

Research Article

Analysis of Freeway Secondary Crashes in Different Traffic Flow States by Three-Phase Traffic Theory

Bo Yang ¹, Yao Wu ^{1,2} and Weihua Zhang ³

¹Jiangsu Key Laboratory of Urban ITS, Southeast University,
Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou #2, Nanjing 210096, China

²Department of Civil Engineering, The University of British Columbia, 6250 Applied Science Lane, Vancouver,
BC V6T 1Z4, Canada

³School of Automobile and Traffic Engineering, Hefei University of Technology, Hefei 230009, China

Correspondence should be addressed to Yao Wu; wuyaomaster@126.com

Received 27 July 2020; Revised 19 August 2020; Accepted 1 September 2020; Published 27 September 2020

Academic Editor: Neng-Chao Lyu

Copyright © 2020 Bo Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The objective of this study is to analyse the relationship between secondary crash risk and traffic flow states and explore the contributing factors of secondary crashes in different traffic flow states. Crash data and traffic data were collected on the I-880 freeway in California from 2006 to 2011. The traffic flow states are categorised by three-phase traffic theory. The Bayesian conditional logit model has been established to analyse the statistical relationship between the secondary crash probability and various traffic flow states. The results showed that free flow (F) state has the best safety performance of secondary crash and synchronized flow (S) state has the worst safety performance of secondary crashes. The traditional logistic regression model has been used to analyse the contributing factors of secondary crashes in different traffic flow states. The results indicated that the contributing factors in different traffic flow states are significantly different.

1. Introduction

Exploring the crash mechanism and contributing factors plays an important role in preventing crash and reducing crash severity to freeway traffic surveillance systems. The occurrence of a crash can generate the turbulence of traffic flow which may lead to further crashes. Secondary crash (SC) occurs within the spatial and temporal impact ranges of the turbulent traffic conditions caused by the primary crash (PC). Previous studies suggested that 2.2% to 3.9% of freeway crashes can result in SC [1–3].

With the widespread use of freeway real-time traffic surveillance systems, researchers have started using high-resolution dynamic traffic flow data to identify the traffic condition before SC occurrences. In general, SC can be affected by various contributing factors, including traffic flow characteristics, geometric design factors, weather conditions, PC characteristics, etc. [4–7]. In addition, many researchers have paid close attention to the identification

method of SC. The common methods include static threshold method (STM) [8–10] and dynamic methods (DM) [11–14].

Although many studies have studied the identification method and crash mechanism of SC, few studies focused on the difference of SC in various traffic conditions. In different traffic conditions, there is a significant difference of traffic flow characteristics that affect the spatial-temporal evolution [15–18] and safety performance [19,20]. The typical divided methods of traffic flow include three-phase traffic theory [15,16], four-phase traffic theory [17], six levels of services [18], etc. It has been proved that the safety performance of SC associated with various traffic conditions has a significant difference [3,21].

In this study, the traffic flow is divided by three-phase traffic theory. The main purpose is to analyse the difference of safety performance for SC in different traffic flow states and explore how contributing factors affect the probability of SC in different traffic flow states. The SC related data were

collected from the I-880 freeway in the United States from 2006 to 2011. The Bayesian conditional logit models have been established to analyse the statistical relationship between the SC probability and traffic flow states. The traditional logistic regression models were established to quantify the effects of various variables on the SC probability in different traffic flow states. This research can help traffic management personnel better understand which traffic flow state is more dangerous for the occurrence of SC and realize the contributing factors of SC in different traffic flow states. The results can be applied to develop effective countermeasures and reduce the SC probability in different traffic flow states.

2. Literature Review

In early studies, STM was usually applied to identify SC. The STM is defined by the fixed spatial and temporal influence areas of traffic flow caused by a prior crash. Numerous studies have utilized the STM to analyse SC, such as Raub [22], Karlaftis et al. [23], Moore et al. [8], Zhan et al. [9], Hirunyanitwattana and Mattingly [10], etc. However, there is an obvious limitation for STM. The determination of the scope in STM is too subjective to have objective and reasonable identification method [24]. In subsequent studies, to overcome the limitation of STM, many researchers adopt DM to identify SC [25–28]. DM has a dynamic boundary of influence area based on speed contour plot [14], shock wave [27], etc.

In recent studies, many researchers have analysed and predicted SC with statistical method or intelligent learning approaches. For example, Wang and Jiang proposed an identification method of SC by the speed contour plot and the spatiotemporal evolution of shockwaves [4]. The results indicated that the identification method based on an integer programming model can reduce the misidentification probability of SC. Kitali et al. used random forest to extract the important variables [5]. The results of Bayesian random effect complementary log-log model showed that some traffic flow variables, the PC types, and severities can significantly affect the probability of SC. In subsequent studies, Kitali et al. utilized the penalized logistic regression model to improve the predictive accuracy of the SC risk model [6]. The results of model indicated that the traffic flow variables and the PC characteristics can significantly affect the probability of SC. Specifically, the traffic flow variables include the occupancy, speed, variation of hourly flow, etc. The PC characteristics include the impact duration, types, occurrence time, etc. Yang et al. confirmed the PC boundary with the clustering method and metaheuristic optimization algorithm [7]. Then, a novel identification method is introduced to identify SC. The results showed that the accuracy of identifying SC can rise to 95% with the market penetration rate increasing from 5% to 25%. Subsequently, Yang et al. summarized and discussed the previous studies from three perspectives, including the identification method of SC, the predictive models of SC risk, and the prevention measures of SC [11]. Goodall predicted the probability of SC by empirical queuing and estimated volumes [12]. It was

found that SC occurred on average once every 10 crashes and 54 disabled vehicles. In the author's previous study, a two-step identification method of SC combined with the speed contour map and the shock wave was applied, and the random effect logit regression was utilized to analyse the contributing factors of SC [13]. The results indicated that the number of significant contributing factors increases with increasing of the threshold value. In addition, the collision type, road surface, speed, and traffic flow can significantly affect the probability of SC.

