Accident Analysis and Prevention 183 (2023) 106973

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Non-stationary crash risk modelling of powered two-wheelers using extreme value analysis of surrogate crash events

Pranab Kar [a,*], Suvin P. Venthuruthiyil [b], Mallikarjuna Chunchu [a]

[a] Indian Institute of Technology Guwahati, India
[b] Indian Institute of Technology Hyderabad, India

ARTICLE INFO

Keywords:
Powered two-wheelers
Sideswipe conflict
Extreme value theory
Surrogate safety measure
Evasive actions
Multilane highways
Non-stationary

ABSTRACT

Evaluating the impact of evasive actions such as braking and steering on the crash risk assessment of vehicles is a scarce endeavor due to the lack of relevant data. This study uses Extreme Value Theory to investigate and model the effect of evasive actions on the sideswipe crash risk of powered two-wheelers (PTWs) moving on multilane rural highways. The crash risk was projected from the observed sideswipe conflicts that were quantified using a surrogate safety indicator called anticipated collision time (ACT). The vehicle trajectory data extracted from traffic videos, collected using an unmanned aerial vehicle, was used as the input for the analysis. The data was denoised using a state-of-the-art trajectory reconstruction method called recursively ensembled low pass filtering. Once the conflicts were identified from the trajectory data, the crash risk models were developed considering five covariates: maximum deceleration rate, maximum yaw rate, and the times spent in decelerating, accelerating, and steering during a sideswipe conflict. These covariates were used to capture the non-stationarity in the traffic conflict extremes. The best performing non-stationary model was selected by comparing the negative log-likelihood values with the stationary-one. The findings suggest that the PTWs experience significant sideswipe crash risk on four-lane (crash risk 0.09%) and six-lane (crash risk 0.17%) highways. The sideswipe crash risk of PTWs increases with the increase in the intensity of braking and steering actions measured in terms of maximum deceleration and yaw rates. Further, this study emphasizes that incorporating the effects of evasive actions in the crash risk estimation and developing non-stationary models could significantly improve the precision of crash frequency estimates. Based on the findings it can be concluded that for the safety improvement of PTWs on multilane highways, lane-restriction should be imposed which can increase the safety margin during sideswipe conflicts.

1. Introduction

Powered two-wheelers (PTWs) are regarded as vulnerable road users (VRUs) since they lack protective gear except for helmets (Damani & Vedagiri, 2021). World Health Organization (WHO) has reported that vulnerable road users, including PTWs, account for 28% of all traffic deaths worldwide (WHO, 2018). Notably, the PTWs constitute a significant proportion of road deaths in most developing countries; for example, 41% of road death victims in India are PTW riders (Haghani et al., 2022). Such a considerable fatality risk associated with PTWs has stimulated the growing interest in PTW safety studies, particularly in the lower- and middle-income countries (LMICs). Despite several attempts to understand PTW behavior, there is a considerable knowledge gap related to PTW safety (Haghani et al., 2022). Remarkably, the PTW behavior in LMICs is highly complex due to the unique traffic characteristics leading to significant safety challenges. Nevertheless, the relevant research from LMICs in this direction is relatively less. As a matter of fact, less than 10% of all accident studies are meant for LMICs (Haghani et al., 2022), even though the fatality rate of road users in LMICs is disproportionately high.

Most safety studies employ crash data to perform crash risk analysis. However, crash data has inherent issues such as under-reporting, inability to capture the driving mechanism that leads to crashes, and requiring an extensive data collection period for the modeling (Venthuruthiyil & Chunchu, 2022b; Zheng et al., 2021). Given these drawbacks of the crash data, researchers opted for the proactive safety approach that uses traffic conflicts as a precursor to crashes. Such an approach can help identify the factors that correlate traffic conflicts to

* Corresponding author.
E-mail addresses: pranabkar69@gmail.com (P. Kar), pv.suvin@ce.iith.ac.in (S.P. Venthuruthiyil), c.mallikarjuna@iitg.ac.in (M. Chunchu).

https://doi.org/10.1016/j.aap.2023.106973
Received 26 August 2022; Received in revised form 25 November 2022; Accepted 14 January 2023
Available online 18 January 2023

crash and remove the ethical dilemma of hoping for crashes to occur to address safety. Given the inherent benefits, there is a growing interest in using the proactive safety approach in modeling crash risk.

To develop crash risk models using a proactive safety approach, the extreme value theory (EVT) emerged as a powerful tool since it enables the extrapolation of rarely observed events (i.e., crashes) from the frequently observed events (i.e., traffic conflicts). This is evident from the recent use of EVT models in crash risk studies (Zheng, Sayed and Essa, 2019; Fu, Sayed and Zheng, 2020; Arun et al., 2021a, Arun et al., 2021b; Arun et al., 2022). Besides, EVT eliminates the mixing of several safety pyramid levels (Hydén, 1987) by identifying the true extremes (i.e., severe conflicts) from the potential conflicts. Therefore, the present study has also chosen EVT for the crash risk estimation of PTWs.

Extreme value theory assumes that the sampled extremes (i.e., severe conflicts) are independent and identically distributed (i.e., drawn from the same population). Crash risk models developed using such extremes are called stationary EVT models. However, in reality, most of the sampled extremes are heterogeneous in nature. Though splitting a heterogonous sample into homogeneous subsamples is one way, treating heterogeneity using statistical modelling is a more efficient strategy (Tarko, 2012).These models are called non-stationary since they capture the heterogeneity in the traffic conflict extremes. Most non-stationary EVT studies consider aggregated variables such as traffic volume and conflict volume to capture the heterogeneity (Fu et al., 2020). Only a few studies have used microscopic driver behavior variables such as spacing between the subject and the leading vehicle, speed of the subject vehicle to model the heterogeneity in the conflict extremes (Ali et al., 2022).

The objective of the present study is to investigate and model the crash risk of PTWs, incorporating the effects of micro-driving behavior variables or covariates for the multilane rural highways of LMICs using EVT. To fulfill this objective, this study selected a suitable conflict indicator to identify conflicts and then chose the factors that are influencing the crash risk of PTWs. Further, this study developed the non-stationary models by using the influencing factors and then selected the best-performing model by comparing them with the stationary model.

The remainder of the paper is organized as follows: Section 2 provides a summary of the studies discussing the crash risk analysis of PTWs based on the proactive safety approach. Section 3 elaborates the approach of crash risk analysis, and Section 4 explains the data collection process. Finally, Section 5 describes the results of the analysis and Section 6 ends with the concluding remarks of the study.

2. Literature review

This section discusses the past studies that models the crash risk using proactive methods and EVT, with a specific consideration to the non-stationarity. For an extensive review of the safety studies involving EVT, the readers may refer to the work of Zheng et al. (2021). Further, this section provides a brief discussion on the various covariates considered in the PTW crash risk modelling. A brief explanation of the limitation of prevalent conflict indicators used to identify the PTW conflicts and the workaround is also provided. Besides, the usefulness of the real field trajectory data is explained along with the need to assess the sideswipe crash risk of PTWs.

Table 1 summarizes the key aspects of non-stationary EVT studies. It is evident from the table that most EVT studies have considered aggregated traffic information (e.g., Songchitruksa & Tarko, 2006; Zheng et al., 2014a; Zheng et al., 2018) as covariates to capture the non-stationarity of the crash mechanisms. Though a few studies (Ali et al., 2022; Cavadas et al., 2020) have employed covariates that capture microscopic driving characteristics, the vehicle conflict data was generated from driving simulator experiments. Such data will lack the realism of actual field dynamics. Besides, most of the covariates considered were related to the closeness or proximity (i.e., Gap/Spacing) between the two conflicting road users. Nevertheless, Guo et al. (2018, 2019) have found that PTWs'crash risk is highly associated with evasive actions such as steering, braking, and acceleration than the proximity between two conflicting vehicles. Therefore, for the proper estimation of PTWs' crash risk, microscopic variables related to the evasive actions should be used. Accordingly, for the crash risk estimation, the present study has developed non-stationary EVT models considering the respective evasive action related microscopic variables.

Prior to modeling the non-stationarity, a suitable conflict indicator needs to be identified. Most studies used proximity measures such as Time-To-Collision (TTC) and Post-Encroachment-Time (PET) for estimating the crash risk (Arun et al., 2021a, Arun et al., 2021b). Tageldin et al. (2015) showed that evasive actions identify the PTW conflicts

Table 1
EVT studies addressing non-stationarity issues.

