COMPREHENSIVE ANALYSIS, EVALUATION, AND APPLICATION OF SAFETY SURROGATE MEASURES

by

Zhaoxiang He

A Dissertation Submitted in

Partial Fulfillment of the

Requirements of the Degree of

Doctor of Philosophy

in Engineering

at

The University of Wisconsin-Milwaukee

May 2024

ABSTRACT

COMPREHENSIVE ANALYSIS, EVALUATION, AND APPLICATION OF SAFETY SURROGATE MEASURES

by

Zhaoxiang He

The University of Wisconsin-Milwaukee, 2024
Under the Supervision of Professor Xiao Qin

This dissertation demonstrates a comprehensive analysis, evaluation, and application of safety surrogate measures. Through five distinct analyses, the research has made several key contributions.

Firstly, the research developed a complete framework to process and analyze the large-scale safety Pilot Model Deployment (SPMD) connected vehicle data, allowing the extraction of meaningful safety insights. By building surrogate safety measures on the vehicle, trip, and link levels, the analyses have been capable of statistically relate those measures to real crash facts, validating their use as proxies for safety assessment. This technique avoids the limitation of conventional strategies that depend entirely on sparse crash data.

The research additionally highlighted the significance of incorporating driver behavior, including but not limited to speed and braking behavior, into the safety evaluation. This enriched data allowed for more comprehensive models which can capture the complex interactions among driver behavior, traffic conditions, and safety outcomes. The findings show the value of using the wealth of information through connected vehicle technologies to enhance road safety evaluation.

Statistical techniques, such as Negative Binomial regression and Extreme Value Theory (EVT) modeling, have been implemented to analyze the surrogate safety measures and predict the chance of extreme safety events. Implementation of these advanced methodologies offers a more rigorous and complicated technique to safety assessment.

Furthermore, the incorporation of a safety index into a route-selection algorithm highlighted the potential for connected vehicle data to inform and optimize transportation decision-making. By guiding drivers toward safer routes, this technique shows how safety considerations could be systematically integrated into navigation systems.

Despite the above contributions, the dissertation also recognized several obstacles that need to be addressed in future studies. These include but not limited to data quality and completeness issues, data processing algorithm issues, limited crash types in the safety measures, and the challenge of low connected vehicle penetration rates. Recommendations for future work include improving data collection methods, enhancing the data processing algorithms, expanding the scope of the safety measures and analysis, and developing effective strategies to address the low connected vehicle penetration rates.

By addressing those obstacles and pursuing those future research directions, the field of surrogate safety measures in road safety studies, especially with more comprehensive connected vehicle data, can keep evolving, supplying increasingly robust and actionable insights to improve road safety.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background

Traditionally, road safety is directly measured by crash frequency and injury severity. However, crash data are often plagued with data availability and quality problems [1]. When issues arise due to 1) small sample sizes and underreporting issue where crashes are not collected because certain reporting thresholds are determined or involving parties chose not to report to the officials, and 2) the lack of details to improve the understanding of crash mechanism, crash-based safety analyses can be inconclusive or even conflicting. The inadequacy of crash-based studies is particularly problematic when evaluating innovative safety treatments, which usually require years of waiting for sufficient observations or several installations, or for specific crash types such as pedestrian or bicycle-related crashes whose observations are rare and collecting an adequate number of observations proves challenging.

Due to these challenges, indirect measures not based on crashes but other crash-related traffic events have been proposed to measure traffic safety. These measures are usually called surrogate safety measures. Based on the predetermined threshold, a surrogate safety measure can be used to identify a traffic conflict, when two or more road users approach each other in space and time with a risk of crash unless they change their movements [1]. Compared to crash-based analysis, surrogate measures are proactive and aim to evaluate road safety without crash data or when crash data is limited [2], [3]. Surrogate measures can be used to assess the safety performance of a highway facility or to estimate the effectiveness of safety treatment. These measures can provide rapid evaluation of innovative intersection designs or new traffic control strategies that usually require more time to accumulate an adequate number of sites and crash history. They can measure the safety improvements for rare crash types such as pedestrian or bicycle crashes [1].

Despite the promising surrogate measures in traffic safety evaluation, many unresolved issues and intrinsic limitations exist, including but not limited to the selection of the appropriate surrogate measures and data collection. Challenges associated with surrogate safety measures from previous studies can be summarized below:

1. Surrogate safety measures can be in different structures and may perform differently regarding road safety. There are no comprehensive comparisons of the existing surrogate safety measures. When evaluating road safety, there is no unified guideline for using each surrogate measure, i.e., under what circumstances should one or more surrogate measures be considered. In addition, there is no reliable method to determine the threshold of a traffic conflict.
2. Another challenge is the integration of human factors into conflict detection and safety evaluation. While automated systems are adept at recording measurable conflict indicators, they often do not account for the psychological or behavioral elements that precede a conflict situation, such as driver distraction or aggression.
3. Surrogate safety measures collected by human observers are time-consuming, labor-intensive, and sometimes inaccurate. Many of these issues have been resolved by the rapid development of computer image and sensing technologies, where the space and time between moving objects can be automatically detected, measured, and tracked. The current progress in sensing technologies contributes to the collection of traffic data in general and surrogate safety measures in particular [1], [4], [5]. However, cost-effective techniques are not always available to develop surrogate measures.

1.2 Research Objectives

The goal of this dissertation is to investigate how to effectively evaluate road traffic safety in the absence of crash data. More specifically, this dissertation aims to:

1. Compare different surrogate safety measures in assessing road safety. Advantages, limitations, and applicable scopes will be evaluated for each type of surrogate safety measures.
2. Build a framework to develop appropriate surrogate safety measures for road safety evaluation.
3. Use emerging data sources like connected vehicle data to calculate surrogate safety measures. Data processing algorithms and Big Data analytics will be proposed.
4. Develop and test the application of the surrogate safety measures. One application can be personal navigation by incorporating road traffic safety.

To achieve the four research objectives, the author organized the dissertation into nine chapters which are elaborated in the next section.

1.3 Dissertation Outline

The remaining chapters of the dissertation are explained as follows:

Chapter 2 provides a comprehensive review of surrogate safety measures. This chapter will present an extensive overview of surrogate safety measures and identify key challenges and opportunities in the field.

Chapter 3 presents the list of research tasks.

Chapter 4 presents a framework to calculate and validate the surrogate safety measures using the emerging data source. Comprehensive data processing algorithm is presented. It also presents the comparison of different surrogate safety measures.

Chapter 5 presents the evaluation of road safety using the surrogate safety measures.

Chapter 6 illustrates the potential routing application of surrogate safety measures.

Chapter 7 summarizes the contributions of the dissertation and presents future research directions.

Chapter 2 Literature Review

This chapter presents a comprehensive review of surrogate safety measures. It aims to provide the readers with an overview of the state-of-the-art field of surrogate safety measures, including an overview of traffic conflicts, surrogate safety measures, data collection techniques, and validation statistical models. In the end, research gaps and directions are discussed.

2.1 Traffic Conflicts

2.1.1 Safety Pyramid Continuum

Hydén introduced a well-known "safety pyramid": a hypothesized continuum for road-user interactions among different events, from normal situations to a fatal accident (see Figure 2-1) [2]. Hydén's safety pyramid serves as a seminal framework for understanding the hierarchy of traffic events, which illustrates the proportionality between non-conflict events and accidents. The pyramid's base represents frequent but less severe incidents — undisturbed traffic flows — while the apex shows rare and severe collisions, including fatalities. It indicates that as we ascend the levels from disturbances to actual crashes, the frequency of events diminishes, but the severity increases. This paradigm shifted the focus from reactive to proactive traffic safety management, aiming to intercept and mitigate risks at the lower levels before they escalate.

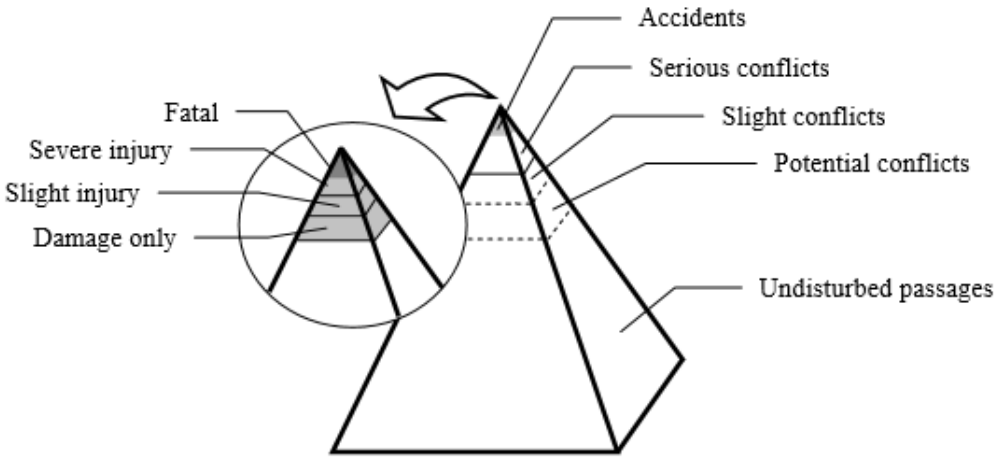


Figure 2-1 "Safety pyramid" (adopted from Hydén, 1987)

The safety pyramid has been revisited and reinterpreted through numerous research [6], [7], [8]. Researchers have debated the quantifiable thresholds among levels: inconsistency exists on the boundaries of the severity hierarchy models, and different groups of traffic events are not clearly identified [9]. Peesapati et al. suggested testing different thresholds and selecting the one that produced the best correlation between conflicts and crashes [10]. To eliminate the inconsistency, researchers incorporated all the traffic events in the safety pyramid and applied the shape of hierarchy models to evaluate traffic safety [3], [4], [11]. However, limited development of parametric models hinders the application of safety hierarchy [11].

Empirical studies have employed the safety pyramid not only to understand but also to anticipate the occurrence of severe traffic incidents [6], [7], [8], [12], [13], [14], [15]. Successive research has sought to enhance the pyramid's predictive capabilities by incorporating traffic volume and road design into its model [12]. Researchers have developed statistical models that correlate lower-level events with higher-level outcomes [8]. For example, some studies have drawn on the pyramid to validate the use of surrogate safety measures as predictors for actual crash likelihood. These studies vary in methodology, with some using regression analysis, while others employ

more contemporary extreme value theory and machine learning techniques to refine the prediction models [16].

2.1.2 Definition & Categorization

Tarko et al. defined: "*a conflict is an observable situation in which two or more road users approach each other in space and time with a risk of crash unless they change their movements*.[1]" Traffic conflicts have been categorized based on the identification methods: evasive action-based and temporal/spatial proximity-based [16]. Evasive action-based conflicts are identified through observable reactions such as braking or swerving, suggesting that a collision was imminent and was avoided through the driver's evasive maneuver [17]. This approach emphasizes the human response aspect of conflict situations. It proposes a counterfactual surrogate event of a crash: a crash would have occurred if an evasive action had not been performed (or not performed entirely). Temporal/spatial proximity-based conflicts, on the other hand, are determined by measurable criteria, such as time-to-collision (TTC) or post-encroachment time (PET), without necessarily observing an evasive action. This approach relies more on the objective assessment of the nearness of road users in time and space, which could lead to a potential collision if their courses remain unchanged.

Each method has its pros and cons [11], [16]. Evasive action-based approaches could be more intuitive and reflect real-world reactions to potential crash scenarios, but they are also more subjective and can suffer from observer bias. In contrast, proximity-based methods offer objectivity and the potential for automated detection through sensors and computer algorithms, yet they might overlook the complexity of human reaction in conflict situations.

The choice of approach to determine traffic conflicts can influence the data collection process [11], [16]. Evasive action-based methods require more field observations or video analysis, while proximity-based conflicts could be accessed by large-scale data collection through in-vehicle sensors and automated processing. Different conflict approaches also have a significant impact on the scope and direction of safety studies. The evasive action-based approach focuses on driver's reaction to avoid crashes, while proximity-based method can result in engineering solutions, such as redesigning the roadways and adopting intelligent transportation systems to monitor the distances between vehicles. Furthermore, the definition can affect the reliability and reproducibility of research findings. For example, proximity-based measures can offer consistent and repeatable data for statistical analysis at the cost of underreporting conflicts that do not meet the strict criteria. However, the missed-out information could be captured by the intuitive evasive action method.

2.1.3 Traffic Conflict Techniques (TCTs)

Traffic Conflict Techniques (TCTs) have been a cornerstone in the field of road safety research, offering a proactive approach to studying and preventing accidents. The origin of TCTs can be traced back to the groundbreaking work by Perkins and Harris at General Motors (GM) in 1967, who developed a method to observe and categorize evasive actions as indicators of potential crashes [18]. Their work laid the foundation for traffic conflicts to be recognized as surrogate safety measures. The initial method focused primarily on severe conflicts, defined as situations where collision was imminent without rapid evasive action. Their approach revolutionized traffic safety studies by highlighting the predictive value of near misses.

The TCT procedure developed by GM includes a formal definition of traffic conflicts, which takes place when "*a driver takes evasive action, brakes or weaves to avoid a collision*" [18]. 24 types of conflicts are introduced and observed in the field based on the definition. The in-field data collection lasts at least ten hours, requiring two observers collecting the 15-min intervals. This TCT is not a comprehensive model and only provides some recommendations on field data collection. Also, some rules to determine a traffic conflict is questionable. For example, the observer should distinguish a conflict based on the observed actions taken by the driver such as the activation of the brake lights, which doesn't necessarily indicate a conflict.

Based on GM methodology and its further research conducted by National Cooperative Highway Research Program (NCHRP), "*Traffic Conflict Techniques for Safety and Operations*" were published by the FHWA, which serves as a guide for transportation engineers to observe and analyze conflicts [17]. The book has two parts: the engineer's guide and the observer manual. The former one demonstrates the full steps of the TCT procedures, and the latter one is a reference for the observer to count conflicts in the field. A major contribution of this book is that it breaks down the occurrence of a conflict into four stages: *1) "One of the vehicles involved in the conflict makes a maneuver; e.g., starting a left turn movement from an exclusive left turn lane"; 2) "The driver of the other vehicle involved senses a danger of collision as a result of a judgement call based on the prevailing conditions"; 3) "The driver of the vehicle sensing danger reacts by taking an evasive action such as weaving or swerving", and; 4) "The other vehicle proceeds through the initiated maneuver without making any corrections on its trajectory".* TCTs have been implemented by several countries over a long time; different techniques adopt different approaches to identify conflicts.

With technological improvements, TCTs have developed from manual observation to state-of-the-art automatic systems [16]. Automated TCTs use various sensors and video analysis tool to identify conflicts, offering a consistent and objective method to detect conflict. Technologies together with radar, lidar, and computer vision algorithms are now integrated into modern TCT applications, enabling more comprehensive studies through different traffic conditions and environments [16]. The automation of conflict detection has additionally facilitated the implementation of real-time traffic safety management [19].

2.2 Surrogate Safety Measures

Surrogate safety measures are the measurement of spatial/temporal proximity to a crash. To apply a surrogate safety measure, Tarko et.al proposed two must-be-met conditions: 1) "*a surrogate safety measure should be based on an observable non-crash event that is physically related to crashes"; 2) "there should be a practical method for converting the non-crash events into corresponding crash frequency and possibly crash severity*" [1]. Based on the two conditions, this section covers collision course, surrogate safety indicators, and related analysis methods.

2.2.1 Collision Course

The earlier definitions of collision course involve traveling on "*present course and at present rates*" [2]. However, it's unlikely to happen strictly with the same speed and direction. The collision course is now defined as "*a situation in which there is at least one (physically) possible trajectory per road user that could lead to their collision at a future instant*." Mohamed & Saunier proposed two approaches for predicting road user trajectories to estimate the collision course [20].

The first approach is called a context-free kinematic method, which assumes a constant velocity, constant angular velocity, or constant acceleration to predict vehicle motion. The constant velocity assumption seems only appropriate for simple conditions such as car-following [21]. As a supplement to this method, Mohamed & Saunier added random noise as the small adjustments made by drivers when they travel straight [20].

The second approach is based on the observed movements of road users rather than the assumed one. Laureshyn et al. recommended the actual trajectories as the substitute for the initially planned trajectories [4]. However, this method becomes problematic when the drivers change their path to avoid the danger. Empirical methods have been applied to predict form of motion patterns using road user trajectories at a given road site [5], [22], [23].

2.2.2 Types of Surrogate Safety Indicators

Surrogate safety indicators can be categorized as: temporal proximal indicators, non-temporal indicators, other indicators, and a composite index [16]. In particular, temporal proximal indicators include time-to-collision family and post-encroachment time family.

2.2.2.1 Temporal Proximal Indicators

Time-to-Collision (TTC) and Post-Encroachment Time (PET) are pivotal metrics in the domain of surrogate safety analysis, primarily because they offer quantifiable and objective measures of traffic interaction.

TTC is defined as the time remaining until a collision would occur if the speed and trajectory of two vehicles (or a vehicle and any other object in its path) remain constant [24]. It's a dynamic measure reflecting the urgency of a situation and is especially useful for evaluating rear-end collision potential. TTC can be calculated using in-road sensors, vehicle-mounted systems, or by

analyzing traffic surveillance footage. It is valued for its predictive capability in identifying high-risk scenarios, especially in high-speed traffic environments.

Variations of TTC have been proposed to deal with the limitations and enhance their utility [25], [26]. Modified TTC (MTTC) includes acceleration or deceleration into the calculation, which represents a realistic situation for driver to avoid a crash [25]. This may be specifically applicable in stop-and-go situation in a congested traffic where speeds aren't constant. Time Exposed Time-to-Collision (TET) and Time Integrated Time-to-Collision (TIT) are extensions of the TTC concept, in which the exposure time or cumulative time below the TTC thresholds is considered [26]. These measures consider not only the closest measurement but also the duration over which a collision is imminent, which is more comprehensive.

PET measures the time difference between the moments one road user (vehicle, cyclist, or pedestrian) vacates a space and the next user enters it (Figure 2-2) [27]. Unlike TTC, PET is not restricted to headway following situations and is particularly useful for assessing conflicts at intersections and crosswalks. It provides insights into the interaction between crossing paths, making it an essential measure for evaluating safety in multi-directional traffic flows.

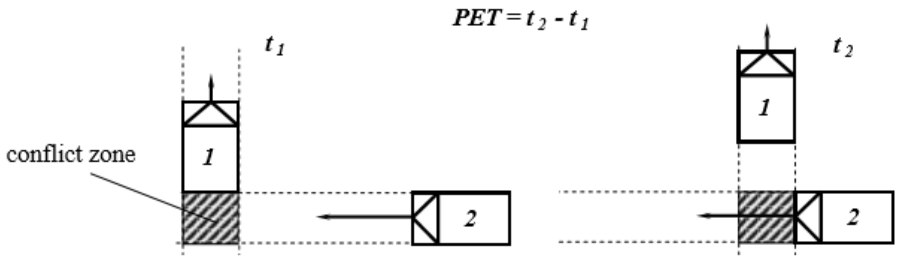


Figure 2-2 Demonstration of PET

Besides PET, post-encroachment time family include but not limited to Gap Time (GT), Encroachment Time (ET), and Time Advantage (TAdv) [4]. GT is "*the time between the entries into the conflict spot of two vehicles*", and ET is "*the time that the first vehicle entering the*

conflict area infringes upon the predicted path of the second vehicle" [4]. Unlike PET, both GT and ET are continuous indicators, the difference between the two is that GT is measured from the front bumper to the front bumper while ET is measured from the rear bumper to the front bumper [4]. TAdv is a predicted PET value which assumes the same course and speed. Similar to GT and ET, TAdv can capture continuous interactions between vehicles [4]. By adjusting these measures to account for dynamic traffic behaviors, researchers and traffic engineers can create more sophisticated models to predict potential conflict points. Mohamed & Saunier proposed a variation of TAdv within the probabilistic framework considering multiple trajectories [20].

2.2.2.2 Non-temporal Indicators

Deceleration-related and distance-related are two types of non-temporal safety indicators. Deceleration Rate to Avoid a Crash (DRAC) was the most commonly used deceleration-related surrogate safety measures, which is defined as the minimum deceleration rate required by the following vehicle to avoid a collision assuming an unchanged speed of the front vehicle. Kuang et al. proposed a Modified DRAC (MDRAC) considering the driver's perception and reaction time [28]. Cunto and Saccomanno developed a Crash Potential Index (CPI) to estimate "*the probability that a given vehicle's DRAC exceeds its Maximum Available Deceleration Rate (MADR) during a given time interval*" [29]. Jerk is the derivative of acceleration with respect to time and captures the "suddenness" of braking. Proportion of Stopping Distance (PSD) is one typical distance-related indicator, which is defined as the ratio between the current distance and the minimally acceptable stopping distance[28]. These non-temporal measures are specifically beneficial in safety evaluation due to the fact they may be indicative of conflict severity and the drivers' subjective assessment of traffic hazard. For instance, excessive deceleration rates and

jerk (the derivative of acceleration with respect to time) values can also correspond to more extreme conflicts, in which drivers perceive a high risk and react more forcefully.

Table 2-1 summarizes the key elements for major surrogate safety indicators [30].

Table 2-1 Key Elements for Major Surrogate Safety Indicators

Indicator	Limitations	Advantages	Collision Type
TTC	Keep constant speeds; Ignore potential conflicts; Collision course; Exist when the leading vehicle is faster	More informative; Vehicle ITS use TTC as an important warning criterion; Applicable for Work Zone safety analysis; Applicable in post-processor such as SSAM	Rear-end, Turning/weaving, Hit objects/parked vehicle, crossing and hit pedestrian
TA	Relying heavily on the subjective judgment of speed and distance; Mainly rely on the evasive action; Other same as TTC	Easy to measure; Manually or by Video analysis; Several manuals have been developed in different countries	Same as TTC
TET	Does not provide the various severity levels of different TTC values below the threshold value; If TTC-value is lower than Highly data-intensive and attainable only in a simulation environment	Suited for application in microscopic simulation studies of traffic	Same as TTC
TIT	Difficult to interpret its meaning for complexity to determine; Not preferable to use in comparative studies in which simulation tools are applied to generate trajectories; Benefits is small due to the uncertainties in driver behavior	Suitable for microscopic simulation studies of traffic	Same as TTC
MTTC	Obtaining the field speed of both users and the distance gap in an evolution process is difficult and has to rely on other approaches; Not fit for lane changing or a head-on collision	More advanced than TTC; Consider driving discrepancies; Severity of the collision could be weighted using CI indicators	Vehicle-vehicle crash, Same as TTC
PET	Only useful in the case of transversal (i.e., crossing) trajectories (Right angle collision); Cannot reflect changes with the dynamics of safety-critical events over	More appropriate than TTC for intersecting conflicts; Can be easily extracted;	Mainly for right angle or crossing crash, Hit pedestrian. Merging/diverging,

	a larger area; Levels of severity as well as the impact of a conflict are not taken into account	Can be easily estimated using photometric analysis in video or simulated environment; Represents the driver behaviors.	head on (to a certain extent)
DRAC	Fails to accurately identify the potential traffic conflict situation; Not suitable for lateral movement	Explicitly considers the role of differential speeds and decelerations in traffic flow	Rear-end, Hit object/parked vehicle, Hit pedestrian, Merging and diverging maneuvers
CPI	Not suitable for lateral movement. Mainly applicable at intersection	Address some of the issues found in DRAC like vehicle braking capability for prevailing road and traffic conditions	Same as DRAC
PSD	Based on evasive actions; PSD provide a higher percentage of vehicles interaction and time exposure to conflict than TTC and DRAC	Single vehicle conflict with fixed or unfixed objects can be evaluated; Easy for observation and calculation	Hit object (on road or roadside), Overturning

More non-temporal measures have been proposed recently. To capture the crash severity, Delta-V was adopted, which is defined as the change of a vehicle' velocity vector during a collision. Researches have verified the assumption that accidents involving higher Delta-V values result in higher injuries [31]. Shelby adopted expected Delta-V, what Delta-V would have been if the road users had collided at their current speeds, to capture conflict severity [32]. Laureshyn et al. introduced Extended Delta-V indicator by considering the time remaining to take an evasive action [33]. The authors recommended the indicator because it represents the severity dimension by combining the collision proximity and its potential consequences [33]. Kuang et al. developed an Aggregated Crash Index (ACI) by accumulating the collision probabilities of all possible outcomes based on four conditions in the car-following scenarios [34].

2.2.2.3 Integration of Different Indicators to a Composite Index

To provide a more comprehensive evaluation of road safety, researchers have tried to combine surrogate safety measures from diverse dimensions such as temporal, spatial, and behavioral factors into one composite index [35]. The integration is based on the premise that no single

indicator can capture the whole picture of the traffic safety risks. One example is the development of the Surrogate Safety Assessment Model (SSAM), which mixes signs like TTC and PET to assess the safety overall performance of intersections [36]. Another instance is the usage of composite measures that consist of required braking rate, maximum available braking rate, and TTC to create an Aggregate Crash Propensity Metric [37].

