Planning Level Regression Models for Prediction of the Number of Crashes on Urban Arterials in Bangladesh

Md. Hadiuzzaman[1], Md. Ahsanul Karim[2], Md. Mizanur Rahman[3], Tanweer Hasan[4]

Received: 09.03.2014 *Accepted: 03.05.2016*

Abstract:

In most of the developing countries, the metropolitan organizations do not assess the safety consequences of alternative transportation systems and one of the reasons is the lack of suitable methodology. The goal of this paper is to develop practical tools for assessing safety consequences of arterial roads in the context of long-term urban transportation plans in Dhaka city, the capital of Bangladesh. The researchers used the generalized linear modeling approach to develop separate models to predict number of crashes for different levels of crash severity for major arterial segments which have the highest crash rates. The models used five independent variables - length of segment, traffic volume, number of access road, design speed and roadway width - all of which are usually collected or predicted by transportation planners. The study reveals that roadway width and design speed are the governing factors for non-lane based traffic on urban arterials for controlling crashes. The crash prediction models presented in this paper can enable Dhaka's transport planners to evaluate the safety impact of alternative road networks with regard to the costs and benefits in long-term planning context.

Keywords: Crash prediction, regression models, negative-binomial distribution, poisson distribution, fatal accident.

Corresponding author e-mail: mhadiuzzaman@ce.buet.ac.bd

1. Assistant Professor, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Bangladesh
2. Travel Demand Forecasting Engineer, Alberta Transportation, Government of Alberta, Edmonton, Canada
3 & 4. Professor, Department of Civil Engineering, Bangladesh University of Engineering and Technology, Bangladesh

1. Introduction

Arterials are the main life-lines of any modern city, and play an important role in a city's growth by facilitating movement of enormous volume of traffic including both passenger and pedestrian. Many research studies in the past have examined the capacity and operation of freeways, investigated the effects of geometric features, drivers, and environmental factors on the number and severity of crashes on freeways [Fazio, Holden and Rouphail, 1993; Kraus et. al, 1993; Khan, Shanmugam and Hoeschen, 1999]. Examples of independent variables used in crash prediction models developed in the aforementioned studies included median width, pavement surface rating, and weather condition. These models can be used for corridor planning and at the project development stage when detailed characteristics of road segments are known. However, at the long range planning stage when decisions are made regarding general characteristics of urban roadways, and also when tradeoffs involving freeways and non-freeway arterials and other modes of transportation are examined, the available crash prediction models are not helpful because they demand data on detailed characteristics of these roadways for which planners do not make forecasts. Consequently, not much effort is made by planners to assess the safety implications of alternative transportation networks, and this is a serious weakness of the long-range planning process. For this purpose, there is a need to understand and recognize the requirements and limitations of the long-range planning process with regard to data availability, and develop tools and criteria that can actually be used by planners for assessing the safety impacts of long-range transportation alternatives. The goal of this paper is to make some contributions toward meeting this need. It should be pointed out that the application scenario for the models presented in this paper is the plan development stage of the long-range transportation planning process when alternative transportation and land use plans are evaluated.

2. Research Objectives

The objective of this research was to develop predictive models for crashes occurring on urban arterials. Separate models are developed for crashes of different severity levels—"Fatal," and "Non-fatal". These crash prediction models are to be used for assessing the safety related consequences associated with alternative types of arterials included in long-term transportation plans. The independent variables used in the models are limited to those which are commonly forecasted by planners for future scenarios. The models presented in this paper can be implemented in the safety conscious planning of alternative roadways for urban areas.

3. Crash Prediction Model Approaches

A crash prediction model can have different alternative forms – simple or complex. It can be in the form of a simple cross-classification table (matrix) where the cell values are crash/accident rates and the classification is based on different factors, or variables, such as the type of roadway, traffic volume, roadway geometry, driveway density, etc. These matrices can vary in complexity depending on the number of factors used for cross-classification.

Another approach is to derive mathematical equations by correlating crash/accident rates with different explanatory/independent variables using a statistical technique such as multiple regression analysis. Whatever the form of the equation may be, these models are developed based on historical crash data. Crash/accident rates are based on the number of crashes and an appropriate exposure measure. Separate models usually are developed for road segments and intersections, which then can be separated by different classes or types. It should be added that the mathematical equations may use "number of crashes" as the dependent variable instead of 'crash rates'; and in that case 'vehicle-miles traveled' (VMT) is used as an independent variable. Many analysts prefer the latter approach [El-Basyouny and Sayed, 2009; Ladron de Guevara , Washington and Oh, 2004; Lord and Persaud, 2004].

