Utilizing Microscopic Traffic and Weather Data to Analyze Real-Time Crash Patterns in the Context of Active Traffic Management

Rongjie Yu, Mohamed A. Abdel-Aty, Mohamed M. Ahmed, and Xuesong Wang

***Abstract*—This paper investigates the effects of microscopic traffic, weather, and roadway geometric factors on the occurrence of specific crash types for a freeway. The I-70 Freeway was chosen for this paper since automatic vehicle identification (AVI) and weather detection systems are implemented along this corridor. A main objective of this paper is to expand the purpose of the existing intelligent transportation system to incorporate traffic safety improvement and suggest active traffic management (ATM) strategies by identifying the real-time crash patterns. Crashes have been categorized as rear-end, sideswipe, and single-vehicle crashes. AVI segment average speed, real-time weather data, and roadway geometric characteristic data were utilized as explanatory variables in this paper. First, binary logistic regression models were estimated to compare single- with multivehicle crashes and sideswipe with rear-end crashes. Then, a hierarchical logistic regression model that simultaneously fits two conditional logistic regression models for the three crash types has been developed. Results from the models indicate that single-vehicle crashes are more likely to occur in snowy seasons, at moderate slopes, three-lane segments, and under free-flow conditions, whereas the sideswipe crash occurrence differs from rear-end crashes with the visibility situation, segment number of lanes, grades, and their directions (up or down). Furthermore, the innovative way of estimating two conditional logistic regression models simultaneously in the Bayesian framework fits the correlated data structure well. Conclusions from this paper imply that different ATM strategies should be designed for three- and two-lane roadway sections and are also considering the seasonal effects.**

***Index Terms*—Active traffic management (ATM), crash-type analysis, intelligent transportation system (ITS), microscopic crash analysis, random effect logistic regression, real-time data.**

I. Introduction

THIS PAPER focuses on investigating the effects of microscopic traffic, weather, and roadway geometric factors on the occurrence of specific crash types for a mountainous freeway. The objective of this paper is to investigate the feasibility of utilizing data from an existing intelligent transportation system (ITS) to identify real-time crash patterns. Understanding various characteristics of different crash types would benefit future active traffic management (ATM) systems to proactively reduce crash occurrences.

This paper focuses on a 15-mile mountainous freeway section of I-70 in Colorado, USA. The COTrip system, developed by the Colorado Department of Transportation (CDOT), is used to provide drivers with information about travel time, congestion, adverse weather conditions, and lane closures due to roadway maintenance. This information is provided as a part of an ITS, which can be accessed online. An automatic vehicle identification (AVI) system was installed to estimate travel times for critical roadway sections, and weather stations were implemented along the freeway section to monitor the adverse weather conditions. In order to explore traffic safety applications in conjunction with the existing ITS and ATM systems, this paper focuses on analyzing crash patterns with the real-time traffic and weather data from the AVI system and weather stations.

Previous studies stated that it is important to analyze the crash by types, particularly when it comes to real-time crash risk assessment [1]. Moreover, recent studies have argued that the hazardous factors influencing crash occurrence vary by crash type [2]–[5]. With the purpose of exploring the potential utilization of microscopic traffic and weather data in preventing crash by types and possibly developing different ATM strategies, this paper develops models to reveal the propensities of different crash types with the aid of real-time traffic and weather data.

The main objectives of this paper are as follows:

1) identifying various crash occurrence mechanisms for the three most common crash types on the studied freeway (i.e., rear-end, sideswipe, and single-vehicle crashes);
2) utilizing random effects logistic regression models in the Bayesian framework to account for the segment unobserved heterogeneity and compare their results to the classic Bayesian logistic regression models;
3) proposing a Bayesian hierarchical logistic regression model to analyze the three crash types simultaneously for more efficient model estimation.

To accomplish these objectives, crash-type propensities were analyzed by comparing each crash type to the other crash types given the crash occurrence. A set of binary logistic regression models and one hierarchical logistic regression model have

Manuscript received March 7, 2013; revised June 19, 2013 and July 22, 2013; accepted July 24, 2013. Date of publication August 26, 2013; date of current version January 31, 2014. The Associate Editor for this paper was M. Brackstone.

R. Yu and X. Wang are with the School of Transportation Engineering, Tongji University, Shanghai 201804, China (e-mail: rongjie.yu@gmail.com; yurongjie@tongji.edu.cn; wangxs@tongji.edu.cn).

M. A. Abdel-Aty is with the Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA (e-mail: m.aty@ucf.edu).

M. M. Ahmed is with the Department of Civil and Architectural Engineering, University of Wyoming, Laramie, WY 82071 USA (e-mail: mahmed@knights.ucf.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2013.2276089

been estimated. Moreover, the Bayesian random effects logistic regression models are compared with the classical Bayesian logistic regression models regarding the model goodness of fit. Furthermore, a hierarchical logistic regression model is introduced to provide a more efficient way to analyze the three crash types. Finally, analyzing results would shed some light on the future of ATM strategies designed for traffic safety improvement.

II. Background

A. Crash-Type Analysis

For the purpose of reducing crash occurrence and alleviating the severity of crashes, various types of statistical models have been developed to unveil the mechanisms of crash occurrence. In general, there are two major types of crash analyses, i.e., aggregate and disaggregate traffic safety studies. Aggregate analyses mainly focus on discovering the hazardous factors that are related to the frequency of total crashes and specific crash type or for each crash severity level. Disaggregate studies benefit from the reliable surveillance systems, which provide detailed traffic and weather data for each crash. This information could help in capturing the microlevel influences of the hazardous factors that lead to different crash types.

