

# Assessment of Interaction of Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather, and Traffic Data

Mohamed M. Ahmed, Mohamed Abdel-Aty, and Rongjie Yu

This study investigated the effect of the interaction between roadway geometric features and real-time weather and traffic data on the occurrence of crashes on a mountainous freeway. The Bayesian logistic regression technique was used to link a total of 301 crash occurrences on I-70 in Colorado with the space mean speed collected in real time from an automatic vehicle identification (AVI) system and real-time weather and roadway geometry data. The results suggested that the inclusion of roadway geometrics and real-time weather with data from an AVI system in the context of active traffic management systems was essential, in particular with roadway sections characterized by mountainous terrain and adverse weather. The modeling results showed that the geometric factors were significant in the dry and the snowy seasons and that the likelihood of a crash could double during the snowy season because of the interaction between the pavement condition and steep grades. The 6-min average speed at the crash segment during the 6 to 12 min before the crash and the visibility 1 h before the crash were found to be significant during the dry season, whereas the logarithms of the coefficient of variation in speed at the crash segment during the 6 to 12 min before the crash, the visibility 1 h before the crash, as well as the precipitation 10 min before the crash were found to be significant during the snowy season. The results from the two models suggest that different active traffic management strategies should be in place during these two distinct seasons.

Traffic detection systems are essential components of any successful intelligent transportation system, and a wider range of vehicle detection devices than ever before is in use on highways, ranging from the popular inductive loops and magnetometers to video- and radar-based detectors. Advances in electronics have had a tremendous impact on enhancement and improvement of detection systems, and new nonintrusive traffic detection devices are more in use these days because of their ease of installation and maintenance, in addition to their accuracy and affordability. One of these nonintrusive detection systems is the automatic vehicle identification (AVI)

system, which is mainly used for toll collection and to estimate travel times along freeways. Traffic data collected from different detection systems, such as inductive loop detectors, were proven by several studies to be useful for real-time safety risk assessment (1–10), and one study by the authors investigated the usefulness of the traffic data collected from AVI systems in real-time safety assessment (11).

The Colorado Department of Transportation developed the COTrip system to provide travelers with important information about travel time, congestion, adverse weather conditions, and lane closures due to occasional avalanche danger, maintenance on the road, or road crashes. This information is provided as a part of an intelligent transportation system and can be accessed through a website. In addition, the real-time information is dynamically disseminated to road users via dynamic message signs. This system estimates the travel time on a segment by monitoring the successive passage times of vehicles equipped with electronic tags at designated locations. AVIs measure space mean speed, which is the harmonic mean of the speed of all vehicles occupying a given stretch of the road over some specified time period.

In previous studies, weather data were estimated from reports for crash cases and from the weather stations of airports in the vicinity of the freeway section for noncrash cases (12, 13). None of these studies, however, had access to actual weather information on the roadway section itself. In this study, real-time weather data were gathered by weather stations installed on the roadway solely for the purpose of collection of real-time information about adverse weather conditions.

Moreover, roadway geometrics were considered in a few studies (6, 14), and their effects were controlled for by a matched case-control framework in other studies (2–9, 11, 12). These studies were mostly conducted on freeways or expressways that feature normal roadway geometries, and therefore, the traffic flow parameters were found to be the most dominant factors that contributed to crash occurrence. Because the roadway section under study features mountainous terrain with relatively steep grades and sharp horizontal curves, the geometric characteristics were considered to examine how the interaction between all these factors contributes to crash occurrence.

This paper investigates the identification of freeway locations with high crash potential with traffic data collected from an AVI system, real-time weather information, and data on geometric features.

---

Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816-2450. Corresponding author: M. M. Ahmed, mahmed@knights.ucf.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2280, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 51–59.  
DOI: 10.3141/2280-06

## BACKGROUND

The safety performance of a transportation facility can be assessed by analysis of crash data, which is one of the most frequently used tools (15). Crash performance functions were conventionally used to establish relationships between traffic characteristics, roadway and environmental conditions, driver behavior, and crash occurrence. Although these models are useful to some extent, the aggregated nature of traffic parameters is not able to identify locations with a high probability of crashes in real time.

Real-time crash analysis captured the researchers' interest in the past few years because it has the ability to identify crashes in real time and thus be more proactive rather than reactive for safety management.

Oh et al. were the first to link real-time traffic conditions and crashes statistically (1). A Bayesian model with traffic data containing average and standard deviation flow, occupancy, and speed for 10-s intervals was used. It was concluded that the 5-min standard deviation of speed contributes the most to the differentiation of precrash and noncrash conditions. Although their sample size of 53 crashes was small, they showed the potential ability to establish a statistical relationship. Moreover, the practical application of their findings is questionable, since the 5 min before a crash is not an adequate time to take any remedial actions.

Lee et al. used the log-linear approach to model traffic conditions leading to crashes, referred to as "precursor conditions," and they added a spatial dimension by using data from upstream and downstream detectors of crashes (16). Moreover, they used the speed profile captured by the detectors to estimate the actual crash time instead of the reported crash time. They refined their analysis in a later study; the coefficient of temporal variation in speed was found to have a relatively longer-term effect on crash potential than density, and the effect of the average variation in speed across adjacent lanes was found to be insignificant (17).

Hourdos et al. developed an online crash-prone condition model using information for 110 live crashes, crash-related traffic events, and other contributing factors visualized from a video traffic surveillance system (e.g., individual vehicle speeds and headways) over each lane in different locations in the study area (10). They were able to detect 58% of the crashes successfully with a 6.8 false decision rate (that is, 6.8% of the crash cases were detected as noncrash cases).

