Accident Analysis and Prevention 64 (2014) 52–61

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks

Zhibin Li[a,1], Seongchae Ahn[b], Koohong Chung[c,*], David R. Ragland[d,2], Wei Wang[a,3], Jeong Whon Yu[b]

[a] School of Transportation, Southeast University, 2 Si Pai Lou, Nanjing 210096, China
[b] Department of Transportation Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, South Korea
[c] School of Civil, Environmental and Architectural Engineering, Korea University, Anam-Dong, Seongbuk-Gu, Seoul 136-713, South Korea
[d] Safe Transportation Research & Education Center, University of California, Berkeley, 2614 Dwight Way #7374, Berkeley, CA 94720-7374, United States

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 30 October 2013
Accepted 6 November 2013

Keywords:
Surrogate safety measure
Rear-end collision
Collision risk
Kinematic wave
Recurrent bottleneck

ABSTRACT

This study presents a surrogate safety measure for evaluating the rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks using aggregated traffic data from ordinary loop detectors. The attributes of kinematic waves that accompany rear-end collisions and the traffic conditions at detector stations spanning the collision locations were examined to develop the rear-end collision risk index (RCRI). Together with RCRI, standard deviations in occupancy were used to develop a logistic regression model for estimating rear-end collision likelihood near freeway recurrent bottlenecks in real-time. The parameters in the logistic regression models were calibrated using collision data gathered from the 6-mile study site between 2006 and 2007. Findings indicated that an additional unit increase in RCRI results in increasing the odds of rear-end collision by 21.1%, a unit increase in standard deviation of upstream occupancy increases the odds by 19.5%, and a unit increase in standard deviation of downstream occupancy increases the odds by 18.7%. The likelihood of rear-end collisions is highest when the traffic approaching from upstream is near capacity state while downstream traffic is highly congested. The paper also reports on the findings from comparing the predicted number of rear-end collisions at the study site using the proposed model with the observed traffic collision data from 2008. The proposed model's true positive rates were higher than those of existing real-time crash prediction models.

Published by Elsevier Ltd.

1. Introduction

Backward moving kinematic waves emanating from freeway bottlenecks force approaching vehicles to abruptly change their traveling speeds. When approaching vehicles do not adjust their speeds in a timely manner, the spacing between vehicles can decrease rapidly and potentially cause rear-end traffic collisions. The objective of this paper is to develop a surrogate safety measure to quantify the likelihood of rear-end collisions in the vicinity of recurrent bottleneck areas by monitoring the changes in traffic conditions induced by backward moving kinematic waves using aggregated traffic data from ordinary loop detectors.

Previous studies have proposed statistical models as the surrogate safety measure to evaluate traffic collision risk using aggregated traffic data obtained from loop detectors on freeways (Lee et al., 2002, 2003; Golob and Recker, 2003; Golob et al., 2004; Abdel-Aty et al., 2004, 2005; Abdel-Aty and Pande, 2005; Pande and Abdel-Aty, 2006; Kockelman and Ma, 2007; Zheng et al., 2010; Hassan and Abdel-Aty, 2011; Hossain and Muromachi, 2011; Abdel-Aty et al., 2012; Xu et al., 2012). For example, Abdel-Aty et al. (2004) developed a logistic regression model to predict the occurrence of traffic collisions using real-time traffic data from a section of freeway. The study reported that the 5-min average occupancy observed at an upstream detector station and the coefficient of variation of speed at a downstream location 5–10 min prior to collisions significantly affect the collision occurrence. Findings from a more recent study by Zheng et al. (2010) indicated that the standard deviation of speed in a 10-min interval could be considered a surrogate safety measure for collisions under congested traffic conditions. These studies estimated collision risk based on statistical data mining approach and did not evaluate how the propagation of kinematic waves affects rear-end type collision risk.

To develop the surrogate safety measure for rear-end collisions, several researchers used different types of hazardous traffic conditions (i.e., situations in which a driver must take evasive action to avoid traffic collision), including time-to-collision (Saccomanno

* Corresponding author. Tel.: +1 510 622 5429.
E-mail addresses: lizhibin@seu.edu.cn (Z. Li), citadel@ajou.ac.kr (S. Ahn), koohong_chung@dot.ca.gov (K. Chung), davidr@berkeley.edu (D.R. Ragland), wangwei@seu.edu.cn (W. Wang), jeongwhon@ajou.ac.kr (J.W. Yu).
[1] Tel: +86 13952097374.
[2] Tel: +1 510 642 0655.
[3] Tel: +86 13905170160.

0001-4575/$ – see front matter. Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.aap.2013.11.003

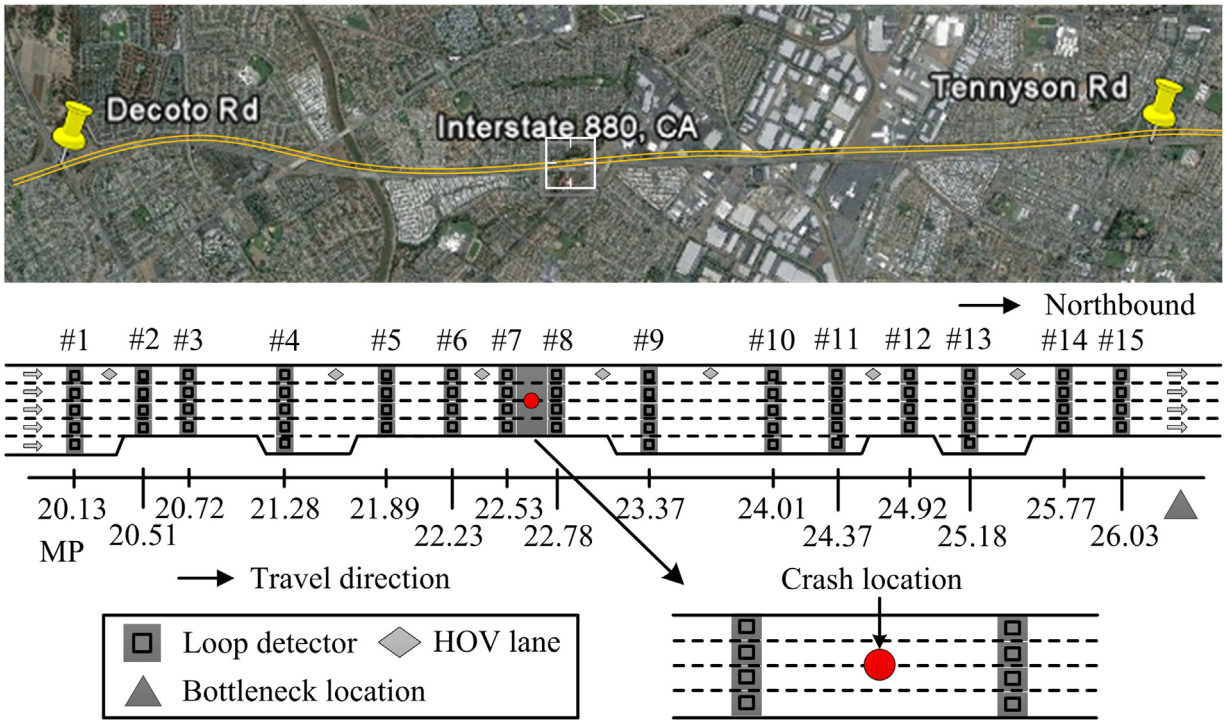

Fig. 1. Schematic of the study site, northbound Interstate 880, Oakland, CA.

et al., 2008; Oh and Kim, 2010), stopping distance index (Oh et al., 2006, 2009), modified time to collision (Ozbay et al., 2008), and individual vehicle speeds and headways (Hourdos et al., 2006). These earlier studies demonstrated that hazardous traffic conditions could be used as a surrogate safety measure. However, these measures are not suitable for monitoring collision risk when only the information from conventional loop detectors is available – monitoring the hazardous conditions specified in the studies mentioned above requires extracting information from individual vehicle trajectories. Additionally, these studies did not consider the collision risk associated with kinematic waves and proximity to recurrent bottleneck areas on freeways.