A composite index can offer a comprehensive evaluation of road safety by captures various aspects of traffic interactions, which will provide a stronger predictive capability of potential crash [35]. However, there also are drawbacks and demanding situations related to composite indices [35], [36], [37]. First, developing a composite index that correctly reflects road safety includes complicated modeling and might require advanced statistical or machine learning techniques. Second, it may require more comprehensive data. Third, it’s a challenge to interpret a composite index. Last but not least, the subjectivity involved in choosing which indicators to include and how to weight them can introduce bias and affect the index's reliability. The interaction effects between different indicators within a composite index also need to be understood and accounted for to ensure the index is meaningful.

2.2.3 Analysis Methods using Surrogate Measures

While it's relatively straightforward to generate surrogate safety measures, interpretation is necessary to perform safety analysis. The core approach of the TCT is to count the number of most severe traffic events. The approach applies a trigger threshold for one or more objectively measured safety indicators. The literature frequently recommends a value of 1.5 seconds on temporal indicators, in particular, TTC [33]. However, instead of categorizing traffic events into either dangerous or not dangerous according to a threshold value, another approach assumes that all indicator values produce different degrees of risk throughout the safety continuum [3], [38],

[39]. Saunier and Mohamed suggested examining the entire TTC curve and using the curve shape to build clusters and distinguish safety-critical events [40]. Zheng used all values of the surrogate measure to build a parametric safety continuum model [8].

2.3 Data Collection Methods

Multiple methodologies and techniques have been proposed and applied to collect data for surrogate safety measure [11], [16], [41]. Traditionally, surrogate safety measures such as traffic conflicts are collected by field observations. The limitations of this method include time-consuming, labor-intensive, and sometimes inaccuracy. Several approaches have been utilized to overcome these deficiencies (e.g., automated techniques and traffic simulations). Furthermore, emerging data such as in-vehicle data from naturalistic driving studies and Safety Pilot Model Deployment (SPMD) study contribute to the development of surrogate safety measures. This section will discuss the four types of data collections methods: field observation, computer vision techniques, in-vehicle data from naturalistic driving study, and traffic microsimulation models.

2.3.1 Field Observations

Field observation techniques are the basic data collection method in traffic safety research [11], [16], [41]. They involve recording traffic interactions either in person or via video recording devices. Specific strategies consist of tally sheets for counting, specified forms for characterizing interactions, and video analysis protocols which could comprise direct or automated coding of behaviors.

The wealthy qualitative data derived from field observations can offer insights into the nuanced behaviors that precede traffic incidents, together with near-misses or aggressive driving maneuvers [11], [16], [41]. Researchers have developed diverse methodologies, each with

standardized protocols to make sure consistency in data collection. Manual methods for field observations have been adapted across different countries to fit local traffic conditions and research objectives.

There are some challenges for field observations [11], [16], [41]. Observer fatigue can cause inconsistencies in data collection, and the subjectivity can introduce bias. Furthermore, the scalability of the data collection is often limited. The reliability of field observations can also be affected by extreme environmental situations, which might inhibit the observer's ability to correctly see and interpret traffic events. To mitigate those demanding situations, a few research hire multiple observers for cross-validation, use established observation protocols, and integrate training session to ensure the consistency.

2.3.2 Computer Vision Techniques

The use of computer vision techniques to collect traffic conflict data relies on computer algorithms to track moving objects and detect traffic conflicts from videos (shown in Figure 2-3) [38], [42], [43], [44]. The conflict detection system includes two parts: 1) a video-processing module to detect and track objects; 2) an interpretation module to extract information and detect traffic conflicts [23], [38]. Despite video cameras are relatively inexpensive and are widely deployed, two issues should be addressed before large-scale analysis [11], [16], [41]. First, the video-processing module needs to be improved to accurately track and classify different types of objects. Secondly, the interpretation module requires more sophisticated conflict framework [43].

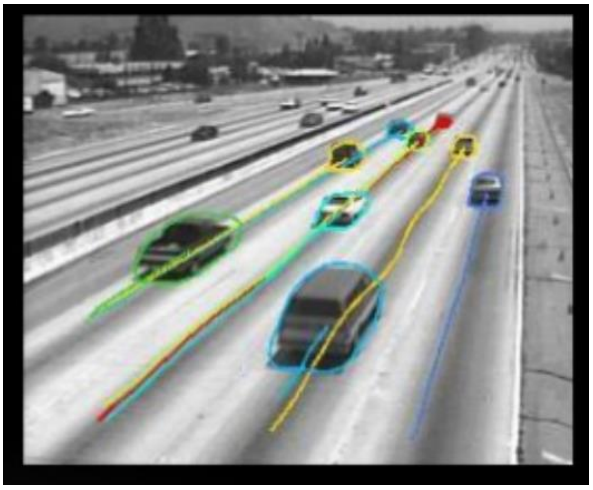


Figure 2-3 Computer Vision
(Source:https://www.slideshare.net/mohamedrajah/computer-vision-11687562)

The fixed video is limited to providing descriptive data on drivers (e.g., age, gender, alcohol levels), detailed descriptions of safety equipment in the car (e.g., seat belt use, airbag provision), and all other in-vehicle driving. However, research shows potential for integrating the fixed video data and crash data when developing the crash continuum pyramid [45].

2.3.3 In-vehicle Data from Naturalistic Driving Study (NDS)

Naturalistic driving studies (NDS) collect in-vehicle data under a real-world environment to understand driving behavior over an extended period [13]. Vehicles are equipped with an array of sensors, cameras, and data recording devices to capture driver behavior, vehicle movements, and environmental situations [13]. This information is then analyzed to identify regular driving patterns, in addition to abnormal or risky behaviors that could cause conflicts or crashes [13].

Several obstacles arise to collect and use the in-vehicle data [11], [13], [14], [15], [16], [41], [46], [47], [48]. Technical challenges in NDS include data management and analysis, given the large-scale data collected [11], [16], [41]. Ensuring the reliability and calibration of in-vehicle sensors over time is critical, as is developing robust algorithms to detect incidents of interest

(e.g., conflict) from the big data. Sampling issues arise in the selection of drivers, vehicles, routes, driver conditions, etc. The study findings are meaningful to general audiences only when the participants are representative. In addition, privacy concerns are paramount in NDS due to the personal nature of the collected data.

In-vehicle data collection continues to evolve, with connected vehicle technologies offering the potential to collect data more seamlessly and across a broader range of vehicles [11], [16], [41]. However, addressing these technical, sampling, and privacy challenges remains critical to the integrity and acceptance of NDS findings.

Future development in data collection technology is likely to provide a considerable amount of data which is very difficult to analyze manually [11], [13], [14], [15], [16], [41]. Machine learning or artificial intelligence techniques can be effective to collect surrogate safety measures. Techniques such as neural networks and fuzzy logic utilize a computer system to take drivers' behavior as input and relate it to the output variables - conflicts and crashes. The collection of in-vehicle data on driver behavior offers an opportunity to use machine learning models to look at car-following and evasive behavior in greater detail [13]. Several research groups have used the naturalistic dataset to develop machine learning algorithms in conflict situations [14], [15]. Chong et al. used the naturalistic dataset collected at Virginia Tech to explore car-following and safety-critical events [49]. The authors developed a fuzzy logic model of car-following with the state layers of relative distance, relative speed, and vehicle speed [49].

2.3.4 Traffic Microsimulation Models

Microsimulation models are effective tools in the traffic safety analysis, with vehicle characteristics, driver behavior patterns and road geometry as inputs, to simulate the complex

dynamics of traffic flow [11], [16], [25], [29], [37], [41], [45], [50], [51], [52]. Regarding surrogate safety measures, the microsimulation model enables proactive safety analysis in the absence of real-worlding testing. It can evaluate the potential safety impact of road design changes, traffic control measures, or the introduction of new vehicle technologies like advanced driver-assistance systems (ADAS). For example, the models can simulate the frequency of hard stops, close following, and different precursor events to collisions, thereby producing surrogate measures such as TTC and PET.

Although traffic microsimulation models provide a controlled environment to test a range of traffic scenarios, some challenges exist [11], [16], [41], [51]. One significant challenge is to ensure the simulated driver behaviors accurately reflect real-world situation, requiring extensive data on real-world's driver behavior to calibrate the simulation models. In addition, powerful computing resources are required to simulate over extensive road networks for long time periods, which is not available in research settings.

2.4 Validation Statistical Methods

The mathematical and statistical methods introduced in the section illustrate the approaches to validating surrogate safety measures. Besides the validation purpose, the statistical models can link the normal traffic events to serious traffic events, providing robust traffic safety measurement when serious events are not sufficient. This section discusses four widely used validation models: traditional regression model, two-phase model, extreme value theory model and casual model. The first two are crash-based models, while the rest two are non-crash-based models.

2.4.1 Traditional Regression Model

Regression techniques can provide the relationship between the traffic conflict counts and crash counts directly. In this situation, traffic conflicts are considered as crash opportunities, which may lead to crashes. Hauer and Garder proposed the relationship in a form as Equation 2-1 [53].

$$\lambda = \sum_i \pi_i c_i \qquad \text{Equation 2-1}$$

where, c_i is the number of observed conflicts of severity level i, and π_i is the crash-to-conflict ratio for conflicts of severity level i.

The authors mentioned that the crash-to-conflict ratio varies by locations and times. Following the relationship, various regression techniques can be used to estimate the ratio [54], [55]. However, it's difficult to ensure a stable crash-to-conflict ratio, especially for varied severity levels [11].

2.4.2 Two-phase Model

As an extension of the traditional regression models, the two-phase model uses estimated conflicts to predict crashes by predicting conflicts from road-related , traffic-related, environment-related factors in the first phase and predicting crashes based on the conflicts in the second phase [12]. Lognormal model and negative binomial model were used for the two phases. The stable relationship between conflicts and crashes in the second phase indicates that identifying the contributing factors for conflicts helps to reduce crashes. However, the influence of various road-related, traffic-related, and environment-related factors on conflicts is complex in the first stage [11]. Furthermore, the model also assumes that the two phases are independent, which may not always be the case in real-world scenarios.

2.4.3 Extreme Value Theory Model

The use of Extreme Value Theory (EVT) in traffic safety research, especially for assessing the relationship between traffic conflict and crash, has been substantially developed in recent years [7], [11], [16], [41], [56], [57]. The focus has been on developing models to extrapolate and estimate the crash risk from the detected traffic conflicts without relying on the crash data. The EVT model has been applied across various road facilities, including freeways, rural roads, signalized intersections, and roundabouts, where signalized intersections are mostly investigated due to the high conflict rate [19], [58], [59], [60]. Moreover, EVT models has been applied to specific crash types, such as rear-end, sideswipe, and lane-changing crashes at signalized intersections [56].

However, there are challenges and related research directions regarding the EVT model [56]. The EVT model needs to be refined to better capture traffic safety dynamics and provide more accurate risk estimate. Critical research areas include but not limited to optimizing the techniques for sampling extremes, determining the optimal sample size or threshold, and selecting the most appropriate traffic conflict measures [7], [11], [16], [41], [56], [57]. Additionally, incorporating covariates and accounting for unobserved heterogeneity are vital to enhancing model robustness [56].

In recent studies, real-time analysis using EVT models becomes an area for improvement [56], [61]. Future studies should focus on more dynamic and real-time focused methods, such as Bayesian dynamic extreme value modelling, to provide timely safety assessments. Moreover, EVT models can provide valuable insights into the safety improvements for connected and autonomous vehicles [56].

2.4.4 Causal Model

Causal models in traffic safety have been applied to understand the dynamics of traffic events and estimate traffic safety. The models work on the premise that traffic conflicts are counterfactually related to crashes, and build relationships between initial conditions, evasive actions, and the probabilities of various outcomes, including crashes. For example, Davis et al. calculates the chance of a crash given preliminary situations and evasive actions by aggregating all the probabilities [62].

Recent studies applied causal models to evaluate traffic safety by including various factors that influence the risk of traffic events. Tarko developed a framework that integrated the delay in a road user's response as a significant factor causing crash following a conflict [63]. Causal models have also been adopted to determine the relationship between crashes and other potential safety measures such as hard braking, congestion, and speed variation, with the incorporation of spatial correlation found to improve model fit significantly. The causal models also can predict real-time crash risk using the proposed Crash Potential Index (CPI), which can be estimated from simulated vehicle trajectory data [64].

Developing the causal models requires a complete observation of traffic events and deep understanding of the mechanisms behind evasive actions, which is a challenge due to the lack of complete observation and understanding of evasive action process [41]. Although the obvious challenge, the continued development and application of the causal models make it a promising area in the traffic safety research, especially with real-time prediction on the connected and autonomous vehicles [41].

2.5 Research Gaps & Directions

A comprehensive review of the literature shows that while there has been a significant improvement in traffic conflict research, gaps nonetheless exist. One obvious gap is the lack of a universally acknowledged definition of what constitutes a traffic conflict, leading to variations in study outcomes and challenges in comparing them. Additionally, most research has targeted on motorized traffic with less attention on vulnerable road users such as pedestrians and cyclists.

Another gap is the integration of human factors into conflict detection and safety evaluation. While automated systems are adept at recording measurable conflict indicators, they often do not account for the psychological or behavioral elements that precede a conflict situation, such as driver distraction or aggression. Research could explore the incorporation of human factor analysis into safety evaluation using conflicts to create a more comprehensive picture of potential crash risk.

Furthermore, research could explore the potential of combining traffic conflict data with emerging sources, such as data from connected vehicles and smart infrastructure, to improve the resolution and predictive power of conflict analysis.

Advancing the analysis methods to include more sophisticated statistical models or machine learning algorithms could address the complexities of traffic conflict interactions, enabling the development of predictive models that could forecast conflicts under a variety of scenarios.

In the dissertation, the author aims to shorten the research gap by 1) exploration of developing surrogate safety measures using emerging connected vehicle data source; 2) comparing and validating different surrogate safety measures using the crash data; 3) applying extreme value

theory method and incorporating drivers' behavior into the safety evaluation; 4) building safe routing application.

Chapter 3 Study Design

This chapter provides a research framework that covers proposed tasks for addressing critical challenges summarized in Chapter 2. The research efforts in this dissertation are divided into the following tasks:

Task 1: Present a comprehensive review of surrogate safety measures. This section presented an extensive overview of surrogate safety measures and identify key challenges and opportunities in the field. TCTs, surrogate safety measures, data collection methods and validation techniques were included (Chapter 2).

Task 2: Apply emerging data sources Safety Pilot Model Deployment (SPMD) to generate surrogate measures. Comprehensive data processing algorithms were developed. Also, different surrogate safety measures were compared. It includes two studies in Chapter 4.

Task 3: Evaluate the road safety for mid-block rear-end crashes using SSMs (Chapter 5).

Task 4: Evaluate the work-zone road safety using SSMs (Chapter 5).

Task 5: Demonstrate the routing application by incorporating road traffic safety (Chapter 6).

Chapter 4 Calculation & Validation of Surrogate Safety Measures

The chapter used the emerging Safety Pilot Model Deployment (SPMD) data to calculate and validate surrogate safety measures. First, the analysis was conducted on several SSMs for rear-end crashes in the mid-block. Then the study scope was expanded to major crashes in both the link and intersection.

4.1 Calculation and Validation of SSMs for Rear-end Crashes in the Mid-block

The goal of this study was to assess roadway link-level surrogate safety measures using the vehicle trajectory data from SPMD. The study's objectives included: 1) developing a framework to process the SPMD dataset using Big Data Analytics; 2) converting raw vehicle motion data from SPMD to surrogate safety measures; and 3) analyzing the statistical relationship between crash records and calculated safety index to compare different SSMs.

4.1.1 Data Collection & Processing

In this section, SPMD vehicle trajectory data, roadway characteristics & traffic data, and crash data were used. The following part discusses the details of the data and data processing steps.

SPMD vehicle trajectory data

SPMD is collected from vehicles equipped with vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication devices under the real-world conditions in Ann Arbor, Michigan [65]. The deployment included approximately 3,000 pieces of onboard vehicle equipment and 30 pieces of roadside equipment in an area of over 73 lane-miles. SPMD data includes driving data (DAS data), BSM data, roadside equipment data, weather data, and network traffic volume data [65]. Driving data provides a vehicle's kinematic and geographic information, and BSM data contains a vehicle's position and motion with the status of

components such as brakes, turning lights, wipers, etc. Compared to video image based data in the NDS, the data such as DAS and BSM from SPMD are relatively easy to process. Moreover, SPMD data can provide real-time safety evaluation in a connected vehicle environment.

The SPMD data contains eight datasets: Data Acquisition System 1 (DAS1), Data Acquisition System 2 (DAS2), Basic Safety Message (BSM), Roadside Equipment (RSE), Network, Weather, Schedule, and RoadWork Activity. Driving data (DAS1 and DAS2) provides a vehicle's kinematic and geographic information, and BSM data contains a vehicle's position and motion with the status of components such as brakes and turning lights. The DAS1 dataset (DataWsu data and DataFrontTargets data) and Roadside Equipment dataset (SPaT data) were used in this study. The DAS1 dataset was collected from around 100 vehicles for October 2012 and April 2013 by the University of Michigan Transportation Research Institute (UMTRI). DataWsu contains subject vehicles' trajectory information (i.e., GPS coordinates) and other related components such as speed, acceleration, and brake status. DataFrontTargets collects front vehicle information (e.g., relative distance and relative speed) with the help of Mobileye's vision-based Advanced Driver Assistance Systems (http://www.mobileye.com/).

Around 2,800 vehicles participated in the SPMD program, including 2,450 equipped with Vehicle Awareness Device (VAD), 300 with Aftermarket Safety Devices (ASD), 19 with Retrofit Safety Devices (RSD), and 67 with Integrated Safety Devices (IDS). Among the vehicles with VAD, 2,350 are cars, 60 are trucks and 85 are transit buses. Vehicles with VAD can only transmit BSM. ASD can transmit and receive BSM but is only installed on cars. Similar to ASD but designed for freight and transit, RSD is installed on 16 trucks and 3 transit buses. Compared to ASD or RSD, IDS can not only send and receive BSM but also connect to the vehicle data processing system. Generally, drivers who reported the most driving within the

SPMD area were selected to the SPMD program. For the 64 cars with IDS, participants were selected based on age and gender to ensure a similar number of drivers in each group.

Two months of the SPMD dataset (Oct. 2012 and Apr. 2013) are free to download in the transportation data sharing platform, or Research Data Exchange (RDE). (https://www.its-rde.net/index.php/rdedataenvironment/10018#). This study used the DataWsu file (12 GB) and the DataFrontTargets file (4.34 GB) in the driving dataset DAS1 collected from around 100 equipped vehicles. In this study, common data fields "Device", "Trip" and "Time" were used to link the two datasets. Table 4-1 contains the primary data elements in the DataWsu file and the DataFrontTargets file, along with a brief description of each.

Table 4-1 Major Data Elements in the *DataWsu* File and *DataFrontTargets* File

Field Name	Type	Units	Description
DataWsu			
Device	Integer	none	A unique numeric ID assigned to each DAS. This ID also doubles as a vehicle's ID
Trip	Integer	none	Count of ignition cycles—each ignition cycle commences when the ignition is in the on position and ends when it is in the off position
Time	Integer	centiseconds	Time in centiseconds since DAS started, which (generally) starts when the ignition is in the on position
GpsValidWsu	Integer	none	Communicates whether a GPS data point is valid or not
GpsTimeWsu	Integer	millisecond	Epoch GPS time received from the remote vehicle that has been targeted by the host vehicle's WSU
LatitudeWsu	Float	deg	Latitude from WSU receiver
LongitudeWsu	Float	deg	Longitude from WSU receiver
AltitudeWsu	Real	m	Altitude from WSU receiver
GpsHeadingWsu	Real	deg	Heading from WSU GPS receiver
GpsSpeedWsu	Real	m/sec	Speed from WSU GPS receiver
SpeedWsu	Real	kph	Speed from vehicle CAN Bus via WSU
TurnSngRWsu	Integer	none	Right turn signal from vehicle CAN Bus via WSU
TurnSngLWsu	Integer	none	Left turn signal from vehicle CAN Bus via WSU
BrakeAbsTcsWsu	Integer	none	Brake, ABS, and traction control from vehicle CAN Bus via WSU
AxWsu	Real	m/sec^2	Longitudinal acceleration from vehicle CAN Bus via WSU
PrndlWsu	Integer	none	Current transmission state (Park, Reverse, Neutral, Drive, Low) from vehicle CAN Bus via WSU
HeadlampWsu	Integer	none	Headlamp state from vehicle CAN Bus via WSU

WiperWsu	Integer	none	Wiper state from vehicle CAN Bus via WSU
ThrottleWsu	Real	none	Throttle position from vehicle CAN Bus via WSU
YawRateWsu	Real	deg/sec	Yaw rate from vehicle CAN Bus via WSU
SteerWsu	Real	deg	Steering angle/position from vehicle CAN Bus via WSU
DataFrontTargets			
TargetId	Integer	none	Numeric ID assigned by the Mobileye sensor to distinguish between the different objects being tracked; the closest obstacle is given a TargetId value of 1
ObstacleId	Integer	none	ID of new obstacle, as assigned by the Mobileye sensor, and its value will be the last used free ID
Range	Integer	m	Longitudinal position of an object, typically the closest object, relative to a reference point on the host vehicle, according to the Mobileye sensor
RangeRate	Real	m/sec	Longitudinal velocity of an object, typically the closest object, relative to the host vehicle, according to the Mobileye sensor
Transversal	Real	m	The lateral position of the obstacle, as determined by the Mobileye sensor
TargetType	Integer	none	Classification of an identified obstacle/target as a car, truck, pedestrian, etc.
Status	Integer	none	Classification of the motion (kinematic state) of an identified obstacle/target as stopped, moving, etc.
CIPV	Integer	none	Field communicating whether an obstacle is the closest in a vehicle's path

Roadway Characteristics & Traffic Data

The data on Michigan Traffic Volume for the year 2013, specifically for state highways designated as trunklines, is accessible through the Government's Open Data Portal (http://maps-semcog.opendata.arcgis.com/), where the Annual Average Daily Traffic (AADT) for these roads can be found. Additionally, the Lane Mile Inventory for these state highways provides insights into the number of lanes, their functional class, and speed limits. However, comprehensive speed limit data across the entire road network is lacking. To address this gap, speed limits have been estimated based on the road's functional classification: primary roads are estimated at 55 mph, secondary roads at 35 mph, and local or city roads at 25 mph. This functional class information for roads is available from the TIGER/Line Shapefiles (https://www.census.gov/geo/maps-data/data/tiger-line.html).

Crash data

Road network data and crash data (Figure 4-1) for Washtenaw County were collected from the Southeast Michigan Council of Governments Open Data Portal (http://maps-semcog.opendata.arcgis.com/). 52,386 crashes occurred in the County from 2011 to 2015, including 17103 rear-end crashes.

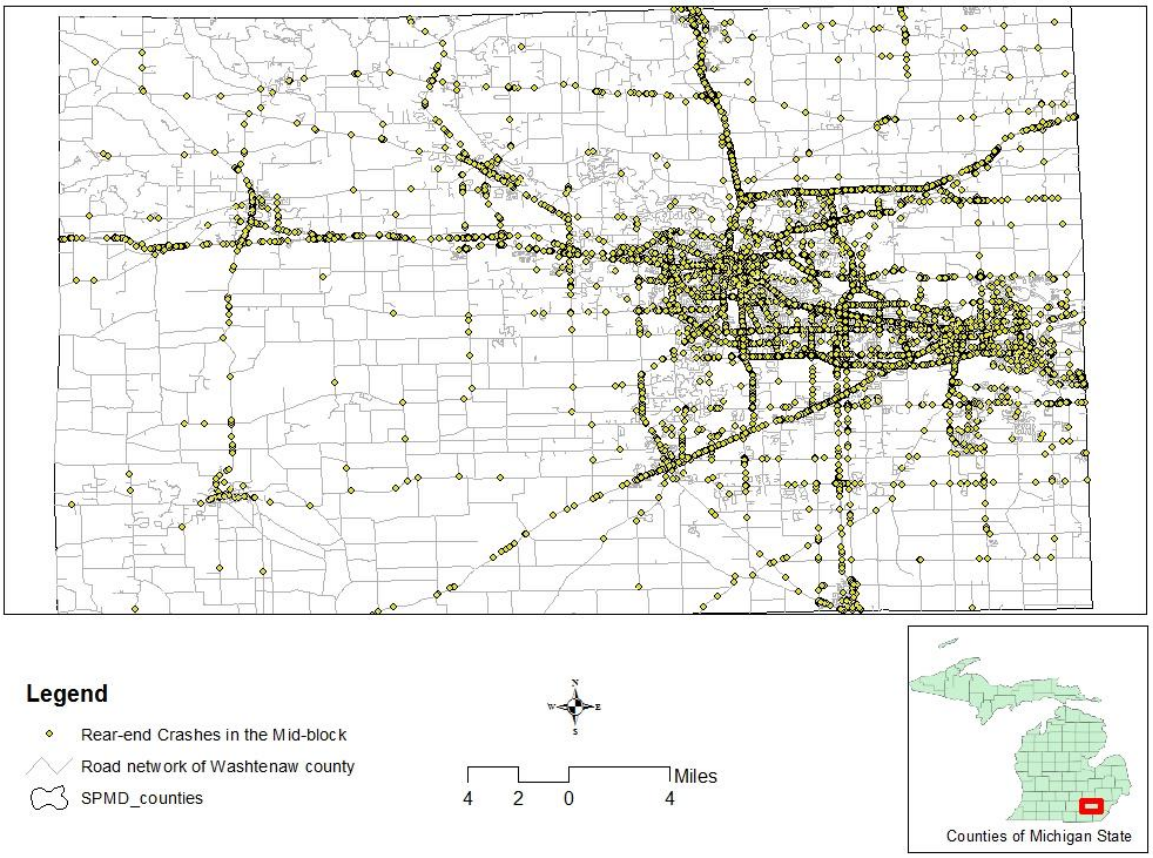


Figure 4-1 Road network & crash points.

