Accident Analysis and Prevention 119 (2018) 58–67

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Likelihood estimation of secondary crashes using Bayesian complementary log-log model

Angela E. Kitali[a,*], Priyanka Alluri[a], Thobias Sando[b], Henrick Haule[a], Emmanuel Kidando[c], Richard Lentz[b]

[a] Department of Civil and Environmental Engineering, Florida International University, 10555 West Flagler Street, EC 3680, Miami, 33174, FL, United States
[b] School of Engineering, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224, United States
[c] Department of Civil and Environmental Engineering, FAMU–FSU College of Engineering, Tallahassee, FL 32310, United States

ARTICLE INFO

Keywords:
Secondary crashes
Complementary log-log
Likelihood
Real-time traffic data

ABSTRACT

Secondary crashes (SCs) occur within the spatial and temporal impact range of a primary incident. They are non-recurring events and are major contributors to increased traffic delay, and reduced safety, particularly in urban areas. However, the limited knowledge on the nature of SCs has largely impeded their mitigation strategies. The primary objective of this study was to develop a reliable SC risk prediction model using real-time traffic flow conditions. The study data were collected on a 35-mile I-95 freeway section for three years in Jacksonville, Florida. SCs were identified based on travel speed data archived by the Bluetooth detectors. Bayesian random effect complementary log-log model was used to link the probability of SCs with real-time traffic flow characteristics, primary incident characteristics, environmental conditions, and geometric characteristics. Random forests technique was used to select the important variables. The results indicated that the following variables significantly affect the likelihood of SCs: average occupancy, incident severity, percent of lanes closed, incident type, incident clearance duration, incident impact duration, and incident occurrence time. The study results have the potential to proactively prevent SCs.

1. Introduction

Traffic incidents are a significant cause of congestion, leading to capacity reduction and deterioration of the quality of service of transportation systems. They account for approximately 25% of all traffic delays (Owens et al., 2010). Moreover, they sometimes cause secondary crashes (SCs) that put other road users and incident first responders' lives at risk. In general, SCs are defined as crashes that occur within the spatial and temporal impact range of a prior incident commonly referred to as a primary incident. More specifically, crashes are considered as SCs if they occur: (a) at the scene of the primary incident; or (b) within the queue upstream of the primary incident; or (c) within the queue in the opposite direction of the primary incident caused due to driver distraction (i.e., rubbernecking effect) (Raub, 1997; Moore et al., 2004; Chang and Rochon, 2011; Wang et al., 2016b).

Statistics indicate that up to 15% of reported crashes are partly or entirely due to primary incidents (Raub, 1997). In a more recent study, Owens et al. (2010) determined that SCs account for 20% of all crashes and 18% of all fatalities on freeways. Moreover, SCs are non-recurrent in nature and lead to significant increased risk of additional crashes, reduced freeway capacity, increased delay, and decreased travel time reliability. Therefore, minimizing the occurrence of SCs is one of the major focus areas for transportation agencies, in particular Traffic Management Centers (TMCs) (Owens et al., 2010).

The primary objective of this study is to develop a SC risk prediction model using real-time traffic flow conditions. The developed SC risk prediction model can be used by the state and local transportation agencies for developing SC mitigation strategies, and hence improving operational and safety performance of freeways.

2. Literature synthesis

Existing literature on SCs has mainly focused on three aspects: methods to identify SCs; characteristics of SCs; and factors that influence the occurrence of SCs. This review is therefore summarized in three sections based on the above three topics.

* Corresponding author.
E-mail addresses: akita002@fiu.edu (A.E. Kitali), palluri@fiu.edu (P. Alluri), t.sando@unf.edu (T. Sando), hhaul001@fiu.edu (H. Haule), ek15f@my.fsu.edu (E. Kidando), r.lentz@unf.edu (R. Lentz).

https://doi.org/10.1016/j.aap.2018.07.003
Received 27 February 2018; Received in revised form 29 May 2018; Accepted 1 July 2018
Available online 07 July 2018
0001-4575/ Published by Elsevier Ltd.

2.1. SC identification

The first and the most critical step in identifying SCs is to determine whether the potential SCs are within the spatial and temporal boundaries of primary incidents. Two major approaches have been used to define the spatiotemporal thresholds: the static method that uses predefined spatiotemporal limits; and the dynamic approach that uses varying spatiotemporal thresholds based on primary incident characteristics and real-time traffic performance measures.

The static method identifies SCs based on some fixed spatial and temporal criteria. Crashes that occurred within a maximum spatial and temporal impact range of a primary incident are identified as SCs. For SCs occurring upstream of the primary incident, the spatial and temporal thresholds employed by previous studies range from 1 to 2 miles and 15 min to 2 h, respectively (Raub, 1997; Karlaftis et al., 1999; Moore et al., 2004; Hirunyatiwattana and Mattingly, 2006; Zhan et al., 2008). On the other hand, SCs occurring on the opposite direction of the primary incident were identified using different thresholds. Chang and Rochon (2011) identified SCs using 30-minutes-0.5-mile threshold in the opposite direction of the primary incident. Some studies considered crashes that occur in the opposite direction within one hour and one mile upstream of previous incidents to be potential SCs (Yang et al., 2014a,b). The static method is limited by the fact that it requires a subjective determination of the fixed spatiotemporal thresholds. It incorrectly assumes that various incident type, which occur at different traffic conditions such as congested and free-flow states, have the same effects on the traffic stream. In fact, incidents occurring during free-flow conditions might not have a long lasting and far reaching effect compared to the incidents occurring during congested traffic flow conditions. In addition, severe incidents tend to lead to a longer incident response and clearance time compared to minor incidents. Thus, both spatial and temporal thresholds should vary based on traffic conditions, geometric characteristics, and certainly incident characteristics.

To overcome the limitation associated with the static approach, recent studies have focused on the use of dynamic methods to detect SCs. Dynamic approaches identify SCs based on varying spatiotemporal thresholds. Sun and Chilukuri (2006,2010) proposed the use of an incident progression curve, a method that uses incident impact duration to estimate the queue length and hence identify SCs that occurred within the queue. The incident progression curve method indicated a 30% improvement in the SCs identification accuracy compared to the static method. Other studies have dynamically identified SCs based on cumulative arrival and departure plots (Zhan et al., 2009) and simulation-based approaches (Chou and Miller-Hooks, 2010).

Other dynamic methods, i.e., speed contour, automatic tracking of moving jams, vehicle probe data, and shock wave principles, determine flexible spatiotemporal thresholds based on the primary incident influence area (Khattak et al., 2011; Imprialou et al., 2013; Yang et al., 2014a, b; Sarker et al., 2015; Zhang et al., 2015; Mishra et al., 2016; Park and Haghani, 2016; Wang et al., 2016a, b; Xu et al., 2016). This approach helps capture the effects of traffic characteristics, e.g., flow, speed, and density that change over time and space, and the effect of the queue formation from the primary incident occurrence. The results from these studies indicate that the proposed dynamic methods provide better accuracy in detecting SCs than conventional static methods.

2.2. SC characteristics

Carrick et al. (2015) compared roadway, environmental, and vehicle characteristics of secondary and normal crashes. SCs were observed to be more likely to occur on freeways and in rainy weather conditions. Another study by Zhang et al. (2015) used a microscopic simulation tool to study the queuing delays associated with SCs. SCs were found to result in longer incident impact duration than normal incidents. Further, the time gap and distance between a primary incident and its SC were observed to significantly affect the total delays. Mishra et al. (2016) concluded that SCs on freeways are more likely to occur during the morning and evening peaks, while the SCs on arterials are more common during the evening peak.

