

Point-Process modeling of Secondary Crashes

Samarth Motagi

Embry-Riddle Aeronautical University

Sirish Namilae (✉ namilaes@erau.edu)

Embry-Riddle Aeronautical University

Audrey Gbaguidi

Embry-Riddle Aeronautical University

Scott Parr

Embry-Riddle Aeronautical University

Dahai Liu

Embry-Riddle Aeronautical University

Research Article

Keywords: Self-exciting Temporal Point Process, Background Rate Function, Queue Times, Non-Stationary Functions

Posted Date: April 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-388055/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Point-Process modeling of Secondary Crashes

Samarth Motagi¹, Sirish Namilaes^{1*}, Audrey Gbaguidi¹, Scott Parr² and Dahai Liu³

¹Dept. of Aerospace Engin., ² Dept. of Civil Engin., and ³College of Aviation
Embry-Riddle Aeronautical University, Daytona Beach, Florida 32114, USA

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 I4 Highway.

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 States¹. In 2016 these figures increased to over 37,000 deaths on United States roadways². 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 secondary³. 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

* Corresponding Author Email: namilaes@erau.edu

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^{4,5,6,7,8}. We have recently utilized epidemiological models to study the spread of fuel shortages during hurricane evacuations⁹. 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 earthquakes¹⁰, 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^{10,11,12,13,14}, several researchers have applied this concept to other clustered societal problems. For example, Mohler et al.⁶ used the self-exciting point process model to understand crime events; Zhao et al.¹⁴ used a self-exciting point process model to predict tweet popularity by modeling information cascades. Dassios et al.¹⁵ modeled the contagion risks or clustering events in finance and insurance while Bertozzi et al.¹⁶ used Hawke's point process models to understand the email communication between individuals in an organization. Towers et al.⁸ employed the same approach to identify the contagion in mass killings and school shootings. Lewis et al.¹⁷ characterized the temporal patterns of violent civilian deaths in Iraq using a self-exciting point process model.

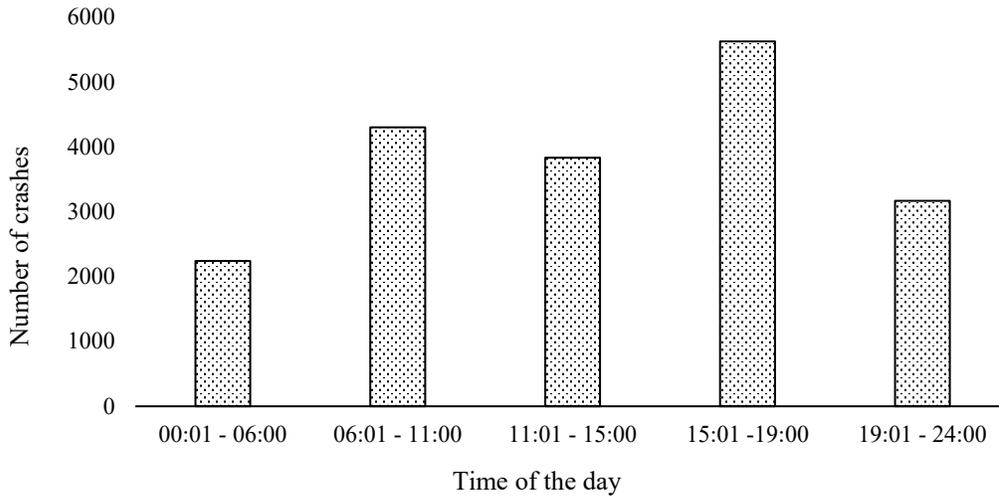
In this paper, a novel approach for identifying secondary crashes using the self-exciting point process model is formulated. To account for periodic variations in traffic events, corresponding periodic background rate variations are introduced. The model is employed to analyze the crash events on a dataset 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.