For the aggregate analysis, Qin *et al.* [3] utilized a hierarchical Bayesian framework to predict crash occurrence in relation to the hourly exposure by crash type. Four crash types were analyzed: 1) single-vehicle; 2) multivehicle same direction; 3) multivehicle opposite directions; and 4) multivehicle intersecting directions. A set of binary regression models was estimated for different crash types and time periods. Through comparing the marginal posterior distributions, it was concluded that the occurrence of crashes during the morning and afternoon periods significantly varies by crash types, and the single- and multivehicle crashes have distinct crash occurrence mechanisms. Moreover, several other studies [4], [6] have also addressed the crash types' propensity through developing safety performance functions (SPFs) for highway intersections. Results demonstrated that the relationship between traffic flow and crash frequency varies by different crash types; better model fit could be achieved by modeling different crash types separately. In general, freeway crash-type analyses have mainly focused on single- and multivehicle crashes [7], [8]. In addition, Geedipally and Lord [9] investigated the effects of modeling single- and multivehicle crashes, separately and jointly. Five years of undivided four-lane highway crash data were utilized. The crash frequency per year per mile was set as the target variable, and only annual average daily traffic was introduced into the models as the covariate. Univariate negative binomial (NB) and bivariate NB models have been employed to estimate the SPFs. Mean absolute deviance, mean squared predictive error, and confidence intervals were used to evaluate the models' performance. The authors concluded that single- and multivehicle crashes should be analyzed separately, and a joint NB model should be utilized.

Researchers have also investigated the effects of drivers' distractions on different crash types. Neyens and Boyle [10] looked at the effect of different distraction sources on crash types of teenage drivers. A multinomial logit model was developed to predict the likelihood of a teenage driver to be involved in each crash type. Three crash types (i.e., angle, rear-end, and collision with fixed objects) and four distraction types (i.e., cell phone, cognitive, passenger, and in-vehicle) were considered. Results concluded that different driver distractions have varying effects on teenage drivers' crash involvement. Similarly, Ghazizadeh and Boyle [11] introduced a multinomial logit model to determine the relationship between distraction types and crash types in Missouri. Conclusions from the model results showed that cell phone- and passenger-related distractions would mostly result in an angle crash, whereas electronic-device-related distractions are more likely to occur before a single-vehicle crash. From these studies, it can be seen that distractions also have varying influences on different crash types.

A more advanced disaggregate crash propensity analysis with traffic data was conducted by Christoforou *et al.* [12]. The study used four-year crash data from the A4-A86 highway section in France; the crash data were split into five distinct types (rear-end with two vehicles, sideswipe, rear-end with more vehicles, multiple-vehicle collision, and single-vehicle crashes). Specific traffic data were obtained from loop detectors and used at 6-min intervals. Univariate probit models were developed for each crash type and a multivariate probit model for all the crash types. Results indicated that two-vehicle rear-end crashes are more likely to happen during daytime with lower values of density and average speed, which represents the critical transitions from free-flow to congestion. Single-vehicle crashes are more probable to occur on straight and flat road segments. The authors concluded that diverse effects of accident-contributing factors exist for each crash type. This paper focuses on conducting disaggregate crash-type propensity analysis with Bayesian models. Microscopic traffic and weather data were achieved to reveal the different crash occurrence scenarios for main crash types.

B. Traffic Safety Analysis With AVI Data

The feasibility of utilizing AVI data in real-time crash risk evaluation models has been investigated by Ahmed and Abdel-Aty [13]. AVI data from expressways in the Orlando, FL, USA, metropolitan area were employed to develop real-time crash prediction models. The authors developed matched case control logistic regression models to classify the crash and noncrash cases. Results of the models demonstrated that AVI data have promising applications in predicting crashes on expressways.

Moreover, Ahmed *et al.* [14] utilized AVI data along with real-time weather information and roadway geometric characteristics to formulate a real-time crash occurrence model. A logistic regression model was estimated with the Bayesian inference technique. The finalized model showed that geometric factors are significant along with the 6-min average speed captured by the AVI system during the 6–12-min interval prior to the crash time, and the 1-h visibility before the crash time was also found to be significant in both the snow and dry seasons. Furthermore, specifically for the snow season, the 10-min precipitation prior to the crash time was also significant.

TABLE I
SUMMARY OF VARIABLES' DESCRIPTIVE STATISTICS

Variables	Type	Coding	Descriptive statistics*
Dependent Variables			
Crash Type	Nominal	1=rear-end crashes	18% (n=121)
		2=single-vehicle crashes	66% (n=442)
		3=sideswipe crashes	16% (n=107)
Explanatory Variables			
Number of lanes	Binary	1=3 lane segment	52% (n=348)
		0=2 lane segment	48% (n=322)
Season	Binary	1=snow season (Oct to Apr)	84% (n=563)
		0=dry season	16% (n=107)
Grade	Binary	Moderate=0 +/- 2%	21% (n=141)
		Steep => 2 to 8%, <-2 to -8%	79% (n=529)
Visibility (in miles)	Continuous	Visbility prior to crash	Mean=1.80; Std Dev.=1.60
Speed (in mph)	Continuous	Mean AVI speed 6 to 12 minutes prior to crash	Mean=49.46; Std Dev.=11.36

* Percentages are provided for the nominal/binary variables; mean and standard deviation values are provided for the continuous variables.

