



**EPIDEMIOLOGICAL MODELS FOR
TRANSPORTATION APPLICATIONS:
SECONDARY CRASHES**

Final Report

JULY 2022

SIRISH NAMILAE¹, DAHAI LIU² and SCOTT PARR³

*¹Dept. of Aerospace Engineering, ²College of Aviation and ³Dept. of Civil
Engineering*

Embry-Riddle Aeronautical University, Daytona Beach, Florida 32114, USA

US DEPARTMENT OF TRANSPORTATION GRANT 69A3551747125

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Epidemiological Models for Transportation Applications: Secondary Crashes		5. Report Date JULY 2022	
		6. Source Organization Code	
7. Author(s) SIRISH NAMILAE, DAHAI LIU AND SCOTT PARR		8. Source Organization Report No. CATM-2022-R6-ERAU	
9. Performing Organization Name and Address Center for Advanced Transportation Mobility Transportation Institute 1601 E. Market Street Greensboro, NC 27411		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. 69A3551747125	
12. Sponsoring Agency Name and Address University Transportation Centers Program (RDT-30) Office of the Secretary of Transportation–Research U.S. Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590-0001		13. Type of Report and Period Covered Final Report: FEB 2019-JUNE 2020	
		14. Sponsoring Agency Code USDOT/OST-R/CATM	
15. Supplementary Notes:			
16. Abstract Secondary crashes or crashes that occur in the wake of a preceding or primary crash are among the most critical incidents occurring on highways, due to the exceptional danger they present to the first responders and victims of the primary crash. In this work, we developed a self-exciting temporal point process to analyze crash events data and classify it into primary and secondary crashes. Our model uses a self-exciting function to describe secondary crashes while primary crashes are modeled using a background rate function. We fit the model to crash incidents data from the Florida Department of Transportation, on Interstate-4 (I-4) highway for the years 2015-2017, to determine the model parameters. These are used to estimate the probability that a given crash is secondary crash and to find queue times. To represent the periodically varying traffic levels and crash incidents, we model the background rate, as a stationary function, a sinusoidal non-stationary function, and a piecewise non-stationary function. We show that the sinusoidal non-stationary background rate fits the traffic data better and replicates the daily and weekly peaks in crash events due to traffic rush hours. Secondary crashes are found to account for up to 15.09% of traffic incidents, depending on the city on the I-4 Highway.			
17. Key Words Point-process model, Secondary crash, Akaike Information Criterion (AIC)		18. Distribution Statement Unrestricted; Document is available to the public through the National Technical Information Service; Springfield, VT.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 33	22. Price ...

TABLE OF CONTENTS

TABLE OF CONTENTS.....	i
EXECUTIVE SUMMARY	1
1. Introduction.....	3
2. Methodology	6
2.1. Traffic and Crash Data.....	6
2.2. Point Process Model Formulation.....	8
2.3. Periodic Variation of the Background Function	10
3. Results.....	12
4. Discussion.....	21
5. Conclusion	26
References.....	28

EXECUTIVE SUMMARY

Secondary crashes occur in the aftermath of a primary crash, increasing the likelihood of subsequent crashes, reducing highway capacity, causing high-density queues, and increasing travel time uncertainty. In this study, we developed a self-exciting temporal point process model to evaluate and categorize the crash event dataset into primary and secondary crashes. The model uses a background rate function to represent primary crashes and a self-exciting function to represent secondary crashes. We applied the model to crash data from the Florida Department of Transportation on Interstate-4 (I-4) highway from 2015 to 2017 to determine the model parameters. Based on the model parameter, the probability of given crash to be a secondary crash and the queue time is calculated. The model investigated crash events in six different cities on I-4. Initially, the model was fit using a stationary background rate. However, the result from the stationary background rate model does not sufficiently fit the data since it is based on the premise that crash events are invariant to any external factors. Therefore to fit the periodic variation of traffic and crash incidents for weekly and daily trends, we modeled a sinusoidal non-stationary function and a piecewise non-stationary function. The goodness-of-fit of the models was assessed using Akaike Information Criterion (AIC) values for each model. When comparing the performance of the stationary and non-stationary background rate model, the AIC values for stationary background rate are greater. This shows that the stationary background rate model has a higher prediction error when compared to other models. We were able to fit the crash data with non-stationary background rate models accurately and generate queue time curves with peaks on Fridays and troughs on Sundays,

which matched the crash data. Furthermore, the sinusoidal background rate model outperforms the piecewise function. Using the sinusoidal non-stationary background rate model, we find that secondary crash events account for 3.38 percent to 15.09 percent of the traffic incidents on the I-4. An average queue time of 82.5 minutes is obtained for the non-stationary background rate model using the sinusoidal function. The results of the point process model compare favorably to those of other models for identifying secondary crashes in literature. The proposed model can be used to create policies and countermeasures that aim to reduce the risk of secondary crashes. Based on the probability distribution of secondary crashes and the average queue time, the proposed model's results can be utilized as a reference to inform Traffic Incident Management (TIM) to clear the traffic incident scene. The exposure of secondary crashes is reduced drastically by clearing the crash scene effectively.

1. INTRODUCTION

Highway crashes can have an immediate and significant impact on the safety of individuals and the mobility of goods. In 2015, motor vehicle accidents resulted in over 35,000 deaths and 2.4 million injuries in the United States [1]. In 2016, these figures increased to over 37,000 deaths on United States roadways [2]. Drivers near the crash area must react rapidly to a dynamic and unpredictable environment in the immediate aftermath of a crash. Vehicles appear to queue up on the highway route as they reach the crash site. Furthermore, the crash scene itself is a distraction to drivers in both directions. This situation can increase the likelihood of yet another crash.

Secondary crashes are incidents that occur as a result of a primary accident. According to estimates, almost ten percent of highway crashes are categorized as secondary [3]. The victims of the primary crash, as well as the first responders dispatched to assist them, are in grave danger in these secondary crashes. Many organizations have missions and forums that encourage drivers and emergency responders to learn about the value of “move over” laws for protecting individuals employed on the side of the road. Comprehending why secondary crashes occur and predicting where and when they occur will help protect vulnerable road users, including primary crash victims and emergency responders.

