Development of Crash Prediction Model for Hill Roads: A Case Study of NH06, Nepal

Bidhan Dahal[1], Anil Marsani[2]

[1]Graduate Student, Dept. of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Lalitpur, Nepal

[2]Assistant Professor, Dept. of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Lalitpur, Nepal

***Abstract*: Crash prediction models (CPMs) have been used in many countries as a useful tool for road safety analysis and design. Each model is different in terms of methodology, data accuracy, variability in highway geometry and explanatory variables used to predict crashes. The model used in this research are based on Generalized Linear Modelling Technique. Two different models; namely Poisson and Negative Binomial models have been used in the analysis. The explanatory variables used are number of horizontal curves, total length of horizontal curves, maximum grade, access density, minimum sight distance within a segment, minimum radius of curvature and presence of hairpin bends within a segment. Out of the explanatory variables, access density, presence of hairpin bends and minimum radius of curvature were found to be significant predictors.**

***Keywords*: CPM, Crash, Safety, Regression, Generalized Linear Model**

1. Introduction

Crash-prediction models are decision-making tools for transportation engineers to provide an estimate of expected crash frequency as a function of various explanatory variables depending on the scope of study. Modeling of crash count data is considered as an important task in road safety. The number of crash occurrences within a given time frame is called the crash frequency, which is used as an indicator of the crash occurrence at highways or certain segments of the roads. CPMs have been developed for various kinds of roads in the past in different countries. The most prominent of the ones developed is the Safety Performance Function (SPF) suggested by Highway Safety Manual (HSM) to be used after applying calibration factor for local conditions. As the manual is applicable only to road segments of homogenous characteristics, researchers have recommended developing indigenous models to predict crash frequencies in developing countries where heterogeneity in traffic composition is observed (Shah and Basu, 2017). As road crash is a rare event, typically, generalized linear models (GLMs) have been used to model crash outcomes based on explanatory variables (average annual daily traffic, lane width, segment length, presence of shoulders, access density etc).

Generally, Poisson and Negative binomial based models have been extensively used for the purpose.

2. Literature Review

The contributing factors that lead to an actual event of crash occurrence are multi-dimensional. They have been generally classified in relevant literature into behavioral factors related to driver behavior and non-behavioral factors related to highway geometry, vehicle and traffic conditions, road side environment, etc. (Caliendo et al., 2007).

Schneider et. al. (2009) developed a crash prediction model for truck crashes on horizontal curves using truck ADT, passenger vehicle ADT, and degree of curvature and segment length. Other studies have developed crash prediction models for horizontal curves using limited variables. Bonneson et al. (2005) developed horizontal curve crash prediction models for multilane highways using radius and speed limit data. similarly, Fitzpatrick et al. (2009) developed a crash prediction model for freeways using single independent variable: degree of curvature and assuming zero degree as the base condition. Likewise, there have been other studies on significant variables affecting crash frequency. 500-ft radius curve was found to be 200% more likely to produce a crash than an equivalent tangent section, and a 1,000-ft radius curve is 50% more likely to produce a crash than an equivalent tangent section (Zegeer et al.,1991).

Although crash prediction models were initially based on MLR (Multiple linear regression) models, but as the data was found to be better fitted with the Poisson distribution, it was started to be used using an advanced modeling technique called the Generalized Linear Models (GLM), instead of the conventional multiple linear regression technique (Caliendo et al., 2007).

Multivariate regression models specifically Poisson regression model and Negative Binomial model have been widely used in the crash prediction models (Lord, D. and Mannering, F., 2010). Negative Binomial (NB) distribution (or Poisson-Gamma) overcomes the problem of mean equal to variance in Poisson distribution, and can be more accurate for over-dispersed data (Geedipally et al.,2012).

The issue of segmentation (segregation of the crash data based on the spatial location of the crash occurrence) in crash modelling has been widely discussed in literature (Koorey, 2009; Fitzpatrick et. al., 2006). Various segmentation approaches have been used to segregate the crash data based on their location. The Highway Safety Manual has prescribed the use of homogeneous segments with respect to AADT, lane width, curvature, number of lanes, driveway density, shoulder width, shoulder type, roadside hazard rating, median width and clear zone width. The manual has suggested the minimum segment length to be no less than 0.10 miles to ensure ease of calculation and consistency in results (AASHTO, 2010).

