FINAL REPORT

THE EFFECT OF SPEED, FLOW, AND GEOMETRIC CHARACTERISTICS ON CRASH RATES FOR DIFFERENT TYPES OF VIRGINIA HIGHWAYS

Nicholas J. Garber, Ph.D.
Professor of Civil Engineering
and
Faculty Research Engineer

Angela A. Ehrhart
Graduate Research Assistant

ABSTRACT

Although considerable progress has been made over the past several years in making highway travel safer, the frequency and severity of speed-related crashes on the nation's highways continue to be of concern. Understanding the factors associated with these crashes enables engineers to identify and implement effective countermeasures to reduce the probability of crashes. A number of studies have been conducted to determine the variation of crash rates as they relate to hourly traffic volumes, geometric characteristics, average speed, and speed variance. However, these studies have not established mathematical relationships that can be used to estimate changes in the crash characteristics as a result of the combined changes in speed, flow, and geometric characteristics. The establishment of direct mathematical models that describe the influence of these factors on crash characteristics would significantly enhance the efforts of traffic engineers to determine suitable countermeasures to reduce the occurrence and severity of crashes. This project develops mathematical relationships that describe the combined influence that traffic and geometric characteristics have on crash occurrences.

This study was limited to roadways in the state of Virginia with speed limits of 89 or 105 km/h (55 or 65 mph). The data were obtained from speed monitoring stations established by the Virginia Department of Transportation (VDOT) and from police accident reports from January 1993 to September 1995.

Using the variables of mean speed, standard deviation of speed, flow per lane, lane width, and shoulder width to predict crash rates, different types of deterministic models, such as multiple linear regression, robust regression, and multivariate ratio of polynomials were fitted to the data. The multivariate ratio of polynomials was found to be the only mathematical model type that was successful in describing any relationship between the combined effects of changes in the speed, flow, and geometric characteristics of the road on crash rates. Based on this study, all of the models show that under most traffic conditions, the crash rate tends to increase as the standard deviation of speed increases. The effect of the flow per lane and mean speed on the crash rate varied with respect to the type of highway.

FINAL REPORT

THE EFFECT OF SPEED, FLOW AND GEOMETRIC CHARACTERISTICS ON CRASH RATES FOR DIFFERENT TYPES OF VIRGINIA HIGHWAYS

Nicholas J. Garber, Ph.D.
Professor of Civil Engineering
and
Faculty Research Engineer

Angela A. Ehrhart
Graduate Research Assistant

INTRODUCTION

Although much progress has been made over the past several years in improving safety on the nation's highways, the frequency and severity of crashes continue to be of concern. Identification of the factors associated with crashes has also continued to be of major interest to traffic engineers. A better understanding of the factors associated with crashes enables engineers to identify and implement effective countermeasures to reduce crash occurrences. Several studies have been conducted to determine the relationships between factors associated with crashes and crash characteristics.[1-9] Although the results of these studies have not been consistent, their data strongly suggest that crash occurrences and severity are greatly influenced by speed characteristics, such as mean speed and speed variance. For example, there is strong evidence to suggest that a higher standard deviation of speed in a traffic stream (which indicates higher variation in vehicle speeds) increases the occurrence of crashes. A number of studies have been conducted to determine the variation of crash rates as they relate to hourly traffic volumes, geometric characteristics, average speed, and standard deviation.[2-7] However, these studies have not established mathematical relationships that can be used to estimate changes in crash characteristics as a result of the combined effects of changes in speed, flow, and geometric characteristics. The establishment of such mathematical models would significantly enhance the efforts of traffic engineers to determine suitable countermeasures to reduce the occurrence and severity of crashes.

PURPOSE AND SCOPE

The purpose of this research was to identify the speed, flow, and geometric characteristics (grade, lane width, and shoulder width) that significantly affect crash rates on different types of roads and to develop mathematical relationships that describe how these characteristics affect crash rates.

The scope of this study was limited to roadways in the state of Virginia with speed limits of 89 or 105 km/h (55 or 65 mph). The data were obtained from speed data collected at speed monitoring stations established by the Virginia Department of Transportation (VDOT) and the associated crashes on those roads.

The specific objectives of the study were as follows:

1.   To review previous research and identify established relationships between crash rates and the characteristics of speed and flow;

2.   To develop deterministic mathematical relationships between crash rates and mean speed, standard deviation of speed, flow per lane, lane width, and shoulder width on different types of highways.

METHODOLOGY

The methodology followed in this study consisted of the following tasks:

1.   A *Literature Review* was conducted to identify the major factors associated with different types of crashes and crash characteristics;

2.   *Data Collection* was done to obtain speed, flow, crash, and geometric data and to define roadway segments;

3.   *Data Reduction* of the speed, flow, and crash data was performed;

4.   *Model Selection Criteria* was established to select the best model;

5.   *Model Development*, using multiple linear regression, robust regression, and multivariate ratio of polynomials was conducted.

Task 1: Literature Review

A literature search on the relationships between crash rates and characteristics of flow, speed, and highway geometry was performed using the Transportation Research Information Service (TRIS) database, the Virginia Transportation Research Council, and the University of Virginia libraries.

Variations with Traffic Volume

The variation of the crash rate as it relates to traffic volumes has been examined in numerous studies.[2-7] The conventional wisdom among the general population is that crash rates should increase with increasing volumes. Previous research on the relationship between crash rates and hourly traffic volume has indicated a "U"-shaped function encompassing broad ranges of traffic volumes.[2-6] This indicates that during low volume periods, such as those that occur during the early morning and late day hours, higher crash rates are observed than when the volumes of the road are greater. A study conducted by Brodsky and Hakkert found contradictory results in the study of crash rates and the volume of traffic.[6] Regression analyses on data obtained from primary and secondary highways show a positive relationship between injury crash rates and traffic volumes, while the relationship is invariant on interstates. These associations, however, were not very strong. The results showed correlations of 0.5 or less. Fatal crashes have a negative correlation with volume, which can be attributed to the relationship between traffic volume and speed and speed variability. The study also found that a modest increase in the density over time has no significant effect on the average total crash rates.

Variations with Speed Characteristics

Research has shown that the greater a vehicle speed deviation is from the average speed, the greater the probability of that vehicle being involved in a crash. This relationship is often referred to as being "U-shaped" and has been verified in research.[7] Studies performed by Garber and Gadiraju have found that speed variance decreases as the average speed increases.[8] They also found that the average speed is dependent on the design speed, and therefore speed variance is also dependent on the design speed. In concurrence with Garber and Gadiraju, a study in France was performed by Lassarre in which models were developed for predicting the severity of crashes based on traffic volume, speed, standard deviation of speed, and the wearing of seat belts.[9] In this study, the models indicate that only a small influence upon safety is brought about by changes in average speeds. Inferences from the data indicate that a greater homogeneity of speeds increases the safety level.

Variations with Geometric Characteristics

Many previous research projects have shown relationships between crashes and geometric characteristics on roadways.[4,10-20] A relationship between the number of lanes and the crash rate has been developed, which suggests that as the number of lanes increases, the crash rate decreases.[10] However the opposite has also been shown to exist.[11] Some four-lane highways have been shown to have higher crash rates than two-lane facilities. This is attributed to roadside development, heavier traffic volumes, and more intersections.[11] Yet other studies have indicated no relationship between the number of lanes and crash rates on roads with similar design standards.[12]

Inconsistent results have also been found for the relationships between crash rates and lane widths. Based on the results of an extensive literature review on the safety aspects of two-lane roads, an inverse relationship was found between crashes and pavement width–in that crash rates tend to decrease with increasing pavement width up to a width of about 7.6 meters (25 feet).[13] Other studies have found no trends between crash rates and lane width or total paved surface width.[14,15,16] Dart and Mann found that the difference in crash rates between 3.4 and 3.7 meters (11 and 12 feet) lanes is insignificant.[17] A study by Garber et al. concerning accident rates of trucks on different lane widths revealed higher truck accident rates on roads with lane widths of 3.1 and 3.2 meters (10.0 and 10.5 feet) than on roads with lane widths of 3.4, 3.5, and 3.7 meters (11.0, 11.5, and 12.0 feet).[18]

Studies exploring the relationship between crash rates and shoulder width have produced contradictory results. Several studies, using statistical analyses, found no relationship between the right shoulder width and crash rates.[16, 19] However, some studies concluded that crash rates decrease as shoulder width increases on two-lane roads.[14, 15] This may be due to methodological problems in some of the studies.

In general, the crash rate tends to gradually grow linearly with increasing grades, but grows exponentially with decreasing grades.[13, 17] These results have been attributed to the operational differences between heavy vehicles and cars on high grades. However, one contradictory study by Raff et al. has found that the crash rate is unrelated to the grade for tangents on freeways, multilane divided and undivided highways, and two-lane highways.[14]

In relating their results on the theory that design speed is a function of the roadway geometrics, Garber and Gadiraju found that drivers travel at higher speeds on highways with better geometric characteristics regardless of the posted speed limit.[8]

Summary of Literature Review

The literature review has identified several inconsistencies in the research regarding relationships between the crash rate and traffic volume. Some studies suggest that the crash rate increases with an increase in the volume, while other studies show the opposite to occur. The literature review has also indicated that the standard deviation of speed affects the crash rate by showing that the crash rate increased as drivers traveled both slower and faster than the average speed; however, no previously developed models included the standard deviation of speed as a variable. Geometric characteristics such as the number of lanes, lane widths less than 11 feet, shoulder widths, and grades greater than 4 percent were also found to have relationships with the crash rate.

