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

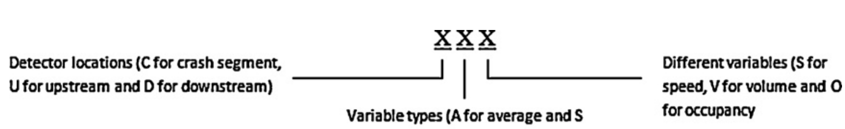

Fig. 1. Nomenclature method for traffic variables.