

Paper No. 02-3790

Analysis of Crash Precursors on Instrumented Freeways

Chris Lee

Department of Civil Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Tel: (519) 888-4567 ext. 6596
Fax: (519) 888-6197
E-mail: chclee@uwaterloo.ca

Frank Saccomanno

Department of Civil Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Tel: (519) 888-4567 ext. 2631
Fax: (519) 888-6197
E-mail: saccoman@uwaterloo.ca

Bruce Hellinga

Department of Civil Engineering
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Tel: (519) 888-4567 ext. 2630
Fax: (519) 888-6197
E-mail: bhellinga@uwaterloo.ca

Words: 5,619 + 6 * 250 = 7,119 words

Paper submitted for the publication
in the Transportation Research Record 2002

Analysis of Crash Precursors on Instrumented Freeways

CHRIS LEE, FRANK SACCOMANNO, AND BRUCE HELLINGA

Department of Civil Engineering, University of Waterloo

Waterloo, Ontario, N2L 3G1, Canada.

This study examines a number of traffic flow characteristics that lead to crashes on urban freeways. Since these characteristics are observed prior to crash occurrence, they are referred to as "crash precursors". The objectives of the study are: 1) to explore factors contributing to changes in crash rate for individual vehicles travelling over an urban freeway; and 2) to develop a probabilistic model relating significant crash precursors to changes in crash potential. The data used to examine crash precursors were extracted from 38 loop detector stations on a 10-km stretch of the Gardiner Expressway in Toronto for a 13-month period. An aggregate log-linear model was developed relating crash rates to the selected crash precursors observed upstream of the crash site. The results of this analysis suggest that the variation of speed and traffic density are statistically significant predictors of crash frequency after controlling for road geometry, weather and time of day. Using the model, the study shows how crash potential can be established based on the precursors obtained from real-time traffic data.

Over the past few decades, many local governments have implemented Instrumented Freeway Traffic Management Systems (IFTMS) in order to reduce freeway crashes. IFTMS allow traffic engineers to monitor real-time traffic flow data, provide information to motorists, and in some cases control the rate at which vehicles enter the freeway. A typical IFTMS consists of road sensors, surveillance cameras, variable message signs, automatic incident detection algorithms, and ramp meters.

A number of researchers have explored the effectiveness of IFTMS in crash reduction. Olmstead (1) found that the IFTMS in Phoenix, Arizona reduced the frequency of crashes by 21~30%. A report from Minnesota Department of Transportation (2) revealed that urban freeway accidents could be reduced by 30%, implying that approximately 900 accidents per year would be prevented by deploying automated freeway management. The case studies of the IFTMS in some European countries (3) showed that the use of advanced technologies such as variable speed limit, lane control signals and automatic queue detection can be very effective in decreasing overall accident rates.

Within an IFTMS, several control policies are performed to minimize traffic disruptions. Among them, timely detection of unexpected incidents – e.g. crashes, stopped cars, debris on the road – and the prompt implementation of incident clearance have been the key components of IFTMS in terms of improving safety. Clearly, this control policy helps reduce the delay and minimize the damages incurred by incidents. However, there has been little work done on how to use an IFTMS as a preemptive warning system rather than just a reactive system. The goal of a reactive system is to minimize the effects of an incident after it has occurred. The goal of a preemptive system is to identify, in real time, traffic conditions known to be associated with high crash frequency, and to intervene in such a way (e.g. variable speed limit) as to modify traffic conditions. In fact, the aforementioned case studies did not explicitly explain how a preemptive warning system could contribute to crash reduction. However, in order to examine the effectiveness of a preemptive warning system in crash reduction, we need to be able to predict crash potential on a real-time basis and identify the important conditions that lead to crashes before crashes actually occur.

To predict crash potential, we need to develop freeway crash prediction models. However, many existing models cannot account for the impacts of variation in real-time traffic flow characteristics on crash occurrence. These models tend to be based on historical crash occurrence data and are used within the planning context to identify “black spots” for the purpose of identifying roadway segments requiring intervention measures (e.g. rumble strips, median barrier, etc.)

Recent developments in Intelligent Transportation Systems (ITS) require prediction models that use real-time traffic flow data and have forward-looking perspective. They should be able to anticipate crash occurrence from the traffic flow conditions that are present prior to the occurrence of the crash. These traffic conditions are referred to in the literature as “crash precursors”.

The study presented in this paper has two basic objectives: 1) to explore factors contributing to changes in crash potential prior to crashes for individual vehicles travelling over an urban freeway; and 2) to develop a probabilistic model relating significant precursors to changes in crash potential. In this paper, the term “crash potential” refers to long-term likelihood that a crash will occur for given traffic, environment, and roadway conditions.

This paper is organized into six sections: The second section reviews the existing freeway crash prediction models and discusses the fundamental relationship between real-time traffic flow and crashes. The third section introduces the framework of the proposed crash prediction model and explains the variables used in the model. The fourth section describes the data used to evaluate the performance of the model. The fifth section shows the results

of model estimation and discusses the implications of the results. Finally, the conclusions and recommendations are presented.

PAST STUDIES IN ANALYSIS OF FREEWAY CRASHES

Statistical models have been widely used to analyze freeway crashes and to explain the relationship between crash involvement and traffic, geometric and environmental factors. These crash prediction models assumes that the number of crashes or crash rate is a linear function of several traffic environmental characteristics such as traffic flow, weather, road geometry, driving condition, etc. (4,5,6). Consequently, these models used linear regression to calibrate this linear relationship.

However, linear regression assumes that the number of crashes is distributed normally. This assumption is not realistic because it fails to explain the random and discrete crash occurrence on freeways (7). Also, due to this assumption, linear regression cannot account for the non-negative nature of accident occurrence. Clearly, there is no clear reason to believe that the relationship is linear.