Abdel-Aty et al. used a matched case-control study to link real-time traffic flow variables collected by loop detectors and the likelihood of a crash (2). The matched case-control study was selected because it has the ability to eliminate the influence of location (i.e., roadway geometry), time, and weather condition. Their model achieved a crash identification rate of more than 69%.

Abdel-Aty and Pande were able to capture 70% of the crashes using the Bayesian classifier-based methodology, a probabilistic neural network, and different parameters of speed only (3). They found that the likelihood of a crash was significantly affected by the logarithms of the coefficients of variation in speed at the nearest crash station and two stations immediately preceding it in the upstream direction measured in 5-min time slices 10 to 15 min before the time of the crash.

Abdel-Aty and Pemmanaboina used principal component analysis and logistic regression to estimate a weather model that determines a rain index on the basis of the rain readings at the weather station in the proximity of the I-4 corridor in Orlando, Florida (13). The archived rain index was used with real-time traffic loop data to model the crash potential by use of a matched case-control logit model. They concluded that the 5-min average occupancy and standard deviation of volume observed at the downstream station and

the 5-min coefficient of variation in speed at the station closest to the crash, all during the 5 to 10 min before the crash, along with the rain index, were the most significant factors to affect crash occurrence.

Ahmed and Abdel-Aty for the first time used data collected from an AVI system in a real-time traffic safety analysis and found that AVI system data are promising as a means to provide a measure of crash risk in real time (11). They used data collected from an AVI system for 78 mi on the central Florida expressway network in Orlando in 2008 and historical crash data obtained for the same period and study area. They concluded that the logarithm of the coefficient of variation in speed at the crash segment during the 5 to 10 min before the crash is the most significant factor affecting the likelihood of a crash on a freeway with tag readers spaced 1 mi, on average, and mostly commuting drivers, whereas the standard deviation of the speed at the crash segment and the average speed at the adjacent downstream segment were found to be the most significant on another freeway section with an average AVI system segment length of 1.5 mi and mixed types of road users.

According to FHWA, weather contributed to more than 22% of the total crashes in 2001 (18). This finding means that adverse weather can easily increase the likelihood of crash occurrence. Several studies, in fact, concluded that during rainfall crashes increase by 100% or more (19, 20), whereas others have found more moderate (but still statistically significant) increases (21, 22).

AVI systems have been widely used for real-time travel time estimation (23, 24). Although one study by the authors used traffic data from an AVI system in a real-time traffic safety application (11), in this study data from an AVI system, real-time weather data, and roadway geometry were used to assess the safety risk on a freeway section that features mountainous terrain.

## DATA PREPARATION

This study used four sets of data: roadway geometry data, crash data, and the corresponding AVI system and weather data. The crash data were obtained from the Colorado Department of Transportation for a 15-mi segment on I-70 for 3 years (2007 to 2009). Traffic data consisted of space mean speed captured by 20 AVI detectors located in both the eastbound and westbound directions along I-70. The processed 2-min space mean speed and the estimated average travel time for each AVI system segment were obtained from the Colorado Department of Transportation. Although the tag readers have the ability to collect data lane by lane, the processed and archived data from the AVI system included only the combined travel time and space mean speed for all lanes. An advanced traveler information system was developed and implemented without consideration of safety applications. The Colorado Department of Transportation also provided weather data recorded by three automated weather stations along I-70 for the same time period. The roadway data were collected from the roadway characteristics inventory and single line diagrams.

AVI system data corresponding to each crash case were extracted by the following process: the location and time of occurrence of each of the 301 crashes were identified. Because the space mean speeds were archived at 2-min intervals, the speeds were aggregated to different aggregation levels of 2, 4, and 6 min to obtain averages and standard deviations and to investigate the best aggregation level that gives a better accuracy in the modeling part of the study. The 6-min aggregation level was found to provide a better fit. Three time slices of the 6 min before the crash time were extracted; for example, if a crash happened on September 16, 2007 (Sunday), at 14:00 and Mile-

post 205.42, data for the corresponding 18-min window from 13:42 to 14:00 recorded by AVI system Segment 34 were used for this crash (the mile marker starts at Milepost 200.8 and ends at Milepost 205.55). Time Slice 1 was discarded in the analysis because it would not provide enough time for a successful intervention to be made to reduce the risk of a crash in a proactive safety management strategy. Moreover, the actual crash time might not be known precisely. Golob and Recker discarded the 2.5 min of traffic data immediately preceding each reported crash time to avoid the uncertainty over the actual crash time (25). In general, with the proliferation of mobile phones and closed-circuit television cameras on freeways, the crash time is usually almost immediately identified.

One-hour speed profiles (about 30 min before and 30 min after the crash time) were also generated to verify the reported crash time. The modeling procedure also required noncrash data. A random selection of data for no-crash cases was also collected from the remaining set of data from the AVI system. These data were extracted for situations in which no crash occurred 2 h before the extraction time and were used to determine the different traffic patterns, weather conditions, and roadway characteristics.

Similarly, weather data for crash cases and noncrash cases were extracted. Automated weather stations continuously monitor the weather conditions, and the weather parameters are recorded when a specific change in the reading threshold occurs and are therefore not recorded according to a specific time pattern. The stations therefore frequently provide readings when the weather conditions change within a short time; if the weather conditions remain the same, the station does not update the readings. However, these readings were aggregated over certain time periods to represent the weather conditions, for example, precipitation, described by rainfall

amount or snowfall liquid equivalent for 10 min and 1, 3, 6, 12, and 24 h, and the estimated average hourly visibility, which provides an hourly measure of the clear distance (in miles) that drivers can see. Visibility in general can be described as the maximum distance (in miles) that an object can be clearly perceived against the background sky. Visibility impairment can be the result of both natural activities (e.g., fog, mist, haze, snow, rain, and windblown dust) and human-induced activities (transportation, agricultural activities, and fuel combustion). The automated weather stations do not directly measure the visibility but rather calculate it from a measurement of light extinction, which includes the scattering and absorption of light by particles and gases.