Study	EVT Model	Conflict Type	Conflict Indicator	Facility Type	Covariates Considered	Remarks
Using Naturalistic Driving Data from Videos						
Songchitruksa & Tarko (2006)	Block Maxima (BM)	Right Angled	Post encroachment time (PET)	Signalized intersection	Total volume, through volume, left-turn volume, conflict volume, & conflicting through volume	Crash frequency estimates & historical crash data match reasonably
Zheng et al. (2014b)	BM & Peak over Threshold (POT)	NA*	PET	Freeway	5-min traffic volume, fraction of oversized vehicles, & number of lane changes	POT performs better than BM
Zheng et al. (2018)	BM	Rear End	Time to Collision (TTC)	Signalized intersection	5-min left-turn volume & number of conflicts below a certain threshold	Number of conflicts below a threshold is a better exposure measure than the volume
Fu et al. (2020)	BM	Rear End	Modified TTC (a variant of TTC), PET, & Deceleration Rate to avoid crash (DRAC)	Signalized intersection	Traffic volume, shock wave area, & platoon ratio	Covariates improve the performance of the models
Using Driving Simulator Data						
(Farah & Azevedo, 2017)	BM & POT	Head on	TTC	Two-lane rural highway	Passing gap & duration, Speeds of passing & conflict vehicles	BM provides more stable results than POT
Cavadas et al. (2020)	BM	Head on & Rear End	TTC & Gap from the passed vehicle (a variant of PET)	Two-lane rural highway	Gaps of subject vehicle with front & opposing vehicles, Speeds of front and passing vehicles	Covariates improve the performance of the models
Ali et al. (2022)	BM & POT	NA*	Gap time for lane-changing (a variant of TTC)	Motorway (Four-lane highway)	Gap, Spacing, & Speed of the subject vehicle during the lane-changing event	POT outperforms BM

Note: NA* indicates the crash risk associated with lane-change maneuver rather than any specific conflict type.

much better than proximity measures such as TTC. Besides, studies (Laureshyn et al., 2010; Tageldin et al., 2015) have already pointed out that a single conflict indicator combining proximity and all evasive actions should be developed. Recently, Venthuruthiyil and Chunchu (2022) proposed a new conflict indicator called Anticipated Collision Time (ACT) that combines proximity and evasive actions using the shortest distance and the closing-in rate between two conflicting vehicles. The detailed description of ACT is provided in Section 3.

Table 1 shows some of the past studies that have obtained conflict information from driving simulator experiments or naturalistic driving data. The driving simulator generates data in a highly controlled environment. Such a controlled environment may not yield realistic results, and this issue might even become larger for the LMICs, where the interactions between PTWs and other vehicle types are intricate. Notably, obtaining accurate microscopic driving data for longer road stretches became a reality due to the recent advancements in image processing-based trajectory extraction tools such as SAVETRAX (Venthuruthiyil & Chunchu, 2022b).

Furthermore, it is to be noted that, on multilane highways, the PTWs frequently change lanes, increasing the chance of a sideswipe crash risk between the two vehicles (Puthan et al., 2021). Guo et al. (2019) found that PTWs apply strong swerving (or steering) evasive action to avoid collisions when they frequently change lanes and overtake vehicles. Unfortunately, most research related to PTWs mainly focused on rear-end collisions, with minimal studies on the sideswipe crash risk. Puthan et al. (2021) recently stressed the need for examining further the sideswipe crashes so that they can be included in the global safety standards of PTWs (ISO 13232). Venthuruthiyil et al. (2022) also found that PTWs experience more sideswipe conflicts in LMICs than the rear-end conflicts.

To summarise, EVT is a powerful tool for developing conflict-crash relationships. For the crash risk estimation of PTWs using EVT, the microscopic variables that define PTWs' evasive actions should be used rather than proximity related variables. Notably, the effect of evasive actions on the sideswipe crash risk modeling of PTWs has rarely been investigated. Given the inherent benefits of ACT, it might be the best possible conflict indicator for PTW conflict identification. Finally, the crash risk assessment of sideswipe conflicts requires further attention to improve the global safety standards of PTWs.

3. Methodology

This section begins with a detailed description of ACT and explains the process of sideswipe conflict identification from ACT profiles. Then, the peak over threshold approach (POT) of extreme value theory was elaborated. Further, this approach includes the threshold estimation of ACT using both graphical plots and statistical methods. Graphical plots include the mean residual life plot and the threshold stability plot. The initial selection of ACT threshold range was obtained from these two plots, and then various POT models were fitted across the threshold range. The best fitted POT model corresponds to the best ACT threshold. Then the crash risk estimation was explained and finally the non-stationary modeling was discussed for the required covariates defining the evasive action of PTW during the sideswipe conflict.

3.1. Description of ACT

ACT is the time remaining to collision based on the shortest distance between the vehicles and the closing-in rate in the shortest distance direction (Fig. 1a). ACT can be computed as:

$$T = \begin{cases} \dfrac{\delta}{\left(\dfrac{\partial\delta}{\partial t}\right)}, & \text{if} \dfrac{\partial\delta}{\partial t} > 0 \\ \infty, & \text{otherwise} \end{cases} \tag{1}$$

where, T is the time after which two vehicles would collide if they move with the same closing-in rate, δ is the shortest distance between the approaching vehicles at a time instant t_1 and $\left(\frac{\partial\delta}{\partial t}\right)$ indicates the rate at which the vehicles approach each other (closing-in rate). The closing-in rate encompasses the effect of all state variables that affect the likelihood of a potential conflict (Fig. 1b).

The closing-in rate of two vehicles is a function of their speed, acceleration, heading angle, and steering. It can be estimated by taking the resultant of speed, acceleration, and steering in the direction of the shortest distance between the vehicles and is shown in the Equation.

$$\frac{\partial\delta}{\partial t} = Rel\left(\vec{v}_{1-2}, \vec{v}_{2-1}\right) + Rel\left(\vec{a}_{1-2}, \vec{a}_{2-1}\right) \times t + Rel\left(\dot{\theta}_1, \dot{\theta}_2\right) \times \delta$$

where, $\vec{v}_{1-2}$, $\vec{a}_{1-2}$ are the components of speed and acceleration of vehicle-1 towards vehicle-2 in the direction of the shortest distance. $\vec{v}_{2-1}$, $\vec{a}_{2-1}$ are the components of speed and acceleration of vehicle-2 towards vehicle-1 in the direction of the shortest distance. $\dot{\theta}_1$, $\dot{\theta}_2$ are the steering rate of vehicle-1 and vehicle-2, respectively. Here, *Rel* is an operator that takes the vector sum of the quantities. The steering rate $\dot{\theta}$ is measured in terms of yaw rate which represents the angular velocity of the road-user rotation around the z-axis or the rate of change of heading angle (θ).

3.2. Identification of sideswipe conflicts

For the multilane highways in LMICs, the sideswipe conflict mostly happens between PTWs and fast-moving vehicles such as passenger cars, especially during a lane-changing or overtaking event. It is to be noted that the sideswipe conflict can also happen between two PTWs. Given the potential benefits of ACT in the proactive safety analysis, this study considered ACT as the SSM to identify sideswipe conflicts involving PTWs.

In the case of ACT, a conflict course exists only when the line drawn from the approaching vehicle's nearest corner crosses through either of the other vehicles' corner points. To identify a conflict as sideswipe, a line (parallel to the heading angle) projected from the approaching

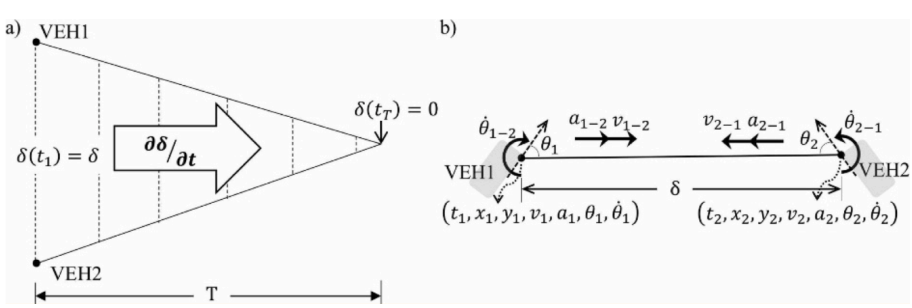


Fig. 1. Simplified illustration of the concept of Anticipated Time to collision (ACT); (a) Closing-in of two vehicles till collision; (b) Factors influencing the closing-in rate of two vehicles. (Source: Venthuruthiyil & Chunchu, 2022b).

vehicle's nearest corner towards the interacting vehicle must touch the side of the interacting vehicle, and both the vehicles should move in the same direction (Fig. 2). The ACT values for a sideswipe conflict can then be calculated for each PTW using Equation (1).

ACT is similar to TTC but functions differently and the conflict identification from continuous ACT profiles is similar to that of TTC profiles (Fig. 3). For identifying the potential conflicts from continuous ACT profiles, the slope of ACT should continuously decrease till it reaches a minimum value indicating crash nearness & then the slope should increase indicating the risk reduction (Fig. 3). The ACT_{min} value signifies the severity of the potential conflict and will be further used in the extreme value analysis. It should be mentioned here that in this study for the same trajectory if multiple conflict points arise which are safety critical then those will be considered rather than considering a single conflict point for a single trajectory. This is pertinent for the longer trips of road users on multilane rural highways.Fig. 4.

