Methodology for Analyzing the Impacts of Highway Engineering Interventions over Traffic Safety Using GIS and Multinomial Logistic Regression

English translation of the paper published in the Brazilian Journal of Development, vol.7, no 1, 2021

(The title reproduces the English title provided in the original peer-reviewed publication)

Original Paper's DOI: https://doi.org/10.34117/bjdv7n1-602

Received: 1 January 2021 | Accepted for publication: 22 January 2021

Rian das Dores de Miranda
Master's Student in Civil Engineering
Fluminense Federal University
Rua Passo da Pátria, 156, São Domingos, Niterói, Rio de Janeiro, 24210-240, Brazil
rian_miranda@id.uff.br

Walber Paschoal da Silva
D.Sc. in Transportation Engineering
Fluminense Federal University
Rua Passos da Pátria, 156, Block D, Room 563C, São Domingos, Niterói, Rio de Janeiro, 24210-240, Brazil
walberpaschoal@id.uff.br

Steven Dutt Ross
D.Sc. in Production Engineering
Federal University of the State of Rio de Janeiro (UNIRIO)
Avenida Pasteur, Urca, Rio de Janeiro, Rio de Janeiro, 22290-240, Brazil
steven.ross@uniriotec.br

ABSTRACT

Since the 1950s, Brazil has prioritized road transport for moving goods and passengers. Consequently, the country's active vehicle fleet reached 65.8 million vehicles, and 96,366 crashes were recorded in 2016. This study presents a methodology for assessing the impact of road improvement projects on traffic safety using a geographic information system (GIS) and the R programming language. The analysis uses data provided by the Brazilian Federal Highway Police and comprises generating kernel density maps, assessing the spatial randomness of crashes using the quadrat count method, and fitting a multinomial logistic regression model to identify factors associated with crash severity. The methodology was applied to a 2.2 km road section in Niterói, Rio de Janeiro, where a road-widening project was completed in 2015. The results identified three crash hotspots and showed that crashes were not randomly distributed in space during either the pre- or post-intervention period. The multinomial logistic regression further indicated that crashes occurring after the intervention had higher odds of being classified in more severe outcome categories than crashes occurring before the intervention.

Keywords: traffic safety; crash hotspots; kernel density map; multinomial logistic regression.

1 INTRODUCTION

The national development strategy implemented under President Juscelino Kubitschek between 1955 and 1960 established road transport as Brazil's primary mode for moving goods and passengers. According to World Bank data (World Bank, 2013, as cited in BBC, 2018), Brazil has the highest reliance on road-based freight and passenger transport among the world's major economies. Road transport accounts for 58% of the country's transportation activity, compared with 53% in Australia, 50% in China, 43% in Russia, and 8% in Canada. Consequently, Brazil's active vehicle fleet reached 65.8 million vehicles, including passenger cars, light commercial vehicles, buses, trucks, and motorcycles (IBPT, 2018).

As the Brazilian vehicle fleet continued to grow – from 59,361,642 vehicles in 2009 to 100,746,553 in 2018 across all categories (IBGE, 2018) – the country recorded a substantial number of road crashes. According to a study by the Ministry of Transport (2018), 96,366 crashes occurred on federal highways in 2016, resulting in 6,398 fatalities and approximately 87,000 injuries. The study also reported that 30.3% of fatal crashes were associated with non-compliance with traffic regulations.

In this context, the R programming language and geographic information systems (GIS) provide important support for road-safety analysis. Their integration enables spatially distributed information to be represented through thematic maps, thereby facilitating the identification of crash hotspots (i.e., locations where crashes are more frequent and more severe). R also supports the construction of generalised linear models that can improve understanding of the factors associated with crash occurrence and severity.

Accordingly, this study aims to present a methodology for investigating the impact of a road improvement project on traffic safety. The specific objectives are to:

- locate crash hotspots in the pre- and post-intervention periods by generating kernel density maps;
- assess whether the spatial distribution of crashes was random before and after the intervention using the quadrat count method; and
- identify factors associated with crashes on the road section under study using multinomial logistic regression models.

The proposed methodology is intended to support road-safety management by transportation agencies, road operators, and public-safety authorities. The procedures described in this study can assist responsible agencies both in identifying crash hotspots and in recognising the main factors associated with crashes on a road. Such information is essential for developing policies aimed at reducing crash frequency and mitigating crash consequences, thereby benefiting society as a whole.

The remainder of this paper is organized as follows. Section 2 reviews the literature underlying the study. Section 3 presents the methodology used to identify crash hotspots and estimate the effects of the intervention on road safety. Section 4 presents a case study and discusses the results obtained from applying the methodology to a 2.2 km road section in Niterói, Rio de Janeiro, Brazil. Section 5 presents the conclusions and recommendations for future research.

2 LITERATURE REVIEW

Transportation agencies worldwide have developed standards and manuals to support the identification of road-crash hotspots. In the United States, the Highway Safety Manual is the main reference in this field and presents methods for estimating crash frequency and severity (AASHTO, 2014). In Brazil, the Ministry of Transport (2002) published the manual titled "PARE Program – Procedures for Treating Critical Road-Crash Locations". The manual reviewed hotspot-identification methods originally presented by the Brazilian National Traffic Department (Portuguese acronym: DENATRAN) in 1987, including the crash-severity technique, and updated their application for contemporary conditions.

GIS technology has been widely applied to road-crash analysis to support hotspot identification and improve understanding of potential contributing factors. Çodur *et al*. (2014) used GIS both to determine the period of the year with the highest crash frequency and to examine the relationship between road-alignment characteristics and the number of crashes recorded along a roadway in Turkey. Truong and Somenahalli (2011), in turn, used GIS to investigate pedestrian-crash risk at bus stops in metropolitan Adelaide, Australia.

Since the early 2000s, GIS has also been widely used to develop kernel density maps for hotspot identification. Anderson (2006) observed that kernel maps provide satisfactory statistical and visual representations of the expected spatial pattern of road crashes. Teodoro *et al*. (2014) found that kernel-estimator results were similar to those produced by the methods in PARE Program and concluded that kernel maps constitute a viable tool for road-safety management.

