Transportation Research Part C 71 (2016) 406–418

Contents lists available at ScienceDirect

Transportation Research Part C

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

Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data

Chengcheng Xu, Ph.D. (Assistant Professor)*, Pan Liu, Ph.D. (Professor), Bo Yang, Ph.D. Candidate (Graduate Research Assistant), Wei Wang, Ph.D. (Professor)

Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing 210096, China
Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Si Pai Lou #2, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 28 February 2016
Received in revised form 8 August 2016
Accepted 23 August 2016
Available online 30 August 2016

Keywords:
Secondary crashes
Real-time
Traffic flow
Random effect logit model
Freeways

ABSTRACT

This study aimed to develop a secondary crash risk prediction model on freeways using real-time traffic flow data. The crash and traffic data were collected on the I-880 freeway for five years in California, United States. The secondary crashes were identified by a method based on speed contour plot. The random effect logit model was used to link the probability of secondary crashes with the real-time traffic flow conditions, primary crash characteristics, environmental conditions, and geometric characteristics. The results showed that real-time traffic variables significantly affect the likelihood of secondary crashes. These traffic variables include the traffic volume, average speed, standard deviation of detector occupancy, and volume difference between adjacent lanes. In addition, the primary crash characteristics, environmental conditions and geometric characteristics also significantly affect the risks of secondary crashes. The model evaluation results showed that the predictive performance of the developed model was deemed satisfactory. The inclusion of traffic flow variables and random effect increases prediction accuracy by 16.6% and 7.7%, respectively. These results have the potential to be used in advanced traffic management systems to develop proactive traffic control strategies to prevent the occurrences of secondary crashes on freeways.

1. Introduction

Secondary crashes are the crashes that result from a prior crash. They usually occur within the spatial and temporal impact ranges of an existing primary crash. The occurrence of secondary crashes on freeways leads to increased risks of additional crashes, reduced freeway capacity, and increased travel time uncertainty. Previous studies suggested that the risks of secondary crashes can be reduced by improved incident management (Sun and Chilukuri, 2010; Zhan et al., 2009). To develop incident management strategies of preventing secondary crashes on freeways, increased attentions have been given to studying the contributing factors to secondary crashes.

A review of literature revealed that significant efforts have been conducted to identify secondary crashes in previous studies (Raub, 1997; Karlaftis et al., 1998; Moore et al., 2004; Hirunyanitiwattana and Mattingly, 2006; Zhan et al., 2008; Chilukuri and Sun, 2006; Sun and Chilukuri, 2010; Zhang and Khattak, 2010; Chou and Miller, 2010). The proposed approaches to identify secondary crashes in previous studies are briefly summarized in Table 1. In early studies, the static

* Corresponding author.
E-mail addresses: xuchengcheng@seu.edu.cn (C. Xu), pan_liu@hotmail.com (P. Liu), ybseu2013@126.com (B. Yang), wangwei@seu.edu.cn (W. Wang).

http://dx.doi.org/10.1016/j.trc.2016.08.015
0968-090X/

Table 1
Summary of secondary crash identification methods in previous studies.

Author	Year	Method
Raub	1997	Static thresholds of 1 mile and 15 min
Karlaftis et al.	1998	Static thresholds of 1 mile and 15 min
Moore et al.	2004	Static thresholds of 2 miles and 2 h
Hirunyanitiwattana and Mattingly	2006	Static thresholds of 2 miles and 1 h
Zhan et al.	2008	Static thresholds of 2 miles and 15 min
Chilukuri and Sun	2006	Incident progression curve
Sun and Chilukuri	2010	Incident progression curve
Zhan et al.	2009	Method based on cumulative arrival and departure plots
Zhang and Khattak	2010	Method based on queue length estimations
Khattak et al.	2011	Method based on queue length estimations
Chou and Miller-Hooks	2010	Simulation-based method
Imprialou et al.	2013	Method based on spatiotemporal impact area of primary crash
Yang et al.	2014a	Method based on spatiotemporal impact area of primary crash
Park and Haghani	2016	Method based on spatiotemporal impact area of primary crash
Sarker et al.	2015	Method based on spatiotemporal impact area of primary crash
Mishra et al.	2016	Method based on spatiotemporal impact area of primary crash
Wang et al.	2016	Method based on spatiotemporal impact area of primary crash

threshold methods were used to identify secondary crashes based on some fixed spatial and temporal criteria. These studies assumed that secondary crashes should occur within a maximum spatial and temporal impact range of a primary incident. For example, Karlaftis et al. (1998) used static spatial and temporal thresholds of 1 mile and 15 min to identify secondary crashes. Any crash occurring within 1 mile upstream of and less than 15 min after a prior crash is defined as a secondary crash. Hirunyanitiwattana and Mattingly (2006) proposed the static thresholds of 1 h and 2 miles upstream of a primary incident for identifying secondary crashes. Other studies conducted by Raub (1997), Moore et al. (2004), and Zhan et al. (2008) used the similar static thresholds. One limitation associated with the static method is that they need a subjective determination of fixed spatial and temporal thresholds. As a result, the static method may not have adequate accuracy in identifying second crashes (Chilukuri and Sun, 2006; Sun and Chilukuri, 2010; Zhang and Khattak, 2010; Chou and Miller, 2010).

To overcome the limitation associated with the static threshold methods, a number of studies proposed various dynamic methods to identify secondary crashes. These dynamic methods include the incident progression curve (Chilukuri and Sun, 2006; Sun and Chilukuri, 2010), queue length estimations (Zhang and Khattak, 2010), cumulative arrival and departure plots (Zhan et al., 2009), and simulation-based method considering shockwave (Chou and Miller, 2010). Sun and Chilukuri (2010) developed incident progression curves to estimate the end of the varying queue throughout the entire incident. The results showed that the method based on incident progression curves can improve the secondary crash identification accuracy by 30%, compared with the static method. In a study conducted by Zhang and Khattak (2010), a dynamic spatial threshold method based on queue length estimations was developed to identify secondary incident. Zhan et al. (2009) proposed a method for detecting secondary crashes based on cumulative arrival and departure plots. In this method, the cumulative arrival and departure plots were used to estimate the maximum queue length and the recovery time of associated queue for incidents with lane blockages.

Recently, a number of studies detected secondary crashes by identifying the spatiotemporal impact area of primary crashes (Imprialou et al., 2013; Yang et al., 2014a; Park and Haghani, 2016; Sarker et al., 2015; Mishra et al., 2016; Wang et al., 2016a, 2016b). Yang et al. (2014a) used the speed contour plot to identify secondary crashes on freeways. The results showed that 50% of secondary crashes occurred within 70 min after and 2 miles upstream of the primary crashes. Imprialou et al. (2013) used the Automatic Tracking of Moving Jams method to detect secondary crashes by defining the actual boundaries of the spatiotemporal influence area of primary crashes. The results showed that the proposed method provides better accuracy in detecting secondary crashes than conventional static threshold methods. Park and Haghani (2016) developed a Bayesian structure equation model to identify spatiotemporal impact area by using the vehicle probe data. Sarker et al. (2015) proposed a method to detect secondary crashes based on spatiotemporal impact area analysis and shockwave principles. The validations results showed that the proposed method can achieve a good detection accuracy ranging from 72% to 91%.