After collecting the data, a series of steps were undertaken to process it. For the SPMD data, the DataWsu data and DataFrontTargets data were used in this study. Initially, each dataset was examined using Python programming language to check the data type and data organization.

The datasets were then imported into Hadoop to conduct a query using Apache Hive, a query language that is built on top of Apache Hadoop. Unfortunately, because of the complicated relationship among columns, the aggregation of two big datasets in Hadoop was extremely slow. Moreover, current Hadoop-GIS does not support advanced spatial analysis such as spatial join, so the two datasets were exported into small files (records with the same device are in one small file) to join them in the PostgreSQL.

The combined data were then imported into ArcGIS to integrate link and intersection information. The combined data were imported back into the PostgreSQL database, and 75 feet buffer zones were created to remove points around the intersections. Finally, a combined dataset was generated for the target type of "car" or "truck." Figure 4-2 shows the process.

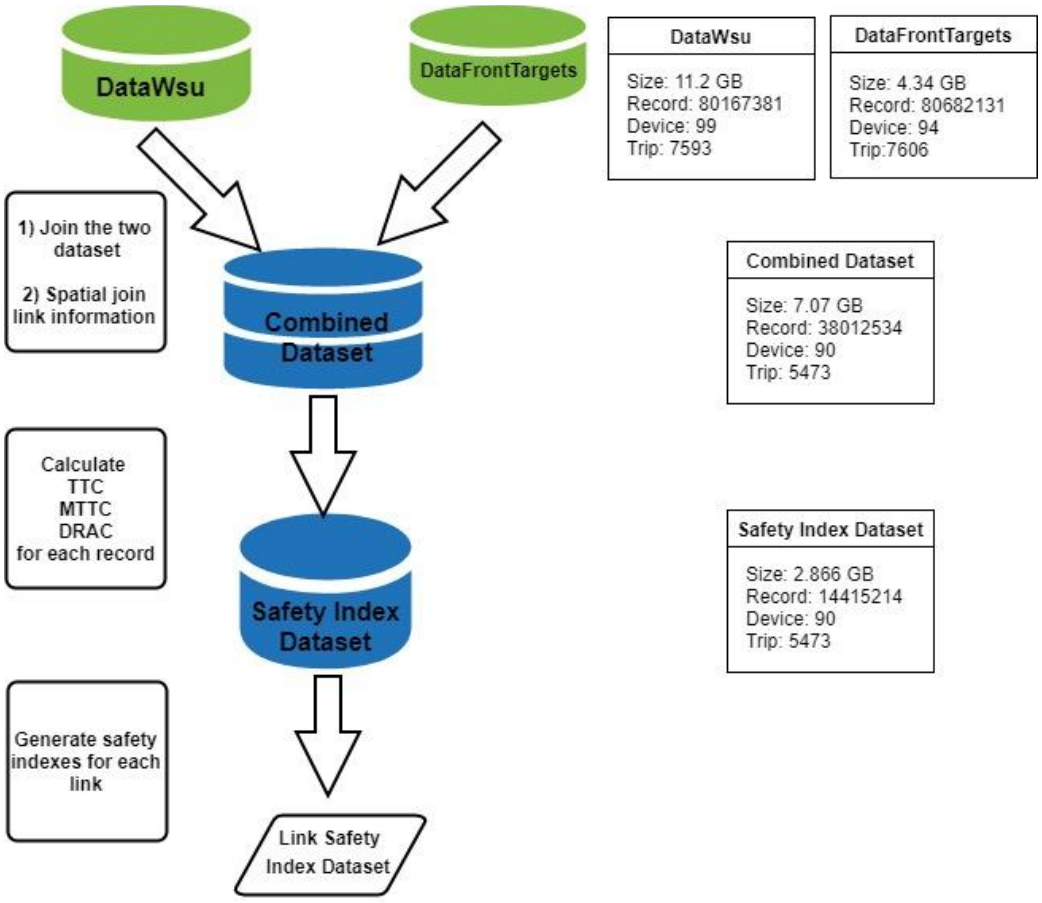


Figure 4-2 Data processing framework.

For the roadway data, roadway attribute information for each segment and intersection was organized in ArcGIS. Then the roadway information for state highway was spatial joined to the roadway shapefile.

For the crash data, according to Michigan's definition of intersection-related crashes, the 75-ft radius was used to remove rear-end crashes at the intersection, and the rest were kept to analyze mid-block crashes.

4.1.2 Methodology: Algorithms to Calculate the SSMs

Vehicle-level safety indexes were calculated within the "range" based on the combined dataset with driving information and front vehicle information (i.e., distance to the front vehicle is less than 250 meters and the lateral distances is less than 3 meters). Then, safety surrogate measures were calculated for each link. Rear-end crashes in the mid-block and the links with safety surrogate measures, are shown in Figure 4-3.

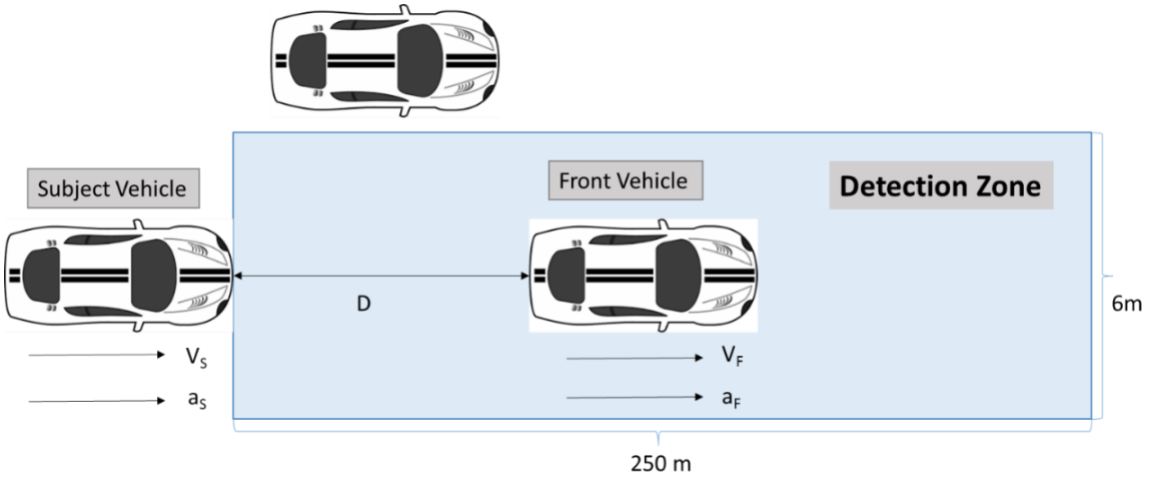


Figure 4-3 Illustration of vehicle motion in SPMD dataset.

Vehicle-level Safety Surrogate Measures

Safety surrogate measures are usually measured at each timestamp between vehicles that are interacting with each other. In Figure 4-3, it is assumed that if the distance to the front vehicle is larger than 250 m, or the lateral difference is greater than 3m, there will be no conflict at this time-spatial point. Let V_F be the front vehicle speed, and V_S be the subject vehicle speed, the relative speed equals ΔV is (V_S - V_F). Let a_F be the front vehicle acceleration and a_S be the subject vehicle acceleration; the relative acceleration Δa equals to (a_S - a_F). D represents the distance between the two vehicles. The SPMD dataset provides all of this information except for a_F, which can be calculated from the front vehicle speed information between two consecutive

time stamps. Equations 4-1 - 4-3 detail surrogate safety measures TTC, MTTC, and DRAC, and the assumptions.

- TTC is defined as "*the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained*" [24].

$$TTC = \frac{D}{\Delta V} \qquad \text{Equation 4-1}$$

When $\Delta V \leq 0$, there is no risk of collision at that moment. Let L be the link length. The upper limit of TTC is set to be $\frac{L}{V_s}$, ensuring that if there is a conflict, it is within the roadway segment link.

- MTTC considers the trajectory parameters of the two consecutive vehicles, including their relative distance, speed, and acceleration [25]. The threshold of MTTC is also assumed to be $\frac{L}{V_s}$.

$$MTTC = \begin{cases} \max(t_1, t_2), & if\ \Delta a > 0 \\ \min(t_1, t_2), & if\ \Delta a < 0\ and\ \Delta V > 0 \\ t_3, & if\ \Delta a = 0\ and\ \Delta V > 0 \\ N/A, & if\ \Delta a \leq 0\ and\ \Delta V \leq 0 \end{cases} \qquad \text{Equation 4-2}$$

Where $t_1 = \frac{-\Delta V + \sqrt{\Delta V^2 + 2\Delta a D}}{\Delta a}$, $t_2 = \frac{-\Delta V - \sqrt{\Delta V^2 + 2\Delta a D}}{\Delta a}$, $t_3 = \frac{D}{\Delta V}$

- Deceleration Rate to Avoid a Crash (DRAC) is "*the minimum deceleration rate required by the following vehicle to avoid a crash with the leading vehicle if the speed of leading vehicle is unchanged during the process*".

$$DRAC = \frac{{V_S}^2 - {V_F}^2}{2(D - V_S * PRT)} \qquad \text{Equation 4-3}$$

Where PRT is the perception-reaction time. Default PRT value used 0.92s. When $V_S - V_F \leq 0$, there is no risk of collision at that moment. When $D \leq V_S * PRT$, the index can be treated as positive infinity, meaning the collision is certain to happen.

Trip-level Safety Surrogate Measures

A trip is defined as the time a vehicle traverses a roadway segment link. After the vehicle-level safety surrogate measures are available, they will be aggregated into safety surrogate measures for each vehicle trip. During a non-stop long trip, a vehicle may traverse the same link multiple times. Thus vehicle-level safety surrogate measures with the same "Trip" field value can be aggregated into multiple trip-level safety indexes. For the same vehicle, if there are multiple front targets at the same timestamp, the closest one ("CIPV =1") is kept.

It is assumed that the safety surrogate measure for each time interval equals the value for the end timestamp of that interval. For trip i in a link, the four safety indexes (SI) are formulated in Equations 4-4 - 4-7 to compute the trip-level safety surrogate measures on a link. For the sake of generality, TTC, MTTC, DRAC and their aggregations are called safety indexes (SI).

- Time Duration (SI_{1i}) is the total time when the surrogate measures exist. Note that the time duration should be less than or equal to the link travel time $\frac{L}{V_s}$. (if the vehicle travels the same link multiple times, it may larger than link travel time)

  $SI_{1i} = \sum_P (t_{i,j} - t_{i,j-1})$ Equation 4-4

- Average Index (SI_{2i}) is the weighted average of surrogate measures over time.

$SI_{2i} = \frac{\sum_P (t_{i,j} - t_{i,j-1}) * Index_{i,j}}{\sum_P (t_{i,j} - t_{i,j-1})}$ Equation 4-5

- Median Index (SI_{3i}) is the median value of surrogate measures over time.

$$SI_{3i} = \underset{P}{\text{median}}\{Index_{i,j}\}$$ Equation 4-6

- Extreme Index (SI_{4i}) is either the minimum TTC or MTTC, or the maximum DRAC, representing the most dangerous situation.

$$SI_{4i} = \min_{P}\{Index_{i,j}\} \; or \; \max_{P}\{Index_{i,j}\}$$ Equation 4-7

where P = {time intervals when safety surrogate measures are available}, j is the index of time stamp.

Link-level Safety Surrogate Measures

One link could include multiple trips: different vehicles can travel on the same link or one vehicle can travel the same link multiple times. Trip-level safety surrogate measures need to be aggregated to a link-level safety surrogate measure. The proposed link-level safety surrogate measures for each link are presented in Equations 4-8 - 4-11.

- Time Duration (SI_1) is the average length of time of all trips in a link

$$SI_1 = \frac{\sum_1^N SI_{1i}}{N}$$ Equation 4-8

- Average Index (SI_2) is the average index of all trips in a link

$$SI_2 = \frac{\sum_1^N (SI_{2i} * SI_{1i})}{\sum_1^N SI_{1i}}$$ Equation 4-9

- Median Index (SI3) is the median index of all trips in a link

$$SI_3 = \underset{1 \le i \le N}{\text{median}}\{SI_{3i}\}$$ Equation 4-10

- Extreme Index (SI_4) is the minimum or maximum index of all trips in a link

$$SI_4 = \min_{1 \le i \le N} \{SI_{4i}\} \text{ or } \max_{1 \le i \le N} \{SI_{4i}\}$$ Equation 4-11

4.1.3 Comparison of SSMs Based on the Crash Record

The comparison & selection of SSMs were conducted using statistical models on number of crashes. The statistical relationship between the link safety surrogate measures and mid-block rear-end crashes was developed using the negative binomial (NB) model, of which the mean is estimated as a log-linear function of the explanatory variables. Only links with observed surrogate measures were selected. Of the 2,772 selected links, the average link length is 352 m, the median length is 213 m, the minimum length is 35 m, and the maximum length is 3,635 m. The surrogate safety measures and other independent variables include vehicle maneuvering actions such as average speed and Brake Duration and segment link length. The link length is treated as an independent variable rather than an exposure variable to a crash because the length is used to set the upper limit for a safety index. Variance inflation factors (VIFs) were calculated for each independent variable to examine the possibility of multicollinearity. In this study, all VIFs are less than ten except for Average Index and Median Index. Small VIF values mean no high correlations exist among the independent variables after excluding Average Index and Median Index. Traffic volume such as AADT is typically included as the traffic exposure to crashes. However, during the variable selection process, AADT was not statistically significant for the selected links and therefore, AADT was excluded from the final NB models. The results of the models are shown in Table 4-2.

Table 4-2 Negative Binomial Models for Mid-block Rear-end Crashes

	TTC		MTTC		DRAC	
	Estimate	**p-value**	**Estimate**	**p-value**	**Estimate**	**p-value**
(Intercept)	-0.990***	$< 2e^{-16}$	-0.624***	$< 2e^{-16}$	-1.237***	$< 2e^{-16}$
Time Duration	0.222**	0.002	0.147***	$< 2e^{-16}$	0.063***	$< 2e^{-16}$
Extreme Index	-0.100***	$< 2e^{-16}$	-0.518***	$< 2e^{-16}$	4×10^{-6}**	0.007

Average Speed	0.082***	$< 2e^{-16}$	0.083***	$< 2e^{-16}$	0.077***	$< 2e^{-16}$
Brake Duration	1.136***	$< 2e^{-16}$	1.040***	$< 2e^{-16}$	1.390***	$< 2e^{-16}$
Link Length	0.001***	$< 2e^{-16}$	0.001***	$< 2e^{-16}$	0.001***	$< 2e^{-16}$
# of Observations	2772		2772		2772	
Dispersion Parameter	0.399		0.419		0.372	
2 log-likelihood	-9615.862		-9542.735		-9701.151	
AIC	9629.900		9556.700		9715.200	
McFadden Pseudo R-squared	0.066		0.075		0.054	

Note: Significant codes: '***' 0.001 '**' 0.01 '*' 0.05.

For TTC, the variables of Time Duration, Extreme Index, Average Speed, Brake Duration, and Link Length are statistically significant at the 1% level or lower. Time Duration and Extreme Index are two safety indexes that affect crash frequency statistical significantly. The positive sign of Time Duration means that the longer the dangerous situation lasts, the more frequently a crash will occur. The negative sign of Extreme Index suggests a larger Minimum TTC is associated with a fewer number of crashes. The positive impact of Average Speed on crashes indicates that a faster speed results in more crashes. The longer brake time within the link means a worse link safety condition (higher crash frequency). The positive effect of Link Length on crashes simply suggests that more crashes may happen if the link is longer.

When compared with TTC, the difference in parameter estimates for MTTC include the lower Time Duration value and larger Extreme Index value. It is found that mean value of Time Duration for MTTC is larger than TTC (1.22s vs. 0.56s) and mean value of Extreme Index for MTTC is lower than TTC (1.86s vs 3.43s), suggesting an MTTC-based conflict is easier to be detected compared with a TTC-based conflict. As an acceleration-based surrogate measure, all safety indexes are statistically significant for DRAC. The positive sign of Extreme Index for DRAC means the increase of the maximum DRAC leads to more crashes. Average Speed and Brake Duration are always statistically significant with the similar parameter estimates,

irrespective of the safety indexes. Excess speed, reflecting the driver's aggressiveness in choosing speed, was calculated and included in the model. The excess speed equals zero when the average speed is less than the speed limit and equals the difference between the two if the average speed is greater than the speed limit. Model results show insignificant safety indexes when replacing average speed with excess speed.

Dispersion parameter was estimated in the NB model to measure the data overdispersion. The respective values of 0.399, 0.419, and 0.372 in the models for TTC, MTTC and DRAC justify the choice of the NB model. Goodness-of-fit measures such as Akaike Information Criterion (AIC) and pseudo R-squared were used to compare the model performance. AIC uses the maximum log-likelihood function with a penalizing term related to the number of variables. A lower AIC value indicates a better fit. McFadden Pseudo R-squared, analogous to the R-squared value for linear regression models, equals to one minus the ratio of the log-likelihood of the full model to the log-likelihood of the intercept-only model. The pseudo R-squared takes a value between 0 and 1, and a higher value indicates better model performance. The relatively low R-square values mean that if crash data are considered as ground truth, it cannot be concluded with confidence which surrogate measure is the best safety metric. Among the three indexes, MTTC is considered as the most statistically significant surrogate safety measure to crashes because it has the highest Pseudo R-squared value and the lowest AIC value. In summary, the proposed safety information can be useful for evaluating segment link safety when roadway geometric and traffic information is limited.

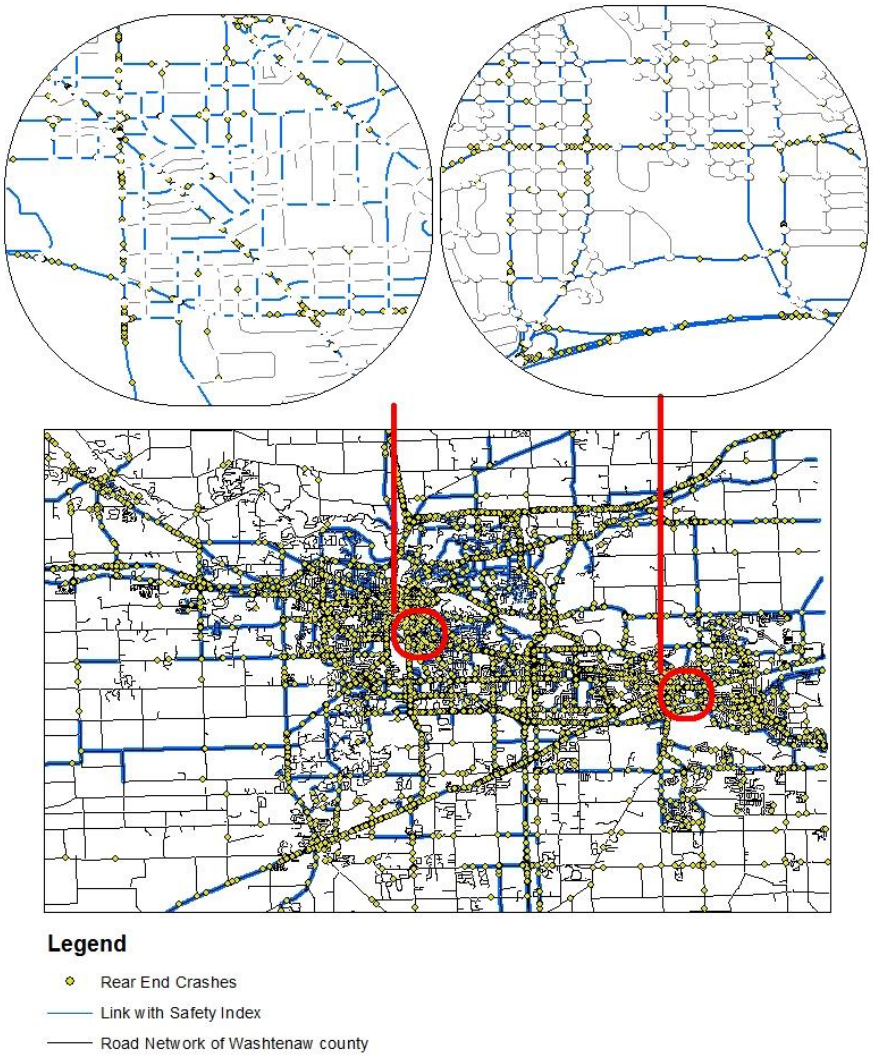


Figure 4-4 Maps with observed safety index and crash points.

4.1.4 Summary

A SPMD dataset was used to evaluate rear-end crashes in the mid-block. Three main objectives were accomplished: 1) develop a framework to process the SPMD big data; 2) construct surrogate safety measures from SPMD data; and 3) analyze the statistical relationship between crash records and the calculated safety index to compare and validate SSMs.

Unlike other studies that adopted surrogate safety measures to identify traffic conflicts, this study attempted to evaluate segment safety using surrogate safety measures. Aggregated surrogate measures were developed for a trip-level and a link-level safety index, as surrogate measures are usually taken at the vehicle level by measuring the time and space between a pair of vehicles. Surrogate measures are treated as safety indexes, which quantify the severity of a potential traffic conflict. For example, a higher value of TTC means a higher safety index. In the real world, however, a TTC of 3s may result in a crash while a TTC of 1s may not; this is dependent on the driver. Thus, arbitrary threshold values were not used in the study to preserve the integrity of the information. The logic of this study is to keep the calculated surrogate safety measures for all the time points and then provide indicators (e.g.: time duration, average index, median, or minimum/maximum index) to summarize these safety indexes on the same trip for each link.

The NB models for mid-block rear-end crashes show the expected impact of explanatory variables on crashes. Among the three models, MTTC has a better goodness of fit when compared with TTC and DRAC. The findings show that augmenting safety analysis with surrogate measures and vehicle performance (i.e.: speed and brake duration from connected vehicles) improves the overall model performance. Such information can be vital when detailed roadway and traffic data are absent.

Some abnormal numbers were produced and removed, and the measurement errors in the dataset were unknown. The study is also less comprehensive because there is no record of the dataset in some columns (e.g.: right turn or left turn signal). The complexity of the dataset means that some of the assumptions or data processing approaches used in this study may not be optimal in all situations; thus, future studies should search for other effective approaches. Future studies can be expanded into the comparison of other safety surrogate measures such as post-encroachment

time (PET), Delta-V, and extended Delta-V. New and emerging safety surrogate measures can be developed with the rich information provided through connected vehicle safety technologies.

4.2 Calculation and Validation of SSMs for Major Crashes in the Link and Intersection

This study attempted to measure road safety in both the link and intersection by using the SPMD data. Algorithms were proposed based on three levels of connected vehicle data that were used to measure road safety: level one contains single vehicle trajectory information; level two includes trajectories of a pair of vehicles following each other (lead and following vehicles); and level three augments level two with signal phasing and timing information. Real-world crash data were used to evaluate the methods on their accuracy, reliability, and effectiveness.

4.2.1 Data Collection & Processing

Besides the rear-end crashes in the mid-block, this section illustrates the calculation of TTC for major crashes in the link and intersection. Data preparation mainly consisted of selecting the study sites, joining data, and cleaning the data. Figure 4-5 shows a general procedural guide. Three levels of data were presented to show how data availability and quality improves the quantification of surrogate safety measures when measuring road safety using connected vehicle data from SPMD.

- *Level 1* includes the basic subject vehicle information (*DataWsu*) such as GPS location, speed, and longitudinal acceleration.
- *Level 2* adds the front vehicle information *(DataFrontTargets)* such as relative distance and relative speed.
- *Level 3* adds signal timing information *(SPaT)*: traffic signal status for each time stamp.

The three levels of data are the most commonly used connected vehicle information. Analytical methods were used to detect traffic conflicts by calculating surrogate safety measures. The detected traffic conflicts were compared with and evaluated by real-life crash data.

