

**IDENTIFICATION OF FAILURE-CAUSED TRAFFIC CONFLICTS IN
TRACKING SYSTEMS: A GENERAL FRAMEWORK**

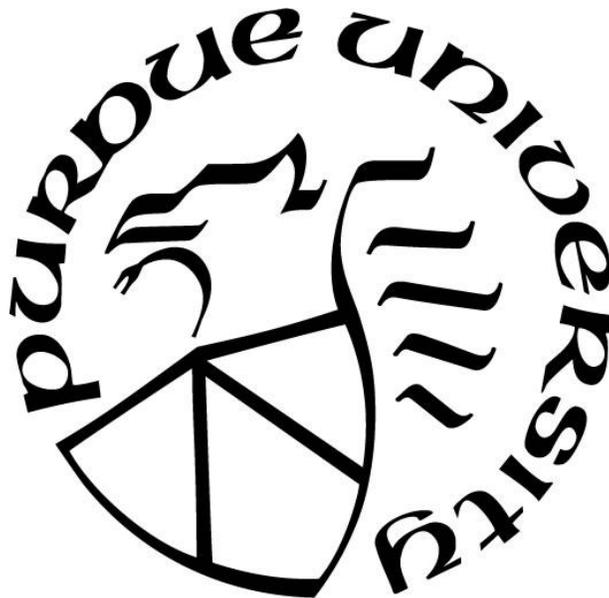
by
Cristhian Lizarazo

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Lyles School of Civil Engineering

West Lafayette, Indiana

December 2020

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Andrew Tarko, Chair

Lyles School of Civil Engineering

Dr. Darcy Bullock

Lyles School of Civil Engineering

Dr. Samuel Labi

Lyles School of Civil Engineering

Dr. Thomas Sellke

Department of Statistics

Approved by:

Dr. Dulcy Abraham

*For my beloved parents
and
Diana Maria,
my sunshine on cloudy days*

ACKNOWLEDGMENTS

First and foremost, I thank my advisor, Professor Andrew Tarko, for trusting me and allowing me to be a member of the Center for Road Safety team, first as an undergraduate research assistant and later as a doctoral student. Your exemplary commitment to excellence and your thoughtful guidance will continue to guide me as I embark on my career. I am very grateful to my other committee members as well: Professor Samuel Labi, Professor Darcy Bullock, and Professor Thomas Sellke. Your valuable support and advice have shed light on ways to improve the outcomes of my research.

I am thankful to my team members at the Center for Road Safety who assisted me greatly as well. Dr. Mario Romero and Mr. Vamsi Bandaru, you were always open to discussing the hurdles of my research, no matter how difficult they were. Dr. Thomas Hall and Dr. Qinzhong Hou, you supported me during our long work hours at Research Park; and I will always remember our late-night conversations about our dreams and hopes. Dr. Christos Gkartzonikas, you are a valued friend, and I know our time at Purdue is just the beginning of a lifelong friendship. Let us see if God, Destiny, or the Universe shows us the path to greater success!

Last but not least, I will be forever grateful to my family for their unconditional love and support. Mom and Dad, I cannot thank you enough for all your support and prayers during my long journey. Diana Maria, you are the love of my life. I could not have done this without you. I cannot thank you enough for being my family and home during this time. I hope we can continue to develop in this world together. I also am deeply grateful to Abuelitos, Tia Luz, and Tia Cecy for showing me that no matter how humble our origins, dreamers can achieve their goals if they persevere. This space is too limited to acknowledge all the individuals who made this dream possible. You all have a special place in my heart, and I will forever remember you.

TABLE OF CONTENTS

LIST OF TABLES	8
LIST OF FIGURES	9
ABSTRACT.....	11
1. INTRODUCTION	13
1.1 Background and Motivation	13
1.2 Research Problem Statement	15
1.3 Scope and Objectives	16
1.4 Dissertation Organization	17
2. TRAFFIC CONFLICTS AS SURROGATE MEASURES OF SAFETY	19
2.1 Background.....	19
2.2 Definition of Traffic Conflicts: An Evolving Concept.....	20
2.3 Validity of Traffic Conflicts as Surrogate Measures	21
2.4 Exceedance Models: Estimation Expected Number of Crashes using Traffic Conflicts..	23
2.5 Counterfactual Definition of Traffic Conflicts	23
2.6 Lomax-based Approach for Estimation of Crashes using Traffic Conflicts.....	25
2.7 Summary	31
3. IDENTIFICATION OF TRAFFIC CONFLICTS: STATE OF PRACTICE	32
3.1 Background.....	32
3.2 Technological Needs for Identification of Traffic Conflicts	33
3.2.1 Road-side Instrumentation: Tracking Systems.....	33
3.2.2 In-vehicle Instrumentation.....	35
3.3 Identification of Traffic Conflicts based on Proximity Measures	38
3.4 Summary	41
4. IDENTIFICATION OF REAR-END TRAFFIC CONFLICTS	42
4.1 Background.....	42
4.2 Data Collection	42
4.3 Detection of Traffic Conflicts Proximity Measures.....	43
4.4 Redefining Identification of Traffic Conflicts	46

4.5	Evaluation of Identified Conflicts.....	49
4.6	Summary.....	57
5.	IDENTIFICATION OF TRAFFIC CONFLICTS: VEHICLE-BICYCLE ENCOUNTERS	59
5.1	Background.....	59
5.2	Data Collection.....	59
5.3	Identification of Traffic Conflicts: Instantaneous Time to Collision.....	61
5.3.1	Calculating iTTC with the Assumption of Fixed Speeds and Directions.....	61
5.3.2	Calculating iTTC with the Assumption of Collision Path between Centroids.....	64
5.3.3	Estimation Expected Number of Crashes instantaneous Time-to-Collision (TTC) ..	65
5.4	Identification of Failure: Longitudinal and Lateral Decomposition of Response	72
5.5	Estimation of Proximity Measures Using Trajectories Without Evasion.....	74
5.6	Evaluation of Traffic Conflicts.....	75
5.7	Summary.....	77
6.	GENERALIZED FRAMEWORK FOR IDENTIFICATION OF TRAFFIC CONFLICTS	79
6.1	Background.....	79
6.2	Estimation of Potentially Dangerous Interactions	80
6.3	Identification of Evasion Using Decomposition of Response	83
6.4	Estimation of Hypothetical Collision Point via Evasion Free Trajectories	87
6.5	Similarity Measures between Trajectories.....	88
6.6	Estimation Time-to-Collision	90
6.7	Summary.....	90
7.	SEVERITY OF HYPOTHETICAL CRASH GIVEN CONFLICT	93
7.1	Background.....	93
7.2	Literature Review.....	94
7.3	Data Collection	96
7.4	Methodology	100
7.5	Model Results	104
7.5.1	Model Crash Reporting.....	104
7.5.2	Model Crash Severity	106
7.6	Summary.....	108
8.	CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS	114

8.1	Conclusions.....	114
8.2	Contributions.....	116
8.3	Limitations and Future Directions	118
	REFERENCES	120
	APPENDIX A. DESCRIPTION DATA REQUEST SHRP2 INITIATIVE	130
	APPENDIX B. ALGORITHM EXTRACTION POTENTIAL CONFLICTS.....	133
	APPENDIX C. DATA POSTPROCESSING SHRP2 INITIATIVE	140
	APPENDIX D. DIAGNOSTICS MULTINOMIAL LOGIT MODEL	141
	VITA.....	148

LIST OF TABLES

Table 4.1. Basic statistics for the three studied types of drivers in the SHRP2 population.....	50
Table 4.2. Basic statistics for the three types of drivers in random sample 1.4% trips	55
Table 4.3. Summary of results Lomax-based analysis for speed threshold 2.5 m/s	55
Table 4.4. Comparison of the population crash estimates and the conflict-based sample estimates expanded to the SHRP2 population	57
Table 5.1. Average number of conflicts per day all intersections InDeV Program.....	76
Table 7.1. Summary of crash analysis research publications	95
Table 7.2. Categories of covariates included in severity analysis	97
Table 7.3. Descriptive statistics dynamic variables severe crash	98
Table 7.4. Descriptive statistics dynamic variables police-reported crash	99
Table 7.5. Descriptive statistics dynamic variables minor collision.....	99
Table 7.6. Descriptive statistics crash reporting model	104
Table 7.7. Frequentist model crash reporting. Reference No-reported crash	105
Table 7.8. Bayesian model crash reporting. Reference No-reported crash	105
Table 7.9. Descriptive statistics severity model.....	107
Table 7.10. Multinomial logit model Frequentist method including intersection indicator	109
Table 7.11. Multinomial logit model Bayesian method including intersection indicator	110
Table 7.12. Multinomial logit model Frequentist method without intersection indicator	111
Table 7.13. Multinomial logit model Bayesian method without intersection indicator	112

LIST OF FIGURES

Figure 1.1 Pyramid of events proposed by Hydén (1987).....	13
Figure 2.1. Direct Graphical Model of crash events depicting conditional independence (Davis et al., 2011)	24
Figure 2.2. Representation of a rear-end conflict with $\tau(t)$ profile (lead car assumed stopped for easier illustration).....	26
Figure 2.3. The effect of long separation threshold on estimates of expected number of crashes	29
Figure 2.4. Separation thresholds versus expected number of crashes estimates.....	30
Figure 3.1. Experimental setup CAMP initiative (Smith et al., 2005).....	36
Figure 3.2. In-vehicle instrumentation SHRP2 Initiative (Campbell, 2012)	37
Figure 3.3. Instrumentation 5th-generation Waymo Driver (Waymo, 2020).....	38
Figure 4.1. Log-log curves for the test sample with the lane exit observations (Tarko, 2018)	44
Figure 4.2. Minimum instantaneous TTC observations in the test sample (Tarko, 2018)	45
Figure 4.3. Log-log curves after removing the lane exit events ($\tau_c = 2$ s) (Tarko, 2018).....	46
Figure 4.4 Example failure free braking with short separation	47
Figure 4.5 Example rear-end conflict in SHRP2 data marked (inside the oval shape)	48
Figure 4.6 Profiles expected crash estimates various speed thresholds (young male drivers)	52
Figure 4.7 Profiles expected crash estimates various speed thresholds (mature male drivers)	52
Figure 4.8 Profiles expected crash estimates various speed thresholds (mature female drivers). 53	53
Figure 4.9 AIC of the expected crash estimates profile models for proper thresholds under various threshold speeds – mature male drivers	53
Figure 4.10 Selection of the proper τ_c threshold based on the information-based performance measure AIC	54
Figure 4.11. Log-log curves for the three studied categories of driver.....	56
Figure 4.12. Expected number of crashes for different τ_c per category of driver	56
Figure 5.1. Data collection planning InDEV program (Johnsson et al., 2018).....	60
Figure 5.2. Data collection setup at intersections (Johnsson et al., 2018)	61
Figure 5.3. Conceptual representation vehicle-bicycle turning conflict	63
Figure 5.4. Relative reference system for estimation of iTTC	64
Figure 5.5. Conceptual representation of turning conflict with collision path between centroids	65

Figure 5.6. Camera location and intersection layout in Spain	66
Figure 5.7. Histogram iTTC relative method (left) and range method (right).....	67
Figure 5.8. Comparison minimum iTTC values relative and range methods.....	67
Figure 5.9. Discrepancy of trajectories with and without lateral evasion.....	68
Figure 5.10. Results for instantaneous TTC relative method	69
Figure 5.11. Results for instantaneous TTC range method	70
Figure 5.12. Log-log curves for instantaneous TTC relative method.....	71
Figure 5.13. Log-log curves for instantaneous TTC range method.....	72
Figure 5.14. Cumulative distribution longitudinal jerk	73
Figure 5.15. Cumulative distribution lateral jerk.....	74
Figure 5.16. Candidate conflict-free trajectories lateral evasion	75
Figure 5.17. Histogram complete estimation Time-to-Collision	76
Figure 6.1. Lateral evasive maneuver for a vehicle turning left in a traffic conflict	81
Figure 6.2. Relative reference system for estimation of iTTC	82
Figure 6.3. Change in point of hypothetical collision point due to evasive maneuver.....	83
Figure 6.4. Change-point-detection acceleration profile	84
Figure 6.5. Unsupervised methods change point detection (Aminikhanghahi & Cook, 2017)....	85
Figure 6.6. Estimation of hypotehtical collision using most similar evasion-free trajectory	88
Figure 6.7. Representation of a rear-end conflict instantaneous Time-to-Collision.....	90
Figure 6.8 Representation of turning conflict Time-to-Collision	91
Figure 7.1 Diagnostics plots Bayesian estimation parameter intersection indicator	108

ABSTRACT

Proactive evaluation of road safety is one of the most important objectives of transportation engineers. While current practice typically relies on crash-based analysis after the fact to diagnose safety problems and provide corrective countermeasures on roads, surrogate measures of safety are emerging as a complementary evaluation that can allow engineers to proactively respond to safety issues. These surrogate measures attempt to address the primary limitations of crash data, which include underreporting, lack of reliable insight into the events leading to the crash, and long data collection times.

Traffic conflicts are one of the most widely adopted surrogate measures of safety because they meet the following two conditions for crash surrogacy: (1) they are non-crash events that can be physically related in a predictable and reliable way to crashes, and (2) there is a potential for bridging crash frequency and severity with traffic conflicts. However, three primary issues were identified in the literature that need to be resolved for the practical application of conflicts: (1) the lack of consistency in the definition of traffic conflict, (2) the predictive validity from such events, and (3) the adequacy of traffic conflict observations.

Tarko (2018) developed a theoretical framework in response to the first two issues and defined traffic conflicts using counterfactual theory as events where the lack of timely responses from drivers or road users can produce crashes if there is no evasive action. The author further introduced a failure-based definition to emphasize conflicts as an undesirable condition that needs to be corrected to avoid a crash. In this case, the probability of a crash, given failure, depends on the response delay. The distribution of this delay is adjusted, and the probability is estimated using the fitted distribution. As this formal theory addresses the first two issues, a complete framework for the proper identification of conflicts needs to be investigated in line with the failure mechanism proposed in this theory.

The objective of this dissertation, in response to the third issue, is to provide a generalized framework for proper identification of traffic conflicts by considering the failure-based definition of traffic conflicts. The framework introduced in this dissertation is built upon an empirical evaluation of the methods applied to identify traffic conflicts from naturalistic driving studies and video-based tracking systems. This dissertation aimed to prove the practicality of the framework

for proactive safety evaluation using emerging technologies from in-vehicle and roadside instrumentation.

Two conditions must be met to properly claim observed traffic events as traffic conflicts: (1) analysis of longitudinal and lateral acceleration profiles for identification of response due to failure and (2) estimation of the time-to-collision as the period between the end of the evasion and the hypothetical collision. Extrapolating user behavior in the counterfactual scenario of no evasion is applied for identifying the hypothetical collision point.

The results from the SHRP2 study were particularly encouraging, where the appropriate identification of traffic conflicts resulted in the estimation of an expected number of crashes similar to the number reported in the study. The results also met the theoretical postulates including stabilization of the estimated crashes at lower proximity values and Lomax-distributed response delays. In terms of area-wide tracking systems, the framework was successful in identifying and removing failure-free encounters from the In-Depth understanding of accident causation for Vulnerable road users (InDeV) program.

This dissertation also extended the application of traffic conflicts technique by considering estimation of the severity of a hypothetical crash given that a conflict occurs. This component is important in order for conflicts to resemble the practical applications of crashes, including the diagnostics of hazardous locations and evaluating the effectiveness of the countermeasures. Countermeasures should not only reduce the number of conflicts but also the risk of crash given the conflict. Severity analysis identifies the environmental, road, driver, and pre-crash conditions that increase the likelihood of severe impacts. Using dynamic characterization of crash events, this dissertation structured a probability model to evaluate crash reporting and its associated severity. Multinomial logistic models were applied in the estimation; and quasi-complete separation in logistic regression was addressed by providing a Bayesian estimation of these models.

1. INTRODUCTION

1.1 Background and Motivation

The reliance of road safety management practice on crash data produces a corrective decision-making process rather than a proactive method that proposes safety-related treatments before crashes occur. Although the value of utilizing crash events in estimating safety is irrefutable (Lamprey et al., 2005; Lamprey et al., 2010), crash data have limitations, which include low quality, lack of reliable insight into the events leading to a crash, and lengthy data collection times (Lord & Mannering, 2010; Tarko et al., 2009). Unlike crash-driven approaches to safety management, surrogate measures of safety offer a much faster approach in which the limitations of crash data do not preclude confirming and solving safety problems.

In fact, the use of surrogate measures of safety is not a competing alternative to crash-based approaches, but rather a complementary approach that helps researchers better understand the cause-effect relationship between road conditions and crash occurrence. Since a surrogate event precedes a crash, there is an obvious causal relationship between the two, as illustrated by the pyramid of events in Figure 1.1 (Hydén, 1987).

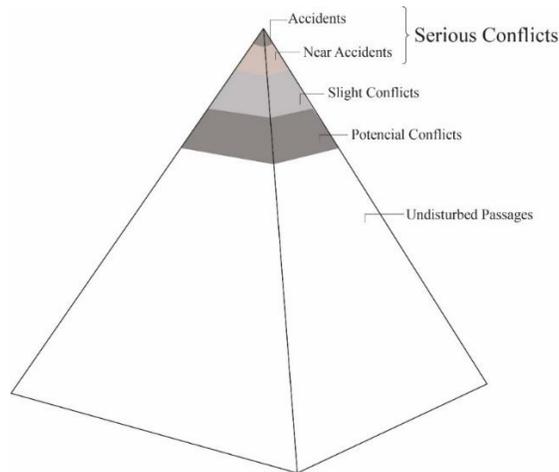


Figure 1.1 Pyramid of events proposed by Hydén (1987).

According to Tarko et al. (2009), in order to be applicable in transportation safety, a surrogate measure must meet the following two conditions: (1) a surrogate measure is based on an observable non-crash event that is physically related, in a predictable and reliable way, to crashes; and (2) there should be a bridge method for estimating crash frequency or severity using non-crash events. Traffic conflicts meet these two conditions for use. Furthermore, Tarko (2018) added another condition to traffic conflicts to strengthen the etiological connection between conflicts and crashes: *a failure of some kind to which a road user responds with a certain delay.*

A traffic conflict is among the most prevalent surrogate measures of safety since, if it is severe (proximity to a crash), it shares common factors with crashes (Mullakkal-Babu et al. , 2017; Zheng et al., 2014). The definition of traffic conflict is still an area for discussion among researchers due to its relevance in identifying events with real values for measuring safety. The operational definition of traffic conflicts can be categorized as either of the following two approaches to identification: (1) approaches based on evasive action and (2) approaches based on temporal or spatial proximity.

According to Parker & Zegeer (1989), the definition of traffic conflicts using the evasive action approach is “...an event involving two or more road users, in which the action of one user causes the other user to make an evasive maneuver to avoid a collision.” Hence, the natures of crashes and conflicts are similar, with the one difference being that, in conflicts, the evasive maneuver is successful. Confidence in this definition relies on its method for identifying traffic conflicts (i.e., conflicts can be identified with certain reliability by trained human observers). This advantage has allowed the evasive action identification method to be widely utilized in the past. However, this definition has multiple limitations, including the lack of distinction between aggressive behavior and a real evasive maneuver that weakens causality (Wu & Jovanis, 2013).

Accurate identification of traffic conflicts using proximity measures has been made possible by the introduction of new sensing technologies. According to Amundsen & Hyden (1977), the general definition of a temporal or spatial proximity-based traffic conflict can be described as “... an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged.” This definition implies that a traffic conflict is identified based on how close in time or space the road users were at the time of the event. A crash is reported when this separation has a zero or negative value.

Naturalistic driving studies and roadside instrumentation provide unique sources of information for identifying these “close call” events. Instrumented vehicles and machine vision algorithms allow accurate detection and tracking of road users and, consequently, extraction of proximity measures for identification of traffic conflicts. Observing these conditions can help researchers evaluate traffic conflicts as appropriate surrogate measures of safety. Traffic conflicts are considered valid surrogates if, by using a methodology to bridge conflicts with crashes, researchers are able to measure a conflict’s safety in terms of the frequency and severity of crashes.

The process of measuring safety with traffic conflicts is still an active area of research; and robust statistical methods and computational approaches help bridge estimation of crashes using traffic conflicts. Based on the work of Davis et al. (2011) and Pearl (2000), Tarko devised a practical method by considering the behavior of the road users in a risky situation that has two alternative outcomes: a crash or a conflict. In this approach, the probability of a crash, with a given failure, depends on the delay in driver response. The distribution of this delay is adjusted, and the probability is estimated using the fitted distribution. Tarko (2019) presented a comprehensive overview of the method with examples that demonstrate its validity.

Although a formal bridging method between conflicts and crashes has been proposed in the literature, a complete framework for proper identification of traffic conflicts by considering the failure mechanism proposed in this theory, has yet to be addressed. This dissertation addresses this missing link (i.e., a practical framework for identifying traffic conflicts with safety values using emerging instrumentation). The safety value of an event is measured in terms of the probability of observing a crash by the delay in the response. As discussed in detail later in this dissertation, appropriate identification of traffic conflicts will be provided by application of the Lomax-based method, described in Tarko (2019), for the proactive estimation of safety. The proposed framework aims to allow practitioners to propose countermeasures without the associated externalities from crash-driven approaches.

1.2 Research Problem Statement

Lack of data for extraction of proximity-based measures has been a common constraint for traffic conflict analysis. This limitation has restricted the application of methodologies to identify traffic conflicts using software-based simulations with questionable results in terms of the fidelity from these algorithms in replicating real-world safety (Archer, 2005; Huang et al., 2010). Multiple

attempts of extracting safety information from dangerous events as proxies for crashes with the help of vehicle instrumentation and machine vision algorithms provide useful data sources. Instrumentation provides additional insights into the failure mechanisms in conflicts, thereby providing a superior representation of causality.

The generalized framework for identification of traffic conflicts applicable to in-vehicle and area-wide instrumentation consistently is based on the theory introduced in Tarko (2018). It intends to better target the data collection efforts and identification process from researchers with the aim of evaluating safety via surrogate measures. The framework was evaluated using in-vehicle instrumentation from the Strategic Highway Research Program SHRP2 (Campbell, 2012). Data from the In-Depth understanding of accident causation for Vulnerable road users program (InDeV) was also used to prove the concept when applied to area-wide detection and tracking systems. This research supplements the theory described in Tarko (2019) bridging conflicts and crashes, in which the definition of a traffic conflict includes a failure of some sort that, if not corrected timely, produces a crash. This dissertation establishes additional characteristics indicating failure and estimation of proximity measures to better calculate the risk of collision.

A hypothetical severity of crashes in conflicting conditions was an additional component included in this analysis. The benefits of estimating severity in hypothetical crashes are twofold: (1) identification of events with more severe outcomes similar to crashes that could be reduced using countermeasures and (2) narrowing the traffic encounters to those which are safety-relevant with hypothetical collisions that are sufficiently severe. As discussed in Tarko (2019), the low perception of risk at low speed may lead to driver acceptance of lower separation values or even mild collisions since the outcome may be negligible. Hence, a warrant of serious outcome should be considered in the estimation method to analyze the risk perception of drivers and crash reporting.

1.3 Scope and Objectives

The primary objective of this dissertation was to develop a framework for identification of traffic conflicts in a general case. This framework was evaluated and tested using in-vehicle instrumentation and area-wide detection systems to show its applicability in emerging tracking technologies. Controlled environments from naturalistic driving studies provided unique opportunities to evaluate the identified conflicts by comparing the estimated expected number of crashes with the reported crashes. In addition, statistical methodologies were applied for

parametric estimation of hypothetical crash severity given a conflict. This framework provides a prospective method to better utilize existing data sources and evaluate road safety relying on traffic conflicts rather than crashes. The specific objectives were as follows:

1. Explore existing methodologies for the identification of traffic conflicts with an emphasis on methods to identify failure-based events.
2. Develop a methodology and computationally efficient algorithm for identification and extraction of rear-end traffic conflicts from in-vehicle instrumentation.
3. Investigate the applicability of the method in conflicts between vehicles and bicycles and evaluate the intrinsic characteristics of bicycles in conflict events compared to motorized road users.
4. Summarize and propose a complete framework for the identification of traffic conflicts using in-vehicle and area-wide tracking systems.
5. Determine the suitability of the framework in line with the theory introduced by Tarko (2019) by evaluating the conflicts using diagnostics from this theoretical background.
6. Investigate the feasibility of predicting the severity of a hypothetical crash using a time-dependent characterization of pre-crash events.

1.4 Dissertation Organization

This dissertation is organized as follows. Chapter 2 includes a comprehensive discussion of traffic conflicts, evolution of the concept, and specific characteristic that makes them valid surrogate measures of safety as well as the methodologies applied to evaluate the predictive validity of these events with special emphasis on the Lomax-based theoretical background developed by Tarko (2018). Chapter 3 provides a general introduction of the technological needs for identification of traffic conflicts and includes a general overview of the required information from area-wide tracking systems, as well as, vehicle instrumentation. Chapters 4 and 5 provide empirical identification of traffic conflicts and discuss the issues from instrumentation and the additional characteristics to identify failure and estimation of proximity measures. In Chapter 4 the proposed framework and method are also applied for identification of rear-end traffic conflicts using in-vehicle instrumentation; and Chapter 5 also expands the framework to redefine conflicts from vehicle-bicycle turning encounters at intersections. Chapter 6 presents the results from the empirical evaluation condensed in a generalized framework replicable for emerging tracking

systems. Chapter 7 describes the estimation of the severity models using information from pre-crash time series characterization of the events. Chapter 8 concludes this dissertation with the major findings, the limitations of the proposed framework, and planned future work.

2. TRAFFIC CONFLICTS AS SURROGATE MEASURES OF SAFETY

2.1 Background

Traffic conflicts are among the most prominent surrogate measures of safety cited in the literature. Development of traffic conflict technique for more than 70 years has provided mature methods that has allowed its passing from the transitional research period to the practical and application stages (Zheng et al., 2014). In general, traffic conflicts enable the evaluation of safety in a wide range of road and traffic conditions. Initial attempts to observe traffic conflicts with informal methodologies were found in McFarland (1956) and Forbes (1957), both of which collected data on near misses as emergency situations or critical incidents that could have resulted in accidents. McFarland (1956) observed near misses with bus drivers and trucks while Forbes (1957) extended the methodology to “accidents that nearly happened” from 200 drivers. These early initiatives corresponded to rather informal observational studies. A formal methodology to observe traffic conflicts was not established until Perkins and Harris in 1968. They provided a standard methodology for extraction of conflicts by human observers. For analysis of their results, the authors did not implement advanced statistical methods but rather observed the correlation between crashes and conflicts at the analyzed intersections.

During the first Traffic Conflict Technique Workshop in 1977, Amundsen and Hyden not only emphasized a correlation between conflicts and crashes, but also postulated the potential applications of traffic conflicts. Similar to crashes, these applications included identification and ranking of hazardous locations for improvement, determining the causes of hazards, and evaluating the effectiveness of countermeasures. These applications were especially appealing where crash data were unavailable or unreliable. An extended discussion in Williams (1981) stated that most of these applications were overly exaggerated. The document contended them by emphasizing the need of a conceptually sound definition and an objective methodology for extraction of traffic conflicts. As additional research was conducted, more criticism emerged in the traditional traffic conflict technique (Cooper, 1977; Glennon et al., 1977).

Three primary issues were identified in the literature for extending practical applications of traffic conflicts: (1) consistency in the definition of a traffic conflict, (2) the predictive validity of a traffic conflict, and (3) the reliability of traffic conflict measurement. The following sections

provide an extensive review of the literature addressing the first two issues while the contribution of the current dissertation is based on the third component.

2.2 Definition of Traffic Conflicts: An Evolving Concept

The definition of traffic conflicts is an evolving concept. The first formal definition for traffic conflicts was stated in the General Motors study by Perking and Harris in 1968. In their original work, Perkins and Harris defined traffic conflicts as evasive actions (brake light indication or a lane change) by drivers or a violation of the uniform traffic code. Specific definitions were developed to describe road-user maneuvers which resulted in weave, cross-traffic, rear-end, and red-light violation conflicts. These definitions allowed extraction of conflicts by human observers relying on the conducted evasion. Many of the subsequent conflict studies followed the same approach with a more detailed classification of evasive actions.

The evasion-based definition of traffic conflicts was constantly criticized in the early 1970s. This definition had some practical and logical issues (Chin & Quek, 1997; Cooper, 1977; Williams, 1981). Practical issues were found with the exhaustive list of possible evasive actions associated with conflicts. Moreover, not all the specified driver actions could be considered as evasive action in nature. Logical issues emerged by considering the argument that evasive action is always applied to avoid a crash. These conditions do not always apply.

A consensus definition was reached in the first workshop on Traffic Conflict Techniques in 1977. In this workshop, a conflict was defined as "...an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remained unchanged"(Amundsen & Hyden, 1977). This workshop was the first initiative towards providing a formal definition. However, criticism emerged towards the unclear distinction between a conflict and a non-conflict situation in practice. The conditions of the definition allowed a wide range of interpretation, especially concerning what is to be considered observable and at a sufficient level of risk (Chin & Quek, 1997).

Chin and Quek (1997) proposed that one way to overcome the ill-defined problem is to concentrate the analysis using only serious cases of conflicts. According to the authors, restricting conflict data to only instances of serious critical encounters thus would provide more common characteristics between conflicts and crashes. In line with these findings, more objective

methodologies to define conflicts have been adopted, which focus on objective assessments of conflict based on, in essence, the shortest separation between interactive road users.

Davis et al. (2011) offered a counterfactual definition of a traffic conflict based on Amundsen and Hyden (1977). The authors developed a causal model considering two characteristics of this definition: (1) each event is decomposed into three stages: an initial condition, the action adopted by the one road user, and a collision-related outcome; and (2) an event is defined as a conflict only if it passes a counterfactual test: if the movements had remained unchanged, then a collision likely would have resulted. However, Guettinger (1982) raised the following theoretical ambiguity question in the definition phase: Is a conflict to be a potential crash or a potential crash that did not, in fact, result in a crash? Based on Davis et al. (2011), theoretically, treating conflicts as potential crashes is useful; but in reality, the researcher must be aware that, almost always, only potential crashes that did not result in crashes are observed in a conflict study. Hence, the authors emphasized a conflict as a partial specification of a complete event, where the conflict together with the actions involved in the parties determines whether or not a collision occurs.

Tarko (2018) extended this counterfactual definition of traffic conflicts as events with a lack of timely response from drivers or road users that produce crashes if there is no evasive action. The author imposed a failure-based definition to emphasize conflicts as an undesirable condition that needs to be corrected to avoid a crash. If there is a timely response, the crash can be avoided. Although this modification of the traffic conflict definition increases the observation period and complexity for identification, it also simplifies the theory and strengthens the crash causality. A failure-based mechanism for identification of traffic conflicts is implemented in this theoretical definition of conflicts.

2.3 Validity of Traffic Conflicts as Surrogate Measures

Some of the published work attempted to confirm the validity of traffic conflicts for safety evaluation by detecting a significant statistical correlation between conflicts and crashes. This validation method was discussed extensively in the literature since correlation, as some authors believe, supports predicting the crash frequency and severity (Archer, 2005; El-Basyouny & Sayed, 2013; Sayed & Zein, 1999). Other authors claimed that a large portion of the crash-conflict correlation is introduced by exposure, and the correlation-based methods should be applied with caution. The mixed results obtained in these studies can be attributed to the following: (1) the

shortcomings of detecting traffic conflicts, (2) under-reporting of crashes, (3) uncontrolled heterogeneity of crashes and conflicts, (4) different observation periods for traffic conflicts and crashes (Zheng et al., 2014).

The most appealing validation methods are those that attempt to validate traffic conflicts without the use of crash data. These methods respond to the need for rapid safety estimation during a period of fast changes in safety. An initial predictive relation between conflicts and crashes was proposed by Hauer (1982), who understood the expectation of crashes as the product of the number of observed conflicts and a ratio that allows translating conflicts into crashes based on the expression:

$$\lambda = \pi \cdot c \quad (2.1)$$

Where λ represents the expected number of crashes, π is the crash-to-conflict ratio and c is the number of observed conflicts. Hauer emphasized the necessity of the crash-to-conflict being stable across the entities where traffic conflicts were observed in expression (2.1). The concept was later extended in Hauer and Garder (1986), where the authors proposed different crash-to-conflict ratios depending on the severity of the conflict. This concept was interpreted as the probability of an event i being a crash or not depending on the probability π_i . Hence, crash-to-conflict ratios may vary based on the conflict's severity as:

$$\lambda = \sum_i \pi_i c_i \quad (2.2)$$

Equation (2.2) assumes a stable probability or crash-to conflict ratios for conflicts of severity i . Hence, a method to identify conflicts that belong to each one of these categories should be proposed. A possible alternative corresponds to a continuous function characterizing the probability of the crash depending on the severity of the conflict (Hu et al., 2004; Saunier & Sayed, 2008):

$$P(\text{crash}|TTC) = \exp \left[- \left(\frac{TTC}{\sigma} \right)^2 \right] \quad (2.3)$$

Where the probability of crash depends on the severity of the conflicts characterized based on the separation measure time to collision (TTC). An alternative approach is to model crashes and conflicts in a continuous dimension, with crashes being at the extreme of the events. These models are referred to as exceedance models.

2.4 Exceedance Models: Estimation Expected Number of Crashes using Traffic Conflicts

An alternative approach to estimate the probability of crash in a conflicting scenario is applying extreme value theory. The methodology, in this case, predicts less likely events (crashes) with more frequent events (conflicts) by extrapolating the tail of an extreme value distribution (Campbell et al., 1996; Songchitruksa & Tarko, 2006) or an exceedance distribution (Tarko, 2012). These extreme value-based methods model crash risk and crash frequency using extrapolation through mathematical limits and finite-level approximations (Smith, 2003). Songchitruksa and Tarko (2006) applied a time-based sampling scheme using the *r-largest* order statistics observed in 15-min intervals. Let N_t be the number of 15-min blocks having similar conditions as period t during the entire period of interest T . The estimated frequency of collisions for a period T is estimated as:

$$\widehat{C}_T = \frac{N_T}{N_t} \sum_{i=1}^{N_t} \widehat{R}_i \quad (2.4)$$

Where \widehat{R}_i , the estimated risk for the *ith* block is obtained by the expression:

$$R_i = \Pr\{Z_i \geq 0\} = 1 - G_i(0) \quad (2.5)$$

Z_i is the maximum negated post-encroachment time (*PET*) of the block and $G_i(\cdot)$ is the extreme value distribution function with fitted parameters. Validation of the experiments proved acceptable estimates of crash frequencies with wide confidence intervals representing low efficiency. In general, randomness was a governing factor in the lack of efficiency with a poor causality in the experiments. Following this line of research, other authors have introduced variations from this approach including bivariate modifications of extreme value distributions (Zheng et al., 2018; Zheng et al., 2019). The low efficiency in these experiments might be attributable to the poor characterization of the pre-conflict event and causal relation between conflicts and crashes.

2.5 Counterfactual Definition of Traffic Conflicts

Rather than extrapolating distribution tails into unobserved events without considering the crash-generation mechanism, another approach considers a crash as a counterfactual outcome of an observed traffic conflict. Davis et al. (2011) described a counterfactual definition of a traffic

conflict understood as “...an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged” (Amundsen & Hyden, 1977). The authors developed a causal model considering two characteristics of this definition. First, decomposing each event into three stages: (1) an initial condition, (2) the action adopted by the one road user, and (3) a collision-related outcome. The second characteristic states a counterfactual test: the non-zero probability of a crash if the driver does not respond. Applying a causal representation as shown in the direct graphical model of Figure 2.1, Davis et. al (2011) included Brill’s rear-end collision model to estimate the probability of observing a crash in a conflict event. In this graph, node u represents values describing initiating conditions, node x denotes the variable characterizing the avoidance action, and node y represents the crash-related outcome, which depends on u and x .

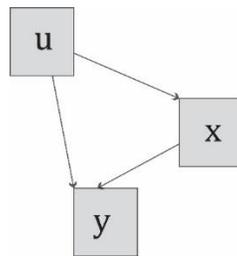


Figure 2.1. Direct Graphical Model of crash events depicting conditional independence (Davis et al., 2011)

Tarko (2018) simplified the method and proposed a counterfactual definition of traffic conflicts as “events with a lack of timely response from drivers or road users that produce crashes if there is no evasive action.” Although this modification of the traffic conflict definition increases the observation period and complexity for identification, it also simplifies the theory and strengthens the crash causality. Based on plausible assumptions, the author demonstrated that a delay in response to failure follows the Lomax distribution – a special case of exceedance distributions. The probability of a crash conditioned on conflict is estimated by fitting Lomax distribution to the delay in these responses. A comprehensive description is provided in the next section of this chapter.

2.6 Lomax-based Approach for Estimation of Crashes using Traffic Conflicts

A collision occurs when the separation between two road users is zero. Hence, a short spatial or temporal separation between two road users with conflicting trajectories represents an undesirable situation that needs to be corrected to avoid a crash. In general, a traffic interaction can be classified as a conflict if the separation between road users is too small to be acceptable, indicating failure, while the speeds warrant a sufficiently serious outcome. The literature reports multiple separation measures with extensive use including TTC or post-encroachment time. Temporal measures of crash nearness are applied in this dissertation for the identification of traffic conflicts. In this case, the collision point is hypothesized and the remaining time to collision is approximated based on what is observed or assumed. Specifically, if one assumes an approximately constant speed during the remaining TTC, then the ratio of distance to the collision point and the current speed, called instantaneous time-to-collision (iTTC) τ , is a satisfactory measure of crash nearness (Tarko, 2019).

Consider the car-following event shown in Figure 2.2. At any time t , the distance between the two vehicles is $D(t)$ and the relative speed is $\Delta v(t)$. The iTTC $\tau(t)$ at time t is:

$$\tau(t) = \frac{D(t)}{\Delta v(t)} \quad (2.6)$$

Let us select a threshold τ_c between two vehicles sufficiently short to be unacceptable to the following drivers. Consequently, instantaneous TTC τ smaller than τ_c must be caused by the driver's error or another type of failure. Once τ falls below τ_c , braking may start at any moment to avoid the collision. Indeed, the expected braking maneuver starts after some delay. Shortly after the beginning of braking, the instantaneous TTC reaches its lowest value τ_m and starts increasing. The risk of collision is eliminated at the moment when $\tau = \tau_m$. The response delay x includes the time before braking and the time taken to remove the hazard. The delay time x may be approximated with the reduction in threshold τ_c calculated as $(\tau_c - \tau_m)$.

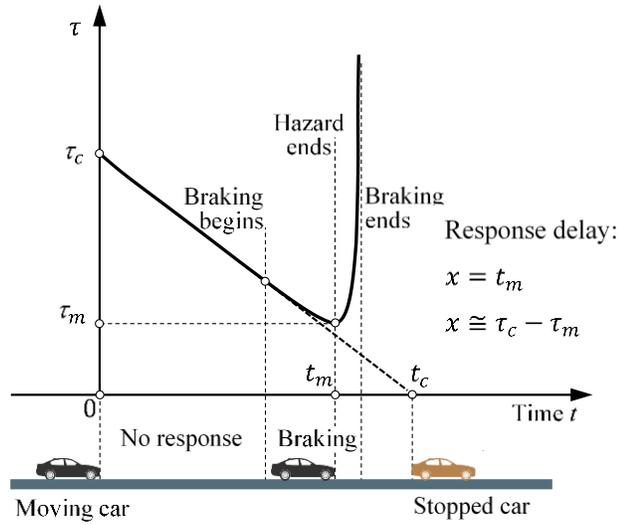


Figure 2.2. Representation of a rear-end conflict with $\tau(t)$ profile (lead car assumed stopped for easier illustration)

Under the assumption of an unintentional failure and the situation is temporarily out of the driver's control, the response delay x may be assumed independent of current τ as long as it is lower than τ_c . Consequently, x follows the exponential distribution density function, and the probability of crash ($x > \tau_c$) can be estimated with a simple survival model (Allison, 2010):

$$P(C|N, \tau_c) = e^{-r(\tau_c)} \quad (2.7)$$

where:

$P(C|N, \tau_c)$ is the probability of crash given no response (N) at time when $\tau = \tau_c$,

r is the response rate (1/s), which represents the propensity to respond, and

τ_c is the threshold of τ too short to be acceptable.

Equation (2.7) represents a homogenous case where the propensity to response r is fixed. However, the propensity to respond is expected to be different for different road users, roads, traffic, and other conditions. To account for the expected but unexplained heterogeneity, r is assumed to vary across incidents according to gamma distribution $f(r)$. Gamma heterogeneity assumes that the crash frequencies on various roads justifies the Negative Binomial widely used to model crash counts (Washington et al., 2011). The assumption of Gamma heterogeneity among response rates r leads to the Lomax-distributed response delays x in failure-caused traffic conflicts.

Parameters k and θ of the Lomax distribution may be estimated with response delays $x = \tau_c - \tau_m$ observed in multiple traffic conflicts. Although in the general case of Lomax parameter k and θ may take negative values, the application to traffic conflicts prompts for restricting these parameters to the non-negative range. Thus, the cumulative distribution of response delay x is:

$$F(x) = 1 - (1 + \theta \cdot x)^{-k} \quad (2.8)$$

where $\theta > 0$ and $k > 0$. A crash occurs given conflict with threshold τ_c if response x is too long: $x > t_c$. The corresponding probability of a crash is (Tarko, 2018; Tarko, 2019):

$$P(C|N, \tau_c) = (1 + \theta \tau_c)^{-k} \quad (2.9)$$

where:

τ_c is the threshold not acceptable by drivers,

k is the shape parameter, and

θ is the scale parameter.