C. *Traffic Safety Analysis With Weather Data*

Weather conditions' influences on traffic safety have been mainly investigated at the aggregate level. Caliendo *et al.* [15] used hourly rainfall data and transformed them into a binary indicator of daily pavement surface status (dry and wet). Miaou *et al.* [16] also utilized a surrogate variable to represent wet pavement conditions. The amount of rainfall and the number of rainy days have been identified to have a positive effect on crash occurrence [17], [18]. However, this study takes a step forward to analyze the relationship between weather conditions and crash occurrence at a disaggregate level; real-time weather data would provide the exact weather conditions for each crash occurrence.

III. DATA PREPARATION

Four data sets were included in this paper: 1) I-70 crash data provided by CDOT (based on crash availability, data from July 2007 to July 2009 and August 2010 to April 2011 were used); 2) roadway segment geometric characteristics data obtained from the Roadway Characteristics Inventory; 3) real-time weather data recorded by six weather stations along the studied roadway section; and 4) real-time traffic data detected by 20 AVI detectors located on the east and west bounds along I-70. By utilizing the real-time data, crash occurrence contributing factors from roadway geometric, weather, and traffic flow characteristics for each crash type could be unveiled. Table I provides the summary descriptive statistics for the dependent and independent variables used in this paper.

A total of 670 crashes were documented within the study period. The 15-mile freeway section, starting at Mile Marker (MM) 205 and ending at MM 220, has been split into 120 homogenous segments (60 in each direction); the homogenous segments have a minimum length of 0.1 mile, and adjacent segments with similar geometric characteristics were combined together (the detailed homogenous segmentation method has been described in a previous study [19]). For the longitudinal grades, two grade levels were defined: Grade [Moderate] (reference condition, grade ranges from 0% to $\pm$ 2%); Grade [Steep] (grade ($>$2% to 8%) and (<-2% to -8%)). Moreover, an interaction variable of grades and grade direction has been created: Grade_Dir [1] (upgrade moderate grades); Grade_Dir [2] (upgrade steep grades); and Grade_Dir [3] (reference condition, downgrade steep grades) (no combinations of downgrade moderate grades segment exist).

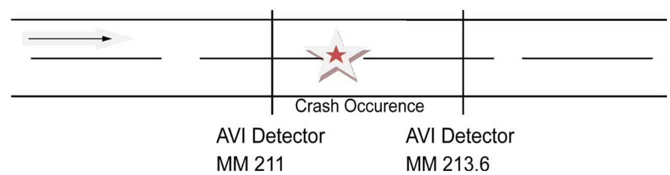

Fig. 1. AVI segment and crash occurrence.

The AVI segment speed data, frequently used to estimate travel times between freeway segments, have been proven to be useful in evaluating real-time traffic safety [14], [20], [21]. The AVI data corresponded to each crash case were extracted using the following procedures. The raw data (2-min space mean speed for each AVI segment) were first aggregated into 6-min intervals, and then each crash was assigned to the AVI segments based on MM. Crash's traffic status is defined as 6–12 min prior to the crash time. For example, if a crash happened at 10:26 at MM 211.3, the corresponding traffic status for this crash is the traffic condition of time interval 10:14 and 10:20 recorded by the AVI segment between MM 211 and MM 213.6 (see Fig. 1). The 6–12-min time interval was chosen in order to avoid errors of crash reporting times and to consider the future applications for crash prediction.

For the weather data, visibility information from six weather stations has been utilized. Crashes have been assigned to the nearest weather station according to the MM. The weather data were matched to each crash based on the reported crash time. The closest weather record prior to the crash time would be extracted and used as the crash occurrence weather condition. Moreover, the sample size requirements have been examined before estimating the logistic regression models; according to Harrel *et al.* [22], logistic regression models require at least ten cases per candidate independent variable. The data sets used in this study meet the desired sample sizes.

IV. METHODOLOGY

To analyze the crash types' propensity, binary logistic regression models and hierarchical logistic regression models were employed in this paper. Binary logistic regression models were developed to provide preliminary analyses; pairs of the crash types were compared for single- and multivehicle crashes and sideswipe and rear-end crashes. Subsequently, a hierarchical logistic regression model was utilized to predict the likelihood of the three crash types simultaneously. Moreover, for the binary logistic regression models, Bayesian random effects logistic regression models were introduced to account for the unobserved heterogeneity at the segment level, which is brought up by variations among the homogenous segments.

The random effects logistic regression models are useful for data with group structure and a binary response variable, whereas the random effects can be utilized to account for a within-group correlation [23]. Huang *et al.* [24] introduced Bayesian random effects logistic regression models to perform the multivehicle crash injury severity analysis. By incorporating driver–vehicle units' correlations in the same multivehicle crashes and the unobserved heterogeneity at the crash units' level, a better model fit has been achieved. In addition, the Bayesian random effects logistic regression model was also employed in crash risk evaluation studies to account for the segment-level unobserved heterogeneity [25]. Moreover, they have been also utilized in the fields of anomaly detection [26], default prediction [27], and clinical trials [28]. In this paper, crashes were considered as clustered at each freeway segment, and random effects logistic regression models are utilized to consider the within-segment correlation.

Due to this, a traditional maximum-likelihood estimation (MLE) method cannot conclude a closed form for the random effects models [26], and the MLE for this kind of models (e.g., mixed logit model) is relatively computational cumbersome [29]. In this paper, the Bayesian inference technique was utilized. Bayesian methods based on Markov chain Monte Carlo simulation provide an easier approach for model estimation. In addition, Bayesian inference approach also provides a coherent and complete way to incorporate prior information, which could benefit future model applications in the implemented ATM systems.

