A Ph.D. Dissertation on

Proactive Safety Assessment of Mixed Traffic on Rural Highways Using Extreme Value Theory

Submitted for the fulfillment of

Doctor of Philosophy

in

Civil Engineering

by

Pranab Kar

(Roll no. 156104039)

Under the guidance of

Dr. C. Mallikarjuna

DEPARTMENT OF CIVIL ENGINEERING

IIT GUWAHATI, ASSAM, INDIA-781039

FEBRUARY, 2023

Certificate

This is to certify that the work presented in the thesis entitled "**Proactive Safety Assessment of Mixed Traffic on Rural Highways Using Extreme Value Theory**" is carried out by Mr. Pranab Kar for the award of Doctor of Philosophy in the Department of Civil Engineering from Indian Institute of Technology Guwahati. The thesis embodies the original work and studies carried out by the student under my supervision and has not been submitted elsewhere for the degree.

Mallikarjuna C.
Professor
Department of Civil Engineering
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India

Date: 02/02/2023

Acknowledgement

Finally, the time has come to write the first page of this dissertation—the most-read page by family, colleagues, and friends looking for their names. I hope I will not forget anybody to thank and that you find your name when you expect it to be here.

This dissertation would not have been possible without the support of several people who have contributed in some way to the success of this research project. It's been a long and exhausting journey but one that has also been highly enlightening. I want to express my profound gratitude to my supervisor Prof. Mallikarjuna Chunchu, for his inspiration, invaluable guidance, and continuous encouragement during this work. His suggestions, comments, and, most importantly, his patience with a beginner like me improved my study experience and made it less complicated. Thank you for genuinely caring about my well-being and challenging me to take a broader perspective on my research. I consider myself blessed to have him as a mentor.

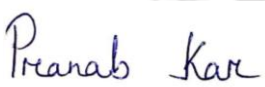
I am sincerely grateful to the chairman of my Doctoral Committee, Prof. Rajib Kumar Bhattacharjya, for his constructive suggestions and timely help during this research work. I would also like to thank the Doctoral Committee members, Dr. Anjan Kumar S., and Dr. Venkatesh T., for their fruitful discussions and invaluable suggestions during the research process.

Civil Engineering Department offered me several warm moments to cherish. I am very thankful to the faculty and staff for the positive environment. I sincerely acknowledge the Indian Institute of Technology Guwahati for allowing me to study at IITG. I thank my research group members Suvin, Bharat, Regulus, Omar bhaiya, Pal bhaiya, with whom I have shared my moments of anxiety and excitement. I wish to thank all my friends and colleagues at IIT Guwahati, Jagdish, Suresh, Vinay, Sanhita, Dhanesh, Nillotpal, Jimmy da, Iba di, Arunabha, Shantanu, Nishant, Bhaskar, Saswathi, Sivasai. I would also like to thank all my colleagues in the field of transportation research, Sandeep, Subhadipto, and Shreya, for all the fruitful conversations we had during our meetings.

At this point, I remember and thank all my teachers and well-wishers for showing me the path to my dreams.

I should pay tribute to my beloved parents for giving me the strength to continue these studies while staying away. I thank them for their persistent faith in my decisions. Thank you, my ever-supportive siblings, Nibedita, Sudipta, Sutapa, Kalpita, Sandeep, Snehamoy, and Swagata. You gave me strength with your emotional support during all the tough times in my life. Rai, Sonai, Tia, Rusha, my baby nieces, your innocent and cute smiles were enough to melt away all my final-year stress. I thank my brother-in-laws Joydeep, Kushalendu, Abhra, and my Mashi, Meso, and Chotomama for their support. Bhatia and Soumya, your weekend calls made me feel happy.

Finally, my utmost gratitude should go to Mother Nature for giving us somewhere to live and something to study. All the philosophers, free thinkers, and writers who taught me to see the beauty of nature and life, giving me the breath of life, should be honored.



Pranab Kar
Date: 02/02/2023

This thesis is dedicated to my beloved family

Abstract

Traffic road safety is one of the significant challenges faced by traffic engineers and transport planners in today's world. Road traffic injuries are the eighth leading cause of death among all age groups and the leading cause of death for children and young adults (WHO, 2018). Therefore, gaining knowledge about the driver behavior that leads to crashes and providing safety countermeasures to avoid these hazards is paramount to improving road safety. Historical crash data are predominantly used to perform the crash risk assessment of traffic streams moving on rural highways. But, inherent issues of crash data such as under-reporting, small sample size, lack of information on the causal mechanism of crashes, and extended data collection period forced researchers to look for an alternative approach. The proactive safety technique uses surrogate safety measures (SSMs) to identify conflicts and then performs the safety assessment based on the observable non-crash events as surrogates. However, the main challenge is establishing the connection between these surrogate events (or conflicts) and the crashes.

For the proactive safety assessment, naturalistic driving data capturing microscopic driving behavior is essential. However, due to economic constraints, collecting such data is challenging, especially for low- and middle-income countries (LMICs) such as India. The present study has collected such data using unmanned aerial vehicles (UAV) for rural highways. From the collected videos, the trajectory of the vehicles was extracted using a semi-automated image processing software named SAVETRAX. Once the trajectories were extracted, the noise embedded in the trajectories was removed using a recursively ensembled low-pass filter (RELP). From the smoothed trajectories, conflict information was obtained for the multiple road users. For the mixed traffic prevalent in rural highways in India, commonly used conflict indicators, namely time-to-collision (TTC) or post-encroachment time (PET), fail to capture the different conflict types associated with the mixed traffic. Hence, the present study employs a multidimensional conflict indicator called anticipated collision time (ACT) for capturing the different conflict types between road users of mixed traffic.

The conflict analysis found that sideswipe conflicts happen more frequently on multilane rural highways operating under low-volume conditions than rear-end conflicts. For the undivided highways passing through mountainous terrains, the run-off-road (ROR) conflict type involving a single-vehicle happens more frequently than the multi-vehicle conflict types such as rear-end,

sideswipe, and head-on. Hence, the present study analyzes the safety of vehicles for the sideswipe and ROR conflict types using the safety metric Crash Risk.

For estimating the crash risk, the present study employs extreme value theory (EVT) that extrapolates the crashes from severe traffic conflicts. For identifying the severe traffic conflicts, the present study uses the ACT_{min} values that indicate the safety margin remaining, obtained from the potential traffic conflicts. Smaller ACT_{min} values indicate severe traffic conflicts. The extremes are sampled from the severe traffic conflicts using the block maxima (BM) and peak-over threshold (POT) approaches of EVT. The parameters were estimated using the maximum likelihood estimation approach. The study found that the POT models outperform BM models because the uncertainty associated with the crash risk estimates is lower for the POT models. Based on the crash risk results, the study found that the powered two wheelers (PTWs) experience significantly higher sideswipe crash risk on multilane highways than other vehicle types. Compared to the four-lane, the crash risk of PTWs is even higher in the case of a six-lane highway. This further proves the vulnerable nature of PTWs under mixed traffic conditions. For the undivided highways passing through mountainous terrains, heavy commercial vehicles (HCVs) experience a significant ROR crash risk than that of other vehicle types.

The crash risk estimates were further improved by considering the microscopic variables related to crashes' causal mechanism, such as evasive action. The variables include braking and steering evasive actions such as maximum braking and steering rates and the braking and steering time for each vehicle type. Non-stationary EVT modelling was carried out by incorporating these variables' effects in the POT model's scale parameter. The sideswipe crash risk increases with the maximum braking and steering rates. Similarly, ROR crash of heavy commercial vehicles increases with the decrease in curve radius, increase in approach length of tangent, and the distance between the midpoints of horizontal and vertical curves. Based on the crash risk assessment, sharp curves with longer approach tangents should not be provided on the horizontal curves. Also, the midpoints of horizontal and vertical curves should overlap.

Contents

List of Figures

List of Tables

List of Abbreviations

WHO	World Health Organization
LMIC	Low-and Middle-Income Countries
HIC	High Income Countries
CAV	Connected and Autonomous Vehicles
EVT	Extreme Value Theory
MORTH	Ministry of Road Transport and Highways
NH	National Highway
SSM	Surrogate Safety Measure
UAV	Unmanned Aerial Vehicle
SAVETRAX	Semi-automated Vehicle Trajectory Extractor
RELP	Recursively Ensembled Low-pass Filter
TTC	Time-to-Collision
PET	Post Encroachment Time
VRU	Vulnerable Road Users
PTW	Powered Two-wheelers
ACT	Anticipated Collision Time
TA	Time-to-accident
TET	Time-exposed TTC
TIT	Time-integrated TTC
MTTC	Modified TTC
PSD	Percent Stopping Distance
DRAC	Deceleration rate to avoid a crash
PN	Poisson
NB	Negative Binomial

PNL	Poisson-Lognormal
SPF	Safety Performance Functions
BM	Block Maxima
GEV	Generalized Extreme Value
POT	Peak-over Threshold
GPD	Generalized Pareto Distribution
SHRP	Strategic Highway Research Program
GPS	Global Positioning System
ML	Median Lane
CL	Center Lane
SL	Shoulder Lane
PC	Passenger Car
HCV	Heavy Commercial Vehicles
LCV	Light Commercial Vehicles
NHAI	National Highway Authority of India
MLE	Maximum Likelihood Estimation Approach
MRLP	Mean Residual Life Plot
TSP	Threshold Stability Plots
AIC	Akaike Information Criterion
LL	Log-likelihood
ROR	Run-off-road

List of Symbols

λ	number of crashes
c	number of conflicts
π	crash-to-conflict ratio
c_i	conflicts corresponding to a severity level (i)
π_i	crash-to-conflict ratio corresponding to a severity level (i)
T	Anticipated Collision Time
δ	the shortest distance between two road users
$\left(\frac{\partial \delta}{\partial t}\right)$	closing-in-rate between two road users
$\dot{\theta}$	Yaw rate or rate of change of heading angle
μ	location parameter
σ	scale parameter
ξ	shape parameter
Z	sampled extremes
R	Crash Risk
u	threshold
N	Estimated annual crashes
d_r	Maximum Deceleration Rate for each sideswipe conflict
θ_r	Maximum Yaw Rate for each sideswipe conflict.

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

Chapter 1 Introduction

1.1 Background

Road safety is one of the significant challenges faced by traffic engineers and transport planners in today's world. Road traffic injuries are the eighth leading cause of death among all age groups and the leading cause of death for children and young adults (WHO, 2018). According to a WHO (2018) report, about 1.35 million people die yearly from road crashes. The percentage of road traffic fatalities is disproportionately higher in low and middle-income countries (LMICs) as compared to the high-income countries (HICs) (WHO, 2018). Therefore, gaining knowledge about the driver behavior that leads to crashes and providing safety countermeasures to avoid these hazards is paramount to improving the road safety scenario.

Historical crash data are predominantly used to perform the safety assessment. The safety countermeasures are developed by modelling the crash data. However, there are inherent issues of crash data such as under-reporting, small sample size, lack of information on the causal mechanism of crashes, and extended data collection period (Mannering & Bhat, 2014; Mannering et al., 2016; Papazikou et al., 2019; Cavadas et al., 2020; Zheng et al., 2021). Further, using crash data to develop better safety strategies is a reactive approach, and the researchers have to wait for crashes to happen. In the future, these problems with crash data will likely worsen because the traffic environment is changing so quickly with the introduction of connected and autonomous vehicles (CAVs) (Tarko, 2018). For instance, it is anticipated that hundreds of years of crash data may be necessary to show that autonomous vehicles could result in a lower death rate than human drivers (Kalra & Paddock, 2016; Guo, 2019).

Hence, proactive safety assessment based on traffic conflicts emerged as a better alternative for capturing the driving mechanism leading to crashes and avoiding certain pitfalls of crash data, such as the exclusion of lower severity events (Cavadas et al., 2020; Tarko, 2021; Zheng et al., 2021). Due to the enormous amount of real-time vehicle data that is likely to be accessible in the coming era of connected and autonomous vehicles, proactive safety assessment will most certainly play a key role (Papadoulis et al., 2019; Virdi et al., 2019; Xie et al., 2019; Mannering et al., 2020). 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 robust statistical approach that can be used to analyze the likelihood of rare, extreme events, such as traffic crashes. The goal of this type of modeling is to identify potential risk factors that may contribute to the occurrence of such events and to develop strategies for mitigating those risks. Extreme value theory is based on the concept that the likelihood of rare events can be predicted by studying the distribution of more common events (i.e., traffic conflicts). Using conflict data, crash risk models can be developed to provide interventions and strategies to reduce crash risk, such as improving road design, implementing safety regulations, and providing driver training. There are a lot of proactive safety studies involving EVT for the HICs (Zheng *et al*., 2019; Fu *et al*., 2020; Arun *et al.,* 2021; Arun *et al.*, 2022). However, such studies are very scarce for LMICs.

The safety of vehicles moving on multilane rural highways in LMICs can vary significantly compared to the HICs. Traffic in LMICs consists of vehicles of varying sizes (i.e., length and width), weights, and kinematic characteristics (i.e., speed, acceleration, braking, steering), sharing traffic lanes. Lane adherence is typically low in LMICs due to the tendency of smaller vehicles to close any gaps left by larger ones to speed up their movement. This is not the situation with HICs, where traffic is homogeneous, and the vehicles adhere to lane discipline. Vehicles in mixed traffic inadvertently undergo intricate interactions with other vehicle classes, increasing the crash likelihood (Dhamani and Vedagiri, 2021). The crash risk assessment of drivers on rural highways in LMICs can be challenging due to the lack of data and reliable information on the various factors influencing the crashes. The availability of limited resources for collecting and analyzing data on road crashes makes it difficult to develop accurate crash risk models. For assessing drivers' crash risk on these roads, collecting data on the influencing factors and analyzing them using appropriate methods is essential. This can help identify areas of high risk and develop strategies to reduce the crash likelihood.

Crash risk models are essential for understanding the factors contributing to traffic crashes and developing strategies to reduce their likelihood. In LMICs, where road safety is one of the significant concerns, the development and use of crash risk models can be precious. These models can help identify areas of high risk and inform the development of targeted interventions to improve road safety. For example, crash risk models can be used to identify locations where speed control measures, guardrails, or traffic lights can help reduce the likelihood of crashes.

Additionally, crash risk models can help evaluate the effectiveness of different safety measures and guide the allocation of resources to the areas and interventions most likely to reduce crash risk. Overall, using crash risk models can help improve the safety of vehicles on the roads of LMICs.

1.2 Need of the Study

World Health Organization (2018) (WHO) reported that 90% of road traffic deaths happen in the LMICs. People in LMICs are three times more likely to die in a road crash than people in HICs (WHO, 2018). Hence, there is a growing need for safety studies, but it is still below par, and only less than 10% of all road safety studies are related to LMICs (Haghani et al., 2022). The proactive safety assessment becomes even more relevant for LMICs since the recorded crash data is heavily under-reported. The tremendous economic constraint for collecting larger datasets is also a reason to advocate proactive safety assessment.

In LMICs such as India, about 3,66,138 road accidents occurred in 2020, out of which 1,31,774 lives were lost (MORTH, 2020). Road traffic deaths are significantly higher on the national highways (NH), including expressways, compared to the other road facilities in India. Reported road traffic deaths were 35.9% on the NH during the year 2020 while comprising only 2% of the total road network (MORTH, 2020). Hence, it is paramount to analyze the overall safety and factors significantly influencing crashes on such highways, particularly in the context of LMICs.

Proactive safety assessment evaluates the potential risks and hazards associated with different types of vehicles on the roads without using police-recorded crash data. It then identifies potential safety issues and implements measures to address them. This type of assessment is essential, especially for the LMICs, because there are several issues with the crash data collection. Due to a lack of standardized reporting procedures and awareness about the importance of reporting crashes, many crashes go unreported. This leads to underestimating the true extent of the problem and hinders efforts to address it. Cultural and social factors may also play a role in the underreporting of crashes. People may hesitate to report crashes due to fear of reprisal or legal consequences. In other cases, a cultural stigma may be associated with being involved in a crash, which can discourage people from coming forward. Overall, the lack of resources, infrastructure, and cultural and social factors can make it challenging to accurately collect and analyze crash data

in LMICs. Hence, the proactive safety approach becomes the more viable option for crash risk assessment and requires further attention.

1.3 Objectives of the Study

The objective of this study is to analyze and model the crash risk experienced by the mixed traffic prevalent on rural highways of LMICs using a proactive safety approach. To fulfill this objective, the following tasks were accomplished.

1. Suitable surrogate safety measure (SSM) or conflict indicator identification for the mixed traffic prevalent in LMICs
2. Naturalistic driving data collection using an unmanned aerial vehicle (UAV)
   2.1. Extraction of detailed trajectory data using SAVETRAX
   2.2. Denoising of trajectory data using a recursively ensembled low-pass filter (RELP)
3. Conflict analysis using the suitable SSM applicable for the mixed traffic in LMICs
4. Crash risk assessment of mixed traffic using EVT
   2.1 Stationary EVT modelling of crash risk using suitable SSM
   2.2 Non-stationary EVT modelling of crash risk using the microscopic driving variables related to the SSM

1.4 Scope of the Study

Road traffic crashes are significantly higher on the NH, including expressways, compared to the other road traffic facilities in India. Further, around 65% of road crashes occur on straight roads, more than 72% of road crashes, and 67% of road traffic deaths happen during sunny and clear weather in India (MORTH, 2020). Approximately 14% of crashes occur on curved roads (MORTH, 2020). Based on the crash data, horizontal curves are one of the most accident-prone locations. The drivers may misperceive the road geometry, leading to driver errors that eventually might lead to crashes (Gaweesh et al., 2019; Yang et al., 2020). Drivers will make even more mistakes if such horizontal curves are superimposed with vertical alignment (Venthuruthiyil, 2021). Keeping these things in mind, this study has collected naturalistic driving data using the on-site videos from the rural highways passing through the flat and mountainous terrains of LMICs such as India. The present study does not address the crash risk associated with the intersections of the rural highways.

1.5 Organization of Thesis

The thesis work is organized into seven chapters. Chapter 1 introduces the research problem and discusses the pertinent need. It also includes the objectives and scope of the study. Chapter 2 contains the literature review of the proactive safety assessment technique. It discusses the different existing surrogate safety measures, their pros and cons, recent developments in the SSMs, and the challenges associated with employing these measures for mixed traffic conditions. It further includes the methods establishing the link between SSMs and crashes. Finally, it summarizes the state of the art in proactive safety assessment and identifies the research gaps. Chapter 3 comprises the data collection, extraction, and the study's research methodology. It mainly discusses the surrogate safety measure used in the study to define traffic conflict and the modelling approach of EVT employed for the crash risk estimation. Chapter 4 includes a detailed analysis of various conflict types on rural highways using the chosen SSM. Chapter 5 discusses the development of stationary crash risk models of mixed traffic prevalent on rural highways using the EVT. Chapter 6 illustrates the non-stationary EVT models developed incorporating the effect of microscopic variables related to driving behavior. Finally, Chapter 7 includes the summary of the work, findings of the study, and further scope of the study.

Chapter 2 Review of Proactive Safety Approach

This chapter presents a detailed review of the various aspects of the proactive safety analysis pertaining to the objectives of the study. Section 2.1 provides a brief introduction to the proactive safety approach. Section 2.2 describes the widely used surrogate safety measures that capture the crash risk and injury severity. It further highlights the importance of multidimensional surrogate safety measures in near-crash identification. Section 2.3 presents the studies that have used EVT in traffic safety applications from the context of crash risk modelling. The section also points out the advantages of EVT over traditional regression models. Section 2.4 reviews the data collection techniques used to perform the crash risk modelling and the need for trajectory data collection for proactive safety analysis. Finally, Section 2.5 points out the research gaps which motivated the current work in this study.

2.1 Proactive Safety Approach

Traffic crashes are random events; even if the traffic conditions remain the same for a given location, the number of traffic crashes could differ yearly. This implies that the actual number of crashes is also an indirect metric, and the true safety metric is the "expected number of crashes," which cannot be measured but can only be calculated using historical crash data or some other approach (Güttinger, 1984; Laureshyn & Varhelyi, 2020). However, due to the rare occurrence of crashes, several years of crash data is required to provide the "expected number of crashes" with reasonable statistical properties (Davis et al., 2011). This has led to the search for an alternative approach that uses non-catastrophic events, such as traffic conflicts, to derive information about catastrophic events, such as crashes (Cavadas et al., 2020; Zheng et al., 2021).

Perkins and Harris (1969) coined "conflict" to describe an unsafe interaction in traffic safety studies. Traffic conflicts represent traffic crash potentials and are defined by the presence of evasive actions such as braking or weaving (i.e., changing lanes) to avoid a crash (Amundsen & Hyden, 1977; Perkins & Harris, 1969). However, there is confusion regarding the position of the conflict in the series of events leading to a crash. **Figure 2.1** illustrates this situation showing two types of conflicts: one representing the complete event (**Figure 2.1** (b)), meaning a potential crash that did not result in a crash (usually referred to as near-crash or severe conflicts). While the other event represents a conflict with significantly lower crash potential (**Figure 2.1** (a)), usually referred to as potential conflicts.

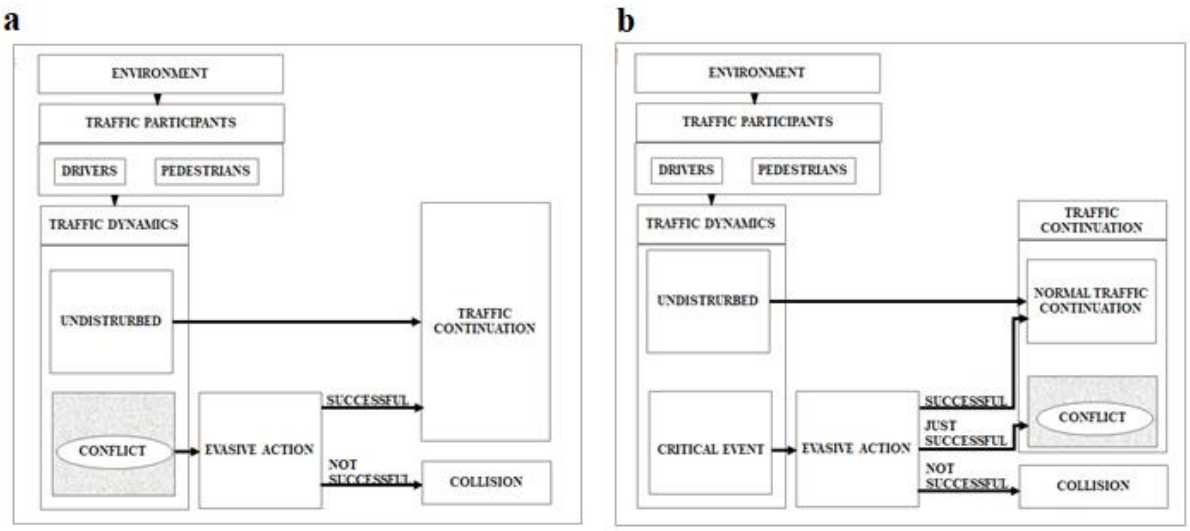


Figure 2.1 Position of Conflict in the Traffic Collision/Crash Process

This concept can be further understood by studying the safety pyramid (Hydén, 1987), describing the hierarchy of road user interactions. **Figure 2.2** represents a series of basic events describing the interaction of road users. Each event happens with a different probability or expectancy associated with varying levels of severity. The safety pyramid concept assumes that serious conflicts represent the interactions with the highest crash expectancy. Further, at that level, a fundamental relationship exists between the frequency of severe conflicts and crashes with a well-defined crash probability (Hydén, 1987). However, the requirements are:

1. Selection of appropriate traffic conflict indicator that can describe the severity of a traffic event
2. Select an appropriate threshold that can define the boundary between slight and severe conflicts. This is essential for establishing the connection between conflicts and crashes.

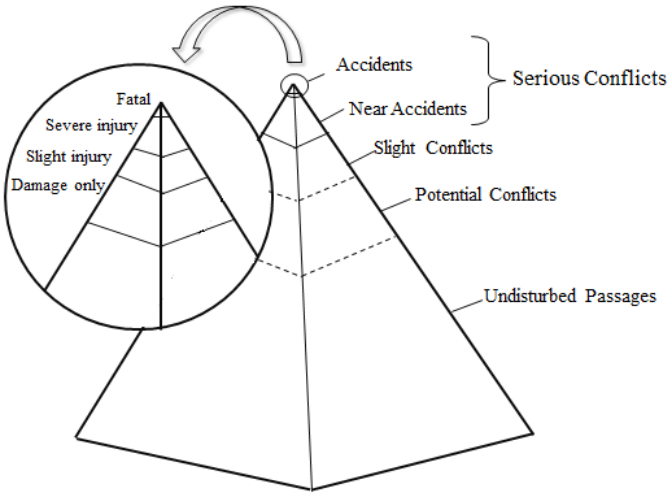


Figure 2.2 Safety hierarchy of traffic events

2.2 Selection of Suitable Traffic Conflict Indicator or Surrogate Safety Measure (SSM)

The proactive safety approach necessitates the selection of a suitable SSM that can be used to describe a traffic event that progresses from a normal interaction to a traffic conflict and eventually to a crash (Arun et al., 2021). The severity of an event can be measured in two different ways: crash risk and injury severity (Laureshyn et al., 2010). Crash risk measures can be broadly classified into two categories: proximity measures and evasive action-based measures. Proximity measures identify critical events based on the closeness or proximity between vehicles in either space or time (Hayward, 1972; Allen et al., 1978). Time proximity is the most used measure since it incorporates spatial proximity and speed (Zheng et al., 2014b). Evasive action-based measures identify critical events based on the magnitude of any evasive action, namely braking, running, or swerving (Davis et al., 2011). The crash risk or crash probability increases with the decrease in proximity between vehicles and the increase in the magnitude of the evasive action (Shelby, 2011).

