Real-time Crash Prediction for Expressway Weaving Segments

Ling Wang[a,*], Mohamed Abdel-Aty[a], Qi Shi[a], Juneyoung Park[a]

[a] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, Florida 32816, United States

ABSTRACT

Weaving segments are potential recurrent bottlenecks which affect the efficiency and safety of expressways during peak hours. Meanwhile, they are one of the most complicated segments, since on- and off-ramp traffic merges, diverges and weaves in the limited space. One effective way to improve the safety of weaving segments is to study crash likelihood using real-time crash data with the objective of, identifying hazardous conditions and reducing the risk of crashes by Intelligent Transportation Systems (ITS) traffic control. This study presents a multilevel Bayesian logistic regression model for crashes at expressway weaving segments using crash, geometric, Microwave Vehicle Detection System (MVDS) and weather data. The results show that the mainline speed at the beginning of the weaving segments, the speed difference between the beginning and the end of weaving segment, logarithm of volume have significant impacts on the crash risk of the following 5-10 minutes for weaving segments. The configuration is also an important factor. Weaving segment, in which there is no need for on- or off-ramp traffic to change lane, is with high crash risk because it has more traffic interactions and higher speed differences between weaving and non-weaving traffic. Meanwhile, maximum length, which measures the distance at which weaving turbulence no longer has impact, is found to be positively related to the crash risk at the 95% confidence interval. In addition to traffic and geometric factors, wet pavement surface condition significantly increases the crash ratio by 77%. The proposed model along with ITS, e.g., ramp metering, Dynamic Message Sign (DMS), and high friction surface treatment can be used to enhance the safety of weaving segments in real-time.

Keyword: expressway weaving segments, real-time crash analysis, multilevel Bayesian logistic regression model, maximum length

* Corresponding author. Tel.:+1 407 823 0300. Email:lingwang@knights.ucf.edu

1 BACKGROUND

Expressways play a vital role in serving mega-cities. They increase the travel speed and reduce the travel time for daily traffic, especially for mid and long trips' traffic. The efficiency and safety of expressways are significantly constrained by bottlenecks. Weaving segments may easily become recurrent bottlenecks during peak hours because the capacity of weaving segments is much lower than that of basic expressway segments when controlling for the free-flow speed, truck percentage, etc. (HCM 2000). Hence, understanding the safety and efficiency of weaving segments are important and need to be addressed.

Weaving is generally defined as the crossing of two or more traffic streams traveling in a same direction along a significant length of highway without the aid of traffic devices (except for guide signs) (HCM 2010). When a merging segment is closely followed by a diverging segment and the two are joined by auxiliary lane(s), a weaving segment is formed. Normally, there are three movements in weaving segments, namely mainline-to-mainline, mainline-to-ramp and ramp-to-mainline movement. The traffic movements are shown in Figure 1.

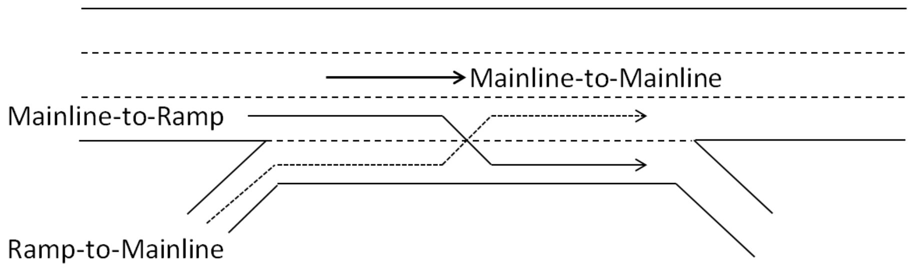

Figure 1 Ramp Weaving Area Traffic Movements

In addition to their importance, weaving segments are also one of the most complicated segments since on- and off-ramp traffic merge, diverge and weave in the limited space. When weaving segment lengths are limited, they are not sufficient for merging and diverging to operate independently. Vehicles entering and exiting expressways have to compete for lane-changing opportunities. This may easily lead to crashes. The occurrence of a crash in weaving segments can bring about serious results. On-ramp vehicles might not be able to get on expressways and then may queue up along ramps or have to change their routes. Off-ramp traffic may have difficulty to get off mainlines and would queue up on mainlines. Moreover, if the crash cannot be cleared in time, the queue may block all traffic, including the non-weaving and weaving traffic. These may significantly reduce the capacity and level of service of weaving segments.

Because of their importance and complexity, weaving segments have already gained considerable attention from researchers since the publication of the first Highway Capacity Manual in 1950. The capacity studies of weaving segments have been a major focus. Researchers have estimated the capacity of weaving segments based on empirical methods, theoretical models, and simulation modeling tools (Stewart et al. 1996, Lertworawanich and Elefteriadou 2001, 2003, 2010). Other studies have also been on level of service (LOS), weaving behavior, geometric design of weaving segments, etc. (Kwon et al. 2000, Roess and Ulerio 2000, 2009, 2010). In addition to the studies above, there were also research on weaving segments' safety. Some researchers analyzed crashes which occurred in different types of weaving segments (Golob et al. 2004, Liu et al. 2009). They discovered that the crash characteristics of different types of weaving segments were not the same, e.g., Type C weaving segments had the lowest crash frequency, Type B was the most dangerous weaving segment. The relationship between geometric design and weaving segment safety was also explored by Qi et al. (2004). Their results showed that shorter segment length, more required lane changes for diverging vehicles, higher diverging traffic volume and lower merging traffic volume may result in higher crash rates. The research above provides guidance when building a new weaving segment, e.g., deciding the number of lanes to satisfy the weaving traffic demand. They can also be used to evaluate the operation or the long-term safety performance of weaving segments. However, with respect to predicting crash risk in real-time or understating crash mechanism based on microscopic data, there were not enough studies for weaving segments.