To overcome these unrealistic assumptions, Poisson regression models were developed. Poisson regression assumes that the crashes are Poisson distributed with some expected value and the occurrences of crashes are independent to each other. For example, Hamerslag et al. (8) used this assumption to consider the probabilistic property of crash occurrence. Furthermore, Oppe (9) suggested that the additive assumption in a regression is not valid and a multiplicative model is more desirable in case that the effects of factors on crashes are mutually independent. In multiplicative models, the factors are converted to logarithmic terms and the relationship is expressed in a linear function of independent variables. This type of model is also called "log-linear model". A number of studies have analyzed the crashes using a Poisson log-linear model (10, 11).

However, these models mainly used a static measure of traffic flow data such as average annual daily traffic (AADT). Thus, they cannot consider the short-term variation of traffic flow in daily traffic and are not suitable for application to real-time operation of an IFTMS.

This problem has motivated the use of real-time traffic flow data in the analysis of freeway crashes. The real-time data can be effectively used to control vehicular flow, and furthermore reduce the crash potential. In an IFTMS, the data are normally obtained from road sensors such as inductive loop detectors which collect traffic flow data at

relatively short time intervals. In this regard, a number of researchers have examined the fundamental relationship between traffic flow characteristics and crashes using real-time data. Hughes and Council (12) suggested that “traffic flow changes” are more important than static measures of flow in the analysis of safety, and the variability in vehicle speeds was found to be the only measure affecting crashes. Corby and Saccomanno (13) concluded that the changes in speed provide the best indicator of flow disruption caused by crashes and this information could be used in early detection of crashes. Feng (14) drew a similar conclusion after reviewing some of the past studies on the relationship between speed and safety using real-time data. In order to improve safety, he suggested that speed variance should be reduced to encourage stable flow. Clearly, their findings imply that capturing the variation of operating speed on a real-time basis is important to predicting likelihood of crashes and analyzing the cause of crashes.

So far, much of the work relating real-time traffic flow characteristics to its effectiveness in reducing freeway crashes has been descriptive in nature. There has been a little work done on the development of statistical models which describe this relationship quantitatively. One study that has presented statistical links between real-time traffic conditions and crashes was carried out by Oh et al. (15), in which lower overall speed variations were found to reduce the likelihood of freeway crashes. Although their findings are based on very limited number of samples, they demonstrated the potential capability of identifying the statistical relationship.

MODEL FRAMEWORK

This study expresses the crash frequency as a function of a variety of traffic and environmental characteristics. Based on previous studies (4-15), the important crash precursors and a number of control features were selected as independent variables. In general, the proposed crash prediction model takes the following form:

$$\text{Crash Frequency} = f(\text{variability of speed, density, road geometry, weather, time of day})$$

Using this functional relationship, the model is calibrated using actual crash data and the effect of each variable on crash potential is examined. In the next subsection, each variable is described in detail.

Variables Used in the Model

Crash precursors represent the traffic flow conditions prior to the crash occurrence. This information can be collected from archived loop detector data. In this study, two types of crash precursors were considered: 1) variation of speed and 2) traffic density.

Variation of speed is measured by the coefficient of variation of speed (CVS) computed over an observation period with fixed duration. In this study, two types of CVS are estimated. The first CVS, CVS_1 , measures the average speed variation experienced by drivers on each lane as they travel from upstream to downstream as shown in Equation 1.

$$CVS_1 = \frac{1}{n} \sum_{i=1}^n \frac{(s_s)_i}{\bar{s}_i} = \frac{1}{n} \sum_{i=1}^n \left(\sqrt{\frac{\frac{\Delta t}{p} \sum_{t=t^*-\Delta t}^{t^*} (s_i(t))^2 - \left(\sum_{t=t^*-\Delta t}^{t^*} s_i(t) \right)^2}{\frac{\Delta t}{p} \left(\frac{\Delta t}{p} - 1 \right)}} \right) / \frac{\sum_{t=t^*-\Delta t}^{t^*} s_i(t)}{\left(\frac{\Delta t}{p} \right)} \quad (1)$$

where,

t^* : actual time of crash;

Δt : observation time slice duration (seconds);

$(s_s)_i$: standard deviation of speed on lane i computed over period Δt ;

\bar{s}_i : average speed on lane i computed over period Δt (km/hour);

p : polling interval of loop detectors (seconds);

$s_i(t)$: speed on lane i at time t (km/hour);

n : total number of lanes.

The second measure, CVS_2 , is the average variation of speed difference across lanes as shown in Equation 2:

$$CVS_2 = \frac{s_{\Delta s}}{ave.(\Delta s)} = \sqrt{\frac{\frac{\Delta t}{p} \sum_{t=t^*-\Delta t}^{t^*} \Delta s(t)^2 - \left(\sum_{t=t^*-\Delta t}^{t^*} \Delta s(t) \right)^2}{\frac{\Delta t}{p} \left(\frac{\Delta t}{p} - 1 \right)}} / \frac{\sum_{t=t^*-\Delta t}^{t^*} \Delta s(t)}{\left(\frac{\Delta t}{p} \right)} \quad (2)$$

where,

$$\Delta s(t) : \text{speed difference across lanes at time } t \left(= \frac{1}{n-1} \sum_{i=1}^{n-1} |s_i(t) - s_{i+1}(t)| \right).$$

* Lane i and lane $i+1$ are adjacent lanes.

Traffic density is also an important crash precursor as it represents average spacing or time/distance headway between two vehicles. Unlike speed, the density at the instant the crash occurs is used instead of the variation of density over time because the density just before the crash has a more dominant effect on the crash occurrence. The density (D) measured in vehicles/km is estimated from speed and volume using Equation 3.

$$D = \frac{1}{n} \sum_{i=1}^n \left(\frac{v_i(t^*)}{s_i(t^*)} \right) \quad (3)$$

where,

$v_i(t^*)$: volume on lane i at the time of crash (vehicle/hour);

$s_i(t^*)$: speed on lane i at the time of crash (km/hour).

