Пут и саобраћај
Journal of Road and Traffic Engineering
www.putisaobracaj.rs

Developing a correlation model between the frequency of road traffic accidents and road geometric design elements

Semachew Molla Kassa [a*], Tarekegn Shirko Lachore[a], Abeje Tilahun Fetene[b], Habtamu Sewnet[b], Muluager Bewket Demlew[c]

[a] *Department of Civil Engineering, School of Civil and Construction, Wachemo University, Hosaena, Ethiopia*
[b] *School of Civil, Hydraulic, and Water Resource Engineering, University of Gondar, Gondar, Ethiopia*
[c] *Faculty of Civil and Water Resource Engineering, Bahir Dar University, Bahir Dar, Ethiopia*

ARTICLE INFO

DOI: 10.31075/PIS.70.04.01
Professional paper
Received: 05.10.2024.
Accepted: 14.10.2024.
Corresponding author:
smakmolla23@gmail.com

ORCID ID
Semachew Molla Kassa: 0000-0001-7920-8522
Tarekegn Shirko Lachore: N.A.
Abeje Tilahun Fetene: 0000-0003-4822-5180
Habtamu Sewnet: 0009-0003-8970-9293
Muluager Bewket Demlew: 0009-0001-7183-9343

***Keywords*:**
Road traffic accident,
Black spot
Accident frequency
Monte Carlo simulation
Principal Component Analysis

ABSTRACT

Road traffic crashes (RTCs) are a global problem that affects all sectors of society, irrespective of age, gender, or socioeconomic status. Specifically, the Wolaita zone in Ethiopia is currently experiencing many road traffic crashes. The study focused on the Halaba-Sodo area, a road segment that frequently witnesses traffic crashes within the Wolaita zone. This research aimed to comprehensively analyze the road geometric design elements that contribute to RTCs, with a particular focus on crash frequency. The factors contributing to RTCs were categorized into three groups: human factors, vehicular factors, and road and environmental factors. To achieve the research objectives, a total of 532 road crashes that occurred over six years (2011-2016 G.C.) were collected from secondary sources such as traffic police records and the transport office database along the entire Halaba-Sodo road. Additionally, field measurements and observations were conducted to gather the necessary data for crash analysis, in addition to the secondary data used in this study. Once the important data was collected using both primary and secondary data collection techniques, statistical analysis was performed using the principal component analysis method. Based on the analysis, significant variables with strong predictive abilities for model development were selected from the three categorized contributing factors. Subsequently, a mathematical model was developed using multiple nonlinear and partial least squares regression models. The Monte Carlo approach was utilized to determine the interactive effect of individual components or variables on the model, through 10,000 repeated simulations of samples within the probability density functions of the input data. The findings of the study revealed that gradient, sight distance, and horizontal curve were highly correlated with accident frequency, with correlation coefficients of 0.692, -0.529, and 0.426, respectively, among other road geometric factors. It was observed that gradient had a positive relationship with crash frequency, whereas sight distance and horizontal curve had a negative relationship. These findings provide valuable insights for engineers, urban planners, and policymakers in developing strategies to mitigate traffic crashes by implementing informed road design improvements.

1. Introduction

Road traffic accidents are a major issue that leads to an increase in property losses, injuries, and fatalities. They can provide serious health, financial, and developmental difficulties for drivers. As of 2019, traffic accidents rank as the 12th most common cause of death overall and the leading cause of mortality for children and youth (ages 5 to 29). As WHO reports, a road traffic crash is one of the causes of death of people and is ranked as one of the top leading causes of death in the world. Millions of people are killed each year by road traffic crashes. Ethiopia is one of the countries categorized as a low-income country in which road traffic crashes are reflected as a serious problem nowadays. The death rate due to a traffic accident is reported to be amongst the highest in the world (World Health Organization, 2020; Mohammed et al., 2019).

Ethiopia is one of the low-income countries and the most serious health issues are road traffic accidents (Fesseha Hailu Mekonnen & Sileshi Teshager, 2014; Abegaz & Gebremedhin, 2019). For instance, according to World Health Organization, (2020) reports, road traffic accidents claimed the lives of 100000 Ethiopians, accounting for 26.7 of all deaths. Therefore, road traffic accidents have been a key public health and development challenge and will continue unless effective measures are taken to control the problem. Road traffic crash also affects the national economy because; most of the people involved in crashes are within the productive age group (Tulu et al., 2013; Wijnen & Stipdonk, 2016). Currently, the rank of road traffic accidents is grouped under the worst diseases like malaria and others that attack human life in different regions of Ethiopia. As stated by (Yismaw & Ahmed, 2015; Zewude & Ashine, 2016), the Wolaita zone is one of the areas experiencing a large number of road traffic crashes in Ethiopia. The study area Halaba-Sodo is one of the road segments that cross the Wolaita zone in which traffic crashes frequently occur.