Following a similar approach, covariates can be included through a link function of parameter k in the distribution, as follows: Let x_1, x_2, x_3, \dots be iid exceedances with Lomax distribution:

$$f(y) = \begin{cases} k\theta(1 + \theta x)^{-k} & \text{if } \theta > 0 \\ 1 - e^{-rx} & \text{if } \theta = 0 \end{cases} \quad (2.10)$$

Where k and θ are the shape and scale parameters, respectively. The likelihood function is:

$$L(k, \theta) = \prod_{i=1}^n (k_i \theta) (1 + \theta x_i)^{-k} \quad (2.11)$$

Then, the log-likelihood function is estimated as:

$$l(k) = \sum_{i=1}^n \log(k_i \theta) - k_i \log(1 + \theta x_i) \quad (2.12)$$

The method of maximum likelihood is explored as the benchmark method to estimate the parameters of the Lomax parameters. The log-likelihood is characterized based on the expression:

$$l^*(\theta, k) = n \log(k\theta) + (-k - 1) \sum_{i=1}^n \log(1 - \theta X_i) \quad (2.13)$$

When maximizing the likelihood function, the parameters k_{MLE} and θ_{MLE} are obtained. As Zhang and Stephens (2009) indicated, the numerical solution of the maximum likelihood in some cases is complex when no local maximum is found. Other methods can be explored including the method of moments (MOM) estimator for θ and k (Hosking & Wallis, 1987). The parameter estimates are:

$$\hat{k}_{MOM} = \left(\frac{\bar{X}^2}{S^2} - 1 \right) / 2; \quad \hat{\theta}_{MOM} = \hat{k}_{MOM} \bar{X} \left(\frac{\bar{X}^2}{S^2} + 1 \right) / 2 \quad (2.14)$$

The Single Parameter Estimate (SPE) method is proposed for estimation of the Lomax parameters (Tarko, 2018). The SPE method focuses on estimating one parameter k rather than both parameters. Considering the local insensitivity of the Generalized Pareto, log-likelihood to its parameters in the vicinity of the sample maximum log-likelihood, multiple pairs of θ and k in the vicinity of the solution tend to deliver estimates of $P(C|N) = 1 - F_x(\tau_c)$ that are within the acceptable estimation error. The method involves selecting a reasonable value of parameter $\theta = 1/\tau_c$ and estimating k using a robust least-square solution to fit k to the log-log line, $\ln(1 - F(x_i)) = -k \ln(1 + x_i/\tau_c)$:

$$k = \frac{-\sum_{i=1}^n \ln\left(1 - \frac{i-0.5}{n}\right) \ln\left(1 + \frac{x_i}{\tau_c}\right)}{\sum_{i=1}^n \left[\ln\left(1 + \frac{x_i}{\tau_c}\right)\right]^2} \quad (2.15)$$

where $(i - 0.5)/n$ approximates $F(x_i)$ corresponding to the observed n response delays sorted from the lowest x_1 to the highest x_n .

The linear relationship between $\ln(1 - F(x_i))$ and $\ln(1 + x_i/\tau_c)$ is expected if the observed response delays x are Lomax distributed. This also provides a convenient check if the assumed threshold τ_c is sufficiently low. Selection of an excessive value of τ_c leads to an overestimation of the number of crashes by including interactions that are aggressive behaviors rather than failure-generated. A sufficiently low and proper τ_c is determined based on the convergence of the estimated expected number of crashes to a stable estimate when the estimation is repeated for a sequence of decreasing threshold values τ_c . The parameter k estimated with assumed parameter $\theta = 1/\tau_c$ simplifies Equation (2.9):

$$P(C|N, \tau_c) = 2^{-k} \quad (2.16)$$

Crashes are the outcomes of multiple independent trials (conflicts) with the probability of a crash in each trial obtained with Equation (2.16). Hence, the expected number of crashes in the period with observed n conflicts is:

$$Q_c = n \cdot 2^{-k} \quad (2.17)$$

where:

Q_c is the estimated number of crashes expected during the observation period.

n is the number of traffic conflicts observed during the observation period.

k is the Lomax distribution parameter calculated with Equation (2.15) from the observed response delays

Selection of threshold τ_c is critical for correct estimation of the expected number of crashes. Tarko (2019) concluded that based on the proposed theory and analytical proof, there exists a proper crash nearness threshold such that any threshold equal to or smaller than the proper threshold provides an unbiased estimate of the expected number of crashes (see Figure 2.3). The convergence occurs through compensation of the reduction in the number of claimed traffic conflicts with the increase of the crash probability when the crash nearness threshold is reduced.

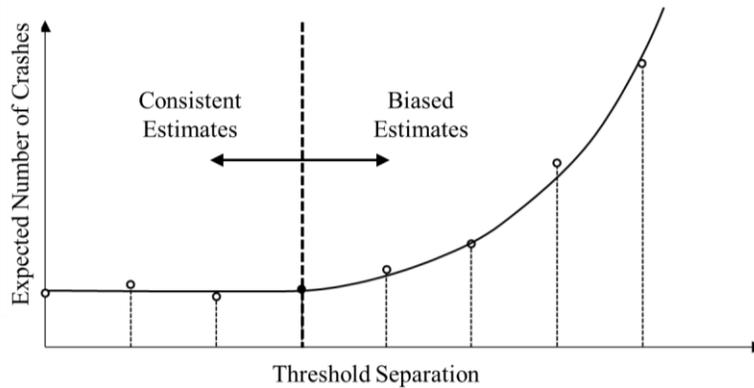


Figure 2.3. The effect of long separation threshold on estimates of expected number of crashes

Based on the described method, the selection of a conservatively short separation threshold guarantees consistent estimates. However, it may lead to the underutilization of the data by rejecting useful events. Tarko (2019) proposed a method relying on the Akaike Information Criterion to estimate a threshold separating biased and unbiased crash estimates. Let us consider m estimates of the expected number of crashes c based on u threshold separations τ_i (see Figure

2.4). The objective of the method is to identify the largest proper threshold that could identify the sequence of stable crash estimates.

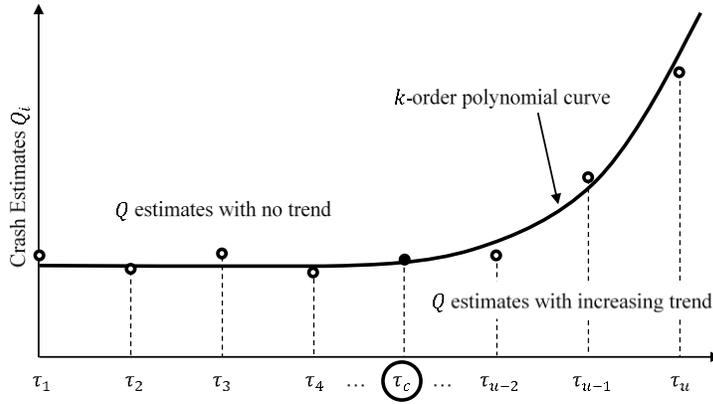


Figure 2.4. Separation thresholds versus expected number of crashes estimates

The search for the proper separation threshold can be implemented by fitting regression models for each candidate threshold τ_p , being $p = 2 \dots (u - 2)$. A threshold τ_c is proper if it separates the stable crash estimates on the left from the increasing trend of estimates on the right. To reduce the dependency of consecutive estimates Q_i at different separation thresholds, the models are fit to the differentials; $y_i = Q_i - Q_{i-1}, i = 2, \dots, u$. The model of the differentials on the left of the tested candidate threshold is $y_i = \varepsilon_i$ with the error term normally distributed and with an expected value of zero. The model of the differentials on the right is a polynomial function of k order:

$$f(\tau_i) = \beta_1(\tau_i - \tau_c) + \beta_2(\tau_i - \tau_c)^2 \dots + \beta_k(\tau_i - \tau_c)^k + \varepsilon_i \quad (2.18)$$

The model parameters β_1, \dots, β_k for a current candidate threshold T_c are estimated by setting them at values that minimize the sum of squares:

$$SS = \sum_{i=2}^c y_i^2 + \sum_{i=c+1}^u [y_i - \beta_1(\tau_i - \tau_c) - \beta_2(\tau_i - \tau_c)^2 \dots - \beta_k(\tau_i - \tau_c)^k]^2 \quad (2.19)$$

Value k defines the order of the fitted polynomial for the ascending trend on the right. Multiple-order polynomials are fit for the same candidate threshold τ_c varying from 1 to $(u - p - 1)$. The model fitting is repeated for all τ_i separations, $i = 2 \dots (u - 2)$. The adequacy of the fitting process is evaluated via the Akaike Information criterion derived by Hurvich and Tsai

(1989). This measure considers the small sample nature of the problem with possible autocorrelation and underlying normal distribution characterized by the expression:

$$AIC = (u - 1) \cdot \ln\left(\frac{SS_{min}}{u - 1}\right) + (u - 1) \cdot \frac{1 + k/(u - 1)}{1 - (k + 2)/(u - 1)} \quad (2.20)$$

The best model is that with the lowest value of the Akaike criterion. The Akaike information criterion minimizes the sum of squared errors by penalizing by the number of parameters involved (which depends on the order of the polynomial) to avoid overfitting.

2.7 Summary

A traffic conflict has been identified as the most applied surrogate measure of safety. This chapter showed how traffic conflict definition adapts to the available methodologies for their extraction. The introduction of new technologies has allowed for the practical implementation of definitions that rely on proximity measures. Objective measurements have served as an incentive to advance methodologies whose primary objective is to confirm predictive validity. Tarko (2018) introduced a theoretical method for estimating the expected number of crashes using a counterfactual definition of traffic conflicts. The author further strengthened the etiological connection between conflicts and crashes by defining traffic conflicts as a failure scenario in which road users respond with a certain delay. The method was validated with different data sources from simulated environments and naturalistic driving studies (Tarko, 2019; Tarko & Lizarazo, 2020). The authors emphasized the need for identifying failure-based traffic conflicts as a necessary condition to evaluate the safety value of the event. This dissertation provides a generalized framework for the identification of failure-based traffic conflicts using different detection and tracking systems that allow a quick estimation of safety using the validated framework shown in Tarko (2018). Identification of conflicts using in-vehicle instrumentation and area-wide detection systems are discussed in Chapter 3 to introduce the framework and evaluate the conflicts detection using diverse sensing systems.

3. IDENTIFICATION OF TRAFFIC CONFLICTS: STATE OF PRACTICE

3.1 Background

In the past, identification of traffic conflicts was conducted by human observers. Extensive training and long data collection periods were required to ensure the consistency and extraction of events with safety value. In practice, human extraction of traffic conflicts popularized the evasion-based definition of conflicts. The recent progress in sensing technologies has encouraged the development of new methods to measure the motion of road users which allow accurate extraction of trajectories and correspondent estimation of separation metrics applied for characterization of these dangerous events.

The precise and accurate detection and tracking of road users is a key component in a wide range of traffic and safety applications (e.g., automated traffic counts, gap acceptance studies, and speed analysis). The prospect of autonomous vehicles, long-anticipated and recently determined as conceivable, has directed many of the attempts toward high-quality detection and tracking of objects in close vicinity of moving vehicles (Premebida et al., 2007; Spinello et al., 2010). Another potential development avenue for research is a system for area-wide tracking and extraction of trajectories from a fixed roadside position.

This chapter is organized in two primary sections. The first section explores the current technologies for detection and tracking of road users using roadside and in-vehicle instrumentation. Video and LiDAR-based stations, jointly with naturalistic driving studies, also are reviewed as well as their potential for extracting trajectories. The second section focuses on methodologies and strategies that have been applied by multiple researchers in an attempt to extract traffic conflicts based on information from trajectories. This review aims to provide a better understanding of the limitations of the existing methods and emphasize the need for a generalized framework that can be adopted by researchers in the identification of conflicts.

3.2 Technological Needs for Identification of Traffic Conflicts

3.2.1 Road-side Instrumentation: Tracking Systems

Video is generally considered in the literature as the most widely used sensing technology for tracking road users. Buch et al. (2011) divided the video-based techniques into two groups: two-dimensional (2D) and three-dimensional (3D). The 2D group includes algorithms applied in the camera domain with neither restitution nor approximation techniques for estimating the depth. The 3D group adapts object prototypes into the video frame in order to approximate 3D models to mainly solve the well-known occlusion problem.

Coifman et al. (1998) proposed a tracking methodology that was restricted to the 2D camera domain. Identification of moving objects was developed through an adaptive model by applying Gaussian distribution to the background pixels whereby each new pixel was classified as a moving object based on a hypothesis test procedure. The video-based method proposed in Coifman et al. (1998) has been extended to cover tracking features under more complex scenes with multiple entrances and exit regions (Mendes & Bianchi, 2015; Saunier & Sayed, 2006). These approaches also considered the possible tracking disruptions produced in urban environments. The method in Saunier and Sayed (2006) first assessed feature tracking through the Kanade-Lucas-Tomasi feature tracker, after which the features were grouped based on a common motion constraint and spatial proximity. In Mendes et al. (2015), modified video-based tracking was implemented by including vehicle size variation during movement.

Jodoin et al. (2016) proposed a feature-based tracking method with a common motion constraint jointly with a finite state machine (FSM) that corrected the wrong associations in the method proposed in Saunier and Sayed (2006). Moreover, the authors were able to identify not only vehicles but also pedestrians and bicyclists. Validation of the method provided metrics to evaluate detection and tracking based on the CLEAR MOT metric in Bernardin and Stiefelhagen (2008). The CLEAR MOT metric includes two parameters in the tracking evaluation: (1) the multiple object tracking precision (MOTP) and (2) the multiple objects tracking accuracies (MOTA). They applied the MOTP metric to estimate a projection plane; and although the authors reported detection accuracy between 71.8% and 89.6%, the tracking errors in the real world were not reflected in the results.

The 3D group of techniques, which provide real-time methodology for classifying and tracking vehicles at intersections was described in Messelodi et al. (2005). Calibration of the camera based on previous data collection was required, and objects were detected with a background subtraction technique. A hybrid tracking method combining region-based and feature-based techniques was proposed. To reduce the computational load and to allow real-time operation of the algorithm, the 3D prototypes were adjusted to the identified vehicles every five frames. In Song & Nevatia (2007), a 3D vehicle model was proposed to solve the occlusion problem, but the misdetections due to improper orientation of the 3D models were not addressed.

The limitations caused by rendering the 3D world onto a 2D projection plane, which is the primary weakness of the video-based techniques, can be mitigated by applying sensors that better characterize the depth of the sensed environment. The most common alternative technology is 24 GHz or 77 GHz radar (Gordon et al., 2012). Techniques relying on radar were shown to be capable of measuring vehicle speed, range, and lateral position at distances 150 meters away from the intersection at the expense of the field of view (FOV), which was narrow and inadequate for area-wide applications (Aoude et al., 2011). Another limitation of these techniques is their difficulty in detecting stopped vehicles (Gordon et al., 2012).

LiDAR senses a 3D environment using light pulses (Schwarz, 2010). LiDAR from a roadside position is commonly used for detection and classification of vehicles passing a certain spot on a road segment (Gallego et al., 2009; Lee & Coifman, 2015). Other authors included roadside LiDAR-based methods for detecting and counting powered two-wheelers and pedestrians (Prabhakar et al., 2013; Premebida et al., 2009; Subirats & Dupuis, 2015). Gallego et al. (2009) produced promising results with a detection precision rate of 99.50% and classification accuracy rate of 93.60%. In Lee and Coifman (2015), LiDAR was applied to validate the performance of vehicle classification stations. Note that all the above LiDAR applications were limited to selected spots. Aijazi et al. (2016) proposed a method to detect vehicles at road intersections using the compact Velodyne VLP-16 LiDAR, with which they were able to detect vehicles using a super-boxel-based approach. Not all the cited LiDAR-based detection studies were able to track objects.

Urban intersections pose a serious challenge for traffic surveillance, particularly for tracking objects. Video algorithms provide potentially good detection and tracking capabilities but are adversely affected by projective congestion, shadows, obstructions, and night conditions that deteriorate the tracking performance considerably. Although attempts have been made to solve

some common issues of using video, such as obstructions and shadows (Guha et al., 2006; Song & Nevatia, 2007), none of these attempts were fully successful. One of the promising approaches to improving tracking is supplementing video cameras with additional sensors.

Light detection and ranging (LiDAR) technology is a promising alternative or supplement to video technology. Due to its initial high cost, LiDAR technology has not received much attention until recently. The successful and already widespread use of LiDAR is most apparent in land surveying for civil engineering planning, design, and asset management (Haugerud & Harding, 2001). Development of a less expensive sensor, such as the newest Velodyne PUCK sensor, which currently can be purchased for less than \$8,000, was mainly due to the potential for sensor applications in autonomous vehicles (Velodyne, 2016). The price of these sensors is expected to decrease with their proliferation in vehicles and incorporation in other applications.

Considering the aforementioned benefits of LiDAR-based algorithms, Tarko et al. (2016) introduced a LIDAR-based station used as roadside instrumentation for road user detection and a tracking algorithm. This system aimed to overcome most of the limitations of automated video-based detection and tracking methodologies and properly identify moving objects, track them, and estimate their dimensions. An extensive evaluation of the system was conducted in Lizarazo (2016), and a detection rate higher than 90% for vehicles within a range of 50 m was reported while the detection of VRUs was close to 70% in the same range of evaluation. The results when considering the MOTP metric reported discrepancies of less than 80 centimeters compared to the results with manual video extraction for trajectories. These results were encouraging and provided a basis to utilize its data for extraction of traffic conflicts.

3.2.2 In-vehicle Instrumentation

Instrumentation on vehicles has been widely used in naturalistic driving studies and research initiatives for autonomous vehicles. According to Regan et al. (2012), there have been more than 40 naturalistic driving initiatives conducted worldwide. These methods provide unique opportunities to complement existing technologies for the extraction of proximity measures and trajectories from probe vehicles. Most of these research studies have been conducted in the United States with a few projects in Japan (Uchida et al., 2010), Israel (Prato et al., 2010), Australia (Regan et al., 2012), and China (Ma et al. 2017).

In the United States, one of the first instrumentation attempts was conducted in the Crash Avoidance Metric Partnership (CAMP) in 1995. This partnership included the Ford Motor Company and General Motors Corporation and targeted the implementation of crash avoidance countermeasures to improve traffic safety. Data from test track studies and instrumented vehicles were collected to understand in-vehicle demand on drivers. The research study tested nine instrumented vehicles, which included state-of-the-art data acquisition systems for lane keeping, car-following, driver eye glance behavior, and object-and-event detection performance while driving on public roads. The general setup of the experiments is shown in Figure 3.1 (Smith et al., 2005).



Figure 3.1. Experimental setup CAMP initiative (Smith et al., 2005)

Considering the promising results from CAMP, NHTSA and VTTI structured the first large scale naturalistic driving study conducted in the United States, the 100-car study (Neale et al., 2005). One hundred vehicles/drivers were recruited to which unobtrusive instrumentation was added to evaluate drivers' behaviors and extract disaggregate information from their interactions with other road users. Based on their findings, the drivers quickly disregarded the presence of the instrumentation. Hence, the resulting measurements were able to demonstrate extreme cases of driving behavior and performance. The instrumentation of this study included accelerometers, doppler radar systems, video-based tracking systems, and a communication network in the vehicle. Additional information that was collected included automated collision detection and GPS data.

The Second Strategic Highway Research Program (SHRP2) was later established by VTTI. The SHRP2 initiative is the largest naturalistic study implemented of its kind to date with a database including more than 5.5 million trips made by 3,400 participant drivers. With nearly 1,549 crashes, this program provides disaggregate measurements for the probe and road users in the vicinity for analysis.

The instrumentation of the vehicles in the SHRP2 initiative is shown in Figure 3.2. The comprehensive data acquisition system utilized has three primary units: (1) the head unit, (2) the

main unit, and (3) the front radar assembly. The head unit assembly holds three cameras capturing video images from the forward roadway scene, the driver's face, and the pedals with instrument cluster interactions. Additional cameras were installed to record the cabin and the area behind the instrumented vehicle. The main unit hosts the computer functions that coordinate the multiple sensor nodes, communication, and data storage. Finally, the radar unit transmits information via Bluetooth wireless to the main unit and captures the relative speed and position of the surrounding objects. Additional instrumentation relevant to this study includes accelerometer data in three axes jointly with the vehicle network data (speed, ABS, steering wheel angle, etc.). This instrumentation can be applied to identification of traffic conflicts (Campbell, 2012).

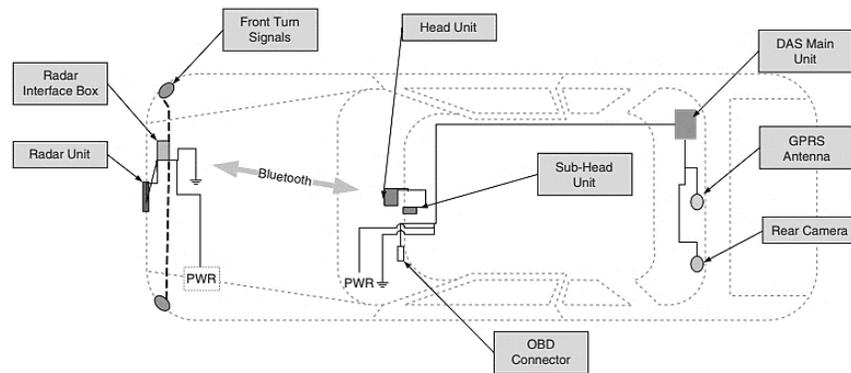


Figure 3.2. In-vehicle instrumentation SHRP2 Initiative (Campbell, 2012)

In addition to naturalistic driving studies, autonomous vehicles can serve as probed vehicles for the extraction of trajectories and tracking road users in close vicinity. These vehicles require extensive instrumentation, including LiDAR, cameras, and radar, which provide tracking information from multiple road users. The instrumentation of the 5th-generation Waymo Driver is shown in Figure 3.3. The system provides proprietary LiDAR and cameras with high-dynamic range and thermal stability to allow better detection in challenging environments. The whole system is complemented with radar, which has the advantages of measuring high-resolution relative velocity with respect to additional objects in rain, fog, and snow conditions (Waymo, 2020). In addition, the Uber driverless system includes ultrasonic sensors for short-range detection.

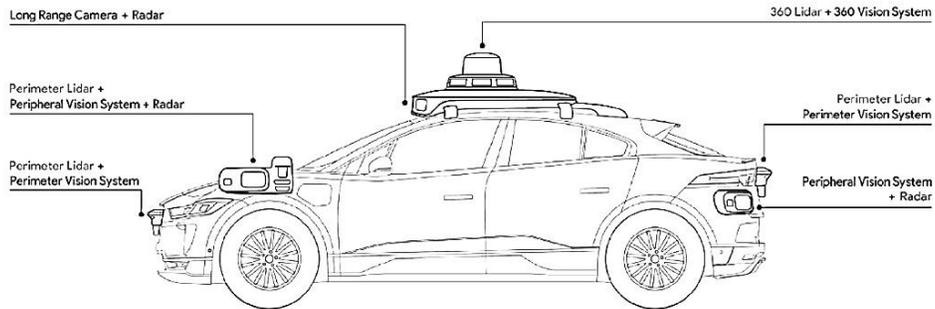


Figure 3.3. Instrumentation 5th-generation Waymo Driver (Waymo, 2020)

3.3 Identification of Traffic Conflicts based on Proximity Measures

As mentioned in the previous chapter, the identification of traffic conflicts has a strong connection with the definition applied in this dissertation. In general, evasion-based approaches were extensively used in human-based observational studies. Proximity in space and time has been applied for a more objective definition of traffic conflicts. These measurements are extracted using the different technologies described in Section 3.2. After obtaining these trajectories, the missing component is the extraction of traffic conflicts.

Objective temporal or spatial separation measures have been introduced in the literature which are obtained from area-wide or in-vehicle technologies. Multiple indicators of extraction and characterization of traffic conflicts include time-to-collision (TTC), post encroachment time (PET), time to accident, gap time, time to departure, braking time, among others. Ismail et al. (2011) broadly categorized these indicators of traffic conflicts into two groups: (1) conflicts measures requiring the presence of a collision course and (2) indicators based on the temporal proximity of road users. In the first case, one of the most representative measures refers to TTC, and in the second case, the most representative measure corresponds to PET. As extensively discussed by Tarko (2018), a counterfactual definition of traffic conflict requires two conditions: (1) S is shorter than a sufficiently short threshold S_c , and (2) the collision would happen if no evasion were performed. In the case of the application of PET, there is no need for a collision course, and conflicts could be cleared by themselves. It violates the postulate of the counterfactual case of a crash. Hence, a proximity measure without a collision point cannot be applied in the proposed method.

When considering methods requiring a collision point, there has been considerable heterogeneity in terms of the spatial or temporal separation values of TTC applied for specification of traffic conflicts. This difficulty has been translated into the definition by users of rather subjective thresholds applied in software programs, such as FHWA's Surrogate Safety Assessment Model (SSAM). SSAM was structured with the primary objective of providing identification, classification, and evaluation of traffic conflicts that occur in simulation models (Gettman et al., 2008). The tool includes an open-standard vehicle trajectory data format that could be obtained from a variety of simulation model vendors/developers including PTV(Vissim), TSS (AIMSUN), Quadston (Paramics), and Riosux Engineering (Texas). Tarko et al (2016) expanded the methodology by including real extracted trajectories into SSAM. The following procedure is applied by the algorithm for the extraction of traffic conflicts:

1. A zone grid is created based on the dimension information given by the header name of the input file, whose individual square zones cover 50-ft by 50-ft (15.25-m by 15.25-m) areas. This split procedure can effectively reduce the number of vehicle-vehicle comparisons necessary for identifying prospective conflicts.
2. Draw an appropriately projected path for each vehicle in the analysis region. One of the most simple and reasonable strategies is to predict the future path of the vehicle based on its current speed and detailed trajectory information. The forward distance is defined as the distance that vehicle *A* can travel at the current speed for the MaxTTC specified by the user in an interval equal to $V1 * \text{MaxTTC}$.
3. When the projected paths for the vehicles of interest are confirmed for the following MaxTTC time interval, the overlapping between two vehicles within each square zone then can be identified after the rectangular perimeter of each vehicle is known. Thus, some conflicting vehicle-pairs are identified for every time-step, and all the previously found conflicting vehicle-pairs would be maintained until the last time-step for analysis.
4. Based on the steps above, a more detailed analysis for each conflicting vehicle-pair can be conducted. This analysis refers to a more accurate estimation of TTC by iteratively shortening the future projection timeline by 0.1 seconds and reprojecting them as before until there are no overlaps.

An important definition of the method is the Max TTC to input for the projection of the path. The analysis considers all the interactions with values lower than MaxTTC regardless of whether

or not it was a controlled action. A more advanced and less deterministic methodology was proposed by Saunier and Sayed (2006). The authors proposed a framework to estimate the probability of observing a collision; and according to the authors, the collision probability for a specific interaction can be estimated by summing the collision probability over all the possible motions that lead to a collision given the radio users state. Hence, the given trajectories for road users A_1 and A_2 with observed trajectories $Q_{1,t \leq t_0}$ and $Q_{2,t \leq t_0}$ (before t_0) and a potential set of extrapolation trajectories H_i and H_j , the probability of observing a collision at point $CP_{i,j}$ is characterized based on the expression:

$$P(\text{Collision}(A_1, A_2) | H_i, H_j) = e^{-\frac{\Delta_{i,j}^2}{2\sigma^2}} \quad (3.1)$$

Being $\Delta_{i,j} = t_{i,j} - t_0$ time between t_0 and collision and σ a normalizing constant close to the road users' reaction time (Hu et al., 2004). In general, extrapolations longer than 3σ provide very low probabilities. The collision probability of two road users A_1 and A_2 at any instant t_0 can be estimated as:

$$P(\text{Collision}(A_1, A_2) | Q_{1,t \leq t_0}, Q_{2,t \leq t_0}) = \sum_{i,j} P(H_i | Q_{1,t \leq t_0}) P(H_j | Q_{2,t \leq t_0}) e^{-\frac{\Delta_{i,j}^2}{2\sigma^2}} \quad (3.2)$$

Being $P(H_i | Q_{1,t \leq t_0})$ the probability of radio user A_1 adopting the hypothesis H_i and similar for the road user A_2 . Since noise and additional factors might potentially alter the estimation of probabilities, the authors proposed taking the average of a small number of the largest values taken by the collision probability over different times t_0 . Hence, let n be the number of largest probabilities, the severity indices for the time interval t_1 to t_2 is estimated as:

$$\text{InteractionSeverityIndex}([t_1 t_2]) = \sum_{\substack{(i,j) \text{ such that } A_i \text{ and } A_j \\ \text{are observed in } [t_1 t_2]}} \text{SeverityIndex}(A_i, A_j) \quad (3.3)$$

This approach and variations of it have been applied in additional studies including Saunier and Sayed (2008) and Saunier et al. (2009). This method provides an interesting framework by avoiding a deterministic approach to traffic conflicts. However, an important component that is missing is considering the presence of failure in the analysis. This dissertation aims to provide a

framework for the identification of failure and the associated separation measures for estimation of the expected number of crashes.

3.4 Summary

The first section of Chapter 3 provided a general description of the technological needs for the extraction of traffic conflicts. A review of area-wide tracking systems and vehicle instrumentation was provided as well to guide the reader regarding available data sources that can be used for extraction of traffic conflicts. The extensive use of video detection algorithms was discussed and its low cost and high resolution. The limitations caused by rendering the 3D world onto a 2D projection plane, which is the primary weakness of the video-based techniques, can be mitigated by applying sensors that better characterize the depth of the sensed environment. The LiDAR-based station for detection in the tracking of road users developed by Tarko et al (2016) also was discussed. Their method provides detection rates higher than 90% for vehicles within a range of 50 m and discrepancies in positions of less than 80 centimeters when compared to the results of manual video extraction. In-vehicle instrumentation provides an additional data source for extraction of traffic conflicts. Detection and tracking of road users in the vicinity of the vehicle enable characterization of their interactions and estimation of proximity measures. Although comprehensive in-vehicle technologies were first conceived in naturalistic driving studies, the introduction of autonomous vehicles is providing unique data for real-time evaluation of hazards using conflicts. An evaluation of these dangerous events can be implemented in algorithms from autonomous vehicles to reduce risk of crash.

The second section of this chapter discussed the adopted methods for the identification of traffic conflicts. A common practice for identification of conflicts includes its definition based on proximity measures. Subjective separation thresholds can be selected, and the associated events can be extracted. Other approaches look at all events as potential conflicts and estimate the probability of collision based on the most likely future states of the road users. None of these methods have the ability to indicate failure, which is a necessary condition included in the counterfactual definition by Tarko (2018).

4. IDENTIFICATION OF REAR-END TRAFFIC CONFLICTS

4.1 Background

Chapter 3 discussed the different methods in the literature for identification of traffic conflicts. There is a general consensus that proximity measures are good proxies for identification of failure. Additional stochastic methods have been used in traffic conflict analysis with the probability of crash increasing with closer proximity.

This chapter introduces a methodology for identification of rear-end traffic conflicts using naturalistic driving data. The issues and corrective measures applied from SHRP2 data sources are discussed as well as the limitations when proximity measures are applied exclusively as an indicator of conflicts. Risk perception at lower speeds and crash reporting are other areas analyzed in this chapter.

The structure and discussion of the results follow Tarko and Lizarazo (2020). The evaluation of identified conflicts was conducted using the method described in Section 2.6. In this dissertation, the conflicts were considered valid if the Lomax-based assumption of response delays was met and the estimation of the expected number of crashes was consistent with the reported number of crashes. Three categories of drivers were considered to demonstrate the method's ability to detect difference between drivers by age and gender as in Tarko and Lizarazo (2020),

4.2 Data Collection

SHRP2 data were used for the identification of rear-end traffic conflicts. With nearly 1,549 crashes documented, this program provided a comprehensive dataset for extraction of conflicts and better validation of these events with a controlled environment of reported crashes. Continuous monitoring of driving in naturalistic studies prevents underreporting of crashes, which is an issue frequently pointed out in the literature on crash-based validation techniques. In-vehicle instrumentation provides a good characterization of rear-end conflicts with braking as a primary evasive maneuver. Other conflict types were not considered due to the lack of instrumentation to measure more complex scenarios.

The SHRP2 program collected two petabytes of data from 5.5 million trips, making the identification of conflicts task overwhelmingly large. As an alternative, an algorithm was developed to search for randomly selected individual trips for potential conflicts based on rather liberal conditions.

The algorithm cycled until it reached 10,000 trips with at least one potential conflict found. The scope of the data extraction is described in Appendix A, and information about the algorithm is in Appendix B. The trips without conflicts that were found when searching for trips with conflicts were included in the sample. This random sampling strategy preserved the SHRP2 population ratio between trips with conflicts and trips without conflicts.

4.3 Detection of Traffic Conflicts Proximity Measures

Identifying traffic conflicts was executed in two phases: (1) identifying potential conflicts based on liberal crash nearness criteria and removing obvious false positives and (2) refining the traffic conflicts detection criteria. In the first phase, the distance to the lead vehicle was continuously measured with the front radar unit. The instantaneous TTC was calculated as the ratio of the distance to the relative speed derived from the temporal changes in the distance. All the car-following periods with an instantaneous TTC lower than two seconds were extracted. After extracting those specific events, additional analyses were conducted to remove false-positive events. The following data collected by the SHRP2 instrumented vehicles were found useful in the second phase:

- Accelerometer-measured longitudinal and lateral acceleration rates of the instrumented vehicle reported 10 times per second.
- Speedometer-measured speed of the instrumented vehicle reported once in one to two seconds at an irregular rate.

The speed measurements were reported at irregular rates with incomplete information. However, the reported values were assumed accurate based on the information obtained from the Virginia Tech team. The gaps between speed measurements were filled with the speed values calculated by integrating the longitudinal acceleration rates. A description of the method applied to complete these values is provided in Appendix C. Any discrepancy between the speeds measured with a speedometer and the speeds derived from the acceleration rates were reconciled by adjusting the acceleration-derived speeds through shifting and scaling that minimized the discrepancy between the measurements. After filling the gaps between the speedometer-measured speeds, these speed sequences were smoothed with a moving average of five values.

Following the method described in Tarko (2018), a small sample of 255 conflicts from randomly selected trips was analyzed to identify the conditions leading to possible false positives. The minimum

instantaneous TTC, τ_m was interpreted as the end of a successful response. Hence, the delay of response x was calculated by subtracting the τ_m value from the τ_c threshold: $x = \tau_c - \tau_m$. According to Section 2, the linear relation between $\ln(1 - F(x_i))$ and $\ln(1 + x_i/\tau_c)$ is expected if the observed exceedances x are Lomax-distributed. This also provides a convenient check if the assumed threshold τ_c is sufficiently small. Hence, the log-log curves $-\ln[1 - F(x)] = f[\ln(1 + x/\tau_c)]$ calculated with the ordered response delays x were inspected. The log-log curves for separation thresholds $\tau_c = 2.0\text{ s}$, $\tau_c = 1.5\text{ s}$ and $\tau_c = 1.0\text{ s}$ are shown in Figure 4.1.

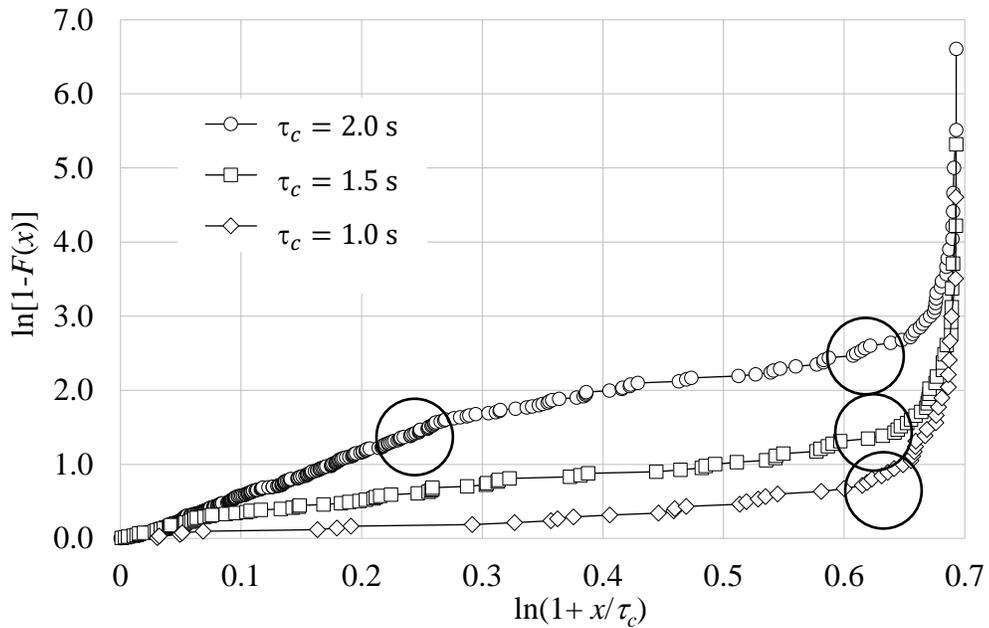


Figure 4.1. Log-log curves for the test sample with the lane exit observations (Tarko, 2018)

The non-linearity of the log-log curves in Figure 4.1 is obvious. For example, the log-log curve for the two-second τ_c threshold is composed of three distinct segments. The first segment ends at point $\ln(1 + x/\tau_c) = 0.27$, the middle segment ends at point $\ln(1 + x/\tau_c) = 0.6$, and is followed by a strongly curved upward segment. The first inflection point is not present in the other two curves with thresholds $\tau_c = 1.5\text{ s}$ and $\tau_c = 1.0\text{ s}$. However, the second inflection point leading to steeply sloped segments remains. The non-linear behavior of the curves shows that the observed exceedances are not Lomax-distributed and a long separation threshold is not the cause.

The non-Lomax distribution indicated by the log-log curves was found to be strongly influenced by a considerable concentration of short τ_m values. Inspecting these cases with video clips revealed that most of these values occurred shortly before the disappearance of the lead vehicle. The lead vehicle signaled in advance the intended lane change and initiated a consistent lateral movement. The following driver could have anticipated that the lane would be cleared soon. This action prompted the following driver to continue closing the gap and passing the lead vehicle at a short time separation. The numerical data obtained from the accelerometers confirmed the followers' gentle braking or no braking at all. These behaviors produced short τ values while the drivers kept the situation under control. The events did not involve traffic failures and were removed from the initially claimed conflicts. Figure 4.2 presents the instrumented vehicles' speeds at the time of their τ_m values. The cases with disappearing lead vehicles are represented by solid circles.

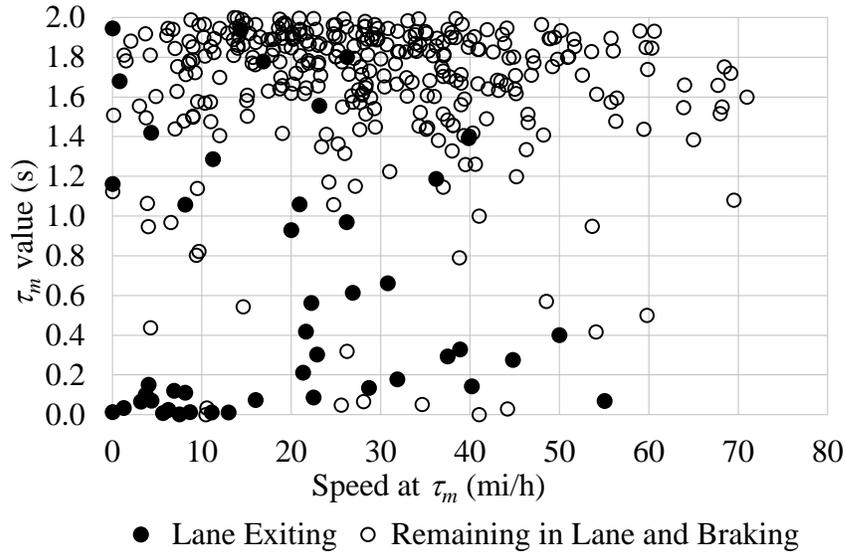


Figure 4.2. Minimum instantaneous TTC observations in the test sample (Tarko, 2018)

Figure 4.3 presents the log-log curve $-\ln[1 - F(x)] = f[\ln(1 + x/\tau_c)]$ for $\tau_c=2.0$ seconds after removing the identified false positives. The strong non-linearity has been eliminated. This result concluded the first phase of refining the detection of traffic conflicts.