In addition to these crash precursors, three external control factors are also considered. Road geometry is included because it affects the number of lane changes. These lane changes are mandatory for some vehicles in the following cases: 1) when the vehicles enter from on-ramps and merge with main traffic stream; 2) when the vehicles enter the exclusive diverging lane(s) in order to exit; and 3) when the number of lanes is reduced or increased. Therefore, in order to consider the impact of lane changes, freeway crashes are divided into two categories: 1) crashes that occurred on a roadway section having the same number of lanes at both the upstream and downstream boundaries and not containing any on-ramps or off-ramps; 2) crashes that occurred on a roadway section with changes in the number of lanes or containing on-ramps or off-ramps.

The weather effect reflects the road condition and visibility of drivers. In this study, the weather effects are simply classified into normal and adverse weather conditions. Adverse weather conditions are defined as the traffic conditions when it rains or snows – road surface is wet or icy, and drivers generally have poor visibility.

Time of day reflects the congested traffic conditions that are experienced during commuting peak periods. As sum, time of day is classified in peak and off peak. The peak period can be determined from the distribution of hourly traffic volume such that the hourly traffic volumes are relatively higher than volumes in other time periods.

Crash Prediction Model

With these variables, a crash prediction model was developed using a log-linear analysis. The log-linear model allows us to investigate the nature of the relationship between selected precursors and frequency of crashes adjusted by the appropriate level of exposure. The distribution of crash frequency is assumed to follow the Poisson distribution because it is similar to the nature of crashes – discrete and non-negative. All the independent variables except exposure are categorical variables. Exposure is continuous variable and is included as scalar covariate. This approach is aggregate in nature.

Alternatively, binary logit model can be used for the crash prediction. In the model, the dependent variable is either crash or no crash, and all independent variables are continuous. However, there are two problems associated with the use of logit model in real-time crash prediction. One is the over-representation of no-crash data. Since the model uses the data measured in very refined time scale – e.g. 5-minute period, and the chances of crash occurrence within this time period are extremely low in daily traffic, the model results will be dominated by the effect of no-crash data. The second problem is the difficulty with determining exposures for continuous traffic flow measures. In order to include exposure in the model, the traffic flow measures must be classified into a number of categorical variables. For these two reasons, a log-linear model was used instead of binary logit model. In this analysis, a first-order log-linear model of crash prediction was developed as shown in Equation 4.

$$\ln(F) = \mathbf{q} + \mathbf{I}_{CVS_1(i)} + \mathbf{I}_{CVS_2(j)} + \mathbf{I}_{D(k)} + \mathbf{I}_{P(l)} + \mathbf{I}_{R(m)} + \mathbf{I}_{W(n)} + \mathbf{b} \cdot \ln(EXP) \quad (4)$$

where,

F : the expected number of crashes over the analysis time frame;

\mathbf{q} : constant;

$\mathbf{I}_{CVS_1(i)}$: effect of the coefficient of variation of speed along the road section (CVS_1) having i levels;

$I_{CVS_2(j)}$: effect of the coefficient of variation of speed difference across lanes (CVS_2) having j levels;

$I_{D(k)}$: effect of density having k levels;

$I_{P(l)}$: effect of time of day having l levels;

$I_{R(m)}$: effect of road geometry having m levels;

$I_{W(n)}$: effect of weather having n levels;

EXP : the exposure in vehicle-kilometers of travel.

Exposure plays a significant and distinctive role in fitting the proposed crash prediction model. Exposure is expressed in terms of vehicle-kilometers of travel during observation time period. The product of AADT and the length of each road section are split into the volume-kilometers according to the proportions of *the occurrence of independent variables in daily traffic*. For example, if the weather condition was normal when certain crash has occurred, the average likelihood of having normal weather condition over a given observation time period was applied to the exposure for the crash. The exposure is estimated as shown in Equation 5.

$$EXP_a = p(CVS_{1a}) \cdot p(CVS_{2a}) \cdot p(D_a) \cdot p(P_a) \cdot p(W_a) \cdot AADT_a \cdot L_a \cdot T \quad (5)$$

where,

EXP_a : exposure of the condition when crash a has occurred;

$p(CVS_{1a})$: proportion of the level of CVS_1 when crash a has occurred;

$p(CVS_{2a})$: proportion of the level of CVS_2 when crash a has occurred;

$p(D_a)$: proportion of the level of density when crash a has occurred;

$p(P_a)$: proportion of the peak period volume if crash a has occurred during the peak period.

Otherwise, proportion of the off-peak period volume;

$p(W_a)$: proportion of normal weather condition if crash a has occurred in normal weather condition.

Otherwise, proportion of adverse weather condition;

$AADT_a$: average annual daily traffic of the section on which crash a has occurred;

L_a : length of the road section on which crash a has occurred;

T : total observation time period.

APPLICATION OF MODEL

To test the performance of the proposed model, this study used incident logs and traffic flow data extracted from loop detectors along a 10-km stretch of the Gardiner Expressway in Toronto, Canada. A total of 38 loop detector stations were located along this stretch of freeway as shown in Figure 1. Among 38 detectors, 19 of them are located upstream of on-ramps/off-ramps or lane transition zones (lane increase/lane reduction). The remainders were located upstream of straight sections.

These data were collected for weekdays over a 13-month period from January 1998 to January 1999. The reason for excluding weekends is that it is normally believed that the traffic pattern during weekends is quite different from the traffic pattern during weekdays – e.g. commuting trips. The traffic flow data from loop detectors include information on speed, volume and occupancy averaged over 20-second time slices.

