COMPARATIVE SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS USING ORDINARY KRIGING AND INVERSE DISTANCE WEIGHTING

D. Sashinka*[*], R. D. Yapa and R. Punchi-Manage

Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

Road traffic accidents are a significant public safety concern, especially in urban areas where congestion and infrastructural limitations increase the likelihood of collisions. This study investigates road traffic accidents within the Kandy Police Division in Sri Lanka by applying geostatistical interpolation techniques to model the spatial distribution of 2,099 RTAs reported from January 2022 to March 2024. The analysis evaluates the suitability of two interpolation methods: Inverse Distance Weighting (IDW), a deterministic approach, and Ordinary Kriging, a model-based geostatistical method that incorporates spatial autocorrelation. For the Kriging analysis, an empirical semivariogram was developed to quantify the spatial dependence structure of the accident data, and accident counts were log-transformed to approximate normality prior to spatial prediction with Ordinary Kriging. Four theoretical models, Spherical, Exponential, Gaussian, and Matérn, were fitted to the empirical semivariogram. The Spherical model outperformed the others, followed the empirical observed points closely, and reached the sill, with a sum of squared errors of 2.59 between empirical and fitted semivariograms. It was therefore selected as the best fit for spatial prediction using Kriging. Both interpolation methods were applied on a regular 10×10 grid across the study area to estimate accident frequencies. The performance of both methods was assessed using cross-validation, and predictive accuracy was evaluated through Mean Error (ME), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Results showed that Ordinary Kriging provided slightly better predictive accuracy than IDW, with lower values for MAE (24.26 and 35.13), RMSE (40.04 and 52.87), and ME (-4.94 and -6.48), respectively. These findings reveal the effectiveness of geostatistical modeling, particularly Kriging, in identifying high-risk areas and supporting data-driven decision making. The findings of this study will support urban planners, traffic engineers, and policymakers in guiding targeted road safety measures, prioritizing infrastructure improvements, and effectively allocating resources in accident-prone zones.

Keywords: crash prediction, Geostatistics, road traffic accidents, spatial interpolation, variogram

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

COMPARATIVE SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS USING ORDINARY KRIGING AND INVERSE DISTANCE WEIGHTING

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 (RTAs) remain a significant global concern, leading to loss of life, injuries, and economic costs. In Sri Lanka, RTAs pose a major public health challenge, accounting for approximately 38,000 incidents annually, with around 3,000 deaths and 8,000 injuries (Sri Lanka's Journey to Road Safety, 2021). The spatial distribution of these incidents is often non-random and influenced by underlying factors such as road geometry and traffic volume (Gill et al., 2017). Therefore, understanding the spatial distribution of these accidents is crucial for enhancing traffic safety measures and infrastructure planning. This study aims to evaluate and compare two spatial interpolation techniques, Ordinary Kriging (OK) and Inverse Distance Weighting (IDW), for predicting RTA counts. By assessing the comparative predictive performance of these geostatistical approaches using real-world accident data, this study contributes to spatially informed decision-making frameworks that prioritize high-risk zones in urban traffic systems.

STUDY AREA & METHODOLOGY

Study area: A total of 2,099 accident records, reported between January 2022 and March 2024 within the Kandy Police Division, were obtained from the Kandy Traffic Police Division. The analysis employed the longitude, latitude, and the number of accidents recorded at each location.

Methodology: A comparative spatial interpolation analysis of RTA counts was conducted using IDW and OK (Bivand et al., 2008). Prior to interpolation, a regular 10 × 10 grid was overlaid on the study area. Within each grid cell, crash frequency was aggregated, and the centroid coordinates of these cells were computed to serve as input points for IDW and OK. For Ordinary Kriging, an empirical semivariogram $\gamma(h)$ was computed to quantify the spatial dependence structure of crash frequency counts. Since the crash frequency data deviated from normality, a log transformation was applied before semivariogram modeling. The semivariance between data points separated by a lag distance h is defined as: $\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$ where $z(x_i)$ denotes the crash frequency at location x_i, and $N(h)$ is the number of data pairs separated by distance h.