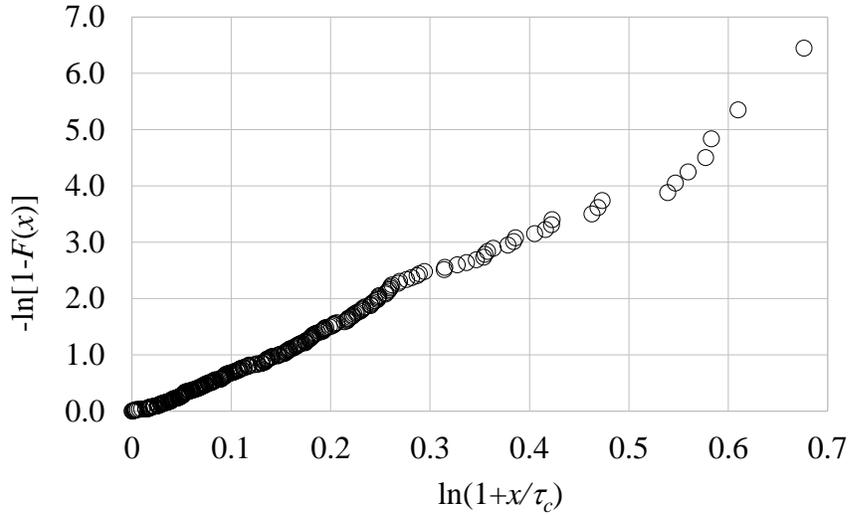


Figure 4.3. Log-log curves after removing the lane exit events ($\tau_c = 2$ s) (Tarko, 2018)

4.4 Redefining Identification of Traffic Conflicts

The second phase included refining traffic conflicts detection as well as conducting a safety analysis on a new sample of 12,742 randomly drawn trips. When randomly searching for trips with conflicts, the conflict-free trips also were preserved. The exposure for the entire population of drivers was relevant in the scope of this dissertation. Hence, the total number of miles driven by all the SHRP2 drivers was obtained. In addition, the crash data information was acquired for validation of the number of crashes predicted using conflicts. The safety analysis was preceded by additional quality control measures. Two additional conditions were observed that were leading to false positives:

1. Sudden reductions of instantaneous TTC τ caused by spurious backward movements of the lead vehicle. These false detections were caused by two scenarios: (1) vehicles moving on the other side of the road in the opposite direction were misinterpreted by the instrumented vehicle as lead vehicles, and (2) vehicles entering a gap in front of the instrumented vehicle were exposing their sides to the radar beam of the instrumented vehicle, which was sliding backward along the side of the entering vehicle.
2. Values of τ smaller than 0.1 seconds, including zero, associated with low speeds and with short ranges often were followed by periods with missing radar readings. The first hypothesis was made that these events involved collisions between vehicles. However, no reports of collisions were found in the SHRP2 data. The final conclusion was that the missing radar readings were caused by relative speeds that were too low and ranges that were too short, which made the radar measurements unreliable.

Furthermore, even if there was physical contact between vehicles in some cases, it was harmless and negligible.

A necessary condition of claiming a conflict is the inevitability of collision if no effective evasive action is executed. Saunier et al. (2010) approached this problem by estimating alternative trajectories from data free of conflicts and applying them in conflict situations. Area-wide extraction provides a more convenient data sources for this analysis. However, considering the longitudinal nature of the data collection process adopted in this dissertation, alternatives needed to be considered to evaluate identification of evasion due to failure. Figure 4.4 and Figure 4.5 provide two different deceleration profiles with iTTC lower than two seconds for distinction. Figure 4.4 shows a case without failure with a controlled deceleration profile, where the driver was aware of the situation and no strong response was present in the event. On the other hand, Figure 4.5 shows a strong response where an evasion was present. A graph-based method was postulated using sequential t-tests, where a collision avoidance was concluded if the follower was braking at a rate at least of 1.5 m/s^2 (Najm & Smith, 2004), and the increase in the braking rate (jerk) was significantly stronger than the average value (one-percent significance level).

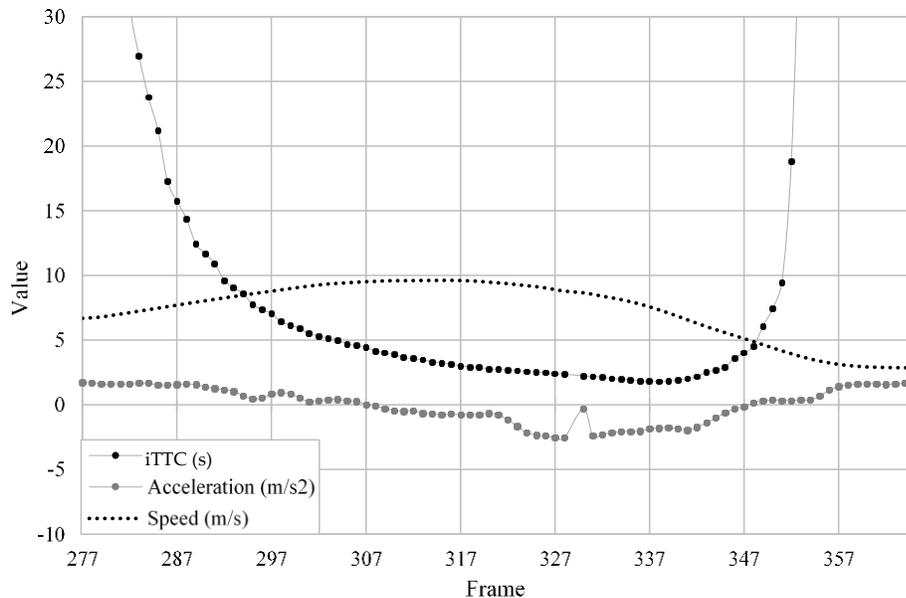


Figure 4.4 Example failure free braking with short separation

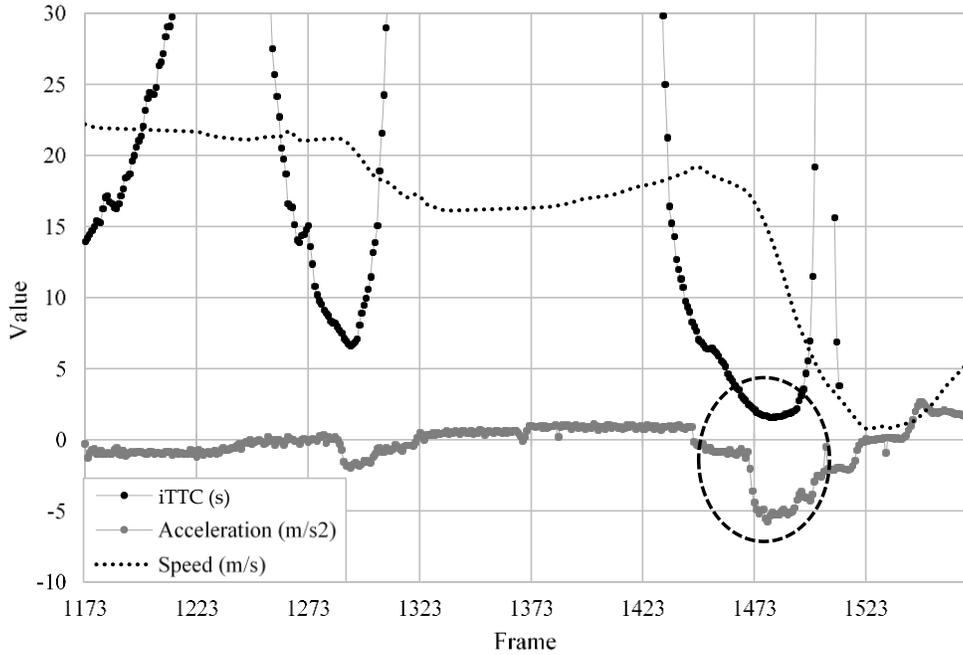


Figure 4.5 Example rear-end conflict in SHRP2 data marked (inside the oval shape)

Another condition of claiming a conflict is the considerable probability of a severe outcome. Thus, events with short separations between vehicles at low impact speeds may be false positives. Two observations justify not considering these events as conflicts. (1) Drivers moving at low speeds may perceive no considerable risk. A small relative speed between vehicles at impact leads to no harm to either property or vehicle occupants. Thus, it is difficult to claim that safety-related failure is possible when there is no hazard. (2) The expected outcome is too minor to be reported by police. The latter problem is less critical than the former because it can be rectified by applying a discrete-outcome model to estimate the probability of a police-reportable outcome. Hence, it is necessary to consider the impact speed at the hypothetical collision. The impact speed can be predicted as follows:

$$\Delta v(t_m + \tau_m) = \dot{R}(t_m) + \tau_m \cdot \ddot{R}(t_m) \quad (4.1)$$

where: $\Delta v(t_m + \tau_m)$ is the relative speed at the time of the hypothetical collision, $\dot{R}(t_m)$ is the relative speed (range rate) at time t_m , $\ddot{R}(t_m)$ is the relative acceleration rate at t_m , t_m is the time at which the τ reaches the minimum value τ_m . Time $t = t_m + \tau_m$ is the instant at which the hypothetical collision happens (good approximation if τ_m is small).

The anticipated impact speed should be sufficiently high to induce a fear of collision in the driver. A procedure similar to the one proposed for estimation of a separation threshold indicating failure can be implemented to determine an impact speed threshold. The speed threshold can be determined as the lowest speed at which estimates of the expected crash frequency for different separation thresholds are stable. It is important to stress that unlike a shorter separation threshold for claiming failure, speed thresholds that are too high lead to bias in the results. In general, speed thresholds that are too conservative tend to neglect valid conflicts associated with potentially serious outcomes and therefore will produce an underestimation of the number of crashes.

The expanded procedure applied in the presented SHRP2 study included the following steps:

1. Set the speed threshold at zero and claim traffic conflicts for a liberally large separation threshold.
2. Estimate the expected number of crashes for gradually decreasing separation thresholds until the crash estimates are stable.
3. If no suitable separation threshold is found in Step 2 before running out of conflicts, increase the speed threshold and repeat claiming conflicts with a liberally large separation threshold.
4. Repeat Steps 2 and 3 until the lowest speed threshold and the corresponding largest separation threshold are found before running out of conflicts.
5. If no thresholds are found, then the sample is too small or there are other issues. Log-log curves should be inspected.

After the quality control and identification of conflicts were completed, the data processing moved to the final phase of traffic conflicts validation via estimation of the expected number of crashes with appropriate log-log curves to ensure the correctness of the results and comparison. This analysis was applied to data obtained for three types of studied drivers.

4.5 Evaluation of Identified Conflicts

As postulated in the introduction of this chapter, the identified conflicts were considered valid if there was a stable prediction of the expected number of crashes following the theory in Tarko (2018). In addition, the estimation of this number followed the crash rates reported in that study. It should be stressed that the estimation, in this case, was conducted using only 1.7% of the trips. A comparison was conducted using the crash rates for the entire population of five million trips.

To evaluate the heterogeneity of the identified conflicts, estimation of the expected number of crashes was conducted using three categories of drivers: (1) young male 16-25, (2) mature male 46-65,

and (3) mature female 46-65. It was hypothesized that the risky nature and limited driving experience of young drivers can result in a higher estimation of expected number of crashes compared to their mature counterparts (Scott-Parker et al., 2013). It typically takes longer for young drivers to develop the necessary perceptual and cognitive skills to safely interact with other road users (Deery, 1999). An underestimation of the hazard and overestimation of their driving skills are among the primary arguments supporting this overrepresentation.

Table 4.1 presents the following information for each driver category: the number of trips, the total time of following other vehicles, and the total distance covered when following other vehicles. Following another vehicle was defined as the presence of a lead vehicle detected by the front radar of the SHRP2 vehicle. The three quantities of (1) number of trips, (2) car-following time, and (3) car-following miles were the alternative measures of exposure to rear-end crash.

The rear-end collisions considered in the analysis were caused by SHRP2 drivers who collided with the lead vehicle (front hit). Events of SHRP2 vehicles being hit by other vehicles were not considered as irrelevant as not caused by the studied drivers. The fifth and sixth columns in Table 4.1 include the counts of all the rear-end crashes caused by the studied drivers during the SHRP2 project and the subset of these crashes reported to the police (police-reported crashes). The rates of police-reported crashes per 100 million miles of following another vehicle (car-following miles) were calculated for each driver category. The young male drivers caused 1,086 police-reported crashes per 100 miles of car following, mature males caused 456 of such crashes, and mature female drivers only 137. As expected, the young male drivers were the least safe drivers while the mature female drivers were the safest. The performance of mature male drivers fell between the performance of the other two categories of drivers.

Table 4.1. Basic statistics for the three studied types of drivers in the SHRP2 population

Driver Category	Number of Trips	Car-following Time (million hrs)	Car-following Miles (mi)	All Rear-end Crashes¹	Police-reported Rear-end Crashes¹
Males 16-25	901,587	190,586	2,209,414	31	24
Males 45-64	443,020	100,656	1,316,178	10	6
Females 45-64	496,351	121,510	1,460,122	2	2

Note: ¹These front-hit crashes were caused by drivers of SHRP2 vehicles when following another vehicle.

The identification process of conflicts followed the framework described in Section 4.4. The evaluation of the conflicts was conducted using three different performance metrics. The first metric evaluated the stabilization of the crash estimates with a reduction in the separation value and an impact speed sufficiently high to induce drivers' fear of collision. The second metric evaluated the linearity of the relation between $\ln(1 - F(x_i))$ and $\ln(1 + x_i/\tau_c)$. Linearity is expected if the observed exceedances were Lomax-distributed. Finally, the third metric evaluated whether the identified conflicts applying the Lomax-based method were capable of producing rear-end crash estimates consistent with the values presented in Table 4.1. The focus should be on the police-reported crashes as they were more consequential for safety and typically were analyzed in past research. Analysis of the SHRP2 reported collisions revealed an average impact speed of 2.925 m/s for police-reported collisions and an average impact speed of 1.65 m/s at non-reported collisions.

A total of 1,882 events were initially claimed as conflicts at the two-second crash nearness threshold (also called the vehicle separation threshold) measured with instantaneous TTC τ_m . Then, the refined procedure described in Section 2 was implemented to identify the largest crash nearness threshold (proper threshold) and the lowest impact speed thresholds that produced stable estimates of expected crashes for separation thresholds shorter than the proper threshold. The two thresholds were determined by decreasing the separation thresholds and increasing the impact speed thresholds until the shape of the expected crash estimates consistently indicated a flat trend below a certain crash nearness threshold as prompted by the theory. Figures 4.6, 4.7, and 4.8 present profiles of the expected crash estimates for the three studied driver categories and the various minimum speed thresholds. The 2.5 m/s speed threshold, at which the profiles became flat, was obvious for young male and mature female drivers (Figure 4.6 and Figure 4.8). It was less obvious for mature male drivers (Figure 4.7). It seems that these drivers perceive a collision undesirable even at speeds lower than 2.5 m/s and maintain a proper level of vigilance. Nevertheless, the speed threshold of 2.5 m/s was selected also for these drivers because the profile model in Figure 4.9 fitted to the expected crash profile for mature male drivers had the lowest AIC value for this speed threshold. The proper separation thresholds for the three studied categories of drivers were determined with the method presented in Section 2.6 (Figure 4.10).

The approximate linearity of the log-log functions for the three categories of drivers in Figure 4.11 confirmed the validity of the selected thresholds. The obtained results supported the common opinion that young drivers are more aggressive and accept higher risk by following other vehicles at

shorter distances than mature drivers. On the other hand, a longer crash nearness threshold for mature drivers indicated a more conservative style of driving.

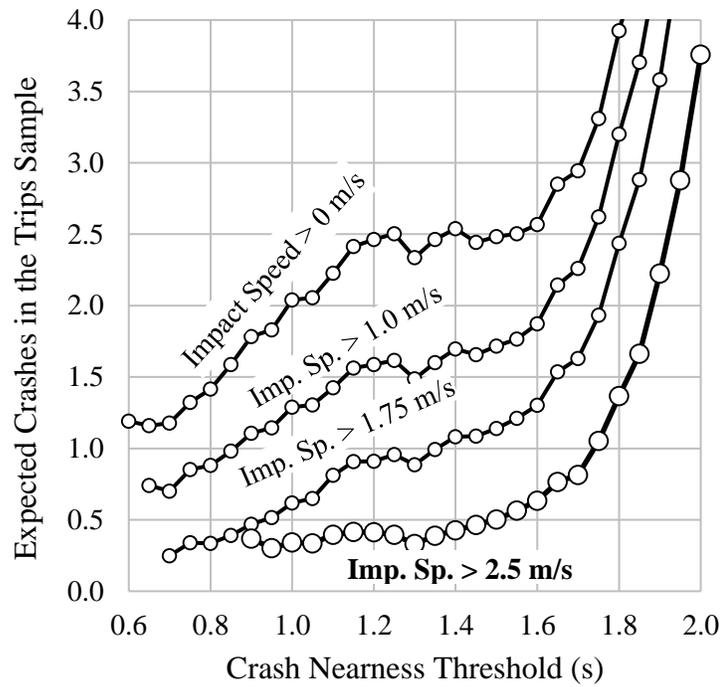


Figure 4.6 Profiles expected crash estimates various speed thresholds (young male drivers)

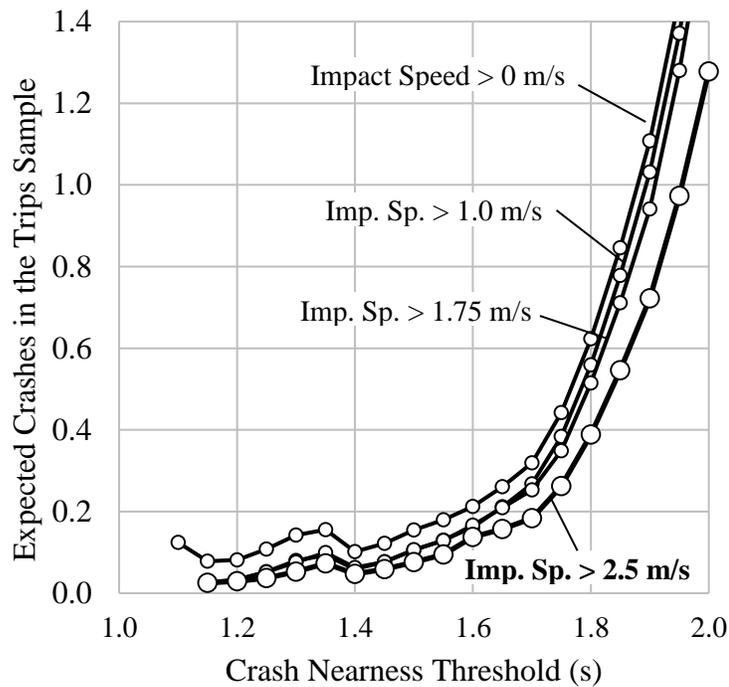


Figure 4.7 Profiles expected crash estimates various speed thresholds (mature male drivers)

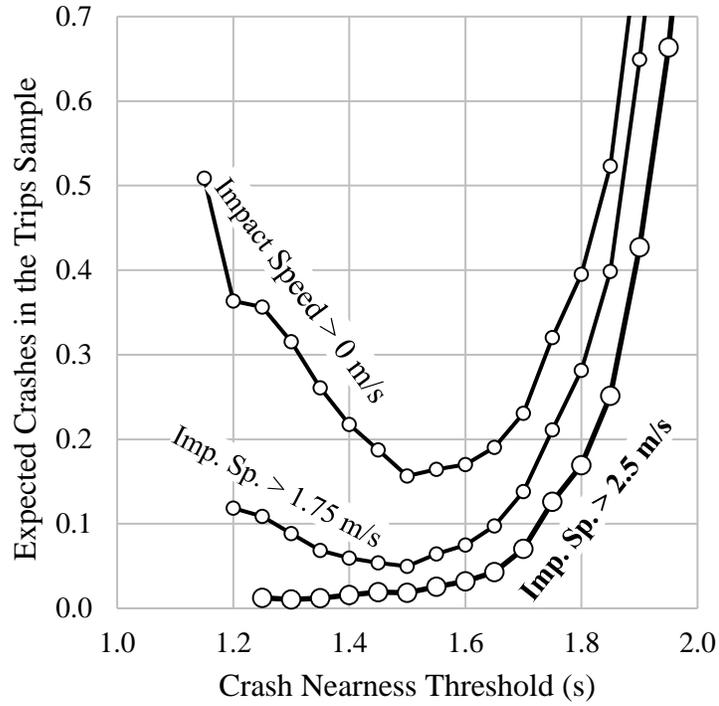


Figure 4.8 Profiles expected crash estimates various speed thresholds (mature female drivers)

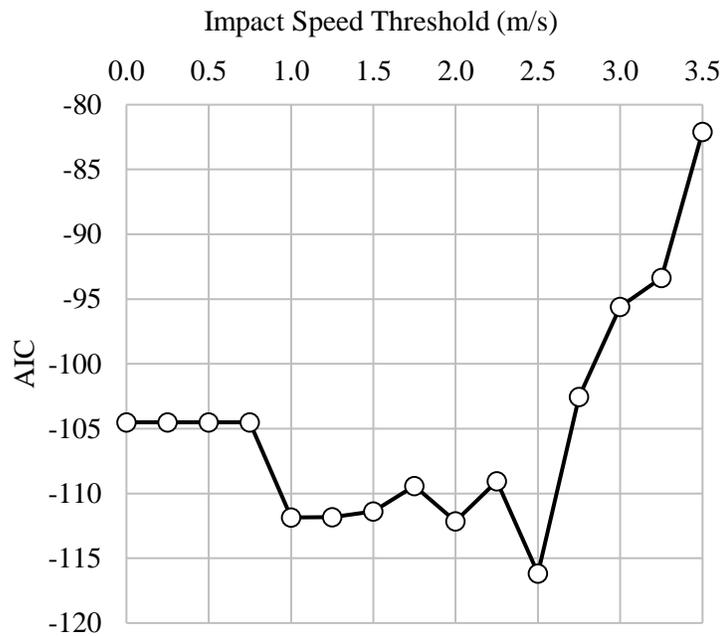


Figure 4.9 AIC of the expected crash estimates profile models for proper thresholds under various threshold speeds – mature male drivers

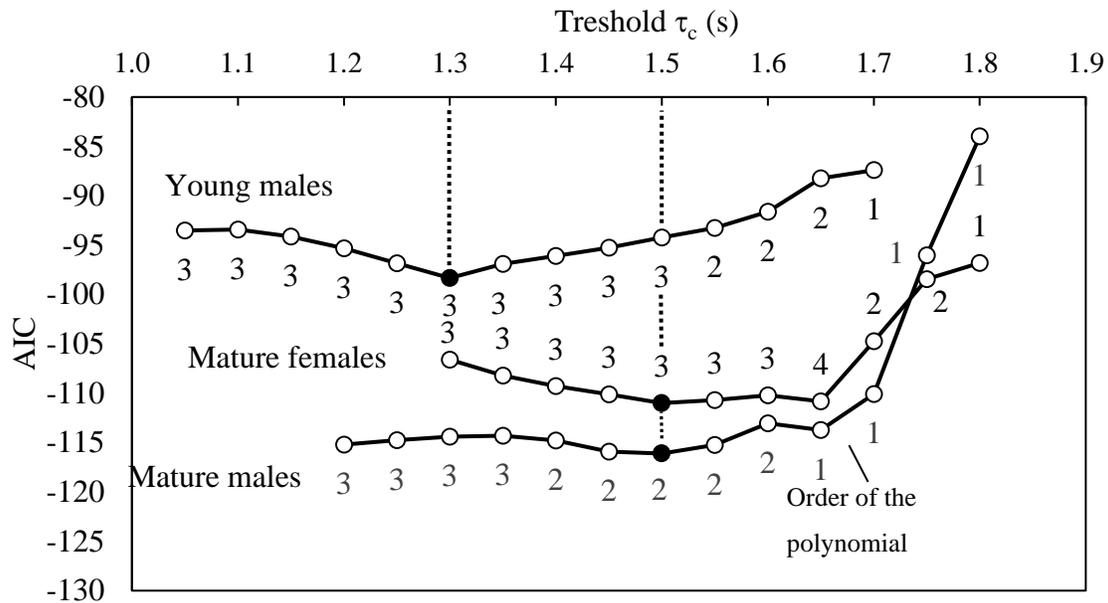


Figure 4.10 Selection of the proper τ_c threshold based on the information-based performance measure AIC

To address the question of whether the identified conflicts jointly with the Lomax-based method could properly estimate the observed crashes of Table 4.1. Basic statistics for the three studied types of drivers in the SHRP2 population and the expansion factors were applied in the estimation. 1.4% of the trips were randomly selected from the 1.84 million trips made by the studied SHRP2 drivers. The information provided for each randomly selected trip included information about the trip, the driver, and the vehicle. Table 4.2 presents the number of trips, the total time of following other vehicles, and the total number of miles traveled when following other vehicles. Although the percentages of trips, hours traveled, and miles traveled in the sample were different, they remained sufficiently close across the driver categories. This close similarity supported the claim that the sampling was indeed random and preserved the population structure. It is also important to note that due to the relatively small samples, none of the randomly selected trips and drivers experienced a rear-end crash.

The estimated number of crashes expected in the sample is provided for each driver category in Table 4.3. The crash nearness threshold for young male drivers was 1.3 s and 1.5 s for mature female and male drivers. Figure 4.11 and Figure 4.12 present the results for the speed threshold 2.5 m/s (5.6 mi/h) applied to all the tested separation thresholds. The solid dots in Figure 4.12 indicate unbiased estimates of the expected number of crashes for all three categories of drivers.

Table 4.2. Basic statistics for the three types of drivers in random sample 1.4% trips

Driver Category	Number of Trips (% Population)	Car-following Time [mln hrs] (% Population)	Car-following Miles [mi] (% Population)	Police-reported Rear-end Crashes (Front Hits)
Males 16-25	12,962 (1.44%)	2,888.5 (1.52%)	37,288 (1.69%)	0
Males 45-64	6,187 (1.40%)	1,574.7 (1.56%)	22,596 (1.72%)	0
Females 45-64	6,876 (1.39%)	1,902.1 (1.57%)	25,142 (1.72%)	0

Table 4.3. Summary of results Lomax-based analysis for speed threshold 2.5 m/s

Driver Category	Separation Threshold (s)	Traffic Conflicts Claimed	Parameter k	Conditional Crash Probability	Expected Number of Crashes
Male 16-25	1.3	75	7.822	0.004418	0.3313
Male 45-64	1.5	71	9.864	0.001074	0.0762
Female 45-64	1.5	70	11.904	0.000261	0.0183

The Lomax estimates of the expected number of rear-end crashes for the three groups of drivers presented in the last column of Table 4.3 apply to the small number of trips randomly selected from the SHRP2 population. These trips are reported in the second column of Table 4.2. The crash estimates had to be expanded before comparing them to the corresponding rear-end crashes in the SHRP2 program. First, the trip-based expansion factor was calculated by dividing the number of trips in the population by the number of trips in the sample. Then, the conflict-based crash estimate in Table 4.3 was multiplied by the trip-based expansion factor, which was repeated for each driver category. These expanded conflict-based estimates are included in Table 4.4 in the Trip-based Expansion column. The same expansion procedure was repeated to obtain the time-based and mile-based expanded estimates.

It is clearly seen that all the estimates in Table 4.4 are much closer to the counts of the police-reported crashes than to all the crashes recorded in the SHRP2 data in Table 4.1. These estimates were made based on the conflicts with predicted impact speeds higher than 2.5 m/s. The proposition was made that drivers involved in conflicts appeared to feel excessive risk and behaved consistently with the theory at higher speed differences. The 2.5 m/s impact speed was also a reasonable demarcation point between crashes not reported to police and crashes reported to police. This consistency may

indicate that, on average, road users can discriminate safe rear-end interactions from consequential ones (police reportable).

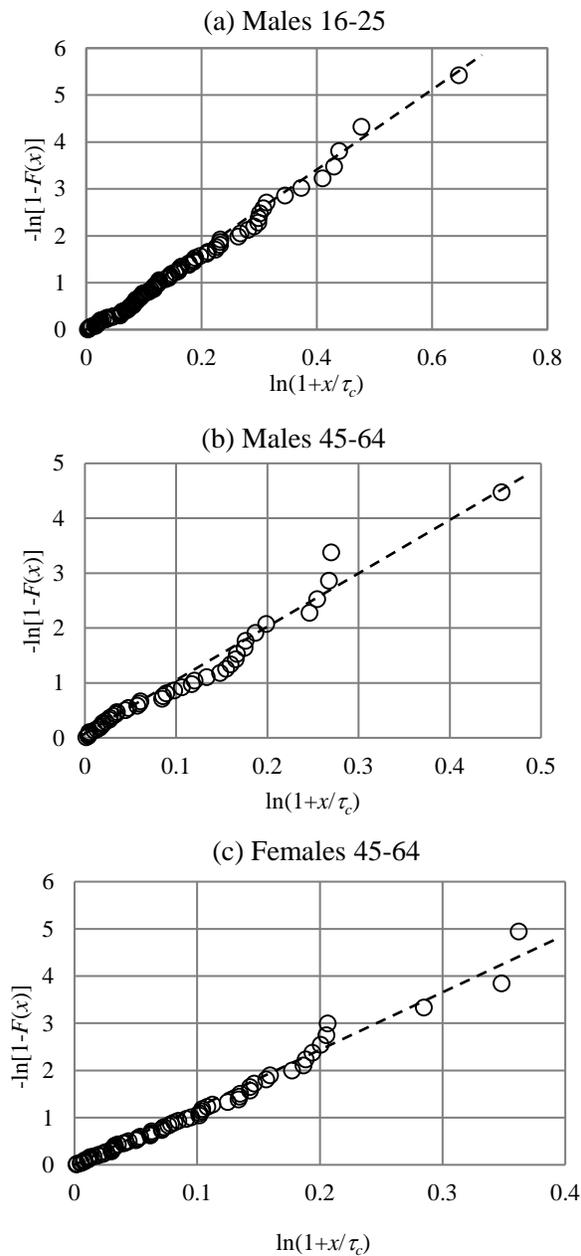


Figure 4.11. Log-log curves for the three studied categories of driver

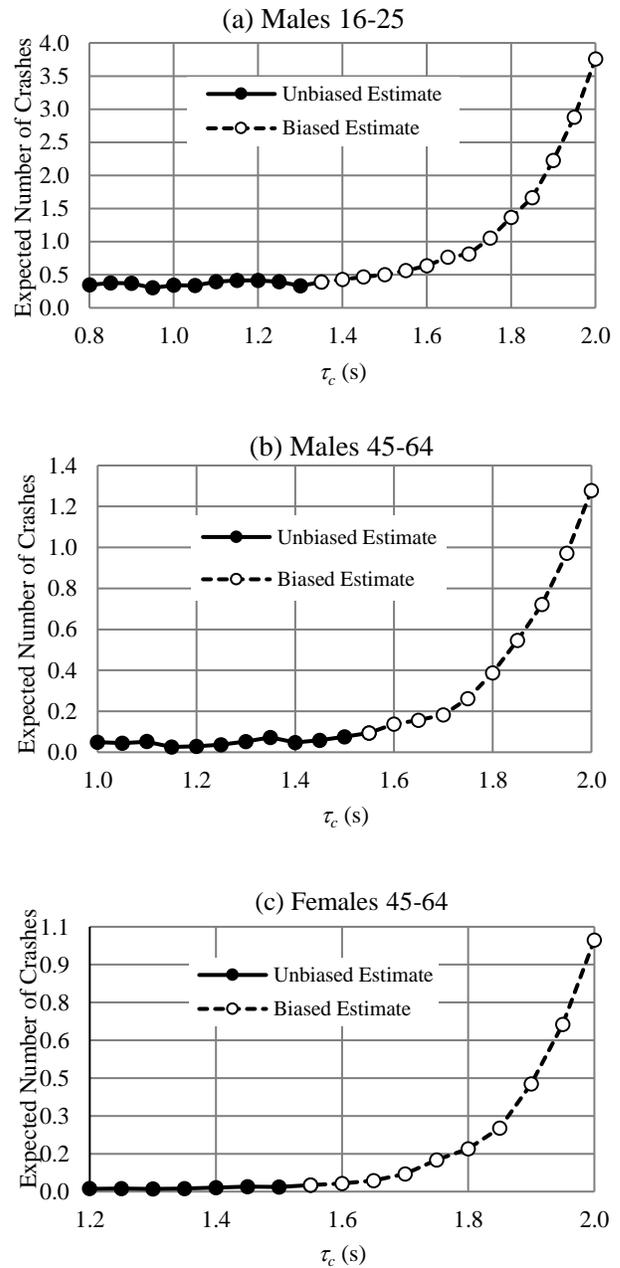


Figure 4.12. Expected number of crashes for different τ_c per category of driver

Table 4.4 shows that all the obtained estimates were well within the 90% confidence interval of the crash-based estimates obtained from all the police-reported rear-end crashes in the SHRP2 period for each of the studied driver categories. Although the car-following time and the car-following distance

were initially believed to be good measures of exposure to rear-end crashes, the results expanded based on the two measures exhibited lower accuracy than the results expanded with the number of trips.

The results supported the redefining of traffic conflicts as successful and proved the validity of the Lomax-based method in Tarko (2018). The process to redefine conflicts provided valuable lessons learned to be transferred into a generalized framework for identification of traffic conflicts. The three metrics were met and the results showed the important characterization and identification of conflicts obtained good prediction on magnitude.

Table 4.4. Comparison of the population crash estimates and the conflict-based sample estimates expanded to the SHRP2 population

Driver Category	Crash-based Estimates in All Trips			Conflict-based Estimates for All Trips		
	Expected Crash Count ¹	90% Conf. Interval		Trip-based Expansion	Time-based Expansion ²	Mile-based Expansion ³
		Lower Limit	Upper Limit			
Male 16-25	24.0	16.55	32.59	23.0	21.8	19.6
Male 45-64	6.0	2.61	10.51	5.4	4.9	4.4
Female 45-64	2.0	0.36	4.74	1.3	1.2	1.1

Notes: ¹Rear-end crashes (front hits by SHRP2 instrumented vehicles) reported to police, ²Total time traveled when a lead vehicle is present (detected by the front radar of a SHRP2 vehicle), ³Total miles traveled when a lead vehicle is present.

4.6 Summary

This chapter introduced a methodology applied for characterization and identification of traffic conflicts using in-vehicle instrumentation from naturalistic driving studies. The limitations and noise measurements from the instrumentation were identified and discarded for proper analysis. Important lessons learned from this study are applicable in a generalized framework for identification of traffic conflicts.

The first component corresponds to the identification of failure using kinematic signatures adopted by drivers. Indication of rapid deceleration was observed in emergencies when the driver was correcting a potentially dangerous encounter to avoid a crash. A graph-based method relying on continuous testing was applied in this dissertation to identify emergencies with a significant change in jerking. The second characteristic discussed corresponds to the issue of risk perception of collision at low speeds. In general, drivers were found to accept short separations when the outcome was not

consequential. This condition was further investigated using different thresholds of hypothetical speed at collision. An iterative process was proposed for estimation of speed that provided stable estimates in the analysis.

Three metrics were applied for evaluation of the identified conflicts. The first metric was the presence of a hypothetical collision speed that provided stabilization of the number of crashes. The second metric included evaluating whether the response delay could be assumed Lomax-distributed. The third metric included a prediction of the expected number of crashes following the actual estimates from the study. The results followed Tarko (2018). The prediction based on traffic conflicts data produced results that closely followed the anticipated trends prompted by the theory in terms of the group of drivers.

The results confirmed that the expected number of rear-end crashes using the claimed traffic conflicts afforded useful information about traffic safety. Young male drivers experienced higher crash frequency compared to their counterparts considering their limited experience and subsequent less-honed driving skills. On the other hand, the safest drivers were between 45 and 64, who were more experienced and possessed adequate perception of hazards. The results were less evident when comparing the obtained frequencies of males and females between 45 and 64.

5. IDENTIFICATION OF TRAFFIC CONFLICTS: VEHICLE-BICYCLE ENCOUNTERS

5.1 Background

Chapter 4 discussed a redefined methodology for the identification of rear-end traffic conflicts using in-vehicle instrumentation from naturalistic driving studies and introduced three important characteristics of traffic conflicts: (1) good approximation of iTTC (constant speed and direction) in rear-end conflicts, (2) definition of potential collision when evasion was identified, and (3) hypothetical speed at collision that provided perception of risk to avoid. Chapter 5 applies some of these characteristics for identification of turning conflicts between vehicles and bicycles. Methodologies for estimation of iTTC in turning encounters are proposed as well as analysis of evasion with braking and change in direction as primary evasive actions. In this analysis, the potential speed at collision is not considered. It is hypothesized that crashes between vehicles and vulnerable road users (VRUs) are always undesirable with a high perception of risk even at low speeds. Data from the InDEV program in Europe is applied for this identification task.

This chapter is structured as follows. The first section provides general information about the collected data and the extraction of trajectories for vehicle-bicycle encounters. Section 2 describes the methodologies applied for the estimation of iTTC in turning maneuvers. Section 3 discusses the limitations of considering proximity measures as the only indicator for failure and the need to identify evasion. Finally, section 4 describes the methodology for the estimation of TTC and lessons learned from the study.

5.2 Data Collection

The European Union structured its InDeV project with the primary objective of developing a comprehensive set of methodologies, concepts, hardware, and software to better understand the safety of VRUs. Considering the desirable benefits of surrogate measures related to the proactive evaluation of safety, their project intended to explore these surrogates to further investigate the safety problems of pedestrians and bicycles (Johnsson et al., 2018).

The two primary data sources of the project included video recording and crash-data collection. The project proposed obtaining a ground-truth measure of safety using crashes that could be validated

using surrogate measures of safety (SMoS) collected from video detectors. The data collection effort was conducted in more than seven countries in Europe, with the number of locations filmed varying from three to four. Semi-automated extraction of trajectories was developed for encounters, which were defined as any clearly defined and countable event that generated an opportunity for an accident to occur (Elvik, 2010). The extraction of trajectories was conducted for encounters between VRUs and motor vehicles.

The large and continuous amount of data collected prevented the research team to extract all vehicle-bicycle encounters for more than 20 locations included in the analysis. The team divided the processing and extraction of encounters into two separate phases. The first phase intended to extract as many encounters as possible in a 24-hour period. In this phase, human observers identified encounters between the relevant road users. The second stage involved the extraction of only critical events identified using automated software in a three-week period. A diagram describing the data collection process is shown in Figure 5.1; and the standard setup is shown in Figure 5.2. The instrumentation included two RGB cameras and a thermal camera.

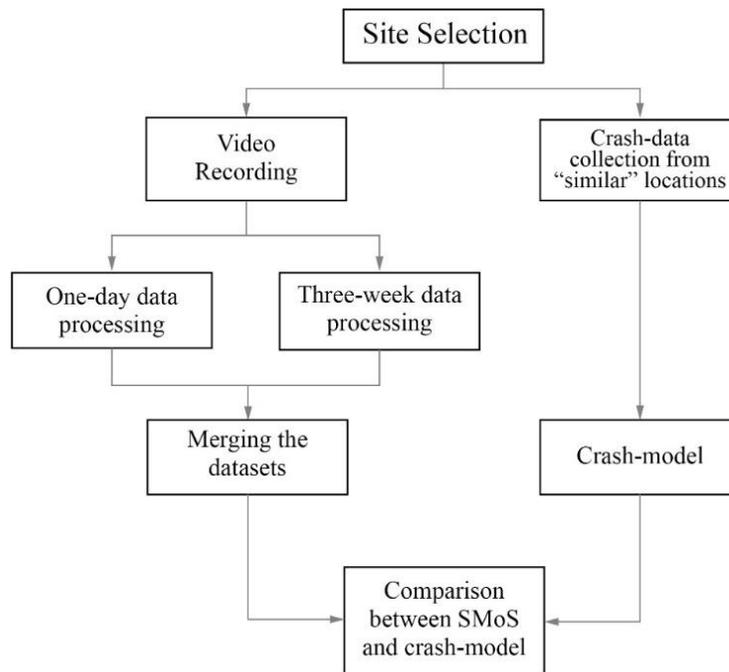


Figure 5.1. Data collection planning InDEV program (Johnsson et al., 2018)



Figure 5.2. Data collection setup at intersections (Johnsson et al., 2018)

5.3 Identification of Traffic Conflicts: Instantaneous Time to Collision

Following a similar procedure to the one implemented in Chapter 4 of this dissertation, iTTC was estimated in order to evaluate if the application of only separation measures indicated failure in vehicle-bicycle encounters. Two different methods for estimating iTTC are described below.

5.3.1 Calculating iTTC with the Assumption of Fixed Speeds and Directions

Let us consider the turning conflict in Figure 5.3. Detection and tracking algorithms provide the mapping trajectories from the vehicle and bicycle, separately. The trajectories are representations from a finite set $I \subset \mathbb{R}$ (I is typically a finite set of time instants at regular intervals δ) to \mathbb{R}^2 (the two-dimensional plane). The trajectories for vehicle C_i can be represented as:

$$I \subset \mathbb{R} \rightarrow \mathbb{R}^2: t \rightarrow \mathbf{C}_i(t) = [x_i(t), y_i(t)] \quad (5.1)$$

Using the relative direction and decomposition of relative speed along this vector, it is possible to estimate instantaneous TTC until the hypothetical collision point. Since observations are repeated at time intervals δ , the vectorial representation of the direction and speed of motion of road user i can be estimated as:

$$\mathbf{d}_i(\mathbf{t}) = \mathbf{C}_i(t + \delta) - \mathbf{C}_i(t - \delta) = [x_i(t + \delta) - x_i(t - \delta), y_i(t + \delta) - y_i(t - \delta)] \quad (5.2)$$

$$\mathbf{V}_i(t) = \frac{C_i(t + \delta) - C_i(t - \delta)}{2\delta} = \left(\frac{x_i(t + \delta) - x_i(t - \delta)}{2\delta}, \frac{y_i(t + \delta) - y_i(t - \delta)}{2\delta} \right) \quad (5.3)$$

The unit direction of travel at time t is:

$$\mathbf{u}_i = \frac{\mathbf{d}_i}{|\mathbf{d}_i|} = (x_{ui}, y_{ui}) \quad (5.4)$$

and the unit vector normal to the direction is:

$$\mathbf{n}_i = (y_{ui}, -x_{ui}) \quad (5.5)$$

An abstract representation of road users from detection and tracking systems include a rectangular model approximating the width and length of the object (Tarko et al., 2016). The same approximation was adopted in the InDeV program. Hence, by obtaining information about the length of road user i represented as l_i and the width w_i , the position of the different corners can be obtained using the following expressions.

$$\begin{aligned} \mathbf{FR}_i &= \mathbf{C}_i + \frac{l_i}{2} \mathbf{u}_i + \frac{w_i}{2} \mathbf{n}_i \\ \mathbf{FL}_i &= \mathbf{C}_i + \frac{l_i}{2} \mathbf{u}_i - \frac{w_i}{2} \mathbf{n}_i \\ \mathbf{BR}_i &= \mathbf{C}_i - \frac{l_i}{2} \mathbf{u}_i + \frac{w_i}{2} \mathbf{n}_i \\ \mathbf{BL}_i &= \mathbf{C}_i - \frac{l_i}{2} \mathbf{u}_i - \frac{w_i}{2} \mathbf{n}_i \end{aligned} \quad (5.6)$$

The vehicle and the bicycle are shown in Figure 5.3. Applying expression (5.6), one can obtain the direction travel \mathbf{d}_1 and \mathbf{d}_2 , respectively. The relative direction of the bicycle in relation to the vehicle is: $\mathbf{d} = \mathbf{d}_2 - \mathbf{d}_1$ and the corresponding relative unit direction is: $\mathbf{u} = \mathbf{d}/|\mathbf{d}|$. The relative speed is $\mathbf{V} = \mathbf{V}_2 - \mathbf{V}_1$.

