

The relationship between urban accidents, traffic and geometric design in Tehran

S. Aftabi Hossein¹ & M. Arabani²

¹*Bandar Anzali Branch, Islamic Azad University, Iran*

²*Department of Technical and Engineering, Guilan University, Iran*

Abstract

Considerable research has been carried out in recent years about crash modelling and to establish relationships between crash occurrences in different types of urban and rural roads and traffic flow parameters, geometric infrastructure characteristics, environmental factors and pavement qualifications.

In this research crash-prediction models for 3 urban arterial highways in Tehran were set up on the basis of accident data observed during a 3-year monitoring period extending between 2006 and 2008. The candidate set of explanatory parameters were: traffic flow (pick hour volume), geometric infrastructure characteristics (section's length, curvature, longitudinal slope, number of lanes) and pavement condition of surface wetness. Statistical analysis is done by SPSS on the basis of nonlinear regression modelling and during the analysis, principal components are identified to assist the principal component analysis method and more important variables recognized that could exhibit best description of crash occurrence. After studying variables for curves, it is shown that significant and efficient variables are section's length (L), pick hour volume (PHV) and longitudinal slope (L.s), whereas for tangent they are section's length (L), pick hour volume (PHV) and longitudinal slope (L.s) and curvature (1/R).

Results indicate that the number of accidents increase with: L, PHV and L.s whereas they decrease with R. In addition, L.s is extremely effective in this research. The relation between speed variations with L.s and consequently traffic congestion, safety reduction and chance of accident occurrence are its efficiency causes. Furthermore, L.s affects driving site distance and driver behaviour such as passing other vehicles too. In addition, the effects of longitudinal surface friction and pavement are increased in plunge slope and they increase the chance of occurrence.

Keywords: crash modelling, traffic flow, geometric design, urban highways.



1 Introduction

In recent years, a lot of studies have been undertaken about accidents, condition of their occurrence and also prediction of occurred crashes counts. The effects of different effective variables on crashes attended in these studies for prevention or to reduce the crashes counts.

The first model of crash occurrence's causes is presented on the basis of Domino theory in 1941. This study peruses every linear set of occurrences leading to crash. Also, the first updated Domino theory is presented in 1974 that represented its supplementary results in 1976. The first prediction's model of multilane roads was presented in 1993. The results of these studies by using generalized linear models show the rates of crash are increased by growth of traffic flow that is introduced by ADT and VH (e.g. Persaud and Dzbik [1]).

The effect of median width on crash rate using a Negative Binomial distribution was surveyed in 1993. The findings of Log-Linear regression method indicated that crash rate decrease with increasing median width (e.g. Knuiman *et al.* [2]).

It was illustrated that there was a strong relation between crash occurrence, rain and geometric characteristics of urban freeway sections in 1995. In this study, negative binomial regression and Poisson models are applied (e.g. Shankar *et al.* [3]) several accident-prediction models with regard to both multilane roads and two-lane roads of rural or urban designation were proposed in 1995. By examining the effect of traffic flow on the crash rate the conclusions reached were that the crash rate increases with increasing AADT on high levels of traffic, while it decreases with AADT on lower traffic volumes (e.g. Hadi *et al.* [4]).

Persaud presented one of the earliest studies for carrying out separate analyses for curves and tangents, albeit limited to two-lane roads in 2000. The dependent variable was crash frequency, while the independent variables were traffic flow and road geometry. For curves, crash frequency was found to increase with: AADT, section length (L) and curvature (1/R). For tangents, the number of accidents per year increases with AADT and L (e.g. Persaud *et al.* [5]).

Martin described the relationship between crash rate and traffic volume per hour (VH) and the influence of traffic on crash severity for French interurban motorways in 2002. A Negative Binomial distribution was used (e.g. Martin [6]).

Golob and Recker used linear and non-linear multivariate statistical analyses to determine how the type of accidents related to traffic flow, weather and lighting conditions in 2003. The study approach was based on the Principal Component Analysis (PCA) in order to identify the most significant variables from a set of original traffic flow variables (e.g. Golob and Recker [7]).

