Accident Analysis and Prevention 50 (2013) 371–376

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

# Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors

Rongjie Yu*, Mohamed Abdel-Aty, Mohamed Ahmed

*Department of Civil, Environmental and Construction Engineering, University of Central Florida, Engineering II – 215 Orlando, FL 32826, United States*

ARTICLE INFO

*Article history:*
Received 16 February 2012
Received in revised form 5 April 2012
Accepted 3 May 2012

*Keywords:*
Mountainous freeway safety
Bayesian inference
Real-time weather data and random effect model

ABSTRACT

Freeway crash occurrences are highly influenced by geometric characteristics, traffic status, weather conditions and drivers' behavior. For a mountainous freeway which suffers from adverse weather conditions, it is critical to incorporate real-time weather information and traffic data in the crash frequency study. In this paper, a Bayesian inference method was employed to model one year's crash data on I-70 in the state of Colorado. Real-time weather and traffic variables, along with geometric characteristics variables were evaluated in the models. Two scenarios were considered in this study, one seasonal and one crash type based case. For the methodology part, the Poisson model and two random effect models with a Bayesian inference method were employed and compared in this study. Deviance Information Criterion (DIC) was utilized as a comparison factor. The correlated random effect models outperformed the others. The results indicate that the weather condition variables, especially precipitation, play a key role in the crash occurrence models. The conclusions imply that different active traffic management strategies should be designed based on seasons, and single-vehicle crashes have different crash mechanism compared to multi-vehicle crashes.



## 1. Introduction

Motor-vehicle crash studies have been a continuously researched topic in the past decades. Researchers have developed various methods, incorporated different types of data and concluded varieties of countermeasures to improve the highway safety condition. In order to gain a better understanding of the crash mechanism, crash-frequency studies are now focusing on more specific problems that can be split into the following categories; crash type based studies (e.g., rear-end, sideswipe and single run-off-roadway crashes), severity based studies (property damage only, injury and fatal crashes), weather related crash studies (rainfall related crashes) and crash-time based studies (peak-hour and non-peak hour crashes). By concentrating on one particular problem with the help of more advanced data collection systems, researchers hope to provide better crash predictions and find out those hazardous factors.

This study focuses on a 15-mile mountainous freeway on I-70 in Colorado. Previous study (Ahmed et al., 2011a) demonstrated a significant seasonal effect on crash frequencies. Snow season (from October through April) have relatively higher crash occurrence and more weather-related crashes than the dry season (from May to September) does. In this study, the same homogeneous segmentation method is applied to the same study area. In addition to the geometric and aggregated traffic data used in the previous work, real-time weather data (visibility, precipitation and temperature) and real-time traffic data (speed, volume and occupancy) are employed in this paper. A season based model and a crash type based model are introduced, and two different types of Bayesian hierarchical random effect methodologies are utilized for each model. Finally the best models would be identified with the aim of providing helpful information to further traffic management strategies for different scenarios.

## 2. Background

Weather condition is relevant to crash frequency and researchers have developed several ways to consider weather influence in crash frequency models. Caliendo et al. (2007) used hourly rainfall data and transformed it into binary indicator of daily pavement surface status (dry and wet). Miaou et al. (2003) also used a surrogate variable to indicate wet pavement conditions. The amount of rainfall and the number of rainy days have been identified to have a positive effect on accident occurrence (Chang and Chen, 2005; Yaacob et al., 2010), results showed that higher precipitation (in terms of days and amount) have greater tendency to be classified with relatively higher accident rates. Daily averaged weather variables like precipitation, snowfall amounts and temperature have been utilized (Malyshkina et al., 2009); conclusions indicated that less safe traffic state is positively correlated with

* Corresponding author.
*E-mail address:* rongjie.yu@knights.ucf.edu (R. Yu).

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2012.05.011

extreme temperatures (low during winter and high during summer), rain precipitation, and snowfall and low visibility distances. More detailed hourly based weather data have been employed (Jung et al., 2010; Usman et al., 2010). However, as stated in Lord and Mannering's (2010) study that "generally the analyst only has precipitation data that is much more aggregated and thus important information is lost by using discrete time intervals – with larger intervals resulting in more information loss".