The *SPaT* (signal phase and timing) data from the RSE dataset are collected at deployed intersections. The *SPaT* data contain information of signal status, and only a portion of the data fields (e.g., timestamp and signal status) in the *SPaT* are used. Figure 4-5 includes details about the data elements for the subject vehicle (*DataWsu),* front vehicle *(DataFrontTargets),* and signal status *(SPaT).*

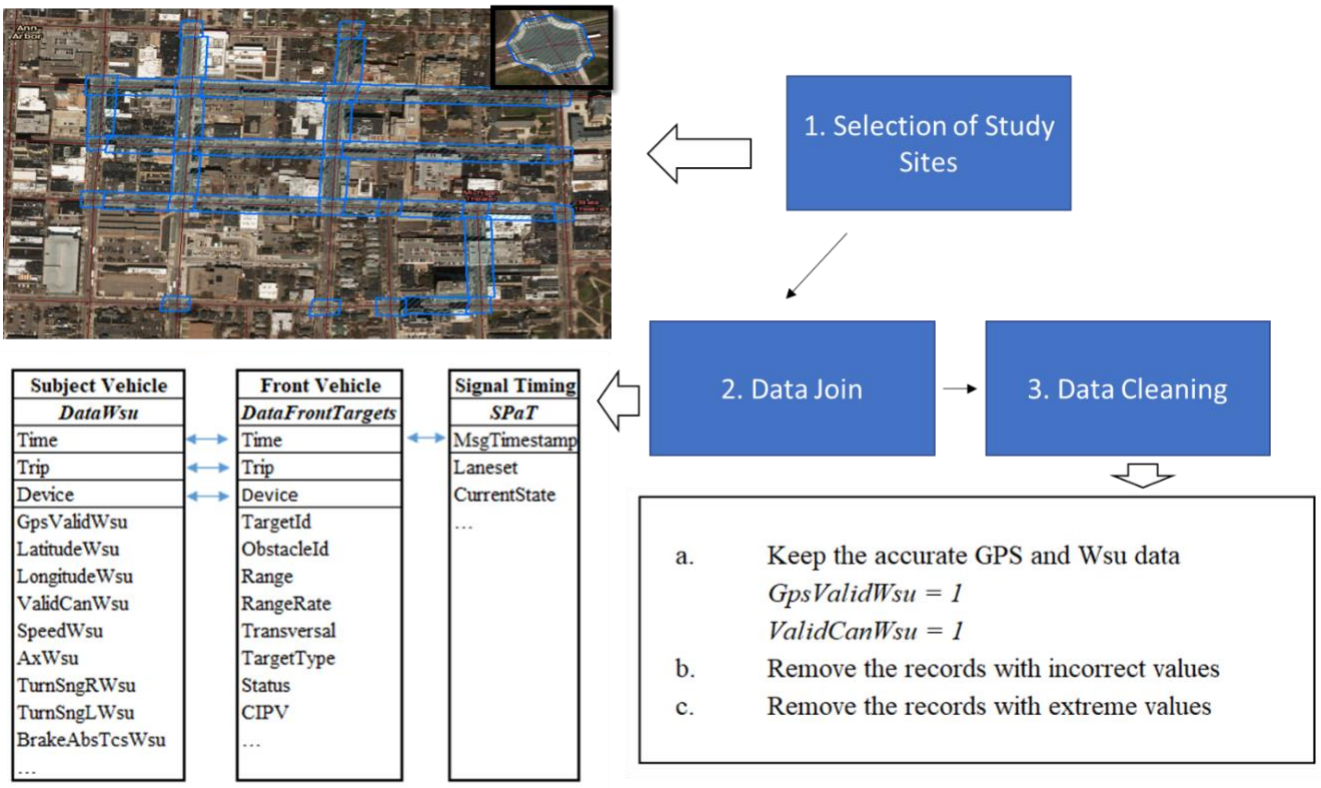


Figure 4-5 Procedures for data preparation.

The first step is to select study sites and extract the subject vehicle data. The study sites should have adequate crash records, which will be used for validation. Twenty segments and twenty intersections in downtown Ann Arbor were chosen as study sites. One intersection (Fuller Rd &

Maiden Ln) with RSE was chosen for preparing the Level 3 data. Polygons were drawn for each segment and for each intersection from stop line to stop line based on the real-world intersections from Google Maps. The polygons were used to extract the "*DataWsu"* data based on the longitude and latitude values. The data extraction left 62 vehicles at the selected intersections and 69 vehicles at the selected segments.

Vehicle data and signal timing data were joined after the subject vehicle information for each study site was prepared. The common data fields of "*Device*," "*Trip*," and "*Time*" were used to link the subjective vehicle and front vehicle information for the 21 intersections and 20 segments. Additionally, signal timing data were joined based on "*MsgTimestamp"* for the intersection with RSE. The data joining process for the *SPaT* data was very complex and will not be discussed in detail here.

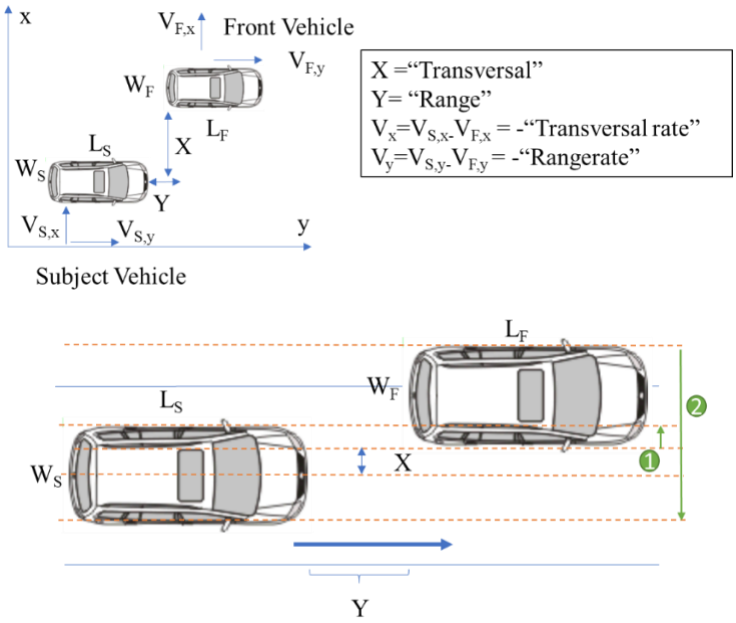
4.2.2 Methodology: Algorithms to process the three levels of data

Three levels of connected vehicle data were prepared for calculating surrogate safety measures. The acceleration of the subject vehicle was used as a surrogate safety measure for the first level data. Several studies have used extreme acceleration to identify dangerous driving behavior [66]. Liu et al. proposed the acceleration thresholds at different speeds [66]. The 95^{th} percentile values in each speed group were used in this study as the threshold to extract dangerous situations.

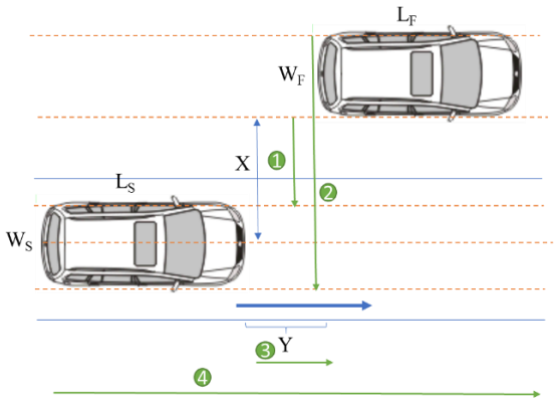
The second level data include information from both the subject vehicle and the front vehicle. Several studies have used this kind of data to calculate surrogate measures for rear-end crashes in a car-following environment. This study also explored the use of surrogate measures for major types of two-vehicle crashes. TTC was used as the surrogate measure, and conflicts can be determined based on a selected threshold. Figure 4-6 demonstrates the collision mechanism for

the derivation of TTC. Four typical crash-prone cases are shown in Figure 4-6: a) car-following, b) lane-change, c) intersection – the front vehicle is within the intersection, and d) intersection – the front vehicle is approaching the intersection.

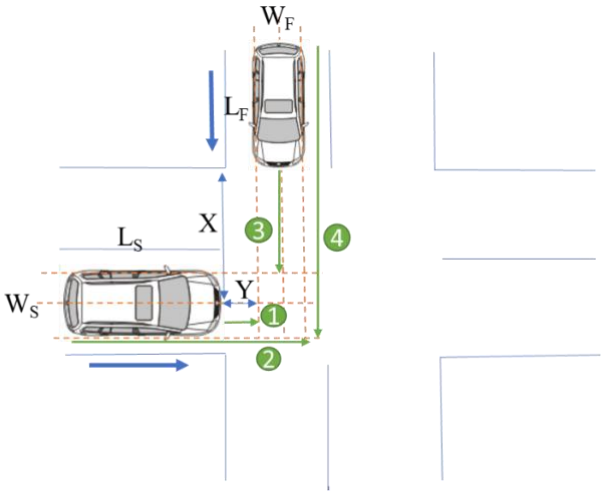
Variables should be defined before introducing each case. The length and width of the subject vehicle and front vehicle are L_S, W_S, L_F and W_F , respectively. Y is the longitudinal gap between the two vehicles and X is the transversal distance. V_x and V_y are the relative transversal speed and longitudinal speed to the front vehicle. X, Y and V_y can be obtained directly from the data as 'Transversal", "Range", and negative "Rangerate", respectively. V_x can be derived from X.



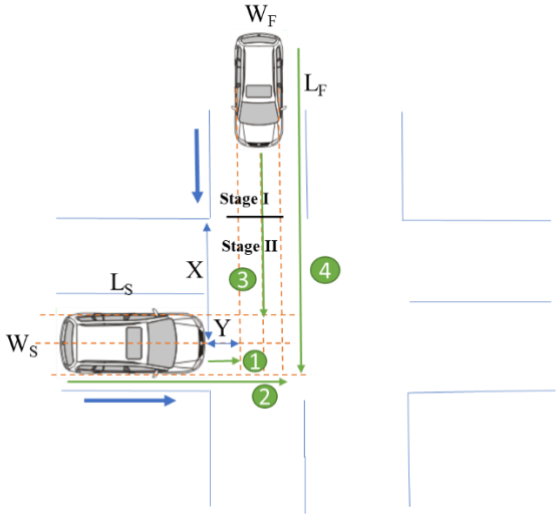
(a) Car- following Case



(b) Lane-change Case



(c) Intersection Case: front vehicle is within the intersection



(d) Intersection Case: front vehicle is approaching the intersection

Figure 4-6 of collision mechanism to calculate TTC.

Two assumptions are made in this study. First, the motion assumption for the segment analysis (subject vehicle in the segment) is that both the subject vehicle and front vehicle keep a constant speed. The assumption for the intersection analysis (subject vehicle in the intersection) is that the subject vehicle keeps a constant speed. The front vehicles will keep the constant speed if they are already in the intersection. If the front vehicles are approaching the intersection, they will remain at a constant speed before entering the intersection. The left turn and right turn speed are assumed to be 15 mph once they arrive at the intersection, and the speed does not change for the through movement. Second, this study considers rear-end collision and side-swipe collision on the segment; and rear-end collision, side-swipe collision, and angle collision are at the intersection.

Case (a) occurs when the front vehicle occupies the same lane and moves slower than the subject vehicle ($|X| < \frac{1}{2}W_s \,\&\, V_y > 0$)(shown in Figure 4-6(a)). A rear-end collision will occur if the front vehicle does not leave the lane on time and the driver takes no action. The maximum distance travelled before the arrival of the subject vehicle equals to *distance 1* ($\frac{1}{2}\mathrm{W_s} - \mathrm{X}$) or *distance 2* ($\frac{1}{2}\mathrm{W_s} + \mathrm{X} + \mathrm{W_F}$). Thus, $TTC = \frac{Y}{V_y}\ \ if\ (V_x < 0\ \&\frac{Y}{V_y} < \frac{\frac{1}{2}W_s - X}{-V_x})\ or\ (V_x > 0\ \&\frac{Y}{V_y} < \frac{\frac{1}{2}W_s + X + W_F}{V_x})or\ V_x = 0$. The potential conflict is a rear-end collision.

Case (b) happens when the front vehicle is in another lane and moves closer to the subject vehicle ($|X| > \frac{1}{2}W_s \,\&\, V_y > 0 \& \frac{X}{V_x} > 0$) (shown in Figure 4-6(b)). A rear-end collision will occur if the distance the front vehicle travelled before the arrival of the subject vehicle ranges from *distance 1* ($X - \frac{1}{2}W_s$) to *distance 2* ($\frac{1}{2}W_s + X + W_F$). A side-swipe collision will occur if the distance the subject vehicle travelled before the arrival of the front vehicle ranges from *distance*

3 (Y) to *distance 4* ($Y + L_S + L_F$). Thus, $TTC = \frac{Y}{V_y},\ if\ \frac{X-\frac{1}{2}W_S}{V_x} < \frac{Y}{V_y} < \frac{\frac{1}{2}W_S+X+W_F}{V_x}$ and $TTC = \frac{X}{V_x}, if\ \frac{Y}{V_y} < \frac{X}{V_x} < \frac{Y+L_S+L_F}{V_y}$.

Case (c) happens around the intersection when two vehicles move closer to each other and the front vehicle is already within the intersection (shown in Figure 4-6(c)). The figure shows a typical right-angle collision case. An angle collision will occur if the distance the subject vehicle travelled before the arrival of the front vehicle ranges from *distance 1*: (Y) to *distance 2*: ($Y + L_S + W_F$) or if the distance the front vehicle travelled before the arrival of the subject vehicle ranges from *distance 3*: ($X - \frac{1}{2}W_S$) to *distance 4*: ($\frac{1}{2}W_S + X + L_F$). Thus, $TTC = \frac{X}{V_x}, if\ \frac{Y}{V_y} < \frac{X}{V_x} < \frac{Y+L_S+W_F}{V_y}$ and $TTC = \frac{Y}{V_y},\ if\ \frac{X-\frac{1}{2}W_S}{V_x} < \frac{Y}{V_y} < \frac{\frac{1}{2}W_S+X+L_F}{V_x}$. In reality, an angle collision around the intersection cannot be a perfect right-angle collision. Given the front vehicle information, it is almost impossible to list all possible trajectories of the front vehicles and all the possible conflicts (type and location). Thus, relative motion is used to calculate the surrogate measure. A constant speed is assumed for both the subject vehicle and front vehicle, and TTC are all based on the relative distance and speed, meaning there is not much difference between actual TTC and TTC derived from the right-angle crash.

Case (d) happens around the intersection when the two vehicles move closer to each other and the front vehicle is not within the intersection (shown in Figure 4-6(d)). Two stages are involved when the subject vehicle is reaching the collision. The first stage is to reach the stop line (intersection boundary), where a constant speed is assumed; the second stage is to cross the intersection. The movement (left turn, through, or right turn) is based on its lane movement information. The minimum TTC is used when the lane is allowed to make multiple movements.

The through movement does not change the speed. Left turn movement and right turn movement are assumed to have a constant speed of 6.7 m/s (15 mph). The longitudinal speed and lateral speed are both assumed to be $\frac{6.7}{\sqrt{2}}$ m/s. The speed of the subject vehicle remains constant. Similarly, if the vehicle makes left turn or right turn movement, the longitudinal speed and lateral speed are assumed to be $\frac{Speed}{\sqrt{2}}$. Movement of the subject vehicle can be determined by *TurnSngRWsu* and *TurnSngLWsu* or by the yaw rate. The left turn movement usually has a yaw rate less than -15 deg/sec, and the right turn has a value of greater than 15 deg/sec.

The front vehicles around the intersection are the only vehicles analyzed in the intersection evaluation. Thus, longitudinal relative distance (i.e.,Y) is set to be less than the intersection length, which is around 50m for the study sites. Rear-end collisions and side-swipe collisions also occur around intersections. A lateral speed of (i.e., $|V_x|$) > 5 m/s, may signal an angle collision. TTC * speed < 50m and $|V_x| \leq 5\ m/s$ may signal a rear-end collision or a side-swipe collision near the intersection.

Level 3 data includes the signal status for each lane set at each time stamp. The lane location for the front vehicle can be estimated using the relative distance (X, Y). Facing a red signal timing, subject vehicle cannot collide with the front vehicle. Note the estimated lane information for each front vehicle is only useful when the vehicle is before the stop line. Vehicles that have already been in the intersection will not be affected by the signal timing.

An algorithm was proposed for using the SPMD data to determine whether the front vehicles were within the intersection and to gather the lane information if the vehicles were not within the intersection. The four-way intersection (Figure 4-7) with signal timing status was presented to illustrate the algorithm. The steps of the algorithm are listed as follows.

1) Extract the GPS coordinates of the four intersection corners: A ($Long_A$, Lat_A), B ($Long_B$, Lat_B), C ($Long_C$, Lat_C), and D ($Long_D$, Lat_D). The length of the intersection polygon is around 50 m for each side.

2) Convert the GPS coordinates of the four intersection corners and locations ($Long_S$, Lat_S) of the subject vehicle (S) into projected coordinates. NAD 83 state plane coordinates, a Michigan-projected coordinate system (Michigan South Zone. 2113), is used in the study. The projected coordinates are A (x_A, y_A), B (x_B, y_B), C (x_C, y_C), D (x_D, y_D), and S (x_S, y_S).

3) Calculate the distance from the subject vehicle to each intersection boundary (AB, CD, BC, AD) using trigonometry. The distances are named as d1, d2, d3, and d4, respectively. For example, calculation of d1 is shown in Equation 4-12.

$$d1 = \sqrt{(x_S - x_B)^2 + (y_S - y_B)^2} \times \sqrt{1 - \left(\frac{(x_A - x_B)(x_S - x_B) + (y_A - y_B)(y_S - y_B)}{\sqrt{(x_A - x_B)^{2+}(y_A - y_B)^2}\sqrt{(x_S - x_B)^{2+}(y_S - y_B)^2}}\right)^2} = \frac{|(y_B - y_A)x_S - (x_B - x_A)y_S + x_B y_A - x_A y_B|}{\sqrt{(x_A - x_B)^{2+}(y_A - y_B)^2}}$$

Equation 4-12

4) Use the "*GpsHeadingWsu*" field to determine the heading of the subject vehicle. The intersection configuration and the SPMD data helped determine that the eastbound heading is from 30 to 90 degrees, the southbound heading is from 90 to 160 degrees, the northbound heading is from 0 to 30 or from 320 to 360 degrees, and the westbound heading is from 200 to 320 degrees. The headings can also be verified by the location (d1, d2, d3, d4) of the subject vehicle.

5) Use the longitudinal relative distance Y and transversal relative distance X to determine the lane information for each front vehicle. Figure 4-7 shows the 12 lanes in the intersection.

Distances from the right intersection boundary to each lane are measured using Google Maps, and are noted as D1, D2, … D12, respectively. The figure shows that a vehicle is moving eastbound with front vehicle information X and Y. X is used to determine that the vehicle is in the southern lanes. $\left(d2 - \frac{1}{2}L_S - Y\right)$ are within the D10, showing the front vehicle is in lane 10. The distance to the stop line (intersection boundary) is (X – d3), which is used to calculate TTC for case (d) above. Once the lane information is determined for the front vehicles, signal timing status can be jointed to each front vehicle.

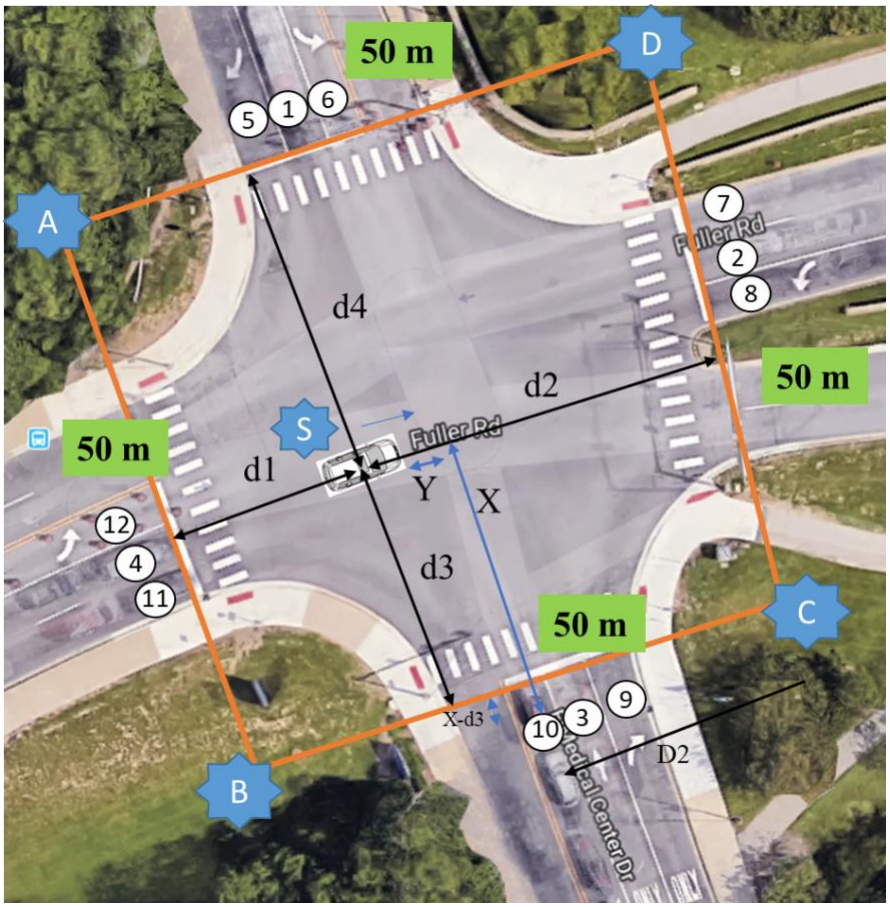


Figure 4-7 Estimating locations of front vehicles from SPMD data.

Figure 4-8 shows a detailed algorithm to process the connected vehicle data using surrogate measures. Level 1 data detected dangerous events with the extremely large value of acceleration.

Level 2 data calculated TTC for each time stamp and identified traffic conflicts based on the threshold of TTC. In the segment, only the closest front vehicle was used to calculate TTC, while in the intersection, minimal TTC for all the front vehicles was used. One conflict was counted when TTC was below the threshold continuously for the same intersection/segment and the same pair of vehicles. Level 3 of data only applied to intersections. Front vehicles presented during red signal light were not considered to calculate TTC. Other than that, Level 3 of data used the same methodology of Level 2 data.

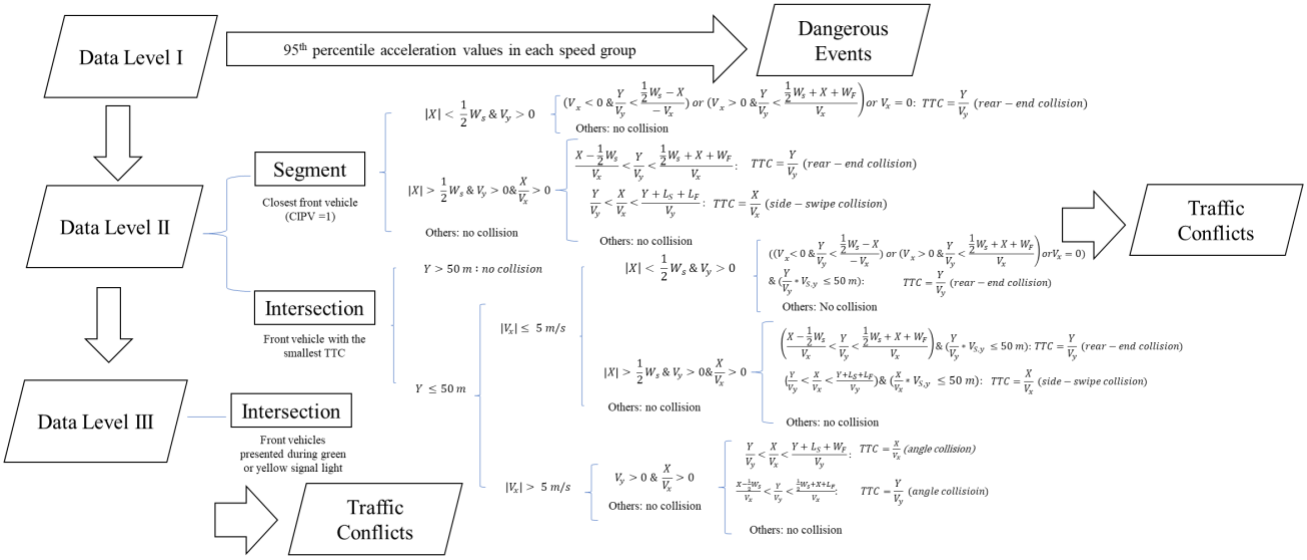


Figure 4-8 Algorithm to process the connected vehicle data.