Four theoretical variogram models were fitted to the empirical semivariogram: Spherical, Exponential, Gaussian, and Matérn. The Spherical variogram model is given by: $\gamma(h) = C_0 + C\left[\frac{3h}{a} - \frac{h^3}{2a^3}\right];\ 0 < h \le a$. The exponential variogram model is given by: $\gamma(h) = C_0 + C[1 + \exp\left(-\frac{h}{a}\right)]$. The Gaussian variogram model is given by: $\gamma(h) = C_0 + C[1 + \exp\left(\left(-\frac{h}{a}\right)^2\right)]$. The Matérn variogram model is given by: $\gamma(h) = C_0 + C[1 - \frac{1}{2^{\upsilon-1}\Gamma(\upsilon)}\left(\frac{h}{a}\right)^{\upsilon} K_{\upsilon}\left(\frac{h}{a}\right)]$. Where: C_0: Nugget (discontinuity at origin), C: Partial sill (structured variance), a: Range (distance beyond which points are uncorrelated), and $C_0 + C$: Total sill, υ is the smoothness parameter, $\Gamma(\upsilon)$ gamma function, K_{υ} is the modified Bessel function of the second kind. The Matérn model reduces to the Exponential model when $\upsilon = 0.5$, and resembles the Gaussian model for larger values of υ. For the Exponential model, the effective range is often approximately $3a$, where around ~95% of the sill is reached. For the Gaussian, the effective range is approximately $\sqrt{3}a$ (Details: Cressie, 1993; Oliver & Webster, 2015).

The optimal model was selected by minimizing the residual sum of squares between the empirical and fitted semivariograms. Using the selected model, Ordinary Kriging was applied to interpolate crash frequencies at unsampled locations. The Kriging estimator $\hat{Z}(x_0)$ for a location x_0 is calculated as: $\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i \, z(x_i)$ where λ_i are the Kriging weights determined by solving the Kriging system based on the variogram. For IDW, the estimated crash frequency $\hat{Z}(x_0)$ at location x_0 is computed as: $\hat{Z}(x_0) = \frac{\sum_{i=1}^{N}\left(\frac{Z(x_i)}{d(x_0,\, x_i)^p}\right)}{\sum_{i=1}^{N}\left(\frac{1}{d(x_0,\, x_i)^p}\right)}$ where $d(x_0, x_i)$ denotes the Euclidean distance between x_0 and x_i, and $p = 2$ is the inverse distance weighting power parameter. Both interpolation techniques were executed on the same 10×10 spatial grid for a fair comparison. The implementations were carried out using the "gstat" package (Gräler et al., 2016; Pebesma, 2004) and the "sf" package (Pebesma, 2018) in R. Cross-validation was conducted for both methods by sequentially excluding each observation, fitting the model to the remaining data, and predicting at the excluded location (Cressie, 1993). The prediction error for each point was computed as $e_i = y_i - \hat{y}_i$ where y_i is the observed crash frequency at location i, and $\hat{y}_i$ is the predicted value. Predictive performance was evaluated using Mean Error (ME), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The implementations of Ordinary Kriging, Inverse Distance Weighting, and Cross-validation methods were carried out using the "gstat" package (Cressie, 1993; Pebesma, 2004) in R (R Core Team, 2023).