However, few researchers have analysed SC combined with the traffic flow states. Park et al. applied stochastic gradient boosting and rule extraction techniques to improve the accurate and comprehensible predictions of SC [21]. The results indicated that the unexpected traffic congestion caused by a crash has a significant effect on the occurrence of SC. Xu et al. used the zero-inflated ordered probit regression model to explore the relationship between the SC risk and traffic related variables, including traffic flow variables, geometric design factors, weather conditional factors, and PC characteristics [3]. The results showed that there is a significant difference of contributing factors between the SC-prone state and the SC-free state.

Although some researchers have considered the traffic flow states into the studies of SC, no researchers have studied the difference of SC mechanism in different traffic flow states divided by classical macroscopic traffic flow theory. The common methods of classical macroscopic traffic flow theory have been widely used in numerous areas of transportation engineering, including six levels of service [18], four-phase traffic theory [17], and three-phase traffic theory [15,16]. In these common theories, the three-phase traffic theory is one of the accepted approaches for modelling freeway traffic flow [15,16]. According to the three-phase traffic theory, freeway traffic flow can be classified into three phases, including free-flow phase (F), synchronized flow phase, and wide moving jams [15,16].

To make up for the shortcomings of previous studies without considering the classical macroscopic traffic flow theory in the analysis of SC, the three-phase traffic theory is applied in this study to explore the difference of SC mechanism in various traffic flow states.

3. Materials and Methods

3.1. Data Sources. Data were obtained from a 34-mile section on the I-880 freeway in the California, United States, between 2006 and 2011. There are 119 loop detector stations along the selected freeway section with an average spacing of 0.5 miles. A total of 3 weather stations are located along the selected freeway section. The weather data were obtained from the National Climate Data Center (NCDC) website which provides hourly weather information. The geometric and traffic data were collected from the nearest loop detector stations to each collision location and obtained from the Highway Performance Measurement System (PeMS) maintained by the California Department of Transportation (Caltrans). Crash data were obtained from the Statewide Integrated Traffic Records System (SWITRS) of the Caltrans.

Specifically, vehicle count, speed, and occupancy were collected in 30 s for each lane. The related crash data included traffic flow variables, environmental factors, geometric design factors, and others. A total of 9,919 crashes were used in this study. There were three types of crashes, including SC, PC, and NC. PC are defined as the crashes that led 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 2. For each crash, to compensate for the possible inaccuracies in the reported collision occurrence time and identify hazardous traffic conditions ahead of the crash occurrence time, traffic data were collected for the 5–10 min prior to crash occurrence [13,19]. Non-crash cases were extracted based on crash locations and times, and the ratio between crash and non-crash is 1:4 [19].

In Table 1, the 30 s raw data of 5-minute intervals for each crash were further converted into the 19 traffic flow variables, in addition to 4 environment variables, 4 geometric characteristics variables, and 5 crash characteristics variables. A total of 32 candidate variables were considered.

3.2. The Identification of Traffic Flow States. Previous studies have suggested that the traffic states defined by the three-phase traffic theory can be identified by the traffic flow characteristics measured from loop detector stations. According to the three-phase traffic theory, the traffic flow is separated into three steady states, including free flow (F), synchronized flow (S), and wide moving jams (J). In addition to the three steady states, there are four traditional states in this study, including the transitional state from free flow to synchronized flow (F→S), the transitional state from synchronized flow to free flow (S→F), the transitional state from synchronized flow to wide moving jams (S→J), and the transitional state from wide moving jams to synchronized flow (J→S). However, previous studies have demonstrated that the wide moving jams generally do not emerge with the free flow phase [25]. Thus, the transitional state from free flow to wide moving jams was not considered in the present study. The identification method of these traffic flow states has been introduced as follows [29–32]:

- (1) The free flow (F) is characterized by high vehicle speeds and low traffic density. The free flow phase can be easily distinguished from congested flow using the time series plot of speed and occupancy.
- (2) The synchronized flow (S) is characterized by weak correlation between flow rate and density, with a correlation parameter lower than 0.2.
- (3) The wide moving jams (J) are identified by strong correlation between density and flow rate, with a correlation parameter greater than 0.5.
- (4) The transitional states between free flow (F) and synchronized flow (S) are identified by a sudden change in the time series plot of speed and occupancy. The reduction in speed overtime is considered an indicator for the transitional state from free flow to synchronized flow (F→S), and vice versa.

- (5) The transitional states between synchronized flow (S) and wide moving jams (J) are identified by the correlation between density and flow rate, with a correlation parameter between 0.2 and 0.5. The reduction in speed overtime is considered an indicator for the transitional state from synchronized flow to wide moving jams (S→J), and vice versa.

3.3. The Identification of Secondary Crash. The method based on speed contour figure was applied to identify SC in this study. This method uses 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. The identification method is introduced in detail as follows [13]:

- (1) 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 has been offered by the figure about whether the queue formations 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.
- (2) 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 were averaged over all the crash-free days.
- (3) 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) The crashes that happened within the spatial and temporal influencing ranges of a prior crash were identified as SC. The crashes that did not lead to SC were identified as NC.

Following the above four steps of identification method, the summary of SC, PC, and NC in different traffic flow states is given in Table 2, respectively. Compared to the previous study by authors, only speed contour figure was

TABLE 1: Candidate variables.