4.2.3 Validation of the calculated SSMs

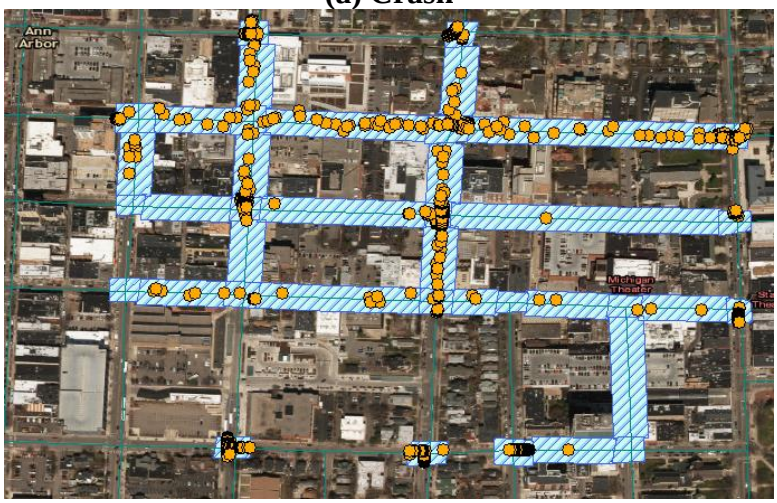
This part verifies the calculated TTC using crash count.

Correlation between number of conflicts and number of crashes

Surrogate safety measures were calculated using the proposed algorithm for each level of data. First, the proposed connected vehicle safety measure was validated using the real-world crash records. 5-year crash data (2010-2015) from the Michigan DOT were used in this study. Extreme traffic events and TTC were calculated for the 20 segments and 20 intersections using level 1 and

level 2 data, respectively. Figure 4-9 shows the crashes and the conflicts detected using a TTC threshold of 3s. The pattern of crashes and conflicts seems consistent via visual observation.

(a) Crash



(b) Conflicts

Figure 4-9 Comparison between crashes and conflicts for the selected sites.

Different thresholds of TTC (1.5 s, 2 s, 2.5 s, 3 s, 3.5 s) were used to identify conflicts. A correlation analysis was conducted between the number of extreme events or conflicts and the number of crashes. Table 4-3 shows the validation results. Overall, the correlation between extreme events and crashes is acceptable despite a large number of extreme points detected. Conflict points vary by the threshold value. Based on the data in the study scope, a TTC value of 1.5 s or 3.5s is more suitable to detect conflicts because both threshold values are highly

correlated with crash count. The surrogate measure computed for the intersection has a higher correlation with the actual crashes, proving the proposed methods can measure intersection safety at a certain level. On the segment, the lower correlation between surrogate measures and the crash may be due to the low volume of traffic along the roadway, requiring further exploration. All crashes happening at the study sites were rear-end crashes; however, some of the calculated conflicts were based on sideswipe and angle collisions, showing this method's ability to measure safety for certain types of crashes in the absence of crash data.

Table 4-3 Validation of the Surrogate Safety Measures

		Level I Data	**Level II Data: TTC Threshold (s)**				
		Extreme Points	**1.5**	**2**	**2.5**	**3**	**3.5**
Segment	# of Extreme Events or Conflicts	6071	10	39	80	121	155
	Correlation with Crash Data	0.32	0.36	0.26	0.34	0.50	0.53
Intersection	# of Extreme Events or Conflicts	3545	5	62	287	549	885
	Correlation with Crash Data	0.75	0.70	0.70	0.75	0.72	0.77
Overall Correlation with Crash Data		0.41	0.45	0.12	0.11	0.33	0.36

More than 9,000 extreme events (level 1 data) were detected for the 20 segments and 20 intersections, but only slightly more than 600 of the events were determined as conflicts (level 2 data) when a threshold of 3s was chosen. This indicates that many false alarms can be avoided when a more comprehensive dataset is used. TTC was calculated for level 3 data with signal information. A couple of conflicts were detected using the level 2 data, but those were treated as false alarms in level 3 data because the dangerous front vehicles were forced to stop at a red signal light. The added signal timing data in level 3 makes the algorithm more realistic, meaning less assumptions need to be made. In summary, it is clear that a distinct improvement exists from level 1 data to level 3 data when it comes to measuring safety.

Because it requires manual effort to extract the intersection polygon based on Google map, the calculation of TTC cannot applies to the whole network. It requires further effort to extract the intersection automatically.

4.2.4 Summary

This study proposed methods for measuring road safety using SPMD connected vehicle data. Surrogate safety measures were calculated for three levels of data: level 1 included subject vehicle information, level 2 included the added front vehicle information, and level 3 involved signal timing information. Algorithms and detailed steps were proposed to calculate surrogate measures. The methods were evaluated by comparing crash data with the traffic conflicts derived from different data levels. According to the results, the proposed algorithm is a promising substitute for crash data when measuring road safety.

However, the data and the algorithm present some limitations that may hinder the use of the connected vehicle data.

1) Data Collection

The camera sensor used to detect front vehicles prevents only rear-end collisions because the sensing range of the camera is limited (e.g., angle is usually 80 degrees and range is from -31 m to 31 m); therefore, not all front vehicles (especially those on the side) can be detected. A side camera sensor is suggested in order to obtain a more accurate safety measure at intersections.

2) Data Quality

GPS data (geo-coordinate) is not always accurate, meaning the results are less reliable. Many records were removed because of the inaccuracy of GPS information ("*Gpsquality*" = 0). Even for some records with "*Gpsquality*" = 1, location of the subject vehicle is not accurate so that there is a shift of the trajectory compared to the real-world map. Additionally, some geo-coordinates fall outside the segment or the intersection, and some intersection-related conflicts are counted as segment-related conflicts. The lane information for the front vehicle may not be correct because of the unreliable GPS information. A video camera should be placed at the intersection, when possible, to check the quality of the GPS data.

The relative distance to the front vehicle ("*Range*") is not always accurate, especially at an intersection. Large values for "*range"* and "*transverse"* tend to exist around intersections, and it seems that the front vehicle often falls outside the road. A video camera at the intersection could perform a data quality check.

Some data fields have either incorrect information or no information. For example, fields for the left-turn signal and right turn signal are always zero. The zero values contradicted to the large yaw rates (> 15 is right turn and <-15 is left turn). Thus, quality check is suggested during the data collection.

3) Data Processing Algorithm

The data processing algorithm involves some assumptions, and still requires improvements. It is difficult to determine the conflicts detected at the segments that might happen at the intersections, and the conflicts detected at the intersections that might happen at the links. Additionally, the algorithm used to calculate TTC for angle collision may not exactly follow the real-world situation. Finally, a detailed intersection configuration is required to estimate the

exact lane information, which adds more work to the big data analysis (e.g., network analysis). In the future, image processing should automatically extract this accurate intersection information.

Several suggestions were proposed to deal with those limitations. The traditional surrogate measure TTC was used to measure safety in this study. In the future, more surrogate safety measures should be evaluated using connected vehicle data, requiring a more comprehensive data collection. Without a more robust dataset, It would be worthwhile to explore SPMD data with machine learning techniques as well, as opposed to analytical methods using surrogate measures. Machine learning techniques could detect certain safety-related patterns from the big data source, and the results can be compared with the crash but the micro-simulation results.

Chapter 5 Road Safety Evaluation using Surrogate Safety Measures

This chapter illustrates how to use surrogate safety measures to evaluate road safety using two study cases: rear-end crashes in the midblock and lane departure events in the work zone.

5.1 Estimating Rear-end Crashes in the Midblock

This study used extreme value theory (EVT) model to estimate the rear-end crashes in the midblock in the absence of crash data. Minimal TTC was used as the surrogate safety measure.

5.1.1 Data Collection & Processing

This section utilizes the same SPMD data in Chapter 4. It employed the minimum TTC as a critical measure in evaluating network safety, specifically concerning rear-end crashes in mid-block segments. The minimum TTC is determined during the interaction between two vehicles, which is defined as the period when the subject vehicle is approaching the front vehicle. An interaction is assumed for our study when the TTC is less than 10 seconds. For the ensuing EVT analysis, the negated value of the minimum TTC is utilized.

The scope of this analysis is confined to state highways. Roadway-related covariates include AADT, link length, the number of lanes, functional class, and speed limit. Meanwhile, driver-related covariates (assessing driver behavior as reckless or cautious) consider the median TTC for the driver, the percentage of brake application, and the variation of speed, quantified by an index of dispersion (the variance to mean ratio). Additionally, the type of front vehicle (passenger car or truck) was included in the analysis.

5.1.2 Methodology: Safety Evaluation using Traffic Conflict Theory

To evaluate road safety, researchers can use summary statistics of SSM, composite safety index such as link-level index in Chapter 4, or traffic conflicts. In this chapter, number of crashes were estimated using the traffic conflict theory. Extreme Value Theory (EVT) were adopted for the estimation.

There are two kinds of EVT models: block maxima (BM) and peak-over-threshold (POT). For BM, the observations are aggregated into fixed intervals over time and space, and then the extremes are extracted from each block by identifying the maxima in each single block. However, for the applied dataset, it's subjective to determine the block. Thus, the POT method with a Generalized Pareto (GP) distribution is used for the EVT modeling. The POT method quantifies the stochastic behavior of processes at extreme levels by considering conflicts surpassing specific thresholds. For instance, in Figure **5-1**, surrogate safety measures surpassing severity thresholds (S_2) are identified as exceedances or risky events [6].These exceedances are then used for EVT modeling to predict the probability of extreme events (S_3).

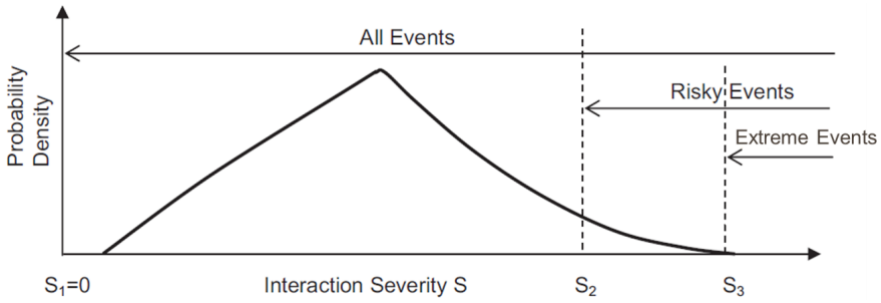


Figure 5-1 Illustration of EVT modelling (Source: Tarko 2012)

EVT focuses on the tail of the distribution, with less frequent exceedances occurring as severity increases. The Generalized Pareto (GP) distribution, selected for its applicability to tail events

surpassing thresholds, is employed to model exceedance distributions. This GP distribution is defined as [6]:

$$P(S > D_2|x_3) = 1 - F(D_2|x_3) \quad \text{Equation 5-1}$$

$$f(S|x_3) = \begin{cases} \frac{1}{\sigma} \times \left(1 + k \times \frac{S-\theta}{\sigma}\right)^{-1-\left(\frac{1}{k}\right)} & for\ (k > 0\ and\ \theta < S)\ or \\ & \left(k < 0\ and\ \theta < S < -\frac{\sigma}{k}\right) \\ \frac{1}{\sigma} \times e^{-S-\frac{\theta}{\sigma}} & for\ k = 0\ and\ \theta < S \end{cases} \quad \text{Equation 5-2}$$

Where

S = transformed risk event severity,

D_2 = threshold collision proximity,

x_3 = exogenous conditions under which the GP distribution is homogeneous,

F = cumulative GP distribution,

f = GP probability density function,

k = shape parameter,

σ = scale parameter.

Using the statistical tool R, the research team applied the Generalized Pareto (GP) distribution with the Peak Over Threshold (POT) method to model exceedances, providing shape and scale parameters for computing conditional probabilities of extreme events. Optimal threshold (S_2) was identified to ensure statistical reliability and model validity.

5.1.3 EVT Results & Discussion

To perform an EVT analysis, a threshold needs to be determined and selected from the observed maximum negated TTC. To determine the optimal threshold, an assessment of mean residual life

and stability plots were performed. A threshold can be determined when the mean residual life plot is almost linear, and the reparametrized scale and shape estimates become constant.

In Figure 5-2, the mean residual life plot of the maximum negated TTC thresholds is linear starting from a negated threshold of around −1.5 s. The stability of GPD reparametrized scale and shape parameters were also analyzed (Figure 5-3). Then diagnostic plots were compared for several negated thresholds around -1.5s and -1.25s was selected for the best fit. Figure 5-4 shows the diagnostic plot for the stationary model using a negated threshold of -1.25s. The figures show that the modeled GPD with a threshold of u= 1.25 s has satisfactory fitting results to the empirical data since the points fall close to the 45° line in the simulated QQ plot.

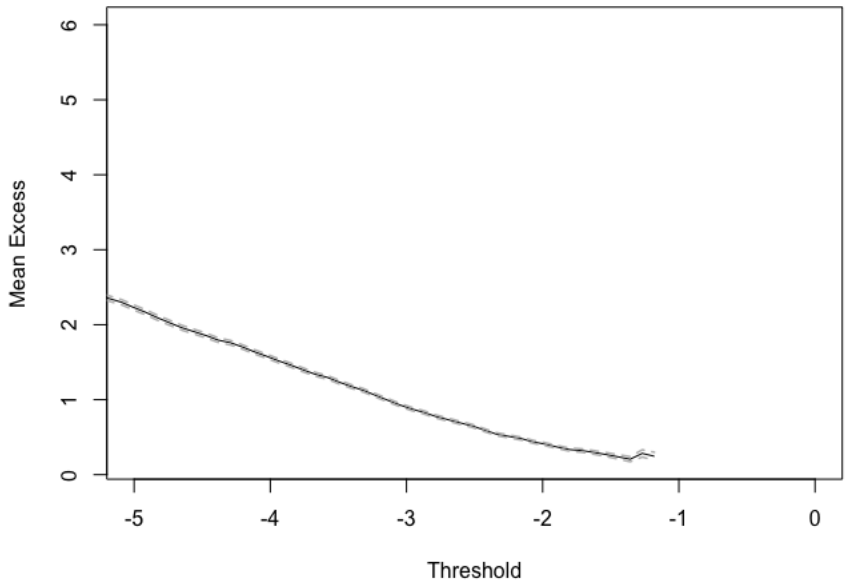


Figure 5-2 Mean residual life for the full data set.

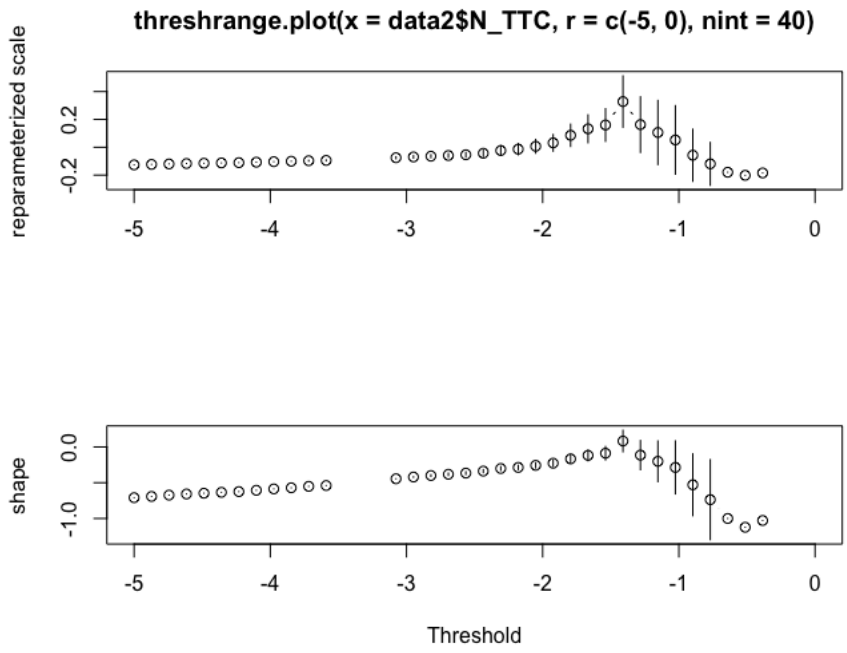


Figure 5-3 Stability plot for GPD model reparametrized (modified by subtracting the shape and multiplied by the threshold) scale parameter and shape parameter for different TTC thresholds.

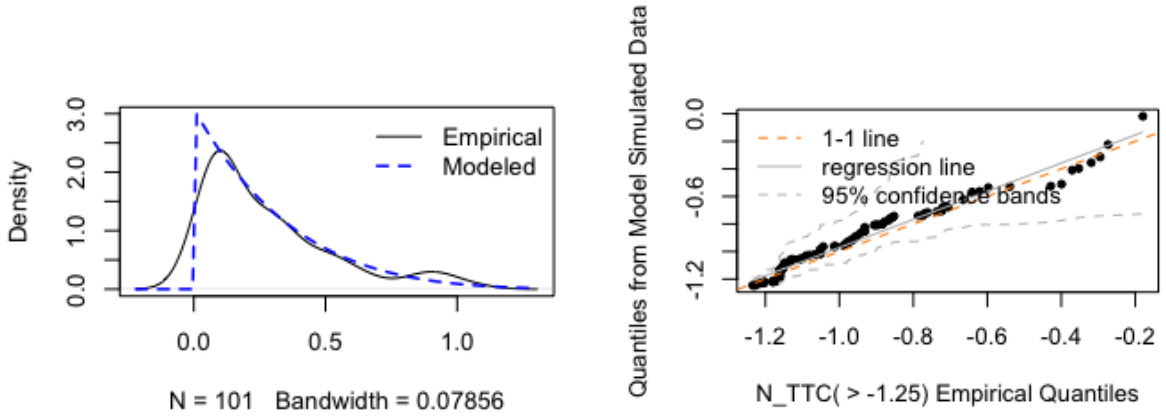


Figure 5-4 (Kernel) Probability density plot (left) and simulated QQ plot (right) for the stationary POT model.

The EVT model is non-stationary and incorporates covariates in the scale parameter estimation. The shape parameter remains unchanged due to the absence of factual evidence indicating non-stationarity in tail behavior [67]; thus, no covariates were included in its estimation.

Before incorporating the covariates in the model, the collinearity of the covariates was calculated (Table 5-1). From the table, it seems that the three driver-related covariates are correlated with each other and only one should be included in the model. Similarly, either AADT or link length should be included. Various combinations of the variables with little collinearity were assessed using the likelihood ratio test to streamline model structures and variable inclusions. The non-stationary model, which incorporates a linear combination of AADT and median of driver's TTC in the scale parameter, exhibits significance with a small p-value of 0.07 in a likelihood ratio test compared to the stationary model. However, when additional terms are included in the linear combination, the model loses significance, as indicated by a larger p-value exceeding 0.1 in a likelihood ratio test when compared to the previous non-stationary model.

Table 5-1 Collinearity of Covariates

	AADT	Num_Lane	Func_class	Spd	Link_Length	median_TTC	pcnt_brake	variation_spd	targettype	N_TTC
AADT	1.00									
Num_Lane	0.36	1.00								
Func_class	-0.59	0.03	1.00							
Spd	0.64	-0.10	-0.71	1.00						
Link_Length	0.28	-0.13	-0.34	0.52	1.00					
median_TTC	0.42	0.01	-0.51	0.60	0.36	1.00				
pcnt_brake	-0.27	0.11	0.38	-0.40	-0.17	-0.60	1.00			
variation_spd	-0.30	0.10	0.45	-0.50	-0.30	-0.75	0.70	1.00		
targettype	-0.02	-0.02	0.01	-0.03	-0.03	-0.08	0.04	0.03	1.00	
N_TTC	-0.15	0.11	0.23	-0.27	-0.20	-0.20	0.18	0.21	-0.01	1.00

Consequently, the final model incorporates a linear combination of AADT and median of TTC in the scale parameter. The optimal threshold of 1.25s results in 101 exceedances (out of a total of 4,931 vehicles) which were used to develop the EVT model. As depicted in Table 5-2, estimation results of the final model reveal a positive scale parameter for AADT, signifying that an increase in traffic volume increases the scale parameter and consequently, the variance of the negated TTC distribution. This increase indicates an increase in the probability of extreme events. Conversely, a negative scale parameter for median of driver's TTC indicates that a more cautious driver has a lower the scale parameter and hence, the variance of the negated TTC distribution. This increase suggests a deduction in the probability of extreme events.

Table 5-2. Estimation results for the POT model

Description	Estimated Value	Standard Error	P-value
Shape Parameter, k	0.27	0.11	<0.01
Scale Parameter, σ_0	0.45	0.11	<0.01
Scale Parameter, σ_1(AADT)	4.38E-06	2E-08	<0.01
Scale Parameter, σ_2 (Median_TTC)	-2.03E-02	8E-03	<0.01
Negative loglikelihood	-30.6		
Number of Observations	4,931		
Threshold (s)	1.25		
Number of Exceedances/risky Events	101		

5.1.4 Summary

This study estimated the likelihood of rear-end collisions at mid-block locations without actual crash data by using minimal Time-to-Collision (TTC) as a surrogate safety metric. An Extreme Value Theory (EVT) model was developed to illustrate the relationship between this probability and various road and driver-related variables. The findings from the EVT model demonstrate how traffic volume and driver aggressiveness impact road safety, indicated by minimal TTC values. Specifically, an increase in Average Annual Daily Traffic (AADT) could positively

affect the likelihood of mid-block rear-end collisions. Conversely, a higher median driver TTC value tends to enhance road safety, aligning with the initial assumptions.

Although the EVT model can predict the probability of extreme events, its application to SPMD data should be approached with caution. Firstly, SPMD data may not fully capture road safety conditions due to the low penetration rate of connected vehicles during the data collection period. The accuracy of this model is expected to improve as more connected vehicles participate in future data collection. Moreover, an algorithm needs to be developed to estimate road safety effectively when connected vehicle penetration is low. Secondly, this safety evaluation is currently limited to rear-end crashes at mid-block locations. Future research should expand the use of the algorithm in Chapter 4 to include major crashes at both links and intersections and apply it at a network scale.

5.2 Evaluating Work-zone Safety under Different Lane Width and Shy Distance

This study used statistical models to evaluate the impact of lane width and shy distance on the work zone lane departure events. Lateral distance — specifically the clearance from the right side of the vehicle — served as the surrogate safety measure.

5.2.1 Data Collection & Processing

The challenge of collecting and modeling safety data in work zones is well-recognized, particularly when employing encroachment-based approaches to identify the relationship between roadside crashes and road conditions. Traditional sources of encroachment data are scarce, leading this study to use vehicles' lateral position as a surrogate measure for safety evaluation. The study focuses on determining the influence of lane width and shy distance on safety by collecting large-scale lateral distance data using Light Detection and Ranging (LiDAR)

sensors. This data was gathered from 17 different work zone sites across Wisconsin, Illinois, and Michigan. It is important to note that records showing significant variability in lateral distance, with a standard deviation greater than 80 cm, were considered as outliers and consequently excluded from the analysis.

The collection was limited to lateral distances from the right side, which restricts the safety analysis to potential right lane (Lane 1) departures. All surveyed sites maintained two lanes of traffic flow during the work zone conditions. The refined dataset for the subsequent analysis includes lateral distance records from the right lane with reliable speed information, resulting in a total of 273,269 vehicles across the 17 different locations.

5.2.2 Exploratory Data Analysis

To illustrate the impact of lane width and shy distance on vehicle lateral position, each site was categorized into a lane-and-shy-distance bin. There are six bins in the dataset: 11 (1-ft lane, 1-ft shy distance), 12 (1-ft lane, 2-ft shy distance), 13 (1-ft lane, 3-ft shy distance), 21 (2-ft lane, 1-ft shy distance), 22 (2-ft lane, 2-ft shy distance), and 23 (2-ft lane, 3-ft shy distance). Table 5-3 presents summary statistics of lateral distance according to lane and shoulder bins. The count of vehicles, minimum, maximum, average, standard deviation, average during day and night are shown for lateral distance. The average lateral distance for daytime and nighttime observations were compared, noting that vehicles tended to have larger lateral distances at night for all sites (e.g., 54.6 inches during the day versus 50.7 inches at night at site WI-1).