Given the difficulties in modeling secondary crashes, we plan to utilize a new modeling approach to study secondary crashes based on a point process model. Secondary crashes exhibit characteristics of social contagion. The traffic and road

conditions developed during the primary crash lead to secondary crashes in a manner similar to contacts with an infected individual leading to further disease spread. Such diffusion of events or information from a primary event has been studied using Hawkes point process models [1-5]. We have recently utilized epidemiological models to study the spread of fuel shortages during hurricane evacuations [6]. These studies point to the success of epidemiological models in examining the dynamics of social contagion.

We propose to utilize a self-exciting point process model commonly used in the study of earthquakes [7], known as Epidemic Type After Shock (ETAS) model to quantify secondary crashes from a traffic dataset. The self-exciting point process models are commonly used to classify the dataset of discrete events into background and offspring events. For example, in case of earthquakes, the background events are the independent mainshock events, while the dependent aftershocks are considered to be secondary events. While there have been several modeling advances, and code developments focusing on seismic modeling [7-11], several researchers have applied this concept to other clustered societal problems. For example, Mohler et al. [3] used the self-exciting point process model to understand crime events; Zhao et al. [11] used a self-exciting point process model to predict tweet popularity by modeling information cascades. Dassios et al. [12] modeled the contagion risks or clustering events in finance and insurance while Bertozzi et al. [13] used Hawke's point process models to understand the email communication between individuals in an organization. Towers et al. [5] employed the same approach to identify the contagion in mass killings and

school shootings. Lewis et al. [14] characterized the temporal patterns of violent civilian deaths in Iraq using a self-exciting point process model.

In recent studies, the impact on the frequency of secondary crashes by a primary crash on highways is evaluated using Zero Inflated Ordered Probit (ZIOP) regression model [15] using real-time traffic flow conditions to correlate the risk of several secondary crashes with real-time traffic, geometric, weather and primary crash characteristics following a primary crash. Kitali et al. [16] used real-time speed data from Blue Toad paired devices to identify secondary crashes. Sarker et al. [17] proposed Generalized Ordered Response Probit (GORP) models that would allow us to predict the frequency of secondary crashes occurrence based on segment and traffic characteristics. Zhang et al. [18] developed a network-based clustering algorithm based on crowdsourced Waze user reports from the primary crash, with any subsequent crashes occurring inside the cluster of primary crash is considered to be secondary crash. Salek et al. [19] presents a method for assessing the likelihood of freeway secondary crashes with Adaptive Signal Control Systems (ASCS) deployed on alternate routes.

In this paper, a novel approach for identifying secondary crashes using the self-exciting point process model is formulated. To account for periodic variations in traffic events, corresponding periodic background rate variations are introduced. The model is employed to analyze the crash events on a dataset corresponding to Interstate Highway-4 (I-4) in Florida, USA. The results from the point process model compare

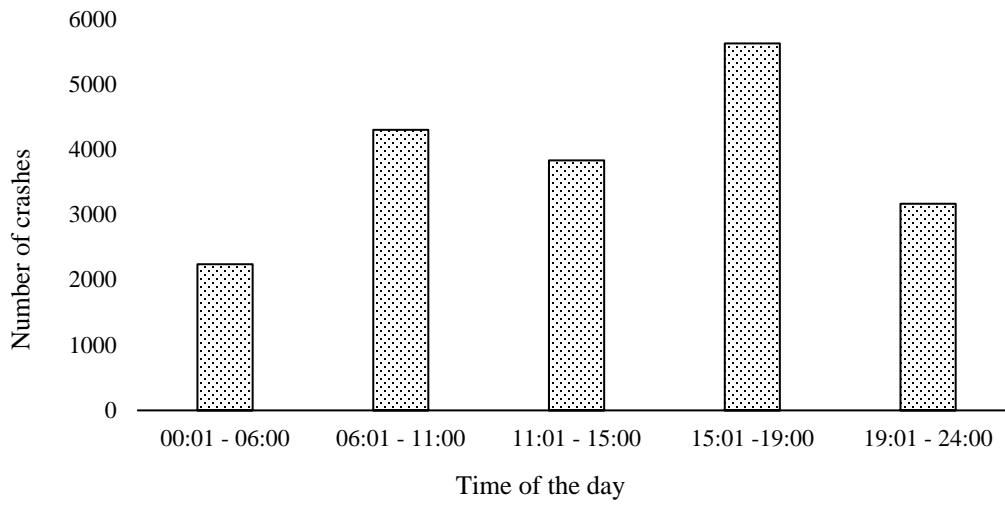
favorably with other models used to analyze secondary crashes. This approach provides a new statistical tool for analyzing this ubiquitous transportation problem.

2. METHODOLOGY

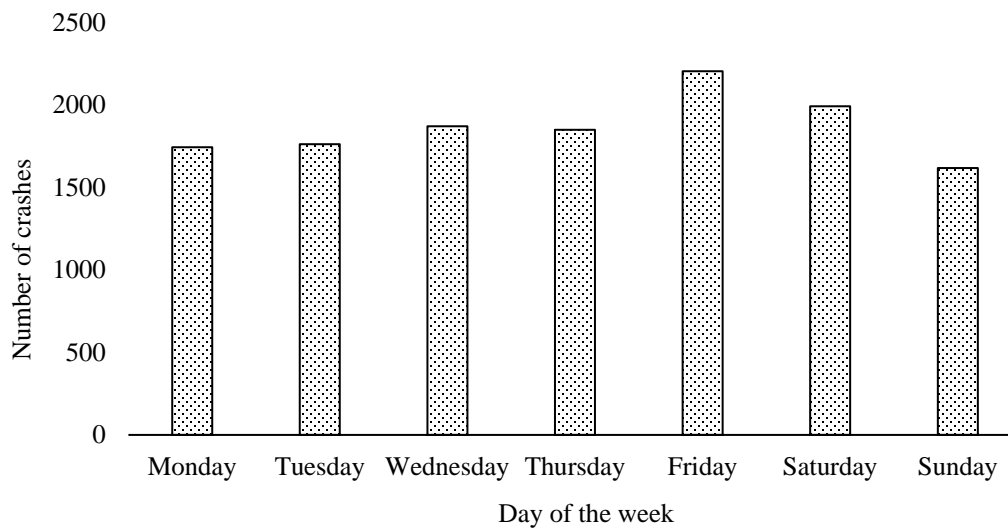
2.1. Traffic and Crash Data

The crash data for this study is obtained from the Florida Department of Transportation (FDOT) Safety Office's GIS Query Tool (SSOGIS) on Interstate4 (I-4) for the years 2015-2017 [20]. The I-4 highway between Daytona Beach, FL and Tampa, FL is entirely located in the state of Florida, USA. All the crashes on this highway are listed in the FDOT Crash Analysis and Reporting database. The data reports several crash characteristics including crash date, crash time, longitude, latitude, weather condition, crash location, speed limit, type of injury, light condition, and surface condition of the road. This data does not identify if the crash events are primary or secondary.