Studies using real-time data have been proven to be a possibly effective way to mitigate crashes on expressways. They can estimate the crash likelihood or explain the crash mechanism, and then enhance the safety by traffic management or control. There has been a great deal of research which linked crash risk with traffic, weather and geometry factors since 1995. The majority of previous studies were for mainlines. Some researchers identified significant traffic precursors and built models to estimate crash likelihood (Oh et al. 2001, Abdel-Aty et al. 2004, Xu et al. 2013). In addition to traffic factors, some studies also took weather into consideration. Visibility, rainfall and binary weather condition variable, which indicated whether the weather is severe or not, were found to be significantly related to crash risk (Madanat and Liu 1995, Lee et al. 2002, Abdel-Aty and Pemmanaboina 2006, Christoforou et al. 2011, Ahmed et al. 2012). Road geometry was not universally implemented in crash risk estimation models. Some of the researchers used matched case-control models to eliminate the impact of geometry design (Abdel-Aty and Pemmanaboina 2006, Zheng et al. 2010, Christoforou et al. 2011, Abdel-Aty et al. 2012). In addition to mainlines, there were also a few studies on ramps (Lee and Abdel-Aty 2006, Wang et al. 2015). The Log-linear model and the Bayesian logistic regression model were used to quantify the impact of ramp type, ramp configuration, time of day, traffic and weather on ramp crash risk.

The real-time crash studies above did not divide the highway into several segment types, e.g., basic segment, weaving segment. However, some researchers discovered that the crash mechanisms of different segment types were not the same. Hossain and Muromachi (2012, 2013a, b) studied basic mainline segments and ramp vicinities. They concluded that congestion index and the speed difference between upstream and downstream had the biggest impact on crash number and crash type of basic segments; on the other hand, ramp flow has the highest influence in determining crash types within ramp vicinities. Though weaving segments are related to ramp vicinity, their study result cannot be directly used as a weaving segment study. First, in these studies, the ramp vicinities were 375 meters upstream and downstream of entrance and exit ramps. The ramp vicinities are not weaving segments if merging and diverging movements can be operated independently. Second, researchers focused on ramp vicinity, hence weaving related variables, e.g., weaving volume ratio, weaving segment length, were not considered in these studies.

From the review above, it is not hard to conclude that there were numerous studies on weaving segments and real-time crash analysis independently. However, real-time crash prediction for weaving segments was not studied. Hence, this study conduct real-time crash prediction for weaving segments. Three types of factors are considered in the model building. Traffic factor is essential and the traffic turbulence is one of the most important factors of crashes (Wang et al. 2015). The geometric characteristics, e.g., segment length, number of lanes involved in weaving, are more site-specified for weaving segments. Exploring the connection between geometric characteristic and crash risk would be helpful in finding hazardous weaving segments. Meanwhile, in addition to the traffic and geometric factors, the weather factors are also important. The frequent lane-changing, deceleration and acceleration make the traffic in weaving segments vulnerable to severe weather, e.g., rain, snow.

The paper is organized into five sections. The second section describes the research methodology which is used in building the model. The third section describes the data, defines the variables and presents the crash characteristics. The fourth section shows the model results and also discusses the findings of the model. The fifth section summarizes the findings, conclusions and limitations of the paper.

2 METHODOLOGY

This study built a multilevel Bayesian logistic regression model to estimate the likelihood of crashes in weaving segments. Compared with traditional regression models, Bayesian models do not treat the coefficients as fixed values but follow distributions. Overestimated odds ratio may occur when the sample size is limited (Nemes et al. 2009). The Bayesian inference can effectively avoid this issue.

Bayesian multilevel model is applied in this study. Multilevel modeling allows multilevel data structures to be properly estimated, and outperforms classical regression model in predictive accuracy (Gelman 2012). It can consider the variation due to the hierarchical structure in the data, and also allows the simultaneous examinations of the effects of segment-level and individual-level variables on outcomes while accounting for the non-independence of observations within a cluster.

In this study, the crash risk is jointly influenced by geometry, traffic and weather. To be more specific, crash risk is impacted by two levels of variables, segment-level and individual-level. The observations within a segment share same geometric information, and other unmeasured factors. Hence, the geometric variables are put in the segment-level. On the other hand, every observation has its own traffic and weather condition, so traffic and weather variables are in the individual-level.

Suppose N observations (individual-level) have been collected, and they are within

M weaving segments (segment-level). With two-level structure data, three sets of equations are formulated: individual-level model, segment-level model and combined model. In this research an event (Y_{ij}) has binary outcome, crash and non-crash, so a binomial logit model is conducted. The multilevel binomial logistic model is well explained in Kim et al. (2007).

The possibilities of crash (Y_{ij}=1) and non-crash (Y_{ij}=0) for this event are converted into p_{ij} (Y_{ij}=1) and 1-p_{ij} (Y_{ij} =0), respectively. The calculation of p_{ij} are as follows:

Individual-level model:

$$Y_{ij} \sim \text{dbern}\ (p_{ij})$$

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{oj} + \sum_{r=1}^{R} \beta_{rj} x_{rij} \qquad (1)$$

Where Y_{ij} follow a Bernoulli distribution whose success probability is p_{ij}. β_{oj} is the intercept at the individual-level model, β_{rj} the regression coefficient of predictor x_{rij}, x_{rij} is the r^th explanatory variable in individual-level for i^th observation in j^th weaving segment, e.g., volume, speed, visibility.