Data for 301 crashes and 880 noncrashes were finally considered in the analysis. Of these, 70 and 231 crashes and the randomly selected 256 and 624 noncrashes occurred during the dry and the snowy seasons, respectively.

### Preliminary Analysis and Results

From the preliminary analysis, the environmental conditions were found to have a strong effect on crash occurrence within that section. According to the meteorological data, the study section had two distinct weather seasons: a dry season, which occurred from May through September and which experienced small amounts of rain, and the snowy season, which occurred from October through April. The crash frequencies during the months of the snowy season were found to be more than double the frequencies during the months of the dry season. Figure 1 shows the 3-year aggregated crash frequency by month and weather for the 15-mi freeway section.

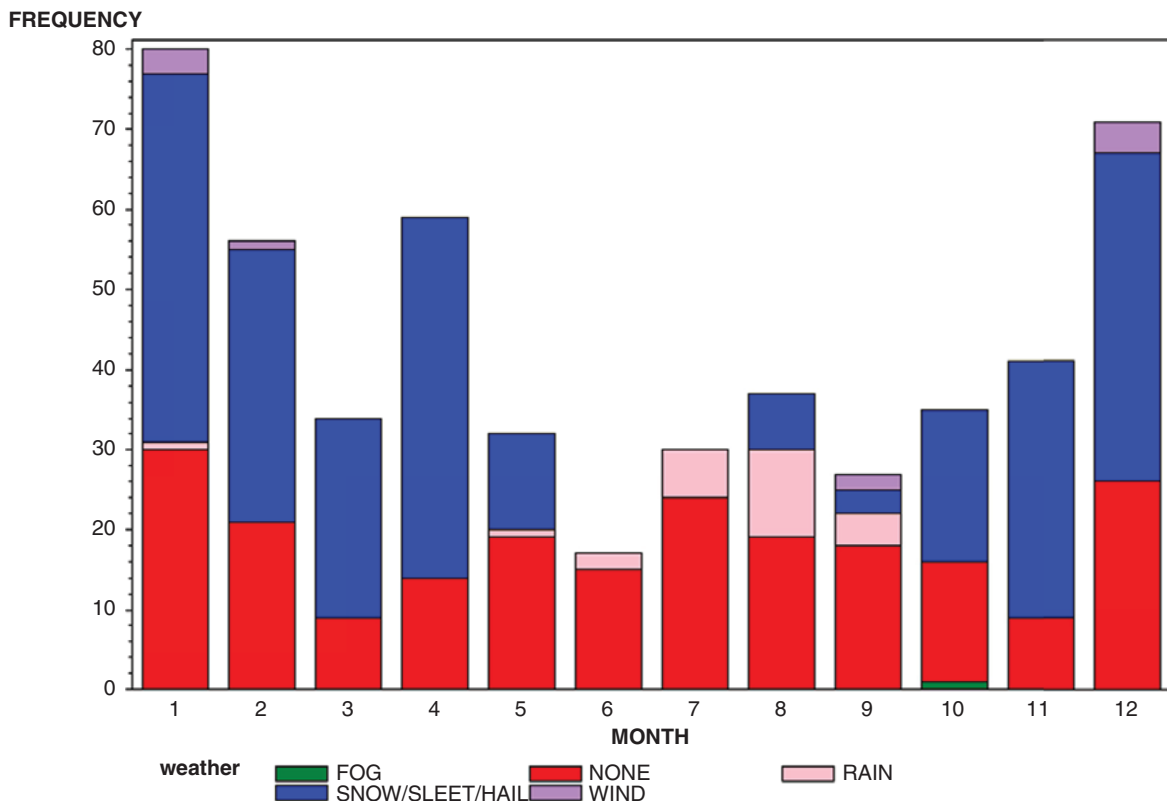


FIGURE 1 Crash frequency by month.

To compare the traffic and environmental factors for crash and noncrash cases as well as between the snowy and dry seasons, a series of statistical tests was conducted. The  $F$ -test showed that the crash and noncrash cases have equal variance, and  $t$ -tests for equal variance were therefore used. The results showed that a significant difference exists between each of the mean of the average speed and the mean of the average visibility 1 h before crash cases and noncrash cases. For example, the 6-min average speed 6 to 12 min before the crash cases for both the snowy and the dry seasons was found to be 48.21 mph, whereas it was found to be 55.71 mph before the noncrash cases ( $t$ -test  $p$ -value =  $6.7 \times 10^{-8}$ ). The mean of the estimated visibilities 1 h before the crash and noncrash cases was found to be significantly higher for noncrash cases than crash cases: the mean estimated visibility for noncrash cases was found to be 1.22 mi, whereas it was found to be 0.95 mi for crash cases. These results indicate that a significant difference between the crash and noncrash cases exists at the 95% confidence level for speed and the different weather-related factors.