Variable category	Symbol	Variables
Traffic flow characteristics	C	Correlation coefficient between occupancy and flow during 5–10 minutes
	Cnt/Spd/Occ	Average 30 s vehicle count/speed/detector occupancy during 5–10 minutes (veh/30 s)/(mile/h)/(%)
	StdC/StdS/StdO	Std. dev. of 30 s vehicle count/speed/detector occupancy during 5–10 minutes (veh/30 s)/(mile/h)/(%)
	Cc/Cs/Co	Coefficient of variation of 30 s vehicle count/speed/detector occupancy during 5–10 minutes (veh/30 s)/(mile/h)/(%)
	Dc/Ds/Do	The difference in average 30 s vehicle count/speed/detector occupancy between 15–20 min and 5–20 min (veh/30 s)/(mile/h)/(%)
	Lcnt/Lspd/Locc	Average difference in 30 s vehicle count/speed/detector occupancy between adjacent lanes during 5–10 minutes (veh/30 s)/(mile/h)/(%)
	Lstdc/Lstds/Lstdo	Std. dev. of difference in 30 s vehicle count/speed/detector occupancy between adjacent lanes during 5–10 minutes (veh/30 s)/(mile/h)/(%)
Environmental characteristics	Vi	Visibility (mile)
	We	1 = bad weather condition; 0 = otherwise
	Rs	1 = otherwise; 0 = dry road surface
	Li	1 = bad light; 0 = otherwise
Geometric characteristics	Lw	Lane width (ft)
	Is	Inner shoulder width (ft)
	Os	Outer shoulder width (ft)
Crash characteristics	Ra	1 = ramp segment; 0 = otherwise
	Tr	1 = the crash including a truck; 0 = otherwise
	Se	1 = the injury crash; 0 = otherwise
	Ho	1 = the hit object crash; 0 = otherwise
	Ss	1 = the sideswipe crash; 0 = otherwise
	Re	1 = the rear end crash; 0 = otherwise

used to identify SC without considering the shockwave in the present study [13], because this study focused on the difference in safety performance of SC associated with various traffic flow states.

3.4. Bayesian Conditional Logit Model. As discussed in the literature review section, although some researchers have considered the traffic flow states into the studies of SC, no researchers have studied the difference in safety performance of SC associated with various traffic flow states divided by classical macroscopic traffic flow theory. The conditional logit model was applied to quantitatively analyse the relative safety performance of SC in different traffic flow states while controlling for the effects of other traffic related variables, such as weather condition, geometric metric design, road pavement, etc. The model can be written as [33–37]

$$y_{ijk} \sim \text{Bernoulli}(p_{ijk}),$$

$$P(y_{itk}) = \frac{1}{\{1 + \exp[-\alpha_i + \sum_{k=1}^K \beta_k x_{ijk}]\}}, \quad (1)$$

where x_{ijk} is the k th unmatched variable for the case ($j=0$) or the j th control in the i th matched set. Therefore, $X = \{x_{ijk}\}$ consists of all the cases, and all matched sets are controlled, where $i = 1, 2, \dots, I$; $j = 0, 1, \dots, J$; $k = 1, 2, \dots, K$. I represents the total number of matched sets; J represents the number of controls in each matched set; and K represents the number of contributing variables.

TABLE 2: The summary of different crash types in different traffic flow states.

Traffic flow states	Normal crash	Primary crash	Secondary crash	Total
F	7584	50	27	7661
S	825	14	16	855
J	273	4	15	292
F→S	367	8	10	385
S→F	336	8	10	354
S→J	195	10	11	216
J→S	145	3	8	156
Total	9725	97	97	9919

α_i is the effect of matching variables on the probability of SC occurrence for each matched set; β_k represents the estimated coefficients for explanatory variables; and x_k is the unmatched contributing variables applied in this study.

A conditional probability is used to account for the selection bias of the matched case-control design. The conditional probability that the first vector of the contributing variables x_{i0} in the i th matched set corresponds to the case, conditional on $x_{i0}, x_{i1}, \dots, x_{iJ}$ being the vectors of contributing variables in the i th matched set, is shown as

$$P_i^c = \frac{\exp(\sum_{k=1}^K \beta_k x_{i0k})}{\exp(\sum_{k=1}^K \beta_k x_{i0k}) + \sum_{j=1}^J \exp(\sum_{k=1}^K \beta_k x_{ijk})}. \quad (2)$$