2.2.1 Crash Risk Measures

Table 1 summarizes the various details of the crash risk measures broadly classified into proximity and evasive-action-based measures. Time-to-collision (TTC) and post-encroachment time (PET) are the most widely used of all proximity measures. TTC is the time for a potential collision if the interacting road users' speed and direction remain unchanged (Hayward, 1972). TTC-like indicators include time to accident (Hydén, 1987), time-exposed TTC (Minderhoud & Bovy, 2001), time-integrated TTC (Minderhoud & Bovy, 2001), and modified TTC (Ozbay et al., 2008).

PET is the time difference between the moment a vehicle leaves the point of a potential collision and the other vehicle arrives at the point of collision (Allen et al., 1978). PET-like indicators include gap time (Gettman & Head, 2003) and headway (Vogel, 2003). However, TTC and PET primarily identify rear-end and crossing conflicts (Arun et al., 2021).

The literature suggested deceleration and angle-based indicators for measuring the magnitude of evasive actions. The deceleration rate to avoid a crash (Cooper & Ferguson, 1976) is the most frequent and extensively used evasive action-based measure. Additional indicators include the deceleration to safety time (Hupfer, 1997), jerk (Tageldin et al., 2015), and yaw rate (Guo et al., 2018). Yaw rate measures the intensity of swerving by the angular velocity of the vehicle's rotation. These indicators tend to perform better than proximity indicators in measuring the crash risk of vulnerable road users (VRUs), such as powered two-wheelers (PTWs) in LMICs like India (Tageldin et al., 2015; Guo et al., 2018).

However, evasive action-based measures are susceptible to providing a vague boundary between potential conflicts and normal events because numerous crashes occur with no evasive action (Zheng et al., 2014b). Further, evasive actions such as braking or changing lanes without closer proximity are mainly precautionary events and do not signify risky situations (Allen et al., 1978; Glauz & Migletz, 1980; Chin & Quek, 1997). Hence, an ideal surrogate safety indicator should integrate proximity and evasive actions to assess the crash risk/crash probability (Johnsson et al., 2018). Recently, Venthuruthiyil & Chunchu (2022) proposed a new indicator named anticipated collision time (ACT) which combines proximity and the evasive maneuvers of the vehicles, such as braking or accelerating and swerving. Existing proximity or evasive action-based measures capture only a few aspects of traffic conflict. ACT combines proximity and evasive action, capturing a more comprehensive evolution of traffic conflict to being a crash. Further, it includes the steering effect of vehicles in the formulation, which improves conflict identification. ACT can also capture multiple conflict types, such as rear-end, sideswipe, head-on, angled, and run-off-road (ROR) conflicts for all highway facility types. As a result, ACT meets all of the criteria for an ideal indicator and can be used to measure the crash risk. More details on ACT are provided in the next chapter.

Table 2.1 Summary of Crash Risk Measures

SSM	Meaning	Strengths	Weaknesses	Remarks
Time-to-Collision (TTC) (Hayward, 1972)	TTC is the time remaining for two vehicles to collide if they continued on their course at the same rates	Easy to measure, interpret	Capture partial aspect of crash risk due to constant relative speed assumption	One of the most widely used SSM particularly for the rear-end conflicts (Arun et al., 2021)
Time-to-Accident (TA) (Hydén, 1987)	Time remaining to collision from the moment the first evasive action is taken by one of the road users	Easy to measure. Considers evasive action.	Suffers from the same issues as that of TTC due to constant relative speed assumption	Hardly used in conflict studies. Swedish traffic conflict technique uses it.
Time-Exposed Time-to Collision (TET) (Minderhoud & Bovy, 2001)	Summation of all moments that a driver approaches a front vehicle with a TTC value below the threshold	Same as TTC	Just an extension of TTC. Faces same issues as that of TTC	Measure of exposure but rarely used in conflict studies
Time Integrated Time-to-Collision (TIT) (Minderhoud & Bovy, 2001)	Integral of the TTC-profile during the time it is below the threshold	Provides a higher degree of risk measure	Difficult to interpret	Hardly used in conflict studies

Modified Time-to-Collision (MTTC) (Ozbay et al., 2007)	Modified TTC based on relative speed and relative acceleration of the conflicting vehicles	Improvement on TTC	Assumes constant relative acceleration between conflicting vehicles. Hence captures partial aspect of crash risk	Mostly used for rear-end conflicts (Charly & Mathew, 2019). Recent studies (Arun et al., 2022; Arun et al., 2021) used MTTC and obtained higher correlation to crashes than that of TTC.
Post-Encroachment Time (PET) (Allen et al., 1978)	the time difference between the moment a vehicle leaves a potential conflict point and the other vehicle arrives at the point.	Easy to measure, interpret	Captures partial aspect of crash risk	One of the most widely used SSM particularly for the crossing conflicts (Arun et al., 2021)
Headway (Vogel, 2003)	Inter-arrival time between two successive vehicles moving on a lane	Easy to measure, interpret	Applicable mostly for lane-based car-following behavior	Rarely used in conflict studies
Percent Stopping Distance (PSD)(Oh et al., 2006)	Ratio between the distance remaining to a potential collision point and the minimum acceptable stopping distance	Considers braking which is an important evasive action. Considers tire-pavement interaction	It does not consider the actions taken by the leader vehicle	Rarely used in conflict studies. Arun et al. (2022) found that PSD provided the wider crash frequency estimates than that of temporal proximity measure such as MTTC
Evasive Action Based Measures				

Deceleration Rate to Avoid Crash (DRAC)	Maximum deceleration required by the follower to avoid collision with the leader (Arun et al., 2022)	Combination of DRAC and proximity measure can capture the complete aspect of crash risk. Used for defining near crashes in naturalistic driving studies (Davis et al., 2011).	Assumes constant relative speed. Not suitable for lateral movement. Conflict and crash thresholds are not clearly defined.	It is the most widely used evasive action-based measure. Weak correlation exists between crashes and conflicts (Arun et al., 2021).
Jerk (Bagdadi & Várhelyi, 2011)	Rate of change of vehicle acceleration with respect to time. It is a measure of smoothness of traffic flow.	Large negative jerk values better identify aggressive drivers from normal drivers as compared to deceleration values (Feng et al., 2017).	Since it is a derivative of acceleration, any noise in the raw data will significantly affect the results. Not useful for measuring the crash risk of PTWs (Guo et al., 2018)	Positive Correlation exists between the number of crashes and the number of critical jerks experienced by the driver (Bagdadi & Várhelyi, 2011).
Yaw Rate (Tageldin et al., 2015)	Angular velocity of the road-user rotation around the z-axis or the rate of change of heading angle	Captures the swerving evasive action. Can identify safety critical events mainly for vulnerable road users such as PTWs (Guo et al., 2018). Useful for side-swipe (sometimes also referred as lane-changing) crash risk assessment (Venthuruthiyil et al., 2022)	Any noise in the raw trajectory data will significantly affect the results.	Combining with proximity and deceleration can capture the complete aspect of crash risk particularly for the heterogeneous traffic prevalent in low- and middle-income countries

Note: It should be mentioned here that there are several other SSMs called crash indices, severity indices developed mainly based on the SSMs mentioned in Table 1. However, they have limited use and can hardly be employed to relate traffic conflicts with crashes.

2.2.3 Injury Severity Measures

Injury severity indicators are mostly speed based measures such as speed and relative speed of the vehicles involved in the conflict (Borsos, 2021). Conflicting speed is used as an injury severity indicator along with the crash risk indicator time-to-accident in the Swedish Traffic Conflict Technique (Svensson & Hydén, 2006). But, studies stated that the consequence of a crash depends on the mass, speed of the participating road users and the angle of collision (Hyden et al., 1982; Hutchinson, 1977; Evans, 2001; Zheng et al., 2014b). Hence, the most widely used injury severity indicator is Delta-V and is defined as the expected change in velocity between the pre- and post-crash trajectories of the conflicting vehicles (Shelby, 2011). The faster the transition from pre- to post-crash, the more severe the outcome. For the proactive studies, Delta-V can only be measured by assuming the future movements of the road users. This makes expected Delta-V a continuous variable because Delta-V values can be calculated for each instant assuming how the interaction would evolve. Hence, it is estimated by assuming a hypothetical collision between two road users at the angle and velocity they have at the point of closest proximity (Laureshyn et al., 2017). Arun et al. (2021) and Arun et al. (2022) stated that the Delta-V threshold of 16 m/s decides the boundary between severe and non-severe conflicts.

2.3 Modelling Approaches for Establishing the relationship between Conflicts and Crashes

Considering the severity indicators discussed in the previous section, one can identify potential conflict situations, which are the frequently observed events in a traffic stream. However, identifying the severe conflicts or near-crashes from potential conflicts is not a straightforward task and requires the selection of an appropriate threshold. There is no consistent approach to determine the threshold for traffic conflicts. One viable approach is to model the severity measure as a continuous dependent variable (Zheng et al., 2021). In this regard, various models have been employed, some of which are based on crash data, while others rely solely on conflict data. They can be broadly classified into two categories: Crash data based and conflict data-based models.

2.3.1 Crash data-based Models

These models employ several statistical methodologies to establish the relationship, with crash frequency as the dependent variable and traffic conflict frequency as one of the independent variables. The first attempt in this regard was done by Hauer (1982) when he estimated the crash frequency using linear regression model. Several studies (Guo et al., 2010; Hauer & Garder, 1986;

Migletz et al., 1985; Sayed & Zein, 1999) developed such models to relate conflicts with crashes. The standard mathematical form is,

$$\lambda = c \times \pi \tag{2.1}$$

where, λ and c denotes the number of crashes and conflicts, respectively, to occur on a road entity during a certain period of time, and π is the crash-to-conflict ratio for that entity, which is estimated using regression models (Hauer, 1982). To represent conflicts with different severity levels as shown in Figure 2, the Equation (2.1) was further extended as

$$\lambda = \sum c_i \times \pi_i \tag{2.2}$$

where c_i indicates the number of conflicts corresponding to a particular severity level (i) and π_i indicates the crash-to-conflict ratio for conflicts of that severity level (i) (Hauer & Garder, 1986).

However, exposure parameters such as traffic volume are one of the main crash indicators (Sacchi & Sayed, 2016). The number of traffic events where a plausible series of events could result in a crash between two road users is known as exposure (Ismail et al., 2011). Hence, for estimating the crashes based on exposure and other traffic variables, many studies (Bagdadi & Várhelyi, 2011; Cafiso et al., 2018; El-Basyouny & Sayed, 2013; Guo et al., 2010; Peesapati et al., 2018; Sacchi & Sayed, 2016; Shahdah et al., 2014; Stipancic et al., 2018) have employed generalized linear regression models. Such models include Poisson (PN), negative binomial (NB), and Poisson-lognormal (PNL) and these models are also known as safety performance functions (SPF). Applying the NB model, El-Basyouny and Sayed (2013) estimated the crash frequency based on the conflict frequency. They modeled the conflict frequency using the traffic volume and geometric variables with the help of PNL model. Peesapati et al. (2018) formulated two sets of NB models to estimate crash frequency: one using PET and the other using exposure and PET. The second model provided better estimation results. Further, Stipancic et al. (2018) used NB models to relate crash frequency with surrogate safety measure such as jerk and traffic flow measures such as average speed, speed variance, and congestion index.

Safety performance functions are widely used to assess the crash risk and provide safety counter measures (Sacchi & Sayed, 2016). However, these models employ crash frequency as the dependent variable and the conflict related variables as the independent variable. Crash data commonly has inherent issues such as mistakes, omissions, and underreporting (Stipancic et al.,

2018) whereas conflict data suffers from inconsistent conflict identification and short observation time (Zheng et al., 2021). Due to this, such models suffer from the issues associated with the both dependent (i.e. crash data) and independent (i.e. conflict data) sides of the SPF.

2.3.2 Conflict data-based Models

These models include extreme value theory models which do not require crash data for the calibration and extrapolates the crash risk from the observed traffic conflicts by identifying true extremes (Cavadas et al., 2020). Further, EVT model eliminates the mixing of several safety pyramid levels (Hydén, 1987) by identifying the true extremes (i.e., severe conflicts) from the potential conflicts (Ali et al., 2022; Cavadas et al., 2020). EVT was first introduced by Campbell et al. (1996) and the modelling methods to relate conflicts and crashes were developed by Songchitruksa and Tarko (2006). 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 approach divides the observed time period into continuous blocks of fixed size and then considers the maximum value (or r largest values) from each time block as an extreme. POT picks the highest values as extremes as long as the values are over a certain threshold. Ali et al. (2022) and Zheng et al. (2014a) found that POT outperforms BM for the smaller study periods. Farah & Azevedo (2017) found that BM provides more stable results than POT. Zheng et al. (2018) stated that the BM method might be preferable since implementing the POT method involves a degree of subjectivity resulting primarily from choosing a threshold value. There is no consensus among the researchers regarding using a particular method. This is evident from the recent use of both BM and POT models in crash risk studies (Zheng et al., 2019; Fu et al., 2020; Arun et al., 2021; Arun et al., 2022).

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, 2012a).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).

Table 2 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 lacks 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.

Table 2.2 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	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. (2014)	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	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	MTTC, PET, & 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

Numerous studies (Songchitruksa & Tarko, 2006; Tarko, 2012; Jonasson & Rootzén, 2014; Zheng et al., 2014a; Farah & Azevedo, 2017; Zheng et al., 2018; Zheng & Sayed, 2019; Zheng et al., 2019b; Fu et al., 2020; Fu & Sayed, 2021) have developed EVT models using only crash risk based measures, such as TTC, PET, DRAC or a combination of TTC, PET & DRAC. But these models do not incorporate the injury severity i.e., the possible consequences of a crash such as vehicular damage, minor injury, serious injury, and fatality. The use of injury severity indicators in modelling the conflict severity requires further attention (Borsos, 2021). The severity of an event can be estimated by combining the crash risk and injury severity indicators (Laureshyn et al., 2010; Johnsson et al., 2018). Recently, a few studies (Arun et al., 2021; Borsos, 2021; Arun et al., 2022) have incorporated injury severity indicator into EVT modelling framework to combine both aspects of the severity. However, these studies estimated the crash risk of homogeneous traffic streams (i.e., cars) for the rear-end or crossing conflicts of shared traffic facilities present in HICs. Hence, the framework still requires further testing, particularly in the context of LMICs.

2.4 Importance of Trajectory Data in the Proactive Safety Modelling

One of the main advantages of proactive safety modelling is that it captures the microscopic driving mechanism which eventually leads to crashes. Different sensor technologies such as presence detectors (Oh et al., 2006; Theofilatos et al., 2019; Xu et al., 2013, 2014, 2015) and video cameras, have been used to gather data for the proactive safety assessment. However, the microscopic driving variables influencing the crash mechanism, such as braking, swerving and other evasive actions, cannot be fully captured using loop detectors and other presence sensors (Roshandel et al., 2015). For collecting the microscopic driver behavior data (or trajectory data), two approaches are widely used: (i) driving simulator study (Choudhary & Velaga, 2017; Pawar et al., 2020) and (ii) naturalistic driving study. In the driver simulator study, the trajectory data is obtained in a controlled environment using vehicle simulators. However, studies (Santos et al., 2005; Chao Wang et al., 2009; Zöller et al., 2019) found that drivers behave differently in a controlled environment as opposed to the naturalistic driving. Hence, naturalistic driving study has emerged as a better alternative to capture the trajectory data. Naturalistic driving study can be further classified into two categories: (i) using probe vehicles (Dingus, Neale, et al., 2006; Dingus et al., 2016; Wu et al., 2014; Wu & Jovanis, 2012) and (ii) using location specific video cameras (Arun et al., 2022; Arun et al., 2021; Autey et al., 2012; Songchitruksa & Tarko, 2006; Tageldin

et al., 2017; Zheng et al., 2018; Zheng & Sayed, 2019b) or unmanned aerial vehicles (UAVs)(Gu et al., 2019; Liu et al., 2021; Venthuruthiyil et al., 2022; Venthuruthiyil & Chunchu, 2022; Wang et al., 2019).

There are few large-scale exhaustive probe vehicle studies such as 100-car naturalistic driving study (Dingus et al., 2006), SHRP 2 naturalistic driving study (Antin et al., 2011), Australian 400-car naturalistic driving study (Regan et al., 2013), and UDRIVE (Barnard et al., 2016). These studies employ various devices such as Global Positioning System (GPS), accelerometer, LiDAR sensors, video cameras, and eye tracking devices to collect data from the driver, vehicle, surroundings and thereby captures the complete driver behavior. However, such data collection process is highly expensive and requires tremendous resources (Singh & Kathuria, 2021), which is not feasible for the LMICs due to the economic constraints. Further, such large datasets create massive storage issues (Ellison et al., 2015; Valero-Mora et al., 2013), and are highly prone to noise and other quality issues which in turn requires tremendous post-processing of data (Ellison et al., 2015; Hallmark et al., 2015). The use of trajectory data extracted from the on-site videos can eliminate some of the issues. Trajectory data can be extracted quickly and accurately due to the advancements in image processing technologies and computing power (Xie et al., 2019a). Such trajectory datasets contain the detailed description of drivers' behavior and therefore can capture the potential crash occurrence with the help of suitable SSMs (Ozbay et al., 2008; Xie et al., 2019b) and these near-crash events can then be used to estimate the crash risk levels (D. Yang et al., 2021). However, the application of trajectory data for the proactive safety assessment in the heterogeneous traffic settings, requires further attention.

2.5 Summary

From the review, it is evident that there are several gaps in the state-of-the-art of proactive safety assessment. The major gaps are:

1. For estimating the crash risk, EVT emerged as a powerful tool as opposed to the widely used SPFs developed using generalized linear regression models such as NB models or linear regression models. The application of EVT models for the crash risk estimation of VRUs in the heterogeneous traffic environment requires further attention. Further, existing EVT models use SSMs that can only capture a partial aspect of the crash risk. The systematic workaround

of using the newly developed SSM called ACT in the crash risk estimation of mixed traffic is seriously lacking in the literature.

2. For capturing the heterogeneity in crash occurrence (also known as non-stationarity), aggregated traffic variables such as exposure, road geometry are mainly incorporated in the EVT models. Developing non-stationary EVT models by considering the microscopic driving characteristics such as braking, accelerating and swerving actions, which significantly influence the crash occurrence, is absent in the literature.
3. Few studies have used injury severity measures (also known as speed related measures) to improve the performance of EVT models. However, the application of suitable injury severity measure in conjunction with the newly developed ACT requires further analysis for the crash risk estimation of mixed traffic.
4. Most proactive safety studies employ driver simulator data or naturalistic driving data (using probe vehicles) to obtain the conflict information. The application of trajectory data extracted from the on-site videos collected using UAVs requires further assessment for the proactive safety modelling especially in the context of LMICs. Besides such datasets contain detailed information of all the microscopic variables such as speeds, spacings, accelerations, steering of each road user.

This research work addresses the gaps listed above. The following chapters discuss contributions of this research work that fills the gaps in the state-of-the-art.

Chapter 3 Data and Methodology

Most crashes occur due to driver error hence understanding the driving mechanism is the key to improve the road safety. Naturalistic driving data is essential to understand and improve the road safety using proactive safety assessment. This can be broadly achieved in two ways: i) individual probe vehicle study and ii) location-based study. For individual probe vehicle study, the naturalistic driving data is captured using various sensors, GPS devices, video cameras, and other eye tracking devices installed in the vehicles. Whereas for location-based study, naturalistic driving data is collected from a specific location using video recording devices such as fixed video cameras, unmanned aerial vehicles (UAVs). The vehicle instrumented data collection approach is highly expensive and needs exhaustive resources. On-site trajectory data collection approach can overcome some of the limitations and therefore was used as a method to collect naturalistic driving data in this study. Section 3.1 describes the trajectory data collection and extraction process. The section further includes the noise removal from trajectory data using state-of-the-art smoothing algorithm.

Once the microscopic variables describing the driving behavior are obtained from the smoothed trajectories, crash risk assessment can be carried out with the help of proactive safety approach. Section 3.2 explains the entire research methodology used in the study. It starts with a brief description of the chosen SSM and the potential conflict identification using the SSM. It then discusses the crash risk modelling approaches used in the study.

3.1 Trajectory Data Collection and Extraction

The traffic videos were collected using an UAV for a duration of two hours (four-lane) and 1.5 hours (six-lane), covering a road stretch of 700 meters. The location of data collection was Kerala, a southern state in India, which is one of the highest contributors to the road traffic crashes (7.6%) in India. Data were collected from the mid-block sections of four-lane, and six-lane divided carriageways with negligible gradient, and no roadside distraction to the vehicular traffic (**Figure 3.1).**

Salem-Kochi Highway (NH 544) Kozhikode-Palakkad Highway (NH 966)

Figure 3.1 Data Collection Sites for the present study

Four-lane divided highways consists of median lane (ML) and shoulder lane (SL) along with the paved shoulder, in each direction. In the case of six-lane divided highways, drivers use ML, center lane (CL), SL, and the paved shoulder for their movement. Each lane has a road width of 3.5 meters with a paved shoulder of 1.5 meter each (**Figure 3.2**).

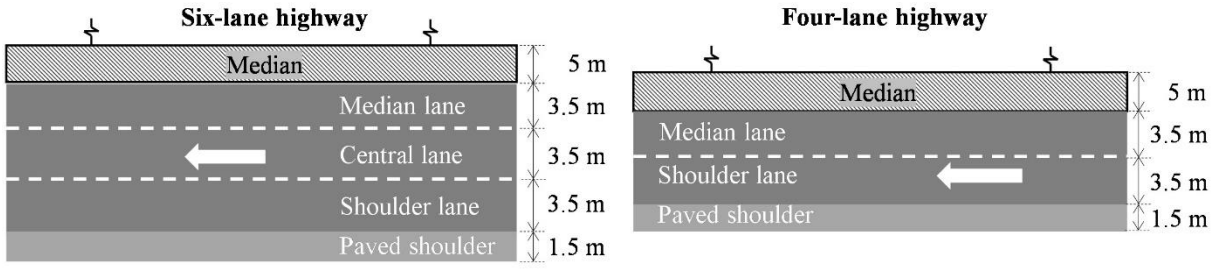


Figure 3.2 Cross-sectional view of the Study Sites

For analyzing the crash risk on horizontal curves, traffic videos were collected using a fixed camera and UAV from a two-lane undivided rural highway passing through mountainous terrain. The location is the Shillong Bypass connecting NH-40 and NH-44 which spans over a road length of 48 km and has a maximum elevation difference of 461.93 m. The naturalistic driving data was collected from fifteen curves for one hour each. **Figure 3.3** shows the aerial view of the few curved sections from where the data was collected in the present study.

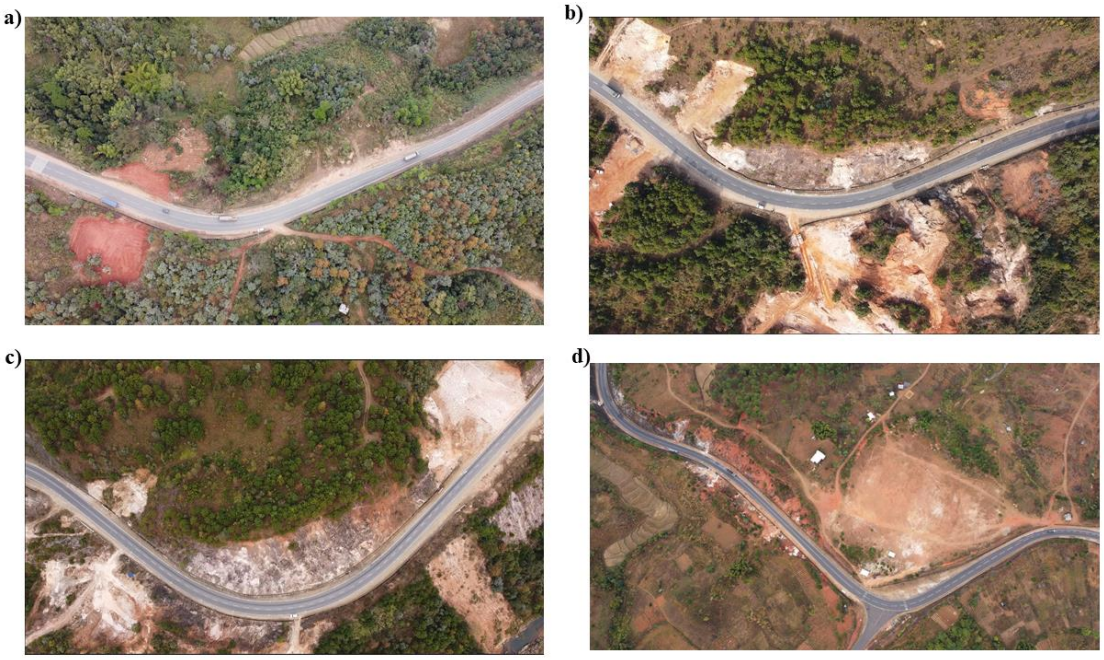


Figure 3.3 Data Collection Sites on Shillong Bypass

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: pre-processor, tracker, and analyzer (**Figure 3.4**). In the pre-processing unit, camera calibration and the tracking boundary were defined. The termination of tracking task is ensured by providing the tracking boundary and can be achieved by drawing a polygon on the video image. The camera calibration involves the transformation of image coordinates to real-world coordinates. This is achieved by using a homography matrix, which indicates the mapping of a set of points from pixel coordinates to cartesian coordinates. It is estimated by considering a minimum of four known points in the field and their corresponding image coordinates. For detailed understanding, the readers can refer to the work of Venthuruthiyil (2021). Further, for the non-planar terrains such as horizontal curves passing through mountainous regions, multiple homographies were estimated. This is accomplished by dividing the entire non-planar terrain into piece-wise planar regions and then estimating a separate homography matrix for each one. For further details, the readers can refer to the study of Bharat et al. (2022). The vehicle detection, classification, and tracking were performed in the tracker module. In the Analyzer module, different corrections to the tracked vehicle's path such as identification and correction of faulty

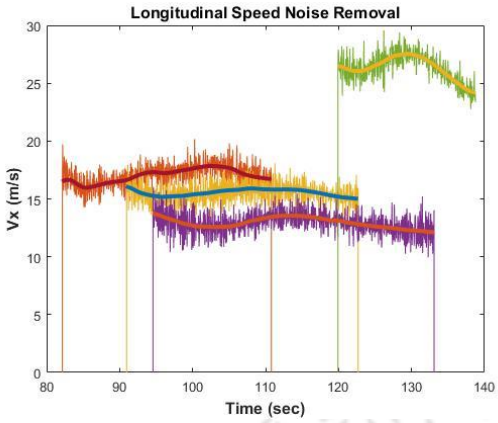
tracks, camera movement compensation, reconstruction of the vehicle path to remove the embedded noise were applied and various vehicle kinematic variables were estimated.

