# Methodological Evolution and Frontiers of Identifying, Modeling and Preventing Secondary Crashes on Highways

Hong Yang [1*], Zhenyu Wang [2], Kun Xie [3], Kaan Ozbay [4], Marianna Imprialou [5]

[1] Assistant Professor, Department of Modeling, Simulation & Visualization Engineering,
Old Dominion University, 4700 Elkhorn Ave, Norfolk, VA 23529, USA
E-mail: hyang@odu.edu; Tel: (757) 683-4529
[*] *Corresponding Author*

[2] Graduate Research Assistant, Department of Modeling, Simulation & Visualization Engineering,
Old Dominion University, 4700 Elkhorn Ave, Norfolk, VA 23529, USA
E-mail: zwang002@odu.edu; Tel: (757) 637-8868

[3] Lecturer, Department of Civil and Natural Resources Engineering, University of Canterbury
20 Kirkwood Ave, Christchurch 8041, New Zealand
E-mail: kun.xie@canterbury.ac.nz; Tel: +64-3-369-2707

[4] Professor & Director, C2SMART Center (A Tier 1 USDOT UTC),
Department of Civil and Urban Engineering; Center for Urban Science and Progress (CUSP)
Tandon School of Engineering, New York University (NYU)
Six MetroTech Center, 4th Floor (RM 404), Brooklyn, NY 11201, USA
E-mail: kaan.ozbay@nyu.edu; Tel: (646) 997-0552

[5] Lecturer (Assistant Professor), Transport Studies Group,
School of Architecture, Civil & Building Engineering,
Loughborough University, Loughborough LE11 3TU, United Kingdom
E-mail: m.imprialou@lboro.ac.uk; Tel: +44 (0)1509 223422

**ABSTRACT**

Secondary crashes (SCs) or crashes that occur within the boundaries of the impact area of prior, primary crashes are one of the incident types that frequently affect highway traffic operations and safety. Existing studies have made great efforts to explore the underlying mechanisms of SCs and relevant methodologies have been evolving over the last two decades concerning the identification, modeling, and prevention of these crashes. So far there is a lack of a detailed examination on the progress, lessons, and potential opportunities regarding existing achievements in SC-related studies. This paper provides a comprehensive investigation of the state-of-the-art approaches, examines their strengths and weaknesses, and provides guidance in exploiting new directions in SC-related research aiming to support researchers and practitioners in understanding well-established approaches so as to further explore the frontiers. Published research focused on SCs since 1997 has been identified, reviewed, and summarized. Key issues concentrated on the following aspects are discussed: (i) static/dynamic approaches to identify SCs; (ii) parametric/non-parametric models to analyze SC risk, and (iii) deployable countermeasures to prevent SCs. Based on the examined issues, needs, and challenges, this paper further provides insights into potential opportunities such as: (a) fusing data from multiple sources for SC identification, (b) using advanced learning algorithms for real-time SC analysis, and (c) deploying connected vehicles for SC prevention in future research. This paper contributes to the research community by providing a one-stop reference for research on secondary crashes.

**KEYWORDS**

Secondary crash, primary crash, road safety, sensor data, crash impact area, highway

## 1. INTRODUCTION

Traffic crashes are the most frequent incidents on highways and the ones with the most severe consequences. Statistically, about 6.3 million highway crashes are reported annually only in the United States, among which more than 32,000 are fatal crashes (NHTSA 2016). These incidents often pose challenging problems in traffic operations and safety. Both transportation agencies and the general public are concerned about their notable direct and indirect impacts. It has been estimated that these highway crashes resulted in almost $1 trillion in economic loss and societal harm in 2010 (Blincoe 2015). The hazardous traffic conditions that are formed due to traffic crashes are often exposing non-involved vehicles and incident responders to a risk of additional crashes: the so-called secondary crashes (SCs). SCs are typically defined as crashes that occur within the spatial and temporal boundaries of the impact area that is formed due to earlier primary crashes (PCs) (Owens *et al.* 2010). This should be distinguished from the "secondary collisions" defined in Xie *et al.* (2018) that are described as different phases of a single crash event. It has been reported that SCs can account for as high as 20% of all crashes and 18% of all fatalities on the United States' freeways (Owens 2010). Considering the significant economic and social costs as well as the potential preventability, SC mitigation has become a priority for transportation agencies around the world.

In fact, many transportation agencies are using SCs as an important indicator to monitor the safety performance of their systems. The frequency of SCs is used as a key factor in assessing a number of safety programs of the Federal Highway Administration (FHWA) and many state/local agencies consider the determination and reduction of SCs in allocating funding for the development of their traffic incident management (TIM) programs (Yang *et al.* 2017b). For example, Arizona Department of Public Safety (AZPDS) used SCs in the agency's strategic plan and launched a specific program for the prevention of SCs (TIM 2017). In addition, the prevention of SCs has been taken into consideration in the planning of the freeway service patrol (FSP) programs in Florida, California, etc. (Lou *et al.* 2011). The success of these programs greatly relies on the knowledge of the inherent mechanisms of when, where and how SCs occur.

SCs have been identified as a serious issue in early Road Safety literature (e.g. Owens (1978)) but during the last two decades tremendous efforts have been made to investigate their characteristics and prevention methods. Researchers developed a variety of methodological approaches and ideas that attempt to address the three fundamental questions regarding SCs: (a) identification; (b) modeling; and (c) prevention leading to substantial progress in the current understanding of these events. This largely supported the exploration of directions for policies and countermeasures that ultimately aim to reduce the risk of SCs. However, there lacks a comprehensive review of the state-of-the-art regarding the progress and lessons learned from existing SC-related research. A thorough investigation on the methodological evolution of identifying, modeling, and preventing SCs can timely support practitioners and researchers in deploying, improving, and extending many of current achievements.

This paper aims to provide a detailed review of contemporary thinking on the SC-related issues and to show how methodological approaches have evolved over time in order to address the problems and explore promising future research directions. To fulfill such goals, this paper starts with the investigation of the state-of-the-art practices regarding the identification, prediction, and prevention of SCs. Following there is a critical synthesis of issues, needs, and challenges regarding current approaches, including suggestions on potential opportunities and future research directions.

## 2. STATE OF THE ART

This state-of-the-art review examines the existing literature on SCs focusing on work published during the last two decades. The databases and search engines used include: Google Scholar, Scopus, Web of Science, and Transportation Research Board's Transport Research International Documentation (TRID), the largest online

bibliographic database of transportation research. The keywords used were “Secondary Crash”, “Secondary Incident”, “Secondary Accident”, and “Secondary Collision” and only reports in English were included. A number of early studies refer to SCs providing statistics about the frequency of their occurrence; for instance Owens (1978) suggested that in their on-the-spot study on a motorway section in the UK 13 out of 40 (32.5%) crashes were found to have occurred as a result of the traffic conditions caused by primary incidents. Fontaine (1995) in an analysis on pedestrian crashes proposes potential mechanisms that may lead to SCs. However, studies that perform more in-depth examinations on SCs start emerging from 1997 onwards and that is why this literature review focuses on this period. Thirty-five relevant references reported in recent two decades have been selected and examined in this paper. The three main thematic areas of this review are the identification of SCs (34 references), modeling and predicting SC risk (19 references), and prevention of SCs (6 references), as presented in Sects. 2.1, 2.2, and 2.3, respectively.

### 2.1. Methodologies for SC Identification

In order to prevent SCs effectively it is necessary to be aware of the frequency of their occurrence. Although crash reports include a number of attributes to describe crashes, in most of the crash reporting systems crashes are not classified as primary or secondary. This could be because this characterization is not always straightforward for police officers who typically complete these documents and may arrive quite a few minutes after a crash occurrence. As a result, SCs need to be identified by research teams by post processing crash and traffic data that show the evolution of the impact (or influence) areas caused by primary crashes. Providing that accurate spatial and temporal crash information is available, if a crash is found to have happened within the influence area of another crash then it is characterized as secondary. A summary of current studies on the identification of SCs is presented in Table 1. Information about authors, major types of data, data facts, identification approaches, and identification criteria are provided. Existing approaches can be grouped into four main categories: (a) static spatial-temporal range-based; (b) queuing theory-based; (c) speed contour map-based; and (d) shockwave-based approaches. Each category is discussed below.

#### *2.1.1. Static spatial-temporal range-based approaches*

The most classical static approach for the identification of SCs defines fixed spatial-temporal thresholds to identify the pairs of PCs and SCs. Given the position and occurrence time $(t_p, s_p)$ of a PC, a crash $C$ is examined using the following criteria:

$$SC = \begin{cases} 1, & if\ [t_c \in (t_p, t_p + \Delta t)] \,\&\, [s_c \in (s_p, s_p + \Delta s)] \\ 0, & others \end{cases} \quad \textbf{(1)}$$

where, $(t_c, s_c)$ denotes the position and occurrence time of the crash $C$ that needs to be examined; $\Delta t$ and $\Delta s$ denote the temporal and spatial impact area (IA) of the PC, respectively; and value 1 means crash $C$ is identified as a SC and 0 if not. The performance of such a static approach mainly relies on the threshold values and their suitability for the study area. The original idea was introduced by Raub (1997), in which a SC has to be located no more than one mile upstream of a PC and has to occur within a period no longer than the PC clearance time + 15 minutes. This serves as the foundation for many subsequent studies that adopted some variations of spatial and temporal threshold values (Karlaftis *et al.* 1999; Latoski *et al.* 1999; Hirunyanitiwattana and Mattingly 2006; Zhan *et al.* 2008; Khattak *et al.* 2009; Jalayer *et al.* 2015; Tian *et al.* 2016). For example, Tian *et al.* (2016) introduced three types of spatial-temporal criteria to identify SCs on interstate highways of Florida: (a) 2 miles, 2 hours; (b) 2miles, clearance time + 15 minutes; and (c) 2 miles, clearance time + 30 minutes. Their identification results varied with different threshold values. It was found that the use of criteria (b) and (c) only identified less than half of SCs compared to (a).

As discussed in Sarker *et al.* (2015), depending on the facility type, traffic conditions, type and characteristics of an incident, rubbernecking has a great potential of inducing SCs not only upstream but also in the opposite

direction of a PC. Thus, SCs on the opposite direction were examined by some studies (Chang and Rochon 2009; Kopitch and Saphores 2011; Green *et al.* 2012; Sarker *et al.* 2015). For example, Chang and Rochon (2009) considered an IA of 2 hours and 2 miles to examine SCs in the same direction of traffic flow, whereas a 0.5-hour and 0.5-mile spatiotemporal range was considered in the opposite direction. Sarker *et al.* (2015) examined SCs occurred in both the upstream and downstream opposite direction of the PC. They have also conducted a sensitivity analysis of different temporal thresholds (30 min. to 300 min.) and spatial thresholds (0.5 mile to 5 miles) on the identification results and confirmed that the static approach can lead to both underestimation and overestimation, depending whether the selected thresholds are conservative.

Overall, static thresholds could be effective for a rough estimation of SCs in a specific study area due to their simplicity, however they are inelastic and thus error prone. Primary crash impact areas may vary significantly depending on weather, traffic conditions or time of the day so the likelihood of misidentification of SCs is significant. Moreover, their transferability is questionable; when static thresholds proposed by Raub (1997) and Moore *et al.* (2004) have been applied to a different study area the false positives (crashes that were characterized as SCs mistakenly) reached 75% and false negatives (SCs that were characterized as primary) were over 40% (Imprialou *et al.* 2014).

#### *2.1.2. Queuing model-based approaches*

To better capture the impact areas of PCs, queuing models have been developed (Sun and Chilukuri 2007; Zhan *et al.* 2009; Sun and Chilukuri 2010; Vlahogianni *et al.* 2010; Zhang and Khattak 2010; Vlahogianni *et al.* 2012; Imprialou *et al.* 2014). Fundamentally, these studies established statistical (quantitative) models to relate a set of variables to the queue length $Q_{(t,s)}$ that approximates the IA of a PC:

$$Q_{(t,s)} = f(\boldsymbol{X}) \tag{2}$$

where, $\boldsymbol{X} = \{x_1, ..., x_n\}$ denotes the vector of contributing factors that affect the queue length, for example, traffic arrival rate, departure rate, capacity of lanes, incident duration, speed information, to name but a few; and $f$ is the mapping function that can be in either simple linear form or more advanced model structures. Given the estimated queue length, crashes located in the estimated boundary of queue will be identified as SCs. For example, Zhan *et al.* (2009) proposed a simple linear equation that used arrival rate, diversion rate, highway capacity adjustment factor, number of lanes, full capacity, and departure rate to calculate the maximum queue length and queue dissipation time. Similarly, Zhang and Khattak (2010) calculated the influence area based on the deterministic queuing models. Later, Sun and Chilukuri (2010) proposed to use the third order polynomial equation to dynamically calculate the incident progression curve based on the time after PC occurrence. Occasionally, given the accessibility of dense traffic surveillance cameras, the queue measurements were directly observed to facilitate the identification of SCs in some studies such as Vlahogianni *et al.* (2010). Although queue-based approaches may offer a more accurate and dynamic representation of impact areas they largely depend on the number and quality of the available predictors. Considering that the factors affecting queue formation and dissipation may vary from case to case, it is likely that the impact areas predicted by these approaches might be inaccurate.

