estimates on the linear road network using Scott's Rule (a, b) and LCV (c, d) with Gaussian (a, c) and Epanechnikov (b, d) kernels

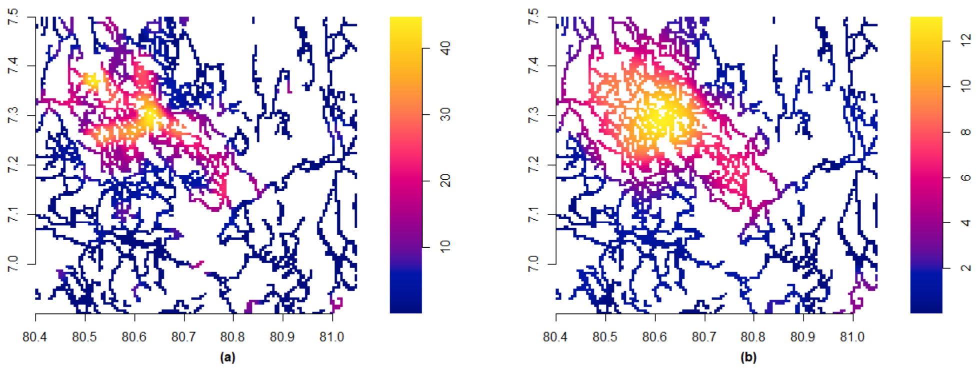

Figure 52: Standard errors of density estimates for the Epanechnikov kernel using the Scott's Rule (a) and LCV method (b)

Figure 2(a) shows that Scott's Rule produces high SE values in the range of approximately 0 to 45. This results from the smaller bandwidth, which captures finer spatial variation but increases estimator variance in high-density areas. Conversely, Figure 2(b) is based on LCV bandwidth. It presents a smoother SE surface with maximum values around 14. This reflects increased smoothing and a reduction in variance. These results indicate that the LCV method provides more precise intensity estimation with lower residual variability compared to Scott's Rule.

CONCLUSIONS

This study demonstrates the effectiveness of Linear Network Point Processes in modeling road traffic accidents within a network-constrained spatial context. By incorporating the geometry of road network and applying appropriate kernel density estimation techniques with optimal bandwidth selection, LNPP method produced more accurate representations of crash intensity. The results showed that the choice of bandwidth has a greater impact on the kernel density estimates than the choice of kernel function. Although both Gaussian and Epanechnikov kernels produced similar spatial patterns, different bandwidths changed the smoothness and accuracy of the results significantly. The Likelihood Cross-Validation method selected a larger bandwidth, which created a smoother and more stable intensity surface with lower variability. In contrast, Scott's Rule produced smaller bandwidths that led to higher variability and less stable estimates. This highlights the importance of selecting an appropriate bandwidth for reliable spatial analysis on linear networks, as it affects the quality of density estimates more than the choice of kernel. Furthermore, the analysis identified Senkadagala, Mahanuwara, Malwatta, Getambe, and Katugasthota Grama Niladhari Divisions as the primary high-risk locations. These findings underscore the importance of adopting network-aware spatial methodologies for accurate hotspot identification.

REFERENCES

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. Chapman and Hall/CRC Press.

Okabe, A., & Sugihara, K. (2012). *Spatial Analysis Along Networks: Statistical and Computational Methods*. Wiley.

R Core Team (2023). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.