More recently, research groups have combined GIS and regression methods in road-safety studies. Andrade *et al*. (2014) used kernel maps and regression models both to identify hotspots along the section of BR-277 located in Paraná State and to assess the effects of road characteristics on fatal crashes. Their study found that length of road in urban area, limited lighting, dual-carriageway segments, and less auxiliary lanes were associated with a higher incidence of fatal crashes. Additionally, crashes were concentrated on dual-carriageway segments, with rural areas recording the highest number of vehicle-to-vehicle collisions, whereas urban areas recorded the highest number of pedestrian crashes. Machado *et al*. (2015) applied the kernel method to identify crash hotspots involving pedestrians and non-motorized road users in São Paulo (Brazil) and Rome (Italy). They also conducted multiple regression analysis to investigate relationships between crashes and potentially contributing variables, including trip-generating facilities and regional demographic characteristics.

As noted by Milton *et al*. (2008), many transportation agencies prioritise understanding the relationship between crash-related variables and crash severity. Researchers worldwide have therefore applied logistic regression to road-crash modelling. This technique is well suited to crash data because it permits mathematical models to be estimated from both categorical and numeric variables.

In this context, logistic regression is a type of generalised linear model well suited to binary categorical response variables. By contrast, Multinomial logistic regression is appropriate when the response variable has more than two unordered categories rather than the two categories used in conventional binary models (also commonly represented through dummy variables).

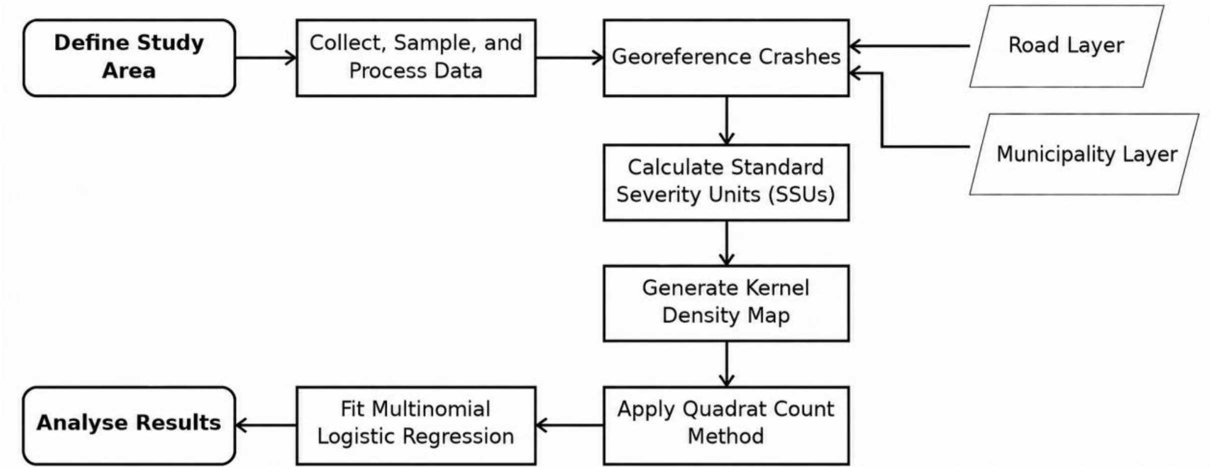
In crash-severity analysis, logistic regression models are applicable when severity is represented by two categories, while the explanatory variables may be numerical or categorical (Kong and Yang, 2010; Alvarenga, 2015). For example, Kong and Yang (2010) used logistic regression to identify the influence of driver age and vehicle speed on pedestrian crashes resulting in injury or death. The authors found a pedestrian-fatality risk of 26% at 50 km/h, 50% at 58 km/h, and 82% at 70 km/h. Furthermore, they identified that pedestrians are more likely to sustain injuries at higher ages. All of these results were statistically significant.

Conversely, multinomial logistic regression is appropriate when the severity outcome has more than two categories (Tay *et al*., 2011; Coutinho *et al*., 2015; Wahab and Jiang, 2019). Tay *et al*. (2011), for example, applied multinomial logistic regression to identify factors associated with pedestrian crashes at three severity levels: minor injury, injury, and fatality. They concluded that fatal crashes were more likely when drivers were male, under the influence of alcohol, or operating heavy vehicles.

3 METHODOLOGY

This study establishes a methodology for investigating the impact of a road improvement project on traffic safety using GIS and multinomial logistic regression. Figure 1 summarizes the procedures adopted in the study.

Figure 1. Methodological workflow



3.1 Data Collection, Sampling, and Processing

Crash datasets for Brazilian federal highways are obtained from the official website of the Brazilian Federal Highway Police (PRF, 2019). These data have been compiled annually since 2007 and include variables describing each crash, such as the date, highway kilometre marker, weather conditions, time of day, road type, crash severity, and the numbers of fatally injured, non-fatally injured, and not injured road users.

To retain only records corresponding to the road section of interest and the specified analysis period, the data are sampled using the R programming language. R is also used to identify and remove records with missing values, following the procedure adopted by Milton *et al*. (2008).

Finally, an explanatory variable named "Intervention" is created to indicate whether each crash occurred before or after the completion of the road improvement works. This classification is based on the "date" variable in the PRF dataset. "Intervention" is coded as "Before" when the crash occurred during the pre-intervention period and "After" when it was recorded during the post-intervention period. Table 1 presents the variable dictionary used in this methodology.

Table 1. Variable dictionary.

Variable	Description	Categories
Month	Month in which the crash occurred	January to December
Day of week	Day of the week on which the crash occurred	Monday to Sunday
Time of day	Time-of-day period when the crash occurred	Dawn; daylight; dusk; night
Weather condition	Weather conditions at the time of the crash	Sunny; clear; windy; cloudy; rain; fog/mist; drizzle
Road alignment	Road-layout characteristic at the crash location	Tangent; curve; bridge; intersection; tunnel
Direction of travel	Direction of travel in which the crash occurred	Increasing; decreasing
Date	Date of the crash	Not applicable
Intervention	Indicates whether the crash occurred in the pre- or post-intervention period.	Before; after

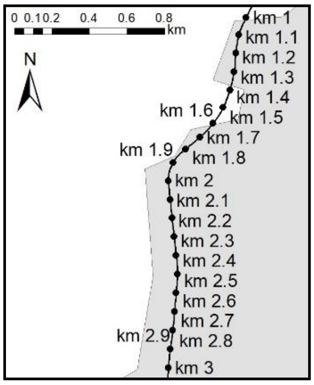
Source: PRF (2017), adapted by the authors.