3.3. Extreme value theory

Considering the conflict indicator discussed in the previous section, one can estimate potential conflict situations, which are the frequently observed events in a traffic stream. However, translating this information into the likelihood of a crash event is not a straightforward task. Extreme Value Theory (EVT) enables the extrapolation of rarely observed events (i.e., crashes) from the frequently observed events (i.e., traffic conflicts), provided that the stochastic behavior of the modeled process is smooth and continuous. EVT has two main approaches: the block maxima (BM) (or minima) using Generalized Extreme Value distribution (GEV) and the Peak over Threshold (POT) using Generalized Pareto distribution (GPD). BM and POT use different sampling methods to model extremes: BM takes the largest value (or r greatest values) in each block of specific duration, while POT takes the highest values over a certain threshold. For the relatively smaller observation periods, the block size of BM method should be as small as possible to obtain a sufficient sample size. This leads to issues in extreme sampling because sometimes the conflict with a large ACT_{min} value would be considered as an extreme since it is the only value available in the block. Studies have already proved that for short-time data, the POT method typically outperforms the BM approach due to its efficiency in sampling extremes (Ali et al., 2022; Zheng et al., 2014a). For the LMICs such as India, collecting the field data for relatively larger observation periods may not be feasible due to the high economic constraints, hence the present study has chosen POT approach for the extreme value modelling.

3.4. Generalized Pareto distribution (GPD) model

Let $X_1, X_2, ..., X_n$ are independently and identically distributed random variables with a common distribution function F. Assume that there exists a threshold u that differentiate extreme and non-extreme events. Then the distribution of the sampled extremes over the threshold u can be defined by the conditional probability as follows:

$$Pr\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \; y > 0$$

Notably, the limiting distributions of **Equation** are the generalized Pareto distribution (GPD), for a sufficiently large enough threshold u.

Now, consider the maximum value among all the events, $M_n = \max\{X_1, X_2, ..., X_n\}$. Suppose there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that $Pr\left\{\frac{M_n - b_n}{a_n} \leqslant z\right\} \rightarrow G(z)$ as $n \rightarrow \infty$ for a non-degenerate distribution function G. Then, G is a member of the GEV family shown in **Equation** (4).

$$G(z) = exp\left\{ -\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} \right\}, \; \xi \neq 0 \tag{4}$$

defined on $\{z : 1 + \xi((z - \mu)/\sigma)\rangle 0\}$, where $-\infty < \mu < \infty, \sigma > 0$, and $-\infty < \xi < \infty$. For a large enough threshold$u$, the limiting distribution function of exceedances $y = X - u$, conditional on $X > u$ is as follows:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{5}$$

defined on $\{y : y > 0 \; and \; (1 + \xi y/\tilde{\sigma})\rangle 0\}$, where $\tilde{\sigma} = \sigma + \xi(u - \mu)$ is

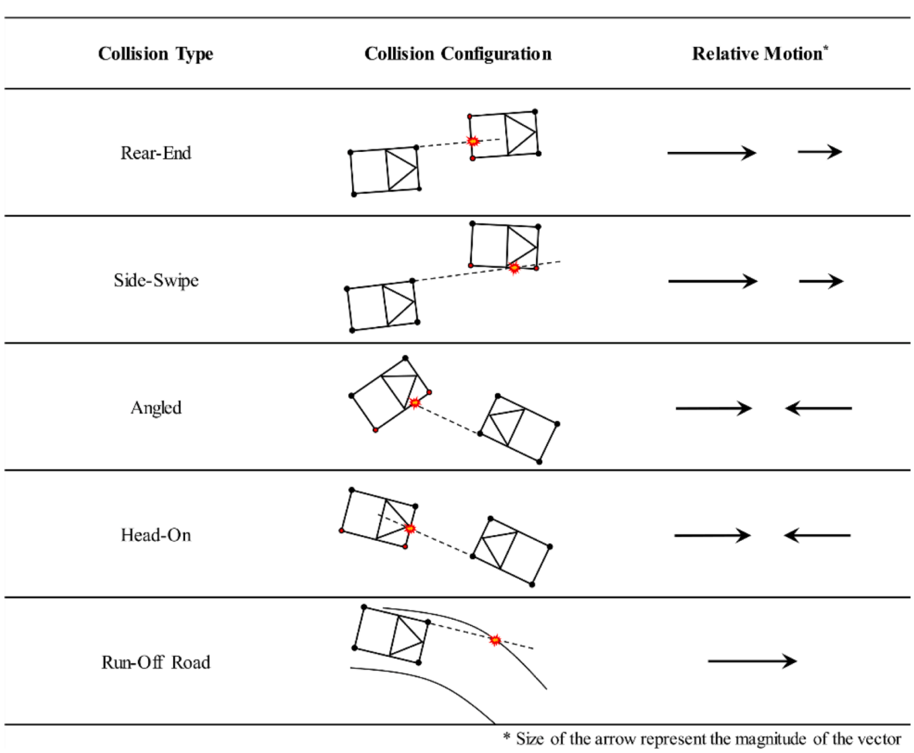


Fig. 2. Criteria for classification of conflict types (Source: Venthuruthiyil & Chunchu, 2022b).

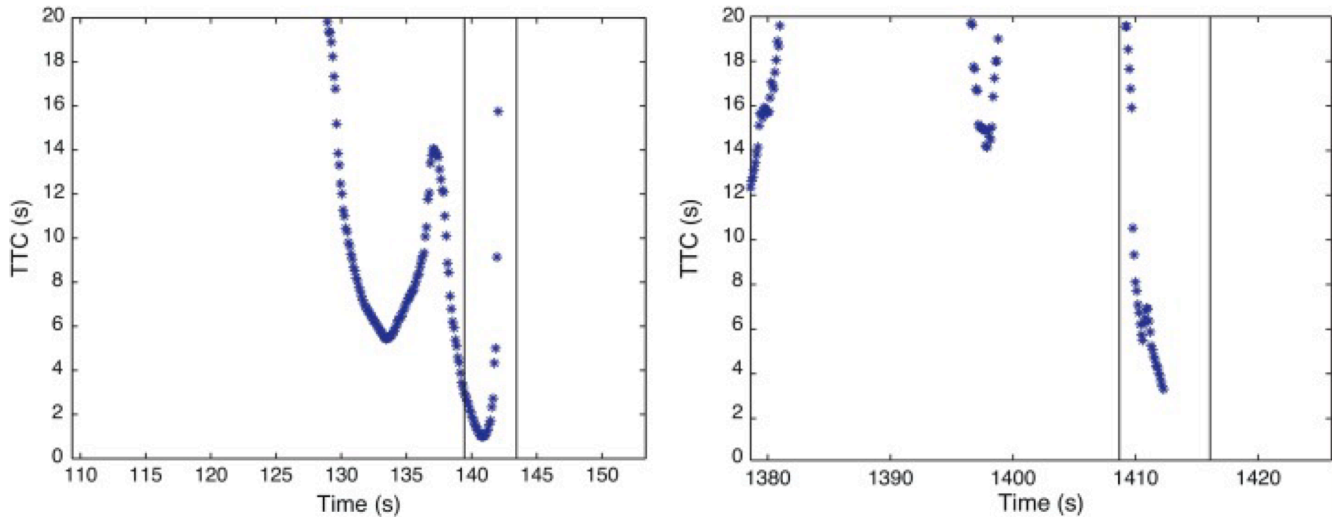


Fig. 3. Illustration of the Potential Conflict Identification from Continuous TTC profiles (Source: Jonasson and Rootzén, 2014).

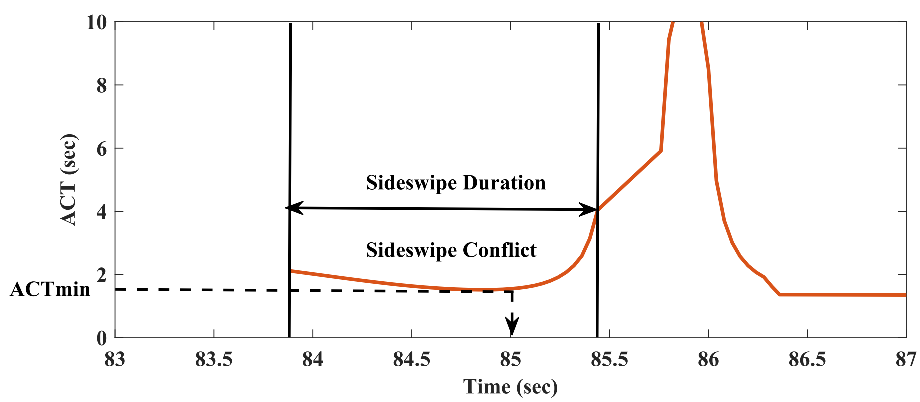


Fig. 4. Sideswipe Conflict Identification from the ACT profile of a PTW Trajectory extracted using SAVETRAX.

the scale parameter, $-\infty < \xi < \infty$ is the shape parameter. The family of distributions defined by **Equation** (5) is the generalized Pareto family.

