IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 20, 2018, accepted January 9, 2019, date of publication January 29, 2019, date of current version March 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2895613

Analysis of Freeway Secondary Crashes With a Two-Step Method by Loop Detector Data

BO YANG[1,2], YANYONG GUO[1,2,3], AND CHENGCHENG XU[1,2]
[1]Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing 210096, China
[2]Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China
[3]Department of Civil Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Corresponding author: Yanyong Guo (yanyong.guo@ubc.ca)

This work was supported in part by the Projects of the National Natural Science Foundation of China under Grant 71701046, in part by the Project of the China Postdoctoral Science Foundation under Grant 2017M571644l and Grant 2018T110427, in part by the Fundamental Research Funds for Central Universities under Grant 2242018R20003 and Grant YBJJ1533, and in part by the Scientific Innovation Research of College Graduates in Jiangsu Province under Grant KYLX_0173.

ABSTRACT A two-step identification method of secondary crashes (SCs) is developed in this paper, and the effects of traffic variables on the SC risks on freeways are captured. Crash-related data were obtained from the I-880 freeway in California, USA. Combined with the speed contour map, the SCs were identified by the real shock wave speed index and the SC identification index. The random effect logit regression was applied to analyze the significantly contributing factors and their mechanism on the SCs. The results showed that the significantly contributing factors are different between the initial SC model and SC models (Threshold = 20mile/h, 15mile/h, 10mile/h, 5mile/h), except for the variable of the standard deviation of 30-s vehicle speed during 5–10 min. In the SC models, the number of significant contributing factors increases with the increase in the threshold value. The results of the elasticity analysis showed that the elasticity values of the hit object crash, wet surface road, the average 30-s vehicle speed during 5–10 min, and the average 30-s vehicle count during 5–10 min are greater than 10%.

INDEX TERMS Freeways, traffic flow, secondary crash, identification method, random effect logit regression.

I. INTRODUCTION

Secondary crashes (SC) happen in the impact ranges of prior crashes, which is regarded as a result of primary crashes (PC). Moreover, SC can result in increased traffic flow fluctuation and more crashes. To reduce the adverse congestion and safety impacts associated with SC, increased attentions have been given to develop advanced management and control strategies to prevent SC. Accordingly, numerous studies have been proceeded to understand the mechanisms of SC and their contributing factors [1]–[3].

In most of the previous studies, SC was identified by the static threshold methods (STM) based on some fixed temporal and spatial range. That is to say SC would happen in a fixed spatial and temporal range of a PC. Many researchers have used STM to analyze the mechanism of SC [4]–[7]. However, the limitation of STM is that the determination of temporal and spatial range is too subjective to give a objective and accuracy identification of SC [1], [8]–[10]. Recently, a number of researchers proposed different dynamic methods (DM) to overcome the limitation of STM, such as automatic tracking of moving jams, speed contour, shock wave principles and vehicle probe data [11]–[14]. The traffic flow characteristics (e.g. density, flow, and speed) which fluctuate over space and time were used in DM. These DM can help to better understand the mechanism of traffic flow and the queue formation process.

In addition to SC identification methods, the characteristics of SC have also been explored in some studies. It was found that some contributing factors are related to the SC risk [6], [14]–[17]. Previous studies utilized random effect logit (REL) regression to analyze the relationship between SC risk and contributing factors, such as traffic flow variables, PC characteristics factors, environmental factors, and geometric design factors.

Numerous studies established probit or logit parametric models to explore the mechanism of SC [12], [13], [16]. Zhang and Khattak [9] used the ordered logistic regression to

The associate editor coordinating the review of this manuscript and approving it for publication was Xu Chen.

analyze the relationship between the occurrence probability of SC and the characteristics of the PC. The result indicated that PC duration and the number of involved vehicles significantly affect the SC risks. Yang *et al.* [15] used the logistic regression to analyze the effects of the PC characteristics on the SC risks. The results indicated that occurrence time and collision type of the PC are main contributing factors to SC. Khattak *et al.* used the logistic regression to establish a SC risk prediction model. It was found that the time of day, adverse weather conditions, and detection source of the PC are related with the occurrence probability of SC [17]. The artificial intelligence methods such as Bayesian neural networks model [12], machine learning algorithms [18], were also applied in analyzing SC,

Although many studies have analyzed SC, limited studies have considered what SC really is, and how to identify SC with more reasonable method. In previous studies, no matter what the method of identification of SC is, the primary task is to identify the influence area when the PC occur at the downstream on freeway. However, the influence area has the spatial and temporal restriction. The traffic is related to various variables, such as traffic flow variables, environment variables, and geometric design variables. When a prior crash occurs, the traffic situation is becoming more complex. As such, the vital questions of identifying SC are (i) how to confirm the spatial and temporal restriction of the influence area when a crash occur, (ii) how to judge the impact of prior crash on traffic flow states, and (iii) how to distinguish that the change of traffic state at the location of SC is affected by the prior crash or other factors such as recurrent congestions. In this study, the primary objective was to explore the SC identification method by traffic flow theory, and to analyze the contributing factors of SC. This study can promote the understanding of the mechanism of SC and help the road safety management authorities to reduce SC risks.