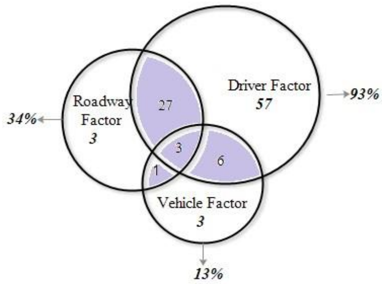

Figure 1. Contributing factors to vehicle crash [1]

Some of the primary geometric design elements that can affect traffic crashs are carriageway, grade, horizontal curvature, shoulder, median, vertical curve, super elevation, and sight distance (Mohammed, 2013). Most of the studies have shown the influence of various geometric design variables on the occurrence of crashs and have concluded that not all variables have the same level of influence in all places. Although it has been clearly shown that very restrictive geometric elements such as very short sight distances or sharp horizontal curves result in considerably higher crash rates and that certain combinations of elements cause an unusually severe crash problem (Mohammed, 2013; Urbanik et al., 1989; Gangwa & Deulkar, 2020) studied assessment on the impacts of geometric design on road safety on MettuGore road and the result confirmed that geometric parameters such as; the radius of a curve, super elevation, gradient, and sight distance are the most significant factors affecting road safety.

Many formerly published types of research on the study area developed several correlation models between the frequency of road traffic crashes and causes and contributing factors to a traffic crash. However, they used conventional methods to analyze the factors contributing to road traffic crashes rather than a clear measurement and model of the related features. Various research papers have studied traffic crashes using different statistical modeling techniques. The most common models used are the regression models, but there are many other techniques that have been used in the modeling by the researchers Chen et al., (2016) & Mohanty & Gupta, (2015) used binary logistic regression. Binary logistic regression is well-suitable for modeling a binomial outcome (takes the value 0 or 1 like having or not having a geometric feature) with one or more explanatory variables. This model is mainly made to describe, explain and predict decision-makers. The study shows that crash rates were found to be significantly (5.0% level) related to road design parameters. Similarly, Ali et al., (2014) & Savova-Mratsenkova & Djonev, (2019) used a multiple linear regression model to analyze the correlation between a traffic crash and independent variables. Because crash-frequency data are non-negative integers, the application of standard ordinary least-squares regression (which assumes a continuous dependent variable) is not appropriate. Given that the dependent variable is a non-negative integer, most of the recent thinking in the field has used the Poisson regression model as a starting point (Mohanty & Gupta, 2015).

Specifically, Poisson models may be negatively affected by low sample means, leading to biased results in small samples. This is because the Poisson regression model assumes that the mean and variance are equal. To address this issue within the data range, the Poisson-gamma/negative binomial model is commonly used in crash-frequency modeling. However, this model is limited in its ability to handle under-dispersed data and may encounter challenges in estimating dispersion parameters when the data has low sample means and small sample sizes. Therefore, this study aims to bridge this gap by conducting a comprehensive analysis and utilizing new statistical approaches, such as principal component analysis, to generate sound and robust conclusions. The primary objective is to develop a correlation model between the frequency of road traffic accidents and the geometric design elements rigorously. Additionally, identifying black spots in this particular road segment is crucial for implementing effective remedial solutions to address these severe issues. Over the years, significant data and methodological concerns have been identified in the literature on crash frequency. The ultimate goal of any model-building technique is to identify the most appropriate and parsimonious model that accurately describes the relationship between a given outcome variable (dependent or response) and a set of independent variables.

The study primarily focuses on the applications of linear and non-linear regression techniques for modeling and analyzing trends. Among various non-linear regression functions, three models were selected for practical use: 2nd-degree polynomial, exponential regression, and generalized linear regression with a log link. Based on the mean squared error (MSE), which is a well-known measure of model accuracy and quality, the non-linear exponential model exhibited the smallest MSE and was selected as the best-fitting model, surpassing simple linear and 2nd-degree polynomial models. Generalized linear regression ranked second best after exponential regression based on the MSE value...

2. Materials and Methods

2.1. Description of the Study Area

The study area Halaba - Sodo road is located in southern Ethiopia about 309 km distance from Addis Ababa. The road section is classified as a trunk road under the number A-7 (A-7-4) according to ERA classification 2013 (H. A. Mohammed, 2013). A-7 (A-7-4) is the number of road segments corresponding to the Halaba – Sodo road segment). Which connects the capital city Addis Ababa with Southern Nations, Nationalities, and Peoples Regional States. The road was constructed by DS4 (design standard four) as expressed in a detailed engineering design and a tender document prepared for supervision. The total length of the road segment assessed by the study was about 72 kilometers. Under this road length, more than three-fourths of the road traverses over flat terrain while the rest is rolling with a very limited mountainous section near Sodo town. The geographic position of the road falls between 7o18 N latitude to 38o05' E longitude and 6o54'N latitude to 37o45' E longitude with an elevation of 1600m to 2100m above sea level. The road traverses through three zones namely; Halaba, Hadiya, and Wolaita zones as shown on the map below in detail. Along its leg, the road connects various towns, villages, schools, worship, and other areas. The study area commonly experienced mixed traffic flow conditions including both motorized and non-motorized modes of transportation. Traffic growth in the area highly increasing from time to time with increasing population density (Levinson & Kumar, 1997; Chang et al., 2021; Batty et al., 2003). The following Figure 2 shows a map of the study area.