3.5. *Threshold selection*

In the POT approach, the distribution of threshold exceedances shown in Equation could be obtained only if the threshold u is large enough to approximate the GPD family. Smaller thresholds may violate the model's asymptotic property, leading to bias, whereas the larger thresholds may result in a few exceedances for the model estimation causing higher variance. In the case of field data, such a threshold is attained by trading-off the bias and variance. Two methods are available in practice to estimate the thresholds (Coles, 2001): 1) Mean Residual Life Plot (MRLP), 2) Threshold Stability Plots (TSP). The first is an exploratory approach used before model estimation, and the second is an evaluation of the stability of the parameter estimated based on the fitting of models to a range of different thresholds. To discuss further, the MRLP depicts the relationship between the mean values of exceedances y and thresholds u. The rationale is that if the GPD is valid for threshold exceedances u_0, then it should be similarly true for other thresholds $u > u_0$, provided the scale parameter σ is appropriately adjusted. Therefore, MRLP should be generally linear above a threshold u_0, where the GPD gives a reliable estimate of the mean exceedances. Similarly, for the TSP, if the GPD is valid for threshold exceedances u_0, then the estimates of the shape and modified scale parameters should be approximately constant above u_0. In the present study, an initial range of thresholds is established using the MRLP and TSP. Further, the best threshold for the analysis was determined by fitting various GPD models and comparing the fit statistics. Zheng et al. (2015) determined the threshold by fitting various GPD models through the initial range of thresholds, and the negative log-likelihood values were taken as the measure. The threshold corresponding to the smallest negative log-likelihood value was selected as the final threshold. The present study uses the goodness-of-fit measure called Akaike Information Criterion (AIC) to compare the fitted models across different ACT thresholds. The AIC is a model fitting measure that combines the likelihood value with a complexity penalty related to the number of model parameters and is shown below (Gilleland & Katz, 2016).

$$\mathrm{AIC}(p) = 2 \times n_p - 2 \times \mathrm{LL} \tag{6}$$

where, LL is the maximum log-likelihood value and n_p is the number of parameters in the p^{th} model.

3.6. *Crash risk modelling*

Once the extreme vehicle conflict events are identified, the crash risk can be modeled using EVT. It is to be noted that the use of EVT for safety assessment is based on the safety continuum, represented by the conflict indicator that places all interactions on the same scale, beginning with the safest (normal events) and progressing to the most dangerous events, which are the crashes. The primary hypothesis of proactive safety assessment is that an unsafe event will end up into a crash if the driver fails to successfully perform the required evasive action. Therefore, an unsafe event is a situation that is closely correlated to the crash. Notably, the ACT gets closer to the extreme values for an unsafe event. Since the GPD samples extremes over a certain threshold, the negative ACT must be considered to represent the extremes. As discussed, ACT is a continuous measure, and for a potential conflict, the ACT profile must contain a clear minimum value (Fig. 1). Here the minimum ACT (ACT_{min}) indicates the amount of safety margin remaining during or after an evasive action showing how close the interaction came before avoiding the crash. It is worth noting that ACT_{min} close to zero does not necessarily mean a crash; however, the crash probability will be higher. A crash occurs when the ACT_{min} value is equal to zero. Now, using the fitted GPD based on the sampled extremes, the risk of a crash can be calculated as:

$$R = \Pr(Z \geqslant 0) = 1 - Pr(Z \leqslant 0) = \left(1 - \xi \frac{u}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{7}$$

where, R is the crash risk or crash probability of a specific conflict type (sideswipe in this study), Z is the sampled maximum of negative ACT values, and $H(\bullet)$ is the GPD. Assuming that the observation period t is representative of a more extended period T(e.g., a year), the estimated annual crashes can be calculated as:

$$N = \frac{T}{t} R \tag{8}$$

3.7. Non-stationary GPD modeling to capture the effect of covariates

As discussed in the literature review section, the crash risk of PTWs is significantly influenced by evasive actions. Such variables influencing the extremes are referred to as covariates of the non-stationary model. EVT theory assumes that the extremes are independent and identically distributed. For non-stationary extremes, ACTs are independent but not identically distributed. In such cases, the standard EVT approach will not work, therefore, the effect of covariates should be incorporated into the GPD model. Coles, (2001) have already reported that, for the POT approach, the parameter estimation procedure could relate the non-stationary extremes to the covariates, which is accomplished by including the covariates into the GPD model parameters (i.e., scale parameter with the log link function as shown in the **Equation** (9)). The scale parameter σ of the GPD distribution can be estimated as follows:

$$\sigma = \beta_0 + \beta_i \times x_i \tag{9}$$

where, x_i is the vector of covariates,β_0 is the intercept and β_i are the coefficients of the vector of covariates. The shape parameter is usually not modified since there is no empirical evidence of non-stationarity in the tail behavior (Coles, 2001). Table 2 shows the list of variables related to evasive actions that are included in the GPD model.

The model parameters were estimated using the Maximum Likelihood Estimation (MLE) approach. Further, the significance of the covariate incorporated in the non-stationary model was determined by comparing the non-stationary model with the stationary-one. This was accomplished by employing the Likelihood Ratio test (Coles, 2001), which evaluates the goodness-of-fit of two competing statistical models based on the ratio of their likelihoods. However, the shape parameter must be kept $\xi > -0.5$ to ensure the regular asymptotic properties of the maximum likelihood estimators (Smith, 1985). Considering the above criteria, the non-stationary models with covariates' effects were identified. Furthermore, the confidence intervals of the estimated crashes were also determined to quantify the uncertainty.

Table 2
List of Covariates for GPD Model.

Acronym	Description
d_r	Maximum Deceleration Rate for each sideswipe conflict
θ_r	Maximum Yaw Rate for each sideswipe conflict. Yaw Rate is the rate of change of heading angle.
t_d	Duration of the braking action for each sideswipe conflict
t_θ	Duration of the steering action for each sideswipe conflict
t_a	Duration of the accelerating action for each sideswipe conflict

4. Data collection

The main purpose of crash risk modeling using surrogate methods is to comprehend the driving mechanism of conflicts eventually leading to crashes. Such studies require extensive vehicle trajectory data to capture the dangerous interactions. In order to do so, this study has collected naturalistic driving data from the recorded traffic video footage from four and six-lane rural highways in India. In India, the four-lane and six-lane highways account for around 22% of the overall National Highway network length (MoRTH, 2021). Notably, PTWs are one of the major constituents of the multilane highways in the LMICs such as India (Bisht & Tiwari, 2022), and most of them primarily operate under free-flow conditions. On these highways, the intricate interactions between PTWs and other vehicle type considerably enhance the crash risk of PTWs (Damani & Vedagiri, 2021).

The traffic videos were collected using an unmanned aerial vehicle (UAV) for a duration of two hours (four-lane) and 1.5 h (six-lane), covering a road stretch of 700 m. The location of data collection was Kerala, a southern state in India. Fig. 5 shows the cross-sectional view of the study sites.

The trajectories were then extracted from the traffic videos using a semi-automated image processing tool called SAVETRAX (Venthuruthiyil & Chunchu, 2020a, 2022a). This tool has three modules: preprocessor, tracker, and analyzer (Fig. 6). In the pre-processing unit, camera calibration and the tracking boundary were defined. Vehicle detection, classification, and tracking were performed in the tracker module. The analyzer module applied different corrections to the tracked vehicle's path, such as identification and correction of faulty tracks and compensation for camera movements were applied and various vehicle kinematic variables were estimated.

Once the trajectories were extracted, the noise embedded in the trajectories was removed using the smoothing technique proposed by Venthuruthiyil & Chunchu (2018, 2020b). First, during the noise removal, all the occluded data points up to a length of 10 m were recovered using the spline technique with an accuracy of 5 cm. Then the heavy-tailed noise observed in the vehicle path was removed using a Recursively Ensembled Low-Pass (RELP) filter, resulting in a realistic and reasonable speed profile. Finally, white Gaussian noise was removed using an adaptive tri-cubic kernel smoothing, and the smoothing parameters were estimated using a grid-search algorithm. The resultant trajectories were smooth, differentiable, and consistent in nature. Fig. 7 displays few PTW trajectories extracted using the SAVETRAX tool and then smoothed using the RELP technique. For an extensive understanding of the smoothing algorithm, the readers can refer to the works of Venthuruthiyil & Chunchu (2018) and Venthuruthiyil & Chunchu (2020b).