In addition to secondary crash identifications, some studies have also been conducted to explore the characteristics of secondary crashes (Hirunyanitiwattana and Mattingly, 2006; Yang et al., 2014a; Khattak et al., 2009; Karlaftis et al., 1998, 1999; Carrick et al., 2015; Xie et al., 2016; Mishra et al., 2016; Jalayer et al., 2015; Zhang et al., 2015; Wang et al., 2016a, 2016b). Hirunyanitiwattana and Mattingly (2006) used the proportional tests to compare the characteristics of secondary crashes and primary crashes on freeways with respect to time of day, roadway classification, severity level and type of accident. The results showed that secondary crashes are more likely to be property-damage-only (PDO) rear-end crashes occurring during peak hours. Yang et al. (2014a) conducted a thorough examination of the spatio-temporal distributions, clearance time, crash type and severity of secondary crashes. The results showed that most secondary crashes occurred within two hours and two miles upstream of primary crashes. They also found that secondary crashes are likely to involve two or more vehicles and to be rear-end crashes. Carrick et al. (2015) compared the roadway, environmental, and vehicle characteristics of secondary and normal crashes. Secondary crashes are more likely to occur on freeways and in rainy weather conditions. Moreover, commercial vehicles are slightly more likely to incur secondary crashes. Jalayer et al. (2015) studied the statistical characteristics of secondary crashes in terms of collision type, severity level, and response duration. Zhang et al. (2015) used microscopic simulation tool to study the queuing delays associated with secondary incidents. The secondary incidents were found to incur longer delays than single incidents. The time gap and distance between a primary incident and its secondary significantly affect the total delays. Mishra et al. (2016) compared the temporal distributions of secondary crashes on Freeways and arterials. The results showed that the occurrence time of secondary crashes on Freeways exhibits both AM and PM peaks. While the occurrence time of secondary crashes on arterials only exhibits a PM peak.

To prevent the occurrence of secondary crashes, a number of studies have been conducted to investigate the relationship between the likelihood of secondary crashes and various contributing factors, such as the characteristics of the primary crashes, weather conditions, geometric conditions, traffic volumes, and roadway functional class (Zhan et al., 2009; Zhang and Khattak, 2010; Kopitch and Saphores, 2011; Khattak et al., 2012; Yang et al., 2014b; Mishra et al., 2016; Wang et al., 2016a, 2016b). Table 2 summarizes the research findings of contributors to secondary crashes in these studies. With regard to the primary crash characteristics, the collision type, occurrence time, crash duration, and number of involved vehicle were found to significantly affect secondary crash likelihood. More specifically, the rear-end crashes with longer durations are more like to incur secondary crashes (Yang et al., 2014b; Mishra et al., 2016). The crashes occurred at the peak-off hours or on the weekends are less likely to incur secondary crashes (Zhan et al., 2009; Khattak et al., 2012; Yang et al., 2014b; Kopitch and Saphores, 2011). Moreover, the risks of secondary crashes increase with an increase in the number of involved vehicles (Zhang and Khattak, 2010; Mishra et al., 2016).

With regard to other contributing factors, adverse weather conditions such as rain and snow were found to significantly increase the risks of secondary crashes (Khattak et al., 2012; Mishra et al., 2016; Wang et al., 2016a, 2016b). The secondary crash risks increase with an increase in the annual average daily traffic (AADT) (Zhang and Khattak, 2010; Khattak et al., 2012; Mishra et al., 2016). Compared with crashes occurred on freeways, crashes on arterials are more likely to incur secondary crashes (Mishra et al., 2016). Curve segments lead to increased risks of secondary crashes (Zhang and Khattak, 2010). The crashes detected by closed circuit television (CCTV) on freeways are more likely to incur secondary crashes (Khattak et al., 2012).

As shown in Table 2, logit model is one of the commonly used methods to predict secondary crashes. It does not make any assumption about the distribution and variances of the dependent variable. Logit model has the advantage of avoiding overfitting problem (Mitra and Washington, 2007). It is also superior to other methods when dealing with unbalanced sample, in which the number of one class is much larger than that of the other classes. In addition, unlike most artificial intelligence models that work as black-boxes, logit model provides an easily readable mathematical function, which defines tangible

Table 2
Research findings of contributors to secondary crashes in previous studies.

Author	Method	Research findings
Zhan et al. (2009)	Logit model	• Secondary crash risks at morning peak are higher than other times • Longer duration crash increases secondary crash risks
Zhang and Khattak (2010)	Ordered logit model	• Longer duration crash increases secondary crash risks • Larger number of involved vehicles increases secondary crash risks • Greater traffic volume (AADT) increases secondary crash risks • Curve segment leads to increased risks of secondary crashes
Khattak et al. (2012)	Logit model	• Adverse weather increases secondary crash risks • Secondary crash risks at peak hours are higher than those at peak-off hours • Larger AADT increases secondary crash risks • Longer duration crash increases secondary crash risks • Primary crashes detected by closed circuit television (CCTV) are more likely to incur secondary crashes
Yang et al. (2014b)	Logit model	• Secondary crash risks at peak hours are higher than those at peak-off hours • Rear-end crashes are more likely to incur secondary crashes • Longer duration crash increases secondary crash risks
Kopitch and Saphores (2011)	Logit model	• Secondary crash risks at peak hours are higher than those at peak-off hours • Secondary crash risks on weekends are lower than those on weekdays
Mishra et al. (2016)	Multinomial logit model	• Larger AADT increases secondary crash risks • Larger number of involved vehicles increases secondary crash risks • Rear-end crashes are more likely to incur secondary crashes • Clear weather reduces the risks of secondary crashes • Secondary crashes are more likely to occur on arterials than on freeways
Wang et al. (2016a, 2016b)	Logit model	• Clear weather reduces the risks of secondary crashes • Larger speed of shockwave increases secondary crash risks • Longer duration crash increases secondary crash risks

relationships between dependent and explanatory variables (Zhang and Xie, 2008). This allows the results to easily be applied in practical engineering applications.

Although numerous studies have been conducted to predict the secondary crash likelihood and to identify the contributing factors, relatively few studies have focused on predicting the likelihood of secondary crashes in real time considering the effects of dynamic traffic flow conditions. Vlahogianni et al. (2012) developed a secondary crash risk model to connect the secondary crash likelihood with primary accident duration, weather conditions, number of vehicles, hourly volume, and hourly speed. The results suggested that the secondary crash likelihood increases with an increase in volume and a decrease in speed. Since hourly volume and speed were used for model development, this model cannot predict secondary crash risks in short time interval, such as 5 min. It is difficult to be used for real-time incident management. In a study conducted by Park and Haghani (2016), traffic condition was classified into congested and uncongested states according to vehicle probe speed data. The risks of secondary crashes were linked with traffic congestions and clearance duration of primary crashes. The results suggested that traffic congestions tend to increase the likelihood of secondary crashes. However, in addition to the average speed, other traffic flow variables such as average occupancy and speed variances, and the differences in traffic flow variables between adjacent lanes may also contribute to the occurrence of secondary crashes. Considering these traffic flow variables in secondary crash prediction will promote a more complete understanding of the relationship between real-time traffic flow conditions and secondary crash risks.