Several recent studies analyzed the collision risks among different traffic states and evaluated the safety impacts of kinematic waves at freeway bottleneck areas. Yeo et al. (2010) evaluated the relative risk of traffic collisions after dividing freeway traffic into four traffic states. Their findings indicated that collisions are about 3.6 times more likely to occur at the back of the queue compared with the free-flowing traffic state; findings from a later study by Xu et al. (2012) further confirmed Yeo's study. Chung et al. (2011) examined the attributes of moving kinematic waves that preceded traffic collisions in the vicinity of a recurrent bottleneck. The findings suggested that sudden and pronounced changes in speed induced by fast backward moving kinematic waves increase the likelihood of traffic collisions, and that the propagation of kinematic waves has a large impact on the probability of such collisions. Li et al. (2013) examined the impacts of downstream queues on the occurrence of traffic collisions. Their study suggested that the likelihood of a collision increases as both the spatial and the temporal proximities to the tail of an expanding or receding queue become smaller.

These previous studies mentioned in the preceding paragraph examined the attributes of kinematic waves that accompanied traffic collisions near freeway recurrent bottlenecks. However, these studies did not attempt to develop a model to quantify risk of rear-end collision during the propagation of kinematic waves in real-time which could be important in developing dynamic traffic control measures for reducing collision risks near freeway bottlenecks. The findings from these studies shed light on developing a surrogate safety measure for rear-end collisions upstream of recurrent bottleneck as proposed in the present study.

The description of the site used in the present study is provided in Section 2, while Section 3 discusses the logic behind developing the proposed surrogate safety measure. Section 4 explains the design of case-control used to estimate the parameters of the proposed surrogate safety measure. Section 5 reports on evaluation of the performance of the proposed measure using empirical data. This paper ends with brief concluding remarks and topics for future research in Section 6.

2. Study site

Fig. 1 shows a 6-mile (10-km) northbound section of the Interstate 880 freeway in Oakland, California, selected as the study site. The 6-mile section is comprised of segments that are 4- to 5-lanes wide, with the median lane of the section reserved for high-occupancy vehicles (carpools and buses) during peak hours. The freeway section is plagued by a recurrent bottleneck at its downstream end (represented in the figure as a gray triangle located downstream of detector station 15). Backward moving kinematic waves frequently emanate from near the bottleneck while it remains active.

The freeway section is equipped with inductive loop detectors installed in all travel lanes. The approximate locations of the 15 detector stations in the study section are shown in Fig. 1. When a collision (represented by the red dot in Fig. 1) occurs between two neighboring detectors (represented by the gray shaded region between stations 7 and 8), data observed from the immediate upstream and downstream detectors are used to develop the surrogate safety measure. The spacing between detector stations ranges from 0.25 to 0.64 mile (0.42–1.02 km) with an average of 0.43 mile (0.69 km). Each detector station reports average flow, occupancy (i.e., dimensionless measure of density), and speed observed during 30-s period ($\Delta t = 30$ s).

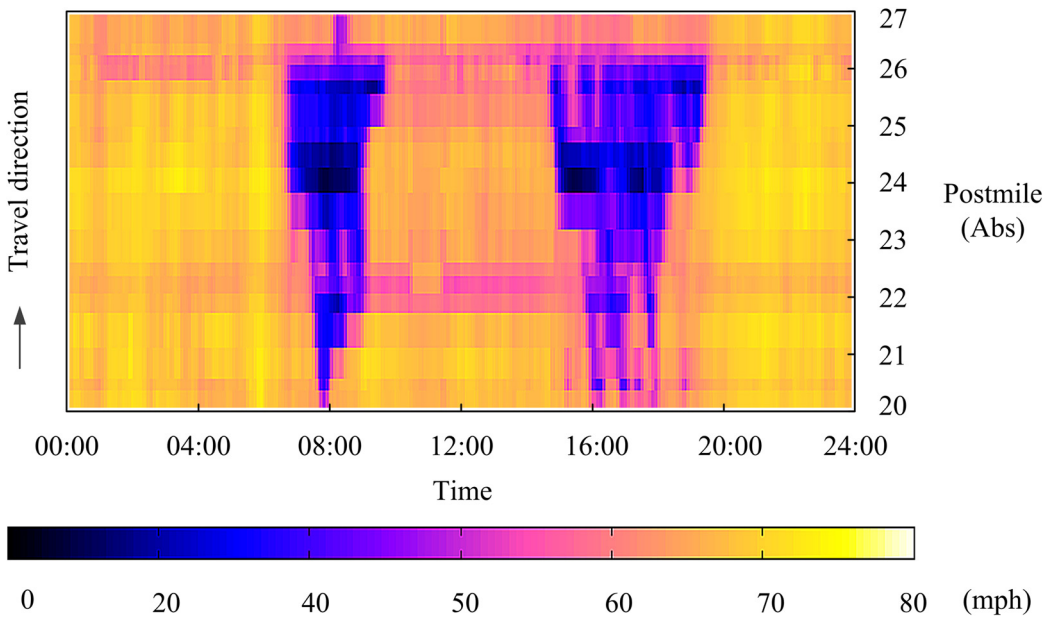

Fig. 2. Speed contour for the study segment (mph): Wednesday, October 24th, 2007; time and space are shown on the *x*- and *y*-axes, respectively, and the color scale represents the estimated time-mean speeds during the day according to the legend in the figure.

The speed contour plot in Fig. 2 illustrates the typical weekday traffic conditions observed at the study site. The bottleneck for the northbound section near milepost 26 typically activates each weekday during both the morning and the afternoon peak periods. The resulting queues from the bottleneck typically increase in length and fully engulf all segments of the study site. Note that detectors 1–15 are all located upstream of the active bottleneck.

Collision data for the years 2006–2008 were obtained for the study site from the Statewide Integrated Traffic Records System (SWITRS). This database, which is maintained by the California Highway Patrol, archives detailed information on each reported collision, including its location to within 0.01 miles; its occurrence time, reportedly to with 1 min; and other relevant descriptive information about the collision. There were 471 rear-end collisions reported during the 3-year period within the study site as shown in Fig. 1.

3. Methodology

3.1. Occurrence condition of rear-end collision

The logic behind the proposed surrogate safety measure can be explained with the help of Fig. 3. Fig. 3(a) shows vehicle trajectories of two consecutive vehicles traveling at freely flowing speed, v_b. Times denoted as t_n and t_{n+1} indicate when the deceleration wave was encountered by the leading, n, (see the black box in Fig. 3(a)) and the following vehicles (see the gray box in Fig. 3(b)), $n+1$, respectively. Upon the arrival of the deceleration wave, the lead vehicle changes its speed from v_b to v_a within t_d (see Fig. 3(b)). If the following vehicle changes its speed within t_d from t_{n+1}, a rear-end collision can be avoided (see Fig. 3(c)). The variables shown in Fig. 3(a)–(c) are used to depict the condition for the occurrence of rear-end collision.

A rear-end collision occurs if:

$$d_a + d_{De} + G_n < d_{De} + d_b \tag{1}$$

where d_a = traveling distance of leading vehicle after speed drop, d_{De} = deceleration distance, G_n = spacing between two vehicles, measured between the rear of one vehicle and the front of the next, $G_n = H_n - L_n$, H_n = distance between consecutive vehicles, measured between corresponding points on the vehicles, L_n = length of vehicle n, d_b = traveling distance of following vehicle before speed drop, $d_a + d_{De} + H_n$ = location of leading vehicle rear at time t' with respect to location x, $d_{De} + d_b$ = location of following vehicle front at time t' with respect to location x.