#### *2.1.3. Speed contour map-based approaches*

Empowered by various sensor technologies, several studies have developed the speed contour map-based approaches to dynamically identify SCs (Chung 2013; Yang *et al.* 2013b; Yang *et al.* 2014b; Dougald *et al.* 2016; Park and Haghani 2016a; b; Goodall 2017; Park *et al.* 2017). The main idea behind these approaches is to establish the speed contour (heat) map based on the speed measurements from various sensor measurements. The time-space diagram is split into grid cells based on certain time intervals (e.g., 5 min., 15 min, etc.) and milepost of sensor stations. In general, each cell is determined to be congested or not based on a criterion similar to this shown

below:

$$V_{(t,s)}^{b} = \begin{cases} 1, & if\ V_{(t,s)} < V_{(t,s)}^{r} \\ 0, & others \end{cases} \tag{3}$$

where, $V_{(t,s)}$ represents its current speed; $V_{(t,s)}^{b} = 1$ means that the cell is congested; $V_{(t,s)}^{b} = 0$ suggests that the cell is not congested; and $V_{(t,s)}^{r}$ depicts its reference speed obtained from historical data. Thus, the IA of a PC can be depicted using the congested cells following the occurrence of the PC. If another crash occurs within the congested cells area will be identified to be a SC. The key premise is the selection of the reference speed, $V_{(t,s)}^{r}$ and for that various approaches have been proposed based on historical sensor measurements. For example, Yang *et al.* (2013a) compared the user-defined percentile speed of historical speed measurements with current speed data from loop detectors to obtain the incident-induced impact area. This approach took recurrent congestion into account and introduced a user defined weighting coefficient to facilitate both conservative and aggressive mapping of the impact area. The binary speed contour plot was then drawn to help identify the pairs of PCs and SCs locating in the impact area. Likewise, many others (Chung 2013; Xu *et al.* 2016) used similar approaches to detect SCs., Dougald *et al.* (2016) and Goodall (2017) adopted the proposed approach in Yang *et al.* (2013a) by using an adjusted assumption on the continuity of the impact area shown on the binary speed contour map. Meanwhile, other studies (Park and Haghani 2016a; b; Park *et al.* 2017) assumed that speed measurements and coefficients of variation are related to features corresponding to the Gaussian distributed traffic pattern, and introduced the Bayesian structure equation model to capture the congested IA of a PC using the INRIX dataset.

#### *2.1.4. Shockwave-based approaches*

Recently, traffic flow theory has been employed for the identification of SCs in several recent studies (Zheng *et al.* 2014; Mishra *et al.* 2016; Wang *et al.* 2016; Sarker *et al.* 2017). In general, the IA of a PC is assumed to be a triangular shape in a spatial and temporal speed contour constituted by the backward forming and discharging shockwaves associated with the occurrence and clearance of the PC. The propagation speeds $\omega_f$ and $\omega_d$ of these two shockwaves are calculated based on the changes of traffic flow and density caused by the PC:

$$\begin{cases} \omega_f = \dfrac{q_{nor} - q_p}{k_{nor} - k_p} \\ \omega_d = \dfrac{q_{nor} - q_{sat}}{k_{nor} - k_{sat}} \end{cases} \tag{4}$$

where, $q_{nor}$, $q_p$, and $q_{sat}$ respectively denote the normal flow before a PC occurs, the flow when a PC occurs, and the flow under road saturation, $k_{nor}$, $k_p$, and $k_{sat}$ represent the density, accordingly. The IA of a PC is defined using the triangular area constituted by three vertexes: $(s_p, t_p)$, $(s_p, t_p + t_{duration})$, and $(s_1, t_1)$, where $(s_p, t_p)$ represents the starting milepost and time of the PC, $t_{duration}$ is the PC duration time, and $(s_1, t_1)$ denotes the milepost and time of the intersection point of the backward forming and discharging shockwaves. For example, Zheng *et al.* (2014) introduced the simple shockwave model with the queuing line and discharging line to describe the IA of a PC. Their model relied on several key assumptions, including that the monthly average hourly traffic volume can represent the traffic flow condition when a PC occurred; and queuing process is stable and constant. Later, with more enhanced assumptions, modified shockwave-based approaches have been proposed. For example, Vlahogianni *et al.* (2012) used the automatic tracking of moving traffic jams model (ASDA) that extracted both traffic information and incident information to define the shockwave-based IA. Sarker *et al.* (2015), Mishra *et al.* (2016), and Sarker *et al.* (2017) collected the clearance time that depends on crash type and severity, number of vehicles involved, number of lanes, availability of shoulder area, and others, to better estimate the density-flow curve of the simple shockwave model for single and bi-directional traffic. To account for the effect when incident

response crews arrive at the PC site, Wang *et al.* (2016) introduced the intermediate speed turning point, and thus generated two-piece back-of-queue shockwave lines to re-estimate the IA.

To further improve the identification performance, in a recent work by Yang *et al.* (2017b), a data-driven analysis framework for the identification of SCs was developed. Clustering methods were firstly introduced to automatically classify unlabeled data archived from probe vehicles, and then intelligent approaches including multi-stage approximation algorithm, genetic algorithm, and ant colony algorithm were developed to estimate the boundary of the IA of a PC to further support the automatic identification of SCs.

**Table 1. Summary of Studies on Identification of SCs**

| Authors | Major data needs | Data facts (Total/PC/SC; period; location) | Method | Identification criteria |
|---|---|---|---|---|
| (Karlaftis *et al.* 1999) | Incident | 741/ 257/ 257; 1992-1995; Borman Expressway, INDOT | Static | 1.5km;15min+clearance |
| (Hirunyanitiwattana and Mattingly 2006) | Incident | 354,161/ 15,442/ 15,442; 1999-2000; California highway | Static | 2 mile; 1 hour |
| (Latoski *et al.* 1999) | Incident | 8,986/ 689/ 689; 1996; Borman Expressway, INDOT | Static | 3 mile; 15min+clearance |
| (Moore *et al.* 2004) | Incident + loop detector | 84,684/ 192/197;1999; Los Angeles Freeway, California | Static | 2 mile; 2 hour |
| (Sun and Chilukuri 2007; 2010) | Incident | 5,514/ 397/ 397; 2003; I-70 and I-270 in Missouri | Dynamic | Incident progression curves |
| (Zhan *et al.* 2008) | Crash data | 7,903/ 352/ 413; 2005-2007; Florida I95, I75, and I595 | Static | 2mile; 15min+clearance |
| (Zhan *et al.* 2009) | Crash data+ SMART data | 7,903/ 221/ 255; 2005-2007; Florida I95, I75, and I595 | Dynamic | Cumulative arrival; departure traffic delay |
| (Khattak *et al.* 2009) | Incident | 38,086/ 736/ 764; 2006; Hampton road, Virginia | Static | 1 mile, duration of PC (+15 min if lane blocked) |
| (Vlahogianni *et al.* 2010) | Incident + monitor + sensor data | 1746/ 279/ 279; 2007-2008; Attica Tollway, Greece | Dynamic | Maximum queue length and queue duration |
| (Chou and Miller-Hooks 2009) | Incident + simulated traffic data | 693/ 27/ 27; 2007; I-693,New York | Dynamic | Simulated speed contour map |
| (Chang and Rochon 2009) | Incident | 19,309/ 702/ 702; 2010; CHART | Static | 2 hours + 2 miles; 0.5 hour +0.5 mile for opposite direction |
| (Kopitch and Saphores 2011) | Incident | 9,549/ 528/ 528; 2008; Orange county, CA | Static | 2 miles upstream and 2 hours |
| (Green *et al.* 2012) | Crash data | 9,330/ 362/ 362; 2009-2010; Kentucky's highway | Static | 80 minutes; 6,000 ft upstream and 1,000 ft downstream |
| (Khattak *et al.* 2012) | Incident | 37,934/ 736/ 764; 2006; Hampton road, Virginia | Dynamic | Segment code; 1 mile, PC duration (+15 min if lane blocked) |
| (Vlahogianni *et al.* 2012) | Incident + monitor + sensor data | 1,465/ 51/ 51; 2007-2010; Attica Tollway, Greece | Dynamic | Dynamic threshold by upstream loop detector using ASDA |
| (Chung 2013) | Crash + sensor data | 6,200/ 182/ 212; 2001-2002; Orange county, California | Dynamic | Dynamic crash impact area using speed contour map |
| (Yang *et al.* 2013b; Yang *et al.* 2014a; Yang *et al.* 2014b) | Crash + sensor data/virtual sensor data | 1,118/ 71/ 100; 2011; 27-mile highway, New Jersey 2011 | Dynamic | Representative speed contour map |
| (Zheng *et al.* 2014) | Crash + hourly volume data + detailed network | 7,034/ 67/ 79; 2010; 1,500-mile freeways in Wisconsin | Dynamic | Shockwave model |
| (Imprialou *et al.* 2014) | Incident + monitor + sensor data | 1,287/ 126/ 17~68; 2007-2009; Attica Tollway, Greece | Dynamic | ASDA, Real influence area method |
| (Jalayer *et al.* 2015) | Crash data | NA/ NA/ NA; 2010-2013; CARE in Alabama | Static | 2 miles; 2 hours |
| (Mishra *et al.* 2016; Sarker *et al.* 2017) | Crash data + lane specific traffic sensor data | 91,325/ 528/ 570; 2010-2012; Shelby county, Tennessee | Dynamic | Dynamic simple shockwave |
| (Wang *et al.* 2016) | Detailed crash data + loop data | 49,753/ 204/ 209; 2010-2012; interstate freeway, California | Dynamic | Spatio-temporal shockwave with 1 speed turning point |
| (Tian *et al.* 2016) | Incident + crash data | NA/ NA/ 326; 2010; Interstate highways, Florida | Static | 2 miles; 2 hours or 15/30 minutes + clearance |
| (Park and Haghani 2016a; b; Park *et al.* 2017) | Incident + INRIX data | 1,150/ 125/ 125; 2012-2013; INRIX data along I-695 corridor | Dynamic | Binary speed contour plot map |
| (Xu *et al.* 2016) | Crash + PEMS data | 8978/ 97/ 113; 2006-2010; I880 freeway, California | Dynamic | Speed contour plot map |
| (Yang *et al.* 2017b; Yang *et al.* 2018) | Crash + probe vehicle data | Simulated incidents and probe vehicle data | Dynamic | Clustered trajectories and optimized boundary of impact area |
| (Goodall 2017) | Incident + RITIS data | 2,466/ 340/ 340; 2014; RITIS on I-66 | Dynamic | Speed contour plot with incident timeline |

***Note***: CARE: crash analysis reporting environment; CHART: coordinated highways action response team; RITIS: Regional Integrated Transportation Information System; INDOT: Indiana Department of Transportation.

## 2.2. Modeling and Predicting SC Risk

A few studies have examined the underlying relationship between SC occurrences and different contributing factors. In general, both parametric and non-parametric models were sought to model the association of SCs with external conditions. With the modeled association, it is expected to provide transportation agencies more insightful information when developing countermeasures to mitigate SC risks. A summary of the modeling practices is provided in Table 2. Information regarding the authors, methods, considered variables, data used, and SC identification methods in each study were presented.