A number of studies have also been conducted to investigate the relationship between the likelihood of SCs and various contributing factors including primary incident characteristics, weather conditions, geometric conditions, traffic volumes, and roadway functional classification (Khattak et al., 2009; Zhang and Khattak, 2010; Khattak et al., 2012; Yang et al., 2014b; Wang et al., 2016a,b). In general, factors contributing to SCs were observed to be: number of vehicles involved in primary incidents, primary incident impact duration, number of lanes blocked, traffic volume, and posted speed limit (Khattak et al., 2009,2012; Chimba et al., 2014; Wang et al., 2016a,b; Xu et al., 2016). Traffic incidents that occurred during the off-peak hours and on weekends are less likely to induce SCs (Khattak et al., 2009,2012; Yang et al., 2014a).

Further, adverse weather conditions such as rain and snow were found to significantly increase the risks of SCs (Khattak et al., 2012; Wang et al., 2016a,b). One common theme among most SCs studies is that they all associate the occurrence of SCs with the primary incident impact duration. Thus, examining the primary incident impact duration is crucial in developing SC occurrence prediction models.

2.3. SC likelihood models

Several studies have analyzed the likelihood of SCs using either non-parametric or parametric models. Several non-parametric models such as neural networks and decision trees have been used to model SC risk. Vlahogianni et al. (2010) developed a Bayesian network for the probabilistic estimation of different influence areas for SCs with respect to various incident and traffic characteristics. Traffic conditions at the time of an incident and incident clearance duration were observed to be the most significant determinants in defining the upstream influence of a crash. Later, Vlahogianni et al. (2012) developed a neural network model with enhanced explanatory power. The study reported that traffic speed, duration of the primary incident, hourly volume, rainfall intensity, and number of vehicles involved in the primary crash as the most significant determinants associated with SC likelihood.

Several other studies developed decision tree models to explore contributing factors based on the prediction results of artificial neural networks algorithm. For example, by identifying SCs based on the binary speed contour plot map using probe vehicles data, Park and Haghani (2016) predicted the likelihood of SCs using Bayesian neural networks model and extracted rules to generate gradient-based decision trees. In turn, the main determinants that influence the occurrence of SCs were shown based on the decision tree. Apart from directly modeling the SC occurrence risk, Wang et al. (2018) used two machine learning algorithms (back-propagation neural network and a least square support vector machine) to model the spatial and temporal gaps between the primary and secondary incidents. It was reported that both algorithms failed to predict the spatial parameter while the back-propagation neural network algorithm outperformed the least square support vector machine algorithm in time prediction.

Most of the studies that developed parametric models used either logit or probit models to analyze the likelihood of SCs (Zhan et al., 2009; Khattak et al., 2009,2012; Yang et al., 2014b; Mishra et al., 2016; Wang et al., 2016a, b). Both logit and probit models are symmetrical in nature, i.e., the likelihood of SC occurrence is presumed to rise up to a probability of 0.5, then decrease toward the asymptote at one (1). In other words, in SC likelihood prediction, symmetric models such as logit or probit models are applicable only when the proportion of normal incidents (~50%) is equal to the proportion of primary incidents (~50%). However, SCs account for less than 20% (Owens et al., 2010) of total incidents, meaning that the proportion of primary incidents is much less than the proportion of normal incidents (i.e., the primary incidents and the normal incidents are asymmetrically distributed). Thus, a model which is asymmetrical around the inflection point is considered to be more reliable in predicting the likelihood of SCs. With this situation at hand, a complementary log-log model (cloglog), which is used as an alternative prediction model over the conventional logit and probit models would be a better model. Unlike the logit and probit models, the cloglog model is asymmetrical with a fat tail as it departs from zero (0) and sharply approaches one (1) (Kitali et al., 2017; Martin and Wu, 2017).

In modeling SCs, most of the previous studies have used general roadway characteristics such as average annual daily traffic (AADT) and speed limit as some of the model variables (Zhang and Khattak, 2010; Khattak et al., 2012; Chimba et al., 2014; Mishra et al., 2016;). These variables limit the reliability of the study findings simply because they are averages and do not reflect real traffic conditions at the time of an incident. Relatively few studies have predicted the likelihood of SCs using real-time data considering the effect of dynamic traffic flow conditions (Vlahogianni et al., 2012; Park and Haghani, 2016; Xu et al., 2016). This study uses the Bayesian cloglog model with a random parameter to study the likelihood of SCs. The model considers incident-related, geometric-related, environmental-related, and real-time traffic-related variables. Use of the model that considers the asymmetrical nature of SCs helps improve the predictive performance of the SC risk model. Further, incorporating real-time traffic flow conditions in lieu of traditional variables (AADT and speed limit) provides a more representative scenario of the actual traffic conditions at the time of the primary incident.

3. Data

The study area includes a 35-mile section on I-95 freeway located in Jacksonville, Florida. It spans from the South to the North ends of Duval County, as shown in Fig. 1. Data used in this study includes speed data from BlueToad devices, incident data from SunGuide database, and real-time traffic data from the Regional Integrated Transportation System (RITIS) for the years 2015-2017. Since BlueToad devices have not yet been extensively deployed, the study area includes only the corridor with these devices.

In addition to these data, several geometric characteristics including median width, type of roadside barrier, and presence of horizontal curve, were extracted from the 2015 Florida Department of Transportation (FDOT) Roadway Characteristics Inventory (RCI), and included in the analysis (FDOT, 2015). The following paragraph discusses the approach employed to identify incidents affected by the horizontal curve.

All incidents that occur on horizontal curves may not necessarily cause queue-visibility problem leading to SCs. For example, the red line in Fig. 2 shows a horizontal curve on I-95. Incidents A and B occurred on I-95 northbound (NB); while incidents C and D occurred on I-95 southbound (SB). Since Incident A occurred at the beginning of the horizontal curve on I-95 NB, it does not cause queue-visibility problem leading to a SC.

On the other hand, Incident B on I-95 NB lanes occurred at the end of the horizontal curve, and may cause queue-visibility problem leading to a SC. Similarly, on I-95 SB lanes, Incident D does not cause queue-visibility problem while Incident C may cause queue-visibility problem. Incidents on horizontal curves were identified using this approach.

Tables 1 and 2 provide the descriptive statistics of the categorical and continuous variables considered in this study, respectively.

3.1. BlueToad devices

BlueToad devices are Bluetooth signal receivers which read the media access control (MAC) addresses of active Bluetooth devices in vehicles passing through their area of influence. These devices act in pairs or network by recording the time when a vehicle passes both devices. This information is used to deduce travel time of the vehicle between a pair of devices. The speed is calculated from the obtained travel time and a known path distance (not Euclidean distance) between the devices.

The study location has 31 BlueToad pair devices (15 Northbound and 16 Southbound) placed approximately every 1.7 miles on the mainline. The posted speed limit on the entire section ranges between 55 mph and 70 mph. This study used data collected by each pair device.

3.2. SunGuide

SunGuide is a software used for incident management to process and archive incident data. The database stores incident attributes including incident ID, incident timeline, and incident type. In the event of a SC, the database links it to the primary incident. Along the 35-mile study corridor, the SunGuide database included a total of 22,492 incidents from 2015–2017. The categories of events included in the SunGuide database are crash, debris on roadway, disabled vehicle, emergency vehicle, flooding, police activity, and vehicle fire. These categories were summarized into three groups: crash, vehicle-related (disabled vehicle, emergency vehicle, and vehicle fire), and others (flooding and debris on roadway).