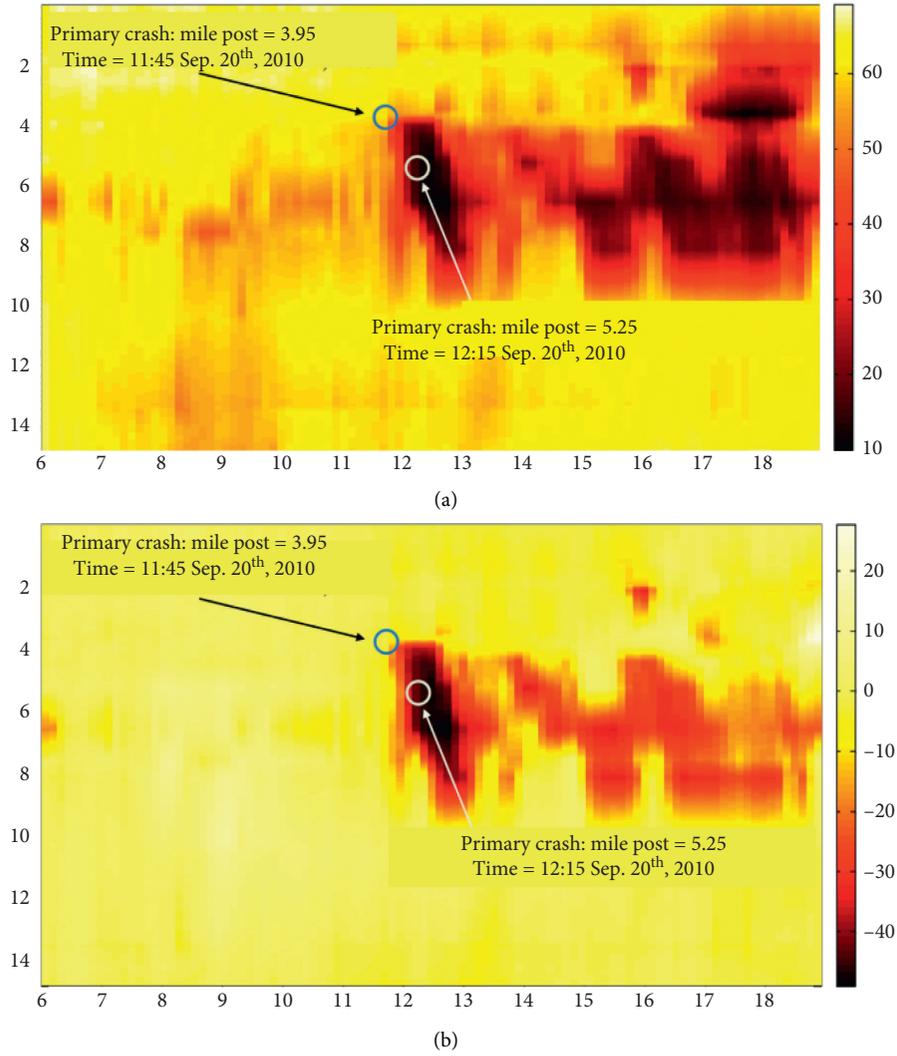


FIGURE 1: Initial identification of SC. (a) Speed contour plot without accounting for the recurrent congestions. (b) Speed contour plot accounting for the recurrent congestions.

Therefore, the probability function of the conditional logit model is expressed as

$$\begin{aligned}
 f(Y|\beta) &= \prod_{i=1}^I f(y_{i0} = 1|\beta) = \prod_{i=1}^I P_i^c \\
 &= \exp \left\{ \sum_{i=1}^I \sum_{k=1}^K (\beta_k x_{i0k}) - \sum_{i=1}^I \log \left[\sum_{j=0}^J \exp \left(\sum_{k=1}^K \beta_k x_{ijk} \right) \right] \right\}. \quad (3)
 \end{aligned}$$

This study applied the Bayesian inference method based on Markov Chain Monte Carlo (MCMC) for the specification of the conditional logit model. Compared to the point estimations of the traditional maximum likelihood estimation (MLE) method, the Bayesian modelling technique regards all unknown parameters as random variables with a prior distribution. The estimates of the mean, standard deviation, and quartiles of the coefficients can be affected by the posterior distribution. Based on the Bayes'

theory, the posterior distribution of parameters can be expressed as

$$f(\beta|Y) = \frac{f(Y, \beta)}{f(Y)} = \frac{f(Y|\beta)\pi(\beta)}{\int f(Y, \beta)d\beta} \propto f(Y|\beta)\pi(\beta), \quad (4)$$

where $f(\beta|Y)$ is the posterior joint distribution of parameters β conditional upon dataset Y , $f(Y, \beta)$ is the joint probability distribution of dataset Y and model parameters β , $f(Y|\beta)$ denotes the likelihood conditional on model parameters β , and The function $\pi(\beta)$ is the prior distribution of model parameters β . The non-informative prior distributions were applied for the model parameters, which can be written as

$$\beta \sim \text{Normal}(0_K, 10^6 I_K), \quad (5)$$

where 0_K represents a $K \times 1$ vector of zeros and I_K represents a $K \times K$ identity matrices. Based on the specification of the

prior distributions for the model parameters β , the posterior joint distribution $f(\beta|Y)$ is expressed as

$$f(\beta|Y) \propto f(Y|\beta)\pi(\beta) = \prod_{i=1}^I f(y_{i0} = 1|\beta) \times \prod_{k=1}^K N\left(\beta_k \mid \mu_k, \sum_k\right) \quad (6)$$

$$\propto \exp\left\{\sum_{i=1}^I \sum_{k=1}^K (\beta_k x_{i0k}) - \sum_{i=1}^I \log\left[\sum_{j=0}^J \exp\left(\sum_{k=1}^K \beta_k x_{ijk}\right)\right] - \frac{1}{2} \sum_{k=1}^K \frac{(\beta_k)^2}{10^6}\right\}.$$

The Markov Chain Monte Carlo (MCMC) method was used to generate realizations from the posterior joint distribution of the model parameters and draw parameters sequentially from equation (6). Compared to the nonstandard conditional distributions in equation (6), the Metropolis-Hasting sampling approach was applied to generate random draws. The inference was used based on the remaining draws after discarding the draws during the burn-in period.

In the present study, only Bayesian conditional logit model was used to analyse the difference in safety performance of SC associated with various traffic flow states. To compare the models, in terms of which one is better in future studies, DIC and AUC values can be adopted. DIC is recognized as Bayesian generalization of AIC (Akaike information criterion) and it was adopted for model comparisons. DIC is a combination of model fit measurement and the effective number of parameters; the smaller DIC indicates a better model fit. AUC value, which is area under the receiver operating characteristic (ROC) curve, was chosen to evaluate and compare these models; larger AUC values indicate a better goodness-of-fit and classification power [38].