(a) (b)

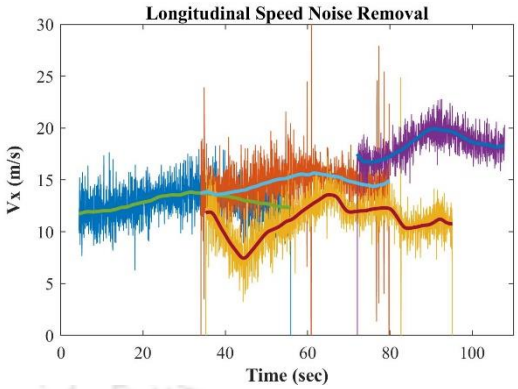
Figure 3.4 Tracking of Vehicles moving on (a) Kozhikode-Palakkad and (b) Salem-Kochi highways Using SAVETRAX

Once the trajectories were extracted, the noise embedded in the trajectories was removed using the smoothing technique proposed by Venthuruthiyil & Chunchu (2018, 2020b). 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 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. **Figure 3.5** shows the noise embedded in the longitudinal and lateral speeds and longitudinal and lateral accelerations of few PTWs moving on the four-lane six-lane rural highway. From the figure, it can be clearly seen that the noise is relatively higher in the longitudinal and lateral accelerations of PTWs. However, the state-of-the-art smoothing algorithm employed in the study, removes the noise satisfactorily in both directions and helps to get smooth trajectories. For an extensive understanding of the smoothing algorithm, the readers can refer to the works of Venthuruthiyil & Chunchu (2018) and Venthuruthiyil & Chunchu (2020b).

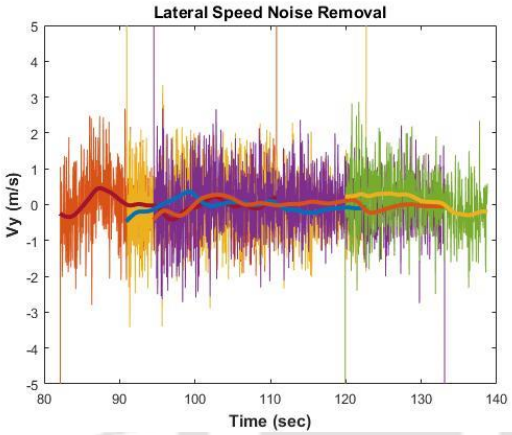
The accuracy of the smoothed trajectories were examined for the different driving conditions in the previous studies (Bharat et al., 2022; Venthuruthiyil and Chunchu, 2018, 2020b).



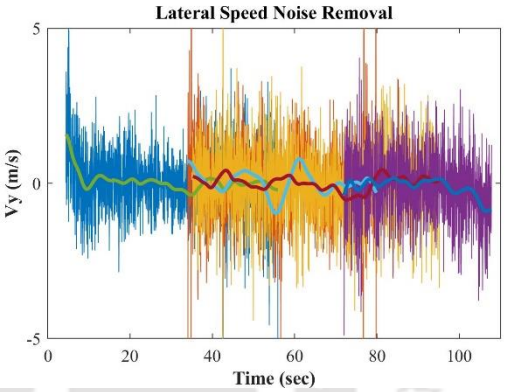
(a)



(b)



(c)



(d)

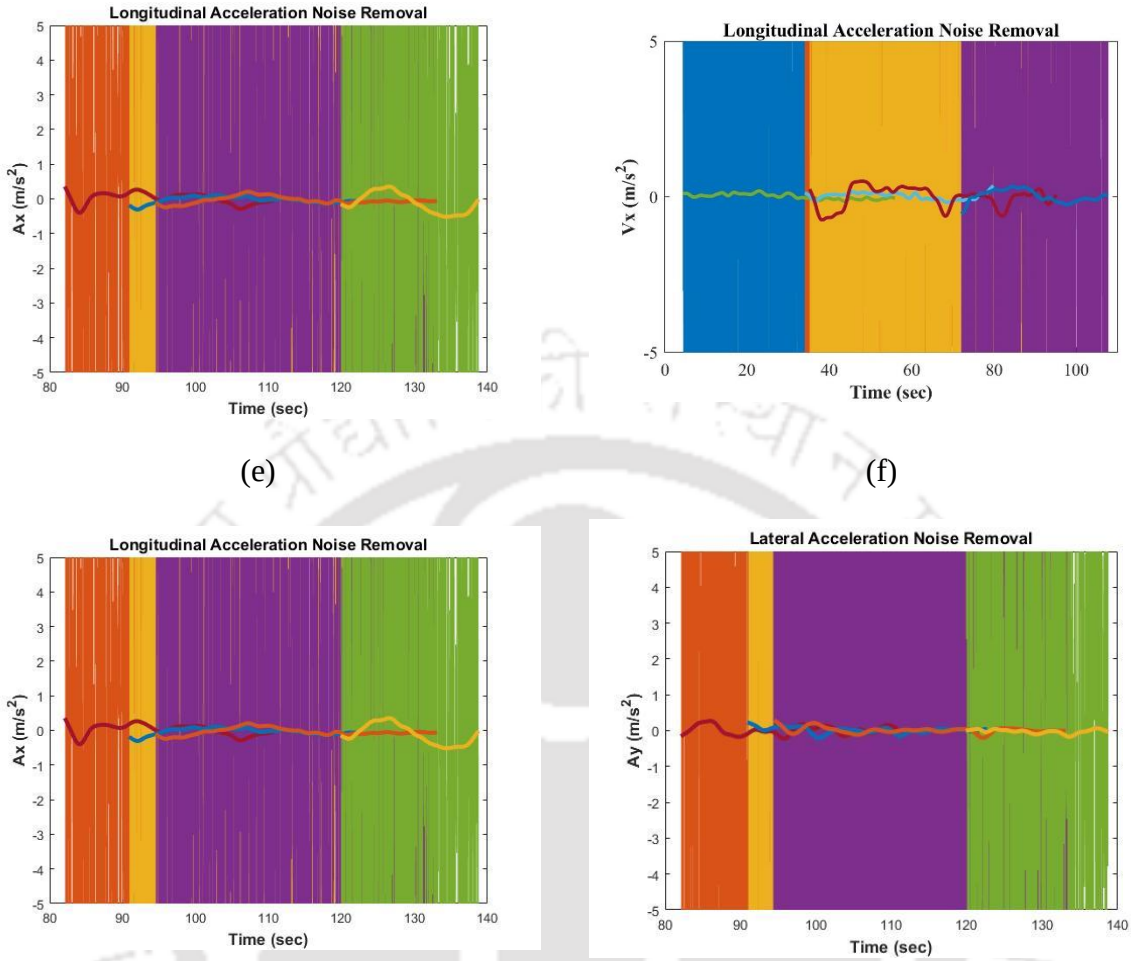
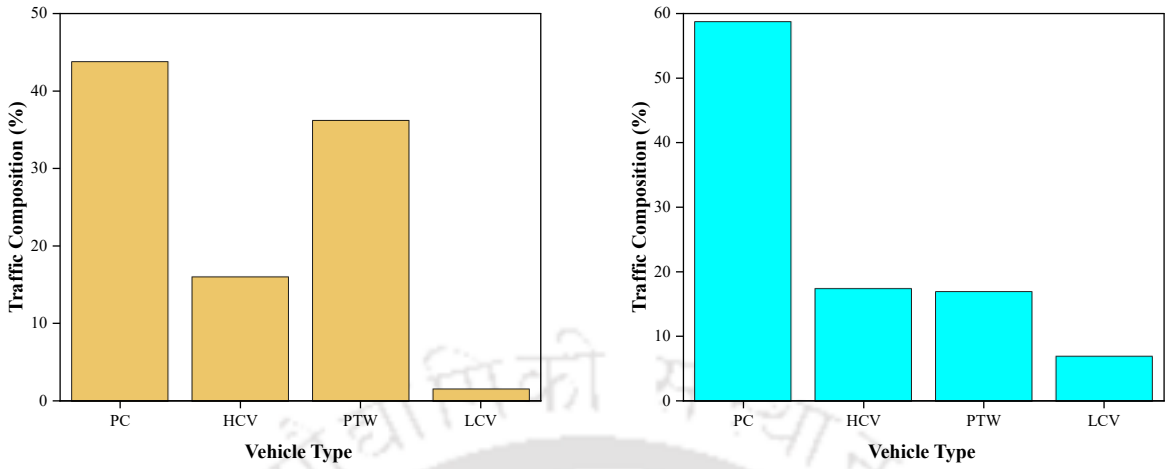


Figure 3.5 Example of Noise Removal of (a-b) Longitudinal Speed (c-d) Lateral Speed (e-f) Longitudinal Acceleration (g-h) Lateral Acceleration of the Trajectories on four-lane and six-lane highways

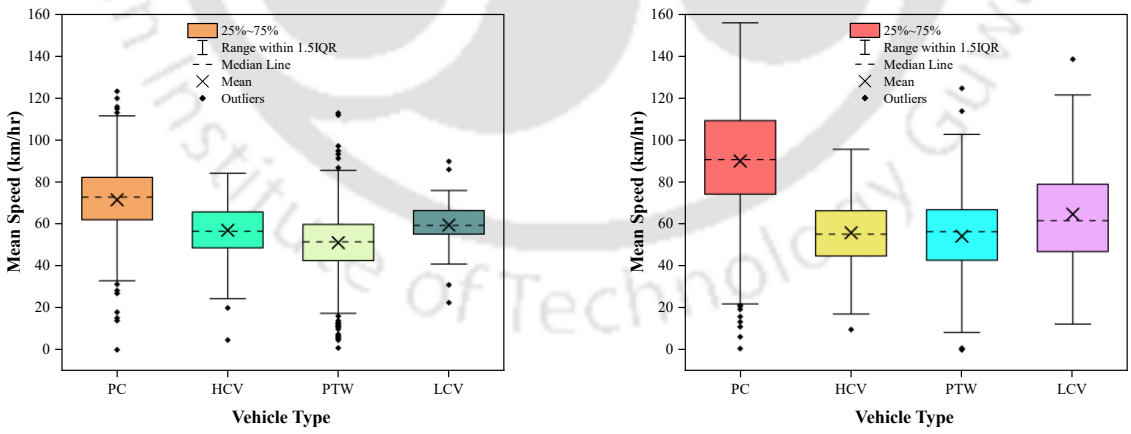
The four-lane and six-lane highways contain 1916 and 1250 trajectories, respectively. The major constituents in the traffic stream were passenger car (PC), heavy commercial vehicle (HCV), light commercial vehicle (LCV), and powered two-wheelers (PTW) (**Figure 3.6**).



(a) (b)

Figure 3.6 Traffic Composition on the (a) four-lane and (b) six-lane divided highways

The speed difference between the vehicle types is significantly high (**Figure 3.7**) and may lead to frequent interactions such as passing and overtaking. However, the number of LCVs is significantly less on both four-lane and six-lane highways (i.e., 38 and 75 trajectories on four-lane and six-lane highways respectively). Therefore, the present study mainly focuses on the crash risk analysis of PC, PTW, and HCV. Further, the speed difference is higher in the case of six-lane highway than that of the four-lane highway (**Figure 3.7**).



(a) (b)

Figure 3.7 Variation of Speed for Each Vehicle Type on (a) four-lane and (b) six-lane divided highways

Figure 3.8 shows the traffic composition of multiple vehicle classes on each lane. PC and HCV occupy both ML and SL in the case of four-lane divided highways, whereas PTWs mostly use the SL. On the other hand, in the case of the six-lane highways, both ML and CL mainly carry PCs and HCVs, with a negligible proportion of PTWs (**Figure 3.8 (b)**). PTWs mostly use the SL, whereas less than 7% of PCs and HCVs move on this lane.

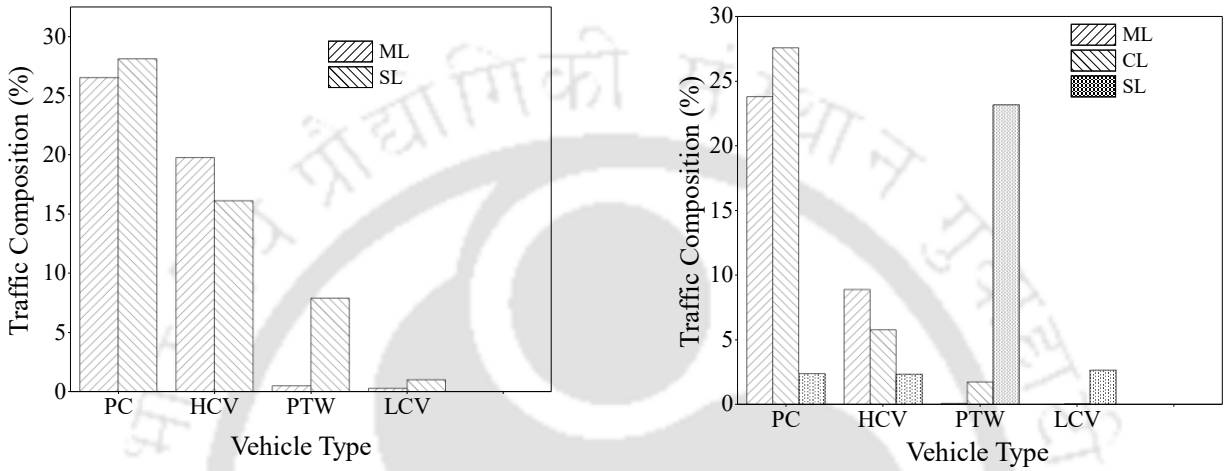


Figure 3.8 Lane utilization of different vehicle types for (a) Four-lane and (b) Six-lane highways

A total of 1762 vehicles were tracked from all sites, including 1067 HCVs, 535 PCs, 132 LCVs, and 21 PTWs. Due to the lower number of PTWs, the present study only focuses on the analysis of HCVs and PCs for the mixed traffic moving on horizontal curves. The National Highway Authority of India (NHAI) provided the geometric information for each curve under examination (Table 3.1).

Table 3.1 Geometry of the curves considered in the study

Curve No	Speed (km/hr)	Direction	Type	R (m)	T_L (m)	Δ	D_C	L_C (m)	e (%)	L_{at} (m)	L_{et} (m)	G_1 (%)	G_2 (%)	G_3 (%)	G_4 (%)	G_5 (%)	L_V	K_{value}	VC_{HC}	L_0 (m)
127	50	Right	Up-grade	100	45	45.21	17.47	33.89	10	63.68	53.49	4.75	4.75	4.75	4.75	4.75	0	0.00	0.00	0.00
171	50	Left	Up-grade	90	50	82.00	19.41	78.81	10	50.82	95.70	5.50	5.50	5.50	5.50	3.56	0	0.00	0.00	0.00
179	50	Left	Down-grade	200	25	23.98	8.74	58.70	5.5	37.55	14.22	-6.00	-6.00	-6.00	-6.00	-6.00	0	0.00	0.00	0.00
183	50	Right	Hog	80	55	81.51	21.84	58.81	10	14.35	103.33	0.50	-1.45	-2.75	-4.05	-6.00	150	23.08	39.21	0.70
189	50	Right	Up-grade	100	45	58.17	17.47	56.52	10	24.98	98.68	4.60	4.60	4.60	4.60	4.60	0	0.00	0.00	0.00
221	50	Left	Down-grade	80	55	85.30	21.84	64.10	10	123.48	54.93	-4.80	-4.80	-4.80	-4.80	-4.80	0	0.00	0.00	0.00
227	40	Right	Sag	80	40	67.78	21.84	54.63	8.8	1.00	5.00	-3.90	-3.88	-3.68	-3.50	-3.50	60	150.00	91.06	2.08
238	50	Right	Down-grade	80	55	120.61	21.84	113.40	10	123.21	43.75	-5.00	-5.00	-5.00	-5.00	-5.00	0	0.00	0.00	0.00
245	50	Left	Hog	150	30	43.17	11.65	83.01	7.4	262.09	44.72	-2.50	-2.50	-3.70	-4.75	-4.75	60	26.67	100	1.68

Note: R = Radius; T_L= Transition Length; L_C = Length of the curve; e = Super-elevation; Δ = Deflection angle; D_C = Degree of curve; L_{at} = Length of approach tangent; L_{et} = Length of exit tangent; G_1 = Gradient at approach tangent; G_2= Gradient at start of curve; G_3= Gradient at middle of the curve; G_4 = Gradient at the end of curve; G_5 = Gradient at exit tangent; L_V = Length of vertical curve; K_{value} = Rate of vertical curve; VC_{HC} = Percentage of vertical curve in a horizontal curve; L_0 = Distance between the mid-points of horizontal and vertical curves

3.2 Modelling of Crash Risk Associated with Mixed Traffic

This section provides a description of the process of potential conflict identification using ACT profiles. Then it describes the EVT approach for crash risk modelling. It ends with discussing the microscopic driving factors influencing the crash risk.

3.2.1 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 (Figure 3.9a). ACT can be computed as:

$$T = \begin{cases} \dfrac{\delta}{\left(\frac{\partial \delta}{\partial t}\right)}, & \text{if } \frac{\partial \delta}{\partial t} > 0 \\ \infty, & \text{otherwise} \end{cases} \tag{3.1}$$

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 (Figure 3b).

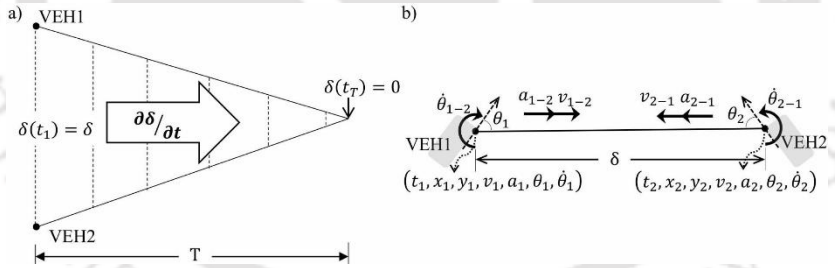


Figure 3.9 Simplified illustration of the concept of Anticipated Time to collision (ACT); (a) Closing-in of two vehicles till collision; (b) Factors influencing the closing-in rate of two vehicles. (**Source:** Venthuruthiyil, 2021)

The widely used measure TTC assumes that the rate at which two vehicles approach each other is only a function of constant relative speed and estimates the collision time based on this assumption. However, collision time can be more accurately estimated if the higher vehicle dynamics of the interacting vehicles, such as acceleration/deceleration and steering rate, can be incorporated into the closing-in rate calculation. Hence, 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 (3.2).

$$\frac{\partial \delta}{\partial t} = Rel\left(\vec{v}_{1-2}, \vec{v}_{2-1}\right) + Rel\left(\vec{a}_{1-2}, \vec{a}_{2-1}\right) \times t + Rel\left(\dot{\theta}_1, \dot{\theta}_2\right) \times \delta \tag{3.2}$$

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 $\dot{\theta}$ 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).

3.2.2 Identification of Potential Conflicts

Given the potential benefits of ACT in the proactive safety analysis, this study considered ACT as the SSM to identify conflicts. In the case of ACT, a conflict course exists only when the line drawn from the approaching vehicle's nearest corner crosses through the either of the other vehicles' corner points. In other words, if two vehicles meet each other at a point based on their heading angles, then there will be a collision course and ACT can be estimated at that time instant. For example, to identify a sideswipe conflict, a line (parallel to the heading angle) projected from the approaching 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 (**Figure 3.10**). The ACT values for a particular type of conflict can then be calculated for each road user using Equation (3.1).

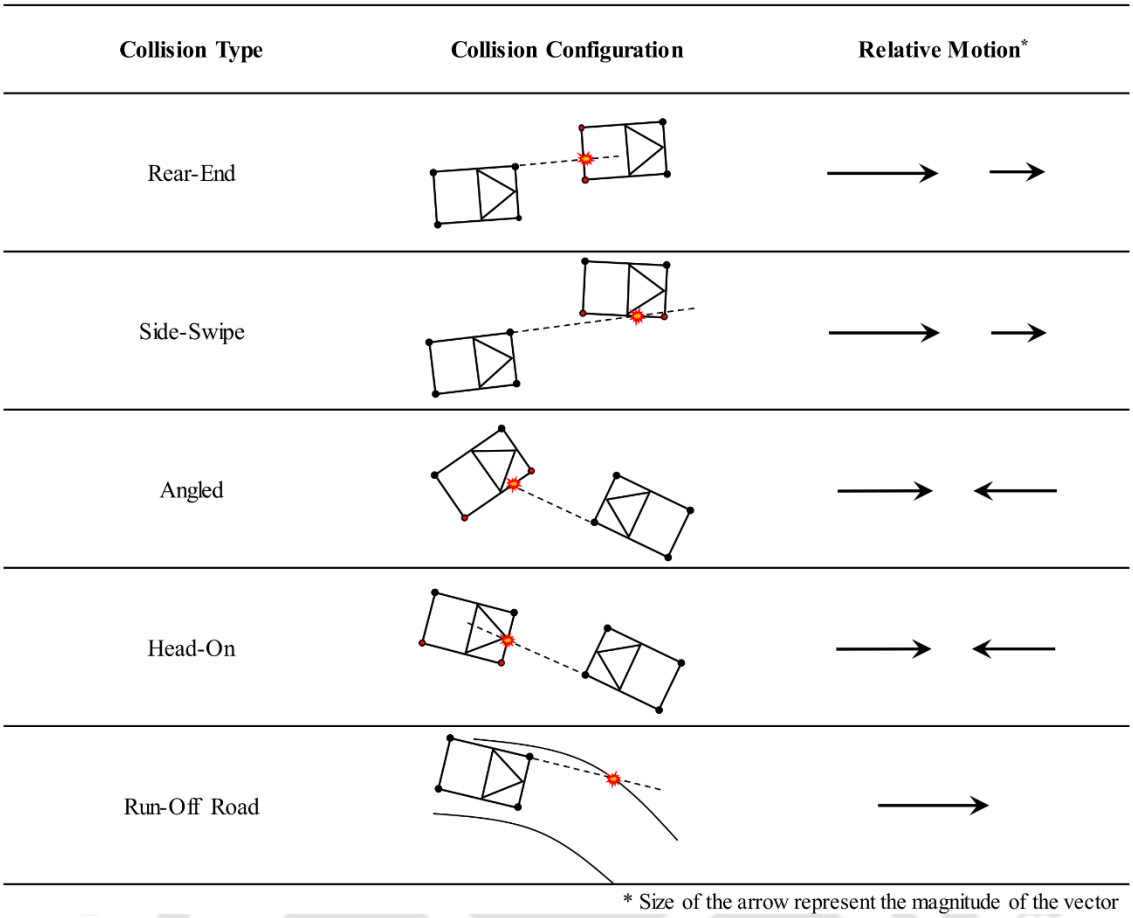


Figure 3.10 Criteria for classification of conflict types (**Source:** Venthuruthiyil, 2021)

ACT is similar to TTC but functions differently and the conflict identification from continuous ACT profiles is similar to that of TTC profiles (**Figure 3.11**). 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 and then the slope should increase indicating the risk reduction (**Figure 3.11**). The ACT_{min} value signifies the least safety margin remaining during or after an evasive action corresponding to a particular conflict and will be further used to estimate the crash risk.