Similarly,  $t$ -tests were used to evaluate weather condition factors in different seasons (dry and snowy seasons). The  $t$ -test results showed that the dry season had a higher visibility and significantly lower rate of precipitation. For visibility, the dry season had a visibility of 1.29 mi, whereas the snowy season had a visibility of 1.09 mi; for 10-min precipitation, the dry season had precipitation of only 0.000543 in., whereas the snow season had precipitation of 0.057 in. The average speeds for the different seasons were also compared. The  $t$ -test result shows that during the dry season the average speed is significantly higher than that during the snowy season and has a smaller standard deviation. These observations also suggest that different active traffic management strategies should be implemented for each season.

### Bayesian Logistic Regression

The study used a Bayesian logistic regression approach to estimate the probability of crash occurrence in each of the dry and the snowy seasons. Bayesian logistic regression has the formulation of a logistic equation and can handle both continuous and categorical explanatory variables. The classical logistic regression treats the parameters of the models as fixed, and unknown constants and the data are used solely to obtain a best estimate of the unknown values of the parameters. In the Bayesian approach, the parameters are treated as random variables and the data are used to update beliefs about the behavior of the parameters to assess their distributional properties. The interpretation of Bayesian inference is slightly different from that in classical statistics; the Bayesian inference derives the updated posterior probability of the parameters and constructs credibility intervals that have a natural interpretation according to their probabilities. Moreover, Bayesian inference can effectively avoid the problem of overfitting that occurs when the number of observations is limited and the number of variables is large.

The Bayesian logistic regression models the relationship between the dichotomous response variable (crash versus no crash) and the explanatory variables of roadway geometry, real-time weather, and traffic. Suppose that the response variable  $y$  has the outcomes  $y = 1$  or  $y = 0$  with respective probabilities  $p$  and  $1 - p$ . The logistic regression equation can be expressed as

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta X \quad (1)$$

where

$$\begin{aligned} \beta_0 &= \text{intercept,} \\ \beta &= \text{vector of coefficients for explanatory variables, and} \\ X &= \text{vector of explanatory variables.} \end{aligned}$$

The logit function relates the explanatory variables to the probability of an outcome  $y = 1$ . The expected probability that  $y$  will be equal to 1 for a given value of the vector of explanatory variables  $X$  can be theoretically calculated as

$$p(y = 1) = \frac{\exp(\beta_0 + \beta X)}{1 + \exp(\beta_0 + \beta X)} = \frac{e^{\beta_0 + \beta X}}{1 + e^{\beta_0 + \beta X}} \quad (2)$$

One advantage of the Bayesian approach over the classical model is the applicability of the choice of parametric family for prior probability distributions. Three different priors can be used: (a) informative prior distributions based on the literature, the knowledge of experts, or information explicitly from an earlier data analysis; (b) weak informative priors that do not supply any controversial information but that are strong enough to pull the data away from inappropriate inferences; or (c) uniform priors or noninformative priors that basically allow the information from the likelihood to be interpreted probabilistically. In this study, uniform priors following a normal distribution were used with initial values for estimation of each parameter from the maximum likelihood method. With the results from this study, different types of prior distributions could be considered for use as priors in further research, once more data become available to update the estimated models.

As discussed earlier, Colorado was found to have two distinct weather seasons, and two models for the snowy and dry seasons were therefore considered. These models were estimated by Bayesian inference with the freeware WinBUGS (26). For each model, three chains of 10,000 iterations were set up in WinBUGS on the basis of the convergence speed and the magnitude of the data set. The deviance information criterion, a Bayesian generalization of the Akaike information criterion, is used to measure the model complexity and fit. The deviance information criterion is a combination of the deviance for the model and a penalty for the complexity of the model. The deviance is defined as  $-2\log(\text{likelihood})$ . The effective number of parameters ( $pD$ ) is used as a measure of the complexity of the model,  $pD = Dbar - Dhat$ , where  $Dbar$  is the posterior mean of the deviance and  $Dhat$  is a point estimate of the deviance for the posterior mean of the parameters. The deviance information criterion is given by  $Dhat + 2pD$  (27). Moreover, receiver operating characteristic (ROC) curve analysis was used to assess the prediction performance.

## RESULTS AND DISCUSSION OF RESULTS

### Model 1. Dry Season

The model for the dry season was estimated with real-time weather data, data from the AVI system, and data for the roadway geometry for crashes that occurred from May to September for the years 2007 through 2009 and for randomly selected noncrashes. Before inferences from a posterior sample can be drawn, the trace, autocorrelation, and density plots were examined visually to ensure that the underlying Markov chains had converged. According to the convergence diagnostics of Brooks and Gelman (28), the mixing in the



chains was found to be acceptable, with no correlation for any of the variables included in the final model detected. After the convergence was ensured, the first 2,000 samples were discarded as adaptation and burn-in. Table 1 shows the mean and standard deviation estimates of beta coefficients, credible interval (CI), hazard ratio, and fit statistics for the dry season model.

All roadway alignment factors included, that is, median width, longitudinal grade, and horizontal curve, were found to be significant. Preliminary analysis of the data indicates that more than 85% of the total crashes occurred on steep grades (grade of  $<-2\%$  or  $>2\%$ ). Steep grades affect the operation and the braking of vehicles on both upgrades and downgrades. The results indicate that the likelihood of a crash increases as the grade increases. The effects of various grades were compared with the effect of a flat grade (reference condition; a flat grade ranges from 0% to  $\pm 2\%$ ). The most hazardous grade was the very steep grade ( $>6\%$  to 8% and  $<-6\%$  to  $-8\%$ ), followed by the steep grade ( $>4\%$  to 6% and  $<-4\%$  to  $-6\%$ ) and moderate grade ( $>2\%$  to 4% and  $<-2\%$  to  $-4\%$ ). In general, trends in the results indicate that steeper grades present a higher crash risk.