4. Results and Discussion

4.1. Safety Performance of Secondary Crash in Different Traffic Flow States. In this study, compared to other studies based on the Bayesian conditional logit model, the Bayesian conditional logit model was used to quantify the difference in the safety performance of SC associated with various traffic flow states divided by three-phase traffic theory. The group variables are separated based on case and control samples. In models, the events (the value of dependent variable is 1) are PC that induced SC and the non-events (the value of dependent variable is 0) are NC that did not induce SC. Because the traffic flow in this study is divided into seven states, including free flow (F), synchronized flow (S), wide moving jams (J), the transitional state from free flow to synchronized flow (F→S), the transitional state from synchronized flow to free flow (S→F), the transitional state from synchronized flow to wide moving jams (S→J), and the transitional state from wide moving jams to synchronized flow (J→S), thus, in this method, the free flow (F) state is considered as the reference level, and the other six

traffic flow states are considered as six independent variables. Finally, the odds ratio can be used to quantify the difference in the safety performance of SC between free flow (F) and the other six traffic flow states. This model did not include other traffic flow variables such as speed and density, because the traffic flow states were highly correlated with traffic flow variables [39].

Three parallel MCMC chains were constructed for Bayesian inference. Each MCMC chain consisted of 10 000 iterations, including an initial “burn-in” period of 4000 iterations [39]. The estimations of each parameter from the MLE method were considered initial values. The initial values for multiple MCMC chains were dispersed throughout the 90% confidence intervals of the estimated parameters from the MLE. The convergence of the posterior distribution samples was checked by the visual inspection of the trace plots, posterior density plots, and autocorrelation function plots. In addition, the Gelman Rubin potential scale reduction (PSR) was also checked. If the PSR was lower than 1.1, the multiple chains were considered converged [39]. The estimation results of the Bayesian conditional logit models are given in Table 3. The 95% credible interval for each parameter in Table 3 indicates that the traffic flow states significantly affect the probability of SC occurrences. The odds ratio for each variable was used to quantify the safety performance of SC in different traffic flow states.

In the Bayesian conditional logit models, as shown in Table 3, the results suggest that the odds ratios of synchronized flow (S), the transitional state from free flow to synchronized flow (F→S), the transitional state from synchronized flow to free flow (S→F), the transitional state from synchronized flow to wide moving jams (S→J), and the transitional state from wide moving jams to synchronized flow (J→S) are significantly greater than free flow (F), and the odds ratio of wide moving jams (J) is not significantly greater than free flow (F). Accordingly, free flow (F) has the best safety performance in terms of the lowest SC likelihood. However, synchronized flow (S) has the highest SC likelihood, followed by the transitional state from synchronized flow to wide moving jams (S→J). The probabilities of SC occurrence associated with F→S, S→F, and J→S are very similar to each other. In detail, the SC probability of synchronized flow (S) is 8.561 times higher than free flow (F). The SC probability of the transitional state from free flow to synchronized flow (F→S) is 2.488 times

TABLE 3: The estimation results of Bayesian conditional logit models.

Traffic flow states	Mean	MC error	2.50%	97.50%	Odds ratio
S	2.258	0.016	1.368	3.190	9.561
J	1.258	0.025	-0.115	2.634	3.519
F→S	1.249	0.018	0.445	2.066	3.488
S→F	1.512	0.019	0.627	2.434	4.535
S→J	2.191	0.018	1.082	3.327	8.943
J→S	1.585	0.025	0.430	2.769	4.881

^aFree flow (F) is the reference level.

higher than free flow (F). The SC probability of the transitional state from synchronized flow to free flow (S→F) is 3.535 times higher than free flow (F). The SC probability of the transitional state from synchronized flow to wide moving jams (S→J) is 7.943 times higher than free flow (F). The SC probability of the transitional state from wide moving jams to synchronized flow (J→S) is 3.881 times higher than free flow (F).

4.2. The Contributing Factors of Secondary Crash in Different Traffic Flow States. To identify how different contributing factors affect the SC probability in different traffic flow states, the traditional logistic regression models were applied [40]. The events are PC that induced SC and the non-events are NC that did not induce SC. In models, 1 is PC and 0 is NC. *P* value of 0.1 was employed for parameter estimate significance in these models. The models were estimated using the software package STATA. To avoid the biased results caused by multicollinearity, the Pearson correlation parameters between different candidate variables were calculated. The highly correlated explanatory variables were avoided to be included into the model simultaneously. The significant variables of the traditional logistic regression models in different traffic flow states are presented in Table 3. The meaning of symbols in Table 4 has been explained in Table 1.

In free flow (F), as shown in Table 4, *Stdc* and *Lstdo* were found to be positively related to the SC probability, but *Spd* and *Ss* were found to be negatively related to the SC probability. The difference of occupancy between adjacent lanes was found to be related to lane-change frequency [41]. The results indicated that the free flow with low speed and more lane changing behaviours can result in the increasing of SC risk. Moreover, the non-sideswipe prior crash can significantly increase the SC risk in free flow. In summary, the preventive measures for SC in free flow (F) are rapid evacuation of congestion and decreasing the interaction between vehicles.

In synchronized flow (S), *Lw*, *Os*, *Occ*, *Lspd*, and *Ho* were found to be positively related to the SC probability, and *Stdo* was found to be negatively related to the SC probability. In synchronized flow, there is a tendency to the synchronization of speeds on each lane and across different lanes [19]. The results showed that the congested flow and synchronization of traffic flow can lead to the increasing of SC likelihood. In addition, larger lane width and outer shoulder width may encourage drivers to take advantage of large space

TABLE 4: The estimation results of the traditional logistic regression models.