II. DATA

The related data used in this study were acquired from the I-880 freeway, in the San Francisco Bay area of California, United States. The length of the selected freeway section is 35-mile, and the length of time period is from 2006 to 2010. 134 loop detectors on two directions were adopted. Traffic data, road design data, and crash data were acquired from the Highway Performance Measurement System (PeMS) maintained by the California Department of Transportation. The related data included traffic flow variables, PC characteristics factors, environmental factors, and geometric design factors. A total of 8,981 crashes were used in this study. There were three types of crashes, including SC, PC, and the normal crashes (NC). PC are defined as the crashes that lead to SC, while NC are defined as the crashes that did not lead to SC. The method of identification of SC and the number of SC, PC, and NC are given in section VI.

Traffic data were collected from the nearest loop detector station to crashes. Specifically, traffic count, speed, and occupancy were collected in 30s for each lane. Moreover, weather data were collected from three weather stations which were located within five miles from the selected freeway segment. In order to match the crash data, weather data was collected 5 to 10 minutes before the occurrence of recorded crash. In Table 1, the 30s raw data of 5-min intervals for each crash were further converted into the 24 traffic flow variables. In addition, 6 crash characteristic variables, 4 environment variables, and 4 geometric characteristics variables were induced in Table 1. A total of 38 candidate variables were considered.

III. METHODOLOGY

In this research, REL regression was used to identify the impact factors of SC on freeway. The impact factors were related to geometric design characteristic, PC characteristic, traffic flow, and environment conditions. In REL regression, the heterogeneity induced by the unobserved factors, such as driver age, driver gender, was explained by a random effect. Neglecting the heterogeneity can induce bias parameter estimates and inconsistent [19], [20]. The REL regression is showed as:

$$y_i \sim \text{Bernoulli}\,(p_i) \tag{1}$$

$$\text{logit}\,(p_i) = \beta_0 + \theta_r + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{ki} x_{ki} \tag{2}$$

where y_i indicates the SC indicator (1 represent a SC occurred caused by a PC, and 0 represent no SC happened) for the i^{th} observation in the sample; p_i indicates the probability of a SC for the ith observation; x_{ki} indicates the value of variable k for sample i; β_k indicates the coefficient of variable k; θ_r is a random term which catches the random effects for freeway segment r. In general, the random effect term θ_r is hypothesized to be normally distributed with mean u_θ and variance Σ_θ [21], [22].

Based on the above specifications, the likelihood function can be written as:

$$f(Y|\boldsymbol{\Theta}) = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N}\left[\left(\frac{e^{\beta_0+\theta_r+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_{ki}x_{ki}}}{1+e^{\beta_0+\theta_r+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_{ki}x_{ki}}}\right)^{y_i} \times \left(1-\frac{e^{\beta_0+\theta_r+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_{ki}x_{ki}}}{1+e^{\beta_0+\theta_r+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_{ki}x_{ki}}}\right)^{(1-y_i)}\right] \tag{3}$$

where Θ indicates the vector of the parameters to be estimated, including the regression parameters β, the random effect θ for different freeway segment, the mean of random effect u_θ, and the variance of random effect Σ_θ. Accordingly, $\Theta = [\beta, \theta, u_\theta, \Sigma_\theta]$.

To estimate the effect of the impact factors on SC probability, the elasticity analysis was utilized. Specifically, the independent variable can be continuous variable or indicator variable. As such, two equations were given to calculated the elasticity. The elasticity of a continuous independent variable x_i is showed as:

$$E_i = \frac{\partial Y_i}{\partial x_i} \times \frac{x_i}{Y_i} = [1 - P(i)]\beta_i x_i \tag{4}$$

TABLE 1. Variables.