### *2.2.1. Parametric Approaches*

Existing studies have used several statistical models to analyze the risk of SC occurrence. Among these studies, the majority of them (e.g., Karlaftis *et al.* (1999); Zhan *et al.* (2008)) have adopted logistic regression models to characterize the dichotomous nature of secondary crash occurrence given the presence of a primary crash: occur or not occur. In general, the risk of having a SC with respect to a set of contributing factors can be represented as follows:

$$P(Y=1 \mid \boldsymbol{X}) = \frac{e^{(\alpha+\boldsymbol{X}\beta)}}{1+e^{(\alpha+\boldsymbol{X}\beta)}} = \frac{1}{1+e^{-(\alpha+\boldsymbol{X}\beta)}} \quad \textbf{(5)}$$

where, $Y=1$ denotes that a SC occurs; $\boldsymbol{X}$ represents the potential contributing factors; $\beta$ depicts the corresponding vector of coefficients; and $\alpha$ is the intercept parameter. By estimating and evaluating the coefficients, the impact of each contributing factor on the SC risk can be determined. For example, Karlaftis *et al.* (1999) reported that the clearance time of PC, season, type of vehicle involved, and lateral location of PC were the most influential variables with regard to SC occurrence. Zhan *et al.* (2008) analyzed SCs that occurred on three corridors in Florida and found that incident visibility and lane blockage durations of PC also significantly affected SC risk. Likewise, a variety of potential contributing factors such as weather information, AADT, traffic flow information, and road geometry information were also considered as the input of logistic regression models in recent studies (Zhan *et al.* 2009; Kopitch and Saphores 2011; Wang *et al.* 2016; Goodall 2017).

Several studies also extended conventional logistic regression models to analyze SC risk. Khattak *et al.* (2009; 2012) proposed a two-level hierarchical prediction approach to account for the issue of measurement error of incident duration information. Incident duration was first estimated using an ordinary least square regression model and then with a logistic regression model based on the estimated duration time and other factors such as weather, road information, and AADT. Second, regarding the infrequent nature of SCs, Yang *et al.* (2014b) introduced the rare-event logistic regression model to account for the bias of estimated coefficients associated with the contributing factors. Finally, in order to account for the heterogeneity caused by unobserved factors, Xu *et al.* (2016) developed a random effects logit model to link the likelihood of SC occurrence with the real-time traffic flow conditions, primary crash characteristics, environmental conditions, and geometric features.

Despite the simplicity of logistic regression models, they are incapable of modeling multiple SCs induced by an individual PC. In order to understand the mechanism of multiple occurrence of SCs, multinomial logit models and other generalized linear models were proposed. For example, Mishra *et al.* (2016) proposed the multinomial logit model as follows:

$$P(m) = \frac{e^{V_m}}{\sum_{k=1}^{M} e^{V_k}} \quad \textbf{(6)}$$

where, $V_m$ denotes the utility of event $m$. The relationship between the probability of one SC, two SCs and contributing factors were examined based on the dataset from the Tennessee Department of Transportation (TDOT). Later, Sarker *et al.* (2017) examined the same dataset using a Poisson model, negative binomial (NB)

model, NB model with heterogeneous dispersion, and NB model with heterogeneous dispersion and unobserved heterogeneity to predict the frequency of SCs (0 SC, 1 SC, 2 SCs, 3 SCs, and 4 or more SCs). Factors, including AADT, traffic composition, land use, number of lanes, right side shoulder width, posted speed limits, and ramp indicator, were found to be key variables effecting SC occurrences.

Other than the aforementioned studies, the proportional test and probit models were also used to examine the likelihood of SCs. For example, Hirunyanitiwattana and Mattingly (2006) identified the time of day and roadway classification to be the contributing factors of SC occurrence through the proportion test on a dataset collected from California highways. Instead of using the logit link function, Khattak *et al.* (2009) proposed three binary probit models to examine the interdependence between PC duration and SC occurrence. Specifically, these models considered three scenarios: (a) model with observed duration; (b) model with observed duration but no closure time; and (c) model with the estimated incident duration. Their findings showed that PC duration, AADT, and number of involved vehicles positively affect the likelihood of SCs.

#### *2.2.2. Nonparametric Approaches*

Several studies also used nonparametric models such as artificial neural networks and decision trees to model SC risk. On the one hand, Vlahogianni *et al.* (2010) developed a Bayesian neural network (BNN) and reported that queuing conditions and primary crash duration observed through the close-circuit television camera system (CCTV) were the most significant determinants. Later, Vlahogianni *et al.* (2012) developed a multi-layer perceptron neural network models to analyze contributing variables that affect SC likelihood. Regarding the issues of such “black-box” models, mutual information and partial derivatives were used to identify potential risk factors. Changes in speed and volume, number of blocked lanes, and percentage of trucks were found to be significant factors whereas rainfall intensity was found to be less influential. On the other hand, decision tree was developed to explore contributing factors based on the prediction results of artificial neural networks. For example, by identifying SCs based on the binary speed contour plot map using probe vehicles data, Park and his colleagues (Park and Haghani 2016a; Park *et al.* 2017) proposed the BNN approach and extracted rules to generate gradient based decision trees. The main effects of the factors that account for SC occurrences were shown based on the decision trees.

### 2.3. Research on SC Prevention

Other than the identification and modeling of SCs, very few studies focused on the prevention of SC occurrences. The primary countermeasures explored in existing studies include the deployment of the active traffic management using changeable or variable message signs (CMS or VMS) variable speed limit control (VSL) and connected vehicles (CVs). For example, Kopitch and Saphores (2011) verified the effectiveness of 11 CMS that provided real-time traffic information about incidents, work zones, congestion, speed limits ahead, and alerts in reducing SC risk. It was found that the effectiveness of CMS increased between 2 and 11.15 miles and decreased between 11.15 and 22.3 miles. Li *et al.* (2014) introduced the strategy of implementing variable speed limit with both weather and traffic flow information to mitigate SC risk. Two surrogate safety measures, including time exposed time-to-collision (TET) and time integrated time-to-collision (TIT), were found to be reduced by 40 to 50 percentage in a case study on I-880 in California during heavy rain conditions. Lately, Yang *et al.* (2017a) examined the impact of connected vehicles on improving the situational awareness of drivers to mitigate SC occurrences. SC risk, measured by the number of simulated conflicts, was found to be significantly reduced if the market penetration rate of CVs on a highway was relatively high in dense traffic conditions.

## Table 2. Summary of Existing Studies on Modeling Secondary Crash Risk

| Reference | Method | Considered variable (significant ones in *Italic* and bold) | Data facts (Total/PC/SC; period; & location) |
|---|---|---|---|
| (Karlaftis *et al.* 1999) | LR | ***Clearance time, season, type of vehicle involved, lateral location***, ramp/median, weekday, winter | 741/257/257;1992-1995; Borman Expressway, INDOT |
| (Hirunyanitiwattana and Mattingly 2006) | Proportional test | ***time of day, roadway classification***, primary collision factors, severity level, and type of accident | 354,161/15,442/15,442;1999-2000; California highway |
| (Zhan *et al.* 2008) | LR | ***number of vehicles; number of lanes; PC duration; time-of-day of PC;*** vehicle rollover; lane closure; environment | 7,903/352/413;2005-2007; Florida I95, I75, and I595 |
| (Zhan *et al.* 2009) | LR | ***primary incident type; primary incident lane blockage duration; time of day; roadway direction***; vehicle type, environment | 7,903/221/255;2005-2007; Florida I95, I75, and I595 |
| (Khattak *et al.* 2009) | Probit models | ***Duration;*** Detection source; Incident type; Response vehicles; AADT; ramp; peak hours; vehicles involved; | 38,086/736/764; 2006; Hampton road, Virginia |
| (Vlahogianni *et al.* 2010) | BNN | ***maximum queue length; duration of queue; PC duration***; Peak time; number of vehicles; distance; vehicle type involved; location of crash; | 1746/279/279;2007-2008; Attica Tollway, Greece |
| (Kopitch and Saphores 2011) | LR | ***Number of vehicles; number of trucks; CMS; road work project***; visibility; precipitation; distance; time of the day; day of the week; | 9549/528/528;2008; Orange county, CA |
| (Khattak *et al.* 2012) | LR (hierarchal) | ***Incident duration; number of vehicles; AADT; peak***; lane closure; right shoulder; ramp; bad weather; detection type; location; emergency medical services | 37,934/736/764;2006; Hampton road, Virginia |
| (Vlahogianni *et al.* 2012) | MLP | ***heavy vehicles involved; travel speed; hourly volume; rainfall***; ***number of blocked lanes***; ***upstream***/downstream ***geometry***; duration; collision type; number of vehicles***;*** alignment; | 1465/51/51;2007-2010; Attica Tollway, Greece |
| (Yang *et al.* 2013b; Yang *et al.* 2014b) | LR (rare event) | ***duration; winter; rear rend;*** Time period; severity; work zone; weekend; lane closure; truck involved | 1,118/71/100; 2011; 27-mile highway, New Jersey 2011 |
| (Mishra *et al.* 2016; Sarker *et al.* 2017) | Linear probability model, LR, ML, and GORP | ***right shoulder width; Speed limit; AADT; type of median;*** segment length; truck traffic; single unit truck; multi-unit truck; PM peak; AM peak; urban land use | 91,325/528/570;2010-2012; Shelby county, Tennessee |
| (Wang *et al.* 2016) | LR | ***shockwave 1,2, and 3 speed; accident process duration; unsafe speed;*** Crash severity; violation category; weather; tow away; road surface; lighting; parties involved; volume; duration | 49,753/204/209;2010-2012; interstate freeway, California |
| (Park and Haghani 2016a; Park *et al.* 2017) | BNN, GBDT; logit model | ***location area; incident type; time of day***; Number of lanes; traffic operation center; number of vehicles; truck involvement; response delay type; severity (guardrail/ramp/normal); require firefighter | 1,150/125/125;2012-2013; INRIX data along I-695 corridor |
| (Xu *et al.* 2016) | LR (random effect) | ***Severity; Collison type; occurrence date;*** sideswipe; mean/std/cov of speed, occupancy, and count; difference between adjacent lanes on speed/occupancy/count; peak; weather; lighting; road surface; lane; width; median width; curve | 8978/97/113;2006-2010; I880 freeway, California |
| (Goodall 2017) | LR | ***Congestion; incident duration***; number of vehicles encounters the incident or its queue | 2,466/340/340;2014; RITIS, I-66 |

***Note****: LR: Logistic regression; ML: multinomial logit model; GORP: generalized response probit framework; BNN: Bayesian neural network; MLP: Multiple layer perceptron network; GBDT: gradient based decision tree; RITIS: Regional Integrated Transportation Information System*

Other than the aforementioned countermeasures, some studies also examined the benefits of service patrol programs in reducing SCs. For example, Karlaftis *et al.* (1999) examined the effect of the Hoosier Helper service patrol program on the Broman Expressway in Indiana. It was found that the program may help reduce SC likelihood by 18.5 percent in winter and by 36.3 percent in other seasons per crash assisted. The delay savings and crash cost savings from secondary crash reduction was $568,080 in 1995 that was 1.38 times of the service patrol program cost. Although there was no quantitative assessment, some other studies also mention the use of service patrol programs as a helpful countermeasure to reduce SC risk. For example, Khattak *et al.* (2012) suggested the improvement of coverage of service patrols and towing service on highway chokepoints that have higher SC occurrence probability.

Mitigation of post-crash impacts of SCs rather than prevention of SCs, has also been discussed by some researchers. Compared with previous studies that only used PC information, Park *et al.* (2016) considered the evolution of PCs and SCs over time to identify an appropriate location for emergency response units. Linear programming approach with relaxed integrality constraint for integer variables was verified to be valid in reducing the expected total delay of crashes in a numerical study with data collected on the highway I-695.

## 3. SYNTHESIS OF ISSUES, NEEDS AND CHALLENGES

### 3.1. Identification of Secondary Crashes

Effective identification is the foundation of all studies on SCs. Its key component is the analysis of the IA of PCs. Despite the progress in existing studies, the depiction of an appropriate IA remains a challenging task that has direct impact to the quality of outputs. Figure 1 demonstrates a hypothetical primary crash A, a secondary crash B and an non-secondary crash (NSC) C and three potential IAs for crash A ( $\Delta t_a, \Delta s_a$ , $\Delta t_b, \Delta s_b$ , and $\Delta t_c, \Delta s_c$ ). The first and the latter IAs could lead to the underestimation or overestimation of SCs but so far it is unclear whether this misidentification is always avoidable. Many research efforts have explored the possible solutions, however, there still exist several challenging issues associated with current approaches that are discussed below.