3.2 Crash Georeferencing

After the preceding steps have been completed, ArcGIS is used to georeference the crash locations. A GIS project is created and populated with the municipal-boundary and road-network layers for all Brazilian territory. The municipal layer is obtained from the IBGE portal (IBGE, 2015), and the road-network layer is extracted from the Ministry of the Environment portal (Brazil, 2008). A projected coordinate system is selected at this stage.

The georeferencing procedure uses the *Construct Points* tool to create equally spaced points along the study section under analysis. To ensure that the points correspond to the recorded crash locations, consecutive points are spaced 100 m apart. This interval reflects the precision of the PRF kilometre variable, which identifies crash locations to the nearest 100 m. The points are then used as references for obtaining the coordinates of each crash location with the ArcGIS *Add XY Coordinates* tool. Finally, the coordinates are inserted into the database. Figure 2 illustrates the georeferencing procedure.

Figure 2. Georeferencing process



km	Latitude	Longitude
1.0	$lat_{1.0}$	$long_{1.0}$
1.1	$lat_{1.1}$	$long_{1.1}$
1.2	$lat_{1.2}$	$long_{1.2}$
⋮	⋮	⋮
2.8	$lat_{2.8}$	$long_{2.8}$
2.9	$lat_{2.9}$	$long_{2.9}$
3.0	$lat_{3.0}$	$long_{3.0}$

Creation of Points in GIS

Coordinate Extraction

Insertion of coordinates into the database

Source: Carvalho and Câmara (2004), adapted by the authors

3.3 Calculation of Standard Severity Units

At this stage, the crash-severity technique proposed in PARE Program is used to determine the number of Standard Severity Units (SSUs) at each kilometre of the study section. The method assigns each crash one of the following severity weights specified in PARE Program: 1 for crashes involving property damage only (PDO); 4 for injury crashes; 6 for injury crashes involving pedestrians; and 13 for fatal crashes. The number of SSUs at each crash location is then calculated using Equation 1 (Brazil, 2002).

$$\mathrm{SSU} = 1 \times \mathrm{PDO} + 4 \times \mathrm{INJ} + 6 \times \mathrm{PED} + 13 \times \mathrm{FAT} \tag{1}$$

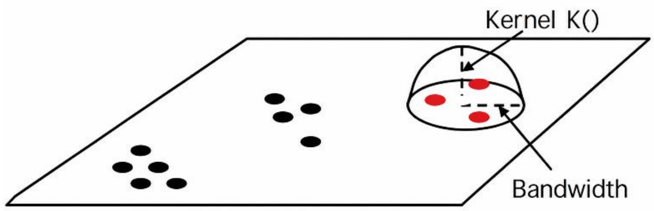
PDO: number of property-damage-only crashes;
INJ: number of injury crashes;
PED: number of injury crashes involving pedestrians;
FAT: number of fatal crashes.

When a crash involves more than one road user, it is classified according to the most severe outcome among those involved.

3.4 Generation of Kernel Density Maps

The potential for crashes to occur along the road section is estimated using the kernel density estimator. According to Carvalho and Câmara (2004), a two-dimensional kernel function can be applied to spatially distributed events, such as road crashes, to produce a raster surface representing the estimated event density per unit area. The function considers all points within the area of influence surrounding a location of interest and weights them according to their distance from that location. Figure 3 illustrates the kernel density estimator.

Figure 3. Conceptual representation of the kernel density estimator



The kernel density estimator is calculated using Equation 2 (Vargas, 2015).

$$\hat{f} = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_x}, \frac{y - Y_i}{h_y}\right) \tag{2}$$

n: sample size;
(h_x, h_y): bandwidths in the x and y directions;
(x, y): coordinates of the location of interest;
(X_i, Y_i): coordinates of event i in the sample;
K: Gaussian kernel interpolation function.

Sheather (2004) noted that kernel density estimates are influenced primarily by bandwidth selection. Vargas (2015) observed that large bandwidths can oversmooth the density surface, whereas small bandwidths restrict the area of influence and limit the number of events captured by the kernel function. To support bandwidth selection, Silverman (1986) proposed the optimal-bandwidth formula shown in Equation 3. For kernel maps, the formula is applied separately in the x and y directions.

$$\mathrm{h} = 1.06 \times \min(\sigma, \mathrm{IQR}) \times \mathrm{n}^{-1/5} \tag{3}$$

σ: standard deviation of the coordinates in the direction being analysed;
IQR: interquartile range of the coordinates in the direction being analysed.

In this study, kernel maps are generated to represent the spatial distribution of crashes before and after the road project. The analysis is intended both to identify crash hotspots and to compare their intensity and location across the two periods. Crashes are weighted by their SSU values so that the kernel estimator assigns greater influence to locations affected by more severe crashes.

3.5 Quadrat Count Method

The quadrat count method is used to investigate whether crashes are randomly distributed along the road. The method divides the study area into a grid of quadrats and counts the number of SSUs within each quadrat. If the spatial distribution of crashes is completely random, the counts should follow a Poisson distribution (Loo and Anderson, 2015). Complete spatial randomness is assessed using a chi-square test at the 5% significance level. The hypotheses are stated below, and the chi-square statistic is given by Equation 4.

H_0: Crashes are completely random in space.

H_1: Crashes are not completely random in space.

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \qquad (2)(4)$$

O_i: observed count in quadrat i;
E_i: expected count in quadrat i.

Loo and Anderson (2015) emphasized that both quadrat size and the number of quadrats influence the chi-square test. A grid with a small number of large quadrats tends to contain many occurrences within each quadrat, whereas a grid with many small quadrats may produce numerous empty cells. To select an appropriate grid configuration, the methodology divides the study area into 1 × 3, 1 × 4, and 1 × 5 grids. Chi-square tests are then conducted for all grid configurations in both the pre- and post-intervention periods.