The relative direction of motion \mathbf{u} and corresponding normal unit vector \mathbf{n} can serve as a new base for a system of coordinates where the relative motion of the bicycle towards the vehicle is along y axis. Let:

$$\mathbf{M} = (\mathbf{n} \ \mathbf{u}) = \begin{pmatrix} y_u & x_u \\ -x_u & y_u \end{pmatrix} \quad (5.7)$$

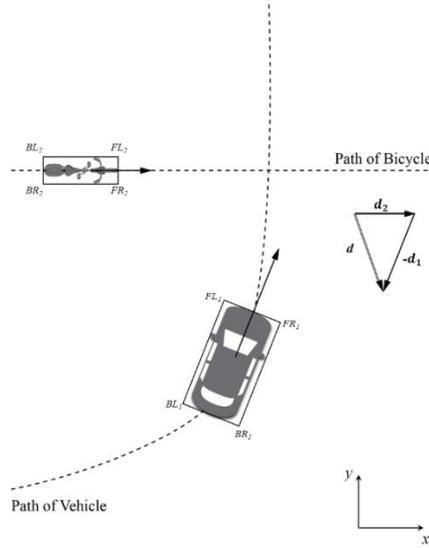


Figure 5.3. Conceptual representation vehicle-bicycle turning conflict

Where \mathbf{u} and \mathbf{n} are vertical vectors. The corners $\mathbf{P} = (x, y)$ of the bicycle and the vehicle in the new x coordinate system can be estimated as (see Figure 5.4):

$$\mathbf{P}^* = \mathbf{M}^{-1} \mathbf{P} \quad (5.8)$$

or

$$\mathbf{P}^* = \begin{pmatrix} y_u & -x_u \\ x_u & y_u \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.9)$$

In a general case, any corner of the vehicle may hit either side of the bicycle. This can be illustrated as the corners' vertical trajectories passing through the bodies of the road users. It is easy to determine the corner that hits the other road user and the corresponding hit side by looking at Figure 5.4, where it is depicted with a solid vertical line that connects the upper corner of the vehicle with the impact point on the side of the bicycle.

The process for estimating the minimum distance between the road users can be done by pairing each corner of the vehicle with a corresponding side of the bicycle and calculating the distance between them along the y-axis. The solution is the corner-side pair separated with the shortest distance among all the pairs. The iTTC is calculated by dividing the found distance by the relative speed V (length of the speed vector V). iTTC does not exist if the two road users "instantaneously" miss each other or when the relative instantaneous speed V is zero. This calculation is repeated for all the times. The outcome includes the iTTC profile and of the separation distance profile.

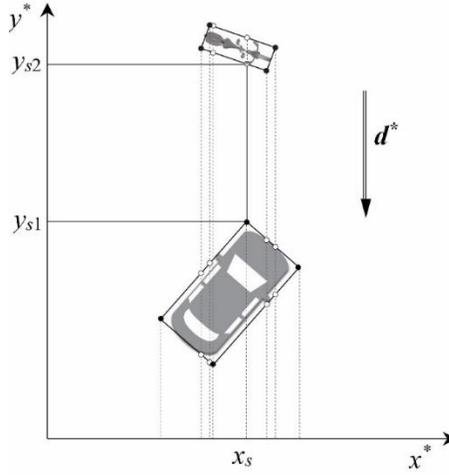


Figure 5.4. Relative reference system for estimation of iTTC

5.3.2 Calculating iTTC with the Assumption of Collision Path between Centroids

The estimation process of the instantaneous TTC based on fixed speeds and accelerations is prone to significant changes between consecutive frames since the estimation applies to frame-specific headings. A more stable approach is shown in Figure 5.5. Instead of considering the direction of travel for each vehicle, a potential collision path can be determined by the direction between the centroids from the two road users and the decomposition of speed along this vector. An advantage of using this approach corresponds to the presence of a continuous iTTC. In the former method iTTC is not reported when two vehicles “instantaneously” miss each other or the relative instantaneous speed V is zero. In this approach, iTTC is not reported only in the case of null relative speed. Following a similar representation of the event, vector d is replaced with the vector connecting centroids $\mathbf{g} = C_2 - C_1$ and the corresponding relative unit direction is characterized based on the expression $\mathbf{v} = \mathbf{g}/|\mathbf{g}|$. The relative direction of motion \mathbf{v} and corresponding normal unit vector \mathbf{m} can serve as a new base for the system of coordinates where the vector between centroids is rotated to be along the y-axis. Hence:

$$\mathbf{N} = (\mathbf{m} \ \mathbf{v}) = \begin{pmatrix} y_v & -x_v \\ x_v & y_v \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.10)$$

The corner of the two road users can be rotated using the expression (5.10). The speed of the vehicles is decomposed along this vector.

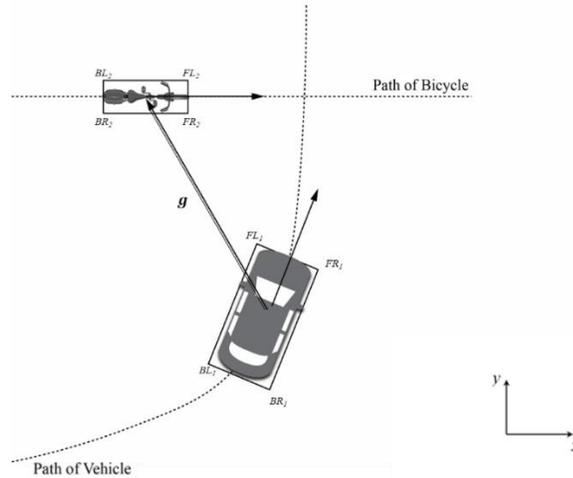


Figure 5.5. Conceptual representation of turning conflict with collision path between centroids

5.3.3 Estimation Expected Number of Crashes instantaneous Time-to-Collision (TTC)

The two methodologies for estimation of instantaneous TTC were applied. Estimation of the expected number of crashes was also conducted to determine if only proximity measures provided a good characterization of failure in vehicle-bicycle encounters. The methods were tested in the intersection with the highest number of encounters in the program. The testing took place in an intersection in Spain for a 24-hour period. The dataset reported a total of 414 turning encounters between bicycles and vehicles. The geometry and characterization of the intersection are shown in Figure 5.6, which corresponds to a four-leg signalized intersection. Conflicting volumes are present due to a bike lane disturbing right turns adopted by vehicles at the intersection.

The two methods for estimation of instantaneous TTC were conducted for the 414 turning encounters. For identification of the events, the method with the assumption of fixed speed and direction is defined as relative method while the additional one based on centroids is defined as range method. This nomenclature is applied as reference for the graphs. The histograms of the minimum iTTC for each one of the encounters are shown in Figure 5.7 for the relative and range method, respectively.



Figure 5.6. Camera location and intersection layout in Spain

A preliminary comparison between the minimum iTTC values found from both methods is shown in Figure 5.8. Based on the results, there is a significant discrepancy. As hypothesized, there were more similar values in lower ranges of minimum iTTC. However, there were some cases with large discrepancies at low values. Investigating the source of discrepancies, large differences were observed when there was a change in direction as an evasive maneuver to avoid a collision. In this case, the hypothetical collision point changes altering significantly the estimation of iTTC when considering the relative method. In general, a more stable behavior of iTTC was found in the range-based method since it does not rely on the heading from the road users but in the vectorial decomposition between centroids. Examples from cases with primary evasive maneuvers as the change of direction from the vehicle and bicycle are shown in Figure 5.9.

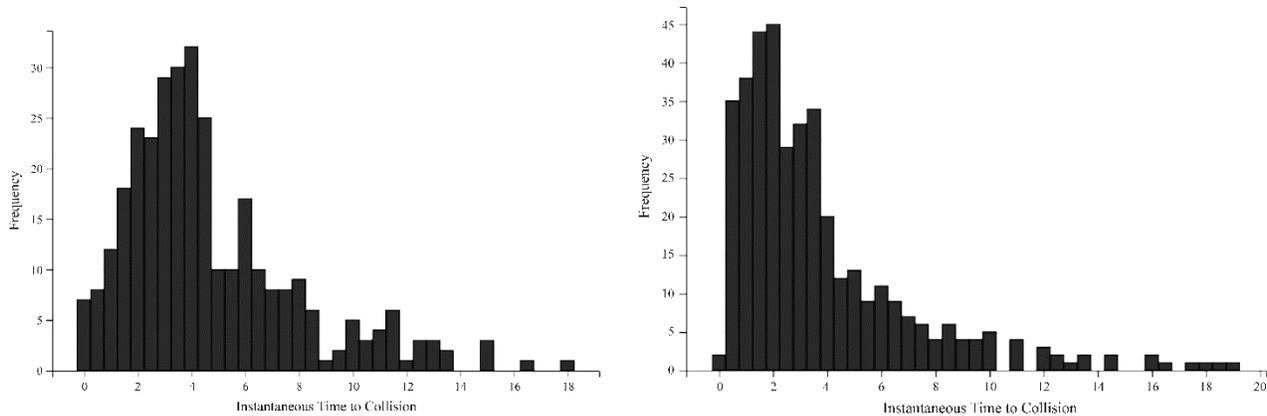


Figure 5.7. Histogram iTTC relative method (left) and range method (right)

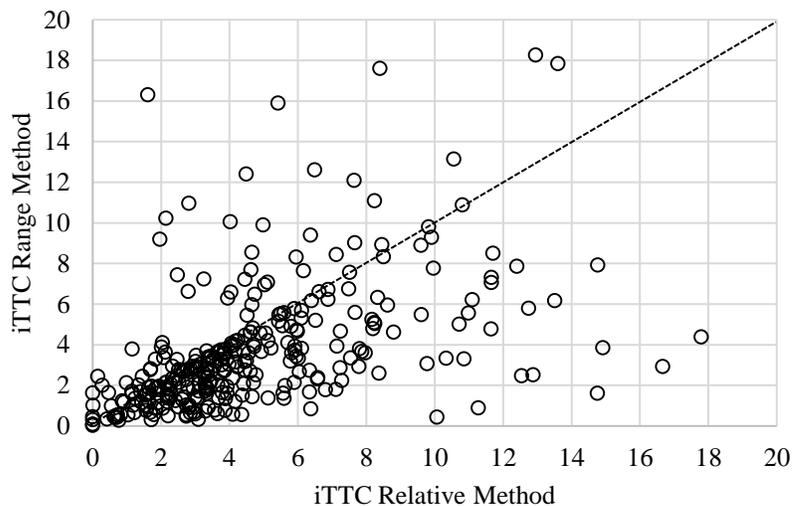


Figure 5.8. Comparison minimum iTTC values relative and range methods

Considering the discrepancy between the estimation of instantaneous TTC between the two methods, the Lomax-based approach was applied to observe the validity of these conflicts for estimation of the expected number of crashes. Similar to the evaluation conducted in the naturalistic driving study, if the conflicts met the primary assumption of failure, there should be stabilization of expected number of crashes and Lomax distributed responses. The results from the estimation of the expected number of crashes, log-log curves, and conditional probabilities of crashes are analyzed separately.

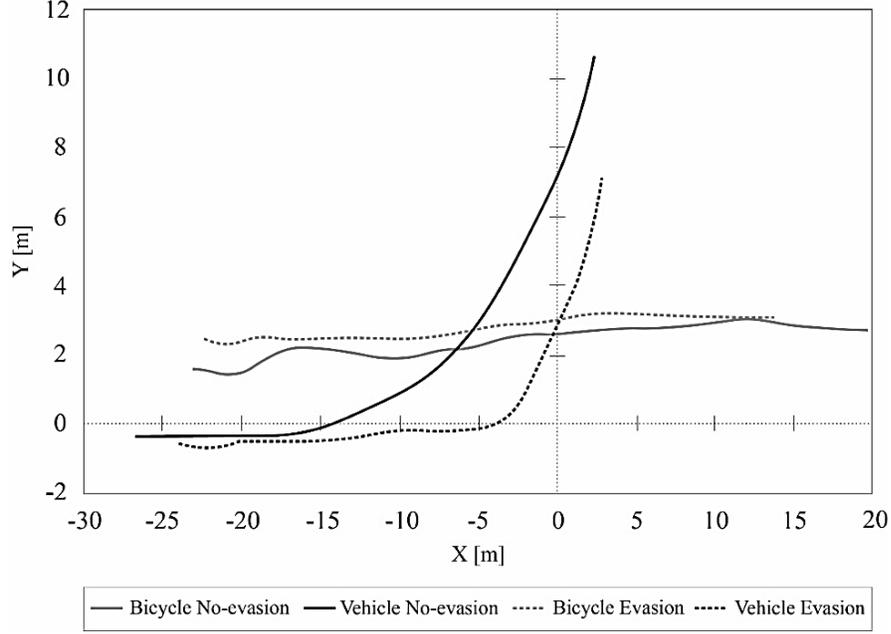


Figure 5.9. Discrepancy of trajectories with and without lateral evasion

Expected Number of Crashes

Following the description in Chapter 2, the estimation of expected number of crashes is conducted using an iterative process of reduction in separation threshold S_c until stabilization in the expected number of crashes is observed. Given a certain separation $\tau_c(s)$ in which events lower than this threshold are considered conflicts, the expected number of crashes was estimated using the expression:

$$Q_c = P(C|N_j) \cdot n_j = 2^{-k_j} \cdot n_j \quad (5.11)$$

Where

$$k_j = \frac{-\sum_{i=1}^n \ln\left(1 - \frac{i-0.5}{n}\right) \ln\left(1 + \frac{x_i}{\tau_c}\right)}{\sum_{i=1}^n \left[\ln\left(1 + \frac{x_i}{\tau_c}\right)\right]^2} \quad (5.12)$$

Being $x_i = \tau_c - \tau_m$ and $P(C|N_j) = 2^{-k_j}$. Based on the results shown in Figure 5.10 and Figure 5.11, the estimation of expected number of crashes does not show stabilization or the presence of unbiased estimates for either of the two methods. In this case, the crash estimates were still decreasing when the number of conflicts fell below a number too small to produce reliable results. The results also

show an excessive number of crashes with close to two crashes in a 24-hour period. In general, improper identification of conflicts meeting failure and longer observation periods should be conducted.

Conditional Crash Probabilities

A stabilization of the expected a number of crashes is supported based on the empirical evidence of a higher probability of crash once the separation of the road users is shorter. Crash is defined when there is no separation between the road users. This trend is observed for the profile probability of the relative method. On the other hand, the range method shows a decreasing probability, contradicting the theory proposed in previous chapters. These results support a better characterization of the crash proximity using the relative method. When there is a monotonic increase in the probability of crash without an interval of stable crash estimation as shown in Figure 5.10, Tarko (2019) indicates that the observation period is too short. Moreover, it is hypothesized that there is a violation of the failure-based definition, with some events adopting low proximity measures considering easier maneuverability by bicycles. The method does not provide good results and it is expected that a longer observation period with better identification of failure may provide better insight as proven in naturalistic driving conflicts.

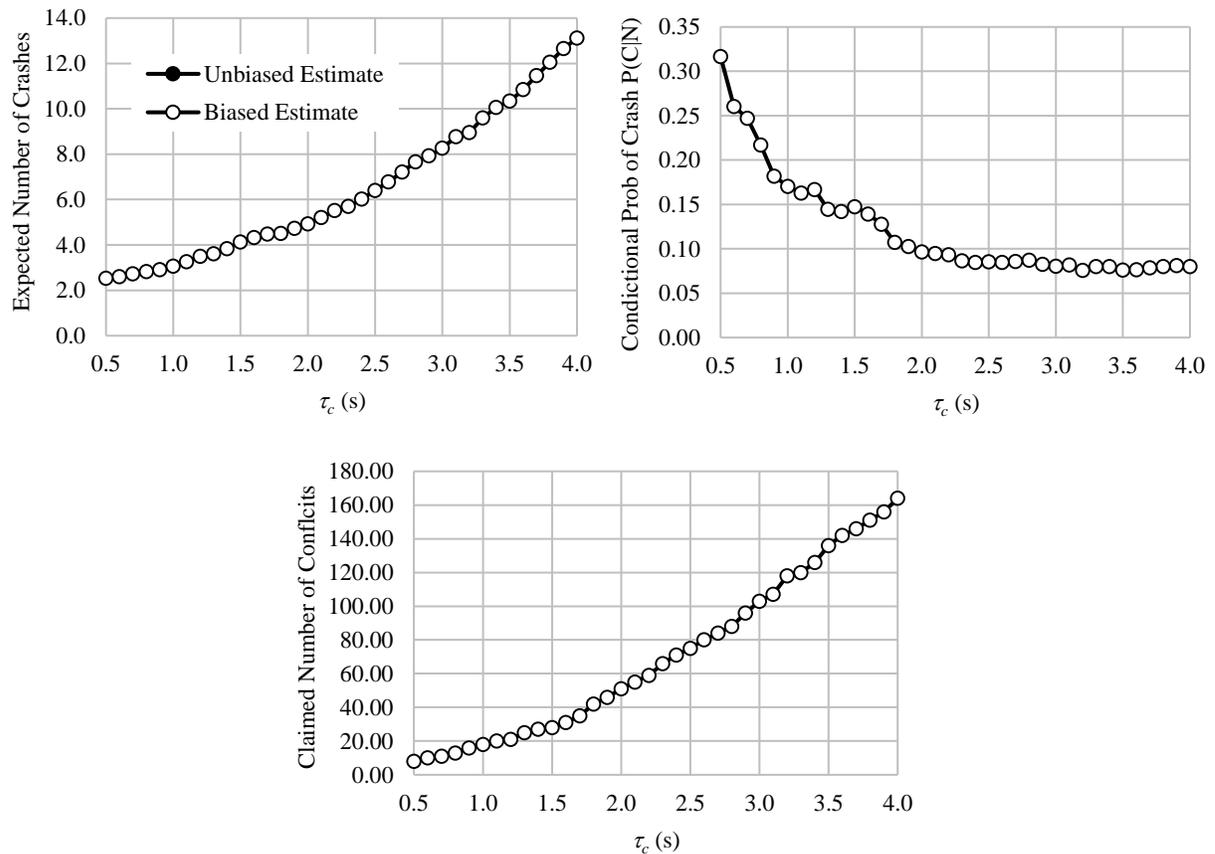


Figure 5.10. Results for instantaneous TTC relative method

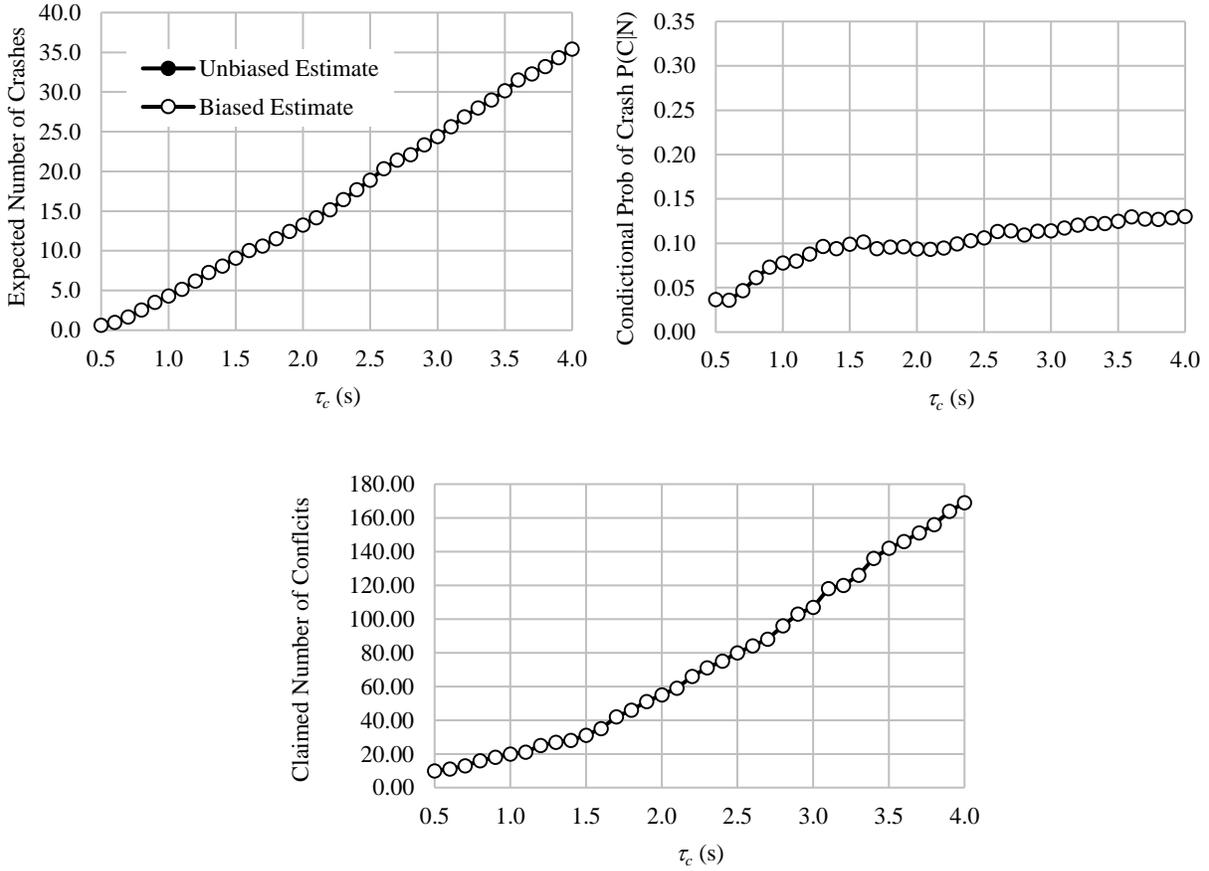


Figure 5.11. Results for instantaneous TTC range method

Log-log Curves

The Lomax distribution of the exceedance values is confirmed using the log-log curves. Following the method described in Section 2.3, it is possible to plot $\ln(1 - F(x_i))$ and $-\ln(1 + \theta x)$ to check the assumption of Lomax distributed delays. Considering the cumulative function of the response delays and the assumption of SPE method $\theta = 1/\tau_c$, it yields the following relation between the n of claimed conflicts for separation τ_c :

$$-\ln [1 - (i - 0.5)/n] = k \ln [1 + x_i/\tau_c] \quad (5.13)$$

Whereas n is the number of claimed conflicts under separation τ_c , i is the index of exceedance x_i being sorted from the lowest to the largest value, τ_c separation to claim conflict, and k is the unknown parameter of the Lomax distribution.

The results from the two methods for the estimation of instantaneous TTC are shown in Figure 5.12 and Figure 5.13. The results prompt the lack of linearity indicating violation of the assumption of Lomax-distributed exceedances. More linearity might be expected at smaller separation threshold τ_c . Events without failure and strong influence of events with short instantaneous TTC supported the shown lack of linearity in these cases.

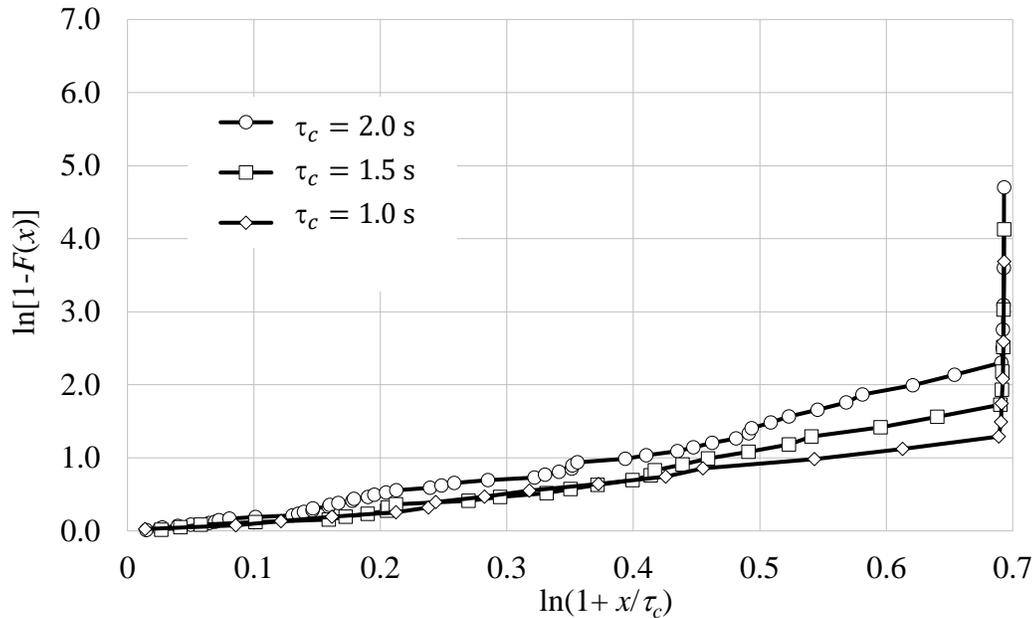


Figure 5.12. Log-log curves for instantaneous TTC relative method

In general, the results supported a significant number of encounters classified as conflicts without failure. It provided overestimation of the expected number of crashes and counterintuitive estimation of the probability of crash given conflict for the case of the range method, as well as the lack of Lomax-distributed responses. Proximity measures and assumption of instantaneous TTC do not apply in more complex scenarios such as turning. The following sections discuss other components considered in the process for identification of traffic conflicts in turning movements to be applied in a general framework.

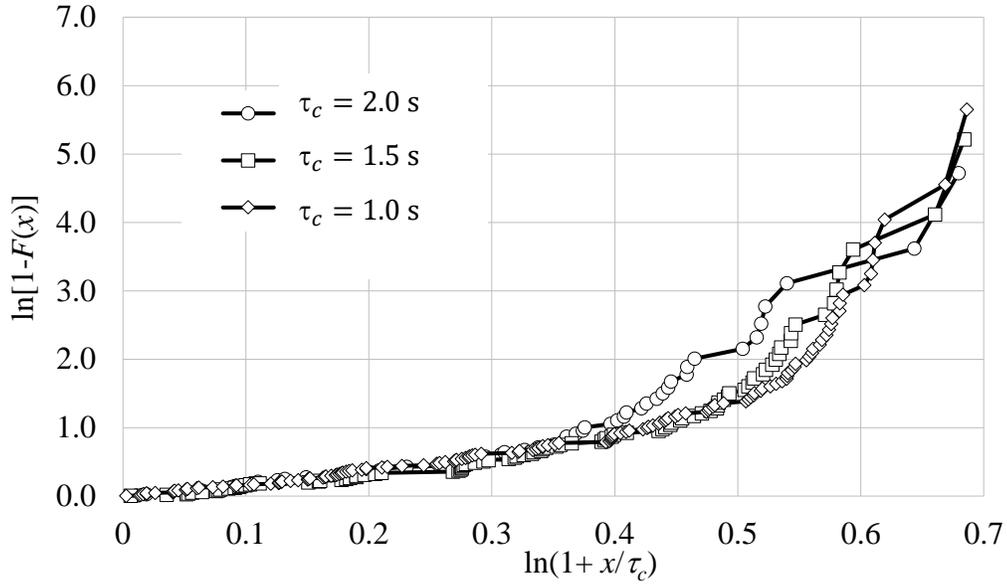


Figure 5.13. Log-log curves for instantaneous TTC range method

5.4 Identification of Failure: Longitudinal and Lateral Decomposition of Response

The sensitivity of the iTTC when there was a change in direction resulted in large discrepancies of minimum iTTC values between the two methods introduced in Section 5.3. In the case of turning encounters, the simplification of braking as the primary evasive maneuver cannot be generalized and the method should be extended to consider a change in direction or lane. In addition to proximity and longitudinal response, the framework should include identification of a lateral response as a potential indication of failure.

A lateral response in emergency conditions has been extensively explored in the literature as post-hoc trigger indicators (Klauer et al. 2006). A post-hoc trigger is defined as a single signature or multiple signatures in the driving performance data stream to identify those points in time when a driver is likely to be involved in an incident, near-crash, or crash. Lateral acceleration was also identified and applied in multiple studies including Guo et al. (2010), and Wu & Jovanis (2012). However, no indication of a significant change or proximity was explored by these authors. Hence, a very severe threshold (lateral motion or greater than 0.7 g) was adopted to reduce false positives.

The adopted process in this method first requires estimation and decomposition of the acceleration vector into the longitudinal and lateral component for each road user. After estimation of lateral and longitudinal acceleration profiles, time-series point detection techniques are explored for

identification of a change in states from “No evasion” to “Evasion.” The methodology proposed in this dissertation relies on graph theory tools. A graph-based framework for change point detection is a nonparametric approach that applies a two-sample test on a graph to find whether there is a change point within the observations. Each possible investigating point τ divides the observations into two time windows: observations before τ and observations after τ . The number of edges in the graphs G (R_G) that connects the observations from these two windows is used as an indicator of a change point, so the smaller edges increase the possibility of a change point. Since the value of R_G depends on time t , the standardized function (Z_G) is defined as:

$$Z_G(t) = -\frac{R_G(t) - E[R_G(t)]}{\sqrt{VAR[R_G(t)]}} \quad (5.14)$$

Where $E[.]$ And $VAR[.]$ are the expectation and the variance, respectively. The change point is accepted if the maxima is greater than 50% of the maximum jerking observed at a specific location with the same road users. The characterization of braking is applied before a specific threshold can be defined. The cumulative histogram of the maximum jerks adopted by each one of the road users during encounters are shown in Figure 5.14 and Figure 5.15. Based on the plots, stronger jerks are adopted by bicycles in the longitudinal axis. However, considering the need for balance on bicycles, the inverse behavior is shown in the case of the lateral component.

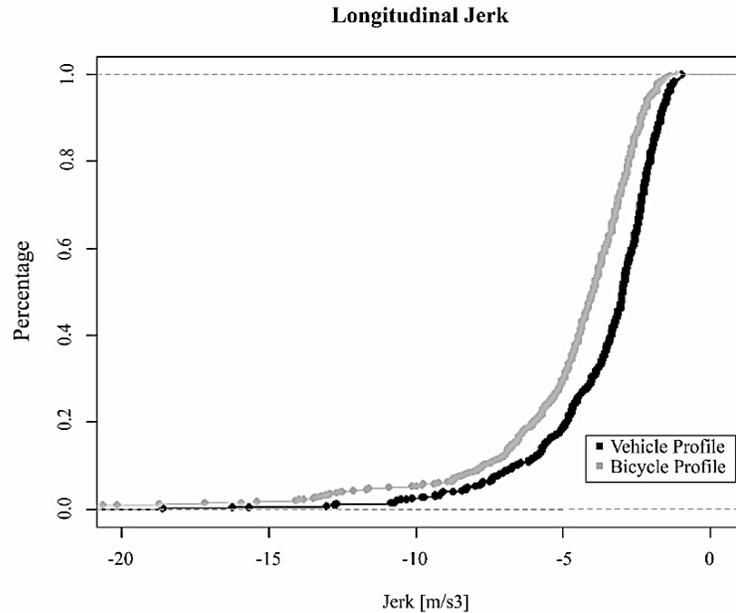


Figure 5.14. Cumulative distribution longitudinal jerk

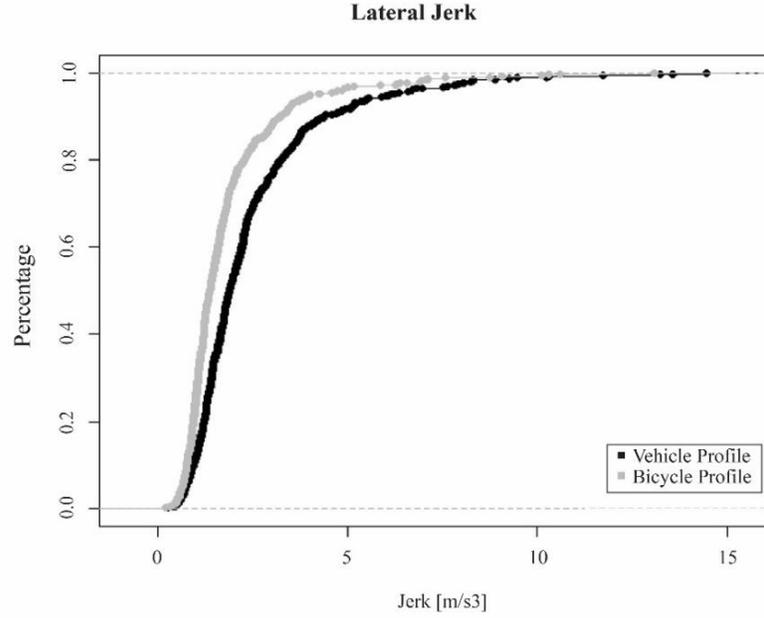


Figure 5.15. Cumulative distribution lateral jerk

5.5 Estimation of Proximity Measures Using Trajectories Without Evasion

In general, evasive maneuvers in turning movements produce a response not only as braking but also changing direction. These changes in direction significantly influence the values of iTTC providing an unstable measurement in more complex scenarios. Assuming that the author could reconstitute the trajectory in the counterfactual scenario of no evasion, it is possible to provide a better estimation of TTC. Using this approach, we can remove the assumption of constant speed and direction from iTTC. Hence, an additional step to estimate the trajectory without evasion should be conducted for a proper estimation of TTC.

Prediction of a hypothetical evasion-free trajectory might follow the stochastic theory of future vehicle motion using probabilistic approaches or reachability theory similar to Saunier et al. (2015). Let trajectory T of road user U characterized based on the sequence of positions $U = ((u_{x,1}, u_{y,1}), \dots, (u_{x,t_e}, u_{y,t_e}), \dots, (u_{x,n}, u_{y,n}))$. In addition, let $U(t_e)$ denote the characterization of the trajectory from road user U before evasion. E.g. past t_e observed positions $U(t_e) = [(u_{x,1}, u_{y,1}), \dots, (u_{x,t_e}, u_{y,t_e})]$. Finding the hypothetical evasion free trajectory requires the ability to predict the road users' future positions based on what we learned from the past. Thus, we need to apply

a method that allows us to estimate a good approximation for $U(-t_e) = [(u_{x,-t_e}, u_{y,-t_e}), \dots, (u_{x,n}, u_{y,n})]$ where $U(-t_e)$ represent the sequence of predictions under no evasion.

The extrapolation of the behavior under no evasion can be approximated using the collected trajectories from the same location. Hence, from the partial trajectory $U(t_e)$ we can compare a set of evasion-free trajectories denoted by the group $H = \{H_1, \dots, H_m\}$ where trajectories H are unaffected. Using this group of unaffected trajectories, we can find the most similar one to complete and extrapolate an evasion free behavior. To develop this approach, metrics should be specified to measure the distance of the partial trajectory $U(t_e)$ with the conflict-free in group H . The Euclidean distance between the points of the trajectories can be the most intuitive measurement. References from multiple candidates selected for a specific example with a bicycle are shown in Figure 5.16. The results show similar hypothetical collision points for the different candidates with a similar estimation of proximity measures.

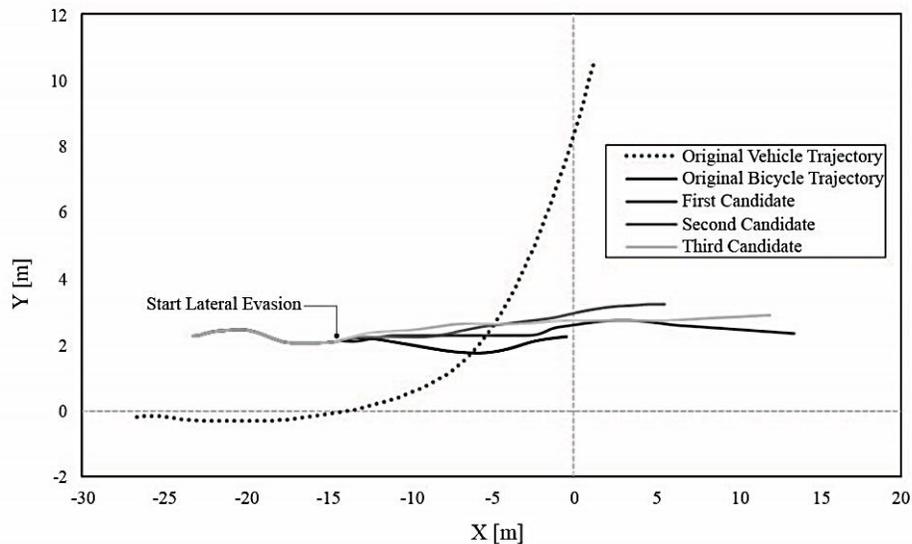


Figure 5.16. Candidate conflict-free trajectories lateral evasion

5.6 Evaluation of Traffic Conflicts

The author intended to evaluate the results using TTC and the Lomax-based approach explained in Section 2.6. However, after developing the process for identification of evasion and estimation of instantaneous TTC, the results supported that in the 414 encounters, 171 had an evasive action with the

presence of hypothetical collision point in 23 of them. The results from the estimation of instantaneous TTC are shown in Figure 5.17.

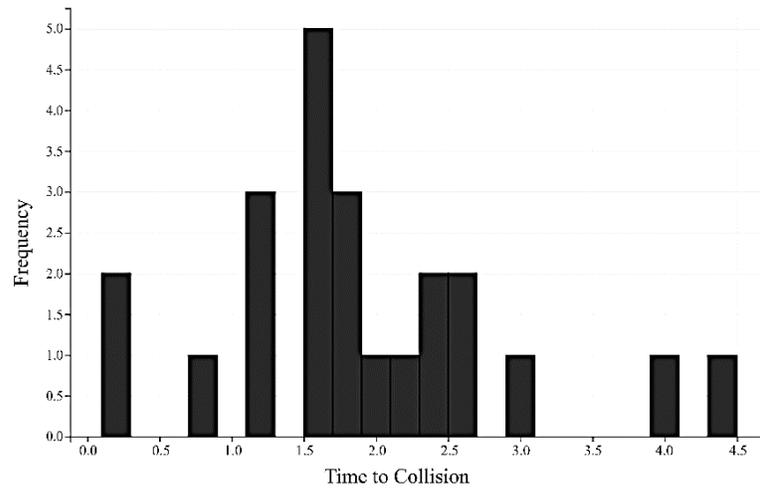


Figure 5.17. Histogram complete estimation Time-to-Collision

The results depict a low number of cases classified as failure conflicts. The same procedure was applied for 20 intersections in the project and the average number of encounters with TTC lower than specific thresholds were calculated. The results are shown in Table 5.1. The results show more realistic estimates than the ones provided when only proximity measures were considered. When adopting a threshold of two seconds, the proposed method reported 2.57 conflicts per day while the method using only iTTC was close to 10.

Table 5.1. Average number of conflicts per day all intersections InDeV Program

Indicator	One-day iTTC	One-day Failure Based trajectory
TTC < 5s	63.3	6.43
TTC < 4s	55.1	5.71
TTC < 3s	35.5	4.71
TTC < 2s	9.9	2.57
TTC < 1.5s	3.7	1.00
TTC < 1.0s	1.4	0.57

The low number of conflicts identified does not allow testing the conflict-based methodology in the identified conflicts. The results were expanded to other countries with an even smaller number of

encounters and associated conflicts. These trends were particularly evident in Scandinavian countries where there were no conflicts identified at some intersections.

5.7 Summary

The scope of this chapter focused on analyzing more complex scenarios, including turning movements as well as conflicts with vulnerable road users. Two preliminary approaches were discussed for estimation of instantaneous TTC. The first method relied on estimation of instantaneous TTC based on the relative direction between the two vehicles. The additional method defined the hypothetical collision path based on the vector defined between the centroids of the objects and decomposing the speeds along this direction. The results support significant discrepancies between these methods when there was a change in direction, altering the position of the hypothetical collision point.

The evaluation of these two estimation methods and identification of the conflicts with the associated failure based on proximity measures was implemented using the Lomax-based approach described in Section 2.4. The evaluation of the conflict was conducted based on three metrics: (1) estimates of the expected number of crashes, (2) stabilization of the estimates at lower τ_c , and (3) the Lomax-distributed delay in the responses. The two methods provided overestimation of the expected number of crashes as well as a lack of stabilization. Counterintuitive results were found in the range-based method for the probability of crash at lower separation thresholds, which indicated that the method violated causation and should not be considered for further analysis. Lack of failure in the encounters and short observation periods were observed in the diagnostics plots.

Empirical observation of the evasive maneuvers adopted by the vehicle and the bicycle in dangerous conditions indicated braking and changing direction as evasive maneuvers adopted by road users to reduce the risk of collision. Hence, the method for identification of evasion was expanded into a longitudinal and lateral decomposition of response. Limited information was provided after response, therefore evasion-free trajectories were adopted to simulate the counterfactual scenario of no response and estimation of TTC. It should be emphasized that instantaneous assumption was not present in the latter approach.