Table 5-3. Summary statistics of lateral distance to edge line by lane width and shy distance bins

Site	Speed Limit (mph)	Lane Width (ft)	Planned Shy Distance (ft)	Measured Shy Distance (in)	Bin	Count	Lateral Distance (in)					
							Min	Max	Ave.[1]	Ave. Day	Ave. Night	Std. Dev.[2]
WI-4	60	12	3	33	23	7,061	8.1	100.1	44	43.2	46.9	11
WI-6	60	12	2	24	22	19,115	4.1	101.7	42.4	41.9	46.1	11.6
WI-7	60	12	2	32	22	13,640	0.2	105.4	42.4	42.0	46.2	10.8
WI-8	60	12	2	34	22	23,647	4.4	103.6	42.0	41.8	44.4	10.7
IL-2	45	12	1	13	21	11,935	12.4	101.7	46.8	45.4	49.8	11.1
IL-3	45	12	1	13	21	19,553	14.3	101.6	51.3	48.6	55.7	11.4
MI-1	60	11	3	30	13	21,667	-10.3	90	31.8	31	38.4	9.9
WI-2	60	11	2	19	12	5,407	1.2	89.4	41.2	41.2	-	11.4
WI-3	60	11	2	24	12	2,116	1.4	80.2	29.7	29.7	-	9.9
MI-2	60	11	2	22	12	8,636	1.6	89.4	33.1	32.8	37	9.5
MI-3	60	11	2	28	12	15,388	0	89.5	37.4	36.8	41	10.6
MI-4	60	11	2	22	12	21,563	-0.9	89.8	40.1	39.1	47	10.2
MI-5	60	11	2	22	12	14,054	-4.8	89	33.3	32.6	36.7	9.3
MI-6	60	11	2	24	12	12,103	-14.1	89.9	33.7	33.4	35.2	9.9
IL-1	45	11	2	24	12	21,989	1	90	34.3	32.1	37.7	10.2
WI-1	55	11	1	12	11	20,043	20.3	89.9	52.5	50.7	54.6	8.8
WI-5	55	11	1	14	11	35,352	-2.2	90	34.2	33.5	39	9.7
All						273,269						

Notes: [1] Average, [2] standard deviation.

The probability density functions (PDFs) for lateral distance to the edge line at each location are plotted in Figure 5-5. Both PDFs and summary statistics indicate that wider lanes correlate with increased lateral distance, whereas wider shy distances correspond to decreased lateral distance. This observation aligns with our initial assumption, which warrants further validation through additional statistical analysis. In addition, similar plots for lateral distance to the barrier were shown in Figure 5-6. Wider lanes or wider shy distances increase the lateral distance to the barrier.

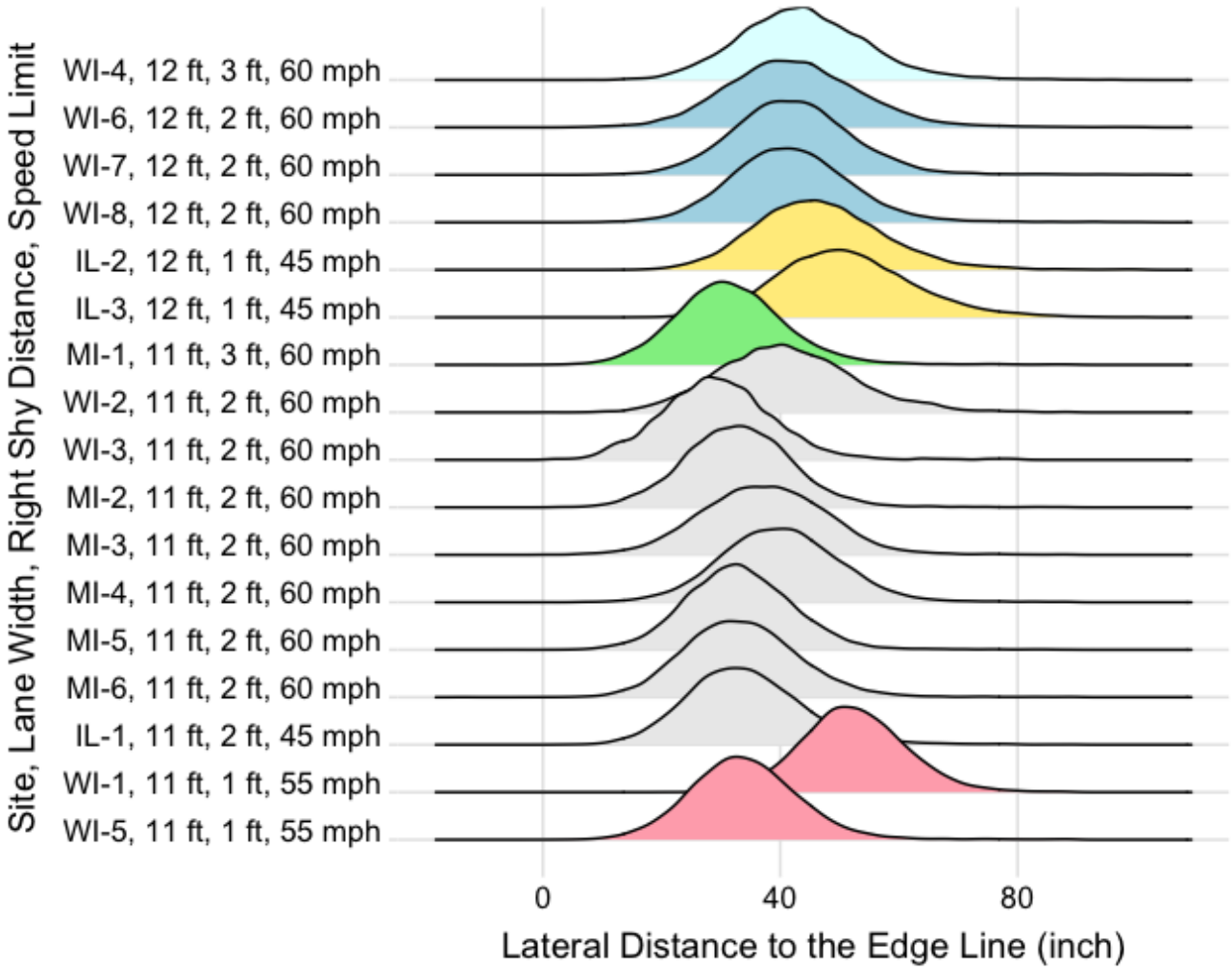


Figure 5-5. Empirical probability density function of lateral distance to the edge line

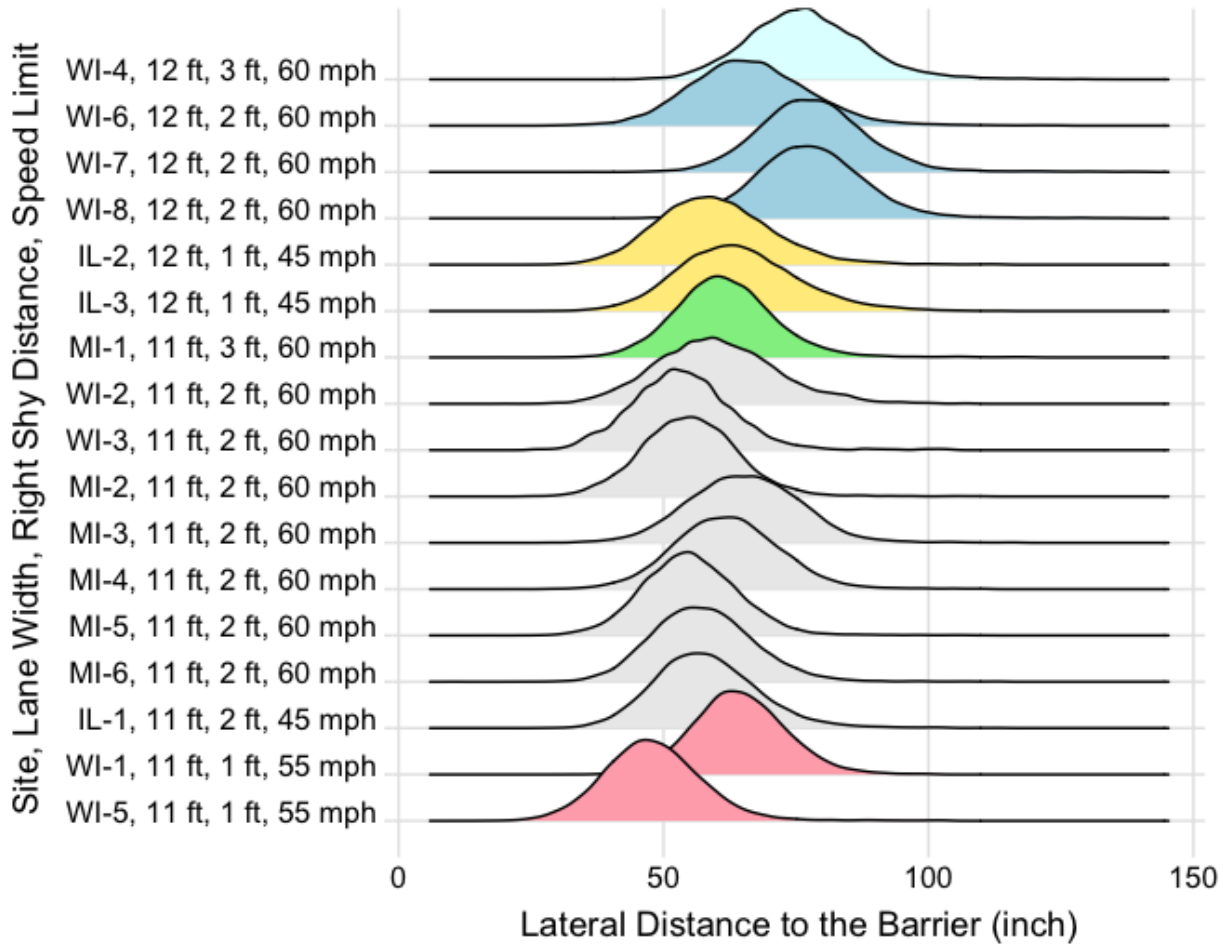


Figure 5-6. Empirical probability density function of lateral distance to the barrier

5.2.3 Modeling Methodology & Analysis

The safety analysis and modeling methodology consists of three components (Figure 5-7). First, all the observations, represented by the average lateral distance to the edge line/barrier were modeled. Second, tail events, represented by the lowest one percentile of the lateral distance observations were modeled. Last, the probabilities of edge line encroachment and barrier contact were modeled. All events and tail events were modeled using the standard linear regression approach. EVT was used to model the probabilities of edge line encroachment and barrier contact.

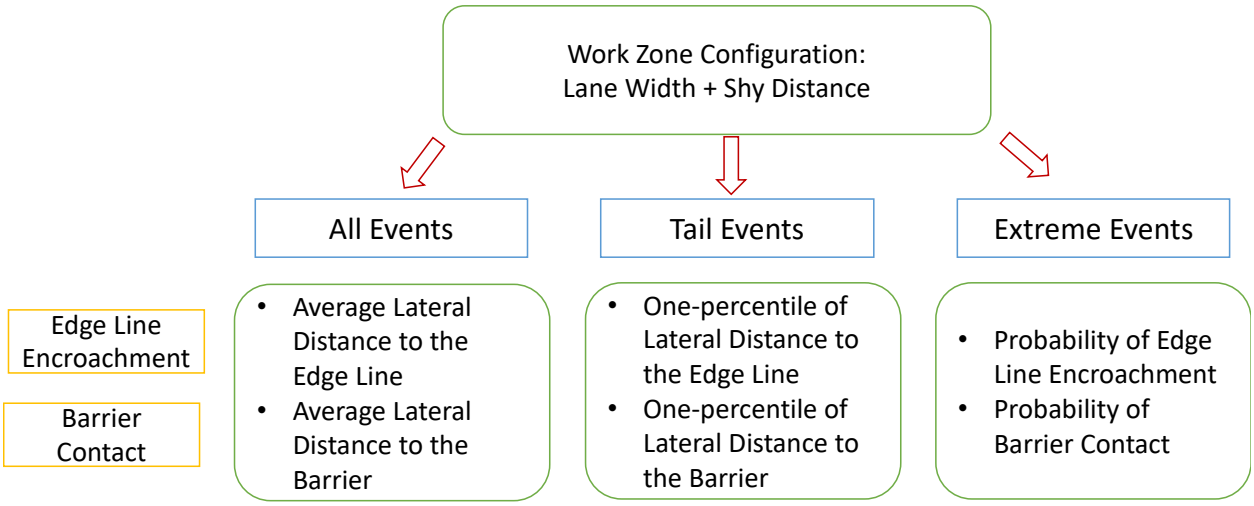
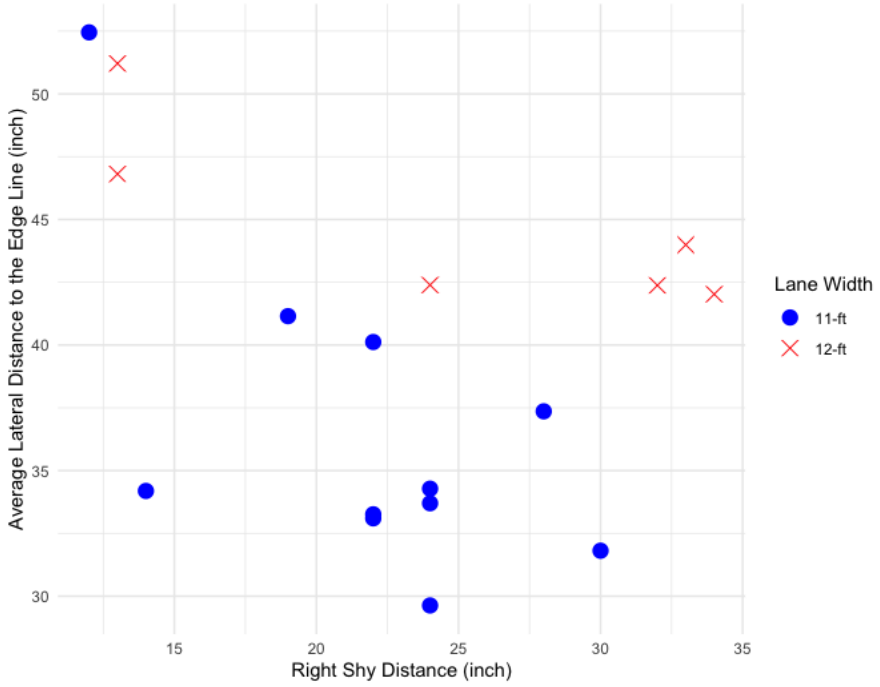


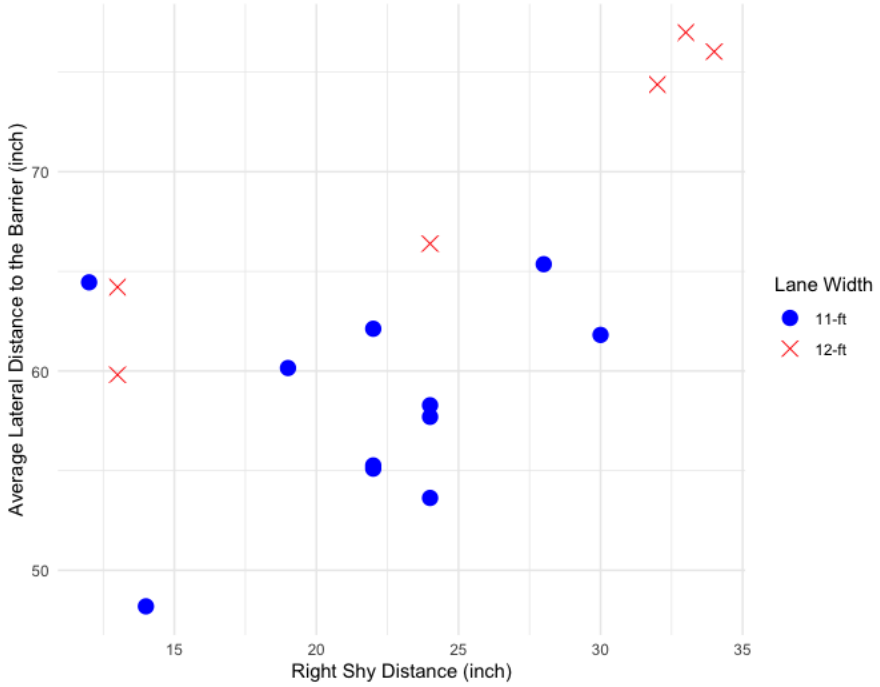
Figure 5-7. Methodology of safety analysis

Modeling of All Events

For all events, average lateral distance to the edge line and average lateral distance to the barrier for different work zone configurations were plotted first. The plots in Figure 5-8 illustrate that vehicles on wider lanes tend to move further towards the edge line and the barrier. Conversely, vehicles on greater shy distance tend to move closer to the edge line but further from the barrier. These observations were further validated using linear regression models, as presented in Table 5-4. In these models, lateral distance serves as the dependent variable, while lane width and shy distance act as independent variables. Both models demonstrate good performance, with lane width and shy distance showing statistical significance (p-value < 0.01). Figure 5-9 illustrates the effects of lane width and shy distance based on the regression model. Vehicles on wider lanes (12 ft compared to 11 ft) tend to drift further towards the edge line and the barrier. Conversely, vehicles on greater shy distances (3 ft compared to 2 ft and 1 ft) tend to gravitate closer to the edge line but farther from the barrier.



(a) Observed average lateral distance to the edge line

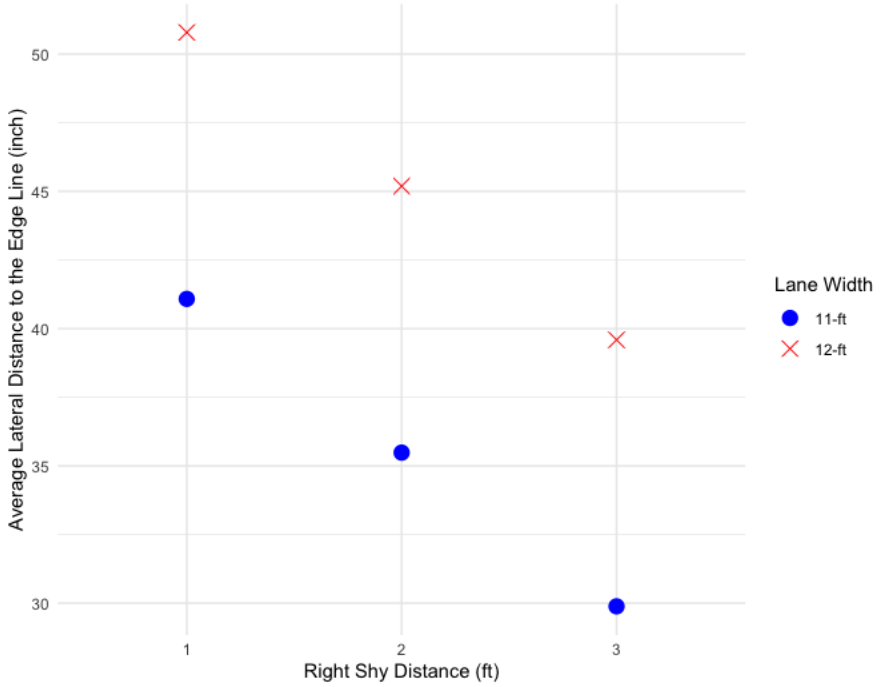


(b) Observed average lateral distance to the barrier

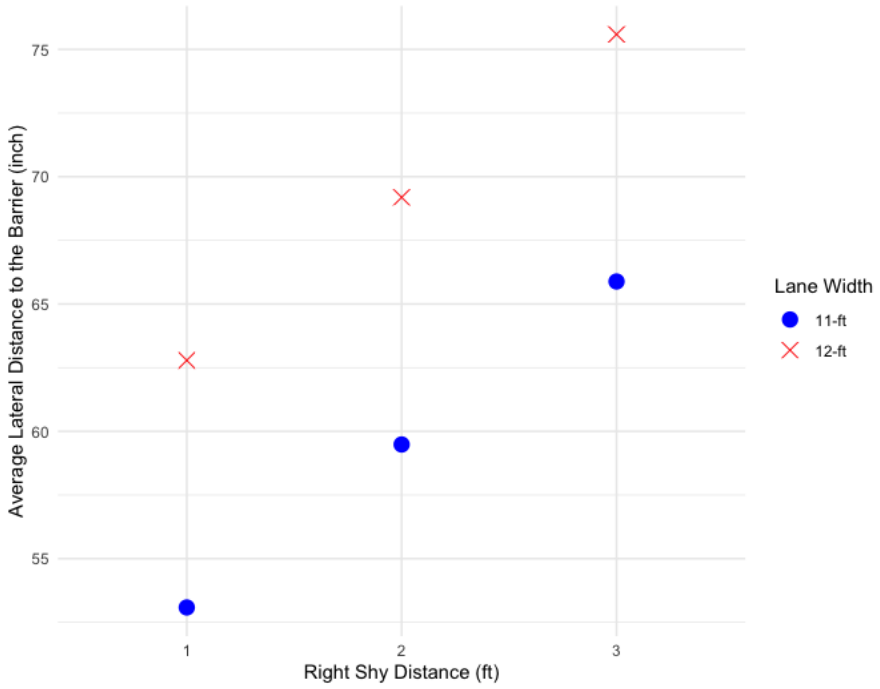
Figure 5-8. Scatter plots of observed average lateral distances

Table 5-4. Regression analysis of average lane position

Dependent Variable	**Average Lateral Distance to Edge line**			
Description	**Coefficients**	**Std. Error**	**t Stat**	**P-value**
Intercept	-60.08	26.63	-2.26	0.04
Lane Width	9.71	2.39	4.06	< 0.01
Shy distance	-0.47	0.17	-2.81	< 0.01
R-square	0.6			
P-value	< 0.001			
Dependent Variable	**Average Lateral Distance to Barrier**			
Description	**Coefficients**	**Std. Error**	**t Stat**	**P-value**
Intercept	-60.08	26.63	-2.26	0.04
Lane Width	9.71	2.39	4.06	< 0.01
Shy distance	0.53	0.17	3.21	< 0.01
R-square	0.71			
P-value	< 0.001			



(a) Average lateral distance to edge line

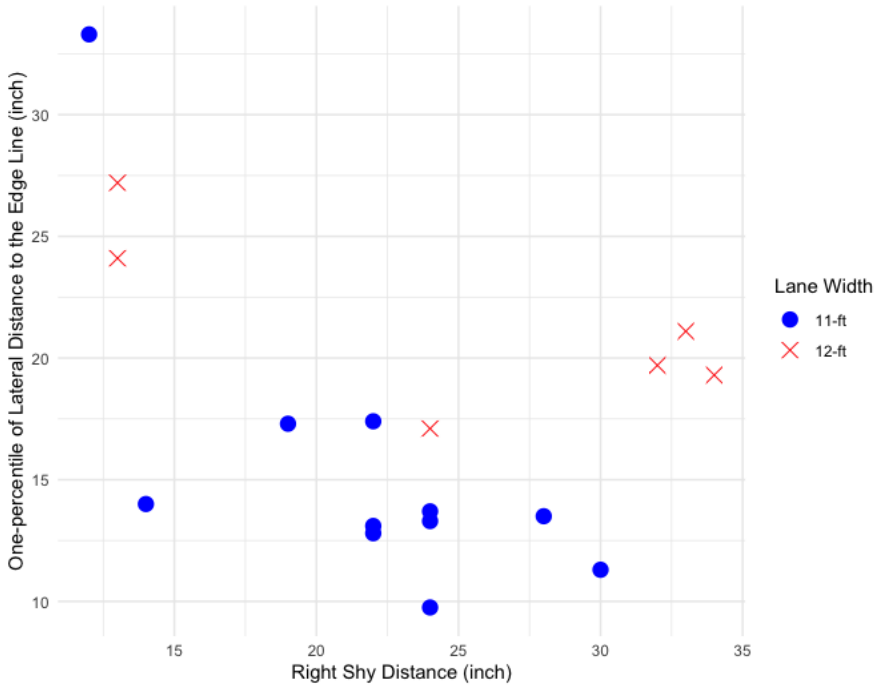


(b) Average lateral distance to barrier

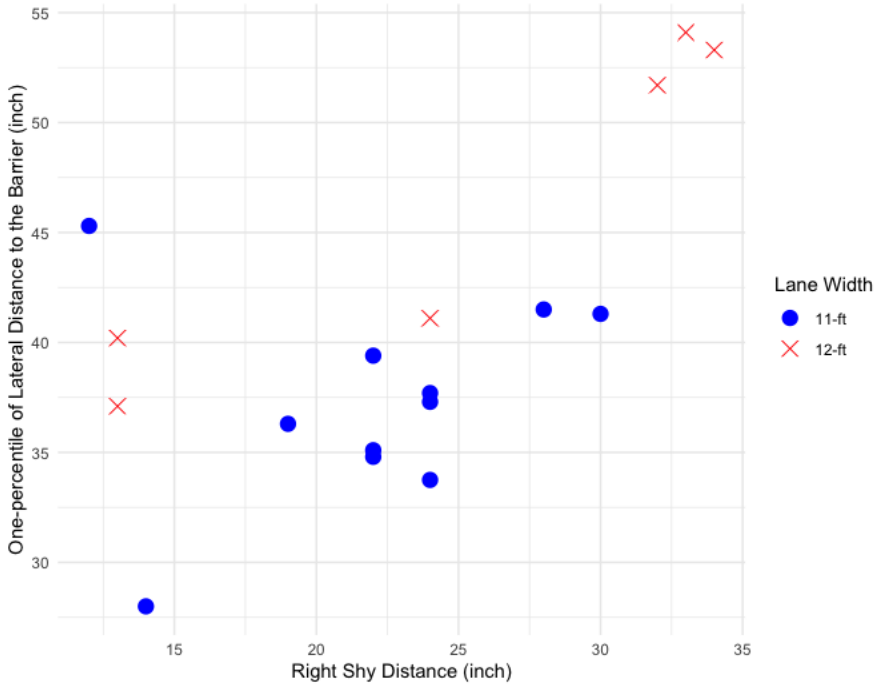
Figure 5-9. Impact of lane width and shy distance on average lateral distances

Modelling of Tail Events

Tail events are represented by the lowest one percentile of the lateral distance observations in the study. They are similar to risky events shown in Figure 5-1, which are more likely to become extreme events (edge line encroachment or barrier contact). The one-percentile lateral distance to the edge line and one-percentile lateral distance to the barrier for different work zone configurations are plotted in Figure 5-10. The plots indicate that the one-percentile of vehicles on wider lanes tends to move farther from the edge line and the barrier. The one-percentile of vehicles on wider shoulders tends to move closer to the edge line but farther from the barrier. These observations were corroborated using linear regression models, as depicted in Table 5-5. Both models demonstrate good performance, with lane width and shy distance showing statistical significance (p-value < 0.01).