Variable Category	Symbol	Variables
Real-time Traffic Conditions	C_{nt15}/ S_{pd15}/ O_{cc15}	Average 30 second vehicle count/speed/detector occupancy during 5 to 20 minutes (veh/30s)/(mile/h)/(%)
	S_{tdc15}/ S_{tds15}/ S_{tdo15}	Standard deviation of 30 second vehicle count/speed/detector occupancy during 5 to 20 minutes (veh/30s)/(mile/h)/(%)
	C_{c15}/ C_{s15}/ C_{o15}	Coefficient of variation of 30 second vehicle count/speed/detector occupancy during 5 to 20 minutes (veh/30s)/(mile/h)/(%)
	C_{nt5}/ S_{pd5}/ O_{cc5}	Average 30 second vehicle count/speed/detector occupancy during 5 to 10 minutes (veh/30s)/(mile/h)/(%)
	S_{tdc5}/ S_{tds5}/ S_{tdo5}	Standard deviation of 30 second vehicle count/speed/detector occupancy during 5 to 10 minutes (veh/30s)/(mile/h)/(%)
	C_{c5}/ C_{s5}/ C_{o5}	Coefficient of variation of 30 second vehicle count/speed/detector occupancy during 5 to 10 minutes (veh/30s)/(mile/h)/(%)
	D_{c5}/ D_{s5}/ D_{o5}	The difference in average 30 second vehicle count/speed/detector occupancy between 15 to 20 min and 5 to 20 min(veh/30s)/(mile/h)/(%)
	L_{dc}/ L_{ds}/ L_{do}	Average difference in 30 second vehicle count/speed/detector occupancy between adjacent lanes during 5 to 20 minutes (veh/30s)/(mile/h)/(%)
PC Characteristics	S_e	1 denotes non-PDO; 0 denotes PDO (property damage only)
	S_w	1 denotes sideswipe crash; 0 denotes non-sideswipe crash
	R_e	1 denotes rear end crash; 0 denotes non- rear end crash
	H_o	1 denotes hit object crash; 0 denotes non-hit object crash
	P_e	1 denotes peak period; 0 denotes otherwise
	D_w	1 denotes weekend; 0 denotes weekday
Environmental Conditions	W_e	1 denotes adverse weather conditions; 0 denotes clear
	R_s	1 denotes wet surface road; 0 denotes otherwise
	L_i	1 denotes no street lights; 0 denotes otherwise
	V_i	1 denotes otherwise; 0 denotes clear
Geometric Characteristics	N_l	Lanes count
	W_{rs}	Road width (ft)
	W_{im}	Inner median width (ft)
	C_S	1 denotes curve road section; 0 denotes otherwise

In this study, the average elasticity of a continuous variable was used to interpret its quantitative effect on the outcome.

As for the indicator variable, the pseudo-elasticity an indicator variable x_i is computing by the following equation [23]:

$$E_i = [\frac{EXP[\Delta(x'\beta)][1 + EXP(x_i\beta_i)]}{EXP[\Delta(x'\beta)][EXP(x_i\beta_i)] + 1} - 1] \times 100 \tag{5}$$

IV. A TWO-STEP IDENTIFICATION METHOD OF SC

A. THE INITIAL IDENTIFICATION OF SC

Traditionally, STM and DM are used to identify a SC. However, STM have the problem of subjective judgment of fixed spatial and temporal thresholds. As an alternative of STM, the DM can determine dynamic thresholds by using queue length estimations or traffic flow simulation. Nevertheless, numerous DM still have drawbacks. For instance, the DM which use queue length estimations need detailed queuing information. However, this information is not always available [10], [6], [15]–[17]. Another DM which use incident progression curve utilize only one identical curve for all SC, which may lead to unreliable results [10].

In order to overcome the drawbacks accompanied with the STM and DM, this study applied a method based on the speed contour figure to initially identify SC. This method utilizes real-time traffic flow data to determine the spatial and temporal influencing range of a prior crash and simultaneously takes the effects of recurrent congestions into account. Subsequently, an initial SC is identified if it is within the spatial and temporal influencing range of the corresponding prior crash. The proposed method is detailed as follows.

1) STEP ONE

The 5-min speed data were extracted to produce a speed contour figure for a prior crash. Specifically, the speed data were extracted from the loop detectors within 10 miles upstream and 10 miles downstream the prior crash during the time interval between 6 hours before and 6 hours after the prior. Figure 1(a) shows an example of a speed contour figure. It can be clearly seen from the figure that congestions and queue formations occur after the prior crash. However, less information is offered by the figure about whether the queue formations were resulted from recurrent congestions or the prior crash. To eliminate the effects of recurrent congestions, the spatial and temporal influencing range of the prior crash should be determined, which is given by the following two steps.