Using the variables shown in Fig. 3(a)–(c), Eq. (1) can be written as below.

$$v_a \cdot (t_{n+1} - t_n) + \frac{(v_b)^2 - (v_a)^2}{2a} + G_n < v_b \cdot (t_{n+1} - t_n) + \frac{(v_b)^2 - (v_a)^2}{2a} \tag{2}$$

$$\Rightarrow \frac{G_n}{v_b - v_a} < t_{n+1} - t_n \tag{3}$$

where v_b is the speed before deceleration, v_a is the speed after deceleration, and a is the deceleration rate (assumed to be constant for all vehicles),

When the left side of the expression in Eq. (3) becomes smaller, the likelihood of a rear-end collision increases (and vice versa). This expression can be used to develop the surrogate safety measure if G_n, v_b to v_a of individual vehicles is available. However, such information is not readily available when the freeway is equipped only with conventional loop detectors. Thus, we used the relationship shown in Eq. (3) and replaced expressions with values that can be estimated using data reported from conventional loop detectors to develop the surrogate safety measure, as explained in Section 3.2.

3.2. Estimating collision risk from detector data

Fig. 3(d) shows the time-space diagram of vehicle trajectories during propagation of a kinematic wave. Suppose the kinematic wave reaches downstream (see dotted line labeled D) at time T_1, and upstream (see dotted line labeled U) at time T_2, respectively. When the traveling time of the backward moving waves between two consecutive detectors are about the same or are longer than the detector data reporting period (which is 30-s at our study site), the sum of the spacing among N vehicles can be roughly estimated using data that are readily available from the loop detectors. Although this estimate can substantially deviate from the true value, it provides adequate information for the purpose of

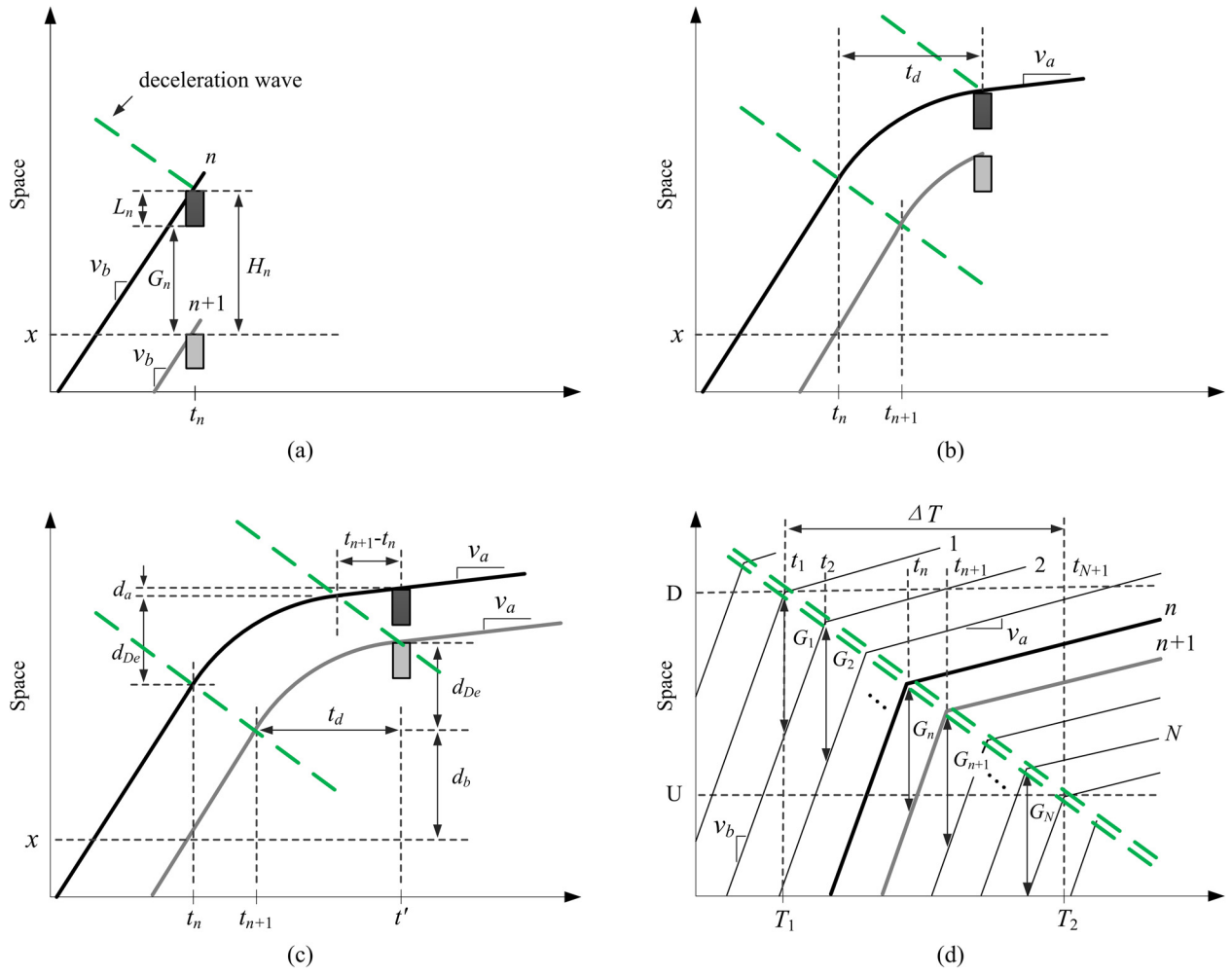

Fig. 3. (a) Arrival of deceleration; (b) description of deceleration trajectories of two vehicles; (c) completion of deceleration; and (d) time-space diagram of vehicle trajectories during propagation of kinematic wave.

developing the surrogate safety measure since the magnitude of the estimate will proportionally vary with the true value.

Eq. (5) is obtained by summing up the relationship shown in Eq. (3) for all N vehicles that reached kinematic wave between T_1 and T_2 shown in Fig. 3(d)

$$\frac{G_1}{v_b - v_a} + \frac{G_2}{v_b - v_a} + \cdots + \frac{G_N}{v_b - v_a} < (t_2 - t_1) + (t_3 - t_2) + \cdots + (t_{N+1} - t_N) \tag{4}$$

$$\Rightarrow \sum_{i=1}^{N} G_i \frac{1}{v_b - v_a} < \sum_{i=1}^{N} (t_{i+1} - t_i) \tag{5}$$

where $\sum_{i=1}^{N} t_{i+1} - t_i$ is the aggregated time period, $\sum_{i=1}^{N} G_i$ is the sum of spacing between consecutive vehicles when kinematic wave reaches the following vehicle within time (T_1, T_2).

Notice how the average value of spacing among vehicles between U and D is reduced as the deceleration wave propagates upstream. The maximum average spacing occurs at time T_1 and the minimum average spacing occurs at time T_2 (see Fig. 3(d)).

The increase in the right side of the expression in Eq. (5) would indicate the propagation of a slow backward moving kinematic wave and the decrease in the left side of the equation would indicate a shorter gap. Assuming the traffic conditions between U and D at time T_1 are homogeneous, the sum of individual information for N vehicles in Eq. (5) can be estimated with the aggregated detector data collected at location U and D. The rear-end collision risk index (RCRI) shown in Eq. (6) is derived from the Eq. (5) as a surrogate safety measure for evaluating the risks of rear-end collisions. The step-by-step deviation of the surrogate measure RCRI is presented in detail in Appendix A.