The incident logs provide information on lane block pattern, type of incident, date and time of detection and the location with respect to the nearest upstream loop detector. The incidents are first detected by the loop detectors based on a number of incident detection algorithms and subsequently confirmed visually by an operator at the traffic control center through closed-circuit TV camera feeds. Once the incident is confirmed, the operator classifies it into one of 18 types by primary cause, such as stopped car, debris on the road, construction, etc. In this study, we are concerned only with those incidents that can be classified as “reportable crashes”. In Ontario, reportable crashes accompany property damage of more than the threshold monetary value (\$700) or presence of personal injury. In general, reportable crashes represent about 40% of all freeway incidents (13).

A total of 234 crashes were reported to the traffic control center for the 13-month period considered in this analysis. To obtain more in-depth information concerning the causes and nature of these 234 crashes, we obtained supplementary police reports for 45 of these crashes cited in the incident log. Police reports provide information on specific location of each crash, number of vehicles involved, causes, underlying circumstances and nature of injuries sustained by each involved person. The incident log was matched to these reports based on the time and the location of crash occurrence. Although not all the crashes cited in the incident log could be matched to the police reports, we

obtained some very valuable insight as to the most common cause of crash on this freeway, which we could not obtain directly from the incident log. According to police reports, a majority of the crashes (88%) was classified as rear-end crashes, the most common type of crash on freeways. These crashes normally occur when the drivers of the following vehicles failed to respond quickly to avoid the downstream vehicles which were either stopped or slowing down quickly due to unexpected congestion. From this finding, it is evident that many crashes on this freeway are associated with abrupt changes in speed and can be avoided if drivers are warned of an impending queue ahead.

From the distribution of hourly traffic volume of the study area, the peak period was set to 7~10 am in the morning and 4~7 pm in the afternoon. The proportion of daily volume during peak period was found to be 36%. For weather effect, the record of hourly variation of weather (16) at a weather station near the study area was obtained. The record contains the weather condition and temperature at each hour of the day and the precipitation of rainfall and snowfall. During the 13-month study period, 86% of hourly weather conditions were normal (i.e. no precipitation was recorded).

Determination of Actual Time of Crashes

In order to calculate the variation of speed and density, the actual time of crash occurrence needs to be determined first. However, the times noted in the incident logs and police reports reflect the time when the incident was reported or verified and do not correspond to the actual time of the crash. Based on previous work (13), we have assumed that the actual time of crash can be extrapolated from the analysis of changes in detector speed profiles.

For example, as illustrated in Figure 2, Crash 4403 was reported in the incident log as having occurred at 10:21:22 a.m. on July 28, 1998. However, analysis of the speed profile at the upstream detector indicates a sudden drop in the speed across all three lanes beginning at 10:19:20 a.m. We assume that this rapid drop in speed is due to the crash. In this case, we would set the time of crash at 10:19:20 a.m. This type of visual speed profile analysis was carried out for all 234 crashes in our sample to determine the actual time of crash occurrence.

Categorization of Variables

In establishing the contingency table for log-linear analysis, we considered how to establish the most appropriate categories for the factors. In this study, the degrees of speed variation and density were categorized as high,

moderate and low. The observation time slice duration prior to crash occurrence was set to 5 minutes. Since the variation of speed is very low under normal traffic condition, it is more likely to observe low variation than high variation in daily traffic. Assuming that the proportions of three levels of variation that occur in daily traffic are 50% (low), 30% (moderate) and 20% (high), the boundary values for each level were determined from 24-hour speed profiles for a typical day.

For this task, 24-hour speed data were divided into 288 five-minute time periods. The quantity of CVS_1 was calculated for each five-minute period and was categorized into one of three levels for the given boundary values. The boundary values were adjusted such that the proportions of three levels of variation are reasonably close to the aforementioned target proportions. As a result, the boundary values of CVS_1 between low and moderate variation, and moderate and high variation were found to be 0.075 and 0.1, respectively. The boundary values of CVS_2 and density were determined in the same manner. The classification of all categorical variables is summarized in Table 1. Based on this categorization, the frequency of crashes was determined for different combinations of CVS_1 , CVS_2 and density with time of day, weather and road geometry controlled.

To visually analyze the property of CVS_1 and CVS_2 , two measures were calculated and plotted against 5-minute average and lane-by-lane speed data on July 20, 1998 when no crash has occurred. As shown in Figure 3, CVS_1 clearly reflects the sudden change in speed caused by non-recurrent congestion at a given location. As the 5-minute average speed sharply decreases, the CVS_1 correspondingly increases. High CVS_1 can take place at anytime during the day but it is noticeable during the peak period between 6 and 7:30 pm. Similarly, CVS_2 reflects the change in speed difference across lanes as shown in Figure 4. As the fluctuation of speed difference between adjacent lanes increases, CVS_2 increases accordingly.

Determination of Exposure

In determining exposure, the average traffic volume on the roadway section where the crash has occurred should be obtained. In this study, each road section is bounded by a pair of successive detector stations. Since we are more concerned with traffic conditions prior to the crash occurrence, traffic volume at the detector upstream of the crash site was used to establish the AADT. We observed that the AADT on all detector stations is almost uniform during weekdays in the study area. Thus it is reasonable to assume traffic volume on typical weekdays of the year as an estimate of the AADT on each section. Average AADT's for different road sections were obtained and these

AADT's were assumed to be valid for all other days. Once AADT was determined, the exposure was calculated using Equation 5.

RESULTS AND DISCUSSIONS

An iterative proportional fitting process yielded maximum likelihood estimates of the expected frequency of crashes from the log-linear model described in Equation 1. This fitting process continues until the difference between the current and previous estimates converges to a small value – in this case, 0.001.