While previous studies have defined relationships between crash rates and parameters such as geometric, environmental, and behavioral characteristics, there is a lack of research defining crash frequency and severity based on the combined effect of speed and flow characteristics. Most notably, the parameter of the standard deviation of speed in a traffic stream is excluded in these relationships.

Task 2: Data Collection

The data collection phase consisted of the following sub-tasks:

1.   Obtaining speed and flow data;

2.   Defining roadway segments;

3.   Obtaining crash and geometric data.

Speed and Flow Data

The first step of this task was to obtain speed and flow data over a period of time on various road types. The Virginia Department of Transportation (VDOT) was contacted and found to have hourly speed and flow data that were collected at 52 locations in Virginia over the period from 1993 to 1995. These data were originally collected to comply with the Federal Highway Administration's requirement of enforcing the National Maximum Speed Limit on all public highways. Therefore, the data consisted of a somewhat sporadic monitoring of urban and rural highways varying from two to eight lanes, with speed limits of either 89 or 105 km/h (55 or 65 mph). The data were obtained on a quarterly basis for each year, although no sites were monitored every quarter. Table 1 shows the number of sites and data points collected for each road type.

Table 1. Data Collected by Highway Type

Highway Type	No. of Sites	No. of Data Points
Freeway with 105 km/h (65 mph) Speed Limit	8	47
Freeway with 89 km/h (55 mph) Speed Limit	9	87
4-Lane Non-Freeway with 89 km/h (55 mph) Speed Limit	20	151
2-Lane Non-Freeway with 89 km/h (55 mph) Speed Limit	15	139

Roadway Segments

Once the speed data were obtained, varying lengths of roadway segments were identified around each monitoring station as the test sites. These segments were based on major intersection and interchanges in order to ensure homogenous traffic type and flow characteristics. Many of the rural two-lane sites were extended in length to ensure that enough crashes occurred for analysis. The sites were identified by the use of maps, personal experience, and visits to the roadways.

Crash Data

After the site locations were defined, the node locations or segment endpoints were sent to VDOT's Central Office where, by use of the Highway and Traffic Records Information System (HTRIS), FR 300 police accident reports were obtained for each test site over the three-year period.

Geometric Data

At each site location, the following geometric characteristics were obtained: number of lanes, lane width, shoulder width, and whether the grade was less than or greater than 4 percent.

Task 3: Data Reduction

In order to facilitate a systematic analysis of the data, the volumes obtained were categorized under different volume ranges. Each of the categories of roads was considered separately, and volume ranges were defined based on the road type's volume. The following volume ranges were defined (number of vehicles/ hour):

- Freeways at 105 km/h (65 mph) 0-500 501-1000 1001-1500 1501-2000

- Freeways at 89 km/h (55 mph) 0-1000 1001-3000 3001-4500 4501-6200

- 4-lane non-freeways 0-500 501-1000 1001-1500 1501-2000 2001-2500

- 2-lane non-freeways 0-75   76-150   151-225   226-300

During the hours of each corresponding volume range, the following descriptive data were determined for each test site: mean speed, standard deviation of speed, and flow per lane.

The FR 300 accident reports obtained from VDOT were divided by direction of interest based on the location of the monitoring station. The crash data were matched to the roadway segments based on corresponding time periods. Only weekday crashes were matched to weekday counts and weekend crashes to weekend counts to ensure that similar speed and traffic characteristics occurred during the crashes. Information pertaining to the number of crashes that occurred during a volume level, the number of vehicles involved, the day of the week the crash occurred, and the type of the collision were taken from the reports. Since single vehicle crashes do not necessarily occur as a result of the interaction of speed variance, flow, and density, the number of crashes (excluding single vehicle crashes) was determined for each site. This allowed for the development of models that did not include single vehicle crashes.

Once all the data were collected and reduced, they were entered into a spreadsheet for analysis. The spreadsheet consisted of the following variables: station number for site identification, volume level, number of hours at a volume level, mean speed (km/h), standard deviation of speed (km/h), flow per lane in vehicles per hour (vph), lane width (m), shoulder width (m), number of crashes, site length (km), number of lanes, and grade (<4% or ≥4%). Since the number of crashes in the data were dependent on the length of the highway section, the number of hours for which the data were collected, and the number of lanes, it was necessary to normalize the variable number of crashes to a variable called CRASHRATE. This was done by dividing the number of crashes by the length, number of hours the data were collected for each volume level, and the number of lanes in one direction. Therefore, the variable crash rate is defined as the crashes per hour, per kilometer, and per lane. Table 2 shows part of the data for rural interstate highways.

Table 2. Rural Interstate Highways Data

STATION	VOLUME	HOURS	MEAN	SD	FPL	LW	SW	CRASH_NO	LENGTH	GRADE	LANE_NO	CRASHRATE
1	1	21	101.269	9.6117	120	3.66	1.525	0	8.5813	0	6	0
1	2	3	102.879	9.91277	265	3.66	1.525	0	8.5813	0	6	0
1	1	23	102.396	9.51993	125	3.66	1.525	2	8.5813	0	6	0.003377752
1	2	1	103.845	9.6117	275	3.66	1.525	0	8.5813	0	6	0
1	1	23	102.879	9.91921	128	3.66	1.525	1	8.5813	0	6	0.001688876
1	2	1	104.65	10.01581	280	3.66	1.525	0	8.5813	0	6	0
1	1	22	103.845	9.48773	143	3.66	1.525	0	8.5813	0	6	0
1	2	2	104.489	9.34122	274	3.66	1.525	0	8.5813	0	6	0
2	1	9	106.582	9.62458	102	3.66	1.525	2	7.5992	0	4	0.014621422
2	2	4	109.158	8.62119	363	3.66	1.525	1	7.5992	0	4	0.0164491
2	3	11	109.48	8.8872	593	3.66	1.525	1	7.5992	0	4	0.005981491
2	1	10	106.582	8.71332	103	3.66	1.525	0	7.5992	0	4	0
2	2	6	109.319	8.80187	404	3.66	1.525	0	7.5992	0	4	0
2	3	8	109.802	8.56359	579	3.66	1.525	0	7.5992	0	4	0
2	1	8	107.226	9.13353	95	3.66	1.525	0	7.5992	0	4	0
2	2	4	109.802	9.26716	335	3.66	1.525	0	7.5992	0	4	0
2	3	9	110.446	8.9194	608	3.66	1.525	3	7.5992	0	4	0.021932133
2	4	2	111.895	8.46538	801	3.66	1.525	1	7.5992	0	4	0.0328982
2	1	8	106.904	8.84695	96	3.66	1.525	1	7.5992	0	4	0.00822455
2	2	4	108.836	8.68434	343	3.66	1.525	0	7.5992	0	4	0
2	3	9	109.319	8.73747	614	3.66	1.525	2	7.5992	0	4	0.014621422
2	4	3	109.963	8.7262	791	3.66	1.525	0	7.5992	0	4	0
3	1	6	101.913	10.64693	161	3.66	3.66	0	4.4436	0	4	0
3	2	6	101.108	9.70025	407	3.66	3.66	0	4.4436	0	4	0
3	3	9	100.947	9.74855	725	3.66	3.66	1	4.4436	0	4	0.012502375
3	4	3	99.82	9.7727	785	3.66	3.66	0	4.4436	0	4	0
4	1	22	103.201	10.75641	151	3.66	1.525	3	13.4113	0	4	0.005083908
4	2	2	104.167	10.48432	254	3.66	1.525	0	13.4113	0	4	0

Task 4: Model Selection Criteria

In order to obtain the best model that fits the observed data, a suitable measurement of quality must be applied in the selection process. Two measurements have been applied in this research, the coefficient of determination (R^2) and Akaike's information criterion (AIC). The R^2 value was used in both the multiple linear and robust regression models to judge the adequacy of each model. The AIC was used to select the best non-linear model of the multivariate ratio of polynomials type of model. The derived model for the AIC is defined as:

$$AIC = \text{-2 ln(L)} + 2k$$

Where

L is the Gaussian likelihood of the model

k is the number of free parameters in the model

Expressed in terms of the sum of square of the errors:[21]

$$AIC = -n \ln\left(\frac{SSE}{n-k}\right) + 2k$$

where

n = number of model residuals

k = number of free parameters in the model
$$SSE = \Sigma(y_i - \hat{y}_i)^2$$

y_i = actual observations

$\hat{y}_i$ = model estimates

The first term in the AIC equation measures the badness of fit, or bias, when the maximum likelihood estimates of the parameters are used. The second term measures the complexity of the model, thus penalizing the model for using more parameters. The goal for selecting the best model is therefore a minimization of the criterion, thus selecting the best fit with the least complexity.

Task 5: Model Development

The procedures used for modeling the data were Linear Regression, Robust Regression and Multivariate Ratio of Polynomials. A correlation analysis was carried out to determine independent variables that were correlated to each other. Correlated variables were not used together in any model. The modeling process began with using the variables of mean speed (MEAN), standard deviation of speed (SD), flow per lane (FPL), lane width (LW), and shoulder width (SW) to predict the crash rate (CRASHRATE). Grade was not a variable in the data collected on freeways with a speed limit of 89 km/h (55 mph), because all sites had grades less than 4 percent. Lane width was not considered a variable in the data collected on freeways with a speed limit of 105 km/h (65 mph), because all sites had a lane width of 3.7 meters (12 feet). All road types were also modeled with single-vehicle crashes excluded, as these crashes are sometimes not related to the interaction among vehicles in the traffic stream as a result of variation in speeds. All modeling was done using the Number Cruncher Statistical Systems (NCSS) software.