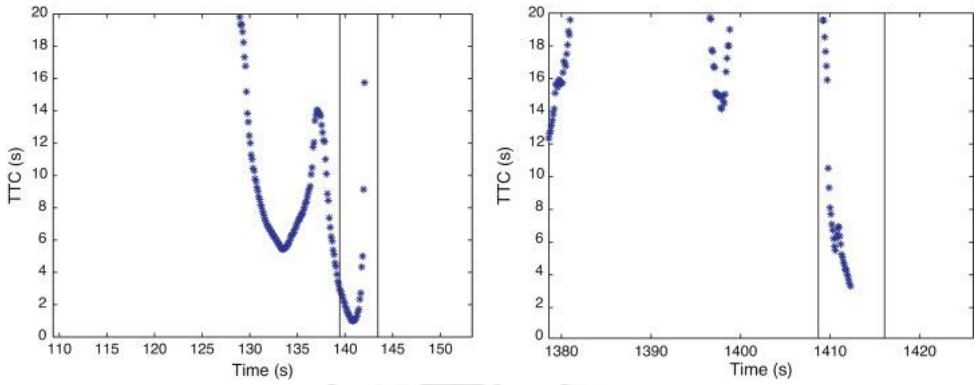
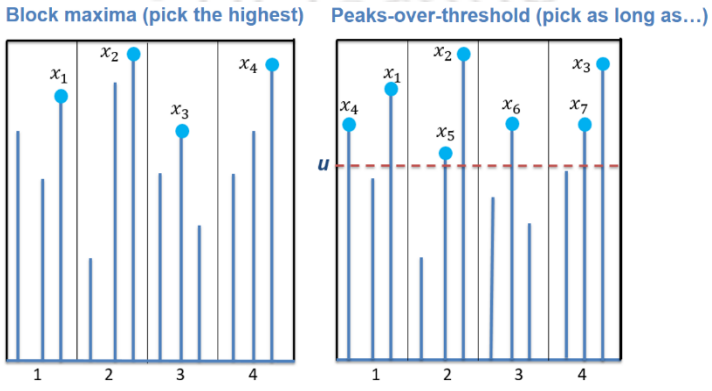


Figure 3.11 Illustration of the Potential Conflict Identification from Continuous TTC Profiles (**Source:** Jonasson and Rootzén, 2014)

3.2.4 Extreme Value Theory Modelling

Considering the surrogate safety indicator discussed in the previous section, one can estimate potential conflict situations and frequently observed events in a traffic stream. However, translating this information into the likelihood of a crash event is not straightforward. EVT enables the extrapolation of rarely observed events (i.e., crashes) from 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 (or minima) using Generalized Extreme Value distribution and the Peak Over Threshold using Generalized Pareto distribution. BM approach divides the sample time into continuous blocks of fixed size and then considers the maximum value (or r largest values) from each time block as an extreme (**Figure 3.12** (a)). POT picks the highest values as extremes as long as the values are over a certain threshold (**Figure 3.12** (b)). Since there is no consensus among the researchers regarding using a particular method, the present study estimated the crash risk using both BM & POT approaches and compared their results.



(a) (b)

Figure 3.12 Simplified illustration of the sampling of conflict extremes using (a) BM method and (b) POT approach

3.2.5 Generalized Extreme Value Distribution Model

The BM method assumes that the observed maxima (or negative of minima) in different time blocks are statistically independent of one another and these maxima follow the generalized extreme value distribution. Let $X_1, X_2, ..., X_n$ are independently and identically distributed random variables with unknown cumulative distribution function and $M_n = \max\{X_1, X_2, ..., X_n\}$ represents the block maxima. If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that $Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} \to G(z)\, as\, n \to \infty$ for a non-degenerate distribution function G (Coles, 2001), then G is a member of the GEV family shown in Equation (3.3).

$$G(z) = exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi \neq 0 \tag{3.3}$$

defined on $\left\{z : 1+\xi\left(\left(z-\mu\right)/\sigma\right) > 0\right\}$, where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter, and $-\infty < \xi < \infty$ is the shape parameter. The parameters of the model were estimated from the observed data using the maximum likelihood estimation (MLE) approach.

The crash risk is defined as the probability of occurring a crash and a crash occurs when the ACT_{min} is equal to zero. Hence, the safety metric crash risk (R) is calculated based on the ACT_{min} values obtained from the time blocks of a fixed size. Using the fitted GEV distribution based on the sampled extremes, the risk of crash can be calculated as

$$R = Pr\left(Z \ge 0\right) = 1 - G\left(0\right) = \left(1 - \xi\frac{\mu}{\sigma}\right)^{-\frac{1}{\xi}} \tag{3.4}$$

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_{min} values, and $G(\bullet)$ is the generalized extreme value distribution.

3.2.6 Generalized Pareto Distribution 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:

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, y > 0 \tag{3.5}$$

Notably, the limiting distributions of **Equation** (3.5) are the generalized Pareto distribution, for a sufficiently large enough threshold u. For a large enough threshold u, the limiting distribution function of exceedances $y = X - u$, conditional on $X > u$ is as follows:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{3.6}$$

defined on $\left\{y : y > 0 \text{ and } \left(1 + \xi y / \tilde{\sigma}\right) > 0\right\}$, where $\tilde{\sigma} = \sigma + \xi(u - \mu)$ is the scale parameter, $-\infty < \xi < \infty$ is the shape parameter. The family of distributions defined by **Equation** (3.6) is the generalized Pareto family.

3.2.7 Threshold Selection

In the POT approach, the distribution of threshold exceedances shown in Equation (3.6) 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_{min} 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).

$$\mathrm{AIC}(p) = 2 \times n_p - 2 \times \mathrm{LL} \tag{3.7}$$

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

3.2.8 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 (**Figure 3.13**). Here the minimum ACT 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:

$$R = \Pr(Z \geq 0) = 1 - \Pr(Z \leq 0) = \left(1 - \xi \frac{u}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{3.8}$$

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:

$$N = \frac{T}{t} R \tag{3.9}$$

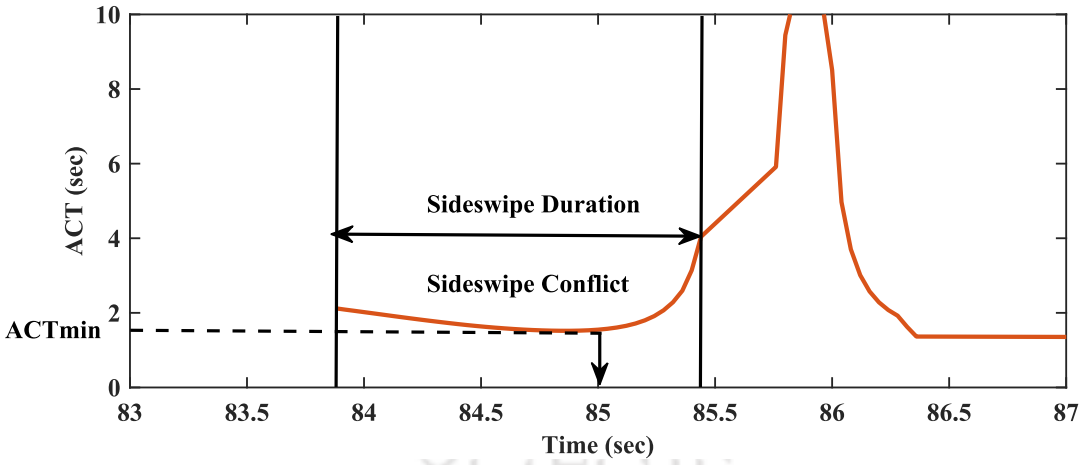


Figure 3.13 Sideswipe Conflict Identification from the ACT profile of a PTW Trajectory extracted using SAVETRAX

However, the shape parameter must be kept $\xi > -0.5$ to ensure the regular asymptotic properties of the maximum likelihood estimators (Smith, 1985). Further, the study measures the precision of the crash risk in terms of confidence bounds. The 95% confidence bounds of the crash risk were calculated assuming that the standard errors are normally distributed for the maximum likelihood

estimator. Lower the upper bound of the crash frequency estimate, more precise the model (Hauer & Garder, 1986). Based on these two criteria, the GEV & GPD models were compared to check which provides better crash risk results.

3.2.9 Incorporating the Effect of Heterogeneity in GEV Model

Base GEV models are developed for the crash risk estimation by assuming that ACT_{min} values (i.e., extremes) are independent and identically distributed. However, traffic conflict extremes (i.e. ACT_{min} values) can be heterogeneous in nature because some variable might influence the extreme behavior. For the BM method, the parameter estimation procedure can automatically capture the heterogeneity by considering the influencing variable, also termed covariate (Coles, 2001). This is accomplished by including the covariates into the location parameter of the GEV model with the identity link function (Songchitruksa & Tarko, 2006; Zheng et al., 2014a; Zheng, Sayed, et al., 2018) as shown in the **Equation** (3.10).

$$\mu = \mu_0 + \mu_1 \times x \tag{3.10}$$

where, μ is the location parameter of the GEV distribution, μ_0 and μ_1 are the parameters and x is the covariate. The covariate x will be calculated by taking the mean of the values corresponding to the selected block interval.

In the current study, two models were developed: the GEV model without considering the effect of covariate (basic model) and the second model considering its effect on the crash risk. The goodness of fit of the second model is calculated by measuring the reduction in the negative log-likelihood (LL) value from the basic model. This actually tells us the significance of the added covariate in the model. If the location parameter corresponding to the covariate has a positive sign then it indicates that the covariate positively influences the crash risk. The parameters of the model were estimated from the observed data using the MLE approach. The advantage of using MLE is that the shape parameter must be kept $\xi > -0.5$ to ensure the regular asymptotic properties of the maximum likelihood estimators (Smith, 1985). Hence, the estimated shape parameters act as a solid tool to compare various competing models with one another (Zheng et al., 2014a).

3.2.10 Incorporating the Effect of Heterogeneity in GPD Model

As discussed 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 (3.11)). The scale parameter σ of the GPD distribution can be estimated as follows:

$$\sigma = \beta_0 + \beta_i \times x_i \tag{3.11}$$

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).

Chapter 4 Conflict Assessment of Mixed Traffic Using ACT

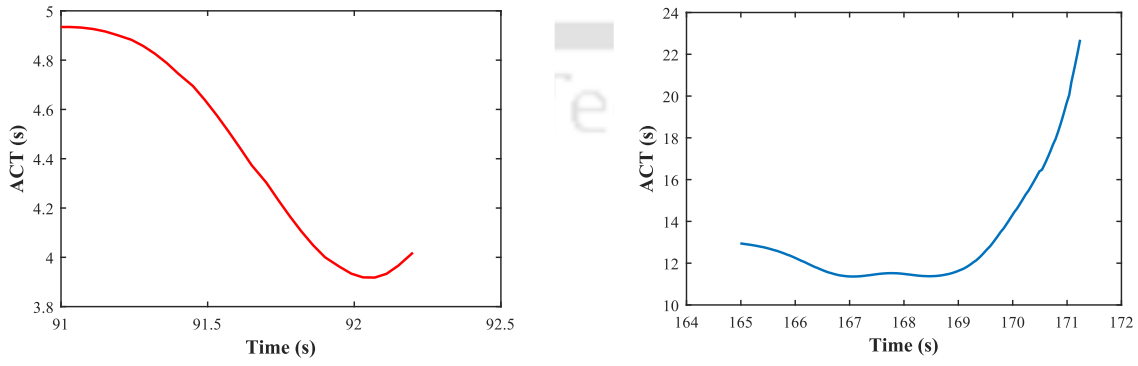
On rural highways, road users experience different conflicting situations in terms of rear-end, sideswipe, angled, head-on, and run-off road. One of the major advantages of ACT is that it can capture multiple conflict types experienced by a road user with the surrounding traffic. This becomes extremely relevant for the mixed traffic conditions prevalent in LMICs. The present chapter deals with the conflict analysis of different vehicle types present on the rural highways in India. Section 4.1 describes the conflict analysis of mixed traffic on multilane highways and Section 4.2 explains the conflict analysis on horizontal curves passing through mountainous terrains.

4.1 Conflict Analysis on Multilane Rural Highways

The present section deals with the rear-end, sideswipe conflict analysis of mixed traffic on multilane rural highways. First, the conflicts were identified based on the methodology described earlier. Then, the crash potential of such conflicts was assessed based on the magnitude of ACT_{min} values.

4.1.1 Rear-end Conflict analysis of Mixed Traffic

Figure 4.1 shows an example of ACT profiles of rear-end conflict for the mixed traffic moving on the multilane rural highways. In the case of **Figure 4.1**, the magnitude of ACT_{min} values differs depending on the vehicle type. The present study identifies such ACT_{min} values and then performs frequency analysis across specific ranges of ACT_{min} to understand the potential of rear-end conflict.



(a) (b)

Figure 4.1 ACT profile of Rear-end Conflict Type of (a) PC and (b) PTW on multilane rural highways

Figure 4.2 represents the frequency distribution of potential conflicts across different values of ACT_{min} corresponding to multiple vehicle types present in the mixed traffic stream on the four-lane rural highway. Compared to the smaller ACT_{min} values, the frequency of potential conflicts is significantly higher for the larger ACT_{min} (i.e., > 5 s). Since larger ACT_{min} values indicate a lower crash potential or risk for a specific conflict type, this shows that potent or risky rear-end conflicts are not frequently happening on such highways

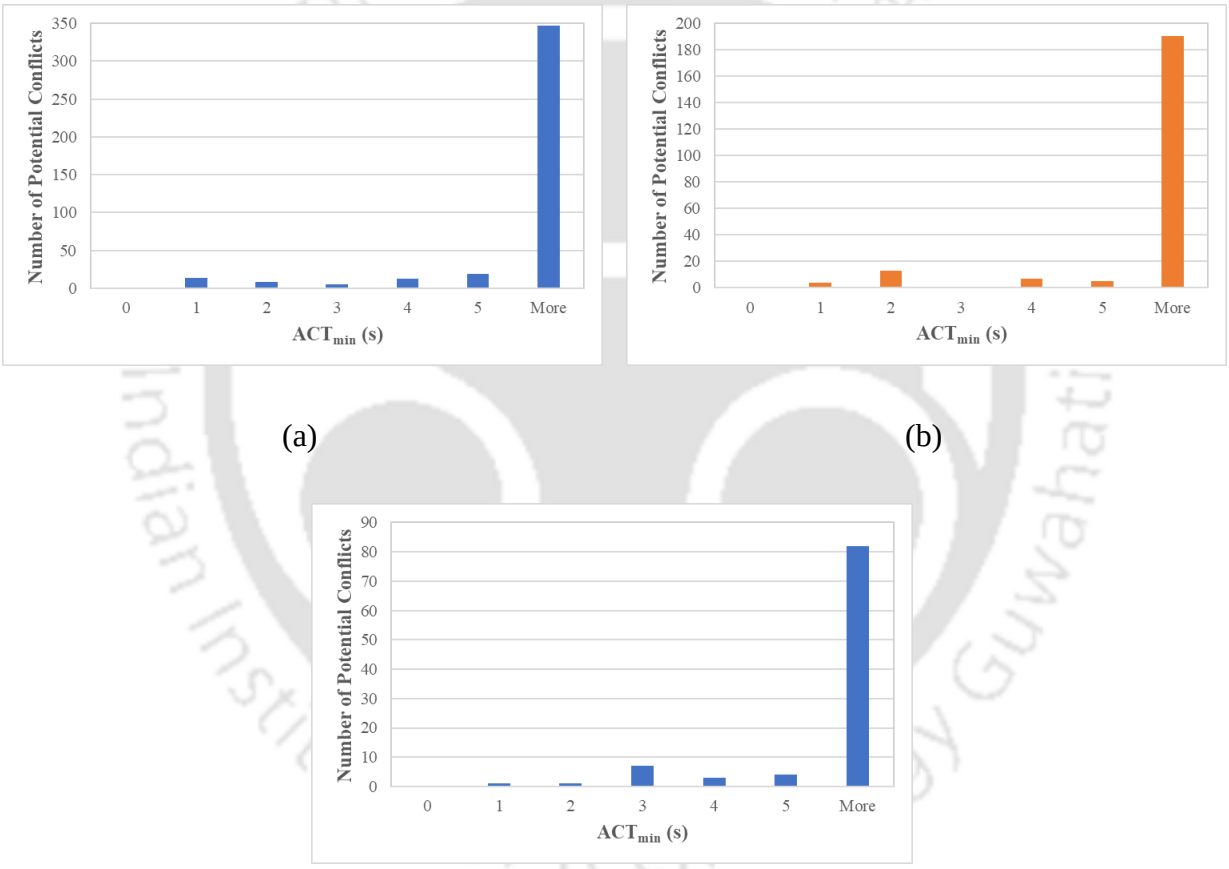


Figure 4.2 Number of Potential Conflicts of (a) PC (b) PTW and (c) HCV on the four-lane rural highway

Figure 4.3 shows the frequency of potential conflicts across different values of ACT_{min} corresponding to multiple vehicle types present in the mixed traffic stream on the six-lane rural

highway. Similar observations were also made for the six-lane rural highway as that of the four-lane highway.

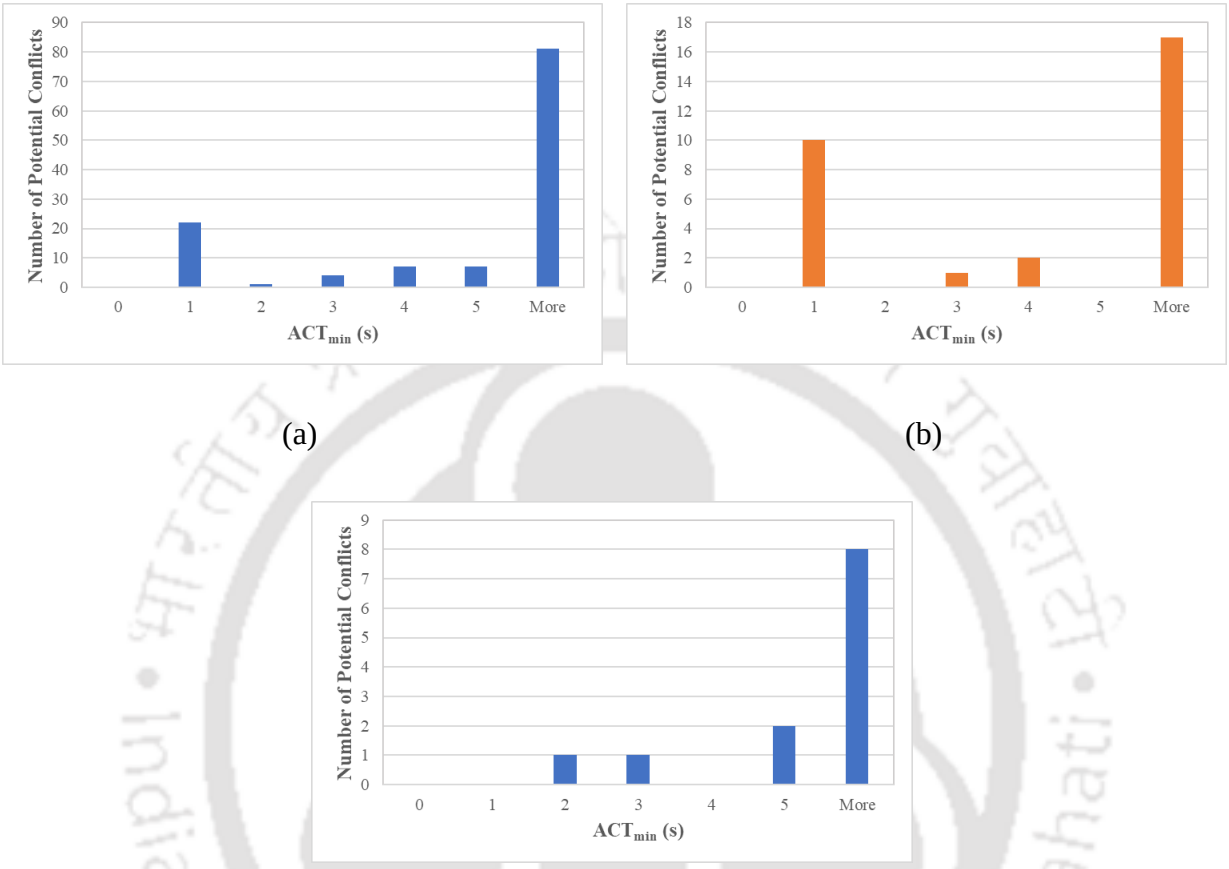
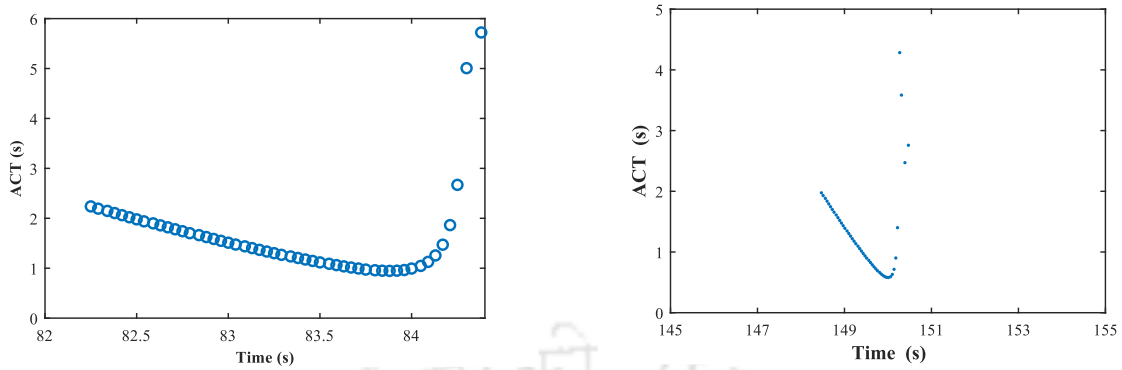


Figure 4.3 Number of Potential Conflicts of (a) PC, (b) PTW, and (c) HCV on the six-lane rural highway

4.1.2 Sideswipe Conflict analysis of Mixed Traffic

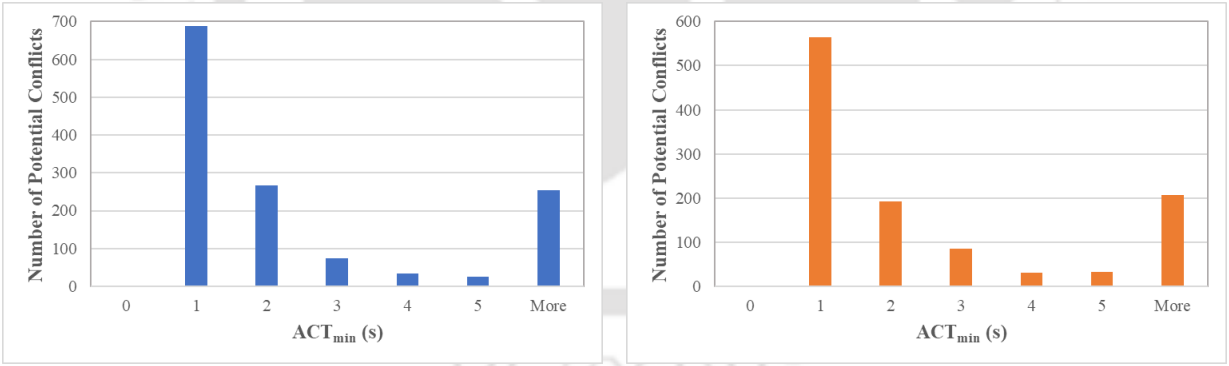
Figure 4.4 shows an example of ACT profiles of sideswipe conflict for mixed traffic moving on multilane rural highways. The ACT profiles in the figure correspond to lower ACT_{min} values (i.e., < 1s) for both PC and PTW vehicle types. Similar to the rear-end conflict, the present study identifies such ACT_{min} values and then performs frequency analysis across specific ranges of ACT_{min} for the sideswipe conflict.



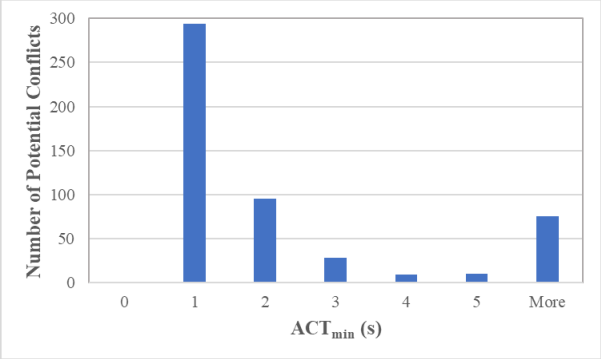
(a) (b)

Figure 4.4 ACT profiles of Sideswipe Conflict for (a) PC and (b) PTW vehicle types on multilane rural highways

Figure 4.5 demonstrates the frequency of potential conflicts across different values of ACT_{min} corresponding to multiple vehicle types present in the mixed traffic stream on the four-lane rural highway. The maximum number of potential sideswipe conflicts corresponds to the ACT_{min} values of less than 1s for all three vehicle types. Since smaller ACT_{min} values indicate a higher crash risk, one can say that risky sideswipe conflicts are more frequently happening on such highways.