Table 1 also shows the hazard ratio for the significant variables. The hazard ratio is the exponent of the beta coefficient, and it represents an estimate of the expected change in the risk ratio of having a crash versus a noncrash. The interpretation of the hazard ratio depends on the measurement scale for the explanatory variable; for interval variables, it represents the change in the risk ratio per unit change in the corresponding factor, whereas for categorical variables, it represents the change in the risk ratio compared with the base case; for example, the hazard ratio of 5.63 for the categorical variable very steep grade means that the likelihood of a crash at very steep grades is 5.63 times the likelihood at the base case of flat grades.

A binary variable, the grade index, was created to represent the direction of the grade at the crash segment (1 = upgrade as a reference and 2 = downgrade). The grade index was found to be significant at the 90% CI with a positive coefficient, which implies that roads with positive grades are slightly safer than roads with negative ones. These results are consistent with the finding from the aggregate models in the literature that steep grades may increase the likelihood of a crash occurrence (29–31).

The results imply that the degree of curvature ( $\beta = -0.246$ ; 95% CI =  $-0.484, -0.024$ ; hazard ratio = 0.78) is significantly associated with the risk of a crash. A unit increase in the degree of curvature is associated with a 22% decrease in the likelihood of a crash, with all other factors remaining constant. A high degree of curvature was found to be associated with a decrease in the likelihood of a crash in previous studies, which may be explained by the fact that the feeling of discomfort along sharp curves might make drivers compensate by driving more cautiously, leading to a lower probability of involvement in a crash (29–32). Median width ( $\beta = -0.046$ ; 95% CI =  $-0.075, -0.019$ ) has a negative coefficient, which means that a wider median is safer because it works as a recovery area for out-of-control vehicles.

The 6-min average speed of the crash segment during the 6 to 12 min before a crash and the average visibility during the last hour before the crash were found to be significant during the dry season. Both variables have negative beta coefficients, which means that the odds of a crash increase as the average speed decreases at the segment of the crash 6 to 12 min before the crash and the average visibility decreases during the 1 h before the crash. The hazard ratio of 0.926 means that the risk of a crash increases 7.4% for each unit decrease in the 6-min average speed, and the hazard ratio of 0.211 means that the risk of a crash increases 79% for each unit distance (mile) decrease in the average visibility measured over 1 h before the crash.

TABLE 1 Parameters and Hazard Ratio Estimates: Dry Season Model

Variable	Parameter Estimates				Hazard Ratio			
	$\beta$ Coefficient		CI		$\beta$ Coefficient		CI	
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
Intercept	2.070	1.37	-0.599	4.830	na	na	na	na
Grade (flat, 0%–2%) (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Grade (moderate, >2%–4%)	0.510	0.554	-0.565	1.640	1.950	1.210	0.568	5.150
Grade (steep, >4%–6%)	1.120	0.485	0.201	2.120	3.470	1.860	1.220	8.330
Grade (very steep, >6%–8%)	1.540	0.604	0.373	2.740	5.630	3.840	1.450	15.600
Grade index (1 = upgrade) (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Grade index (2 = downgrade)	0.658	0.354	-0.023	1.350	2.060	0.755	0.977	3.860
Degree of curvature	-0.246	0.116	-0.484	-0.024	0.787	0.091	0.616	0.976
Median width	-0.046	0.014	-0.075	-0.019	0.955	0.014	0.928	0.981
Average speed	-0.076	0.020	-0.115	-0.037	0.926	0.019	0.891	0.964
Visibility	-1.750	0.636	-3.070	-0.568	0.211	0.141	0.046	0.566
pD: number of effective variables	9.803	na	na	na	na	na	na	na
DIC	297.762	na	na	na	na	na	na	na
ROC	0.783	na	na	na	na	na	na	na
Sensitivity	75.71	na	na	na	na	na	na	na

NOTE: SD = standard deviation; DIC = deviance information criterion; na = not applicable. Summary statistics [mean (SD)]: degree of curvature = 1.33 (1.49); median width (ft) = 25.96 (15.11); average speed (mph) = 56.4 (7.94); visibility (mi) = 1.29 (0.95).

## Model 2. Snow Season

Another model was estimated for the crash and noncrash cases during the snowy season to examine whether the same variables in the model for the dry season have the same effect on the likelihood of a crash. Comparisons between the two models present interesting findings. On the one hand, the same geometric variables were significant; on the other hand, it was noticeable that all the coefficients increased because the hazard ratios increased as a result of the interaction between the snowy, icy, or slushy pavement conditions during the snowy season, which were exacerbated by the steep grades.

As shown in Table 2, the hazard ratio for the very steep grade (>6% to 8% and <-6% to -8%) during the snowy season increased to 9.67, whereas it was 5.63 during the dry season. This increase means that the change in the risk ratio almost doubled during the snowy season. Similar findings were obtained for degree of curvature and median width. Another interesting observation from the parameter estimate for grade index was that the hazard ratio decreased and the variable became insignificant, which may indicate that during the snowy season steep grades become hazardous in both the upgrade and the downgrade directions.

Although only the 1-h visibility was significant in the dry season model, in the snowy season model, both 1-h visibility and the 10-min precipitation, described by the amount of rainfall or snowfall liquid equivalent, were significant. These results are consistent with the preliminary analysis that rates of precipitation are significantly higher during the snowy season than during the dry season. A one-unit increase in precipitation increases the risk of a crash by 165%. Moreover, it can be implied from the results that a one-unit decrease in visibility during the snowy season increases the likeli-

hood of a crash by 88%, whereas the increase in likelihood is 79% during the dry season.