The logistic regression models were estimated to predict the probability of a specific crash type relative to the whole crash data. Suppose that the rear-end crashes have the outcomes $y = 1$ or $y = 0$ with respective probabilities p and $1 - p$. The random effects logistic regression can be set up as

$$y \sim \text{Binomial}(p)\ \text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \boldsymbol{X\beta} + u_j(i)$$

where β_0 is the intercept, $\boldsymbol{X}$ is the vector of the explanatory variables, $\boldsymbol{\beta}$ is the vector of coefficients for the explanatory variables, and u_j is the random effects variable defined in the model, which represents the segment-specific random effects in this paper. The segment random effects would account for the unobserved heterogeneity (e.g., geometric factors such as median width and curvature, which are identical at the homogeneous segments). The random effects are set to follow a normal distribution $u_j \sim N(0, \tau)$, $j = 1, 2, 3 \ldots 120$, where τ is the precision parameter and it was specified a gamma prior as $\tau \sim$ Gamma (0.001, 0.001). For the explanatory variables, noninformative priors were set to follow normal distribution (Normal (0, 0.001)).

Full Bayesian inference was employed in this paper. For each model, three chains of 15 000 iterations were set up in WinBUGS [30]; 5000 iterations were used in the burn-in step. To prove the superiority of the random effects logistic regression models and the importance of accounting for segments' heterogeneity, the results of the models have been compared with the results from the classic Bayesian logistic regression models. The overall classification rates, deviance information criterion (DIC), and Brier scores (BSs) are utilized to assess the model goodness of fit and discrimination ability. The DIC, recognized as Bayesian generalization of Akaike information criterion, is a combination of the measure of model fitting and the effective number of parameters. The smaller DIC indicates a better model fit and, according to Spiegelhalter *et al.* [31], differences of more than 10 might definitely rule out the model with higher DIC. Differences between 5 and 10 are considered substantial. BS averages the squared differences between pairs of prediction probabilities and the subsequent binary observations, which was frequently used to compare models [32]. The BS falls between 0 and 1; the smaller the score, the better the predictive ability of the model.

Moreover, with the aim of analyzing the three crash types together, a hierarchical logistic regression approach has been proposed to establish two conditional logistic regression models simultaneously that share the common error term. Assume π_1, π_3, and π_3 are the probabilities of rear-end, single, and sideswipe crashes, respectively. Therefore, the two conditional logistic regression models are: 1) the probability π_2 of single-vehicle crashes and 2) the conditional probability $\pi_3/(\pi_1 + \pi_3)$ of a sideswipe crash, given the multivehicle crashes. The full model was fitted as

$$\log\frac{\pi_2}{\pi_1 + \pi_3} = \alpha_1 + \boldsymbol{x\beta} + \varepsilon_i$$
$$\log\frac{\pi_3}{\pi_1} = \alpha_2 + \boldsymbol{x\beta} + \varepsilon_i$$

where α_1 and α_2 are the intercepts, $\boldsymbol{X}$ is the vector of the explanatory variables, $\boldsymbol{\beta}$ is the vector of coefficients for the explanatory variables, and ε_i is the common error for the two models. The proposed model was also run in the Bayesian framework with also three chains of 15 000 iterations and 5000 iterations used in the burn-in step.

V. Results and Discussion

For the modeling procedures, the total crash data were initially classified as single-vehicle crashes and multivehicle crashes. Then, the multivehicle crashes were further categorized into rear-end crashes and sideswipe crashes. Separate binary logistic regression models were performed for each crash type, and then a hierarchical logistic regression model was applied to estimate the propensity of the three crash types simultaneously.

A. *Single-Vehicle Versus Multivehicle Crash Model*

A previous study [5] suggested comparing the single-vehicle with the multivehicle crashes at the disaggregate level. Among the 670 total crashes, 66.27% are single-vehicle crashes (444

TABLE II
PARAMETER ESTIMATES FOR SINGLE-VEHICLE VERSUS MULTIVEHICLE CRASH MODEL

Variables	Bayesian logistc regression		Bayesian random effects logistic regression	
	Estimate	(2.5%, 97.5%)	Estimate	(2.5%, 97.5%)
Intercept	-1.474	(-2.51, -0.421)	-1.279	(-2.392, -0.284)
Season[snow]	1.249	(0.786, 1.737)	1.368	(0.878, 1.855)
No_lanes[3 lanes]	0.432	(0.075, 0.792)	0.488	(0.005, 0.994)
Speed	0.028	(0.012, 0.044)	0.024	(0.009, 0.042)
Grade[Steep]	-0.629	(-1.11, -0.185)	-0.779	(-1.455, -0.146)

TABLE III
CLASSIFICATION RESULTS AND MODEL FITS FOR THE SINGLE-VEHICLE CRASH MODEL

	Bayesian logistc regression			Bayesian random effects logistic regression		
	Predict			Predict		
Observe	0	1	Total	0	1	Total
0	128 (56.63%)	98	226	155 (68.58%)	71	226
1	157	287 (64.64%)	444	134	310 (69.82%)	444
Total	285	385	670 (61.94%)	289	381	670 (69.40%)
Brier Score		0.20472			0.18365	
DIC		813.305			803.752	