In our study, the intercept β_{oj} is assumed to vary across weaving segment as a function of geometric variables, i.e., segment short length, base length, number of lanes, configuration, but the slope coefficient β_{rj} is assumed not to vary across segments. Then, the intercept β_{oj} and variable coefficient β_{rj} are calculated as follows,

Segment-level model:

$$\beta_{oj} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} w_{qj}$$

$$\beta_{1j} = \gamma_{10}$$

$$\vdots$$

$$\beta_{pj} = \gamma_{p0} \qquad (2)$$

Where γ_{00} is the intercept of segment-level model, γ_{0q} is the regression effect of q^th segment-level variable, and Q is the number of segment-level variables.

Substituting Equation (2) into Equation (1), the combined model is as follows,

Combined model:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} w_{qj} + \sum_{r=1}^{R} \gamma_{r0} x_{rij} \qquad (3)$$

Three chains of 10,000 iterations were set up in Winbugs. In order to eliminate the influence of the starting values, the first half of the iteration (burn-in step) is discarded and the second half is used in the analysis (Gelman et al. 2014). In model estimation, with no prior knowledge of the value of parameters for weaving segments, the priors are set as non-informative with zero mean and a large variance, i.e., Normal (0, 10^6) (Xu et al. 2014a).

Model convergences were checked by examining the MCMC trace plots of exploratory variables (Spiegelhalter et al. 2003, Yu and Abdel-Aty 2013b). Deviance information criterion (DIC) and Area Under the Curve (AUC) were used to measure the model goodness of fit. A smaller DIC or a higher AUC indicates better model fitting.

3 EXPERIMENTAL DESIGN AND DATA COLLECTION

3.1 Study Area and Data

The researchers study the 22-mile SR 408 in Central Florida. Four datasets are collected, i.e., crash, traffic, weather, and geometry data. The data are from July 2013 to April 2015. However, due to the absence of traffic data in April, 2014, only the other 21 months data are used.

The crash data are from Signal Four Analytics. It is an interactive, web-based system designed to support the crash mapping and analysis needs of law enforcement, traffic engineering, transportation planning agencies, and research institutions in the state of Florida (University of Florida 2015). It provides information for all reported crashes, e.g., crash time and location, type and severity. One hundred and sixty five crashes are identified in the studied weaving segments during the study period. The traffic data are provided by the Central Florida Expressway Authority (CFX). The traffic data, which include traffic count, lane occupancy and speed, are automatically archived at one-minute interval by MVDS.

As for the weather data, they are collected from the National Climate Data Center (NCDC) which records the weather for Orlando Executive Airport (ORL). The airport is about 0.5 miles north of the middle of SR 408. Its weather data are continuously monitored. If weather condition does not change, the data are recorded every one hour. Once the weather parameters change, the weather station records the new weather state at once. The weather dataset includes weather type, wind direction and speed, temperature, visibility, hourly precipitation, etc.

The geometric data are collected manually by using ArcGIS map. There were 17 segments in which off-ramp is closely followed by on-ramp on the studied expressway. Among these 17 segments, the configuration of one segment is different from others. The number of lanes for on-ramp is 2 for this segment, but all the others are 1. This special case is excluded from the study.

3.2 Experimental Design

The two segment lengths which are relevant to this paper are illustrated in Figure 2. Within short length (Ls), lane changing is prohibited or discouraged. However, this does not mean lane changing only happens within this length. Some lane changing takes place over the solid white lane and is within base length (Lb) (HCM 2010). Hence, the study area is within Lb, and crashes happened in this area are collected.

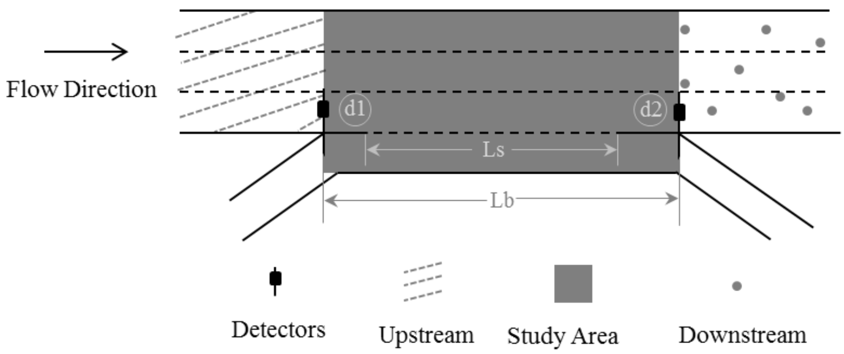

Figure 2 Segment Length and Experimental Design

The location of traffic detectors is illustrated by Figure 2. The d1 detector can detect traffic of all lanes which are at the beginning of the segments, including mainline and ramp. The d2 can also detect all lanes at the end of segments. Because of the high coverage of the MVDS system on SR 408, all studied segments had two detectors, which were d1 and d2. Hence, the traffic data for all weaving segments were available.