$$\text{RCRI} = \frac{[\bar{V}_U(t - \Delta T,\ t) - \bar{V}_D(t - \Delta T,\ t)] \cdot \bar{O}_U(t - \Delta T,\ t)}{1 - \bar{O}_U(t - \Delta T,\ t)} \tag{6}$$

$$\bar{V}_U(t - \Delta T, t) = \frac{\sum_{m=1}^{M} \sum_{j=1}^{J} V_U^m(t_j - \Delta t, t_j)}{M \cdot J} \tag{7}$$

$$\bar{V}_D(t - \Delta T, t) = \frac{\sum_{m=1}^{M} \sum_{j=1}^{J} V_D^m(t_j - \Delta t, t_j)}{M \cdot J} \tag{8}$$

$$\bar{O}_U(t - \Delta T, t) = \frac{\sum_{m=1}^{M} \sum_{j=1}^{J} O_U^m(t_j - \Delta t, t_j)}{M \cdot J} \tag{9}$$

where $V_U^m(t_j - \Delta t, t_j)$ is the speed in lane m at upstream detector during time interval Δt, $V_D^m(t_j - \Delta t, t_j)$ is the speed in lane m at downstream detector during time interval Δt, $O_U^m(t_j - \Delta t, t_j)$ is the occupancy in lane m at upstream detector during time interval Δt, $\bar{V}_U(t - \Delta T, t)$ is the average speed at upstream detector during time period ΔT, $\bar{V}_D(t - \Delta T, t)$ is the average speed at upstream detector during time period ΔT, $\bar{O}_U(t - \Delta T, t)$ is the average speed

at upstream detector during time period ΔT, J = number of time intervals in one time period ($J = \Delta T/\Delta t$, $\Delta t = 30$ s), M = number of traveling lanes. For some road sections with unequal lane numbers at upstream and downstream detector locations, the minimal lane number is considered as M.

The RCRI represents rear-end collision risk induced by kinematic waves. Higher values of RCRI would indicate higher risk of rear-end collisions; lower values of RCRI indicate lower likelihood of rear-end collision risk (including negative RCRI values). The speeds of kinematic waves were calculated by monitoring the changes of traffic states at detector locations upstream of the recurrent bottleneck (Mauch and Cassidy, 2002; Chung et al., 2007; Zheng et al., 2010). In our study site the wave travel time between two consecutive detectors upstream of the bottleneck ranges from 2.5 to 9 min and traffic data from a 5-min period, ΔT, were used for subsequent analysis.

3.3. Logistic regression model

The logistic regression model was calibrated in this study to quantify the likelihood of rear-end collisions. The model form is shown in Eq. (10): Y takes a value "1" if a rear-end collision occurs during ΔT, and "0" otherwise. $P(Y=1)$ denotes the probability of a rear-end collision occurring. The surrogate measure RCRI was derived assuming stationary conditions upstream and downstream of kinematic waves. However, the traffic situations could not be stationary in the aggregated time period ΔT. Thus, the measures of traffic variables and RCRI in Eq. (6) only reflect the average situation during ΔT and could contain system uncertainties or variations. It is assumed that a large variation in traffic could create more dangerous conditions and increase collision potential. As a consequence, in addition to the RCRI, the standard deviation of occupancy at upstream and downstream detector locations, which represent the variation in traffic, were also included in the logistic regression model.

The logistic regression model can be used to predict the likelihood of rear-end collision given real-time traffic data from loop detectors near recurrent bottleneck areas:

$$\begin{aligned}\text{Logit}(P(Y=1)) &= \log\frac{P(Y=1)}{1-P(Y=1)} \\ &= \beta_0 + \beta_1 \cdot RCRI + \beta_2 \cdot \sigma(O_U) + \beta_3 \cdot \sigma(O_D) \end{aligned} \quad (10)$$

$$\sigma(O_U) = \sqrt{\frac{1}{M \cdot J}\sum_{m=1}^{M}\sum_{j=1}^{J}\left(O_U^m(t_j - \Delta t, t_j) - \bar{O}_U(t - \Delta T, t)\right)^2} \quad (11)$$

$$\sigma(O_D) = \sqrt{\frac{1}{M \cdot J}\sum_{m=1}^{M}\sum_{j=1}^{J}\left(O_D^m(t_j - \Delta t, t_j) - \bar{O}_D(t - \Delta T, t)\right)^2} \quad (12)$$

where β_0, β_1, β_2, β_3 = coefficients of explanatory variables that need to be estimated, $\sigma(O_U)$, $\sigma(O_D)$ = standard deviation of occupancy at upstream and downstream detector locations.

4. Case–control design

A case–control study (Schlesselman and Stolley, 1982; Gross and Jovanis, 2007) was used to identify the hazardous traffic conditions that contribute to rear-end collisions. Samples were divided into a case group (i.e., traffic conditions that accompanied rear-end collisions) and a control group (i.e., traffic conditions that did not accompany rear-end collisions). The effects of confounding variables were controlled to better understand the relationship between rear-end collisions and traffic variables, and to prevent the effects of confounding variables from contaminating the analysis. The description of the case-control design used in the present study is provided in this section.

4.1. Case (collision)

Since our objective is to explore the influence of hazardous traffic conditions on rear-end collision likelihood, we excluded analysis of those collisions that could have been attributed to other prominent causes. These included collisions that occurred: during inclement weather (31 records), within on- or off-ramps (55 records), near work zones (4 records), with alcohol or drug involvement (9 records), or related to pedestrian, motorcycle, animal, fixed object or other objects (27 records). We also excluded several collisions that occurred during times when nearby detectors were malfunctioning.

After excluding the traffic collisions described above, a total of 341 rear-end collisions were used in the analysis. The samples for analysis contain 231 property damage only crashes, 86 complaint of pain injuries, 23 visible injuries, and 1 fatal crash. Among these collisions, 70.9% are due to unsafe speed, 18.9% are due to following too closely, and 7.5% are related to unsafe lane changes. A total of 95.9% of rear-end crashes occurred during daytime hours, and 74.9% of rear-end crashes occurred during peak hours.

The traffic data during the 5-min period prior to the collision were selected as the "cases" in our study. The collision times reported in SWITRS were further confirmed by evaluating traffic data from detectors spanning the collision location. The collision time reported in SWITRS was regarded as approximate time of collision. Then, the sudden and pronounced increase in upstream occupancy and the decrease in downstream occupancy (decrease in upstream speed and increase in downstream speed) shortly before or after the reported time of event in SWITRS were considered the actual occurrence time of a collision.

4.2. Control (no collision)

The traffic data from the control group represent normal traffic conditions in the study freeway section that did not result in any rear-end collisions. The confounding factors were controlled for the selection of traffic data in the control group. In previous studies, some of the confounding factors considered included the road geometry, speed limit, time of day, day of week, season, weather, and traffic conditions (Lee et al., 2002; Abdel-Aty et al., 2004; Pande and Abdel-Aty, 2006; Zheng et al., 2010).

Road geometry, speed limit, day of week, season, weather, and road surface condition were considered as confounding factors in the present study. The time of day was not considered as a confounding factor since the traffic conditions during the same time of the day are reproducible near the recurrent bottleneck. Assuming that the change of weather and road surface condition within a particular day is smaller than for different days, we selected non-crash data from the same day of crashes for the control group to reduce the confounding impacts of the two factors. After controlling for the confounding variables, four samples of 5-min aggregated traffic data from the detector where the data for the case were obtained were randomly selected to be the control group. The candidate time periods for the control were entire day when the traffic collision was occurred except the 5-min period used for the case. A control-to-case ratio of 4:1 was implemented for this study since the statistical power generally does not increase significantly beyond a 4:1 ratio (Schlesselman and Stolley, 1982). The model was evaluated via repeated model developments using different samples of control. The findings are presented in Section 5.

5. Results and discussion

The data from years 2006 and 2007 were used to estimate the parameters in the proposed model (see Eq. (10)) and the data from year 2008 was used for model validation. Since the traffic data in the control group may have been biased as a result of road construction and incidents, it is essential to evaluate the consistency of the statistical model prior to the model interpretation (Zheng et al., 2010). The proposed model was evaluated by re-sampling controls for each case and repeating the model development processes based on the newly drawn samples.

Ten different sets of controls were sampled from the poll of candidate controls and were used for the model development, resulting in a control-to-case ratio of 4:1. Table 1 shows the modeling results from 10 different runs. The first and second columns show the run and variables. The third column shows the odds ratio associated with each predictor. The odds are defined as the probability of a collision occurring divided by the probability of the collision not occurring, i.e., $P(Y=1)/(1-P(Y=1))$. The odds ratio for a contributing variable is defined as the relative amount by which the odds of the outcome (probability of rear-end collision) increase or decrease when the value of the explanatory variable is increased by 1 unit.