Traffic variables always play a vital role in crash occurrence studies. Kononov et al. (2011) used Annual Average Daily Traffic (AADT) as the only variable to develop the safety performance function and the results indicated that when some critical traffic density is reached, the crash occurrence likelihood would increase at a faster rate with an increase in traffic. Besides, in a spatially disaggregate road casualty analysis, Noland and Quddus (2004) used proximate employment variables to represent the different traffic flow scenarios and the results indicated that traffic flow has a high influence on increasing casualties. Furthermore, an AAA Foundation for Traffic Safety (1999) study focused on congestion and crashes concluded that a U-shaped model can explain the relationship between the two; crash rates are high at low levels of congestion and rapidly decrease as the volume to capacity (v/c) ratios increase, however they will increase again as the peak levels of congestion turn up. With the help of the data mining method of Classification and Regression Tree (CART), Chang and Chen (2005) concluded that AADT were the key determinant for freeway accident frequencies. However, similar to the weather related variables, using only aggregated traffic data such as AADT would lead to the loss of the most valuable information of pre-crash traffic status.

Random effect models have been widely used in crash frequency studies (Shankar et al., 1998; Miaou and Lord, 2003; Guo et al., 2010; Yaacob et al., 2010). Researchers have benefited from its advantage of handling temporal and spatial correlations (Lord and Mannering, 2010). With the random effect being added to the Negative Binomial model, the formulation would have a better ability to account for unobserved heterogeneity across spatial and temporal correlations (Chin and Quddus, 2003).

The Bayesian inference method is a frequently adopted way to predict crash occurrence in recent studies. A Hierarchical Bayes model was built to estimate area-based traffic crashes (Miaou et al., 2003). Shively et al. (2010) employed a Bayesian nonparametric estimation procedure in their study. A 5 X ST-level hierarchy structure was proposed to deal with multilevel traffic safety data (Huang and Abdel-Aty, 2010). Guo et al. (2010) included three types of Bayesian models in consideration of different complexities; fixed effect model, mixed effect model and conditional autoregressive (CAR) model have been compared. Furthermore, previous work (Ahmed et al., 2011a) on the same freeway segment employed Bayesian hierarchical models to account for seasonal and spatial correlations.

Run-off-road crashes (also recognized as single-vehicle crashes) take up to 30.8% in the overall crash occurrences. Shankar and Mannering (1996) worked on the injury severities of the statewide single-vehicle motorcycle crashes in Washington. Lee and Mannering (2002) analyzed run-off-roadway accidents on a highway segment in Washington State and provided potential countermeasures. Jung et al. (2010) assessed the effects of precipitation on the severity of single-vehicle crashes on Wisconsin interstate highways. Ivan et al. (1999) modeled single-vehicle and multi-vehicle crashes separately, aimed at identifying different causality factors for those two types of crashes. Geedipally and Lord (2010) investigated the influence of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals. Their conclusion indicated that single-vehicle and multi-vehicle crashes are correlated and modeling them separately will result in better model fittings.

## 3. Data preparation

Four data sets were included in this study, (1) one year of crash data (from August 2010 to August 2011) provided by Colorado Department of Transportation (CDOT), (2) road segment geometric characteristic data captured from the Roadway Characteristics Inventory (RCI), (3) real-time weather data recorded by 6 weather stations along the study roadway segment and (4) real-time traffic data detected by 30 Remote Traffic Microwave Sensor (RTMS) radars. To the best of our knowledge, this is the first time that real-time weather and traffic data have been employed in a study to estimate safety performance functions. By utilizing real-time data, contributing factors from roadway geometric, weather and traffic flow characteristics of crashes could be unveiled.

A total of 251 crashes were documented within the study period. The 15-mile segment, starting at Mile Marker (MM) 205 and ends at MM 220, have been split into 120 homogenous segments (60 in each direction), the homogenous segmentation method has been described in a previous study (Ahmed et al., 2011a).

Six weather stations were implemented with the purpose of providing real-time weather information to motorists. Information about temperature, visibility and precipitation had been recorded. The weather data is not recorded continuously, once the weather condition changes and reaches a preset threshold, a new record will be added to the archived data. Crashes have been assigned to the nearest weather station according to the Mile Marker. For each crash, based on the reported crash time, the closest weather record prior to the crash time has been extracted and used as the crash time weather condition.