After excluding incidents on ramps (5941), incidents with missing coordinates (78), incidents with no matched BlueToad pairs (9740), incidents along the section without BlueToad pairs (3253), the remaining data consisted of a total of 3480 incidents. These incidents were categorized into three categories: normal incidents, primary incidents, and SCs. Note that normal incidents are those that did not lead to any SCs; primary incidents are those that lead to SCs; and SCs are those that occurred due to primary incidents. Section 4.1 discusses the method used to identify SCs. Of the 3480 incidents included in the analysis, 1024 were further dropped from further analysis. Of these 1024, a total of 819 incidents were missing some of the attributes, and the remaining 205 incidents were marked as non-tertiary SCs, meaning that these are SCs that did not result in additional SCs. In total, 293 incidents (88 being tertiary SCs) were identified as SCs accounting for 8.42% of the 3480 incidents that were included in the analysis. The 242 primary incidents that induced SCs represents 7.6% of all non-secondary incidents (3480-293 = 3187). These results indicate that approximately one in every thirteen normal incidents was associated with a SC.

3.3. RITIS

Traffic data such as speed, volume, and occupancy, were extracted from RITIS maintained by the Center for Advanced Transportation Technology (CATT) laboratory. Data retrieved from the RITIS database are used as potential variables that could influence the likelihood of occurrence of SCs. This is because the RITIS database provides high-resolution raw traffic data including speed (miles per hour-mph), volume (vehicles/30 s), and detector occupancy (the percent time that the sensor is occupied at 30-sec intervals). There are 375 RITIS detector stations along the selected freeway corridor (183 on NB and 192 on SB directions). The average spacing between detectors is approximately half a mile.

In this study, the spatiotemporal thresholds of 30-min and 1-mile radius are used to capture traffic conditions before the occurrence of the incident. Note that traffic data just before the incident occurrence account for potential inaccuracies in the reported incident time. The 30-min temporal threshold is hence selected based on existing literature. Previous studies including Ahmed et al. (2012b), Yu et al. (2016), etc., have used 30 min as the temporal parameter to capture traffic conditions prior to the incident occurrence.

Regarding the spatial threshold, previous studies have used detectors closest to the incidents to capture traffic conditions in the vicinity of the incidents. For example, Ahmed et al. (2012b) used distance from

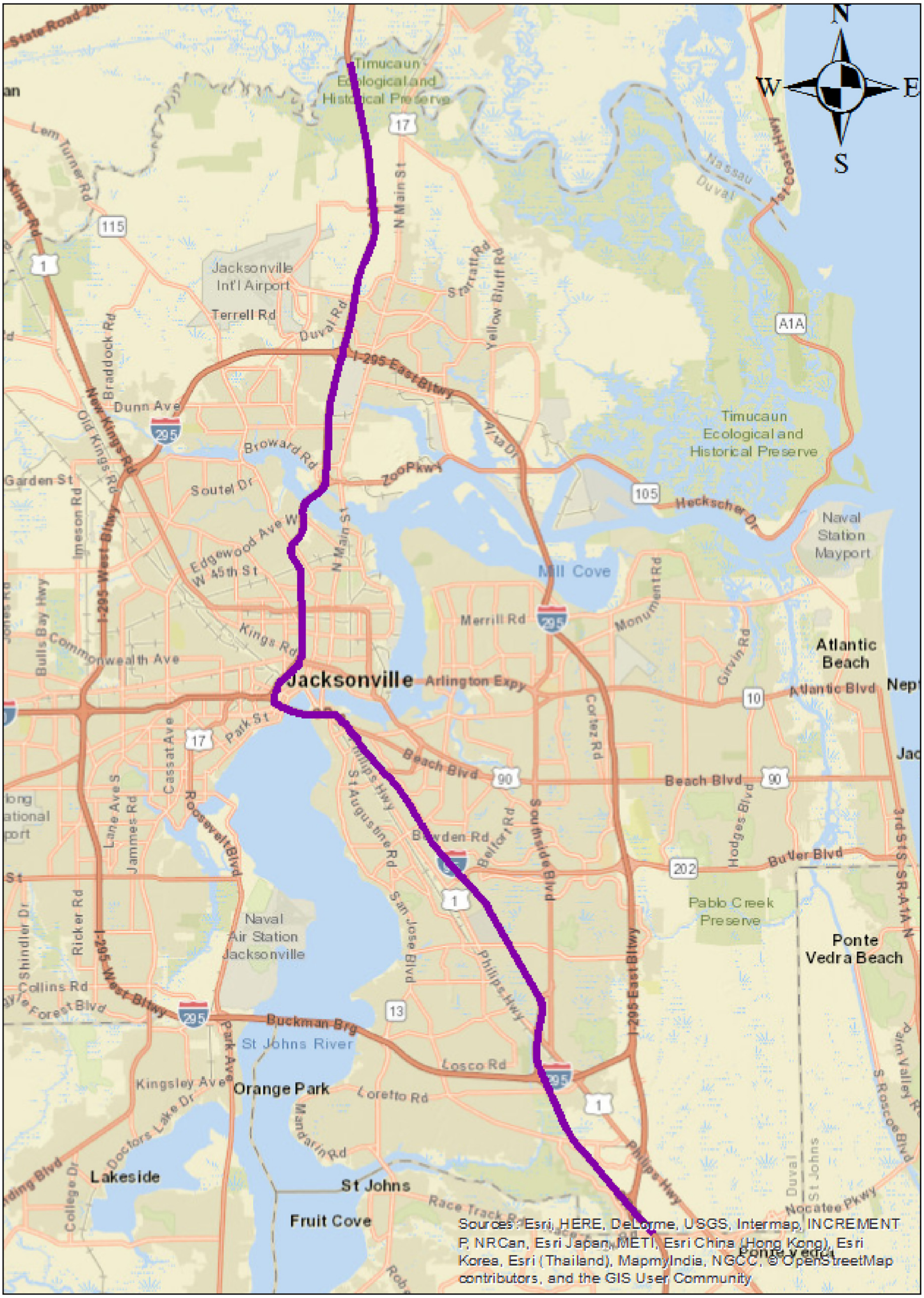

Fig. 1. Study Area.

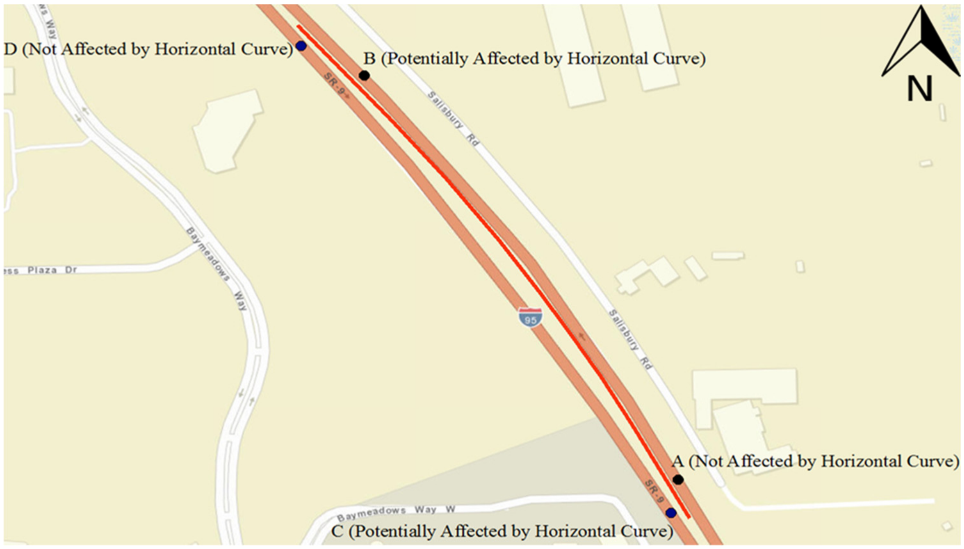

Fig. 2. Identification of Incidents Affected by the Horizontal Curved Segments.