In this paper, the researchers applied statistical technique for developing crash prediction models. Though simple cross-classification approach avoids mathematical formulation of an equation, this approach can be somewhat crude because the values of the variables used for cross-classifying the crash rates must be grouped into categories using specific ranges of values. Further, the selection of variables to use for cross-classification usually is done intuitively, and the strength of correlation of individual independent variables with the crash rates, the response/dependent variable, usually is not determined. On the contrary, statistical procedures such as an analysis of log-likelihood values can be performed to determine if the relationship among those two types of variables is significant or not [Turner and Roozenberg, 2005]. A prediction model in the form of an equation requires a rigorous statistical analysis for determining the appropriate mathematical formulation. The underlying statistical distributions may be simple or fairly complex, such as Normal, Poisson, or Negative Binomial.

4. Review of Crash Prediction Models

Most of the previous attempts to develop planning level crash prediction models were focused on freeway or highway facilities. In most cases freeway crash prediction models were developed for evaluating the effectiveness of safety countermeasures. The literature review found several studies that developed freeway crash prediction models for long-range planning applications [Persaud and Dzbik, 1993; Chatterjee et. al, 2003; Kononov and Allery, 2004] while the majority developed models for the purpose of either identifying highly hazardous locations, traditionally called black spots, or performing before and- after analysis [Fazio et. al, 1993; Kraus et. al, 1993; Shankar et. al, 1996; Resende and Benekohal, 1997; Khan et. al, 1999; Konduri and Sinha, 2002]. These applications considered in most freeway crash studies were suitable for corridor and short-range planning involving specific road segments. Though few crash studies were devoted to urban arterials [El-Basyouny and Sayed, 2009; Ladron de Guevara et al., 2004; Lord and Persaud, 2004], their limited extent and applicability for developing countries requires for more research in this field for sustainable transport.

The most common model form used in the aforementioned studies is:

$$E(Y) = b_0 * VKT^{b_1} * exp \sum_i b_i x_i \quad (1)$$

where E (Y) is the predicted collision frequency (over a period); b_0 is intercept, b_1 and bi are model parameters; VKT is the external exposure variable, and X_i represents the explanatory variable.

5. Data Collection

The model consists of several independent or explanatory variables, encompassing elements from road geometry to traffic condition. For this study, the variables which have been considered are arterial segment length, 85thpercentile speed, traffic volume, roadway width and number of access points per kilometer

5.1 Crash data

Data on traffic crashes in Dhaka city were collected from the First Information Reports (FIR) of police stations in the city for the 8 years (1998 to 2005). The only official source of crash data in Dhaka is the traffic division of the Dhaka Metropolitan Police (DMP). In Dhaka very few crashes between non-motorized vehicles are reported to police. This is because damages from these slow-moving vehicles are usually minor and compensation costs are settled immediately.

5.2 Link Length

In analyzing crashes, the length of link includes the area of influence around a hazard. The link (or mid-block) as defined in this paper includes the roadway segment between two intersections. All crashes that occur within a specified radius of 250 feet from the center of an intersection has been considered as intersection crashes and thus excluded from the present research which is focused mainly on prediction of link crash.

5.3 Traffic Volume

Traffic volume data were collected at each link on weekdays for two hours with the help of a video camera. Time of data collection was so selected that the morning or afternoon peak was included within the study hour. The hourly traffic volume and the composition of various types of vehicles at the study sections were obtained from the videos of the traffic stream.

5.4 Design Speed

Speed data were collected both manually and automatically (using radar-gun). The 85th percentile speeds were determined from spot speed measurement using SPSS program.

5.5 Number of Access Roads

Number of access roads on each link was observed by traversing the entire length of the road segment and the number of major access point per kilometer for every section is obtained.

5.6 Roadway Width

In case of developing countries like Bangladesh, vehicles usually do not follow lane discipline. So instead of considering lane based data total roadway width has been considered as one of the model parameters which seem logical for the current situation.

Table 1 presents the descriptive statistics of the data/variables.