RESULTS AND DISCUSSION

***Semivariograms and Ordinary Kriging*:** Prior to applying ordinary kriging, the distribution of accident counts per quadrant was assessed for normality. The counts showed a right-skewed distribution, which violated the assumption of normality required for kriging. Therefore, a log-transformation with offset (log(n+1)) was applied to approximate normality. The selection of an appropriate theoretical model for fitting the empirical semivariogram was carried out by evaluating the performance of four standard models: Spherical, Gaussian, Exponential, and Matérn (Fig. 1). The Sum of Squared Errors (SSE) for each model was as follows: Spherical – 2.590; Exponential – 2.349; Matérn – 2.349; Gaussian – 4.548. Based on the SSE values, the Exponential and Matérn models provided the lowest errors. However, visual inspection of the semivariogram showed that only the Spherical model reached a sill, with its curve passing close to most points, whereas the Exponential and Matérn models continued to increase without leveling off. Beyond approximately 15,000 units, the Spherical model levels off and indicates spatial independence. Therefore, the Spherical model was identified as the most appropriate for capturing the spatial structure in the data. Since a log transformation was applied to the accident counts, the predictions were back-transformed to the original count scale. The predicted counts from the Spherical model and residuals from cross-validation are visualized (Fig. 2). The predicted counts ranged from 1-220, while the residuals ranged from -102.26 to 191.24.

Inverse Distance Weighted Interpolation: IDW interpolation was used to predict accident counts (Fig. 3). For IDW, predicted counts and cross-validated residuals are shown (Fig. 3). Predicted values varied between 8-154, and residuals ranged from -68.36 to 251.56. It highlights the similar general hotspots and spatial trends as Kriging. However, predicted accident counts are slightly lower for IDW compared to Kriging.

Cross-Validation Method: The performance of the two interpolation methods was evaluated using cross-validation. Residuals were computed as the difference between observed and predicted accident counts (Figs 2 and 3). Most of the residuals are close to zero. There are a few locations that have large residuals which need careful attention from decision-makers.

Table 1: Cross-validation performance metrics for Ordinary Kriging and IDW

	Ordinary Kriging	Inverse Distance Weighting
Mean Error (Mean Bias)	-4.94	-6.48
Mean Absolute Error (MAE)	24.26	35.13
Root Mean Squared Error (RMSE)	40.04	52.87

Performance metrics: The performance of the overall Ordinary Kriging and IDW methods was evaluated using the ME, MAE, and RMSE, calculated from the residuals obtained through cross-validation (Table 1). All three performance metrics showed slightly lower values for Ordinary Kriging than for IDW, which shows marginally better predictive accuracy. The mean error (bias) for Kriging and for IDW slightly negative. This indicates a very slight tendency to underestimate the actual values for both methods, with Kriging showing the smallest bias.

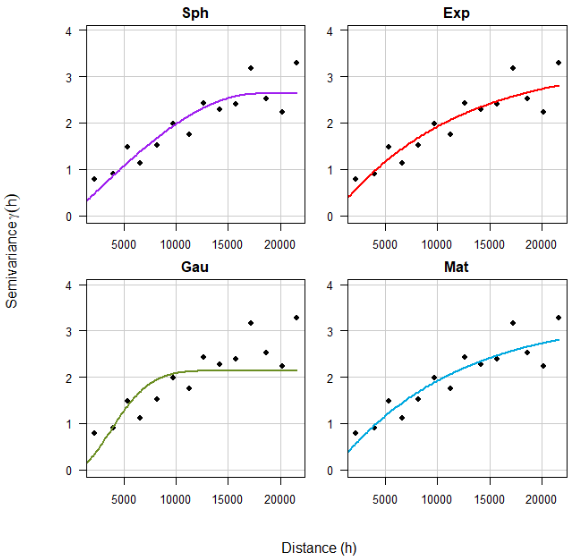

Figure 1: Experimental semivariogram with fitted Spherical, Exponential, Gaussian, and Matérn models