Golob evaluated the safety effects of changes in freeway traffic flow, in a subsequent paper in 2004. In this study, three crash characteristics were used in the analysis, namely crash type (rear end, sideswipe or hit object, number of vehicles involved); crash location (e.g. left lane, interior lanes, right lane, shoulder); crash severity (injuries and fatalities per vehicle) (e.g. Golob *et al.* [8]).

Also, Hauer developed statistical road safety modelling by using the Negative Binomial distribution. The selected variables were geometric characteristics and traffic flow [9]).

Caliendo and Parisi implemented a comprehensive study about PCA method and selecting effective principal variables in crash occurrences in 2005. They advised this method to other researchers because of its simple and uncomplicated trait, different to Fuzzy logic, etc. Results for curves presented that crashes are increased by an increase in the section's length (L), curvature ($1/R$), average daily traffic (ADT) and different in design speed between tangents and curves. Also for tangents, crashes are increased by an increase in the section's length (L) and average daily traffic (ADT) (e.g. Caliendo and Parisi [10]).

Caliendo, Guida and Parisi discussed in a new paper about correlated variables with traffic flow, geometric infrastructure characteristics, surface of pavement and rainy condition in 2006. The Poisson, Negative Binomial and Negative Multinomial regression models, applied separately to tangents and curves were used to model the frequency of accident occurrence. For curves it is shown that significant variables are L , $1/R$ and AADT, whereas for tangents they are L , AADT and junctions. The effect of rain precipitation was analysed on the basis of hourly rainfall data and assumptions about drying time. It is shown that a wet pavement significantly increases the number of crashes (e.g. Caliendo *et al.* [11]).

Tobit regression was applied for recognition of principal parameters that affect accidents in 2008. Results show that factors correlated with qualification of pavement (side friction coefficient, rough index of pavement, rut of pavement, pavement's degree) and geometric infrastructure characteristics (type and width of median, width of shoulder, counts of ramps, bridges, horizontal and vertical curves) affect crashes counts (e.g. Panagiotis *et al.* [12]).

The above studies show that there are several variables that affect accidents and their attention points have concentrated on rural roads and there is little attention to the accidents of urban highways. This paper surveys accidents of urban arterial highways in Tehran. Identifying more important variables that they increase crash counts in urban arterial highways is the object of the paper for the costs of accidents will be decreased after identifying each variable on the basis of available data and the amount of their effects. Surveying the above studies shows that more important parameters of crash prediction models are traffic and geometric design parameters. According to experience of these studies, researchers tried to exhibit simpler models with a higher probability ratio and they applied various linear and nonlinear models to present many logical models. In this paper, nonlinear models are applied for modelling because of their greater flexibility in adapting estimations on collected data. So logistic analysis is made by accident data observed of three urban arterial highways in Tehran during a 3-year monitoring period extending between 2006 and 2008 until more important causes of crash occurrences and their effect amount on crashes will be identified in such highways. The aim is identifying more effective variables on crash occurrence of urban highways.

2 Data description

Crash statistics are collected from three urban arterial highways in Tehran (4 or 5-lane median-divided) that their characteristics are shown in table 1. During the period of observation, crash data, traffic flow and geometric characteristics were collected. Accident data were collected and extracted from the official reports of the Tehran's Driving Police (Rahvar). For each accident a variety of details was recorded, including date location of accident, horizontal alignment (tangent or curve), vertical alignment (upgrade or downgrade), weather and pavement surface conditions (dry or wet), type and severity of accidents.

Table 1: Characteristics of highways under study.

