Scientific
Research
Publishing

# GIS-Based Methodology for Crash Prediction on Single-Lane Rural Highways

**Márcia Macedo[1]\* , Emilia Kohlman Rabbani[1] , Maria Maia[1], Marlos Macedo[2] , Bianca Ferreira[3]**

[1]Post-Graduate Program in Civil Engineering, UPE, Recife, Brazil
[2]Post-Graduate Program in Systems Engineering, UPE, Recife, Brazil
[3]Graduate Program in Architecture and Urban Planning, UNICAP, Recife, Brazil
Email: \*marcia.macedo@poli.br

## Abstract

Due to the need to update the current guidelines for highway design to focus on safety, this study sought to build an accident prediction model using a Geographic Information System (GIS) for single-lane rural highways, with a minimum of statistically significant variables, adequate to the Brazilian reality, and improve accident prediction for places with similar characteristics. A database was created to associate the accident records with the geometric parameters of the highway and to fill in the gaps left by the absence of geometric highway plans through geometric reconstitution or semi-automatic extraction of highways using satellite images. The Generalized Estimating Equation (GEE) method was applied to estimate the coefficients of the model, assuming negative distribution of the binomial error for the count of observed accidents. The accident frequency and annual average daily traffic (AADT) were analyzed, along with the spatial and geometric characteristics of 215 km of federal single-lane rural highways between 2007 and 2016. The GEE procedure was applied to two models having three variations of distinct homogeneous segmentation, two based on segments and one based on the kernel density estimator. To assess the effect of constant traffic, two more variations of the models using AADT as an offset variable were considered. The predominant correlation structure in the models was the exchangeable. The principal contributing factors for the occurrence of collisions were the radius of the horizontal curve, the grade, segment length, and the AADT. The study produced clear indicators for the design parameters of roadways that influence the safety performance of rural highways.

## Keywords

Roads, GEE, Single-Lane Rural Highways, GIS, Crash Prediction

## 1. Introduction

Although most highway accidents occur on straight stretches of road, it is on curves where accidents with greater severity occur [1]. Curves, particularly flat ones, concentrate 54% of the fatal accidents that occur on rural highways in Brazil [2]. They are more dangerous for drivers because of the additional centripetal forces exerted on the vehicle and the greater attention required on the part of the driver [3].

Due to accident severity, flat curves have been a focus for many researchers. Most studies have focused on the relationship between the characteristics of the curve and its safety performance, including design attributes [4], such as signage and markings [5] [6], and strategies to improve safety [7] [8]. In this scenario, the Accident Prediction Models (APM) emerge as tools that are capable of modeling the relevant factors for traffic accidents. They are statistical models that relate the frequency of traffic accidents to geometric and operational attributes of the road. These models lack a large amount of data [9].

Among the barriers encountered in the APM development process is a lack of documentation on road networks and projects that almost always consider only the possibility of a shorter route, better flow, and lower costs, without taking accident dynamics and their relationship with the geometric characteristics of the roads into account [10]. Better planning could be done through geoprocessing and remote sensing techniques, utilizing Geographic Information Systems (GIS) and interpretation of satellite images. Automatic or semi-automatic extraction of roads from satellite images may be the most convenient way to overcome the lack of project documentation for road safety in Brazil [11].

Another limitation in the development of APMs is in the segmentation of sections with similar geometric characteristics (homogeneous sections). This homogeneous division is necessary to establish the spatial relationships between the accident and the place where the accident occurred. In Brazil, the characterization between tangent and curve, for example, is made based on visual inspection, which may lead to errors in the identification of straight and curved sections. In this case, the attribution of an accident to a particular section of the highway may be incorrect. In order to avoid this, it is necessary to identify parameters that can characterize the segment correctly.

This study seeks to develop a database capable of associating accident records to the geometric parameters of the highway, obtained by a geometric reconstitution process when vector data is available or through semi-automatic extraction of highways from satellite images when it isn't. Spatial modeling and analysis tools will be used to extract spatial elements, such as lane width, shoulder width, superelevation, and curve radius from digital terrain models, satellite images, and from the geometric design, complementing any information unavailable in the traditional accident database. The homogeneous segments will be analyzed and classified using an analytical method (HSM) and a spatial method (Kernel-KDE density). The goal is to build an accident prediction model appropriate

for Brazil using GIS for single lane rural highways, with a minimum of statistically significant variables, and to improve accident prediction for places with similar characteristics.

## 2. Background

### 2.1. Accident Prediction Models

Numerous studies have examined the impact of road characteristics on accident frequency [12]-[21]. Most studies used the traditional statistical models of Multiple Regression, Poisson, and Negative Binomial Regression [19] [22]-[30].

The problem with traditional models is that they assume that the residuals between observations are independent. Disregarding this hierarchical structure, when present, may result in models with biased estimates of parameters and biased standard errors. When working with longitudinal data (samples measured more than once over time) or grouped, this assumption of independence between variables may not make sense. There are several methodologies available to solve this problem, with perhaps the best known, in the non-Gaussian context, being the Generalized Estimating Equations (GEE) methodology. The GEE model showed better results, however, for horizontal curves, stretches in which accident causes have been poorly studied [31]. One of the principal characteristics of GEE is its ability to unify several statistical techniques that are usually studied separately. This makes it possible to increase the number of assumptions admitted and to examine more than just the linear relationships between the explanatory variables and the response. This type of model allows the potential interactions between variables to be evaluated and is capable of modeling databases with longitudinal, spatial, or multilevel structures.

### 2.2. Variables Involved in Modeling

Various models have outlined accidents on horizontal curves based on variables that include the length of the curve, degree of curvature, and grade. Almost all models used the traffic volume for each segment, based on estimated AADT counts. These studies indicate that a greater central angle [32] [33] a greater slope [21] [34] and greater superelevation increases accident frequency [33], while greater radius reduces it [33] [35]. Considering the same variables and the influence of geometric parameters on accident severity, studies indicate that greater slope increases severity [36] [37] while greater radius [36] and greater superelevation reduce it [38].

The inclusion of spatial relationships in a safety analysis can be an important consideration for a more accurate and comprehensive approach. The spatial relationship of a curve to adjacent curves, including distance to adjacent curves, direction of rotation of adjacent curves, radius of adjacent curves, and length of adjacent curves, as well as the vertical curvature, are also important characteristics that can influence the safety of a horizontal curve or a series of curves [3].

Based on the variables found in the literature and their influence on accident

frequency, the following variables were selected *a priori* for this study: horizontal curvature (radius, degree of curve, deflection angle, curve length), lane width, shoulder width and type, traffic volume, and grade. Qualitative spatial variables, such as land use (rural or urban), road-track profile (flat, wavy, or mountainous), layout (straight or flat), day of the week, climatic conditions, and accident type will be used to assist in the selection of homogeneous sections.

## 3. Materials and Methods

### 3.1. GIS-Based Accident Prediction

The study methodology was developed using three principal steps: 1) construction of a database from data collection and semi-automatic extraction of highways from vector bases and/or satellite images, 2) homogeneous segmentation of highways, and 3) accident frequency modeling.

#### 3.1.1. Data Collection

The traffic accident data and information were collected for this study through electronic spreadsheets obtained from the traffic accident reports of the Federal Highway Police Department (DPRF), covering the period from January 1, 2007 to December 31, 2016, for highway BR-232, between km 141 and km 356.

Road sections from the National Transportation Plan (PNV) were obtained from DNIT (2016). These sections of road have not undergone any constructive changes during the period analyzed.

Traffic volumes (AADT) were obtained from the National Traffic Control Plan (PNCT), available at DNIT (2016) for the years 2014, 2015, and 2016. For the previous years (2007 to 2013), the AADT values were taken from the ANTT Annual Report (2015).