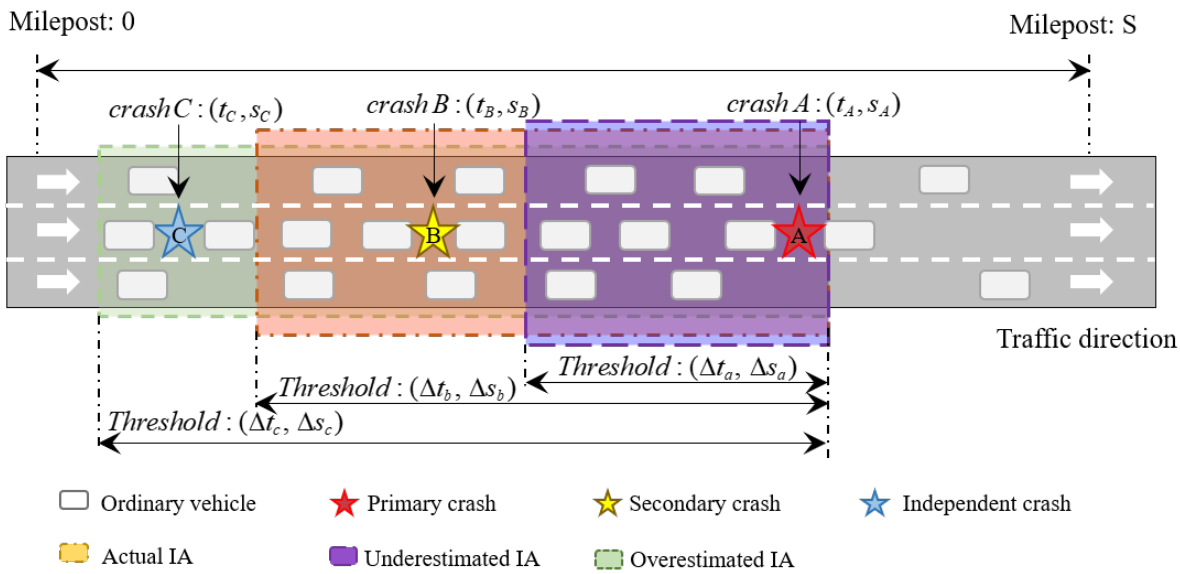


**Figure 1. Conceptual demonstration of a PC, SC, and NSC.**

Theoretically, secondary crashes do not only occur on highways, but also on arterials and other types of roadways. Existing studies are mainly focused on SCs occurred highways. This is mainly due to the fact that relevant data for traffic state estimation are often difficult to obtain from arterials. In addition, an objective defitintion of a crash IA on arterials is more complex because of the cofounded effects associated with the queue

caused by intersections and traffic controls.

### *3.1.1. Quantifying issues regarding the impact area of PCs*

The review of existing research has shown that there is no consistent criterion on how to define the spatio-temporal threshold values used in the static approaches. These threshold values are critical in representing the IA of a PC. However, current selections can significantly affect the identification results and are often too subjective. For example, according to Tian *et al.* (2016), less than 50% of actual SCs would be detected by just reducing the temporal threshold. Consequently, a fixed spatio-temporal threshold applied to all scenarios with various traffic conditions, roadway geometry, weather, etc. is highly prone to two issues as shown in Figure 1: (a) overestimation – the thresholds in the green box are too large and the independent crash C is incorrectly identified to be a SC; and (b) underestimation – the thresholds in the purple box are too small to correctly include crash B as a SC.

Regarding the queuing model-based approaches that intend to dynamically estimate the IA, the major issues are the models' reliability and simplified assumptions. Firstly, current assumptions are often too simple to reflect actual traffic conditions. For example, Zhang and Khattak (2010) assumed that the spatial impact of a PC only exists within its incident duration period. This assumption can be easily violated if the arrival rate of upstream traffic exceeds the downstream departure rate after the clearance of the PC. Meanwhile, each crash scenario is subject to a unique queueing process in equation (2) because of different traffic patterns, road geometry, incident characteristics, etc. The reliability of the selected queuing model $f$ is critical. Using an unsuitable queuing model can fail to capture the actual IA of a PC. For example, the purple triangle area in Figure 2 cannot identify crash B as a SC. Existing studies have developed a number of queuing models, but it is difficult to argue one is better than the others because most of their model practices are limited by the available data. In addition, it is impractical to build many queuing models for each segment.

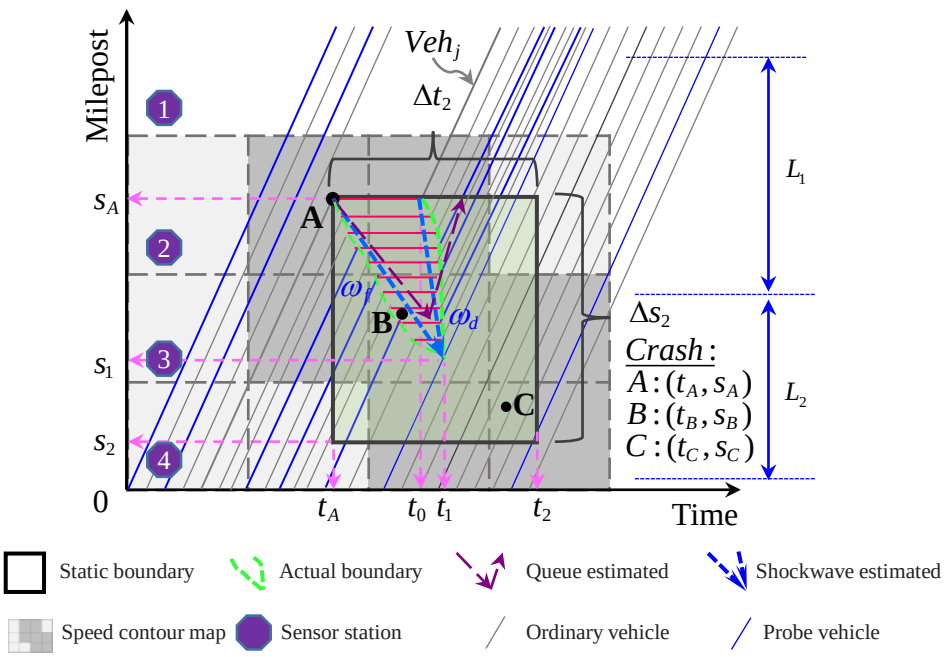


**Figure 2. Illustration of current modeling issues.**

On the other hand, the reference speed is the key premise of speed contour map-based approaches but there still remains gaps in appropriately defining it. Some research proposed to use the speed reduction factor to classify congestion pattern and non-congestion pattern, but the fixed factor cannot be easily transferred to different

time periods and highway segments. A small reduction factor suggests an aggressive threshold to define congestion, however, there lacks the examination of its impact on the identification performance. As shown in Figure 2, the darker cells represent the congested area. Since no lighter cells exist between crash A and C (indicating a significant speed increase between the two events), crash C is incorrectly identified to be a SC. Besides, more complex models such as the GMM have been proposed to estimate the reference speed, but the computational complexity is too high. Finally, there might exist blank cells due to the incomplete (or insufficient) sensor station data and thus greatly affects the identification performance. As an improvement, Yang *et al.* (2014a) introduced virtual sensors to collect crowdsourcing data from third parties such as the Bing map. Nevertheless, extensive comparison between the actual sensor stations and the virtual sensors is necessary. Also, the large-scale deployment of the virtual sensors requires time-consuming labeling work and is subject to the license agreement with the data providers.

Last but not least, the shockwave-based approaches can dynamically estimate the impact area of a PC given the changes of traffic flow and density. Existing studies all used simplified models similar to equation (4) to compute the shockwaves due to the PC but the simple triangular area enclosed by the two shockwaves cannot well depict the dynamic progression of the traffic states. As shown in Figure 2, the deterministic blue triangle area is drawn based on the back-of-queue and front-of-queue propagation waves. Constant shockwave propagation speeds were assumed in references (Mishra *et al.* 2016; Wang *et al.* 2016; Sarker *et al.* 2017) other than additional unrealistic assumptions such as constant arrival and discharging volumes. In fact, many factors such as the proportion of trucks, road geometry, weather, and the non-uniform arrival and discharging flow rates can result in non-constant shockwave propagation speeds and form a non-triangular IA (e.g., the area enclosed by green curves in Figure 2). Meanwhile, current approaches only examined PCs that induced a queue, and thus the scenario that a PC occurs within an existing queue cannot be easily tackled. As the density of the traffic is often difficult to be measured, it also requires the analysts to first estimate the density in order to estimate the shockwaves. This would accumulate errors in implementing the shockwave-based approaches.

To further illustrate the limitations of current studies on identifying SCs in actual transportation scenarios, the speed heat map based on actual sensor stations were shown. The case study of crashes occurred on the Interstate highway I-50 in California is shown in Figure 3. Each cell depicts the speed information of sensor stations in 5-min intervals. It should be noted that sensors are not uniformly distributed (66 sensors distributed along the 44.7-mile highway section on the westbound, with an average interval of 0.68 mile; and 67 sensors distributed with an average interval of 0.67 mile on the eastbound). A yellow or red cell denotes a short segment with the relatively low speed and implies potential congestion, whereas a green cell represents high-speed conditions. The blue dots represent the spatio-temporal position of the crashes and the cyan lines show the corresponding incident duration. Since no correct label was provided, two crashes located in the same congested region are denoted as the pairs of PCs and SCs by simply checking their relative positions. There are several notable issues. First, the missing data led to the large empty area shown in Figure 3(a). The availability and accessibility of sensor station data cannot be easily warranted. In addition, the gaps between sensor stations are not uniformly and densely distributed. For example, adjacent sensor stations (e.g., the upper part of Figure 3(b)) can have a large gap of about 5 miles that will result in incorrect representation of the speed variation along the segment between the two sensor stations. Second, another issue is that the IA of a PC is irregular and cannot be simply approximated by using the triangular shape obtained through the shockwave based approach or the queueing based approach as shown in Figure 3(a). Third, multiple SCs induced by one individual PC need to be examined as shown in Figure 3(b). Last but not least, some cells inside the IA are green while its surroundings are yellow or red cells. This requires additional imputation to exclude errors associated with the speed contour map caused by missing data or incorrect measurements. In summary, more research endeavor is expected to improve

current approaches for identifying SCs with a high performance in practices.

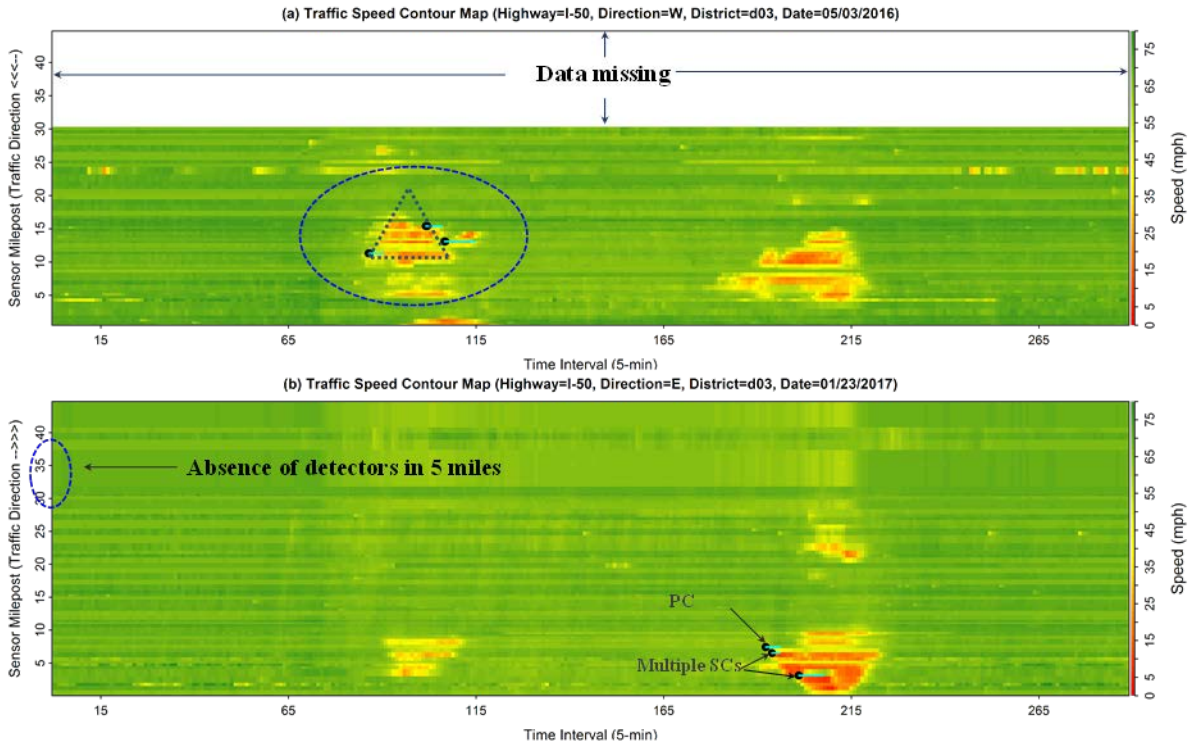


**Figure 3. Case study on the identification of SCs on the highway I-50.**

### *3.1.2. Data related issues and challenges in identification*

Data related issues also significantly affect the performance of SC identification. The major data related issues can be grouped into the following categories: (a)crash and/or incident data issues; and (b) traffic data issues associated with sensors and/or probe vehicles.

