A real-time crash risk estimation framework for signalized intersections

Yasir Ali[1], Md. Mazharul Haque[2,*], Simon Washington[3]

[1]*School of Civil & Environmental Engineering, Faculty of Engineering, Queensland University of Technology, Brisbane 4001, Qld Australia, y2.ali@qut.edu.au*
[2]*School of Civil & Environmental Engineering, Faculty of Engineering, Queensland University of Technology, Brisbane 4001, Qld Australia, m1.haque@qut.edu.au*
[3]*Advanced Mobility Analytics Group Pty Ltd, Brisbane 4001, Qld, Australia, simon@amagroup.io*

1. Introduction

Existing crash risk models for signalized intersections mainly suffer from three issues: (i) they are reactive because of the dependence on crash data, (ii) they utilize limited information available in the crash data for crash risk assessments, and (iii) they lack real-time capabilities. A fundamental issue in crash risk evaluation requires an unusually large number of crashes to occur at an urban intersection before that location can be assessed for safety issues. Similarly, the current safety assessments are mainly based on police-reported crash data, which suffer from significant shortcomings such as under-reporting, low sample means, limited behavioral information, and omitted variable bias in the data. Along the same lines, the current reactive road safety assessment cannot assess real-time risks at signalized intersections, and as such, real-time risk mitigation strategies could not be developed.

Statistical models that often rely on historical crash data have been frequently applied to understand causal relationships as well as to estimate crash risk in real-time. However, these studies lead to two main challenges. First, given that these models are applied in a classification context to predict whether a crash would occur or otherwise, it can only provide “yes” or “no” outcomes rather than insights into the risky conditions. Second, loop detectors and similar other data collection procedures provide little to no information about micro-driving behavior such as hard braking, acceleration, and swerving.

Another related issue is that much of the existing research focuses on real-time crash risk prediction on motorways, perhaps because of ease of data accessibility (usually collected by sensors and detectors). Comparatively, signalized intersections have received less attention despite being one of the major sources of crashes and injuries. For instance, during 2018, fatal crashes at signalized intersections accounted for about 20% of the total crashes in Queensland, Australia [1].

A thorough review of real-time crash risk literature uncovers several limitations. First, the definition of real-time is ambiguous. Some studies call “real-time” to the use of trajectory data [2], while others assume “real-time” as the prediction in a short time interval, like 5-min [3]. Second, the intersections have received less attention for real-time safety analysis relative to highway or motorway sections. Third, much of the existing studies on real-time safety rely on loop detector data, which provide insufficient information about the crash mechanism and its contributing factors. Finally, crash prediction models generally require an unusually large number of crashes over multiple years, limiting the applications for real-time safety analysis. To this end, Extreme Value Theory (EVT) approach has been frequently applied and received significant attention in the literature because of the following peculiarities. First, the EVT approach does not require historical crash records to estimate crash risks. Second, crashes estimated using EVT rely on traffic conflict and fit well within Hydén’s pyramid, suggesting that traffic extremes identified through conflicts can provide insights into the crash mechanism. The suitability of EVT models for crash predictions is confirmed by several studies (e.g., Ali, et al. [4] and Zheng and Sayed [5]). It is worth noting here that Zheng and Sayed [5] analyzed crash risk at a signal cycle; however, their study did not shed light on how crash risk varies across periods and whether these differences are statistically significant.

Motivated by these research gaps, this study presents a framework to estimate crash risk in real-time at signalized intersections. This framework leverages the power of the automated video analysis technique developed

[2] * Corresponding author. Tel.: + +61 7 3138 7195;
E-mail address: m1.haque@qut.edu.au

at the Queensland University of Technology (QUT) to extract conflict data and microscopic traffic information. In particular, this study employs an EVT approach to answer the following research questions:

1. How can EVT be leveraged to obtain real-time crash risk?
2. How crash risk varies with different time periods of the day (e.g., peak vs. off-peak hours)?
3. Does incorporating a covariate in the model yield better fit and crash estimates?

2. Modelling Framework

A Bayesian hierarchical modeling approach is adopted to estimate the crash risk in a signal cycle. The GEV distribution function is employed to model traffic extremes. Ensuring the positive scale parameter of GEV, it is reparametrized as GEV(μ, ϕ, ξ), where $\phi = log\sigma$. Assume z_{ij} be the i^{th} cycle maximum at site j, with j = 1 ,…, s, and i = 1, …,nj. Given that z_{ij} corresponds to the maximum value of a traffic conflict indicator for cycle i at site j, a GEV distribution indicates parameters for each site as $\mu_{ij}, \phi_{ij}, \xi_{ij}$ and GEV distribution function can be written as

$$G(z_{ij} < z|\mu_{ij}, \phi_{ij}, \xi_{ij}) = exp\left(-\left[1 + \xi_{ij}\left(\frac{z-\mu_{ij}}{exp(\phi_{ij})}\right)\right]^{-1/\xi_{ij}}\right) \tag{1}$$