Variables	Coefficient	S.D.	$P> z $	2.50%	97.50%
Free flow (F)					
<i>Spd</i>	-0.144	0.022	0.0001	-0.187	-0.100
<i>Stdc</i>	0.472	0.146	0.001	0.185	0.759
<i>Lstdo</i>	13.093	4.187	0.002	4.886	21.300
<i>Ss</i>	-1.306	0.533	0.014	-2.351	-0.261
Cons	2.511	1.487	0.091	-0.403	5.425
Synchronized flow (S)					
<i>Lw</i>	3.295	1.005	0.001	1.325	5.264
<i>Os</i>	0.274	0.126	0.029	0.027	0.520
<i>Occ</i>	12.822	4.849	0.008	3.318	22.325
<i>Stdo</i>	-9.711	4.097	0.018	-17.742	-1.681
<i>Lspd</i>	0.087	0.043	0.043	0.003	0.171
<i>Ho</i>	1.403	0.829	0.091	-0.222	3.027
Cons	-48.560	13.309	0.0001	-74.645	-22.476
Wide moving jams (J)					
<i>Stdo</i>	20.025	8.668	0.021	3.037	37.014
<i>Lcnt</i>	-1.685	0.576	0.003	-2.813	-0.556
Cons	-6.809	1.697	0.0001	-10.135	-3.483
The transitional state from free flow to synchronized flow (F→S)					
<i>Is</i>	-0.275	0.126	0.029	-0.521	-0.029
<i>Dc</i>	-0.764	0.280	0.006	-1.313	-0.215
<i>Occ</i>	26.222	9.524	0.006	7.556	44.888
<i>Co</i>	8.109	2.639	0.002	2.936	13.281
<i>Lstdo</i>	-23.091	10.220	0.024	-43.122	-3.060
Cons	-8.399	2.144	0.0001	-12.602	-4.197
The transitional state from synchronized flow to free flow (S→F)					
<i>C</i>	-4.475	1.635	0.006	-7.681	-1.270
<i>Ds</i>	0.077	0.047	0.100	-0.015	0.170
<i>Ss</i>	1.459	0.789	0.065	-0.088	3.005
Cons	-1.254	0.912	0.169	-3.041	0.534
The transitional state from synchronized flow to wide moving jams (S→J)					
<i>Is</i>	-0.241	0.118	0.041	-0.473	-0.010
<i>Dc</i>	-0.695	0.288	0.016	-1.258	-0.131
<i>Stdo</i>	9.106	4.366	0.037	0.548	17.663
Cons	-1.586	1.119	0.156	-3.780	0.607
The transitional state from wide moving jams to synchronized flow (J→S)					
<i>Tr</i>	7.849	3.412	0.021	1.161	14.536
<i>Cs</i>	-27.593	14.976	0.065	-56.945	1.759
Cons	0.442	2.240	0.844	-3.948	4.831

to pass the congested flow which is caused by a prior crash. Thus, the SC probability will increase in larger lane width and outer shoulder width situation. If the prior crash is a hit object crash, the SC probability will also increase. In summary of the preventive measures for SC in synchronized flow (S), relieving the traffic congestion as soon as possible, reducing the number of lanes, and reducing outer shoulder width can decrease the SC risk while a prior crash occurs, especially a hit object prior crash.

In wide moving jams (J), *Stdo* was found to be positively related to the SC probability, and *Lcnt* was found to be negatively related to the SC probability. The results indicated that the synchronization of vehicles between different adjacent lanes can significantly increase the SC likelihood. The large variation of occupancy in wide moving jams can

significantly increase the SC likelihood. Therefore, making full use of available lanes and relieving the traffic congestion quickly, while a prior crash occurs, can help to reduce the SC risk in wide moving jams (J).

In the transitional state from free flow to synchronized flow ($F \rightarrow S$), Occ and Co were found to be positively related to the SC probability, but Is, Dc, and Lstdo were found to be negatively related to the SC probability. The results implied that the significant tendency from free flow to synchronized flow can result in more congestion, less space, and more lane-change behaviours; the SC likelihood will increase with the traffic flow more and more congested in $F \rightarrow S$. In addition, the evacuation of congested flow can benefit from the larger inner shoulder width, and the larger inner shoulder width can provide more space for drivers to apply crash avoidance measures. Thus, the larger inner shoulder width can decrease the SC likelihood.

In the transitional state from synchronized flow to free flow ($S \rightarrow F$), Ds and Ss were found to be positively related to the SC probability, and Cc was found to be negatively related to the SC probability. Similar to the results of $F \rightarrow S$, the more congested the transitional state tends to be, the larger the SC probability will get. In addition to the traffic flow characteristics, if the prior crash is a sideswipe crash, the SC probability will also increase. In summary of the preventive measures for SC in $S \rightarrow F$ state, relieving the traffic congestion quickly can help to reduce the SC risk while a prior crash occurs, especially a sideswipe prior crash.

In the transitional state from synchronized flow to wide moving jams ($S \rightarrow J$), Stdo was found to be positively related to the SC probability, but Is and Dc were found to be negatively related to the SC probability. Synchronized flow and wide moving jams are both congested flow. In this transitional state, more congested flow will decrease the SC probability. This result is opposite to the results in $F \rightarrow S$ and $S \rightarrow F$. It is because the less available space for drivers will lead to less dangerous driving behaviours. In this transitional state, similar to the results of $F \rightarrow S$, the larger inner shoulder width can provide more space for drivers to apply crash avoidance measures.

In the transitional state from wide moving jams to synchronized flow ($J \rightarrow S$), Tr was found to be positively related to the SC probability, and Cs was found to be negatively related to the SC probability. In previous studies, it has been proved that the prior crash including a truck is found to be a significant factor of SC [11]. In this transitional state, the results showed that the prior crash including a truck also significantly affects SC risk.