In order to evaluate the overall goodness-of-fit of the model, a Pearson chi-square test was performed. This test measures how close the expected frequencies are to the observed frequencies for any combination of crash precursors and control features. In terms of overall performance of the model, Pearson chi-square value was found to be low (224 for 205 degrees of freedom, p -value = 0.17) which means the difference between observed and predicted crash frequencies is not significant at a 95% significance level. In other words, this indicates that the model yields a good fit to the data. Also, since the over-dispersion parameter ($= 224/205 = 1.1$) is very close to 1, our assumption of Poisson distribution is valid.

Table 2 summarizes the estimated parameters of the log-linear model. In Table 2, the cell of higher category value for each categorical variable was set as the “aliased” cell. This cell serves as the basis against which log-linear parameters are applied to obtain crash frequency for any combination of precursors. All factors were found to be significant at a 95% confidence level based on their Z -values. The constant term in exponential expression ($= \exp(\mathbf{q})$) gives the crash frequency for base case factors (i.e. aliased cells of the contingency table) for a given exposure. From these results, we observed the followings:

1. As the variation of speed along the section and the variation of speed difference across lanes increase, crashes are more likely to occur. This result reflects the fact that as the variation of speed increases, drivers have to adjust their speed more frequently and they are more likely to make misjudgment in keeping a safe separating distance from other cars.

2. As the density increases, crashes are more likely to occur. This result is reasonable from a theoretical standpoint because as average spacing between vehicles decreases, the chances of crashes will subsequently increase.
3. Crashes are more likely to occur during the peak period than off-peak period. This is due to heavy traffic congestion during peak period, which induces more fluctuation of traffic flow.
4. In the road sections with more frequent lane changes (i.e. merging and diverging sections), crashes are more likely to occur, compared to the straight sections. As vehicles are merging or diverging, the drivers have to change their lanes more frequently leading to increased risk of collision with other vehicles.
5. In normal weather condition, crashes are more likely to occur. This result is inconsistent with our expectation. Two possible reasons for this result are: 1) The drivers traveling on this road generally pay less attention to their driving under ideal weather conditions whereas they tend to be more cautious in adverse weather conditions; 2) Some of the weather data used in the model may not represent real weather conditions at the time of crashes. Since weather data are available on an hourly basis, the data cannot capture the short-term variation of weather condition within one-hour period. This might have caused the overestimation of exposure measure under adverse weather condition.
6. As the exposure increases, crashes are more likely to occur. In agreement with our prior expectation, the total number of crashes within a given category of crash precursors increases with vehicle-kilometers of travel. Also, since the estimated coefficient (\mathbf{b}) is less than 1, the relationship between crash frequency and exposure is non-linear. This result is reasonable because in reality, as traffic volume approaches to capacity, the rate of increase in crash frequency gradually decreases rather than remains constant.

This model can be used to predict crash potential based on the current traffic flow conditions. For example, if the variation of speed (CVS_1 , CVS_2) is observed to be high over the last 5-minute period, the density is high, the current time is during peak period, the road section is a merge section, and the exposure is 1,000,000 vehicles-km of

travels over a 13-month period, the expected frequency of crashes for these base case factors can be estimated as follows:

$$F = \exp(\mathbf{q} + \mathbf{I}_{CVS_1=3} + \mathbf{I}_{CVS_2=3} + \mathbf{I}_{D=3} + \mathbf{I}_{P=1} + \mathbf{I}_{W=1} + \mathbf{I}_{R=1} + \mathbf{b} \cdot \ln(EXP))$$

$$= \exp(1.3386 + 0 + 0 + 0 + 0 + 0 + 0 + 0.2870 \cdot \ln(1)) = 3.8 \text{ crashes}$$

Once the frequency is estimated, the crash rate is obtained as follows:

$$\text{Crash Rate} = \frac{F}{EXP} = \frac{3.8 \text{ crashes/13-month}}{1 \times 10^6 \text{ veh} \cdot \text{km/13-month}} = 3.8 \times 10^{-6} \text{ crashes/veh-km}$$

Since crash rates represent the probability of observing crashes per unit vehicle-kilometer, the crash rates can be viewed as an objective measure of crash potential. From above expression, it is evident that as the variation of speed and the density increases, the crash potential also increases with all other external factors controlled.

CONCLUSIONS AND RECOMMENDATIONS

This paper has investigated the potential for crashes along a stretch of freeway as the function of real-time traffic flow characteristics called crash precursors. From the results of statistical analysis, we observed that the use of real-time traffic data to predict crash potential on freeways is promising. In agreement with previous studies, high variation of speed and high density appear to be the most hazardous traffic condition leading to crashes. The results also indicate that external factors such as road geometry, weather and time of day must be controlled to examine pure effects of crash precursors on crash potential.

The log-linear model developed in this study displays a number of advantages. First, the model is useful in observing general trends of crash potential in the function of selected categorical factors. Second, this model is able to account for exposure as a scalar covariate in the expression. Third, although we have presented a very simple formulation (single order factors), the model is also able to account for a number of higher order interaction terms in

explaining crash frequency. However, we were able to establish a parsimonious model that clearly relates crash potential to variation of traffic flow in real-time.

This study provides meaningful insights into the effects of real-time traffic factors, obtained prior to crash occurrence, to provide an estimate of crash potential over time at different location. Loop detector data which are not readily available elsewhere provides an opportunity to assess this relationship over time.