Investigation of the relationship between real-time traffic flow conditions and the risks of secondary crashes will help to increase the predictive performance of the secondary crash risk model. Moreover, such relationship can be used to develop real-time secondary risk prediction models, which will play an important role in incident management for preventing secondary crashes. Recent studies suggested that dynamic traffic management systems (DTMS), such as the variables limit system and ramp metering system, have the potential to reduce collision risks on freeways (Abdel-Aty et al., 2007, 2012; Yu and Abdel-Aty, 2014; Chen and Ahn, 2015). The real-time secondary crash risk prediction models can also be used to develop proactive traffic control strategies in DTMSs to reduce the risks of secondary crashes on freeways. The central idea is to proactively eliminate the hazardous traffic flow conditions prone to secondary crashes by adjusting the speed limits or flow rate on freeway mainlines. The hazardous traffic flow conditions can be identified using a real-time secondary crash risk prediction model that links the risks of secondary crashes with dynamic traffic flow variables. Compared with simply alerting drivers upstream of a crash, the applications of DTMSs are expected to more effectively prevent the occurrences of secondary crashes. Because the proactive traffic control strategies can help to determine the best speed for drivers or flow rate on mainline to minimize secondary crash risks. The study presented in this paper fills gaps in understanding the effects of real-time traffic flow conditions on the likelihood of secondary crashes.

The primary objective of this study was to develop a secondary crash risk prediction model incorporating the effects of real-time traffic flow conditions. The high-resolution traffic data collected from loop detectors were used for the model development. This study has the potential to contribute to the field of secondary crash analysis by: (1) exploring the traffic flow conditions prone to secondary crashes on freeways; and (2) using high-resolution traffic flow data to develop a real-time secondary crash risk prediction model. This fills the knowledge gap that previous studies generally developed the secondary crash risk prediction model without considering the effects of real-time traffic flow conditions. The developed model in this study is expected to improve the prediction accuracy of secondary crashes.

2. Data sources

The used traffic and crash data were obtained from a 35-mile section on the I-880 freeway in the state of California, United States from 2006 to 2010. There were 134 loop detectors stations along the selected freeway section on two directions. The average spacing between loop detector stations is about 0.5 miles for both directions. Crash data were obtained from the Statewide Integrated Traffic Records System (SWITRS) that is maintained by the California Department of Transportation (Caltrans). The extracted crash characteristics included the date, time, crash severity, collision type, road surface conditions, weather conditions, and lighting conditions.

A total of 9188 crashes were identified and used for further data analysis. 4846 crashes occurred on the southbound, and 4342 occurred on the northbound. Three types of crashes were identified, including the secondary crashes, the primary crashes and the normal crashes. To be more specific, the secondary crashes are the crashes that are induced by primary crashes. The primary crashes are defined as the crashes that lead to secondary crashes, while the normal crashes are defined as the crashes that did not lead to any secondary crashes. The number of identified secondary crashes, primary crashes and normal crashes are 113, 97, and 8978 respectively. The method for identifying these three types of crashes is given in Section 4.1.

The high-resolution traffic data were extracted from the Highway Performance Measurement System (PeMS) maintained by the Caltrans. The PeMS database provides 30-s raw loop detector data for each lane, including count, speed, and detector occupancy. For each crash, the authors extracted the raw traffic data from the nearest detector station to the crash location. For each crash, traffic data were extracted for the time interval between 5 and 10 min prior to crash occurrence. The purpose was to account for the potential inaccuracies in the reported crash time (Golob and Recker, 2004). The 30-s raw traffic data from the nearest loop detector station for each crash were further aggregated into 5-min intervals and converted into the 12

Table 3
Candidate variables for secondary crash risk prediction model development.

Variable category	Variable	Description
Real-time traffic conditions	AvgCnt	Average vehicle count during 5-min period (veh/30 s)
	AvgSpd	Average vehicle speed during 5-min period (mile/h)
	AvgOcc	Average detector occupancy during 5-min period (%)
	DevCnt	Std. dev. of vehicle count during 5-min period (veh/30 s)
	DevSpd	Std. dev. of vehicle speed during 5-min period (mile/h)
	DevOcc	Std. dev. of detector occupancy during 5-min period (%)
	CovCnt	Coefficient of variation of count during 5-min period (veh/30 s))
	CovSpd	Coefficient of variation of speed during 5-min period (mile/h)
	CovOcc	Coefficient of variation of occupancy during 5-min period (mile/h)
	DifCnt	Vehicle count difference between adjacent lanes (veh/30 s)
	DifSpd	Vehicle speed difference between adjacent lanes (veh/30 s)
	DifOcc	Occupancy difference between adjacent lanes (veh/30 s)
Primary crash characteristics	Severity	1 = Injury crashes; 0 = PDO
	Sideswipe	1 = Sideswipe crash; 0 = otherwise
	Rear-end	1 = Rear end crash; 0 = otherwise
	Peak	1 = Peak period; 0 = otherwise
	Dayweek	1 = Weekend; 0 = weekday
Environmental conditions	Weather	1 = Adverse weather conditions; 0 = clear
	Roadsurf	1 = Road surface is wet; 0 = otherwise
	Lighting	1 = No street lights or street lights not functioning; 0 = otherwise
Geometric characteristics	Lane	Number of lanes
	Width_s	Road surface width (ft)
	Width_m	Inner median width (ft)
	Curve	1 = Curve section; 0 = otherwise

traffic flow variables presented in Table 3. The geometric variables for each crash were extracted based on its occurrence location. As shown in Tables 3 and 4 candidate geometric variables were considered.

3. Methodology

Statistical methods used in this study are briefly discussed in this section. The proportionality tests were conducted to compare the characteristics between the primary crashes and normal crashes. The Bayesian random effect logit model was used to predict the probability of secondary crashes induced by the primary crashes on freeways.

3.1. Bayesian random effect logit model

The Bayesian random effect logit model was used to develop a secondary crash risk prediction model, in which the likelihood of secondary crashes was linked with real-time traffic variables, primary crash characteristics, weather conditions and geometric characteristics. The random effect was included to account for the heterogeneity caused by the unobserved factors, such as the work zones, design features, and pavement conditions. Previous studies suggested that overlooking the unobserved heterogeneity may lead to inconsistent and bias parameter estimates (Washington et al., 2003; Anastasopoulos and Mannering, 2009). Accordingly, the random effect logit model is expected to produce more accurate parameter estimates and better predictive performance. It can be expressed as follows:

$$y_n \sim \text{Bernoulli}(p_n) \tag{1}$$

Table 4
Proportional tests between primary and normal crashes.

	Normal crashes	Primary crashes	Secondary crashes	Tests between primary and normal crashes	
				Z value	Sig.
Rear-end	0.58	0.73	0.73	3.31	0.001
Sideswipe	0.22	0.11	0.18	3.23	0.001
Others	0.20	0.15	0.10	1.22	0.23
1–6 am	0.09	0.02	0.00	4.62	0.000
7–12 am	0.31	0.38	0.40	1.49	0.14
13–18 pm	0.39	0.43	0.32	0.81	0.42
19–24 pm	0.21	0.16	0.28	1.23	0.22

$$\mathrm{logit}(p_n) = \beta_0 + \theta_r + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_i x_{in} \tag{2}$$

where y_i represents the secondary crash indicator (=1 if a secondary crash is induced by a primary crash, and 0 indicates that no secondary crash occurred) for the nth observation in the sample; p_n denotes the probability of a secondary crash induced by a primary crash; x_{in} denotes the value of variable i for sample n; β_i is the coefficient of variable x_i; θ_r is the random effect which captures the heterogeneity effects for freeway segment r. The random effect θ_r is assumed to be normally distributed as $\theta_r \sim (0, \Sigma_\theta)$.