5. Conclusion

In this study, the traffic flow is divided by three-phase traffic theory. The main purpose is to analyse the difference of safety performance for SC in different traffic flow states and explore how contributing factors affect the probability of SC in different traffic flow states. The SC related data were collected from the I-880 freeway in the United States from 2006 to 2011. The Bayesian conditional logit models have

been established to analyse the statistical relationship between the SC probability and traffic flow states. The traditional logistic regression models were established to quantify the effects of various variables on the SC probability in different traffic flow states.

More specifically, the results of the Bayesian conditional logit model have been summarized as follows:

- (1) F has the best safety performance in terms of the lowest SC likelihood
- (2) S has the highest SC likelihood, followed by $S \rightarrow J$
- (3) The probabilities of SC occurrence associated with $F \rightarrow S$, $S \rightarrow F$, and $J \rightarrow S$ are very similar to each other
- (4) J is not significantly greater than F

In addition, the results of the traditional logistic regression model have been summarized as follows:

- (1) In free flow (F), rapid evacuation of congestion and decreasing the interaction between vehicles can reduce the SC risk while a prior crash occurs, especially a non-sideswipe prior crash
- (2) In synchronized flow (S), relieving the traffic congestion as soon as possible, reducing the number of lanes, and reducing outer shoulder width can decrease the SC risk while a prior crash occurs, especially a hit object prior crash
- (3) In wide moving jams (J), making full use of available lanes and relieving the traffic congestion quickly, while a prior crash occurs, can help to reduce the SC risk
- (4) In the transitional state from free flow to synchronized flow ($F \rightarrow S$) and the transitional state from synchronized flow to wide moving jams ($S \rightarrow J$), larger inner shoulder width can decrease the SC likelihood
- (5) In the transitional state from synchronized flow to free flow ($S \rightarrow F$), relieving the traffic congestion quickly can help to reduce the SC risk while a prior crash occurs, especially a sideswipe prior crash
- (6) In the transitional state from wide moving jams to synchronized flow ($J \rightarrow S$), the prior crash including a truck also significantly affects SC risk

This research can help traffic management personnel better understand which traffic flow state is more dangerous for the occurrence of SC and realize the contributing factors of SC in different traffic flow states. The results can be applied to develop effective countermeasures and reduce the SC probability in different traffic flow states. However, there are still several issues and potential future studies as follows:

- (1) More PC characteristics need to be considered in the analysis of SC risk. In this study, only four crash types were taken into account in models
- (2) In this study, the traffic flow is only divided by three-phase traffic theory. More macroscopic traffic flow theories should be studied in the future

- (3) Additional research is needed to test for the transferability of the research findings to other freeways
- (4) In future studies, more models should be used to compare with the Bayesian conditional logit model and the traditional logistic regression model in this study

Data Availability

The weather data were obtained from the National Climate Data Center (NCDC) website which provides hourly weather information. The geometric and traffic data were collected from the nearest loop detector stations to each collision location and obtained from the Highway Performance Measurement System (PeMS) maintained by the California Department of Transportation (Caltrans). Crash data were obtained from the Statewide Integrated Traffic Records System (SWITRS) of the Caltrans.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was sponsored by the Projects of the National Natural Science Foundation of China (71701046 and 51878236).

References

- [1] C. Zhan, A. Gan, and M. Hadi, "Identifying secondary crashes and their contributing factors," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2102, no. 1, pp. 68–75, 2009.
- [2] H. Zhang and A. Khattak, "What is the role of multiple secondary incidents in traffic operations?" *Journal of Transportation Engineering*, vol. 136, no. 11, pp. 986–997, 2010.
- [3] C. Xu, S. Xu, C. Wang, and J. Li, "Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression," *Physica A: Statistical Mechanics and Its Applications*, vol. 524, pp. 121–129, 2019.
- [4] Z. Wang and H. Jiang, "Identifying secondary crashes on freeways by leveraging the spatio-temporal evolution of shockwaves in the speed contour plot," *Journal of Transportation Engineering Part A-Systems*, vol. 146, no. 2, Article ID 04019072, 2020.
- [5] A. E. Kitali, P. Alluri, T. Sando, H. Haule, E. Kidando, and R. Lentz, "Likelihood estimation of secondary crashes using Bayesian complementary log-log model," *Accident Analysis & Prevention*, vol. 119, pp. 58–67, 2018.
- [6] A. E. Kitali, P. Alluri, T. Sando, and W. Wu, "Identification of secondary crash risk factors using penalized logistic regression model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 11, pp. 901–914, 2019.
- [7] H. Yang, Z. Wang, K. Xie, and D. Dai, "Use of ubiquitous probe vehicle data for identifying secondary crashes," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 138–160, 2017.
- [8] J. E. Moore, G. Giuliano, and S. Cho, "Secondary accident rates on los angeles freeways," *Journal of Transportation Engineering*, vol. 130, no. 1, pp. 280–285, 2004.
- [9] C. Zhan, L. Shen, and M. Hadi, "Understanding the characteristics of secondary crashes on freeways," in *Proceedings of the Presented at 87th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2008.
- [10] W. Hirunyanitawattana and S. P. Mattingly, "Identifying secondary crash characteristics for the california highway system," in *Presented at 85th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2006.
- [11] H. Yang, Z. Wang, K. Xie, K. Ozbay, and M. Imprialou, "Methodological evolution and frontiers of identifying, modeling and preventing secondary crashes on highways," *Accident Analysis & Prevention*, vol. 117, pp. 40–54, 2018.
- [12] N. J. Goodall, "Probability of secondary crash occurrence on freeways with the use of private-sector speed data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2635, no. 1, pp. 11–18, 2017.
- [13] B. Yang, Y. Guo, and C. Xu, "Analysis of freeway secondary crashes with a two-step method by loop detector data," *IEEE Access*, vol. 7, pp. 22884–22890, 2019.
- [14] M.-I. M. Imprialou, F. P. Orfanou, E. I. Vlahogianni, and M. G. Karlaftis, "Methods for defining spatiotemporal influence areas and secondary incident detection in freeways," *Journal of Transportation Engineering*, vol. 140, no. 1, pp. 70–80, 2014.
- [15] B. S. Kerner and H. Rehborn, "Experimental properties of complexity in traffic flow," *Physical Review E*, vol. 53, no. 5, pp. 4275–4278, 1996.
- [16] B. S. Kerner, "Experimental features of self-organization in traffic flow," *Physical Review Letters*, vol. 81, no. 17, pp. 3797–3800, 1998.
- [17] N. Wu, "A new approach for modeling of Fundamental Diagrams," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 10, pp. 867–884, 2002.
- [18] Transportation Research Board, *Highway Capacity Manual*, Transportation Research Board of the National Academies, Washington, DC, USA, 2010.
- [19] C. Xu, P. Liu, W. Wang, and Z. Li, "Safety performance of traffic phases and phase transitions in three phase traffic theory," *Accident Analysis & Prevention*, vol. 85, pp. 45–57, 2015.
- [20] C. Xu, P. Liu, and F. W. Zhang, "Development of a real-time crash risk prediction model incorporating the various crash mechanisms across different traffic states," *Traffic Injury Prevention*, vol. 16, no. 1, pp. pp28–35, 2015.
- [21] H. Park, A. Haghani, S. Samuel, and M. A. Knodler, "Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion," *Accident Analysis & Prevention*, vol. 112, pp. 39–49, 2018.
- [22] R. A. Raub, "Occurrence of secondary crashes on urban arterial roadways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1581, no. 1, pp. 53–58, 1997.
- [23] M. G. Karlaftis, S. P. Latoski, N. J. Richards et al., "Empirical analysis of secondary crash causes," in *Presented at 77th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 1998.
- [24] C.-S. Chou and E. Miller-Hooks, "Simulation-based secondary incident filtering method," *Journal of Transportation Engineering*, vol. 136, no. 8, pp. 746–754, 2010.