#### 3.1.2. Highway Network Digital Processing

To acquire information on the stretches of the highway that did not have a geometric design, the methodology developed by Macedo *et al.* [11] was used, which consists of the extraction of geometric characteristics of roads from satellite images based on pattern classifiers. The main steps of this approach are 1) detecting the road and 2) filtering elements of interest and obtaining the road network. In this process, an attempt was made to outline the principal road guideline.

For the DNIT highways base, a semi-automated process developed by Macedo *et al.* [11] was chosen, involving the combination of: 1) vertex reduction techniques using ArcMap's ArcTool Box; 2) development of an algorithm to identify curves in AutoCad Civil 3D; 3) visualization of results using satellite images as a reference; and 4) creation of an alignment in AutoCad Civil 3D.

Based on the reconstruction of the alignment, a table was created containing all of the curve information (radius, angle, transition, deflection, degree of the curve, coordinates, length) and these were exported to an Excel table.

#### 3.1.3. Database Construction

The database stores information on the road network, the environment, and

road safety factors, including traffic accidents and traffic volume, which have been linked in order to combine the variables and assist in homogeneous segmentation.

The highway base was divided into kilometers using dynamic segmentation, making it possible to identify any information using the highway kilometer marks, such as accident data, traffic volumes, and the environment. From this georeferenced base with all the information attached, a dBase file was converted using ArcGis software into a points file. To connect this data, the tables containing other information use the highway to which they belong and the kilometer marks. This information was either point or linear. Accidents, access locations, signs, etc. were stored as point information, while geometric design, traffic, etc. were tabular information associated with linear features.

To ensure that the two sets of data were compatible, a combination of two techniques was used to create a common field in which to merge the datasets. With the first technique, a kilometer reference field (KM_REF) was created in the highway data table, as well as a conversion table, developed to create a common field. The conversion table recognizes the reference kilometer and associates the accident data to its corresponding kilometer. With the second technique, a spatial junction was performed between the tables, that is, each accident was spatially assigned to the segment to which it belonged. These two techniques allowed for the recognition and attribution of more than 99.6% of the accidents from the traffic accident reports of the DPRF, covering the period from January 1, 2007 to December 31, 2016, for highway BR-232, between km 141 and km 356.

## 3.2. Homogeneous Segmentation

The roads and all associated information were divided into homogeneous segments in three different ways: two by the methodology proposed by HSM [39] and another based on kernel density. They are listed below:

- Segmentation method 1: Based on HSM, segments are between intersections with a minimum length of 160 m.
- Segmentation method 2: Variation of the HSM method, with 50 m before and 50 m after curves, avoiding short segments and minimizing the problem of incorrect location of accidents.
- Segmentation method 3: Division of segments based on Kernel density and all variables used in the stepwise procedure are explanatory within each segment with their original values.

## 3.3. Statistical Modeling

The proposed model is classified as a Generalized Estimating Equations model, which can be interpreted as an extension of the Generalized Linear Models for panel data and incorporates a variety of variables in addition to just traffic volumes. The initial function was proposed by Liang and Zeger [40]:

$$\mu_i = \beta_0 * \left(\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_n X_n\right) + \varepsilon \tag{1}$$

Em que: $\mu_i$ = predicted annual rate of accidents; $\beta_0, \beta_1, \cdots, \beta_n$ = regression parameters; $X_{1i}, X_{2i}, \cdots, X_{qi}$ = the variables of interest; $\varepsilon$ = specification error

The choice of method is mainly due to the possibility of combining quantitative and categorical variables, not only as dummy variables (binary - 0 or 1), but as multinomial variables (having more than two ordinal categorical variables). The dependent variable is of the count type (number of accidents that occur in a given segment) and the linking function is a negative binomial.

To adjust a generalized linear model, the vector ($\beta$) of parameter estimates was determined. These coefficients were estimated from the observed data.

In this study, the first step was to verify whether the estimated coefficients were significant, that is, whether there was a statistically significant association between the explanatory variables and the response variable. Wald's chi-square statistical test was used to assess the adherence of the accident distribution between the actual and predicted data. The $\chi^2$ calc value was obtained from experimental data, taking into account both observed and expected values.

As this is an alternative hypothesis, in which the observed accident frequencies are different from the predicted frequencies, there was a need to verify the association between groups by comparing the calculated $\chi^2$ data with the tabulated $\chi^2$ data. The tabulated $\chi^2$ depends on the number of degrees of freedom and the level of significance adopted.

The hypothesis that the model fits the data is rejected if the p-value associated with the test statistic is less than the level of significance $\alpha$. Thus, for level of significance $\alpha$, a decision is made by comparing the two $\chi^2$ values:

If $\chi^2$ calculated $\geq \chi^2$ tabulated $\rightarrow$ the model is rejected

If $\chi^2$ calculated $\leq \chi^2$ tabulated $\rightarrow$ the model is accepted

The higher the $\chi^2$ value, the more significant the relationship between the dependent variable and the independent variable.

The quality-of-fit indications are based on the Wald Hypothesis Test values in the different models. The Wald test is used to test the null hypothesis that the estimated $\beta_j$ parameter is equal to zero.

Two statistical elements were considered when analyzing the quality of the fit of each model generated: 1) the Quasi-likelihood Information Criterion (QIC) and 2) the accumulated residue test (CURE Plot).

The QIC is a modification of the Akaike information criterion (AIC) in the GEE procedure. The comparison of the models is done using the maximum likelihood logarithm, which is the one that best fits the observed data. The QIC is expressed by Equation (2).

$$QIC = 2 * LIK + 2K \tag{2}$$

where: LIK = is the maximized likelihood log, $k$ = is the number of regression coefficients, and $r$ = number of parameters estimated for the calculation of $E_i$.

According to this criterion, the best model is the one having the lowest QIC value. Several other information criteria are available in the spatial statistics

tools, most of which are variations of the QIC, with changes in the way they penalize parameters or observations.

The CURE method to assess the quality of the fit is based on the study of residuals, that is, the difference between the number of accidents observed in a location and the value expected for the same location in the same time period, considering that residuals assume an abnormal distribution. The CURE Plot graph is used to examine residuals after estimating the parameters of the model and assessing whether the chosen function fits each explanatory variable over the entire range of values represented. The trend of residuals with respect to AADT (or other variables) can be assessed in relation to variance. An upward or downward deviation is a sign that the model consistently predicts fewer or more accidents, respectively, than were counted. It is therefore desirable that the cumulative graph of residuals oscillate close to zero or between two additional curves formed by the acceptable limits ($\pm 2\rho^*$) for cumulative residuals.

To validate the model, the Root Mean Square Error (RMSE) was used. RMSE is commonly used to express the accuracy of numerical results with the advantage that it presents error values in the same dimensions as the variable analyzed.

## 4. Results

### 4.1. Study Area

The scope of the analysis was highway BR 232, between km 141 and km 356, latitudes 8°02'30"S and 8°39'27"S and longitudes 36°11'56"W and 37°48'57"W (Figure 1). The 255 km stretch of rural federal highway runs through the municipalities of São Caetano, Pesqueira, Arcoverde, Cruzeiro do Nordeste, and Custódia, in northeastern Brazil.

Information was obtained from the Federal Highway Patrol Database for the years between 2007 and 2016, which contains the Incident Records and Police Reports, as well as from the DNIT highway base, the OSM cartographic base, and the digital terrain model provided by the Condepe/Fidem Agency.

The AADT values considered for the years 2014, 2015, and 2016 were obtained from the National Department of Transport Infrastructure (DNIT), including both volumetric and classificatory traffic counts. For previous years (2007 to 2013), as there was no active collection point in the study area, the AADT from the ANTT Annual Report (2015) was considered, as shown in Table 1.

The lack of standardization of police reports and the lack of rigor in filling them out reduce their reliability and their usefulness for studies. An analysis therefore had to be carried out to identify any absences or inconsistencies in the information recorded in the reports. Tables that did not contain all of the necessary information, such as location, type, and accident date, were excluded from the sample.