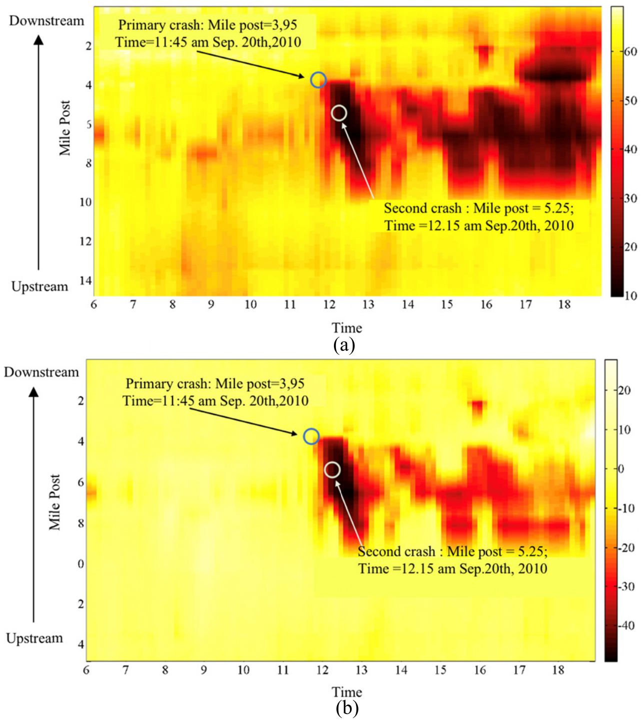


FIGURE 1. **Initial identification of SC.**

2) STEP TWO

The 5-min speed data for the same time and same location in step one, however, from crash-free days, were extracted for the whole year in this step. For instance, the prior crash in Figure 1(a) happened at 11:45 am on September 20, 2010 at milepost 3.95. Following this step, the speed data for the same time and location in Figure 1(a) were collected from all crash-free days in 2010. Subsequently, the speed data for each time and location was averaged over all the crash-free days.

3) STEP THREE

To eliminate the potential effects of recurrent congestions, the average speed in step two was subtracted from the speed data for each time and location in step one. A new speed contour figure was developed using the differences between speeds in step two and step one for various times and locations. The new speed contour figure as shown in Figure 1(b) was then used to determine the spatial and temporal influencing range of the prior crash.

4) STEP FOUR

The crashes that happened within the spatial and temporal influencing ranges of PC were identified as initial SC. The crashes that did not lead to initial SC were identified as NC.

Following the above four steps of initial identification method, the numbers of SC, PC, and NC in the dataset are 97, 97, and 8787 respectively.

B. IDENTIFICATION OF SC BY SHOCKWAVE

Subsequently, the method based on shockwave is used for accurate identification of SC. Obviously, the SC is the crash that is caused by PC. In early studies of STM and DM, the aim is to identify PC influence area. In recent studies, the shockwave theory was used to identify PC influence area based on traffic flow theory. However, all the above mentioned methods have ignored the fact that PC affect the transition of traffic flows, and the transition of traffic flows affect the occurrence of SC. Namely, PC indirectly affects the SC. If the traffic flow state is stable or the transition of traffic flows is not affected by the prior crash, the crash in PC influence area, which identified by the initial methods mentioned above, can no longer be identified as a SC. It is because that these phenomena also occur in the situation of non-SC.

The location of crashes can be seemed as a bottleneck on freeway. When a crash occurs, a shock wave maybe developing from the crash location to the upstream. If $V_{real} < 0$ in Eq. (6), the shock wave is developing. If $V_{real} \geq 0$, it indicates that the upstream traffic flow is not affected. Therefore, when $V_{\mathrm{real}} \geq 0$, although the upstream crashes are in speed contour map, they cannot be identified as SC. Only in this case that a crash locates in the upstream of PC and the speed contour map, meanwhile, the crash is affected by the shock wave, and the SC can be identified. Although the shock wave will lead to traffic flow fluctuation, it still has spatial and temporal restriction. Therefore, there is in need of a $\varepsilon_{threshold}$ in Eq. (7). In this study, the value of $\varepsilon_{threshold}$ is 20 mile/h, 15 mile/h, 10 mile/h and 5 mile/h were adopted. In Eq. (8), if $V_{real-threshold} \leq V_{secondary-primary} \leq V_{real+threshold}$, a SC is identified. As shown in Figure 2, crash B occur in the speed contour map and the upstream from the crash A, the value of $V_{secondary-primary}$ is between $V_{real-threshold}$ and $V_{real+threshold}$. It indicates that the crash B is affected by the shock wave which is generated by crash A. On the other hand, although crash C and crash D are in the speed contour map and the upstream from the crash A, they are not affected by the shock wave. Crash C occurs after the shock wave and crash D occurs before the shock wave. There are no direct and real-time relationship between the crash A (PC) and traffic flow states at the location of crash C and D.