Methodology

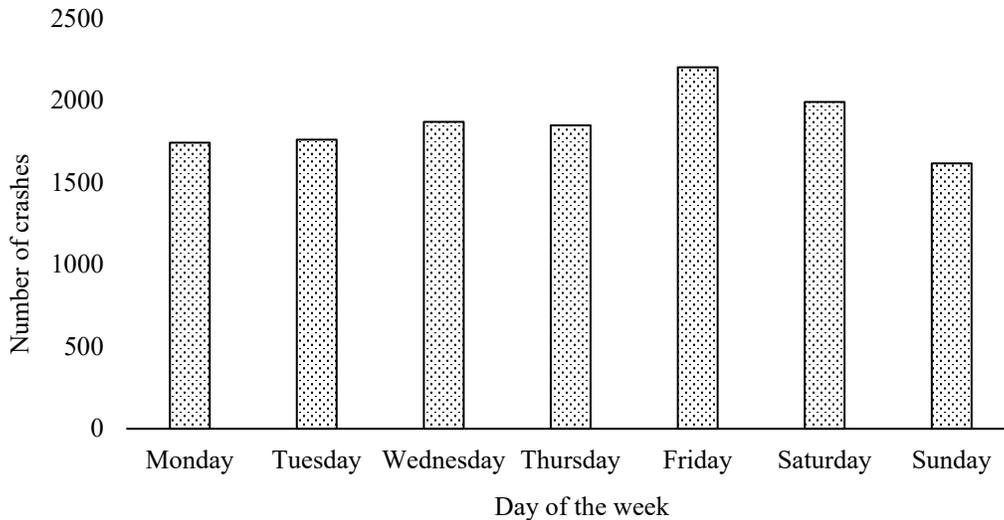
Traffic and Crash Data

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 2015-2017¹⁸. 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 I4 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.

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 queuing¹⁹. 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 function²⁰ is:

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

To obtain the maximum likelihood estimate (MLE) of θ , the Davidon-Fletcher-Powell method is used as a gradient-based nonlinear optimization procedure²⁰. A numerical approximation of the integral terms in the log-likelihood function is first computed. A probabilistic approach previously used for declustering an earthquake catalog¹¹ 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.

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.¹⁷ 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.²¹ 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 use two functions: (a) a step function and (b) a sinusoidal function. A step function is 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:

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

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:

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

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:

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

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 Jalilian¹² is 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.

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)¹⁸. The interstate highway I-4 is located entirely within the state of Florida, spanning 132.3 miles²². 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 are chosen because their corresponding data contains more than 500 crash events. The other cities modeled here have at least 200 crash events.

Spatial and temporal thresholds from a primary incident are often used to analyze secondary crashes. Recent studies by multiple investigators^{23,24,25} indicate that the spatial threshold is about 2-miles. Therefore, we analyze the data within a threshold of 2-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)²⁶ 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 2-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, no 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 |

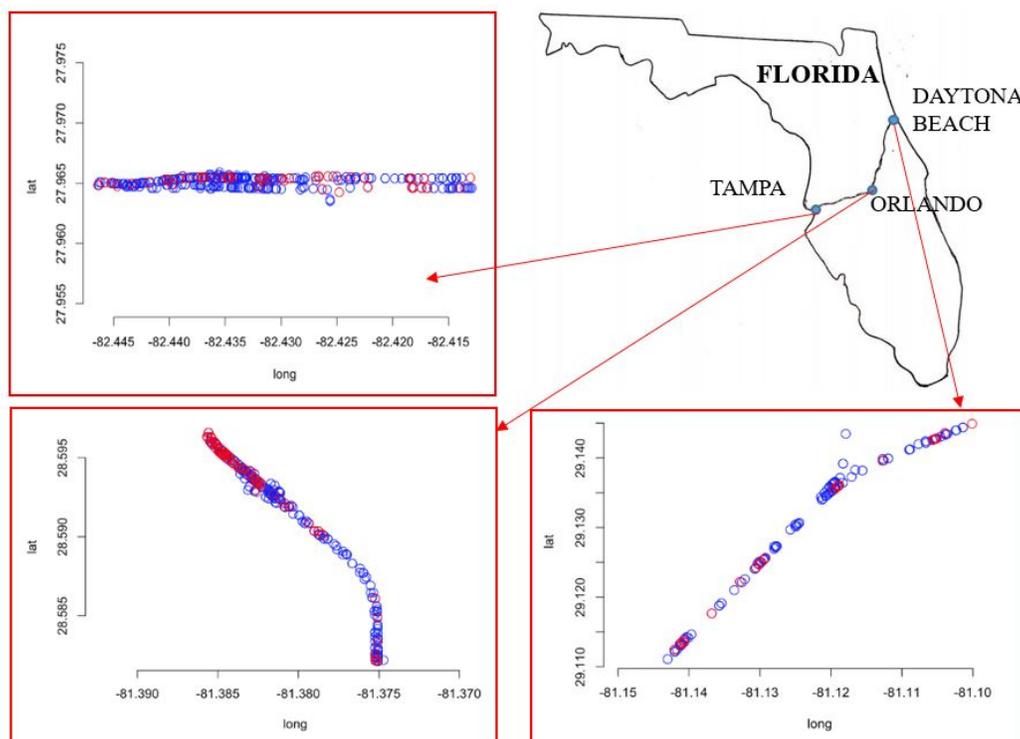


Figure 2: Primary and secondary crashes classified from the model for Tampa, Orlando and Daytona Beach along the I-4 highway. Secondary crashes are shown in red.