TABLE IV
PARAMETER ESTIMATES FOR SIDESWIPE VERSUS REAR-END CRASH MODEL

Variables	Bayesian logistic regression		Bayesian random effects logistic regression	
	Estimate	(2.5%, 97.5%)	Estimate	(2.5%, 97.5%)
Intercept	-0.661	(-1.271, -0.078)	-0.912	(-1.814, -0.128)
Visibility	-0.232	(-0.423, -0.051)	-0.236	(-0.45, -0.037)
No_lanes[3 lanes]	1.326	(0.712, 1.947)	1.66	(0.803, 2.654)
Grade_Dir [1]	0.828	(0.01, 1.661)	1.264	(0.101, 2.574)
Grade_Dir [2]	0.574*	(-0.069, 1.209)	0.556*	(-0.374, 1.491)

single-vehicle crashes), which makes the studied freeway interesting as it differs from other freeways where the multivehicle crashes are the majority. Table II shows the significant explanatory variables in the single-vehicle versus multivehicle crash model. The snow season index is significant with a positive sign, which indicates a positive association between the snow season index and single-vehicle crash occurrence. Single-vehicle crashes are more likely to happen within the snow seasons, during which road surface conditions are not safe due to precipitation and low temperatures; the probability of drivers losing control is relatively high during snow seasons compared with dry seasons. Moreover, the real-time speed parameter has a positive sign, which means that segments with higher operation speed are more probable to have single-vehicle crashes. This phenomenon can be understood as single-vehicle crashes are likely to happen under free-flow conditions, whereas the multivehicle crashes typically occur during the congestion periods. Furthermore, single-vehicle crashes are more probable to happen at the three-lane segments since the three-lane segment index has a positive sign.

For the longitudinal grades, compared with the moderate grades, the steep grade index is significant with a negative sign. This demonstrates that flat grade segments are the most hazardous segments for the single-vehicle crashes. A similar study [12] also concluded that "the single-vehicle crashes seem to be more probable (compared to all other crash types) on straight and flat road segments," which is consistent with the result of our model. Nevertheless, grade directions (up or down) seem to have no sufficient influence on the single-vehicle crash occurrence.

Evaluation and comparison of the model goodness of fits have been provided in Table III. Considering the classification rates, the Bayesian random effects logistic regression model is superior to the Bayesian logistic regression model (69.40% compared with 61.94%). Furthermore, for the BSs, with the added segment random effects, the BS can be reduced by 10.3%. In addition, the DIC value has been reduced by 10 after considering the segment-level unobserved heterogeneity, which is a significant model improvement.

B. Sideswipe Versus Rear-End Crash Model

There are a total of 226 multivehicle crashes documented during the study period; 45.58% of multivehicle crashes are sideswipe crashes, whereas the other 54.42% are rear-end crashes. In this model, the crash propensity between sideswipe and rear-end crashes is investigated. Sideswipe crashes were labeled as 1, and rear-end crashes were marked as 0 in the logistic regression models. Table IV provides both the point estimates and credible intervals for the significant explanatory variables in the sideswipe versus rear-end crash model. For the real-time weather information, visibility turned out to be significant with

TABLE V
CLASSIFICATION RESULTS AND MODEL FITS FOR THE SIDESWIPE CRASH MODEL

	Bayesian logistc regression			Bayesian random effects logistic regression		
	Predict			Predict		
Observe	0	1	Total	0	1	Total
0	78 (63.41%)	45	123	91 (73.98%)	32	123
1	35	68 (66.02%)	103	27	76 (73.78%)	103
Total	113	113	226 (64.60%)	118	108	226 (73.89%)
Brier Score	0.21524			0.17537		
DIC	289.896			286.332		

a negative sign, which indicates that sideswipe crashes are more likely to happen under bad visibility conditions. During bad visibility conditions, lane-changing maneuvers are much more difficult compared with good visibility situations, which would lead to more sideswipe crashes. The binary index of snow season is not significant in this model, which indicates that multivehicle crashes share the same seasonal effect; both sideswipe and rear-end crashes are more likely to happen during dry seasons.

For the roadway geometric characteristics, the number of lanes and the interaction variable of grades and grade direction were found to be significant. The three-lane segment index has a positive sign, which demonstrates that lane-changing maneuvers are more frequent at the three-lane segments compared with the two-lane segments. Three-lane segments provide larger exposure for sideswipe crashes and, naturally, the sideswipe crashes have relatively higher risk at these locations. For the longitudinal grade, both the grade values and the directions of slopes have sufficient effects on the sideswipe crash occurrence, which demonstrate that sideswipe crashes are more likely to happen at the upgrade slope segments. Referring to the downgrade steep slopes, upgrade segments with moderate grades are the most hazardous ones for sideswipe crash occurrence, followed by the upgrade segments with steep grades. In other words, rear-end crashes are most likely to happen at the downgrade steep grades. Drivers who failed to stop the vehicles promptly at downgrade steep grades would experience high probability of rear-end crashes.

In order to evaluate the results provided by the two different models, classification rates, DIC, and BS have been listed in Table V. The Bayesian random effects logistic regression model again outperformed the Bayesian logistic regression model with higher accuracy rate (73.89%), smaller DIC value (286.332), and lower BS value (0.17537).

C. Hierarchical Logistic Regression Model

Results from the previous models are capable of identifying crash occurrence hazardous factors for each crash type. However, in the context of traffic safety management, each crash type's probability needs to be calculated simultaneously. Previous binary models are not succinct enough in terms of crash-type determination, and a full model that analyzes the three crash types is then developed.

The response variable in the full model would be a nominal variable, which has three levels to represent different crash types. Intuitively, multinomial and conditional logit models can be utilized to analyze the nominal variable. However, these models require the probability of having a specific type of crash to be independent of the presence or characteristics of the other crash types (Independence from Irrelevant Alternatives [IIA] assumption [33], [34]) since the three crash types are not exactly independent from each other. For example, drivers who change lanes abruptly to avoid a rear-end crash can end up either in a sideswipe crash or a runoff-road single-vehicle crash. A common error shared by the three crash types needs to be considered in the full model. One way to account for this is to assume crash-type error in the generalized logit link, which leads to a hierarchical logistic regression model. Moreover, the nested logit model is also considered and estimated (the model results are not shown in this paper for brevity). The inclusive value parameter was not significant in the nested logit model, which indicates that the data did not support the nested structure. With considering data features and the IIA issues, the hierarchical logistic regression model with Bayesian inference technique has been adopted.