In future work, the model could be applied to driver warning systems which influence drivers' behaviour to reduce the variation of speed. The warning information to drivers is based on the crash potential estimated by the model on a real-time basis. For example, if the current traffic condition is hazardous, the model will predict high crash potential and the warning information, e.g. maintain the speed at certain level, will be displayed to alert drivers a high risk of crashes. Assuming that all drivers follow the warning information, the traffic condition will be stabilized later. Then the model will predict low crash potential and the warning information will be withdrawn. However, the detailed strategy of reducing speed variance in real traffic situation is beyond the scope of this paper. Also, to improve the performance of the model, it is recommended that additional "potential" precursors such as the presence of a queue, the direction of shock wave (queue formation/dissipation) be investigated to reflect changes in traffic flow under congested traffic condition.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science and Engineering Research Council (NSERC), Canada. The authors are indebted to Ms. Lisa Maasland and Mr. Jim Smith at Metropolitan Toronto Transportation for providing the data used herein.

REFERENCES

1. Olmstead, T. Freeway Management Systems and Motor Vehicle Crashes: A Case Study of Phoenix, Arizona. *Accident Analysis and Prevention* 33, 2001, pp. 433-447.
2. Preston, H. Potential Safety Benefits of Intelligent Transportation System (ITS) Technologies. Semisequicentennial Transportation Conference Proceedings, Center for Transportation Research and Education, Iowa State University, 1996.
3. Innovative Traffic Control: Technology and Practice in Europe. Federal Highway Administration, U.S. Department of Transportation, August 1999.
4. Lundy, R. A. Effect of Traffic Volumes and Number of Lanes on Freeway Accident Rates. *Highway Research Record* 99, 1965, pp. 139-156.
5. Ivey, D. L., L. I. Griffin III, T. M. Newton, R. L. Lytton and K. C. Hankins. Predicting Wet Weather Accidents. *Accident Analysis and Prevention*. Vol .13, 1981, pp. 83-99.
6. Liu, G. X. and A. Popoff. Provincial-Wide Travel Speed and Traffic Safety Study in Saskatchewan. In *Transportation Research Record 1595*, TRB, National Research Council, Washington, D.C., 1997, pp 1-8.
7. Miaou, S. P. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions. *Accident Analysis and Prevention* Vol. 26 (4), 1994, pp. 471-482.
8. Hamerslag, R., J. P. Roos and M. Kwakernaak. Analysis of Accidents in Traffic Situations By Means of Multiproportional Weighted Poisson Model. In *Transportation Research Record 567*, TRB, National Research Council, Washington, D.C., 1976, pp 29-36.
9. Oppe, S. The Use of Multiplicative Models for Analysis of Road Safety Data. *Accident Analysis and Prevention*. Vol .11, 1978, pp. 101-115.
10. Chira-Chavala, T. and D. E. Cleveland. Causal Analysis of Accident Involvements for the Nation's Large Trucks and Combination Vehicles. In *Transportation Research Record 1047*, TRB, National Research Council, Washington, D.C., 1986, pp. 56-64.
11. Saccomanno, F. F. and C. Buyco. Generalized Loglinear Models of Truck Accident Rates. In *Transportation Research Record 1172*, TRB, National Research Council, Washington, D.C., 1987, pp. 23-31.

12. Hughes, R. and F. Council. On Establishing the Relationship(s) between Freeway Safety and Peak Period Operations: Performance Measurement and Methodological Considerations. Presented at *the 78th Annual Meeting of Transportation Research Board*, Washington, D.C., 1999.
13. Corby, M. J. and F. F. Saccomanno. Analysis of Freeway Accident Detection. In *Transportation Research Record 1603*, TRB, National Research Council, Washington, D.C., 1998, pp. 80-89.
14. Feng, C. Synthesis of Studies on Speed and Safety. Presented at *the 80th Annual Meeting of Transportation Research Board*, Washington, D.C., 2001.
15. Oh, C., J. Oh, S. Ritchie and M. Chang. Real-Time Estimation of Freeway Accident Likelihood. Presented at *the 80th Annual Meeting of Transportation Research Board*, Washington, D.C., 2001.
16. Monthly Meteorological Summary. Atmospheric Environment Service, Environment Canada, 1999.

LIST OF TABLES AND FIGURES

TABLE 1. Classification of Categorical Variables

TABLE 2. Estimated Parameters of Log-linear Model

FIGURE 1. Schematic drawing of Gardiner Expressway and number of crashes

FIGURE 2. Determination of actual time of crash occurrence based on detector data

FIGURE 3. CVS₁ at Station dw0080dwg (July 20, 1998)

FIGURE 4. CVS₂ at Station dw0080dwg (July 20, 1998)

TABLE 1 Classification of Categorical Variables

Variables	Category Name	Category Value	Criteria
Longitudinal Variation of Speed (CVS_1)	Low	1	$CVS_1 \leq 0.075$
	Moderate	2	$0.075 < CVS_1 \leq 0.1$
	High	3	$CVS_1 > 0.1$
Lateral Variation of Speed (CVS_2)	Low	1	$CVS_2 \leq 0.35$
	Moderate	2	$0.35 < CVS_2 \leq 0.5$
	High	3	$CVS_2 > 0.5$
Density (D)	Low	1	$D \leq 15$
	Moderate	2	$15 < D \leq 25$
	High	3	$D > 25$
Time of Day (P)	Off-Peak Period	0	Time = 0-6 am, 10 am-4 pm, 7pm-12am
	Peak Period	1	Time = 7-10 am or 4-7 pm
Road Geometry (R)	Straight Section	0	No on-ramp/off-ramp, No change in lanes
	Merge/Diverge	1	On-ramp or off-ramp, change in lanes
Weather (W)	Normal	0	Clear
	Adverse	1	Rain, Snow