- [25] H. Zhang, M. Cetin, and A. J. Khattak, "Joint analysis of queuing delays associated with secondary incidents," *Journal of Intelligent Transportation Systems*, vol. 19, no. 2, pp. 192–204, 2015.
- [26] H. Park and A. Haghani, "Real-time prediction of secondary incident occurrences using vehicle probe data," *Transportation Research Part C: Emerging Technologies*, vol. 70, pp. 69–85, 2016.
- [27] J. Wang, L. Boya, Z. Lanfang, and D. R. Ragland, "Modeling secondary accidents identified by traffic shock waves," *Accident Analysis and Prevention*, vol. 87, pp. 141–147, 2016.
- [28] C. Xu, P. Liu, B. Yang, and W. Wang, "Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 406–418, 2016.
- [29] Y. Toshiyuki and V. Shankar, "Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects," *Accident Analysis and Prevention*, vol. 36, pp. 869–879, 2004.
- [30] W. Knospe, L. Santen, A. Schadschneider et al., "Single-vehicle data of highway traffic: microscopic description of traffic phases," *Physical Review E*, vol. 65, pp. 1–16, 2002.
- [31] J. Carroll, "Generalization of canonical correlation analysis to three or more sets of variables," in *Presented at 76th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 1968.
- [32] V. Burg, *Nonlinear Canonical Correlation and Some Related Techniques*, DSWO Press, Leiden, Netherlands, 1988.
- [33] N. Bruce, D. Pope, and D. Stanistreet, *Quantitative Methods for Health Research: A Practical Interactive Guide to Epidemiology and Statistics*, John Wiley & Sons, Hoboken, NJ, USA, 2008.
- [34] Y. Guo, P. Liu, Y. Wu, and J. Chen, "Evaluating how right-turn treatments affect right-turn-on-red conflicts at signalized intersections," *Journal of Transportation Safety & Security*, vol. 12, no. 3, pp. 419–440, 2020.
- [35] Y. Guo, Z. Li, P. Liu, and Y. Wu, "Modeling correlation and heterogeneity in crash rates by collision types using full bayesian random parameters multivariate Tobit model," *Accident Analysis & Prevention*, vol. 128, pp. 164–174, 2019.
- [36] Y. Guo, Z. Li, and T. Sayed, "Analysis of crash rates at freeway diverge areas using bayesian tobit modeling framework," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 4, pp. 652–662, 2019.
- [37] Y. Guo, Z. Li, P. Liu, and Y. Wu, "Exploring risk factors with crashes by collision type at freeway diverge areas: accounting for unobserved heterogeneity," *IEEE Access*, vol. 7, pp. 11809–11819, 2019.
- [38] R. Yu, X. Wang, K. Yang, and M. Abdel-Aty, "Crash risk analysis for Shanghai urban expressways: a Bayesian semi-parametric modeling approach," *Accident Analysis & Prevention*, vol. 95, pp. 495–502, 2016.
- [39] C. Xu, Y. Wang, P. Liu, W. Wang, and J. Bao, "Quantitative risk assessment of freeway crash casualty using high-resolution traffic data," *Reliability Engineering and System Safety*, vol. 169, pp. 299–311, 2018.
- [40] Z. Li, W. Wang, R. Chen, P. Liu, and C. Xu, "Evaluation of the impacts of speed variation on freeway traffic collisions in various traffic states," *Traffic Injury Prevention*, vol. 14, no. 8, pp. 861–866, 2013.
- [41] D. C. Gazis, R. Herman, and G. H. Weiss, "Density oscillations between lanes of a multilane highway," *Operations Research*, vol. 10, no. 5, pp. 658–667, 1962.