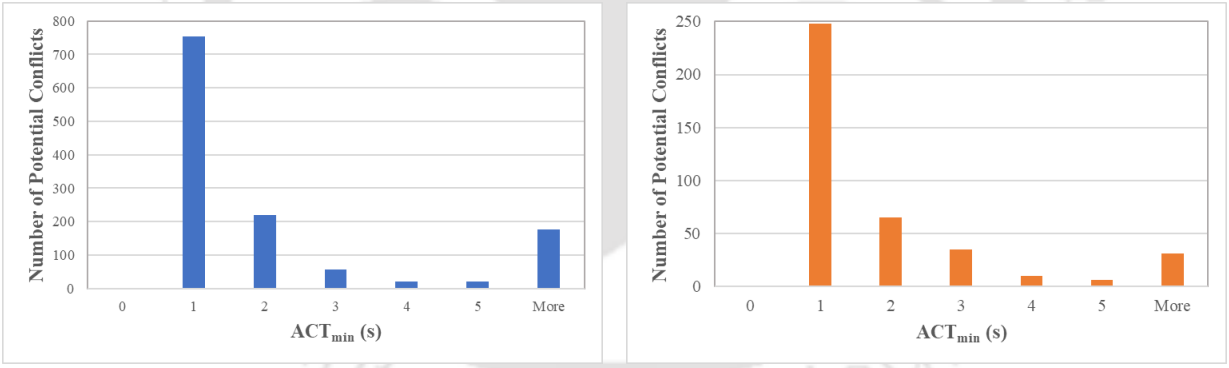
(a) (b)



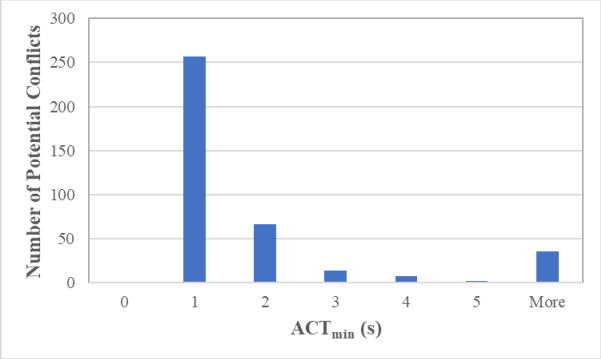
(c)

Figure 4.5 Number of Potential Conflicts of (a) PC (b) PTW and (c) HCV on the four-lane rural highway

Figure 4.6 shows the frequency of potential conflicts across different values of ACT_{min} corresponding to multiple vehicle types present in the mixed traffic stream on the six-lane rural highway. Similar observations were also made for the six-lane rural highway as that of the four-lane highway.



(a) (b)

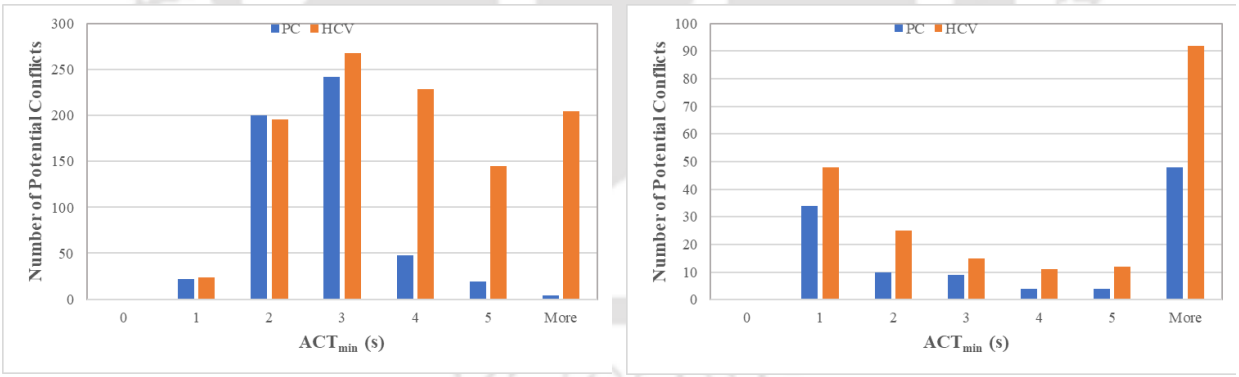


(c)

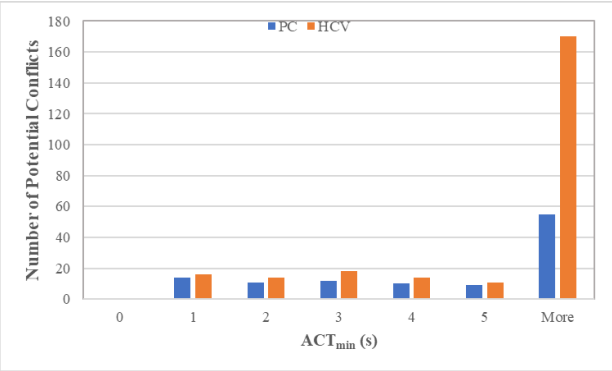
Figure 4.6 Number of Potential Conflicts of (a) PC, (b) PTW, and (c) HCV on the six-lane rural highway

4.2 Conflict Analysis on Horizontal Curves

Various types of conflicts occur on horizontal curves, including single-vehicle (ROR) and multi-vehicle (rear-end, sideswipe, and head-on) conflicts. The study has identified each conflict type prevalent on the mountainous terrain utilizing the previously outlined methodology. Subsequently, the severity of each traffic conflict was evaluated based on the magnitude of the ACT_{min} value.



(a) (b)



(c)

Figure 4.7 Number of Potential Conflicts of PC and HCV for (a) ROR (b) Sideswipe and (c) Rear-end types on the horizontal curves

Figure 4.7 (a) shows that the ROR conflict type happens more frequently on the horizontal curves passing through mountainous terrains. Compared to the ROR type, the sideswipe and rear-end cases are relatively less on the curved sections (**Figure 4.7** (b-c)). Lower sideswipe and rear-end cases could be because of the low volume traffic conditions and hence vehicular interactions are very less. Due to the mountainous terrain, ROR conflicts are happening at relatively higher frequency. For the ROR conflict type, lower ACT_{min} values (i.e., < 1s) are less, which indicates that drivers are experiencing less severe conflicts on the curves. Hence, the ROR crash risk might be lower on these curves even though it is happening more frequently than that of sideswipe and rear-end types.

The present study collected the historical crash data on these curves. The data corresponds to the period between December 2013 to March 2021. The crash records include the study site, crash type, subject vehicle, conflicting vehicle, injury type, number of crashes, and time of the day. To maintain the temporal consistency between conflict and crash data, the present study separated the day-time crash data from that of the night-time. **Figure 4.8** shows that during the day-time, HCVs mostly involve in ROR crashes. The multi-vehicle crashes such as sideswipe, rear-end, and head-on types are significantly lower than that of ROR type. Further, the day-time crash data is insufficient to perform a crash risk modeling of a specific vehicle type on each curve.

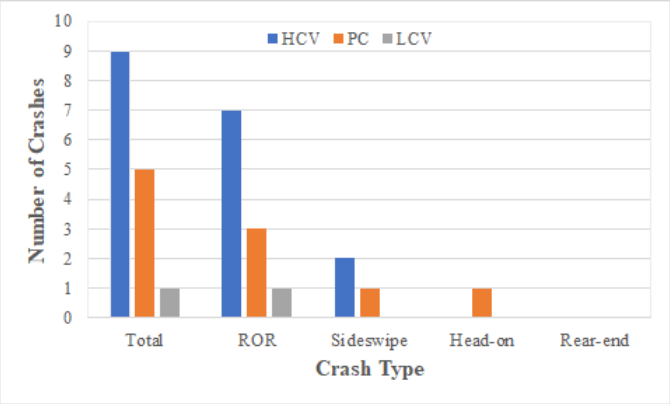


Figure 4.8 Vehicle-Type Specific Day-time Crashes on the Nine Curves, between 2013 to 2021

4.3 Summary

The present chapter analyzes the different conflict types of the mixed traffic prevalent on the rural highways passing through the flat and mountainous terrains in India. This analysis was carried out with the help of the multi-dimensional SSM called ACT.

For the multilane rural highways passing through flat terrains, it was found that the risky rear-end conflicts are significantly lower than that of sideswipe conflicts for all the three vehicle types present in the mixed traffic. This could be because sideswipe conflicts happen more frequently under low-volume conditions due to the frequent passing and lane-changing opportunities available for drivers. When the volume is higher, vehicles tend to undergo car-following behavior more regularly, giving rise to more rear-end conflicts. Hence, for the low-volume conditions prevalent on the multilane highways in India, the crash risk assessment of sideswipe conflict type requires further attention.

For the rural highways passing through mountainous terrains in India, both the rear-end and sideswipe conflict types are happening less frequently due to the low volume traffic conditions. On such terrains, ROR conflict type i.e., single-vehicle conflict happens more frequently than that of multi-vehicle conflicts. Further, for the ROR conflict type, the risky situations with high crash likelihood are also very less on such terrains for the different vehicle classes.

Chapter 5 EVT Modelling of Mixed Traffic in Rural Highways

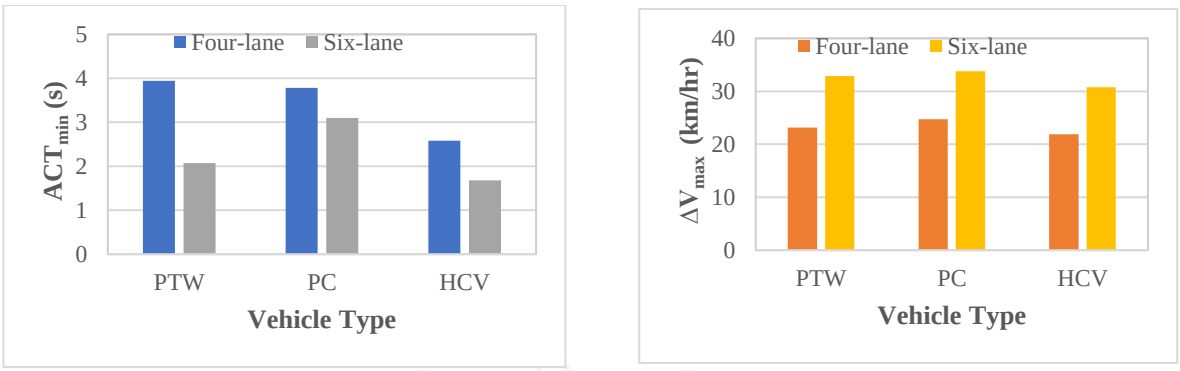
The crash risk assessment of mixed traffic has been understudied due to a lack of pertinent data. In recent years, proactive methods have gained significant attention in transportation safety analysis due to their numerous advantages. This chapter presents modeling and evaluating the crash risk of mixed traffic on rural highways using a novel, multidimensional proactive safety indicator called Anticipated Collision Time (ACT). For analysis, detailed trajectory data were collected using an unmanned aerial vehicle from midblock sections of rural highways. A safety metric called Crash Risk was derived from the observed conflict risk, which is further used to assess the system's safety performance under study. Using Extreme Value Theory (EVT), the conflict risk was mapped to crash risk. The extreme events were identified using the BM and POT approaches. Later, the GEV and GPD models were developed for each location separately by extracting the conflicts from the vehicle trajectories. Section 5.1 describes the EVT modelling of sideswipe conflict type prevalent on multilane rural highways. Section 5.2 explains the results of ROR conflict type prevalent on horizontal curves of a highway passing through mountainous terrain. Section 5.3 provides the summary of the crash risk analysis of mixed traffic for both multilane highways and horizontal curves.

5.1 EVT Modelling of Sideswipe Conflict Type

The present section starts with the preliminary analysis of ACT_{min} and then describes the crash risk results of BM and POT approaches. It ends with discussing the comparison between BM and POT models.

5.1.1 Preliminary Analysis of ACT_{min}

Figure 5.1 shows the mean values of safety margin (i.e., ACT_{min}) and maximum speed difference (i.e., ΔV_{max}) during sideswipe conflict, for each vehicle type on the four-lane and six-lane rural highways. Compared to the four-lane highway, drivers on a six-lane highway maintains a smaller safety margin on average during sideswipe conflict (Figure 5.1 (a)). Further, the vehicles on six-lane highway maintains a significantly higher maximum speed difference as compared to the four-lane highway during sideswipe conflict (Figure 5.1 (b)).



(a) (b)

Figure 5.1 Mean Values of ACT_{min} during sideswipe conflict for the four-lane and six-lane rural highways

Table 5.1 shows the descriptive statistics of the safety margin during sideswipe conflict for each vehicle type on the four-lane and six-lane rural highways. On average, PTWs maintain larger safety margin on the four-lane highway with the mean value of 4.0 seconds. Whereas PTWs on six-lane highway maintain a smaller safety margin of 2.0 seconds. For the PC, the safety margin reduces by 0.5 seconds on the six-lane highway from the four-lane highway. Similarly, for the HCV, the safety margin reduces by 0.9 seconds compared to the four-lane highway indicating the fact that PTWs suffer the most reduction in the safety margin on the wider roads. This might lead to the higher sideswipe crash risk of PTWs on the six-lane highway as compared to the four-lane highway. This difference has been tested statistically using a t-test at 95% confidence ($p = 0.012$, $p = 6.5 \times 10^{-6}$, and $p = 0.001$, for PC, PTW, and HCV, respectively). This could be happening because the drivers on a six-lane highway feel a false sense of safety due to the additional lane available to evade any risky situations they might encounter. Consequently, such perception makes them over-aggressive, and they maintain smaller gaps while passing and changing lanes.

Table 5.1 Descriptive Statistics of ACT_{min} during Sideswipe Conflict

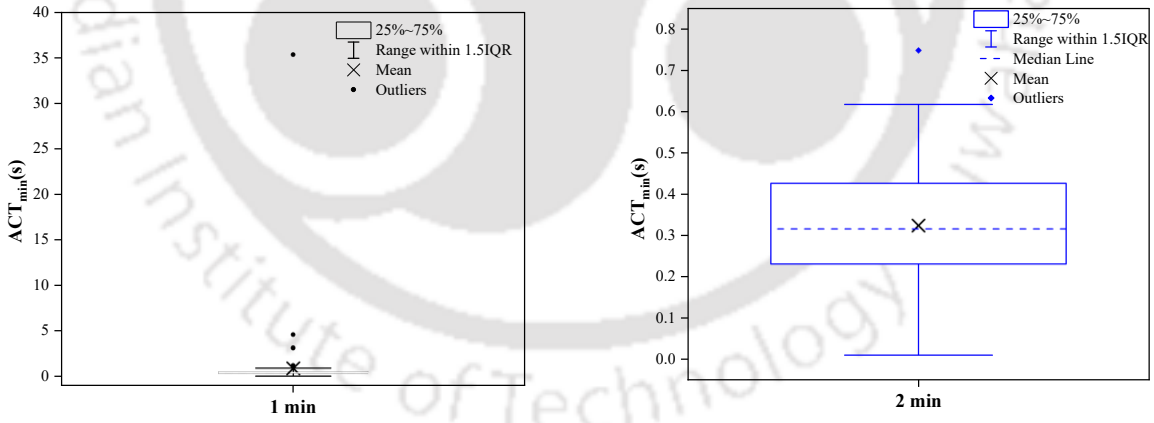
Study Sites	Variable	Potential Conflicts			Mean			SD			Min			Max			Correlation		
		PTW	PC	HCV	PTW	PC	HCV	PTW	PC	HCV	PTW	PC	HCV	PTW	PC	HCV	PTW	PC	HCV
Four-lane	ACT_{min} (sec)	1113	1340	511	3.94	3.78	2.58	7.73	6.93	4.67	0.002	0.01	0.02	47.77	47.77	46.39	0.46	0.48	0.47
	ΔV_{max} (km/hr)				23.17	24.72	21.88	18.94	19.17	15.26	0.02	0.14	0.07	88.46	103	85.33			
Six-lane	ACT_{min} (sec)	395	1248	381	2.07	3.10	1.68	4.66	6.76	3.26	0.008	0.01	0.02	43.73	48.97	32.45	0.38	0.46	0.45
	ΔV_{max} (km/hr)				32.89	33.81	30.79	25.25	24.67	20.24	0.03	0.006	0.07	106.52	107.33	100.70			

Note: SD, Min, and Max denote standard deviation, minimum, and maximum values of the variables. Correlation means the linear association between negative ACT_{min} values and ΔV_{max}.

The correlation test was conducted to evaluate the strength of the association between the safety margin and maximum speed difference. Correlation values (**Table 5.1**) indicate that there is a moderate linear relationship between the safety margin and maximum speed difference during sideswipe conflict. Further, the positive sign indicates that as the maximum speed difference increases, the safety margin decreases during sideswipe conflict. However, Table 5.1 further demonstrated that the potential conflicts contain wide range of values for the safety margin. Hence, to identify severe conflicts, the present study uses the previously described BM and POT approaches of extreme value theory.

5.1.2 Block Interval Selection

The present study considered blocks of size such as 1 min, 2 min, 3min, and 4 min for the sampling of extremes. Figure 5.2 shows the variation of ACT_{min} extremes of PC corresponding to the different block sizes. Figure 5.2 (a) shows that 1 min block interval contains ACT_{min} extremes ranging from 0.01s to 35s. Large ACT_{min} extremes represent normal interactions instead of severe conflicts (Tarko, 2018). Such values arise because that might be the only observation available in the case of smaller block intervals (i.e., 1 min in this case). This can be addressed by taking slightly larger i.e., 2 min block interval as shown in Figure 5.2 (b).

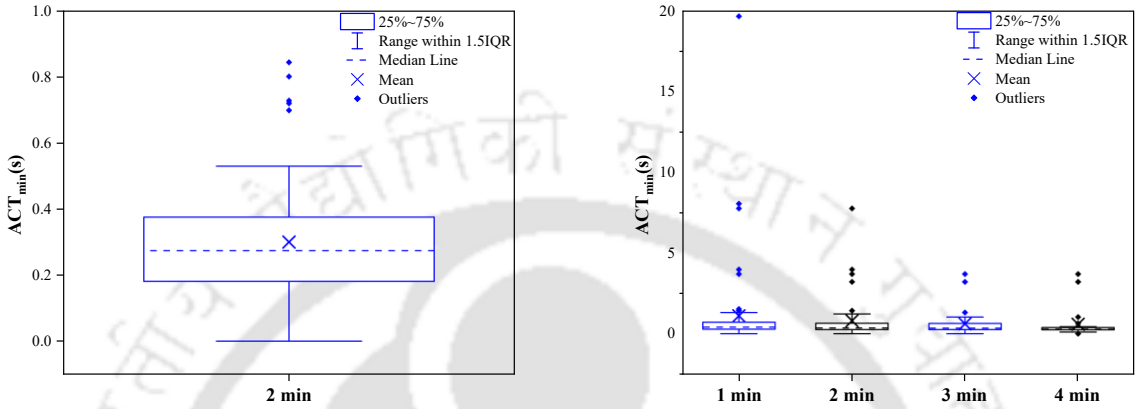


(a) (b)

Figure 5.2 Sampled Extremes for the Sideswipe Conflict of PC corresponding to (a) 1 min and (b) 2 min block intervals

Similarly, the present study has chosen the block interval of 2 min for the PTW, since it mainly contains smaller ACT_{min} values representing traffic conflict extremes (Figure 5.3 (a)). For the

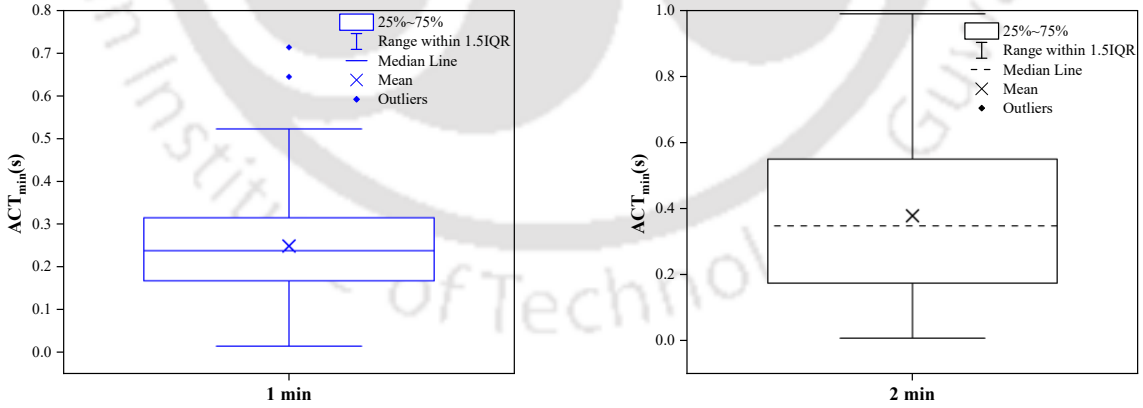
sideswipe conflict of HCV, appropriate sample of extremes cannot be obtained for the chosen block intervals. Figure 5.3 (b) shows that a few relatively larger ACT_{min} values exist even with the increase in the block interval. Therefore, the BM modelling of sideswipe crash risk of HCV cannot be carried out for the given dataset.



(a) (b)

Figure 5.3 Sampled Extremes for the Sideswipe Conflict of (a) PTW and (b) HCV

Similarly, for the PC and PTW moving on the six-lane highway, the block intervals of 1 min and 2 min, respectively, contain the severe sideswipe conflicts (Figure 5.4).



(a) (b)

Figure 5.4 Sampled Extremes for the Sideswipe Conflict of (a) PC and (b) PTW on six-lane highway

5.1.3 Crash Risk Results Using GEV Models

Table 5.2 shows the statistics corresponding to different GEV models fitted to the block maxima data of PC and PTW moving on the four-lane and six-lane rural highways. For PTWs, ξ <- *0.5*, violating the asymptotic properties of MLE (Smith, 1985). Further, the uncertainty associated with the crash risk estimates of PTW is significantly high.

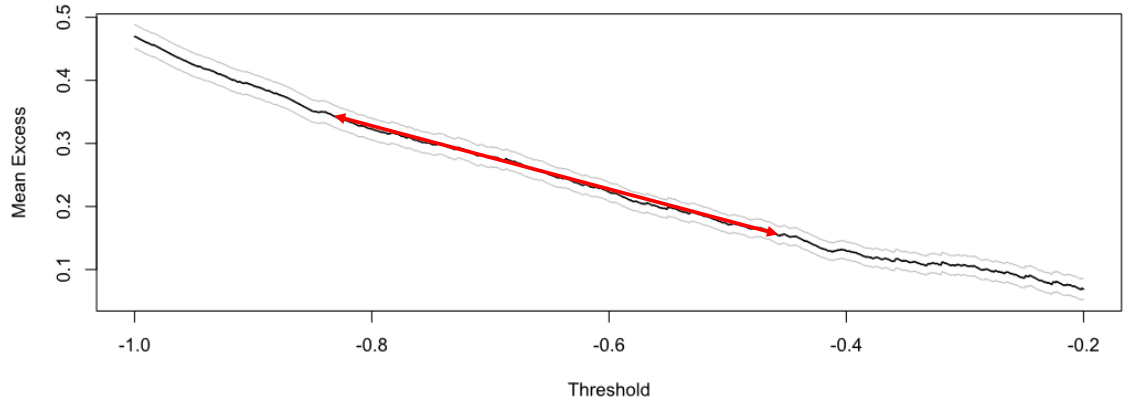
Table 5.2 Fitted Statistics and Parameters of the GEV Model for PC and PTW

Parameters	**GEV Models**			
Vehicle Type	PC		PTW	
Study Site	Four-lane	Six-lane	Four-lane	Six-lane
Sample Size	57	70	57	36
μ (SE)	-0.37 (0.023)	-0.28 (0.017)	-0.34 (0.030)	-0.41 (0.067)
σ (SE)	0.16 (0.016)	0.13 (0.012)	0.21 (0.024)	0.32 (0.075)
ξ (SE)	-0.35* (0.085)	-0.45* (0.06)	-0.59** (0.091)	-0.76** (0.315)
Log-likelihood (LL)	-26.95	-49.56	-18.44	0.01
AIC	-47.90	-93.12	-30.88	6.03
Crash Risk	0.63%	0.13%	0.52%	0.57%
Estimated Annual Crashes	11	4	9	17
Upper 95% CI	204	297	366	1216
Lower 95% CI	0	0	0	0

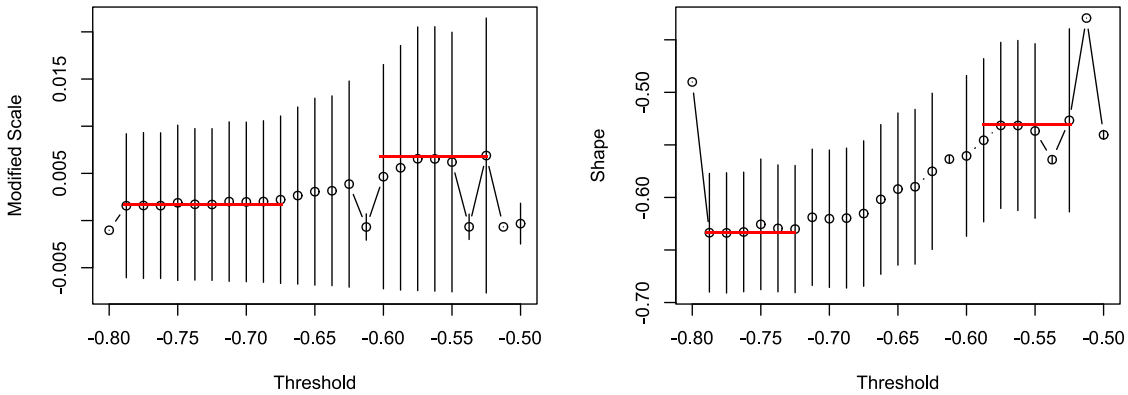
Note: SE indicates the standard error estimates of model parameters. * indicates the model that fulfils the usual asymptotic properties whereas ** indicates that of violating the asymptotic properties. Sample size indicates the number of severe conflicts (i.e., extremes).

5.1.4 Threshold Selection

As stated earlier, subjectivity is involved in selecting the ACT_{min} threshold using the MRLP and TSP methods. Figure 5.5 (a) shows that the MRLP with 95% confidence bounds is approximately linear for the threshold range of (-0.8, -0.5). TSP (**Figure 5.5** (b)) 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).