Multiple Linear Regression

The data were first applied to the multiple linear regression model, which expresses the mean of the response variable as a linear relationship of two or more predictor variables. The models were fitted to the data by the method of least squares and a t-test performed to test the resulting ∃'s (regression coefficients) against the null hypothesis of being equal to zero (H_0: ∃ = 0) at the 5 percent significance level. The adequacy of the fit was judged based on the R^2 value and the normality of the residuals.

Robust Regression

The data were next applied to robust regression because the robust regression lessens the normality restriction by identifying outliers in the data. Having outliers present in the data tends to draw the least squares fit too much in the direction of the outliers, which can lead to distortions in the estimated regression coefficients and t-tests. The distortion causes the residuals to be smaller than they should be, making them difficult to identify. Robust regression makes residuals larger and easier to spot. Assigning a weight to each observation based on a curve called an influence function controls the residual size. The models were fitted to the data by the method of least squares, after which weights were applied to the residuals and new regression coefficients were estimated until little change was noticed in the residuals. T-tests were then performed to test the resulting β 's (regression coefficients) against the null hypotheses of being equal to zero (H_0: β_j =0) at the 5 percent significance level. The adequacy of the fit was judged based on the R^2 value and the normality of the residuals.

Multivariate Ratio of Polynomials

The multivariate ratio of polynomial procedure is a heuristic process, which searches through hundreds of potential curves looking for a model that best fits the data. The models developed through this process offer a larger variety of surfaces than the usual polynomial models; however, care must be taken not to use the model outside of the range of the data. In addition, the model should be studied graphically to determine that the model behaves as expected between data points.

The program, NCSS, contains a function called the multivariate ratio of polynomials search. This function provides a shortcut to the normally slow iterative process whereby an approximate solution is found quickly because a large number of models may be searched in a short period of time. The procedure provides the option to enter one dependent and up to four independent variables. Both the dependent and independent variables may be transformed in the following forms for searching: $1/Y^2, 1/Y, 1/\sqrt{Y}, \ln(Y), Y^2$. The procedure also provides options for which models may be searched. The models searched in this research involved only the numerator polynomial of the hierarchical type of model up to the second power. The reports generated listed the variables included in each model, the variable transformations, number of parameters in the model, and the R^2 value of the model. Since there is no direct R^2 value for non-linear regression, a pseudo R^2 value is used by NCSS. This value is defined by the following form:

$$_p R^2 = \frac{Model\ SS - Mean\ SS}{Total\ SS - Mean\ SS}$$

where

Mean SS is the sum of squares due to the mean $= \sum (\bar{y})^2$

Model SS is the sum of squares due to the model $= \sum (\hat{y}_i)^2$

Total SS is the total uncorrected sum of squares of the dependent variable $= \sum (y_i)^2$

where

$\hat{y}_i =$ model estimates

$\bar{y} =$ mean of the observations

$y_i =$ actual observations

The $_pR^2$ value in this form reveals how the model performs after removing the influence of the mean of the dependent variable. The $_pR^2$ value is used in the same manner as in a multiple regression, with values closer to 1.0 indicating more explanatory power in the model. Although the value does not have the same definition in traditional sense, it serves well for comparative purposes.

The data were analyzed by combining the variables SD, FPL, MEAN, SW, and LW in 2-, 3-, and 4-variable groups. A maximum number of 4 variables were entered at a time as per the program's allowance.

The data were entered into the multivariate ratio of polynomials search function in groups described above. The combinations of variables associated with the top three $_pR^2$ values in the search results were then entered into the multivariate ratio of polynomials fit function, which generated a report detailing each model selected from the search procedure. The models developed were examined graphically in terms of their ability to represent the collected data.

RESULTS

Multiple Linear Regression

The results of fitting the data using linear regression are shown in Table 3.

Table 3. Multiple Linear Regression

Road Type	Variables		R^2
Freeways at 65 mph	MEAN		0.1352
grades <4%	MEAN		0.1891
no single-vehicle crashes	——		——
Freeways at 55 mph	SD	FPL	0.2888
no single-vehicle crashes	SD	FPL	0.3338
4-lane non-freeways	FPL	SW	0.1767
grades <4%	FPL	SW	0.1715
no single-vehicle crashes	FPL		0.1681
no single & grade <4%	FPL		0.1652
2-lane non-freeways	——		——

Table 3 shows that the multiple linear regression models developed for all road types have low R^2 values. The best model developed was for data collected on freeways with a speed limit of 89 km/h (55 mph), with single-vehicle crashes excluded. No model could be developed for 2-lane non-freeways in which the regression coefficients were not equal to zero. Examination of all residual plots indicated that the residuals were not normally distributed and that the normality assumption inherent in the development of a linear regression model was violated. It was therefore concluded that linear relationships cannot be used to describe the data obtained for all road types.

11

Robust Regression

The results obtained from using the robust regression procedure are shown in Table 4.

Table 4. Robust Regression

Road Type	Variables				R^2
Freeways at 65 mph	——				——
Freeways at 55 mph	MEAN	SD	FPL	LW	0.7400
no single-vehicle crashes	SD	FPL			0.5399
4-lane non-freeways	FPL	LW	SW		0.3004
grades <4%	FPL	LW			0.3140
no single-vehicle crashes	FPL	LW			0.3555
no single & grade <4%	FPL	LW			0.3811
2-lane non-freeways	——				——

This table shows improvement in the R^2 values when compared to the multiple linear regression models, especially for the freeways with a 89 km/h (55 mph) speed limit. However, examination of all residual plots indicated that the residuals were not normally distributed. Since non-normality of the residuals violates an assumption in linear regression analysis, robust regression was determined to be an unacceptable type of model to describe the relationships of the data.

Multivariate Ratio of Polynomials

Table 5 shows the number of models developed that graphically represent the observed data for each road type.

Table 5. Models Examined

Roadway Type	No. of Models Developed
Freeway with 105 km/h (65 mph) Speed Limit	3
Freeway with 89 km/h (55 mph) Speed Limit	3
4-Lane Non-Freeway with 89 km/h (55 mph) Speed Limit	13
2-Lane Non-Freeway with 89 km/h (55 mph) Speed Limit	4

All of the models for each road type show similar trends. The best model of each road type, defined as the model with the lowest AIC, is described in detail in the following section. This description is followed by a summary of all of the models.

Freeways with 105 Km/h (65 mph) Speed Limit: Model 1

The best model (Model 1) for these highways uses the variables SD, FPL, and MEAN to predict the CRASHRATE, with the data collected at sites with grades greater than 4 percent that were excluded from the analysis (in keeping with the results of the literature review). The estimated model is:

$$LN(CRASHRATE) = ((2629.697)-(0.424)*(SD^2)-(5.427E\text{-}04)*(SD^4)-(2254323)*(1/(FPL^2))+(4.490)*(SD^2)*(1/(FPL^2))-(5.397E\text{+}08)*(1/(FPL^2))^\wedge 2-(510.682)*(SQRT(MEAN))+(5.171E\text{-}02)*(SD^2)*(SQRT(MEAN))+(224565.2)*(1/(FPL^2))*(SQRT(MEAN))+(24.69)*(SQRT(MEAN))^\wedge 2)$$