The results indicate that for all runs, the RCRI and standard deviation of upstream and downstream occupancy are statistically significant at a 95% confidence level, and their odds ratios were consistent throughout these runs. The average odds ratio of RCRI is 1.211, and of standard deviation of upstream and downstream occupancy is 1.195 and 1.187, respectively. This indicates that an additional unit increase in RCRI increases the odds of rear-end collision occurrence by an average of 21.1%, a unit increase in standard deviation of upstream occupancy increases the odds by 19.5%, and a unit increase in standard deviation of downstream occupancy increases the odds by 18.7%. The increase of RCRI and standard deviation of upstream and downstream occupancy increase the likelihood of rear-end collisions. Those estimates are consistent with intuition since a large RCRI represents a higher rear-end collision risk (see Eq. (6)) and a large variation in occupancy represents more volatile traffic conditions (i.e., large changes in speeds) in a freeway section.

Based on the average estimates of coefficients in the logistic regression models, the likelihood of a rear-end collision under a particular traffic condition can be calculated as

$$P(Y=1)=\frac{\exp\left(-3.095+0.191\cdot RCRI+0.178\cdot\sigma(O_U)+0.172\cdot\sigma(O_D)\right)}{1+\exp\left(-3.095+0.191\cdot RCRI+0.178\cdot\sigma(O_U)+0.172\cdot\sigma(O_D)\right)} \quad (13)$$

Graphical representation of the proposed surrogate safety measure RCRI and the collision likelihood are shown in Fig. 4. Fig. 4(a) depicts the flow-occupancy relation observed at the study site. Fig. 4(b) shows RCRI using Eq. (6). Fig. 4(c) and (d) together shows how collision likelihood changes with respect to the changes in occupancies and their standard deviations.

In estimating RCRI for each cell in Fig. 4(b), only the average values of speed and occupancies from the upstream and downstream locations were used. The traffic conditions at upstream and downstream locations were assumed to be stationary (i.e., the traffic state did not change from free-flow to congested or vice versa). As discussed in Section 3.2 of this paper, the purpose of RCRI was to monitor the changes in magnitude of the expressions shown in Eq. (5). RCRI alone would not capture the effect of variances in occupancy in rear-end collision likelihood. To this end, we developed a logistic regression model to estimate the rear-end collision likelihood as shown in Eq. (13).

Fig. 4(c) shows the estimated collision likelihood assuming constant value of 5 for both $\sigma(O_U)$ and $\sigma(O_D)$. The darker areas in the Fig. 4(c) represent higher rear-end collision likelihoods and the sum of the likelihoods in Fig. 4(c) in all the grids equals 1. The model predicts the lowest likelihood when the upstream traffic is more congested than the downstream (i.e., upstream reports higher occupancy) or traffic conditions at both locations are comparable (i.e., similar occupancy).

In general, increased RCRI and collision likelihood are observed when downstream conditions are more congested than upstream conditions – since approaching vehicles are forced to reduce speed to avoid rear-end collisions. The highest RCRI and collision likelihood are predicted to occur when upstream traffic is at its capacity state (i.e., occupancy near 15) while the downstream is congested

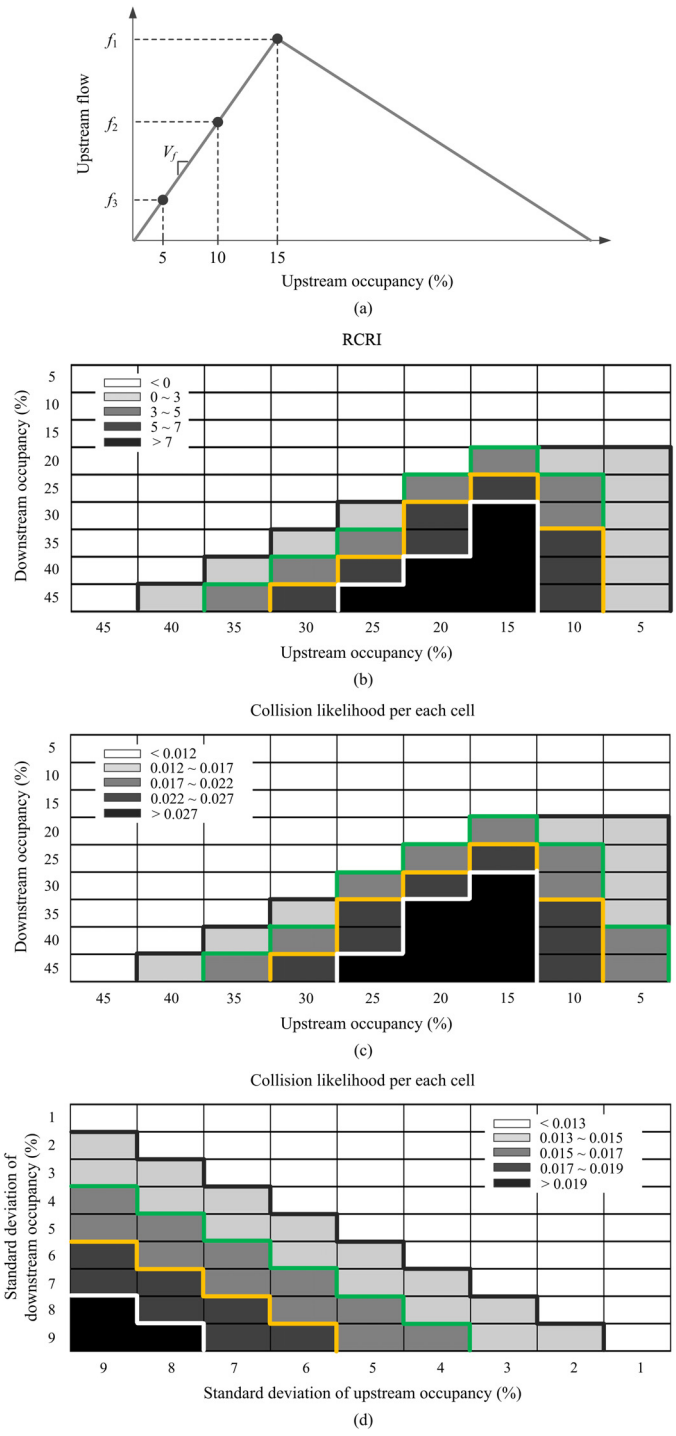

Fig. 4. (a) Fundamental diagram of traffic at upstream location; (b) sensitivity analyses for RCRI with occupancy combinations; (c) sensitivity analyses for collision likelihood with occupancy combinations; and (d) sensitivity analyses for collision likelihood with combinations of standard deviation of occupancy.

Table 1
Results from logistic regression model (4:1 control-to-case ratio).