Highway name	Curve sections (C)		Tangent sections (T)		Total length
	Total length	Count	Total length	count	
Chamran	5198	22	6727	19	11925
Navab	1620	9	5367	19	6987
Saeidi	376	3	921	4	1297

Table 2 shows some 12,891 accidents were considered in this study, 32 of which were fatal, 1,348 were injury accidents and 11,601 led to just financial damage. During the period of observation 1,380 severe accidents (fatal and injury crashes) equivalent 10.63% of all crashes and 11,601 non severe accidents equivalent to 89.37% of all crashes were registered. 34.4% of all crashes occurred on curves and 65.6% of all crashes occurred on tangents. So, the chance of crash occurrence on tangents is 1.91 crash occurrence on curves. The database does not include accidents taking place on the ramps of junctions, service areas, shoulders and at tollbooths or since such accidents are not due to traffic flow and infrastructure characteristics. Pedestrians and bicycles are forbidden to use this infrastructure and they were not involved in accidents (e.g. Noland and Oh [13]).

Table 2: Crash count separately on curves and tangents.

Highway name	Year	Occurred crash count						
		North-south roadway			South-north roadway			Total highway crash
		T	C	Total	T	C	Total	
Chamran	2006	941	486	1427	946	505	1451	2878
	2007	1032	553	1585	1061	561	1622	3207
	2008	472	237	709	489	269	758	1467
Navab	2006	471	246	717	469	243	712	1429
	2007	514	262	776	512	264	776	1552
	2008	464	194	658	465	279	744	1402
Saeidi	2006	70	41	111	73	39	112	223
	2007	126	66	192	124	66	190	382
	2008	114	61	175	114	62	176	351
Total		4204	2146	6350	4253	2288	6541	12891

For the purpose of the subsequent statistical analysis, homogeneous road sections were separated. The characteristics of these three highways were collected from the Transportation Department of Tehran's municipality which include longitudinal slope, different segments of curves and tangents, length and width or curvature of each segment, number of lanes, location of junctions and such. Segments for each carriageway have constant horizontal curvature and longitudinal slope. For these segments the following major variables did not change: width and number of lanes, type and width of shoulders and medians.

In this research among the different available parameters, we attended to traffic flow parameters and horizontal and vertical dimension characteristic parameters. For example, many complicated calculations and particular statistics are necessary to survey the effect of drivers caution in crash occurrence that it is not easily included in modelling. So a variable such as present junctions count in the path are not included because low junctions count in the three involved highways. The wet effect variable is not included in modelling because it is not effective on increasing and decreasing crash counts.

2.1 Traffic flow

The initial studied parameter in modelling is related to traffic flow. In this study, pick hour traffic volume (PHV) is used for entering the effect of traffic flow in crash occurrence. Traffic flow was extracted from the data file of the Transportation Department of Tehran's municipality in each of the three highways involved, in input and output points and in different sections. Amount of PHV is measured in each divided section of the three highways on the basis of these explanatory statistics of pick hour traffic volume and involved in the analysis.

In many sections, where traffic flow evaluating was difficult and it was necessary to control the accuracy of vehicle traffic on a highway; it was assumed that the total number of vehicles leaving a carriageway at a given exit point was equal to the number of vehicles entering the other carriageway through the corresponding controlled access point in the opposite direction. Table 3 gives the maximum and minimum registered PHV statistics in each of the three highways.

2.2 Geometric design characteristics

A lot of research results show that changing the geometric design characteristics can reduce the accident occurrence chance (e.g. Ma *et al.* [14]). Geometric design characteristics of each of the three highways involved are extracted on the basis of final data that are collected by the Transportation Department of Tehran's municipality. The plans of each of the three highways involved are supplied by Autodesk Inc. Auto Cad 2009 software and the analysis is done for parameters and necessary characteristics. These geometric characteristics that are presented in separate horizontal and vertical dimension characteristics were: tangent length, curve length, curvature and longitudinal slope. Table 3 shows collected data of horizontal and vertical geometric characteristics too.

Table 3: Traffic and geometric design characteristic data.