Incident related data have been frequently used in identifying SCs (Zhan *et al.* 2009; Sun and Chilukuri 2010; Vlahogianni *et al.* 2010; Zhang and Khattak 2010; Vlahogianni *et al.* 2012). It should be noted that both crash and incident data are prone to issues such as measurement error and difficulty in real-time archiving. Critical information such as incident duration and severity is often difficult to be timely and accurately obtained. As shown in Figure 3, measurements such as incident duration, number of lanes, and reduced capacity are often needed to estimate the incident induced queues. Additionally, crash occurrence time and location are two of the most frequently misreported attributes in crash databases (Imprialou and Quddus 2017). If any of these measurements is not available or accurate, the estimated IA can be biased, resulting in incorrect identification of SCs. Likewise, incident information will also affect the estimation of shockwaves that need to consider the responses of rescue crews, the evolution of lane closures, etc. Thus, precise and real-time incident data restricts the use of many existing approaches.

**Table 3. Incident data in queuing models**

| Authors | Data | Model |
|---|---|---|
| Sun and Chilukuri (2010) | Time after PC occurrence | Three-order polynomial equation |
| Zhan *et al.* (2009) | HCM factor, reduced capacity, full capacity, number of lanes, arrival rate, departure rate, duration | Linear equation |
| Zhang and Khattak (2010) | Arrival rate, departure rate, reduced capacity, normal capacity | Linear equation |
| Vlahogianni *et al.* (2010) | CCTV observation, arrival rate, departure rate | Observation |

Other than crash and incident data, sensor data have also been frequently applied in many recent studies (Yang *et al.* 2013b; Dougald *et al.* 2016; Park *et al.* 2017) but are also restricted to their availability, accessibility, and precision. First, not all highway sections are instrumented with traffic sensors and/or continually archive sensor measurements. However, the performance of speed contour map-based approach greatly relies on the reference speed for classifying congested and non-congested periods of each segment. This requires large amounts of historical traffic data to obtain reliable reference speed and deal with recurrent congestion scenarios. The lack of historical data can significantly limit the application of the speed contour map-based approach for identifying SCs. On the other hand, many of current highways only have sparse-distributed sensors due to the high costs associated with the installation and maintenance. As shown in Figure 2, the gap between sensor stations 1 and 2 can be very large and using the speed measurement from either sensor may not accurately capture the speed when crash A occurs at the middle of the segment. Similarly, accurate estimation of the propagation of shockwaves requires very sensitive measurements of the traffic state changes in terms of flow and density that cannot be easily acquired reliably in current sensor deployment practices. The high-density (e.g. 0.1-mile space) installation of sensor stations on all highways is impractical.

Compared with sensor stations, probe vehicle data (measured by devices that use GPS, Wi-Fi, and or Bluetooth) can provide more detailed and spatiotemporally disaggregated information of individual vehicles but low penetration rate remains an issue. As shown in Figure 2, the blue lines represent the trajectory of probe vehicles and the grey lines denote the trajectory of ordinary vehicles (that are often unobservable). Few studies experimentally studied the probe vehicle data from INRIX that were aggregated by highway segments (Park and Haghani 2016a; b; Park *et al.* 2017). However, such measurements cannot perfectly capture the prevailing traffic conditions, especially when the PC (e.g., A) occurs in the middle of the link $L_1$ or $L_2$ (see example in Figure 2). The real-time probe data of individual vehicles can significantly improve the estimation performance of shockwave as well as speed contour map-based approaches. Nevertheless, it takes time to accumulate an acceptable market penetration rate and such solutions are also subject to the relatively high cost of data acquisition and processing.

#### *3.1.3. Verification of SCs*

As it has been discussed in section 2.1, there are different approaches that can be used for SCs identification. A key issue is also the further verification of the identified SCs. Up to date, there are no well-established approaches in the literature for the verification of the identification performance. Nonetheless, two viable methods can be considered: (a) manual verification by reviewing police reports; and (b) reviewing surveillance camera recordings. There are still challenges in widely implementing these approaches firstly due to limited data availability. Additionally, even police reports and video recordings exist they are often difficult to be obtained due to privacy concerns and/or restrictions. Moreover, manually reviewing of police reports and/or video recordings can be time consuming. Although the police narrative is an important resource for verifying SCs, it should be also mentioned

that many crash databases do not contain it nor do transportation agencies' guidelines require the inclusion of them in crash database (Montella *et al.* 2013). The unavailability of the electronically documented narratives in a crash database creates additional challenges in reviewing, i.e., it might be difficult to read and interpret illegible handwriting. In literature, based on the detailed police reports, Zheng *et al.* (2015) have attempted to use police narratives to verify the identified SCs. Text was extracted from police narratives based on optical character recognition and then relationship keywords and event keywords were collected to assess the likelihood of a secondary crash. However, the success of the proposed automatic text mining procedure relies on an appropriate threshold to eliminate false negatives while minimizing false positivesthus, more sophisticated language processing techniques for complex police narratives were proposed for future studies.

An early study by Fries *et al.* (2007) has demonstrated the benefits and costs of traffic camera recordings in incident detection and verification. Theoretically, if the studied roadway is under constant surveillance through cameras, one can playback the archived videos in verifying secondary crashes. With the videos, and the time and location as well as the queuing process of the crashes can be accurately extracted (Vlahogianni *et al.* 2010). However, to link two crashes as a pair of primary and secondary crashes, still requires analysts to subjectively determine whether the later crash was related to the interrupted traffic caused by the earlier crash. The level of interruption is often difficult to be distinguished visually, especially when there was no severe congestion. Additionally, it could be difficult to track the relationship between a pair of crashes that occurred far away from each other, in particular when no continuous video coverage along the road section was available.

### 3.2. Modeling and Analysis of SC Risk

Existing research on modeling SC risk have been conducted since 1999 and the major problems associated with the modeling practices are discussed below.

#### *3.2.1. Model specification issues*

Despite the simplicity, many parametric models cannot be used to predict multiple secondary crashes occurred in the presence of a PC. Existing practices mainly assumed that each PC can be paired with only one SC when assembling the data set for modeling. Naturally, models such as the logistic regression can be used to work with the assumed dichotomous outcomes. However, it has been frequently reported in literature that some PCs can induce more than one SC (e.g., Khattak *et al.* (2009); Mishra *et al.* (2016); Sarker *et al.* (2017)). Thus, the sequence of these SCs cannot be easily depicted in these parametric models. Although ordered and/or multinomial logit/probit models can be an alternative, they require modelers to define at least three categories of the outcomes, for example, no SC, one SC, and two or more SCs. In addition, users should be aware of the imbalance issue of these defined categorical outcomes. In particular, the higher-order category is expected to have a considerably small proportion in regard to the total number of crashes. Moreover, many PCs do not cause any SC. Therefore, any candidate model should take into account such imbalance issue and excessive none-event scenarios.

Another major issue is that current parametric modeling practices lack the validation component. Existing practices mainly focused on the explanation of risk factors contributing to SC occurrences. Most of them have made great efforts in calibrating the predictive models but few have specifically validated the developed models. The key findings regarding the contributing factors were only limited to specific datasets and may not have good transferability. In addition, there exists the risk of overfitting the models with the individual dataset. As a result, despite that some common factors such as longer incident durations were found to significantly affect SC occurrences (Hirunyanitiwattana and Mattingly 2006; Yang *et al.* 2014b), inconsistent findings frequently occurred among studies. For example, incident visibility was found to be the critical factor in Zhan *et al.* (2008), whereas it was not significant in Kopitch and Saphores (2011). Thus, it is strongly suggested that more validation tests should be sought to examine the predictive performance of the proposed models.

On the other hand, nonparametric models such as BNN have better predictive performance but the flip side of this is the somewhat limited explanatory capability of the contributing factors (Vlahogianni *et al.* 2010; Vlahogianni *et al.* 2012). These contributing factors are subject to model structures of black-box character. Due to the complex and nonlinear transfer and activation functions, no solid and theoretical interpretation of the relationships between the input risk factors and SC occurrences is available. Alternatively, several techniques including mutual information, partial derivatives, and decision tree have been used to enhance the capability to determine the contributing factors (Vlahogianni *et al.* 2010). Nevertheless, it should be noted that current nonparametric approaches are also subject to many of the issues associated with parametric approaches, for example, exclusion of multi-SC scenarios, imbalance of events and event-free scenarios, limited validation tests, etc. In addition, they often employ more complex model structures and the parameter calibration is computationally expensive. Currently, there lacks guidance on the selection of the model hyper-parameters such as the number of hidden layers, hidden units, and their initial weighting values. This means that the end user must design a suitable model tuning strategy by a costly trial-and-error analysis. Therefore, it is difficult to use them in real-time prediction of SC occurrences. The balance between the computational complexity and the prediction accuracy needs to be carefully addressed to facilitate the implementation of these models in practice.

Finally, model variable selection was not well considered in current SC modeling practices. Although many contributing factors can be considered as the input variables, simply including all factors may lead to multi-collinearity issues and can result in counterintuitive findings. Currently, only few studies have conducted the variable selection but the performance need to be further investigated. For example, Zhan *et al.* (2009) proposed a forward conditional criterion to add one best-fit variable at a time during the regression process. Nevertheless, correlations between the independent factors may still exist. In addition, since some key factors may not be available, latent variables should be considered in the modeling structures. For example, random effect models can be used to address the impact of unobserved variables and result in improved prediction performance compared with standard logit models (Xu *et al.* 2016). However, the model calibration is complex and cannot be easily transferred to other datasets. Guidance on the adoption and estimation of different variables in SC prediction models would be particularly useful for future analyses.

#### *3.2.2. Data related issues and challenges in modeling*

Despite of the aforementioned model specification issues, data related issues also significantly impedes the development of reliable models for predicting SC risks. The major data related issues can be grouped into the following categories: (a) offline data issues; (b) real-time data issues; and (c) obtaining and incorporating contributing risk factors.

Contributing risk factors derived from offline data sources such as police crash reports are prone to issues such as imprecision, measurement error, and missing data. Firstly, the spatial and temporal scales of many modeled variables were too abstract to reflect the actual conditions in the presence of a PC. For example, the frequently used AADT in models (e.g., Khattak *et al.* (2009); Mishra *et al.* (2016)) only represents traffic flow conditions at an aggregated level and cannot exactly reveal the traffic conditions in each crash scenario. Secondly, risk factors like incident duration time are subject to measurement errors due to inaccurate reports on incident time, delayed information communication, etc. Although some studies (e.g., Khattak *et al.* (2012)) have proposed estimation algorithms to approximate incident durations, they cannot warrant the performance due to their own model issues (e.g., simplified model structures and assumptions). Finally, not all risk factors such as accurate incident location and detailed road geometric characteristics are fully available in each modeled dataset. This forces analysts to only consider the available variables while excluding some more important ones in modeling. A detailed offline data selection and acquisition plan, thus, is expected to guide analysts to acquire the necessary data for developing more reliable prediction models.

On the other hand, some studies used real-time traffic data in models. However, the accessibility remains an issue. Real time traffic monitoring systems are not always available in different regions and time periods. Thus, not all crash scenarios can be linked with the prevalent traffic flow data. For example, Vlahogianni *et al.* (2010) reported that more than 50 percent of the crashes cannot be linked to the avaiable hourly lane volume and travel speed information from the dense closed-circuit television camera system on the Attica Tollway in Greece.