To characterize traffic extremes, suitable covariates are included using an identity link function as

$$\begin{cases} \mu_{ij} = \alpha_{\mu 0} + \alpha_{\mu 1}X + \varepsilon_{\mu j} \\ \phi_{ij} = \alpha_{\phi 0} + \alpha_{\phi 1}X + \varepsilon_{\phi j}, \\ \quad \xi_{ij} = \alpha_{\xi 0} + \varepsilon_{\xi j} \end{cases} \tag{2}$$

where, $\alpha_{\mu 0}$, $\alpha_{\phi 0}$, and $\alpha_{\xi 0}$ are intercept terms corresponding to three model parameters,$\alpha_{\mu 1}$ and $\alpha_{\phi 1}$ are parameter estimates for the covariates X, and $\varepsilon_{\mu j}$, $\varepsilon_{\phi j}$, and $\varepsilon_{\xi j}$ are random error terms. To characterize the latent process, this layer assigns priors to parameters $\alpha_{\mu 0}$, $\alpha_{\phi 0}$, $\alpha_{\mu 1}$ and $\alpha_{\phi 1}$. As no prior information on how GEV parameters vary is available, uninformative priors for these parameters are adopted, which are assumed to follow a normal distribution with mean zero and large variance, i.e., $N(0,10^6)$.

3. Data

The proposed modeling framework is applied to rear-end crash estimation at the signal cycle level of a signalized intersection. Traffic movement data at a four-legged signalized intersection in Southeast Queensland, Australia (i.e., Logan Rd – Kessels Rd intersection) were captured by video cameras. A total of 96 hours of video data was recorded. An advanced computer vision technique was developed to extract the required information from raw video recordings. This platform utilized a deep neural network to detect signal timing, extract vehicle trajectories and conflict measures like time-to-collision. This method consists of several steps: camera calibration, object detection and tracking, prototype generation, prototype matching event generation, and conflict identification, as shown in Figure 1.

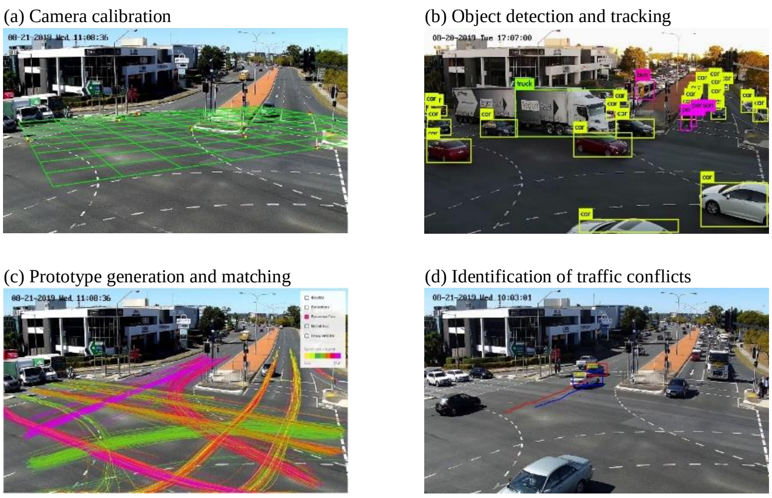

Figure 1: Trajectory data extraction process using artificial intelligence techniques

Using automated analysis, traffic conflicts are automatically collected. In this study, modified time-to-collision (MTTC) indicator is used. This platform was developed by Advanced Mobility Analytics Group (AMAG), which offers a world-leading platform in transport analytics from video data. This platform utilizes time-to-collision (TTC) measures with a TTC threshold of 3 s for identifying traffic conflicts.

4. Results

A Bayesian hierarchical model is developed including traffic volume as a covariate added to the location parameter. This model was estimated using two separate chains with different initial values. The total iterations were set as 50,000, while the first 20,000 were considered burn-in samples and thus removed. The posterior estimates were obtained from the remaining 30,000 iterations. Table 1 indicates the model estimation results. Three models are compared using DIC, and it is found that the models with covariate perform better than the stationary model.

Table 1: Summary of model estimation results

Parameter	location		scale		shape	DIC
	μ_0	μ_V	σ_0	σ_V	ξ_0	
mean	-0.5572	0.0020	0.2571	—	-0.4644	1086.74
s.d.	0.006	0.0002	0.0037	—	0.0075	
2.50%	-0.5301	0.0018	0.2503	—	-0.4771	
97.50%	-0.5057	0.0011	0.2644	—	-0.449	

μ_0 *and* μ_V *denote location intercept and traffic volume parameters, respectively*