The $_pR^2$ value for Model 1 is 0.7597, and the AIC is 1.3865. Examination of the normal probability and residual plots indicates the residuals are normally distributed. Graphical representations of Model 1 are shown in Figures 1, 2, 3, and 4.
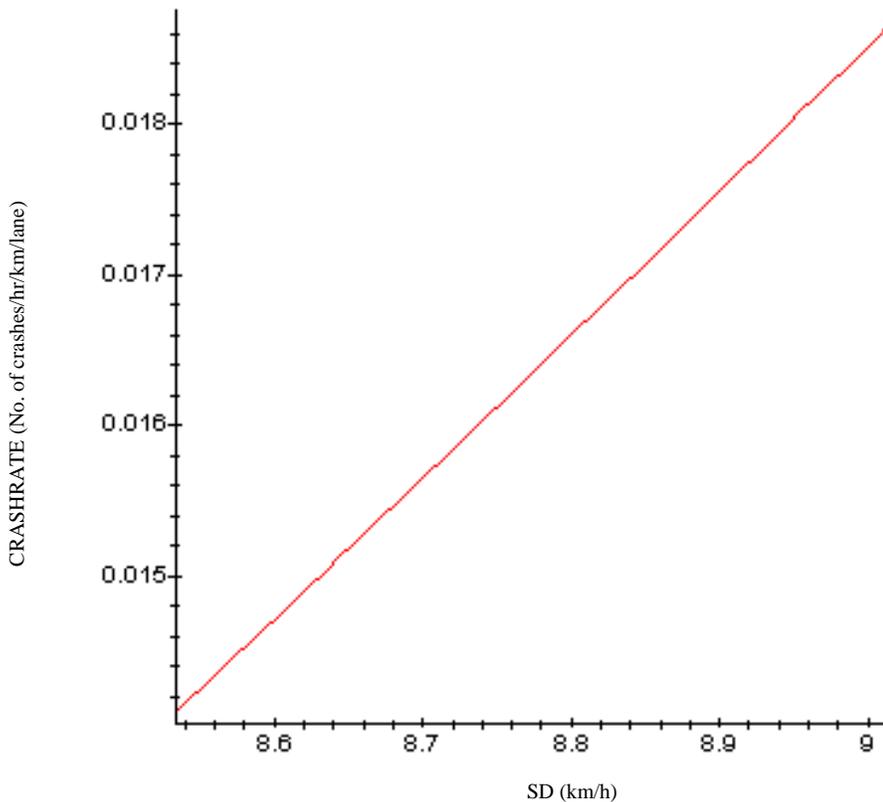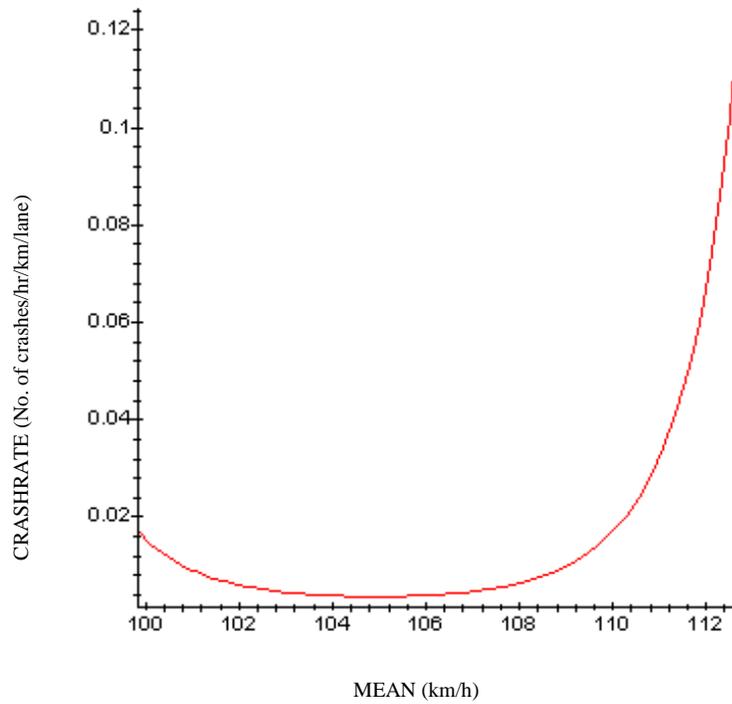

Figure 1. Model 1, CRASHRATE vs. SD for 104.65 Km/H (65 Mph) Freeways (MEAN=109.48 Km/H (68 Mph), FPL=350 Vph)
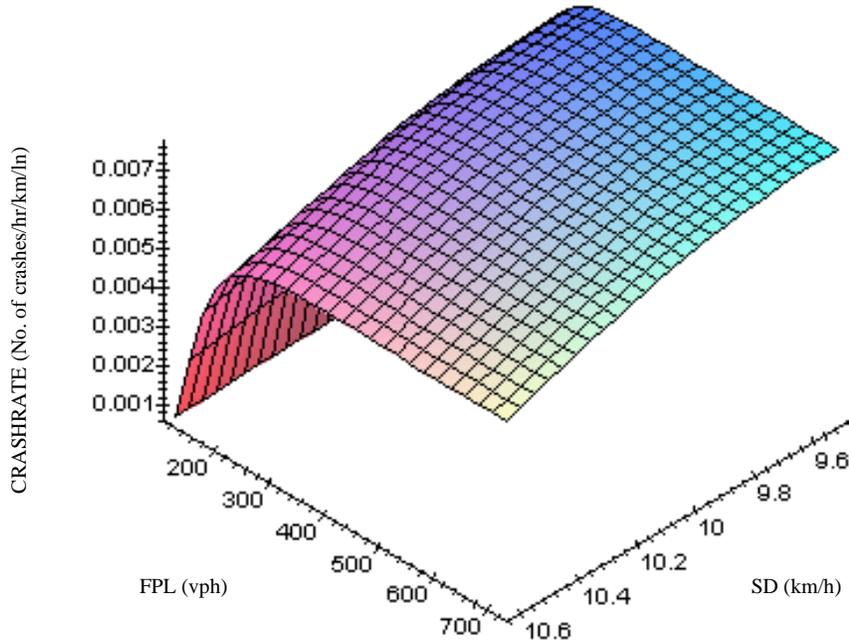
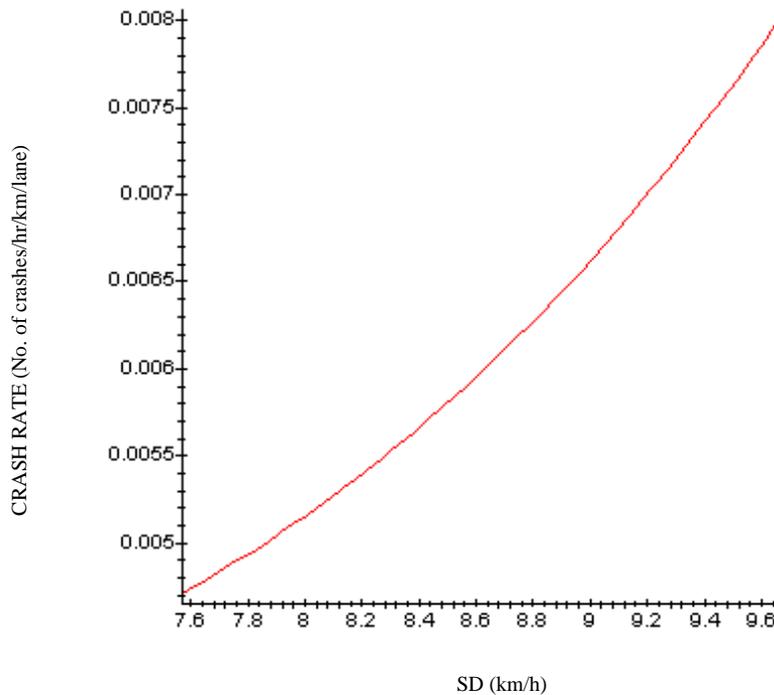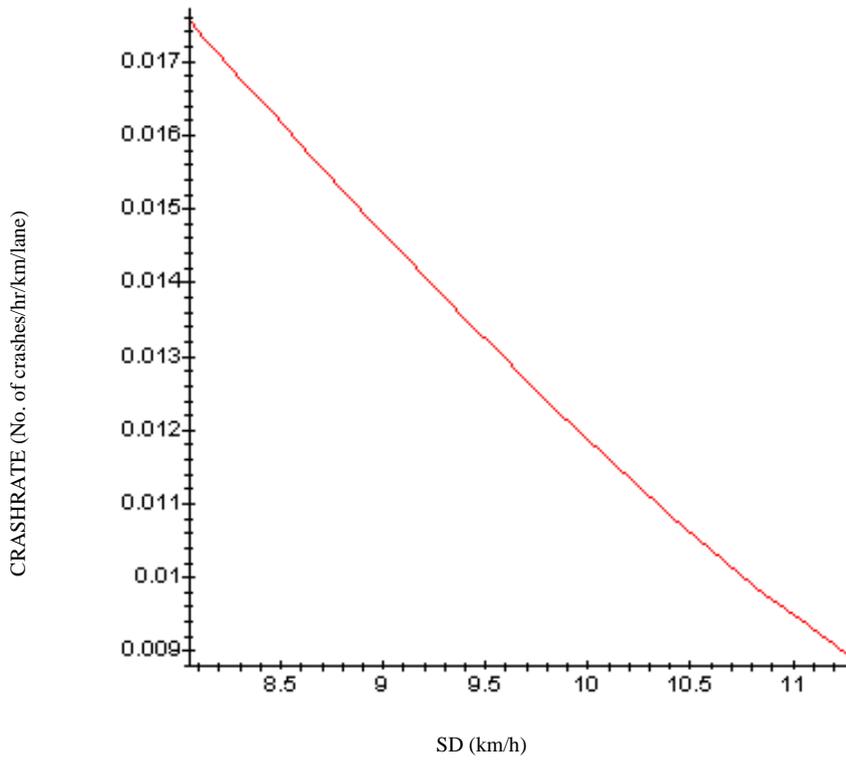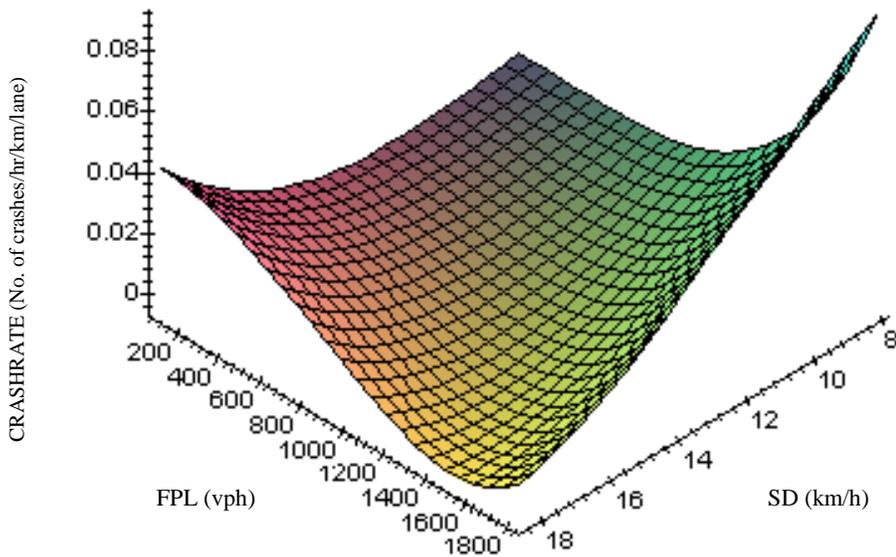**Figure 2. Model 1, CRASHRATE vs. FPL and SD for 104.65 Km/H (65 Mph)
Freeways   (MEAN=104.65 Km/H (65 Mph))**

Figure 1 shows that the crash rate increases as the standard deviation of speed increases up to a standard deviation of 9 km/h (5.6 mph). Figure 2 shows that the crash rate increases as the flow per lane increases until a flow rate of approximately 300 vph is reached, after which the crash rate begins to decrease. Figure 2 also shows that for flows between 200 and 700 vph and standard deviations higher than 9.8 km/h, there is a small decrease in the crash rate as the standard deviation of speed increases.