Meanwhile, real-time information archived from data sources such as loop detectors and probe vehicles (e.g., INRIX data) are also often not precise enough to match the conditions of a PC. As shown in Figure 4, one PC occurred at 9:32 and the orange box denotes its impact area. In case of the loop detector scenario (Figure 4(a)), it is unclear which loop detector should be considered for extracting the traffic flow data for modeling (e.g., flow, speed, and/or occupancy). The downstream sensor station 1, the upstream sensor station 2, or the others? Also, should the measurements of one lane or multiple lanes be used in the models? Likewise, the measurements of which period should be considered? For example, Xu *et al.* (2016) used the traffic data collected between 5 and 10 minutes prior to the PC occurrence to count for the potential inaccuracies in reported crash time, whereas Vlahogianni *et al.* (2012) extracted traffic information measured 10 minutes after the PC occurred to better capture post incident traffic conditions. It is not known which application is more effective. Thus, there should be a clear guidance on the use of the traffic measurements. In case of the probe vehicle scenario (Figure 4(b)), similar problems also present. Current studies used the link-level traffic measurements aggregated from probe vehicle data (e.g., INRIX data) in modeling SC occurrences (Park and Haghani 2016a; Park *et al.* 2017). Should the measurement of $L_1$ or $L_2$ be used? Which period should be considered? In fact, none of the link measurements can perfectly represent the traffic conditions under the PC scenario, especially when the number of probe vehicles is small and a PC occurred in the middle of a long link. Not only the traffic conditions at the onset of a PC affect the risk of SCs, but also the incoming traffic post the occurrence of PC. In other words, the upstream varying traffic (e.g., the ones approaching link $L_4$ in Figure 4(b)) will also dynamically change the SC risk. However, the inclusion of these traffic information in the prediction model will be difficult as the arrival traffic changes spatio-temporally.

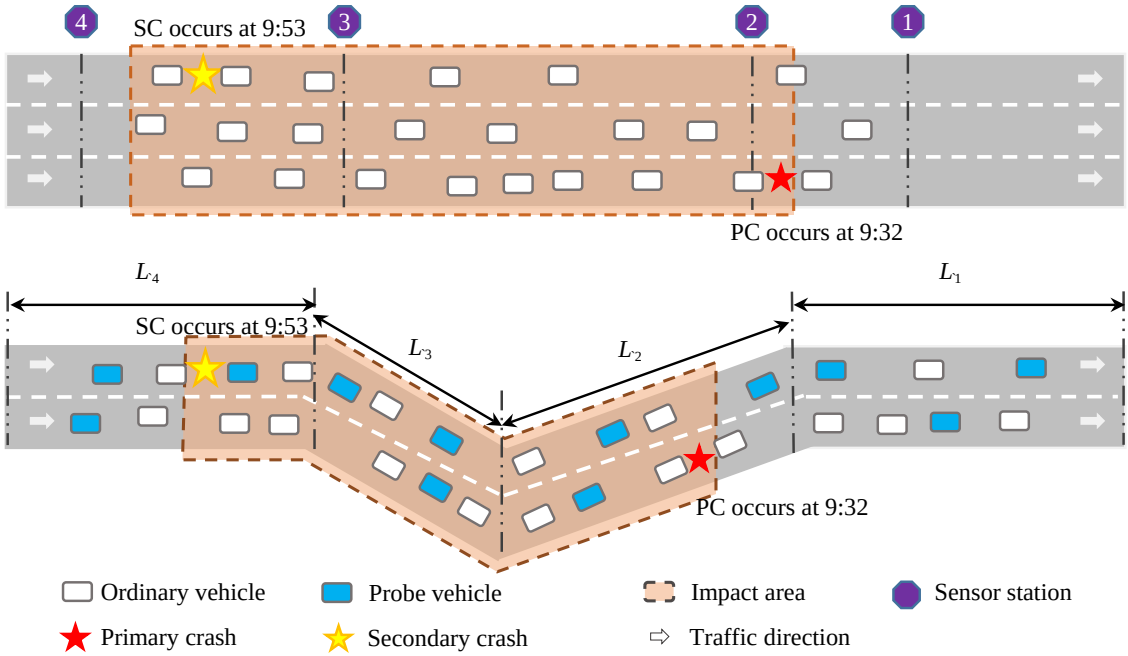


**Figure 4. Collecting real-time traffic measurements: (a) sensor stations; and (b) probe vehicles.**

Another major issue involves the use of limited risk factors to predict SC occurrences. The inconsistent use of different variables in current modeling practices implies that it is impossible to collect and incorporate all possible factors in the model. As shown in Figure 5, despite the agreement on certain factors such as incident duration, many factors were found to have inconsistent impacts on SC risk. Based on the existing papers, the most frequently used risk factor is time of the day, but only half of the studies reported it to be significantly related to SC occurrences. Likewise, the number of lanes was found to be negatively related to SC occurrences in two references whereas it was identified to be insignificant in the other three references. Most studies find that longer incident duration leads to higher risk of SCs (Vlahogianni *et al.* 2012; Yang *et al.* 2014b; Wang *et al.* 2016; Park *et al.* 2018). It should be noted that some factors such as visibility have been rarely used in current studies and should be considered in future studies. In summary, the selection of candidate risk factors regarding the prediction of SC risk need to be carefully considered as no clear guidance is available.

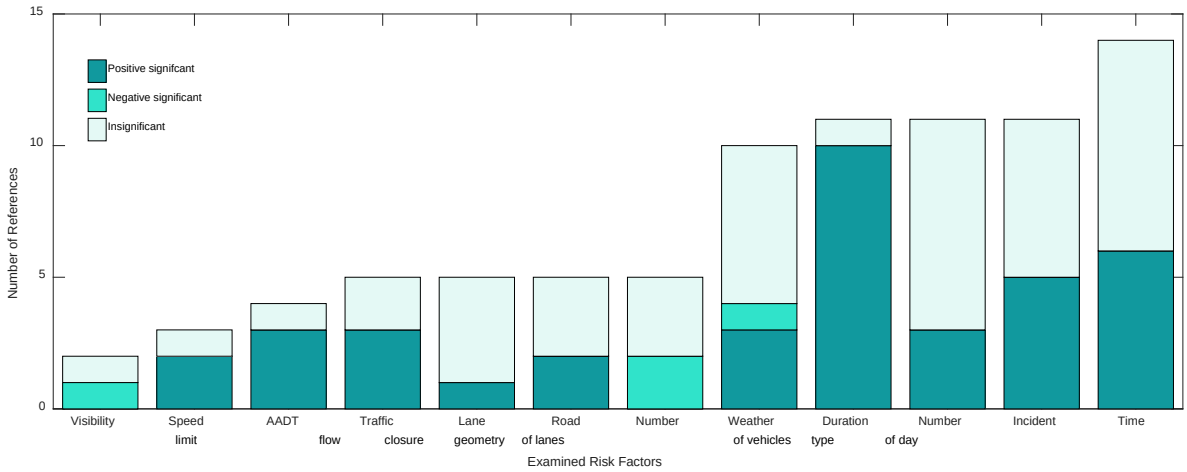


**Figure 5. Significance of contributing risk factors in existing studies.**

### 3.3. Development of Countermeasures for Reducing SC Risk

Based on the identification and modeling of SC occurrences, few countermeasures for mitigating SC risk have been investigated. Existing studies mainly explored the potentials of the advanced warning, traffic control, and effective incident management in reducing SC risk: (a) by taking advantage of real-time traffic and incident information, advanced warning aims to provide drivers a more informed driving environment upstream; (b) empowered by the ITS solutions, active traffic control can help reduce the fluctuation of traffic flow operations; and (c) responsive traffic incident management programs help minimize the impact of incidents post their occurrence. Although the potential benefits of these countermeasures are perceivable, there are still many challenges to make them more readily deployable.

One major challenge is where, when, and how should the information of the PC be transmitted to alert drivers. Firstly, the effect of the incident information will decrease with the increase in its propagating distance along a roadway. For example, Kopitch and Saphores (2011) found that the impact of the CMS decreased after 11.15 miles. Nevertheless, defining the range of information sharing will be difficult because each PC may have a different spatio-temporal impact on traffic. For example, given a PC in Figure 6, should the crash information be posted on the nearest CMS or all the upstream CMSs? Should the CMS be supplemented with radio information to alert drivers about the incidents and to provide them ample time to react to downstream traffic changes? How long should the message last? If out-of-date information was frequently provided, drivers may disregard CMS information. Similarly, the implementation of the VSL under incident condition also needs to resolve the spatio-temporal coverage issues. When connected vehicles are used as the medium to share incident information, an

immediate question to ask is its effectiveness under low penetration condition. For example, Figure 6 shows that only a limited number of CVs are present and their impacts on the overall traffic flow may not be notable.

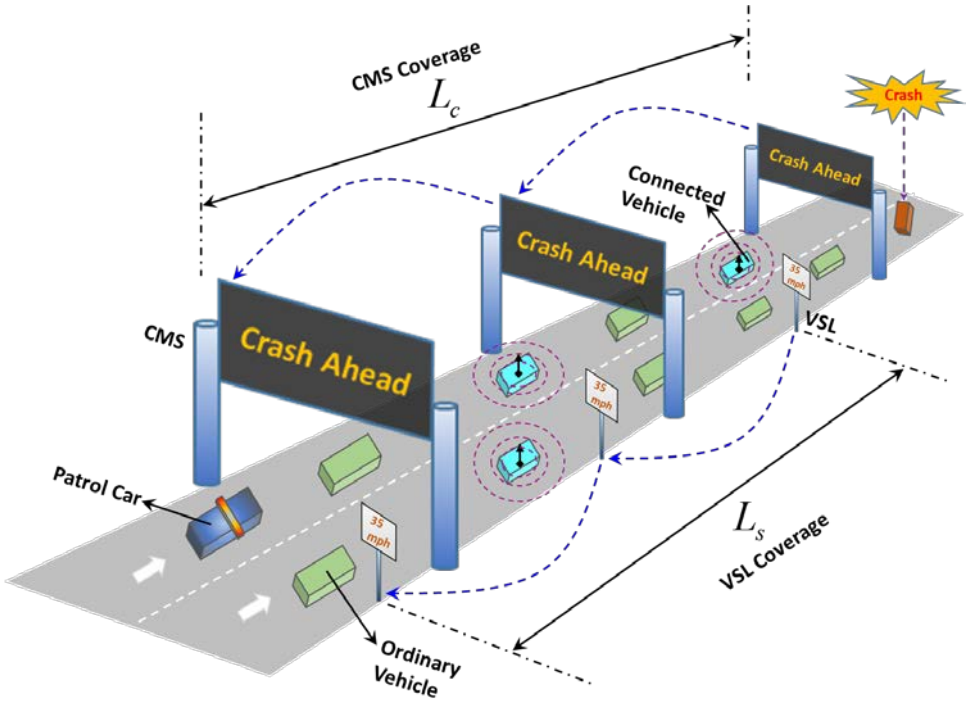


**Figure 6. Possible countermeasures to prevent secondary crashes.**

In fact, most of existing highways only have a limited number of CMS and VSL. CMS and VSL are often sparsely distributed along the highways. In addition, not all vehicles and roadways are equipped with sufficient sensor devices or roadside units to facilitate information sharing. Adding additional infrastructure and hardware is also subject to high cost. These issues together largely restrict the widely use of existing countermeasures for reducing SC risk.

Obviously, quickly clearing a PC from the roadway will minimize its impact on traffic flow. However, this requires an effective highway service patrol and incident response program. Two approaches, including installing closed circuit televisions system in Karlaftis *et al.* (1999) and allocating emergency response with optimized distributions in Park *et al.* (2016), have been examined. The implementation of such countermeasures is challenging due to the limited resources available (e.g., patrol vehicles, personnel, traffic surveillance systems, etc.). The allocation of these resources should be well investigated. In addition, each PC may occur at different condition and consequently result in different impact. Any fixed incident response program may not work effectively. For example, a patrol vehicle in Figure 6 may get stuck in a long queue and delay the process of crash clearance. Thus, research on the preparedness of incident management deserves more efforts.

## 4. POTENTIAL OPPORTUNITIES AND FUTURE RESEARCH DIRECTIONS

Based on the knowledge gaps with respect to the identification, modeling, and prevention of SCs in existing literature, this paper also examined some potential opportunities and research directions that deserve more investigations in future work.