Table 1
Descriptive Statistics of Categorical Variables.

Parameter	Factor	Count	Percent (%)
Roadway alignment	Straight	1,857	75.61
	Curved	599	24.39
Incident occurrence time	Off-peak	1,196	48.70
	Peak	1,260	51.30
Incident type	Other	130	5.29
	Vehicle-related	1,280	52.12
	Crash	1,046	42.59
Emergency Medical Services (EMS) involved	No	2,266	92.26
	Yes	190	7.74
Towing involved	No	2,077	84.57
	Yes	379	15.43
Number of responding agencies	1	1,208	49.19
	2–3	965	39.29
	> 3	283	11.52
Percent lane closed (%)	0–25	2,098	85.42
	> 25	358	14.58
Shoulder blocked	No	1,052	42.83
	Yes	1,404	57.17
Primary incident severity	Minor	2,424	98.70
	Moderate/severe	32	1.30
Detection method[a]	On-site	2,121	86.36
	Off-site	335	13.64
Roadside barrier	Guard rail	1,238	50.41
	Barrier wall	1,218	49.59
Lighting condition	Daytime	2,274	92.59
	Nighttime	182	7.41

[a] Identifying SCs on a closed-circuit television (CCTV) camera at a TMC is considered an off-site approach; while identifying SCs on site by the incident responders including police, road rangers, etc. is considered an on-site approach.

four detectors (two upstream of the incident and two downstream of the incident) to capture traffic conditions near the incident. Selecting a fixed number of detectors would not result in a consistent impact area for incidents, especially when detectors are not uniformly spaced. In this research, instead of selecting a fixed number of detectors, a spatial threshold of 1-mile is used to locate *all* detectors that are within 1 mile of the incident. Moreover, using the 1-mile radius, an average of four and six detectors were assigned to northbound and southbound incidents, respectively. This number is consistent with previous studies (Pande et al., 2011; Ahmed et al., 2012b; Yu et al., 2016).

Table 2
Descriptive Statistics of Continuous Variables.

Variable	Minimum	Mean	Median	SD	Maximum
Primary incident impact duration (min)[1]	30	97	75	87	855
Incident clearance duration (min)	0	44	27	46	417
Average vehicle speed (mph)	5	56	61	15	78
SD of vehicle speed (mph)	0	9	7	6	35
Average EHV (veh/hour)	0	38	32	39	624
SD of EHV (veh/hour)	0	17	11	36	465
Average detector occupancy (%)	0	10	7	7	47
SD of detector occupancy (%)	0	5	3	3	20
Median width (ft)	16	39	40	28	150

Note: SD = standard deviation, EHV = Equivalent hourly volume.
[1] Time taken for the traffic to come back to normal after the occurrence of the primary incident.

4. Methodology

The primary objective of this study is to develop a SC risk prediction model based on real-time traffic flow conditions. The objective was achieved using the following steps. Potential SCs were first identified using real-time speed data from BlueToad paired devices. Most important variables contributing to SCs were next screened using Random Forests approach. Finally, the Bayesian random effect complementary log-log (cloglog) model was used to predict the probability of SCs. The following sections discuss the study methodology in detail.

4.1. Dynamic method to identify SCs

In this study, SC identification method focuses on identifying the impact range of the primary incident using speed data archived by the BlueToad paired devices and detecting SCs occurring within the impact range of the primary incident. This method aims to better capture the effects of traffic flow characteristics such as speed that change over space and time, and affect the queue formation as a result of a primary incident.

The raw speed data from BlueToad paired devices were extracted to determine the spatiotemporal effects of primary incidents under the prevailing traffic conditions. Since the analysis is sensitive to the choice of appropriate time interval used in measuring traffic flow parameters (speed, in this research), the speed data aggregated in 15-min intervals was used in this study. Note that natural traffic flow at shorter time

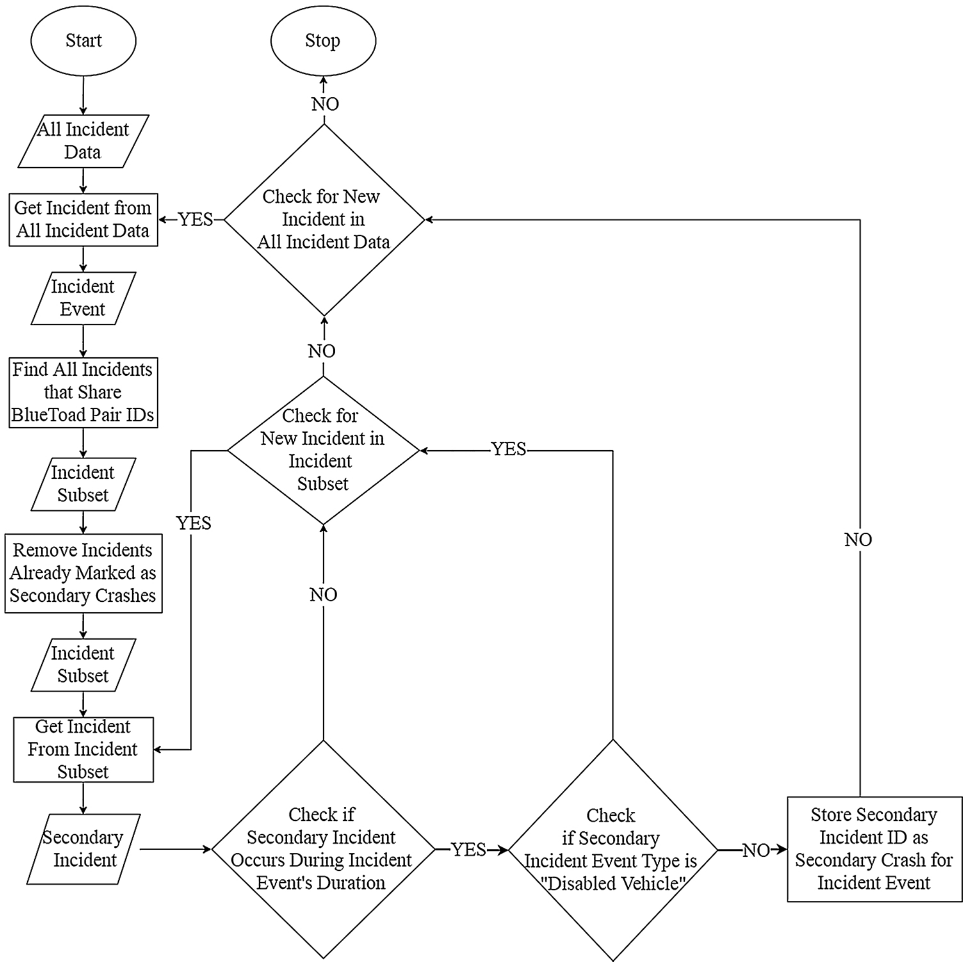

Fig. 3. Identification of Secondary Crashes using Speed Data.

intervals will contain a large amount of noise (Guo et al., 2018), and previous literature has recommended using a minimum of 15-min measurement intervals in order to obtain stable traffic flow rates (Smith and Ulmer, 2003).