Fifteen radar detectors were available for each direction to provide speed, volume and occupancy information. RTMS data corresponding to each crash case was extracted using the following process: the raw data were first aggregated into 5-min intervals, then each crash was assigned to the nearest downstream radar detector, and the crash's traffic status is defined as 5–10 min prior to the crash time. For example if a crash happened at 15:25, at the Mile Marker 211.3. The corresponding traffic status for this crash is the traffic condition of time interval 15:15 and 15:20 recorded by RTMS radar at Mile Marker 211.8. Similarly, upstream and downstream traffic statuses were also extracted for each crash case. To avoid confusing pre and post crash conditions, 5–10 min traffic variables prior to the reported crash time were extracted. Average, standard deviation and coefficient of variance of speed, volume and occupancy during the 5-min interval were calculated to represent the pre-crash traffic statuses. These traffic variables are named in a specific way as Fig. 1 shows. For example, DAO stands for the average

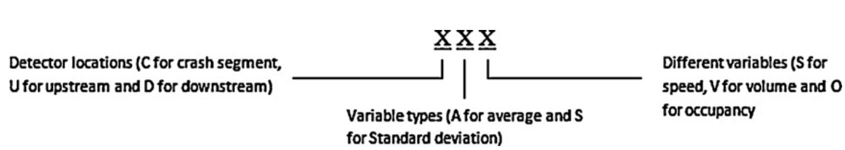


**Fig. 1.** Nomenclature method for traffic variables.

**Table 1**
Summary of variables descriptive statistics for the seasonal model.

| Variables | Description | Mean | Std dev. | Minimum | Maximum |
|---|---|---|---|---|---|
| Crash Frequency | Crash frequency counts for the segment | 1.09 | 1.95 | 0 | 13 |
| Av_visibility | Average visibility during the crashes | 3.97 | 1.85 | 0.1 | 7.1 |
| Av_temp | Average temperature during the crashes | 38.92 | 16.09 | 6.0 | 77.0 |
| Av_1hourprecip | Average value of 1 h precipitation (rain/snow) before the crash | 0.039 | 0.17 | 0 | 2.23 |
| S_1hourprecip | Standard deviation of 1 h precipitation (rain/snow) before the crash | 0.16 | 0.32 | 0 | 3.72 |
| CAS | Average speed for the crash segment | 53.81 | 10.59 | 7.45 | 68.0 |
| Season | Dry = 0, snow = 1 | 0.5 | 0.5 | 0 | 1.0 |
| Grade | Longitudinal grade, eight categories: upgrade: 0–2% = 1, 2–4% = 2, 4–6% = 3, 6–8% = 4; Downgrade: 0–(−2)% = 5, (−2)–(−4)% = 6, (−4)–(−6)% = 7, (−6)–(−8)% = 8 | 4.45 | 2.40 | 1 | 8 |
| VMT | Daily vehicle mile traveled | 6582 | 4419 | 2267 | 23409 |

occupancy captured by radar located at downstream of the crash location.

## 4. Methodology

Bayesian hierarchical models were employed in this study. Hierarchical modeling is a statistical technique that allows multilevel data structures to be properly specified and estimated (Gelman and Hill, 2007). This modeling approach have been applied in many previous studies (Shankar et al., 1998; Chin and Quddus, 2003; Guo et al., 2010; Ahmed et al., 2011a) and suggested to be utilized to model the multilevel traffic safety data (Huang and Abdel-Aty, 2010). In this study, for one specific segment, two distinct seasons' crashes or two different types of crashes were considered in the models. Two levels of data (e.g. segment level and seasonal level) are modeled, which means the data structural used in this study naturally is hierarchical.

Crash occurrence along the freeway can be assumed to follow Poisson process. The Poisson model has played an important role in crash-frequency studies. However it has been blamed of lacking the ability to handle over-dispersion problems (Lord and Mannering, 2010). Multiplicative gamma distributed random effects were introduced into the Poisson model, which implies a negative binomial marginal sampling distribution (Ntzoufras, 2009). The hierarchical model can be setup as follows:

$$Y_{it} \sim \text{Poission}(\lambda_{it}) \text{ for } t = 1, 2$$

$$\log \lambda_{it} = \log e_{it} + \boldsymbol{X_{it}} \boldsymbol{\beta} + \gamma_1 u_{it} + \gamma_2 b_i$$