In order to reduce traffic data noise, they were aggregated into 5-minute intervals. The same as previous studies (Hossain and Muromachi 2013b, Xu et al. 2013, Yu and Abdel-Aty 2013a), the researchers also extracted the traffic data which were 5-10 minutes prior to crash occurrence. The reason why 5-10 minute data are used is as follows. First, compared to the traffic data which are 10-15 minutes prior to cases, it can provide more accurate crash-precursor condition. Second, compared to the traffic data which are 0-5 minute prior to cases, it can provide sufficient time for the traffic management center to analyze, react and announce warning information to the drivers. What is more important, the recorded crash time is normally the time when drivers call the policeman. Hence, the recorded crash time is actually the time after crash. If the 0-5 minute data are used in the model estimation, some traffic condition is already impacted by the crash and is not crash-precursor condition any more. It is assumed that a crash is reported in 5 minutes after its occurrence on the studied expressway which is with high traffic volume. Therefore, the traffic which is 5-10 minute before the recorded crash time is not influenced by crash yet.

For the weather data, three parameters are selected, i.e., weather type, hourly precipitation, and visibility. The former two parameters are combined into a binary predictor, named road surface condition. If weather type includes TS (thunderstorm), RA (rain) and DZ (drizzle), or hourly precipitation higher than 0, they indicate it rains. Combining these two parameters can provide more accurate and complete rain information. Pavement surface condition of an event is wet if the weather type or hourly precipitation shows that it rained in the past one hour.

Not all crashes, which happen within a segment which is formed by merging closely followed by diverging, can be regarded the weaving segment crash. HCM 2010 proposed a parameter, named the maximum weaving segment length. It is simplified into maximum length in the following of this paper. It is the length at which weaving turbulence no longer has impact on the operation within the segment, or alternatively, on the capacity of the weaving segment. The maximum length represents weaving influence length. The value of this parameter is not fixed, but changeable according to the geometric and traffic condition. The HCM 2010 also expressed that when the short length of a segment is larger than the maximum length, the segment is not a weaving segment but a regular merging area followed by a diverging area. The calculation of this parameter is given in the next section. The study's objective is weaving segments, so we only chose cases which happen within weaving segments. In order to achieve it, the maximum lengths of all cases are dynamically calculated and compared with their short length in a 5-minute interval. If an observation's maximum length is larger than its short length, it is kept; if an observation's maximum length is less than its short length, it is discarded.

After the processes above, all datasets were combined together. One hundred and twenty five crashes and 1,250 non-crash cases are filtered out in the study. The non-crash cases are randomly chosen from the non-crash datasets by SAS. All observations happened in weaving segments and have complete traffic, geometric and weather information. Meanwhile, in order to ensure the purity of the non-crash observations, we ensured that no crashes happened within 5 hours before and after a non-crash event.

3.3 Variable Definition

The definitions and acronyms of variables which can be obtained from the traffic, geometric and weather data are shown in Table 1.

The speed standard deviation is the speed changes over time. In order to obtain this value, the average speed of a mainline section for every one minute is firstly obtained, then standard deviation of speed is calculated based on a 5-minute interval.

Table 1 List of Variables

Data	Symbol	Description
Traffic*	Bm_spd	Average speed at the beginning of weaving segments (mile/h)
	Bm_vol	Vehicle count per lane at the beginning of weaving segments (vehicles)
	Bm_occ	Average lane occupancy at the beginning of weaving segments (%)
	Bm_std_spd	speed standard deviation at the beginning of weaving segments (mile/h)
	Onr_spd	Average speed for on-ramp (mile/h)
	Onr_vol	Total vehicle count for on-ramp (vehicles)
	Onr_occ	Average lane occupancy for on-ramp (%)
	Em_spd	Average speed at the end of weaving segments (mile/h)
	Em_vol	Vehicle count per lane at the end of weaving segments (vehicles)
	Em_occ	Average lane occupancy at the end of weaving segments (%)
	Em_std_spd	speed standard deviation at the end of weaving segments (mile/h)
	Offr_spd	Average speed for off-ramp (mile/h)
	Offr_vol,	Total vehicle count for off-ramp (vehicles)
	Offr_occ	Average lane occupancy for off-ramp (%)
	V_{FF}	Mainline-to- mainline vehicle count (vehicles)
	Volume	Total traffic count in the weaving segment (vehicles)
	VR	Weaving volume ratio, weaving volume over total traffic count (%)
	Bm_Em_spd	Speed difference between the beginning and end of weaving segment, equal to Bm_spd- Em_spd (mile/h)
Geometry	Ls	Short length, distance between the end points of any barrier markings (solid white lines) that prohibit or discourage lane changing (feet)
	Lb	Base length, distance between points in the respective gore areas where the left edge of the ramp-traveled way and the right edge of the freeway-traveled way meet (feet)
	N_{WL}	Number of lanes from which a weaving maneuver may be made with one or no lane changes (lane)
	N	Number of lanes within the weaving segment (lane)
	LC_{RF}	Minimum number of lane changes that must be made by a single weaving vehicle moving from the on-ramp to the expressway (lane)
	LC_{FR}	Minimum number of lane changes that must be made by a single weaving vehicle moving from expressway to off-ramp (lane)
Weather	Visibility	The distance at which an object or light can be clearly discerned (miles)
	Surface	1=if the pavement surface condition is wet; 0=otherwise

* All traffic data are measured in a 5-minute interval and in the weaving segment

In addition to the variables shown in Table 1, there are three variables, i.e., LC, LC_{min}, and L_{max}. These three variables are obtained from equations (4) to (6). All other variables used in the computation are already defined in Table 1. Among these three variables, the calculations of LC_{min} and L_{max} are defined by the Highway Capacity Manual 2010.