Each incident was first matched to a specific BlueToad pair located along the roadway segment based on geographical coordinates (i.e., latitudes and longitudes). The speed data at the time of each incident was retrieved from the matched BlueToad pair. Historical speed data for the BlueToad device pairs with matched incidents were used to establish recurrent speed profile of the section under normal traffic conditions. Average speed at 15-min intervals was used to establish the speed profile. Furthermore, a 95% confidence interval (CI) was established to define the upper and lower bounds of the speed profile to account for the recurrent speed variations.

Following the establishment of the recurrent speed for each of the incident-related BlueToad device pairs, aggregated 15-min speed data from the time of the incident occurrence were collected in order to identify upstream BlueToad pairs affected by the incident. As illustrated in Fig. 3, for each of the incident subset, the retrieved vehicle speeds since the incident reported time were compared with the recurrent speeds of the respective pairs. Note that the incident subset in Fig. 3 refers to a set of incidents that occurred within a similar BlueToad pair device.

This process was implemented on both upstream direction and opposite direction of the prior incident, to identify pairs affected by the occurrence of the incident. A BlueToad pair is considered to be affected by the occurrence of an incident when the speeds from the incident reporting time are lower than the defined boundary of recurrent speeds. Once all the BlueToad pairs affected by the incident subset (both NB and SB directions) were identified, the BlueToad pairs were then checked to identify whether there was another incident occurring within the affected BlueToad pairs. All incidents identified within the affected BlueToad pairs were checked to determine whether they occurred within the time that the current speed falls below the average speed of the respective BlueToad pair. Next, all identified incidents within the incident impact duration of the prior incident were checked and all incidents that were crashes, are considered as SCs and were retrieved.

A crash is identified as "secondary" if it occurred within the impact area of a primary incident, i.e., it occurred within the primary incident impact duration and location where speed dropped below the recurring speeds along this section. This applies for crashes occurring both on the upstream and opposite direction of the primary incident. This approach was implemented in the open source statistics software "*R*".

Fig. 4 describes the example of incidents that occurred on Monday, August 29th, 2016 along I-95 NB in 15-min speed intervals. On this particular day, ten incidents that occurred along the study corridor resulted in significant congestion, i.e., average speeds dropped below the recurring speeds along this corridor. Three of these eleven incidents were identified as SCs. The first SC (S1) occurred within the incident impact duration of the first primary incident (P1). This primary incident (P1) occurred at 1451 h and affected eight BlueToad pairs on the upstream direction (8.5 miles). It is worth noting that the speed along the BlueToad pair #6 came back to normal much earlier than the rest of the pairs. Due to congestion caused by the primary incident (P1), drivers might have detoured to other parallel routes (e.g., I-295). BlueToad pair

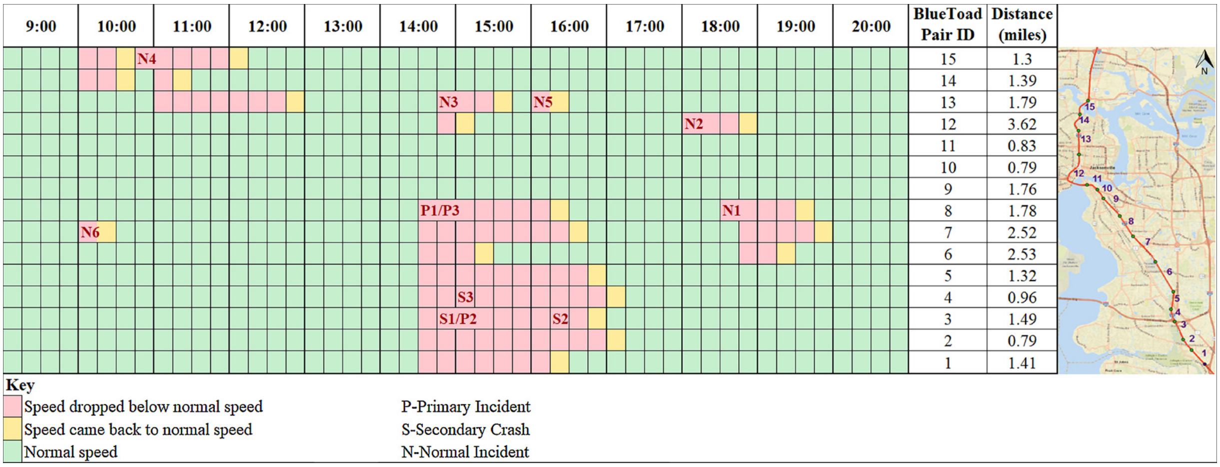

Fig. 4. Detection of Secondary Crashes.

#6 is the only pair with an exit in the middle of the two Bluetooth devices. A SC (S1) occurred 16 min later and 7.8 miles upstream of the primary incident. The SC S1 resulted in a significant drop in speeds on the pair that it occurred (#5) plus two other pairs (#3 and #4) on its upstream direction. Thirty-six minutes later, the same primary incident (P1) resulted in another SC (S3) at about 7.01 miles upstream of P1. This incident resulted in a significant non-recurring congestion on the pair that it occurred (#4) and three other pairs on the upstream direction (#1, #2, and #3).

Incident S1 turned out to be a tertiary crash, meaning that it is a SC that became a primary incident (P2) to another SC (S2), representing cascading events. Incident P2, which occurred at 1507 h along BlueToad pair #3 affected one additional pair #2, on the upstream direction. Secondary crash S2 occurred on the same pair (#5) as its primary incident 85 min later and affected one pair on the upstream direction (#4). It can be inferred from this observation that apart from resulting in non-recurring congestion, SCs can also lead to additional crashes, herein referred to as tertiary crashes.

In general, several of the SCs were found to contribute to additional crashes. Six of the incidents in Fig. 4 were normal incidents meaning that they did not result in SCs. For example, the normal incident N1, occurred at 1857 h on pair #8 and resulted in a significant drop in speed along two additional pairs (#6 and #7) on the upstream direction.

4.2. Random forests approach to select important variables

The Random Forests (RF) model is a non-parametric statistical method that is based on decision-trees (Breiman, 2001). More recently, traffic safety researchers are increasingly using the RF approach to select the important variables before applying other statistical models (Abdel-Aty and Haleem, 2011; Ahmed et al., 2012b; Haleem et al., 2015; Theofilatos, 2017). Unlike the classification and regression tree models, RF models can provide unbiased error estimates and does not require a cross-validation test (Breiman, 2001). During the tree-growing process, one-third of the training cases are left out and not used in the growing of the tree, conventionally referred to as out-of-bag (OOB) data.

The principle of RF is to aggregate many binary decision trees coming from two random perturbation mechanisms, i.e., the use of bootstrap samples and the random choice of a subset of explanatory variables at each node. The RF algorithm is used in this study to estimate prediction performance and quantify variable importance based on the OOB error.

RF use OOB samples to measure the prediction strength of each variable by constructing a different variable-importance measure. These values are in turn used to generate the accuracy plot that test to see how worse the model would perform without each variable. The use of OOB randomization to compute the variable importance using mean decrease accuracy plot tends to spread the importance more uniformly (Hastie et al., 2008). Note that several other studies have employed similar plots to identify important variables (Yu and Abdel-Aty, 2014; Theofilatos, 2017). Thus, the mean decrease accuracy (MDA) plot provided by the *R* package "randomForest" was used to select the important variables (Liaw and Wiener, 2015). A higher accuracy value represents a higher variable importance.