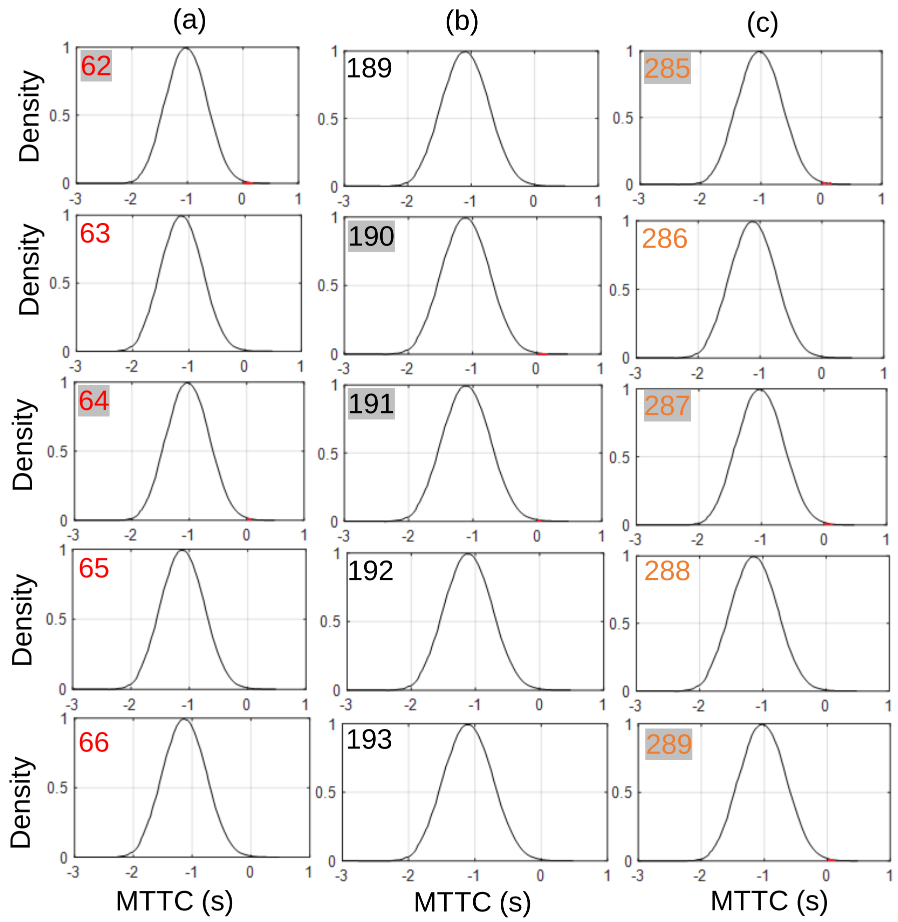

Figure. 2: Estimation GEV distributions for some representation cycles of one day for (a) morning peak hours, (b) off-peak hours, and (c) evening peak hours. Note that numbers within each subfigure indicate cycle number, and the highlighted cycle number indicates positive crash risk.

This model is used for generating GEV distributions for each cycle with its corresponding traffic volume. A typical example of some representative cycles for one day can be seen in Figure 2. The shape of the distribution is of paramount importance as it provides insights into crash-prone conditions. More specifically, the tail of a GEV distribution ending after the negated MTTC = 0 indicates a positive crash risk. Figure 1 shows that cycles 62, 64, 190, 191, 285, 287, 289 have positive crash risk (and are risky cycles) as their distributions have crossed the negated MTTC = 0 points. Further, a statistical analysis of GEV distributions is performed using a Kolmogorov–Smirnov test to compare the equality of distributions across signal cycles. At a 95% confidence level, it is observed that cycles with positive crash risk are significantly different from other cycles (p-value < 0.05), further confirming the difference in crash risk in different cycles.

5. Conclusions

This study applied a real-time crash estimation framework to estimate rear-end crash risk at a signal cycle level. Leveraging the benefits of Extreme Value Theory, a Block Maxima approach was applied by considering each cycle as a block from which the maximum value of negated MTTC was selected as an extreme. Extremes from all cycles were used as an input to a Bayesian hierarchical model. The proposed modelling framework was applied to a total of 96 hours of video data. GEV distributions were generated for each signal, and the shape of a GEV distribution was utilized to assess crash risk. It was found that some cycles were safe while the others were risky, as indicated by the tail of the GEV distribution that ended after negated MTTC = 0.

References

[1] DTMR, "2018 Summary Road Crash Report, Queensland Road Fatalities," Customer Services, Safety & Regulation Division, Department of Transport and Main Roads, Brisbane, Australia: Queensland Transport2019.

[2] L. Wang, M. Abdel-Aty, J. Lee, and Q. Shi, "Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors," *Accident Analysis & Prevention,* vol. 122, pp. 378-384, 2019.

[3] J. Yuan and M. Abdel-Aty, "Approach-level real-time crash risk analysis for signalized intersections," *Accident Analysis & Prevention,* vol. 119, pp. 274-289, 2018.

[4] Y. Ali, M. Haque, and Z. Zheng, "An Extreme Value Theory approach to estimate crash risk during mandatory lane-changing in a connected environment," *Analytic Methods in Accident Research,* p. 100193, 2021.

[5] L. Zheng and T. Sayed, "A novel approach for real time crash prediction at signalized intersections," *Transportation Research Part C,* vol. 117, p. 102683, 2020.