Figure 3 shows that as the mean speed increases, the crash rate decreases slightly until the mean speed reaches a value that is approximately equal to the speed limit–and then the crash rate begins to increase. Also, the rate of increase of the crash rate is higher as the mean speed increases further away from the speed limit. Figure 4 reveals the same trend as Figure 3 in that the crash rate increases as the mean speed deviates from the posted speed limit. Figure 4 also shows the increase in the crash rate as the standard deviation of speed increases. These results clearly show that changes in crash rates are not necessarily caused by any one independent factor, but that these changes are a result of the combined effects of these independent factors.

**Figure 3. Model 1, CRASHRATE vs. MEAN for 104.65 km/h (65 mph)
Freeways (SD=9.499 km/h (5.9 mph), FPL=790 vph)**
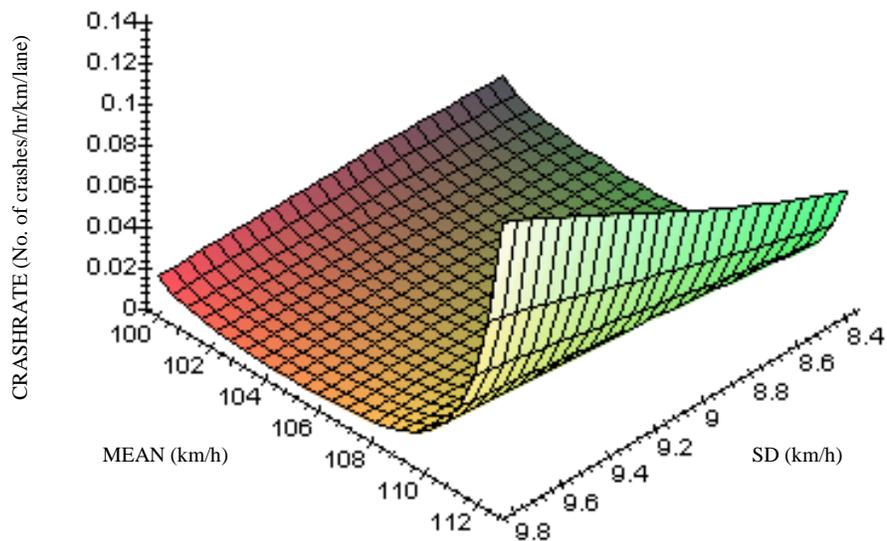
**Figure 4. Model 1, CRASHRATE vs. MEAN and SD for
104.65 km/h (65 mph) Freeways (FPL=790 vph)**

Freeways with 88.55 Km/hr (55 mph) Speed Limit: Model 2

Model 2 uses the variables SD, FPL, and MEAN to predict the CRASHRATE. The estimated model is:

CRASHRATE = $((0.355) - (1.591E\text{-}03)*(SD^2) + (8.651E\text{-}07)*(SD^4) - (2.071E\text{-}08)*(FPL^2) - (1.256E\text{-}10)*(SD^2)*(FPL^2) + (8.527E\text{-}15)*(FPL^4) - (6.509E\text{-}05)*(MEAN^2) + (1.725E\text{-}07)*(SD^2)*(MEAN^2) + (3.143E\text{-}12)*(FPL^2)*(MEAN^2) + (2.875E\text{-}09)*(MEAN^4))$

The $_pR^2$ value for Model 2 is 0.5156 and the AIC is -562.11. Examination of the normal probability and residual plots indicates that the residuals are normally distributed.   Graphical representations of Model 2 are shown in Figures 5, 6, 7, 8, and 9.

SD (km/h)

**Figure 5. Model 2, CRASHRATE vs. SD for 88.55 km/h (55 mph)
Freeways (MEAN=96.6 km/h (60 mph), FPL=680 vph)**

**Figure 6. Model 2: CRASHRATE vs. SD for 88.55 km/h (55 mph)
Freeways (MEAN=91.77 km/h (57 mph), FPL=1150 vph)**

**Figure 7. Model 2, CRASHRATE vs. FPL and SD for 88.55 km/h (55 mph)
Freeways (MEAN=88.55 km/h (55 mph))**

Figure 5 shows that the crash rate increases as the standard deviation of speed increases to about 9.6 km/h at a low value of the flow per lane. Figure 6 shows that at a relatively high flow per lane, the crash rate decreases as the standard deviation of speed increases. Figure 7 shows the same trends as Figures 5 and 6 regarding the resultant effect on the crash rate of the interaction of the flow per lane and the standard deviation of speed. Figure 7 also shows that the crash rate decreases as the flow per lane increases until a flow rate between 1,200 and 1,400 vph is reached, after which the crash rate begins to increase.
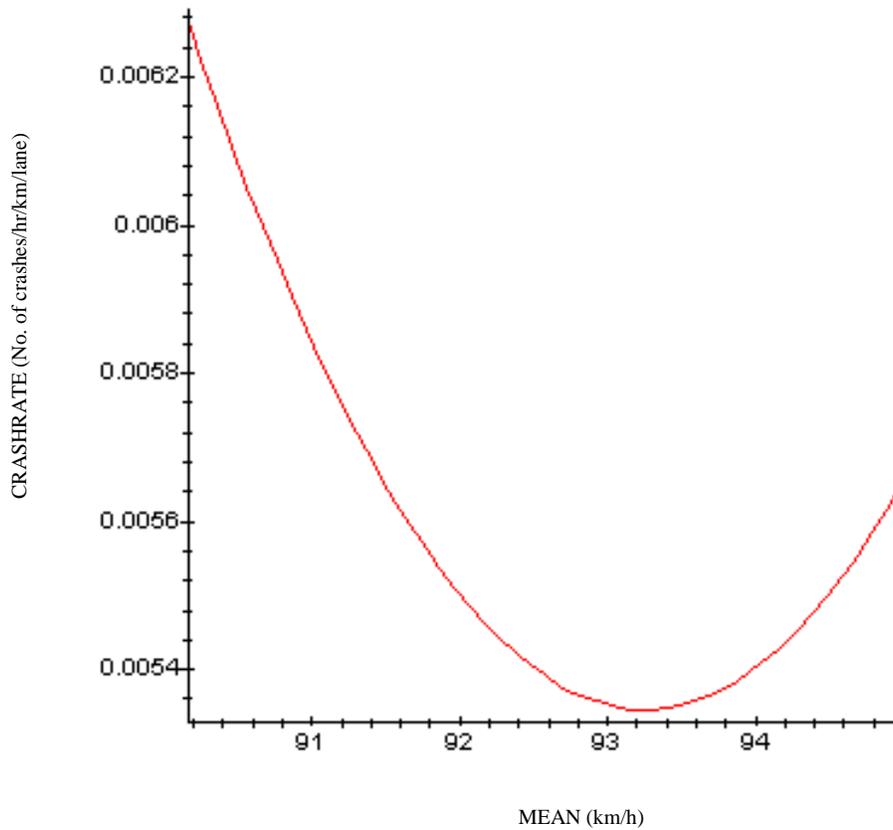
**Figure 8. Model 2, CRASHRATE vs. MEAN for 88.55 km/h (55 mph)
Freeways (SD=9.34 km/h (5.8 mph), FPL=125 vph)**
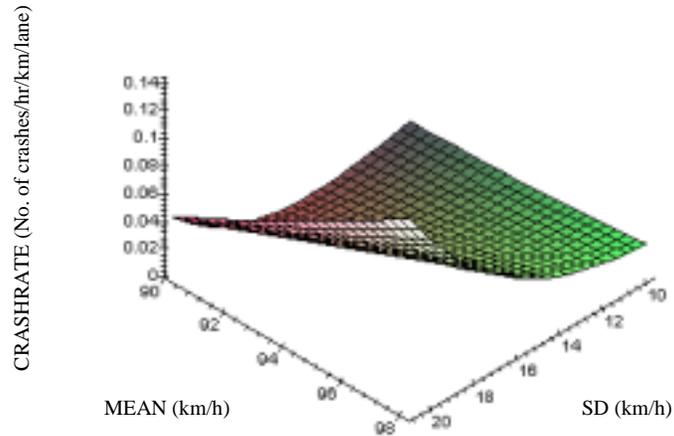
**Figure 9. Model 2, CRASHRATE vs. MEAN and SD for
88.55 km/h (55 mph) Freeways (FPL=1185 vph)**

Figure 8 shows that the crash rate decreases as the mean speed increases to a speed of approximately 93.4 km/h (58 mph), and then the crash rate begins to increase. Figure 9 shows that the crash rate increases as the standard deviation of speed increases. Again, these results show that the manner in which the crash rate on an interstate changes depends on the combined effects of the independent factors considered, and not on a single factor.