TABLE 2 Estimated Parameters of Log-linear Model

Parameters	Estimate	Standard Error*	Z-Value**
Constant (q)	1.3386	.1985	6.74
$I_{CVS_1=1}$ ($CVS_1 \leq 0.075$)	-1.4307	.1903	-7.38
$I_{CVS_1=2}$ ($0.075 < CVS_1 \leq 0.1$)	-1.2750	.1826	-6.98
$I_{CVS_1=3}$ ($CVS_1 > 0.1$)***	0	.	.
$I_{CVS_2=1}$ ($CVS_2 \leq 0.35$)	-1.8417	.2300	-8.01
$I_{CVS_2=2}$ ($0.35 < CVS_2 \leq 0.5$)	-.8271	.1508	-5.48
$I_{CVS_2=3}$ ($CVS_2 > 0.5$)***	0	.	.
$I_{D=1}$ ($D \leq 15$)	-1.3911	.2097	-6.63
$I_{D=2}$ ($15 < D \leq 25$)	-.4289	.1498	-2.86
$I_{D=3}$ ($D > 25$)***	0	.	.
$I_{P=0}$ ($P = 0$)	-.6049	.1437	-4.21
$I_{P=1}$ ($P = 1$)***	0	.	.
$I_{W=0}$ ($W = 0$)	.9648	.1765	5.47
$I_{W=1}$ ($W = 1$)***	0	.	.
$I_{R=0}$ ($R = 0$)	-.5271	.1356	-3.89
$I_{R=1}$ ($R = 1$)***	0	.	.
b (Exposure)	.2870	.1869	8.26

* A measure of the dispersion of the coefficient.

** A standardized measure of the parameter coefficient.

*** Aliased cells.

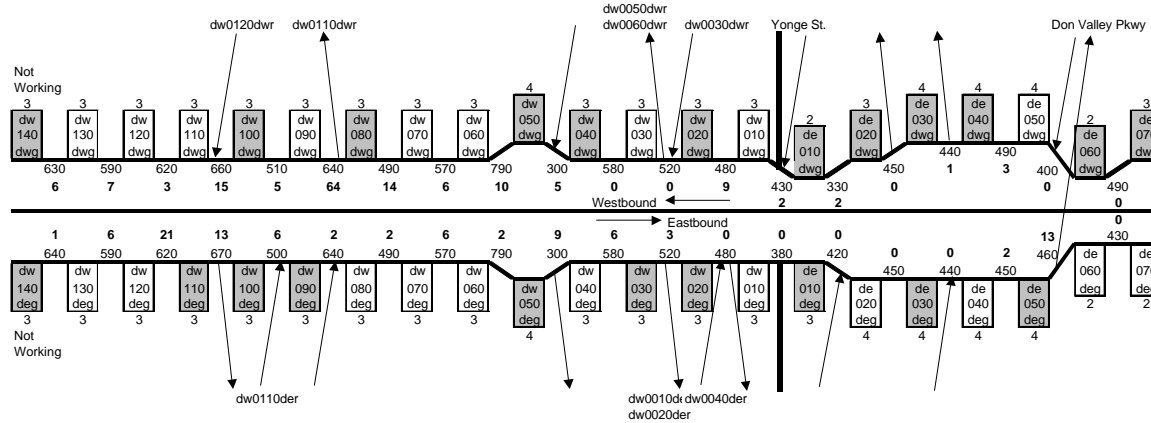


FIGURE 1 Schematic drawing of Gardiner Expressway and number of crashes

Note:

1. The arrows pointing outward indicate off-ramps and the arrows pointing inward indicate on-ramps.
2. The letters inside the squares denote the station ID.
3. The numbers shown above or below the station ID are the number of lanes.
4. The numbers shown between two successive detectors are the distance in meter.
5. Shaded detector stations are the stations where traffic is influenced by merging or diverging vehicles.
6. The bold numbers shown above or below the distance are total number of crashes in 13 months.

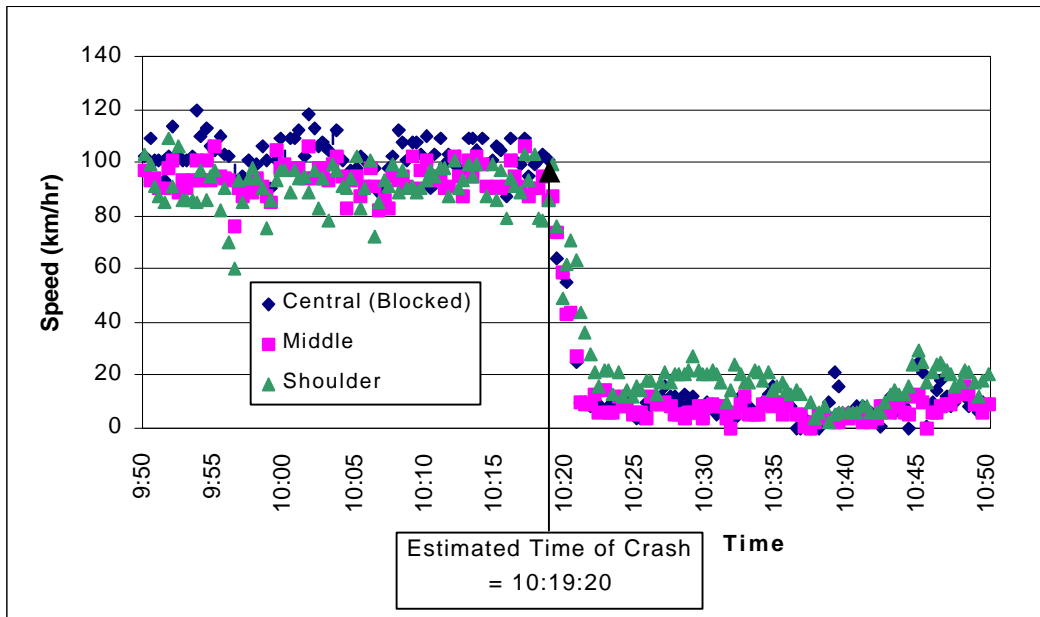


FIGURE 2 Determination of actual time of crash occurrence based on detector data.