The Bayesian approach based on Markov chain Monte Carlo (MCMC) simulations was used to estimate the random effect logit model. The non-informative prior distributions were used. The inference was made on the basis of the remaining draws after discarding the draws during the burn-in period.

To evaluate the effects of traffic flow variables on the likelihood of secondary crashes, the elasticity analysis was conducted. The elasticity of a continuous independent variable represents the percentage change in the dependent variable resulting from a 1% change in the independent variable (Washington et al., 2003). The elasticity of a continuous independent variable x_i is given as:

$$E_i = \frac{\partial Y}{\partial x_i} \times \frac{x_i}{Y} = (1 - P)\beta_i x_i \tag{3}$$

Although each observation in the dataset has an elasticity that depends on the value of x_i and the estimated probability of a secondary crash P, it is customary to report the average elasticity in the sample. In the following analysis, both the average and standard deviation of the elasticity are given. Note that Eq. (3) cannot be used to calculate the elasticity for an indicator variable. The elasticity of an indicator variable x_i is estimated by computing a pseudo-elasticity using the following equation (Washington et al., 2003):

$$E_i = \left\{ \frac{e^{\Delta(x'\beta)} \times (1 + e^{x_i \beta_i})}{e^{\Delta(x'\beta)} \times e^{x_i \beta_i} + 1} - 1 \right\} \times 100 \tag{4}$$

3.2. Receiver Operating Characteristic (ROC) curve

The ROC curve is frequently used to compare the predictive performance of different models (Egan, 1975). A model of binary outcome (event = 1 and non-event = 0) classifies an observation as an event (a crash) if the predicted probability of the observation exceeds a pre-specified threshold. Otherwise, it will be classified as a non-event. The predictive performance of a model of binary outcome can be measured with two complementary indicators: the true positive rate (sensitivity) that is the proportion of events predicted as an event, and true negative rate (specificity) that is the proportion of non-events predicted as a non-event. Both sensitivity and specificity depend on the same probability threshold varying between 0 and 1. The ROC curves illustrate the relationship between the true positive rate (sensitivity) and the false alarm rate (1-specificity) for thresholds from 0 to 1. To develop a ROC curve, one needs to calculate the sensitivity and (1-specificity) for multiple thresholds varying between 0 and 1.

4. Data analysis and results

4.1. Identification of secondary crashes

The conventional static method of identifying secondary crash suffers from the subjective determinations of fixed spatial and temporal thresholds. The dynamic methods compensate for the inadequacies of the static method by determining dynamic thresholds based on queuing estimations or traffic simulations. However, a number of the dynamic methods still have some shortcomings. For example, the methods based on queue length estimations need detailed queuing information which is not commonly available (Chou and Miller, 2010; Khattak et al., 2009, 2011; Hirunyanitiwattana and Mattingly, 2006; Yang et al., 2014a, 2014b). The methods based on incident progression curve use an identical progression curve for all secondary crashes, resulting in inaccurate secondary crash identification results (Chou and Miller, 2010).

To overcome the shortcomings with the existing static and dynamic methods, this study used a method based on the speed contour plot to identify secondary crashes. The central idea is to determine the spatial and temporal impact range of a prior crash using real-time traffic flow data while accounting for the effects of recurrent congestions. A secondary crash is then identified if it is within the spatial and temporal impact range of this prior crash. The detailed procedure of this method is as follows:

(1) The 5-min speed data were extracted to develop a speed contour plot for a prior crash. More specifically, the speed data were extracted for the time interval between 6 h before the prior crash and 6 h after the prior crash from the loop detectors within 10 miles upstream the prior crash. Fig. 1(a) illustrates an example of a speed contour plot for a prior crash, in which congestions and queue formations were clearly observed after the prior crash. However, it is difficult to

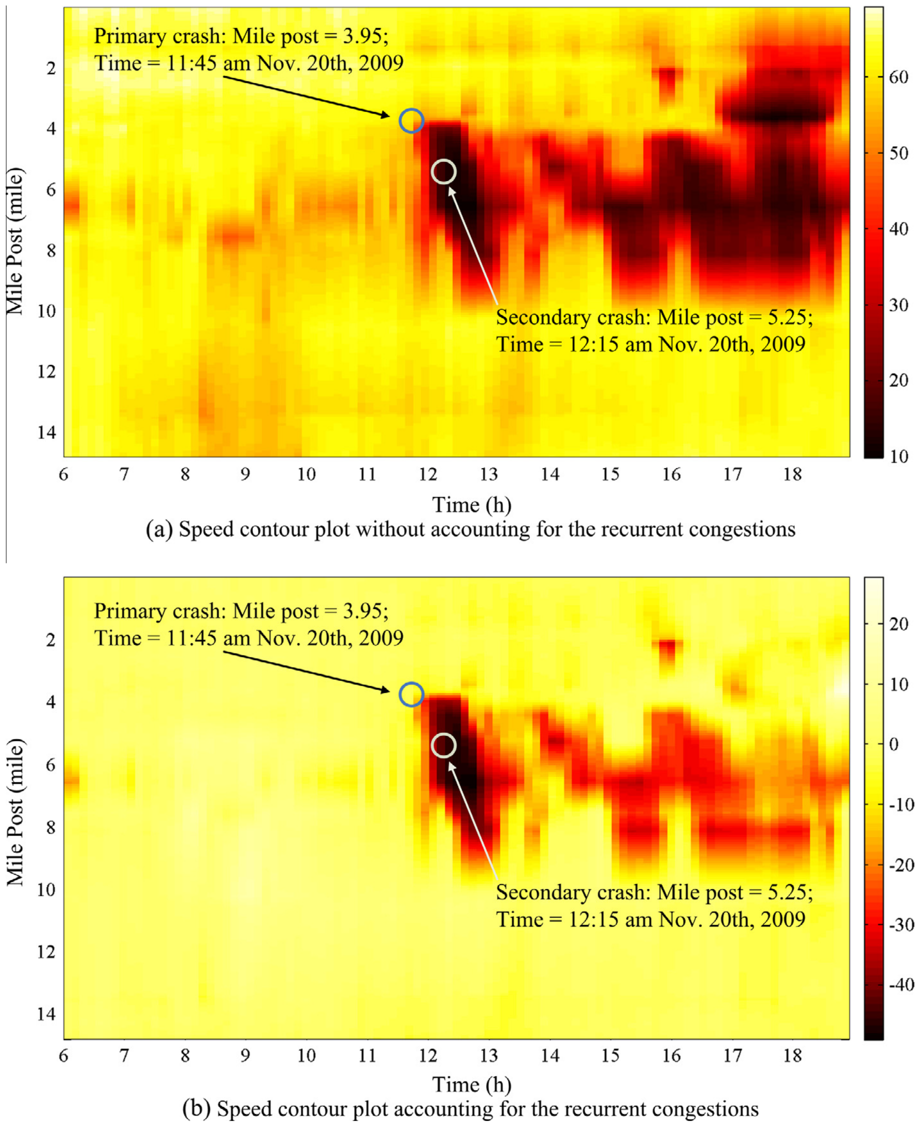

Fig. 1. Identification of secondary crashes.

determine whether the queue formations were caused by recurrent congestions or the prior crash. To account for effects of the recurrent congestions, the following two steps were further used to identify the spatial and temporal impact range of the prior crash.