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, no 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 | mu (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 |
| Sinusoidal | 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 |

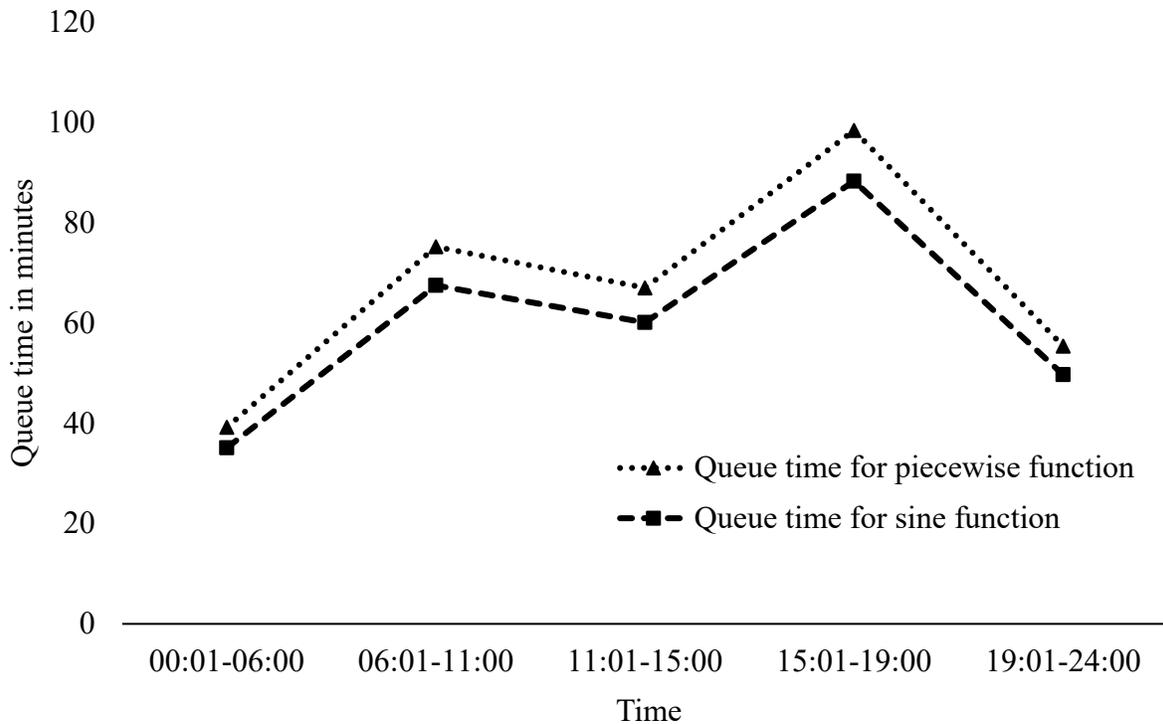


Figure 3: Queue time for Orlando region during different time of the day.

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-hour²⁷ 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

No 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)²⁶ 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 outperform the piecewise function. Similarly, the sinusoidal model exhibits lowest AIC for the hourly variation as well.

Table 4

Comparison of models using AIC values.

| Cities | AIC values for stationary background rate | AIC values for weekly trend | | AIC values for daily trend | |
|---------------|---|-----------------------------|----------------|----------------------------|----------------|
| | 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 Beach | 1198.58 | 1192.69 | 1168.87 | 1178.41 | 1163.56 |

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^{23-25,28-29} 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 methods³⁰⁻⁵⁰ 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.²³; 7.5% in Chung³¹; 6-7% in Kalair et al.³²; 8.42% in Yang et al.³³; 5.22 % in Zhan et al.²⁵; 5.53 % in Kopitch et al.²⁹; 3.23% in Zhan et al.³⁴ 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 ²⁸ , Karlaftis et al ³⁶ . | Static | 1-mile and 15 minutes |

| | | |
|---|---------|---|
| Moore et al. ²³ . | Static | 2-mile and 2 hours; same for opposite direction |
| Hirunyanitiwattana et al. ²⁴ . | Static | 2-mile and 1 hour |
| Zhan et al. ²⁵ . | Static | 2-mile and 15 minutes |
| Chang et al. ³⁷ | Static | 2-mile, 2 hours and 0.5-mile, 0.5 hour (opposite direction) |
| Sun et al. ³⁰ | Dynamic | Incident progression curve |
| Zhang et al. ³⁸ | Dynamic | Queue length estimations |
| Zhan et al. ³⁴ | Dynamic | Cumulative arrival and departure plots |
| Chou et al. ³⁹ | Dynamic | Simulation-based methods |
| Yang et al. ⁴⁰ , Xu et al. ⁴¹ , Kitali et al. ⁴² , Chung et al. ³¹ | 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.