The method was successful in removing a significant number of encounters without failure. This approach resulted in only 23 events with hypothetical collision points and corresponding estimation of TTC. The results were more realistic in terms of the expected number of conflicts to observe at a

specific intersection. Due to the low number of events, it was not possible to implement an estimation of the expected number of crashes.

In general, the methodology was proven effective in removing events not meeting the failure and hypothetical collision criteria stipulated in Section 2.4. The low number of identified events restricted the study to estimating the expected number of crashes. Additional data sources should be included to provide validation of the method.

6. GENERALIZED FRAMEWORK FOR IDENTIFICATION OF TRAFFIC CONFLICTS

6.1 Background

The previous chapters provided lessons learned towards adopting a generalized framework for the identification of traffic conflicts with safety value. These chapters emphasized the need or presence of failure as a necessary component of the counterfactual definition provided by Tarko (2018). This definition was linked with an estimation of risk and the associated probability of observing a crash given this failure. Tarko (2018) indicated a very short separation threshold as an indicator of failure for most drivers. However, noisy measurements from sensing instrumentation and heterogeneity across drivers supported adding other kinematic signatures in terms of the strength of the response that indicates failure and corrective action to avoid collision. In addition, low perception of risk and short separation in small speed ranges should be considered for a proper estimation of the expected number of crashes using conflicts in the case of vehicle-vehicle encounters.

This chapter introduces a general framework for identification of traffic conflicts with safety value, which means that the probability of observing a crash greater than 0 and the conflict cannot clear itself without evasive maneuver. This method can be applied for in-vehicle instrumentation as well as an area-wide detection system relying on video or LiDAR-based systems. Discussion of the stages involved in identification is included in this chapter as a guide in the practical implementation of the method. The phases of the method include:

1. Identification of potentially dangerous interactions that due to low instantaneous TTC values are candidates for further analysis.
2. Analysis of longitudinal and lateral acceleration profiles for interactions in Phase I for identification of response due to failure.
3. Estimation of TTC by identification of evasion point and estimation of hypothetical collisions using counterfactual evasion-free trajectory.

The structure of this chapter follows these three components.

6.2 Estimation of Potentially Dangerous Interactions

Data collection periods for safety analysis may be longer than those adopted in operational studies. These findings were further confirmed in Chapter 5 where data collection in a 24-hour period was limited for identification of sufficient conflicts for estimation of the expected number of crashes in vehicle-bicycle encounters. These extended periods require analysis of thousands of trajectories and interactions between road users for identification of traffic conflicts. Application of the complete framework may require long processing times and analysis of trajectories without the potential for collision.

Processing times can be reduced by first identifying potential interactions and road users based on simplified assumptions of instantaneous TTC, which seem to provide some indication of nearness that should be further analyzed. In general, refinement of identification in tracking algorithms usually requires estimation of the distance and relative directions of the objects on the field of view (Tarko et al. 2016). Hence, the estimation of instantaneous TTC can be conducted in real-time with an estimation of TTC in post-processing.

Let us consider the right-angle conflict in Figure 6.1 with a more complex evasive maneuver. This case shows the path of vehicle 2 with an intended maneuver of turning left. Failure is caused by the driver of vehicle 1 who does not realize vehicle 2 is approaching. Similar to Section 5.3, the tracking algorithms provide the mapping trajectories from both vehicles. The trajectories are representations from a finite set $I \subset \mathbb{R}$ (I is typically a finite set of time instances at regular intervals δ) to \mathbb{R}^2 (the two-dimensional plane). The trajectories for vehicle C_i can be represented as:

$$I \subset \mathbb{R} \rightarrow \mathbb{R}^2: t \rightarrow \mathbf{C}_i(t) = [x_i(t), y_i(t)] \quad (6.1)$$

Using the relative direction and decomposition of relative speed along this vector, it is possible to estimate instantaneous TTC until the hypothetical point of collision. Since observations are repeated at time intervals δ , the vectorial representation of the direction and speed are estimated with expression (5.2) and expression (5.3), respectively. The unit direction of travel at time t is:

$$\mathbf{u}_i = \frac{\mathbf{d}_i}{|\mathbf{d}_i|} = (x_{ui}, y_{ui}) \quad (6.2)$$

and the unit vector normal to the direction is:

$$\mathbf{n}_i = (y_{ui}, -x_{ui}) \quad (6.3)$$

Using a rectangular representation of the road users with the length of vehicle i represented as l_i and the width w_i , the position of the different corners can be obtained applying the expression (5.6).

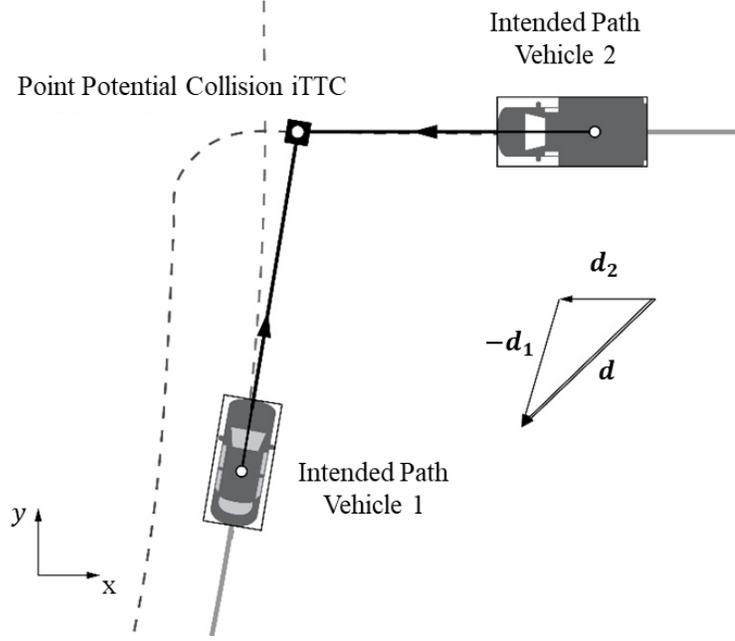


Figure 6.1. Lateral evasive maneuver for a vehicle turning left in a traffic conflict

Recapitulating the procedure described in Section 5.3.1, the obtained directions for vehicle 1 and vehicle 2 are \mathbf{d}_1 and \mathbf{d}_2 , respectively. The relative direction of vehicle 2 in relation to vehicle 1 is: $\mathbf{d} = \mathbf{d}_2 - \mathbf{d}_1$ and the corresponding relative unit direction is $\mathbf{u} = \mathbf{d}/|\mathbf{d}|$. The relative speed is $\mathbf{V} = \mathbf{V}_2 - \mathbf{V}_1$.

The relative direction of motion \mathbf{u} and corresponding normal unit vector \mathbf{n} can serve as a new base for a system of coordinates where the relative motion of the second vehicle towards the first vehicle is along y axis. Let

$$\mathbf{M} = (\mathbf{n} \ \mathbf{u}) = \begin{pmatrix} y_u & x_u \\ -x_u & y_u \end{pmatrix} \quad (6.4)$$

where the \mathbf{u} and \mathbf{n} are vertical vectors. The corners $\mathbf{P} = (x, y)$ of the two vehicles in the new coordinates are (Figure 6.2):

$$\mathbf{P}^* = \mathbf{M}^{-1}\mathbf{P} \quad (6.5)$$

Or

$$\mathbf{P}^* = \begin{pmatrix} y_u & -x_u \\ x_u & y_u \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (6.6)$$

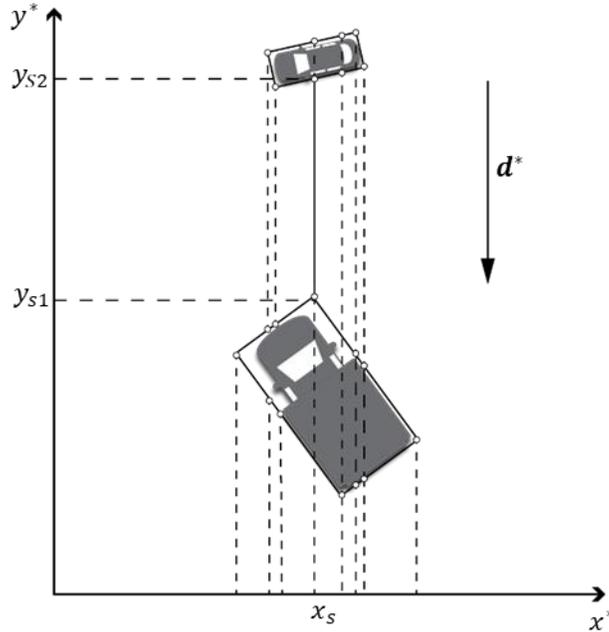


Figure 6.2. Relative reference system for estimation of iTTC

According to Figure 6.2, it is possible to estimate the shortest separation between the vehicles denoted by the straight line. The process for estimating the minimum distance between vehicles is done by pairing each corner of a vehicle with a corresponding side of the other vehicle and calculating the distance between them along the y axis. The solution is the corner-side pair separated with the shortest distance among all the pairs. The iTTC is calculated by dividing the found distance by the relative speed V (length of the speed vector V). As a reminder, iTTC does not exist if the two vehicles “instantaneously” miss each other or when the relative instantaneous speed V is zero. This calculation is repeated for all times. The outcome includes the ITTC profile and the separation distance profile.

iTTC provides a good representation of TTC if the assumption of constant relative velocity (direction and relative speed) is weakly violated. However, this method should be redefined in more complex cases as shown in Figure 6.3. In this case, the intended maneuver for vehicle 2 is turning left. Hence, the assumption of constant speed is violated since braking is expected in the counterfactual case of turning without evasion. Moreover, in order to avoid a collision, the second

vehicle abruptly changed direction. This change produced a displacement of the hypothetical collision point and a significant change in the iTTC profile. Hence, more advanced methodologies relying on estimation of hypothetical collision using evasion-free trajectories are further examined in the remainder of this chapter to provide an estimation of TTC in these scenarios.

6.3 Identification of Evasion Using Decomposition of Response

The scenario shown in Figure 6.3 depicts an evasive action conducted by the second vehicle in order to avoid a collision. Moreover, the assumption of constant speed and direction does not hold considering the intended maneuver of turning left in the counterfactual scenario. In order to find the best characterization of this counterfactual action, identification of the evasion point is an important step. The methodology in this case relies on the estimation of a significant change in the lateral acceleration profile.

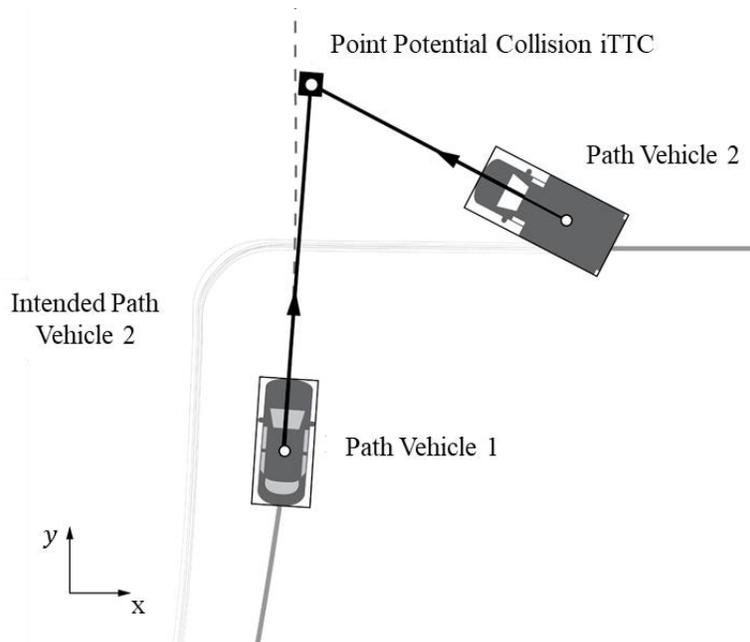


Figure 6.3. Change in point of hypothetical collision point due to evasive maneuver

Estimation of the longitudinal and lateral acceleration profiles requires obtaining heading θ_i and acceleration a_i from the position and velocities derived in expression (5.3). Since vectorial representation of the acceleration is conducted using a global XY reference system, using the

heading, the acceleration is decomposed in a local coordinate system referring to the longitudinal and lateral components of the vehicle as follows:

$$[a_{Longitudinal}, a_{Lateral}] = \begin{bmatrix} \cos(\theta_i(t)) & \sin(\theta_i(t)) \\ -\sin(\theta_i(t)) & \cos(\theta_i(t)) \end{bmatrix} \begin{bmatrix} a_{x,i}(t) \\ a_{y,i}(t) \end{bmatrix} \quad (6.7)$$

Change point detection methods are explored to detect the transition between two different states in the lateral and longitudinal acceleration profiles. The two states represent the change from the non-response to response conditions. The longitudinal and lateral acceleration profiles are characterized by a time-series data stream and the sequence of elements extracted are rate $1/\delta$:

$$a_{i,j}(t) = \{a_{i,1}, a_{i,2}, \dots, a_{i,j}, \dots\} \quad (6.8)$$

Being a_i the i th component of the acceleration representing the longitudinal and lateral components. Using the vectorial representation of these two components, detection of evasion can be defined as a hypothesis testing problem between two alternatives, the null hypothesis H_0 : "No reaction occurs" and the alternative hypothesis H_A : "A reaction occurred":

$$\begin{aligned} H_0: \mathbb{P}_{T_m} = \dots = \mathbb{P}_{T_k} = \dots = \mathbb{P}_{T_n} \\ H_A: \text{There exist } m < k^* < n \text{ such that } \mathbb{P}_{T_m} = \dots = \mathbb{P}_{T_k} \neq \mathbb{P}_{T_{k^*+1}} = \dots = \mathbb{P}_{T_n} \end{aligned} \quad (6.9)$$

Where \mathbb{P}_{T_i} is the probability density function of the sliding window start at point T_i and k^* is a change point. These trends are visualized in Figure 6.4.

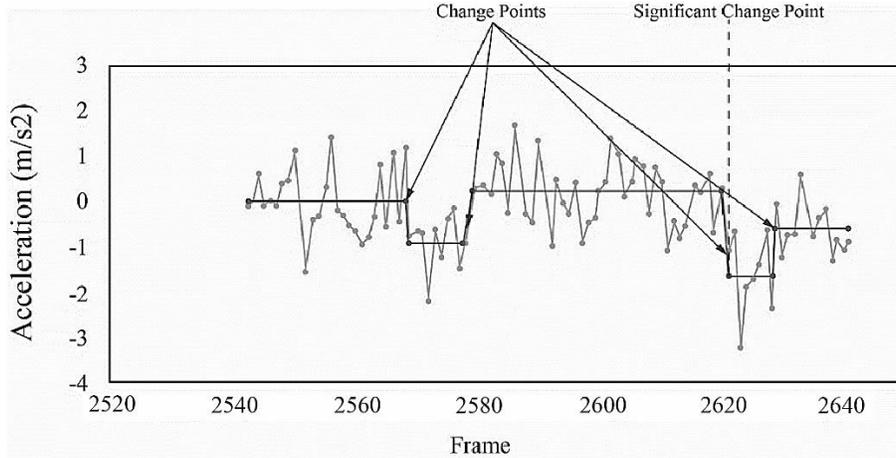


Figure 6.4. Change-point-detection acceleration profile

Different machine learning algorithms have been adapted to cope with the necessity to identify these transition states. Supervised and unsupervised methods have been proposed in the literature for change point detection. Considering the nature of the problem to segment and identify

the transition between non-evasion and evasive action from drivers, unsupervised learning methods are applied in this component. Aminikhanghahi and Cook (2017) summarized multiple unsupervised machine learning algorithms applied for change point detection, which are shown in Figure 6.5.

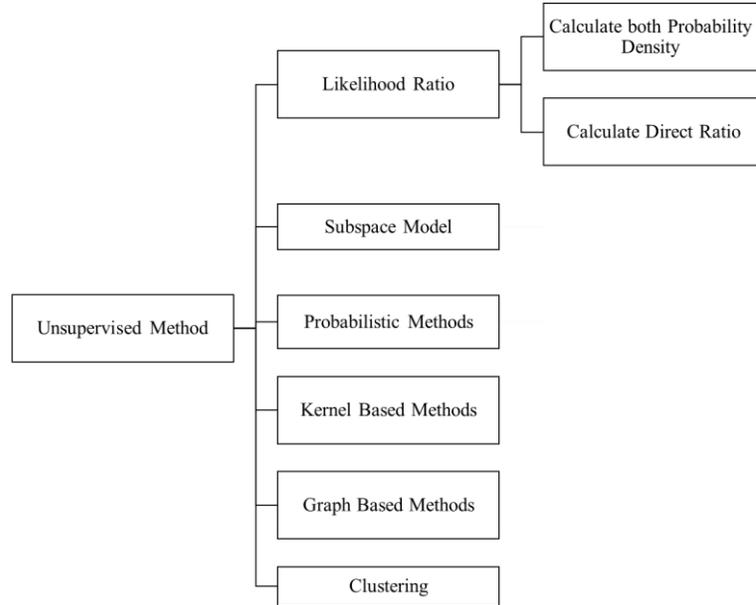


Figure 6.5. Unsupervised methods change point detection (Aminikhanghahi & Cook, 2017)

From a historical perspective, the likelihood ratio methods were the first applied to test the difference in the probability density function in a window before and after the point of analysis. They used parametric and non-parametric characterization of the density functions (Kawahara & Sugiyama, 2012). Nonparametric methodologies were preferred considering more flexible functional representation as compared to pre-specified parametric models. Non-parametric approaches use the ratio of the probability densities of the windows directly without needing to perform density estimation, making it a simpler approach.

Kawahara and Sugiyama (2012) modeled the density ratio between two consecutive intervals χ and χ' by non-parametric Gaussian Kernel model characterized based on the expression:

$$g(\chi) = \frac{p(\chi)}{p'(\chi')} = \sum_{l=1}^n \theta_l K(X, X_l) \quad (6.10)$$

$$K(X, X') = \exp\left(-\frac{\|X - X'\|^2}{2\sigma^2}\right) \quad (6.11)$$

In this expression $p(\chi)$ characterizes the probability distribution of interval χ , $\theta = (\theta_1, \dots, \theta_n)^T$ are parameters trained based on sample, X is the sliding window, and $\sigma > 0$ is the kernel parameter. The training phase estimates the parameter θ that minimize the dissimilarity measure. Hence, estimating the density ration estimator for samples χ_t and χ_{t+n} in the testing phase, the higher the dissimilarity measure, the more likely the point is a change point.

Kernel-based methods also have been applied in unsupervised learning with the intention of testing the homogeneity of the data in time series using sliding windows (Cook & Krishnan, 2015). Following the objective of kernel functions, non-linear transformation in a Kernel Hilbert space are applied using a feature map $\Phi(X) = k(X, \cdot)$. The kernel Fisher discriminant ratio is proposed as a measure of homogeneity between windows. Given the two windows of observation, the kernel Fisher discriminant ratio (KFDR) between two samples is defined as:

$$KFDR(X_1, X_2) = \frac{n_1 n_2}{n_1 + n_2} \langle \hat{\mu}_2 - \hat{\mu}_1, (\hat{\Sigma}_w + \gamma I)^{-1} (\hat{\mu}_2 - \hat{\mu}_1)_H \rangle \quad (6.12)$$

Where γ is a regularization parameter and:

$$\hat{\Sigma}_w = \frac{n_1}{n_1 + n_2} \hat{\Sigma}_1 + \frac{n_2}{n_1 + n_2} \hat{\Sigma}_2 \quad (6.13)$$

The easiest way to determine whether a change point exists between two windows is to compare the KFDR ratio with a threshold value. However, likelihood ratio methods and kernel function distort the original units of jerk with transformations.

Considering the cumbersome transformation and evaluation of the thresholds in the previous methods, graph theory tools were applied in this dissertation for identification of the change points described in Section 5.4. A graph-based framework for change point detection is a nonparametric approach that applies a two-sample test on a graph to determine whether there is a change point within the observations. Each possible investigating point τ divides the observations into two-time windows. Observations before τ and observation after τ . The number of edges in the graph $G(R_G)$ that connects the observations from these two windows is used as an indicator of the change point so that the smaller edges increase the possibility of a change point. Since the value of R_G depends on time t , the standardized function (Z_G) is defined as:

$$Z_G(t) = -\frac{R_G(t) - E[R_G(t)]}{\sqrt{VAR [R_G(t)]}} \quad (6.14)$$

The change point is accepted if the maxima is greater than 50% of the maximum jerk observed at a specific location with the same road users. The advantage of this method relies on the definition of the threshold using the units of the acceleration profiles, which are easier to interpret by the final user.

The method to identify a change point is conducted in the lateral and longitudinal components separately. The frame of the first change point is retained for the next phase of the method denoted by T_e .

6.4 Estimation of Hypothetical Collision Point via Evasion Free Trajectories

Complex evasive maneuvers provide a significant change in the estimation of iTTC as explained in the scenario shown in Figure 6.3. Based on the definition of a traffic conflict proposed in Tarko (2018), the counterfactual outcome represented as a crash can be obtained by removing evasion and simulating the trajectory without a response. This method allows estimation of TTC instead of iTTC since no assumptions are imposed in terms of the constant relative speeds between the road users. This methodology also provides a better characterization of the proximity measure and risk estimation of the conflicts.

In order to predict the trajectory adopted by the vehicle in the “what-if” scenario of no evasion, Lefevre et al (2014) discussed two possible groups of methods. The first group uses empirical knowledge of dynamics from vehicles to simulate the most probable response; and the second group, which is more appealing in area-wide detection systems, is the characterization of evasion-free trajectories collected from the same location and type of movement. Hence, from the partial trajectory before evasion for road user $C_i(t_e)$, it is possible to conduct a comparison from a set of evasion-free trajectories $H = \{H_1, \dots, H_n\}$ to determine the most similar one. Relying on the assumption that the user will continue along a similar path, the movement of the road user in this “what-if” scenario can be extrapolated, and estimation of the hypothetical collision point can be applied using the simulated trajectory.

Figure 6.6 illustrates the application of the method in the complex scenario described in Section 6.2, which identifies the evasive maneuver represented by a significant change in the

lateral jerk. After removing this change in the lateral jerk, the trajectory before frame T_e is applied in order to estimate the most similar evasion-free trajectory that allows extrapolation.

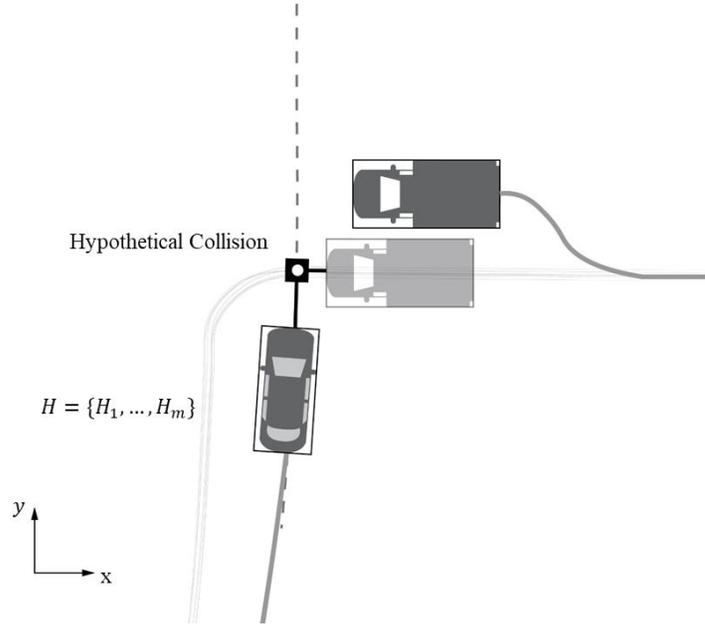


Figure 6.6. Estimation of hypothetical collision using most similar evasion-free trajectory

6.5 Similarity Measures between Trajectories

Different similarity metrics are reported in the literature in order to obtain the most similar trajectory in H from the partial one before evasion. Among the most common similarity metrics, authors have used Euclidean distance, the modified Hausdorff metric (Atev et al., 2010), the longest common subsequence metric (Buzan et al., 2004), and its translation and rotation invariant version of the quaternion-based rotationally Invariant LCS (Hermes et al., 2009).

The Euclidean distance between the points of the trajectories is the most intuitive measure. Consecutive points reported at a constant rate $1/\delta$ are compared and the average cumulative distance between points is estimated as:

$$R_{ij} = \frac{\sum_{m=1}^K \left| |X_i - H_j| \right|^2}{K}, 1 \leq i, j \leq K \quad (6.15)$$

Being K the number of centroids compared between trajectories X and H . Implementation of this method was conducted in this dissertation since the processing time and noise from sensors was not considerable. However, in the case of processing time constraints and high signal to noise

measures, Buzan et al. (2004) discussed the Longest Common Subsequence (LCSS) as the most appropriate similarity measure in these conditions.

Let U and V be two trajectories of moving objects with norm n and m respectively, where $U = ((u_{x,1}, u_{y,1}), \dots, (u_{x,n}, u_{y,n}))$ and $V = ((v_{x,1}, v_{y,1}), \dots, (v_{x,m}, v_{y,m}))$. For trajectory U , let the $\text{Head}(U)$ be the sequence $\text{Head}(U) = ((u_{x,1}, u_{y,1}), \dots, (u_{x,n}, u_{y,n-1}))$. Given an integer δ and a real number $0 < \epsilon < 1$, $LCSS_{\delta,\epsilon}(U, V)$ is defined as:

$$LCSS_{\delta,\epsilon}(U, V) = \begin{cases} 0 & \text{if } U \text{ or } V \text{ is empty} \\ 1 + LCSS_{\delta,\epsilon}(\text{Head}(U), \text{Head}(V)) & \text{if } |u_{z,n} - v_{z,m}| < \epsilon \text{ and } n - m \leq \delta \\ \max(LCSS_{\delta,\epsilon}(\text{Head}(U), V), LCSS_{\delta,\epsilon}(U, \text{Head}(V))) & \text{Otherwise} \end{cases} \quad (6.16)$$

The constant δ controls how far in time we can go to match a given point from one trajectory to another trajectory, and the constant ϵ refers to the matching threshold or bounding error. The advantages of the metric include a lower processing time since the metric does not require finding a global minimum as compared to heuristic with the Euclidean distance between neighboring points. In addition, this metric tends to be relatively robust to noise and possible outliers. Based on LCSS, it is possible to define two similarity metrics, S1 and S2. The similarity metric S1 between two trajectories U and V , given is defined as:

$$S1(\delta, \epsilon, U, V) = \frac{LCSS_{\delta,\epsilon}(U, V)}{\min(n, m)} \quad (6.17)$$

Translation can be applied to find similar and parallel trajectories located in shifted locations in space. A translation is defined as a function $f_{c,d}(U) = ((u_{x,1} + c, u_{y,1} + d), \dots, (u_{x,n} + c, u_{y,n} + d))$. It is possible to define the values c and d that can maximize S1. Hence, it is possible to define similarity measure S2 as:

$$S2(\delta, \epsilon, A, B) = \max_{f_{c,d} \in F} S1(\delta, \epsilon, A, f_{c,d}(B)) \quad (6.18)$$

This metric ranges from 0 to 1. The authors of the report that it is possible to find $LCSS_{\delta,\epsilon}(U, V)$ in $O(\delta(n + m))$ time and $S2(\delta, \epsilon, U, V)$ in $O((n + m)^3 \delta^3)$ time.

6.6 Estimation Time-to-Collision

This section discusses the estimation of TTC in line with the limitations of the application of instantaneous TTC in more complex approaches. Recapitulating Figure 6.7, in the case of longitudinal responses from naturalistic driving studies, the counterfactual behavior of the driver in a rear-end collision is approximated in the dashed line with a prediction of their motion without changing speed or paths. According to Tarko (2019), the iTTC is equivalent to the actual TTC until the braking begins. Hence, the iTTC may be a reasonable approximation of TTC in the case of the assumption of constant relative speed.

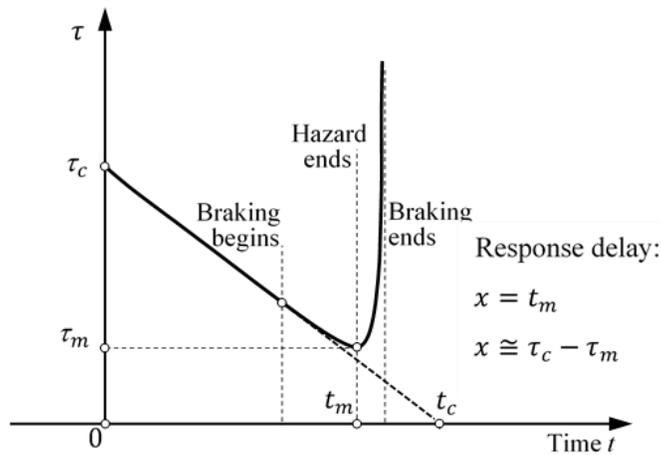


Figure 6.7. Representation of a rear-end conflict instantaneous Time-to-Collision

However, in the case of a change in path or speed, such as the one shown in Figure 6.8, estimating T requires predicting the originally intended trajectories under the lack of evasion. Extrapolating the trajectory by learning from evasion-free trajectories can be adopted, and estimation of the hypothetical time to crash T_c implemented. TTC is estimated as $TTC = T_c - t_e$. One should emphasize that in the case of a lateral response, beginning of evasion and an end of hazard is located in the same point t_e as compared to braking (see Figure 6.8).

6.7 Summary

This chapter summarized the lessons learned from the methods applied in naturalistic driving studies and the InDeV program for identification of traffic conflicts. It provides a framework that

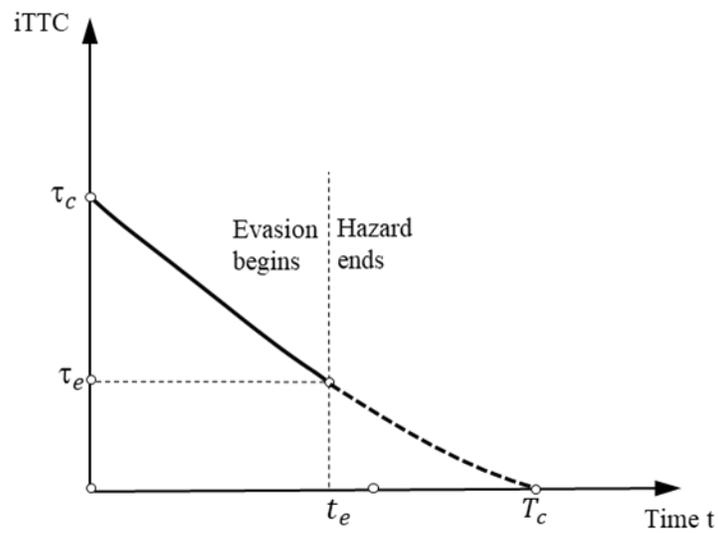


Figure 6.8 Representation of turning conflict Time-to-Collision

can be easily reproduced by other researchers to evaluate safety with surrogate measures. This chapter discussed the following three primary components for identification of traffic conflicts:

1. Identification of potential road users' interactions that due to low instantaneous TTC values are candidates for further analysis.
2. Analysis of longitudinal and lateral acceleration profiles for interactions in Phase I for identification of response due to failure.
3. Estimation of TTC by identification of evasion point and estimation of hypothetical collision using counterfactual evasion free trajectory.

Different methodologies were described targeting the components for identification of evasion in longitudinal and lateral components. In addition, similarity metrics between trajectories as well as estimation of TTC were discussed to provide the user with a holistic review of the lessons learned from Chapters 4 and 5. In general terms, the methodology can be summarized in the following steps:

1. Identify potential pairs in conflicting situation based on iTTC. State a liberal threshold of iTTC for identification of potential pairs (< 5 seconds)

2. Among all the detected pairs, find the ones with a significant change in lateral or longitudinal acceleration (jerk indicating potential avoidance action).
3. Find another vehicle free of evasion whose trajectory is close to the trajectory of the vehicle performing evasion.
4. In the pair of vehicles formed in step 1, replace the evading vehicle with the vehicle free of evasion found in step 3 and check if the two vehicles would collide.
5. If yes, then check the necessary three conditions of a conflict a claim or reject a conflict. Perform steps 2-5 for all the remaining potential pair of vehicles in conflicting conditions from step 1.

7. SEVERITY OF HYPOTHETICAL CRASH GIVEN CONFLICT

7.1 Background

Defining the severity of a hypothetical crash in the counterfactual scenario is an important component in the estimation of the expected number of crashes. The characterization of severity allows targeting not only locations or conditions that provide a high frequency of crashes but also the ones that lead to severe interactions with more serious consequences. The application of severity analysis in the following framework is two-fold: The first application evaluates the perception of hazards by drivers. Chapter 4 addressed non-stable crash estimates in low-speed scenarios. Under the lack of a hazard, it was observed that drivers were able to accept dangerous separation values in time and space at low speeds since they are able to stop almost immediately, thereby removing the risk. In this case, an iterative process was applied to identify the minimum speed that provided convergence in the crash estimates for the sequence of decreasing separation thresholds as theoretical stipulations.

The second application of severity evaluates reporting and further categorization of the severity in a hypothetical crash. A comprehensive safety analysis based on conflicts should include estimation of not only frequency but also severity. These two estimates provide short-term safety evaluation based on conflicts translated into economic loss. Severity analysis of crashes has been extensively investigated in the literature considering the need to further understand the environmental, vehicle, and driver conditions that may increase the consequences of these events.

Introducing severity into the evaluation of safety using conflicts can be conducted using two different approaches (Tarko, 2019): (1) predicting the frequency of crashes for each one of the severities or (2) predicting the conditional probability for each severity outcome. Following the second approach can include calculating the probability of the hypothetical crash being at a specific level of severity using statistical models. The individual characteristic of the events, including environmental and crash attributes, can be considered in the severity definition. The expression characterizing the estimated frequency of crashes for severity S_i is:

$$Q_{c,s_i} = n \cdot P(C|N, \tau_c) \cdot P(s_i) \quad (7.1)$$

where:

Q_{c,s_i} is the estimated frequency of crashes for severity S_i .

n is the number of traffic conflicts observed during the observation period.

$P(C|N, \tau_c)$ is the is the probability of crash (C) given no response (N) at time when $\tau = \tau_c$

$P(s_i)$ is the probability of outcome severity estimated using a probability model of collision outcome severity.

The objective of this chapter is to provide a parametric characterization of $P(s_i)$, representing the probability model of collision outcome severity. This chapter discusses the methodologies reported in the literature for parametric estimation of crash severity and the data sources applied to structure the statistical models. The remaining sections include a description of the most significant covariates and a discussion of the primary findings.

7.2 Literature Review

The nature of crashes as multi-causal events has produced extensive research of the covariates affecting their severity and probable consequences. Sobhani et al. (2011) identified and categorized crash severity research as evaluating the influence of (1) the road and environmental characteristics (transportation research), (2) the dynamics during pre-crash (crash analysis research), and (3) the influence of human intrinsic characteristics (medical research). This literature review investigated different documents with special emphasis on crash analysis research since the dynamics before a crash are the most influential contributors towards the observed severity.

Traditional transportation research evaluates the influence of road and environmental characteristics using multiple statistical models depending on the nature of the dependent variable binary and multiclass classification methods applied. The most common methodologies include logistic regression models (Jung et al., 2010; Liu & Dissanayake, 2009; Rana et al., 2010), ordered probit models, (Christoforou et al., 2010; Quddus et al., 2010; Tarko et al., 2010), proportional odds models (Quddus et al., 2010; X. Wang & Abdel-Aty, 2008; Wang et al., 2009), and machine learning algorithms including support vector machine (Li et al., 2008).

The area-wide and in-vehicle tracking systems described in Chapter 4 allow researchers to obtain dynamic information about pre-crash conditions. These covariates are more relevant for the estimation of hypothetical severity in the case of the counterfactual scenario of a crash

(Munyazikwiye et al., 2016). Table 7.1 is a review of the relevant crash analysis publications, emphasizing their data sources and important variables as well as their methodologies.

Among the most important predictors included in crash research analysis are Delta-V (Kononen et al., 2010), jerk, kinetic energy (Munyazikwiye et al., 2016), and occupant impact velocity (OIV). Delta-V is defined as the difference between impact velocity and separation velocity while OIV corresponds to the largest difference in velocity, lateral and longitudinal are treated, respectively, at impact. These variables are included to structure the crash reporting and severity models in this chapter.

The most common methods applied for the estimation of severity using dynamic information include logistic regressions (Farmer et al., 1997; Prato et al., 2010), fuzzy logic methods (Munyazikwiye et al., 2016), and mathematical models based on Newtonian mechanics (Buzeman et al., 1998a, 1998b; Wood & Simms, 2002). This chapter applies logistic regression considering the extensive use of this method in the literature.

Table 7.1. Summary of crash analysis research publications

Publication	Data Source	Categorization of Severity	Key Predictors	Source of Predictors	Model
Kononen et. al., 2011	NASS-CDS database	Injury Severity Score (ISS)	Delta-V	WinSmash	Logistic Regression
Munyazikwiye et al., 2016	Acceleration data	Dynamic crash (cm)	Jerk and Kinetic energy	Computation from data	Fuzzy logic
Gabauer & Gabler, 2008	EDR data from NHTSA	AIS scale	Delta-V and Occupant Impact Velocity (OIV)	WinSmash/CRASH3	Binary Logistic
Tsoi & Gabler, 2015	EDR data from NASS-CDS 2000-2013 cases	Maximum abbreviated injury scale (MAIS)	Occupant impact velocity (OIV), Acceleration severity index (ASI), Vehicle pulse index (VPI), Maximum delta-v (delta-v)	Manual for Assessing Safety Hardware and ISO/TR 12353-3, U.S.	Binary Logistic
Viano & Parenteau, 2010	1996-2007 NASS-CDS	MAIS	Delta-V	Direct Computation	Basic Classification

7.3 Data Collection

The analysis of severity requires further investigation of the observed rear-end crashes with a vehicle in front from the SHRP2 study. The SHRP2 initiative defines three different severity categories of crashes, which are described below (Hankey, 2016).

I - Severe Crash: This category includes any crash that includes an airbag deployment; any injury of driver, pedal cyclist, or pedestrian; a vehicle rollover; a high Delta V; or a crash that requires vehicle towing. Injury, if present, should be sufficient to require a doctor's visit, including those self-reported and those apparent from video. A high Delta V is defined as a change in the speed of the subject vehicle in any direction during impact greater than 20mph (excluding curb strikes) or acceleration on any axis greater than $\pm 2g$ (excluding curb strikes).

II - Police-reportable Crash: This category includes any police-reportable crash that does not meet the requirements for a Level I crash that includes sufficient property damage which is police-reportable (minimum of ~\$1,500 damage, as estimated from video). This category also includes crashes that reach acceleration on any axis greater than $\pm 1.3g$ (excluding curb strikes). If there is a police report, this strike will be noted. Most large animal strikes and sign strikes are included here.

III - Minor Collision: Most other crashes not included above are Level III crashes, which are defined as including physical contact with another object but with minimal damage. This category includes most road departures (unless criteria for a more severe crash are met), small animal strikes, all curb and tire strikes potentially in conflict with oncoming traffic, and other curb strikes with an increased risk element (e.g., would have resulted in worse collision had curb not been there, usually related to some kind of driver behavior or state).

The definitions of these severities provide the opportunity to estimate a model evaluating the likelihood of crash reporting to the police in the first stage. The second stage includes a more disaggregated analysis for each one of the severities. Multiple parameters were considered for structuring the models and are summarized in Table 7.2. These factors span crash specific characteristics jointly with disaggregate vehicle, driver, and environmental characteristics.

A preliminary analysis showed a strong influence from the event-specific characteristics in determining the reporting and severity of the crash. The influence of these covariates is expected and beneficial in the attempt of the author to implement these models in detection and tracking

systems using roadside instrumentation or in-vehicle systems. In general, event-specific characteristics are factors that can be extracted using trajectory information from the road users.

Table 7.2. Categories of covariates included in severity analysis

Category	Variables
Event Specific Characteristics	<ul style="list-style-type: none"> ▪ Adopted speed of Vehicle before Collision ▪ Difference of vehicle speeds before Collision ▪ Acceleration X Axis at Collision ▪ Acceleration Y Axis at Collision ▪ Acceleration Z Axis at Collision ▪ Yaw Rate X Axis at Collision ▪ Yaw Rate Y Axis at Collision ▪ Yaw Rate Z Axis at Collision
Vehicle Characteristics	<ul style="list-style-type: none"> ▪ Weight Instrumented Vehicle ▪ Weight additional vehicle ▪ Class Instrumented Vehicle ▪ Class additional vehicle ▪ Airbag Deployment ▪ Mechanical defects (Brake system, Wheels , Tires)
Driver Characteristics	<ul style="list-style-type: none"> ▪ Gender ▪ Age of driver ▪ Income Level ▪ Education ▪ Driver Engaging Secondary Task (Cellphone, Children, among others) ▪ Driver Impairment (alcohol, drugs)
Environmental Characteristics	<ul style="list-style-type: none"> ▪ Infrastructure (Roadway Alignment, Roadway Sight Distance, among others) ▪ Lighting ▪ Weather ▪ Surface Conditions ▪ Traffic Flow ▪ Traffic Density ▪ Traffic Control ▪ Construction Zone ▪ Alignment

Damage analysis can be conducted to include additional information from the crash events. Crashes can be decomposed in the approach period and the separation period. As specified by McHenry (1969), the approach period is the time when the maximum crush of a crash occurs. Hence, the velocities of the two vehicles are assumed to be equal at the moment of the maximum

crash as is denoted by V_c . The initial kinetic energy of the two-vehicle system is characterized based on the expression:

$$KE_0 = \frac{1}{2}M_1(V_1)^2 + \frac{1}{2}M_2(V_2)^2 \quad (7.2)$$

From conservation of momentum, the common velocity V_c can be obtained as:

$$V_c = \frac{M_1V_1 + M_2V_2}{M_1 + M_2} \quad (7.3)$$

The change in kinetic energy of the system during the approach period is:

$$E_A = KE_0 - KE_C = \frac{1}{2}M_1V_1^2 + \frac{1}{2}M_2V_2^2 - \frac{1}{2}(M_1 + M_2)V_c^2 = \frac{1}{2}\left(\frac{M_1M_2}{M_1 + M_2}\right)(V_2 - V_1)^2 \quad (7.4)$$

Since $V_2 - V_1$ is reported from the radar measurement, the change of kinetic energy in the system can be estimated and included for modeling. The preliminary statistics of the dynamic variables and the energy released in the crash are categorized by severity in Table 7.3 through 7.5.