### 4.1. Fusing data from different sources for SC identification

With the development of intelligent transportation systems (ITS), more and more data sources are becoming available. This provides many opportunities to revisit the estimation of the IA associated with PCs. As shown in

Figure 7(a), other than the conventional sensor stations, probe vehicles and CVs are emerging as the new sources for obtaining high-resolution traffic measurements to estimate the IA of PCs. If additional traffic measurements were available for the part between two spaced inductive loops, it would be very useful for the characterization of the evolution of traffic states. As discussed earlier, it is often impractical to install additional sensors to bridge the large gap between existing sensor stations, for example, stations 1 and 2 in Figure 7(a). Alternatively, the virtual sensor stations introduced by Yang *et al.* (2014a) can be considered to provide traffic speed estimation for the highway section divided into shorter segments. Rather than relying on map information, probe vehicles and/or CVs that collect valuable information from in-vehicle devices such as smartphones, GPS, Wi-Fi receivers, and Bluetooth can be integrated with traditional sensor measurements. For example, the original two sections $L_1$ and $L_2$ in Figure 7(a) covered by sensor stations 1, 2 and 3 can be divided into five short segments associated with six blue virtual sensors shown as the triangular markers. The speed of each segment can be re-estimated based on the fusion of data from both sensor stations, probe vehicles, and/or CVs. For instance, the blue dots in Figure 7(b) represent the sampled trajectory points with speed measurements. Integrating these measurements with archived sensor data can obtain enhanced speed estimation for each cell between two virtual sensors shown in Figure 7(c). For example, the speed information of virtual sensors 2 and 3 can be estimated based on traffic information of sensors 1 and 2 and all probe points located on the road segment. Various approaches such as linear, nonlinear interpolation, and shockwave based can be considered. Thus, the IA of the PC can be better estimated with the smaller cells constructed by the virtual sensors. Nevertheless, an immediate task that deserves attention is to investigate the appropriate data fusion algorithms that facilitate the aggregation of the measurements from different sources.

Although there are already some actual CV datasets available, their usability requires further exploration. For example, the Safety Pilot Model Deployment (SPMD) program[1] collected connected vehicle data from approximately 3,000 onboard vehicle equipment and 30 roadside equipment (RSE) in Michigan (Xie *et al.* 2017). However, no studies have examined their potentials in SC identification. The problems such as the necessary market penetration rates of CVs deserves specific investigation in the context SC analysis. Also the problem of missing data because some segments may not have any measurement from probe or CVs in some extreme conditions needs to be also addressed. How to impute the speed for these measurements also needs more attention. It is expected that an efficient and valid solution to these issues can leverage the value of these emerging data sources and significantly increase the performance of SC identification and risk prediction.

[1] https://www.its.dot.gov/factsheets/safetypilot_modeldeployment.htm

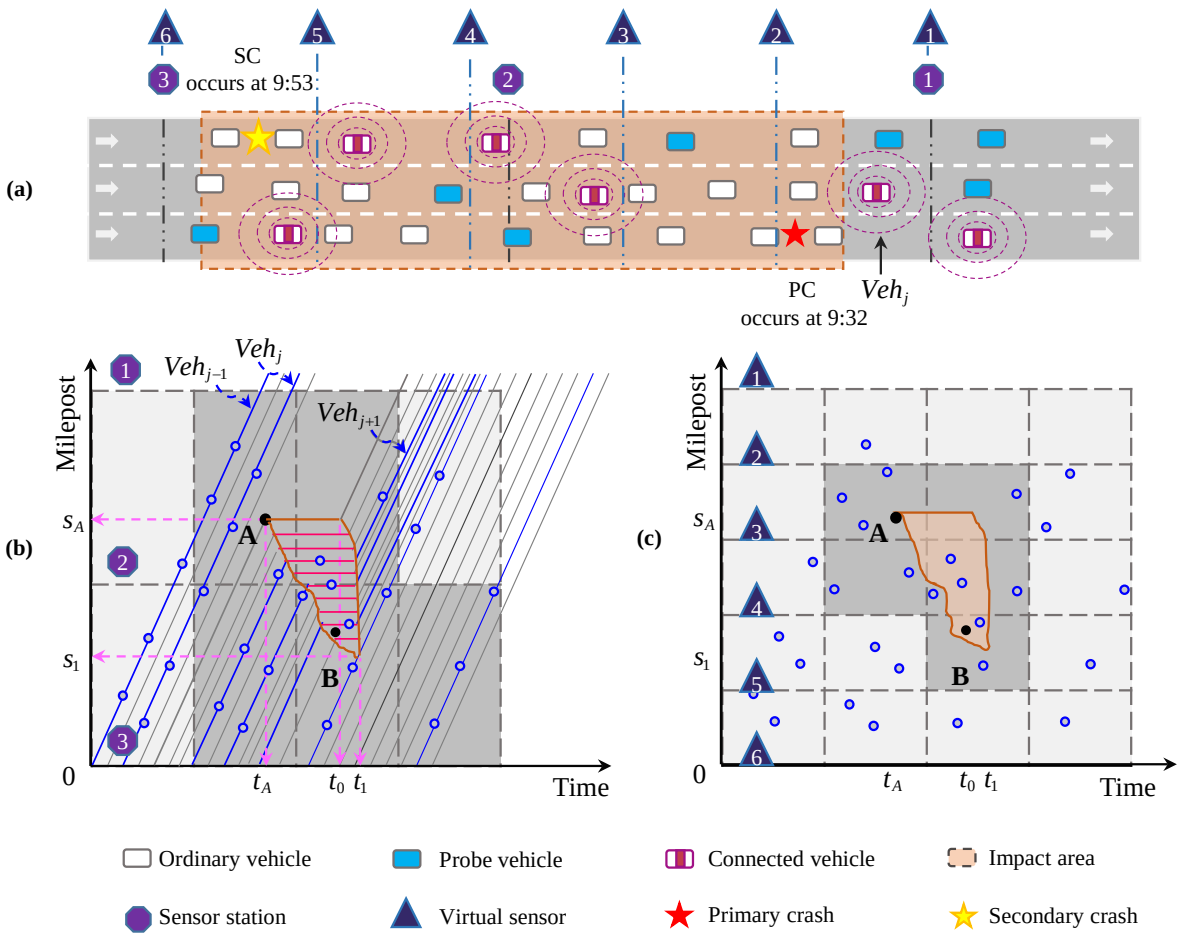


**Figure 7. Data integration: (a) road system with multiple data sources; (b) actual sensor stations; and (c) virtual sensor stations.**

### 4.2. Using advanced learning algorithms for real-time SC analysis

Advanced learning algorithms have drawn increased attention to many researchers in transportation field (Halim *et al.* 2016). Many approaches such as genetic algorithms (GA), supervised learning, dimensionality reduction, reinforcement learning, and deep learning have been explored. They have been used for addressing problems associated with driving safety, vehicle crash prediction, etc. With the increasing possibility of acquiring massive high-resolution sensor data, many learning algorithms are promising in capturing the nonlinear relationship between SC risk and limited contributing factors.

The performance of some approaches such as the BNN has already been examined and were found to outperform those simple parametric algorithms (Vlahogianni *et al.* 2012). However, the use of advanced learning algorithms for real-time analysis is not yet entirely solved yet. Firstly, the selection of the advanced learning algorithms needs to be carefully conducted. For example, should convolutional neural network (CNN) or support vector machine (SVM) be selected under datasets with different sizes? Secondly, the computational complexity of each algorithm needs to be alleviated while maintaining the performance for real-time SC analysis. Thirdly, the scope of the real-time analysis needs to be defined. A spatio-temporal window has to be applied to extract necessary data as the input for advanced learning approaches. However, it is obscure to choose the appropriate window size and analysis duration. For example, should the traffic flow information be updated every five minutes? A small time interval such as 30 seconds for data collection and processing may be too frequent and result in unnecessary computation burden, whereas a large time interval may not provide satisfied performance in capturing

the changes of traffic states. If these issues can be resolved, some other approaches such as survival analysis can also be considered to model the risk of SCs. In other words, there exist potentials for adapting advanced learning approaches for real-time SC analyses but special attentions should be made in deploying them appropriately and efficiently.

### 4.3. Deploying CVs for SC prevention

Compared with CMS and VSL signs, the emerging CV technologies that use wireless short-range communication offer real-time information sharing channels. Thus, CVs not only provide valuable data sources for SC identification and prediction, but also grant promising solutions to prevent SCs by timely sharing safety messages with approaching vehicles. CVs with a reasonably high penetration rate facilitate the construction of a high-resolution speed contour map. Given a sufficient number of CVs, even the precise impact boundary (e.g., the yellow boundary) shown in Figure 7 (c) can be approximated. More importantly, CVs enable the real-time information to alert incoming traffic and maintain a more stable traffic flow. For example, TIM agencies can adaptively reduce the speed limit upstream of the PC and transmit such information to remind target drivers as shown in Figure 8. The benefit of using CVs to mitigate SC risk has been lately examined in Yang *et al.* (2017a). PC information can be shared with CVs through safety message, thus drivers can optimize their driving behavior such as car following and lane changing for safety. Meanwhile, drivers with conventional vehicles can passively interact with surrounding vehicles, and assess information such as CMS and VSL to adjust their behavior. It was shown that CVs can help drivers make more informed driving responses to reduce SCs risk even with a relative low market penetration. However, it should be noticed that the impact of CVs in low volume conditions with an extremely low penetration rate (such as 1%) might be too trivial to prevent SC occurrences. Thus, there will be a transition period that information dissemination will mainly rely on other media (e.g., CMS) to support mixed traffic environment with few CVs.

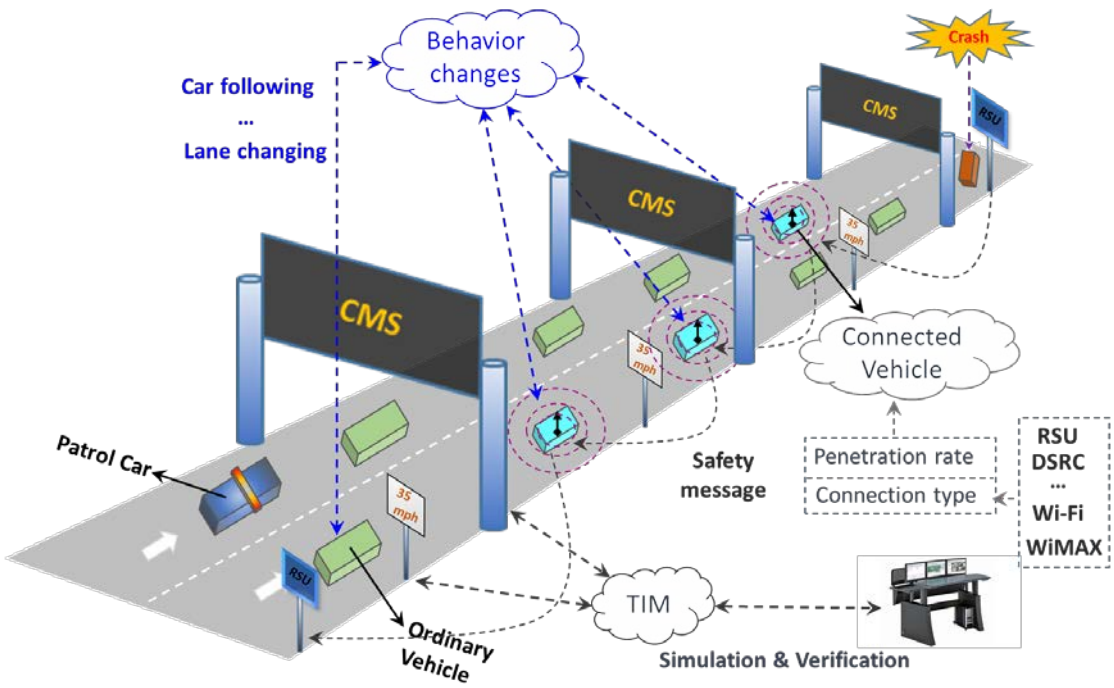


**Figure 8 . Deploying CVs for SC prevention.**

Despite the potential of CVs for preventing SCs, there still exists issues such as possible communication delays and information package loss that need to be further addressed. It needs to be noted that these issues will

be more severe in heterogeneous CV networks. When dedicated short-range communication (DSRC) or road side units (RSUs) are assumed to be the major techniques for transmitting information about PCs, communication techniques such as Wi-Fi and WiMAX can also be implemented to construct the heterogeneous communication network to support more CVs. Since it is impractical to examine countermeasures such as dynamic CMS and VSL information in actual crash conditions, simulation software such as SUMO, Paramics, and Vissim can be used with communication simulators (e.g., NS3) to conduct simulation and validation tests for both communication mechanism and traffic interaction mechanism. Driving simulation can be used to train drivers to respond to the information from CVs, CMS, and VSL appropriately. It is expected to help researchers address the impact of issues such as handoff delay and assess the deployment of CVs for SC prevention.