Run	Variable	Odds ratio	Std. err.	Z	$p>Z$	95% confidence interval		χ^2
1	RCRI	1.184	0.072	2.790	0.005	1.051	1.333	294.71
	$\sigma(O_U)$	1.226	0.046	5.400	<0.001	1.138	1.320	
	$\sigma(O_D)$	1.165	0.037	4.820	<0.001	1.095	1.240	
2	RCRI	1.228	0.072	3.520	<0.001	1.095	1.378	296.95
	$\sigma(O_U)$	1.161	0.042	4.160	<0.001	1.082	1.245	
	$\sigma(O_D)$	1.202	0.038	5.890	<0.001	1.131	1.279	
3	RCRI	1.256	0.079	3.620	<0.001	1.110	1.422	323.12
	$\sigma(O_U)$	1.266	0.052	5.740	<0.001	1.168	1.372	
	$\sigma(O_D)$	1.160	0.040	4.270	<0.001	1.083	1.241	
4	RCRI	1.203	0.074	2.990	0.003	1.066	1.358	274.94
	$\sigma(O_U)$	1.152	0.040	4.090	<0.001	1.077	1.233	
	$\sigma(O_D)$	1.183	0.037	5.340	<0.001	1.112	1.259	
5	RCRI	1.343	0.091	4.360	<0.001	1.177	1.534	324.69
	$\sigma(O_U)$	1.284	0.052	6.170	<0.001	1.186	1.390	
	$\sigma(O_D)$	1.128	0.038	3.570	<0.001	1.056	1.205	
6	RCRI	1.219	0.076	3.190	0.001	1.080	1.377	315.54
	$\sigma(O_U)$	1.249	0.050	5.590	<0.001	1.155	1.350	
	$\sigma(O_D)$	1.178	0.040	4.860	<0.001	1.103	1.259	
7	RCRI	1.153	0.056	2.930	0.003	1.048	1.268	275.60
	$\sigma(O_U)$	1.143	0.035	4.330	<0.001	1.076	1.215	
	$\sigma(O_D)$	1.209	0.034	6.780	<0.001	1.145	1.278	
8	RCRI	1.165	0.071	2.520	0.012	1.035	1.313	321.04
	$\sigma(O_U)$	1.198	0.046	4.750	<0.001	1.112	1.291	
	$\sigma(O_D)$	1.222	0.041	5.950	<0.001	1.144	1.306	
9	RCRI	1.180	0.069	2.840	0.004	1.053	1.323	262.26
	$\sigma(O_U)$	1.140	0.041	3.680	<0.001	1.063	1.223	
	$\sigma(O_D)$	1.188	0.037	5.500	<0.001	1.117	1.263	
10	RCRI	1.176	0.072	2.650	0.008	1.043	1.326	296.07
	$\sigma(O_U)$	1.127	0.042	3.210	0.001	1.048	1.213	
	$\sigma(O_D)$	1.235	0.043	6.100	<0.001	1.154	1.322	
Average	RCRI	1.211	0.073	3.141	0.004	1.076	1.363	298.49
	$\sigma(O_U)$	1.195	0.044	4.712	<0.001	1.111	1.285	
	$\sigma(O_D)$	1.187	0.039	5.308	<0.001	1.114	1.265	

(i.e., occupancy above 40). This is equivalent to traffic traveling at 50 mph or greater speed approaching the rear of a platoon of vehicles traveling at or less than 15 mph.

Following the cells horizontally from the left side of the figure, increases in RCRI and rear-end collision likelihood are shown as the upstream occupancy decreases until it reaches the cell representing occupancy 15–10. Then, there is a reduction in RCRI and collision likelihood with respect to the reductions in upstream occupancy even though the downstream traffic remains more congested. This decrease in RCRI and collision likelihood can be explained using the fundamental diagram shown in Fig. 4(a).

The black dots labeled f_1, f_2 and f_3 marks three different freely flowing traffic states. Traffic in all three of these states is traveling at speed V_f, but the corresponding occupancies are different. Thus, when traffic traveling in these three different states suddenly encounters slow moving traffic, the faster moving traffic must reduce its speed by the same amount but under different surrounding conditions for the three different states. The distance between the adjacent vehicles in traffic state f_2 will be greater than of f_1, thus, the rear-end traffic collision likelihood of f_2 will be less than f_1. Similarly, the rear-end traffic collision likelihood of f_3 will be less than f_2 (see Eq. (A1)–(A5) in Appendix A).

In constructing Fig. 4(d), the upstream and downstream occupancies were first set to be 15 and 40: this combination of occupancies corresponds to the highest collision likelihood estimated in Fig. 4(c). Then, the $\sigma(O_U)$ and $\sigma(O_D)$ were varied from 1 to 9 to evaluate how the collision likelihood further changes. The darker areas in the Fig. 4(d) represent higher rear-end collision likelihoods and the sum of the likelihoods in Fig. 4(d) in all the grids equals 1. The figure shows increase in the standard deviation of occupancy at either upstream or downstream location increases the rear-end collision likelihood.

To evaluate the validity of the proposed model, the estimated rear-end collision likelihood was compared with empirical data. Since our proposed model estimates the rear-end collision likelihood based on traffic conditions observed at two neighboring detector stations, the traffic conditions in 2008 were evaluated to count the frequency of different traffic states for 5-min periods. The results are shown in Fig. 5(a), which presents the percentage of each traffic condition occurring – the sum of all the percentages shown in Fig. 5(a) adds up to 1. Occurrences of different traffic states varied markedly. For example, in 2008, the traffic condition A_1 (see Fig. 5(a)), in which both the upstream and downstream occupancy in a section was between 5 and 10, occurred about 84 times per day, while traffic condition A_2 (see Fig. 5(a)), in which the upstream occupancy was between 10 and 15 while the downstream occupancy was between 30 and 35, occurred only 2.5 times. Thus, even if the likelihood of collision of A_1 is much less than A_2, taking the frequencies of each traffic state into account, the estimated magnitude of number of collisions for the cell representing A_1 could be greater than that for A_2 if the product of collision likelihood associated with traffic state A_1 and its frequency is greater than that of A_2. An example of such a case is shown in Fig. 5(b). A large number of collisions was estimated for a freely flowing traffic condition B_1, (see Fig. 5(b)) with its occupancy rate at less than 5. This traffic state corresponds to nighttime traffic conditions (i.e., from 23:00 to 5:00) with the lowest traffic demand. It does not, however, influence the application of our model as a surrogate safety measure for

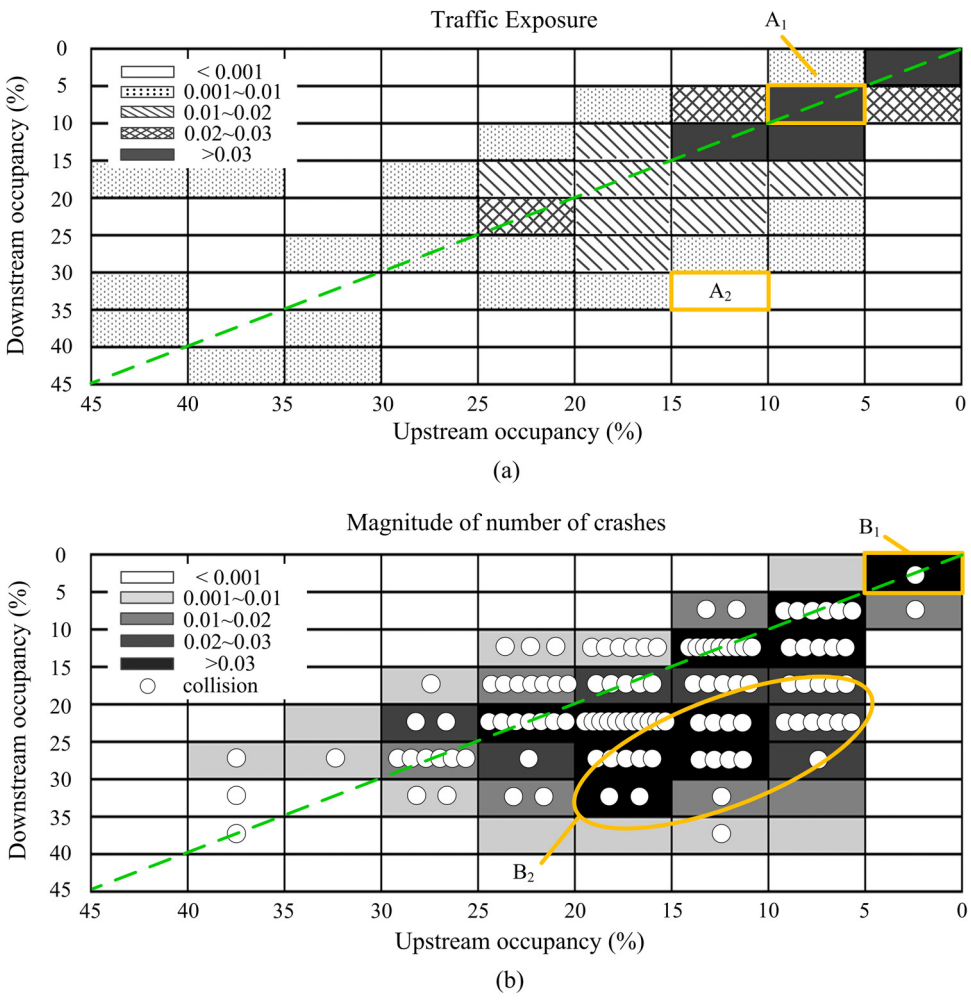

Fig. 5. (a) Traffic exposure map with occupancy combinations; and (b) collision map with occupancy combinations for model validation.

evaluating the rear-end collision risk near freeway recurrent bottlenecks, as is explained next.