6. Model Form

The models used in the studies are called generalized linear models and typically have a negative binomial or Poisson error structure. Generalized linear models were first introduced to road crash studies in [Maycock and Hall, 1984] and extensively developed in [Hauer , Ng and Lovell, 1989]. The aim of the modeling exercise is to develop relationships between the crash rate (as the dependent variable), and predictor variables. Typically the models have the following form:

$$A = b_0\, x_1^{b1}\, x_2^{b2}\, x_3^{b3}\, x_4^{b4}\, x_5^{b5} \qquad (1)$$

Where A is the annual mean number of crash rate, x_1, x_2, x_3, x_4 and x_5 are explanatory variables and b_1, b_2, b_3, b_4 and b_5 are the coefficients. In the modeling process, a log-linear transformation is made (refer to Equation 2). This is the reason the models are called generalized linear models even though the final model form is multiplicative.

$$\log A = \log (b_0\, x_1^{b1}\, x_2^{b2}\, x_3^{b3}\, x_4^{b4}\, x_5^{b5}) \qquad (2)$$
$$= \log b_0 + b_1 \log x_1 + b_2 \log x_2 + b_3 \log x_3 + b_4 \log x_4 + b_5 \log x_5$$

The models also may have the following form:

$$A = b_0\, x_1^{b1}\, x_2^{b2}\, x_3^{b3}\, x_4^{b4}\, x_5^{b5}\, e^{b6} \qquad (3)$$

In the modeling process, the log-linear transform is:

$$\log A = \log (b_0\, x_1^{b1}\, x_2^{b2}\, x_3^{b3}\, x_4^{b4}\, x_5^{b5}\, e^{b6}) \qquad (4)$$
$$= \log b_0 + b_1 \log x_1 + b_2 \log x_2 + b_3 \log x_3 + b_4 \log x_4 + b_5 \log x_5 + b_6$$

The link function of the previous equations may be identity or log.

7. Model Development

For the development of models, the 'Crash Frequency Method' has been used in determining the crash rates. The crash data has been normalized to convert the crash frequencies to 'Crash Rate' in terms of crash per mile per year. In this study, crash data for eight years on five major arterials and a sum total of eighteen roadway segments or links of Dhaka metropolitan were considered of which sixty percent data have been used in the model development stage and remaining data for model validation. Figure 1 shows the selected five arterials of Dhaka city having the highest crash record.

Table 1. Descriptive statistics of the data/variables

Statistics	Average	Minimum	Maximum	Standard Deviation
Number of access roads	8.4	2.0	24.0	7.0
Link length (Miles)	0.8	0.2	2.0	0.6
Roadway Width (Meter)	27.7	9.7	40.0	7.0
Traffic Volume (PCU/hr)	3238.5	1120.0	5641.0	1234.3
Design Speed (Mile/hr)	42.2	35.0	45.3	3.0
Accident Rate (Accident Per Mile Per Year)	7.9	2.1	18.3	4.6
Fatal Accident Rate (Fatal Accidents per Mile Per Year)	3.1	0.6	7.8	1.9
Non-Fatal Accident Rate (Non-Fatal Accidents per Mile Per Year)	4.8	1.2	12.7	3.5