| City | Average queue time (minutes) and % secondary crashes | | |
|----------------|--|---------------------------|----------------------------|
| | 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 crashes³⁵. 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.

Conclusions and Summary

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 find 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.

Reference

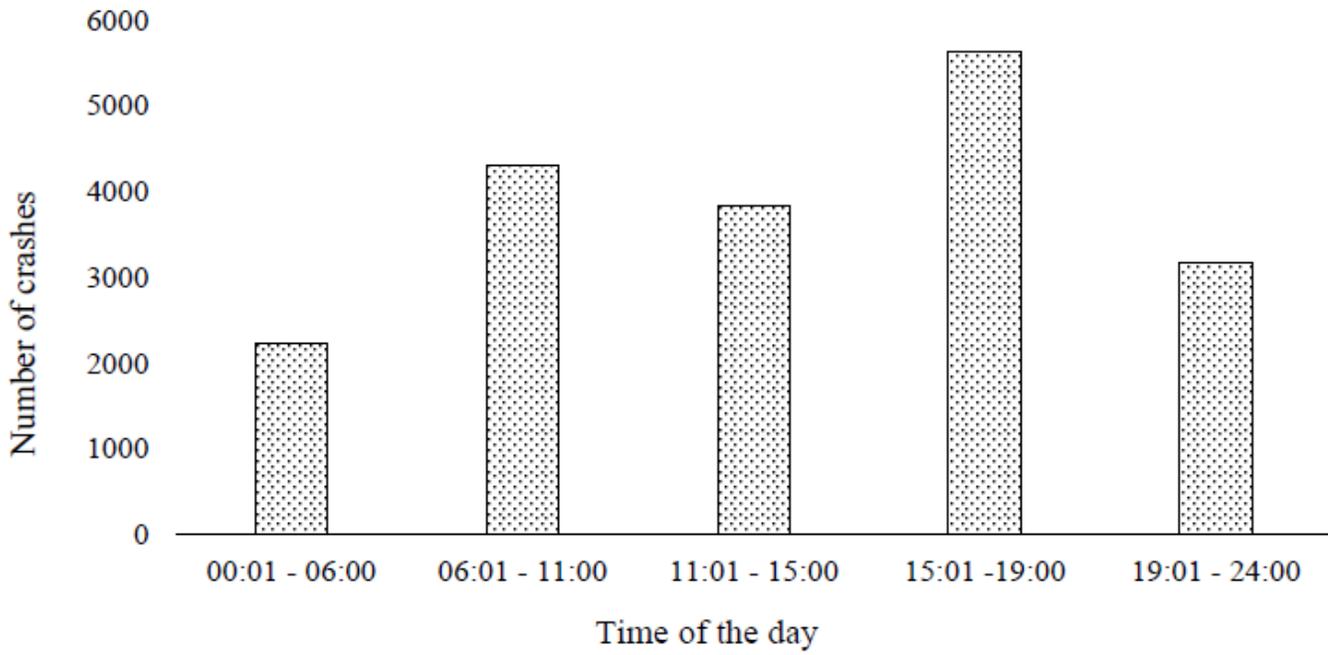
1. National Highway Traffic Safety Administration (NHTSA). 2015 Motor Vehicle Crashes: Overview NHTSA report. *U.S. Department of Transportation* 1–9 <https://crashstatsnhtsa.dot.gov/Api/Public/ViewPublication/812318> (2015).
2. NHTSA. Traffic Safety Facts 2017: A Compilation of Motor Vehicle Crash Data. *Dot Hs 812 806* <https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812806> (2019).
3. Goodall, N. J. Probability of secondary crash occurrence on freeways with the use of private-sector speed data. *Transp. Res. Rec.* **2635**, 11–18 (2017).
4. Hawkes, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90 (1971).
5. Aït-Sahalia, Y., Cacho-Diaz, J. & Laeven, R. J. A. Modeling financial contagion using mutually exciting jump processes. *J. financ. econ.* **117**, 585–606 (2015).
6. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. & Tita, G. E. Self-exciting point process modeling of crime. *J. Am. Stat. Assoc.* **106**, 100–108 (2011).
7. Mitchell, L. & Cates, M. E. Hawkes process as a model of social interactions: a view on video dynamics. *J. Phys. A Math. Theor.* **43**, 45101 (2009).
8. Towers, S., Gomez-Lievano, A., Khan, M., Mubayi, A. & Castillo-Chavez, C. Contagion