The extracted trajectories from four-lane and six-lane highways include 690 and 212 PTW trajectories, respectively. The other major constituents in the traffic stream were Passenger Car (PC), Heavy Commercial Vehicles (HCV), and Light Commercial Vehicle (LCV). From the reconstructed trajectories, the variables such as acceleration, deceleration rate, and yaw rate were estimated. The yaw rate is the angular velocity of the road user's rotation around the z-axis or the rate of change of the heading angle. The maximum value of the yaw rate during a traffic conflict measures the sudden swerving or change of direction of drivers on the road whereas maximum deceleration rate indicates the hard braking during a conflict (Guo et al., 2018). Fig. 8 shows an example of the sideswipe conflict obtained from the ACT profile of a PTW and their corresponding Acceleration and Yaw Rate profiles. It is evident from the figure that the PTW has applied both the braking and swerving actions during a conflict, which are two crucial evasive behavior.

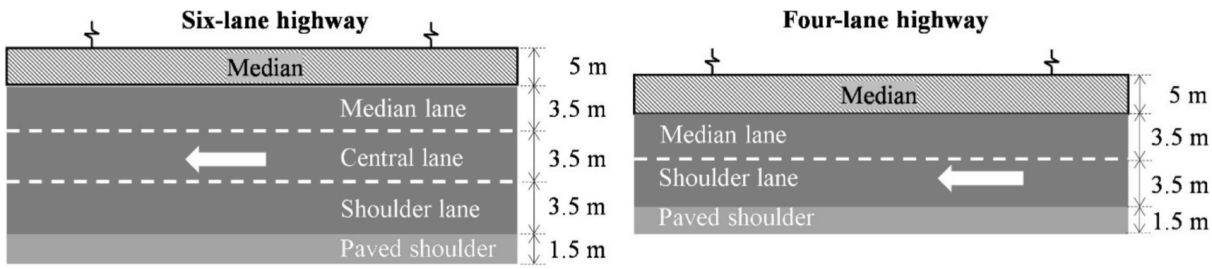


Fig. 5. Data Collection Sites for the present study.

Fig. 6. Tracking of Vehicles moving on (a) Four-lane and (b) Six-lane highways Using SAVETRAX.

5. Results & analysis

5.1. Descriptive Statistics of ACT & covariates

The present study identified 1113 and 395 sideswipe conflicts from the four and six-lane highways, respectively. On average, the ACT_{min} values for the four-lane highway show proximity of 4 s (Table 3), and for the six-lane highway, 2 s. As closer proximity indicates a larger crash risk, the six-lane highway could be more prone to the sideswipe crash risk than the four-lane highway in the study area.

Further, the yaw rate of PTWs is significantly closer to the critical threshold value reported in the literature compared to deceleration or acceleration rates. Note that Lee et al. (2011) found that the critical threshold of deceleration, acceleration, and yaw rate values for the risky situations are 6.5 m/s^2, 7.5 m/s^2, and **1.33 degrees/second**, respectively. Further, it was found that, on average, PTWs spend more time applying steering than the brakes during a sideswipe conflict on both highways.

5.2. Results of GPD modeling

As stated earlier, subjectivity is involved in selecting the ACT threshold using the MRLP and TSP methods. Fig. 9a shows that the MRLP with 95% confidence bounds is approximately linear for the threshold range of (-0.8, -0.5). TSP (Fig. 2b) demonstrates that there are two threshold ranges, where each of the scale and shape parameters are approximately constant. However, the present study has chosen the relatively smaller threshold range since the smaller threshold represents severe conflicts (i.e., true extremes) (Tarko, 2018). The modified scale parameter is approximately constant for the threshold range of (-0.6, -0.5), and the shape parameter is also nearly constant for the threshold range of (-0.6, -0.5). Based on the observations from these plots, the initial threshold range was selected, which is (-0.6, -0.5).

Nevertheless, GPD models were fitted across different thresholds to choose the best threshold. The threshold corresponding to the smallest AIC value is considered the best threshold. Fig. 10 depicts the variation of AIC values with the change in ACT threshold for both four-lane and six-lane highways. The figure clearly shows that as the threshold decreases (i.e., ACT becomes smaller), the model fit becomes better. The best thresholds for both four-lane and six-lane highways came out to be 0.5 and 0.6 s, respectively.

Based on the threshold, normal interactions are separated from the severe conflicts. From Table 4 it can be seen that, for the four-lane highway, out of 1113 sideswipe conflicts, 266 conflicts are severe in nature. Similarly, for the six-lane highway, there are 147 severe conflicts out of 395. This means that proportion-wise, there are more severe conflicts relative to the total number of interactions on the six-lane highway than that on the four-lane highway.

Table 4 shows the GPD model parameters and statistics for the best thresholds, where the values in parenthesis represent standard errors. Both the thresholds resulted in the shape parameter values of $\xi > -0.5$. Therefore, the MLE estimates possess the regular asymptotic properties of the EVT. The crash risk (or crash probability) of PTWs on the four-lane and six-lane highways came out to be 0.13% and 0.74%, respectively. The estimated annual sideswipe crashes of PTW related to the crash probabilities are 2 and 22, respectively, indicating the higher sideswipe crash risk of six-lane highway. However, the precision of the crash risk estimates (measured in confidence bounds) is higher for the four-lane highway than that of the six-lane highway.

5.3. Non-stationary modelling results

Table 5 shows the estimation results of the best-fitted models. The identity link function for the scale parameter reduced negative log-likelihood values when introduced the covariates into the stationary GPD model. The greater the reduction, the better the model fits. However, the statistical significance of such reduction for each model with a covariate was verified using a likelihood ratio test. The non-stationary model will be better fitted than the base stationary model if the likelihood ratio between the stationary and non-stationary models exceeds the chi-square value for the specified degrees of freedom.

Table 5 shows that, for the four-lane highway, the covariates, such as deceleration rate and the steering time, provide the best non-stationary model fit. Whereas for the six-lane highway, deceleration rate, yaw rate,

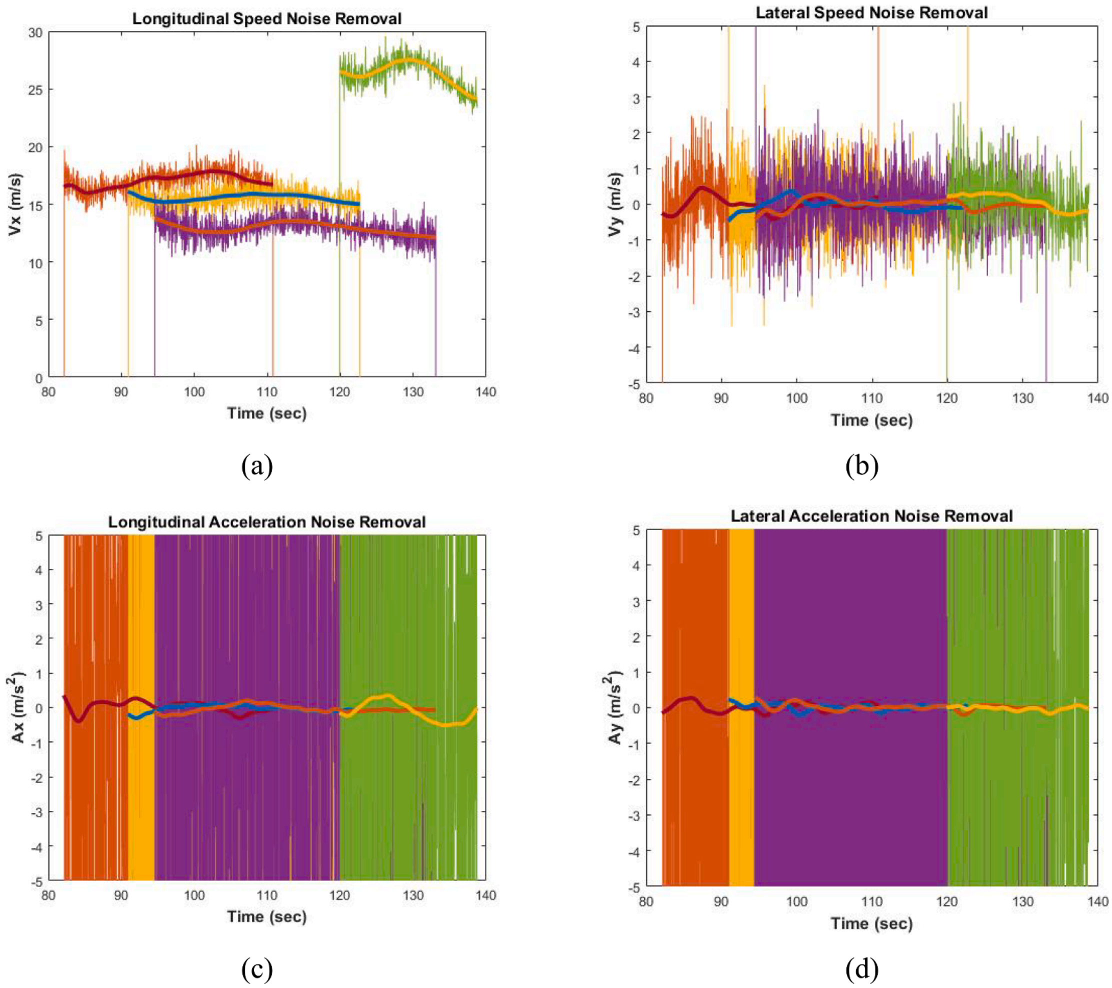


Fig. 7. Example of Noise Removal of (a) Longitudinal Speed (b) Lateral Speed (c) Longitudinal Acceleration (d) Lateral Acceleration of the few PTW Trajectories on four-lane highway.