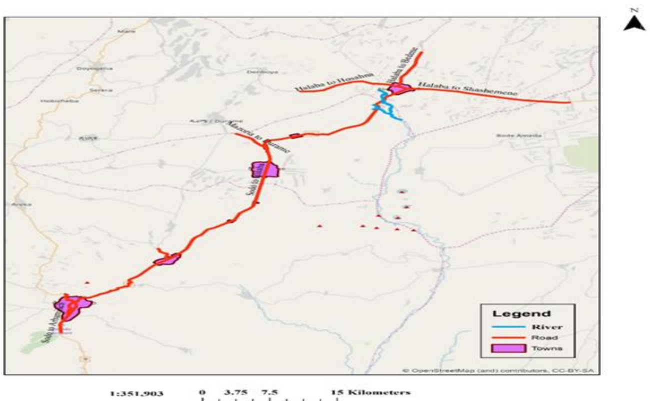

Figure 2. The study area map

2.2. Methods of Data Collection

To achieve a satisfactory outcome and align with the research objective, the data collection method and sources play a crucial role in this study. Considering the strategic importance and data requirements, both primary and secondary data were collected. The primary data was gathered through measurements, while the secondary data was sourced from traffic police records, the transport office database, and the ERA Sodo district office.

2.3. Road traffic crash data

The research was conducted using a total of 532 road crash data recorded over six years (2011 – 2016). The accident records were obtained from five traffic police stations, namely Wolaita Zone, Sodo Zuria Woreda, Damot Gale Woreda, Shone Woreda, and Halaba Liyu Woreda traffic police stations, covering the entire length of the road. Data was also collected from the transport office. To ensure data reliability, both data sources (traffic police records and transport office database) were linked for both rural and urban areas of the road section. Additionally, a portion of the data collected in 2017 was used to validate the model.

2.4. Road geometry data

The road geometry data was collected from the ERA Sodo district, in addition to the features documented in the traffic police accident database, through actual field measurements and visual inspections. Field measurements of the geometric elements were taken in places where adequate data from the Ethiopia Road Authority was not available, or where geometric problems and frequent traffic accidents occurred (black spots). The following common geometric elements were collected through the aforementioned methods: carriageway width, shoulder width, curve radius, length of the curve, gradient, super elevation, and sight distance. The sight distance data, along with other elements that required updated measurements, were collected using the manual method. Since using a survey instrument to measure the required variables is expensive, an alternative method was employed. For instance, sight distance was measured by using a sighting rod prepared to the driver's eye height (1.07m) and a target rod prepared to the object height for stopping sight distance (0.15m).as recommended in (Gemechu & Tulu, 2021; Wedajo et al., 2017).

2.5. Variables included in the study

Data collected from both primary and secondary sources are processed and analyzed by using a statistical technique known as principal component analysis (PCA). PCA is a modern powerful statistical tool for data analysis (Trcolea et al., 2011). The primary objective of PCA was to enhance the significance of variables by reducing the dimensionality. Eigenvalues, eigenvectors, squared cosines, contributions, correlation matrices, and correlation circles are the selected features utilized for identifying significant variables from the rest. Moreover, any redundant variables were eliminated based on these findings. In Principal Component Analysis, a multitude of variables are condensed into a principal component that explains the highest variability of the response (dependent) variable. This is advantageous as it reduces redundancy and facilitates faster computation for data modeling. The principal component analysis is often performed before a regression, to have a better overview of the variables and to avoid using redundant variables. The Table 1 below shows the variables, were imported for principal component analysis and their corresponding detail description.

Table 1. Candidate Variables Considered in the Principal component analysis

Variable	Description
Road environment-related factors	
Gradient of road	Algebraic difference of vertical grades in %
Sight distance	Sight distance in meter
Super elevation	An elevation difference of pavement edge in %
Width of pavement	Width of pavement in meter
Shoulder width Inner	Width of the shoulder in meter
Shoulder width Outer	Width of the shoulder in meter
Radius of curvature	The radius of curvature in meter
Length of curve	Length of the curve in meter
Traffic control device	1 if there is a traffic control device, = 0 otherwise

Based on the above criteria of PCA, significant predictor variables were selected and statistically insignificant variables were excluded. With these selected variables on principal component analysis, both the regression model and simulation have been processed.

2.6. Multiple non–linear regression

When a linear regression model cannot adequately represent the relationship between variables, the non – linear approach is an appropriate statistical technique to fit the relation (Hocking, 2013; Motulsky & Ransnas, 1987). Non-linear functions better characterize the relationships between crash frequencies and explanatory variables (Lao et al., 2014; Xie & Zhang, 2008). Non–linear regression is defined as a curvilinear relationship between response and predictor variables. Nonlinear regression can be solved by using different computer programs. Under various methods solver and XLSTAT.

$$Y = \beta 1 + e^{(\beta 2 + \beta 3 * X1 + \beta 4 * X2 + \ldots + \beta n + 2Xn)} + \varepsilon \tag{1}$$

Where:
Y - frequency of road traffic accident; X1, X2 …
X_n - contributing factors to a road traffic accident;
β1, β2, β3.... parameters to be fit in the model equation;
ε - uncorrelated random errors with mean zero and variance σ^2.