4-Lane Non–Freeways: Model 3

Model 3 uses the variables SD, FPL, MEAN, and SW to predict CRASHRATE; only the data collected at sites with a lane width greater than 3.4 meters (11 feet) are included.

The estimated model is:

$LN(CRASHRATE) = ((-48.034)+(2123.118)*(1/(SD^2))-(21753.58)*(1/(SD^2))\verb|^|2 +(1.045E-02)*(LN(FPL))-(82.175)*(1/(SD^2))*(LN(FPL))+(0.539)*(LN(FPL))\verb|^|2+(388332.4)*(1/(MEAN^2))-(9795407)*(1/(SD^2))*(1/(MEAN^2))-(23324.54)*(LN(FPL))*(1/(MEAN^2))-(4.011E+08)*(1/(MEAN^2))\verb|^|2+ (2.197)*(SW^2)-(21.263)*(1/(SD^2))*(SW^2)-(0.109)*(LN(FPL))* (SW^2)-(12767.63)*(1/(MEAN^2))*(SW^2)+(2.816E-03)*(SW^4))$

The $_pR^2$ value for Model 3 is 0.7609 and the AIC is –62.819. Examination of the normal probability and residual plots indicates that the residuals are normally distributed. Graphical representations of Model 3 are shown in Figures 10, 11, 12, 13 and 14.
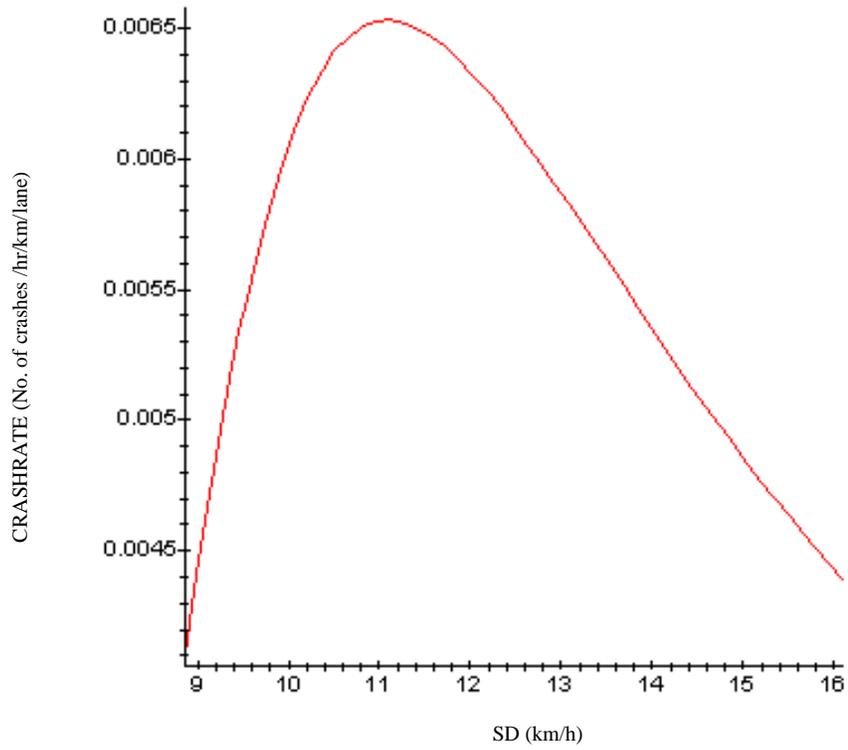
Figure 10. Model 3, CRASHRATE vs. SD for 88.6 km/h (55 mph) 4-Lane Non-Freeways (MEAN=88.6 km/h (55 mph), FPL=120 vph, SW=2.4 meters (8 feet))
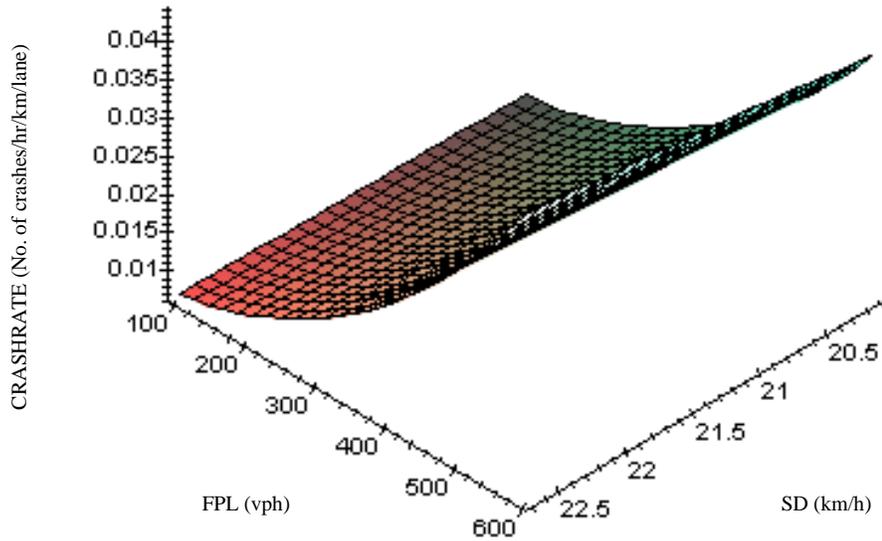

Figure 11. Model 3, CRASHRATE vs. FPL and SD for 88.6 km/h (55 mph) 4-Lane Non-Freeways (MEAN=78.9 km/h (49 mph), SW=3.1 meters (10 feet))
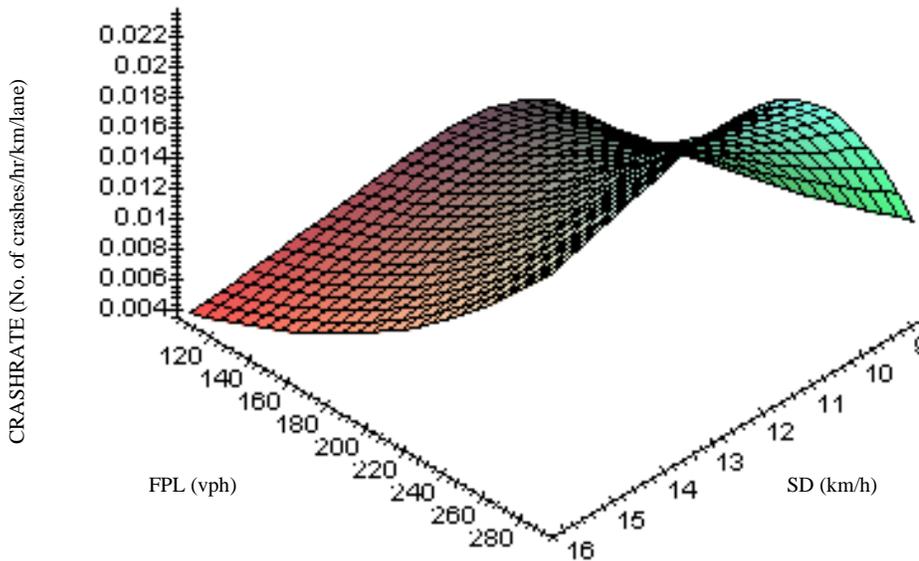
Figure 12. Model 3, CRASHRATE vs. FPL and SD for 88.6 km/h (55 mph) 4-Lane Non-Freeways (MEAN=88.6 km/h (55 mph), SW=2.4 meters (8 feet))

Figure 10 shows that the crash rate increases with the standard deviation of speed–until a standard deviation of speed of about 11.3 km/h (7 mph) occurs, after which the crash rate begins to decrease. Figure 11 shows a large range of the flow rate over a short range of the standard deviation of speed, while Figure 12 shows the opposite. Figure 11 shows that overall, the crash rate increases with an increase in the flow rate. Figure 12 combines the trends of Figures 10 and 11. It shows that the crash rate increases as the flow per lane increases. Figure 12 also shows the crash rate increases as the standard deviation of speed increases up to a value of 11.3 km/h (7 mph), after which the crash rate begins to decrease.