3.6 Multinomial Logistic Regression

Multinomial logistic regression is used to estimate a mathematical model that describes crash severity from the explanatory variables recorded in the sample. In this model, crash severity is the response variable and has four levels: 1) PDO crash (weight = 1), 2) injury crash (weight = 4), 3) injury crash involving pedestrians (weight = 6), and 4) fatal crash (weight = 13). Each crash can be assigned to only one category, and PDO crashes is used as the reference category. The primary objective of this analysis is to examine the effect of the "Intervention" variable on the severity of crashes observed on the road section. The multinomial regression model is expressed by Equation 5 (Hosmer *et al*., 2013).

$$g_k(x) = \ln[\Pr(k \mid x) \,/\, \Pr(0 \mid x)] = \beta_{k0} + \beta_{k1}x_1 + \ldots + \beta_{kp}x_p$$
$$\text{for } k \in \{0, 1, 2, \ldots, K\} \text{ and } i \in \{0, 1, 2, \ldots, p\} \qquad (5)$$

k: crash-severity category;
K: number of categories in the response variable;
p: number of explanatory-variable terms in the model, including indicator variables for the non-reference categories of categorical explanatory variables;
x: vector containing the values of the explanatory variables;
$Pr(k \mid x)$: conditional probability of severity category k given x;
$Pr(0 \mid x)$: conditional probability of the reference severity category given x;
β_{ki}: model parameter for severity category k and explanatory variable i.

The model is developed using the four-stage procedure described by Alvarenga (2015): (1) stepwise variable selection, (2) the likelihood-ratio test, (3) the Hosmer-Lemeshow test, and (4) the Wald test.

Accordingly, a null hypothesis is rejected when the test p-value is below 0.05 and is not rejected otherwise.

The stepwise method develops regression models by adding or removing variables in successive stages. At each step, model fit is evaluated using the Akaike Information Criterion (AIC). The AIC penalizes models containing a large number of variables and thereby favours models that combine lower complexity with satisfactory predictive performance (Alvarenga, 2015; Murphy, 2012). AIC values are calculated using Equation 6.

$$\mathrm{AIC} = -2[\log(\mathrm{L}) - \mathrm{p}] \tag{6}$$

L: Log likelihood of the estimated model;
p: number of model parameters.

According to Alvarenga (2015), lower AIC values indicate better model fit. The stepwise procedure therefore stops when the current model has the minimum AIC among the models estimated in the preceding steps (Burnham and Anderson, 2004; Hastie *et al*., 2017). At that point, any additional inclusion or removal of variables would increase the AIC and reduce the quality of the resulting model.

The statistical significance of the set of variables selected by the stepwise procedure is evaluated using the likelihood-ratio test, which compares the fit of the estimated model with that of the null model. The hypotheses are stated below, and the test statistic is given by Equation 7 (Hosmer *et al*., 2013; Washington *et al*., 2003).

H_0: The explanatory variables do not collectively improve model fit relative to the null model.

H_1: The explanatory variables collectively improve model fit relative to the null model.

$$\mathrm{G} = -2 \times [\mathrm{LL}(\beta_R) - \mathrm{LL}(\beta_U)] \tag{7}$$

$LL(\beta_R)$: log-likelihood of the restricted model without the variables;
$LL(\beta_U)$: log-likelihood of the unrestricted model containing the variables.

The goodness-of-fit between the model and the sample data is evaluated using the Hosmer–Lemeshow test. The test uses the Pearson chi-square statistic shown in Equation 8, with the hypotheses stated below (Hosmer *et al*., 2013; Fagerland *et al*., 2008).

H_0: The model fits the data adequately.

H_1: The model does not fit the data adequately.

$$\chi^2 = \sum_{\mathrm{j}=1}^{J} \sum_{k=0}^{K-1} \frac{(\mathrm{O_{jk}} - \mathrm{E_{jk}})^2}{\mathrm{E_{jk}}} \tag{8}$$

J: number of groups into which the sample is divided;
O_{jk}: observed frequency for severity category k in group j;
E_{jk}: expected frequency for severity category k in group j.

Finally, the Wald test is used to evaluate the statistical significance of each estimated model parameter, β_{ki}. The test statistic follows the standard normal distribution and is presented in Equation 9. The corresponding hypotheses are stated below.

H₀: $\beta_{ki} = 0$.
H₁: $\beta_{ki} \neq 0$.

$$W_j = \frac{\hat{\beta}_{ki}}{SE(\hat{\beta}_{ki})} \cap N(0,1) \tag{9}$$

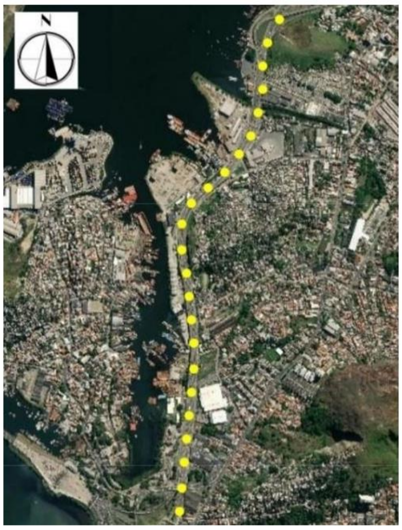
$\hat{\beta}_{ki}$: estimated parameter for severity category k and explanatory variable i;
$SE(\hat{\beta}_{ki})$: standard error of $\hat{\beta}_{ki}$.