	Data description	Chamran		Navab		Saeidi	
		min	max	min	max	min	max
Traffic flow	PHV (vph)	2506	7562	1885	5190	3416	4823
Geometric design characteristic	Tangent length (km)	0.104	0.767	0.128	0.897	0.044	0.376
	Curve length (km)	0.066	0.601	0.055	0.455	0.095	0.173
	Curvature	0.33	4.00	0.50	3.33	1.49	3.33
	Longitudinal slope (%)	-4.96	4.96	-3.85	3.85	-2.32	2.32
	Number of lanes	3	5	3	5	3	3

All curves are simple circular form in horizontal alignment and whereas curves are adapted to simple curve forms, all curves in geometric characteristic's plan are defined as simple curves. In vertical alignment, because of large noticeable differences between each couple consecutive longitudinal slope and the choice of condition of longitudinal slope, vertical curves are ignored.

3 Methodology and modelling

The aim of the crash prediction model presentation is to identify the more important variables that must be entered in the model with surveying descriptive variables of crash occurrence till the best description of crash is expressed. With these models, effectiveness amount of each effective principal variable in crash occurrence can be determined and the risk of crash occurrence can be decreased with reducing negative effects. The importance of data analysis tools will be identified by this introduction.

In this study SPSS software is applied to data analysis and necessary outcomes presentation. It is one of the professional statistic software which is a comprehensive and flexible analyser. Before data analysis being cleared which variables have a more explanatory ability different to the other variables. So the required regression model should be identified so it is decrypted in 3-2.

3.1 Principal variables choice

In this study, principal variables in modelling are produced in order to the PCA method. Thus it is possible to select principal variables by this method because of avoiding complicated methods such as severe arithmetic logics for example fuzzy logic and such. There are consecutive steps in this method that the principal variables are resulted after carrying out these steps (e.g. Caliendo *et al.* [10]).

A 3-year monitoring period extending from 2006 to 2008 was perused on three urban arterial highways. The geometric design of these three highways is

assumed as a set of tangent and simple curve (without link curve) segments and the other necessary data were collected in addition to the crash data and geometric design characteristics during the studying period. Crash data are categorized separately in similar sets on the basis of similarity of tangent and curve segments for each direction after collecting.

By categorizing data, final variables are: Y (occurred crash count), L (curve section length whether tangent), $1/R$ (curvature), n (number of lane). Y as occurred crash count in each section is identified as a dependent variable on the basis of the other independent variables.

Collected data must be standardized. So the Z standard matrix for primary variables is introduced.

$$Z_{ij} = (x_{ij} - x_j) / S_j, i=1, 2, \dots, n \text{ \& } j=1, 2, \dots, q \quad (1)$$

where, x_j and S_j are respectively, mean and standard deviation of the generic variable x_{ij} . The elements in Z matrix have zero mean and unit variance. The covariance between two variables z_k and z_j (for $k \neq j = 1, 2, \dots, n$) is the correlation coefficient.

Next the covariance matrix (R) was computed, which contains the correlations between the original variables:

$$R = 1/n Z' Z \quad (2)$$

where Z' is given by the matrix transpose of Z.

From the so-called characteristic equation of the covariance matrix:

$$\text{Det}(R - \lambda I) = 0 \quad (3)$$

where I is the identity matrix containing unit values, the eigenvalues λ_h were calculated.

The variances accounted for by the principal components were computed with the eigenvalues λ_h .

The eigenvectors v_h were identified by means of the matricial equation:

$$(R - \lambda_h I) v_h = 0 \quad (4)$$

In addition, the S matrix was determined, which contains the correlation coefficients between the original variables and the principal axes:

where V is the matrix of the eigenvectors and L is the diagonal matrix whose elements are the eigenvalues. Next the principal components were rotated by the quartimax method in order to assist in interpreting the results.

This method permits the maximizing of the sum of the fourth power of elements contained in U, which is:

$$U = S T \quad (5)$$

where, T is the rotation matrix.

The multiple correlation coefficients (ρ^2) between original variables and rotated principal axes were computed as the addition of the square of the elements for each row contained in U. Higher ρ^2 are associated with the most significant original variables.