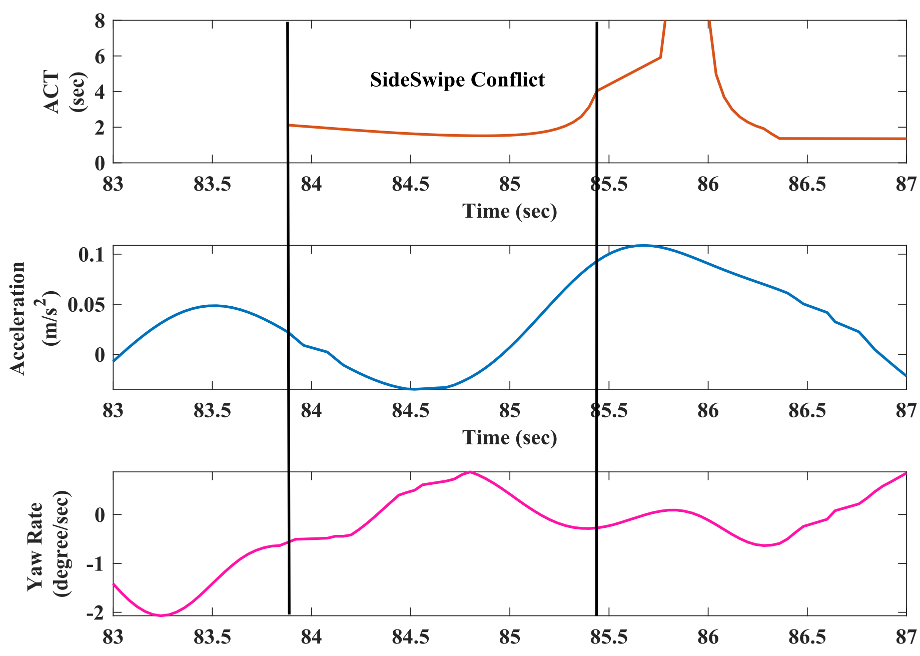


Fig. 8. ACT, Acceleration/Deceleration, and Yaw Rate Profiles of a Sideswipe Conflict.

Table 3
Statistics of sideswipe conflict data.

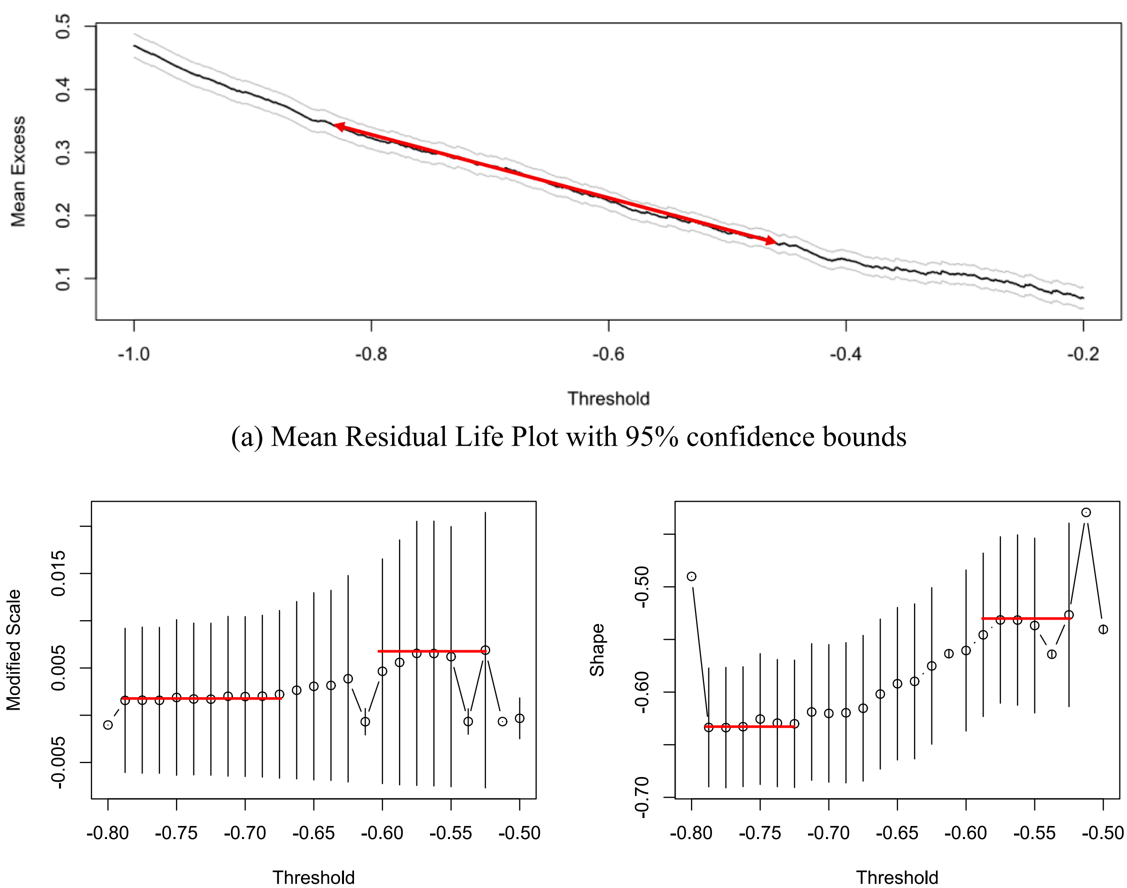
Variables	Mean		SD		Min		Max	
	4-lane	6-lane	4-lane	6-lane	4-lane	6-lane	4-lane	6-lane
ACT_{min} (s)	4.0	2.0	7.7	4.6	0.0	0.0	47.8	43.7
Maximum Deceleration Rate (m/s^2)	0.3	0.4	0.6	1.0	0.0	0.0	10.8	6.8
Maximum Yaw Rate (degree/s)	**0.9**	**1.1**	2.1	4.0.0	0.0	0.0	39.7	59.5
Braking Time (s)	0.5	0.5	0.6	0.7	0.0	0.0	3.5	5.5
Steering Time (s)	1.0	1.0	0.6	0.8	0.0	0.0	6.2	7.2
Acceleration Time (s)	0.5	0.5	0.7	0.7	0.0	0.0	4.8	4.8

Note: SD, Min, and Max denote standard deviation, minimum, and maximum values of the variables.

and steering time provided the best-fitted non-stationary models compared to the base model. For the multiple combinations of these three covariates, multiple non-stationary models such as M7, M8, M9, and M10 were developed (See Table 4). Further, these models were verified against the stationary model and the best-fitted non-stationary models with a single covariate. Table 5 shows that the non-stationary models such as M7, M8, and M9 do not perform better than the M2 model for the four-lane highway. Therefore, the M2 model was selected as the best-fitted model for the four-lane highway to estimate the crash risk. For the six-lane highway, model M2 was compared with model M10, which considers all three covariates, and it was found that model M10 performs the best. Therefore, the non-stationary model M10 was selected to estimate the sideswipe crash risk for the six-lane highway.

It should be pointed out here that, for both locations, all the GPD models with covariates related to the evasive actions perform better than the stationary ones. As stated earlier, the best performing model for the six-lane highway contains the covariates namely the deceleration rate, yaw rate, and steering time during a sideswipe conflict. However, this is not the case with the four-lane highway, where the model with the deceleration rate performed the best. The possible explanation might be that the sideswipe conflict mostly happens during lane-change/overtaking maneuvers, which will be frequent on a multilane highway in LMICs due to the high-speed difference between the road users. In our case, since six-lane highways, on average, has a higher speed difference than the four-lane highway, lane changes could be more frequent. Such a situation might increase the PTW crash risk with lane-changing/overtaking vehicles. However, PTWs try to evade laterally to avoid a crash with the lane-changing/overtaking vehicles through the steering. In the case of four-lane highways, the lateral movement might be more restricted which could be resulting in the less significant steering effect of PTWs on the crash risk.