The logarithm of the coefficient of variation of the speed at the crash segment at Time Slice 2 (6 to 12 min before the crash) was significant. The logarithm of the coefficient of variation of the speed has a positive beta coefficient, which means that the risk of a crash increases as the variation of the speed increases. The increase in the standard deviation coupled with the decrease in the average speed 6 to 12 min before the crash (since the coefficient of variation of speed includes the standard deviation as the numerator and the average speed as the denominator) may increase the likelihood of occurrence of a crash.

To implement the estimated model in a real-time application, a sensitivity analysis was conducted. Figure 2 and Table 3 show the sensitivity and the specificity for the final models. Sensitivity is the proportion of crashes that were correctly identified as crashes by the estimated Bayesian logistic regression models, whereas specificity is the proportion of noncrashes that were correctly identified as noncrashes by the estimated Bayesian logistic regression models (33). The sensitivities were found to be 75.71% and 80.09% for the dry and snowy season models, respectively, whereas the two models achieved specificities of 66.41% and 67.79% at cutoff points equal to 0.20 and 0.25 for the dry and the snowy seasons, respectively. The cutoff was chosen for each model to reduce the rates of false-positive results (incorrect classification as crashes), which were 33.59% and 32.21% for the dry and snowy season models, respectively.

Different classification accuracies can be obtained by use of a change in the threshold, depending on the management strategy. The threshold selected for application should be chosen carefully because large numbers of false alarms might affect drivers' com-

TABLE 2 Parameters and Hazard Ratio Estimates: Snow Season Model

Variable	Parameter Estimates				Hazard Ratio			
	$\beta$ Coefficient		CI		$\beta$ Coefficient		CI	
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
Intercept	1.596	0.510	0.600	2.541	na	na	na	na
Grade (flat, 0%–2%) (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Grade (moderate, >2%–4%)	0.820	0.354	0.147	1.533	2.420	0.905	1.158	4.631
Grade (steep, >4%–6%)	0.927	0.341	0.279	1.612	2.691	0.952	1.261	4.951
Grade (very steep, >6%–8%)	2.203	0.361	1.533	2.928	9.671	3.730	4.634	18.69
Grade index (1 = upgrade) (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Grade index (2 = downgrade)	0.009	0.188	-0.369	0.381	1.031	0.191	0.688	1.456
Degree of curvature	-0.301	0.067	-0.434	-0.175	0.742	0.049	0.648	0.839
Median width	-0.053	0.008	-0.069	-0.038	0.948	0.008	0.933	0.963
Precipitation	0.881	0.418	0.149	1.774	2.652	1.268	1.161	5.892
Visibility	-2.207	0.342	-2.862	-1.533	0.117	0.041	0.057	0.216
log CV speed	0.501	0.225	0.056	0.944	1.693	0.388	1.058	2.576
pD: number of effective variables	9.506	na	na	na	na	na	na	na
DIC	802.028	na	na	na	na	na	na	na
Area under ROC curve	0.84	na	na	na	na	na	na	na
Sensitivity	80.09	na	na	na	na	na	na	na

NOTE: CV = coefficient of variation. Summary statistics [mean (SD)]: degree of curvature = 1.39 (1.52); median width (ft) = 24.50 (15.45); visibility (mi) = 1.09 (0.47); precipitation (in.) = 0.05 (0.29); log CV speed = 0.24 (0.38).

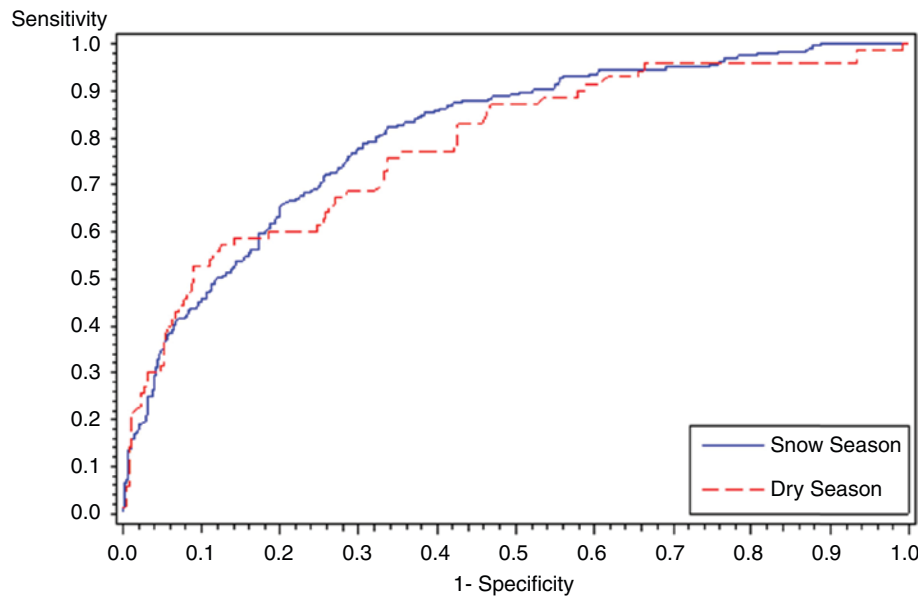


FIGURE 2 ROC curve analysis (dry and snowy season models).