Then in order to PCA method, the results obtained were plotted as correlation circle diagram. The correlation's deal of each variable is identified by these diagrams. When the points representing the original variables are closer to the principal axes as well as the circumference a high correlation is shown. Furthermore, if the original variables are closer to each other a close correlation between these variables is revealed.

In order to the correlation's deals of variables in difference to each other, n variable has maximum correlation with the other variables. Results of table 4 show that it is possible to describe crash occurrence by selected variables up to 92% among the available 6 independent variables and 1 dependent variable. By these 5 variables crash occurrence is predictable by acceptable sureness and the effect of 6th variable is not noticeable. Thus, it is possible to model occurrence in order to this table's results without n variable.

Table 4: Total variance values.

Component	Tangent components			Curve components		
	Initial Eigenvalues		Rotated components matrix	Initial Eigenvalues		Rotated components matrix
	Variance	Cumulative	ρ^2	Variance	Cumulative	ρ^2
ACC	34.456	34.456	0.816	25.147	25.147	0.860
L	23.053	57.509	0.930	21.554	46.702	0.978
C	-	57.509	-	19.997	66.699	0.813
PHV	18.084	75.593	0.891	13.604	80.302	0.933
L.S	16.427	92.020	0.967	12.410	92.713	0.986
n	7.980	100.000	0.999	7.287	100.000	0.992

U matrix exhibits ρ^2 values as shown in table 4 that it is possible to consequence the principal variables on the basis of its results. The values of ρ^2 in tangent sections are between 0.816 up to 0.999 and in curve sections are between 0.813 up to 0.992 on the basis of multiple correlation coefficients as shown in table 4. Minimum effect on crashes is applied by PHV. Since there is a little difference between ρ^2 for this variable and the other variables, it is withdrawn the omission. L.s variable has the best correlation with crash counts while it has the least correlation with the other variables. The dependence of L.s variable to crash occurrence is 93%. Figure 1 is presented on the basis of two first parameters of U matrix which explains the correlation's amount each variable to axis and to each other. This diagram identifies which variables should stay in modelling.

Consequences show that the crash count variable (ACC) in tangent sections is almost independent of first and second axis. PHV and L variables have maximum dependence to first and second axis respectively. While L.s and n variables show dependence to second axis slightly. Consequently, the best correlation with crash counts is belonged to L.s that it is confirmed in previous ratiocination. In addition, consequences show that (ACC) in curve sections is

tended to second axis. The L.s variable is dependent to this axis slightly. The PHV variable is dependent to first axis while n variables show dependence to this axis slightly. The L and $1/R$ are almost independent from two axes. Also, there is dependence between L and $1/R$ variables.

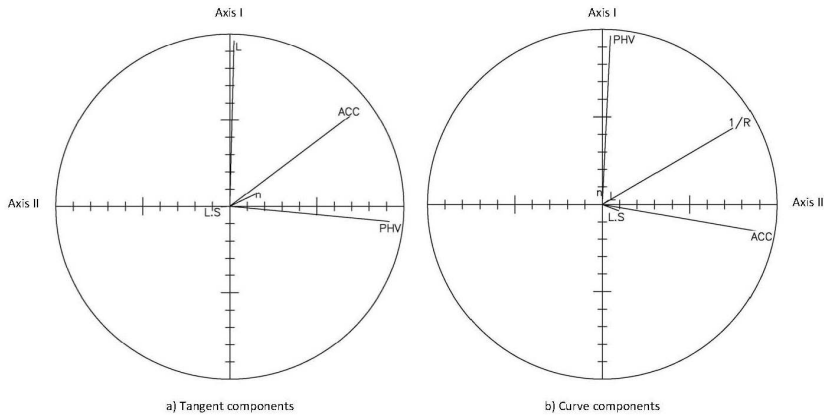


Figure 1: Correlation circle diagram.

The remarkable result is that the L.s variable has most effect on occurred crash counts in tangents and curves amount the explanatory variables. There are several reasons to affect crash occurrence by longitudinal slope:

- Amount of maximum longitudinal slope and length of a section are effective on amount of speed travelling of vehicles.
- Control of vehicles is more severe in plunge slopes because of longitudinal friction coefficient and pavement condition.
- Sight distance in convex vertical curves (at collision location of two consecutive longitudinal slopes) is decreased.