The white circles in Fig. 5(b) represent the occupancy observed upstream and downstream of the collision location during the 5-min period preceding the collision. There were 116 rear-end collisions reported, and each circle marks the traffic condition observed at the upstream and downstream locations 5 min prior to the collision. In the figure, the color of each cell represents the magnitude of the number of rear-end collisions after accounting for the frequency of different traffic states, with darker regions representing a larger number of collisions, and lighter cells representing fewer collisions.

The findings shown in Fig. 5 indicate that predictions of rear-end collisions in the model are generally consistent with the observations, except for freely flowing traffic conditions with low demand (see B_1 in Fig. 5(b)). Most collisions occur when traffic is moderately congested with the upstream and downstream occupancy between 20 and 35, as shown in Fig. 5(b). This could be attributed to the kinematic waves frequently emanating from an active bottleneck to propagate toward upstream in congested traffic. Congested traffic also has relatively large exposure, increasing the number of collisions. Though the exposure of a condition in which upstream traffic is free-flowing and downstream traffic is congested, B_2 (see Fig. 5(b), which presents the traffic near tails of queues) is relatively small, our safety surrogate measure predicted a large number of collisions which was consistent with the observations. As expected, few collisions occur in the traffic state wherein speed at downstream is higher than upstream, and a large number of collisions are observed when the upstream speed is higher.

The proposed model would predict traffic collisions when the predicted probability of a crash exceeds a pre-specified threshold value which can be determined based on a false positive rate (i.e., predicting traffic collision rate when there was no traffic collision) that the operator is willing to accept. In general, the true positive rate (i.e., correctly predicting traffic collision) increases as the false positive rate increases. The true positive rate of previous real-time crash prediction models varies from 50% to 75% when the false positive rate is set between 20% and 30% (see a summary of predictive performance of previous models in Xu et al. (2013)). In the model proposed in our study, the true positive rate is 71.2% if the false positive rate is set at 20%; and the true positive rate is 84.6% if the false positive rate is set at30%.

The model developed in this study can be used to assist in the development of dynamic traffic control measures such as variable speed limits to reduce the rear-end collision risk near freeway bottlenecks. By monitoring the real-time traffic data obtained from loop detectors, the risk of rear-end collisions can be calculated using the model proposed in this paper. When the estimated collision likelihood exceeds a pre-specified threshold, the traffic control can be initiated to reduce the speeds of vehicles that are approaching the risky kinematic waves. Reducing vehicle speeds in a timely manner is expected to reduce the risk of rear-end collisions near freeway recurrent bottlenecks.

6. Conclusions

The proposed model estimates the risk of rear-end collisions related to kinematic waves near freeway recurrent bottleneck areas. By aggregating the trajectories of vehicles that reach a kinematic wave during the time that the wave passes from the downstream to upstream, the rear-end collision risk index (RCRI) was developed for the proposed model using data that are readily available from conventional loop detectors. Findings indicate that a unit increase in RCRI results in increasing the odds of a rear-end collision by 21.1%. An additional unit increase in standard deviation of occupancy at upstream and downstream locations increases the odds by 19.5% and 18.7% respectively.

The likelihood of rear-end collisions is highest when traffic approaching from upstream is at capacity state while traffic downstream is highly congested. The propagation of kinematic waves increases the rear-end collision potential. The proposed model was used to predict rear-end collision risk along the 6-mile freeway section study site. The estimated number of collisions based on the model was compared with observed traffic collision data from the year 2008. The estimated rear-end collisions are consistent with the observations.

The important parameters affecting the collision risk identified in our proposed model and previous studies are the same although they differ in approach – the proposed model is grounded in traffic flow theory principles as opposed to statistical data mining. The speed difference between upstream and downstream locations and the average occupancy were identified as contributing factors in Abdel-Aty et al. (2004), Abdel-Aty and Pande (2005), Pande and Abdel-Aty (2006), Hassan and Abdel-Aty (2011), and Xu et al. (2013). The standard deviation of occupancy was reported as a contributing factor in Pande and Abdel-Aty (2006), Hassan and Abdel-Aty (2011), and Hossain and Muromachi (2011).

Since the proposed model can be used to evaluate the rear-end collision risk using 5-min aggregated data, the model can be useful in developing real-time traffic control strategies for reducing rear-end traffic collisions, such as ramp metering and variable speed limits. The proposed model did not consider the effect of kinematic waves on lane change behavior. Incorporating lane change behavior and developing a variable speed limit control strategy using the proposed model are subjects of future study.

Acknowledgements

This work was partially supported by National Key Basic Research Program (NKBRP) of China (No. 2012CB725400), as well as the National Research Foundation of Korea grant funded by the government of Korea (MEST) (NRF-2011-0029443). Authors also thank Professor Pan Liu at the Southeast University of China.

Appendix A. Derivation of surrogate safety measure RCRI

Here we present the step-by-step derivation of the surrogate safety measure RCRI for the evaluation of rear-end collision risk related to kinematic waves. The relationship among density (K), occupancy (O), and average vehicle length ($\bar{L}$) for homogenous traffic conditions is illustrated in Daganzo (1997), as $K = O/\bar{L}$. Then left term of Eq. (5) can be written as

$$\sum_{i=1}^{N} G_i \frac{1}{v_b - v_a} = \sum_{i=1}^{N} (H_i - L_i)\frac{1}{v_b - v_a} = \left\{\sum_{i=1}^{N} H_i - \sum_{i=1}^{N} L_i\right\} \cdot \left\{\frac{1}{v_b - v_a}\right\} = \left\{\sum_{i=1}^{N} \frac{1}{\bar{K}} - N \cdot \bar{L}\right\} \cdot \left\{\frac{1}{v_b - v_a}\right\}$$

$$= \left\{\sum_{i=1}^{N} \left(\frac{\bar{L}}{\bar{O}}\right) - N \cdot \bar{L}\right\} \cdot \left\{\frac{1}{v_b - v_a}\right\} = \left\{\frac{N \cdot \bar{L}}{\bar{O}} - N \cdot \bar{L}\right\} \cdot \left\{\frac{1}{v_b - v_a}\right\} = N \cdot \bar{L} \cdot \left\{\frac{1 - \bar{O}}{\bar{O}}\right\} \cdot \left\{\frac{1}{v_b - v_a}\right\} \quad \text{(A1)}$$

where $\bar{K}$ is the average density (number of vehicles occupying a road lane per unit of length), $\bar{O}$ is the average occupancy (proportion of time that the detector is occupied).