Recent research (Cafiso et al.,2018; Green, 2018) have gone to great depths on investigating the statistical implications of various segmentation strategies on the performance of the crash prediction models. Cafiso et al. (2018) has discussed that while crash-based segmentation is likely to identify optimal segments for safety analysis, it is less practical than a fixed segment based on roadway data. After comparative analysis of various segmentation approaches based on goodness of fit, Green (2018) found out that the segmentation approach with fixed length of 650 m, coinciding with the maximum length of an interchange area, and selected to be just longer than the longest horizontal curve, gave the best results.

3. Methodology

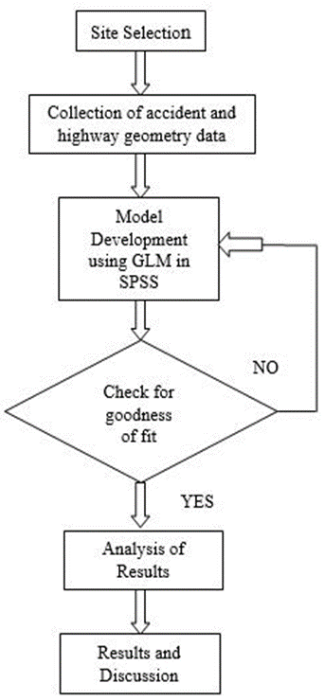

Fig 1. Methodological Framework

A. Site Selection

The hill road section of NH06, namely BP highway was considered for the purpose of the case study Although initially envisioned as a bypass road, due to shorter travel time, BP highway has been exposed to traffic overload. The road crash data in all four sections of the highway from 2008 to 2016 indicate that there have been 1308 casualties; out of which 241 cases have been fatal. In the study, Section II (Khurkot-Sindhuli) has been considered for model development and a certain portion of Section III (Nepalthok-Khurkot) was used for model validation. These critical sections have not been used in previous studies even though they have multiple crash prone-locations with varying geometric features. The sections have been chosen because they possess a combination of horizontal curves and straight segments which is expected to aid in a more comprehensive analysis. The site map of the sections is shown in Figure 2.

Fig. 2. Site Map on Google Earth

B. Collection of crash and highway geometry data

The crash data was collected from Department of Roads, Dhulikhel-Sindhuli-Bardibas Road Project Office and Area Police Office, Khurkot. The data in which the exact location of the crash site was not included was confirmed with the use of the accident form and public enquiry. Four of the crash locations of 2014 was not included in the analysis as the locations could not be confirmed as they were from 2014 and the crashes were 'Damage Only'. The final sorted crash data was then plotted in Google Earth. The highway geometry data of each section was obtained from the as-built drawings of the sections. The sight distance data was obtained by taping and access density was obtained from site. The sections were sub-divided into fixed segments of 700 m.

C. Model formulation and related procedures

The explanatory variables used in the model development taken from literature review are as follows:

- Minimum Radius of Curvature is the radius of the sharpest curve in the segment.
- Curve Density is the number of horizontal curves per kilometer. The value is found out by counting the total number of curves in the segment.
- Total Length of Curves in a section is calculated from the as-built drawing. The curves with degree of curvature greater than 3.5 degrees is excluded from the analysis as they have been found to behaves as a straight segment (Khan et al., 2012).
- Access Density is the number of access points per kilometer.

- Minimum Sight Distance is the minimum value out of the sight distances of the horizontal curves in a particular segment. It was measured in site with measuring tape.
- Maximum Grade is the maximum value of vertical grade within a segment. It is obtained from the as-built drawings.
- Presence of hairpin bends is used as a categorical variable to account for the risk of crash occurrence on sharp turns.

Although Average Annual Daily Traffic (AADT) is also considered an essential metric in crash prediction, but due to the poor quality of data available, the segmentwise AADT data could not be obtained. So, AADT is excluded from the analysis.