The crash dataset used in this research contains information on 6367, 6663 and 6133 crash events on I-4 for the years 2015, 2016 and 2017 respectively. A temporal distribution of the crash data indicates that the highest number of crashes occurred during the evening rush-hour, followed by morning rush-hour. Also, the highest number of crashes occurred on Fridays, and the lowest on Sundays. The histograms in Figure 1 show those variations.



(a)



(b)

Figure 1. Histograms showing the distribution of crashes according to (a) time and (b) day of the week.

2.2. Point Process Model Formulation

In a self-excitation point process model, recent prior events increase the probability of another event happening in the near future. Primary crashes are often sites for reduced mobility which causes a shock wave in traffic flow. The shock waves are the byproduct of traffic congestion and queuing [21]. When a driver enters through a shockwave, one experiences a sudden change in vehicle speed, which often causes secondary accidents. Secondary accidents are also caused due to rubbernecking phenomena as the driver approaches the vicinity of the crash location site. Given a time crash database consisting of N crash events (t_i) , $i = 1, \dots, N$, representing respectively the time of the i^{th} crash event, the conditional intensity function is given by:

$$\lambda(t) = \mu + \sum_{i:t_i < t} g_{A,\alpha}(t - t_i) \quad (1)$$

λ is the limiting expected rate of crash events, given the history $H_t = \{(t_i): t_i < t\}$ of all crash events in a region S , up to time t . The model classifies crash events in the catalog data into two categories, primary crashes and secondary crashes. Primary crashes are assumed to occur independent of time at a rate $\mu > 0$.

$g_{A,\alpha}(t - t_i)$ is the probability density function (PDF) of the occurrence time of a secondary crash at time t_i and is modeled as

$$g_{A,\alpha}(t - t_i) = \begin{cases} A * \alpha * \exp[-\alpha * (t - t_i)], & t > t_i \\ 0, & t \leq t_i \end{cases} \quad (2)$$

The conditional intensity function $\lambda(t|H_t)$ can now be expressed as,

$$\lambda(t) = \mu + A \sum_{i: t_i < t} \alpha e^{-\alpha(t-t_i)} \quad (3)$$

Here, α indicates the decay rate in time for the occurrence of secondary crashes while A represents the amount of excitation generated by the collection of prior events. Note that $g_{A,\alpha}$ is the function describing the time lag $t - t_i$ between a secondary crash and corresponding primary crash. The model is function of three unknown parameters $\theta = (\mu, A, \alpha)$. These parameters are determined through maximum likelihood estimation while fitting the model to the crash data. Given a time sequence of crash data consisting of N crash events $\{(t_i), i = 1, \dots, N\}$, during a time interval $[0, T]$ and in a region S , the log-likelihood function [22] is:

$$l(\theta|H_t) = \sum_{i=1}^N \log \lambda_{\theta}(t_i|H_{t_i}) - \int_0^T \iint_S \lambda_{\theta}(t|H_t) dt \quad (4)$$

To obtain the maximum likelihood estimate (MLE) of θ , the Davidon-Fletcher-Powell method is used as a gradient-based nonlinear optimization procedure [22]. A numerical approximation of the integral terms in the log-likelihood function is first computed. A probabilistic approach previously used for declustering an earthquake catalog [8] is used here to decluster the crash data and to obtain the primary crash rate μ . The probability that crash event i triggered crash event j is:

$$p_{ij} = \begin{cases} \frac{g_{A,\alpha}(t_j-t_i)}{\lambda(t_j|H_{t_j})} & , t_j > t_i \\ 0 & , t_j \leq t_i \end{cases} \quad (5)$$

Hence, the probability of crash event j being a secondary crash is $p_j = \sum_{i=1}^{j-1} p_{ij}$.

Consequently, the probability of crash event j being a primary crash is:

$$1 - p_j = \frac{\mu}{\lambda(t_j|H_{t_j})} \quad (6)$$

Once the MLE of θ is found and the probability values p_j of all N crash events are found, the crash events can be classified into primary crashes and their related secondary crashes. The maximum queue time for secondary crashes can then be extracted from the model. The queue time is the maximum time lag between primary crashes, and their related secondary crashes.

2.3. Periodic Variation of the Background Function

The model in equation (3) assumes a constant background rate. However, the crash data varies periodically on a weekly and daily basis, as shown in Figure 1. To examine this variation, a non-stationary background rate is introduced. Other researchers have used a similar approach. For example, when Lewis et al. [14] inspected the data on violent death in Iraq using a stationary background rate, the results failed to capture the upward trend over time; that issue was addressed using a non-stationary background rate model. Fox et al. [23] used a nonstationary background rate to investigate daily and weekly email activity trends.

To capture the periodic traffic and crash variations through the background rate, we used two functions: (a) a step function and (b) a sinusoidal function. A step function was used to model piecewise variation in background rate. This allows the background

rate to jump to different stationary levels at different time periods. The conditional intensity function $\lambda(t|H_t)$ with non-stationary background rate as a step function can be expressed as:

$$\lambda(t) = \mu_{step}(t) + A \sum_{i:t_i < t} \alpha \cdot e^{-\alpha(t-t_i)} \quad (7)$$

Here, when modeling the variation in crash rates for different days of the week (see fig 1(b)), the background rate is expressed as:

$$\mu_{step}(t) = \mu_i \text{ for } i = 1 \text{ to } 7, \text{ representing the 7 days of the week.}$$

Similarly, when modeling the rush-hour vs non rush hour traffic:

$\mu_{step}(t) = \mu_i$ for $i = 1$ to 3 , representing morning rush-hour, non rush-hour and evening rush-hour.