2.7. The method used for the simulation

Once the model equation was developed as per the objective of the study, simulation of the correlation between the frequency of traffic accidents and geometric design elements contributing to the accident was performed by Monte Carlo (MC) simulation method. Numerous options are available to use Monte Carlo simulations in computers. One can use any high-level programming language like C, C++, Java, or one of the .NET programming languages introduced by Microsoft to develop a computer program.

2.8. Road traffic crash black spot identification

According to Gangwa & Deulkar, (2020) Investigated several literature reviews of black spot analysis to identify the most appropriate method for black spot selection. Yet, they come up with the conclusion of no universally accepted definition and identification of a black spot. Locations are an assessment of risk and the likelihood of a crash occurring at a location. The most common assumption for a black spot location is that there should be any road environmental or geometric issues resulting in the repetition of crashs. Therefore, most of the scholars accepted the idea to accept dangerous sites (black spots) based on the priority value (Dereli & Erdogan, 2017; Bhavsar et al., 2021). This was due to restricted funding for hot spot safety work does put a limit to the number of sites that may be treated. Therefore, it is necessary to prioritize between sites and safety measures in order to utilize limited funds as effectively as possible.

To get a black spot on this particular road, six years of crash data record was distributed along the road to give a rough location of crashs. Then, a dangerous site was selected based on a priority value (P), calculated using the following formula similar to (Chen et al., 2016). The author has used the accident frequency method, accident density method, accident rate method, accident severity method, and quality control methods to identify black spot before decided to use priority method. For prioritizing those black spots, the ratios of accident costs by the degree of severity established by TRL was used, and the weight given for fatal accident is 5, for serious injury is 3, and for light injury and property damage are 2 and 1 respectively.

$$P = \frac{(1 + W + 2*X + 3*Y + 5*Z)}{D} \quad (2)$$

P = prioritization weight,
W = total number of property damage,
X = total number of light injuries,
Y = total number of serious injuries,
Z = total number deadly injuries,
D = distance of black spot section in km

3. Results and discussion

3.1. Black Spot Identification

The most common location for a black spot is a road segment that experiences repeated crashes. In accordance with the criteria outlined in chapter three for identifying black spots, 24 road sections were defined and grouped. Black spots were selected based on a priority value, with a distance of no more than 1.4km between each spot and a minimum value of 13. The geometric element used for analysis was measured at the top 10 black spots, which are listed in the table below.

Table 2. Top 10 black spots selected based on the priority value

No.	Selected Black spots	FTC	km	W	X	Y	Z	P
1	Kawo Shapa	13	0.5	2	0	5	6	94
2	Bortuwa Shapa (River)	11	0.6	1	3	3	4	60
3	Shone Adilo River	7	0.7	2	2	1	2	27.1
4	DalboWogene Village	7	0.9	1	2	1	3	25.6
5	Kokate Maracare	7	0.8	2	1	2	2	25
6	Dalbo St. Gebriel church (300m to Sodo)	5	1.1	1	0	2	3	20
7	Halaba Bilate River	6	0.9	2	1	1	2	18.9
8	In front of the LEWI International Hotel	4	0.8	1	0	2	1	15
9	DalboWogene Abatuna Limat	6	1.4	1	1	2	2	13.6
10	Kokate Forest Cooperative Nursery (River)	5	1.2	0	1	3	1	13.3

Table 2. below identifies the selected locations and parameters used to identify the black spots. The basic parameters used for black spot identification were Frequency of traffic crashes (FTC), Black spot (km), Total number of property damage (W), Total number of light injuries (X), Total number of serious injuries (Y), Total number of dead injuries (Z), Prioritization weight (P). The black spots shown below were the locations with geometric problems and traffic crashes frequently experienced in not more than 1.4 km in each spot.

The purpose of prioritizing these locations based on their priority value was to measure geometric features that were used in the analysis. In addition to the locations listed in the table above, there are several other spots that have experienced frequent traffic crashes. However, according to the traffic police records, the causes and contributing factors for these crashes are unrelated to road geometry. Examples of places that have experienced frequent traffic crashes due to human factors and road environments include rural villages, markets, schools, and church areas, specifically in the Mazoria, Shone, Buge, Gacheno, and Boditi town sections. Therefore, the causes and contributing factors beyond the crashes were analyzed in the following section. The figure below indicates the topographical show of the black spots among selected locations. Location is shown in Figure 3 (a) is named Kawo Shapa and Figure 3 (b) is named Bortuwa Shapa in which traffic crash occurs frequently. Both locations have sharp horizontal and vertical curves in the same area.

(a)

(b)

Figure 3. Topographical shows of sample black spots (a) Kawo Shapa, (b) Bortuwa Shapa

3.2. Analysis of crashes caused by road and environmental factors

The table below, Table 3 presents the distribution of road alignment in relation to total traffic crashes. It is evident that the majority of traffic crashes, accounting for 73.5%, transpired on straight and level roads. Following this, 9.2% of crashes took place on steep downward roads (known as Escarp), with the remaining 26.5% transpiring on curved and sloped areas. These findings suggest that human factors have a greater impact on road traffic crashes than road geometry.