- in mass killings and school shootings. *PLoS One* **10**, e0117259 (2015).
9. Islam, S., Namilae, S., Prazenica, R. & Liu, D. Fuel shortages during hurricanes: Epidemiological modeling and optimal control. *PLoS One* **15**, e0229957 (2020).
 10. Ogata, Y. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**, 9–27 (1988).
 11. Zhuang, J., Ogata, Y. & Vere-Jones, D. Stochastic declustering of space-time earthquake occurrences. *J. Am. Stat. Assoc.* **97**, 369–380 (2002).
 12. Jalilian, A. ETAS: an R package for fitting the space-time ETAS model to earthquake data. *J. Stat. Softw.* **88**, 1–39 (2019).
 13. Fox, E. W. Estimation and inference for self-exciting point processes with applications to social networks and earthquake seismology. (2015).
 14. Zhao, Q., Erdogdu, M. A., He, H. Y., 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* 1513–1522 (2015).
 15. Dassios, A. & Zhao, H. Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Probab.* **18**, (2013).
 16. Bertozzi, A. L., Coronges, K., Fox, E. W., Schoenberg, F. P. & Short, M. B. Self-Exciting Point Process Models for E-mail Networks.
 17. Lewis, E., Mohler, G., Brantingham, P. J. & Bertozzi, A. L. Self-exciting point process models of civilian deaths in Iraq. *Secur. J.* **25**, 244–264 (2012).
 18. Florida Department of Transportation (FDOT). State Safety Office Geographic Information System (SSOGis) Crash Query Tool User Manual Version 1.2.
 19. Fundamentals of Transportation/Shockwaves. https://en.wikibooks.org/wiki/Fundamentals_of_Transportation/Shockwaves.
 20. Ogata, Y. Space-time point-process models for earthquake occurrences. *Ann. Inst. Stat. Math.* **50**, 379–402 (1998).
 21. Fox, E. W., Short, M. B., Schoenberg, F. P., Coronges, K. D. & Bertozzi, A. L. Modeling e-mail networks and inferring leadership using self-exciting point processes. *J. Am. Stat. Assoc.* **111**, 564–584 (2016).
 22. Interstate-4. https://en.wikipedia.org/wiki/Interstate_4.
 23. Moore, J. E., Giuliano, G. & Cho, S. Secondary accident rates on Los Angeles freeways. *J. Transp. Eng.* **130**, 280–285 (2004).
 24. Hirunyanitiwattana, W. & Mattingly, S. P. *Identifying secondary crash characteristics for California highway system*. (2006).
 25. Zhan, C., Shen, L., Hadi, M. A. & Gan, A. *Understanding the characteristics of secondary crashes on freeways*. (2008).
 26. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
 27. Santosa, S. P., Mahyuddin, A. I. & Sunoto, F. G. Anatomy of injury severity and fatality