(a) **One-percentile of lateral distance to the edge line**



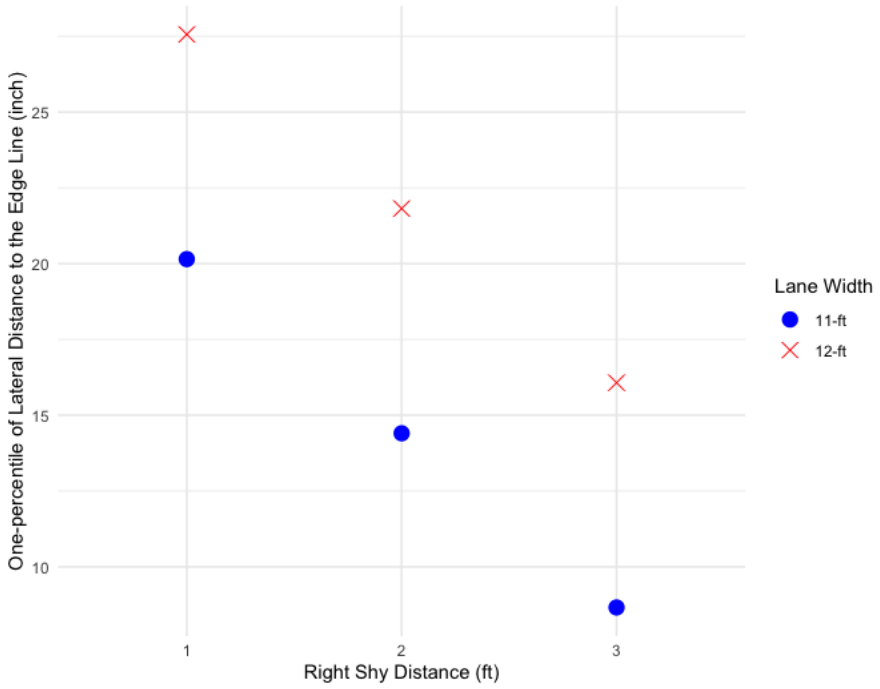
(b) One-percentile of lateral distance to the barrier

Figure 5-10. Scatter plots of observed one-percentile lateral distances

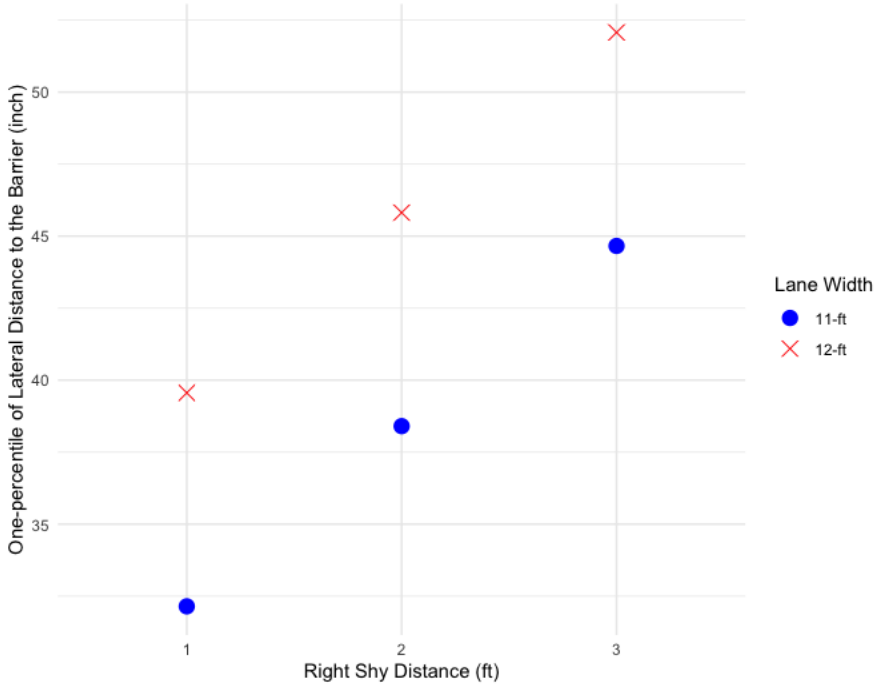
Table 5-5. Regression analysis of tail lane position

Dependent Variable	**One Percentile of Lateral Distance to Edge line**			
Description	**Coefficients**	**Std. Error**	**t Stat**	**P-value**
Intercept	-55.64	26.42	-2.11	0.05
Lane Width	7.41	2.37	3.13	<0.01
Shy distance	-0.48	0.16	-2.9	<0.01
R-square	0.52			
P-value	<0.010			
Dependent Variable	**One Percentile of Lateral Distance to Barrier**			
Description	**Coefficients**	**Std. Error**	**t Stat**	**P-value**
Intercept	-55.64	26.42	-2.11	0.05
Lane Width	7.41	2.37	3.13	<0.01
Shy distance	0.52	0.16	3.16	<0.01
R-square	0.64			
P-value	< 0.001			

Figure 5-11 shows the effects of lane width and shy distance based on the regression models for tail events. Specifically, the one-percentile of vehicles on wider lanes (12 ft compared to 11 ft) tend to move farther from the edge line and the barrier. The one-percentile of vehicles on wider shy distances (3ft compared to 2 ft and 1 ft) tend to move closer to the edge line but farther from the barrier.



(a) **One-percentile of lateral distance to the edge line**



(b) **One-percentile of lateral distance to the barrier**

Figure 5-11. Impact of lane width and shy distance on risky events.

Probability of Edge Line Encroachment and Barrier Contact

When modeling lateral distance, the focus was on extreme low values, so the negative values of lateral distance to the edge line was utilized in the POT. Optimal threshold (S_2) was identified to ensure statistical reliability and model validity. Two types of lane departure events were evaluated: edge line encroachment and barrier contact. For edge line encroachment events, where lateral distance is less than zero, S_3 was set to 0. For barrier contact events, where lateral distance is less than the negative of shy distance, S_3 was set to the negative of shy distance. The conditional probability of extreme events given risky events was estimated using the GP distribution. The probability of extreme events was calculated by multiplying the estimated conditional probability by the empirical probability of risky events (the number of exceedances over the number of total observations).

The study initially estimated the probability of extreme events using individual EVT models for each site. However, it became challenging to discern the safety impact of lane width and shy distance because many zeros appeared in the probability estimates (shown in Table 5-6).

Table 5-6 Results of individual EVT models for each site

Site	Count	Threshold (inch)	Number of exceedances	Neg. log likelihood	Probability of shoulder encroachment	Probability of barrier encroachment
10	20,043	28	32	55.56	8.15E-18	0
12	5,407	16	37	79.64	3.50E-05	2.07E-08
13	2,116	11	35	70.54	2.83E-04	3.91E-08
14	7,061	18	33	61.91	2.03E-12	0
26	35,352	7	30	59.59	4.99E-05	0
28	19,115	10	31	52.91	0	0
29	13,640	14	32	64.69	1.33E-05	1.38E-11
30	23,647	12	30	53.89	0	0
54	21,667	6	44	88.37	2.32E-04	1.92E-07
55	8,636	10	34	67.37	7.63E-06	0

56	15,388	8	37	68.74	6.75E-05	0
57	21,563	11	32	66.32	2.11E-05	0
58	14,054	10	47	84.80	8.87E-05	2.48E-06
59	12,103	9	30	73.00	3.09E-04	2.24E-05
68	21,989	9	32	62.70	0	0
69	11,935	21	40	72.73	0	0
70	19,553	22	35	66.93	0	0

The study then developed a comprehensive EVT model utilizing data from all sites to capture the impact of all the variables of interest. More exceedances were included compared to individual EVT models. This model is non-stationary and incorporates covariates in the scale parameter estimation. Previous studies have indicated that the shape parameter remains stationary, thus no covariates were included in its estimation [67].

The optimal threshold was determined first. In Figure 5-12 the mean residual life plot of the maximum negated thresholds is linear starting from a negated threshold of around −12 inch. The stability of GPD reparametrized scale and shape parameters were also analyzed (Figure 5-13). Then diagnostic plots were compared for several negated thresholds around -12 inch and -12 inch was selected for the best fit. Figure 5-14 shows the diagnostic plot for the stationary model using a negated threshold of -12 inch. The figures show that the modeled GPD with a threshold of u= 12 inch has satisfactory fitting results to the empirical data since the points fall close to the 45° line in the simulated QQ plot.

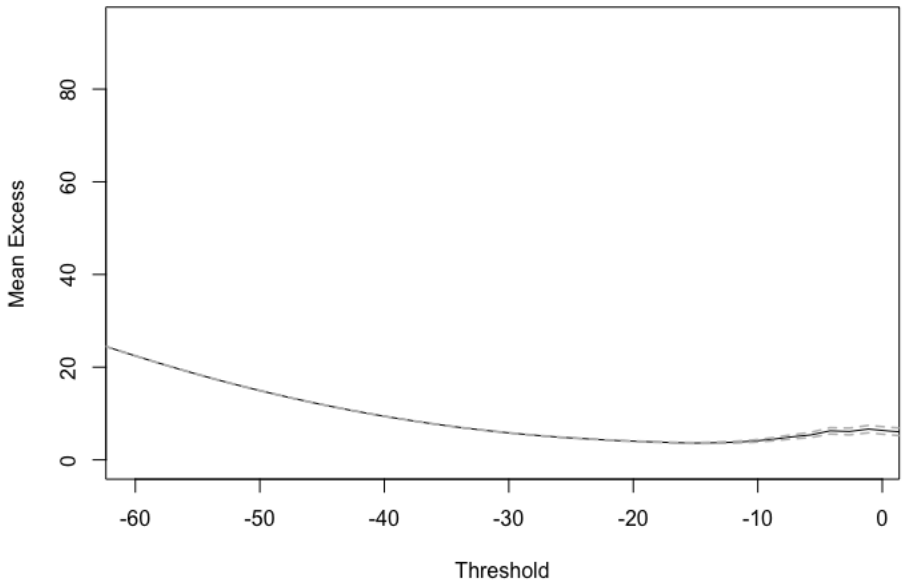


Figure 5-12 Mean residual life for the full data set.

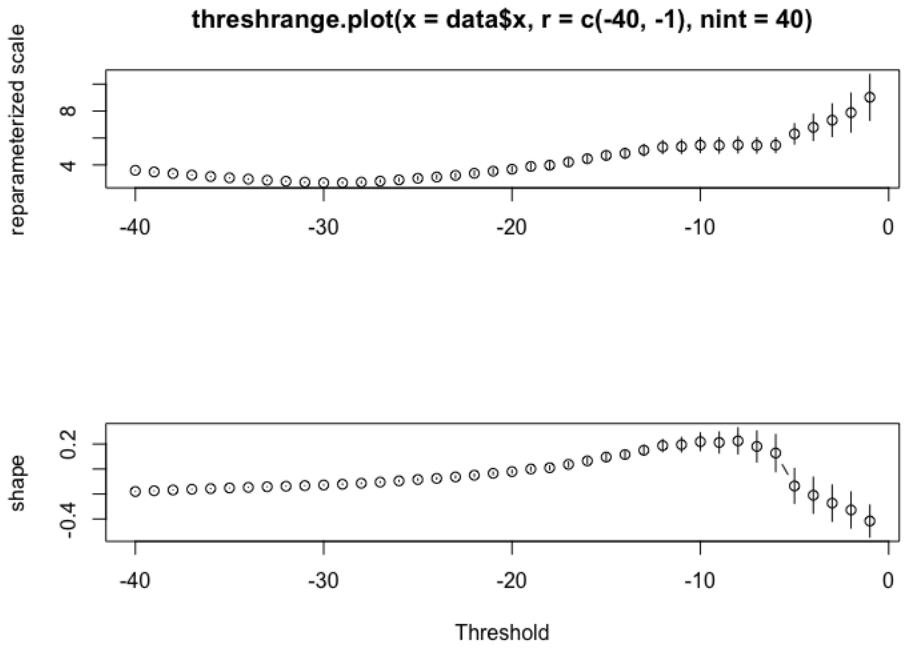


Figure 5-13 Stability plot for GPD model reparametrized (modified by subtracting the shape and multiplied by the threshold) scale parameter and shape parameter for different lateral distance thresholds.

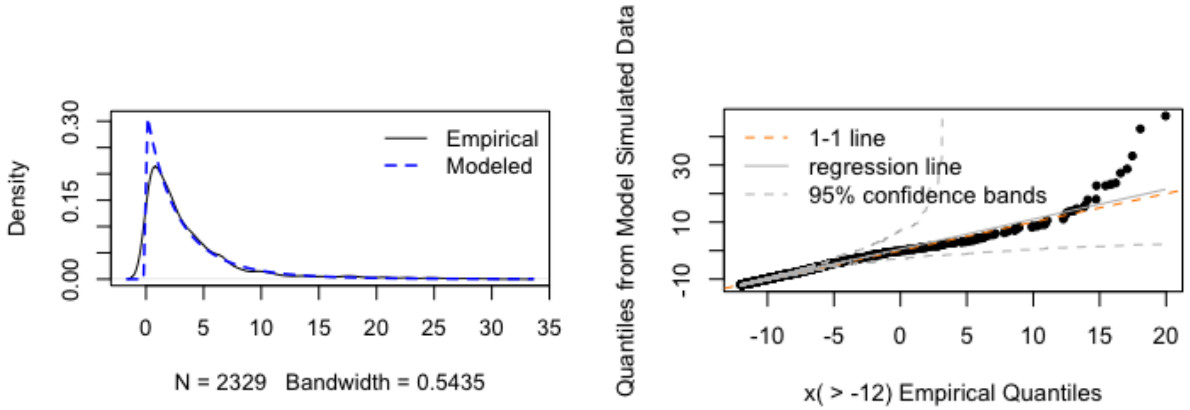


Figure 5-14 (Kernel) Probability density plot (left) and simulated QQ plot (right) for the stationary POT model.

Covariates used are location-level variables such as lane width, shy distance, road curvature, and speed limit, along with vehicle-level variables including time of day (day or night), free flow conditions, vehicle type (passenger car or truck), and adjacency to other vehicles. Various combinations of these variables were assessed using the likelihood ratio test to streamline model structures and variable inclusions. The non-stationary model, which incorporates a linear combination of lane width and shy distance in the scale parameter, exhibits significance with a small p-value of 0.02 in a likelihood ratio test compared to the stationary model. However, when additional terms such as the interaction between lane width and shy distance or other covariates like time of day are included in the linear combination, the model loses significance, as indicated by a larger p-value exceeding 0.1 in a likelihood ratio test when compared to the previous non-stationary model.

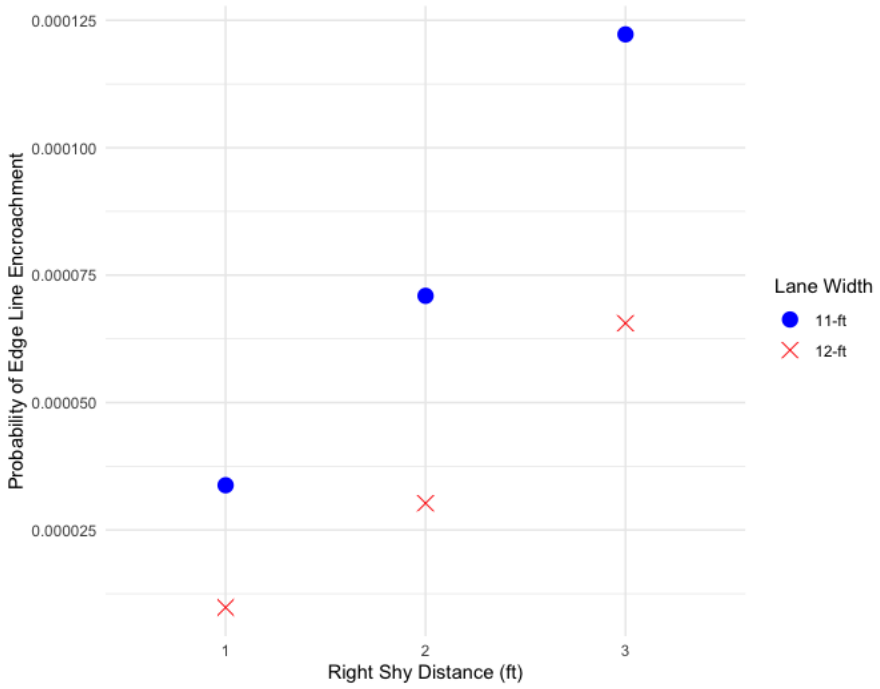
Consequently, the final model incorporates a linear combination of lane width and shy distance in the scale parameter. The optimal threshold of 12 inches results in 1,147 exceedances (out of a total of 273,269 vehicles) which were used to develop the EVT model. As depicted in Table 5-7, estimation results of the final model reveal a negative scale parameter for lane width, signifying

that an increase in lane width reduces the scale parameter and consequently, the variance of the lateral distance distribution. This reduction indicates a decrease in the probability of extreme events. Conversely, a positive scale parameter for shy distance indicates that an increase in shy distance augments the scale parameter and hence, the variance of the lateral distance distribution. This increase suggests a rise in the probability of extreme events.

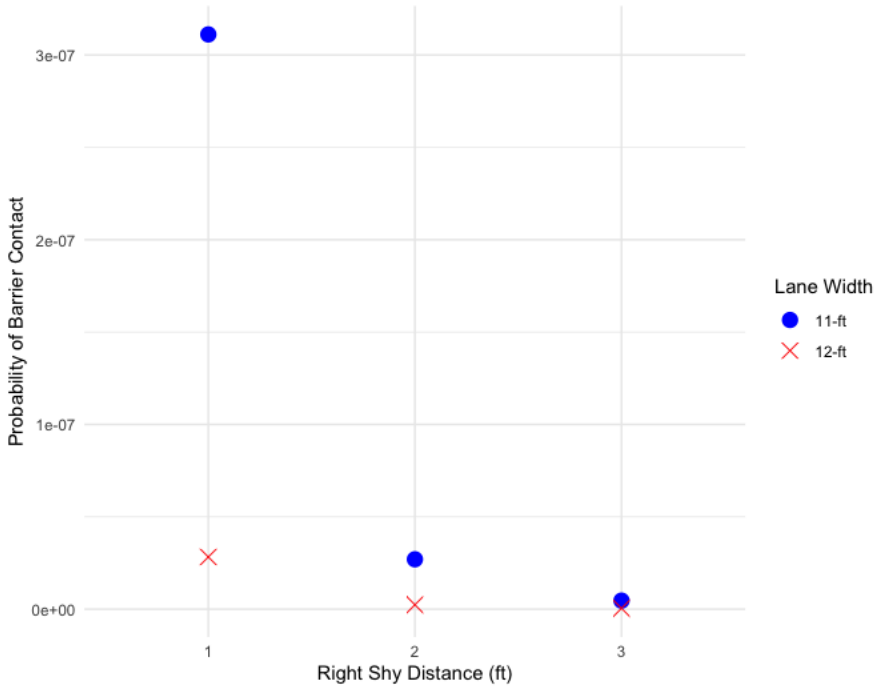
Table 5-7. Estimation results for the POT model

Description	**Estimated Value**	**Standard Error**	**P-value**
Shape Parameter, k	0.01	0.03	
Scale Parameter, σ_0	7.59	2.92	<0.01
Scale Parameter, σ_1 (lane width)	-0.51	0.27	0.06
Scale Parameter, σ_2 (shy distance)	0.04	0.01	<0.01
Negative loglikelihood	2,354.117		
Number of Observations	273,269		
Threshold (inch)	12		
Number of Exceedances/risky Events	1,147		

Figure 5-15 illustrates the impact of lane width and shy distance on the probability of edge line encroachment and barrier contact. An increase in lane width leads to a reduction in the probability of edge line encroachment and barrier contact. An increase in shy distance results in an increase in the probability of edge line encroachment and a decrease in the probability of barrier contact. For a 11-ft lane, the probability of barrier contact approaches zero for a 3-ft right shy distance. Similarly, for a 12-ft lane, the probability of barrier contact approaches zero for a 2-ft shy distance.



(a) **Probability of edge line encroachment**



(b) Probability of barrier contact

Figure 5-15. Impact of lane width and shy distance on extreme events.

5.2.4 Summary

In summary, the configuration of work zones, specifically lane width and shy distance, significantly influences vehicle lane departures. This impact was evaluated using vehicles' lateral position in the right lane as a surrogate measure for three levels of events: all events, tail events, and extreme events. The analysis focused on two types of lane departure: edge line encroachment and barrier contact. All events and tail events were modeled by linear regression of the average lateral distance of all vehicles and one-percentile lateral distance respectively. Extreme Value Theory (EVT) modeling was used to estimate the probabilities of edge line encroachment and barrier contact.

The results for both types of lane departure were consistent across all three levels of events. Narrower lanes were found to contribute to shoulder encroachment and barrier encroachment, while wider shy distances were associated with increased shoulder encroachment and reduced barrier encroachment. These findings align with our initial assumptions and highlight the importance of considering work zone configuration in mitigating lane departure incidents.

The analysis only focused on the vehicles' right departures in the right lane. Future research should collect lateral distance data from both sides and estimate lane departures in both directions for vehicles in both lanes. Also, few work zone sites had 1-ft shy distance for a short section (a few hundred feet), which might affect the findings.

Chapter 6 Safe Routing Application using Surrogate Safety Measures

This chapter introduces the development of a safe routing application that relies on surrogate safety measures. The crash risk for each road segment was previously estimated using the SPMD data in Chapter 5. However, due to the low penetration rate of connected vehicle technology and the specificity of crash types that can be recorded, constructing a safe routing application solely on this basis is not feasible.

To overcome the data limitations, the study was faced with two options: either to expand the collection of in-vehicle data or to devise an algorithm capable of enabling safe routing without relying on such data. The author chose to pursue the latter approach. As a result, this chapter details the process of creating a safe routing application that operates independently of in-vehicle data collection. This entails the innovative use of available data and surrogate safety measures to estimate crash risks, providing drivers with safer route options even when direct vehicle-derived safety data is unavailable.

A safety index or a safety hazard index for routing should be easy to implement and be able to handle data limitations. Therefore, a novel safety index was proposed (Figure 6-1) which required only road traffic density data and speed data.

Figure 6-1 Framework of the proposed methods.

Among all of the safety surrogate measures discussed, DRAC has the minimum deceleration rate required to avoid a collision with the leading vehicle, and also has the maximum available deceleration rate (MADR) that a vehicle can adopt [68]. It's assumed that a crash would happen if DRAC exceeds MADR. MADR can be determined by multiple factors, including vehicle type (vehicle weight), pavement condition (dry/wet), and so on [68]. A safety hazard index can be expressed as the ratio of DRAC to MADR; a lower value indicates a safer road. Two kinds of data are required to determine this index: the existing road information such as link traffic density and average link speed, and the traveler information which allows the index to vary by user. Respective indexes are used for links and intersections based on crash mechanism. An index is proposed for different turning movements at an intersection (left turn, through turn, and right turn). The remainder of this section describes the indexes for links and intersections.

6.1 Safety Hazard Index for a Roadway Link

The most common cause of a rear-end crash, the sudden braking of the lead vehicle, was considered for the purpose of simplicity. Suppose the subject vehicle A and its preceding vehicle B operate in the same lane with the speed of v_A and v_B respectively. Suddenly, the leading vehicle B stops, and, its speed becomes zero. Suppose D_L is the expected closest distance between the two consecutive vehicles, Δt is the perception-reaction time for the driver in following vehicle (vehicle A). Human behaviors such as driver inattention, impairment and distraction are not considered. Then DRAC for vehicle A can be formulated in Equation 6-1.

$$DRAC_A = \frac{{v_A}^2}{2(D_L - v_A \Delta t)} = \frac{v^2}{2(D_L - v\Delta t)} \qquad \text{Equation 6-1}$$

where, v_A can be estimated by the average link speed v.