(2) The research team further extracted 5-min speed data for the same time and location in the step (1) from crash-free days in one year. For example, the prior crash in Fig. 1(a) occurred at the time of 11:45 am on November 20, 2009 and the milepost of 3.95. Then the speed data were collected for the same time interval and location in Fig. 1(a) from all crash-free days in 2009. Then the speed data for each time and location was averaged over all the crash-free days.

(3) To account for the potential effects of recurrent congestions, we subtracted the average speed over crash-free days in step (2) from the speed data for each time and location in step (1). The differences between speeds in step (2) and step (1) for various times and locations were then used to develop a new speed contour plot, which was used to identify the spatial and temporal impact range of the prior crash. Fig. 1(b) illustrates the modified speed contour for identifying the spatial and temporal impact range of a prior crash.

(4) The crashes that occurred in the spatial and temporal impact ranges of primary crashes were identified as secondary crashes. The crashes that did not lead to secondary crashes were then identified as normal crashes. Using the above identification method, the numbers of identified secondary, primary, and normal crashes in the dataset are 113, 97, and 8978 respectively.

Previous studies suggested that the speed contour based method can accurately determine the spatiotemporal impact ranges of traffic flow disturbances or traffic congestions (Chen et al., 2004; Kerner et al., 2004; Li and Bertini, 2010; Yang et al., 2014a; Hojatia et al., 2014). Accordingly, the used speed contour method is expected to accurately identify primary, normal, and secondary crashes. To validate this point, we further compared the results of secondary crash identification with those of the latest publications about secondary crash identifications. The ratio between the identified secondary crashes and all crashes in this study is 1.23%, which is consistent with the findings of the latest publications in this area that this ratio is around 1–1.5% (Park and Haghani, 2016; Sarker et al., 2015; Mishra et al., 2016; Wang et al., 2016a, 2016b).

4.2. Characteristics of primary and normal crashes

Fig. 2(a) illustrates the distributions of primary, normal and secondary crashes by collision type. The rear-end collision is the predominant collision type for both primary and secondary crashes. 73.2% of the primary crashes and 72.6% of the secondary crashes were rear-end collisions. The proportion of rear-end collision in normal crashes is 58.2%, which is about 15% lower than that in primary crashes. The proportionality test was used to investigate whether the difference in rear-end collision proportion between primary and normal crashes is significant. The significance level of 0.05 was used in the test. As shown in Table 4, the difference in the proportion of rear-end collision between primary and normal crashes is statistically significant (p-value = 0.001). As expected, the proportion of sideswipe collision in normal crashes is significantly greater than that in primary crashes (see Table 4). These results are consistent with the findings in previous studies that primary crashes leading to secondary crashes are more likely to be rear-end crashes (Yang et al., 2014b; Mishra et al., 2016).

Fig. 2(b) illustrates the distributions of primary and normal crashes by occurrence time. From 0:01 to 6:00 am, the proportion of normal crashes is 8.8%, which is about four times the proportion of primary crash. The proportional test indicates a significant difference between these two proportions (see Table 4). Due to the low traffic volume, crashes occurred during this time period usually have short queue lengths, resulting in low risks of secondary crashes. From 6:01 to 12:00 am, the proportions of both primary and secondary crashes are slight greater than that of the normal crashes. For other time periods, the proportions of primary and secondary crashes are not significantly different from that of normal crashes. These results are consistent with the findings in previous studies (Hirunyanitiwattana and Mattingly, 2006; Yang et al., 2014b).

4.3. Secondary crash risk prediction model

To identify how real-time traffic flow conditions affected the likelihood of secondary crashes on freeways, Bayesian logit models were used to develop secondary crash risk prediction models, in which the events are the primary crashes that lead to secondary crashes and the non-event are the normal crashes that did not lead to secondary crashes. In previous studies, the real-time traffic flow conditions were generally not considered in the secondary crash risk assessment. For the purpose of comparison, we also developed a reduced model in which the real-time traffic flow variables were not included. To account for the possible correlations between candidate variables presented in Table 3, the research team calculated the Pearson correlation parameters between different pairs of candidate variables and generated several combinations which included the maximum number of uncorrelated variables. Stepwise logit analysis was then conducted to select variables from each combination. The log likelihood at the convergence of each model was compared. The model with the highest log likelihood was considered the best model.

Table 5 gives the estimation results of the models with and without traffic variables. The following likelihood ratio test was conducted to identify whether the inclusion of real-time traffic flow data significantly improved the goodness-of-fit of the secondary crash risk prediction model. The test statistic is given by (Washington et al., 2003):

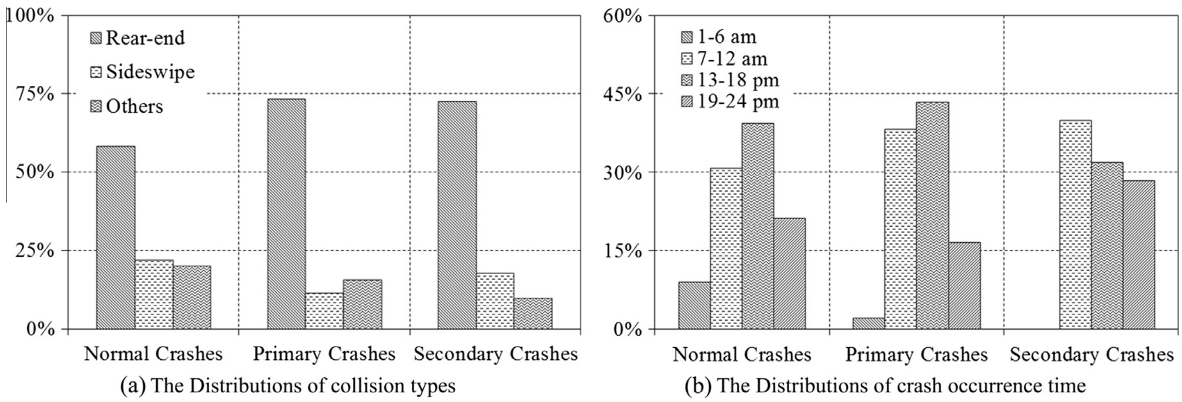

Fig. 2. Comparisons of crash characteristics between primary and normal crashes.

Table 5
The secondary crash risk prediction models with and without traffic variables.

Variables	Model without traffic variables				Model with traffic variables			
	Mean	S.D.	2.5%	97.5%	Mean	S.D.	2.5%	97.5%
Constant	−2.63	0.62	−3.65	−1.59	−5.12	0.98	−6.78	−3.51
Severity	−0.55	0.26	−0.99	−0.13	−0.49	0.26	−0.95	−0.08
Sideswipe	−0.83	0.33	−1.39	−0.31	−0.69	0.34	−1.27	−0.17
Dayweek	−2.18	0.64	−3.34	−1.24	−1.30	0.65	−2.47	−0.34
Roadsurf	0.44	0.27	0.10	0.79	0.65	0.28	0.17	1.11
Lane	−0.40	0.16	−0.67	−0.14	−0.27	0.18	−0.56	−0.04
AvgCnt	–	–	–	–	0.23	0.04	0.17	0.31
AvgSpd	–	–	–	–	−0.03	0.01	−0.04	−0.02
DevOcc	–	–	–	–	2.54	1.23	0.36	4.43
DifCnt	–	–	–	–	0.14	0.06	0.03	0.24
Log-likelihood	−518.00				−486.70			

$$\chi^2 = -2[LL(\beta_{without}) - LL(\beta_{with})] \tag{5}$$

where $LL(\beta_{\text{without}})$ is the log-likelihood at convergence of the model without traffic variables; and $LL(\beta_{\text{with}})$ is the log-likelihood at convergence of the model with traffic variables. This test statistic is chi-squared distributed with the degrees of freedom equal to the difference in the number of parameters between these two models. As shown in Table 6, the test results indicate that the inclusion of real-time traffic flow information significantly improves the goodness-of-fit of the secondary crash risk prediction model (p-value < 0.0001).