The hierarchical logistic regression model compares the single-vehicle crashes with the multivehicle crashes and sideswipe crashes with rear-end crashes simultaneously. Results of the parameter estimations are shown in Table VI; the results are identical with the binary logistic regression models.

TABLE VI
LIKELIHOOD OF SINGLE-VEHICLE AND SIDESWIPE CRASHES USING A HIERARCHICAL LOGISTIC REGRESSION MODEL

Variables	Mean	Std. Error	2.5%	97.5%
Single Crashes				
Intercept	**-1.637**	0.519	-2.603	-0.478
Visibility	-0.065	0.056	-0.175	0.046
Speed	**0.025**	0.008	0.009	0.039
No_lanes[3 lanes]	**0.477**	0.187	0.114	0.839
Season[snow]	**1.114**	0.252	0.598	1.594
Grade_Dir[1]	**0.55**	0.249	0.069	1.036
Grade_Dir[2]	-0.253	0.194	-0.629	0.13
Sideswipe Crashes				
Intercept	-0.905	0.66	-2.207	0.436
Visibility	**-0.266**	0.105	-0.476	-0.065
Speed	0.009	0.01	-0.01	0.03
No_lanes[3 lanes]	**1.325**	0.319	0.706	1.95
Season[snow]	-0.159	0.384	-0.9	0.6
Grade_Dir[1]	**0.759***	0.428	-0.074	1.607
Grade_Dir[2]	**0.516***	0.326	-0.138	1.167

* Significant at 90 percentile.
DIC: 2036.98.
Brier Score: 0.1928.

TABLE VII
CLASSIFICATION ACCURACY RATE COMPARISON

Authors	Prediction accuracy of crashes
Oh et al. (2001) [35]	55.8%
Abdel-Aty et al. (2004) [36]	69.4%
Oh et al. (2005) [37]	35.2%
Abdel-Aty and Pande (2005) [38]	73.9%
Hossain and Muromachi (2010) [39]	63.3%
Xu et al. (2013) [40]	61.0%

For the single-vehicle crash occurrence, average AVI speed, three-lane segments, snow season, and steep grades are positively related; the visibility condition and grade direction are not significant. For the sideswipe crashes, compared with the rear-end crashes, bad visibility conditions, three-lane segments, and upgrade with both moderate and steep grades would increase the probability of sideswipe crash occurrence, whereas speed and season are not significantly associated with sideswipe crashes. Moreover, the hierarchical model correctly classified 71.85% of single-vehicle crashes and 71.84% of sideswipe crashes. Comparisons of the classification accuracy rates have been made with several disaggregate traffic safety studies listed in Table VII. It can be concluded that the classification accuracy rates of this study are relatively high. In addition, overall BS for the hierarchical model is 0.1928, which is comparable with the results provided by the binary models, but the full model is believed to provide more efficient estimations.

VI. CONCLUSION

This paper has conducted disaggregate crash-type propensity analysis for a mountainous freeway, which suffered large amount of single-vehicle crashes. Distinct crash occurrence mechanisms have been found from the estimated models.

1) For the average speed, single-vehicle crashes are more likely to happen with higher speeds, whereas the multivehicle crashes would probably occur at congested segments.
2) Single-vehicle crashes are more likely to happen during snowy seasons with slippery road surface, whereas the multivehicle crashes mostly occur during the dry seasons.
3) The visibility conditions differentiate the rear-end crashes from the sideswipe crashes. Rear-end crash occurrence is positively associated with visibility, whereas sideswipe crashes have a negative relationship with the visibility condition.
4) Rear-end crashes tend to occur at two-lane segments, whereas the three-lane segments are more likely to have single-vehicle and sideswipe crashes.
5) Diverse results have been found for the influences of longitudinal grades on crash occurrence. For the rear-end crashes, downgrade segments with steep grades are the most dangerous ones, whereas the sideswipe crashes are more likely to occur at upgrade flat slopes. Nevertheless, for the single-vehicle crashes, flat grade segments are the most risky ones no matter the slope direction.

From the methodological point of view, Bayesian random effects logistic regression models have been proven to be superior to the classic Bayesian logistic regression models. With the randomly distributed segment effects, heterogeneity among the homogenous segments can be accounted for in the models. Moreover, from the model goodness-of-fit perspective, with the random effects added, accuracy values of the models' prediction significantly increased with lower BSs and DIC values. Furthermore, through the Bayesian inference technique, results from the aforementioned models could be used as prior information to update the developed models in future system implementation.

In addition, the hierarchical logistic regression model fits the data structure well, which provides an efficient way to analyze the three crash types' propensity. The IIA requirement for the multinomial logit model has been violated due to the correlated crash data. Results of the hierarchical model are identical with the binary logistic regression models, which also indicate that modeling the two binary models simultaneously is an appropriate approach to deal with such data set.

In addition to the models' results and the methodological contributions, incorporating real-time traffic and weather variables has the benefits of explaining different characteristics for each crash type. Furthermore, the results can be helpful in designing ATM systems. Different traffic management strategies should be in place during two distinctive seasons [41] and the three- and two-lane sections of the freeway. For example, within the snowy season, the main purpose of the ATM system should be to decrease the single-vehicle crash occurrence; speed limits should be lowered during adverse weather conditions to prevent runoff road crashes, and a weather warning system can be used to deliver messages about the weather and road surface condition to the drivers. On the other side, during dry seasons, a variable speed-limit system can be introduced to smooth the flow during recurrent congestion for the purpose of reducing multivehicle crashes; in addition, lane-changing maneuvers should be restricted under bad visibility situations to alleviate the sideswipe crash occurrence probability.