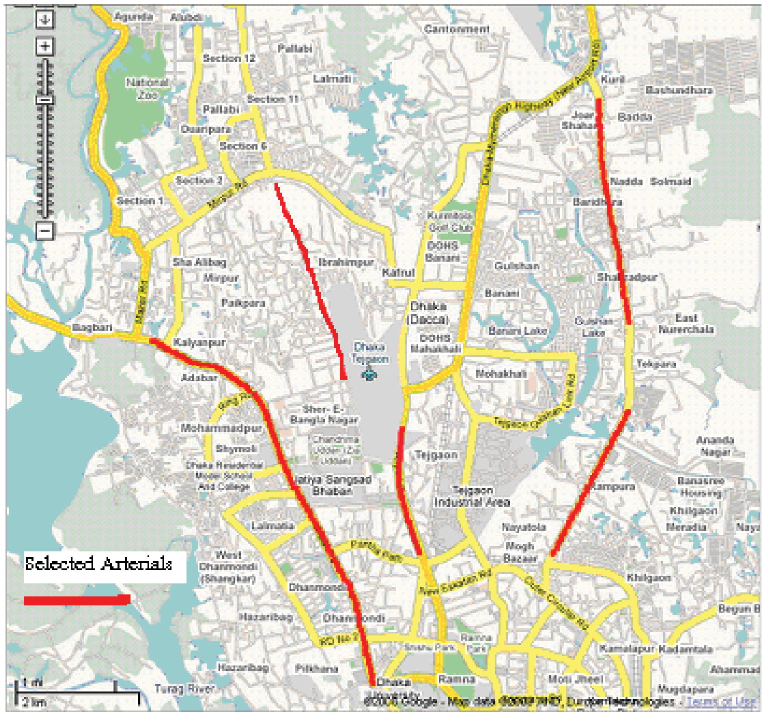

Figure 1. Selected arterial segments

Three different types of crash models (i.e. total, fatal and non-fatal) have been developed using SPSS and GENMOD of SAS for different link function and regression approach based on crash data distribution. Most of the previous study reveals that crash data typically follows either negative-binomial or Poisson distribution. The 'Poisson' model was used where the variance in crash numbers is roughly equal to or less than the mean over the majority of the traffic flow range. However, when the variability is generally higher than the mean the 'negative binomial' model is used. The Negative binomial model is a mixture of the Poisson and gamma distributions. In this study, it was observed that though for metropolitan Dhaka total and non-fatal crash follow negative-binomial distribution, but fatal crash follows normal distribution. Probably this is due to the fact that in the study area, pedestrians represent around 72% of road traffic fatalities and the pedestrian's exposure is more or less same for each roadway segment [ARC, 2005]. As a result, the fatal crash rate has been found nearly same and follows normal distribution instead of negative binomial or Poisson distribution. Using statistical software SPSS and SAS, a total of fourteen models were developed. Models which have significant statistical parameters have been listed in Table 2. For the models developed in SAS, only the models that have significant parameters, the ratio of deviance to the degrees of freedom close to one and a relatively high estimated R_{α}^2, are presented in this section. The deviance is a measure of discrepancy between actual and estimated values, which was calculated in the GENMOD procedure. The R_{α}^2 indicator is used to measure the level of explanatory ability of each model [Miaou, 1996].

$$R_{\alpha}^2 = 1 - \frac{K}{K_{\max}} \tag{5}$$

The term k =dispersion parameter estimated in the negative binomial model while k_{max}=dispersion parameter estimated in the same negative binomial model with only an intercept term and a dispersion parameter. This measure can be used equivalently to the coefficient of determination (R^2).

8. Model Validation

In the model validation phase, forty percent of the crash data were used. Both coefficient of determination (R2) and coefficient of variation of root mean square er-

Table 2. Regression statistics and models for different crash severity

Crash Severity	Model No.	Model
Total Crashes	1.	$N = 0.0007458 * (Access)^{-0.173} * (Length)^{-0.234} * (Width)^{2.144} * (Volume)^{0.167} * (Speed)^{0.435}$ $R_\alpha^2 = 0.652$
	2.	$N = \exp\{0.00043 * (Access)^{-0.138} * (Length)^{-0.205} * (Width)^{1.481} * (Volume)^{0.161} * (Speed)^{0.767}\}$ $R_\alpha^2 = 0.633$
	3.	$N = 0.0007458 * (Access)^{-0.173} * (Length)^{-0.234} * (Width)^{2.144} * (Volume)^{0.167} * (Speed)^{0.435}$ $R_\alpha^2 = 0.838$
	4.	$N = 6.714x10\text{-}9 * (Access)^{-1.170} * (Length)^{-0.607} * (Width)^{1.998} * (Volume)^{0.613} * (Speed)^{2.757} * \exp(0.2 * link)$ $R_\alpha^2 = 0.919$
Fatal Crash	1.	$N = 3.11x10^{-17} * (Access)^{-0.206} * (Length)^{0.019} * (Width)^{3.488} * (Volume)^{0.334} * (Speed)^{7.094}$ $R_\alpha^2 = 0.594$
	2.	$N = \exp\{1.75x10^{-39} * (Access)^{-0.831} * (Length)^{1.020} * (Width)^{9.898} * (Volume)^{-0.469} * (Speed)^{17.704}\}$ $R_\alpha^2 = 0.625$
	3.	$N = 3.11x10^{-17} * (Access)^{-0.206} * (Length)^{0.019} * (Width)^{3.488} * (Volume)^{0.334} * (Speed)^{7.094}$ $R_\alpha^2 = 0.878$
	4.	$N = 3.261x10\text{-}8 * (Access)^{-0.149} * (Length)^{0.050} * (Width)^{-8.210} * (Volume)^{0.554} * (Speed)^{6.404} * \exp(0.765 * Width)$ $R_\alpha^2 = 0.951$
	5.	$N = 1.31x10^{-15} * (Access)^{-0.131} * (Length)^{0.059} * (Width)^{4.323} * (Volume)^{-0.069} * (Speed)^{6.339}$ $R_\alpha^2 = 0.921$
	6.	$N = 9.89x10^{-8} * (Access)^{-0.046} * (Length)^{-0.040} * (Width)^{-7.910} * (Volume)^{0.437} * (Speed)^{6.209} * \exp(0.740 * Width)$ $R_\alpha^2 = 0.961$
Non-fatal Crash	1.	$N = 697.15 * (Access)^{-0.369} * (Length)^{-0.125} * (Width)^{1.559} * (Volume)^{-0.322} * (Speed)^{-1.815}$ $R_\alpha^2 = 0.672$
	2.	$N = \exp\{3.568x10^{-12} * (Access)^{-0.663} * (Length)^{-1.085} * (Width)^{2.045} * (Volume)^{0.284} * (Speed)^{4.989}\}$ $R_\alpha^2 = 0.551$
	3.	$N = 697.15 * (Access)^{-0.369} * (Length)^{-0.125} * (Width)^{1.559} * (Volume)^{-0.322} * (Speed)^{-1.815}$ $R_\alpha^2 = 0.814$
	4.	$N = 1.169x10\text{-}5 * (Access)^{-1.906} * (Length)^{-0.699} * (Width)^{1.334} * (Volume)^{0.365} * (Speed)^{1.762} * \exp(0.308 * link)$ $R_\alpha^2 = 0.922$