Table 3. Frequency of crash and corresponding layout of the road

Road layout	Frequency of accident	Percentage shared on accident
Crash on the Straight and level road	391	73.5
Crash on a curved road	45	8.5
Crash on Escarp (steep downward) road	49	9.2
Crash on Slope Road	47	8.8

In addition to the road, the surrounding environment also contributes to road traffic collisions (RTCs) to some extent. Approximately 74% of accidents occur during daylight hours, while the remaining 24% occur at night. Similarly, out of the total number of crashes, 10% are caused by animals present on or near the road, including animal-drawn carts. Furthermore, environmental factors such as fog, rain, and other adverse conditions account for approximately 11% of all accidents. Table 4 presents the road geometry data imported into xl-stat for the principal component analysis purpose.

Table 4. Principal component analysis of road and its environment-related factors

Variable	Observations	Min	Max	Mean	Std. deviation
Width of pavement (m)	10	7	9	7.76	0.77
Shoulder width Inner (m)	10	1.6	2	1.75	0.15
Shoulder width Outer (m)	10	1.5	2.2	1.76	0.23
The radius of curvature (m)	10	41	320	167.4	100.28
Length of the curve (m)	10	45.6	339.	159.1	100.68
Sight distance(m)	10	106	210	134.4	29.71
Super elevation (%)	10	2.5	7.5	3.92	1.88
The gradient of the road (%)	10	3.8	6	5.04	0.63
Traffic control device	10	0	1	0.7	0.48
Traffic volume	10	1344	1611	1472	73.31
Frequency of Accident	10	4	13	7.1	2.8

The following Table 5 shows the eigenvalue result of road & its environment-related factors.

Table 5. Eigenvalue result of road & its environment-related factors

	F1	F2	F3	F4	F5	F6	F7	F8	F9
Eigenvalue	4.75	3.15	1.56	1.42	0.55	0.29	0.22	0.03	0.03
Variability (%)	39.58	26.24	13.01	11.87	4.59	2.39	1.81	0.29	0.23
Cumulative %	39.58	65.81	78.83	90.69	95.28	97.68	99.48	99.77	100

Table 6 shows the correlation between the selected predictor & response variables. Magnitudes of correlation coefficient of approximate 1 and -1 indicate a high correlation between variables. Correlation between dependent and independent variables determined by simulation result rather than PCA result. Based on redundancy criteria and other above-stated outputs of PCA, variables listed in the table 6 below were selected for the model development. The acronym GR, SE, SD, RC, PW, SW, and FTA, in the table 6 below stands for a gradient of the road, super elevation, sight distance, the radius of curvature, width of pavement, shoulder width, and frequency of traffic crash respectively.

Table 6. Correlation matrix of road & its environment-related factors

Variables	GR	SE	SD	RC	PW	SW	FTA
GR	1.000	-0.253	-0.736	-0.336	-0.865	-0.908	0.826
SE	-0.253	1.000	-0.167	-0.051	0.24	0.279	-0.36
SD	-0.736	-0.167	1.000	0.522	0.587	0.695	-0.663
RC	-0.336	-0.051	0.522	1.000	0.263	0.203	-0.725
PW	-0.865	0.24	0.587	0.263	1.000	0.828	-0.676
SW	-0.908	0.279	0.695	0.203	0.828	1.000	-0.732
FTA	0.826	-0.36	-0.663	-0.725	-0.676	-0.732	1.000

Based on PCA, the model prediction involved selecting significant variables while removing less significant variables from the list of crash contributing factors. For the development of the model, the following road factors were chosen: sight distance, road gradient, super elevation, pavement width, shoulder width, and radius of curvature. Traffic volume, curve length, and traffic control device were excluded based on the PCA criteria mentioned above.

The traffic volume in the areas where the crashes occurred was found to be relatively similar. In order to enhance the significance of the variable, the average shoulder width was calculated by merging the inner and outer shoulder widths during the model development process. This was done because the two variables were collinear with each other and their contributions were almost identical. Therefore, the shoulder width on either side could effectively represent this factor.

3.1. Model Evaluation, Interpretation, and Simulation Development

Once the relevant variables were selected based on the criteria of principal component analysis, the subsequent step involved developing a mathematical model that demonstrates the relationship between the frequency of road traffic crashes and geometric design elements. After the model was constructed, a Monte Carlo simulation was conducted to evaluate the performance of the functions.

3.1.1. Evaluation of model and simulation result

This section provides an overview of the multiple nonlinear regression model and simulation results regarding road factors. The model presents the outcome of the dependent variable (traffic crash frequency) and independent variables (road gradient, sight distance, super elevation, pavement width, shoulder width, and curvature radius) in a well-structured manner below.

3.1.2. The model parameters

Table 7 contains the parameters (coefficients) of the model for the frequency of traffic crash projection. This table gives the value of each parameter after fitting it to the model.

Table 7. Model parameters

Parameter	β1	β2	β3	β4	β5	β6	β7	β8
Value	2.806	0.83	0.436	-0.062	-0.007	-0.002	-0.012	-0.032

The equation of the model developed to simplify the simulation can be rewritten in the form of; FTA = 2.806 + Exp (0.830 + 0.436 GR – 0. 062 SE – 0.007 SD – 0.002 RC – 0.012 WP – 0.032 SW). As described below, once its goodness was checked the equation was used as input in simulation development.