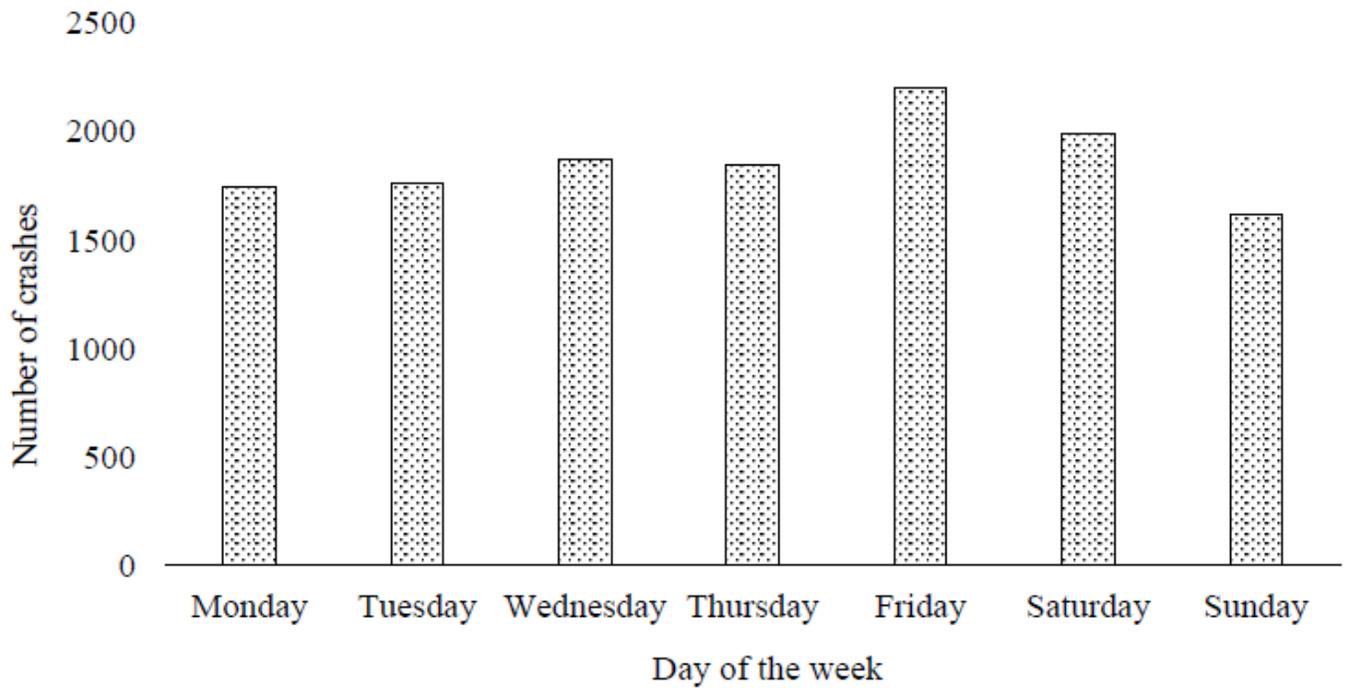
- in Indonesian traffic accidents. *J. Eng. Technol. Sci.* **49**, 412–422 (2017).
28. Raub, R. A. Occurrence of secondary crashes on urban arterial roadways. *Transp. Res. Rec.* **1581**, 53–58 (1997).
 29. Kopitch, L. & Saphores, J.-D. M. *Assessing effectiveness of changeable message signs on secondary crashes.* (2011).
 30. Sun, C. C. & Chilukuri, V. Dynamic incident progression curve for classifying secondary traffic crashes. *J. Transp. Eng.* **136**, 1153–1158 (2010).
 31. Chung, Y. Identifying primary and secondary crashes from spatiotemporal crash impact analysis. *Transp. Res. Rec.* **2386**, 62–71 (2013).
 32. 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.)* **70**, 80–97 (2021).
 33. Yang, H., Ozbay, K. & Xie, K. Assessing the risk of secondary crashes on highways. *J. Safety Res.* **49**, 143-e1 (2014).
 34. Zhan, C., Gan, A. & Hadi, M. Identifying secondary crashes and their contributing factors. *Transp. Res. Rec.* **2102**, 68–75 (2009).
 35. US Department of Transportation, Federal Highway Administration. (2019). National TIM Responder Training Program Quarterly Update. Retrieved 2020. https://transops.s3.amazonaws.com/uploaded_files/Talking TIM 052219 Webinar PPT - TIM Trng Quarterly Update - Final.pdf.
 36. Karlaftis, M. G., Richards, N. J., Latoski, S. P. & Sinha, K. C. An empirical analysis of secondary crash causes. in *Proceedings of the 77th Annual Meeting of the Transportation Research Board* (1998).
 37. Chang, G. L. & Rochon, S. Performance evaluation and benefit analysis for chart—the real-time incident management system (year 2010). *Final Report. Univ. Maryl.* **20**, (2011).
 38. Zhang, H. & Khattak, A. What is the role of multiple secondary incidents in traffic operations? *J. Transp. Eng.* **136**, 986–997 (2010).
 39. Chou, C.-S. & Miller-Hooks, E. Simulation-based secondary incident filtering method. *J. Transp. Eng.* **136**, 746–754 (2010).
 40. Yang, H., Bartin, B. & Ozbay, K. Mining the characteristics of secondary crashes on highways. *J. Transp. Eng.* **140**, 4013024 (2014).
 41. 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.* **71**, 406–418 (2016).
 42. Kitali, A. E. *et al.* Likelihood estimation of secondary crashes using Bayesian complementary log-log model. *Accid. Anal. Prev.* **119**, 58–67 (2018).
 43. Imprialou, M.-I. M., Orfanou, F. P., Vlahogianni, E. I. & Karlaftis, M. G. Methods for defining spatiotemporal influence areas and secondary incident detection in freeways. *J. Transp. Eng.* **140**, 70–80 (2014).
 44. Park, H. & Haghani, A. Real-time prediction of secondary incident occurrences using