A database was created that grouped detailed information on lane widths, shoulder conditions, road curvature, grade, and AADT on the 215 km stretch of
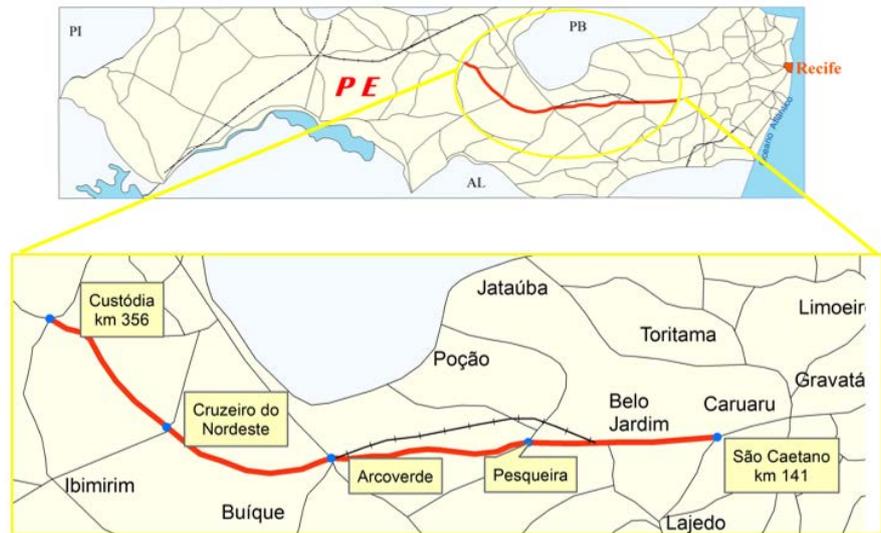
**Figure 1.** BR 232 Highway, between km 141 and km 356.

**Table 1.** Annual average daily traffic (2007 to 2016) for the section of BR 232-PE under study.

| Year | Volume (AADT) (vehicles/day) |
|------|------------------------------|
| 2007 | 5,997 |
| 2008 | 5872 |
| 2009 | 6220 |
| 2010 | 6317 |
| 2011 | 5989 |
| 2012 | 6480 |
| 2013 | 6317 |
| 2014 | 6684 |
| 2015 | 6720 |
| 2016 | 6530 |

rural highway in Pernambuco. This was achieved using geoprocessing tools to extract relevant attributes from the road network, spatial characteristics of the surroundings, and traffic flow, which were then combined with the accident database created for the study. The accident data included in the database contained all accidents registered over a 10-year period, from January 2007 to December 2016.

## 4.2. Homogeneous Segmentation

Two groups of variables were considered, one related to spatial variables (group 1) and the other to roadway geometry (group 2). The spatial variables considered were: accident cause, age group, accident type, day of the week, time, layout, condition, cause 1 (with injured victims, without victims, with fatal victims, ignored), road type, land use, period of the day (full daytime, full nighttime). The second group of variables included: lane width, shoulder width and type, segment length, grade and superelevation, curve radius and curve length, including the length of the transition spiral, if any.

For segmentation 1 (Figure 2), the results were verified using a sample of homogeneous road sections, selected to have a minimum length of 160 m. According to this criterion, of the 253 straight sections identified, 200 were selected. For the curved stretches, 88 out of a total of 226 were selected, meeting the criterion of a minimum radius of 100 m. Because of this significant reduction in the number of curves, tests were also carried out using a minimum radius of 50 m, totaling 115 stretches.

For segmentation 2, a variation of the homogeneous HSM segmentation of the HSM, the homogeneous stretches contiguous to the curves were excluded to a distance of at least 50 meters from the curve start and end points (Figure 3). This partial exclusion of the sections contiguous to the curves was done to isolate the influences of the curves when considering the accident history for the calibration procedure. It is necessary to differentiate accidents into those occurring on curves and on straight sections, however, this differentiation is performed in an approximate manner based on the km where the accident occurred. The accidents that occurred within these areas were added to their respective curved sections.

For homogeneous segmentation considering spatial criteria, grouping was performed by sub-sections according to the road surface type, land use, terrain type, roadway layout, and grade. Through the Query Builder tool, a consultation
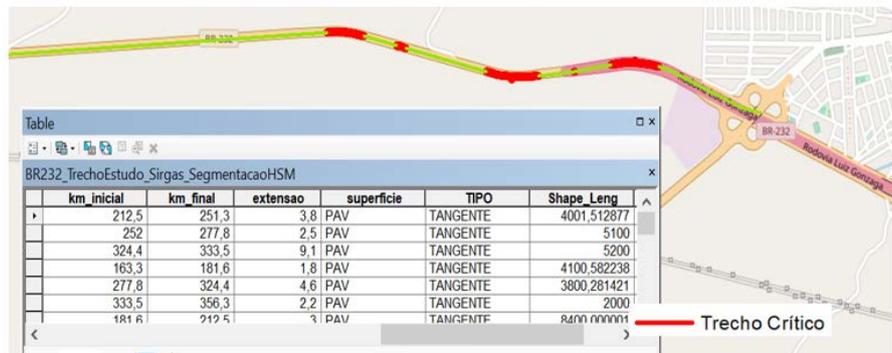


**Figure 2.** Table of homogeneous HSM segments.
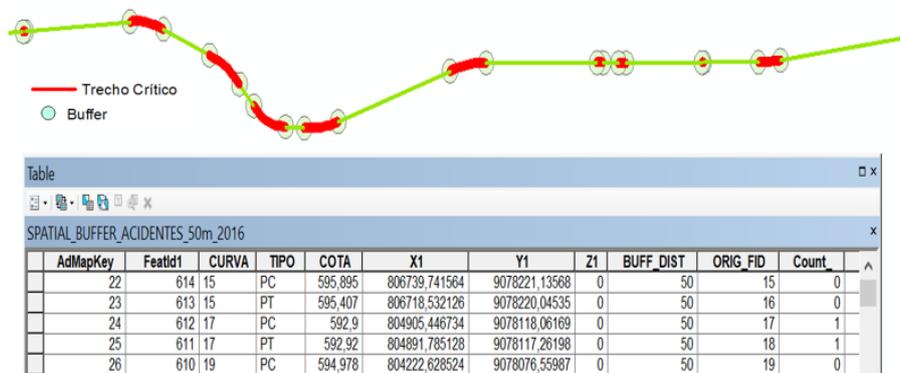


**Figure 3.** Table of segments contiguous to the curves tao a distance of 50 m.

was made to identify the accidents associated with each group and where they occurred, over the entire period of analysis.

At first, to ensure that segmentation was carried out according to the spatial characteristics without considering accident frequency, a Risk Index was created. According to the characteristics most often presented in the literature and their respective ranges, values were established ranging from 1 to 3, where 1 is low risk, 2 is medium risk, and 3 is high risk for accidents (Table 2). Table 3 shows the estimated risk index values for the category variables Day of the Week and Age Group.

The risk index ranges from 3 to 8, with 3 having the lowest risk and 8 the highest. For example, a stretch 1880 m long with an AADT of 4800 vpd on a downward slope has a risk index of 5, whereas a stretch with an AADT of 4800 vpd on a downward slope with a 500 m radius curve has a risk index of 7, according to the composition presented in Table 4.