As mentioned in Section 3, overlooking the unobserved heterogeneity may lead to inconsistent and bias parameter estimates in the model. Accordingly, a random effect was further introduced in the above model with traffic variables to account for the heterogeneity caused by the unobserved factors. Similarly, a likelihood ratio test was conducted to compare the models with and without random effect. The test result indicates that the inclusion of a random effect can significantly improve the goodness-of-fit of the model with traffic variables (see Table 6). Table 7 gives the estimation results of the secondary crash risk model with random effect. This model has nine independent variables, including four traffic flow variables, three variables for primary crash characteristics, one environmental variable, and one geometric variable.

As shown in Table 7, the average traffic volume (represented by AvgCnt) is positively associated with the risks of secondary crashes, indicating that the crashes occurred in high-volume traffic are more likely to induce secondary crashes. This result is consistent with the findings of previous studies that the risks of secondary crash increase with an increase in AADT (Zhang and Khattak, 2010; Khattak et al., 2012; Mishra et al., 2016). Larger traffic volume indicates smaller time headway between vehicles. Generally speaking, smaller time headway gives drivers less space for taking crash avoidance maneuver, resulting in increased risks of secondary crashes. Accordingly, when a crash occurred in high volume traffic conditions, the ramp metering systems can be used to reduce the upstream volume rate on mainline to reduce the risks of secondary crashes. In addition, gradually reducing the speed in the upstream can also be used to decrease the traffic volume.

The negative parameter of the average speed (represented by AvgSpd) indicates that the risks of secondary crashes increase as the average speed decreases. The decreasing speed represents an increase in traffic density and queue formations. The disturbances caused by the primary crashes are easier to propagate in this queuing traffic conditions, leading to high risks of secondary crashes. To prevent secondary crashes in this condition, traffic control strategies should be developed to accelerate the dissipation of queue in traffic flow. For example, the variable speed limit system can be used to gradually increase the speed downstream the crash location and gradually decrease the speed upstream the crash location at the same time.

Interestingly, the standard deviation of detector occupancy (represented by DevOcc) and the difference in traffic volume between adjacent lanes (represented by DifCnt) also significantly affect the risks of secondary crashes. Specifically, the standard deviation of detector occupancy is positively associated with the occurrence of secondary crashes. The standard deviation of detector occupancy is indicative of oscillating traffic conditions in which vehicles accelerate and brake frequently, thereby increasing the risks of secondary crashes. In this scenario, gradually reducing the upstream speed may reduce the risks of secondary crashes, because the approaching traffic will meet the oscillating traffic conditions at a lower speed.

The positive parameter of the traffic variable DifCnt indicates that the risks of secondary crashes increase with an increase in volume difference between adjacent lanes. A possible explanation is that the imbalances of traffic volume between lanes may encourage drivers to change lanes more frequently, resulting in increased risks of secondary crashes. Alerting upstream

Table 6
Results of likelihood ratio tests.

Tests	LL_R	LL_U	χ^2	d.f.	p-value
Likelihood ratio test between models with and without traffic variables	−518	−486	64	4	<0.0001
Likelihood ratio test between models with and without random effect	−486	−478	16	1	<0.0001

Table 7
Results of the secondary crash risk prediction model with random effect.

Variables	Mean	S.D.	2.5%	Median	97.5%	Elasticity
Constant	−4.94	0.90	−6.39	−5.01	−3.35	–
Severity	−0.50	0.26	−0.94	−0.49	−0.07	−25.58 (1.91[a])
Sideswipe	−0.69	0.33	−1.28	−0.68	−0.18	−34.83 (2.75)
Dayweek	−1.32	0.64	−2.45	−1.28	−0.37	−59.93 (4.14)
Roadsurf	0.65	0.28	0.18	0.65	1.10	29.86 (3.84)
Lane	−0.31	0.17	−0.64	−0.29	−0.05	−1.22 (0.22)
AvgCnt	0.23	0.04	0.17	0.23	0.31	2.29 (0.78)
AvgSpd	−0.03	0.01	−0.04	−0.03	−0.02	−1.84 (0.46)
DevOcc	2.56	1.30	0.32	2.62	4.53	0.13 (0.18)
DifCnt	0.13	0.06	0.03	0.13	0.24	0.55 (0.24)
Random effect variance	0.40	0.21	0.05	0.41	0.74	–
Log-likelihood	−478.30	–	–	–	–	–

[a] Note: The standard deviation of elasticity for each variable.

drivers of a crash and prohibiting lane changing within the impact area of the primary crash may help to alleviate the risks of secondary crashes.

The elasticity analysis was further conducted to compare the effects of different traffic flow variables on the secondary crash likelihood. Elasticity tells how many times the secondary crash probability changes if the explanatory variable changes by 1% while the other variables remain fixed. Unlike the marginal effects, the elasticity is dimensionless. Accordingly, it is more convenient for comparing the effects of different variables. The elasticity analysis results of the four traffic flow variables are also given in Table 7. The average elasticity values for these four traffic flow variables (AvgCnt, AvgSpd, DevOcc, and DifCnt) are 2.29, −1.84, 0.13, and 0.55, respectively. It means that the one-percent increase in these four traffic flow variables is associated with the 2.29%, −1.84%, 0.13%, and 0.55% increases in the secondary crash probability, respectively. The traffic volume and speed have relatively higher effects on secondary cash risks, compared with the other two traffic flow variables.

Among the variables for primary crash characteristics, the severity level, collision type and occurrence date of primary crashes were found to be statistically significant. The estimated parameter of the severity level (represented by Severity) is negative, indicating that the injury crashes are less likely to incur secondary crashes. A possible explanation is that freeway agencies usually arrive at the locations of injury crashes more quickly, resulting in lower risks of secondary crashes. Besides, injury crashes often occurred in less congested conditions. The average elasticity for severity level is −25.58. Compared with other types of crashes, the injury crashes have lower likelihood of inducing secondary crashes.

The parameter of sideswipe crashes is negative, which is consistent with the results of the previous studies that rear-end crashes are more likely to incur secondary crashes. The elasticity of −34.83 suggests that if the primary crash is a sideswipe crash, the probability of inducing a secondary crash is 34.83% lower than that of other types of crashes. The negative parameter of occurrence date (represented by Dayweek) indicates that compared with weekdays, crashes occurring in the weekends have lower likelihood of incurring secondary crashes, which is consistent with the findings in previous study (Kopitch and Saphores, 2011). According to the elasticity, the likelihood of secondary crashes on weekends is about 59.93% lower than likelihood of secondary crashes on weekdays.

As expected, the parameter for the number of lanes (represented by Lane) is negative, indicating that crashes occurring on freeway segment with larger number of lanes have lower likelihood of incurring secondary crashes. This finding is intuitive because larger number of lanes decreases the negative impacts of the primary crashes on traffic flow operation, and leaves more space for taking crash avoidance maneuver. The average elasticity of number of lanes is −1.22.