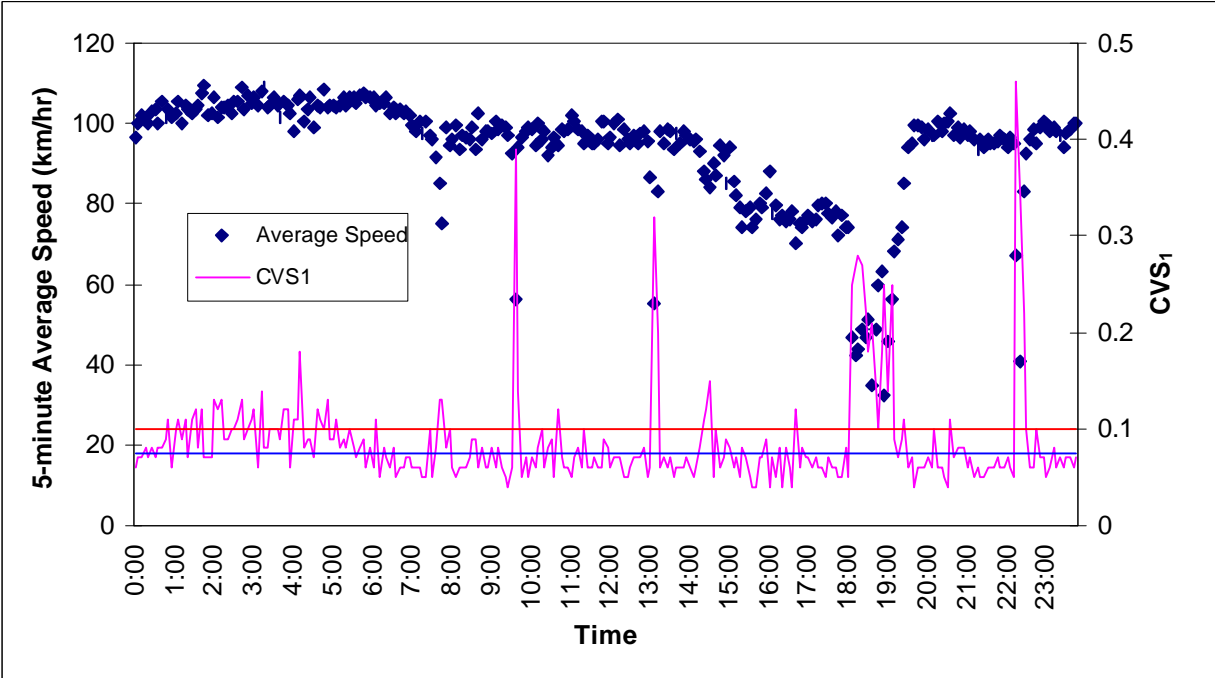


FIGURE 3 CVS₁ at Station dw0080dwg (July 20, 1998).

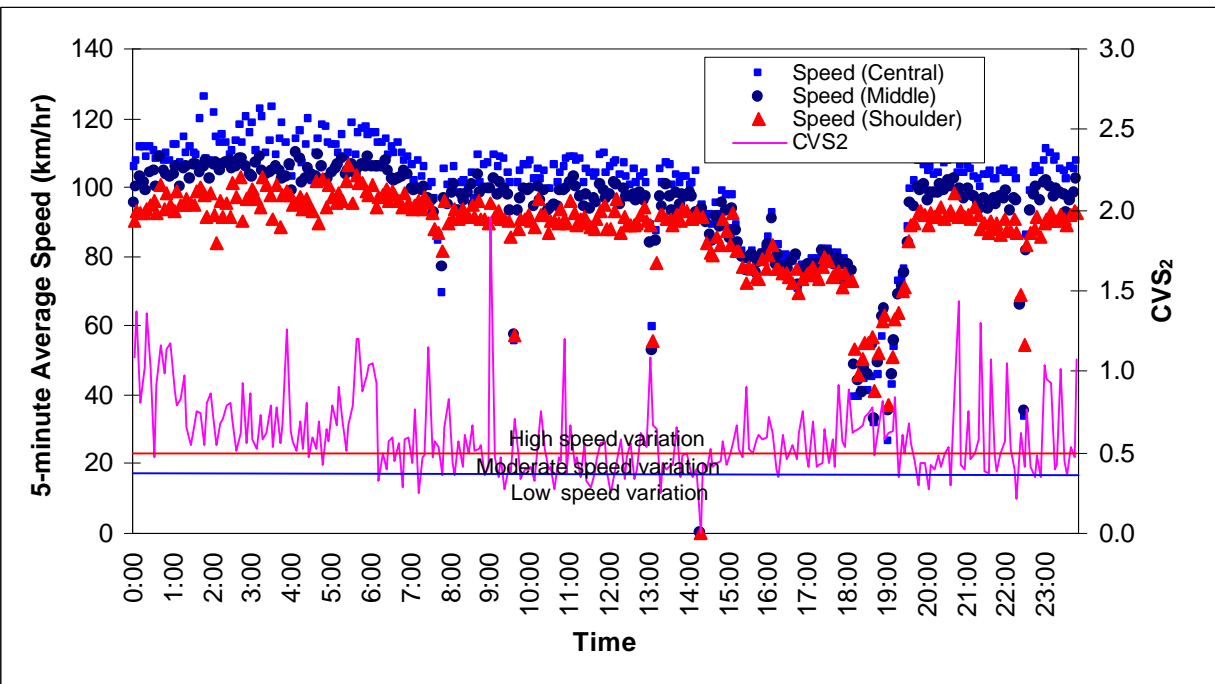


FIGURE 4 CVS₂ at Station dw0080dwg (July 20, 1998).