As shown in Fig. 3(d), the variables in the right side of Eq. (A1) can be replaced by the aggregated detector data, which is:

$$N \cdot \bar{L} \cdot \left\{\frac{1 - \bar{O}}{\bar{O}}\right\} \cdot \left\{\frac{1}{v_b - v_a}\right\} = N(T_1,\ T_2) \cdot \bar{L} \cdot \left\{\frac{1 - O_U(T_1,\ T_2)}{O_U(T_1,\ T_2)}\right\} \cdot \left\{\frac{1}{V_U(T_1,\ T_2) - V_D(T_1,\ T_2)}\right\}$$

$$= \frac{N(T_1,\ T_2) \cdot \bar{L} \cdot (1 - O_U(T_1,\ T_2))}{O_U(T_1,\ T_2) \cdot (V_U(T_1,\ T_2) - V_D(T_1,\ T_2))} \quad \text{(A2)}$$

where $N(T_1, T_2)$ = number of vehicles that reach kinematic wave within time (T_1, T_2), $O_U(T_1, T_2)$ = average occupancy at upstream location within time (T_1, T_2), $V_U(T_1, T_2)$ = average speed at upstream location within time (T_1, T_2), $V_D(T_1, T_2)$ = average speed at downstream location within time (T_1, T_2).

Right term of Eq. (5) can be written as

$$\sum_{i=1}^{N}(t_{i+1} - t_i) = t_2 - t_1 + t_3 - t_2 + \cdots + t_{N+1} - t_N$$

$$= t_{N+1} - t_1 = T_2 - T_1 = \Delta T \quad \text{(A3)}$$

where ΔT is the time period that a wave propagating from downstream detector to upstream detector.

Then Eq. (A4) can be obtained after rewriting Eq. (5) with Eqs. (A2) and (A3)

$$\frac{N(T_2 - \Delta T,\ T_2) \cdot \bar{L} \cdot (1 - O_U(T_2 - \Delta T,\ T_2))}{O_U(T_2 - \Delta T,\ T_2) \cdot (V_U(T_2 - \Delta T,\ T_2) - V_D(T_2 - \Delta T,\ T_2))} < \Delta T \quad \text{(A4)}$$

Eq. (A4) shows the occurrence condition of rear-end collisions during the propagation of kinematic wave from downstream to upstream detector locations. However, Eq. (A4) contains information, such as the number of vehicles in the section and the average vehicle length, which cannot be obtained easily from loop detector data. In addition, the wave propagation time ΔT between two consecutive loop detectors varies over time and across locations. Thus, it is difficult to calculate the exact value of the left and right sides of Eq. (A4). However, the potential of occurrence of rear-end collision could be inferred from the Eq. (A4): during the propagation of the kinematic wave, for a detector aggregation period, an increase in the expression in the left side of Eq. (A4) decreases the rear-end collision potential, while a decrease in its value increases the rear-end collision potential.

When the arrival time of kinematic wave at detector stations does not coincide with the detector aggregation period, the expression in the left side of Eq. (A4) can be underestimated. However, as illustrated earlier, the results of our analysis remain unchanged: the objective is to develop the surrogate safety measure, not to obtain the exact value. Notice that Eq. (A4) is valid only when $V_U > V_D$. A reciprocal transformation was made to Eq. (A4) to obtain a consistent collision risk function in the interval $[-\infty, +\infty]$. Average vehicle length $\bar{L}$ is assumed to be fixed and is not considered in the

evaluation of rear-end collision risk. The expression in the left side of Eq. (A4) shows the sum of rear-end collision risk for N vehicles that reach the kinematic wave within (T_1, T_2). Thus the rear-end collision risk for a vehicle that reaches a kinematic wave can be estimated by

$$\frac{O_U(T_2 - \Delta T,\ T_2) \cdot (V_U(T_2 - \Delta T,\ T_2) - V_D(T_2 - \Delta T,\ T_2))}{1 - O_U(T_2 - \Delta T,\ T_2)} \quad \text{(A5)}$$

The RCRI for a freeway section at time t is obtained by replacing $V_U(T_2 - \Delta T, T_2)$, $V_D(T_2 - \Delta T, T_2)$ and $O_U(T_2 - \Delta T, T_2)$ in Eq. (A5) with Eqs. (7)–(9).

References

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M.F., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case–control logistic regression. Transportation Research Record 1897, 88–95.

Abdel-Aty, M., Uddin, N., Pande, A., 2005. Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways. Transportation Research Record 1908, 51–58.

Abdel-Aty, M., Pande, A., 2005. Identifying crash propensity using specific traffic speed conditions. Journal of Safety Research 36 (1), 97–108.

Abdel-Aty, M., Hassan, H., Ahmed, M., Al-Ghamdi, A., 2012. Real-time prediction of visibility-related crashes. Transportation Research Part C: Emerging Technologies 24, 198–288.

Chung, K., Rudjanakanoknad, J., Cassidy, M.J., 2007. Relation between traffic density and capacity drop at three freeway bottlenecks. Transportation Research Part B: Methodological 41, 82–95.

Chung, K., Jang, K., Oum, S., Kim, Y.H., Song, K.H., 2011. Investigation of attributes of kinematic waves preceding traffic collisions. In: The 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations. Pergamon Press, Oxford.

Golob, T.F., Recker, W.W., 2003. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. Journal of Transportation Engineering 129 (4), 342–353.

Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Freeway safety as a function of traffic flow. Accident Analysis and Prevention 36 (6), 933–946.

Gross, F., Jovanis, P.P., 2007. Estimation of the safety effectiveness of lane and shoulder width: case–control approach. Journal of Transportation Engineering 133 (6), 362–369.

Hassan, H., Abdel-Aty, M., 2011. Exploring visibility-related crashes on freeways based on real-time traffic flow data. In: The 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Hourdos, J., Garg, V., Michalopoulos, P., Davis, G., 2006. Real-time detection of crash-prone conditions at freeway high-crash locations. Transportation Research Record 1968, 83–91.

Hossain, M., Muromachi, Y., 2011. Understanding crash mechanisms and selecting interventions to mitigate real-time hazards on urban expressways. Transportation Research Record 2213, 53–62.

Kockelman, K.M., Ma, J., 2007. Freeway speeds and speed variations preceding crashes, within and across lanes. Journal of the Transportation Research Forum 46 (1), 43–62.

Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Transportation Research Record 1784, 1–8.

Lee, C., Hellinga, B., Saccomanno, F., 2003. Real-time-crash prediction model for application to crash prevention in freeway traffic. Transportation Research Record 1840, 67–77.

Li, Z., Chung, K., Cassidy, M.J., 2013. Collisions in freeway traffic: the influence of downstream queues and interim means to address it. In: The 92th Annual Meeting of the Transportation Research Board, Washington, DC.

Mauch, M., Cassidy, M.J., 2002. Freeway traffic oscillations: observations and predictions. In: Proceedings of the fifteenth International Symposium on Transportation and Traffic Theory, pp. 653–674.

Oh, C., Kim, T., 2010. Estimation of rear-end crash potential using vehicle trajectory data. Accident Analysis and Prevention 42 (6), 1888–1893.

Oh, C., Oh, J., Min, J., 2009. Real-time detection of hazardous traffic events on freeways methodology and prototypical implementation. Transportation Research Record 2129, 35–44.

Oh, C., Park, S., Ritchie, S.G., 2006. A method for identifying rear-end collision risks using inductive loop detectors. Accident Analysis and Prevention 38 (2), 295–301.

Ozbay, K., Yang, H., Bartin, B., Mudigonda, S., 2008. Derivation and validation of new simulation-based surrogate safety measure. Transportation Research Record 2083, 105–113.

Pande, A., Abdel-Aty, M., 2006. Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. Transportation Research Record 1953, 31–40.

Saccomanno, F., Cunto, F., Guido, G., Vitale, A., 2008. Comparing safety at signalized intersections and roundabouts using simulated rear-end conflicts. Transportation Research Record 2078, 90–95.

Schlesselman, J.J., Stolley, P.D., 1982. Case–Control Studies: Design, Conduct, Analysis. Oxford University Press, Oxford, United Kingdom.

Xu, C., Liu, P., Wang, W., Li, Z., 2012. Evaluation of the impacts of traffic states on crash risks on freeways. Accident Analysis and Prevention 47 (1), 162–171.

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

Yeo, H., Jang, K., Skabardonis, A., 2010. Impact of traffic states on freeway collision frequency. In: The 89th Annual Meeting of the Transportation Research Board, Washington, DC.

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