Figure 4. presents the relationship between observed and predicted values of the frequency of road traffic crashes. Comparison of predicted and observed values was also used to evaluate the model's ability to predict. The value of predicted and observed data on the frequency of traffic crashes closed in a given confidence interval assures the quality of the model. As shown on the graph both observed and predicted values of frequency of traffic crashes are very close in curve fitting.

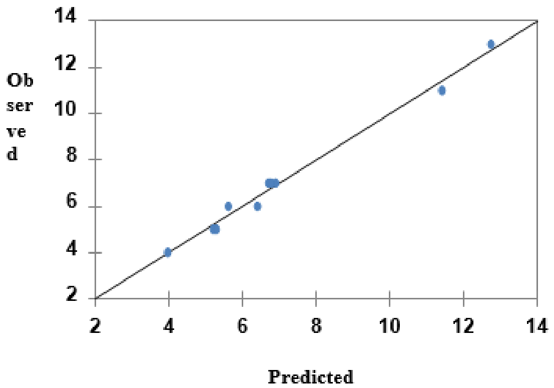

Figure 4. Predicted versus observed frequency of crash

3.2.3. The goodness of fit statistics for frequency of traffic crash

Table 8 shows the goodness of fit for road factors with the dependent variable (frequency of traffic crashes). The performance of the model is measured by the coefficient of the model (R2) square error (RMSE) of the average of the analysis expressed in the original measurement unit. The higher the R2 (closer to 1) and the lower the RMSE, the higher the precision and accuracy of the model to predict the frequency of traffic crashes. R2 = 0.987 shows that 98.7% of the variation of the dependent variable is explained by the explanatory/independent/ variables. In the same way, the variation of the dependent variable unexplained by the independent variable is very few in percentage.

Table 8. Goodness of fit statistics

Observations	DF	R^2	SSE	MSE	RMSE	Iterations
10	2	0.987	0.836	0.418	0.646	64

Therefore, analysis of the model shows that the model is well fitted with (R2 = 0.987) and could be used for prediction or simulation purposes
.

3.3.4. Monte Carlo simulation result

The simulation was used to study the interactive effect of individual components or variables to determine which ones are more important. As described in the above section, the simulation was run 10,000 times. Conducting sensitivity analysis is used to determine which parameters of the model are most critical (the “sensitive parameters”) in determining the solution. Table 9 shown below is the sensitivity result of the 10,000 times simulation.

Except for the gradient of the road, all variables correlated negatively to a road traffic crash. A gradient of a road, sight distance, and radius of curvature have a high contribution to a traffic crash with an absolute value of 47.85%, 27.91%, and 18.09% respectively. The simulation result shows that the width of the shoulder and pavement contributes less effect on the frequency of traffic crashes with absolute value contributions of 0.44% and 0.14% respectively.

Table 3. Sensitivity (frequency of traffic crash)

Distributions	Correlation	Contribution	Contribution (Absolute)
Gradient	0.692	47.85%	47.85%
Sight distance	-0.529	-27.91%	27.91%
Radius of curvature	-0.426	-18.09%	18.09%
Super elevation	-0.236	-5.57%	5.57%
Shoulder	-0.066	-0.44%	0.44%
Width pavement	-0.037	-0.14%	0.14%

The above chart allows for identifying which of the explanatory variables contribute more to the projection model. Based on the correlation coefficient, it is possible to order the predictor variables according to their influential strength. The gradient of the road → sight distance → radius of curvature → Super elevation → Shoulder width → Width of pavement are parameters in influential order based on sensitivity analysis.
Figure 5. similarly shows the contribution of road geometry to the frequency of traffic crashes. Their contribution is also similar to that of sensitivity analysis shown in the above Table.

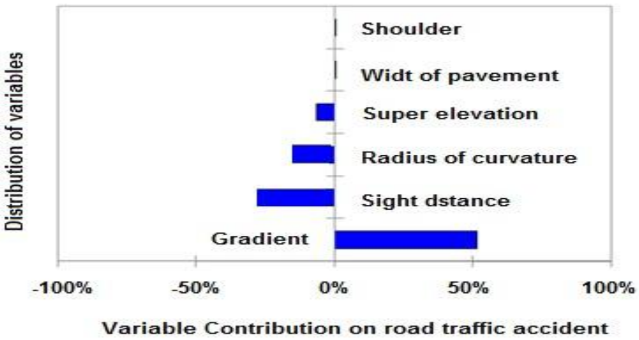

Figure 5. Sensitivity (Frequency of traffic crash)

3.4. Effect of variable change in frequency of road traffic crash

The effect of all individual variables on the frequency of crashes is determined by making all other variables constant and changing the value of a single required variable. A unit change in variables had a significant effect on crash frequency. Accordingly, when the gradient increased from 3% to 6%, the frequency of the crash increased by 4. When the sight distance increased from 80m to 100m frequency of the crash decreased by 1 and when the sight distance increased from 80m to 200m crashes decreased by 3. Likewise, when the radius of the curve increased from 50m to 180m number of crashes was reduced by 2. Similarly, Super elevation has also the same effect on crashes as a radius of curvature. Pavement and shoulder width between 6 – 10m and 1.2 – 2.5m respectively made almost similar contributions to the number of crashes.