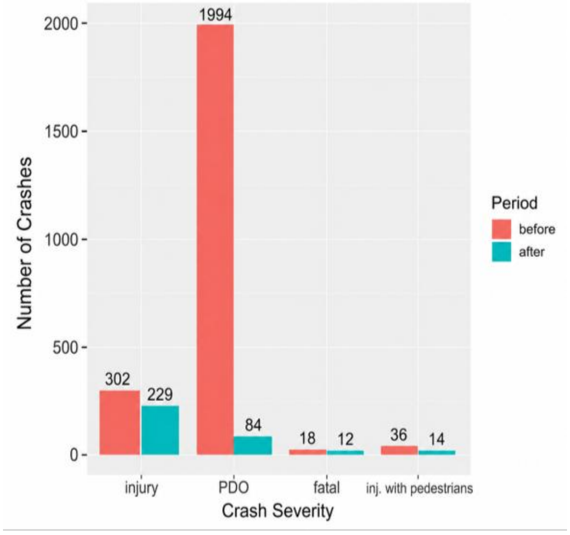
4 RESULTS AND DISCUSSION

The methodology described in Section 3 was applied to the section of federal highway BR-101 known as Avenida do Contorno, between km 319.8 and km 322 in Niterói, Rio de Janeiro, Brazil. This road section underwent a widening project that opened on 7 August 2015. This project increased the number of lanes from two to three in each direction, constructed shoulders, installed LED lighting along the road and increased the posted speed limit from 60 km/h to 80 km/h (Grandes Construções, 2015; Arteris, 2015). Figure 4a shows the road section under study and the crash locations georeferenced using the SIRGAS 2000 / UTM zone 23S coordinate reference system.

The PRF data contained 2,728 crashes recorded between January 2011 and November 2019. Of these, 39 records were excluded from the analysis due to missing information. Among the remaining 2,689 records, 2,350 crashes occurred before the intervention and 339 after it. In the pre-intervention period, the severity distribution comprised 1,994 PDO crashes, 302 injury crashes, 36 injury crashes involving pedestrians, and 18 fatal crashes. In the post-intervention period, the corresponding values were 84, 229, 14, and 12, respectively. Figure 4b compares crash frequencies across the two periods.

Figure 4. (a) Georeferenced crashes on the study section; (b) crash frequency before and after the intervention.





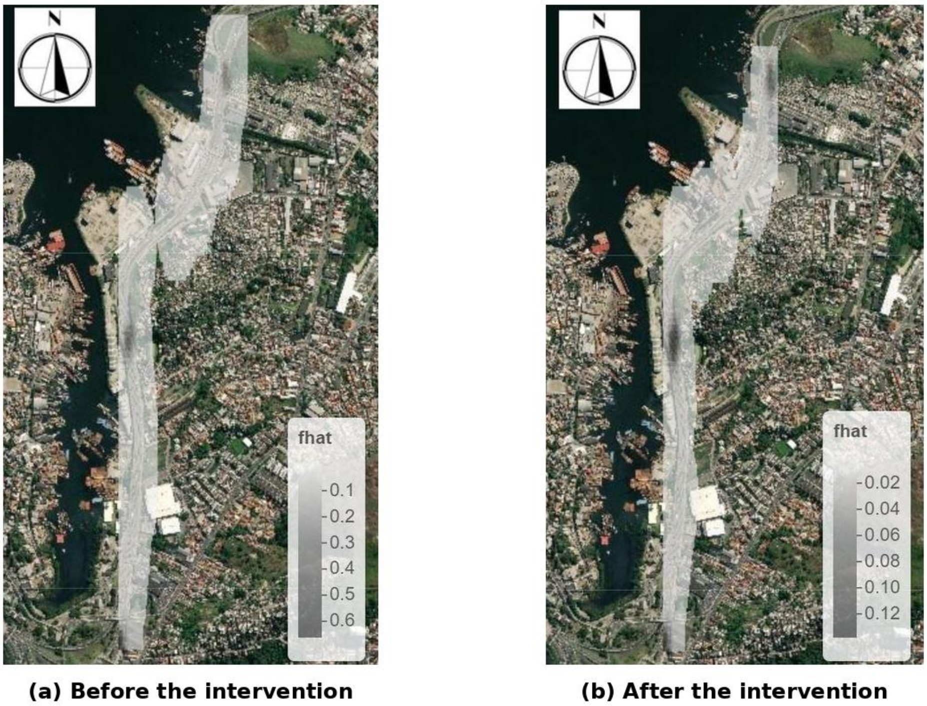
(a) **(b)**

The crash-severity technique was then applied to identify locations with the highest and lowest SSU totals. Kilometres 320, 321, and 322 recorded the highest SSU totals throughout the study period, at 854, 1,322, and 2,410, respectively. The SSU total at each of these locations decreased between the pre- and post-intervention periods: from 611 to 243 at km 320, from 911 to 411 at km 321, and from 1,914 to 496 at km 322. These reductions were primarily associated with the decline in PDO crashes during the post-intervention period. The lowest SSU totals occurred at km 321.3 and km 321.4, with each location recording an SSU total of 1. At each location, the single recorded case was a PDO crash that occurred before the intervention.

Next, kernel maps were generated from the SSU values. For the pre-intervention period, Equation 3 produced bandwidths of 7.36 m in the x direction and 147.70 m in the y direction. For the post-intervention period, the corresponding bandwidths were 10.48 m and 183.57 m. The x-direction bandwidths were considerably smaller than the y-direction values because the longitudinal coordinates showed low variance among crashes recorded between km 321 and km 322, where most of the road section's SSUs were concentrated. Very small bandwidths can adversely affect the performance of the kernel function by restricting the number of events within its area of influence and creating discontinuities in the map.

To preserve continuity in the kernel density surfaces, an x-direction bandwidth of 50 m was adopted for both periods. The resulting maps are shown in Figure 5.

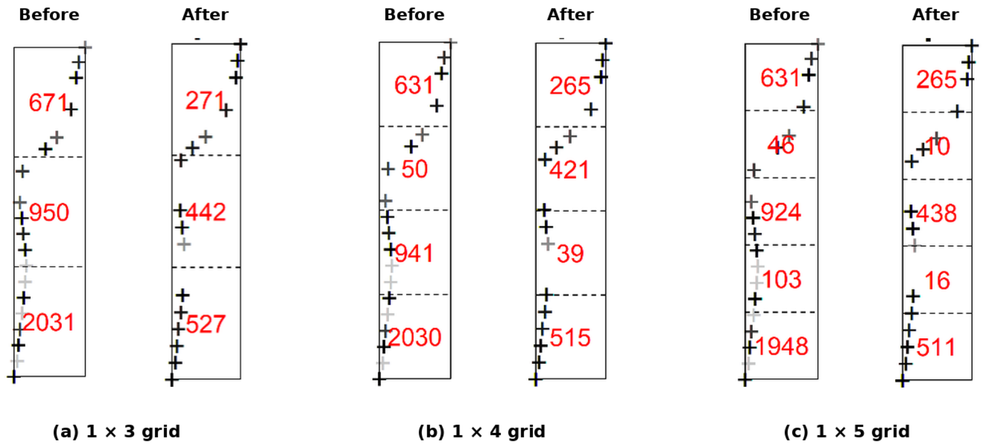
Figure 5. Kernel density maps: (a) before the intervention; (b) after the intervention



The results indicate that the crash hotspots were located at km 320, km 321, and km 322 in both periods. However, kernel-estimator values were higher before the intervention, indicating greater crash potential than in the post-intervention period. In both maps, the maximum density occurred at km 322, with an estimator value of 0.60 before the intervention and 0.12 after it.