$$LC = \begin{cases} 0 & if\ LC_{RF} = 1\ and\ LC_{FR} = 1 \\ 1 & if\ LC_{RF} = 0\ \ or\ \ LC_{FR} = 0 \end{cases} \tag{4}$$

$$LC_{min}=(LC_{RF}\times \text{Onr\_vol})+ (LC_{FR}\times \text{Offr\_vol}) \tag{5}$$

$$L_{max}=(5728(1+VR)^{1.6}-1566N_{WL})/1000 \tag{6}$$

Equation (4) integrates LC_{RF} and LC_{FR} and generates a new binary variable. In our study, both LC_{RF} and LC_{FR} for all weaving segments only have two values which are 0 and 1. Meanwhile, their values are not equal to 0 simultaneously. By integrating, we can use one variable to represent two important parameters, LC_{RF} and LC_{FR}. Figure 3 shows the weaving segment configuration types based on the newly defined LC. In the studied area, 10 weaving segments' LC is 0, and 6 weaving segments' LC is 1.

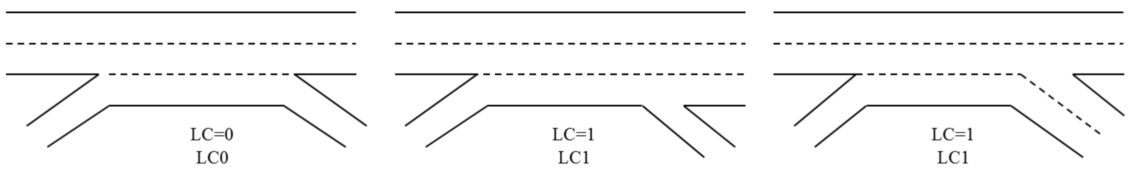

Figure 3 Configuration of weaving segments

LC_{min} stands for the minimum rates of lane changing that must exist for all weaving vehicles to complete their weaving maneuvers successfully in a 5-minute interval. The segments with higher LC_{min} have higher crash risk when controlling for the configurations. However, when the configurations are different, it's not comparable because the crash mechanisms for different configurations may vary.

L_{max} is the maximum length and can be also called weaving influence length. L_{max} is decided by weaving volume ratio (VR) and number of lanes from which weaving maneuvers may be made (N_{WL}). The higher VR, the higher weaving influence length.

3.4 Crash Characteristics

There were 85 crashes happened in LC0 weaving segments, and 80 in LC1 weaving segments. The number of crashes per segment was 8.5 for LC0 and 13.3 for LC1. In the study period, the average ADT of LC1 weaving segments is 1.186 times of that of LC0, the million vehicle-miles traveled VMT of LC1 is 1.108 times of that of LC0. Meanwhile, the segment length of these two types is almost the same. However, the average crash number of LC1 weaving segment is 1.564 times of that of LC0, which is significantly higher than traffic and VMT ratio. This indicates that LC1 may have a higher risk than LC0. The result is similar to a previous study by Liu et al. (2009).

There are one possible reason for this phenomenon. For LC0, the auxiliary lane is almost fully occupied by weaving vehicles and the lane which is close to the auxiliary lane shared by weaving and non-weaving vehicles. But for weaving segment, in which the minimum lane change for off-ramp vehicle is 0, there is one through lane for weaving vehicles. In addition, the two lanes which are adjacent to the through lane are also used by weaving vehicles. Hence, more lanes are involved in weaving movements, and then more non-weaving vehicles are affected by weaving vehicles.

The crash severity, number of vehicles involved in crash and crash type information of each weaving configuration are shown in Table 2. The Chi-square test indicated that weaving configuration does not have a significant impact on crash severity and number of vehicles involved in the crash. However, it clearly demonstrates that weaving configuration has significant impact on crash type at the 95% confidence interval.

Table 2 Crash Characteristics

	LC0	LC1	Chi-square	p-value
Crash Severity				
Injury	25	23	0.0087	0.9255
PDO*	60	57		
Number of Vehicle Involved				
1	22	13	2.2879	0.1304
More than 1	63	67		
Crash Type				
Rear End	28	48	12.2520	0.0066
Sideswipe	22	13		
Off Road	18	9		
Other	17	10		

* Property Damage Only

From Table 2, we can find 76 out of 165 (46.1%) crashes were rear-end crashes, and rear-end crash has the highest likelihood of occurrence in weaving segments. A previous paper by Golob et al. (2004) also discovered the similar result. LC1 weaving segments tend to have more rear-end crashes as shown. For LC0, the weaving vehicles change lane as soon as they get the opportunity, and they tend to use the beginning portion of the auxiliary lane (Kwon et al. 2000). At the beginning portion, the speed difference between merging and diverging vehicles does not vary significantly. The merging vehicles are in low speed due to the speed limitation of the on-ramps, and the diverging vehicles also are at low speed since they have to adjust to off-ramps' speed limit. However, for LC1, in addition to the weaving interactions at the beginning portion, a significant number of entering vehicles meet the exiting vehicles at the end of the weaving segments. For the weaving segment, in which the minimum lane change for on-ramp traffic is 0, a large number of entering vehicles do not change their lane and keep on the through lane where all exiting vehicles have to pass in a low speed. For the weaving segment, in which the minimum lane change for off-ramp traffic is 0, plenty of exiting vehicles take the through lane which all entering vehicles have to use. There exists a big speed difference between these entering and exiting vehicles for LC1. Entering vehicles are in high speed to follow speed limits on mainlines, but exiting vehicles are in low speed to follow the speed limits on ramps separately. Under this situation, a rear-end crash may happen.