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



(b) Threshold Stability Plot

Figure 5.5 Graphical Plots of the 4-lane highway for Threshold Selection of PTW

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. Figure 5.6 depicts the variation of AIC values with the change in ACT_{min} threshold for both four-lane and six-lane highways. The figure clearly shows that as the threshold decreases (i.e., ACT_{min} 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 seconds, respectively.

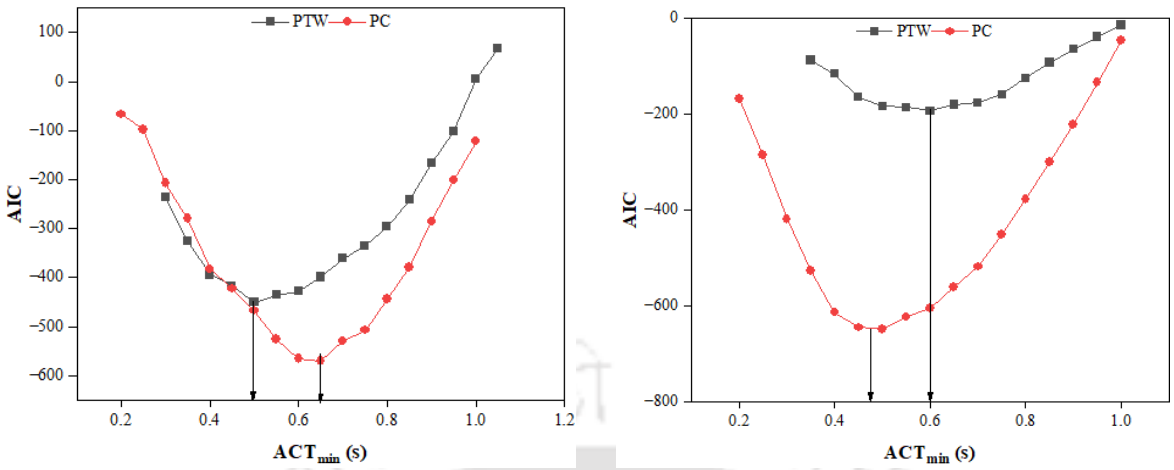


Figure 5.6 Effect of ACT_{min} threshold on the goodness of fit of GPD model for PTW and PC on (a) four-lane highway and (b) six-lane highway

Using the similar procedure, the ACT_{min} threshold of PC came out to be 0.65s and 0.45s respectively for the four-lane and six-lane highways.

5.1.5 Crash Risk Results Using GPD Models

Based on the threshold, normal interactions are separated from the severe conflicts. From **Table 5.3**, it can be seen that, for the four-lane highway, out of 1113 sideswipe PTW conflicts, 266 conflicts are severe in nature. Similarly, for the six-lane highway, there are 147 severe conflicts out of 395 PTW conflicts. This means that, proportion-wise there are more severe PTW conflicts relative to the total number of interactions on the six-lane highway than that on the four-lane highway. However, for the PC, the fraction of severe conflicts relative to the total conflicts is higher in the case of four-lane highway than that of six-lane highway. **Table 5.3** 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 estimate 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. However, the crash risk associated with the PC came out to be 0.02% for both four-lane and six-lane highways. This indicates that PTWs are experiencing significantly higher sideswipe crash risk than that of PCs on multilane rural highways.

Table 5.3 Parameters and Statistics of the GPD Model

Parameters	**GPD Models**			
Vehicle Type	PC		PTW	
Study Site	Four-lane	Six-lane	Four-lane	Six-lane
Sample Size	1340	1248	1113	395
Threshold	0.65	0.45	0.5	0.6
Exceedances	425	324	266	147
σ (SE)	0.29 (0.016)	0.22 (0.014)	0.25 (0.009)	0.28 (0.013)
ξ (SE)	**-0.43*** (0.034)	**-0.47*** (0.04)	**-0.48*** (0.115)	**-0.40*** (0.183)
Log-likelihood (LL)	-287.08	-324.55	-228	-98
AIC	-570.16	-645.11	-450	-191
Crash Risk	0.02%	0.02%	0.13%	0.74%
Estimated Annual Crashes	1	1	2	22
Upper 95% CI	43	110	**96**	**287**
Lower 95% CI	0	0	0	0

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

5.1.6 Comparison of the Crash Risk Results of BM and POT Models

The present study analyzes the crash risk of the mixed traffic using both BM and POT models. Compared to the POT approach, BM approach estimates the crash risk of mixed traffic with wider confidence bounds due to the limited sample of extremes. Since wider confidence bounds indicate less precise crash risk estimates, POT models outperform BM models for the mixed traffic streams. For the conflict information obtained under the homogeneous traffic conditions, studies (Ali et al., 2022; Zheng et al., 2014a) have already found that POT models outperform BM models. The similar finding from the present study under the mixed traffic conditions further strengthen this concept.

5.2 EVT Modelling of ROR Conflict Type

The present section discusses the crash risk of results of ROR conflict type using GPD models. It provides the crash risk analysis of nine horizontal curves passing through the mountainous terrain of Guwahati-Shillong bypass in India.

5.2.1 GPD Model Results for ROR Conflict Type on Horizontal Curves

Table 5.4 provides the crash risk estimates of PC and HCV associated with ROR conflict type for the horizontal curves passing through mountainous terrain. Out of the nine curves, HCVs experience risky situations on five curves. The ACT_{min} thresholds vary from 2.5 seconds to 5.0 seconds. This has contributed to the significant ROR crash risk of HCVs. However, the precision of the crash risk estimates is lower due to the smaller sample sizes.

Table 5.4 ROR Crash Risk Results of Horizontal Curves Using GPD Models

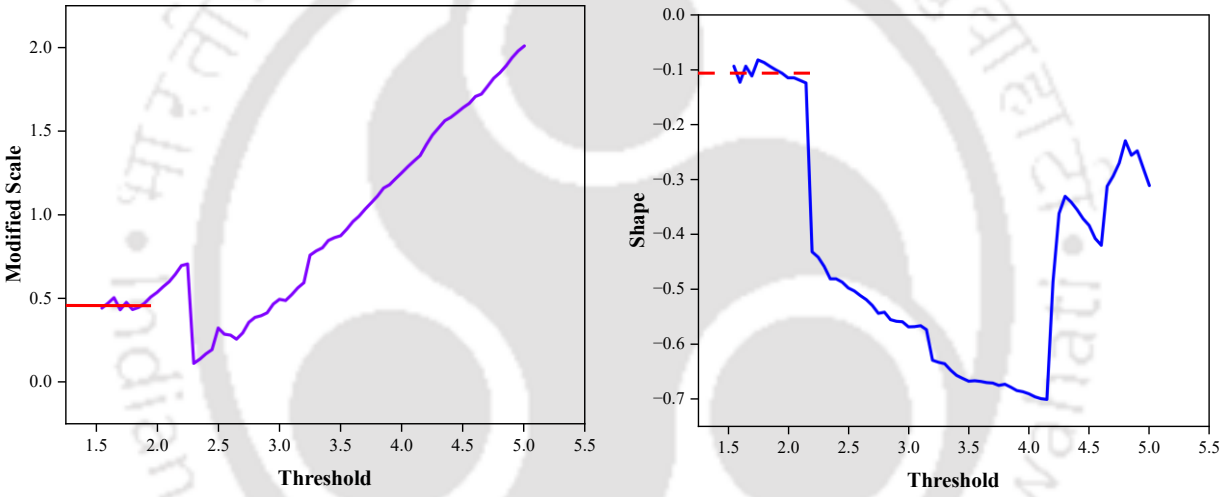
Curve No	Sample Size		ACT_{min} Threshold		Exceedances		Crash Risk		95% Confidence Bounds	
	PC	HCV	PC	HCV	PC	HCV	PC	HCV	PC	HCV
C127	65	110	U*	U*	U*	U*	U*	U*	--	--
C171	44	117	2.6	3.0	37	55	0.29%	**0.17%**	**[0 640]**	**[0 812]**
C179	40	64	U*	U*	U*	U*	U*	U*	--	--
C183	34	112	U*	2.5	U*	55	U*	**0.11%**	--	**[0 152]**
C189	62	121	U*	U*	U*	U*	U*	U*	--	--
C221	70	98	U*	5.0	U*	96	U*	**0.09%**	--	**[0 495]**
C227	69	136	U*	4.0	U*	105	U*	**0.2%**	--	**[0 746]**
C238	66	145	U*	5.0	U*	105	U*	**0.01%**	--	**[0 425]**
C245	85	164	U*	U*	U*	U*	U*	U*	--	--

Note: U* indicates the undefined results obtained due to the absence of severe ROR conflicts in the sample size.

5.2.2 Stationary Modelling of ROR Crash Risk of HCV for Combined Data

For obtaining the crash risk, the study first determines the ACT_{min} threshold using stability plots and the best-fitted GPD model across various thresholds. Figure 5.7 (a) shows that the scale

parameter remains nearly constant for the ACT_{min} ranging between 1.5 s to 2.0 s. For the shape parameter, the same thing happens for the ACT_{min} ranging between 1.5 s to 2.2 s (Figure 5.7 (b)). Therefore, the threshold lies between 1.5 s to 2.0 s. Figure 5.8 demonstrates that the maximum log-likelihood value (LL) decreases with the increase in threshold until it reaches a minimum point and then increases with the increase in threshold. The lowest LL value indicates the best-fitted model and the corresponding threshold as the best threshold, which is 2 seconds for the ROR conflict. The parameters of the GPD model corresponding to the threshold of 2 seconds are $\sigma = 0.53$, and $\xi = -0.11$. The estimated crash risk, $R, is\ 0.76\%$, which is expected to produce 3 crashes per year. However, the uncertainty associated with the annual expected crashes lies between $[0, 6]$ indicating the variability in the crash frequency estimates.



(a) (b)

Figure 5.7 Threshold Stability Plots of Modified Scale and Shape Parameters

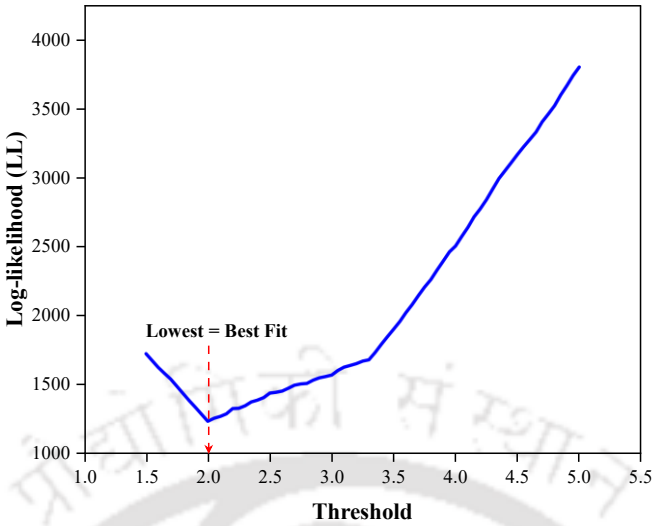


Figure 5.8 Goodness-of-fit of GPD models for Various Thresholds

Further, from the historical crash data, it was found that seven ROR crashes of HCV happened on the selected curves over eight years. Since the observed crash frequency is a point measure with no variance, Songchitruksa and Tarko (2006) estimated the Poisson confidence bounds for the observed crash frequency and validated the crash risk models. Then they compared it with the confidence bounds obtained from the crash risk estimate. The crash risk models can be considered accurate if the bounds are similar. Recently, this approach has been employed in several other studies to validate crash risk models (Arun et al., 2021; Wang et al., 2019; Zheng and Sayed, 2019a, 2019b). The Poisson confidence interval (λ) can be defined as:

$$\left\{\lambda: \frac{1}{2T}\chi^2_{2y_0,1-\frac{\alpha}{2}} \leq \lambda \leq \frac{1}{2T}\chi^2_{2(y_0+1),\frac{\alpha}{2}}\right\} \quad (1)$$

Where y_0 is the observed crash frequency in period T, λ is the estimated crash frequency for period T, and α is the significance level. Based on the relation, the 95% confidence bound for the observed crash frequency is $\frac{1}{16}\chi^2_{14,0.975} \leq \lambda \leq \frac{1}{16}\chi^2_{16,0.025}$, which lies between $[0.352, 1.803]$. Hence, the stationary crash risk model overestimates, with the confidence bound $[0, 6]$. To further improve the crash risk prediction, non-stationary models were developed with the geometric variables and the evasive actions as covariates.

5.3 Summary

The present study analyzed the crash risk associated with the mixed traffic prevalent on rural highways using GEV and GPD models. It was found that GPD models provide more precise estimates of the crash risk for mixed traffic. Further, based on the crash risk results of such models, it was found that PTWs are the vulnerable road users as compared to the other vehicle types prevalent on the multilane rural highways. Compared to the four-lane highways, PTWs are more prone to sideswipe crash risk on the six-lane highways (i.e., Crash Risk is 0.74%). It is because PTWs have higher maneuverability, which leads to more frequent lane changes or passing operations on such highways. Hence, they have been associated with the higher sideswipe crash risk. Vehicles maintain a smaller safety margin on the six-lane highway than on the four-lane highway. On a six-lane highway, situations like aggressive driving, smaller gaps, and higher speed differential may arise due to the false sense of safety that results from the more spatial freedom available to the road users. These situations can lead to more sideswipe crashes when road users make a trivial error while driving.

HCVs experience significant ROR crash risk on the horizontal curves passing through mountainous terrains in India. The ROR crash risk of HCVs varies between 0.01% to 0.17% across the different length of horizontal curves. For the combined data on curves, the ROR crash risk of HCVs came out to be 0.76%, indicating the risk associated with ROR conflict while driving during the observation period. Larger σ value $\sigma = 0.53$ indicate the wide spread of ROR conflicts across various severity levels rather than more conflicts in the high severity level with a narrow peak. Such spread of ROR traffic conflicts was also evident in the frequency distribution of ACT_{min} values (Figure 4.8(c)).

Chapter 6 Non-stationary EVT Modelling of Mixed Traffic in Rural Highways

Stationary Crash Risk models are developed for the safety assessment by assuming that ACT_{min} values (i.e., extremes) are independent and identically distributed. However, traffic conflict extremes (i.e. ACT_{min} values) can be heterogeneous in nature because some variables might influence the extreme behavior. To capture such heterogeneity the present study performs non-stationary EVT modelling for the mixed traffic moving on rural highways. Section 6.1 provides the non-stationary EVT modelling of sideswipe conflict using GPD models. Section 6.2 explains the non-stationary GPD modelling of ROR conflict. Section 6.3 summarizes the crash risk results of non-stationary GPD models for the sideswipe and ROR conflict types.

6.1 Non-stationary EVT Modelling of Sideswipe Conflict

This section presents a preliminary analysis of covariates influencing the extreme behavior of ACT_{min} values and then it describes the corresponding crash risk results of the non-stationary GPD models.

6.1.1 Preliminary analysis of Covariates Associated with the Evasive Actions of Mixed Traffic

For the mixed traffic conditions, Guo et al. (2018, 2019) found that variables such as peak yaw rate, deceleration rate during a traffic conflict better captures the conflict severity or crash risk. Venthuruthiyil and Chunchu (2022b) further stated that the time of evasive action is important in measuring the severity of traffic conflicts. However, the effect of these microscopic variables related to a traffic conflict has not been explored in the previous studies. Therefore, for the crash risk estimation, the present study has developed non-stationary EVT models considering the covariates shown in Table 6.1.

Table 6.1 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

From the reconstructed trajectories, 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. It measures the swerving maneuvers of vehicles on the road (Guo et al., 2018). **Figure 6.1** shows an example of the sideswipe conflict obtained from the ACT profile of a PTW and the 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 actions.

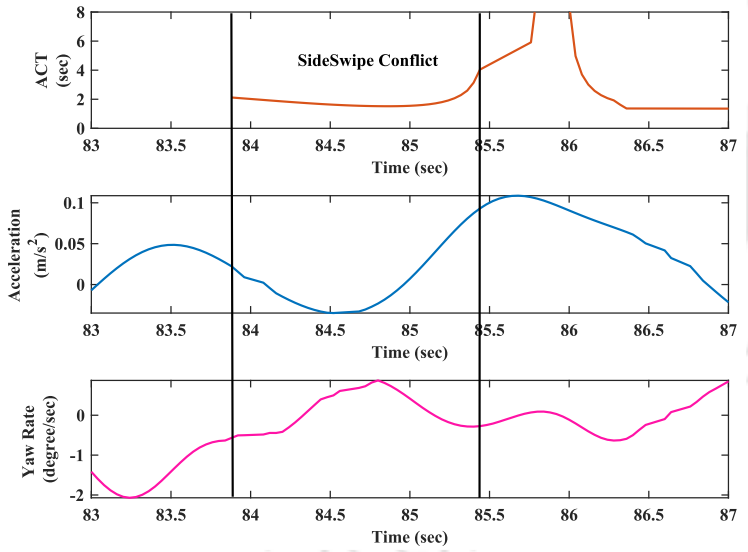
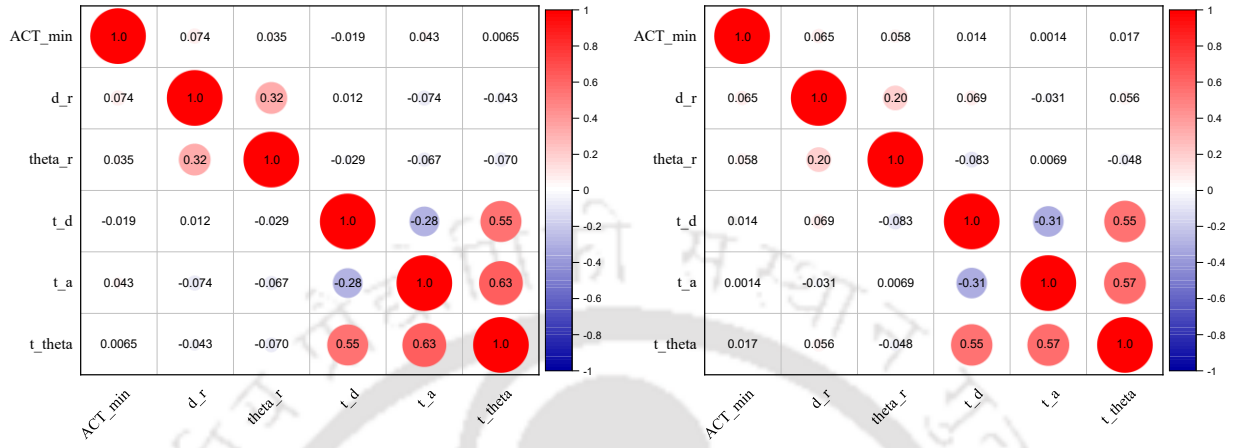


Figure 6.1 ACT, Acceleration/Deceleration, and Yaw Rate Profiles of a Sideswipe Conflict

Table 6.2 demonstrates that the yaw rate of mixed traffic 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, drivers in mixed traffic spend more time applying steering than the brakes during a sideswipe conflict on both highways.



(a) (b)

Figure 6.2 Correlation Plot of ACT_{min} and covariates for PTW on (a) four-lane and (b) six-lane highways

Figure 6.2 shows that there is no correlation between ACT_{min} and the covariates chosen in Table 6.1. However, there is a moderate correlation between deceleration rate and steering rate. This indicates that drivers in general are applying both braking and steering together while evading a sideswipe conflict.

Table 6.2 Statistics of Covariates Capturing the Evasive Actions

Study Sites	Variable	Mean			SD			Min			Max		
		PTW	PC	HCV	PTW	PC	HCV	PTW	PC	HCV	PTW	PC	HCV
Four-lane	d_r (m/s^2)	0.3	0.5	0.31	0.6	0.6	0.37	0.0	0.0	0.01	10.8	7.2	2.5
	θ_r *(degrees/s)*	0.9	1.9	1.99	2.1	3.71	2.47	0.0	0.0	0.0	39.7	68.29	22.46
	t_d *(s)*	0.5	0.5	0.42	0.6	0.7	0.55	0.0	0.0	0.0	3.5	5.92	4.48
	t_θ *(s)*	1.0	0.9	0.76	0.6	0.8	0.6	0.0	0.0	0.0	6.2	6.24	3.96
	t_r (s)	0.5	0.51	0.39	0.7	0.7	0.54	0.0	0.0	0.0	4.8	6.28	4.52
Six-lane	d_r (m/s^2)	0.4	0.64	0.21	1.0	1.14	0.42	0.0	0.002	0.003	6.8	9.87	5.39
	θ_r *(degrees/s)*	1.1	1.88	1.42	4.0	5.65	2.63	0.00	0.00	0.03	59.5	96.91	32.97
	t_d *(s)*	0.5	0.52	0.4	0.7	0.73	0.47	0.0	0.0	0.0	4.8	5.46	2.92
	t_θ *(s)*	1.0	0.94	0.74	0.8	0.88	0.55	0.0	0.0	0.29	7.2	6.64	3.63
	t_a (s)	0.5	0.5	0.35	0.7	0.69	0.48	0.0	0.0	0.0	4.8	6.64	3.63

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

6.1.2 Non-stationary Modelling Results of PTW

In the first stage, non-stationary models with single covariate were developed and then compared with the stationary model. **Table 6.3** 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 analysis. The non-stationary model can be said to 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 6.3 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, 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. Further, these models were verified against the stationary model and the best-fitted non-stationary models with a single covariate. **Table 6.3** 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 6.4 shows the estimation results of the best-performing GPD models. The shape parameter for both models was less than -0.5, indicating 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 6.2**) substituted in the scale function 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 6.4** shows that the sideswipe crash risk 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. Further, the study measures the precision of the crash risk in terms of 95% confidence bounds assuming that the standard errors are normally distributed for the maximum likelihood estimator. Past studies have mentioned that lower the upper bound of the crash frequency estimate, more precise the model (Hauer and Garder, 1986). The uncertainty in the estimates is reduced with the non-stationary model for both four-lane and six-lane highways. Further, the negative shape parameter for all the GPD models indicates that the GPD distribution has a finite upper limit. This is meaningful since the study areas will have a finite number of crashes for any given period. Also, the scale parameter (σ) indicates the spread of traffic conflicts across various severity levels. Smaller σ values ($\sigma_{4lane} = 0.248,\ \sigma_{6lane} = 0.27$) indicate that more conflicts in the high severity level with a narrow peak. Any disturbance in the traffic stream or driver error would lead to a sideswipe conflict being a crash.

Table 6.3 Statistics of the Fitted Non-stationary Models for PTW

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	-230.58	-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	**5.06***	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***	

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

Table 6.4 Estimation Results of the best fitted non-stationary models for PTW

Site	**Sample Size**	**Model**		**Maximum Likelihood Estimation Results**								
			LL	β_0	β_1	β_2	β_3	σ	ξ	*Crash Probability*	*Crash Frequency*	*95% confidence bounds*
4-lane	266	M2	-230.58	0.24 (0.01)	0.02 (0.01)	-	-	0.248	-0.47 (0.04)	0.09%	2	[0, 28]
6-lane	147	M10	-106.78	0.28 (0.00)	0.02 (0.00)	0.01 (0.00)	-0.04 (0.00)	0.27	-0.42 (0.02)	0.17%	5	[0, 19]

Note: values in the parentheses indicates standard errors

6.1.3 Non-stationary Modelling Results of PC

Table 6.5 shows that the covariates deceleration rate and steering time offer the best non-stationary model fit for the four-lane highway. In comparison to the base model, for the six-lane highways, the deceleration rate, yaw rate, and acceleration time provided the best-fitted non-stationary models (**Table 6.7**). Multiple non-stationary models, such as M2′, M3′, M4′, M5′, M6′, M7′, and M10′ were constructed, considering various combinations of the covariates. For both highways, the GPD models with covariates related to the evasive actions perform better than the stationary ones. The best performing model for the four-lane highway (i.e., M4′ in **Table 6.6**) contains the covariates namely the deceleration rate and steering time during a sideswipe conflict of PC. However, this is not the case with the six-lane highway, where the model only with the yaw rate performed the best (i.e., M6′ in **Table 6.8**). This is happening because the sideswipe conflict mainly occurs during lane change/overtaking actions, which are more prevalent on the six-lane highway in LMICs due to the high-speed difference between the road users. Compared to the six-lane highway, there are lesser lane-changing operations on the on the four-lane highway, and therefore, the steering effect of PCs on the crash risk may not be as significant.

Table 6.6 and **Table 6.8** show the estimation results of the best-performing GPD models of PC for both four-lane and six-lane highways. 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. **Table 6.6** and **Table 6.8** show that the sideswipe crash risk increases with the intensity of the evasive action, such as braking and steering. Further, the time spent in steering evasion negatively affects the sideswipe crash risk. For the non-stationary GPD models, the uncertainty in the sideswipe crash risk of PCs has significantly improved.

Table 6.5 Statistics of the Best Fitted Non-Stationary GPD Models for PC on four-lane highway

Model Number	**M1′**	**M2′**	**M3′**	**M4′**	
Model type	*Stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	
Description	Base case	$\sigma = \beta_0 + \beta_1 \times d_r$	$\sigma = \beta_0 + \beta_1 \times t_\theta$	$\sigma = \beta_0 + \beta_1 d_r + \beta_2 t_\theta$	
Site	LL1	LL2	LL3	LL4	
Four-lane	-287.07	-290.57	-291.64	-295.46*	
Likelihood ratio test results					
Site	M1′&M2′	M1′&M3′	M1′&M4′	M2′&M4′	M3′&M4′
Four-lane	6.98	9.12	9.12	9.78	7.64

Note: $\chi^2_{(0.95,df=1)} = 3.841$, $\chi^2_{(0.95,df=2)} = 5.991$. The * indicates the best performing non-stationary model out of all models.