4.3. Bayesian framework to model SCs

The Bayesianrandom-effect complementary log-log (cloglog) model was used to predict the probability of SCs. Specifically, this model was used to develop a SC risk prediction model, in which the likelihood of SCs was linked with real-time traffic variables, primary incident characteristics, environmental conditions, and geometric characteristics. It is worth noting that, unlike previous studies which used the conventional logit and probit model (symmetrical models), this study uses the cloglog model to account for the asymmetric distribution of the response variable (only 8.0% of all crashes are SCs). The normal random effect parameter was included to account for the heterogeneity caused by the unobserved factors such as work zones, design features, and pavement conditions, among other factors. Failure to account for the unobserved variation in the data may lead to inconsistent and bias parameter estimates (Mannering et al., 2016; Xu et al., 2016).

In this study, the response variable is binary in nature, i.e., y_i represents the SC indicator (1 indicates a SC is induced by a primary incident (i), and 0 indicates that no SC crash occurred). π_n denotes the probability of a SC induced by a primary incident; **X** denotes the vector of explanatory variables used in the study, $\boldsymbol{\beta}$ is the coefficients vector for explanatory variables **X**. The random-effect cloglog model can be presented using Eqs. (1) and (2).

$$y_i \sim \mathrm{Binomial}(\pi_i) \tag{1}$$

$$\mathrm{cloglog}(\pi_i) = \log(-\log(1-\pi_i)) = \boldsymbol{\beta}\mathbf{X} + \varepsilon_i \tag{2}$$

ε_i is the normal random effect variable in the model, which represents the incident level random error.

The Bayesian inference for cloglog regression follows the usual procedure for all Bayesian analysis. In Bayesian inference, a prior distribution for all unknown parameters has to be defined. Normally, two categories of priors are used in the Bayesian approach; informative and non-informative priors. Informative priors are based on the literature, expert knowledge, or information retrieved explicitly from a previous data analysis (Ahmed et al., 2012a). On the other hand, non-informative priors, also called "vague" priors, are often used in the absence of reliable prior information regarding model parameters (Huang et al., 2008; Kitali et al., 2017). For this study, there is no prior knowledge of the expected effect; hence, non-informative priors are used. The most commonly used priors are normal distributions with a zero mean, expressing the prior doubt of the relationship between the predictor variable and the response variable, and large variance. Thus, the coefficients of the predictor variables were set up with non-informative priors following normal distributions with a zero mean and a variance of one, i.e., Normal (mean = 0, SD = 1) for predictor parameters, intercept, and random parameter. The first 10,000 iterations were discarded as burn-in sample and 4 chains of 20,000 iterations were set up. The specification of the priors is then followed by estimation of the likelihood function. The likelihood function for the cloglog regression can be expressed using Eq. (3).

$$\text{Likelihood} = \prod_{i=1}^{n} [\pi(x_i)^{y_i}(1-\pi(x_i))^{(1-y_i)}] \tag{3}$$

Where $\pi(x_i)^{y_i}$ is the probability of the event for the ith incident, which has covariate vector $\boldsymbol{X}$.

The priors and the likelihood function are then used to estimate the posterior distribution of the study parameters (Eq. (4)). The Bayes theorem is normally applied while estimating the posterior distribution of all parameters.

$$\text{Posterior} = \text{Prior} \times \text{Likelihood} \tag{4}$$

The proposed model was implemented through Rstanarm, an open source "*R*" package. The ratios of the Monte Carlo (MC) errors relative to standard deviations of the estimates, trace, density, and autocorrelation plots were monitored to achieve parameter estimation convergence. As a rule of thumb, the MC error was maintained at less than 5% of the posterior standard deviation for a parameter to converge (Huang et al., 2008). The 95% BCI was used to determine the significance of the predictor variables, which provides probability interpretations with normality assumptions that the true parameter is inside the region with measured probability (Huang et al., 2008).

5. Results and discussion

5.1. Variable importance

RF algorithm was used to estimate the importance of each of the predictor variable by monitoring the change in the prediction error when OOB data for the respective variables are permuted while all other remaining variables are left unchanged (Liaw and Wiener, 2015). Forests were grown using 1200 trees and by randomly selecting 6 predictor variables at each node for splitting, since these combinations yield stable results with minimum OOB error rate of 0.025. Fig. 5 shows the final results of the variable importance ranking where the MDA was used as the selection criterion.

The cut-off value of 10 for the MDA was chosen to identify the important variables that yield meaningful parameter estimates. The following 16 variables were identified as important, and were included in the model: incident impact duration, incident clearance time, standard deviation of EHV, mean of occupancy, incident type, standard deviation of occupancy, mean of EHV, mean of speed, standard deviation of speed, number of responding agencies, incident severity, towing involved, median width, percent lane closed, EMS involved, and incident time.

5.2. Variable correlation

At least some of the variables (e.g., traffic-related variables) identified by the RF technique are considered to be correlated. Using Pearson correlation, a correlation matrix was built to identify and

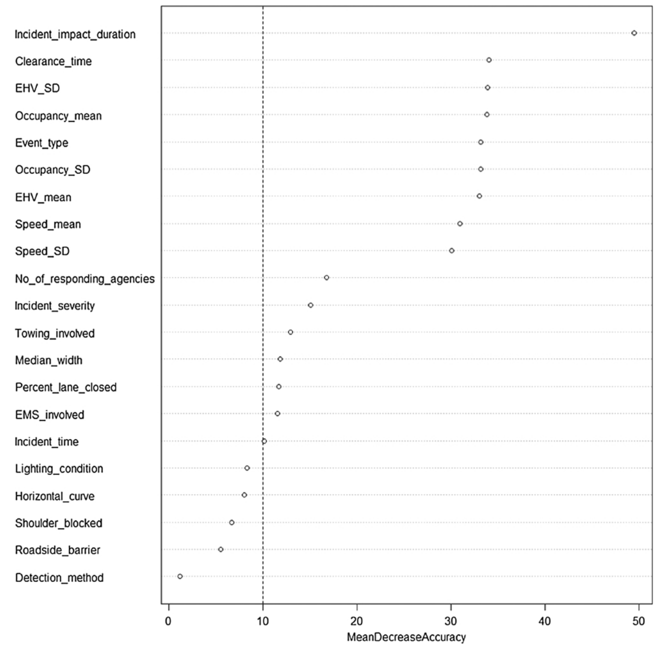

Fig. 5. Variable Importance Ranking using Random Forests Technique.

exclude highly correlated variables. A correlation threshold of 0.5 was used to identify highly correlated variables (Kobelo et al., 2008; Dissanayake and Roy, 2013). The standard deviation of occupancy variable was dropped from the analysis since it is highly correlated with two variables, mean of occupancy (0.8) and mean of speed (−0.7). Furthermore, the mean of EHV was also removed from further analysis since it is also correlated with the standard deviation of EHV variable (0.9) and mean of occupancy (0.4). Finally the remaining 13 variables were used as input variables in modeling the SC likelihood.

5.3. Model results

Table 3 summarizes the findings of the final Bayesian cloglog model with a random parameter. Seven (7) of the 13 variables are significant at the 95% Bayesian Credible Interval (BCI). Note that the predictor variable in the model is considered to be significant at the 95% BCI when the values of the 2.5% and 97.5% percentiles do not include zero (0), i.e., they are both either negative or positive. The significant variables include average occupancy, incident severity, percent of lanes closed, incident type, incident clearance duration, incident impact duration, and incident occurrence time.