ror (CV_RMSE) values between predicted and actual count of crashes have been calculated to identify the best model for each crash type. Table 3 shows the measured statistical parameter. CV_RMSE has been calculated using the following equation (6).

From the table 3, it is clear that Model 2 for total crash, Model 6 for fatal crash and Model2 for non-fatal crash prediction have high coefficient of determination and lowest root mean square error. So, these three models have been finally proposed for fatal, non-fatal and total crash prediction on urban arterial of developing countries. These three models have been enlisted below with the corresponding R^2 values and P-values for the explanatory variables in order of Access, Length, Width, Volume, and Speed:

$$CV\_RMSE = \frac{\sqrt{\sum_j \left(Model_j - Count_j\right)^2 / \left(Number\_of\_counts - 1\right)}}{\left(\sum_j Count_j / Number\_of\_Counts\right)} \times 100 \qquad (6)$$

Table 3. RMSE and R^2 Values between Observed and Predicted Crash

Crash Severity	Model no.	R^2	CV_RMSE
Total Crash	Model 1	0.967	12.474
	Model 2	0.969	11.124
	Model 3	0.967	12.474
	Model 4	0.915	17.464
Fatal Crash	Model 1	0.841	32.302
	Model 2	0.961	14.858
	Model 3	0.841	32.302
	Model 4	0.977	13.193
	Model 5	0.915	24.278
	Model 6	0.971	12.712
Non-fatal Crash	Model 1	0.923	18.641
	Model 2	0.980	9.415
	Model 3	0.923	18.642
	Model 4	0.899	21.623

Total Crash:

$N = \exp\{0.00043 * (Access)^{-0.138} * (Length)^{-0.205} * (Width)^{1.481} * (Volume)^{0.161} * (Speed)^{0.767}\}$ (1)

(R_α^2=0.633; P-values in order: <0.001, <0.001, <0.001, <0.001, <0.001)

Fatal Crash:

$N = 9.89 \text{ x } 10\text{-}8 * (Access)^{-0.046} * (Length)^{-0.040} * (Width)^{-7.910} * (Volume)^{0.437} * (Speed)^{6.209} * e^{(0.740*Width)}$ (2)

(R_α^2=0.961; P-values in order: <0.001, <0.001, <0.001, >0.001, <0.001)

Non-fatal Crash:

$N = \exp\{3.568 \text{x} 10^{-12} * (Access)^{-0.663} * (Length)^{-1.085} * (Width)^{2.045} * (Volume)^{0.284} * (Speed)^{4.989}\}$ (3)

(R_α^2=0.551; P-values in order: <0.001, <0.001, <0.001, <0.001, <0.001)

where,

N = crash per mile per year

Access = number of access points on the arterial segment

Length = length of arterial segment or link (mile)

Width = Width of roadway (meter)