The wet road surface (represented by Roadsurf) is positively associated with the risks of secondary crashes, indicating that the crashes occurred on wet road surface are more likely to induce secondary crashes. Due to the reduced friction between pavement and tires, upstream vehicles require longer distance to decelerate. The corresponding elasticity of 29.86 suggests that the wet road surface conditions increase the probability of secondary crashes by 29.86%. To better reduce the risks of secondary crashes in adverse weather, transportation agency should identify crashes and alert upstream drivers about the crash occurrence in time.

4.4. Predictive performance

The ROC curve was developed to evaluate the predictive performance of the developed secondary crash risk prediction model presented in Table 7. As shown in Fig. 3, the areas under the ROC curve is 0.837, indicating that the developed model provides good predictive performance. Table 8 gives the secondary crash prediction accuracy at different false alarm rates. The developed model was able to correctly identify 66.0% at the false alarm rate of 0.2. Accordingly, this model has adequate accuracy to identify secondary crash in advance and can be used in practical applications to reduce the risks of secondary crashes.

To investigate whether the inclusion of real-time traffic flow data increases the prediction accuracy of secondary crashes, the ROC curves were also developed for the models with and without traffic variables presented in Table 5 (see Fig. 3). The

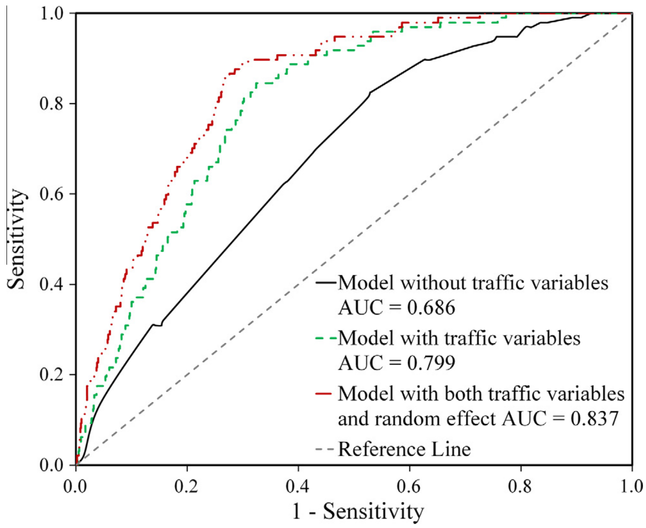

Fig. 3. ROC curves of three different models.

Table 8
Comparison of predictive performance between different models.

1-specificity	Sensitivity of the models		
	Without traffic variables	With traffic variables	With random effect
0.1	22.0%	30.9%	42.3%
0.2	38.9%	53.6%	66.0%
0.3	52.0%	76.3%	85.6%
0.4	65.5%	88.0%	90.7%
0.5	78.9%	91.8%	94.8%

ROC curve for the model with traffic variables is always to the above of the ROC curve for the model without traffic variables, indicating that the secondary crash prediction accuracy of the model with traffic variables is better than that of the model without traffic variables, no matter at what false alarm rates. Table 8 also gives the secondary crash prediction accuracy of the models with and without traffic variables at different false alarm rates. It can be found that the inclusion of real-time traffic flow data increases the prediction accuracy of the secondary crash risk prediction model by an average of 16.6%.

To investigate whether the inclusion of a random effect increases model predictive performance, the authors also compared the ROC curves of the model with random effect and the model with traffic variables. The ROC curve for the model with both traffic variables and random effect is to the above of the ROC curve for the model with only traffic variables (see Fig. 3), indicating that the secondary crash prediction accuracy of the model with random effect in Table 7 is better than that of the model with traffic variables in Table 5. Table 8 compares the secondary crash prediction accuracy of these two models at different false alarm rates. Accounting for the unobserved heterogeneity by including a random effect can further increase the prediction accuracy of secondary crashes by an average of 7.7%.

5. Summary and conclusions

This study investigated the effects of real-time traffic flow variables on the likelihood of secondary crashes. A method based on speed contour plot was first developed to identify secondary crashes on freeways. This method can help to identify the spatial and temporal impact ranges of primary crashes while accounting for the effects of recurrent congestions. The proportionality test was used to compare the characteristics of primary and normal crashes. The random effect logit model was then used to link the probability of secondary crash occurrences with the real-time traffic flow conditions, primary crash characteristics, environmental conditions, and geometric characteristics.

The results showed that the real-time traffic flow variables significantly affect the risks of secondary crashes. The occurrence of a primary crash leads to turbulent traffic conditions, which propagates upstream the crash occurrence location. The upstream drivers might involve in a secondary crash if they travel with high speed and short following distance, when meeting such turbulent traffic conditions. Accordingly, the traffic flow conditions are the contributing factors to secondary crashes. Investigation of the relationship between real-time traffic conditions and secondary crash risks can promote a better understanding the characteristics of secondary crashes and improve prediction accuracy. However, most of the previous studies used static and aggregate traffic data such as AADT, which cannot capture the impacts of sudden variations in traffic conditions on secondary crash risks. The results of this study suggested that incorporating real-time traffic variables increases the prediction accuracy by 16.6%.

The crash characteristics analysis showed that rear-end collision is the predominant collision type for both primary and secondary crashes. The vehicles from upstream stations are required to decelerate abruptly when approaching relatively lower speed traffic where the primary crash occurred. This may partly explain why the percentages of rear-end collision in primary and secondary crashes are higher than that in normal crashes. Accordingly, drivers are required to reduce speed and keep a longer car following distance before meeting hazardous traffic conditions, especially crashes. With regard to the occurrence time, the proportions of primary and secondary crashes occurred in the morning are slight greater than that of the normal crashes, indicating that crashes occurred in the morning period especially the morning peak are more likely to incur secondary crashes.

The impact of unobserved heterogeneity is an important modeling issue that should be considered in the secondary crash risk assessment. However, this issue has been largely overlooked in previous studies. Theoretically, the available explanatory variables only explain part of the variance in the risks of secondary crashes. A number of unobserved variables, such as the work zones, vehicle conditions, pavement conditions, driver behaviors, may have significant effects on secondary crash risks and introduce heterogeneity. Accordingly, overlooking these unobserved heterogeneity may lead to inconsistent and bias parameter estimates (Washington et al., 2003). To account for the unobserved heterogeneity, the random effect logit model was used in this study. The evaluation results showed that the inclusion of random effect increases the prediction accuracy by 7.7%.

The results of this study have the potential to help incorporate secondary crash prevention strategies in the existing DTMSs, such as the ramp metering systems and variable speed limit systems. For example, when a crash occurred, the developed model in this study can be immediately used to estimate the probability of a secondary crash using real-time traffic flow data. If the high risks of secondary crashes are identified, traffic flow control strategies can be applied to alleviate the risks of secondary crashes by reducing the upstream volume and smoothing turbulent traffic conditions. These applications can help reduce the secondary crash risks immediately through the use of DTMSs. Before the results of this study are used in practical engineering, additional efforts are still need to understand the connection between these traffic control methods and secondary crash risks, and to test the transferability of the developed model using data collected from other freeways. The authors suggest that further studies may focus on these issues.