- vehicle probe data. *Transp. Res. Part C Emerg. Technol.* **70**, 69–85 (2016).
45. Sarker, A. A., Naimi, A., Mishra, S., Golias, M. M. & Freeze, P. B. Development of a secondary crash identification algorithm and occurrence pattern determination in large scale multi-facility transportation network. *Transp. Res. Part C Emerg. Technol.* **60**, 142–160 (2015).
 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, D. R. Modeling secondary accidents identified by traffic shock waves. *Accid. Anal. Prev.* **87**, 141–147 (2016).
 48. Wang, J., Xie, W., Liu, B. & Ragland, D. R. Identification of freeway secondary accidents with traffic shock wave detected by loop detectors. *Saf. Sci.* **87**, 195–201 (2016).

Figures



(a)



(b)

Figure 1

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

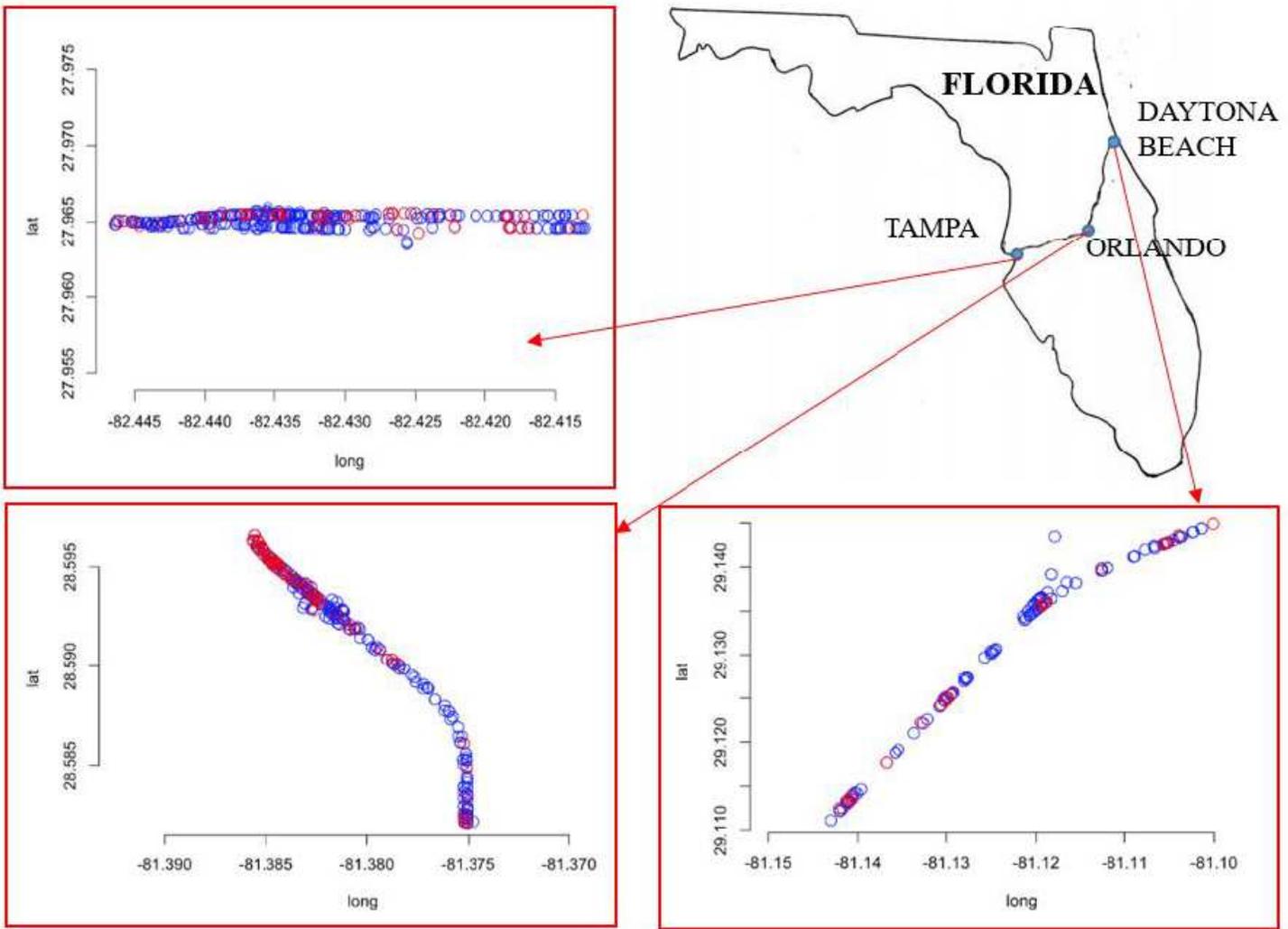


Figure 2

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.