The Kernel estimating technique was applied, based on the index, in order to identify the areas with similar spatial characteristics, as shown in Figure 4. When

**Table 2.** Estimated values for calculating the risk index.

| Variables | Categories | Estimated values |
|---|---|---|
| AADT [41] | ≤5500 vpd | 1 |
| | >5500 vpd | 2 |
| Curve radius (m) [42] | ≤600 | 3 |
| | 600 - 1500 | 2 |
| | >1500 | 1 |
| Grade (%) [36] | Negative, | 3 |
| | Positive, or Zero | 1 |
| Segment length (m) [43] | ≤200 | 1 |
| | 200-1000 | 2 |
| | ≥1000 | 3 |

**Table 3.** Estimated values for calculating the risk index for the category variables day of week and age group.

| Variables | Categories | Estimated values |
|---|---|---|
| Day of Week [44] | Weekday | 1 |
| | Weekend | 2 |
| Age Group (years) [45] | 18 - 30 | 3 |
| | 30 - 50 | 1 |
| | >50 | 2 |

**Table 4.** Example composition of the risk index.

| Variables | AADT ≤5500 vpd | Curve radius (m) ≤600 | Grade (%) Negative | Segment length (m) ≥1000 | Weekend |
|---|---|---|---|---|---|
| Values | 1 | - | 3 | 1 | 5 |
| | 1 | 3 | 3 | - | 7 |
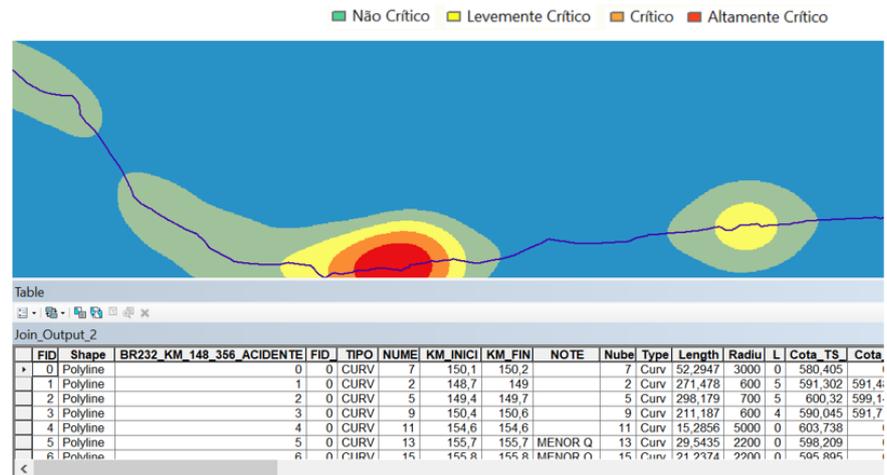
**Figure 4.** Table of homogeneous segments, considering spatial criteria.

crossing the spatial variables with the geometric variables, for example, the "road layout" and "grade" variables, the Kernel estimating technique was also applied to identify and verify the differences in concentrations between the road layout and the presence of a rising or descending slope. The procedure was repeated for various combinations of clusters.

After segmentation of the homogeneous stretches, the accidents that fit within the selected segments of highway were associated with them.

With the database structured in this manner, it was possible to compare the distribution of accident severity and accident frequency on curved stretches, considering the slope of the terrain. The results show that approximately 68% of accidents occur on straight stretches and 32% on curved stretches, however, attention is drawn to the accident severity. Of the accidents that occurred on straight stretches (220), 29% (64) were serious and 9% (9) were fatal, compared to 35% (37) and 18% (19), respectively, for curved stretches, which had a total of 103 accidents (**Figure 5**). The analyses also show that approximately 41% of accidents in curved stretches occurred on a descending slope, with 40% of the total being serious accidents and 19% fatal, while the percentage was less than 1% for all cases on straight stretches (**Figure 6**).

## 4.3. Calibration of the Proposed Model

The models developed were calibrated using the GEE technique, assuming errors with a Negative Binomial distribution because of the presence of a large number of observations with zero value and, therefore, high dispersion. In the SPSS software, version 23.0.0, this analysis can be found in the procedures: Analyze >> Generalized Linear Models >> Generalized Estimating Equations.

There was insufficient data to build a model for varying shoulder width values and traffic volume. The lane width was also constant throughout the study section. Although there was a single point in the entire section studied where traffic volume was counted, AADT was considered in the model, because its importance

**Figure 5.** Distribution of accidents on straight sections and curves from 2007 to 2016.
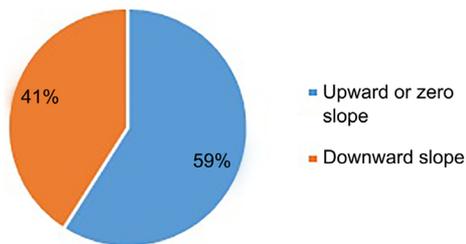


**Figure 6.** Distribution of accidents occurring on straight section and curves with downward slope for the period from 2007 to 2016.

is consolidated in the literature. Two adjusted models were then prepared, with three variations corresponding to the homogeneous segmentations. (1, 2, 3) for each model. They are:

Model 1—dependent variable (frequency); age group, day of the week, AADT (categorical variables); radius, grade, length (covariables); lane width and shoulder width (only for type 1 and 2 segmentation).

Model 2—dependent variable (frequency); grade (categorical variable); radius, length, AADT (covariables); lane width and shoulder width (only for type 1 and 2 segmentation).

To evaluate the effect of constant traffic, two variations of models 1 and 2, called models 3 and 4, were considered, using the same segmentations 1, 2, and 3, with AADT included as an offset variable. The model terms were factorially combined so that all combinations between variables could be evaluated. A summary of the estimated models is described in Table 5.

The significance of the parameter coefficients and the deviance were observed, in order to analyze whether the variables were significant for the model. With QIC, the correlation structures were evaluated and the best global model was selected with the CURE Plot. A significance level of 5% was used, meaning that

Table 5. Summary of estimated models.

| Model | Approach | Variables | Distribution Considered | Correlation Structure |
|---|---|---|---|---|
| 1 | For homogeneous segments, starting from the null model, the other variables were included one by one | AADT<br>Length<br>Radius<br>Lane Width<br>Shoulder Width<br>Day of the Week<br>Age Group<br>Grade | Negative Binomial | Exchangeable |
| 2 | For homogeneous segments, starting from the null model, the other variables were included one by one | AADT<br>Length<br>Radius<br>Lane Width<br>Shoulder Width<br>Grade | Negative Binomial | Exchangeable |
| 3 | For homogeneous segments with VDMA as an offset, the other variables were included, one by one | Length<br>Radius<br>Lane Width<br>Shoulder Width<br>Day of the Week<br>Age Group<br>Grade | Negative Binomial | Exchangeable |
| 4 | For homogeneous segments with VDMA as an offset, the other variables were included, one by one | Length<br>Radius<br>Lane Width<br>Shoulder Width<br>Grade | Negative Binomial | Exchangeable |

variables with a p-value greater than 5% were not considered to be significant. In the analysis of deviance, a Chi-square distribution of 5% significance was used. Therefore, if the contribution of the variable to the deviance was less than 1.96, the variable should not be included in the model.

When adding the lane width and shoulder width variables, neither model obtained satisfactory results. In both cases of the model tested, the parameter associated with the lane width and shoulder width variables was not statistically significant for $\alpha = 5\%$.

This result might be related to the constant values for all of the elements of the sample. The calibration results for models 1, 2, 3, and 4 are shown in Tables 6-9. Differences in the signs of the coefficients may indicate, depending on the segmentation, an opposite influence of the variable on the expected number of accidents estimated by the model.

The choice of working correlation matrix represents intra-individual dependency. The best structure should be sought, using the lowest QIC as a criterion. The QIC values found by adjusting models 1, 2, 3, and 4 with other correlation matrices are shown in Tables 10-13.

It can be seen that, according to the QIC parameter, the exchangeable correlation structure was the one that best fit the longitudinal data for the models generated.