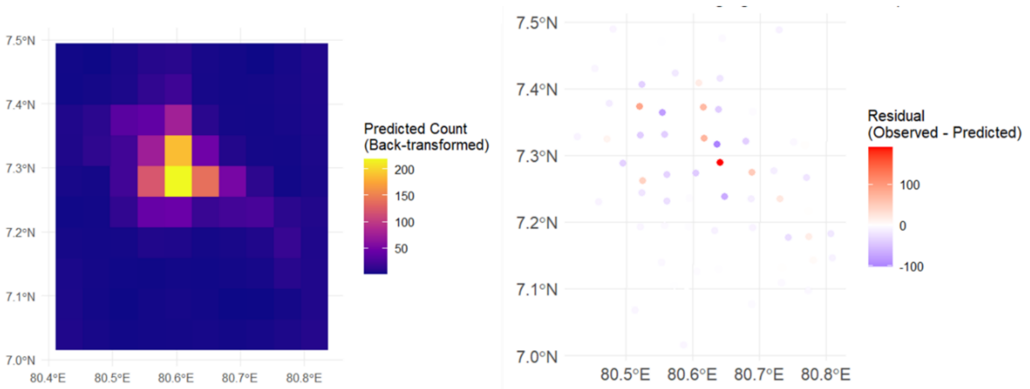

Figure 2: Ordinary kriging prediction (Spherical model) and residuals from cross-validation

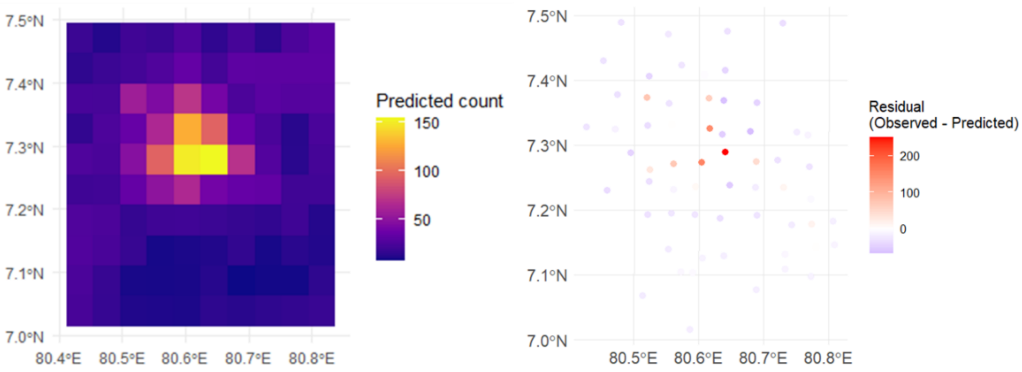

Figure 3: Inverse weighted distance prediction and residuals from cross-validation

CONCLUSIONS/ RECOMMENDATIONS

Road traffic accidents pose a serious public safety and public health issue. They result in substantial loss of life, injury, and economic burden worldwide. In densely populated and traffic-congested urban environments such as Kandy, understanding the spatial dynamics of RTAs is vital for informed decision-making in urban planning and traffic control. This study utilized two geostatistical interpolation methods, Inverse Distance Weighting and Ordinary Kriging, to model and predict the spatial distribution of RTAs. The analysis demonstrated that both methods are effective for spatial prediction of accident hotspots, but Ordinary Kriging showed slightly superior accuracy. The study identified hotspot locations, which were mostly concentrated in the Senkadagala, Mahanuwara, Malwatta, Getambe, and Katugasthota Grama Niladari Divisions. The spatial predictions generated through this analysis can serve as a foundation for traffic safety interventions, such as the installation of signage, redesign of hazardous intersections, or targeted law enforcement. The spatial predictions of accident counts provide insights into the distribution of crash risk and highlight hotspot locations for further investigation. Future studies could incorporate additional variables (e.g., road type, weather conditions) to improve the predictive accuracy of these accident hotspots.

References

Bivand, R., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. Springer.

Cressie, N. A. C. (1993). *Statistics for spatial data* (revised ed.). Wiley.

Gill, G., Sakrani, T., Cheng, W., & Zhou, J. (2017). Investigation of roadway geometric and traffic flow factors for vehicle crashes using spatiotemporal interaction. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42*(2/W7), 1163–1166.

Oliver, M. A., & Webster, R. (2015). *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer International Publishing.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, *30*, 683-691.

Pebesma, E., 2018. Simple Features for R: Standardized support for spatial vector data. *The R Journal 10*(1), 439-446.

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