Table 7.3. Descriptive statistics dynamic variables severe crash

Crash Severity	Variable	N	Mean	Std Dev	Min	Max
I – Severe Crash	Speed Before Collision GPS (m/s)	31	10.100	7.570	0.000	27.900
	Difference in speed Before Collision (m/s)	28	-6.351	4.869	-24.256	-1.071
	Acceleration X Axis at Collision	31	-2.609	0.653	-4.836	-1.642
	Change Kinetic Energy (Kg m ² /s ²)	28	32196.66	84887.35	391.96	449849.27
	Acceleration Y Axis at Collision	31	0.487	0.684	0.052	3.822
	Acceleration Z Axis at Collision	31	-1.204	0.223	-1.966	-0.974
	Yaw Rate X Axis at Collision	31	13.742	26.312	0.000	149.915
	Yaw Rate Y Axis at Collision	31	26.089	54.669	0.000	265.359
Yaw Rate Z Axis at Collision	31	14.602	47.079	0.000	265.359	

Table 7.4. Descriptive statistics dynamic variables police-reported crash

Crash Severity	Variable	N	Mean	Std Dev	Min	Max
II - Police-reportable Crash	Speed Before Collision GPS (m/s)	47	4.256	3.102	0.167	14.500
	Difference in speed Before Collision (m/s)	45	-2.925	1.787	-7.126	4.750
	Change Kinetic Energy (Kg m ² /s ²)	45	4355.28	3616.51	0	16190.45
	Acceleration X Axis at Collision	47	-1.537	0.393	-2.742	-0.370
	Acceleration Y Axis at Collision	47	0.220	0.203	0.032	0.882
	Acceleration Z Axis at Collision	47	-1.167	0.237	-2.253	-0.977
	Yaw Rate X Axis at Collision	47	6.753	10.378	0.000	62.112
	Yaw Rate Y Axis at Collision	47	13.029	9.865	0.000	53.332
	Yaw Rate Z Axis at Collision	47	7.819	8.387	0.000	39.023

Table 7.5. Descriptive statistics dynamic variables minor collision

Crash Severity	Variable	N	Mean	Std Dev	Min	Max
III - Minor Collision	Speed Before Collision GPS (m/s)	40	2.050	2.264	0.000	10.306
	Difference in speed Before Collision (m/s)	37	-1.654	0.978	-4.356	-0.139
	Change Kinetic Energy (Kg m ² /s ²)	37	1489.93	1818.04	7.39	7590.23
	Acceleration X Axis at Collision	39	-0.964	0.329	-2.333	-0.378
	Acceleration Y Axis at Collision	39	0.134	0.096	0.015	0.484
	Acceleration Z Axis at Collision	39	-1.067	0.050	-1.180	-0.989
	Yaw Rate X Axis at Collision	39	18.428	59.447	0.000	347.309
	Yaw Rate Y Axis at Collision	39	19.820	69.207	0.000	436.412
	Yaw Rate Z Axis at Collision	39	15.634	68.132	0.000	426.006

The values of the results are in line with the expectations. In general, there is a clear distinction in terms of the speed, difference in speed, and acceleration value. These characteristics appear to exert extensive influence on the severity definition. Additional modeling techniques are presented in the next section.

7.4 Methodology

The primary objectives are estimation of the probability of a specific crash being reported and defining its potential severity. In line with these objectives, two different discrete choice models were structured to estimate the crash reporting and the crash, independently. A crash was considered as reported if it was inside Severity I and Severity II. The reporting model is a classification problem with two categories while the severity model includes three categories.

Estimation using discrete choice models has been extensively used by different authors to evaluate the severity of collisions as reported in Section 7.2. Traditional ordered probability models can be applied considering the ordinal nature of injury data. Ordered probability models are derived by defining an unobserved variable z for modeling the ordinal ranking of data. This unobserved variable z is usually specified by a linear function characterized by the following expression (Washington et al., 2011):

$$z = \beta X + \varepsilon \quad (7.5)$$

Using expression (7.5) and assuming the random disturbance ε as normally distributed across observations with a mean equal to zero and a variance equal to 1, an ordered probit model can be estimated with the probability of each ordered category being selected characterized by the equations:

$$\begin{aligned} P(y = 1) &= \Phi(-\beta X) \\ P(y = 2) &= \Phi(\mu_1 - \beta X) - \Phi(\beta X) \\ P(y = I) &= 1 - \Phi(\mu_{I-1} - X) \end{aligned} \quad (7.6)$$

Where Φ is the cumulative normal distribution. This is the simplest representation for evaluating an ordered probit model. However, ordered probit models refer to potential problems for modeling crash data (Savolainen et al., 2011). The first problem is its susceptibility for addressing possible underreported crashes. The second problem is the restriction that traditional ordered probability models place on the method that the variables influence outcome probabilities.

The shift in thresholds in an ordered model is constrained to move in the same direction. This produces a major restriction on evaluating the variables influencing each type of severity, which is a primary objective of this dissertation.

Since the major specifications of the ordered probit model may affect the estimation of the factors influencing each of the injury-severity categories, a multinomial logit model is considered a more appropriate technique for crash severity estimation (Savolainen, et al. 2011). Multinomial logit models are the traditional technique that estimates the probability of three or more possible outcomes by not considering the order. Using the general form of the multinomial logit model with the disturbance terms independently and identically distributed as generalized extreme value-distributed, the probabilities for determining the severity of a crash is given by the expression:

$$P(i) = \frac{e^{U_i}}{\sum_{i=1}^m e^{U_i}} \quad (7.7)$$

Where $P(i)$ denotes the probability of a crash being of a specific severity and u_i are the indirect utility functions for crash severity (Washington et al., 2011). The indirect utility is defined by a linear function that determines the injury outcome i for observation n as:

$$U_{in} = \beta_i X_{in} + \varepsilon_{in} \quad (7.8)$$

Where β_i is the vector of estimable parameters, X_{in} is a vector of observable characteristics that influence the injury severity, and ε_{in} is the disturbance term that includes unobserved effects in the model. Using the frequentist approach, estimation of the parameter vectors β is conducted by maximizing the log-likelihood function:

$$LL = \sum_{n=1}^N \left(\sum_{i=1}^I \delta_{in} \left[\beta_i X_{in} - LN \sum_{\forall I} e^{\beta_i X_{in}} \right] \right) \quad (7.9)$$

Being I is the total number of outcomes, δ_{in} equals to one if the observed discrete outcome of observation n is i or zero; otherwise, and the additional parameters are as previously specified. Estimation problems for multinomial logit models are observed in the case of complete and quasi complete separation of the observations (Albert & Anderson, 1984). In this case, there is an absence of a finite maximum of the likelihood function. Following the notation in Albert and Anderson, (1984), a complete separation of sample points is defined where exists a vector $\alpha \in \mathbb{R}^p$, such that for all $i \in E_j$ and for $j, t = 1, \dots, g (j \neq t)$

$$(\alpha_j - \alpha_t)^T x_i > 0 \quad (7.10)$$

This means that vector α can allocate all the observations for a specific discrete category I . Whenever there is a complete separation of the data points, the maximum likelihood estimates $\hat{\alpha}$ does not exist. In addition, the vector $\alpha \in \mathbb{R}^v$ provides a quasi-complete separation of the sample data points if for all $i \in E_j$ and for $j, t = 1, \dots, g (j \neq t)$

$$(\alpha_j - \alpha_t)^T x_i \geq 0 \quad (7.11)$$

With equality for at least one (i, j, t) triplet. Let us denote $j(i)$ as the value of j which $i \in E_j$. Let $Q(\alpha)$ refer to the set of index values i that satisfy the previous equality. Hence, the points x_i are defined as quasi-separated with respect to α . Complete and quasi-complete separation is a prevalent problem in the analysis of data with a modest sample size and a large number of covariates such as the one presented in Section 7.3 (Ghosh et al., 2018). A potential solution is to remove some covariates, but it can result in omitting important variables related to the discrete model (Zorn, 2005). Other solutions involve penalizing the maximum likelihood or modified versions of the likelihood estimation to restrict solutions on a finite scale (Heinze & Schemper, 2002). Bayesian estimation can be implemented as a convenient solution with a prior specification on the regression coefficients. Frequentist and Bayesian estimations of the models are included in the analysis for comparison.

Bayesian econometric and statistical analysis has been introduced in different problems in transportation. The advantages of Bayesian approaches include consideration of parameters as random variables, computability with the likelihood principle, the possibility of introducing prior knowledge, and more information derived from predictive posteriors. Bayesian analysis requires the specification of a prior distribution over the coefficient estimates. Normal priors are applied in the estimation process on β as:

$$\pi(\beta) = N(\beta, \Omega_\beta) \quad (7.12)$$

Based on the results in Gosh et al. (2018), in logistic regression under a multivariate normal prior for β , normal priors provided the fastest convergence of the sampler, reasonable scales for the posterior draws of β , and comparable predictive performance than other priors. The posterior density of the parameter β is:

$$p(\beta|Y) \propto L(Y|\beta) \pi(\beta) \quad (7.13)$$

Although in rare cases the explicit forms of the posterior density may be acquired based on Bayes theorem, there are still general situations when the analytical characterization of the posterior distribution is too difficult to complete. Markov Chain Monte Carlo (MCMC) methods are introduced to the stochastic sampling procedure when $p(\beta|Y)$ is not exactly known, which is a quite common situation if the likelihood function $L(Y|\beta)$ is too complicated. The general idea of MCMC is to construct a Markov chain whose equilibrium distribution is just the focused posterior distribution, and the parameters of posterior distribution would be obtained through repeated sampling for simulation.

When the MCMC methods are selected for the analysis of posterior distribution, one most important problem appears that requires a systematic way of iteratively building a Markov chain that converges to the posterior distribution. Here samples of the posterior distribution are obtained using the Metropolis-Hastings approach of Strawderman and Gamerman (2000), which is the most prevalent method for MCMC sampling. For the implementation of the Metropolis-Hastings method in this dissertation, consider a general transition density $q(\beta^{(curr)}, \beta^{(cand)})$, from which a candidate $\beta^{(cand)}$ is drawn given the current guess. Additionally, during this process, an acceptance rate for this draw is defined as:

$$\alpha(\beta^{(curr)}, \beta^{(cand)}) = \min \left\{ 1, \frac{p(Y|\beta^{(cand)})p(\beta^{(cand)})q(\beta^{(curr)}, \beta^{(cand)})}{p(Y|\beta^{(curr)})p(\beta^{(curr)})q(\beta^{(cand)}, \beta^{(curr)})} \right\} \quad (7.14)$$

Starting from a proposed or arbitrary value $\beta^{(0)}$ for the algorithm, $\beta^{(cand)}$ is drawn from $q(\beta^{(t-1)}, \beta)$ at the t^{th} step and would be accepted as $\beta^{(t)}$ with probability $\alpha(\beta^{(t-1)}, \beta^{(cand)})$, whereas the old one is preserved $\beta^{(t)} = \beta^{(t-1)}$ with probability $1 - \alpha(\beta^{(t-1)}, \beta^{(cand)})$. Given the equilibrium distribution of this Markov chain would be the posterior distribution, $\beta^{(t)}$ will be sampled from the posterior density when t is large enough. Typically, to make the Markov chain converge quickly enough, proper parameters of the transition density may be selected, which is known as tuning. After draws from the goal posterior distribution have been simulated through MCMC methods, the results point and interval estimates for the Multinomial Logit Bayesian model can be constructed and further evaluated.

7.5 Model Results

7.5.1 Model Crash Reporting

The first model characterizes police reporting of crashes. The models were estimated using the methodology described in Section 7.4. Frequentist and Bayesian estimation were conducted to account for possible quasi-complete separation of the data. The summary statistics of the significant covariates are shown in Table 7.6. Crashes not being reported were considered as the reference for modeling. The results include the estimates of the parameters using the frequentist approach in Table 7.7 and the Bayesian approach in Table 7.8.

Table 7.6. Descriptive statistics crash reporting model

Crash Severity	Variable	N	Mean	Std Dev	Min	Max
I – No Reported Crash	Speed Before Collision GPS (m/s)	40	2.050	2.264	0.000	10.306
	Change Kinetic Energy (Kg m ² /s ²)	37	1489.930	1818.040	7.387	7590.230
	Acceleration X Axis at Collision	39	-0.964	0.329	-2.333	-0.378
	Instrumented Passenger/Front SUV	40	0.400	0.496	0.000	1.000
II –Reported Crash	Speed Before Collision GPS (m/s)	78	6.578	6.030	0.000	27.900
	Change Kinetic Energy (Kg m ² /s ²)	73	15034.170	53814.620	0.000	449849.270
	Acceleration X Axis at Collision	78	-1.963	0.733	-4.836	-0.370
	Instrumented Passenger/Front SUV	79	0.291	0.457	0.000	1.000

Based on the results shown in Table 7.7 and Table 7.8, the parameter estimates obtained using the frequentist and Bayesian approaches are similar. The issue of quasi-complete separation was not shown in this model. Among the most significant parameters influencing reporting, the author found a change in kinetic energy, the speed before collision, longitudinal acceleration, and vehicle typology as the most significant parameters.

Table 7.7. Frequentist model crash reporting. Reference No-reported crash

Variable	Parameter Estimate	Std. Error	z-Value	Pr(> z)
<i>Reported Crash</i>				
Intercept	-5.700	1.351	17.792	<.0001
Speed before Collision (m/s)	0.146	0.135	1.174	0.279
Change Kinetic Energy (Kg m2/s2)	3.09E-04	1.37E-04	5.120	0.024
Longitudinal Acceleration at Collision (m/s2)	-4.121	1.039	15.741	<.0001
Instrumented Passenger/Front SUV	-1.283	0.765	2.810	0.094
AIC Covariates (Intercept Only)	71.141 (138.671)			
Log-likelihood (Intercept Only)	-30.571 (-68.335)			

Table 7.8. Bayesian model crash reporting. Reference No-reported crash

Variable	Parameter Estimate	Std. Error	95% PD Int*	
<i>Reported Crash</i>				
Intercept	-6.014	1.367	-8.754	-3.557
Speed before Collision (m/s)	0.102	0.146	-0.181	0.389
Change Kinetic Energy (Kg m2/s2)	3.67E-04	1.57E-04	5.50E-05	6.44E-04
Longitudinal Acceleration at Collision (m/s2)	-4.394	1.050	-6.556	-2.510
Instrumented Passenger/Front SUV	-1.669	0.850	-3.436	-0.173

*95 % Confidence Interval Posterior Distribution

The signs of these estimates follow the results of past research studies. Higher kinetic energy and speed before collision as well as longitudinal acceleration at impact increase the probability of the crash being reported. On the other hand, when the instrumented vehicle being classified as a passenger car rear-ends an SUV, the probability of the crash being reported was reduced. This is an interesting finding since the party not at fault (front vehicle) exerts a strong influence on crash reporting. The damage of larger vehicles not at fault is less evident and their drivers are more likely to accept the event without reporting it. The probability of a crash being reported can be estimated based on the expression:

$$P(\text{Reported}) = [\exp(-6.014 + 0.102 \text{ Speed}_{\text{Collision}} + 0.00036 \Delta \text{ Energy} - 4.394 \text{ Long}_{\text{Acc}} - 1.669 \text{ Pass}_{\text{Suv}}) + 1]^{-1} \quad (7.15)$$

7.5.2 Model Crash Severity

The analysis of severity was expanded by including a more disaggregate classification of severity. This model was considered a preliminary evaluation in a methodology including more severity levels. In general, it is desirable for estimation of severity models to use benchmark severity scales such as KABCO or AIS. The summary statistics of the variables included in the models separated by severity level are shown in Table 7.9.

A similar estimation process as in the crash reporting model was reproduced for severity. The first stage included the estimation of a multinomial logistic regression using the frequentist method. When structuring the model, quasi-complete separation of the model was observed when including the variable Intersection Indicator. Despite the model showing a good fit, significantly higher parameter estimates and standard errors were observed for the variables Longitudinal Acceleration and Intersection Indicator as shown in Table 7.10 for severe crashes. Considering the issue of quasi-complete separation, Bayesian estimation was conducted and is shown in Table 7.11. The results supported the lower parameter estimates from these variables. The diagnostics of the model are shown in Appendix B. An example of the parameter estimate for Intersection Indicator is shown in Figure 7.1. Trace plots, correlation plots, and posterior distribution are included for diagnostics. The trace plots show that the Marko chain stabilized and appears constant over the iterations. In addition, it shows a good mixing of the chain reflecting a good exploration of the tails and mode areas. The autocorrelation plot indicates a quick autocorrelation reduction among the posterior samples supporting an appropriate estimation of the parameters. The results also support the variable Intersection Indicator as not significant when applying Bayesian estimation.

A second set of models using frequentist and Bayesian methods were estimated when removing Intersection Indicator. The models are shown in Table 7.12 and Table 7.13. In this case, the parameter estimates from these two approaches were similar. The goodness of fit measures showed a significant reduction of AIC. The significant parameter estimates were similar to those found in the previous section.

Table 7.9. Descriptive statistics severity model

Crash Severity	Variable	N	Mean	Std Dev	Min	Max
I – Severe Crash	Speed Before Collision GPS (m/s)	31	10.100	7.570	0.000	27.900
	Change Kinetic Energy (Kg m ² /s ²)	28	32196.66	84887.35	391.96	449849.27
	Acceleration X Axis at Collision	31	-2.609	0.653	-4.836	-1.642
	Measurement Y Axis at Collision (m/s ²)	31	26.089	54.669	0.000	265.359
	Indicator Contiguous Lanes >3	32	0.281	0.457	0.000	1.000
	Instrumented Passenger/Front SUV	32	0.375	0.492	0.000	1.000
	Intersection Indicator	32	0.688	0.471	0.000	1.000
II –Police Reportable Crash	Speed Before Collision GPS (m/s)	47	4.256	3.102	0.167	14.500
	Change Kinetic Energy (Kg m ² /s ²)	45	4355.28	3616.51	0.000	16190.45
	Acceleration X Axis at Collision	47	-1.537	0.393	-2.742	-0.370
	Gyroscope Measurement Y Axis at Collision (m/s ²)	47	13.029	9.865	0.000	53.332
	Indicator Contiguous Lanes >3	47	0.511	0.505	0.000	1.000
	Instrumented Passenger/Front SUV	47	0.234	0.428	0.000	1.000
	Intersection Indicator	47	0.809	0.398	0.000	1.000
III – Minor Collision	Speed Before Collision GPS (m/s)	40	2.050	2.264	0.000	10.306
	Change Kinetic Energy (Kg m ² /s ²)	37	1489.93	1818.04	7.38	7590.23
	Acceleration X Axis at Collision	39	-0.964	0.329	-2.333	-0.378
	Gyroscope Measurement Y Axis at Collision (m/s ²)	39	19.820	69.207	0.000	436.412
	Indicator Contiguous Lanes >3	40	0.575	0.501	0.000	1.000
	Instrumented Passenger/Front SUV	40	0.400	0.496	0.000	1.000
	Intersection Indicator	40	0.925	0.267	0.000	1.000

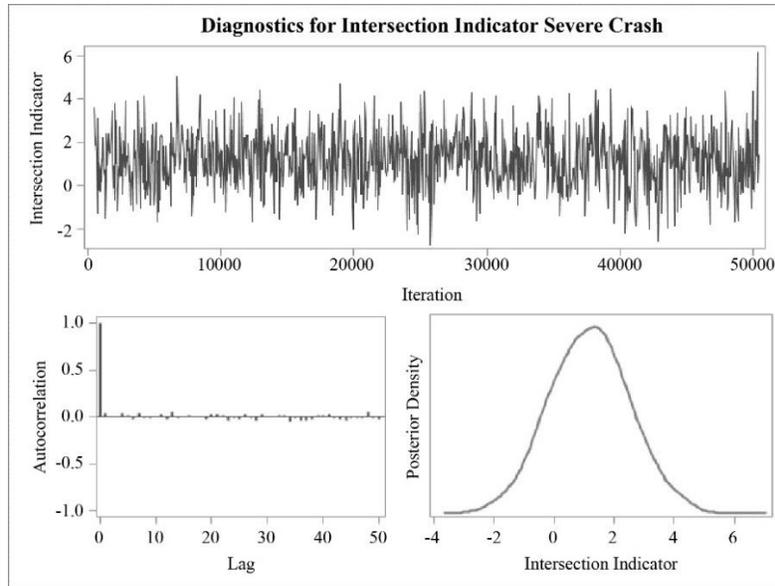


Figure 7.1 Diagnostics plots Bayesian estimation parameter intersection indicator

The parameter estimates for the severe crash category showed the covariates' speed before collision, the change in kinetic energy, and the longitudinal acceleration as the factors that increased the severity of the crash. On the other hand, the Yaw rate in the Y-axis, an indicator of contiguous lanes, as well as crashes involving SUVs in front reduced severity. In general, dissipating the energy in other directions reduced the severity of the crash as indicated by the gyroscope and more lanes variables. Finally, the same pattern was found in the case of the typology of vehicles from the analysis.

7.6 Summary

This chapter described the models applied for the estimation of severity of hypothetical crash given a conflict. Parametric characterization of the probability of a crash being reported or defined into three specific severities was conducted. Different environmental, vehicle, driver, and crash covariates were included in the models to better identify the factors that increase the likelihood of a crash being reported or being more severe. Following the results from the literature, the pre-crash specific dynamics were the most significant variables in the analysis. Acceleration at impact was the most important covariate that defined crash reporting and multiclass severity. Kinetic energy, acceleration, and type of vehicle were the predictors that were found most important in the estimation task.

Table 7.10. Multinomial logit model Frequentist method including intersection indicator

Variable	Parameter Estimate	Std. Error	Wald Chi-Square	p-value
<i>I. Severe Crash</i>				
Intercept	-146.600	91.597	2.560	0.110
Speed before Collision (m/s)	1.708	0.959	3.174	0.075
Change Kinetic Energy (Kg m2/s2)	9.92E-04	5.98E-04	2.748	0.097
Longitudinal Acceleration at Collision (m/s2)	-63.012	37.636	2.803	0.094
Gyroscope Measurement Y Axis at Collision (m/s2)	-1.913	1.128	2.873	0.090
Indicator Contiguous Lanes >3	-15.328	9.518	2.594	0.107
Instrumented Passenger/Front SUV	-4.180	3.151	1.760	0.185
Intersection Indicator	35.702	23.640	2.281	0.131
<i>II. Police Reportable Crash</i>				
Intercept	-4.213	1.603	6.905	0.009
Speed before Collision (m/s)	0.163	0.144	1.278	0.258
Change Kinetic Energy (Kg m2/s2)	3.26E-04	1.42E-04	5.308	0.021
Longitudinal Acceleration at Collision (m/s2)	-4.057	1.121	13.100	0.000
Gyroscope Measurement Y Axis at Collision (m/s2)	-0.003	0.007	0.204	0.652
Indicator Contiguous Lanes >3	-0.407	0.726	0.315	0.575
Instrumented Passenger/Front SUV	-1.168	0.798	2.146	0.143
Intersection Indicator	-1.412	1.150	1.506	0.220
<i>III. Minor Collision (Reference)</i>				
AIC at convergence = 98.722		AIC (constants only) = 233.951		
SC at convergence = 141.487		SC (constants only)= 239.297		
Log-likelihood at Convergence= -33.361		Log-likelihood at Convergence= -114.976		

Table 7.11. Multinomial logit model Bayesian method including intersection indicator

Variable	Parameter Estimate	Std. Error	95% HPD Interval	
<i>I. Severe Crash</i>				
Intercept	-17.741	3.342	-24.113	-11.279
Speed before Collision (m/s)	0.471	0.062	0.347	0.586
Change Kinetic Energy (Kg m ² /s ²)	4.28E-04	1.60E-05	3.96E-04	4.57E-04
Longitudinal Acceleration at Collision (m/s ²)	-10.406	1.612	-13.677	-7.654
Gyroscope Measurement Y Axis at Collision (m/s ²)	-0.323	0.095	-0.512	-0.152
Indicator Contiguous Lanes >3	-1.846	0.339	-2.558	-1.228
Instrumented Passenger/Front SUV	-2.382	0.428	-3.190	-1.504
Intersection Indicator	1.194	1.317	-1.577	3.622
<i>II. Police Reportable Crash</i>				
Intercept	-3.323	1.283	-5.580	-0.441
Speed before Collision (m/s)	0.286	0.050	0.193	0.387
Change Kinetic Energy (Kg m ² /s ²)	2.88E-04	4.30E-05	2.10E-04	3.74E-04
Longitudinal Acceleration at Collision (m/s ²)	-3.721	0.570	-4.854	-2.612
Gyroscope Measurement Y Axis at Collision (m/s ²)	-0.011	0.004	-0.019	-0.003
Indicator Contiguous Lanes >3	-0.734	0.335	-1.355	-0.077
Instrumented Passenger/Front SUV	-2.002	0.226	-2.445	-1.569
Intersection Indicator	-1.885	0.706	-3.308	-0.503
<i>III. Minor Collision (Reference)</i>				

Table 7.12. Multinomial logit model Frequentist method without intersection indicator

Variable	Parameter Estimate	Std. Error	Wald Chi-Square	p-value
<i>I. Severe Crash</i>				
Intercept	-20.620	5.313	15.062	0.000
Speed before Collision (m/s)	0.266	0.183	2.109	0.146
Change Kinetic Energy (Kg m2/s2)	5.17E-04	1.98E-04	6.81E+00	9.10E-03
Longitudinal Acceleration at Collision (m/s2)	-12.756	3.017	17.875	<.0001
Gyroscope Measurement Y Axis at Collision (m/s2)	-0.341	0.193	3.113	0.078
Indicator Contiguous Lanes >3	-2.633	1.484	3.150	0.076
Instrumented Passenger/Front SUV	-0.993	1.456	0.465	0.496
<i>II. Police Reportable Crash</i>				
Intercept	-5.346	1.380	15.010	0.000
Speed before Collision (m/s)	0.137	0.139	0.972	0.324
Change Kinetic Energy (Kg m2/s2)	3.20E-04	1.43E-04	4.98E+00	2.56E-02
Longitudinal Acceleration at Collision (m/s2)	-4.192	1.113	14.181	0.000
Gyroscope Measurement Y Axis at Collision (m/s2)	-0.003	0.007	0.248	0.618
Indicator Contiguous Lanes >3	-0.648	0.693	0.875	0.350
Instrumented Passenger/Front SUV	-1.325	0.779	2.891	0.089
<i>III. Minor Collision (Reference)</i>				
AIC at convergence = 111.742		AIC (constants only) = 233.951		
SC at convergence = 149.162		SC (constants only)= 239.297		
Log-likelihood at Convergence= -41.871		Log-likelihood at Convergence= -114.976		

Table 7.13. Multinomial logit model Bayesian method without intersection indicator

Variable	Parameter Estimate	Std. Error	95% HPD Interval	
<i>I. Severe Crash</i>				
Intercept	-19.634	3.771	-26.766	-12.146
Speed before Collision (m/s)	0.266	0.183	-0.093	0.637
Change Kinetic Energy (Kg m ² /s ²)	6.02E-04	1.82E-04	2.87E-04	9.90E-04
Longitudinal Acceleration at Collision (m/s ²)	-12.253	2.303	-16.978	-8.071
Gyroscope Measurement Y Axis at Collision (m/s ²)	-0.353	0.172	-0.697	-0.042
Indicator Contiguous Lanes >3	-2.921	1.353	-5.352	0.027
Instrumented Passenger/Front SUV	-1.349	1.455	-4.059	1.508
<i>II. Police Reportable Crash</i>				
Intercept	-5.176	1.329	-7.644	-2.639
Speed before Collision (m/s)	0.126	0.142	-0.149	0.397
Change Kinetic Energy (Kg m ² /s ²)	3.57E-04	1.55E-04	6.10E-05	6.72E-04
Longitudinal Acceleration at Collision (m/s ²)	-4.251	1.095	-6.335	-2.165
Gyroscope Measurement Y Axis at Collision (m/s ²)	-0.015	0.014	-0.044	0.008
Indicator Contiguous Lanes >3	-0.842	0.679	-2.346	0.324
Instrumented Passenger/Front SUV	-1.444	0.742	-2.869	-0.015
<i>III. Minor Collision (Reference)</i>				

When estimating the models using crash dynamic characteristics, the procedure had issues with quasi-complete separation when applying logistic regression. This issue was addressed by proposing a Bayesian methodology with normal priors to avoid overestimation of the parameters and associated standard errors. This issue was found in the multiclass definition of severity, but there were no issues when evaluating the likelihood of a crash being reported.

The model for assessing crash reporting is transferable to other studies. However, the definition of more disaggregate categories was conducted using a severity scale pre-specified by the SHRP2 program. Estimation of this model is more desirable using benchmark scale severities, such as KABCO or AIS, to allow comparison with other research initiatives.

In summary, this chapter provided an interesting framework for the estimation of models evaluating reporting and severity. Naturalistic driving studies provide disaggregate information that allows researchers to evaluate these two dimensions of safety in real conditions. The introduction of instrumented vehicles in the future will better guide this estimation and can provide more data from these events.

8. CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

8.1 Conclusions

This dissertation discussed an important issue for practical implementation of traffic conflicts technique, namely, identification of conflicts using emerging area-wide and in-vehicle instrumentation technologies. Recently developed detection and tracking systems are providing researchers with new data sources for accurate extraction of trajectories and associated proximity measures between road users. These systems provide the conditions and means to transition traffic conflict technique from the research phase to the implementation phase by transportation engineers.

Surrogate measures of safety have been explored over the years with the ever-growing need to provide a proactive rather than reactive approach to safety issues. The validity of crashes as measures of safety is irrefutable; and they are useful for pinpointing road conditions that increase hazards for the purpose of mitigating them. However, the limitations related to the dubious quality of reporting, lack of reliable insight into the events leading to a crash, and long data collection times have caused researchers to reevaluate their validity, especially in the era of autonomous vehicles when safety will be changing rapidly.

A traffic conflict is a prevalent surrogate measure of safety. Traffic conflicts meet the conditions of crash surrogacy in the following ways according to Tarko (2018): (1) they are more frequent than crashes; (2) there is a method of estimating the corresponding crash frequency and severity; and more importantly, (3) there is an etiological connection between conflicts and crashes introduced by a failure to which a road user responds with delay.

The predictive validity of traffic conflicts is still an active area of research. Davis et al. (2011) explored a new perspective on dangerous traffic interactions from the counterfactual point of view proposed by Pearl (2000). Tarko (2018) elaborated this approach into a practical method by considering the behavior of the road users in a risky situation bringing two alternative outcomes: a crash or a conflict. Tarko (2019) delivered a comprehensive overview of the method with examples that demonstrated the method's validity.

The method introduced in Tarko (2018) expanded the benchmark definition of the traffic conflict proposed in Amundsen and Hyden (1977), who theorized that space and time proximity was the only condition associated with the risk of crash. However, the counterfactual approach in

Tarko (2018) defined a failure as a necessary condition that simplifies analysis and strengthens the connection with a crash. In this method, failure is indicated by a separation too short to be acceptable. However, as experimental validation has shown, the heterogeneity of the drivers, risky driving habits, and sensor noise undermine the proximity measures as sufficient.

The experimental validation in this dissertation strongly supported failure as a necessary condition for identifying traffic conflicts that meet the theoretical stipulations in Tarko (2018). Lack of failure from the analyzed encounters was confirmed to cause overestimation of the expected number of crashes, the non-Lomax-distributed response delays, and the lack of convergence in the expected number of crashes at lower separation thresholds. This dissertation found that abnormally strong avoidance behavior is a necessary indicator of failure. Strong avoidance behavior was observed in the presence of a significant change in acceleration or jerking by drivers or bicyclists to correct for dangerous conditions. In the case of rear-end traffic conflicts, strong avoidance was only observed in the longitudinal component of movement while in turning conflicts a lateral response (change in direction) was adopted to avoid a crash.

Traffic interactions at low speed were found to be especially challenging in the case of vehicle-vehicle encounters because they make hypothetical collisions non-consequential and the collision nearness does not produce the perception of considerable hazard among the involved drivers. Eliminating the danger even at very short separations between vehicles undermines the notion of failure. Hence, in vehicle-vehicle encounters, it was applied a speed threshold above which the presence of risk and the potential for related failure is applicable. On the other hand, crashes between vehicles and vulnerable road users (VRUs) are always undesirable with a high perception of risk even at low speeds.

By applying convenient instantaneous TTC values and associated assumptions, the results showed good estimates in the case of rear-end traffic conflicts. Since the collision points in these events are located in the same direction of the longitudinal axis of the instrumented vehicle, the assumption of constant speed and direction in the counterfactual scenario of no evasion provided a good estimation of the hypothetical collision point.

The same assumptions did not show good estimates for the case of turning encounters. Evasive maneuvers for vehicle-bicycle interactions included lateral responses producing significant changes in the location of the hypothetical collision point among consecutive frames under the iTTC assumptions. Hence, a redefined procedure for the estimation of a hypothetical

collision point under the lack of response should be considered to estimate proximity measures. Different models can be applied to predict the future position in the counterfactual scenario of no evasion. An example could be a Kalman filter-based constant acceleration model. An additional method was proposed in this dissertation that corresponds to extrapolating the trajectory by learning from evasion-free trajectories collected from the same location and type of movement. This approach is especially appealing in area-wide systems that collect multiple trajectories from drivers in the same intersection or road segment.

Finally, an important component to provide comprehensive safety analysis relying on traffic conflicts refers to the ability of the method to estimate crash frequency and severity. Using measurements from crash events reported in the SHRP2 initiative, two models evaluating the probability of a crash being reported were estimated in Chapter 7. The models showed the pre-crash specific dynamics as the most significant variables in the analysis; and acceleration at impact was the most important covariate. Kinetic energy, type of vehicle, and number of lanes were among the predictors found to be important in the estimation task.

Due to the low sample, the severity model showed quasi-complete separation issues, which caused biased estimates and high standard errors when applying the frequentist approach in the estimation of the logistic model. By proposing a Bayesian methodology with normal priors, the results showed convergence in the parameter estimates and low standard errors. Due to the internal severity definition by the SHRP2 program, only the crash reporting model is transferable to other studies. Further investigation of severity analysis using commonly used scales in the literature should be conducted.

8.2 Contributions

The primary contribution of this dissertation is a generalized framework for identification of traffic conflicts in line with the failure-based definition proposed in Tarko (2018). Previous studies focused on methods for estimating the expected number of crashes without further discussing the definitions of conflicts and the methodologies applied for their identification. The analysis in this dissertation successfully proved the methods for the extraction of rear-end and turning conflicts using in-vehicle instrumentation and area-wide detection systems.

As discussed in Tarko (2018), failure should be an important component of conflicts that end in crashes as an outcome if there is not a timely response. The original method hypothesized

separation as the only necessary condition that indicated failure, but the empirical results showed the necessity to identify strong avoidance behavior as an indication of failure. Strong avoidance behavior was identified with the presence of a significant change in the lateral or longitudinal components of the acceleration profile. An additional contribution included an analysis of the method for extracting conflicts from vehicle-bicycle encounters. Analysis of conflicts for vulnerable road users is more challenging compared to motorized users.

Tarko (2019) discussed the limitation for applying instantaneous TTC in the general case where speeds may be changing and paths may be curved. These conditions require predicting the trajectories in the counterfactual scenario of no evasion (Laureshyn et al., 2010). Following the postulates in Saunier et al. (2007), this dissertation introduced a methodology that allows extrapolating the trajectories using the most similar evasion-free trajectory adopted for the same type of movement and location. Additional characteristics of similarity, such as the type of vehicle, can be adopted in the future. This extrapolation allows estimation of TTC without approximations.

Empirical evaluation of the framework was conducted using two different data sources from area-wide tracking systems and in-vehicle instrumentation. The results from the SHRP2 study were particularly encouraging. The appropriate identification of traffic conflicts resulted in the estimation of expected number of crashes meeting the theoretical postulates from Tarko (2018), which included stabilization of the estimates at lower separation thresholds, Lomax-distributed response delays, and reasonable estimates of the expected number of crashes.

Finally, this dissertation introduced a severity model that could be potentially used for estimation of the hypothetical severity of traffic conflicts. This dissertation's analysis included a model evaluating crash reporting and severity. Based on the author's knowledge of the literature, this was the first attempt to provide a model estimating the probability of a crash being reported to the police based on pre-crash, vehicle, and environmental conditions. Moreover, the analysis of severity constitutes a unique opportunity for researchers to structure a safety management system relying on traffic conflicts. It also allows better identification of locations that may expect a higher number and more consequential crashes. Crash reporting is also an important component towards comparing the estimation of the expected number of crashes with the reported crashes.

8.3 Limitations and Future Directions

The limitations of this dissertation and future research directions are discussed in this section. The first limitation was found with the approximation of instantaneous TTC in rear-end conflicts analysis from SHRP2. Prediction of trajectory under the counterfactual scenario of no evasion was characterized by extracting additional braking scenarios from the same driver instead. This method allows estimation of TTC without any approximations.

The general framework proved to be effective while identifying and distinguishing events without failure in the InDeV study. A more complete evaluation for identification of conflicts using vehicle-bicycle encounters was restricted by the availability of data. Longer data collection times are required for estimation of the expected number of crashes meeting the theoretical conditions discussed in Tarko (2018). The framework was codified and used for evaluation of the LIDAR-based system described in Section 3.2.1. An extensive evaluation of the method will be implemented in future work.

Chapter 4 presented an a priori division of drivers into different categories. In order to meet the objective of providing a complete safety management system based on conflicts, there should be a method for identifying the sources of heterogeneity and estimating their effects on crash expectancy (safety effects estimation). Heterogeneity, in terms of the driver, vehicle, and road infrastructure, was considered to better understand how conflicts resemble crashes regarding the factors affecting them. Future research should include methods to address the parametric characterization of heterogeneity in the estimation method.

In line with the crash reporting and severity models, future work will address modeling and obtaining the values of the variables included in the models in hypothetical collisions. An additional limitation to be addressed is to include an analysis of a benchmark severity scale transferable and comparable with other studies.

Finally, extension of the method to the calibration and training of autonomous vehicles is possible. These additional data sources provide opportunities for better characterization of conflict events. Failure can be easily identified based on reported disengagement from this technology, removing ambiguity to identify emergency conditions. In additional violations to comfort based thresholds will provide an indication of undesirable situation to be corrected creating additional opportunities to apply the method for estimation of expected number of crashes. Rapid estimation of safety is needed specially in the transition period with human driven and autonomous vehicles

(Chen et al., 2020; Li et al., 2020). This area could be explored in conjunction with the on-going Center for Road Safety project titled “Guidelines for Evaluating Safety Using Traffic Conflicts: Proactive Crash Estimation on Roadways with Conventional and Autonomous Vehicle Scenario.”