The quadrat count method was then applied to assess the spatial randomness of crashes. After the study area was divided into 1 × 3, 1 × 4, and 1 × 5 grids, SSUs were found to be concentrated in the southernmost quadrat, primarily because of crashes at km 322. This pattern was observed both before and after the intervention, as illustrated in Figure 6. The chi-square test produced a p-value of 2.2×10^{-16} regardless of the grid configuration or analysis period. The null hypothesis was therefore rejected, indicating that crashes were not randomly distributed along the study section.

Figure 6. Quadrat count method applied before and after the intervention using three configurations: (a) 1 × 3 grid; (b) 1 × 4 grid; (c) 1 × 5 grid.



Finally, a multinomial logistic regression model was fitted to examine the effect of the "Intervention" variable on crash severity. The analysis began with the selection of statistically relevant explanatory variables using the stepwise method. Forward, backward, and bidirectional stepwise selection all produced the same final model, which included the variables "Intervention", "Direction of travel", and "Time of day". Table 2 presents the forward stepwise process. As described in the methodology, the model selected at iteration 3 was retained because it had the lowest AIC among the models considered.

Table 2. Forward stepwise selection results.

Iteration	Model	AIC
1	Weight ~ Intervention	2966.09
2	Weight ~ Intervention + Direction of travel	2927.26
3	Weight ~ Intervention + Direction of travel + Time of day	2918.55
4	Weight ~ Intervention + Direction of travel + Time of day + Road alignment	2920.376

The likelihood-ratio and Hosmer–Lemeshow tests were subsequently applied to evaluate, respectively, the joint statistical significance of the selected variables and the model's goodness of fit. The likelihood-ratio statistic was $G = 579.6$, with $p = 2.2 \times 10^{-16}$. The null hypothesis was therefore rejected, indicating that the selected variables were jointly statistically significant. The Hosmer–Lemeshow test produced $\chi^2 = 3.367$ and $p = 0.7615$. Consequently, the null hypothesis that the model fits the sample data adequately was not rejected.

The Wald test was then used to estimate the model parameters and assess their individual statistical significance. At the 5% significance level, the results in Table 3 show that the estimated coefficients for the "After" category of the "Intervention" variable were statistically significant for injury crashes, injury crashes involving pedestrians, and fatal crashes. The estimated coefficients for the "After" intervention category were also positive for all of these crash-severity categories. Thus, relative to the pre-intervention period, crashes occurring after the road works had higher odds of being classified as injury crashes, injury crashes involving pedestrians, or fatal crashes rather than as PDO crashes. One possible explanation is that the road widening, along with the increase in the posted speed limit, improved traffic flow and enabled higher operating speeds, thereby increasing the severity of the crashes that did occur.

The Wald test also indicates that crashes occurring in the São Gonçalo-bound direction (decreasing kilometre direction) had lower odds of injury, injury involving pedestrians and fatalities than crashes in the Rio de Janeiro-bound direction (increasing kilometre direction). Crashes occurring during daylight also had significantly lower odds of being fatal than crashes occurring at dawn.

Table 3. Wald test results.

Crash Severity	Variable	β_{ki}	$SE(\beta_{ki})$	W_{ki}	p-value
Injury	Intercept	−1.25561	0.27250	−4.6078	4.07×10^{-6}
	Time of day: Dusk	−0.37468	0.35748	−1.0481	0.2945849
	Time of day: Night	−0.38824	0.28746	−1.3506	0.1768329
	Time of day: Daylight	−0.12721	0.27721	−0.4589	0.6463250
	Direction of travel: Decreasing	−0.63261	0.11512	−5.4954	3.90×10^{-8}
	Intervention: After	2.84567	0.14332	19.8556	$< 2.2 \times 10^{-16}$
Injury involving pedestrians	Intercept	−3.54510	0.73735	−4.8079	1.53×10^{-6}
	Time of day: Dusk	0.80434	0.82843	0.9709	0.3315849
	Time of day: Night	−0.39771	0.78426	−0.5071	0.6120765
	Time of day: Daylight	−0.03467	0.74814	−0.0463	0.9630374
	Direction of travel: Decreasing	−0.64556	0.29669	−2.1759	0.0295634
	Intervention: After	2.22008	0.33663	6.5951	4.25×10^{-11}
Fatal	Intercept	−3.08846	0.62848	−4.9141	8.92×10^{-7}
	Time of day: Dusk	−1.24269	1.18158	−1.0517	0.2929286
	Time of day: Night	−0.19260	0.65824	−0.2926	0.7698320
	Time of day: Daylight	−1.64679	0.70260	−2.3439	0.0190857
	Direction of travel: Decreasing	−1.39284	0.38842	−3.5860	0.0003358
	Intervention: After	2.70067	0.39637	6.8135	9.53×10^{-12}

Reference categories: Crash severity: PDO Time of day = Dawn; Direction of travel = Increasing; Intervention = Before.

5 CONCLUSIONS AND RECOMMENDATIONS

This study presented a methodology for investigating the impacts of a road improvement project on traffic safety. The objectives were addressed through a case study of Avenida do Contorno in Niterói, Rio de Janeiro, Brazil, where a road-widening project was completed on 7 August 2015. The application demonstrated the usefulness of the methodology for road-safety analysis and produced three principal outputs: kernel density maps for the pre- and post-intervention periods, diagrams derived from the quadrat count method, and a multinomial logistic regression model.