Figure 6 shows a simulation result of the correlation between predictor variables with crash frequency in the scatter plot. The scatter diagram shows the correlation between predictor variables and the response variable (frequency of crash). The line that fits well indicates the correlation of variables with the frequency of crashes and vice versa. Gradient, sight distance, and radius of curvature correlated strongly with road traffic crashes. While, super elevation, the width of the shoulder, and pavement correlated weakly with road traffic crashes. The magnitude of the correlation coefficient indicates the amount of independent variables' influences on the dependent variable.

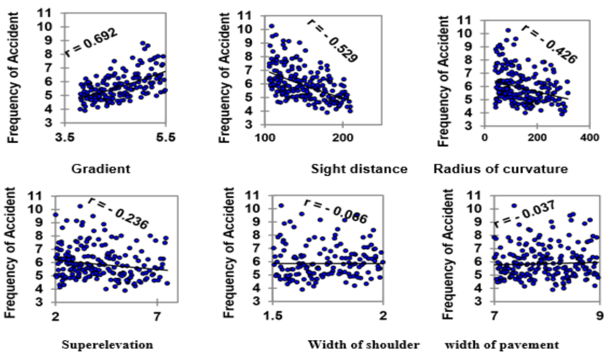

Figure 6. Scatter plots of variables with the frequency of crash

3.5. Model Validation

This method evaluates the accuracy of crash frequency calculated from the equation by analysis and the differences observed. Validation of the models against additional years of crash data for the same stretch was used in the estimation. This validation is used to assess the model's ability to forecast accidents across the road in a given time. Traffic crash data for 2010 has been collected from traffic records in the same method in the study section and calculated the crash frequency using three model equations developed for the same highway. The result of both the predicted and observed frequency of accidents emanated within the assumed confidence interval. Observed and predicted crash frequency assures the validity of the model (Gangwa & Deulkar, 2020) Used estimation and validation samples from the same general area and came up with a good validation process for the model. Basso et al., (2018) used validation of the models against additional years of crash data for the same stretch used in the prediction and additional segment of the same highway. The authors' results in both methods declare the validity of the predicted model using a multiple linear regression model. A high R2 value is needed for the goodness-of-fit while R2 with a low value shows either some predictor variables are missing or the relationship between the dependent and independent variables is not linear or does not follow a normal distribution. Similarly, a P-value below 0.01 (99% confidence interval) was considered statistically significant and was undertaken to suggest possible engineering measures. The P-value of the simulation result is less than 0.01 which shows that the correlation is typically considered significant and the model developed was valid. This was due to the strong correlations having low p – p-values because the probability that they have no relationship is very low.

4. Conclusion

This study identified the primary road-related contributing factors for black spots. These factors include horizontal curves, sight distance, steep gradients, super elevation, pavement, and shoulder width. The highest number of crashes occurred on straight and level roads rather than other types of terrain. However, steep gradients, limited sight distance, and sharp horizontal curves had a higher frequency of crashes compared to straight sections with the same length and traffic composition. Among all road geometric factors, gradient, sight distance, and horizontal curve showed a strong correlation with crash frequency, with correlation coefficients of 0.692, -0.529, and -0.426, respectively. Road gradients were found to have a significant positive impact on crash occurrence, accounting for approximately 47.85% of the total selected road geometric elements. Specifically, an increase in gradient from 3% to 6% resulted in a 4% increase in crash frequency.

A particularly significant effect on crash frequency was observed for road gradients ≥ 7%. The negative correlation coefficients for sight distance and horizontal curves indicate their negative contribution to road traffic crashes. Sharp horizontal curves and limited sight distance were found to have a significant influence on crash rates, especially for radii below 200 m and sight distances below 100 m. Super elevation, shoulder width, and pavement width showed lower correlations with crash frequency compared to the aforementioned variables. However, they still had a negative impact on crash frequency. In terms of pavement width, a width of 7.0 m is considered sufficient for two-way two-lane roads, assuming equivalent traffic volume. Similarly, shoulder widths between 1.5 and 2.5 m have a similar effect on crash occurrence. The influence of contributing factors can be determined by both the direction and magnitude of the correlation coefficient. Variables with a lower magnitude of correlation coefficient have a smaller impact on crash frequency, while those with a higher magnitude have a greater impact.

Data availability statement

All data are available within the article.

Conflict of interest

The author declares there is no conflict of interest.

References

[1] Abegaz, T., & Gebremedhin, S. (2019). Magnitude of road traffic accident related injuries and fatalities in Ethiopia. PloS One, 14(1), e0202240.

[2] Ali, E. K., El-Badawy, S. M., & Shawaly, E.-S. A. (2014). Young drivers behavior and its influence on traffic accidents. Journal of Traffic and Logistics Engineering, 2(1), 45–51.

[3] Basso, F., Basso, L. J., Bravo, F., & Pezoa, R. (2018). Real-time crash prediction in an urban expressway using disaggregated data. Transportation Research Part C: Emerging Technologies, 86, 202–219.

[4] Batty, M., Besussi, E., & Chin, N. (2003). Traffic, urban growth and suburban sprawl.