REFERENCES

[1] A. Pande and M. Abdel-Aty, "Assessment of freeway traffic parameters leading to lane-change related collisions," *Accid. Anal. Prev.*, vol. 38, no. 5, pp. 936–948, Sep. 2006.
[2] D. Kim, Y. Lee, S. Washington, and K. Choi, "Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models," *Accid. Anal. Prev.*, vol. 39, no. 1, pp. 125–134, Jan. 2007.
[3] X. Qin, J. Ivan, N. Ravishanker, J. Liu, and D. Tepas, "Bayesian estimation of hourly exposure functions by crash type and time of day," *Accid. Anal. Prev.*, vol. 38, no. 6, pp. 1071–1080, Nov. 2006.
[4] T. Jonsson, J. Ivan, and C. Zhang, "Crash prediction models for intersections on rural multilane highways: Differences by collision type," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2019, pp. 91–98, 2007.
[5] R. Yu, M. Abdel-Aty, and M. Ahmed, "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors," *Accid. Anal. Prev.*, vol. 50, pp. 371–376, Jan. 2013.
[6] T. Jonsson, C. Lyon, J. Ivan, S. Washington, I. van Schalkwyk, and D. Lord, "Investigating differences in the performance of safety performance functions estimated for total crash count and for crash count by crash type," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2102, pp. 115–123, 2009.
[7] J. Ivan, R. Pasupathy, and P. Ossenbruggen, "Differences in causality factors for single and multi-vehicle crashes on two-lane roads," *Accid. Anal. Prev.*, vol. 31, no. 6, pp. 695–704, Nov. 1999.
[8] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis," *Accid. Anal. Prev.*, vol. 34, no. 2, pp. 149–161, Mar. 2012.

[9] S. Geedipally and D. Lord, "Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma models," *Accid. Anal. Prev.*, vol. 42, no. 4, pp. 1273–1282, Jul. 2010.

[10] D. Neyens and L. Boyle, "The effect of distractions on the crash types of teenage drivers," *Accid. Anal. Prev.*, vol. 39, no. 1, pp. 206–212, Jan. 2007.

[11] M. Ghazizadeh and L. Boyle, "Influence of driver distractions on the likelihood of rear-end, angular, and single-vehicle crashes in Missouri," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2138, pp. 1–5, 2009.

[12] Z. Christoforou, S. Cohen, and M. Karlaftis, "Identifying crash type propensity using real- time traffic data on freeways," *J. Safety Res.*, vol. 42, no. 1, pp. 43–50, Feb. 2011.

[13] M. Ahmed and M. Abdel-Aty, "The viability of using automatic vehicle identification data for real-time crash prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 459–468, Jun. 2012.

[14] M. Ahmed, M. Abdel-Aty, and R. Yu, "Assessment of the interaction between crash occurrence, mountainous freeway geometry, real-time weather and AVI traffic data," *J. Transp. Res. Board*, vol. 2280, pp. 51–59, 2012.

[15] C. Caliendo, M. Guida, and A. Parisi, "A crash-prediction model for multilane roads," *Accid. Anal. Prev.*, vol. 39, no. 4, pp. 657–670, Jul. 2007.

[16] S. Miaou, J. Song, and B. Mallick, "Roadway traffic crash mapping: A space–time modeling approach," *J. Transp. Stat.*, vol. 6, pp. 33–58, 2011.

[17] L. Chang and W. Chen, "Data mining of tree-based models to analyze freeway accident frequency," *J. Safety Res.*, vol. 36, no. 4, pp. 365–375, 2005.

[18] W. Yaacob, M. Lazim, and Y. Wah, "Evaluating spatial and temporal effects of accidents likelihood using random effects panel count model," in *Proc. Int. Conf. Sci. Social Res.*, Kuala Lumpur, Malaysia, 2010, pp. 960–964.

[19] M. Ahmed, H. Huang, M. Abdel-Aty, and B. Guevara, "Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway," *Accid. Anal. Prev.*, vol. 43, no. 4, pp. 1581–1589, Jul. 2011.

[20] M. Ahmed, R. Yu, and M. Abdel-Aty, "Safety applications of automatic vehicle identification and real-time weather data on freeways," in *Proc. 18th World Congr. Intell. Transp. Syst.*, Orlando, FL, USA, 2011, pp. 1–11, [CD-ROM].

[21] M. Ahmed, M. Abdel-Aty, and R. Yu, "A Bayesian updating approach for real-time safety evaluation using AVI data," *Transport. Res. Rec.*, vol. 2280, pp. 51–59, 2012.

[22] F. Harrel, K. Lee, R. Califf, D. Pryor, and R. Rosati, "Regression modeling strategies for improved prognostic prediction," *Stat. Med.*, vol. 3, pp. 143–152, 1984.

[23] G. Wong and W. Mason, "The hierarchical logistic regression model for multilevel analysis," *J. Amer. Stat. Assoc.*, vol. 80, no. 391, pp. 513–524, Sep. 1985.

[24] H. Huang, H. Chin, and M. Haque, "Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis," *Accid. Anal. Prev.*, vol. 40, no. 1, pp. 45–54, Jan. 2008.

[25] R. Yu and M. Abdel-Aty, "Investigating the different characteristics of weekday and weekend crashes," *J. Safety Res.*, vol. 46, pp. 91–97, Sep. 2013.