## 5. CONCLUSIONS

As indicated in the preceding discussion, while a substantial amount of research efforts have been invested in SC-related studies, there still exists several formidable problems in terms of SC identification, prediction, and prevention. In the past two decades, the advances in SC identification techniques can now enable analysts to dynamically estimate the impact area based on crash and traffic information. Despite the increasingly attention to the identification of SCs, their verification is still challenging as no practical solutions are available for large-scale analysis. Meanwhile, other than the frequently used parametric approaches, innovative non-parametric approaches have also been introduced to improve the performance of predicting SC occurrence. Their predictive performance and model transferability were not well testified. Moreover, the investigations on the prevention of SCs have been inherently limited by the available data, expensive cost, and thus still need continuous endeavor.

The anticipated availability of high-quality real-time data emerging from probe vehicles sourcing data from GPS, Wi-Fi, Bluetooth, and/or connected vehicles holds considerable promise for the future development of methods for SC analysis. Combined with the intelligent learning approaches, the real-time identification, prediction, and prevention of SCs will be possible. One can expect that the refinements of advanced algorithms and data fusion approaches can provide new insights into the mechanism of SC occurrences as well as their prevention.

## 6. ACKNOWLEDGEMENT

The authors appreciate the valuable comments and suggestions from the anonymous reviewers for helping improve the paper. The contents in this paper do not reflect views of any sponsoring agencies.

## 7. REFERENCES

Blincoe, L.J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A, 2015. The economic and societal impact of motor vehicle crashes, 2010 (revised). Annals of Emergency Medicine. Washington, DC: National Highway Traffic Safety Administration., pp. 194-196.

Chang, G.-L., Rochon, S., 2009. Performance evaluation and benefit analysis for chart. Final Report - Year 2007. University of Maryland, College Park.

Chou, C.-S., Miller-Hooks, E., 2009. Simulation-based secondary incident filtering method. Journal of Transportation Engineering 136 (8), 746-754.

Chung, Y., 2013. Identifying primary and secondary crashes from spatiotemporal crash impact analysis. Transportation Research Record: Journal of the Transportation Research Board (2386), 62-71.

Dougald, L., Goodall, N.J., Venkatanarayana, R., 2016. Traffic incident management quick clearance guidance and implications. Virginia Transportation Research Council.

Fontaine, H., 1995. A typological analysis of pedestrian accidents. 7th workshop of International Committee for

Traffic Conflict techniques (ICTCT).

Fries, R., Chowdhury, M., Ma, Y., 2007. Accelerated incident detection and verification: A benefit to cost analysis of traffic cameras. Journal of Intelligent Transportation Systems 11 (4), 191-203.

Goodall, N.J., 2017. Probability of secondary crash occurrence on freeways with the use of private-sector speed data. Transportation Research Record: Journal of the Transportation Research Board (2635), 11-18.

Green, E.R., Pigman, J.G., Walton, J.R., Mccormack, S., 2012. Identification of secondary crashes and recommended countermeasures to ensure more accurate documentation. Transportation Research Board 91st Annual Meeting. Washington, DC.

Halim, Z., Kalsoom, R., Bashir, S., Abbas, G., 2016. Artificial intelligence techniques for driving safety and vehicle crash prediction. Artificial Intelligence Review 46 (3), 351-387.

Hirunyanitiwattana, W., Mattingly, S.P., 2006. Identifying secondary crash characteristics for california highway system. Transportation Research Board 85th Annual Meeting. Washington, DC.

Imprialou, M.-I.M., Orfanou, F.P., Vlahogianni, E.I., Karlaftis, M.G., 2014. Methods for defining spatiotemporal influence areas and secondary incident detection in freeways. Journal of transportation engineering 140 (1), 70-80.

Imprialou, M., Quddus, M., 2017. Crash data quality for road safety research: Current state and future directions. Accident Analysis & Prevention.

Jalayer, M., Baratian-Ghorghi, F., Zhou, H., 2015. Identifying and characterizing secondary crashes on the alabama state highway systems. Advances in Transportation Studies (37), 129-140.

Karlaftis, M.G., Latoski, S.P., Richards, N.J., Sinha, K.C., 1999. Its impacts on safety and traffic management: An investigation of secondary crash causes. Journal of Intelligent Transportation Systems 5 (1), 39-52.

Khattak, A., Wang, X., Zhang, H., 2009. Are incident durations and secondary incidents interdependent? Transportation Research Record: Journal of the Transportation Research Board (2099), 39-49.

Khattak, A., Wang, X., Zhang, H., 2012. Incident management integration tool: Dynamically predicting incident durations, secondary incident occurrence and incident delays. IET Intelligent Transport Systems 6 (2), 204-214.

Kopitch, L., Saphores, J.-D.M., 2011. Assessing effectiveness of changeable message signs on secondary crashes. Transportation Research Board 90th Annual Meeting. Washington, DC.

Latoski, S.P., Pal, R., Sinha, K.C., 1999. Cost-effectiveness evaluation of hoosier helper freeway service patrol. Journal of Transportation Engineering 125 (5), 429-438.

Li, Z., Li, Y., Liu, P., Wang, W., Xu, C., 2014. Development of a variable speed limit strategy to reduce secondary collision risks during inclement weathers. Accident Analysis & Prevention 72, 134-145.

Lou, Y., Yin, Y., Lawphongpanich, S., 2011. Freeway service patrol deployment planning for incident management and congestion mitigation. Transportation Research Part C: Emerging Technologies 19 (2), 283-295.

Mishra, S., Golias, M., Sarker, A., Naimi, A., 2016. Effect of primary and secondary crashes: Identification, visualization, and prediction. CFIRE 09-05. National Center for Freight and Infrastructure Research and Education (CFIRE), University of Wisconsin, Madison.

Montella, A., Andreassen, D., Tarko, A., Turner, S., Mauriello, F., Imbriani, L., Romero, M., 2013. Crash databases in australasia, the european union, and the united states: Review and prospects for improvement. Transportation Research Record: Journal of the Transportation Research Board (2386), 128-136.

Moore, J.E., Giuliano, G., Cho, S., 2004. Secondary accident rates on los angeles freeways. Journal of Transportation Engineering 130 (3), 280-285.

NHTSA, 2016. Quick facts 2015. https://crashstats.nhtsa.dot.gov/. Accessed: August 18, 2017.

Owens, D., 1978. Traffic incidents on the m1 motorway in hertfordshire. No. TRRL SR 390. Transport and Road

Research Laboratory, Crowhorne, Berkshire.

Owens, N., 2010. 2010 traffic incident management handbook update. Rep. No. FHWA-HOP-10-013, Federal Highway Administration, Office of Transportation Operations, Washington, DC.

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., Trego, T., 2010. Traffic incident management handbook.

Park, H., Gao, S., Haghani, A., 2017. Sequential interpretation and prediction of secondary incident probability in real time.

Park, H., Haghani, A., 2016a. Real-time prediction of secondary incident occurrences using vehicle probe data. Transportation Research Part C: Emerging Technologies 70, 69-85.

Park, H., Haghani, A., 2016b. Use of clustering model and adjusted boxplot model for identification of secondary incidents. Transportation Research Board 95th Annual Meeting. Washington, DC.

Park, H., Haghani, A., Samuel, S., Knodler, M.A., 2018. Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion. Accident Analysis & Prevention 112, 39-49.

Park, H., Shafahi, A., Haghani, A., 2016. A stochastic emergency response location model considering secondary incidents on freeways. IEEE Transactions on Intelligent Transportation Systems 17 (9), 2528-2540.

Raub, R., 1997. Occurrence of secondary crashes on urban arterial roadways. Transportation Research Record: Journal of the Transportation Research Board (1581), 53-58.

Sarker, A.A., Naimi, A., Mishra, S., Golias, M.M., Freeze, P.B., 2015. Development of a secondary crash identification algorithm and occurrence pattern determination in large scale multi-facility transportation network. Transportation Research Part C: Emerging Technologies 60, 142-160.

Sarker, A.A., Paleti, R., Mishra, S., Golias, M.M., Freeze, P.B., 2017. Prediction of secondary crash frequency on highway networks. Accident Analysis & Prevention 98, 108-117.

Sun, C., Chilukuri, V., 2007. Secondary accident data fusion for assessing long-term performance of transportation systems. InTrans Project Reports. 164. Center for Transportation Research and Education (CTRE), Iowa State University.

Sun, C., Chilukuri, V., 2010. Dynamic incident progression curve for classifying secondary traffic crashes. Journal of Transportation Engineering 136 (12), 1153-1158.

Tian, Y., Chen, H., Truong, D., 2016. A case study to identify secondary crashes on interstate highways in florida by using geographic information systems (gis). Advances in Transportation Studies (2), 103-112.

TIM, 2017. Agency case studies in arizona. http://nchrptimpm.timnetwork.org/?page_id=73#citation1. Accessed: August 12, 2017.

Vlahogianni, E., Karlaftis, M., Golias, J., Halkias, B., 2010. Freeway operations, spatiotemporal-incident characteristics, and secondary-crash occurrence. Transportation Research Record: Journal of the Transportation Research Board (2178), 1-9.

Vlahogianni, E.I., Karlaftis, M.G., Orfanou, F.P., 2012. Modeling the effects of weather and traffic on the risk of secondary incidents. Journal of Intelligent Transportation Systems 16 (3), 109-117.

Wang, J., Liu, B., Zhang, L., Ragland, D.R., 2016. Modeling secondary accidents identified by traffic shock waves. Accident Analysis & Prevention 87, 141-147.

Xie, K., Ozbay, K., Yang, H., 2018. Secondary collisions and injury severity: A joint analysis using structural equation models. Traffic injury prevention 19 (2), 189-194.

Xie, K., Yang, D., Ozbay, K., Yang, H., 2017. Identifying high-risk locations using large-scale connected vehicle data. Road Safety & Simulation International Conference 2017 (RSS 2017). The Hague, Netherlands.

Xu, C., Liu, P., Yang, B., Wang, W., 2016. Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data. Transportation research part C: emerging technologies 71, 406-418.

Yang, H., Bartin, B., Ozbay, K., 2013a. Investigating the characteristics of secondary crashes on freeways. Transportation Research Board 92nd Annual Meeting. Washington, DC.

Yang, H., Bartin, B., Ozbay, K., 2013b. Mining the characteristics of secondary crashes on highways. Journal of transportation engineering 140 (4), 04013024.

Yang, H., Ozbay, K., Morgul, E., Bartin, B., Xie, K., 2014a. Development of online scalable approach for identifying secondary crashes. Transportation Research Record: Journal of the Transportation Research Board (2470), 24-33.

Yang, H., Ozbay, K., Xie, K., 2014b. Assessing the risk of secondary crashes on highways. Journal of safety research 49, 143. e1-149.

Yang, H., Wang, Z., Ozbay, K., Xie, K., Zhu, Y., 2018. An improved methodological framework based on probe vehicle data for detecting secondary crashes. Transportation Research Board 97th Annual Meeting. Washington, DC.

Yang, H., Wang, Z., Xie, K., 2017a. Impact of connected vehicles on mitigating secondary crash risk. International Journal of Transportation Science and Technology 6 (3), 196-207.

Yang, H., Wang, Z., Xie, K., Dai, D., 2017b. Use of ubiquitous probe vehicle data for identifying secondary crashes. Transportation Research Part C: Emerging Technologies 82, 138-160.

Zhan, C., Gan, A., Hadi, M., 2009. Identifying secondary crashes and their contributing factors. Transportation Research Record: Journal of the Transportation Research Board (2102), 68-75.

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

Zhang, H., Khattak, A., 2010. What is the role of multiple secondary incidents in traffic operations? Journal of Transportation Engineering 136 (11), 986-997.

Zheng, D., Chitturi, M.V., Bill, A.R., Noyce, D.A., Year. Secondary crash identification on a large-scale highway system. In: Proceedings of the Transportation Research Board 93rd Annual Meeting.

Zheng, D., Chitturi, M.V., Bill, A.R., Noyce, D.A., 2015. Analyses of multiyear statewide secondary crash data and automatic crash report reviewing. Transportation Research Record: Journal of the Transportation Research Board (2514), 117-128.