Figure 13 shows that the crash rate decreases as the mean speed increases at a low flow rate and at a moderate standard deviation of speed. Figure 14 shows the same trend of a decreasing crash rate with an increasing mean speed; however, the effect is not as great at low standard deviations of speed. Figure 14 also shows that at standard deviations of speed higher than about 11 km/hr, the crash rate decreases as the standard deviation of speed increases; however, this effect is not as great at low mean speeds. These results also show that the crash rate is dependent on the combined effects of the independent variables used.
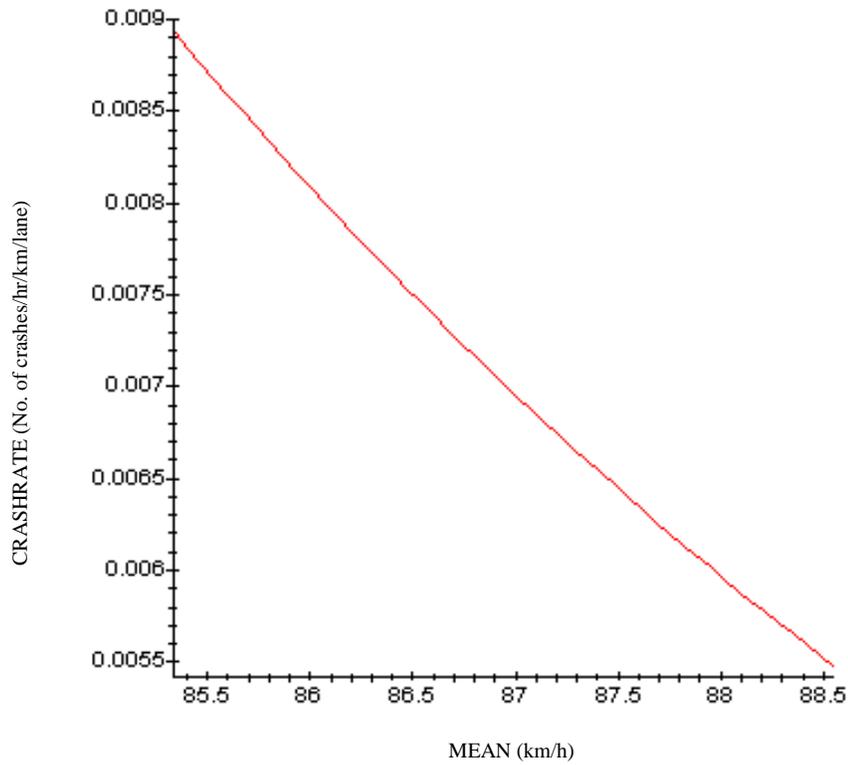
Figure 13.   Model 3, CRASHRATE vs. MEAN for 88.6 km/h (55 mph) 4-Lane Non-Freeways (SD=15.3 km/h (9.5 mph), FPL=135 vph, SW=2.4 meters (8 feet))
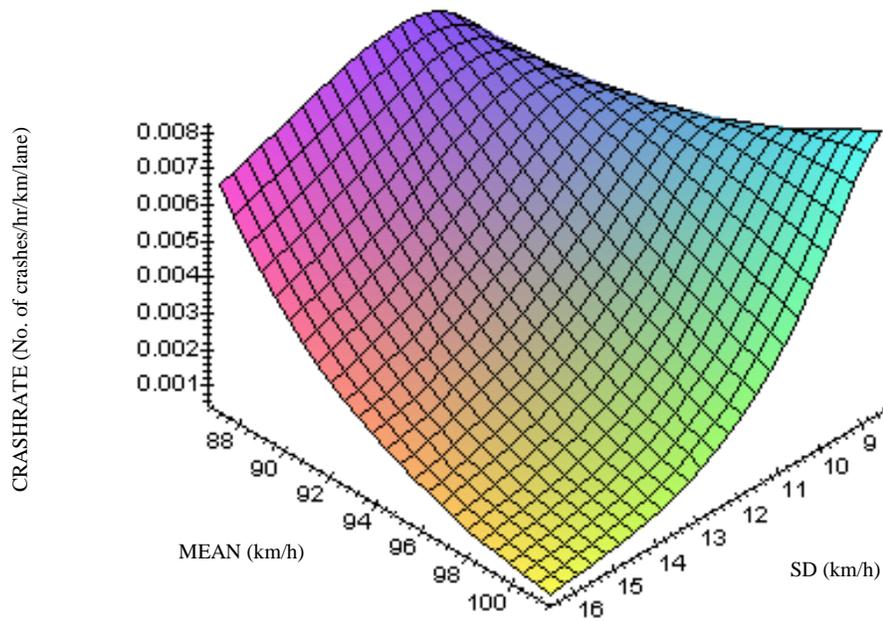

Figure 14. Model 3, CRASHRATE vs. MEAN and SD for 88.6 km/h (55 mph) 4-Lane Non-Freeways (FPL=135 vph, SW=2.4 meters (8 feet))

2-Lane Non-Freeways

Model 4 uses the variables SD, FPL, LW, and SW to predict CRASHRATE. Only data collected at sites with grades less than 4 percent are included in the analysis.   The estimated model is:

LN(CRASHRATE) = $((44.323)-(25755.82)*(1/(SD^2))+(93793.11)* (1/(SD^2))$^2-(8.686E-03)$*(FPL^2)+(0.106)*(1/(SD^2))*(FPL^2)-(1.687E-08)*(FPL^4)+(469.071)*(1/SQRT(LW))+(44529.25)*(1/(SD^2))*(1/SQRT(LW))+ (1.445E-02)*(FPL^2)*(1/SQRT(LW))-(956.114)* (1/SQRT(LW))$^2-(93.415)*SW-$(660.808)*(1/(SD^2))*SW+(5.626E-05)* (FPL^2)*SW+(152.084)*(1/SQRT(LW))*SW+(3.475)*SW^2)$

The $_pR^2$ value for Model 4 is 0.9864 and the AIC is -48.715. Examination of the normal probability and residual plots indicates that the residuals are normally distributed. Graphical representations of Model 4 are shown in Figures 15 and 16.
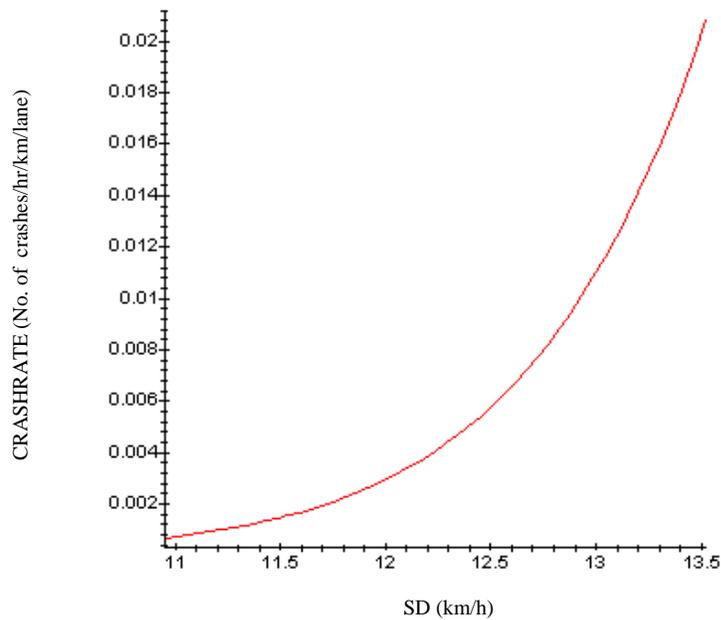

Figure 15. Model 4, CRASHRATE vs. SD for 88.6 km/h (55 mph) 2-Lane Non-Freeways (FPL=40 vph, LW=3.4 meters (11 feet), SW=1.8 meters (6 feet))
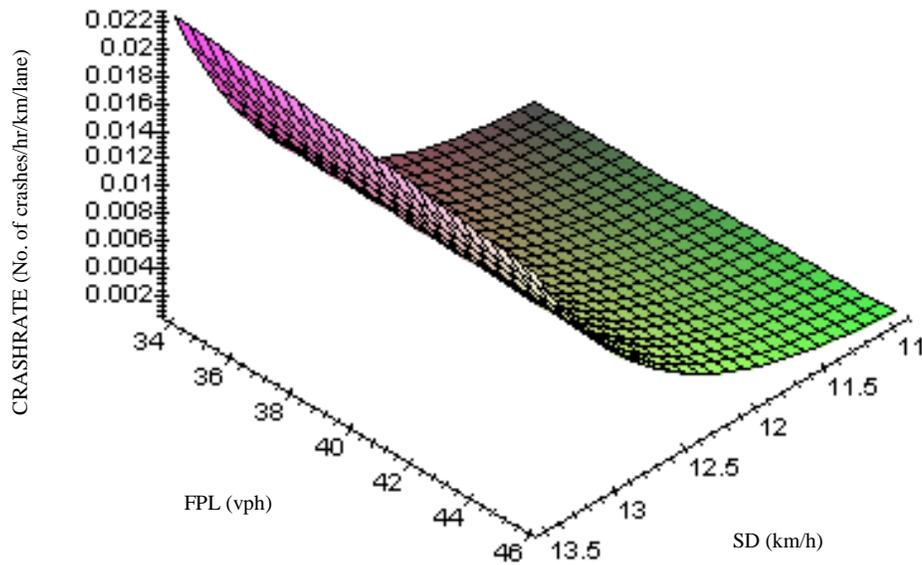
Figure 16. Model 4, CRASHRATE vs. FPL and SD for 88.6 km/h (55 mph) 2-Lane Non-Freeways (LW=3.4 meters (11 feet), SW=1.8 meters (6 feet))

Figure 15 shows that the crash rate increases as the standard deviation of speed increases. Figure 16 shows the same trend as Figure 15 and shows that the crash rate decreases slightly as the flow per lane increases.

Summary of Results

In an attempt to establish mathematical models to describe the influence of speed, flow, and geometric characteristics on crash characteristics, three types of models were explored. The linear and robust regression models were found unsuitable for describing any relationship between speed and crash characteristics. The multivariate ratio of polynomials type of model, however, was successful in describing such relationships. Although these mathematical models cannot be used to predict a numerical value for the crash rate outside the range that was considered in the study, these models show that there is a relationship between crash rates and the independent variables of standard deviation of speed, mean speed, and flow per lane. These models also show that the crash rate is not solely dependent on any one of the independent variables, but on a complex interaction of these independent variables.