As shown in Table 3, compared to incidents that occurred during off-peak hours, incidents that occurred during peak hours are observed to have a higher likelihood of resulting in SCs (mean = 0.639, 95% BCI (0.252, 1.074)). This observation implies that incidents occurring during congested time periods are more likely to induce traffic. Similar findings was also found by (Hirunyatiwattana and Mattingly, 2006; Mishra et al., 2016).

Congested traffic is characterized with smaller gaps between vehicles providing drivers with lesser space for maneuvering to avoid a crash. Accordingly, similar to the travel time messages posted on dynamic message signs, safety messages about the risk of SCs can also be posted based on different levels of prevailing traffic congestion, especially during peak hours. This scenario is also supported by the positive parameter of the average occupancy represented by occupancy mean parameter (mean = 0.084, 95% BCI (0.054, 0.121)). Increase in average occupancy represents an increase in traffic density, traffic volatility, and queue formation. The disturbances induced by the primary incidents are easier to propagate in this queuing traffic conditions, leading to a higher risk of SCs. According to the data shown in Table 3,

Table 3
Posterior Estimates of Bayesian cloglog Regression Model.

	Parameter	Factor	Mean	Median	MCSE	SD	BCI (%) 2.5	BCI (%) 97.5
	Intercept		−7.432	−7.324	0.02810	1.205	−10.070	−5.367
Geometric characteristics	Median width (ft)		−0.002	−0.002	0.00002	0.003	−0.009	0.004
Traffic flow characteristics	Occupancy mean		0.084	0.083	0.00033	0.017	0.054	0.121
	EHV SD (veh/hr)		−0.007	−0.007	0.00004	0.005	−0.020	0.001
	Speed SD (mph)		−0.025	−0.025	0.00010	0.015	−0.055	0.003
Primary/ normal incident characteristics	Incident severity	Minor						
		Moderate/severe	1.099	1.076	0.00758	0.552	0.092	2.257
	Percent lane closed (%)	0–25						
		> 25	1.133	1.119	0.00417	0.282	0.615	1.727
	Number of responding agencies	1						
		2–3	−0.067	−0.067	0.00164	0.232	−0.518	0.395
		> 3	−0.117	−0.113	0.00302	0.426	−0.983	0.711
	Towing involved	No						
		Yes	0.180	0.179	0.00187	0.265	−0.328	0.713
	Emergency Medical Services (EMS) involved	No						
		Yes	−0.616	−0.612	0.00259	0.366	−1.353	0.088
	Incident type	Other						
		Vehicle related	0.870	0.818	0.00569	0.805	−0.529	2.613
		Crash	2.461	2.398	0.00570	0.807	1.079	4.241
	Incident clearance duration (min)		0.007	0.007	0.00003	0.002	0.003	0.011
	Incident occurrence time	Off-peak						
		Peak	0.639	0.630	0.00148	0.209	0.252	1.074
	Incident impact duration (min)		0.008	0.008	0.00004	0.001	0.005	0.011

Note: Underlined variables are significant, EHV = Equivalent hourly volume, SD = Standard Deviation, MCSE = Monte Carlo Standard Error, BCI = Bayesian Credible Interval.

the more the segments upstream of the prior incident is occupied, the longer it will take for the traffic flow to come back to normal. In this case, occupancy can also be used as one of the input parameters used to display the real-time information about the risk of SCs on DMSs and on connected vehicles. This situation is further explained by the positive parameter of time taken for speed to come back to normal, represented by the primary incident impact duration, (mean = 0.005, 95% BCI (0.003, 0.007)), which indicates that the risk of having a SC increases with primary incident impact duration.

Similarly, the positive parameter of the incident clearance duration indicates that the risks of SCs increase with an increase in incident clearance time (mean = 0.008, 95% BCI (0.005, 0.011)). As expected, the percentage of lanes closed is also identified as one of the significant predictor variables that influence the risk of SCs occurrence. More specifically, incidents resulting in more than 25% of lane closure have a higher likelihood of resulting in SCs (mean = 1.133, 95% BCI (0.615, 1.727)) compared to incidents involving less than 25% of lane closure. Note that the percentage of lanes closed was used instead of the number of lanes closed since it is a more representative variable. The percent of lanes closed is an indicator of the severity of the primary incident as severe incidents tend to result in an increased number of lanes closed. This fact is proven by the positive value of the primary incident severity coefficient (mean = 1.099, 95% BCI (0.092, 2.257)).

Incident type is also a statistically significant predictor of the likelihood of the occurrence of SCs; incidents that are crashes have a higher likelihood of resulting in SCs (mean = 2.461, 95% BCI (1.079, 1.727)) compared to those involving other incident types such as debris on roadway.

6. Summary and conclusions

Secondary crashes (SCs) are crashes that occur within the spatial and temporal impact range of a primary incident. This study investigated the effect of real-time traffic, incident, environmental, and geometric related variables on the likelihood of SCs.

As a first step toward achieving the study objective, potential SCs were identified using real-time speed data from BlueToad paired devices. This method was able to identify the spatial and temporal impact ranges of primary incidents, while accounting for the effects of traffic flow characteristics, both on the upstream and opposite directions. Random forests technique was next used to screen for the important variables. Highly correlated variables were then identified and excluded from further analysis. Finally, Bayesian random effect complementary log-log (cloglog) model was used to link the probability of SC occurrence with the real-time traffic flow variables, primary incident characteristics, environmental, and geometric characteristics. Note that the cloglog model was employed because of the asymmetrical nature of the SCs data, and the random parameter was incorporated in the model to account for unobserved heterogeneity.

The results indicated that several primary incident characteristics and real-time traffic variables influence the occurrence of SCs. The following seven variables were found to be significant at the 95% Bayesian credible interval (BCI): time taken for the traffic flow speed to come back to normal (incident impact duration), incident clearance duration, incident occurrence time, average occupancy, incident severity, percent of lanes closed, and incident type.

As can be inferred from the study findings, prevention of SCs is a function of primary incident severity, how quickly the primary incident is cleared, and how quickly information about the occurrence and location of traffic incidents is disseminated to the upstream drivers. To prevent the risk of SC occurrence, traffic management strategies should be developed to accelerate the dissipation of queue upstream of the primary incident. The likelihood of a SC occurrence can be estimated using real-time traffic data in combination with primary incident characteristics. Warnings can be sent to drivers approaching a primary crash scene in real-time through various means including dynamic message signs (DMSs), information sharing technologies such as WAZE application, and the emerging technologies such as connected vehicles, giving them an opportunity to take necessary precautions (such as detour and/or drive with caution) to avoid being involved in a crash. Furthermore, when the conditions associated with a high likelihood of SCs prevail, responding agencies such as highway patrol, emergency medical services, towing agencies, etc. could be better prepared to respond to SCs, if they were to occur. These strategies will help to potentially reduce the frequency and severity of SCs.

7. Limitations and future work

This study used speed data from BlueToad devices, incident data from SunGuide database, and real-time traffic data from RITIS database to estimate the likelihood of SCs. Speed data extracted from the BlueToad devices was used to determine the spatiotemporal impact range of primary incidents, and hence, to identify SCs. However, the BlueToad devices, with an average spacing of 1.7 miles, may not be able to precisely capture the speed changes over space. Although incidents were found to usually have longer spatial impacts in the upstream direction, affecting multiple BlueToad pairs, using speeds along shorter sections could potentially capture the realistic traffic dynamics, including queue formation and dissipation.

As BlueToad pairs become more prevalent and as crowdsourced travel speed data become more readily available, future studies can incorporate virtual detectors that use crowdsourced traffic information in the middle of the BlueToad pairs in order to obtain additional traffic speed data. Moreover, with the use of crowdsourced traffic data, the study locations do not have to be limited to the corridors with active BlueToad devices.