$$u_{it} \sim N(0, \sigma_u^2)$$

$$b_t \sim N(0, \sigma_b^2)$$

Where $Y_{it}$ is the crash count at segment $i$ ($i$ = 1,· · ·, 120 (60 segments on each direction)) during season $t$ ($t$ = 1 for dry season, 2 for snow season) or for certain number of vehicles involved in the crash ($t$ = 1 for single vehicle crashes, 2 for multi vehicle crashes). $\boldsymbol{X_{it}}$ represent the risk factors and $\boldsymbol{\beta}$ is a vector of regression parameters. Two random effects are defined in the model, $u_{it}$ is the segment-season specific random effect and $b_t$ is the segment only specific random effect. Both random effects are set to follow a normal distribution $b_i \sim N(0, 1/a)$, where $a$ is the precision parameter and it was specified a gamma prior as $a \sim \text{gamma}(0.001, 0.01)$.

Full Bayesian inference was employed in this study. The key 'hierarchical' part of these models is that $\varnothing$, the random effects $(u_{it}, b_t)$ is unknown and thus has its own prior distribution, $p(\varnothing)$. The joint prior distribution is (Gelman et al., 2004)

$$p(\varnothing, \theta) = p(\varnothing)p(\theta|\varnothing),$$

and the joint posterior distribution can be defined as

$$p(\varnothing, \theta|y) \propto p(\varnothing, \theta)p(y|\varnothing, \theta) = p(\varnothing, \theta)p(y|\theta).$$

Based on the above formulation, three models were considered in this paper: the fixed effects model with $(\gamma_1, \gamma_2) = (0, 0)$; the over-dispersed Poisson model with no correlation with $(\gamma_1, \gamma_2) = (1, 0)$; and the over-dispersed correlated Poisson model $(\gamma_1, \gamma_2) = (0, 1)$. For each model, three chains of 20,000 iterations were set up in WinBUGS (Lunn et al., 2000), 5000 iterations were used in the burn-in step.

## 5. Modeling results and discussions

### 5.1. Seasonal model

As stated and proved in the previous work (Ahmed et al., 2011a), significant seasonal effect exists on the chosen freeway segment. Totally 240 observations (120 segments × 2 seasons) were entered in the above defined models. For each observation, it represents the crash frequency for a specific homogenous segment in one season. For a segment with more than one crash occurrence, mean values of weather variables and traffic status variables from different crashes were calculated for this segment. Zero crash occurrence segments use the seasonal average values of the weather variables and traffic status variables in the final data set. Descriptive statistics of variables entered into the final models are summarized in Table 1.

Daily vehicle mile traveled (VMT) were estimated by multiplying segment length and AADT to represent crash exposure for each segment. One hour precipitation were adopted rather than the 10 min precipitation because rain or snow affect crash occurrence by influencing road surface condition, long-period precipitation can better reflect the road surface conditions.

Table 2 provides the estimations of the significant parameters for the seasonal model. Although three candidate models were considered, similar results for the significant parameters have been achieved. Geometric characteristic parameter (Grade index) has shown a consistent effect in the model as in the previous study results, which identify Grade −6 to −8% as the most hazardous slope. Moreover, the trends of Grade indexes indicate that the steeper slopes experience a higher crash frequency; and upgrade segments are safer than those downgrade segments with the same slope range. Also the LogVMT variable has an identical significant positive effect on crash frequency, which means the larger VMT