Acknowledgements

This research was sponsored by the National Natural Science Foundation of China (Grant No. 51508093), and Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (No. 5151101143). The authors would like to thank the editor and the anonymous reviewers for their constructive comments and valuable suggestions to improve the quality of the article.

References

Abdel-Aty, M., Dilmore, J., Gayah, V., 2007. Considering various ALINEA ramp metering strategies for crash risk mitigation on freeways under congested regime. Transp. Res. Part C: Emerg. Technol. 15 (2), 113–134.
Abdel-Aty, M., Hassan, H., Ahmed, M., Al-Ghamdi, A., 2012. Real-time prediction of visibility related crashes. Transp. Res. Part C: Emerg. Technol. 24, 288–298.
Anastasopoulos, P., Mannering, F., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.
Carrick, G., Srinivasan, S., Bejleri, I., 2015. Descriptive analysis and characterization of secondary traffic crashes in Florida. Presented at 94th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Chen, D., Ahn, S., 2015. Variable speed limit control for severe non-recurrent freeway bottlenecks. Transp. Res. Part C: Emerg. Technol. 51, 210–230.
Chen, C., Skabardonis, A., Varaiya, P., 2004. Systematic identification of freeway bottlenecks. Transp. Res. Rec. 1867, 46–52.
Chou, C., Miller, H.E., 2010. Simulation-based secondary incident filtering method. J. Transp. Eng. 136 (8), 746–754.
Chilukuri, V., Sun, C., 2006. The use of dynamic incident progression curve for classifying secondary accidents. Presented at 85th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Egan, J.P., 1975. Signal Detection Theory and ROC Analysis. Series in Cognition and Perception. Academic Press, New York.
Golob, T., Recker, W., 2004. A method for relating type of crash to traffic flow characteristics on urban freeways. Transp. Res. Part A 38 (1), 53–80.
Hirunyanitiwattana, W., Mattingly, S.P., 2006. Identifying secondary crash characteristics for the California highway system. Presented at 85th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Hojatia, T., Ferreira, L., Washington, S., Charles, P., Shobeirinejad, A., 2014. Modeling total duration of traffic incidents including incident detection and recovery time. Accid. Anal. Prev. 71, 296–305.
Imprialou, M.I.M., Orfanou, F.P., Vlahogianni, E.I., Karlaftis, M.G., 2013. Methods for defining spatiotemporal influence areas and secondary incident detection in freeways. J. Transp. Eng. 140 (1), 70–80.
Jalayer, M., Baratian-Ghorghi, F., Zhou, H., 2015. Identifying and characterizing secondary crashes on the Alabama state highway systems. Adv. Transp. Stud. 37, 129–140.
Karlaftis, M.G., Latoski, S.P., Richards N.J., Sinha, K., 1998. Empirical analysis of secondary crash causes. Presented at 77th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Karlaftis, M.G., Latoski, S., Richards, P., Nadine, J., Sinha, K.C., 1999. Its impacts on safety and traffic management: an investigation of secondary crash causes. J. Intell. Transp. Syst. 5 (1), 39–52.
Kerner, B.S., Rehborn, H., Aleksic, M., Haug, A., 2004. Recognition and tracing of spatial-temporal congested traffic patterns on freeways. Transp. Res. Part C Emerg. Technol. 12 (5), 369–400.

Khattak, A., Wang, X., Zhang, H., Cetin, M., 2011. Primary and Secondary Incident Management: Predicting Durations in Real-Time Research Report No. VCTIR 11-R11. Virginia Center for Transportation Innovation and Research.
Khattak, A., Wang, X., Zhang, H., 2009. Are incident durations and secondary incidents interdependent? Transp. Res. Rec. 2099, 39–49.
Khattak, A., Wang, X., Zhang, H., 2012. Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays. IET Intel. Transp. Syst. 6 (2), 204–214.
Kopitch, L., Saphores, J., 2011. Assessing the effectiveness of changeable message signs on secondary crashes. Presented at 90th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Li, H., Bertini, R.L., 2010. Comparison of algorithms for systematic tracking of patterns of traffic congestion on freeways in Portland, Oregon. Transp. Res. Rec. 2178, 101–110.
Mishra, S., Golias, M., Sarker, A., Naimi, A., 2016. Effect of Primary and Secondary Crashes: Identification, Visualization, and Prediction Research Report No. CFIRE 09-05. University of Wisconsin-Madison.
Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accid. Anal. Prev. 39, 459–468.
Moore, J.E., Giuliano, G., Cho, S., 2004. Secondary accident rates on Los Angeles freeways. J. Transp. Eng. 130 (3), 280–285.
Park, H., Haghani, A., 2016. Real-time prediction of secondary incident occurrences using vehicle probe data. Transp. Res. Part C: Emerg. Technol. 70, 69–85.
Raub, R., 1997. Occurrence of secondary crashes on urban arterial roadways. Transp. Res. Rec. 1581, 53–58.
Sarker, A.A., Naimi, A., Mishra, S., Golias, M.M., Freeze, P.B., 2015. Development of a secondary crash identification algorithm and occurrence pattern determination in large scale multi-facility transportation network. Transp. Res. Part C: Emerg. Technol. 60, 142–160.
Sun, C., Chilukuri, V., 2010. Dynamic incident progression curve for classifying secondary traffic crashes. J. Transp. Eng. 136 (12), 1153–1158.
Vlahogianni, E.I., Karlaftis, M.G., Orfanou, F.P., 2012. Modeling the effects of weather and traffic on the risk of secondary incidents. J. Intell. Transp. Syst. 16 (3), 109–117.
Wang, J., Liu, B., Zhang, L., Ragland, D.R., 2016a. Modeling secondary accidents identified by traffic shock waves. Accid. Anal. Prev. 87, 141–147.
Wang, J., Xie, W., Liu, B., Fang, S., Ragland, D.R., 2016b. Identification of freeway secondary accidents with traffic shock wave detected by loop detectors. Saf. Sci. 87, 195–201.
Washington, S., Karlaftis, M., Mannering, F., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC, Boca Raton, FL.
Xie, K., Ozbay, K., Yang, H., 2016. A joint analysis of secondary collisions and injury severity levels using structural equation models. Presented at 95th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Yang, H., Bartin, B., Ozbay, K., 2014a. Mining the characteristics of secondary crashes on highways. J. Transp. Eng. 140 (4), 70–80.
Yang, H., Ozbay, K., Bartin, B., 2014b. Assessing the risk of secondary crashes on highways. J. Saf. Res. 49, 143–149.
Yu, R., Abdel-Aty, M., 2014. An optimal variable speed limits system to ameliorate traffic safety risk. Transp. Res. Part C: Emerg. Technol. 46, 235–246.
Zhan, C., Gan, A., Hadi, M., 2009. Identifying secondary crashes and their contributing factors. Transp. Res. Rec. 2102, 68–75.
Zhan, C., Shen, L., Hadi, M., Gan, A., 2008. Understanding the characteristics of secondary crashes on freeways. Presented at 87th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
Zhang, H., Khattak, A., 2010. What is the role of multiple secondary incidents in traffic operations? J. Transp. Eng. 136 (11), 986–997.
Zhang, H., Cetin, M., Khattak, A.J., 2015. Joint analysis of queuing delays associated with secondary incidents. J. Intell. Transp. Syst. 19 (2), 192–204.
Zhang, L., Xie, C., 2008. Travel mode choice modeling with support vector machines. Transp. Res. Rec. 2076, 141–150.