[5] Bhavsar, R., Amin, A., & Zala, L. (2021). Development of model for road crashes and identification of accident spots. International Journal of Intelligent Transportation Systems Research, 19, 99–111.

[6] Chang, Y. S., Jo, S. J., Lee, Y.-T., & Lee, Y. (2021). Population density or populations size. Which factor determines urban traffic congestion? Sustainability, 13(8), 4280.

[7] Chen, C., Wang, Y., Ma, C., & Zhang, W. (2016). How expressway geometry factors contribute to accident occurrence? A binary logistic regression study. Periodica Polytechnica Transportation Engineering, 44(4), 215–221.

[8] Dereli, M. A., & Erdogan, S. (2017). A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods. Transportation Research Part A: Policy and Practice, 103, 106–117.

[9] Fesseha Hailu Mekonnen, F. H. M., & Sileshi Teshager, S. T. (2014). Road traffic accident: The neglected health problem in Amhara National Regional State, Ethiopia.

[10] Gangwa, M., & Deulkar, V. (2020). A Review on Effect of Highway Geometric Elemnets on Accident Modeling. International Journal of Scientific Research & Engineering Trends.

[11] Gemechu, S. M., & Tulu, G. S. (2021). Safety effects of geometric design consistency on two-lane rural highways: The case of Ethiopia. American Journal of Traffic and Transportation Engineering, 6(4), 107–115.

[12] Hocking, R. R. (2013). Methods and applications of linear models: Regression and the analysis of variance. John Wiley & Sons.

[13] Hsiao, H., Chang, J., & Simeonov, P. (2018). Preventing emergency vehicle crashes: Status and challenges of human factors issues. Human Factors, 60(7), 1048–1072.

[14] Lao, Y., Zhang, G., Wang, Y., & Milton, J. (2014). Generalized nonlinear models for rear-end crash risk analysis. Accident Analysis & Prevention, 62, 9–16.

[15] Levinson, D. M., & Kumar, A. (1997). Density and the journey to work. Growth and Change, 28(2), 147–172.

[16] Manual, H. S. (2010). American Association of State Highway and Transportation Officials. Washington, DC, 19192.

[17] Mohammed, A. A., Ambak, K., Mosa, A. M., & Syamsunur, D. (2019). A review of traffic accidents and related practices worldwide. The Open Transportation Journal, 13(1).

[18] Mohammed, H. (2013). The influence of road geometric design elements on highway safety. International Journal of Civil Engineering and Technology, 4(4), 146–162.

[19] Mohammed, H. A. (2013). The influence of road geometric design elements on highway_2013. International Journal of Civiil Engineering & Technology (IJCIET), 4(4), 146–162.

[20] Mohanty, M., & Gupta, A. (2015). Factors affecting road crash modeling. Journal of Transport Literature, 9(2), 15–19.

[21] Motulsky, H. J., & Ransnas, L. A. (1987). Fitting curves to data using nonlinear regression: A practical and nonmathematical review. The FASEB Journal, 1(5), 365–374.

[22] Porter, R. J., Donnell, E. T., & Mason, J. M. (2012). Geometric design, speed, and safety. Transportation Research Record, 2309(1), 39–47.

[23] Savova-Mratsenkova, M., & Djonev, G. (2019). Determining the values of the Coefficient of Restitution in the meanwhile of a crash between two vehicles. 618(1), 012058.

[24] Srisurin, P., & Chalermpong, S. (2021). Analyzing Human, Roadway, Vehicular and Environmental Factors Contributing to Fatal Road Traffic Crashes in Thailand. Engineering Journal, 25(10), 27–38.

[25] Trcolea, C., Parıs, A. S., & Voıcu, P. (2011). Principal Component Analysis Applied to Agricultural Equipments. Tarım Makinaları Bilimi Dergisi, 7(3), 305–308.

[26] Tulu, G. S., Washington, S., & King, M. (2013). Characteristics of police-reported road traffic crashes in Ethiopia over a six year period. 1–13.

[27] Urbanik, T., Hinshaw, W., & Fambro, D. (1989). Safety effects of limited sight distance on crest vertical curves. Transportation Research Record, 1208, 23–35.

[28] Wedajo, T., Quezon, E. T., & Mohammed, M. (2017). Analysis of road traffic accident related of geometric design parameters in Alamata-Mehoni-Hewane section. Int. J. Sci. Eng. Res, 8(1), 874–881.

[29] Wijnen, W., & Stipdonk, H. (2016). Social costs of road crashes: An international analysis. Accident Analysis & Prevention, 94, 97–106.

[30] World Health Organization. (2020). European regional status report on road safety 2019. World Health Organization. Regional Office for Europe.

[31] Xie, Y., & Zhang, Y. (2008). Crash frequency analysis with generalized additive models. Transportation Research Record, 2061(1), 39–45.

[32] Yismaw, A., & Ahmed, M. E. (2015). The Causes of Road Traffic Accidents in Bahir Dar City, Ethiopia. International Journal of African and Asian Studies, 11, 99–107.

[33] Zewude, B. T., & Ashine, K. M. (2016). Statistical Modeling on Determinants of Traffic Fatalities and Injuries in Wolaita Zone, Ethiopia. Global Journal of Human-Social Science Research.