4 MODEL ESTIMATION

A multilevel Bayesian logistic regression model is built to estimate the crash risk for weaving segments using traffic, weather and geometric variables. The ratio of crash over non-crash is 1:10. The whole dataset, which includes 125 crash and randomly selected 1,250 non-crash observations, is randomly split into training and validation datasets with a ratio of 70:30.

Beginning with all variables considered, each variable is tested whether it was statistically significant to the target variable. The insignificant variables are eliminated from the next model building step. Later, in order to select the most significant and not highly correlated variables, Random Forests is used to rank the variable's importance and Pearson correlation test was done. Random Forests are a combination of tree predictors and are robust with respect to noise. One important implementation of Random Forests is estimating the variable importance (Breiman 2001). If two variables are found to be highly correlated (coefficient>0.4), the variable which is more important is chosen for further analysis.

The variables selected above are put in the model estimation, the result shows that only speed at the beginning of weaving segment, volume, speed difference between the beginning and the end of a weaving segment, maximum length, weaving configuration and pavement surface condition are significant in the presence of other variables.

The T-Test shows that speed difference, volume and maximum length have significant positive impact on crash. The value of these variables which are 5-10 minutes before the crash is significantly higher than normal condition. On the other hand, speed at the beginning of the weaving segment in normal condition is significantly higher than disruptive condition at the 95% confidence interval. As for the categorical variables, when surface condition is wet or configuration is LC1, which indicate there is no need for on-ramp or off-ramp traffic to execute lane changing, the crash risk is high. The impact of each variable on crash risk is in Table 3. The Sensitivity, Specificity and Accuracy are calculated when the cut-off-point is 0.08.

Table 3 Real-time Crash Prediction Model for Weaving Segment

Variables	Mean	Std	95% CI	
Bm_spd	-0.111	0.012	(-0.135, -0.088)	
Bm_Em_spd	0.064	0.028	(0.008, 0.119)	
log(Volume)	0.536	0.117	(0.308, 0.767)	
Surface	0.571	0.328	(-0.108, 1.188)*	
L_{max}	0.300	0.087	(0.145, 0.473)	
LC	0.702	0.312	(0.063, 1.333)	
	$\bar{D}$	p_D	DIC#	
	475.740	9.088	484.828	
	AUC	**Sensitivity**	**Specificity**	**Accuracy**
Training	0.773	0.678	0.716	0.712
Validation	0.701	0.676	0.700	0.698

* *variable significant at 90% interval*

DIC=$\bar{D}+p_D$

In order to check whether the ratio of crash over non-crash has impact on coefficient and model performance, other ratios (1:8 and 1:15) were tried. The model results showed that the significant variables and their signs remained the same for three ratios (1:8, 1:10 and 1:15). There were minor but insignificant changes for coefficients when the ratio changed. For example, the coefficient of speed difference between the beginning and the end of the weaving segment were 0.047 and 0.055 when the ratio were 1:8 and 1:15 respectively. The small variation of the coefficient might because of minor difference between population and random sample. The randomly selected unbiased sample cannot 100% represent the whole non-crash population. Previous studies also stated that the estimated regression coefficients for predictors could provide a valid estimation of the log odds ratio with regardless of crash and non-crash ratio (Andersen and Skovgaard 2010, Vittinghoff et al. 2011).

The average speed of mainlines which are at the beginning of weaving segments (Bm_spd) is found to be negatively related to crash risk. According to HCM 2010, high speed on weaving segments reflects smooth traffic flow and few weaving maneuvers, thus it was found to negatively affect crash risk. Meanwhile, in this study, the correlation coefficient between speed and occupancy at the beginning of weaving segment is -0.66. Low average speed indicates high occupancy which results in high crash risk (Abdel-Aty et. Al, 2005). Another study on crash risk at ramp vicinity by Hossain and Muromachi (2013a) also found that the speed has a significantly negative impact on crash risk.

The speed difference between the beginning and the end of the weaving segment (Bm_Em_spd) has a positive impact on crash risk. The higher speed difference, the higher crash risk. When the speed at the beginning of weaving segment is higher than that at the end, vehicles have to decelerate. If drivers are distracted or cannot react in time, it is easy to have a rear-end crash. A previous research by Hossain and Muromachi

(2013b) shows that the speed difference can best explain the crash risk and type for basic freeway segment. The studied weaving segment includes lanes which do not involve in weaving. These lanes are similar to lanes at freeway basic segment, and the crashes on these lanes may have the similar indicators as freeway basic segment.

The logarithm of volume in 5-minute interval is with a positive coefficient, indicating high volume might increase the crash risk on a weaving segment. It is easy to be understood. High volume means high exposure of a single-vehicle crash. Meanwhile, high volume also indicates the interactions between vehicles are high, and then result in high exposure of a multi-vehicle crash.

Wet pavement surface condition increases crash risk in weaving segment. It leads to smaller friction and result in longer braking distances. Meanwhile, vehicles are more likely to lose control. The impact of wet pavement surface on weaving segment is even more severe than on basic segment. The on- and off-ramp vehicles have to execute lane changing along with deceleration and acceleration. The complicated traffic condition enlarges the wet pavement surface's impact.