REFERENCES

- Aijazi, A. K., Checchin, P., Malaterre, L., & Trassoudaine, L. (2016). Automatic detection of vehicles at road intersections using a compact 3D Velodyne sensor mounted on traffic signals. In *Intelligent Vehicles Symposium* (pp. 662–667).
- Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. <https://doi.org/10.1093/biomet/71.1.1>
- Allison, P. D. (2010). *Survival Analysis Using SAS: A Practical Guide*. *Survival*.
- Aminikhangahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-016-0987-z>
- Amundsen, F., & Hyden, C. (1977). Proceedings of first workshop on traffic conflicts. In *Norway and LTH Lund*. Lund, Sweden.
- Aoude, G. S., Desaraju, V. R., Stephens, L. H., & How, J. P. (2011). Behavior classification algorithms at intersections and validation using naturalistic data. In *IEEE Intelligent Vehicles Symposium, Proceedings* (pp. 601–606). <https://doi.org/10.1109/IVS.2011.5940569>
- Archer, J. (2005). Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections. *Royal Institute of Technology, Department of Infrastructure, Division for Transport and Logistics, Centre for Transport Research*. [https://doi.org/ISBN 91-7323-119-3](https://doi.org/ISBN%2091-7323-119-3)
- Atev, S., Miller, G., & Papanikolopoulos, N. P. (2010). Clustering of vehicle trajectories. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2010.2048101>
- Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. *Eurasip Journal on Image and Video Processing*, 2008. <https://doi.org/10.1155/2008/246309>
- Buch, N., Velastin, S. A., & Orwell, J. (2011). A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2011.2119372>
- Buzan, D., Sclaroff, S., & Kollios, G. (2004). Extraction and clustering of motion trajectories in video. In *Proceedings - International Conference on Pattern Recognition*. <https://doi.org/10.1109/icpr.2004.1334287>

- Buzeman, D. G., Viano, D. C., & Lövsund, P. (1998a). Car occupant safety in frontal crashes: A parameter study of vehicle mass, impact speed, and inherent vehicle protection. *Accident Analysis and Prevention*. [https://doi.org/10.1016/S0001-4575\(98\)00020-7](https://doi.org/10.1016/S0001-4575(98)00020-7)
- Buzeman, D. G., Viano, D. C., & Lövsund, P. (1998b). Injury probability and risk in frontal crashes: Effects of sorting techniques on priorities for offset testing. *Accident Analysis and Prevention*. [https://doi.org/10.1016/S0001-4575\(98\)00018-9](https://doi.org/10.1016/S0001-4575(98)00018-9)
- Campbell, K. L. (2012). The SHRP2 naturalistic driving study: Addressing driver performance and behavior in traffic safety. *Trb News*, 282.
- Campbell, K. L., Joksch, H. C., & Green, P. (1996). *A bridging analysis for estimating the benefits of active safety technologies*.
- Chen, S., Leng, Y., & Labi, S. (2020). A deep learning algorithm for simulating autonomous driving considering prior knowledge and temporal information. *Computer-Aided Civil and Infrastructure Engineering*. <https://doi.org/10.1111/mice.12495>
- Chin, H. C., & Quek, S. T. (1997). Measurement of traffic conflicts. *Safety Science*. [https://doi.org/10.1016/S0925-7535\(97\)00041-6](https://doi.org/10.1016/S0925-7535(97)00041-6)
- Christoforou, Z., Cohen, S., & Karlaftis, M. G. (2010). Vehicle occupant injury severity on highways: An empirical investigation. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2010.03.019>
- Coifman, B., Beymer, D., McLauchlan, P., & Malik, J. (1998). A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C: Emerging Technologies*, 6(4), 271–288. [https://doi.org/10.1016/S0968-090X\(98\)00019-9](https://doi.org/10.1016/S0968-090X(98)00019-9)
- Cook, D. J., & Krishnan, N. C. (2015). *Activity learning: Discovering, recognizing, and predicting human behavior from sensor data*. *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*. <https://doi.org/10.1002/9781119010258>
- Cooper, P. (1977). State-of-the-art. Report on traffic conflicts research in Canada. In *Proc. 1st Workshop on Traffic Conflicts (1977)*.
- El-Basyouny, K., & Sayed, T. (2013). Safety performance functions using traffic conflicts. *Safety Science*. <https://doi.org/10.1016/j.ssci.2012.04.015>
- Elvik, R. (2010). Exploratory study of mechanisms by which exposure influences accident occurrence. *Transportation Research Record*. <https://doi.org/10.3141/2148-09>

- Farmer, C. M., Braver, E. R., & Mitter, E. L. (1997). Two-vehicle side impact crashes: The relationship of vehicle and crash characteristics to injury severity. *Accident Analysis and Prevention*. [https://doi.org/10.1016/S0001-4575\(97\)00006-7](https://doi.org/10.1016/S0001-4575(97)00006-7)
- Forbes, T. W. (1957). Analysis of “near accident” reports. *Highway Research Board Bulletin*, (152), 23–37.
- Gabauer, D. J., & Gabler, H. C. (2008). Comparison of roadside crash injury metrics using event data recorders. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2007.08.011>
- Gallego, N., Mocholi, A., Menendez, M., & Barrales, R. (2009). Traffic Monitoring: Improving Road Safety Using a Laser Scanner Sensor. *2009 Electronics, Robotics and Automotive Mechanics Conference (CERMA)*, 281–286. <https://doi.org/10.1109/CERMA.2009.11>
- Gettman, D., Pu, L., Sayed, T., & Shelby, S. (2008). Surrogate Safety Assessment Model and Validation: Final Report. *Publication No. FHWA-HRT-08-051*.
- Ghosh, J., Li, Y., & Mitra, R. (2018). On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*. <https://doi.org/10.1214/17-BA1051>
- Glennon, J. C., Glauz, W. D., Sharp, M. C., & Thorson, B. A. (1977). Critique of the traffic-conflict technique. *Transportation Research Record*.
- Gordon, T., Bareket, Z., Kostyniuk, L., Barnes, M., Hagan, M., Kim, Z., ... Skabardonis, A. (2012). Site-Based Video System Design and Development. *SHRP2 Report*. Retrieved from <http://www.trb.org/Main/Blurbs/166323.aspx%5Cnhttps://trid.trb.org/view/1226172>
- Guettinger, V. A. (1982). From Accidents to Conflicts: Alternative Safety Measurement. In *Short-Term and Area-Wide Evaluation of Safety Measures*.
- Guha, P., Mukerjee, A., & Venkatesh, K. S. (2006). Appearance based multiple agent tracking under complex occlusions. In *Pricai 2006: Trends in Artificial Intelligence, Proceedings* (Vol. 4099, pp. 593–602).
- Guo, F., Klauer, S. G., Hankey, J. M., & Dingus, T. A. (2010). Near crashes as crash surrogate for naturalistic Driving Studies. *Transportation Research Record*. <https://doi.org/10.3141/2147-09>
- Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis and Prevention*. [https://doi.org/10.1016/0001-4575\(82\)90014-8](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 and Prevention*. [https://doi.org/10.1016/0001-4575\(86\)90020-5](https://doi.org/10.1016/0001-4575(86)90020-5)

- Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*. <https://doi.org/10.1002/sim.1047>
- Hermes, C., Wohler, C., Schenk, K., & Kummert, F. (2009). Long-term vehicle motion prediction. In *IEEE Intelligent Vehicles Symposium, Proceedings*. <https://doi.org/10.1109/IVS.2009.5164354>
- Hosking, J. R., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*. <https://doi.org/10.2307/1269343>
- Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. <https://doi.org/10.1109/TSMCC.2004.829274>
- Huang, H., Abdel-Aty, M., & Darwiche, A. (2010). County-level crash risk analysis in Florida. *Transportation Research Record, 2148*, 27–37. <https://doi.org/10.3141/2148-04>
- Hydén, C. (1987). The Development of a Method for Traffic Safety Evaluation: the Swedish Traffic Conflict Technique. *Bulletin Lund University of Technology*. <https://doi.org/10.1002/2016GC006399>
- Ismail, K., Sayed, T., & Saunier, N. (2011). Methodologies for aggregating indicators of traffic conflict. *Transportation Research Record*. <https://doi.org/10.3141/2237-02>
- Jodoin, J. P., Bilodeau, G. A., & Saunier, N. (2016). Tracking All Road Users at Multimodal Urban Traffic Intersections. *IEEE Transactions on Intelligent Transportation Systems, 17*(11), 3241–3251. <https://doi.org/10.1109/TITS.2016.2545245>
- Johnsson, C., Laureshyn, A., Varhelyi, A., D’Agostino, C., & Farah, H. (2018). *Surrogate measures of safety*.
- Jung, S., Qin, X., & Noyce, D. A. (2010). Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2009.07.020>
- Kawahara, Y., & Sugiyama, M. (2012). Sequential change-point detection based on direct density-ratio estimation. *Statistical Analysis and Data Mining*. <https://doi.org/10.1002/sam.10124>
- Klauer, S. G., Klauer, S. G., Dingus, T. a., Dingus, T. a., Neale, V. L., Neale, V. L., ... Ramsey, D. J. (2006). The Impact of Driver Inattention On Near Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *Analysis*. <https://doi.org/DOT HS 810 594>

- Kononen, D. W., Flannagan, A. C., & Wang, S. C. (2011). Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2010.07.018>
- Lamprey, G., Singh, L., Labi, S., & Sinha, K. C. (2010). Systematic framework for incorporating safety in network-level transportation planning and programming. *Journal of Transportation Engineering*. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000070](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000070)
- Lamprey, Godfrey, Labi, S., & Sinha, K. C. (2005). Investigating the sensitivity of optimal network safety needs to key safety management inputs. In *Transportation Research Record*. <https://doi.org/10.3141/1922-08>
- Lee, H., & Coifman, B. (2015). Using LIDAR to Validate the Performance of Vehicle Classification Stations. *Journal of Intelligent Transportation Systems*, 19(4), 355–369. <https://doi.org/10.1080/15472450.2014.941750>
- Li, X., Lord, D., Zhang, Y., & Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2008.04.010>
- Li, Y., Chen, S., Dong, J., Steinfeld, A., & Labi, S. (2020). Leveraging Vehicle Connectivity and Autonomy to Stabilize Flow in Mixed Traffic Conditions: Accounting for Human-driven Vehicle Driver Behavioral Heterogeneity and Perception-reaction Time Delay, (arXiv preprint arXiv:2008.04351).
- Liu, L., & Dissanayake, S. (2009). Factors affecting crash severity on gravel roads. *Journal of Transportation Safety and Security*. <https://doi.org/10.1080/19439960903381669>
- Lizarazo, C. G. (2016). *Classification of road users detected and tracked with LiDAR at intersections*. ProQuest Dissertations and Theses.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291–305. <https://doi.org/10.1016/j.tra.2010.02.001>
- Ma, Z., Zhang, H., Chien, S. I., Wang, J., & Dong, C. (2017). Predicting expressway crash frequency using a random effect negative binomial model: A case study in China. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2016.10.012>
- McFarland, R. (1956). Human factors in highway transport safety. In *SAE Technical Papers*. <https://doi.org/10.4271/560064>

- McHenry, R. R. (1969). An analysis of the dynamics of automobiles during simultaneous cornering and ride motions. *Computer-Aided Design*. [https://doi.org/10.1016/S0010-4485\(69\)80082-5](https://doi.org/10.1016/S0010-4485(69)80082-5)
- Mendes, J. C., Bianchi, A. G., & Júnior, A. R. (2015). Vehicle Tracking and Origin-destination Counting System for Urban Environment. In VISAPP (Ed.), *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*. (pp. 600–607).
- Messelodi, S., Modena, C. M., & Zanin, M. (2005). A computer vision system for the detection and classification of vehicles at urban road intersections. *Pattern Analysis and Applications*, 8(1–2), 17–31. <https://doi.org/10.1007/s10044-004-0239-9>
- Mullakkal-Babu, F. A., Wang, M., Farah, H., Van Arem, B., & Happee, R. (2017). Comparative assessment of safety indicators for vehicle trajectories on highways. *Transportation Research Record*. <https://doi.org/10.3141/2659-14>
- Munyakwiye, B. B., Karimi, H. R., & Robbersmyr, K. G. (2016). Fuzzy logic approach to predict vehicle crash severity from acceleration data. In *iFUZZY 2015 - 2015 International Conference on Fuzzy Theory and Its Applications, Conference Digest*. <https://doi.org/10.1109/iFUZZY.2015.7391892>
- Najm, W. G., & Smith, D. L. (2004). Modeling driver response to lead vehicle decelerating. In *SAE Technical Papers*. <https://doi.org/10.4271/2004-01-0171>
- Neale, V. L., Dingus, T. A., Klauer, S. G., & Goodman, M. (2005). An overview of the 100-car naturalistic study and findings. *Traffic Safety*.
- Parker, M., & Zegeer, C. (1989). *Traffic Conflict Techniques for Safety and Operations: Engineers Guide. Report FHWA-IP-026*.
- Perkins, S., & Harris, J. (1968). Traffic conflict characteristics-accident potential at intersections. *Highway Research Record*, (225), 35–43.
- Prabhakar, Y., Subirats, P., Lecomte, C., Violette, E., & Bensrhair, A. (2013). A lidar-based method for the detection and counting of Powered Two Wheelers. In *IEEE Intelligent Vehicles Symposium, Proceedings* (pp. 1167–1172). <https://doi.org/10.1109/IVS.2013.6629624>
- Prato, C. G., Toledo, T., Lotan, T., & Taubman - Ben-Ari, O. (2010). Modeling the behavior of novice young drivers during the first year after licensure. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2009.09.011>

- Premebida, C., Ludwig, O., & Nunes, U. (2009). LIDAR and vision-based pedestrian detection system. *Journal of Field Robotics*, 26(9), 696–711. <https://doi.org/10.1002/rob.20312>
- Premebida, C., Monteiro, G., Nunes, U., & Peixoto, P. (2007). A Lidar and vision-based approach for pedestrian and vehicle detection and tracking. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* (pp. 1044–1049). <https://doi.org/10.1109/ITSC.2007.4357637>
- Quddus, M. A., Wang, C., & Ison, S. G. (2010). Road traffic congestion and crash severity: Econometric analysis using ordered response models. *Journal of Transportation Engineering*. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000044](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000044)
- Rana, T. A., Sikder, S., & Pinjari, A. R. (2010). Copula-based method for addressing endogeneity in models of severity of traffic crash injuries: Application to two-vehicle crashes. *Transportation Research Record*. <https://doi.org/10.3141/2147-10>
- Regan, M., Williamson, A., Grzebieta, R., & Tao, L. (2012). Naturalistic Driving Studies: Literature Review and Planning for the Australian Naturalistic Driving Study. *Australasian College of Road Safety National Conference*.
- Saunier, N., & Sayed, T. (2006). A feature-based tracking algorithm for vehicles in intersections. In *Third Canadian Conference on Computer and Robot Vision, CRV 2006* (Vol. 2006). <https://doi.org/10.1109/CRV.2006.3>
- Saunier, N., & Sayed, T. (2008). Probabilistic framework for automated analysis of exposure to road collisions. *Transportation Research Record*. <https://doi.org/10.3141/2083-11>
- Saunier, N., Sayed, T., & Ismail, K. (2010). Large-scale automated analysis of vehicle interactions and collisions. *Transportation Research Record*. <https://doi.org/10.3141/2147-06>
- Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2011.03.025>
- Sayed, T., & Zein, S. (1999). Traffic conflict standards for intersections. *Transportation Planning and Technology*. <https://doi.org/10.1080/03081069908717634>
- Scott-Parker, B., Watson, B., King, M. J., & Hyde, M. K. (2013). Revisiting the concept of the “problem young driver” within the context of the “young driver problem”: Who are they? *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2013.05.009>

- Smith, K., Gatica-Perez, D., Odobez, J.-M., & Ba, S. (2005). Evaluating Multi-Object Tracking. *Management*, 00(c), 36–36. <https://doi.org/10.1109/CVPR.2005.453>
- Smith, R. L. (2003). Statistics of extremes, with applications in environment, insurance, and finance. In *Extreme Values in Finance, Telecommunications, and the Environment*. <https://doi.org/10.1201/9780203483350.ch1>
- Sobhani, A., Young, W., Logan, D., & Bahrololoom, S. (2011). A kinetic energy model of two-vehicle crash injury severity. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2010.10.021>
- Song, X., & Nevatia, R. (2007). Detection and Tracking of Moving Vehicles in Crowded Scenes. *IEEE Workshop on Motion and Video Computing*, (Mcmc), 4. <https://doi.org/10.1109/WMVC.2007.13>
- Songchitruksa, P., & Tarko, A. P. (2006). The extreme value theory approach to safety estimation. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2006.02.003>
- Spinello, L., Arras, K., Triebel, R., & Siegwart, R. (2010). A Layered Approach to People Detection in 3D Range Data. *Aaai*, (2007), 1625–1630. Retrieved from <http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/download/1877/2267%5Cn/Downloads/PresnetationCD%5Cnhttp://www2.informatik.uni-freiburg.de/~spinello/AAAI2010.html>
- Strawderman, R. L., & Gamerman, D. (2000). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. *Journal of the American Statistical Association*. <https://doi.org/10.2307/2669581>
- Subirats, P., & Dupuis, Y. (2015). Overhead LIDAR-based motorcycle counting. *Transportation Letters - The International Journal of Transportation Research*, 7(2), 114–117. <https://doi.org/10.1179/1942787514Y.0000000038>
- Tarko, A. (2019). *Measuring Road Safety with Surrogate Events* (1st ed.). Cambridge, M.A.: Elsevier Ltd.
- Tarko, A., Davis, G., Saunier, N., Sayed, T., & Washington, S. (2009). Surrogate measures of safety. *Transportation Research Board of the National Academies.*, 20(3), 1–13.
- Tarko, A. (2012). Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2011.07.008>

- Tarko, A., & Lizarazo, C. G. (2020). Validity of Failure-caused Traffic Conflicts as Surrogates of Rear-end Collisions in Naturalistic Driving Studies. *Accident Analysis and Prevention, In Review*.
- Tarko, A. (2018). Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution – A theoretical and numerical exploration. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2018.01.008>
- Tarko, A., Ariyur, K. B., Romero, M. A., Bandaru, V. K., & Lizarazo, C. G. (2016). *T-Scan: Stationary LiDAR for Traffic and Safety Applications—Vehicle Detection and Tracking*. <https://doi.org/10.5703/1288284316347>
- Tarko, A., Bar-Gera, H., Thomaz, J., & Issariyanukula, A. (2010). Model-based application of abbreviated injury scale to police-reported crash injuries. *Transportation Research Record*. <https://doi.org/10.3141/2148-07>
- Tsoi, A. H., & Gabler, H. C. (2015). Evaluation of Vehicle-Based Crash Severity Metrics. *Traffic Injury Prevention*. <https://doi.org/10.1080/15389588.2015.1067693>
- Uchida, N., Kawakoshi, M., Tagawa, T., & Mochida, T. (2010). An investigation of factors contributing to major crash types in Japan based on naturalistic driving data. *IATSS Research*. <https://doi.org/10.1016/j.iatssr.2010.07.002>
- Velodyne. (2016). *VLP-16 datasheets*. Retrieved from http://velodynelidar.com/docs/datasheet/63-9229_Rev-C_VLP16_Datasheet_Web.pdf
- Viano, D. C., & Parenteau, C. S. (2010). Injury risks in frontal crashes by delta V and body region with focus on head injuries in low-speed collisions. *Traffic Injury Prevention*. <https://doi.org/10.1080/15389581003751623>
- Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2008.06.001>
- Wang, Z., Chen, H., & Lu, J. J. (2009). Exploring impacts of factors contributing to injury severity at freeway diverge areas. *Transportation Research Record*. <https://doi.org/10.3141/2102-06>
- Washington, S., Karlaftis, K., & Mannering, F. (2011). *Statistical and Econometric Methods for Transportation Data Analysis* (2nd ed.). Boca Raton, FL: CRC Press.
- Waymo. (2020). *Safety Report Waymo. Waymo*. <https://doi.org/10.1016/b978-0-08-037539-7.50012-0>

- Williams, M. J. (1981). Validity of the traffic conflicts technique. *Accident Analysis and Prevention*. [https://doi.org/10.1016/0001-4575\(81\)90025-7](https://doi.org/10.1016/0001-4575(81)90025-7)
- Wood, D. P., & Simms, C. K. (2002). Car size and injury risk: A model for injury risk in frontal collisions. *Accident Analysis and Prevention*. [https://doi.org/10.1016/S0001-4575\(01\)00003-3](https://doi.org/10.1016/S0001-4575(01)00003-3)
- Wu, K. F., & Jovanis, P. (2013). Screening naturalistic driving study data for safety-critical events. *Transportation Research Record*. <https://doi.org/10.3141/2386-16>
- Zheng, L., Ismail, K., & Meng, X. (2014). Traffic conflict techniques for road safety analysis: Open questions and some insights. *Canadian Journal of Civil Engineering*. <https://doi.org/10.1139/cjce-2013-0558>
- Zheng, L., Ismail, K., Sayed, T., & Fatema, T. (2018). Bivariate extreme value modeling for road safety estimation. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2018.08.004>
- Zheng, L., Sayed, T., & Essa, M. (2019). Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2018.12.007>
- Zorn, C. (2005). A Solution to separation in binary response models. *Political Analysis*. <https://doi.org/10.1093/pan/mpi009>

APPENDIX A. DESCRIPTION DATA REQUEST SHRP2 INITIATIVE

SHRP2 – Data Use License Description Data request

Summary of Phase I

In the first phase, VTTI randomly selected 255 trips with at least one conflict from the entire population of 5,411,204 trips. A trip could be selected only once (drawing without returning). The search for conflicts in the selected trips was facilitated with an algorithm written in Matlab and provided by the Purdue team. IDs of trips drawn without conflicts were saved. VTTI delivered to Purdue the IDs of 255 trips with at least one traffic conflict and 1320 trip IDs without conflicts. In addition to the two lists of trip IDs, the following information was provided for each trip (with and without conflicts):

- (1) trip information included in the trip summary table,
- (2) driver information included in the driver summary table,
- (3) vehicle information included in the vehicle summary table,
- (4) event information included in the event summary table (if available).

For each conflict identified with the Matlab script, the following information was provided:

- (1) ID of the trip during which the conflict occurred,
- (2) time-series data obtained with the Matlab algorithm,
- (3) two video frames associated with a conflict (frame IDs extracted with the algorithm).

For 200 trips randomly selected from the trips without conflicts the following information was provided:

- (1) trip ID,
- (2) two randomly selected video frames at approximately 1/3 of the trip time and 2/3 of the trip time.

Purdue inspected the provided data. The evaluation provided some changes in the methodology and the algorithm described as follows.

Methodology of Phase II

The data scope and processing in the second phase are similar to those in Phase I. This time, however, VTTI is supposed to provide Purdue with IDs of 10,000 trips with at least one traffic conflict and the IDs of all the conflict-free trips that were drawn randomly when searching for the trips with conflicts.

The data provided for each trip (with and without conflicts) should include trip, driver, vehicle and event information. In addition, a conflict-free car-following sequence will be extracted with the provided Matlab script from each trip that includes at least one traffic conflict. This additional information is needed to characterize conflict-free behavior of the driver. For each trip with at least one conflict the following information is to be provided:

- (1) ID of the trip with a traffic conflict,
- (2) Conflict time-series data obtained with the Matlab algorithm,
- (3) Conflict-free car-following time series data obtained with the Matlab algorithm,
- (4) Two video frames associated with a conflict (frame IDs extracted with the algorithm).

For 10,000 conflict-free trips randomly selected when searching for trips with conflicts, the following information is requested (same as in phase I):

- (1) trip ID,
- (2) two randomly selected video frames at approximately 1/3 of the trip time and 2/3 of the trip time.

The evaluation of the traffic conflict data obtained with the Matlab algorithm in Phase I indicated contradictions between data collected with radar, speedometer, and IMU (acceleration). If the Purdue team finds such contradictions in the new data obtained in Phase II, Purdue will request video clips of 40 seconds concurrent with the contradicting sequence of data to clarify the source of the contradictions. This new piece of information is needed to help confirm and reject the event as a conflict.

Algorithm

A modified Matlab script for identification of potential conflicts is summarized as follows.

In the first step, the missing values of Speed_network and Speed_GPS variables are obtained through interpolation between available values.

In the second step, the algorithm identifies the sequence of time frames with a lead vehicle using the processed radar variable IS_LEAD_VEHICLE.

For each of the identified sequences, the following applies:

- (1) Calculate TTC every 0.1 seconds,
- (2) Find time frames with $TTC < 1.5$ seconds,
- (3) Find the time frame with the lowest TTC,
- (4) Check if the time frame of the lowest TTC is NOT in the boundary when the Lead Vehicle Indicator Variable changes from “is lead vehicle” to “is not lead vehicle”.
- (5) Check if at least one of the values of acceleration between point A and point C (See Figure A.1) is lower than $-0.1g$ and speed at point A is greater than zero.

- (6) If the conditions in steps 4 and 5 are met, extract time series data 30 seconds before point A and 10 seconds after point B.
- (7) Extract video frames at point A and one second after this point.
- (8) In case a conflict is identified in the trip, extract 3 minutes of time series data with a lead vehicle.

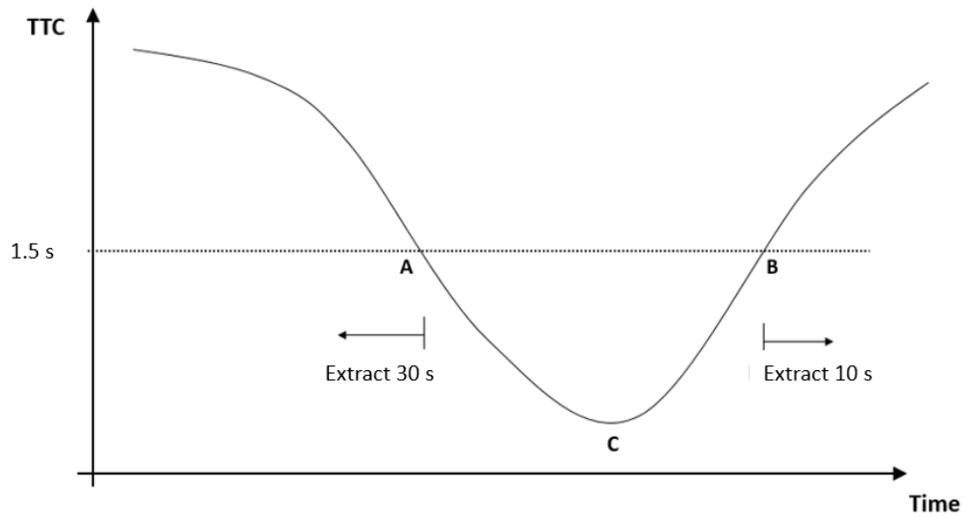


Figure A.1. Example TTC values that indicate a potential conflict

APPENDIX B. ALGORITHM EXTRACTION POTENTIAL CONFLICTS

```
%% Load the file
fileName = 'Event_ID_159987044debug.csv';
rawData = textread(fileName, '', 'delimiter', ',', ...
    'emptyvalue', NaN, 'headerlines', 1);

[r,c] = size(rawData);
temp = rawData;
timeStamps = rawData(:,1);
maxTimeValue=max(timeStamps);
validTimeStamps = false(size(timeStamps));

tic

%Load the video
v = VideoReader('E:\File_ID_1236_Front.mp4');

% spaces where manually converted to -9999 in csv. This is to distinguish
% between blank readings and Numeric 0. Blank readings are taken as NaN for
% easier Matlab representation.
temp(temp == -9999) = nan;
dataCorrected = temp;

%% Lead vehicle identification
% Find out if lead vehicle exists and note down the ID of lead vehicle
% columns 127 to 134 (TRACK1_IS_LEAD_VEHICLE to TRACK8_IS_LEAD_VEHICLE)
isLeadVehicleArray = dataCorrected(:, 127:134);
leadVehicleIndices = isLeadVehicleArray == 1;
targetIDs = dataCorrected(:, 63:70);
leadVehicleIDs = zeros(size(targetIDs));
leadVehicleIDs(leadVehicleIndices) = targetIDs(leadVehicleIndices);

% Sometimes, data has two lead vehicles in the same frame. We choose the
% lead vehicle to be the nearest one
leadVehicleIndicesCol = sum(leadVehicleIndices,2);
duplicates = find(leadVehicleIndicesCol == 2);
distanceX = dataCorrected(:, 71:78);
distanceLead = zeros(size(distanceX));
distanceLead(leadVehicleIndices) = distanceX(leadVehicleIndices);
distanceLeadColumn=sum(distanceLead,2);
temp = distanceLead(duplicates,:);
temp(temp == 0) = inf;
[distanceLeadColumn(duplicates),pos] = min(temp,[],2);

%%
% The ID of the target vehicle is not important, what is important is
% whether there is a lead vehicle or not. So collapse the 8 columns into
% one.
leadVehicleIDsColumn = sum(leadVehicleIDs,2);
for i = 1:length(duplicates)
    leadVehicleIDsColumn(duplicates(i)) =
leadVehicleIDs(duplicates(i),pos(i));
    t = leadVehicleIndices(duplicates(i),pos(i));
```

```

        leadVehicleIndices(duplicates(i),:) = false;
        leadVehicleIndices(duplicates(i),pos(i)) = t;
end
nonZeroVehicleIDIndices = leadVehicleIDsColumn ~= 0;
nonZeroVehilceIds      = leadVehicleIDsColumn(nonZeroVehicleIDIndices);

%%
% now pick out stretches where a lead vehicle exists

diffInVals = diff(nonZeroVehilceIds);
indicesOfIDjumps = find(diffInVals~=0);
indicesOfIDjumps2 = [0;indicesOfIDjumps;length(nonZeroVehilceIds)]; % Insert
a 1 for finding out the number of readings with first ID
timeStretches = diff(indicesOfIDjumps2);
% validTimeStreches = timeStretches>149; % 150 = 15 seconds (10 hz data rate)

noOfStretches    = length(indicesOfIDjumps2)-1;
timeStretchVals = zeros(noOfStretches,2);
temp = 1:1:(noOfStretches+1);
timeStretchVals(:,1) = indicesOfIDjumps2(temp(1):temp(end-1))+1;
timeStretchVals(:,2) = indicesOfIDjumps2(temp(2):temp(end));

temp            = false(size(nonZeroVehilceIds));
stretchNumber   = zeros(size(nonZeroVehilceIds));

%Identify the number of following events
for i = 1:1:noOfStretches
    temp(timeStretchVals(i,1) : timeStretchVals(i,2)) = true;
    stretchNumber(timeStretchVals(i,1) : timeStretchVals(i,2)) = i;
end

validTimeStamps(nonZeroVehicleIDIndices) = temp;
stretchNumberVec = zeros(size(validTimeStamps));
stretchNumberVec(nonZeroVehicleIDIndices) = stretchNumber;
%% Vehicle nearness based conflict
% Once the vehicle following is established, we look for conflicts. First
% type of conflict identification is based on nearness.

%Interpolate network speed
aux1=0;
speedMovingNet = dataCorrected(:,54) ./3.6;
if nansum(speedMovingNet)==0
    speedMovingNet = dataCorrected(:,53) ./3.6;
end
if nansum(speedMovingNet)==0
    aux1=1;
end
%Empty if there are not reported speed values
if aux1==0
    %speedMoving_inter = speedMoving;
    tSpeedMoving = isnan(speedMovingNet);
    f_indx = find(~tSpeedMoving, 1);
    prevReading = speedMovingNet(f_indx);
    noReadingCount=0;
    for k = f_indx:1:length(speedMovingNet)

```

```

    if (tSpeedMoving(k))
        noReadingCount = noReadingCount + 1;
    elseif(noReadingCount == 0)
        continue
    else
        interpVals = 1:1:noReadingCount;
        interpMultip = ((speedMovingNet(k))-
prevReading)/(noReadingCount+1);
        speedMovingNet((k - noReadingCount):(k-1)) = ...
            (interpVals*interpMultip) + prevReading;
        prevReading = speedMovingNet(k);
        noReadingCount = 0;
    end

end

speedMovingNet( leadVehicleIndicesCol < 1) = 0;

speedDiffLead    = dataCorrected(:, 87:94);
speedDiff        = zeros(size(speedDiffLead));
speedDiff(leadVehicleIndices) = speedDiffLead(leadVehicleIndices);
speedDiffColumn = sum(speedDiff,2);

negativeSpeedDiff = speedDiffColumn < 0;
ttc = distanceLeadColumn ./ (-speedDiffColumn) ;
ttc(~negativeSpeedDiff) = inf;
%% Add ttc estimation into raw data
dataCorrected=[dataCorrected ttc];

%% ttc threshold for vehicle nearness is 2.0 seconds or less
ttcNear = ttc < 2.0;
speedThreshold = 0;
%% Vehicle deceleration rates
% we look at time periods when the deceleration is high
accelData = dataCorrected(:,5);
accelData( leadVehicleIndicesCol < 1) = inf;

%Theshold Deceleration. Add condition for nearness when deceleration rate
%is greater than 0.1
isSuddenDecelTTC=accelData < -0.1 ;

% suddenDeceleration is defined when deceleration rate is greater than
0.6
isSuddenDecel = accelData < -0.6 ;

% Bring the video frames for weather extraction
videoFrames = dataCorrected(:,60);

%% Loop through the time stretches
% Loop through the time stretches that have a lead vehicle and find out
% periods that satisfy the various conflict criteria.

outputReadings          = false(size(timeStamps));
outputReadingsVec       = zeros(size(timeStamps));
stetchCount             = 0;
conflictStrtchesFull    = false(size(stretchNumberVec));

```

```

conflictStretches      = false(size(stretchNumberVec));

dataCorrected=[dataCorrected stretchNumberVec speedMovingNet];

%Initialize variables for non-conflict time series
stretchLengthCount=0;
extractNonConflict=0;
lastConflictEvent=0;

for i = 1:1:noOfStretches
    stretchFrameNos      = find(stretchNumberVec == i);
    stretchLength        = length(stretchFrameNos);
    outputReadingsPartial = outputReadings(stretchFrameNos);

    % Bring frames from videos
    exVideoFrames=videoFrames(stretchFrameNos);

    % check based on nearness
    isTtcDanger      = ttcNear(stretchFrameNos);
    %Check where is the minimum TTC
    exTtc      = ttc(stretchFrameNos);
    exTtcD      = exTtc(isTtcDanger);
    findTtcD=find(isTtcDanger,1);
    [value, index] = min(exTtcD(:));
    [row, col] = ind2sub(size(exTtcD(:)), index);
    % Find location of min TTC
    boundCondition=0;
    if isempty(value)==0
        FrameNoMin = find(exTtc == value);
        if FrameNoMin(end,1)>=(size(exTtc,1)-2)
            boundCondition=1;
        end
    end
    %Check special case
    isSplCaseT      = isTtcDanger(1) == true;

    diffIsTtcDanger      = diff(isTtcDanger);
    indDiffIsTtcDanger = find(diffIsTtcDanger~=0);

    oddNosT = 1:2:length(indDiffIsTtcDanger);
    oddNosT = oddNosT + isSplCaseT*1;
    oddNosT = oddNosT(1:(end - isSplCaseT));

    stretchSpeedMovinng = speedMovingNet(stretchFrameNos);
    speedChk = stretchSpeedMovinng(indDiffIsTtcDanger(oddNosT)) >
speedThreshold;

    %Is desire Deceleration of 0.1

    isDesireDeccl1 = isSuddenDecelTTC(stretchFrameNos);
    isDesireDeccl2=isDesireDeccl1(isTtcDanger);
    isDesireDeccl3=isDesireDeccl2(1:row);
    % check based on driver behavior
    stretchIsSuddenDecel = isSuddenDecel(stretchFrameNos);
    isSplCaseD = stretchIsSuddenDecel(1) == true;

```

```

diffIsDecelDanger = diff(stretchIsSuddenDecel);
indDiffIsDecDanger = find(diffIsDecelDanger~=0);

oddNosD = 1:2:length(indDiffIsDecDanger);
oddNosD = oddNosD + isSplCaseD*1;
oddNosD = oddNosD(1:(end - isSplCaseD));

decelChk = stretchSpeedMovinnng(indDiffIsDecDanger(oddNosD)) >
speedThreshold;

if (sum(decelChk) > 0) || (sum(speedChk) > 0 && sum(isDesireDeccl3)>0
&& boundCondition==0)
    stretchCount = stretchCount + 1;
    conflictStrtchesFull(stretchFrameNos) = true;
    rowSpeed=-999;
    rowDecel=-999;
    startFramesT=[];
    startFramesD=[];
    endFramesT=[];
    endFramesD=[];
    if (sum(speedChk) > 0 && sum(isDesireDeccl3)>0 &&
boundCondition==0)

frameSpeedChk=stretchFrameNos(indDiffIsTtcDanger(oddNosT(speedChk)));
    rowSpeed = zeros(size(frameSpeedChk));
    for k = 1:1:length(frameSpeedChk)
        rowSpeed(k) = find(timeStamps==frameSpeedChk(k));
    end
    %Extract 30 seconds before and 10 seconds after
    startFramesT = rowSpeed - 300;
    endFramesT = rowSpeed + 100;
    startFramesT(startFramesT<0)=0;
    endFramesT(endFramesT>maxTimeValue)=maxTimeValue;
end
if sum(decelChk) > 0

frameDecelChk=stretchFrameNos(indDiffIsDecDanger(oddNosD(decelChk)));
    rowDecel= zeros(size(frameDecelChk));
    for k = 1:1:length(frameDecelChk)
        rowDecel(k) = find(timeStamps==frameDecelChk(k));
    end
    %Extract 30 seconds before and 10 seconds after
    startFramesD = rowDecel - 300;
    endFramesD = rowDecel + 100;
    startFramesD(startFramesD<0)=0;
    endFramesD(endFramesD>maxTimeValue)=maxTimeValue;
end
startFramesTWeather=indDiffIsTtcDanger(oddNosT(speedChk));
startFramesDWeather = indDiffIsDecDanger(oddNosD(decelChk));
startFrames = [startFramesT;startFramesD];
startFramesPos=startFrames(startFrames>0);
startFramesWeather=[startFramesTWeather;startFramesDWeather];
endFrames = [endFramesT;endFramesD];
endFramesPos=endFrames(startFrames>0);
[startFramesPos, orderI] = sort(startFramesPos);

```

```

startFramesWeather=sort(startFramesWeather);
startFramesWeatherMod=startFramesWeather+10;

startFramesWeatherMod(startFramesWeatherMod>stetchLength)=stetchLength;
endFramesPos = endFramesPos(orderI);

startFramesF = startFramesPos;
endFramesF = endFramesPos;

if length(startFrames) > 1
    diffStPeriods = startFrames(2:end) - endFrames(1:end-1);
    smallGaps = find(diffStPeriods < 600);
    startFramesF(smallGaps + 1) = [];
    endFramesF(smallGaps) = [];
end

for k = 1:1:length(startFramesF)
    outputReadings(startFramesF(k):endFramesF(k)) = true;
    outputReadingsVec(startFramesF(k):endFramesF(k))=1;
end
for k=1:1:length(startFramesWeather)
    videoFrameID=exVideoFrames(startFramesWeather(k));
    video = read(v,videoFrameID);
    imwrite(video, strcat('E:\File_ID_1236_Front_Frame_',
num2str(videoFrameID), '_1.jpg'));

    videoFrameID=exVideoFrames(startFramesWeatherMod(k));
    video = read(v,videoFrameID);
    imwrite(video, strcat('E:\File_ID_1236_Front_Frame_',
num2str(videoFrameID), '_2.jpg'));
end
%Extract nonconflict. Go back to previous car following events
if extractNonConflict==0
    currentStretch=i-1;
    while currentStretch>0
        stretchFrameNoCon = find(stretchNumberVec ==
currentStretch);
        stetchLengthNoCon = length(stretchFrameNoCon);
        % check based on nearness
        isTtcDanger = ttcNear(stretchFrameNoCon);
        % Check the car following is greater than 30 seconds
        if stetchLengthNoCon>=300 && sum(isTtcDanger)==0

stetchLengthCount=stetchLengthCount+stetchLengthNoCon;
        % Extract 4 minutes of car following conflict-free
        if (stetchLengthCount-stetchLengthNoCon)<=2400
            outputReadings(stretchNumberVec ==
currentStretch) = true;
            outputReadingsVec(stretchNumberVec ==
currentStretch)=2;
        end
    end
    currentStretch=currentStretch-1;
end
extractNonConflict=1;
end

```

```

        lastConflictEvent=i;
    end
end
%In case the car following was not found before the conflict event
%check after for extraction
if stretchLengthCount<=2400 && lastConflictEvent~=0
    for i = lastConflictEvent+1:1:noOfStretches
        stretchFrameNos        = find(stretchNumberVec == i);
        stretchLength           = length(stretchFrameNos);
        isTtcDanger              = ttcNear(stretchFrameNos);
        if stretchLength>300 && sum(isTtcDanger)==0
            stretchLengthCount=stretchLengthCount+stretchLength;
            if (stretchLengthCount-stretchLength)<=2400
                outputReadings(stretchNumberVec == i) = true;
                outputReadingsVec(stretchNumberVec == i)=2;
            end
        end
    end
end
end
toc
% dataNeeded contains the sets of readings that we are interested in for
% the given trajectory without latitude (36) and longitude (41)
end
dataNeeded =
horzcat (dataCorrected (outputReadings,1:35),dataCorrected (outputReadings,37:40
),dataCorrected (outputReadings,42:end));
outputReadingsVec=outputReadingsVec (outputReadingsVec~=0);
dataNeeded= [dataNeeded, outputReadingsVec];
%% Visualization
% some plots for visualization
% ttcNeeded = ttc(outputReadings);
% ttcNeeded(ttcNeeded>10) = 10;
% scatter (timeStamps (outputReadings), ttcNeeded,2)

```

APPENDIX C. DATA POSTPROCESSING SHRP2 INITIATIVE

The analysis of traffic conflict requires complete and properly coordinate information of the following variables every 0.1 seconds:

1. Longitudinal speed of the instrumented vehicle
2. Longitudinal acceleration of the instrumented vehicle
3. Range and range rate between the instrumented and front vehicle. Range rate is derived from the change rate of range using centered approach.

After analyzing the time series data reported by VTTI the following issues were found:

1. Longitudinal speed is reported at non-constant rates with missing values for up to 2 seconds. Repeated values of speed for periods of up to 5 seconds.
2. Missing relation between longitudinal speed and acceleration. Acceleration values should be scaled to match the reported values of speed.