The kernel density maps identified crash hotspots at km 320, km 321, and km 322 in both periods. Kernel estimates at the hotspots were higher before the intervention than after it, indicating that the project contributed both to a reduction in SSUs at these locations and to a more dispersed distribution of SSUs along the road section. Additionally, the quadrat count method showed that crashes were not randomly distributed in either period. The diagrams in Figure 6 suggest that SSUs were concentrated in the quadrats containing the hotspots.

Finally, the multinomial logistic regression indicated that the variables “Intervention”, “Direction of travel”, and “Time of day” were statistically significant predictors of crash severity. Relative to the pre-intervention period, crashes recorded after the intervention had higher odds of being classified as injury crashes, injury crashes involving pedestrians, or fatal crashes rather than as PDO crashes. This increase in the conditional severity of post-intervention crashes may have resulted from higher operating speeds enabled by improved traffic flow after the addition of a traffic lane in each direction. In addition, the São Gonçalo-bound direction had lower odds of all non-reference severity categories than the Rio de Janeiro-bound direction, suggesting riskier behaviour among drivers and motorcycle riders travelling toward the state capital. Daylight conditions were also associated with lower odds of fatal outcomes than dawn, potentially because of differences in visibility between the two periods.

Based on these findings, traffic authorities should prioritize enforcement at the identified hotspots, particularly km 322, which recorded the highest SSU total on the study section. Future studies should collect operating-speed data for both the pre- and post-intervention periods. Such data would permit a direct assessment of the relationship between vehicle speed and crash severity and could support decisions by road operators concerning measures such as speed-camera installation or reductions in the posted speed limit.

ACKNOWLEDGMENTS

This study was financed in part by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), Finance Code 001.

REFERENCES

AASHTO. **Highway Safety Manual User Guide. National Cooperative Highway Research Program 17-50.** Washington, DC. 2014. Available at: <http://onlinepubs.trb.org/onlinepubs/nchrp/docs/nchrp17-50_userguide.pdf> Accessed on: 28 April 2019.

Alvarenga A. M. T. **Modelos lineares generalizados: aplicação a dados de acidentes rodoviários [Generalized linear models: application to road crash data].** Master's dissertation. Faculty of Sciences, University of Lisbon. Lisbon, Portugal. 2015. (In Portuguese).

Anderson, T. K. Comparison of spatial methods for measuring road accident ‘hotspots’: a case study of London. **Journal of Maps.** v. 16, n. 2, p. 55-63. 2006. https://doi.org/10.1080/jom.2007.9710827

Andrade L.; Vissoci J. R. N.; Rodrigues C. G.; Finato K.; Carvalho E; Pietrobon R.; Souza E. M.; Nihei O. K; Lynch C. and Carvalho M. D. B. Brazilian Road Traffic Fatalities: A Spatial and Environmental Analysis. **PLoS ONE:** e87244, v. 9, n. 1, p. 1-9. 2014. https://doi.org/10.1371/journal.pone.0087244

ARTERIS. **Av. do Contorno de Niterói | BR-101 [Contorno Avenue in Niterói | BR-101].** Rio de Janeiro. 2015. **Youtube.** Available at: <https://www.youtube.com/watch?v=G-nCt7KV9_A>. Accessed on: 16/03/2020. (In Portuguese).

Mesquita, L. Crise revela dependência de transporte rodoviário que é 'mais barato e dá voto' [Crisis reveals dependence on road transport, which is “cheaper and wins votes”]. **BBC Brasil,** London, 24 May 2018. Available at: <https://www.bbc.com/portuguese/brasil-44247460> Accessed on: 20 April 2019. (In Portuguese).

Brazil. Procedimentos para o tratamento de locais críticos de acidentes de trânsito [Procedures for treating road crash hotspots]. Ministry of Transport. **PARE Program.** Brasília, DF. 2002. (In Portuguese).

Brazil. Rodovias Federais, Estaduais e Municipais do Brasil - PNLT (2008) [Federal, state, and municipal highways of Brazil - PNLT (2008)]. Plano Nacional de Logística de Transportes (PNLT) [National Transport Logistics Plan (PNLT)]. **Shapefile.** Ministry of Transport. 2008. Available at: <http://mapas.mma.gov.br/i3geo/datadownload.htm#> Accessed on: 15 April 2019. (In Portuguese).

Brazil. Avaliação das Políticas Públicas de Transportes, Segurança nas Rodovias Federais [Evaluation of Transportation Public Policies: Federal Highway Safety]. **Ministry of Transport, Ports, and Civil Aviation.** Brasília, DF. 2018. Available at:

<https://www.infraestrutura.gov.br/images/2018/documentos/APT_Seguranca_Rodovias_Federais_Final_8.pdf> Accessed on: 28 April 2019. (In Portuguese).

Burnham K. P. and Anderson D. R. Multimodel Inference: Understanding AIC and BIC in Model Selection. **Sociological Methods & Research.** v. 33, n. 2, p.261-304. 2004. https://doi.org/10.1177/0049124104268644

Çodur M.; Atalay A.; Tortum A.; and Doğru A. Geographical Information Systems Aided Traffic Accident Analysis Case Study: City Of Erzurum North Ring Road. In: International Conference on Traffic and Transport Engineering. Belgrade. 2014, **Electronic Conference Proceedings...** p. 336-342. 2014. Available at: < https://trid.trb.org/view/1409523>. Accessed on: 08/04/2019

Coutinho C. H. L.; Cunto F. J. C. and Ferreira S. M. P. Análise da severidade dos acidentes com motocicletas utilizando modelos probit e logit ordenados [Analysis of motorcycle crash severity using ordered probit and logit models]. **Revista Transportes.** v. 23, n. 4, p. 60-66. 2015. https://doi.org/10.14295/transportes.v23i4.926. (In Portuguese).

Carvalho, M.S. and Câmara, G. Análise de Eventos Pontuais [Analysis of Point Events]. In: Druck, S. and Monteiro, A.V.M. (eds.) Análise Espacial de Dados Geográficos [Spatial Analysis of Geographic Data]. Brasília, **EMBRAPA.** 2004. Available at: <http://www.dpi.inpe.br/gilberto/livro/analise/> Accessed on: 20 May 2019. (In Portuguese).