Freeways with 105 Km/hr (65 mph) Speed Limit: Model 1

The three models that were developed for these roads exhibit significant similarities in the relationship between the crash rate and the standard deviation of speed, the flow per lane, and the mean speed. These models indicate the following:

- An increase in the crash rate with an increase in the standard deviation of speed up to a value of about 9 km/h (5.6 mph);

- An increase in the crash rate as the flow per lane increases to a value of approximately 300 vph;

- An increase in the crash rate as the mean speed deviates from the posted speed limit. The crash rates are higher when the mean speed is less than the posted speed limit. The crash rates decrease to a minimum when the mean speed is approximately equal to the posted speed limit–then crash rates continue to increase significantly as the mean speed increases to above the posted speed limit. This confirms results of earlier research.[7]

Freeways with 89 km/h (55 mph) Speed Limit: (Model 2)

The three models that were developed for these roads also exhibit significant similarities in the relationship between the crash rate and the standard deviation of speed, the flow per lane, and the mean speed. These models indicate the following:

- An increase in the crash rate with an increase in the standard deviation of speed for a flow rate less than 1000 vph and a decrease for a flow rate greater than 1100 vph;

- For a given standard deviation of speed, the crash rate decreases as the flow per lane increases to approximately 1200 vph, after which the crash rate begins to increase with the flow rate.

Four-Lane Non-Freeways (Model 3)

The thirteen models developed for these roads also exhibit significant similarities in the relationship between the crash rate and the standard deviation of speed and the flow per lane. These models indicate the following:

- An increase in the crash rate as the standard deviation of speed increases to a value of about 11 km/h (7 mph), after which the crash rate decreases with an increasing standard deviation of speed.

- All models show an increase in the crash rate as the flow per lane increases.

Two-Lane Non-Freeways (Model 4)

The four models developed for these roads also exhibit significant similarities in the relationship between the crash rate and the standard deviation of speed and the flow per lane. These models indicate the following:

- An increase in the crash rate as the standard deviation of speed increases;

- A slight decrease in the crash rate as the flow per lane increases, although this effect is not very significant.

DIRECTION FOR FUTURE RESEARCH

Four areas of this research have been identified as needing further investigation. They are:

1. *Model types*–This study only looked at deterministic models. Crashes, however, do not necessarily occur under specific conditions; rather, they occur at random. Since this research has shown that combined changes in speed, flow, and geometric characteristics significantly affect crash rates, future efforts should be devoted to stochastic models. These stochastic models can then be used to predict the changes in the probability of a crash occurring as a result of changes in speed, flow and geometric characteristics.

2. *Data collection sites*–The data used in this research were collected only in the state of Virginia. It therefore cannot be concluded that the same results would be obtained if the data were collected elsewhere. Data should be collected at various locations encompassing different weather, driver, and terrain characteristics to ensure that the results are not location-specific. The data were also collected at sites previously established for speed monitoring for submittal to the Federal Highway Administration to certify enforcement of the National Maximum Speed Limit. Some of these sites may have specific characteristics that could have influenced the results. The use of sites selected randomly may contribute to eliminating any biases in the data. This elimination of bias is, however, extremely difficult to achieve due to the possible lack of the necessary data at sites which are selected randomly.

3. *Data reduction*–The results show that crash rates may not be linear with respect to changes in the flow per lane. Therefore, it may be more appropriate to develop volume categories that are not evenly spaced as was done in this study, but rather use the results of the trends developed to define the volume categories.

4 *Model Validation*–It will be useful for these models to be validated using different data sets to see whether the trends obtained are similar to those obtained in this study.

CONCLUSIONS

- The results of the study show that in estimating the effect of changes in traffic and geometric characteristics on crash rates, the type of road should be taken into consideration.

- The results also show that crash rates are not dependent solely on any single factor, but on the complex interaction of all the independent variables considered. For example, on interstate highways, the crash rate may increase or decrease as the standard deviation of speed increases, depending on the flow rate. Therefore, any model that relates the crash rate to only one of these independent variables will not be capable of fully explaining the effect of changes in the independent variable on the crash rate. This may be the reason for the inconsistencies of the results in different studies as noted in the literature survey.

- Although the models developed in this study show the effect of combined changes in mean speed, standard deviation of speed, flow per lane, and in some cases shoulder width on the crash rate, these relationships are very complicated. Also, these relationships cannot be directly used to predict the accident rate values on roads outside Virginia, because of the limitation of the data. However, these relationships give a good indication of the combined effect of these independent variables on crash rates.

- Because the results of the study show that the combined effect of the traffic, flow, and geometric characteristics on the crash rate is not the same for different types of roads, it is essential that any future study conducted in this area must be road-type specific.

RECOMMENDATIONS

- The developed models of the multivariate ratio of the polynomial type of model show distinct trends on the resultant effect of changes in the mean speed, standard deviation of speed, flow per lane, and geometric characteristics on the crash rate. These trends can be used to develop countermeasures to control the occurrence of crashes.

- One use of these trends is the development of a set of procedures that can be used by police officers or engineers to control speed and flow characteristics. The procedures could indicate when police officers should lay emphasis on speed control or when traffic should be diverted to or from a highway without increasing the crash rate. For example, the trends on a rural freeway indicate that as the mean speed increases beyond 109 km/h (68 mph), and flows are less than 400 vph per lane, the crash rate increases. These conditions may be used to determine when speed limits should be strictly enforced.

- On urban freeways, the trends indicate that the crash rate is lowest when flows are between 1000 and 1200 vph. Under these conditions, traffic could be diverted to or from the freeway to keep the flow within this safe range.

- On 2-lane non-freeways, where the trend shows that the crash rate increases with an overall increase in the standard deviation, having officers enforce minimum and maximum speed limits could increase safety.

ACKNOWLEDGEMENTS

REFERENCES

1. Rural and Urban Crashes: A Comparative Analysis. August 1996. *U.S. Department of Transportation Technical Report.* DOT HS 808 450.

2. Pfundt K. 1969. Three Difficulties in the Comparison of Accident Rates. *Accident Analysis and Prevention.* 1, pp. 253-259.

3. Gwynn, David W. 1967. Relationship of Accident Rates and Accident Involvements with Hourly Volumes. *Traffic Quarterly.* July. Vol. XXI. No. 3,   pp. 407-418.

4. Hakkert, A., I. Hocherman, and A Mensah. 1996. Levels of Safety on Interurban Roads. *Report No. TRR* 1553. National Research Council. Washington, D.C.: Transportation Research Board, pp.95-102.

5. Hall, J. W. and O. J. Pendleton. 1990. Rural Accident Rate Variations with Traffic Volume. *Report No. TRR* 1281. Washington, D.C.: National Research Council. Transportation Research Board. pp. 62-70.

6. Brodsky, H. and A. S. Hakkert. 1983. Highway Accident Rates and Rural Travel Densities. *Accident Analysis and Prevention.* Vol. 15. 1, pp. 73-84.

7. Speed and Accidents. 1970. *Research Triangle Institute.* Research Triangle Institute. Durham, NC: Vol 2. June.

8. Garber, N. J. and Ravi Gadiraju. 1988. Factors Affecting Speed Variance and Its Influence on Accidents. *AAA Foundation for Traffic Safety.* Falls Church, VA: July.

9. Lassarre, S. 1996. The Introduction of the Variables Traffic Volume, Speed, and Belt-Wearing into a Predictive Model of the Severity of Accidents. *Accident Analysis and Prevention.* Vol. 18. No. 2. pp. 129-134.

10. Lundy, R. 1965. Effect of Traffic Volumes and Number of Lanes on Freeway Accident Rates. *Highway Research Record 99*. HRR. Washington, D.C.: National Research Council. pp. 138-156.

11. Traffic Control and Roadway Elements—Their Relationship to Highway Safety. 1963. *Automotive Safety Foundation.*

12. Accident Rates as Related to Design Elements of Rural Highways. 1968. *National Cooperative Highway Research Program Report 47*. HRB. Washington, D.C.: National Research Council.

13. Choueiri, E. M., R. Lamm, J. H. Kloeckner, and T. Mailaender. 1994. Safety Aspects of Individual Design Elements and Their Interactions on Two-Lane Highways: International Perspective Report No. TRR 1445. Washington, D.C.: National Research Council. Transportation Research Board. pp. 34-46.

14. Raff, M. 1953. Interstate Highway Accident Study. *Highway Research Board Bulletin 74*. HRB. Washington, D.C.: National Research Council. pp. 18-43.

15. Perkins, E. 1957. Relationship of Accident Rate to Highway Shoulder Width. *Highway Research Board Bulletin 151*. HRB. Washington, D.C.: National Research Council. pp. 13-14.

16. Cost and Safety Effectiveness of Highway Design Elements. 1978. *National Cooperative Highway Research Program Report 197*. Washington, D.C. National Research Council. Transportation Research Board.

17. Dart, Jr., O., and L. Mann, Jr. 1970. Relationship of Rural Highway Geometry to Accident Rates in Louisiana. *Highway Research Record 312*. Washington, D.C.: National Research Council. pp. 1-16.

18. Garber, Nicholas J., Surbhi Patel, and Ravi Kalaputapu. 1993. The Effect of Trailer Width and Length on Large-Truck Accidents. *Virginia Transportation Research Council.* VTRC Report No. 93-R11. Charlottesville, VA.: March.

19. Cirillo, J., S. Dietz, and R. Beatty. 1969. Analysis and Modeling of Relationships Between Accidents and the Geometric Traffic Characteristics of the Interstate System. *Bureau of Public Roads.* August.

20. Zegeer, C., R. Stewart, F. Council, and T. Neuman. 1994. Accident Relationships of Roadway Width on Low-Volume Roads. Report No. TRR 1445. Washington, D.C. National Research Council. Transportation Research Board, pp. 160-168.

21. *SAS/ETS User's Guide,* Version 6. 1993. Second Edition. Cary, NC.