Maximum length (L_{max}) is an estimated factor which measures the distance at which weaving turbulence no longer has an impact on operation and capacity (HCM 2010). This study finds that maximum length also has significant impact on crash risk. This variable is associated with two factors, i.e., weaving ratio (VR), number of lanes from which a weaving may be made (N_{WL}). The increase of maximum length is mainly contributed by the increase of weaving ratio. When the weaving ratio increase and the whole volume at weaving segment does not change, there are more on-ramp and/or off-ramp vehicles. These vehicles would lead to more turbulence comparing to mainline-to-mainline vehicles, and further result in high crash risk.

As for the configuration (LC), the model result confirms what has been discussed in the crash characteristics section. Weaving segments (LC1) in which there is no need for on- or off-ramp traffic to change lane have an increased crash risk, since the interactions between weaving and non-weaving vehicles are likely to increase and more rear-end crashes would occur due to high speed difference at the end of the weaving segments.

Previous real-time crash studies are mainly on mainline and ramp vicinity. Weaving segment belongs to mainline and also is at ramp vicinity. Hence, the result found in this study is similar to the previous studies, e.g., the impacts of average speed and speed difference on crash risk. However, due to the special traffic condition at weaving segment, more factors which are related to weaving are found, e.g., maximum length and configuration.

5 CONCLUSIONS AND DISCUSSION

Weaving segments are potential recurrent bottlenecks which affect the efficiency and safety of expressways during peak hours. Meanwhile, they are one of the most complicated segments, since on- and off-ramp traffic merges, diverges and weaves in the limited space. One effective way to improve the safety of weaving segments is to study crash likelihood using real-time crash data with the objective of, identifying hazardous conditions and reducing the risk of crashes by Intelligent Transportation Systems (ITS) traffic control. In order to provide effective predictors for real-time weaving segment crash risk, this study collected almost two years' MVDS traffic, geometry and weather data of 16 weaving segments.

The multilevel Bayesian logistic regression model shows that weaving segment configuration is an important factor. Weaving segment (LC1), in which there is no need for on- or off-ramp traffic change lane, is with high crash risk. For LC1, the minimum lane change rate is low and lane changing maneuver is much less. However, there exists high speed differences between the on- and off-ramp traffic. This indeed increase the rear-end crash risk.

In addition to geometric factors, several traffic related parameters are found to have significant impact on crash risk. The mainline speed at the beginning of the weaving segments, logarithm of volume have significant impacts of the crash risk of the following 5-10 minutes for weaving segments. Meanwhile, Speed difference plays a important role in estimating crash risk. If the speed difference increases 1 mph, the crash ratio increases by 6.6%; if the difference increase 10 mph, the crash ratio increases by 89.6%. The low speed at the end of weaving segment may be due to congestion at the downstream of segment, or because of the disturbance generated by merging and diverging. Under high speed difference condition, if Dynamic Message Sign (DMS) is used to inform drivers at the beginning of weaving segment to slow down, the crash risk would be significantly reduced.

Maximum length, which measures the distance at which weaving turbulence no longer has impact, is found to be positively related to the crash risk at the 95% confidence interval. Decreasing maximum length is also an option to decrease crash likelihood. Weaving ratio (VR) has the most important impact on maximum length. If VR changes from 0.2 to 0.1, the maximum length decrease by 758 feet and crash ratio decrease by 20.3%. Ramp metering can be implemented in decreasing on-ramp traffic and weaving ratio, and then improving the safety of weaving segment in real-time.

Previous weaving segment safety study did not explore the impact of maximum length on crashes. However, as a new proposed parameter in HCM 2010, it is very important. First, it is the dynamic threshold which changes according to weaving ratio and the number of lanes from which a weaving may be made. When the short length of a segment is more than maximum length, the segment is not weaving segment. Previous weaving segment safety papers did not compare maximum length with short length. It may have one disadvantage. When a segment is 2500 feet and the maximum length is only 2000 feet, this segment is not weaving segment but merging followed by diverging. The weaving maneuver actually seldom happens in this segment. Second, this parameter is much more important in real-time crash estimation than short length and base length. The segment length (short and base length) cannot determine whether weaving segments are prone to have a crash or not. A short weaving segment may be safer than a long weaving segment when the weaving influences length of this short segment is much shorter than that of a long segment.

Besides traffic and geometric factors, wet pavement surface condition significantly increases the crash ratio by 77%. Frequent lane changing along with deceleration and acceleration within weaving segment makes the safety condition is sensitive to pavement surface condition. Wet road surface can reduce pavement friction and result in skidding or hydroplaning, and then result in a crash. High friction surface is a good treatment to relieve this impact.

Based on the proposed model, the crash hazard for weaving segments can be identified. ITS, e.g., ramp metering and DMS, and high friction surface treatment can be used to enhance the safety of weaving segments in real-time. There are limitations of this study. The weaving segment sample size is only 16 though the crash sample size is enough for real-time safety study. All the studied weaving segments are from one expressway, their geometric designs do not vary too much, e.g., the speed limits of these weaving segments are only with two values (55 and 65 mph). More geometric parameters can be explored by identifying more weaving segments with different geometric design in the future.

ACKNOWLEDGEMENT

The authors thank the Central Florida Expressway Authority (CFX) for funding this research and providing data. The authors also thank the Southeastern Transportation Center UTC consortium for partial funding of this research.

REFERENCE

Abdel-Aty, M., Hassan, H.M., Ahmed, M., 2012. Real-time analysis of visibility related crashes: Can loop detector and avi data predict them equally? Transportation Research Board 91st Annual Meeting. Washington, DC.