In order to mitigate these issues, the following steps are applied in order to complete and coordinate in time of the reported readings:

1. Apply shift between -1.5 to 1.5 seconds and apply the following process:
 - 1.1. Identify the frames with reported speed values in the vector $k = \{k_1, k_2, \dots, k_i \dots k_T\}$.
 - 1.2. Extract the reported speed for frames k_i and k_{i+1} with the acceleration values contained between these frames.
 - 1.3. Using the speed on k_i to estimate speed on frame k_{i+1} using the acceleration values a between frames k and $k+1$.
 - 1.4. Estimate a global scale factor γ for acceleration values that minimize the corresponding squared error (SE) between estimated and reported speeds on each $k+1$ frame

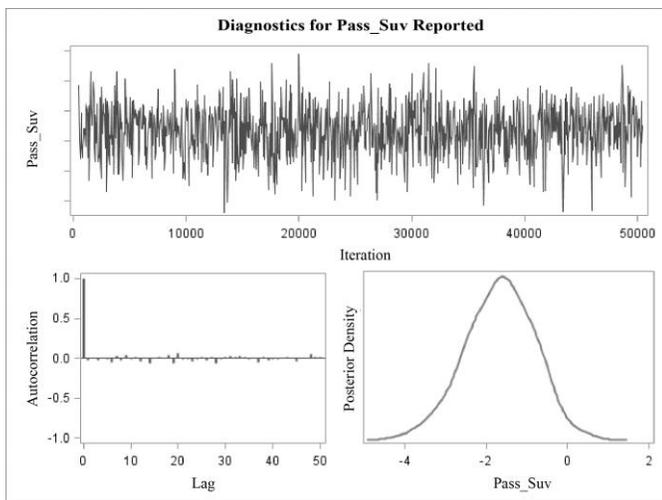
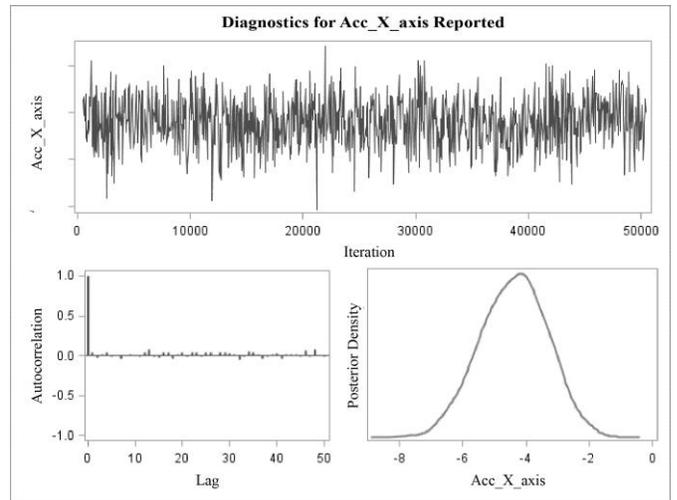
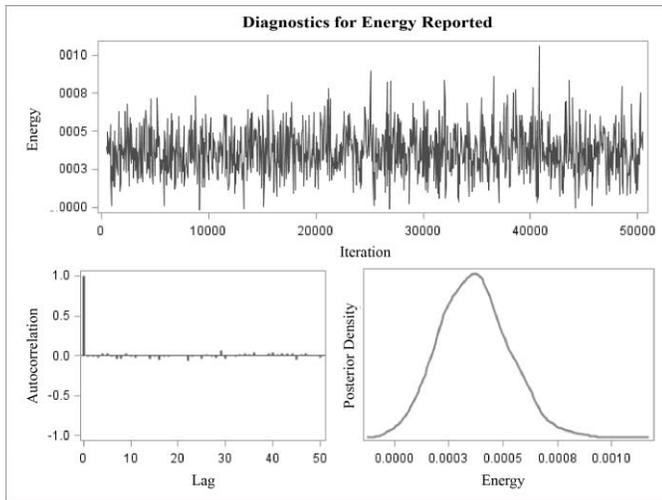
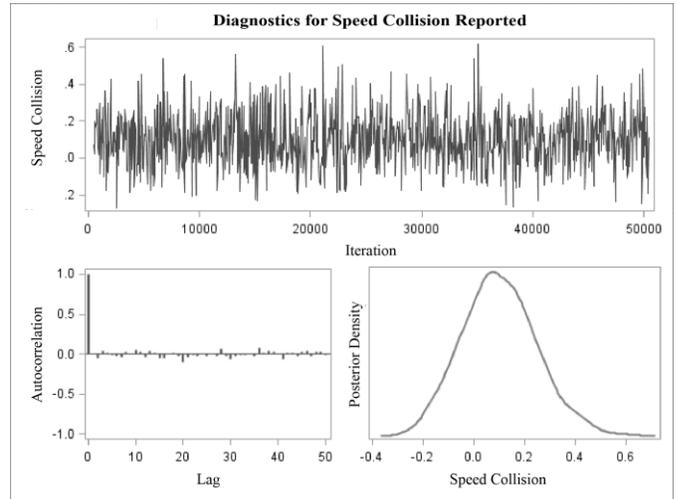
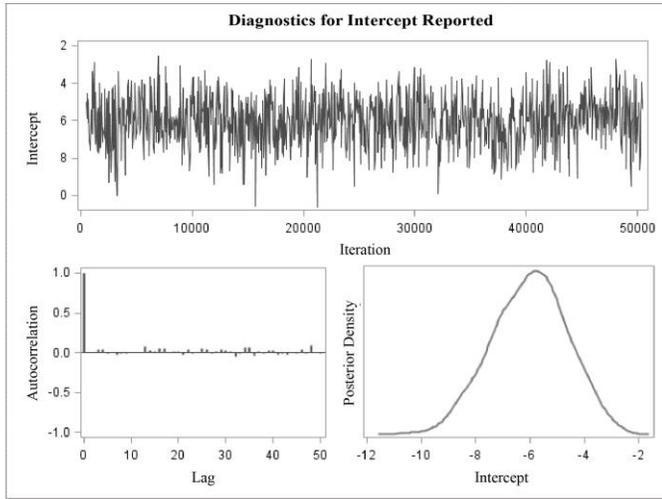
$$SE_m = \sum_{i=1}^T (v_i^{k+1} - \tilde{v}_i^{k+1})^2$$

Obtain the vector SE_m and γ_m where m corresponds to the different time shifts.

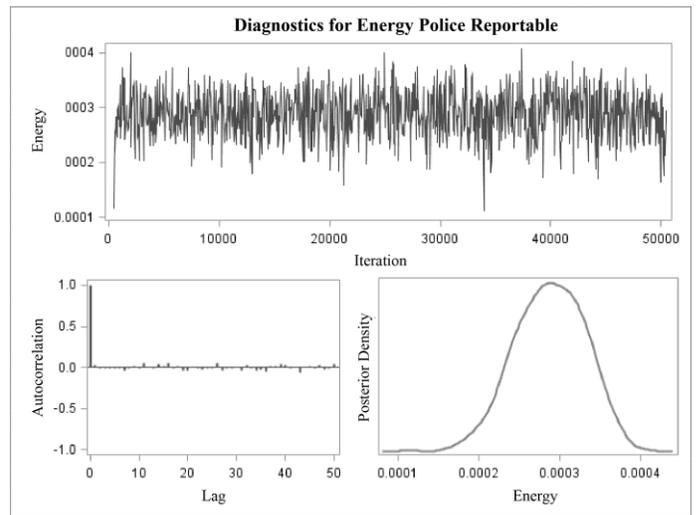
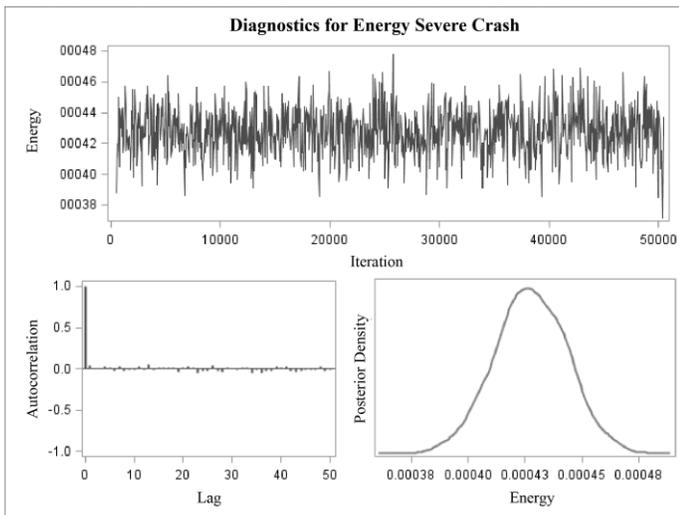
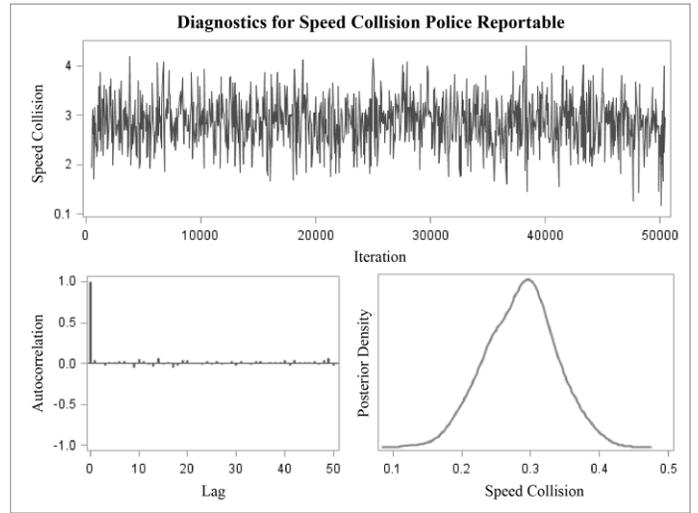
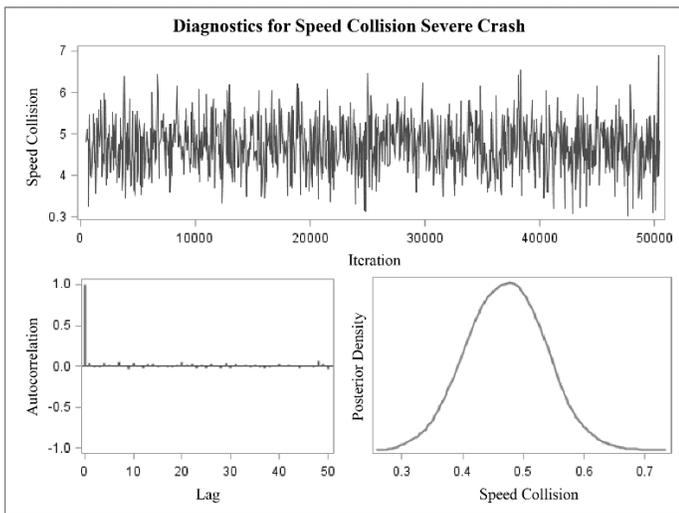
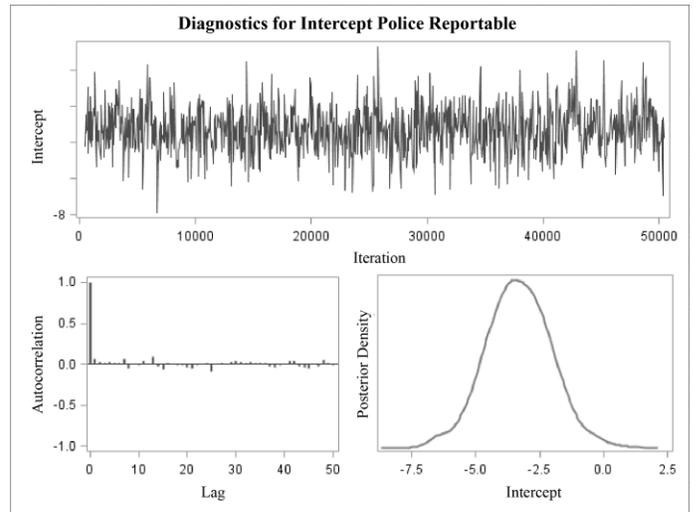
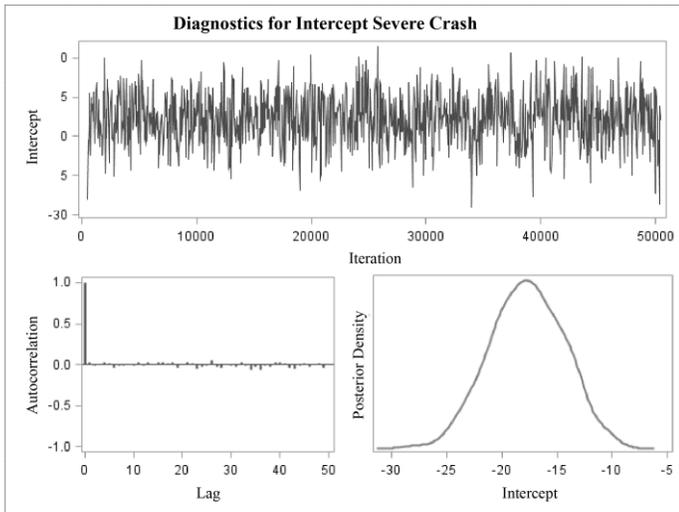
- 1.5 Estimate the minimum value of SE_m and use this one as the correspondent time shift between speed and acceleration
2. Estimate individual scaling factors to remove jumps in speed profile
3. Apply moving average to smooth speed values

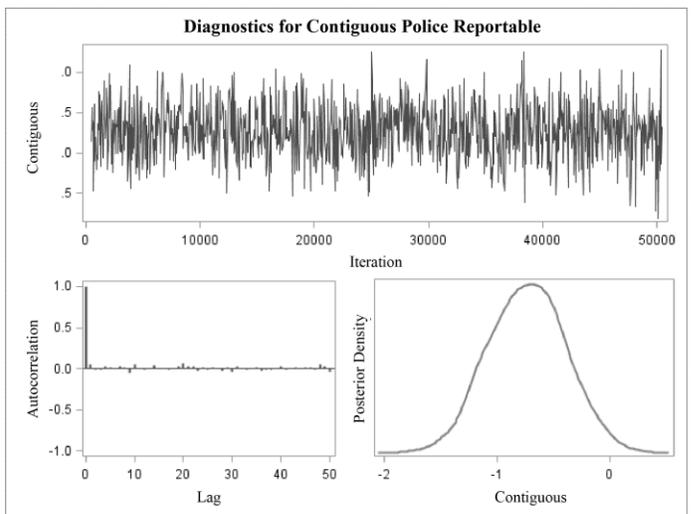
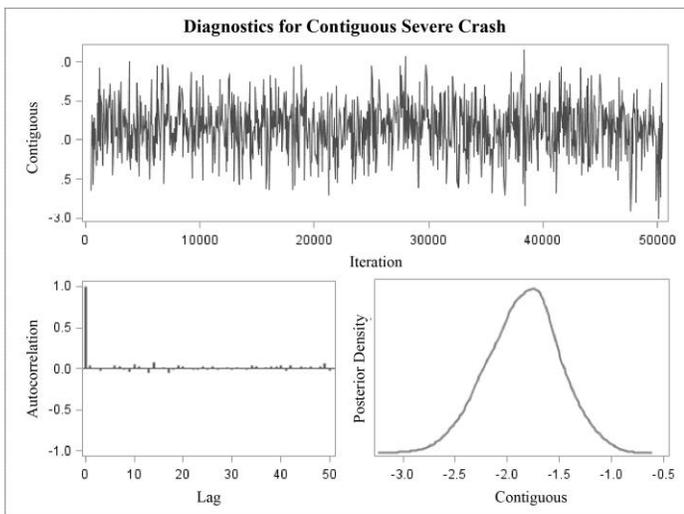
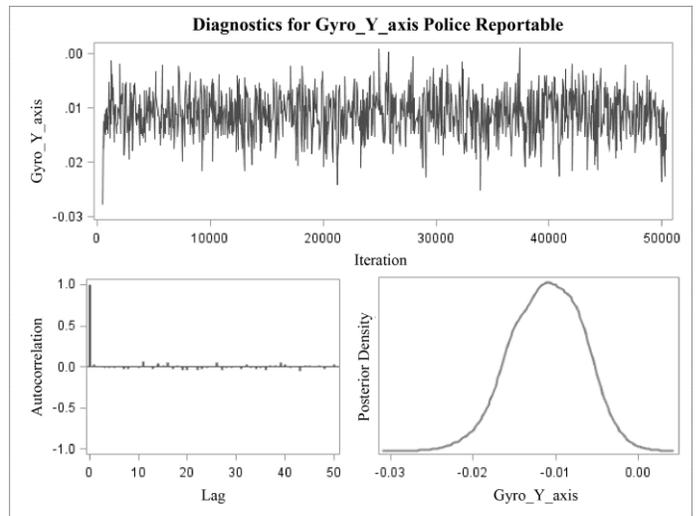
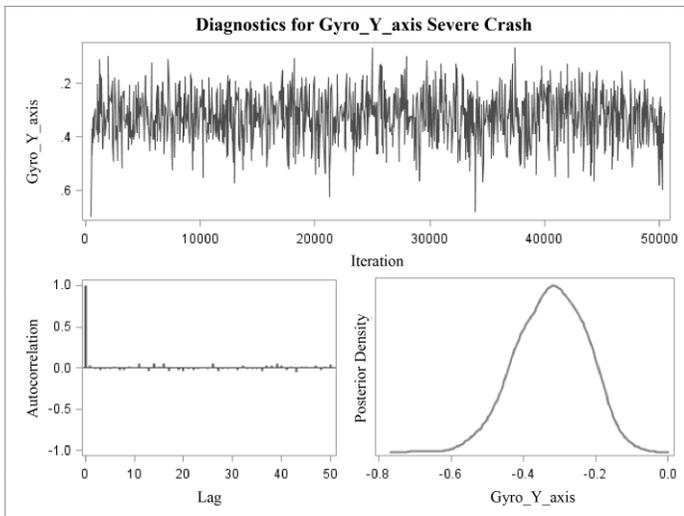
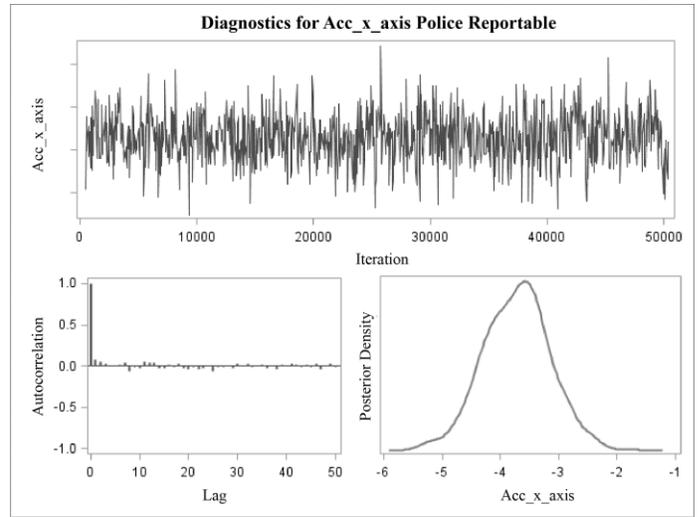
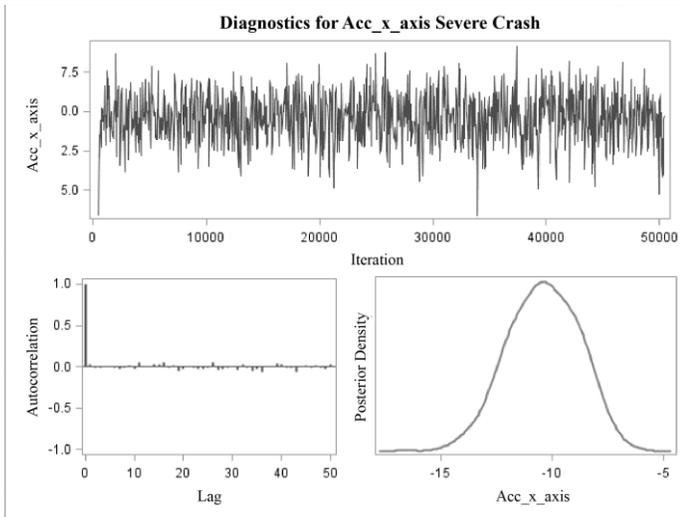
APPENDIX D. DIAGNOSTICS MULTINOMIAL LOGIT MODEL

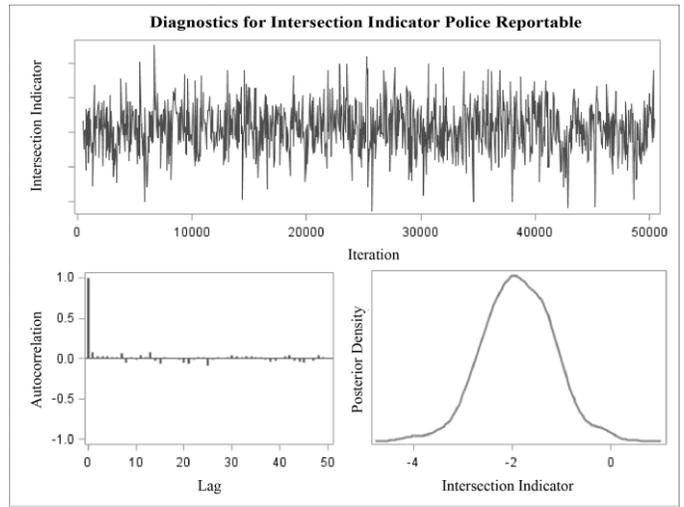
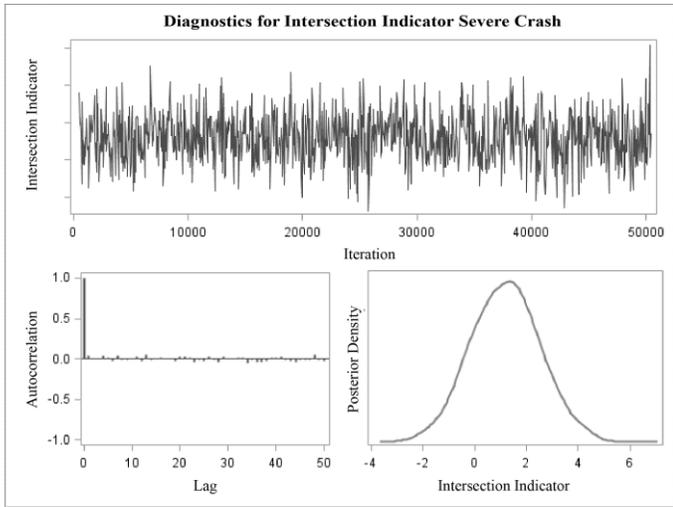
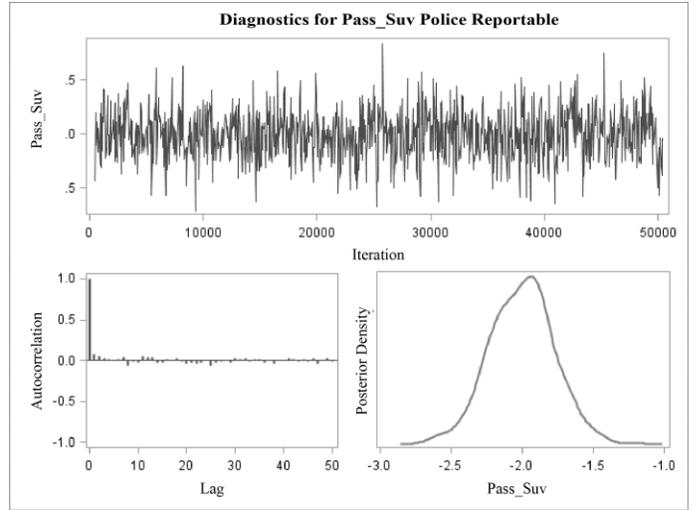
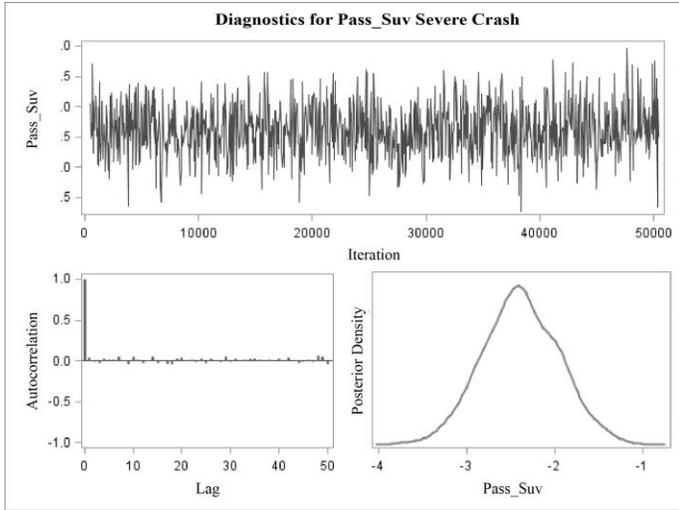
D.1. Diagnostics Bayesian Parameters Crash Reporting Model



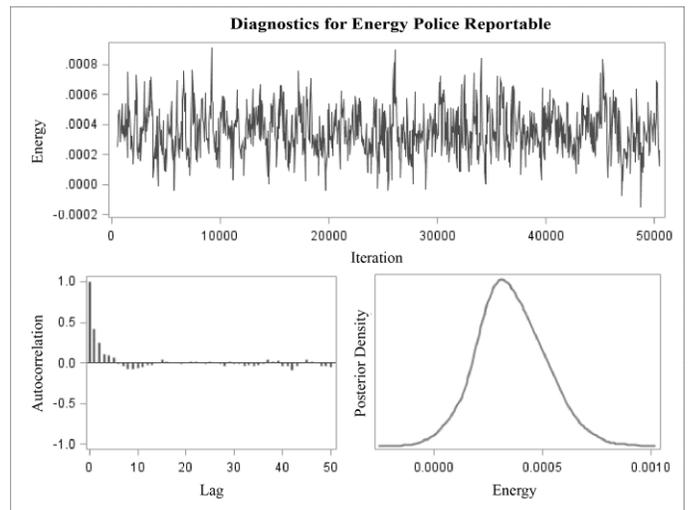
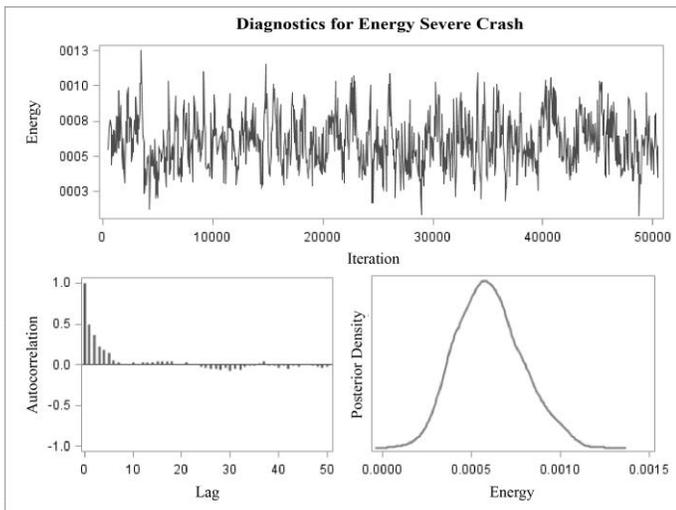
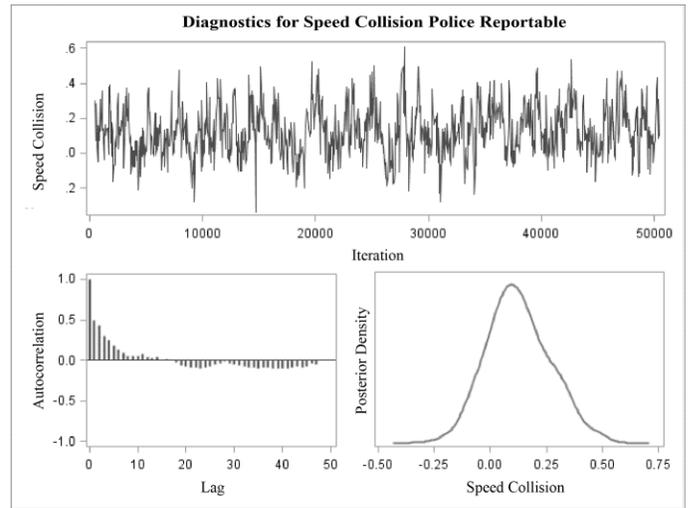
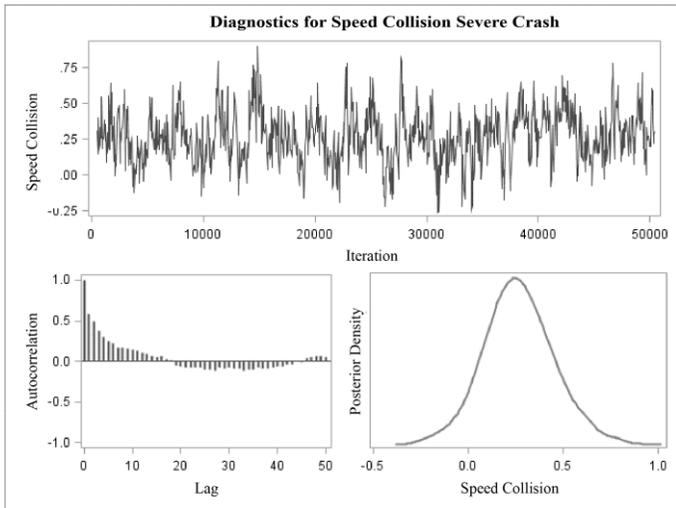
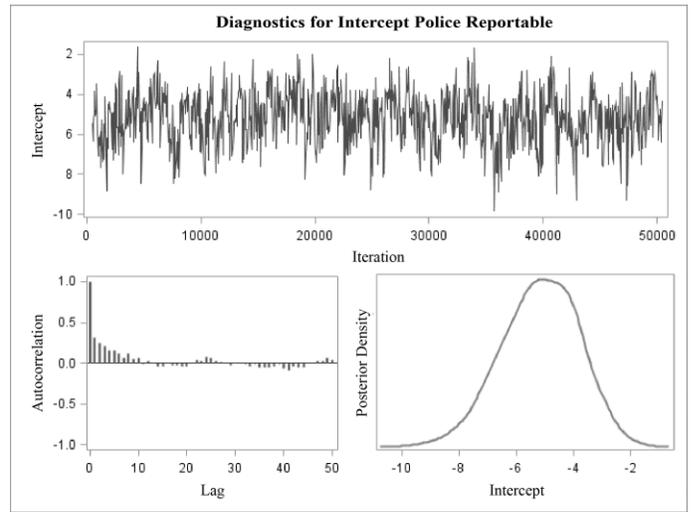
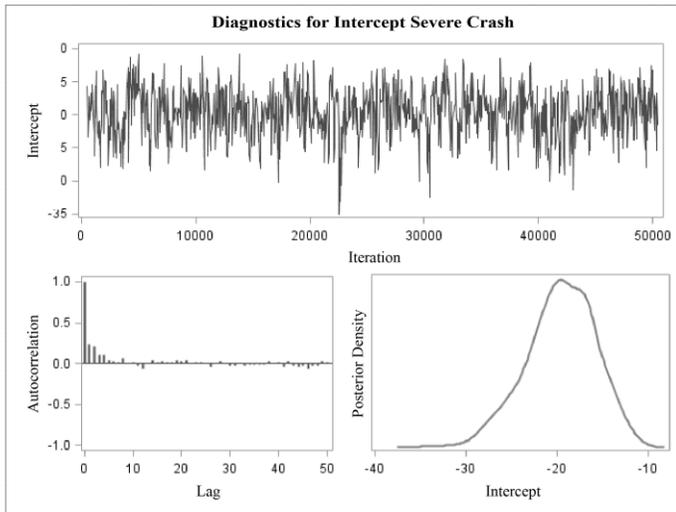
D.2. Diagnostics Bayesian Parameters Crash Severity Model with Intersection Indicator

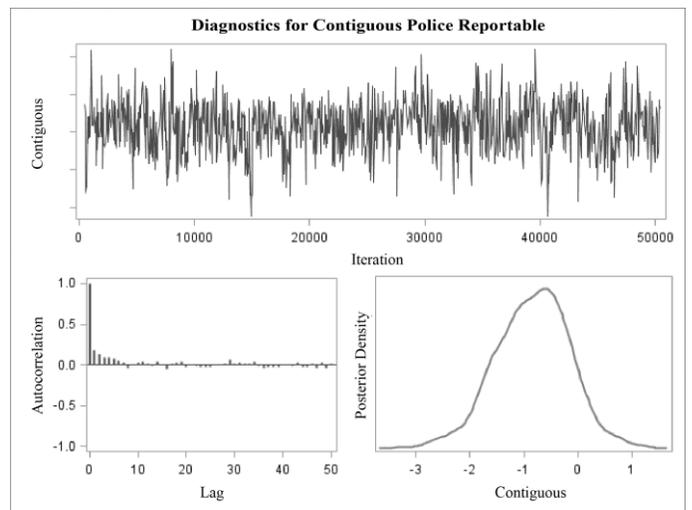
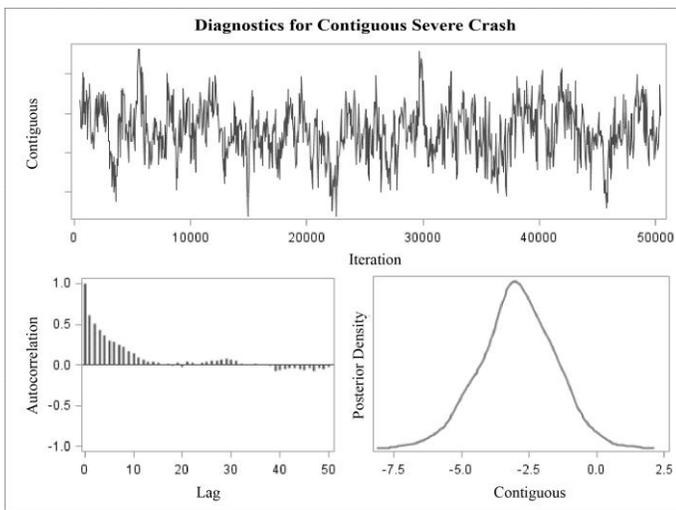
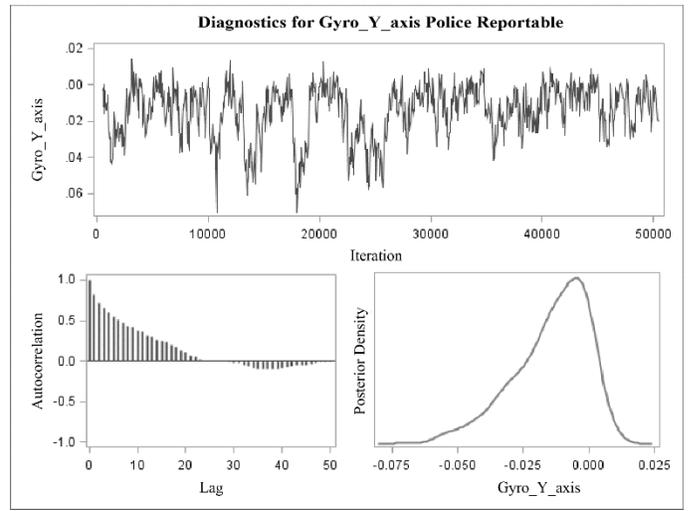
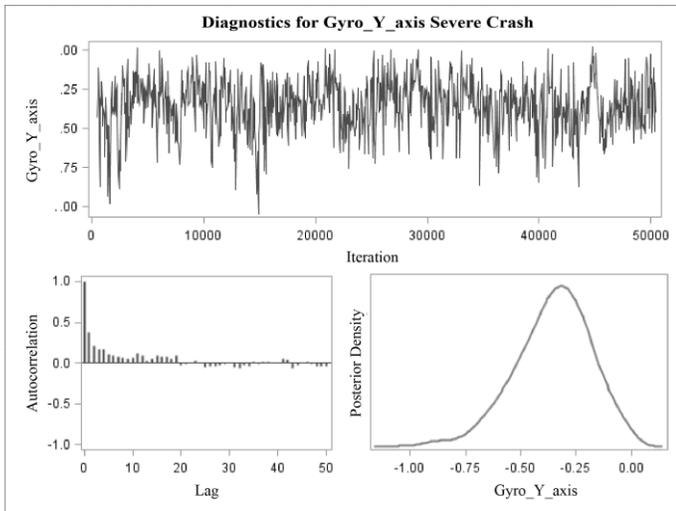
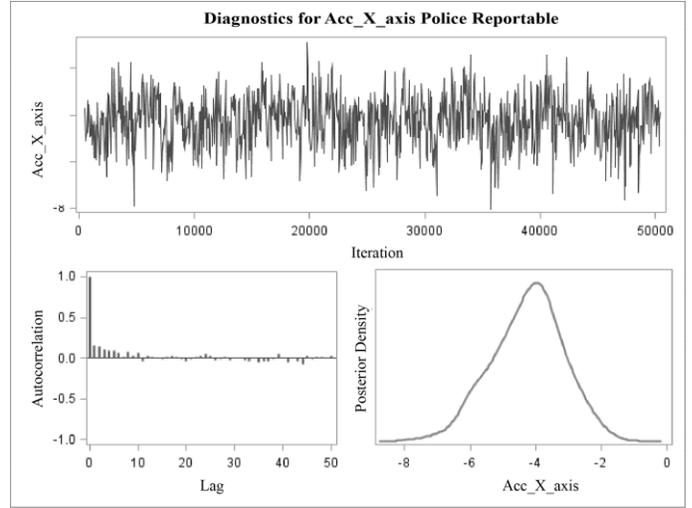
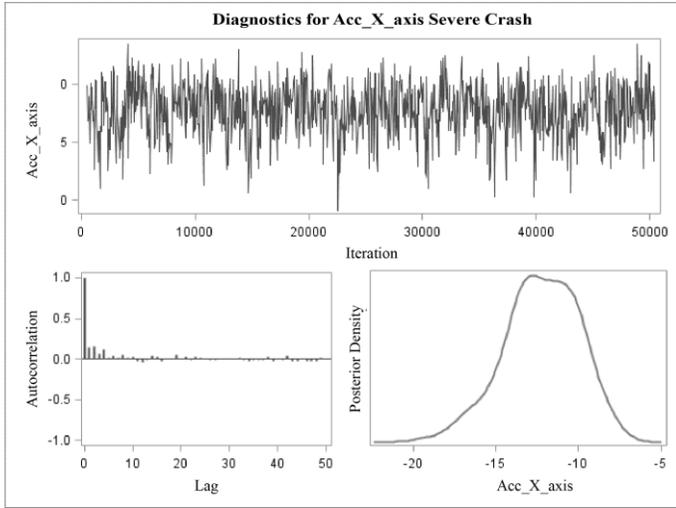


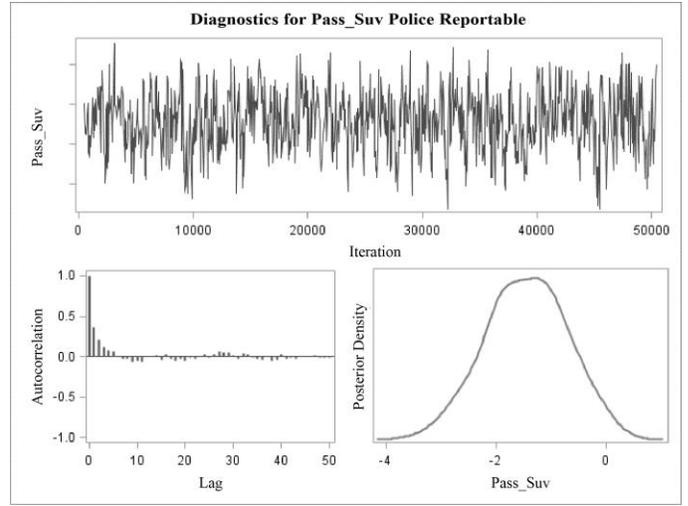
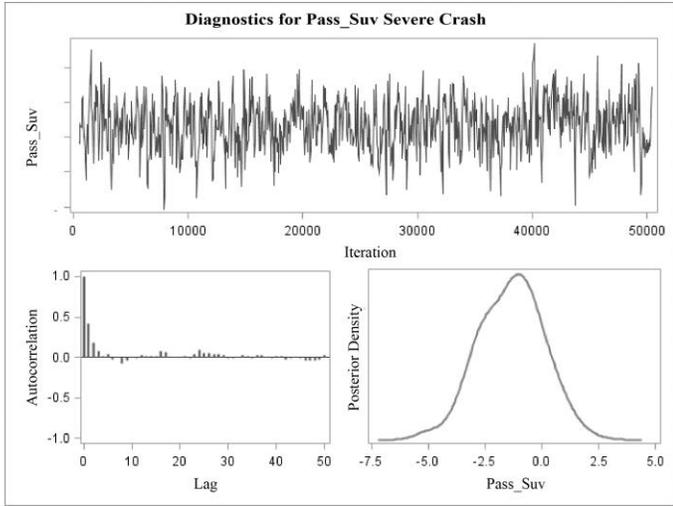




D.3. Diagnostics Bayesian Parameters Crash Severity Model without Intersection Indicator







VITA

Born into a family of dreamers in Bogota, Colombia, Cristhian Lizarazo learned early on the critical role of education in achieving greater success. Cristhian received his bachelor's degree in Civil Engineering from the National University of Colombia in August 2015. Inspired to pursue innovative solutions in transportation, he chose Purdue University for his graduate studies. It was here where he was afforded the tremendous opportunity to join the Center for Road Safety, first as an undergraduate research assistant and later as a graduate researcher. Cristhian received his master's degree in Civil Engineering in December 2016 with an emphasis in the Transportation and Infrastructure Systems area, and began pursuing his PhD immediately thereafter. In a proactive move to enhance his breadth of knowledge for his dissertation, he also received a master's degree in Applied Statistics in May 2020.

During his five years of graduate studies, Cristhian has been involved in more than 10 projects whose primary objective was to reduce the frequency and severity of road crashes. His research interests include detection and tracking of road users using emerging technologies, instrumented vehicles, and the connection of these two new data sources with traffic safety. His career goal is to advance transportation engineering practice in the era of autonomous vehicles when road safety will change at a never before seen pace.

A Ross Fellow at Purdue and Ponce de Leon Awardee from the Colombian Society of Engineers, Cristhian expects to receive his PhD in Civil Engineering from Purdue University in December 2020.