The real shock wave speed index (RSWSI) can be calculated by:

$$V_{real} = \frac{q_i - q_j}{k_i - k_j} = \frac{120 \times Q_i - 120 \times Q_j}{\frac{120 \times Q_i}{V_i} - \frac{120 \times Q_j}{V_j}} \tag{6}$$

where V_{real} is the real shock wave speed index (RSWSI) (mile/h); q_i(vehicle/h), q_j(vehicle/h), Q_i(vehicle/30s) and Q_j(vehicle/30s) are the traffic volume of crash i and j; k_i(vehicle/mile) and k_j(vehicle/mile) are the density of crash i and j; V_i(mile/h) and V_j(mile/h) are the density of crash i and j. Crash j occurs at the upstream of crash i.

$$V_{real \pm threshold} = V_{real} \pm \varepsilon_{threshold} \tag{7}$$

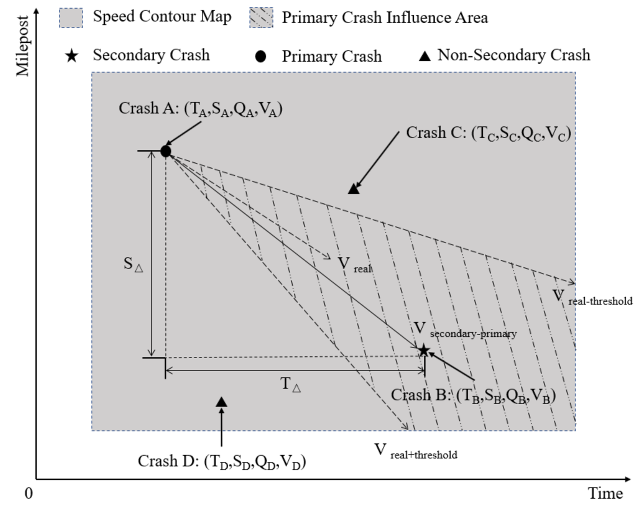


FIGURE 2. **Identification of SC by shockwave.**

where $V_{real\pm threshold}$ is spatial and temporal restriction of shock wave (mile/h); $\varepsilon_{threshold}$ is the threshold value (mile/h).

$$V_{secondary-primary} = \frac{S}{T} = \frac{S_i - S_j}{T_i - T_j} \tag{8}$$

where $V_{secondary-primary}$ is SC identification index (SCII) (mile/h); S_i (mile), S_j(mile), T_i (h) and T_j (h) are the location and time of crash A and crash B; S (mile) and T (h) are the difference of the location and time between crash A and crash B.

The number of SC, PC, and NC in section VI-A are further identified by the method proposed in this section. The final number of SC, PC, and NC are given in Table 2. It reveals that the number of SC and PC are smaller with the threshold decreasing.

TABLE 2. **The distribution of SC crashes in different identification method.**

Identification method	PC	SC	NC	Total
Initial SC	97	97	8787	8981
SC (Treshold=20mile/h)	44	44	8787	8875
SC (Treshold=15mile/h)	36	36	8787	8859
SC (Treshold=10mile/h)	27	27	8787	8841
SC (Treshold=5mile/h)	14	14	8787	8815

V. RESULTS OF REL REGRESSIONS

REL regressions were applied to identify how different types of variables affected the SC probability. The events are PC that induce SC and the non-events are NC that did not induce SC. In REL regression, 1 is PC, 0 is NC. *P*-value of 0.1 was employed for parameter estimate significance in these models. The REL regression were estimated using the software package STATA. The results of REL regressions are presented in Table 3.

In Table 3, the results (initial SC) showed that the average 30s vehicle occupancy during 5-10 minutes (O_{cc5}) is found to be positively related to the SC risks, implying that the crashes occurred in high-occupancy traffic have high probability to lead to SC. The positive parameter of the standard deviation of 30s vehicle speed during 5-10 minutes (S_{tdc5}) and the average difference in 30s vehicle count between 15-20 minutes and 5-20 minutes (L_{dc}), and the negative parameter of variation of 30s vehicle count during 5-10 minutes (C_{c5}) show that the SC risks increase as S_{tdc5}, L_{dc} increases and C_{c5} decreases. The increasing S_{tdc5}, L_{dc} and the decreasing C_{c5} represent an increase disturbances and instability in traffic state leading to increased SC risks. The turbulent traffic conditions can lead to higher SC probability. The negative coefficient with lanes counts (N_l) suggests that the SC risks increase with an decrease in N_l. It implied that lane changing with less lanes on multilane freeway can increase the SC risks.

Further, elasticity analysis was applied to quantify the impacts of significant variables on the SC probability. As shown in Table 3, the average elasticity for the five traffic variables (O_{cc5}, S_{tdc5}, C_{c5}, L_{dc} and N_l) are 0.786, 2.263, −3.799, 1.267 and 1.369, respectively, indicating that a 1% increase in these five traffic variables leads to 0.786%, 2.263%, −3.799%, 1.267%, and 1.369% increases in the SC probability, respectively. Moreover, the absolute value of coefficient of variation of vehicle count is the largest among the five, indicating that the impact of variation of vehicle count on SC risks is greater than other traffic variables.