**Table 6.** Estimated $\rho$ and SD values of model 1 for the different segmentations.

| | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Intercept | −3.820 (<0.0001) 1.6127 | 6.240 (0.0003) 1.9789 | 6.392 (<0.0001) 1.3723 | 5.030 (<0.0001) 1.4557 |
| AADT | −0.96 (0.230) 0.1577 | 0.028 (0.0233) 0.04422 | 1.4003 (<0.0001) 0.1198 | 0.520 (0.230) 0.1341 |
| Length | −0.475 (0.227) 0.0455 | 0.326 (0.200) 0.9718 | 0.702 (<0.0001) 0.0278 | 0.736 (0.977) 0.0133 |
| Radius | −0.011 (0.0054) 0.0189 | 0.280 (0.941) 1.2216 | 0.211 (0.0003) 0.0122 | 0.008 (0.402) 0.0105 |
| Lane Width | 0.1423 (0.0010) 0.1245 | 1.8788 (0.0172) 0.0123 | | |
| Shoulder Width | 1.049 (0.0749) 0.1522 | 1.164 (0.0500) 1.1156 | | |
| Day of the Week | 0 −0.249 | 0 0.383 | 0 0.710 (0.524) | 0 0.670 (0.678) |
| Age Group | 0.314 −0.120 0 | 0.172 0.024 0 | 0.967 (0.730) 0.172 (0.908) 0 | 0.227 (0.200) −0.340 (0.814) 0 |
| Grade | −0.060 (0.0015) 0.1756 | 0.0342 (<0.0001) 0.2716 | −0.3320 (0.0001) 0.1595 | 0.320 (0.527) 0.0322 |
| QIC | 19427.12 | 3247.23 | 2474.16 | 971.43 |

Number of observations in the database = 428.

**Table 7.** Estimated $\rho$ and SD values of model 2 for the different segmentations.

| | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Intercept | −15.2239 (<0.0001) 1.7233 | 8.1931 (0.0003) 1.5799 | −15.2277 (<0.0001) 1.9456 | −17.2512 (<0.0001) 1.032 |
| AADT | 1.3072 (<0.0001) 0.9978 | 0.7289 9 (0.0233) 1.241 | 1.4003 (<0.0001) 1.092 | 1.3997 (<0.0001) 0.0255 |
| Length | −0.232 (0.328) 0.0022 | 0.328 (0,118) 0.3421 | 0.211 (0,0003) 0.4467 | 1.472 (<0.0001) 0.0023 |
| Radius | 508.331 (0.0054) 0.0342 | −931.75 (0.0273) 0.0122 | 0.2111 (0.0003) 0.0342 | 284.2822 (0.0037) 0.0112 |
| Lane Width | 0.1423 (0.00 10) 0.0342 | 1.8788 (0.0172) 0.9711 | | |
| Shoulder Width | 3.1423 (0.00 10) 0.0034 | −3.0280 (0.0500) 0.0017 | | |
| Grade | 0.0076 (0.0015) 0.0678 | 0.0342 (<0.0001) 0.0774 | −0.3320 (<0.0001) 0.0227 | 0.0041 (0.0008) 0.0129 |
| QIC | 4226.10 | 2321.33 | 5310.22 | 600.30 |

Number of observations in the database = 428.

**Table 8.** Estimated $\rho$ and SD values of model 3 for the different segmentations.

| | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Intercept | −4.9712 (0.001) 1.4503 | −11.6339 (0.0020) 2.5998 | −7.4989 (0.001) 2.0095 | −4.3113 (<0.0001) 1.5015 |
| Length | −0.2861 (0.017) 0.0889 | 0.3998 (0.1980) 0.0981 | 0.3968 (0.001) 0.1185 | 0.2472 (0,977) 0.0906 |
| Radius | −90.6663 (0.022) 39.0627 | 0.7796 (0,9918) 0.4249 | 0.3160 (0.011) 0.1257 | 0.7349 (0.402) 0.2232 |
| Lane Width | 0.6927 (0.0041) 0.2209 | 0.3434 (0.0172) 0.1125 | | |
| Shoulder Width | 0.0413 (0.0829) 0.0088 | 1.2743 (0.0690) 0.2274 | | |
| Day of the Week | 0 −0.518 | 0 0.453 | 0 0.718 (0.526) | 0 0.720 (0.528) |
| Age Group | 0.619 −0.232 0 | 0.322 0.044 0 | 0.683 (0.5200) 0.234 (0.878) 0 | 0.442 (0.248) −0.284 (0.927) 0 |
| Grade | −0.597 (0.0435) 0.8430 | 0.0277 (0.009) 0.0106 | −0.2566 (0.037) 0.0123 | 0.3101 (0.726) 0.3929 |
| QIC | 24456.28 | 12428.16 | 5927.13 | 2822.14 |

Number of observations in the database = 428.

**Table 9.** Estimated $\rho$ and SD values of model 4 for the different segmentations.

| | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Intercept | −12.9435 (<0.0001) 2.5355 | 8.7422 (0.001) 1.7090 | −7.5992 (0.001) 1.8989 | −12.2009 (<0.0001) 2.9594 |
| Length | 0.3987 (0.010) 0.0966 | 0.3998 (0.001) 0.0936 | 0.3219 (0.003) 0.1119 | 0.4772 (<0.0001) 0.1127 |
| Radius | 1.2715 (0.004) 0.4407 | 0.2229 (0.043) 0.1042 | 0.4993 (0.003) 0.1737 | 1.2982 (0.0039) 0.5127 |
| Lane Width | 0.2898 (0.006) 0.1069 | 0.0209 (0.036) 0.0106 | | |
| Shoulder Width | 3.0014 (0.010) 0.0103 | −3.0010 (0.050) 0.1327 | | |
| Grade | 0.02883 (0.034) 0.0106 | 0.1527 (0.007) 0.1521 | −0.2121 (0.0003) 1.1197 | 0.3291 (0.011) 0.1302 |
| QIC | 14,844.37 | 7827.16 | 6424.12 | 1973.22 |

Number of observations in the database = 428.

**Table 10.** Values for QIC adjusting model 1 with other correlation structures.

| | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Exchangeable | 19,427.12 | 3247.23 | 2474.16 | 971.43 |
| Independent | 22,319.77 | 3441.29 | 2929.17 | 797.99 |
| AR(1) | 24,212.67 | 3835.33 | 3003.22 | 1098.22 |
| Non-Structured | 24,832.02 | 3913.88 | 3567.19 | 2756.34 |

Table 11. Values for QIC adjusting model 2 with other correlation structures.

|  | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Exchangeable | 4226.10 | 2321.33 | 5310.22 | 600.30 |
| Independent | 4231.70 | 4428.22 | 5397.27 | 1736.97 |
| AR(1) | 6969.90 | 8885.72 | 6211.12 | 1798.72 |
| Non-Structured | 9444.12 | 9144.90 | 6474.16 | 2224.18 |

Table 12. Values for QIC adjusting model 3 with other correlation structures.

|  | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Exchangeable | 24,456.28 | 12,428.16 | 5827.13 | 2822.14 |
| Independent | 24,731.76 | 14,323.29 | 5944.22 | 2939.99 |
| AR(1) | 25,694.43 | 13,825.14 | 5922.47 | 3495.88 |
| Non-Structured | 29,675.14 | 19,222.74 | 7282.19 | 4322.15 |

Table 13. Values for QIC adjusting model 4 with other correlation structures.

|  | Segmentation 1 | Segmentation 2 | Segmentation 3 | Segmentation 3 (adjusted) |
|---|---|---|---|---|
| Exchangeable | 14,844.37 | 7827.16 | 6424.12 | 1973.22 |
| Independent | 14,931.12 | 8622.27 | 6528.12 | 1995.43 |
| AR(1) | 15,786.44 | 9528.77 | 6599.32 | 1998.52 |
| Non-Structured | 18,767.67 | 9812.22 | 6896.19 | 1999.34 |

With this correlation structure, it can be said that the correlation between any two observations within a group are constant. The adjusted Segmentation 3 offered the best result for all models, however, most parameters were not statistically significant ($p > 0.05$).

The CURE Plot graphs of the models are presented in Figure 7 and Figure 8. For Models 1 and 2, it is possible to observe that the curve for cumulative residuals oscillates around 0 and does not cross the upper nor the lower acceptable limit. For models 3 and 4, the cumulative residual curve oscillates around 0 but exceeds the upper limit. Therefore, the best accident prediction model is model 2, because it presented the lowest QIC value (600.30).