On the basis of discussed ratiocinations, effective and included variables in urban arterial highway accidents modelling are specified in each tangent and curve sections. So the parameters of crash prediction model for tangent sections are section length (L), pick hour volume of traffic (PHV) and longitudinal slope (L.s). Also the parameters for curve sections are section length (L), pick hour volume of traffic (PHV) and longitudinal slope (L.s) and curvature ($1/R$).

3.2 Selecting the regression model

Next significant step in data modelling is selecting one of linear or nonlinear regression models to present an adapted model for collected data that it can presents best adapted estimates on statistic. In linear regression models, used approximations are coefficients of a linear equation which contains one or more independent variable which express the best estimated value of independent variable. Also, for non-linear regression, coefficients of a nonlinear equation

such as exponential, power, logarithmic and etc. is presented which show good approximation of nonlinear variable (e.g. Turner *et al.* [15]).

In this research, after testing the linear regressions, we did not discover good consequences, so regarding to nonlinear models which lead to better results; the crash counts was modelled. Also nonlinear regression models were used in modelling test and final model was selected as a combination of some models.

Other noticeable point in regression model is the distribution type which data are matched in them. For data distribution, there are many data distribution method which can use the best one referring to this point that used data are matched in one of these methods. Some of these usual mathematical distributions are Normal distribution, Gamma distribution, Binomial distribution, Multinomial distribution, Poisson distribution and other liked them. For recognizing the used data matched in such type of distribution, we can specify the type of distribution by Q-Q diagram in SPSS software which expresses the type of data distribution.

3.3 Presentation of prediction crash model

Registered crashes data during the 3-year monitoring period extending from 2006 to 2008 was perused on three urban arterial highways and was analysed for presentation of prediction crash model. These statistics contain curves and tangents accident data separately. According to the last parts, modelling is carried out for curves and tangents separately.

3.3.1 Crash prediction model for tangents

Tangents accidents data in 42 parts of tangents (84 parts in couple directions of roadway) are considered in three urban arterial highways with 13015 meters total length. Regarding to analysis, principal variables of model are section length (L), Pick hour volume of traffic (PHV) and longitudinal slope (L.s).

Data analysis for presentation of model coefficients is done with SPSS software in 25 steps. The results of analysis express that we have more accurate estimations to compare with another models. The value of R^2 coefficient was increased up to 37.8% which is more accurate with comparison the last values.

Then, for using model coefficients with more certainty, we have to reduce the value of R^2 . In this research, for approaching better values in regression coefficients to compare with the last step coefficients, this restriction has been reduced by Min Residual Method. The regression model of tangent accident prediction is obtained by Min Residual Method, after 50 step processing by software, with reducing the value of residues:

$$ACC = 44.927 \left\{ \begin{array}{l} Exp[0.347Ln(L)] + Exp(0.610L.s) \\ + Exp(1.910PHV^{0.333}) - 22.391 \end{array} \right\} \quad (6)$$

For example, crash amount in a tangent section is up to 35 accident in year with $L=0.490$ Km length, $L.s=0.60$ longitudinal slope and $PHV=3.898 \times 10^{-3}$ vph Pick hour volume of traffic.

Described model for crash count prediction shows that accident occurrence is increased in tangent sections of urban arterial highways referred to the value and

the sign of coefficients of model variables with increasing in the section length (L), Pick hour volume of traffic (PHV) and longitudinal slope (L.s) parameters.

3.3.2 Crash prediction model for curves

Curve accidents data in 34 parts of tangents (68 parts in couple directions of roadway) are considered in three urban arterial highways with 7194 meters total length. Regarding to analysis, principal variables of model are section length (L), Pick hour volume of traffic (PHV), longitudinal slope (L.s) and curvature (1/R).