The results (threshold=20mile/h) also showed that non-PDO (S_e) and sideswipe (S_w) crashes have negative effects on SC risks. The average 30s vehicle count during 5-10 minutes (C_{nt5}) and the standard deviation of 30s vehicle count during 5-10 minutes (S_{tdc5}) is positively related to the SC risks, and the average 30s vehicle speed during 5-10 minutes (S_{pd5}) is negatively correlated with the SC risks, suggesting that the crashes occurred in high-flow and low-speed traffic are more likely to lead to a SC. In addition, the results also implied that a crash happened on weekday (D_w), wet surface road (R_s) and low visibility (V_i) can increase the SC risks.

A SC probability increases 3.011%, −1.325% and 0.189% for a 1% increases in C_{nt5}, S_{pd5} and S_{tdc5}. In addition, the probability of a SC involving a non-PDO crash, a sideswipe crash, weekday, wet surface road and low visibility increases by −24.519%, −34.323%, 4.775%, 35.877% and 1.459%, respectively.

The results of the model (threshold=15mile/h) is similar to the results of the model (threshold=20mile/h). The results (threshold=15mile/h) also showed that sideswipe (S_w) crashes have negative effects on SC risks. The average 30s vehicle count during 5-10 minutes (C_{nt5}) and the standard deviation of 30s vehicle count during 5-10 minutes (S_{tdc5}) is positively associated with the SC risks, and the average 30s vehicle speed during 5-10 minutes (S_{pd5}) is negatively associated with the SC risks, indicating that the crashes happened in high-flow and low-speed traffic have high probability of leading to a SC. In addition, the results also implied that the crash happened on wet surface road can increase the SC risks.

A SC probability increases 2.932%, 12.354% and 0.019% for a 1% increases in C_{nt5}, S_{pd5} and S_{tdc5}. In addition, the

TABLE 3. Results of REL regressions.

Identification method	Variables	Mean	S.D.	2.50%	97.50%	Elasticity
Initial SC	O_{cc5}	4.443	1.026	2.433	6.453	0.786(1.846)
	S_{tdc5}	0.425	0.215	0.003	0.848	2.263(4.442)
	C_{c5}	-7.324	1.631	-10.52	-4.128	-3.799(7.062)
	L_{dc}	0.178	0.105	-0.028	0.385	1.267(2.804)
	N_l	-0.268	0.153	-0.568	0.031	-1.369(2.214)
	Constant	-3.901	0.708	-5.289	-2.513	
			Log likelihood=-497.7045			
			Prob>chi2=0.0000			
			Wald chi2(5)=48.23			
SC (Threshold=20mile/h)	S_e	-0.477	0.259	-0.985	0.032	-24.519(1.789)
	S_w	-0.683	0.324	-1.319	-0.047	-34.323(2.702)
	C_{nt5}	0.225	0.043	0.14	0.309	3.011(4.236)
	S_{pd5}	-0.023	0.005	-0.034	-0.013	-1.325(1.316)
	S_{tdc5}	2.274	1.197	-0.072	4.62	0.189(0.501)
	V_i	0.139	0.084	-0.026	0.303	1.459(1.822)
	D_w	-1.342	0.593	-2.505	-0.179	4.775(0.604)
	R_s	0.8	0.285	0.242	1.358	35.877(5.684)
	Constant	-6.915	1.017	-8.908	-4.922	
			Log likelihood=-480.84795			
			Prob>chi2=0.0000			
			Wald chi2(8)=84.11			
SC (Threshold=15mile/h)	S_w	-0.532	0.325	1.17	0.106	-27.003(1.931)
	C_{nt5}	0.239	0.044	0.153	0.325	2.932(3.948)
	S_{pd5}	-0.026	0.006	-0.037	-0.015	12.354(10.970)
	S_{tdc5}	2.583	1.198	0.235	4.931	0.019(0.049)
	R_s	0.74	0.278	0.194	1.285	32.106(6.212)
	Constant	-5.944	0.667	-7.25	-4.637	
			Log likelihood=-454.45436			
			Prob>chi2=0.0000			
			Wald chi2(5)=74.97			
SC (Threshold=10mile/h)	C_{nt5}	1.099	0.479	0.16	2.038	10.136(4.835)
	H_o	1.222	0.499	0.243	2.201	49.155(12.671)
	Constant	-7.214	0.549	-8.29	-6.138	
			Log likelihood=-178.10451			
			Prob>chi2=0.0064			
			Wald chi2(2)=10.10			
SC (Threshold=5mile/h)	H_o	1.999	1.051	-0.062	4.059	66.378(23.138)
	R_s	0.913	0.542	-0.149	1.976	40.068(7.197)
	Constant	-8.345	1.007	-10.318	-6.371	
			Log likelihood=-98.463765			
			Prob>chi2=0.0218			
			Wald chi2(2)=7.65			

probability of a SC involving a sideswipe crash, and occurring on wet surface road increases by −27.003% and 32.106%, respectively.