Acknowledgments

This research was sponsored by the Florida Department of Transportation (FDOT) and conducted as a cooperative effort by the Florida International University (FIU) and the University of North Florida (UNF).

References

Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accid. Anal. Prev. 43 (1), 461–470.

Ahmed, M., Abdel-Aty, M., Yu, R., 2012a. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. Transp. Res. Rec. J. Transp. Res. Board 2280, 51–59.

Ahmed, M., Abdel-Aty, M., Yu, R., 2012b. Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data. Transp. Res. Rec. J. Transp. Res. Board 2280, 60–67.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Carrick, G., Srinivasan, S., Beljeri, I., 2015. Descriptive analysis and characterization of secondary traffic crashes in Florida. In: Presented at the 94th Annual Meeting of the Transportation Research Board. Washington D.C..

Chang, G., Rochon, S., 2011. Performance Evaluation and Benefit Analysis for CHART. Coordinated Highways Action Response Team, Maryland Department of Transportation.

Chimba, D., Kutela, B., Ogletree, G., Horne, F., 2014. Impact of abandoned and disabled vehicles on freeway incident duration. J. Transp. Eng. 140 (3).

Chou, C., Miller-Hooks, E., 2010. Simulation-based secondary incident filtering method. J. Transp. Eng. 136 (8), 746–754.

Dissanayake, S., Roy, U., 2013. Crash severity analysis of single vehicle run-off-road crashes. J. Transp. Technol. 4, 1–10.

FDOT, Florida Department of Transportation-Transportation Statistics Office, 2015. Roadway Characteristic Inventory (RCI) Database. Tallahassee, Florida. .

Guo, J., Liu, Z., Huang, W., Wei, Y., Cao, J., 2018. Short-term traffic flow prediction using fuzzy information granulation approach under different time intervals. IET Intell. Transp. Syst. 12 (2), 143–150.

Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Accid. Anal. Prev. 81, 14–23.

Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition. Springer series in statistics, pp. 593–594 Chapter 15.

Hirunyatiwattana, W., Mattingly, S.P., 2006. Identifying secondary crash characteristics for the California highway system. In: Presented at the 85th Annual Meeting of the Transportation Research Board, CD-ROM. Washington D.C..

Huang, H., Chin, H.C., Haque, M.M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. Accid. Anal. Prev. 40 (1), 45–54. http://dx.doi.org/10.1016/j.aap.2007.04.002.

Imprialou, M.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.

Karlaftis, M.G., Latoski, S.P., Richards, N.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. http://dx.doi.org/10.1080/10248079908903756.

Khattak, A., Wang, X., Zhang, H., 2009. Are incident durations and secondary incidents interdependent? Transp. Res. Rec. J. Transp. Res. Board 2099, 39–49. http://dx.doi.org/10.3141/2099-05.

Khattak, A.J., Wang, X., Zhang, H., Cetin, M., 2011. Primary and Secondary Incident Management: Predicting Durations in Real-Time Research Report No. VCTIR 11-R11. Old Dominion University and Virginia Center for Transportation Innovation and Research.

Khattak, A., Wang, X., Zhang, H., 2012. Incident management integration tool: dynamically predicting incident durations, secondary incident occurrence and incident delays. IET Intell. Transp. Syst. 6 (2), 204–214.

Kitali, A.E., Kidando, E., Sando, T., Moses, R., Ozguven, E.E., 2017. Evaluating aging pedestrian crash severity with bayesian complementary log–log model for improved prediction accuracy. Transp. Res. Rec. J. Transp. Res. Board 2659, 155–163. http://dx.doi.org/10.3141/2659-17.

Kobelo, D., Patrangenaru, V., Mussa, R., 2008. Safety analysis of Florida urban limited access highways with special focus on the influence of truck lane restriction policy. J. Transp. Eng. 134 (1), 297–306.

Liaw, A., Wiener, M., 2015. Breiman and Cutler's Random Forests for Classification and Regression. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.

Martin, J.-L., Wu, D., 2017. Pedestrian fatality and impact speed squared: cloglog modeling from French national data. Traffic Inj. Prev. 19 (1), 94–101.

Mishra, S., Golias, M.M., Sarker, A.A., Naimi, A., 2016. Effect of Primary and Secondary Crashes: Identification, Visualization, and Prediction Research Report No. CFIRE 09-05.

Moore, J.E., Giuliano, G., Cho, S., 2004. Secondary accident rates on Los Angeles freeways. J. Transp. Eng. 130 (3), 280–285.

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., Trego, T., 2010. Traffic Incident Management Handbook, Report No. FHWA-HOP-10-013. Washington D.C. .

Pande, A., Das, A., Abdel-Aty, M., Hassan, H., 2011. Estimation of real-time crash risk: are all freeways created equal? Transp. Res. Rec. J. Transp. Res. Board 2237, 60–66.

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. J. Transp. Res. Board 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. http://dx.doi.org/10.1016/j.trc.2015.08.011.

Smith, B.L., Ulmer, J.M., 2003. Freeway traffic flow rate measurement: investigation into impact of measurement time interval. J. Transp. Eng. 129 (3), 223–229.

Sun, C.C., Chilukuri, V., 2006. The use of dynamic incident progression curve for classifying secondary accidents. In: Presented at the 85th Annual Meeting of the Transportation Research Board. Washington D.C..

Sun, C.C., Chilukuri, V., 2010. Dynamic incident progression curve for classifying secondary traffic crashes. J. Transp. Eng. 136 (12), 1153–1158.

Theofilatos, A., 2017. Incorporating real-time traffic and weather data to explore road accident likelihood and severity in urban arterials. J. Saf. Res. 61, 9–21.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J., Halkias, B., 2010. Freeway operations, spatiotemporal-incident characteristics, and secondary-crash occurrence. Transp. Res. Rec. J. Transp. Res. Board (2178), 1–9.

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. http://dx.doi.org/10.1080/15472450.2012.688384.

Wang, J., Boya, L., Lanfang, Z., 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.

Wang, J., Liu, B., Fu, T., Liu, S., Stipancic, J., 2018. Modeling when and where a secondary accident occurs. Accid. Anal. Prev. http://dx.doi.org/10.1016/j.aap.2018.01.024. Available online 2 February 2018.

Xu, C., Liu, P., Yang, B., Wang, W., 2016. Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data. Transp. Res. Part C Emerg. Technol. 71, 406–418. http://dx.doi.org/10.1016/j.trc.2016.08.015.

Yang, H., Bartin, B., Ozbay, K., 2014a. Mining the characteristics of secondary crashes on highways. J. Transp. Eng. 140 (4).

Yang, H., Ozbay, K., Xie, K., 2014b. Assessing the risk of secondary crashes on highways. J. Saf. Res. 49, 143–149.

Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Saf. Sci. 63, 50–56.

Yu, R., Yang, K., Abdel-Aty, M., 2016. Crash risk analysis for Shanghai urban expressways: a Bayesian semi-parametric modeling approach. Accid. Anal. Prev. 95, 495–502. http://dx.doi.org/10.1016/j.aap.2015.11.029.

Zhan, C., Shen, L., Hadi, M., Gan, A., 2008. Understanding the characteristics of secondary crashes on freeways. In: Presented at the 87th Annual Meeting of the Transportation Research Board. Washington D.C..

Zhan, C., Gan, A., Hadi, M., 2009. Identifying secondary crashes and their contributing factors. Transp. Res. Rec. J. Transp. Res. Board 2102, 68–75. http://dx.doi.org/10.3141/2102-09.

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.