Table 6.6 Estimation Results of the best performing non-stationary model for PC on four-lane highway

Site	**Threshold (s)**	**Sample Size**	**Model**	**Maximum Likelihood Estimation Results**							
				LL	β_0	β_1	β_2	ξ	*Crash Probability*	*Crash Frequency*	*95% confidence bounds*
Four-lane	0.65	425	M4′	-295.46	0.32 (0.02)	0.02 (0.01)	-0.04 (0.007)	-0.44 (0.03)	0.004%	0	[0, 22]

Note: values in the parentheses indicates standard errors

Table 6.7 Statistics of the Best Fitted Non-Stationary GPD Models for PC on six-lane highway

Model Number	**M1′**	**M5′**	**M6′**	**M7′**	**M8′**	**M9′**	**M10′**
Model type	*Stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*	*Non-stationary*
Description	Base case	$\sigma = \beta_0 + \beta_1 \times d_r$	$\sigma = \beta_0 + \beta_1 \times \theta_r$	$\sigma = \beta_0 + \beta_1 \times t_a$	$\sigma = \beta_0 + \beta_1 d_r + \beta_2 \theta_r$	$\sigma = \beta_0 + \beta_1 \theta_r + \beta_2 t_a$	$\sigma = \beta_0 + \beta_1 d_r + \beta_2 \theta_r + \beta_3 t_a$
Site	LL1	LL2	LL3	LL4	LL5	LL6	LL7
Six-lane	-287.07	-327.68	-328.76*	-327.55	-329.26	-330.00	-295.46*
Likelihood ratio test results							
Site	M1′&M5′	M1′&M6′	M1′&M7 ′	M6′&M8′	M6′&M9′	M6′&M10′	
Six-lane	6.26	8.4	5.98	1.08	2.49	3.18	

Note: $\chi^2_{(0.95,df=1)} = 3.841$, $\chi^2_{(0.95,df=2)} = 5.991$.The * indicates the best performing non-stationary model out of all models.

Table 6.8 Estimation Results of the best performing non-stationary model for PC on six-lane highway

Site	**Threshold (s)**	**Sample Size**	**Model**	**Maximum Likelihood Estimation Results**						
				LL	β_0	β_1	ξ	*Crash Probability*	*Crash Frequency*	*95% confidence bounds*
Six-lane	0.45	324	M6′	-328.76	0.20 (0.01)	0.004 (0.002)	-0.45 (0.04)	0.04%	1	[0, 12]

Note: values in the parentheses indicates standard errors

6.2 Non-stationary EVT Modelling of ROR Crash risk for All Curves

The variables such as R, L_C, d_r, θ_r, e, Δ, D_C, T_L, L_{at}, L_{et}, G_1, G_2, G_3, G_4, G_5, L_V, K_{value}, VC_{HC}, and L_0 were used to develop univariate non-stationary models (Table 6.9).

Table 6.9 Goodness-of-fit Statistics of Non-stationary Models with one Covariate

Model Number	Model type	Description	Goodness-of-fit (LL)		Likelihood Ratio=2(LL1-LLnon-stationary)
M1	*Stationary*	Base case	LL1	1235.92	
M2	***Non-stationary***	**$\sigma=\beta_0+\beta_1*d_r$**	**LL2**	**1223.26**	**25.34**
M3	*Non-stationary*	$\sigma=\beta_0+\beta_1*\theta_r$	LL3	1234.68	2.48
M4	***Non-stationary***	**$\sigma=\beta_0+\beta_1*R$**	**LL4**	**1222.51**	**26.82**
M5	*Non-stationary*	$\sigma=\beta_0+\beta_1*T_L$	LL5	1235.1	1.65
M6	*Non-stationary*	$\sigma=\beta_0+\beta_1*\Delta$	LL6	1234.98	1.89
M7	*Non-stationary*	$\sigma=\beta_0+\beta_1*L_C$	LL7	1235.92	0.01
M8	*Non-stationary*	$\sigma=\beta_0+\beta_1*e$	LL8	1235.22	1.41
M9	*Non-stationary*	$\sigma=\beta_0+\beta_1*D_C$	LL9	1235.6	0.65
M10	***Non-stationary***	**$\sigma=\beta_0+\beta_1*L_{at}$**	**LL10**	**1232.36**	**7.13**
M11	*Non-stationary*	$\sigma=\beta_0+\beta_1*L_{et}$	LL11	1234.31	3.23
M12	*Non-stationary*	$\sigma=\beta_0+\beta_1*G_1$	LL12	1234.28	3.29
M13	*Non-stationary*	$\sigma=\beta_0+\beta_1*G_2$	LL13	1234.18	3.49
M14	*Non-stationary*	$\sigma=\beta_0+\beta_1*G_3$	LL14	1234.76	2.33
M15	*Non-stationary*	$\sigma=\beta_0+\beta_1*G_4$	LL15	1234.34	3.17
M16	*Non-stationary*	$\sigma=\beta_0+\beta_1*G_5$	LL16	1234.67	2.51
M17	*Non-stationary*	$\sigma=\beta_0+\beta_1*L_V$	LL17	1234.92	2.01
M18	*Non-stationary*	$\sigma=\beta_0+\beta_1*K_{value}$	LL18	1235.81	0.23
M19	*Non-stationary*	$\sigma=\beta_0+\beta_1*VC_{HC}$	LL19	1235.89	0.07
M20	***Non-stationary***	**$\sigma=\beta_0+\beta_1*L_0$**	**LL20**	**1230.53**	**10.79**

Note: $\chi^2_{0.95,df=1} = 3.841$ should be less than the likelihood ratio values, then the model better fits the data.

Table 6.9 shows that the non-stationary models (M2, M4, M10, M20) developed considering covariates such as d_r, R, L_{at}, L_0 significantly improve the fit. The remaining covariates failed to improve the performance of the stationary crash risk model and therefore were not considered for

further analysis. Model M4 performed the best among all the univariate non-stationary models and was used as the reference model for further assessment of the bivariate non-stationary models.

The study developed the non-stationary models considering the linear combination of two covariates simultaneously. All the non-stationary models with two covariates performed better than model M4 (Table 6.10).

Table 6.10 Goodness-of-fit Statistics of Bivariate Non-stationary Models

Model Number	Model type	Description	Goodness-of-fit (LL)		Likelihood Ratio=2(LL4-LLnon-stationary)
M4	*Non-stationary*	$\sigma=\beta_0+\beta_1*R$	LL4	1222.515	
M21	***Non-stationary***	**$\sigma=\beta_0+\beta_1*R+\beta_2*d_r$**	**LL21**	**1197.64**	**49.75**
M22	*Non-stationary*	$\sigma=\beta_0+\beta_1*R+\beta_2*L_{at}$	LL22	1210.77	23.49
M23	*Non-stationary*	$\sigma=\beta_0+\beta_1*R+\beta_2*L_0$	LL23	1199.89	45.25
M24	*Non-stationary*	$\sigma=\beta_0+\beta_1*d_r+\beta_2*L_{at}$	LL24	1216.3	12.43
M25	*Non-stationary*	$\sigma=\beta_0+\beta_1*d_r+\beta_2*L_0$	LL25	1214.58	15.87
M26	*Non-stationary*	$\sigma=\beta_0+\beta_1*L_{at}+\beta_2*L_0$	LL26	1218.88	7.27

However, model M21 involving *R*, and d_r provided the best fit among the bivariate non-stationary models (Table 6.10). Therefore, it will be further compared with the trivariate non-stationary models. It should be mentioned here that all the non-stationary models with the radius of the curve as one of the explanatory variables perform significantly better than those without the curve radius. This is because the radius of the curve is one of the most critical road geometry parameters associated with ROR crashes at horizontal curve sections on two-lane rural highways (Donnell et al., 2019; Xin et al., 2017).

Similarly, the trivariate non-stationary model (M28), including curve radius, maximum deceleration rate, and distance between the center points of horizontal and vertical curves, outperformed the other trivariate non-stationary models (Table 6.). Hence, this non-stationary model was compared with the non-stationary model (M31), considering the linear combination of all four covariates. It was found that adding the length of the approach tangent to the M28 model resulted in a further improvement in the performance. Therefore, the model M31 with four covariates was selected as the best-fitted non-stationary model for HCV's ROR crash risk estimation. From this model, it was found that as the curve radius increases, the ROR crash risk

decreases ($\beta_1 = -0.01$). Whereas if the deceleration rate on the curve and length of approach tangent increases, the ROR crash risk increases ($\beta_2 = 0.22; \beta_3 = 0.004$). Also, as the distance between the midpoints of horizontal and vertical curves increases, the ROR crash risk increases ($\beta_4 = 0.007$).

Table 6.11 Goodness-of-fit Statistics of Non-stationary Models with Three and Four Covariates

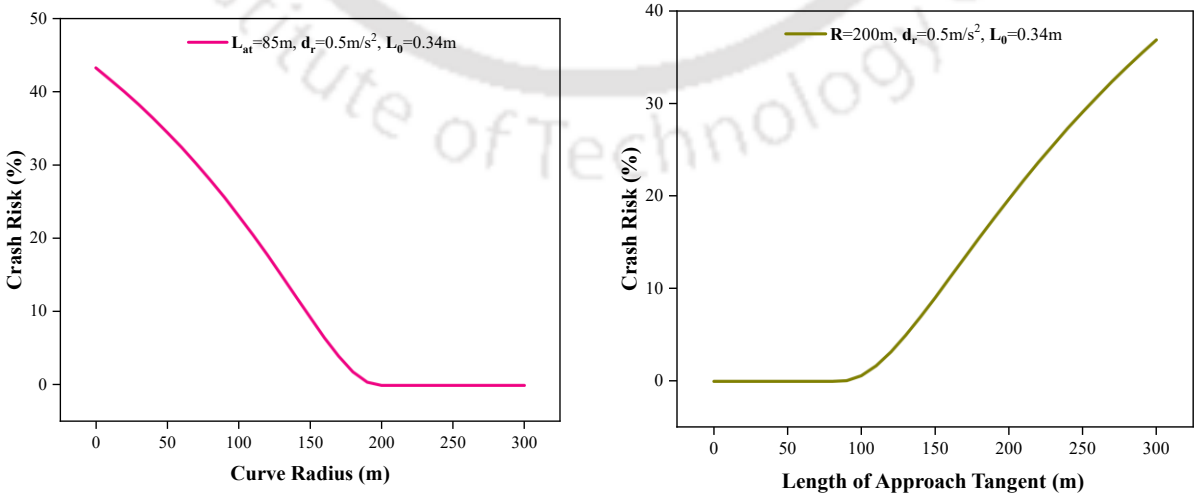
Model Number	Model type	Description	Goodness-of-fit (LL)		Likelihood Ratio=2(LL21-LLnon-stationary)
M21	*Non-stationary*	$\sigma=\beta_0+\beta_1*R+\beta_2*d_r$	LL21	1197.64	
M27	*Non-stationary*	$\sigma=\beta_0+\beta_1*R+\beta_2*d_r+\beta_3*L_{at}$	LL27	1189.71	15.86
M28	***Non-stationary***	**$\sigma=\beta_0+\beta_1*R+\beta_2*d_r+\beta_3*L_0$**	**LL28**	**1178.45**	**38.38**
M29	*Non-stationary*	$\sigma=\beta_0+\beta_1*R+\beta_2*L_{at}+\beta_3*L_0$	LL29	1195.79	3.7
M30	*Non-stationary*	$\sigma=\beta_0+\beta_1*d_r+\beta_2*L_{at}+\beta_3*L_0$	LL30	1200.26	No Improvement
		Model with Four Covariates			
M31	***Non-stationary***	**$\sigma=\beta_0+\beta_1*R+\beta_2*d_r+\beta_3*L_{at}+\beta_4*L_0$**	**LL31**	**1174.69**	**7.52**

Table 6.11 shows that the estimated crash risk is 0.34% with shape parameter ($\xi = -0.394$) and scale paramter ($\sigma = 0.882$). Further, the 95% confidence bounds of the crash frequency estimates of the non-stationary model $[1, 2]$ matches closely with that obtained from the observed crash frequency $[0.35, 1.8]$. This highlights the validity of the non-stationary model M31 for predicting crash risk.

Table 6.12 Parameters of the Best Fitted Non-stationary GPD Model

Parameters	β_0 (SE)	β_1 (SE)	β_2 (SE)	β_3 (SE)	β_4 (SE)	σ	ξ (SE)	LL	Crash Risk (R)	Estimated Crash bounds	Observed Crash bounds
Values	1.92 (0.03)	-0.01 (0.00)	0.22 (0.05)	0.004 (0.00)	0.007 (0.00)	0.882	-0.394 (0.00)	1174.69	0.34%	[1, 2]	[0.35, 1.8]

The study further analyzed the design variables' sensitivity for the best-fitted non-stationary model. The effect of the design variables on the ROR crash risk was studied while keeping the other covariates in the model constant. With the increase in curve radius, ROR crash risk decreases. Figure 6.3 (a) shows that curves with smaller radii are associated with significantly higher crash risk. Elvik (2013) performed an extensive study on the effect of curve radius on crashes across multiple countries and found that sharp curves with radii less than 200 m are prone to crashes. Therefore, such radii should be avoided while designing horizontal curves because drivers may not expect or anticipate such upcoming sharp curves. It is true, mainly when horizontal curves are overlapped with vertical curves and the sight distance is inadequate. For a curve with a specific radius, the crash risk increases with the length of the approach tangent (Figure 6.3 (b)). Therefore, horizontal curves following longer tangents will experience significant crash risk. This is because drivers will travel with high operating speeds on longer tangents giving rise to considerable crash risk. This will be even more dangerous if sharp curves are provided at the end of longer tangents (Figure 6. (c)). Hence, AASHTO Green book clearly states that "sharp curves should not be introduced at the ends of long tangents." As the distance between the midpoints of horizontal and vertical curves increases, crash risk increases (Figure 6.3 (d)). Drivers may misperceive the road geometry if there is a more considerable distance between the midpoints of horizontal and vertical curves. It can be seen from Figure 6.3 (d) that curves with an offset of more than 1 m are associated with significantly higher crash risk. Hence, the midpoints of horizontal and vertical curves should overlap. For exceptional circumstances, if such practice is necessary, then warning signs should be provided on the curves to alert the drivers so that they can anticipate the crash risk.



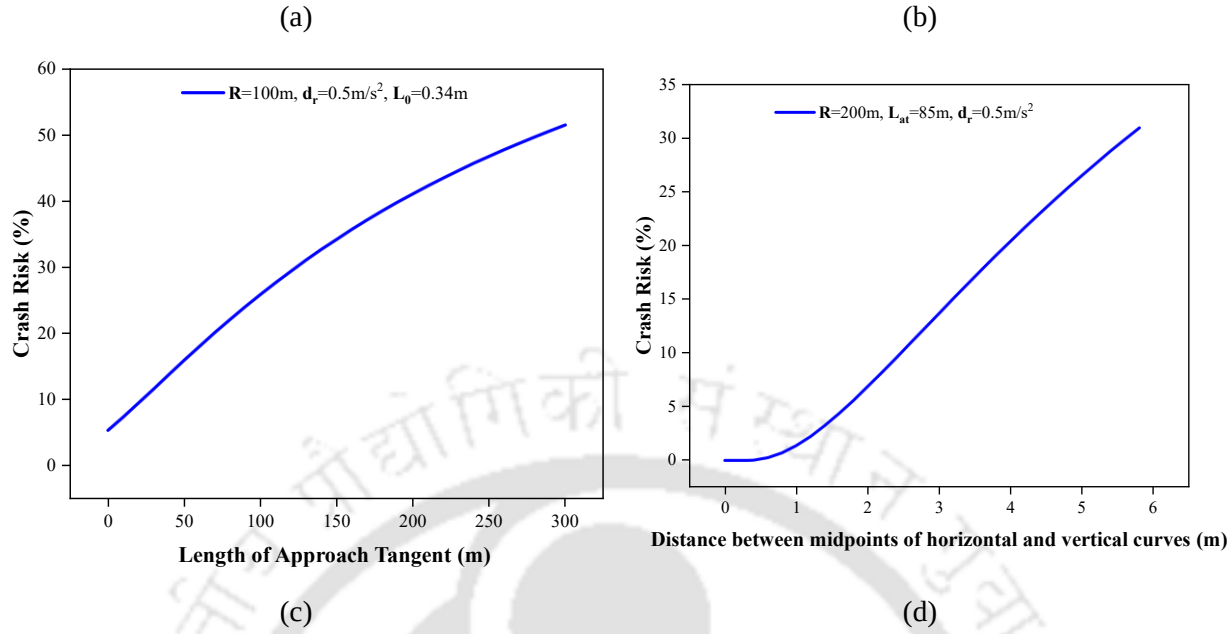


Figure 6.3 Sensitivity Analysis of Design Variables on Crash Risk

6.3 Summary

The present study analyses and models the effect of microscopic driving variables such as braking and steering associated with the safety margin (i.e., ACT_{min}) for the sideswipe and ROR crash risks on multilane highways and the curves of undivided rural highways passing through mountainous terrain. This was carried out by incorporating the effect of the influencing covariates on the scale parameter of GPD models as an identity link function and such models are called non-stationary GPD models.

For the multilane highways, it was found that both braking and steering influences the sideswipe crash risk of the mixed traffic. The sideswipe crash risk increases with the increase in the braking and yaw rates of the vehicles. For the six-lane highways, the yaw rate seems to be the major influencing variable for the sideswipe crash risk. This is happening because sideswipe conflicts predominantly arise due to the frequent lane changing or overtaking operations on such highways. Compared to the four-lane highways in LMICs, such actions become more predominant in six-lane highways and hence the steering rate becomes more significant.

For the horizontal curves of undivided highway passing through mountainous terrain, the ROR crash risk of HCV depends on the curve radius, length of approach tangent, deceleration rate, and distance between the midpoints of horizontal and vertical curves. The non-stationary model

matches closely with the observed crash frequency. As the curve radius increases, crash risk decreases whereas crash risk increases with the increase in approach tangent length and distance between horizontal and vertical curves mid points.

Chapter 7 Summary and Conclusions

7.1 Summary

This thesis put forward a systematic framework for the crash risk assessment of the mixed traffic prevalent on rural highways in LMICs. The key findings obtained from the previous chapters of the thesis will be beneficial for the traffic safety operations of rural highways. This study assessed the crash risk of mixed traffic on rural highways passing through flat and mountainous terrains in India using a multidimensional conflict indicator named ACT. For the safety evaluation, the safety metric such as crash risk was calculated by using the EVT theory.

The major findings derived based on the conflict assessment using ACT are:

1. Sideswipe conflicts with higher crash potential happen more frequently as compared to the rear-end conflicts on Indian multilane rural highways operated under low volume conditions.
2. Single-vehicle conflicts, namely ROR, happen more frequently on the horizontal curves of undivided highways passing through mountainous terrain than that of multi-vehicle crashes such as rear-end, sideswipe, and head-on types.
3. The frequency of potential conflicts estimated for multiple conflict types using ACT_{min} values corroborates with the frequency trends obtained from crash data.

The main findings drawn from the stationary EVT modelling of crash risk are:

1. During sideswipe conflict, on average, PTW riders have a significantly lower safety margin available on the six-lane highway as compared to the four-lane highway. Whereas for the PC, the safety margin available is relatively larger as compared to the PTW on both the four-lane and six-lane highways.
2. The safety margin (i.e., ACT_{min}) during a sideswipe conflict decreases with the increase in the maximum speed difference between two road users.
3. 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. PTW and PC experience higher crash risk on the six-lane highway as compared to the four-lane highway. The crash risk of PTW is significantly higher than that of PC on both the highways.

4. The uncertainty in the crash risk estimates are significantly higher in the case of BM models than that of POT models for the mixed traffic.
5. Based on the POT models, HCVs experience ROR crash risk on horizontal curves passing through mountainous terrain. The threshold is 2 seconds for the study location.

The main findings obtained from the non-stationary EVT modelling of crash risk are:

1. On average, during a sideswipe conflict, vehicles 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. The sideswipe crash risk of vehicles increases with the increase in the intensity of the maximum deceleration and maximum yaw rates.
3. Non-stationary POT models for the sideswipe conflict type have significantly reduced the uncertainty in the crash risk estimates compared to the stationary POT models.
6. The ROR crash risk of HCV decreases as the curve radius increases. A previous study based on crash data (Xin et al., 2017) also found that the ROR crashes on a horizontal curve decrease with the increase in curve radius. Also, the crash risk increases with the length of the approach tangent. This conforms with the findings of the study of Donnell et al. (2019).
7. Non-stationary POT models with geometric variables and evasive actions provided the most accurate prediction results when validated with police-recorded crash data.

7.2 Conclusions of the Study

The main conclusions drawn from the findings are:

1. For the multilane rural highways, more attention should be given to the sideswipe conflicts due to their high crash likelihood.
2. Drivers evade a sideswipe conflict mainly by making extreme steering evasion rather than applying hard brakes. Therefore, for the sideswipe conflict identification, 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.

3. PTWs experience significantly higher sideswipe (sometimes also referred as lane-changing) crash risk on the six-lane highway as compared to the four-lane highway. This means that providing additional lane (or widening the roads from four-lane to six-lane) does not fully improve the safety 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.
4. 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.
5. For the horizontal curves, the applicability of ACT_{min} as a crash precursor seems justified since the present study found corroborating results between the frequency of potential conflicts and crashes.
6. HCVs experience ROR crash risk on the horizontal curves, and the curve should be provided with a larger radius and smaller tangent lengths to counteract such risk.
7. Center points should coincide for curves superimposed with vertical curves to minimize the ROR crash risk of HCVs. This agrees with the codal provisions suggesting that there should be no distance between the center points of horizontal and vertical curves. Hence, care must be taken while designing a horizontal curve overlapped with a vertical curve.
8. The effect of evasive actions and road geometry need to be incorporated in the ROR crash risk models of HCVs. Otherwise, the crash risk prediction will lead to overestimating crash frequency.

7.3 Contributions of the Study

The contribution of this study is manifold.

1. This study adds to the relatively little literature available on the crash risk assessment of mixed traffic using a newly developed multidimensional proactive measure called ACT. This measure helps in identifying the risky situations by integrating the steering effect on the proximity

calculation, which is a very important parameter for the lane-changing operations of mixed traffic.

2. By estimating the safety metric (i.e., Crash Risk) using the EVT approach, the present study provided a systematic workaround to perform risk assessment without using police recorded crash data. This can be useful in conducting real-time threat assessment on rural highways.
3. 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.
4. Using crash data, most studies assessed the effect of road geometry on crashes at horizontal curves. The present study contributed towards incorporating the impact of geometric variables and the evasive actions of drivers moving on horizontal curves into the crash risk modeling without using the crash data. This is useful for the safety assessment of drivers on horizontal curves, particularly for the LMICs where the reported crash data could be highly erroneous.

7.4 Future Scope of the Study

The present study can be further extended in numerous ways.

1. The non-stationary EVT models can be further improved by collecting the trajectory data for extended periods of the day. Finally, more data should be collected on the six-lane highway to further understand the crash risk dynamics of mixed traffic on such highways. Though the non-stationary EVT models outperform the stationary models, the former need further validation using historical crash data.
2. The crash severity of sideswipe conflict in terms of vehicular damage or injury can also be incorporated in identifying the severity of traffic conflicts. This can be achieved by using speed-based measures such as maximum speed difference, delta V etc. that can capture the vehicular damage and injury to the drivers. However, the crash severity levels should be well established beforehand by collecting sensor-based data for extended periods of time over longer road stretches.

3. 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 composition.
4. 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.
5. Finally, the effect of traffic stream characteristics on the crash risk should be studied by collecting data from different locations with varying traffic compositions.
6. The non-stationary ROR model should be further improved by collecting trajectory data for more extended periods. The transferability analysis of the developed crash risk model should be carried out on other horizontal curves.

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. *Analytic Methods in Accident Research*, *33*, 100193. https://doi.org/10.1016/J.AMAR.2021.100193

Allen, B. L., Shin, B. T., & Cooper, P. (1978). Analysis of Traffic Conflicts and Collisions. *Transportation Research Record: Journal of the Transportation Research Board*, *667*, 67–74.