TABLE 3 Classification Results

Variable	Predicted		Total
	0 (Noncrash)	1 (Crash)	
<b>Dry Season Model</b>			
Actual			
0 (noncrash)			
Frequency	170	86	256
Percentage	52.15	26.38	78.53
Row (%)	66.41 <sup>a</sup>	33.59 <sup>b</sup>	na
Column (%)	90.91	61.87	na
1 (crash)			
Frequency	17	53	70
Percentage	5.21	16.26	21.47
Row (%)	24.29 <sup>c</sup>	75.71 <sup>d</sup>	na
Column (%)	9.09	38.13	na
Total			
Frequency	187	139	326
Percentage	57.36	42.64	100.00
<b>Snow Season Model</b>			
Actual			
0 (noncrash)			
Frequency	423	201	624
Percentage	49.47	23.51	72.98
Row (%)	67.79 <sup>a</sup>	32.21 <sup>b</sup>	na
Column (%)	90.19	52.07	na
1 (crash)			
Frequency	46	185	231
Percentage	5.38	21.64	27.02
Row (%)	19.91 <sup>c</sup>	80.09 <sup>d</sup>	na
Column (%)	9.81	47.93	na
Total			
Frequency	469	386	855
Percentage	54.85	45.15	100.00

<sup>a</sup>Specificity.

<sup>b</sup>False-positive rate.

<sup>c</sup>False-negative rate.

<sup>d</sup>Sensitivity.

pliance with the system and hence reduce its effectiveness. Nevertheless, the objectives of advanced traffic management systems of achieving reductions in turbulence to improve operation can still be achieved even with a high percentage of false alarms. Conditions with false alarms are still not ideal, and a reduction of the flow turbulence could lead to operational benefits, although it might not lead to a crash. Strategies that are part of intelligent transportation systems, such as variable speed limits, could be introduced so that drivers would have no knowledge of the occurrence of a false alarm.

ROC curves were also generated as another way to assess the performance of the models. The area under the ROC curve shows how well the model discriminates the response variable between the crash ( $y = 1$ ) and noncrash ( $y = 0$ ) case. This discrimination is similar to the misclassification rate, but the ROC curve calculates sensitivity (true positive rate) and  $1 - \text{specificity}$  (false-positive rate) values for many cutoff points. The exact areas under the ROC curves were found to be 0.783 and 0.840 for the dry and the snowy seasons, respectively, which indicate that the models can provide good discrimination.

## CONCLUSION

Real-time crash prediction models that depend only on traffic parameters are useful for freeways with normal geometries and at locations that do not encounter severe weather conditions. Most of the previous studies found that the use of traffic turbulence (e.g., speed variance) defined by traffic parameters is a more dominant way to discriminate between crash and noncrash cases; therefore, the matched case-control design was an adequate technique to account for the small variability in roadway geometry and weather. This study illustrates that the same traffic turbulence can affect the driver differently on roadway sections with special geometries and under different weather conditions. A roadway geometry in mountainous terrain and adverse weather could exacerbate the effect of

traffic turbulence, and therefore, the inclusion of these factors is vital in the context of active traffic management systems.

Although all previous studies used data from loop detectors (which provide time mean speed, flow, and lane occupancy), this study shows that traffic data collected from an AVI system and real-time weather data provide good measures of a crash risk in real time.

Preliminary analysis of the data and findings from a previous study (29) indicate that the risk of a crash during the snowy season is 82% higher than the risk of a crash during the dry season; therefore, two models were considered in this study to examine the effect of the interaction between geometric features, weather, and traffic data on crash occurrence. Although all geometric factors included in the models were significant during the dry and snowy seasons, the coefficient estimates indicate that the likelihood of a crash could be doubled during the snowy season because of the interaction between the snowy, icy, or slushy pavement conditions and the steep grades. The hazard ratio for the very steep grades (>6% to 8% and <-6% to -8%) during the snowy season increased to 9.71, whereas it was 5.63 during the dry season.

The same conclusion can be made for visibility: a reduction in visibility of one unit was found to increase the risk of a crash by 88% during the snowy season, whereas the value was 79% during the dry season. The precipitation in the 10 min before the time of a crash was significant only in the model of the snowy season: a one-unit increase in precipitation increased the risk of a crash by 169%. The logarithm of the coefficient of variation of the speed at the crash segment during the 6 to 12 min before the crash was found to be significant during the snowy season, whereas the 6-min average speed at the crash segment 6 to 12 min before the crash was found to be significant during the dry season.

The results from this study suggest that the inclusion of roadway and weather factors in real-time crash prediction models is essential, in particular, with roadways that feature challenging roadway characteristics and adverse weather conditions. Different active traffic management strategies should also be in place during these two distinct seasons, and more resources should be devoted during the snowy season.

The findings of this study also indicate that traffic management authorities can benefit from the use of data from AVI systems and real-time weather data not only to ease congestion and enhance operations but also to mitigate increased safety risks.

## ACKNOWLEDGMENT

The authors thank the Colorado Department of Transportation for providing the data that were used in this study and for funding this research.