**Table 2**
Parameters estimates for the seasonal model.

| Model | Fixed effect | | | Random effect (uncorrelated) | | | Random effect (correlated) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Dry season | | | Snow season | | |
| | Mean | 2.5% | 97.5% | Mean | 2.5% | 97.5% | Mean | 2.5% | 97.5% | Mean | 2.5% | 97.5% |
| LogVMT | 0.77 | 0.6 | 0.92 | 0.7 | 0.4 | 1.0 | 1.0 | 0.7 | 2.0 | 0.6 | 0.3 | 0.9 |
| Av_visibility | −0.13 | −0.21 | −0.05 | −0.06 | 0.2 | 0.05 | – | – | – | −0.2 | −0.3 | −0.02 |
| Av_temp | −0.02 | −0.03 | −0.008 | −0.03 | −0.002 | −0.05 | 0.06 | 0.02 | 0.1 | −0.04 | −0.08 | −0.002 |
| Av_1hourprecip | 1.27 | −0.25 | 2.65 | 4.0 | 2.0 | 7.0 | – | – | – | 5.0 | 3.0 | 8.0 |
| S_1hourprecip | −0.76 | −1.54 | 0.11 | −2.0 | −3.0 | −0.9 | – | – | – | −2.0 | −4.0 | −1.0 |
| CAS (Avg. Speed) | −0.034 | −0.04 | −0.02 | −0.03 | −0.05 | −0.02 | −0.07 | −0.1 | −0.03 | −0.007 | −0.03 | 0.01 |
| Season [snow] | – | – | – | 2.0 | 1.0 | 3.0 | – | – | – | – | – | – |
| Season [dry] | – | – | – | – | – | – | – | – | – | – | – | – |
| Grade[1] | −1.52 | −2.58 | −0.65 | −2.0 | −3.0 | −0.5 | −2.0 | −3.0 | −0.6 | −2.0 | −3.0 | −0.6 |
| Grade[2] | −0.37 | −0.76 | 0.008 | −0.4 | −1.0 | 0.2 | −0.7 | −1.0 | −0.09 | −0.7 | −1.0 | −0.09 |
| Grade[3] | −0.85 | −1.25 | −0.46 | −0.8 | −1.0 | −0.3 | −0.8 | −1.0 | −0.3 | −0.8 | −1.0 | −0.3 |
| Grade[4] | −0.32 | −0.78 | 0.11 | −0.3 | −1.0 | 0.4 | −0.5 | −1.0 | 0.2 | −0.5 | −1.0 | 0.2 |
| Grade[5] | −1.04 | −1.67 | −0.47 | −1.0 | −2.0 | −0.2 | −1.0 | −2.0 | −0.4 | −1.0 | −2.0 | −0.4 |
| Grade[6] | −1.23 | −1.84 | −0.69 | −1.0 | −2.0 | −0.3 | −1.0 | −2.0 | −0.3 | −1.0 | −2.0 | −0.3 |
| Grade[7] | −0.45 | −0.92 | 0.00 | −0.3 | −1.0 | 0.5 | −1.0 | −0.4 | 0.3 | −1.0 | −0.4 | 0.3 |
| Grade[8](reference) | – | – | – | – | – | – | – | – | – | – | – | – |

increase the likelihood of more frequent crashes because of the higher exposure.

Several real-time weather variables have been included in the final models. Average visibility within the segment was significant with a negative sign, which indicates that a good visibility condition will decrease the crash occurrence. Two precipitation variables were included in the models, 1-h precipitation's mean value and its standard deviation. Average 1-h precipitation volume has a positive coefficient means larger precipitation increase the crash hazardousness. While the standard deviation of 1-h precipitation volume has a negative coefficient indicates that segments suffered sudden rain or snow are more dangerous than those segments suffering continuous precipitation. This means that drivers are driving much carefully through those consistent high precipitation areas, which might be because of warning signs in these frequent precipitation segments. Average temperature has a distinct coefficient sign in the third model; it has a positive coefficient in the dry season's model and a negative coefficient in the snow season's model. However this interesting result shows that less safe state is positively correlated with extreme temperatures (low during winter and high during summer) as found before in Malyshkina et al. (2009). The distinct effect of temperature in two seasons can only be captured by the random effect correlated model since it only reflects a negative influence on crash occurrence in the other two models.

For the real-time traffic variables, only the 5-min average speed of the crash segment during 5–10 min prior to the crash time was found to be significant. The CAS has a negative sign, which means that the crash occurrence likelihood increases as the average speed decreases 5–10 min before the crash occurrence. This result has been proved in several real-time crash prediction models (Ahmed et al., 2011b, 2012).

The DIC, recognized as Bayesian generalization of AIC (Akaike information criterion), was used to compare the performance of the three candidate models (Table 3). $\bar{D}$ is the measure of model fitting, $p_D$ is the effective number of parameters and DIC is a combination of these two measures. The two random effect models have relatively lower DIC than the fixed effect model, which implies that the overdispersion problem does exist and cannot be handled by the Poisson model. For the two random effect models, the correlated random effect model has a lower DIC value (516.8 vs. 534.3) with less effective number of parameters (47.6 vs. 48.4). Furthermore, the correlated random effect model provides two sets of distinct estimated parameters for the two seasons. This is beneficial for establishing different freeway management strategies to accommodate for distinct seasonal safety conditions.