The results obtained from the validation demonstrate that the best model for accident prevention is Model 2, because the root mean square error of the model adjustment ($\Delta$RMSE) is closest to zero, with a value of −0.082 (Table 14).

It is worth mentioning that the parameters obtained for the variables Day of the Week and Age Group agree with the values found in the simulations for variable selection. Taking the age group of those over 50 as a reference, young people between 18 and 30 are 22.7% more likely to be involved in fatal accidents while adults between 30 and 50 are 34% less likely to be involved in accidents. On weekends, the chance of accidents occurring is 67% higher than during the week.
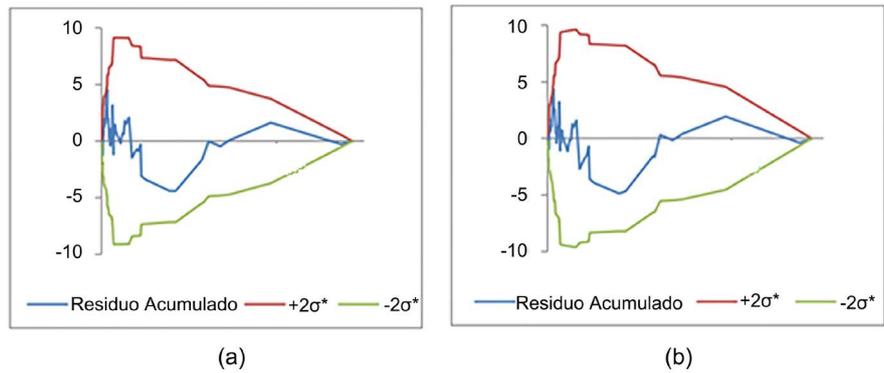
Figure 7. CURE Plot for the models developed (1 and 2). (a) Model 1; (b) Model 2.
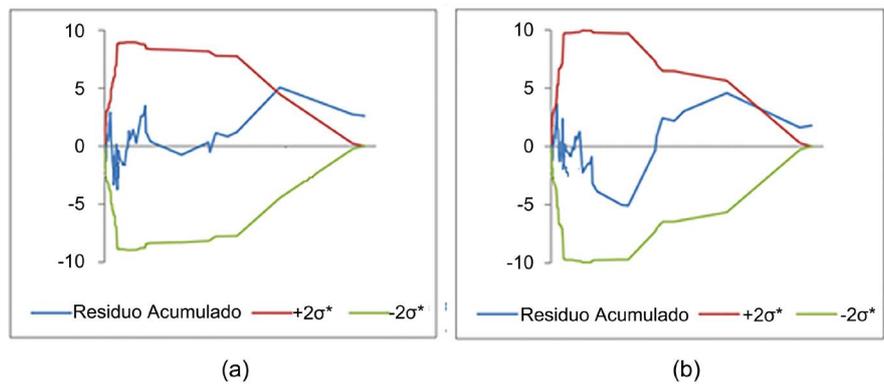


Figure 8. CURE Plot for the models developed (3 and 4). (a) Model 3; (b) Model 4.

Table 14. Model validation parameters.

| Models | Validation | | Adjusted | ΔRMSE |
| --- | --- | --- | --- | --- |
| | Average | RMSE | RMSE | |
| 1 | 0.807 | 0.973 | 1.101 | −0.082 |
| 2 | 0.843 | 1.156 | 1.168 | −0.112 |
| 3 | 0.799 | 1.112 | 1.472 | −0.173 |
| 4 | 0.873 | 1.114 | 1.267 | −0.129 |

For Segmentation 1, based on HSM, the selected variables have larger standard errors than those selected for other segmentation approaches. This is likely because, on highways, homogeneous segments change only at intersections, producing very long segments, where a great number of them have zero acidentes, and with considerable variation within individual segments in the other variables that cannot be adequately modeled.

For the model estimated for Segmentation 2, which includes 50 m of roadway on each end of a curve, the results were also significant. However, they tend to underestimate the number of accidents for low AADT values and overestimate accidents for higher AADT values.

Initially, the segmentation producing the worst results in the number of va-

riables that can be included in the model was Segmentation 3, in which all variables are explanatory for each segment. Therefore, variable categories were created, based on fixed value ranges, to improve the statistical power of the model. These categories were defined by attempts to obtain the best fit of the model and statistical significance for the main parameter estimates.

Finally, the GEE model was defined in order to predict the occurrence of accidents in a segment considering the AADT, curve radius, segment length, and grade, as shown in Equation (3):

$$\mu_i = e^{\left(\beta_0 + \beta_1 \text{AADT} + \beta_2 R + \beta_3 \text{Greide} + \beta_4 L + \varepsilon\right)} \tag{3}$$

where: $\mu_i$ = frequency of expected accidents per year; $\beta_0$ = intercept; $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ = parameters; $R$ = Curve radius (m), L = segment length (m), Grade = Grade (negative, positive, or zero), and $\varepsilon$ = error term.

Table 15 shows the model effects of all of the independent variables. The variable categories have no absolute values, but define the value of the parameter estimate ($\beta_n$ in column 3). The exponent of the estimated parameter (and $\beta_n$ in column 5) can be interpreted as a form of relative risk value for any declared variable category. This means that the following interpretations can be made based on each of the variables, considering that all other variables in the model have been kept constant.

The study showed that curves with a radius less than or equal to 600 m have a 3.2 times greater risk of accidents than curves with a radius greater than 2200 m (relatively straight). It also showed that sections with a downward slope have a risk of accidents 1.6 times greater than upward sloping or level road sections. Straight stretches longer than 1000 m on a downward slope, followed by a curve, have a risk of accident 2.2 times greater.

Equation (3) was solved for all variable categories in the model. The average value of the rate for accidents with victims in curved sections per kilometer was low: 0.048. This reflects the low frequency of accidents in these stretches. However, the causes may be related to the low traffic flow on rural roads, to the fact that there is a single point Where traffic data is collected in the studied section, or even the underreporting of this type of accident. Therefore, it was more significant to present the model's results from the combination of road characteristics, including the radii of the curves. For the sample mean of 0.048 accidents per km, the value 1.0 was defined.

To visualize the data more easily, a color code was applied: green represents an expected value for accidents with victims below the sample average (less than 1.0), yellow represents scenarios with a risk between the average and double the average (between 1 and 2), and orange represents scenarios in which the risk is two to three times the average value (between 2 and 3). The red color represents an extreme risk condition in which the predicted accident value was more than three times the sample average. Table 16 shows changes in the expected risk level for accidents in curves based on the predicted value.

From these results, it can be concluded that radii between 600 and 1500 meters

**Table 15.** Estimated parameters and effects of the log-linear negative binomial predictive model.

| Variables | Categories | Estimated parameter ($\beta_n$) | Estimated error parameter | Exponent of the estimated parameter ($e\beta_n$) | Statistical significance (Wald test) |
|---|---|---|---|---|---|
| Intercept | | −7.249 | 0.525 | 0.003 | p ≤ 0.001 |
| AADT | ≤5500 vpd | −0.605 | 0.134 | 0.546 | p ≤ 0.001 |
| | >5500 vpd | 0.000 | - | 1.000 | - |
| Curve radius (m) | ≤600 | 1.163 | 0.314 | 3.200 | p ≤ 0.001 |
| | 600 - 1500 | 0.539 | 0.203 | 1.716 | p ≤ 0.01 |
| | >1500 | 0.000 | - | 1.000 | - |
| Grade (%) | Negative | 0.470 | 0.428 | 1.600 | p ≤ 0.05 |
| | Positive or zero | 0.000 | - | 1.000 | - |
| Segment length (m) | ≤200 | 0.423 | 0.501 | 1.527 | p ≤ 0.001 |
| | 200 - 1000 | 0.930 | 0.167 | 1.213 | p ≤ 0.05 |
| | ≥1000 | 0.788 | 0.144 | 2.200 | p ≤ 0.01 |