Alternatively, to account for weekly and daily traffic periodicity, a sinusoidal function is used to represent the non-stationary background rate. The conditional intensity function $\lambda(t|H_t)$ can then be expressed as:

$$\lambda(t) = \mu(t) + A \sum_{i:t_i < t} \alpha \cdot e^{-\alpha(t-t_i)} \quad (8)$$

Here, the best fit function for the weekly trend can be expressed as:

$$\mu(t) = \mu_{\sin}(t) = \mu_o \times \{P \times \sin(Q \times t + R) + S\} \quad (9)$$

Similarly, when modeling the rush-hour vs non rush-hour, daily variation in the background function is expressed as:

$$\mu(t) = \mu_{\cos}(t) = \mu_o \times \{P_h \times \cos(Q_h \times t + R_h) + S_h\} \quad (10)$$

Here, P , Q , R , S and P_h , Q_h , R_h , S_h are the parameters for background rate that fit the weekly and daily crash trend respectively.

The R code for the ETAS model obtained from Jalilian [9] was adapted to identify the secondary crash events in the present work. The code was developed to analyze an earthquake catalog using the stochastic declustering approach. We modified the code for a temporal-only self-exciting point process for the current application. The input data contains the date and time of the crash events. We also utilized a separate Matlab code for post-processing to classify potential secondary crashes from the dataset based on queue time and corresponding probability values.

3. RESULTS

The proposed approach is demonstrated on a dataset consisting of all crashes on Interstate-Highway-4 (I4) in Florida, USA for three years (2015 to 2017) [20]. The interstate highway I-4 is located entirely within the state of Florida, spanning 132.3 miles [24]. In the west, the I-4 begins in Tampa and ends in the east in Daytona Beach. The model is investigated in six different cities on I-4: Tampa, Plant City, Kissimmee, Orlando, Sanford, and Daytona Beach. Tampa and Orlando are major cities and were chosen because their corresponding data contains more than 500 crash events. The other cities modeled here have at least 200 crash events each.

Spatial and temporal thresholds from a primary incident are often used to analyze secondary crashes. Recent studies by multiple investigators [25-27] indicate that the spatial threshold is about two miles. Therefore, we analyze the data within a

threshold of two miles of the highway for the central parts of the above mentioned cities. Initially, the model is implemented with a constant background rate. It is then used to examine the periodic variation of weekly and daily trends by employing non-stationary background rates using sinusoidal and piecewise functions. Finally, the Akaike Information Criterion (AIC) [28] is used to identify the best-fit models.

Table 1 shows the optimized parameters for the constant background rate model, based on the maximum likelihood estimate. Here, A is the measure of excitation generated from the previous events and, α represents the rate of decay in the triggering function in equation (3). Based on the functional form of the triggering function (g), the reciprocal of α gives the average time-window over which secondary crash can occur following a primary crash. For example, from Table 1, in case of Tampa $\alpha^{-1} = 0.0956$ days = 137.7 minutes. This implies that if a crash occurs within 137.7 minutes and a two mile distance from the primary crash location, the crash is related to the primary crash and can be classified as secondary crash. The queue time evaluated from the model is used to classify the crash dataset into the primary and secondary crashes. The classification is conducted based on the probability values obtained using equation (6). The number of secondary crashes shown in Table 1 is calculated using a post-processing code. The post-processing code finds the time difference between successive events, and if the event with corresponding time difference is lesser than the queue time and corresponding threshold probability, the event is considered a secondary crash. The average queue time of secondary crash events for all the cities

studied here is 119.3 minutes for the stationary background rate. Figure. 2 shows the primary and the secondary crash locations for a two mile stretch for Tampa, Orlando, and Daytona Beach.

Table 1

Estimated parameter values, number of secondary crashes and average time of secondary event represented for various cities on I-4 for constant background rate.

City	Total crashes	mu (μ)	A	Alpha (α)	Secondary crashes	Secondary event avg. time (min)
Tampa	613	0.495	0.114	10.459	64 (10.44%)	137.7
Plant city	322	0.259	0.115	22.375	27 (8.38%)	64.3
Kissimmee	207	0.178	0.052	15.323	9 (4.34%)	93.9
Orlando	583	0.392	0.262	12.465	114 (19.55%)	115.5
Sanford	248	0.205	0.090	8.676	19 (7.66%)	166.0
Daytona Beach	242	0.190	0.138	10.401	25 (10.33%)	138.4
Average						119.3