Table 6 shows the estimation results of the best-performing GPD models. The shape parameter for both models was less than −0.5, meaning that the estimators from MLE fulfill the regular asymptotic properties of EVT and thus are more reliable. The crash probabilities or crash risks were calculated using the mean values of the covariates (shown in Table 3) substituted in Equation along with the estimated parameter values. The confidence bounds were calculated assuming that the model parameters estimated from MLE follow normal distribution under regularity conditions. Table 6 shows that the sideswipe crash risk



(a) Mean Residual Life Plot with 95% confidence bounds

(b) Threshold Stability Plot

Fig. 9. Graphical Plots of the 4-lane highway for Threshold Selection.

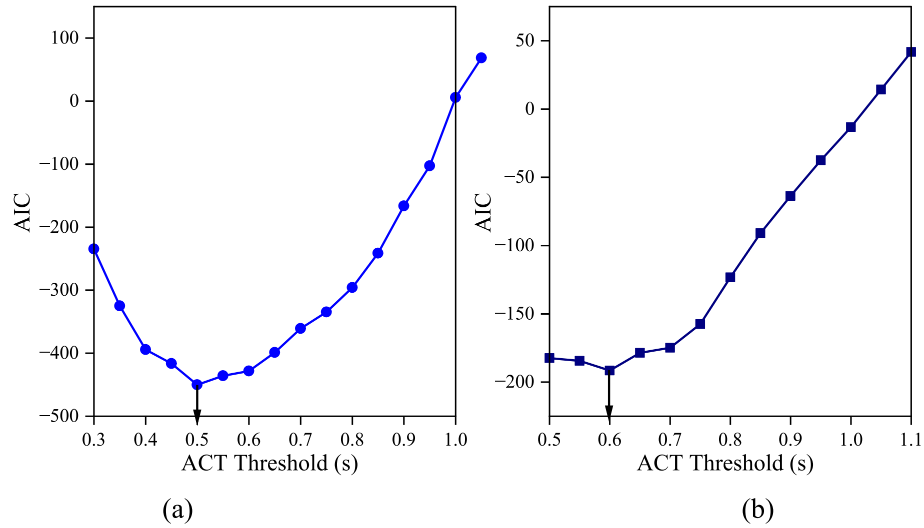


Fig. 10. Effect of ACT threshold on the goodness of fit of GPD model (a) 4-lane highway; (b) 6-lane highway.

Table 4
Parameters and statistics of the GPD model.

Parameters	GPD Model	
	four-lane	six-lane
Threshold (s)	0.5	0.6
σ (SE)	0.25 (0.009)	0.28 (0.013)
ξ (SE)	**−0.48*** (0.115)	**−0.40*** (0.183)
Exceedances	266	147
Log-likelihood	−228	−98
AIC	−450	−191
Crash Probability	0.13%	0.74%
Estimated Annual Crashes	2	22
Upper 95% CI	**96**	**287**
Lower 95% CI	0	0

Note: * indicates that the GPD model satisfies the criteria that $\xi > -0.5$.

increases with the intensity of the evasive action, such as braking and steering. Also, the time spent in steering evasion negatively affects the sideswipe crash risk. Notably, for the non-stationary GPD models, the uncertainty in the sideswipe crash risk of PTWs has significantly improved.

6. Summary & conclusions

This study investigates the sideswipe crash risk of PTWs by considering various covariates influencing the sideswipe conflict. The study, therefore, developed non-stationary models using the POT approach to capture the impact of such covariates. The findings of this study are:

1. On average, during a sideswipe conflict, PTWs spend more time performing steering evasion than braking evasion. Besides, only the intensity of the steering evasion measured in terms of maximum yaw rate reaches safety critical values i.e. exhibits extreme driving evasions.
2. Based on the ACT_{min} threshold, higher percentage of severe sideswipe conflicts are occurring on the six-lane highway than that on the four-lane highway. Also, PTW riders experience higher crash risk on the six-lane highway. The sideswipe crash risk of PTWs increases with the intensity of the maximum deceleration and yaw rates.
3. Non-stationary GPD models have significantly reduced the uncertainty in the crash risk estimates compared to the stationary models.

The main conclusions of this study are:

1. PTWs evade a sideswipe conflict mainly by making extreme steering evasion rather than applying hard brakes. Therefore, for the sideswipe conflict identification of PTWs, steering evasion must be considered. Otherwise, the locations may give false negatives in terms of sideswipe conflict identification which will eventually lead to the erroneous identification of risk-free locations.
2. PTWs experience significantly higher sideswipe (sometimes also referred as lane-changing) crash risk on the six-lane highway as

Table 5
Statistics of the fitted non-stationary models.

Model Number	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Model type	*Stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*
Description	Base case	$\sigma = \beta_0 + \beta_1 \times d_r$	$\sigma = \beta_0 + \beta_1 \times \theta_r$	$\sigma = \beta_0 + \beta_1 \times t_d$	$\sigma = \beta_0 + \beta_1 \times t_a$	$\sigma = \beta_0 + \beta_1 \times t_\theta$	$\sigma = \beta_0 + \beta_1 d_r + \beta_2 \theta_r$	$\sigma = \beta_0 + \beta_1 d_r + \beta_2 t_\theta$	$\sigma = \beta_0 + \beta_1 \theta_r + \beta_2 t_\theta$	$\sigma = \beta_0 + \beta_1 d_r + \beta_2 \theta_r + \beta_3 t_\theta$
Site	LL1	LL2	LL3	LL4	LL5	LL6	LL7	LL8	LL9	LL10
4-lane	−228.05	−229.98	−229.24	−228.19	−228.97	−230.05	−230.21	−231.36	−230.91	——
6-lane	−98.74	−102.76	−101.88	−99.97	−99.39	−101.22	−104.33	−104.66	−103.94	−106.78
Likelihood ratio test results										
Site	M1&M2	M1&M3	M1&M4	M1&M5	M1&M6	M2&M7	M2&M8	M2&M9	M2&M10	
4-lane	**3.84***	2.39	0.29	1.83	**3.99***	0.46	2.76	1.86	——	
6-lane	**8.04***	**6.29***	2.46	1.30	**4.95***	3.14	3.79	2.35	**8.04***	

Notes: $\chi^2_{(0.95,df=1)} = 3.841, \chi^2_{(0.95,df=2)} = 5.991$. The * values indicate the best-fitted non-stationary models.

Table 6
Estimation Results of the best fitted non-stationary models.

Site	Sample Size	Model	Maximum Likelihood Estimation Results								
			LL	β_0	β_1	β_2	β_3	ξ	Crash Probability	Crash Frequency	95% confidence bounds
4-lane	266	M2	−229	0.24 (0.01)	0.02 (0.01)	–	–	−0.47 (0.04)	0.09%	2	[0, 28]
6-lane	147	M10	−106	0.28 (0.00)	0.02 (0.00)	0.01 (0.00)	−0.04 (0.00)	−0.42 (0.02)	0.17%	5	[0, 19]

Note: values in the parentheses indicates standard errors.

compared to the four-lane highway. This means that providing additional lane (or widening the roads from four-lane to six-lane) did not fully improve the safety situations of PTWs. The workaround could be limiting the aggressive and frequent overtaking and lane-changing of vehicles on such highways and providing additional lanes/service roads for the exclusive movement of PTWs.

3. Non-stationary GPD model developed considering evasive actions such as braking and steering outperforms the stationary model. Hence, the effect of evasive actions needs to be incorporated in the crash risk models because the parameters such as deceleration rate and yaw rate significantly improve the models' precision. Such crash risk models will be very useful in identifying the accident-prone locations and hence will be beneficial in conducting real-time threat assessment on multilane highways without the help of crash data.

The contribution of this study is manifold:

1. This paper adds to the relatively little literature available on the PTWs' sideswipe crash risk assessment using a newly developed multidimensional proactive measure called ACT. This measure helps in identifying the risky situations for PTW riders by integrating the steering effect on the proximity calculation which is a very important parameter for the lane-changing operations of PTWs.
2. As discussed in the literature review, most non-stationary crash risk models consider aggregated variables such as traffic volume, conflict volume, lane change duration, number of lane changes which influences the crash occurrences. Crash risk models developed considering the microscopic driving characteristics are very scarce in the literature. The present study has contributed to this scarce literature by considering the microscopic driving actions such as braking and steering in developing the crash risk models.