**Table 16.** Changes in the predicted level of accident risk on curves.

| Curve radius (m) | Grade (%) | Segment length (m) | AADT (vpd) | |
|---|---|---|---|---|
| | | | ≤5500 | >5500 |
| ≤600 | Negative | <200 | 2.69 | 3.25 |
| | | 200 - 1000 | 1.53 | 1.85 |
| | | >1000 | 2.71 | 3.28 |
| ≤600 | Positive | <200 | 1.09 | 1.33 |
| | | 200 - 1000 | 0.62 | 0.75 |
| | | >1000 | 1.10 | 1.34 |
| 600 - 1500 | Negative | <200 | 1.29 | 1.56 |
| | | 200 - 1000 | 0.73 | 0.89 |
| | | >1000 | 1.30 | 1.57 |
| | Positive | <200 | 0.52 | 0.64 |
| | | 200 - 1000 | 0.30 | 0.36 |
| | | >1000 | 0.53 | 0.64 |
| >1500 | Negative | <200 | 0.61 | 0.74 |
| | | 200 - 1000 | 0.35 | 0.42 |
| | | >1000 | 0.62 | 0.75 |
| | Positive | <200 | 0.25 | 0.30 |
| | | 200 - 1000 | 0.14 | 0.17 |
| | | >1000 | 0.25 | 0.31 |

should be preferred in all scenarios for the design of new roads in order to reduce the frequency of accidents in curves. The results also show that long downward-sloping stretches followed by curves with radii less than 600 m offer

the greatest risk for accidents. If highways with a radius less than 600 m were converted into highways having radii greater than 600 m, accidents with victims in curves would decrease by about 18%. Roads with radii smaller than 600 m on a downward slope would see a reduction of 27%. With the model results and the historical accident numbers for the analyzed segments, the calibration procedure was carried out by dividing the actual total value by the calculated estimated value. The value obtained was 2.35 for Segmentation 1 and 1.75 for Segmentation 3.

## 5. Conclusions

The structuring of the database with a GIS was focused on the utilization of accident data, compared through the types of accidents that occurred, accident rates, accident indices, the situation of those involved, climatic conditions, vehicles, and with regard to the referenced period. The database structure sought to visualize the geometric parameters, mainly those of curves, not only through blueprints that do not always reflect the constructed reality, but through a semi-automated process proposed in this study combining several current and available databases. Geoprocessing techniques, such as reducing the excessive number of vertices, reconstructing curved elements, and smoothing segments, were necessary to improve the geometric quality of the road base.

The results are consistent when comparing the homogeneous segmentation between the Kernel map approaches and the statistical methods. This result was expected, because both methods work with the average severity of each accident. The discovery that homogeneous segmentation based on the Kernel estimator provides good results, shows that it is possible to create a hierarchy and establish geometric characteristics that have the greatest influence on the occurrence and severity of traffic accidents on rural single-lane Brazilian highways.

This model can be used to provide information about future revisions to the curve parameter selection guidelines, based on the principal road design parameters available in the Brazilian database. The modeling results can be used for curve selection, based on the reduction of accident risk.

The study produced clear indicators for the highway design parameters that influence the safety performance of rural highways. The exponents of the parameter estimates were statistically significant at $p \leq 0.1$ and the majority was significant at $p \leq 0.05$. Although the accident rate per kilometer on curves was low, the model highlights the severity of accidents on these stretches. It was concluded that radii between 600 and 1500 meters should be preferred in all scenarios for the design of new roads to reduce the frequency of accidents.

The carrying out of this study made it possible to verify that the rural roads in the state of Pernambuco are still 3.3 times more prone to accidents with fatalities than those in urban areas. Approximately 58% of fatal road accidents occur on horizontal curves, according to visual inspection when filling out accident reports, meaning that the true number may be higher. The analysis represents an

important step towards the revision of curve design guidelines. An approach to the design of curves based on the management of accident results may involve defining an increase in radius values and in the transition sections to meet the accident safety target for curves. As future study, the area of analysis is to be expanded and the methodology applied to other regions with similar characteristics to northeastern Brazil, as well as to other developing countries, not for the transferability of the model, but to fit the model and variables of interest to the regional level and subsequently adapt it to the national level.

## Acknowledgements

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Radimsky, M., Matuszkova, R. and Budik, O. (2016) Relationship between Horizontal Curves Design and Accident Rate. *Jurnal Teknologi*, **78**, 75-78. https://doi.org/10.11113/jt.v78.8493

[2] Departamento Nacional de Infraestrutura de Transportes DNIT (2010) Manual de projeto e práticas operacionais para segurança nas rodovias. Instituto de Pesquisas Rodoviárias, Rio de Janeiro.

[3] Findley, D.J., Hummer, J.E., Rasdorf, W., Zegeer, C.V. and Fowler, T.J. (2012) Modeling the Impact of Spatial Relationships on Horizontal Curve Safety. *Accident Analysis & Prevention*, **45**, 296-304. https://doi.org/10.1016/j.aap.2011.07.018

[4] Strathman, J.G., Dueker, K.J., Zhang, J. and Williams, T. (2001) Analysis of Design Attributes and Crashes on the Oregon Highway System. Publication FHWA-OR-RD-02-01, Federal Highway Administration, U.S. Department of Transportation, Washington DC.

[5] Lyles, R.L. and Taylor, W. (2006) Communicating Changes in Horizontal Alignment. NCHRP Report 559, Transportation Research Board, National Research Council, Washington DC.

[6] Charlton, S.G. (2007) The Role of Attention in Horizontal Curves: A Comparison of Advance Warning, Delineation, and Road Marking Treatments. *Accident Analysis and Prevention*, **39**, 873-885. https://doi.org/10.1016/j.aap.2006.12.007

[7] Mcgee, H.W. and Hanscom, F.R. (2006) Low-Cost Treatments for Horizontal Curve Safety. Publication FHWA-SA-07-002, Federal Highway Administration, U.S. Department of Transportation, Washington DC.

[8] Elvik, R. (2013) International Transferability of Accident Modification Functions for Horizontal Curves. *Accident Analysis & Prevention*, **59**, 487-496. https://doi.org/10.1016/j.aap.2013.07.010

[9]     Organisation for Economic Co-operation and Development (OECD/ITF) (2016) Road Safety Annual Report 2016. OECD Publishing, Paris.

[10]    Abdulhafedh, A.A. (2017) Novel Hybrid Method for Measuring the Spatial Autocorrelation of Vehicular Crashes: Combining Moran's Index and Getis-Ord G*i Statistic. *Open Journal of Civil Engineering*, **7**, 208-221. https://doi.org/10.4236/ojce.2017.72013

[11]    Macedo, M.R.O.B.C., Maia, M.L.A., Kohlman Rabbani, E.R. and Lima Neto, O.C.C. (2020) Remote Sensing Applied to the Extraction of Road Geometric Features Based on OPF Classifiers, Northeastern Brazil. *Journal of Geographic Information System*, **12**, 15-44. https://doi.org/10.4236/jgis.2020.121002

[12]    Ye, X., Pendyala, R., Shankar, V. and Konduri, K. (2013) A Simultaneous Model of Crash Frequency by Severity Level for Freeway Sections. *Accident Analysis and Prevention*, **57**, 140-149. https://doi.org/10.1016/j.aap.2013.03.025

[13]    Yu, R. and Abdel-Aty, M. (2013) Multi-Level Bayesian Analysis for Single- and Multi-Vehicle Freeway Crashes. *Accident Analysis and Prevention*, **58**, 97-105. https://doi.org/10.1016/j.aap.2013.04.025

[14]    Castro, M., Paleti, R. and Bhat, C.R. (2012) A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections. *Transportation Research Part B*, **46**, 253-272. https://doi.org/10.1016/j.trb.2011.09.007

[15]    Park, E.-S., Carlson, P., Porter, R. and Anderson, C. (2012) Safety Effects of Wider Edge Lines on Rural, Two-Lane Highways. *Accident Analysis and Prevention*, **48**, 317-325. https://doi.org/10.1016/j.aap.2012.01.028

[16]    Zegeer, C.V., Stewart, J.R., Council, F.M., Reinfurt, D.W. and Hamilton, E. (1991) Cost-Effective Geometric Improvements for Safety Upgrading of Horizontal Curves. Publication FHWA-RD-90-074, Federal Highway Administration, U.S. Department of Transportation, Washington DC.