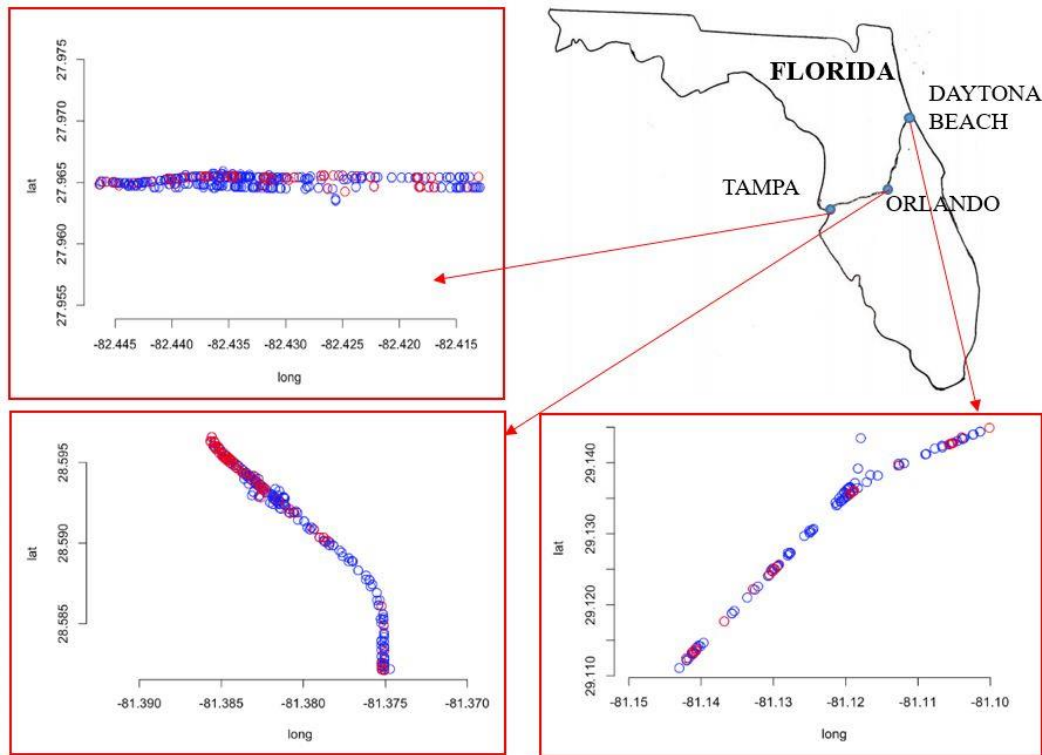


Figure 2 Primary and secondary crashes classified from the model for Tampa, Orlando and Daytona Beach along the I-4 highway. Secondary crashes are shown in red.

The point process model with a non-stationary background rate using a piecewise function is examined next. Here, the background rate is set to vary at different stationary levels on different days in a week as described earlier. Figure. 1(b) shows that the number of crashes on Friday is 136% of that on Sunday. Similarly, a non-stationary background point process model using a sinusoidal function shown in equations 8 and 9 is simulated to capture this periodic variation. Table 2 shows the estimated parameters for the cities considered for the piecewise background rate and sinusoidal background rate variations, respectively. MLE based optimization is again

used to estimate the parameters. The queue time reported in the Table 2 is a weighted average of the values for the seven days of the week. Once the average queue time is calculated for each city, the number of secondary crashes is evaluated using the post-processing described earlier.

Table 2

Estimated parameter values, number of secondary crashes and average time of secondary event represented for various cities on I-4 for piecewise and sinusoidal background rate for the different days of the week.

	City	Total crashes	μ (μ)	A	Alpha (α)	secondary crashes	Secondary event avg. time (min)
Piecewise	Tampa	613	0.491	0.121	9.511	60 (9.78%)	127.9
	Plant city	322	0.262	0.105	26.378	23 (7.14%)	46.1
	Kissimmee	207	0.179	0.050	16.025	8 (3.86%)	75.8
	Orlando	583	0.405	0.238	15.411	95 (16.29%)	78.9
	Sanford	248	0.206	0.089	8.715	18 (7.25%)	139.6
	Daytona Beach	242	0.190	0.138	10.412	23 (9.50%)	116.8
	Average						97.5
Si	Tampa	613	0.588	0.089	14.706	50 (8.15%)	82.7

Plant city	322	0.305	0.101	27.664	23 (7.14%)	43.9
Kissimmee	207	0.208	0.046	16.466	7 (3.38%)	73.8
Orlando	583	0.476	0.225	17.073	88 (15.09%)	71.2
Sanford	248	0.241	0.075	10.951	16 (6.45%)	112.8
Daytona Beach	242	0.220	0.134	10.933	22 (9.09%)	110.4
Average						82.4

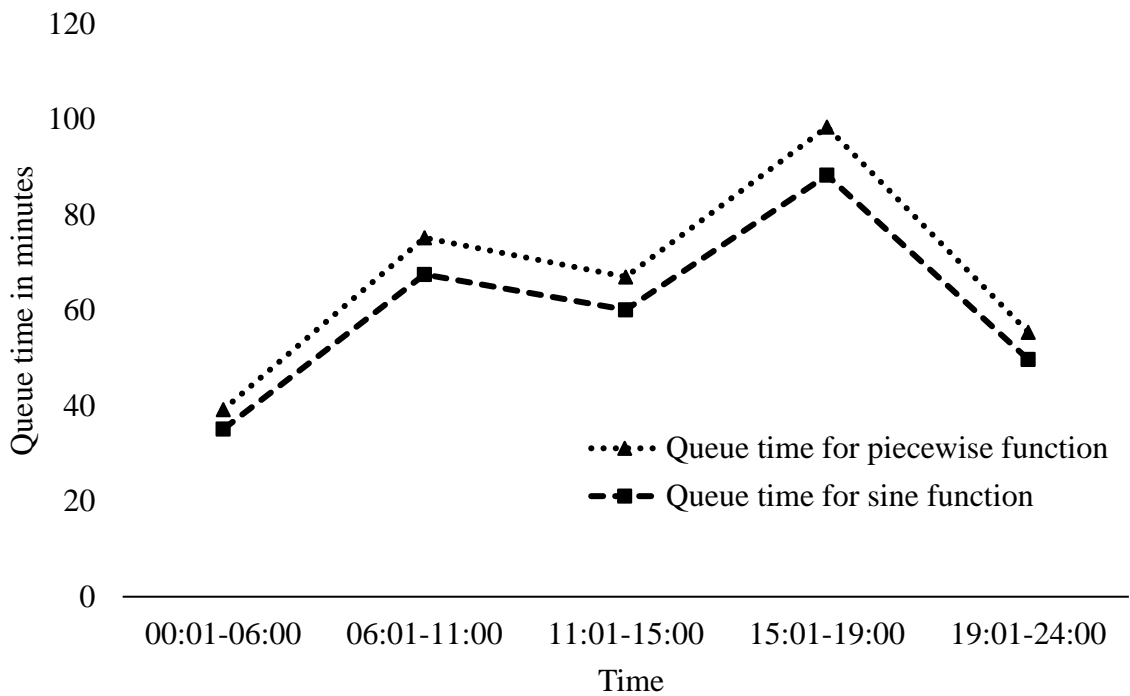


Figure 3 Queue time for Orlando region during different time of the day.