Abdel-Aty, M., Pemmanaboina, R., 2006. Calibrating a real-time traffic crash-prediction model using archived weather and its traffic data. Intelligent Transportation Systems, IEEE Transactions on 7 (2), 167-174.

Abdel-Aty, M., Uddin, N., Pande, A., 2005. Improving safety and security by developing a traffic accident prevention system. First International Conference on Safety and Security Engineering Proceedings: Rome, Italy

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F.M., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transportation Research Record: Journal of the Transportation Research Board 1897 (1), 88-95.

Ahmed, M.M., Abdel-Aty, M., Yu, R., Lee, J., 2012. Exploring the feasibility of using airport data in real-time risk assessment. Transportation Research Board 92nd Annual Meeting. Washington, DC.

Andersen, P.K., Skovgaard, L.T., 2010. Regression with linear predictors. New York: Springer Science & Business Media.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5-32.

Christoforou, Z., Cohen, S., Karlaftis, M.G., 2011. Identifying crash type propensity using real-time traffic data on freeways. Journal of Safety Research 42 (1), 43-

50.
Gelman, A., 2006. Multilevel (hierarchical) modeling: What it can and cannot do. Technometrics 48, 432-435.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2014. Bayesian data analysis. London: Chapman & Hall/CRC
Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Safety aspects of freeway weaving sections. Transportation Research Part A: Policy and Practice 38 (1), 35-51.
Highway Capacity Manual, 2000. Transportation Research Board. Washington, D.C.
Highway Capacity Manual, 2010. Transportation Research Board. Washington, D.C.
Hossain, M., Muromachi, Y., 2012. A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accident Analysis & Prevention 45, 373-381.
Hossain, M., Muromachi, Y., 2013a. A real-time crash prediction model for the ramp vicinities of urban expressways. IATSS Research 37 (1), 68-79.
Hossain, M., Muromachi, Y., 2013b. Understanding crash mechanism on urban expressways using high-resolution traffic data. Accident Analysis & Prevention 57, 17-29.
Kim, D.-G., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. Accident Analysis & Prevention 39 (1), 125-134.
Kwon, E., Lau, R., Aswegan, J., 2000. Maximum possible weaving volume for effective operations of ramp-weave areas: Online estimation. Transportation Research Record: Journal of the Transportation Research Board 1727 (1), 132-141.
Lee, C., Abdel-Aty, M., 2006. Temporal variations in traffic flow and ramp-related crash risk. Applications of Advanced Technology in Transportation. The Ninth International Conference.
Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Transportation Research Record: Journal of the Transportation Research Board 1784 (1), 1-8.
Lertworawanich, P., Elefteriadou, L., 2001. Capacity estimations for type b weaving areas based on gap acceptance. Transportation Research Record: Journal of the Transportation Research Board 1776 (1), 24-34.
Lertworawanich, P., Elefteriadou, L., 2003. A methodology for estimating capacity at ramp weaves based on gap acceptance and linear optimization. Transportation Research Part B: Methodological 37 (5), 459-483.
Liu, P., Chen, H., Lu, J.J., Cao, B., 2009. How lane arrangements on freeway mainlines and ramps affect safety of freeways with closely spaced entrance and exit ramps. Journal of Transportation Engineering 136 (7), 614-622.
Madanat, S., Liu, P.-C., 1995. A prototype system for real-time incident likelihood prediction. ITS-IDEA Program Project Final Report.
Nemes, S., Jonasson, J.M., Genell, A., Steineck, G., 2009. Bias in odds ratios by logistic regression modelling and sample size. BMC medical research methodology 9

(1), 56.

Oh, C., Oh, J.-S., Ritchie, S.G., Chang, M., 2001. Real-time estimation of freeway accident likelihood. 80th Annual Meeting of the Transportation Research Board, Washington, DC.

Qi, Y., Liu, J., Wang, Y., 2014. Safety performance for freeway weaving segments. SWUTC/14/600451-00045-1

Roess, R.P., Ulerio, J.M., 2000. Weaving area analysis in year 2000 highway capacity manual. Transportation Research Record: Journal of the Transportation Research Board 1710 (1), 145-153.

Roess, R.P., Ulerio, J.M., 2009. Level of service analysis of freeway weaving segments. Transportation Research Record: Journal of the Transportation Research Board 2130 (1), 25-33.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. Winbugs Version 1.4 User Manual. MRC Biostatistics Unit, Cambridge

Stewart, J., Baker, M., Van Aerde, M., 1996. Evaluating weaving section designs using integration. Transportation Research Record: Journal of the Transportation Research Board 1555 (1), 33-41.

University of Florida. About Signal Four Analytics, Accessed June 27, 2015. https://s4.geoplan.ufl.edu/

Vittinghoff, E., Glidden, D.V., Shiboski, S.C., Mcculloch, C.E., 2011. Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models. New York: Springer Science & Business Media.

Wang, L., Shi, Q., Abdel-Aty, M., Kuo, P., 2015. Predicting crashes on expressway ramps with real-time traffic and weather data. Transportation Research Board 94nd Annual Meeting. Washington, DC.

Xu, C., Tarko, A.P., Wang, W., Liu, P., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Analysis & Prevention 57, 30-39.

Yu, R., Abdel-Aty, M., 2013a. Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes. Accident Analysis & Prevention 58, 97-105.

Yu, R., Abdel-Aty, M., 2013b. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention 51, 252-259.

Zheng, Z., Ahn, S., Monsere, C.M., 2010. Impact of traffic oscillations on freeway crash occurrences. Accident Analysis & Prevention 42 (2), 626-636.