The results of the model (threshold=10mile/h) and the results of the model (threshold=5mile/h) showed that the hit object (H_o) crashes have positive effects on the SC risks. The average 30s vehicle count during 5-10 minutes (C_{nt5}) in the model (threshold=10mile/h) and the wet surface road (R_s) in the model (threshold=5mile/h) have positive effects on SC risks.

In elasticity analysis, the results of the model (threshold=10mile/h) showed that the SC probability enhances 10.136% for a 1% increases in C_{nt5}. The probability of a SC involving a hit object crash increases by 49.155%. The results of the model (threshold=10mile/h) showed that the crash probability of a SC involving a hit object crash and occurring on wet surface road increases by 66.378% and 40.068%, respectively.

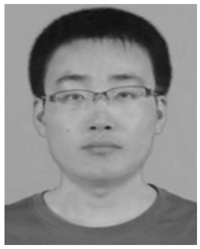
VI. DISCUSSION AND CONCLUSIONS

In this study, a two-step identification method of SC was proposed, and the REL regressions were used to capture the effects of traffic variables, such as traffic flow variables, geometric design variables, environment variables, on the SC risks on freeways.

More specifically, compared to the other identification methods of SC, only in this case that a crash locates in the upstream of PC and the speed contour map, meanwhile, the crash is affected by the shock wave, then the SC can be identified. The real shock wave speed index (RSWSI) and SC identification index (SCII) are used to identify the SC in this study. In the initial SC model, the average 30s vehicle occupancy during 5-10 minutes (O_{cc5}), the standard deviation of 30s vehicle speed during 5-10 minutes (S_{tdc5}), and the average difference in 30s vehicle count between 15-20 minutes and 5-20 minutes (L_d) have been proved that can enhances the SC risks. The coefficient of variation of 30s vehicle count during 5-10 minutes (C_{c5}) and lanes counts (N_l) can decrease the SC risks. In the SC models (Threshold=20mile/h, 15mile/h, 10mile/h, 5mile/h), the average 30s vehicle count during 5-10 minutes (C_{nt5}), the standard deviation of 30s vehicle speed during 5-10 minutes (S_{tdc5}), low visibility (V_i), wet surface road (R_s), and the hit object crashes(H_o) can increase the SC risks. Non-PDO crashes (S_e), sideswipe crashes (S_w), the average 30s vehicle speed during 5-10 minutes (S_{pd5}), weekday (D_w) can decrease the SC risks. The REL regression estimate results revealed that the significantly contributing factors are different between the initial SC model and the SC models (Threshold=20mile/h, 15mile/h, 10mile/h, 5mile/h), except for the standard deviation of 30s vehicle speed during 5-10 minutes (S_{tdc5}). In the SC models, the number of significant contributing factors increases with the threshold value increase. In the elasticity analysis, the elasticity values of the hit object crash(H_o), wet surface road (R_s), the average 30s vehicle speed during 5-10 minutes (S_{pd5}) and the average 30s vehicle count during 5-10 minutes (C_{nt5}) are greater than 10%. The results in this research can help road safety management authorities prevent the SC on freeway, reduce the SC risks, as well as better understand the significantly contributing factors and their mechanism on SC.