Data analysis for presentation of model coefficients is done with SPSS software in 7 steps. The results of analysis express that we have more accurate estimations to compare with another models. The value of R^2 coefficient was increased up to 38.2% which is more accurate the comparing with the last values.

Then, for using model coefficients with more certainty, we have to reduce the value of R^2 . In this section like last section, for approaching better values in regression coefficients with comparison with the last step coefficients, this restriction has been reduced by Min Residual Method. By Min Residual Method, after 50 step processing by software, with reducing the value of residues, the regression model of curve accident prediction is obtained:

$$ACC = -1.333 \left[\ln(5.949 \times 10^{-6}L) + 0.972L.s - 0.239L.s^2 \right] + 3.967\left(\frac{1}{R}\right) - 0.233PHV \quad (7)$$

For example, crash amount in a curve section is up to 23 accident in year with $L=0.139\text{Km}$ length, $L.s=0.59$ longitudinal slope, $1/R=1.61$ curvature and $PHV=3.898 \times 10^{-3}\text{vphh}$ Pick hour volume of traffic.

Described model for crash count prediction in curve sections of urban arterial highways show that accident incident is increased referred to the value and the sign of coefficients of model variables with increasing in the section length (L), Pick hour volume of traffic (PHV) and longitudinal slope (L.s) parameters and decreasing in the curve radius (R) parameter.

4 Conclusions

In this study, accidents are occurred in urban arterial highways of Tehran are studied for first time and a logic and trusty model of occurred crashes behaviour in these highways is introduced. The purpose was including the different effective phases in crash occurrence such as traffic flow and geometric design characteristic parameters of these highways to present a model which raises road safety and reduces harms with studying of parameters effectiveness. Necessity of exhibition of this research is that increase of safety in urban highways with identifying of important causes of accident.

In this research, crash-prediction models for 3 urban arterial highways in Tehran were set up on the basis of accident data analysis observed during a 3-year monitoring period extending between 2006 and 2008. These statistics contains separate data of crash occurrences for each tangent and curve sections of highways. For each accident a variety of details was recorded, including location of accident, horizontal alignment (tangent or curve), vertical alignment (upgrade or

downgrade), weather and pavement surface conditions (dry or wet), type and severity of accidents. Finally, the parameters which are screened to study accidents quantitatively are: crash occurred count (ACC), section's length (L), pick hour volume (PHV) and longitudinal slope (L.s), and curvature (1/R).

Observations show that some 12,891 accidents were considered in this study, 32 of which were fatal, 1,348 were injury accidents and 11,601 were led to just financial damage. So in summary, during the period of observation 1,380 severe accidents (fatal and injury crashes) equivalent 10.63% of all crashes and 11601 no severe accidents equivalent 89.37% of all crashes were registered. 34.4% of all crashes occurred on curves and 65.6% of all crashes occurred on tangents. So the crash occurrence chance in tangents is 1.91 crash occurrence in curves.

Effective principal variables in crash occurrence are selected with PCA method. The variables are regarded which are more independent and affective and lesser correlated in comparing to the other variables. Analysis consequences show that the principal variables of tangents are: section's length (L), pick hour volume (PHV) and longitudinal slope (L.s), whereas for tangent they are section's length (L), pick hour volume (PHV) and longitudinal slope (L.s) and curvature (1/R).

After screening the principal variables, data analysed by SPSS software with nonlinear regression models and final crash prediction model is presented for each tangent and curve sections separately.

Also R^2 values are observed 37.8% and 38.2% respectively in tangent and curve sections that they are suitable the compare with many similar models. Min residual method is applied to decrease the R^2 values and better assurance to have a suitable model.

Presented models show that crash occurrence is increased with the increase in each of section's length (L), pick hour volume (PHV) and longitudinal slope (L.s) variables whereas it is increased with the decrease of curvature (1/R).