The dependent variables for the purpose of the study are taken as number of crashes and number of fatal and severe crashes. GLM based Poisson and Negative Binomial Regression Models were used as predictive models. The goodness of fit of data was checked using Log likelihood, Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). Comparisons were made between the models and better fitting model (Poisson Regression in our case) was selected.

4. Model development

As shown in Table 1, seven predictor variables were used in the model development out of which six are continuous while one (presence of hairpin bends) is categorical variable.

Out of 56 segments, 27 had hairpin bends present whereas in 26 of the segments, hairpin bends were absent. (Table 2)

Table 3 indicates that the average crashes considering all of the sections is 1.73. Similarly, the descriptive statistics about the predictor variables are also tabulated.

Out of the two models considered, Poisson Regression Model had better goodness of fit based on values of Log Likelihood (AIC, AICC, BIC, CIA (smaller better) as shown in Table 4.

Table 5 shows the parameter estimation of the selected model based on maximum likelihood method.

5. Data Analysis and Results

Poisson Distribution was chosen for model development based on all of the goodness of fit indicating parameters. That would imply that the data is not over dispersed enough for negative binomial distribution to be a better option. Out of the explanatory variables, access density, minimum radius of curvature and presence of hair pin bends fell within the 95%

Table 1
Variable and coding information

Variable	Coding	Variable Type
Access points per km	Access_points km	Continuous
Minimum Radius of Curvature	Min_Radius_of_Curvature	Continuous
Total length of Horizontal Curve	Length_Horizontal_Curve	Continuous
Curve Density (Number of Curves per km)	Curve Density Curve km	Continuous
Minimum horizontal sight distance	Min_Horizontal_Sight_Distance	Continuous
Max_Gradient	Maximum Gradient within a segment	Continuous
Presence of Hairpin Bends	1 for YES	Categorical
	0 for NO	

Table 2
Categorical Data Information

			N	Percent
Factor	Presence of Hair pin Bends	0 (NO)	29	51.8%
		1 (YES)	27	48.2%
		Total	56	100.0%

Table 3
Continuous Data Information

		N	Min.	Max.	Mean	Std. Dev.
Response Variable	Crash_No	56	0.0	7.0	1.73	1.89
Covariate	Access_points km	56	0.0	12.9	3.24	3.54
	Min_ Radius_of_Curvature	56	13	150	25.38	22.95
	Length_Horizontal_Curve	56	36	570	388.59	118.96
	Curve Density (Curve/ km)	56	1.4	35.7	17.57	7.62
	Max_Gradient	56	3%	10%	8%	1.83%
	Min_Horizontal_Sight_Distance	56	15	145	31.88	20.84

Table 4
Goodness of fit: Poisson VS Negative Binomial (Poisson Selected)

	Poisson Regression	Negative Binomial Regression
Log Likelihood	**-90.039**	-93.471
AIC	**196.078**	202.943
AICC	**199.142**	206.007
BIC	**212.281**	219.146
CAIC	**220.281**	227.146

Table 5
Parameter Estimation: Poisson Regression

Parameter	B	Std. Error	Hypothesis Test			Exp (B)
			Wald Chi-Square	df	Sig.	
(Intercept)	**1.302**	.7529	3.678	1	**.084**	3.68
[Presence of Hairpin Bends=0]	**-.525**	.2625	.591	1	**.045**	.591
[Presence of Hairpin Bends=1]	**0[a]**		1			1
Access_points km	**.099**	.0295	1.104	1	**.001**	1.10
Min_Radius_of_Curvature	**.010**	.0046	1.010	1	**.039**	1.01
Length_Horizontal_Curve	**-.001**	.0016	.999	1	**.608**	.999
CurveDensityCurvekm	**-.023**	.0277	.978	1	**.412**	.978
Max_Gradient	**-.035**	.0623	.966	1	**.576**	.966
Min_Horizontal_Sight_Distance	**-.007**	.0050	.993	1	**.173**	.993

confidence interval becoming statistically significant predictors. For every unit increase in access density, the number of crashes in a particular segment increases by 10.4% whereas the absence of hair-pin bends in a certain segment decrease the number of crashes in the segment by 40.9%. The final model obtained is:

Total Crashes/year =

0.2* EXP [(1.302 - 0.525* (HairPin_Absent) + 0.099 * (Access_pointskm) +0.01 * (Min_Radius_of_Curvature)]

Explanation:

a) Here, if the right hand side of the equation is not multiplied by 0.2, the value gives total predicted crashes for 5 years.
b) The value for HairPin_Absent is 1 if hairpin bends are absent in the segment and 0 if at-least one hairpin bend is present.