Similar piecewise and sinusoidal functions for background rate variation are used to examine the periodic variation in daily crash trends for rush-hour and other periods as reported in Figure 1(a). These are described in equations 8 and 10. Table 3 shows the corresponding number of secondary crashes and average queue times. The queue time varies for each time period and the average queue time shown in Table 3 is a weighted average value for rush hour and other time-periods.

Figure. 3 shows the variation of queue time for different time-periods for the Orlando region. Here the queue time is higher for the morning rush hour period and the evening rush hour periods. The higher traffic rates during rush hour [29] lead to an increase in the number of crashes at these times as shown in Figure 1(a). This in turn leads to higher queue times during rush hour. The underlying data for parametrizing equations (8) and (9) is based on Figure 1(a), therefore, the queue time follows the same trend for both sinusoidal and piecewise background rate variations. The results for the data corresponding to other cities are similar to Figure 3.

Table 3

Number of secondary crashes and average time of secondary event for piecewise and sinusoidal background rates, considering rush-hour and non-rush hour variations.

City	Piecewise background rate		Sinusoidal background rate	
	Secondary crashes	Secondary event avg. time (min)	Secondary crashes	Secondary event avg. time (min)
Tampa	53 (8.64%)	96.1	51 (8.31%)	88.2
Plant city	22 (6.83%)	36.6	22 (6.83%)	37.5
Kissimmee	7 (3.38%)	66.7	6 (2.89%)	57.9
Orlando	87 (14.92%)	67.0	81 (13.89%)	60.1
Sanford	15 (6.04%)	109.6	15 (6.04%)	95.1
Daytona Beach	21 (8.67%)	92.6	21 (8.67%)	86.4
Average		78.1		70.8

Akaike Information Criterion (AIC) [28] is used to find the goodness-of-fit for the proposed models. The AIC estimates the relative amount of data loss when a statistical model is fit to a given dataset. Given a set of candidate models, the model with a relatively lower AIC value is considered to be the best fit. The AIC is given by:

$$AIC = 2k - 2\ln(L) \quad (11)$$

Where k is the number of model parameters and, L represents the maximum value of the likelihood function for the model. While comparing the models with a different number of parameters, a penalty of $2k$ is added to reject any overfitting by the model. The AIC values for each model, with stationary and non-stationary background rates for weekly and daily traffic trends for each city, are shown in Table 4. The AIC values for stationary background rate are higher compared to non-stationary background rate. This suggests that the stationary background rate model has a relatively higher prediction error compared to other models. Further, the sinusoidal background rate models outperformed the piecewise function. Similarly, the sinusoidal model exhibits lowest AIC for the hourly variation as well.

Table 4

Comparison of models using AIC values.

	AIC values for stationary background rate	AIC values for weekly trend		AIC values for daily trend	
Cities	μ	$\mu_{step}(t)$	$\mu_{sin}(t)$	$\mu_{step}(t)$	$\mu_{cos}(t)$
Tampa	2037.24	2010.92	1894.67	2043.00	2001.81
Plant city	1570.23	1417.07	1381.89	1446.74	1424.88
Kissimmee	1185.62	1131.92	1099.97	1147.41	1125.74
Orlando	1944.56	1710.30	1638.90	1768.66	1720.58
Sanford	1289.25	1257.19	1218.55	1278.02	1253.69
Daytona	1198.58	1192.69	1168.87	1178.41	1163.56
Beach					

4. DISCUSSION

Several studies listed in Table 5 model secondary crashes using different approaches. The key parameters in these studies are the temporal and spatial boundaries of primary incidents that define secondary crashes. The studies based on the static method [25, 27, 30-31] model fixed and predetermined spatial and temporal thresholds for secondary crash

identification. However, these studies often do not yield a uniform threshold for secondary crashes since the traffic conditions, geometric characteristics, and incident attributes vary for each crash. Various dynamic methods [32] using varying spatio-temporal thresholds were proposed in recent studies to improve the identification of secondary crashes. In the context of the current research, the stationary background rate approach in equation 3 can be compared to the static methods. In contrast, the model with a non-stationary background rate for piecewise and sinusoidal functions is similar to the dynamic approaches.

The queue time calculated from the three models is compared in Table 6. The average queue time of 119.3 minutes is obtained for stationary background rate. The AIC values are higher for the model with a stationary background rate than the other two models as it fails to capture the periodic variation of the crash trend. Average queue time of 97.5 minutes is obtained for the non-stationary background rate model using the piecewise function. For sinusoidal function a queue time of 82.5 minutes is obtained. The AIC values are the least for sinusoidal function compared to the other two models. These queue time values are comparable to the literature in Table 5.

The secondary crash percentage from the models with stationary and non-stationary background rates are shown in Table 6. In general, the percentage of secondary crashes as well as queue time obtained from the stationary background rate model are higher than those obtained with models with non-stationary background rates. Recent studies show a similar percentage of secondary crashes with 8.8% in Moore et al. [25]; 7.5% in Chung [33]; 6-7% in Kalair et al. [34]; 8.42% in Yang et al. [35]; 5.22 % in Zhan et al. [27]; 5.53 % in Kopitch et

al. [31]; 3.23% in Zhan et al. [36] for different urban locations. The results from the current model are comparable to the literature. The point process model provides an alternate reliable modeling approach for classifying secondary crashes and devising policies and procedures to reduce the resulting loss of life and property damage.

Table 5

Literature review on identifying secondary crashes

Authors	Method	Spatio-temporal threshold
Raub [30], Karlaftis et al [37].	Static	1-mile and 15 minutes
Moore et al [25].	Static	2-mile and 2 hours; same for opposite direction
Hirunyanitiwattana et al [26].	Static	2-mile and 1 hour
Zhan et al [27].	Static	2-mile and 15 minutes
Chang et al. [38]	Static	2-mile, 2 hours and 0.5-mile, 0.5 hour (opposite direction)
Sun et al. [32]	Dynamic	Incident progression curve
Zhang et al. [39]	Dynamic	Queue length estimations
Zhan et al. [36]	Dynamic	Cumulative arrival and departure plots
Chou et al. [40]	Dynamic	Simulation-based methods
Yang et al. [41], Xu et al. [42], Kitali et al. [16], Chung et al. [33]	Dynamic	Spatio-temporal impact area methods based on speed contour plot

Imprialou et al. [43]	Dynamic	Automatic tracking of moving jams
Park et al. [44]	Dynamic	Vehicle probe data
Sarker et al. [45], Mishra et al. [46], Junhua et al. [47], Wang et al. [48]	Dynamic	Shockwave principles

Table 6

Comparison of average time of secondary crash event for each city on I-4.