Fagerland M. W.; Hosmer D. W. and Bofin A. M. Multinomial goodness-of-fit tests for logistic regression models. **Statistics in Medicine.** v. 27, n. 21, p. 4238-4253. 2008. https://doi.org/10.1002/sim.3202

GRANDES CONSTRUÇÕES. Rodovias: Nova Safra de Concessões Promete Bons Frutos [Highways: New batch of concessions promises strong results]. **Grandes Construções: Construção, Infraestrutura, Concessões e Sustentabilidade.** n. 62, p. 20-54. 2015. Available at: <https://issuu.com/sobratema/docs/gc_62/52>. Accessed on: 08/04/2019. (In Portuguese).

Hastie T.; Tibshirane R. and Friedman J. **The Elements of Statistical Learning.** Ed. 2. Stanford, California: Spring. 2017.

Hosmer D. W.; Lemeshow S. and Sturdivant R. X. **Applied Logistic Regression.** Ed. 3. Hoboken, New Jersey: John Wiley & Sons, Inc. 2013

IBGE. Municípios Brasileiros [Brazilian Municipalities]. **Shapefile.** 2015. Available at: <ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2015/Brasil/BR/> Accessed on: 15 April 2019. (In Portuguese).

IBGE. Cidades – **Frota de veículos 2018** [Cities – Vehicle Fleet 2018]. 2018. Available at: <https://cidades.ibge.gov.br/brasil/pesquisa/22/28120?ano=2018> Accessed on: 15 April 2019. (In Portuguese).

IBPT. Real Frota Circulante no Brasil é de 65,8 Milhões de Veículos, Indica Estudo [Study finds that Brazil's actual active fleet totals 65.8 million vehicle.]. **Brazilian Institute of Planning and Taxation.** São Paulo, SP. 2018. Available at: <http://www.automotivebusiness.com.br/abinteligencia/pdf/estudo_frota_completo.pdf> Accessed on: 20 April 2019. (In Portuguese).

Kong C. and Yang J. Logistic regression analysis of pedestrian casualty risk in passenger vehicle collisions in China. **Accident Analysis and Prevention.** v. 42, n. 4, p. 987-993. 2010. https://doi.org/10.1016/j.aap.2009.11.006

Loo B. P. Y. and Anderson T. K. **Spatial Analysis Methods of Road Traffic Collisions.** Ed. 1. Florida: CRC Press. 2015.

Machado C. A. S.; Giannotti M. A.; Neto F. C.; Tripodi A.; Persia L. and Quintanilha J. A. Characterization of Black Spot Zones for Vulnerable Road Users in São Paulo (Brazil) and Rome (Italy). **ISPRS International Journal of Geo-Information.** v. 4, n. 2, p. 858-882. 2015. https://doi.org/10.3390/ijgi4020858

Milton C. J.; Shankar V. N. and Mannering F. L. Highway accident severities and the mixed logit model: An explanatory empirical analysis. **Accident Analysis and Prevention.** v. 40, n. 1, p. 260-266. 2008. https://doi.org/10.1016/j.aap.2007.06.006

Murphy K. P. **Machine Learning: A Probabilistic Perspective.** Ed. 3. The MIT Press Cambridge, Massachusetts. London, England. 2012.

PRF. **Dados Abertos: 2011-2019 [Open Data: 2011-2019].** 2019. Available at: <https://portal.prf.gov.br/dados-abertos-acidentes> Accessed on: 10 February 2020. (In Portuguese).

PRF. **Dicionário de Variáveis – Acidentes [Variable Dictionary – Crashes].** 2017. Available at: <https://arquivos.prf.gov.br/arquivos/index.php/s/mheQDcXUx2veHnU#pdfviewer> Accessed on: 10 February 2020. (In Portuguese).

Sheather S. J. Density Estimation. **Statistical Science.** v. 19, n. 4, p. 588-597. 2004. doi:10.1214/088342304000000297

Silverman B. W. **Density Estimation for Statistics and Data Analysis.** London: Chapman and Hall. 1986.

Tai R.; Choi J.; Kattan L. and Khan A. A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity. **International Journal of Sustainable Transportation.** v. 5, n. 4, p. 233-249. 2011. https://doi.org/10.1080/15568318.2010.497547

Teodoro A. B.; Alcantara F. A. and Barbosa H. M. Comparação entre dois Métodos para Identificação de Locais Críticos de Acidentes de Trânsito [A Comparison Between Two Methods of Crash Hotspot Identification]. In: XXVIII Congresso de Ensino e Pesquisa em Transportes (ANPET) [XXVIII Transport Teaching and Research Congress (ANPET)]. Curitiba. 2014, **Electronic Conference Proceedings...** Curitiba: ANPET 2014. p. 1-13. 2014. Available at: <http://redpgv.coppe.ufrj.br/index.php/es/produccion/articulos-cientificos/2014-1/866-comparacao-entre-dois-metodos-para-identificacao-de-locais-criticos-de-acidentes-de-transito/file>. Accessed on: 08/04/2019. (In Portuguese).

Truong L. T. and Somenahalli S. V. C. Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops. **Journal of Public Transportation.** v. 14, n. 1, p. 99-114. 2011. http://doi.org/10.5038/2375-0901.14.1.6

Vargas S. A. V. **Previsão da distribuição da densidade de probabilidade da Geração de Energia Eólica usando técnicas não paramétricas [Forecasting the probability density distribution of wind energy generation using non-parametric techniques].** Doctoral thesis. Department of Electrical Engineering, PUC-Rio. 2015. (In Portuguese).

Wahab L. and Jiang H. A multinomial logit analysis of factors associated with severity of motorcycle crashes in Ghana. **Traffic Injury Prevention.** v. 20, n. 5, p. 521-527. 2019. https://doi.org/10.1080/15389588.2019.1616699

Washington S. P.; Karlaftis M. G. and Mannering F. L. **Statistical and Econometric Methods for Transportation Analysis.** Ed. 2. London: Chapman and Hall/CRC. 2003.