The model was validated using independent data set from Section III. The value of coefficient of determination (R-Squared) obtained after cross-validation was 0.615 i.e. the model is able to explain 61.5% of variability of the dependent variable around its mean.

6. Conclusion and Recommendations

Based on the results obtained, the access density is the most significant variable in crash frequency determination. The problem of unmanaged and haphazard access road opening around the highway has been a growing phenomenon in the last few years which should be controlled with the help of local authorities. Along with proper regulation, hill roads have to be designed predicting the fact that a number of access roads may prop up after the construction which may be difficult to manage after project completion.

As suggested by the model, hairpin bends also play a significant role in the occurrence of crashes. Hairpin bends have to be designed with proper consideration to the possibility of crash hazards. Maneuverability in sharp curves can be made safer by making the vertical grades milder and efficient use of traffic signs.

Nepal still lacks a proper accident database management system. The exact locations of crash points are very tough to find which makes prediction modelling a difficult task. Government funding has to be increased on providing traffic officials will all the essential equipment and trainings that they require for accurate record-keeping.

Predictive analysis of road crashes in developing countries like Nepal has a good scope and potential given the lack of readily applicable international models that suit indigenous local conditions. More predictive models need to be formulated using other sections of the road network to check the consistency of the relationships and to introduce other combinations of predictor variables which can be beneficial in road safety decision-making.

References

[1] Basu, S. and Saha, P., 2017, "Regression Models of Highway Traffic Crashes: A Review of Recent Research and Future Research Needs", *Procedia Engineering 187*, pp.59 – 66.
[2] Caliendo, C1., Guida, M., and Parisi A., 2007, "A crash-prediction model for multilane roads", *Accident Analysis & Prevention* Volume 39, Issue 4, July 2007, pp. 657-670
[3] Greibe, Paul. ,2003, "Accident prediction models for urban roads", *Accident Analysis & Prevention*, Volume 35, Issue 2, pp. 273-285
[4] Abdulhafedh, A., 2016, "Crash Frequency Analysis", *Journal of Transportation Technologies*, 6, pp 169-180.
[5] Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick, 2005, "Roadway Safety Design Synthesis", Report No. *FHWA/TX-05/04703*—1, Texas Department of Transportation.
[6] Fitzpatrick, K., Lord, D., and Park, B, 2009, "Evaluating Safety Effects of Ramp Density and Horizontal Curve for Freeways Using Texas Data".
[7] Zegeer, C., Stewart, R., 1992, "Safety Effects of geometric improvements on horizontal curves", *Transportation Research Board. No. 1356*, Washington D.C, pp.11- 19.
[8] Geedipally, S.R., Lord, D., and Dhavala, S.S., 2012, "The Negative-Binomial Lindley Generalized Linear Model: Characteristics and Application Using Crash Data", *Accident Analysis and Prevention*, pp 45, 258-265.
[9] "Highway Safety Manual", 2010, American Association of State Highway and Transportation Officials, Washington D.C.
[10] Koorey, G, 2009, "Road Data Aggregation and Sectioning Considerations for Crash Analysis" *Transportation Research Record*, 2103(1), 61–68.
[11] Cafiso, S., D'Agostino, C., Persaud, B., 2018, "Investigating the influence of segmentation in estimating safety performance functions for roadway sections", *Journal of Traffic and Transportation Engineering*.
[12] Green, Eric R., 2018, "Segmentation strategies for road safety analysis", *Ph.D. Dissertation*, Department of Civil Engineering, University of Kentucky.