**Table 3**
Model comparison.

| Models | $\bar{D}$ | $p_D$ | DIC |
|---|---|---|---|
| Fixed effect | 565.2 | 14.8 | 579.9 |
| Random effect (uncorrelated) | 485.8 | 48.4 | 534.3 |
| Random effect (correlated) | 469.2 | 47.6 | 516.8 |

### 5.2. Single-vehicle vs. multi-vehicle crash model

In this study, single-vehicle (SV) and multi-vehicle (MV) crashes are analyzed with the consideration of differences between the two crash types. The same three modeling approaches are applied and a summarized descriptive statistics of the variables can be found in Table 4.

The coefficient estimates of the beta coefficients of the significant parameters are shown in Table 5. Only real-time weather variables and RTMS traffic variables were proved to be significant in the three candidate models. A binary MV indicator was created and included in the models to identify whether the different characteristics of SV and MV crashes exist.

For the real-time weather variables, only the two precipitation descriptive variables present a significant influence on crash occurrence. Average 1-h precipitation and standard deviation of 1-h precipitation have the consistent beta coefficient signs as the seasonal model. As explained above, these imply that crashes are more likely to happen at the segments with a sudden high precipitation (rain or snow).

In the fixed effect and uncorrelated random effect model, average speed at the crash segment and average occupancy at the upstream segment have been indicated as significant factors. Lower speed at the crash segment and higher occupancy at the upstream segment 5–10 min before the crash time increase the likelihood of crashes. This could be an indication of queuing. The binary indicator of SV and MV crashes was significant in both models, which represents that these two crash types have distinct crash mechanisms.

More real-time traffic variables were found significant in the correlated random effect model. For the SV crashes, except for the two precipitation variables, only the average speed prior to the crash time was significant. This implies that SV crashes are more influenced by the weather and are not influenced by the upstream and downstream traffic status which always turned out to be run-off road crashes. For the MV crashes, except for the average speed

**Table 4**
Summary of variables descriptive statistics for single-vehicle and multi-vehicle crash model.

| Variables | Description | Mean | Std dev. | Minimum | Maximum |
|---|---|---|---|---|---|
| Crash Frequency | Crash frequency counts for the segment | 1.09 | 1.7 | 0 | 12 |
| Av_1hourprecip | Average value of 1 h precipitation (rain/snow) before the crash | 0.061 | 0.24 | 0 | 2.23 |
| S_1hourprecip | Standard deviation of 1 h precipitation (rain/snow) before the crash | 0.24 | 0.35 | 0 | 3.72 |
| CAS | Average speed for the crash segment | 52.2 | 10.8 | 7.0 | 68.0 |
| Multi | Single vehicle = 0, multi vehicle = 1 | 0.5 | 0.5 | 0 | 1.0 |
| DAO | Average occupancy of downstream detector | 4.48 | 3.54 | 1.28 | 33.78 |
| UAO | Average occupancy of upstream detector | 4.15 | 1.93 | 1.52 | 19.96 |
| DCO | Coefficient of variance of speed at downstream detector | 0.10 | 0.046 | 0.025 | 0.41 |

**Table 5**
Parameters estimates for single vehicle vs. multi-vehicle model.

| Model | Fixed effect | | | Random effect (uncorrelated) | | | Random effect (correlated) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Single vehicle | | | Multi vehicle | | |
| | Mean | 2.5% | 97.5% | Mean | 2.5% | 97.5% | Mean | 2.5% | 97.5% | Mean | 2.5% | 97.5% |
| Av_1hourprecip | 8.75 | 7.27 | 10.22 | 8.0 | 7.0 | 10.0 | 10.0 | 9.0 | 20.0 | 7.0 | 5.0 | 10.0 |
| S_1hourprecip | −6.27 | −7.34 | −5.18 | −6.0 | −7.0 | −5.0 | −8.0 | −10.0 | −6.0 | −5.0 | −7.0 | −4.0 |
| CAS (Avg. Speed) | −0.022 | −0.03 | −0.01 | −0.03 | −0.04 | −0.01 | −0.03 | −0.04 | −0.007 | −0.03 | −0.05 | −0.005 |
| UAO (Avg. Occ. Up) | 0.03 | −0.009 | 0.076 | 0.06 | −0.007 | 0.1 | – | – | – | 0.1 | 0.02 | 0.3 |
| DCO (Coe. Spd. Dn) | – | – | – | – | – | – | – | – | – | 5.0 | 0.8 | 9.0 |
| DAO (Avg. Occ. Dn) | – | – | – | – | – | – | – | – | – | −0.1 | −0.2 | −0.01 |
| Multi | 0.28 | 0.03 | 0.53 | 0.3 | −0.01 | 0.5 | – | – | – | – | – | – |