## REFERENCES

- Oh, C., J. Oh, S. Ritchie, and M. Chang. Real-Time Estimation of Freeway Accident Likelihood. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
- Abdel-Aty, M., N. Uddin, A. Pande, M. F. Abdalla, and L. Hsia. Predicting Freeway Crashes from Loop Detector Data by Matched Case-Control Logistic Regression. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1897, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 88–95.
- Abdel-Aty, M., and A. Pande. Identifying Crash Propensity Using Specific Traffic Speed Conditions. *Journal of Safety Research*, Vol. 36, No. 1, 2005, pp. 97–108.
- Abdel-Aty, M., and A. Pande. Comprehensive Analysis of Relationship Between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1953, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 31–40.
- Abdel-Aty, M., N. Uddin, and A. Pande. Split Models for Predicting Multivehicle Crashes During High-Speed and Low-Speed Operating Conditions on Freeways. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1908, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 51–58.
- Abdel-Aty, M., A. Pande, C. Lee, V. Gayah, and C. Santos. Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety on Freeways. *Journal of Intelligent Transportation Systems*, Vol. 11, No. 3, 2007, pp. 107–120.
- Abdel-Aty, M. A., A. Pande, A. Das, and W. J. Knibbe. Assessing Safety on Dutch Freeways with Data from Infrastructure-Based Intelligent Transportation Systems. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2083, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 153–161.
- Abdel-Aty, M., and A. Pande. Classification of Real-Time Traffic Speed Patterns to Predict Crashes on Freeways. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
- Pande, A., and M. Abdel-Aty. Assessment of Freeway Traffic Parameters Leading to Lane-Change Related Collisions. *Accident Analysis and Prevention*, Vol. 38, No. 5, 2006, pp. 936–948.
- Hourdos, J. N., V. Garg, P. G. Michalopoulos, and G. A. Davis. Real-Time Detection of Crash-Prone Conditions at Freeway High-Crash Locations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1968, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 83–91.
- Ahmed, M., and M. Abdel-Aty. The Viability of Using Automatic Vehicle Identification Data for Real-Time Safety Risk Assessment. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, 2012, pp. 459–468.
- Hassan, H., and M. A. Abdel-Aty. Exploring Visibility Related Crashes on Freeways Based on Real-Time Traffic Flow Data. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
- Abdel-Aty, M., and R. Pemmanaboina. Calibrating a Real-Time Traffic Crash-Prediction Model Using Archived Weather and ITS Traffic Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, No. 2, 2006, pp. 167–174.
- Abdel-Aty, M., and M. F. Abdalla. Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes: Generalized Estimating Equations for Correlated Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1897, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 106–115.
- Abdel-Aty, M., and A. Pande. Crash Data Analysis: Collective vs. Individual Crash Level Approach. *Journal of Safety Research*, Vol. 38, No. 5, 2007, pp. 97–108.
- Lee, C., F. Saccomanno, and B. Hellinga. Analysis of Crash Precursors on Instrumented Freeways. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 1–8.
- Lee, C., B. Hellinga, and F. Saccomanno. Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1840, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 67–77.
- Goodwin, L. *Weather Related Crashes on U.S. Highways*. FHWA, U.S. Department of Transportation, 2002. [http://ops.fhwa.dot.gov/Weather/best\\_practices/CrashAnalysis2001.pdf](http://ops.fhwa.dot.gov/Weather/best_practices/CrashAnalysis2001.pdf). Accessed Dec. 2010.
- Brodsky, H., and S. Hakkert. Risk of a Road Accident in Rainy Weather. *Accident Analysis and Prevention*, Vol. 20, No. 2, 1988, pp. 161–176.



20. National Transportation Safety Board. *Fatal Highway Accidents on Wet Pavement—The Magnitude, Location, and Characteristics*. Report HTSB-HSS-80-1. National Technical Information Service, Springfield, Va., 1980.
21. Andreescu, P., and D. Frost. Weather and Traffic Accidents in Montreal, Canada. *Climate Research*, Vol. 9, No. 3, 1998, pp. 225–230.
22. Andrey, J., and R. Olley. Relationships Between Weather and Road Safety, Past and Future Directions. *Climatological Bulletin*, Vol. 24, No. 3, 1990, pp. 123–137.
23. Tam, M., and W. Lam. Application of Automatic Vehicle Identification Technology for Real-Time Journey Time Estimation. *Information Fusion*, Vol. 12, No. 1, 2011, pp. 11–19.
24. Dion, F., and H. Rakha. Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data for Low Sampling Rates. *Transportation Research Part B: Methodological*, Vol. 40, No. 9, 2006, pp. 745–766.
25. Golob, T., and W. Recker. A Method for Relating Type of Crash to Traffic Flow Characteristics on Urban Freeways. *Transportation Research Part A: Policy and Practice*, Vol. 38, No. 1, 2004, pp. 53–80.
26. Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS—A Bayesian Modeling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing*, Vol. 10, No. 4, 2000, pp. 325–337.
27. Spiegelhalter, D., N. Best, B. Carlin, and V. Linde. Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal of the Royal Statistical Society B*, Vol. 64, No. 4, 2003, pp. 583–616.
28. Brooks, S., and A. Gelman. Alternative Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, Vol. 7, 1998, pp. 434–455.
29. Ahmed, M., H. Huang, M. Abdel-Aty, and B. Guevara. Exploring a Bayesian Hierarchical Approach for Developing Safety Performance Functions for a Mountainous Freeway. *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011, pp. 1581–1589.
30. Shankar, V., F. Mannering, and W. Barfield. Effect of Roadway Geometrics and Environmental Factors on Rural Accident Frequencies. *Accident Analysis and Prevention*, Vol. 27, No. 3, 1995, pp. 371–389.
31. Chang, L., and W. Chen. Data Mining of Tree-Based Models to Analyze Freeway Accident Frequency. *Journal of Safety Research*, Vol. 36, No. 4, 2005, pp. 365–375.
32. Anastasopoulos, P., A. Tarko, and F. Mannering. Tobit Analysis of Vehicle Accident Rates on Interstate Highways. *Accident Analysis and Prevention*, Vol. 40, No. 2, 2008, pp. 768–775.
33. Agresti, A. *Categorical Data Analysis*, 2nd ed. John Wiley & Sons, Inc., Hoboken, N.J., 2002.

---

*All opinions and results are those of the authors.*

*The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.*