REFERENCES

[1] C. C. Sun and V. Chilukuri, "Dynamic incident progression curve for classifying secondary traffic crashes," *J. Transp. Eng.*, vol. 136, no. 12, pp. 1153–1158, 2010.
[2] C. Zhan, A. Gan, and M. Hadi, "Identifying secondary crashes and their contributing factors," *Transp. Res. Rec.*, vol. 2102, no. 1, pp. 68–75, 2009.
[3] R. Raub, "Occurrence of secondary crashes on urban arterial roadways," *Transp. Res. Rec.*, vol. 1581, pp. 53–58, Jan. 1997.
[4] M. G. Karlaftis, S. P. Latoski, N. J. Richards, and K. Sinha, "Empirical analysis of secondary crash causes," presented at the 77th Annu. Meeting Transp. Res. Board, CD-ROM, Washington, DC, USA, 1998.
[5] J. E. Moore, II, G. Giuliano, and S. Cho, "Secondary accident rates on Los Angeles freeways," *J. Transp. Eng.*, vol. 130, no. 3, pp. 280–285, 2004.
[6] W. Hirunyanitiwattana and S. P. Mattingly, "Identifying secondary crash characteristics for California highway system," presented at the 85th Annu. Meeting Transp. Res. Board, CD-ROM, Washington, DC, USA, 2006.
[7] C. Zhan, L. Shen, M. A. Hadi, and A. Gan, "Understanding the characteristics of secondary crashes on freeways," presented at the 87th Annu. Meeting Transp. Res. Board, CD-ROM, Washington, DC, USA, 2008.
[8] V. Chilukuri and C. Sun, "Use of dynamic incident progression curve for classifying secondary accidents," presented at the 85th Annu. Meeting Transp. Res. Board, CD-ROM, Washington, DC, USA, 2006.
[9] H. Zhang and A. Khattak, "What is the role of multiple secondary incidents in traffic operations?" *J. Transp. Eng.*, vol. 136, no. 11, pp. 986–997, 2010.
[10] C.-S. Chou and E. Miller-Hooks, "Simulation-based secondary incident filtering method," *J. Transp. Eng.*, vol. 136, no. 8, pp. 746–754, 2010.
[11] H. Zhang, M. Cetin, and A. J. Khattak, "Joint analysis of queuing delays associated with secondary incidents," *J. Intell. Syst.*, vol. 19, no. 2, pp. 192–204, 2015.
[12] H. Park and A. Haghani, "Real-time prediction of secondary incident occurrences using vehicle probe data," *Transp. Res. C, Emerg. Technol.*, vol. 70, pp. 69–85, Sep. 2016.
[13] W. Junhua, L. Boya, Z. Lanfang, and D. R. Ragland, "Modeling secondary accidents identified by traffic shock waves," *Accident Anal. Prevention*, vol. 87, pp. 141–147, Feb. 2016.
[14] C. Xu, P. Liu, B. Yang, and W. Wang, "Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data," *Transp. Res. C, Emerg. Technol.*, vol. 71, pp. 406–418, Oct. 2016.
[15] H. Yang, B. Bartin, and K. Ozbay, "Mining the characteristics of secondary crashes on highways," *J. Transp. Eng.*, vol. 140, no. 4, pp. 1–10, 2014.
[16] H. Yang, K. Ozbay, and K. Xie, "Assessing the risk of secondary crashes on highways," *J. Saf. Res.*, vol. 49, pp. 143–149, Jun. 2014.
[17] M. G. Karlaftis, S. P. Latoski, N. J. Richards, and K. C. Sinha, "Its impacts on safety and traffic management: An investigation of secondary crash causes," *J. Intell. Transp. Syst.*, vol. 5, no. 1, pp. 39–52, 1999.
[18] J. Wang, B. Liu, T. Fu, S. Liu, and J. Stipancic, "Modeling when and where a secondary accident occurs," *Accident Anal. Prevention*, 2019. doi: 10.1016/j.aap.2018.01.024.
[19] Y. Guo, Z. Li, Y. Wu, and C. Xu, "Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities," *Accident Anal. Prevention*, vol. 115, pp. 118–127, Jun. 2018.
[20] Y. Guo, Z. Li, Y. Wu, and C. Xu, "Evaluating factors affecting electric bike users' registration of license plate in China using Bayesian approach," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 59, pp. 212–221, Nov. 2018.
[21] Y. Guo, P. Liu, Y. Wu, and J. Chen, "Evaluating how right-turn treatments affect right-turn-on-red conflicts at signalized intersections," *J. Transp. Saf. Secur.*, pp. 1–22, Sep. 2018. 10.1080/19439962.2018.1490368.
[22] Y. Guo, T. Sayed, and M. H. Zaki, "Exploring evasive action–based indicators for PTW conflicts in shared traffic facility environments," *J. Transp. Eng. A, Syst.*, vol. 144, no. 11, 2018, Art. no. 04018065.
[23] S. P. Washington, M. G. Karlaftis, and F. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*. Boca Raton, FL, USA: CRC Press, 2003.



BO YANG received the B.S. degree in transportation engineering from the Hefei University of Technology, in 2013. He is currently pursuing the Ph.D. degree with the School of Transportation, Southeast University. His research interests include traffic safety on freeway and crash risk predict.

YANYONG GUO received the M.S. degree in transportation engineering from Chang'an University, Xi'an, China, in 2012, and the Ph.D. degree in transportation engineering from Southeast University, Nanjing, China, in 2016. From 2014 to 2015, he was a Visiting Ph.D. Student with The University of British Columbia, where he is currently a Postdoctoral Fellow. His research interests include road safety evaluations, traffic conflicts techniques, advanced statistical techniques for safety evaluation, and pedestrian and cyclist behaviors. He received the China National Scholarship, in 2014, and the Best Doctoral Dissertation Award by the China Intelligent Transportation Systems Association, in 2017.

CHENGCHENG XU received the Ph.D. degree from Southeast University, in 2014, where he has been an Associate Professor, since 2014. His current research interests include traffic safety and crash risk predict.