**Fig. 2.** Multi-vehicle crash model illustration.

at the crash segment, average occupancy of upstream and downstream and coefficient of variance in speed at downstream came out to be significant. Coefficient of variance of downstream speed (DCO) has a positive beta coefficient, while downstream average occupancy with a negative sign and upstream average occupancy have a positive sign. However, the four traffic variables that are significant in the multi-vehicle model can be explained as: congestion happening which result in a queuing area upstream. This causes high occupancy upstream and low speed within the congested area; vehicles pass through the congested segment and start to speed up at different rates leading to low occupancy and high variation of speed downstream, which makes the downstream segment a turbulent area. These conditions together, result in a high probability of multi-vehicle crashes within the congested area, i.e. between the queue and turbulent areas, which is illustrated in Fig. 2.

The same model evaluation methods were employed for the SV and MV models (Table 6). Again the two random effect models have significantly lower DIC compared to the fixed effect model. Within the two random effect models, the correlated random effect model is superior. Although the correlated random effect model have a larger number of effective parameters (46.1 vs. 43.7), it has lower DIC (556.8 vs.564.4) and more important is that the two sets of parameters are useful to identify the hazardous factors of the SV and MV crashes.

**Table 6**
Model comparison.

| Models | $\bar{D}$ | $p_D$ | DIC |
|---|---|---|---|
| Fixed effect | 598.4 | 5.9 | 604.4 |
| Random effect (uncorrelated) | 520.7 | 43.7 | 564.4 |
| Random effect (correlated) | 510.7 | 46.1 | 556.8 |

## 6. Conclusion

Crash occurrence on a mountainous freeway is highly influenced by the weather conditions. Distinct seasonal weather conditions reflect on the crash frequencies and crash contributing factors. To fully account for the weather influence on crash occurrence, real-time weather data were used in this study. In most previous studies, weather data were estimated from aggregated weather records or crash reports, which end up in the loss of important information. In this study, weather data were recorded by up to 6 weather stations along the 15-mile freeway segment. These weather stations provide real-time information about the adverse weather conditions, which were demonstrated by the models of being highly related to crash occurrence.

In addition to the real-time weather data, real-time traffic variables prior to the crash time have been included in the models. Unlike the fixed value of speed limits, etc., incorporating real-time traffic variables have the benefit of explaining different characteristics of crashes under various scenarios. Furthermore, those significant variables are helpful to design the active traffic management system in future studies.

For the methodological part, random effect models have been proven to be superior to the fixed effect models since they can handle the overdispersion problem of the data. Moreover, the correlated random effect models provide better fitting, and different sets of parameters for distinct scenarios can help the researchers understand more the diverse crash occurrence mechanisms.

The results of this study suggest that real-time weather information and traffic statuses are essential to address the crash frequency models, particularly for mountainous freeways with adverse weather conditions. Also, different strategies of freeway management should be implemented during these two distinctive seasons and to address different SV and MV crash characteristics. SV and MV models share some significant variables such as precipitation and average speed, which demonstrate the correlated mechanisms of these two crash types. Besides, the correlated random effect model outperformed the other models also indicate that SV and MV crashes are correlated within the same segment. However, careful comparisons of the model results show that SV crashes are more related to the weather condition and vehicle speeds, while more traffic variables are found to be statistically significant in the MV model. These findings indicate that the two crash types should be modeled with the concern of different characteristics.