The present study can be further extended in numerous ways:

1. The effect of road geometry, drivers' and land use characteristics on the sideswipe crash risk should be studied by collecting data from different locations with varying traffic compositions.
2. The effect of temporal features on crash mechanisms should be analyzed by collecting trajectory data during different times of the day and seasons of the year.
3. The EVT models can be further improved by considering the effect of speed-based measures such as maximum speed difference, delta V etc. that can capture the vehicular damage and injury to the drivers.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Pranab Kar, Suvin P. Venthuruthiyil, and C. Mallikarjuna; data collection: Pranab Kar and Suvin P. Venthuruthiyil; analysis and interpretation of results: Pranab Kar, Suvin P. Venthuruthiyil, and C. Mallikarjuna; draft manuscript preparation: Pranab Kar, Suvin P. Venthuruthiyil, and C. Mallikarjuna. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ali, Y., Haque, M.M., Zheng, Z., 2022. An Extreme Value Theory approach to estimate crash risk during mandatory lane-changing in a connected environment. Anal. Methods Acc. Res. 33, 100193 https://doi.org/10.1016/J.AMAR.2021.100193.
Arun, A., Haque, M.M., Bhaskar, A., Washington, S., Sayed, T., 2021a. A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. Anal. Methods Acc. Res. 32, 100180 https://doi.org/10.1016/j.amar.2021.100180.
Arun, A., Haque, M.M., Washington, S., Sayed, T., Mannering, F., 2021b. A systematic review of traffic conflict-based safety measures with a focus on application context. Anal. Methods Acc. Res. 32, 100185 https://doi.org/10.1016/J.AMAR.2021.100185.
Arun, A., Haque, M.M., Washington, S., Sayed, T., Mannering, F., 2022. How many are enough?: Investigating the effectiveness of multiple conflict indicators for crash frequency-by-severity estimation by automated traffic conflict analysis. Transport. Res. Part C: Emerging Technol. 138, 103653 https://doi.org/10.1016/J.TRC.2022.103653.
Bisht, L.S., Tiwari, G., 2022. Safety effects of paved shoulder width on a four-lane divided rural highway in India: A matched case-control study. Saf. Sci. 147, 105606 https://doi.org/10.1016/j.ssci.2021.105606.
Cavadas, J., Azevedo, C.L., Farah, H., Ferreira, A., 2020. Road safety of passing maneuvers: A bivariate extreme value theory approach under non-stationary conditions. Accid. Anal. Prev. 134, 105315.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London, 10.1007/978-1-4471-3675-0.
Damani, J., Vedagiri, P., 2021. Safety of motorised two wheelers in mixed traffic conditions: Literature review of risk factors. J. Traffic Transport. Eng. (English Edition) 8 (1), 35–56. https://doi.org/10.1016/J.JTTE.2020.12.003.
Farah, H., Azevedo, C.L., 2017. Safety analysis of passing maneuvers using extreme value theory. IATSS Res. 41 (1), 12–21. https://doi.org/10.1016/j.iatssr.2016.07.001.
Fu, C., Sayed, T., Zheng, L., 2020. Multivariate Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. Anal. Methods Acc. Res. 28, 100135 https://doi.org/10.1016/J.AMAR.2020.100135.
Gilleland, E., & Katz, R. W. (2016). in2extRemes: Into the R package extRemes. Extreme value analysis for weather and climate applications. https://doi.org/10.5065/D65T3HP2.
Guo, Y., Sayed, T., Zaki, M.H., 2018. Exploring evasive action-based indicators for PTW conflicts in shared traffic facility environments. J. Transport. Eng., Part A: Syst. 144 (11) https://doi.org/10.1061/jtepbs.0000190.
Guo, Y., Sayed, T., Zaki, M.H., 2019. Evaluating the safety impacts of powered two wheelers on a shared roadway in China using automated video analysis. J. Transport. Safety Secur. 11 (4), 414–429. https://doi.org/10.1080/19439962.2018.1447058.
Haghani, M., Behnood, A., Dixit, V., Oviedo-Trespalacios, O., 2022. Road safety research in the context of low- and middle-income countries: Macro-scale literature analyses, trends, knowledge gaps and challenges. Saf. Sci. 146, 105513 https://doi.org/10.1016/J.SSCI.2021.105513.
Hydén, C., 1987. The development of a method for traffic sfaety evaluation: The Swedish Traffic-Conflicts Technique [Lund Institute of Technology]. Lund Institute of Technology, 10.1007/978-3-642-82109-7_12.
Jonasson, J.K., Rootzén, H., 2014. Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach. Accid. Anal. Prev. 62, 102–109. https://doi.org/10.1016/J.AAP.2013.09.013.
Laureshyn, A., Svensson, Å., Hydén, C., 2010. Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. Accid. Anal. Prev. 42 (6), 1637–1646. https://doi.org/10.1016/j.aap.2010.03.021.

Lee, S.E., Simons-Morton, B.G., Klauer, S.E., Ouimet, M.C., Dingus, T.A., 2011. Naturalistic assessment of novice teenage crash experience. Accid. Anal. Prev. 43 (4), 1472–1479. https://doi.org/10.1016/J.AAP.2011.02.026.

MoRTH. (2021). *Basic Road Statistics in India 2017-18*. http://www.morth.nic.in.

Puthan, P., Lubbe, N., Shaikh, J., Sui, B., Davidsson, J., 2021. Defining crash configurations for Powered Two-Wheelers: Comparing ISO 13232 to recent in-depth crash data from Germany, India and China. Accid. Anal. Prev. 151, 105957 https://doi.org/10.1016/J.AAP.2020.105957.

Smith, R. L. (1985). *Maximum likelihood estimation in a class of nonregular cases*. *72*(1), 67–90. http://biomet.oxfordjournals.org/.

Songchitruksa, P., Tarko, A.P., 2006. The extreme value theory approach to safety estimation. Accid. Anal. Prev. 38 (4), 811–822. https://doi.org/10.1016/J.AAP.2006.02.003.

Tageldin, A., Sayed, T., Wang, X., 2015. Can time proximity measures be used as safety indicators in all driving cultures? case study of motorcycle safety in China. Transp. Res. Rec. 2520, 165–174. https://doi.org/10.3141/2520-19.

Tarko, A.P., 2012. Use of crash surrogates and exceedance statistics to estimate road safety. Accid. Anal. Prev. 45, 230–240. https://doi.org/10.1016/j.aap.2011.07.008.

Tarko, A.P., 2018. Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution – A theoretical and numerical exploration. Accid. Anal. Prev. 113, 63–73. https://doi.org/10.1016/J.AAP.2018.01.008.

Venthuruthiyil, S.P., Chunchu, M., 2018. Trajectory reconstruction using locally weighted regression: a new methodology to identify the optimum window size and polynomial order. Transportmetrica A: Trans. Sci. 14 (10), 881–900. https://doi.org/10.1080/23249935.2018.1449032.

Venthuruthiyil, S.P., Chunchu, M., 2020a. SAVETRAX: A semi-automated image processing based vehicle trajectory extractor. 99th Annual Meeting of Transportation Research Board.

Venthuruthiyil, S.P., Chunchu, M., 2020b. Vehicle path reconstruction using Recursively Ensembled Low-pass filter (RELP) and adaptive tri-cubic kernel smoother. Transport. Res. Part C: Emerg. Technol. 120, 102847 https://doi.org/10.1016/J.TRC.2020.102847.

Venthuruthiyil, S.P., Chunchu, M., 2022a. SAVETRAX: A Tool for Automatic Extraction of Multitude of Microscopic Traffic Data from Different Camera Platforms. Working Paper.

Venthuruthiyil, S.P., Chunchu, M., 2022b. Anticipated Collision Time (ACT): A two-dimensional surrogate safety indicator for trajectory-based proactive safety assessment. Transport. Res. Part C: Emerg. Technol. 139, 103655 https://doi.org/10.1016/J.TRC.2022.103655.

Venthuruthiyil, S.P., Samalla, S., Chunchu, M., 2022. Association of Crash Potential of Powered Two Wheelers (PTW) With the State of Traffic Stream. 8th Road Safety and Simulation International Conference.

WHO, W. H. O. (2018). Global status report on road safety 2018. In *Geneva, Switzerland, WHO*.

Zheng, L., Ismail, K., Meng, X., 2014a. Freeway safety estimation using extreme value theory approaches: A comparative study. Accid. Anal. Prev. 62, 32–41. https://doi.org/10.1016/J.AAP.2013.09.006.

Zheng, L., Ismail, K., Meng, X., 2014b. Traffic conflict techniques for road safety analysis: Open questions and some insights. Can. J. Civ. Eng. 41 (7), 633–641. https://doi.org/10.1139/CJCE-2013-0558.

Zheng, L., Ismail, K., Meng, X., 2015. Evaluation of peak over threshold approach for road safety estimation. J. Transport. Safety Security 7 (1), 76–90. https://doi.org/10.1080/19439962.2014.904029.

Zheng, L., Sayed, T., Tageldin, A., 2018. Before-after safety analysis using extreme value theory: A case of left-turn bay extension. Accid. Anal. Prev. 121, 258–267. https://doi.org/10.1016/j.aap.2018.09.023.

Zheng, L., Sayed, T., Essa, M., 2019. Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. Accid. Anal. Prev. 123, 314–323. https://doi.org/10.1016/j.aap.2018.12.007.

Zheng, L., Sayed, T., Mannering, F., 2021. Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. Anal. Methods Acc. Res. 29, 100142 https://doi.org/10.1016/J.AMAR.2020.100142.