Amundsen, F., & Hyden, C. (1977). Proceedings: First Workshop on Traffic Conflicts. *Proceedings: First Workshop on Traffic Conflicts*. https://www.semanticscholar.org/paper/Proceedings%3A-first-Workshop-on-Traffic-Conflicts%2C-Peterson/03db15dddf95d905883752d3e0b639306684bfa5

Antin, J., Lee, S., Hankey, J., & Dingus, T. (2011). Design of the In-Vehicle Driving Behavior and Crash Risk Study. *Design of the In-Vehicle Driving Behavior and Crash Risk Study: In Support of the SHRP 2 Naturalistic Driving Study*. https://doi.org/10.17226/14494

Arun, A., Haque, M. M., Bhaskar, A., Washington, S., & Sayed, T. (2021). A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Analytic Methods in Accident Research*, *32*, 100180. https://doi.org/10.1016/j.amar.2021.100180

Arun, A., Haque, M. M., Washington, S., Sayed, T., & Mannering, F. (2021). A systematic review of traffic conflict-based safety measures with a focus on application context. *Analytic Methods in Accident Research*, *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. *Transportation Research Part C: Emerging Technologies*, *138*, 103653. https://doi.org/10.1016/J.TRC.2022.103653

Autey, J., Sayed, T., & Zaki, M. H. (2012). Safety evaluation of right-turn smart channels using automated traffic conflict analysis. *Accident Analysis and Prevention*, *45*, 120–130. https://doi.org/10.1016/j.aap.2011.11.015

Bagdadi, O., & Várhelyi, A. (2011). Jerky drivingâ€"An indicator of accident proneness? *Accident Analysis and Prevention*, *43*, 1359–1363. https://doi.org/10.1016/j.aap.2011.02.009

Barnard, Y., Utesch, F., van Nes, N., Eenink, R., & Baumann, M. (2016). The study design of UDRIVE: the naturalistic driving study across Europe for cars, trucks and scooters. *European Transport Research Review*, *8*(2), 1–10. https://doi.org/10.1007/S12544-016-0202-Z/FIGURES/4

Bharat Kumar Anna, V. A., Venthuruthiyil, S. P., & Chunchu, M. (2022). Vehicle trajectory data extraction from the horizontal curves of mountainous roads. *Transportation Letters*. https://doi.org/10.1080/19427867.2022.2125487

Borsos, A. (2021). Application of Bivariate Extreme Value models to describe the joint behavior of temporal and speed related surrogate measures of safety. *Accident Analysis & Prevention*, *159*, 106274. https://doi.org/10.1016/J.AAP.2021.106274

Cafiso, S., D'Agostino, C., Kieć, M., & Bak, R. (2018). Safety assessment of passing relief lanes using microsimulation-based conflicts analysis. *Accident Analysis & Prevention*, *116*, 94–102. https://doi.org/10.1016/J.AAP.2017.07.001

Campbell, K. L., Joksch, H. C., & Green, P. E. (1996). *A Bridging Analysis for Estimating the*

Benefits of Active Safety Technologies Task One under Contract No. DTNH22-93-D-07000 Crash Avoidance Research Technology Support-Simulation Models Final Report.
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. *Accident Analysis and Prevention*, *134*. https://doi.org/10.1016/J.AAP.2019.105315
Charly, A., & Mathew, T. V. (2019). Estimation of traffic conflicts using precise lateral position and width of vehicles for safety assessment. *Accident Analysis and Prevention*, *132*, 1–10. https://doi.org/10.1016/j.aap.2019.105264
Chin, H. C., & Quek, S. T. (1997). Measurement of traffic conflicts. *Safety Science*, *26*(3), 169–185. https://doi.org/10.1016/S0925-7535(97)00041-6
Choudhary, P., & Velaga, N. R. (2017). Modelling driver distraction effects due to mobile phone use on reaction time. *Transportation Research Part C: Emerging Technologies*, *77*, 351–365. https://doi.org/10.1016/J.TRC.2017.02.007
Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London. https://doi.org/10.1007/978-1-4471-3675-0
Cooper, D. F., & Ferguson, N. (1976). Traffic Studies at T-Junctions - A Conflict Simulation Record. *Traffic Engineering & Control*, *17*(17), 306–309. https://trid.trb.org/view/66554
Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. *Accident Analysis and Prevention*, *43*(6), 1907–1919. https://doi.org/10.1016/J.AAP.2011.05.001
Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2636–2641. https://doi.org/10.1073/PNAS.1513271113/SUPPL_FILE/PNAS.201513271SI.PDF
Dingus, T. A., Klauer, S. G., Neale, V. L. (Vicki L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R. (Zachary R., Jermeland, J., Knipling, R. R. (Ronald R. ., & Institute, V. P. I. and S. U. T. (2006). *The 100-Car Naturalistic Driving Study, Phase II - Results of the 100-Car Field Experiment*. United States. Department of Transportation. National Highway Traffic Safety Administration. https://doi.org/10.21949/1503647
Dingus, T. A., Neale, V. L., Klauer, S. G., Petersen, A. D., & Carroll, R. J. (2006). The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking. *Accident Analysis & Prevention*, *38*(6), 1127–1136. https://doi.org/10.1016/J.AAP.2006.05.001
El-Basyouny, K., & Sayed, T. (2013). Safety performance functions using traffic conflicts. *Safety Science*, *51*(1), 160–164. https://doi.org/10.1016/j.ssci.2012.04.015
Ellison, A. B., Greaves, S. P., & Bliemer, M. C. J. (2015). Driver behaviour profiles for road safety analysis. *Accident Analysis & Prevention*, *76*, 118–132. https://doi.org/10.1016/J.AAP.2015.01.009
Evans, L. (2001). Causal influence of car mass and size on driver fatality risk. *American Journal of Public Health*, *91*(7), 1076. https://doi.org/10.2105/AJPH.91.7.1076
Farah, H., & Azevedo, C. L. (2017). Safety analysis of passing maneuvers using extreme value theory. *IATSS Research*, *41*(1), 12–21. https://doi.org/10.1016/j.iatssr.2016.07.001
Feng, F., Bao, S., Sayer, J. R., Flannagan, C., Manser, M., & Wunderlich, R. (2017). Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic

driving data. *Accident Analysis and Prevention*, *104*, 125–136. https://doi.org/10.1016/J.AAP.2017.04.012
Fu, C., & Sayed, T. (2021). Multivariate Bayesian hierarchical Gaussian copula modeling of the non-stationary traffic conflict extremes for crash estimation. *Analytic Methods in Accident Research*, *29*, 100154. https://doi.org/10.1016/J.AMAR.2020.100154
Fu, C., Sayed, T., & Zheng, L. (2020). Multivariate Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. *Analytic Methods in Accident Research*, *28*, 100135. https://doi.org/10.1016/J.AMAR.2020.100135
Gaweesh, S. M., Ahmed, M. M., & Piccorelli, A. V. (2019). Developing crash prediction models using parametric and nonparametric approaches for rural mountainous freeways: A case study on Wyoming Interstate 80. *Accident Analysis & Prevention*, *123*, 176–189. https://doi.org/10.1016/J.AAP.2018.10.011
Gettman, D., & Head, L. (2003). Surrogate Safety Measures from Traffic Simulation Models. *Transportation Research Record*, *1840*, 104–115. https://doi.org/10.3141/1840-12
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
Glauz, W. D., & Migletz, D. J. (1980). APPLICATION OF TRAFFIC CONFLICT ANALYSIS AT INTERSECTIONS. *National Cooperative Highway Research Program Report*, *219*. https://trid.trb.org/view/153539
Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., & Yuan, J. (2019). Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. *Accident Analysis & Prevention*, *123*, 159–169. https://doi.org/10.1016/J.AAP.2018.11.010
Guo, F. (2019). Statistical methods for naturalistic driving studies. In *Annual Review of Statistics and Its Application* (Vol. 6, pp. 309–328). Annual Reviews. https://doi.org/10.1146/annurev-statistics-030718-105153
Guo, F., Klauer, S. G., Hankey, J. M., & Dingus, T. A. (2010). Near crashes as crash surrogate for naturalistic Driving Studies. *Transportation Research Record*, *2147*, 66–74. https://doi.org/10.3141/2147-09
Guo, Y., Sayed, T., & Zaki, M. H. (2018). Exploring Evasive Action–Based Indicators for PTW Conflicts in Shared Traffic Facility Environments. *Journal of Transportation Engineering, Part A: Systems*, *144*(11). https://doi.org/10.1061/jtepbs.0000190
Güttinger, V. A. (1984). *Conflict Observation in Theory and in Practice*. 17–24. https://doi.org/10.1007/978-3-642-82109-7_3
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. *Safety Science*, *146*, 105513. https://doi.org/10.1016/J.SSCI.2021.105513
Hallmark, S. L., Tyner, S., Oneyear, N., Carney, C., & McGehee, D. (2015). Evaluation of driving behavior on rural 2-lane curves using the SHRP 2 naturalistic driving study data. *Journal of Safety Research*, *54*, 17.e1-27. https://doi.org/10.1016/J.JSR.2015.06.017
Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis and Prevention*, *14*(5), 359–364. https://doi.org/10.1016/0001-4575(82)90014-8
Hauer, E., & Garder, P. (1986). Research into the validity of the traffic conflicts technique. *Accident Analysis & Prevention*, *18*(6), 471–481. https://doi.org/10.1016/0001-4575(86)90020-5
Hayward, J. C. (1972). NEAR-MISS DETERMINATION THROUGH USE OF A SCALE OF

DANGER. *51st Annual Meeting of the Highway Research Board*, 24–35. https://www.semanticscholar.org/paper/NEAR-MISS-DETERMINATION-THROUGH-USE-OF-A-SCALE-OF-Hayward/d7dc3003871814c9ac84d9c75456aa8089fc70fd

Hupfer, C. (1997). Deceleration to safety time (DST) - a useful figure to evaluate traffic safety? *The 10th ICTCT Workshop, Lund*.

Hutchinson, T. P. (1977). Intra-accident correlations of driver injury and their application to the effect of mass ratio on injury severity. *Accident Analysis & Prevention*, *9*(3), 217–227. https://doi.org/10.1016/0001-4575(77)90023-9

Hydén, C. (1987). The development of a method for traffic sfaety evaluation: The Swedish Traffic-Conflicts Technique [Lund Institute of Technology]. In *Lund Institute of Technology*. https://doi.org/10.1007/978-3-642-82109-7_12

Hyden, C., Gaarder, P., & Linderholm, L. (1982). An updating of the use and further development ofthe traffic conflicts technique. *Proceedings of the Third International Workshop on Traffic Conflicts Techniques*, 42–48. https://trid.trb.org/view/188279

Ismail, K., Sayed, T., & Saunier, N. (2011). Methodologies for aggregating indicators of traffic conflict. *Transportation Research Record*, *2237*(2237), 10–19. https://doi.org/10.3141/2237-02

Johnsson, C., Laureshyn, A., & De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. *Transport Reviews*, *38*(6), 765–785. https://doi.org/10.1080/01441647.2018.1442888

Jonasson, J. K., & Rootzén, H. (2014). Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach. *Accident Analysis and Prevention*, *62*, 102–109. https://doi.org/10.1016/J.AAP.2013.09.013

Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, *94*, 182–193. https://doi.org/10.1016/J.TRA.2016.09.010

Laureshyn, A., De Ceunynck, T., Karlsson, C., Svensson, Å., & Daniels, S. (2017). In search of the severity dimension of traffic events: Extended Delta-V as a traffic conflict indicator. *Accident Analysis & Prevention*, *98*, 46–56. https://doi.org/10.1016/J.AAP.2016.09.026

Laureshyn, A., Svensson, Å., & Hydén, C. (2010). Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. *Accident Analysis and Prevention*, *42*(6), 1637–1646. https://doi.org/10.1016/j.aap.2010.03.021

Laureshyn, A., & Varhelyi, A. (2020). *The Swedish traffic conflict technique observer's manual* (Issue July).

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

Liu, T., Li, Z., Liu, P., Xu, C., & Noyce, D. A. (2021). Using empirical traffic trajectory data for crash risk evaluation under three-phase traffic theory framework. *Accident Analysis & Prevention*, *157*, 106191. https://doi.org/10.1016/j.aap.2021.106191

Mannering, F., Bhat, C. R., Shankar, V., & Abdel-Aty, M. (2020). Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic Methods in Accident Research*, *25*, 100113. https://doi.org/10.1016/j.amar.2020.100113

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, *1*, 1–22. https://doi.org/10.1016/J.AMAR.2013.09.001

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, *11*, 1–16. https://doi.org/10.1016/J.AMAR.2016.04.001

Migletz, J., Glauz, W. D., Bauer, K. M., & Midwest Research Institute (Kansas City, M. . (1985). *Relationships between traffic conflicts and accidents. Volume 2, Final technical report*. Turner-Fairbank Highway Research Center. https://doi.org/10.21949/1503647

Minderhoud, M. M., & Bovy, P. H. L. (2001). Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis and Prevention*, *33*(1), 89–97. https://doi.org/10.1016/S0001-4575(00)00019-1

MORTH. (2020). *Road Accidents in India 2020*.

Oh, C., Park, S., & Ritchie, S. G. (2006). A method for identifying rear-end collision risks using inductive loop detectors. *Accident Analysis and Prevention*, *38*(2), 295–301. https://doi.org/10.1016/j.aap.2005.09.009

Ozbay, K., Associate Professor, P. D., Yang, H., Research Assistant, G., Bartin, B., Research Associate, P. D., & Mudigonda, S. (2007). *Derivation and Validation of a New Simulation-based Surrogate Safety Measure*.

Ozbay, K., Yang, H., Bartin, B., & Mudigonda, S. (2008). Derivation and validation of new simulation-based surrogate safety measure. *Transportation Research Record: Journal of the Transportation Research Board*, *2083*(1), 105–113. https://doi.org/10.3141/2083-12

Papadoulis, A., Quddus, M., & Imprialou, M. (2019). Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accident Analysis & Prevention*, *124*, 12–22. https://doi.org/10.1016/J.AAP.2018.12.019

Papazikou, E., Quddus, M., Thomas, P., & Kidd, D. (2019). What came before the crash? An investigation through SHRP2 NDS data. *Safety Science*, *119*, 150–161. https://doi.org/10.1016/J.SSCI.2019.03.010

Pawar, N. M., Khanuja, R. K., Choudhary, P., & Velaga, N. R. (2020). Modelling braking behaviour and accident probability of drivers under increasing time pressure conditions. *Accident Analysis & Prevention*, *136*, 105401. https://doi.org/10.1016/J.AAP.2019.105401

Peesapati, L. N., Hunter, M. P., & Rodgers, M. O. (2018). Can post encroachment time substitute intersection characteristics in crash prediction models? *Journal of Safety Research*, *66*, 205–211. https://doi.org/10.1016/J.JSR.2018.05.002

Perkins, S. R., & Harris, J. I. (1969). Traffic Conflict Characteristics-Accident Potential at Intersections. *HRB Rec. 225, Highway Res. Bd*, *225*, 35–44.

Regan, M. A., Williamson, A., Grzebieta, R., Charlton, J., Lenne, M., Watson, B., Haworth, N., Rakotonirainy, A., Woolley, J., Anderson, R., Senserrick, T., & Young, K. (2013, August). The Australian 400-car Naturalistic Driving Study: innovation in road safety research and policy. *Australasian Road Safety Research Policing Education Conference, 2013*.

Roshandel, S., Zheng, Z., & Washington, S. (2015). Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis. *Accident; Analysis and Prevention*, *79*, 198–211. https://doi.org/10.1016/J.AAP.2015.03.013

Sacchi, E., & Sayed, T. (2016). Conflict-based safety performance functions for predicting traffic collisions by type. *Transportation Research Record*, *2583*, 50–55. https://doi.org/10.3141/2583-07

Santos, J., Merat, N., Mouta, S., Brookhuis, K., & De Waard, D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F: Traffic Psychology and*

Behaviour, *8*(2), 135–146. https://doi.org/10.1016/J.TRF.2005.04.001
Sayed, T., & Zein, S. (1999). Traffic conflict standards for intersections. *Transportation Planning and Technology*, *22*(4), 309–323. https://doi.org/10.1080/03081069908717634
Shahdah, U., Saccomanno, F., & Persaud, B. (2014). Integrated traffic conflict model for estimating crash modification factors. *Accident Analysis & Prevention*, *71*, 228–235. https://doi.org/10.1016/J.AAP.2014.05.019
Shelby, S. G. (2011). Delta-V as a Measure of Traffic Conflict Severity. *Transportation Research Record*, 1–19.
Singh, H., & Kathuria, A. (2021). Analyzing driver behavior under naturalistic driving conditions: A review. *Accident Analysis and Prevention*, *150*, 105908. https://doi.org/10.1016/j.aap.2020.105908
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. *Accident Analysis & Prevention*, *38*(4), 811–822. https://doi.org/10.1016/J.AAP.2006.02.003
Stipancic, J., Miranda-Moreno, L., Saunier, N., & Labbe, A. (2018). Surrogate safety and network screening: Modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. *Accident Analysis & Prevention*, *120*, 174–187. https://doi.org/10.1016/J.AAP.2018.07.013
Svensson, Å., & Hydén, C. (2006). Estimating the severity of safety related behaviour. *Accident Analysis and Prevention*, *38*(2), 379–385. https://doi.org/10.1016/j.aap.2005.10.009
Tageldin, A., Sayed, T., & Shaaban, K. (2017). Comparison of time-proximity and evasive action conflict measures case studies from five cities. In *Transportation Research Record* (Vol. 2661, pp. 19–29). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.3141/2661-03
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. *Transportation Research Record*, *2520*, 165–174. https://doi.org/10.3141/2520-19
Tarko, A. P. (2012a). Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis and Prevention*, *45*, 230–240. https://doi.org/10.1016/j.aap.2011.07.008
Tarko, A. P. (2012b). Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis & Prevention*, *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. *Accident Analysis & Prevention*, *113*, 63–73. https://doi.org/10.1016/J.AAP.2018.01.008
Tarko, A. P. (2021). A unifying view on traffic conflicts and their connection with crashes. *Accident Analysis & Prevention*, *158*, 106187. https://doi.org/10.1016/J.AAP.2021.106187
Theofilatos, A., Chen, C., & Antoniou, C. (2019). Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction. *Transportation Research Record*, *2673*(8), 169–178. https://doi.org/10.1177/0361198119841571/ASSET/IMAGES/LARGE/10.1177_0361198119841571-FIG1.JPEG
Valero-Mora, P. M., Tontsch, A., Welsh, R., Morris, A., Reed, S., Touliou, K., & Margaritis, D. (2013). Is naturalistic driving research possible with highly instrumented cars? Lessons learnt in three research centres. *Accident Analysis & Prevention*, *58*, 187–194.

https://doi.org/10.1016/J.AAP.2012.12.025
Venthuruthiyil, S. P. (2021). *Trajectory-Based Proactive Safety Assessment of Road Traffic* [Indian Institute of Technology Guwahati]. https://www.gyan.iitg.ac.in:8080/xmlui/handle/123456789/2199
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: Transport Science*, *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. *Transportation Research Part C: Emerging Technologies*, *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. *Transportation Research Part C: Emerging Technologies*, *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*.
Virdi, N., Grzybowska, H., Waller, S. T., & Dixit, V. (2019). A safety assessment of mixed fleets with Connected and Autonomous Vehicles using the Surrogate Safety Assessment Module. *Accident Analysis & Prevention*, *131*, 95–111. https://doi.org/10.1016/J.AAP.2019.06.001
Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident Analysis and Prevention*, *35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2
Wang, Chao, Quddus, M., & Ison, S. (2009). The effects of area-wide road speed and curvature on traffic casualties in England. *Journal of Transport Geography*, *17*(5), 385–395. https://doi.org/10.1016/J.JTRANGEO.2008.06.003
Wang, Chen, Xu, C., & Dai, Y. (2019). A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accident Analysis & Prevention*, *123*, 365–373. https://doi.org/10.1016/J.AAP.2018.12.013
WHO, W. H. O. (2018). Global status report on road safety 2018. In *Geneva, Switzerland, WHO*. https://www.google.com/search?q=global+status+report+on+road+safety+2018&newwindow=1&rlz=1C1CHBD_enIN934IN934&sxsrf=APq-WBumb7x0fiAkKlsq2b9G6PqeKTBHyQ%3A1650233337882&ei=-Y9cYpLDNYXB3LUP_NmOwA8&gs_ssp=eJzj4tVP1zc0TMvKM08pLzMyYPTSSM_JT0rMUSguSSwpLVYoSi3ILy
Wu, K. F., Aguero-Valverde, J., & Jovanis, P. P. (2014). Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accident Analysis and Prevention*, *72*, 210–218. https://doi.org/10.1016/J.AAP.2014.07.005
Wu, K. F., & Jovanis, P. P. (2012). Crashes and crash-surrogate events: Exploratory modeling

with naturalistic driving data. *Accident Analysis and Prevention*, *45*, 507–516. https://doi.org/10.1016/j.aap.2011.09.002
Xie, K., Ozbay, K., Yang, H., & Li, C. (2019). Mining automatically extracted vehicle trajectory data for proactive safety analytics. *Transportation Research Part C: Emerging Technologies*, *106*, 61–72. https://doi.org/10.1016/J.TRC.2019.07.004
Xie, K., Yang, D., Ozbay, K., & Yang, H. (2019). Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accident; Analysis and Prevention*, *125*, 311–319. https://doi.org/10.1016/J.AAP.2018.07.002
Xu, C., Liu, P., Wang, W., & Li, Z. (2014). Identification of freeway crash-prone traffic conditions for traffic flow at different levels of service. *Transportation Research Part A: Policy and Practice*, *69*, 58–70. https://doi.org/10.1016/J.TRA.2014.08.011
Xu, C., Liu, P., Wang, W., & Li, Z. (2015). Safety performance of traffic phases and phase transitions in three phase traffic theory. *Accident Analysis & Prevention*, *85*, 45–57. https://doi.org/10.1016/J.AAP.2015.08.018
Xu, C., Tarko, A. P., Wang, W., & Liu, P. (2013). Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis and Prevention*, *57*, 30–39. https://doi.org/10.1016/j.aap.2013.03.035
Yang, D., Ozbay, K., Xie, K., Yang, H., Zuo, F., & Sha, D. (2021). Proactive safety monitoring: A functional approach to detect safety-related anomalies using unmanned aerial vehicle video data. *Transportation Research Part C: Emerging Technologies*, *127*, 103130. https://doi.org/10.1016/J.TRC.2021.103130
Yang, G., Ahmed, M., Gaweesh, S., & Adomah, E. (2020). Connected vehicle real-time traveler information messages for freeway speed harmonization under adverse weather conditions: Trajectory level analysis using driving simulator. *Accident Analysis & Prevention*, *146*, 105707. https://doi.org/10.1016/J.AAP.2020.105707
Zheng, L., Ismail, K., & Meng, X. (2014a). Freeway safety estimation using extreme value theory approaches: A comparative study. *Accident Analysis and Prevention*, *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. *Canadian Journal of Civil Engineering*, *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. *Journal of Transportation Safety and Security*, *7*(1), 76–90. https://doi.org/10.1080/19439962.2014.904029
Zheng, L., Ismail, K., Sayed, T., & Fatema, T. (2018). Bivariate extreme value modeling for road safety estimation. *Accident Analysis and Prevention*, *120*, 83–91. https://doi.org/10.1016/J.AAP.2018.08.004
Zheng, L., & Sayed, T. (2019a). Comparison of Traffic Conflict Indicators for Crash Estimation using Peak Over Threshold Approach. *Transportation Research Record*, *2673*(5), 493–502. https://doi.org/10.1177/0361198119841556
Zheng, L., & Sayed, T. (2019b). Bayesian hierarchical modeling of traffic conflict extremes for crash estimation: a non-stationary peak over threshold approach. *Analytic Methods in Accident Research*, *24*, e100106–e100106. https://doi.org/10.1016/j.amar.2019.100106
Zheng, L., Sayed, T., & Essa, M. (2019a). Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. *Accident Analysis and Prevention*, *123*, 314–323. https://doi.org/10.1016/j.aap.2018.12.007

Zheng, L., Sayed, T., & Essa, M. (2019b). Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. *Analytic Methods in Accident Research, 23*. https://doi.org/10.1016/J.AMAR.2019.100100

Zheng, L., Sayed, T., & Mannering, F. (2021). Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Analytic Methods in Accident Research*, *29*, 100142. https://doi.org/10.1016/J.AMAR.2020.100142

Zheng, L., Sayed, T., & Tageldin, A. (2018). Before-after safety analysis using extreme value theory: A case of left-turn bay extension. *Accident Analysis and Prevention*, *121*, 258–267. https://doi.org/10.1016/j.aap.2018.09.023

Zöller, I., Abendroth, B., & Bruder, R. (2019). Driver behaviour validity in driving simulators – Analysis of the moment of initiation of braking at urban intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *61*, 120–130. https://doi.org/10.1016/J.TRF.2017.09.008

List of Publications

Journal Articles:

1. **Pranab Kar,** Suvin P. Venthuruthiyil, and Mallikarjuna Chunchu (2023). "Non-stationary crash risk modelling of powered two-wheelers using extreme value analysis of surrogate crash events." Accident Analysis & Prevention, 183, 106973. doi: 10.1016/j.aap.2023.106973.
2. **Pranab Kar,** Suvin P. Venthuruthiyil, and Mallikarjuna Chunchu (2023). "Assessing the Crash Risk of Mixed Traffic on Multilane Rural Highways Using Proactive Safety Approach." Accident Analysis & Prevention, 188, 107099. doi: 10.1016/J.AAP.2023.107099.
3. **Pranab Kar,** Suvin P. Venthuruthiyil, and Mallikarjuna Chunchu. "Crash Risk of Estimation of Heavy Commercial Vehicles on Horizontal Curves in Mountainous Terrain Using Proactive Safety Method." Accident Analysis & Prevention (Under Review).

Conference Proceedings:

1. **Pranab Kar,** Suvin P. Venthuruthiyil, and Mallikarjuna Chunchu. "Proactive Safety Assessment of Sideswipe Crash Risk of Powered Two Wheelers on Multilane Rural Highways in Low- and Middle-Income Countries" Transportation Research Board 102nd Annual Meeting, 2023, Washington DC, United States.

2. **Pranab Kar,** and Mallikarjuna Chunchu. “Effect of steering on the sideswipe crash Risk of Powered Two Wheelers on Multilane Rural Highways for the Connected Autonomous Vehicles” International Conference on Traffic and Granular Flow, 2022, Delhi.