In addition, the results also shed some light on the policy implication to bring down the crash occurrence. For the studied road segment, results indicated that crash occurrence in the snow season have clear trends associated with adverse weather situations (bad visibility and large amount of precipitation); weather warning systems can be employed to improve road safety during the snow season. Furthermore, different traffic management strategies should be developed according to the distinct seasonal influence factors. In particular, sites with steep slopes need more attention from officials and decision makers especially during snow seasons to control the excess crash occurrence. Moreover, distinct strategy of freeway management should be designed to address the differences between SV and MV crash characteristics. For MV crashes as they are more associated with roadway traffic conditions, speed management system like a Variable Speed Limit system can be introduced. For example, if a queue has been detected and backed up upstream of the roadway section, in order to reduce the multi-vehicle crash occurrence, the speed limits downstream of the queuing area should be lowered and warning messages like "gradual speed increase" should be displayed. These examples are provided here only to show how policy development can benefit from the crash frequency models with real-time traffic and weather information, more efforts and research are needed to reach efficient and driver friendly policies.

While the previous studies have addressed the effects of roadway geometrics on crash frequencies, this study makes a step forward by analyzing the weather effects and traffic status effects on crash frequencies under different scenarios.

## Acknowledgment

The authors wish to thank the Colorado Department of Transportation for funding this work and providing the data.

## References

AAA Foundation for Traffic Safety, 1999. Did you know? Congestion and crashes. http://www.aaafoundation.org/pdf/congest.pdf (accessed 25.3.12).

Ahmed, M., Abdel-Aty, M., Yu, R., 2012. Assessment of the iteraction between crash occurence, mountainous freeway geometry, real-time weather and avi traffic data. In: Compendium of papers CD-ROM, Transportation Research Board 2012 Annual Meeting, Washington, D.C.

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011a. Exploring a bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis and Prevention 43, 1581–1589.

Ahmed, M., Yu, R., Abdel-Aty, M., 2011b. Safety applications of automatic vehicle identification and real-time weather data on freeways. In: Presented at the 18th World Congress on Intelligent Transport Systems, Orlando.

Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis and Prevention 39, 657–670.

Chang, L.Y., Chen, W.C., 2005. Data mining of tree-based models to analyze freeway accident frequency. Journal of Safety Research 36, 365–375.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized interesections. Accident Analysis and Prevention 35, 253–259.

Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models. Accident Analysis and Prevention 42, 1273–1282.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2004. Bayesian Data Analysis, 2nd ed. Chapman &Hall/CRC.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multievel/Hierarchical Models. Cambridge University Press.

Guo, F., Wang, X., Abdel-Aty, M., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis and Prevention 42, 84–92.

Ivan, I., Pasupathy, R., Ossenbruggen, P., 1999. Differences in causality factors for single and multivehicle crashes on two-lane roads. Accident Analysis and Prevention 31, 695–704.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and bayesian analysis in traffic safety. Accident Analysis and Prevention 42, 1556–1565.

Jung, S., Qin, X., Noyce, D., 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis and Prevention 42, 213–224.

Kononov, J., Lyon, C., et al., 2011. Relating flow speed and density of urban freeways to functional form of an SPF. In: Compendium of Papers CD-ROM, Transportation Research Board 2011 Annual Meeting, Washington, DC.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34, 149–161.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44, 291–305.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10 (4), 325–337.

Malyshkina, N., Mannering, F., Tarko, A., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accident Analysis and Prevention 41, 217–226.

Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and bayes versus empirical bayes. Transportation Research Record 1840, 31–40.

Miaou, S.P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. Journal of Transportation and Statistics 6, 33–58.

Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis and Prevention 36 (6), 973–984.

Shankar, V., Albin, R., Milton, J., Mannering, F., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transportation Research Record 1635.

Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of Safety Research. 27 (3), 183–194.

Ntzoufras, I., 2009. Bayesian Modeling Using Winbugs. Wiley.

Shively, T., Kockelman, K., Damien, P., 2010. A bayesian semi-parametric model to estimate relationships between crash counts and roadway characteristics. Transportation Research Part B 44, 699–715.

Usman, T., Fu, L., Miranda-Moreno, L., 2010. Quantifying safety benefit of winter road maintenance: accident frequency modeling. Accident Analysis and Prevention 42, 1878–1887.

Yaacob, W.F., Lazim, M.A., Wah, Y.B., 2010. Evaluating spatial and temporal effects of accidents likelihood using random effects panel count model. In: Presented at the 2010 International Conference on Science and Social Research, Kuala Lumpur, Malaysia.