Volume = Traffic Volume (Passenger Car Unit/hour)

Speed = Design Speed of arterial segment (mile/hour)

9. Conclusion

The primary objective of this research was to develop planning level crash prediction models for assessing the long-range safety impact of alternative arterial networks of urban areas. A set of regression models was developed following the generalized linear modelling technique. The authors used five independent variables - length of segment, traffic volume, number of access road, design speed and roadway width. It is expected that all of these variables are usually known by the transportation planners as these are regularly collected or predicted. The developed models are not meant to be used for analyzing detailed design features of urban road during project development stage, which usually is the responsibility of country's transportation and planning ministry. Though all the previous models for freeway show crash rates largely influenced by interchanges, the present study reveals that roadway width and design speed are the governing factors for non-lane based traffic on urban arterials for controlling crashes.

The developed models can actually be used by the metropolitan planners for assessing the safety impacts of alternative arterials. However, it should be noted that these models are most suitable in the plan development stage of the long-range transportation planning process when alternative transportation and land use plans are evaluated.

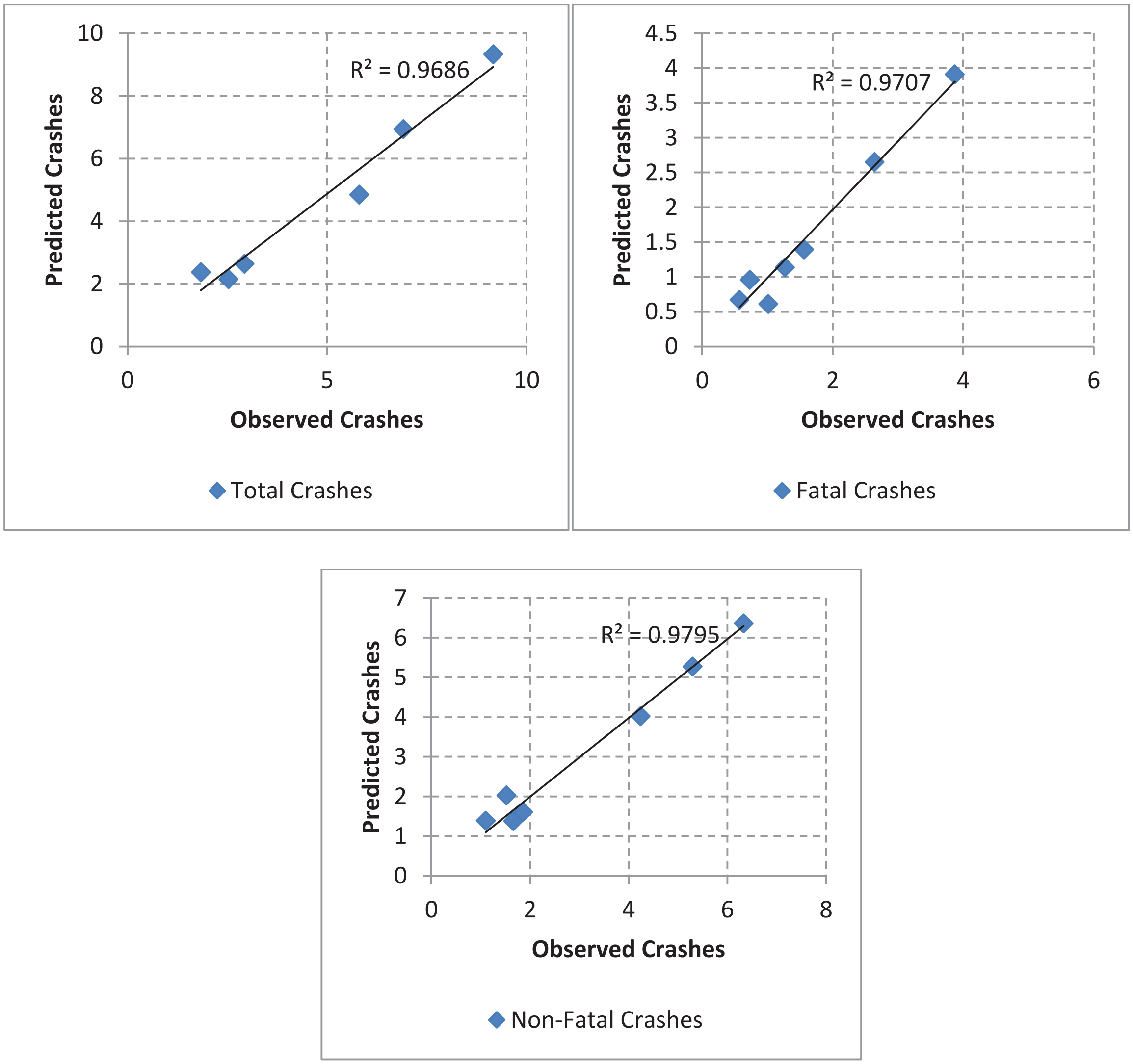

Figure 2. Predicted values against the observed values for the proposed models

10. References

- Accident Research Center (2005) "Key road safety facts in Bangladesh, Preliminary Brief Report", Published by Accident Research Center, Bangladesh University of Engineering and Technology.

- Chatterjee, A., Everett, J. D., Reiff, B., Schwetz, T. B., Seaver, W. L. and Wegmann, F. J. (2003) "Tools for assessing safety impact of long-range transportation plans in urban areas", Center for Transportation Research, Univ. of Tennessee, Knoxville, Rep. Prepared for Federal Highway Administration and available through Travel Model Improvement Program Clearing House, Tenn..

- Fazio, J., Holden, J. and Rouphail, N. M. (1993) "Use of freeway conflict rates as an alternative to crash rates in weaving section safety analyses", Transportation Research Record, Transportation Research Board, Washington, D.C., Vol. 1401, pp. 61–69.

- Hauer, E., NG, J. C. N. and Lovell, J. (1989) "A study of intersection accident exposure", Proc. 9thAustralian Road Research Board Conference, Vol. 9 (5), pp. 151-160.