[26] M. Mok, S. Sohn, and Y. Ju, "Random effects logistic regression model for anomaly detection," *Exp. Syst. Appl.*, vol. 37, no. 10, pp. 7162–7166, Oct. 2010.

[27] S. Sohn and H. Kim, "Random effects logistic regression model for default prediction of technology credit guarantee fund," *Eur. J. Oper. Res.*, vol. 183, no. 1, pp. 472–478, Nov. 2007.

[28] K. Singh, A. Bartolucci, and S. Bae, "The Bayesian multiple logistic random effects model for analysis of clinical trial data," *J. Data Sci.*, vol. 8, pp. 495–504, 2010.

[29] M. Burda, M. Harding, and J. Hausman, "A Bayesian mixed logit–probit model for multinomial choice," *J. Econometrics*, vol. 147, no. 2, pp. 232–246, Dec. 2008.

[30] D. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, "WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility," *Stat. Comput.*, vol. 10, no. 4, pp. 325–337, Oct. 2000.

[31] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, *WinBUGS User Manual*. Cambridge, U.K.: MRC Biostatistics Unit, 2003.

[32] D. Wilks, *Statistical Methods in the Atmospheric Sciences*. New York, NY, USA: Academic, 2011.

[33] T. Fry and M. Harris, "A Monte Carlo study of tests for the independence of irrelevant alternatives property," *Transp. Res. B*, vol. 30, no. 1, pp. 19–30, Feb. 1996.

[34] T. Fry and M. Harris, "The dogit ordered generalized extreme value model," *Australian New Zealand J. Stat.*, vol. 47, no. 4, pp. 531–542, Dec. 2005.

[35] C. Oh, J. Oh, and S. Ritchie, "Real-time estimation of freeway accident likelihood," presented at the 80th Annu. Meet. Transportation Research Board, Washington, DC, USA, 2001.

[36] M. Abdel-Aty, N. Uddin, F. Abdalla, A. Pande, and L. Hisa, "Predicting freeway crashes based on loop detector data using matched case-control logistic regression," *Transp. Res. Rec.*, vol. 1897, pp. 88–95, 2004.

[37] C. Oh, J. Oh, and S. Ritchie, "Real-time hazardous traffic condition warning system: Framework and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 3, pp. 265–272, Sep. 2005.

[38] M. Abdel-Aty and A. Pande, "Identifying crash propensity using specific traffic speed conditions," *J. Safety Res.*, vol. 36, no. 1, pp. 97–108, 2005.

[39] M. Hossain and Y. Muromachi, "Evaluating location of placement and spacing of detectors for real-time crash prediction on urban expressways," presented at the 89th Annu. Meet. Transportation Research Board, Washington, DC, USA, 2010.

[40] C. Xu, A. Tarko, W. Wang, and P. Liu, "Predicting crash likelihood and severity on freeways with real-time loop detector data," *Accid. Anal. Prev.*, vol. 57, pp. 30–39, Aug. 2013.

[41] R. Yu and M. Abdel-Aty, "Utilizing support vector machine in real-time crash risk evaluation," *Accid. Anal. Prev.*, vol. 51, pp. 252–259, Mar. 2013.

Rongjie Yu received the B.Sc. degree in civil engineering from Tongji University, Shanghai, China, in 2010 and the M.Sc. degree in transportation engineering and the Ph.D. degree from the University of Central Florida, Orlando, FL, USA, in 2012 and 2013, respectively.

He is now an assistant research fellow at the School of Transportation Engineering, Tongji University. His research interests include traffic safety analysis, statistical modeling, variable speed limits, and intelligent transportation systems.

Mohamed A. Abdel-Aty received the Ph.D. degree from the University of California, Davis, CA, USA in 1995.

He is a Professor of transportation engineering with the University of Central Florida (UCF), Orlando, FL, USA. He is also the Deputy of the Center for Advanced Transportation Systems Simulation, UCF. He has published more than 300 papers (165 in journals). His main expertise and interests are in the areas of traffic safety, travel demand analysis, and intelligent transportation systems. His research in the real-time prediction of traffic crashes on freeways is well recognized. He and his research team have moved freeway management from reactive incident detection to proactive incident prediction.

Prof. Abdel-Aty is a Registered Professional Engineer in Florida. He is an Editor of *Accident Analysis and Prevention*. He is a member of the Editorial Board of the *Journal of Intelligent Transportation Systems*. He was the recipient of UCF's Distinguished Researcher of the Year Award in 2003 and UCF's Outstanding Graduate Teacher Award in 2007.

Mohamed M. Ahmed received the B.Sc. degree in civil engineering from Al-Azhar University, Cairo, Egypt, in 2001 and the M.Sc. and Ph.D. degrees in transportation engineering from the University of Central Florida (UCF), Orlando, FL, USA, in 2009 and 2012, respectively.

He is an Assistant Professor of transportation engineering at the University of Wyoming, Laramie, WY, USA. His research interests include traffic safety analysis, intelligent transportation systems, and statistical and data mining applications in transportation engineering.

He was the recipient of the UCF's 2010–2011 Award for Excellence in Graduate Student Teaching.

Xuesong Wang received the B.Sc. and M.Sc. degrees in civil engineering from Tongji University, Shanghai, China, in 2000 and 2003, respectively, and the Ph.D. degree in transportation engineering from the University of Central Florida, Orlando, FL, USA, in 2006.

He is an Associate Professor of transportation engineering with Tongji University. His research interests include traffic safety analysis, safety planning, and traffic safety education.