The remarkable result in this study was the effect of longitudinal slope variable (L.s) on crash occurrence. This variable has least dependence to the other variables. With simple explanation, L.s has the best effectiveness in the crash occurrence of urban arterial highway. First reason that L.s has the most effect on occurred crash count is the relation between speed variation and longitudinal slope. The traffic congestion is changed by speed variation obviously that is a clear reason for relation between longitudinal slope, decreasing safety and consequently decreasing the chance of crash occurrences. Second reason is that the effect of longitudinal slope amount is increased and pavement condition in more plunge longitudinal slope and the chance of crash occurrence is increased too. The third reason is that the sight distance in convex vertical curves (at collision location of two consecutive longitudinal slope) is decreased and consequently the risk if crash occurrence is increased.

The amount of R^2 as an index of model's confidence is noticeable in researches. More parameters might exist that can effect crash occurrence and it is possible to increase the amount of this index by entering some of them. Driver's behaviour is one of effective variables on crash occurrence that it is not comfortable to be studied. So it is recommended to other researchers in the

future opportunities to notice the parameters and causes are related to driver's behaviour, environment condition and the other variables are related to transportation infrastructures.

The above presented results exhibit some of more important causes leading to crash occurrence on urban highways. Effective amounts of crash variables can be determined by the models shown and it can reduce amounts of their negative effects and consequently crash counts too. Moreover, the amounts of these variables can be noticed in the design of new roads.

References

- [1] Persaud, B., Dzbik, L., Accident prediction models for freeways, *Transportation Research*, Rec. 1401, pp. 55–60, 1993.
- [2] Knuiman, M.W., Council, F.M. and Reinfurt, D.W., Association of median width and highway accident rates, *Transportation Research*, Rec. 1401, 1993.
- [3] Shankar, V., Mannering, F. and Barfield, W., Effect of roadway geometrics and environmental factors on rural freeway accident frequencies, *Accident Analysis and Prevention*, 27(3), pp. 542–555, 1995.
- [4] Hadi, M.A., Aruldas, J., Chow, L. and Wattleworth, J.A., Estimating safety effects of cross-section design for various highway types using Negative Binomial regression, *Transportation Research*, Rec. 1500, 1995.
- [5] Persaud, B., Retting, R.A. and Lyon, C., Guidelines for the identification of Hazardous Highway Curves, *Transportation Research*, Rec. 1717, pp. 14–18, 2000.
- [6] Martin, J., Relationship crash rate and hourly traffic flow on interurban motorways, *Accident Analysis and Prevention*, 34, pp. 619–629, 2002.
- [7] Golob, T.F., Recker, W.W., Relationship among urban freeway accidents, traffic flow, weather, and lighting conditions, *Transportation Engineering*, 129, pp. 342–353, 2003.
- [8] Golob, T.F., Recker, W.W. and Alvarez, V.M., Toll to evaluate safety effect of changes in freeway traffic flow, *Transportation Engineering*, 130(2), 2004.
- [9] Hauer, E., Statistical road safety modelling, *Proc. of 3rd Int Conf. of TRB*, 2004.
- [10] Caliendo, C., Parisi, A., Principal component analysis applied to crash data on multilane roads, *Proc. of 3rd Int Conf. of SIIV*, 2005.
- [11] Caliendo, C., Maurizio, G. and Parisi, A., A crash-prediction model for multilane roads, *Accident Analysis and Prevention*, 39, pp. 657–670, 2007.
- [12] Panagiotis, Ch., Andrew P. and Fred, L. Mannering Tobit analysis of vehicle accident rates on interstate highways, *Accident Analysis and Prevention*, 40(2), pp. 768–775, 2008.
- [13] Noland, R.B. and Oh, L., The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data, *Accident Analysis and Prevention*, 36, pp. 525–532, 2003.



- [14] Ma, J., Kockelman, K.M., Damien, P., A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity, using Bayesian Methods, *Accident Analysis and Prevention*, 2008.
- [15] Turner, S., Persaud, B., Chou, M., Transferability of Overseas Crash Prediction Models to New Zealand, IPENZ, 2007.
www.ipenz.org.nz/ipenztg/papers/2007/A3a_TurnerPersaudChou.pdf