[17]    Lee, J. and Mannering, F. (2002) Impact of Roadside Features on the Frequency and Severity of Runoff-Roadway Accidents: An Empirical Analysis. *Accident Analysis and Prevention*, **34**, 149-161. https://doi.org/10.1016/S0001-4575(01)00009-4

[18]    Zegeer, C.V. and Deacon, J.A. (1987) Effect of Lane Widht, Shoulder Widht, and Shoulder Type on Highway Safety. In: *Relationship between Safety and Key Highway Features: A Synthesis of Prior Research*, State of the Art Report 6, Transportation Research Board, Washington DC, 1-21.

[19]    Vogt, A. and Bared, J. (1998) Accident Models for Two-Lane Rural Segments and Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, **1635**, 18-29. https://doi.org/10.3141/1635-03

[20]    Karlaftis, M. and Golias, I. (2002) Effects of Road Geometry and Traffic Volumes on Rural Roadway Accident Rates. *Accident Analysis & Prevention*, **34**, 357-365. https://doi.org/10.1016/S0001-4575(01)00033-1

[21]    Shankar, V., Mannering, F. and Woodrow, B. (1995) Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis & Prevention*, **27**, 371-389. https://doi.org/10.1016/0001-4575(94)00078-Z

[22]    Hadi, M.A., Aruldhas, J., Chow, L.F. and Wattleworth, J.A. (1995) Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression. Transportation Research Center, University of Florida, Gainesville.

[23]    Persaud, B., Retting, R. and Lyon, C. (2000) Guidelines for Identification of Ha-

zardous Highway Curves. *Transportation Research Record*, **1717**, 14-18. https://doi.org/10.3141/1717-03

[24] Cafiso, S., Di graziano, A., Di Silvestro, G., La Cava, G. and Persaud, B. (2010) Development of Comprehensive Accident Models for Two-Lane Rural Highways Using Exposure, Geometry Consistency and Context Variables. *Accident Analysis and Prevention*, **34**, 357-365.

[25] Quddus, A.M., Chao, W. and Stephen, G.I. (2010) Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. *Journal of Transportation Engineering, ASCE*, **136**, 424-435. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000044

[26] Chiou, Y., Lan, L.L. and Chen, W. (2010) Contributory Factors to Crash Severity in Taiwan's Freeways: Genetic Mining Rule Approach. *Journal of the Eastern Asia Society for Transportation Studies*, **8**, 1865-1877.

[27] Haleem, K., Abdel-aty, M. and Mackie, K. (2010) Using a Reliability Process to Reduce Uncertainty in Predicting Crashes at Unsignalized Intersections. *Accident Analysis and Prevention*, **42**, 654-666. https://doi.org/10.1016/j.aap.2009.10.012

[28] Mustakim, F. and Fujita, M. (2011) Development of Accident Predictive Model for Rural Roadway. *World Academy of Science, Engineering and Technology*, **58**, 126-131.

[29] Eluru, N. (2013) Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables. *Accident Analysis and Prevention*, **55**, 1-11. https://doi.org/10.1016/j.aap.2013.02.012

[30] Boodlal, L., Donnell, E.T., Porter, R.J., Garimella, D., Le, T.Q., Croshaw, K., Himes, S., Kulis, P. and Wood, J. (2015) Factors Influencing Operating Speeds and Safety on Rural and Suburban Roads. Report No. FHWA-HRT-15-030, Federal Highway Administration, Office of Safety Research and Development, McLean.

[31] Costa, J.O., Freitas, E.F., Jacques, M.A.P. and Pereira, P.A.A. (2016) Collision Prediction Models with Longitudinal Data: An Analysis of Contributing Factors in Collision Frequency in Road Segments in Portugal. RS5C-Road Safety on 5 Continents.

[32] Kiran, B.N., Kumaraswamy, N. and Sashidhar, C. (2017) A Review of Road Crash Prediction Models for Developed Countries. *American Journal of Traffic and Transportation Engineering*, **2**, 10-25.

[33] Garnaik, M.M. (2014) Effects of Highway Geometric Elements on Accident Modelling. Thesis Master of Technology in Transportation Engineering, Department of Civil Engineering, National Institute of Technology, Rourkela.

[34] Agbelie, B.R.D.K. (2016) A Comparative Empirical Analysis of Statistical Models for Evaluating Highway Segment Crash Frequency. *Journal of Traffic and Transportation Engineering*, **3**, 374-379. https://doi.org/10.1016/j.jtte.2016.07.001

[35] Andriola, C.L. (2018) Análise da frequência e severidade de acidentes viários em curvas de rodovias de pista simples: O caso da BR 116. Masters Dissertation, Civil Engineering Graduate Program, Federal University of Rio Grande Do Sul, 201.

[36] Anastasopoulos, P.C., Shankar, V.N., Haddockc, J.E. and Mannering, F.L. (2012) A Multivariate Tobit Analysis of Highway Accident Injury-Severity Rates. *Accident Analysis & Prevention*, **45**, 110-119. https://doi.org/10.1016/j.aap.2011.11.006

[37] Chikkakrishna, N.K., Parida, M. and Jain, S.S. (2017) Identifying Safety Factors Associated with Crash Frequency and Severity on Nonurban Four-Lane Highway Stretch in India. *Journal of Transportation Safety & Security*, **9**, 32-30.

https://doi.org/10.1080/19439962.2016.1150927

[38] Sameen, M.I. and Pradhan, B. (2016) Forecasting Severity of Traffic Accidents Using Road Geometry Extracted from Mobile Laser Scanning Data. *The 37th Asian Conference on Remote Sensing* (*ACRS*), Sri Lanka, 17-21 October 2016, 1-6.

[39] American Association of State and Highway AASHTO (2014) Transportation Officials. Highway Safety Manual, Washington, EUA.

[40] Liang, K. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22. https://doi.org/10.1093/biomet/73.1.13

[41] Dong, C., Nambisan, S.S., Richards, S.H. and Ma, Z. (2015) Assessment of the Effects of Highway Geometric Design Features on the Frequency of Truck Involved Crashes Using Bivariate Regression. *Transportation Research Part A*: *Policy and Practice*, **75**, 30-41. https://doi.org/10.1016/j.tra.2015.03.007

[42] Cruz, P., Echaveguren, T. and González, P. (2017) Estimación del potencial de rollover de vehículos pesados usando principios de confiabilidad. *Revista ingeniería de construcción*, **32**, 5-14. https://doi.org/10.4067/S0718-50732017000100001

[43] Erdogan, S., Yilmaz, I., Baybura, T. and Gullu, M. (2008) Geographical Information Systems Aided Traffic Accident Analysis System Case Study: City of Afyonkarahisar. *Accident Analysis and Prevention*, **40**, 174-181.
https://doi.org/10.1016/j.aap.2007.05.004

[44] Souza, B.F. and Silva, J.P. (2017) Análise Espacial dos acidentes de trânsito em Passos (MG). *Ciência et Praxis*, **10**, 19-27.

[45] Mendonça, M.F.S., Silva, A.P.S.C. and Castro, C.C.L. (2017) Análise espacial dos acidentes de trânsito urbano atendidos pelo Serviço de Atendimento Móvel de Urgência: Um recorte no espaço e no tempo. *Revista Brasileira de Epidemiologia*, **20**, 727-741. https://doi.org/10.1590/1980-5497201700040014