Average queue time (minutes) and % secondary crashes			
City	Constant background rate	Piecewise background rate	Sinusoidal background rate
Tampa	137.7 (10.44%)	127.9 (9.78%)	82.7 (8.15%)
Plant city	64.3 (8.38%)	46.1 (7.14%)	43.9 (7.14%)
Kissimmee	93.9 (4.34%)	75.8 (3.86%)	73.8 (3.38%)
Orlando	115.5 (19.55%)	78.9 (16.29%)	71.2 (15.09%)
Sanford	166.0 (7.66%)	139.6 (7.25%)	112.8 (6.45%)
Daytona Beach	138.4 (10.33%)	116.8 (9.50%)	110.4 (9.09%)
Average	119.3	97.5	82.4

The percentage of the secondary crash for the Orlando region is higher compared to other cities in the study. There are likely a number of factors which contribute to this

phenomenon including, construction and maintenance, unique lane geometry of the Orlando section, high traffic volumes, and urban setting. However, one factor that is unique to Orlando and likely plays a significant role in the secondary crashes is the prevalence of unfamiliar drivers. Orlando is one of the top tourist destinations in the USA, receiving approximately 75 million domestic and international visitors annually. Unfamiliar drivers tend to be more distracted, as they seek route guidance via roadside signs or GPS system. Such conditions likely contribute to secondary crashes.

Traffic incident management (TIM) programs are designed to detect, respond to, and clear traffic incident scenes. These multi-disciplinary efforts aim to safely and quickly treat victims, clear vehicle wreckage, and restore the flow of traffic. Effective TIM programs reduce incident duration and affect, while improving safety for crash victims, the motoring public, and emergency responders. The adoption of a planned and coordinated approach to protect the incident scene, modify the flow of traffic, and separate crash victims and responders from the motorist may have a profound impact on the severity and frequency of secondary crashes [49]. The results of this research including probability distribution of secondary crashes and the secondary event average time could inform TIM toward a targeted clearance time. Decreasing the time needed to clear a primary incident reduces the exposure of the scene to secondary crashes. This research found that secondary crashes, on average occurred between 37 and 96 minutes after the primary incident, when accounting for the natural fluctuations in traffic and incident patterns. This may suggest that reducing the primary incident impact duration below this threshold, could significantly reduce the likelihood of secondary crashes. While the model

is parameterized to a specific highway in this work, the approach is generic and can be used for the any crash and traffic datasets.

5. CONCLUSION

A novel approach to identify secondary crashes is investigated using temporal self-exciting point process model. The proposed approach is used to model crash events on I-4 highway in Florida, USA, spanning a period of three years. The model has the ability to analyze and classify crash events data using either a stationary background rate, a sinusoidal or a piecewise non-stationary background rates, for periodic variation in crash events data. After fitting the model to the data and optimizing the model parameters, queue time values resulting from primary crashes are determined and used to classify the crash events into primary and secondary events. The stationary background rate model failed to appropriately fit the data, since it is based on the assumption that the crash events are invariant to any external factors. However, the data from I-4 exhibits periodic variation with weekly and daily trends based on rush hours period inherent to traffic on highways. Using non-stationary background rate models, we were able to accurately fit the crash data and obtain queue time curves with peak on Fridays and trough on Sundays similar to the crash data. Using the sinusoidal non-stationary background rate model, we found that 3.38% to 15.09% of the traffic incidents from the crash data on I-4 are secondary. The proposed models in this work can be used to create policies and countermeasures that aim to reduce the risk of secondary crashes.

REFERENCES

1. Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*. 1971;58(1):83–90.
2. Ait-Sahalia Y, Cacho-Diaz J, Laeven RJA. Modeling financial contagion using mutually exciting jump processes. *J financ econ*. 2015;117(3):585–606.
3. Mohler GO, Short MB, Brantingham PJ, Schoenberg FP, Tita GE. Self-exciting point process modeling of crime. *J Am Stat Assoc*. 2011;106(493):100–8.
4. Mitchell L, Cates ME. Hawkes process as a model of social interactions: a view on video dynamics. *J Phys A Math Theor*. 2009;43(4):45101.
5. Towers S, Gomez-Lievano A, Khan M, Mubayi A, Castillo-Chavez C. Contagion in mass killings and school shootings. *PLoS One*. 2015;10(7):e0117259.
6. Islam S, Namilae S, Prazenica R, Liu D. Fuel shortages during hurricanes: Epidemiological modeling and optimal control. *PLoS One*. 2020;15(4):e0229957.
7. Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc*. 1988;83(401):9–27.
8. Zhuang J, Ogata Y, Vere-Jones D. Stochastic declustering of space-time earthquake occurrences. *J Am Stat Assoc*. 2002;97(458):369–80.
9. Jalilian A. ETAS: an R package for fitting the space-time ETAS model to earthquake data. *J Stat Softw*. 2019;88(1):1–39.
10. Fox EW. Estimation and inference for self-exciting point processes with applications to social networks and earthquake seismology. UCLA; 2015.

11. Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J. Seismic: A self-exciting point process model for predicting tweet popularity. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. 2015. p. 1513–22.
12. Dassios A, Zhao H. Exact simulation of Hawkes process with exponentially decaying intensity. *Electron Commun Probab*. 2013;18.
13. Bertozzi AL, Coronges K, Fox EW, Schoenberg FP, Short MB. Self-Exciting Point Process Models for E-mail Networks.
14. Lewis E, Mohler G, Brantingham PJ, Bertozzi AL. Self-exciting point process models of civilian deaths in Iraq. *Secur J*. 2012;25(3):244–64.
15. Xu C, Xu S, Wang C, Li J. Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression. *Phys A Stat Mech its Appl* [Internet]. 2019;524:121–9. Available from: <https://doi.org/10.1016/j.physa.2019.03.036>
16. Kitali AE, Alluri P, Sando T, Haule H, Kidando E, Lentz R. Likelihood estimation of secondary crashes using Bayesian complementary log-log model. *Accid Anal Prev*. 2018;119:58–67.
17. Sarker AA, Paleti R, Mishra S, Golias MM, Freeze PB. Prediction of secondary crash frequency on highway networks. *Accid Anal Prev* [Internet]. 2017;98:108–17. Available from: <http://dx.doi.org/10.1016/j.aap.2016.09.019>
18. Zhang Z, Liu Y, Han LD, Freeze PB. Secondary Crash Identification using

- Crowdsourced Waze User Reports. *Transp Res Rec J Transp Res Board*. 2021;2675(10):853–62.
19. Salek MS, Jin W, Khan SM, Chowdhury M, Gerard P, Basnet SB, et al. Assessing the likelihood of secondary crashes on freeways with Adaptive Signal Control System deployed on alternate routes. *J Safety Res* [Internet]. 2021;76:314–26. Available from: <https://doi.org/10.1016/j.jsr.2020.12.020>
 20. Florida Department of Transportation (FDOT). State Safety Office Geographic Information System (SSOGis) Crash Query Tool User Manual Version 1.2.
 21. Fundamentals of Transportation/Shockwaves [Internet]. Available from: https://en.wikibooks.org/wiki/Fundamentals_of_Transportation/Shockwaves
 22. Ogata Y. Space-time point-process models for earthquake occurrences. *Ann Inst Stat Math*. 1998;50(2):379–402.
 23. Fox EW, Short MB, Schoenberg FP, Coronges KD, Bertozzi AL. Modeling e-mail networks and inferring leadership using self-exciting point processes. *J Am Stat Assoc*. 2016;111(514):564–84.
 24. Interstate-4 [Internet]. Available from: https://en.wikipedia.org/wiki/Interstate_4
 25. Moore JE, Giuliano G, Cho S. Secondary accident rates on Los Angeles freeways. *J Transp Eng*. 2004;130(3):280–5.
 26. Hirunyanitiwattana W, Mattingly SP. Identifying secondary crash characteristics for California highway system. 2006.
 27. Zhan C, Shen L, Hadi MA, Gan A. Understanding the characteristics of secondary

- crashes on freeways. 2008.
28. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974;19(6):716–23.
 29. Santosa SP, Mahyuddin AI, Sunoto FG. Anatomy of injury severity and fatality in Indonesian traffic accidents. *J Eng Technol Sci.* 2017;49(3):412–22.
 30. Raub RA. Occurrence of secondary crashes on urban arterial roadways. *Transp Res Rec.* 1997;1581(1):53–8.
 31. Kopitch L, Saphores J-DM. Assessing effectiveness of changeable message signs on secondary crashes. 2011.
 32. Sun CC, Chilukuri V. Dynamic incident progression curve for classifying secondary traffic crashes. *J Transp Eng.* 2010;136(12):1153–8.
 33. Chung Y. Identifying primary and secondary crashes from spatiotemporal crash impact analysis. *Transp Res Rec.* 2013;2386(1):62–71.
 34. Kalair K, Connaughton C, Alaimo Di Loro P. A non-parametric Hawkes process model of primary and secondary accidents on a UK smart motorway. *J R Stat Soc Ser C (Applied Stat.)* 2021;70(1):80–97.
 35. Yang H, Ozbay K, Xie K. Assessing the risk of secondary crashes on highways. *J Safety Res.* 2014;49:143-e1.
 36. Zhan C, Gan A, Hadi M. Identifying secondary crashes and their contributing factors. *Transp Res Rec.* 2009;2102(1):68–75.
 37. Karlaftis MG, Richards NJ, Latoski SP, Sinha KC. An empirical analysis of secondary

- crash causes. In: Proceedings of the 77th Annual Meeting of the Transportation Research Board. 1998.
38. Chang GL, Rochon S. Performance evaluation and benefit analysis for chart—the real-time incident management system (year 2010). Final Report Univ Maryl. 2011;20.
 39. Zhang H, Khattak A. What is the role of multiple secondary incidents in traffic operations? *J Transp Eng.* 2010;136(11):986–97.
 40. Chou C-S, Miller-Hooks E. Simulation-based secondary incident filtering method. *J Transp Eng.* 2010;136(8):746–54.
 41. Yang H, Bartin B, Ozbay K. Mining the characteristics of secondary crashes on highways. *J Transp Eng.* 2014;140(4):4013024.
 42. Xu C, Liu P, Yang B, Wang W. Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data. *Transp Res part C Emerg Technol.* 2016;71:406–18.
 43. Imprialou M-IM, Orfanou FP, Vlahogianni EI, Karlaftis MG. Methods for defining spatiotemporal influence areas and secondary incident detection in freeways. *J Transp Eng.* 2014;140(1):70–80.
 44. Park H, Haghani A. Real-time prediction of secondary incident occurrences using vehicle probe data. *Transp Res Part C Emerg Technol.* 2016;70:69–85.
 45. Sarker AA, Naimi A, Mishra S, Golias MM, Freeze PB. Development of a secondary crash identification algorithm and occurrence pattern determination in large scale multi-facility transportation network. *Transp Res Part C Emerg Technol.*

2015;60:142–60.

46. Mishra S, Golias M, Sarker A, Naimi A. Effect of primary and secondary crashes: Identification, visualization, and prediction. 2016.
47. Junhua W, Boya L, Lanfang Z, Ragland DR. Modeling secondary accidents identified by traffic shock waves. *Accid Anal Prev*. 2016;87:141–7.
48. Wang J, Xie W, Liu B, Ragland DR. Identification of freeway secondary accidents with traffic shock wave detected by loop detectors. *Saf Sci*. 2016;87:195–201.
49. US Department of Transportation, Federal Highway Administration. (2019). National TIM Responder Training Program Quarterly Update. Retrieved 2020 [Internet]. Available from: https://transops.s3.amazonaws.com/uploaded_files/Talking TIM 052219 Webinar PPT - TIM Trng Quarterly Update - Final.pdf.