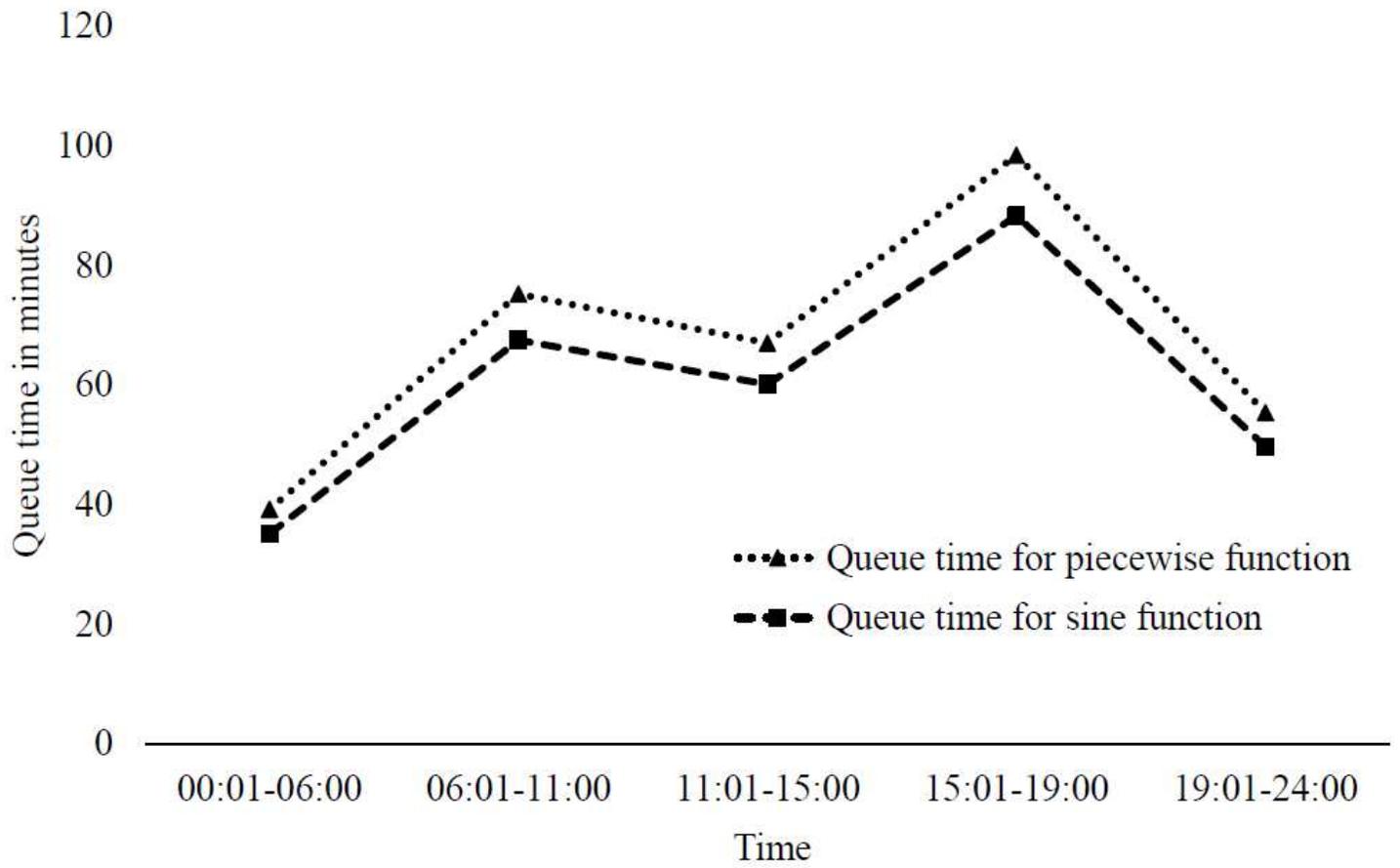


Figure 3

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