- El-Basyouny, K. and Sayed, T. (2009). "Urban arterial accident prediction models with spatial effects". Transportation Research Record: Journal of the Transportation Research Board, Vol. 2102, pp. 27-33.

- Khan,S.,Shanmugam, R. and Hoeschen, B. (1999)

"Injury, fatal, andproperty damage accident models for highway corridors", TransportationResearch Record, Transportation Research Board, Washington,D.C., Vol. 1665, pp. 84–92.

- Kononov, J. and Allery,B. K. (2004) "Explicit consideration of safetyin transportation planning and project scoping", Proc., 83rd AnnualMeeting of the Transportation Research Board, Washington, D.C..

- Konduri, S. and Sinha, K.C. (2002) "Statistical models for predictionof freeway incidents", Proc., 7th Int. Conf. on Applications of AdvancesTechnology in Transportation Engineering, Cambridge, Mass., pp. 167–174.

- Kraus,J. F., Anderson,C. L., Arzemanian, S., Salatka, M., Hemyari, P. and Sun, G. (1993) "Epidemiological aspects of fatal and severe injury urban freeway crashes", Accident Analysis and Prevention, Vol. 25(3), pp. 229–239.

- Ladrón de Guevara, F., Washington, S. and Oh, J. (2004) "Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona". Transportation Research Record: Journal of the Transportation Research Board, Vol. 1897, pp. 191-199.

- Lord, D., and Persaud, B. N. (2004). Estimating the safety performance of urban road transportation networks. Accident Analysis & Prevention, Vol. 36 (4), pp. 609-620.

- Maycock, G. and Hall, R.D. (1984) "Accidents at four-arm roundabouts", Transport and Road Research Laboratory Report LR1120.

- Miaou, S.P. (1996) "Measuring the goodness-of-fit of accident prediction models", Publication No. FHWA-RD-96–040, FHWA, U.S. Department of Transportation, Washington, D.C..

-Persaud, B.N.and Dzbik,L. (1993)"Accident prediction models forfreeways", Transportation Research Record, Transportation Research Board, Washington, D.C., Vol. 1401, pp. 55–60.

- Resende, P. T. V. and Benekohal, R. F. (1997) "Development of volume to-capacity based accident prediction models", Proc., Traffic Congestion and Traffic Safety in the 21stCentury, Chicago, pp. 215–221.

- Shankar, V., Mannering, F. and Barfield, W. (1996) "Statistical analysis of accident severity on rural freeways", Accident Analysis and Prevention, Vol. 28 (3), pp. 391–401.

- Turner, S. and Roozenburg, A. (2005) "Accident prediction models for signalized intersection", IPENZ-TG Technical Conferences, Hyatt Regency, Auckland, September.