The index is calculated as the ratio of $DRAC_A$ to $MADR_A$. If the subject vehicle A is the lead vehicle, the DRAC for vehicle B is exactly the same as it is for vehicle A; however, $MADR_B$ may be different. Therefore, the larger value of the two considered as the safety hazard index for the link.

In Equation 6-1, the expected closest distance between the two successive vehicles D_L is unknown. Two traffic density scenarios (sparse and dense) are proposed to calculate this distance [69]. The density of 50 vehicle/mile/lane was used to distinguish the two traffic states. The number of arrival vehicles along the link in a sparse traffic state is assumed to follow a Poisson distribution [70]; thus, the gap between two consecutive vehicles follows a negative exponential distribution which is expressed in Equation 6-2.

$$f(s) = e^{-ds} \qquad \text{Equation 6-2}$$

where, *s* is the distance between two successive vehicles on the same lane, and *d* is the vehicle density along the link.

The expected closest distance D_L is calculated as Equation 6-3.

$$D_L = \int_{s=0}^{s=l} f(s)\, ds = \int_{s=0}^{s=l} e^{-ds}\, ds = \frac{1}{d}(1 - e^{-dl}) \qquad \text{Equation 6-3}$$

where, *l* is the link length.

The distance between two consecutive vehicles in a dense traffic state is assumed to follow the Gaussian Unitary Ensemble (GUE) distribution [71], which is shown in Equation 6-4.

$$f(s) = \frac{32s^2d^3}{\pi^2} e^{-\frac{4s^2d^2}{\pi}} \qquad \text{Equation 6-4}$$

The expected closest distance D_L is defined in Equation 6-5.

$$D_L = \int_{s=0}^{s=l} sf(s)\,ds = \int_{s=0}^{s=l} \frac{32s^3d^3}{\pi^2} e^{-\frac{4s^2d^2}{\pi}} ds = \frac{1}{d}\left(1 - \frac{4l^2d^2}{\pi} e^{-\frac{4l^2d^2}{\pi}} - e^{-\frac{4l^2d^2}{\pi}}\right) \qquad \text{Equation 6-5}$$

The expression of the safety hazard index I_{link} for any link is presented in Equation 6-6.

$$I_{link} = \max\left\{\frac{DRAC_A}{MADR_A}, \frac{DRAC_B}{MADR_B}\right\} = \max\left\{\frac{v^2}{2(D_L - v\Delta t)}/MADR_A, \frac{v^2}{2(D_L - v\Delta t)}/MADR_B\right\} \qquad \text{Equation 6-6}$$

where, $D_L = \begin{cases} \frac{1}{d}(1 - e^{-dl}), \text{for sparse traffic condition} \\ \frac{1}{d}\left(1 - \frac{4l^2d^2}{\pi} e^{-\frac{4l^2d^2}{\pi}} - e^{-\frac{4l^2d^2}{\pi}}\right), \text{for dense traffic condition} \end{cases}$

6.2 Safety Hazard Index at an Intersection

The intersection-related crash is more complicated than the link-based crash, as there are several types of intersections (uncontrolled, stop control, signal control) and collisions can occur in various situations (stop control violation, signal violation, conflicts between left-turn flow and approaching flow, etc.). It is not practical, due to data limitations, to develop indexes for different intersections. Figure 6-2 presents a typical angle collision process irrespective of intersection types. For simplicity, only angle collision was discussed.

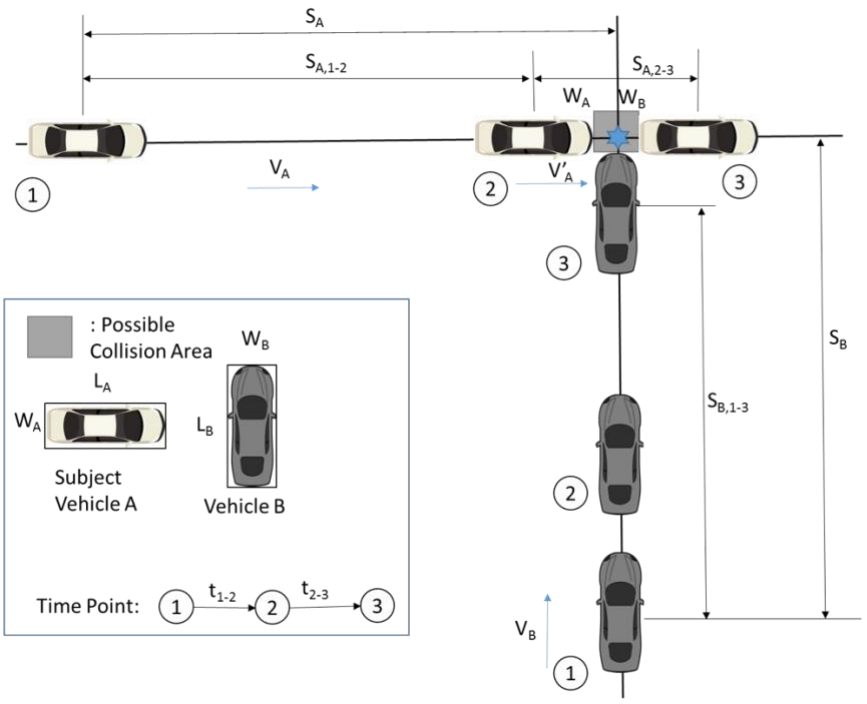


Figure 6-2 Vehicle collision modeling at an intersection.

Suppose the subject vehicle A and another vehicle B which is the closest to A around the intersection on the other leg approach with the speed of v_A and v_B, respectively. Two crash scenarios are considered: 1) vehicle B decelerates but fails to avoid colliding with vehicle A, and 2) vehicle A decelerates but still collides with vehicle B. In Figure 6-2, measured from time point 1, the initial distance from vehicle A and vehicle B to the intersection are S_A and S_B. The sum of S_A and S_B is D_I, defined as the expected closest distance between two vehicles around the intersection. Vehicle B is assumed to decelerate while vehicle A keeps the same speed before reaching the intersection or the possible collision area (time point 2). The possible collision area is determined by the length and width of the approaching vehicles. To obtain DRAC, a critical time point when vehicle A just leaves the collision zone and vehicle B enters the area is proposed (time point 3). Time interval between time point 1 and 2 is expressed in Equation 6-7.

$$t_{1-2} = \frac{S_{A,1-2}}{v_A} = \frac{S_A - \frac{L_A + W_B}{2}}{v_A} \quad \text{Equation 6-7}$$

where, $s_{A,1-2}$ is the distance A travels from time point 1 to 2.

When vehicle A enters the intersection, the speed may change (noted as v'_A). For a left turn movement, vehicle's speed will significantly decrease. The time interval between time point 2 and 3 is expressed in Equation 6-8.

$$t_{2-3} = \frac{s_{A,2-3}}{v'_A} = \frac{L_A+W_B}{v'_A} \quad \text{Equation 6-8}$$

where, $s_{A,2-3}$ is the distance A travels from time point 2 to 3.

Under the critical situation that vehicle B doesn't collide with vehicle A, the distance vehicle B travels from the start of deceleration to the time vehicle A leaves the collision zone (from time point 1 to 3) is expressed in Equation 6-9.

$$S_{B,1-3} = v_B\Delta t + v_B(t_{1-3} - \Delta t) - \frac{1}{2}DRAC_B(t_{1-3} - \Delta t)^2 = v_B t_{1-3} - \frac{1}{2}DRAC_B(t_{1-3} - \Delta t)^2$$

Equation 6-9

where, $t_{1-3} = t_{1-2} + t_{2-3} = \frac{S_A - \frac{L_A+W_B}{2}}{v_A} + \frac{L_A+W_B}{v'_A}$ is the time interval between time point 1 and 3, and Δt is the perception-reaction time of the driver to decelerate.

The minimum value of S_A is $\frac{L_A+W_B}{2}$, or t_{1-2} is nonnegative in Equation 6-7. Thus,

$$0 < S_{B,1-3} \leq D_I - \min(S_A) - \frac{L_B+W_A}{2} = D_I - \frac{L_A+W_B}{2} - \frac{L_B+W_A}{2} = D_I - \frac{L_A+W_B+L_B+W_A}{2}$$

Equation 6-10

Given the assumption that vehicles are randomly distributed along the street, the expected value of $S_{B,1-3}$ equals to $\frac{1}{2}(D_I - \frac{L_A+W_B+L_B+W_A}{2})$. Replace $S_{B,1-3}$ in Equation 6-9 and DRAC for vehicle B can be expressed in Equation 6-11.

$$DRAC_B = \frac{2v_B t_{1-3} - (D_I - \frac{\sum L}{2})}{(t_{1-3} - \Delta t)^2} \quad \text{Equation 6-11}$$

where, $t_{1-3} = \frac{\frac{1}{2}(D_I - \frac{3L_A+3W_B-L_B-W_A}{2})}{v_A} + \frac{L_A+W_B}{v'_A}$ and $\sum L = L_A + W_B + L_B + W_A$.

If vehicle A decelerates to avoid the collision, DRAC for vehicle A can be calculated by exchanging A and B in Equation 6-11, as shown in Equation 6-12.

$$DRAC_A = \frac{2v_A t_{1-3}' - (D_I - \frac{\sum L}{2})}{(t_{1-3}' - \Delta t)^2} \quad \text{Equation 6-12}$$

where, $t_{1-3}' = \frac{\frac{1}{2}(D_I - \frac{3L_B+3W_A-L_A-W_B}{2})}{v_B} + \frac{L_B+W_A}{v'_B}$, v'_B is the speed of B around the intersection.

In both Equations 6-11 and 6-12, D_I, the expected closest distance between two vehicles around the intersection is unknown. According to Chandra, a similar assumption can be made that vehicles around the intersection are uniform randomly distributed under light traffic conditions [69], which means the probability of n vehicles around the intersection is

$$P(N = n) = \frac{(ds)^n e^{-ds}}{n!} \quad \text{Equation 6-13}$$

where, N is the random number of vehicles around the intersection, s is the total link-based distance around the intersection (sum of the lengths of the streets converging to the intersection), and d is the vehicle density around the intersection (number of vehicles per unit length).

The probability that subject vehicle A misses a vehicle, within the total distance of s around the intersection is the probability when n=0, namely $P(N=0)=e^{-ds}$. Then the expected closest distance around the intersection D_I is calculated as,

$$D_I=\int_{s=0}^{s=L}P(N=0)\,ds=\int_{s=0}^{s=L}e^{-ds}\,ds=\frac{1}{d}(1-e^{-dL}) \quad \text{Equation 6-14}$$

where, L approximately equals to the summation of the length of all street legs around the intersection ($L\approx\sum l$).

For dense traffic conditions, the link-based distance between two consecutive vehicles around the intersection is assumed to follow Gaussian Unitary Ensemble (GUE) distribution [69] and formulated in Equation 6-15,

$$f(s)=\frac{32s^2d^3}{\pi^2}e^{-\frac{4s^2d^2}{\pi}} \quad \text{Equation 6-15}$$

The expected closest distance D_I is expressed in Equation 6-16.

$$D_I=\int_{s=0}^{s=L}sf(s)\,ds=\int_{s=0}^{s=L}\frac{32s^3d^3}{\pi^2}e^{-\frac{4s^2d^2}{\pi}}\,ds=\frac{1}{d}(1-\frac{4L^2d^2}{\pi}e^{-\frac{4L^2d^2}{\pi}}-e^{-\frac{4L^2d^2}{\pi}}) \quad \text{Equation 6-16}$$

In summary, the safety index for each intersection is expressed as Equation 6-17.

$$I_{\text{intersection}}=\max\{\frac{2v_At_{1-3}{}'-(D_I-\frac{\sum L}{2})}{(t_{1-3}{}'-\Delta t)^2}/MADR_A,\frac{2v_Bt_{1-3}-(D_I-\frac{\sum L}{2})}{(t_{1-3}-\Delta t)^2}/MADR_B\} \quad \text{Equation 6-17}$$

where, $D_I=\begin{cases}\frac{1}{d}(1-e^{-dL}),\text{for sparse traffic condition}\\ \frac{1}{d}(1-\frac{4L^2d^2}{\pi}e^{-\frac{4L^2d^2}{\pi}}-e^{-\frac{4L^2d^2}{\pi}}),\text{for dense traffic condition}\end{cases}$,

$t_{1-3}{}'=\frac{\frac{1}{2}(D_I-\frac{3L_B+3W_A-L_A-W_B}{2})}{v_B}+\frac{L_B+W_A}{v'_B}$,

$$t_{1-3} = \frac{\frac{1}{2}(D_I - \frac{3L_A + 3W_B - L_B - W_A}{2})}{v_A} + \frac{L_A + W_B}{v'_A},$$

and $\sum L = L_A + W_B + L_B + W_A$.

Safety index for right turn movement at the intersection is assumed to be zero because the right-turning vehicle is unlikely to be involved in an angle collision. The difference for through and left-turn movement can be simplified to be the speed for the subject vehicle when entering the intersection. For a through movement, it's assumed that $v'_A = v_A$. When the vehicle makes a left turn, it has to slow down and 15mph is assumed to be the speed. Note that turning movement is only considered for the subject vehicle A. For vehicle B, v'_B.is assumed to be v_B.

6.3 A Review of Safety Hazard Index

Default values for the user-customized variables are considered when applying the safety index. The perception-reaction time uses 2 seconds as the average value for older drivers (>51 years old) and 1 second for younger drivers [72]. According to the Green Book [73], the design vehicle dimension for a passenger car is 19 feet ×7 feet, and 30 feet × 8 feet for a single-unit truck. Mean MADR values are adopted from previous research: 8.45 m/s^2 for a passenger car on dry pavement, 6.82 m/s^2 for a passenger car on wet pavement, 6.34 m/s^2 for a truck on dry pavement, and 5.12 m/s^2 for a truck on wet pavement [74].

The proposed index was reviewed through qualitatively assessing the relationship between the modeled safety index and the variables in Table 6-1. An increase in vehicle density increases the safety index due to the increased DRAC for both links and intersections. The relationship between crash rates and traffic density is explained in the previous literature [75]. Previous studies have also proven that vehicle speed increases accident probability [76], which is also

shown in the index. Intersection speed, also proposed in this index, differentiates the speed for different turning movement. Although no literature focuses on this speed, the assumption is that higher speed results in less time spent in the intersection, therefore reducing the probability of a crash. Perception-reaction time has been shown to be closely related to the minimum stopping sight distance [77]. Higher perception-reaction time leads to longer stopping sight distance, which increases the crash rate; this coincides with the proposed index. The longer the length of a truck, the more time it takes to cross the intersection, hence increasing the chance of a collision. Furthermore, it is obvious that the deceleration rate would be lower for trucks and lower on wet pavement, which could also contribute to crash occurrence. Overall, most variables in the proposed model can be validated through previous studies, which means it is a sufficient index for safety assessment.

Table 6-1 Relationship Between Indexes and Variables

Variable		**Safety Hazard Index**				
		MADR	**Link**		**Intersection**	
			DRAC	**Safety Index**	**DRAC**	**Safety Index**
Density (Increase)		-	↑	↑	↑	↑
Speed (Increase)	Link Speed (Increase)	-	↑	↑	↑	↑
	Intersection Speed (left-turn to though movement: Increase)	-	↓	↓	↓	↓
Age (Young to Old)	Perception-reaction time (Increase)	-	↑	↑	↑	↑
Vehicle Type (Passenger Car to Truck)	Length (Increase)	-	-	-	↑	↑
	MADR (Decrease)	↓	-	↑	-	↑
Pavement Condition (Dry to Wet)	MADR (Decrease)	↓	-	↑	-	↑

NOTE: "-" = constant, "↑" = increase, and "↓" = decrease.

6.4 Case Study

The proposed method was tested in a case study. A multi-objective shortest path model (which can also consider safety) was utilized instead of a typical route-finding model, which focuses only on the route with the shortest time.

The first objective was to find several routes with shorter travel time. The shortest route-finding tool in ArcGIS can achieve this goal [78]. The second objective was to identify a non-inferior solution by balancing both travel time and safety. Most studies summed up the safety index of all links and intersections along one route in order to obtain safety performance information and then select a route based on the safety performance. Besides considering the overall safety performance of the route, the links and intersections with a high safety hazard index were avoided in order to identify the preferred route in the proposed application.

The proposed route-finding method was tested on a street network near the University of Wisconsin-Milwaukee campus. The Wisconsin Information System for Local Roads (WISLR) GIS map provided information on link traffic volume, speed, and link length for the routing analysis. The speed limit was estimated based on the roadway functional class, and the average link speed was substituted by randomly assigning ± 5 mph to the speed limit. Link vehicle density was calculated using the average daily traffic. One origin and one destination (university campus) were selected. Five candidate routes with short travel time were chosen with ArcGIS (see Figure 6-3).

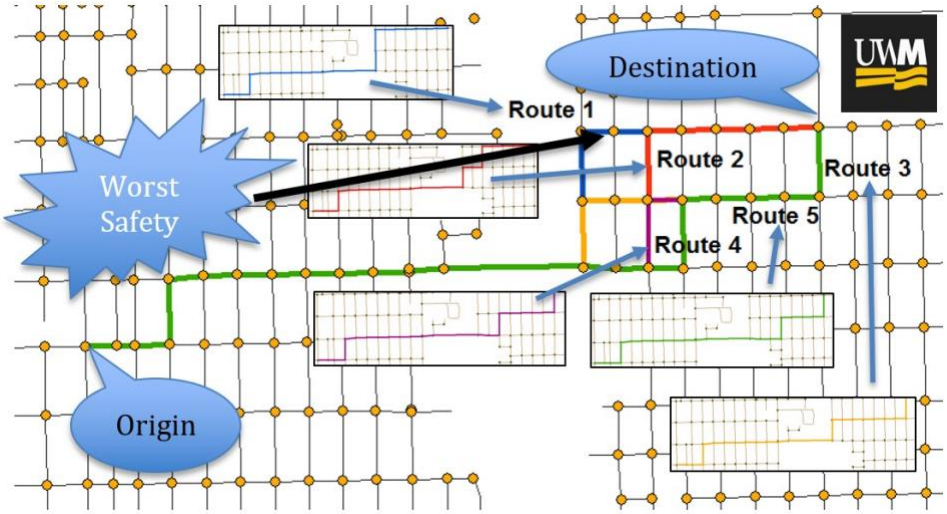


Figure 6-3 Alternative routes.

Route 5 was chosen based on it having the lowest travel time. The safety index was also calculated and compared along these routes. The safety index recommends non-inferior routes depending on the user-customized information. A lower value in the safety hazard index indicates a safer route. Three customized options, “truck”, “old”, and “wet”, which correspond to vehicle type, age, and pavement condition, were also applied. The default options of passenger car, young and dry, were presented for comparison. The overall safety index and the highest index (worst safety) were calculated for all customized options (see Table 6-2). Route 3 is recommended for truck drivers because it has the lowest overall safety index and avoids the most dangerous route (route 1). Route 3 is also the best choice for older drivers, wet pavement conditions and default customer information. Although travel times in routes 3 and 5 are almost equal, it is much safer to use route 3 because the summation of all options in the safety index decreases a great deal. Aside from choosing the safest route, the links or intersections with the worst safety (highest safety index) were also identified. Table 6-2 shows the dangerous link/node that exists in route 1 for all four customized options.

Table 6-2 Comparison Among Different Routes

Route	Safety Index					Travel Time (s)
	Option / Index	Truck	Old	Wet	Default	
1	Sum	11.67	13.61	12.19	9.65	234.52
	Highest	1.81	2.03	1.95	1.46	
2	Sum	9.9	12.73	10.31	8.24	236.73
	Highest	0.65	2.03	0.7	0.53	
3	Sum	8.84	10.72	9.15	7.51	232.82
	Highest	0.65	2.03	0.7	0.53	
4	Sum	9.17	11.34	9.51	7.84	238.54
	Highest	0.65	2.03	0.7	0.53	
5	Sum	9.75	12.18	10.12	8.32	232.63
	Highest	0.65	2.03	0.7	0.53	

6.5 Summary

This study developed a method for incorporating safety into the pathfinding process by developing and applying a novel safety index to the shortest route-finding algorithm. The ratio of DRAC to MADR was adopted as the safety hazard index, in which a lower value indicates a safer condition. The index was formulated on the basis of the collision mechanism along the roadway link and at the intersection. Besides the required roadway information (e.g., link speed, vehicle density), user-specific information (e.g., vehicle type, age, and pavement condition) can be included in the safety index model. A qualitative review of the index, which considers the findings from previous literature, supports the index as a sufficient proxy for traffic safety. A real roadway network was used to apply the proposed safety index. Three objectives were established in the search for the safest and shortest route: shorter travel time, lower overall index, and avoidance of the highest index. Safe routes with different user-customized information were obtained in the application example.

Model assumptions and limitations should be taken into consideration when this index model is applied. The model deals with typical intersections and urban roadways, and not all collision types were considered in developing the index (only rear-end collision and angle collisions were considered). These assumptions were made when the shortest distance between two successive vehicles was acquired, but this result may not always correspond with real-world situations. Future work will expand use of the model to other types of collisions and crash severities and will develop solutions for the multi-objective shortest-path problem. A systematic safe pathfinding method that considers all types of crashes in the safety index is proposed to be built for practical navigation applications in a real-world road network. At the same time, with increased penetration rate of connected vehicle and development of data collection techniques, real-time safety index can be a better option using the calculated surrogate safety measures from the collected data.

Chapter 7 Conclusions & Future Research

This chapter summarizes the contributions of the dissertation and presents future research directions.

7.1 Summary & Conclusions

The dissertation demonstrates a comprehensive analysis, evaluation, and application of surrogate safety measures by utilizing the SPMD connected vehicle data and vehicles' lateral distance data. It has three parts with five studies: calculation and validation of surrogate safety measures (Study 1 & Study 2), evaluating road safety using surrogate safety measures (Study 3 & Study 4), and safety route application using safety surrogate measures (Study 5).

In Study 1, SPMD data was harnessed to evaluate mid-block rear-end collisions, achieving three objectives: establishing a method to process large-scale SPMD data, creating surrogate safety measures from this data, and linking these measures to crash records to validate their effectiveness. This study diverged from prior research by focusing on evaluating road safety using link-level indices, rather than the conventional vehicle-level approach that measures vehicle distances. It used a range of indicators to summarize the surrogate safety measures, such as time-to-collision (TTC), without imposing arbitrary thresholds. Negative binomial (NB) models, which included variables like speed and braking duration from connected vehicles, showed that the Modified TTC (MTTC) had a better fit than other measures, demonstrating that surrogate measures could help to evaluate road safety when detailed road and traffic data are not available.

In Study 2, SPMD data was used to expand the first study by including major crashes at both link and intersection. The study introduced surrogate safety measures across three data levels: Level 1 focused on the subject vehicle, Level 2 incorporated information from the front vehicle, and

Level 3 included signal timing data. Detailed algorithms and steps for calculating these surrogate safety measures were proposed. The effectiveness of these methods was proved by comparing actual crash data with traffic conflicts derived from each of the three data levels. The findings suggest that the proposed algorithms offer a viable alternative to traditional crash data for road safety assessment.

In Study 3, the minimal Time-to-Collision (TTC) metric was used as a surrogate for assessing the probability of rear-end collisions at mid-block locations in the absence of actual crash data. An Extreme Value Theory (EVT) model was formulated to explore the correlation between this probability and various road and driver characteristics. The results of the EVT model revealed how factors such as traffic volume and driver aggressiveness influence road safety as reflected by minimal TTC values. Notably, an increase in Average Annual Daily Traffic (AADT) was found to potentially raise the likelihood of rear-end collisions at mid-blocks. On the other hand, higher median TTC values of each driver generally improves road safety, which corroborates the initial hypotheses of the study.

In Study 4, the lateral position of vehicles within the right travel lane was analyzed as an indicator of the safety implications of lane width and edge buffer zone, termed "shy distance." Data from over 250,000 vehicles were collected at 17 locations across three states, all featuring two-lane work zones with lane widths of either 11 or 12 feet, and shy distances ranging from 1 to 3 feet. The safety analysis was focused on vehicles' right departure in the right lane, as the data collection devices were installed only on the right-side barriers. Linear regression was used to model the average and the one-percentile lateral distances, revealing that both lane width and shy distance had a significant effect on lateral positioning. Vehicles in 12-foot lane or with narrower shy distance maintained a greater distance from both the edge line and barrier. EVT modeling

was employed to predict the likelihood of vehicles encroaching on the edge line or contacting the barrier, finding that narrower lanes increase these risks, while wider shy distances lead to more edge line encroachments but fewer barrier contacts.

In an effort to integrate safety into the route selection process, Study 5 introduced a safety hazard index based on the ratio of Deceleration Rate to Avoid a Crash (DRAC) to Maximum Available Deceleration Rate (MADR), with a lower index value signifying safer conditions. This index was used to assess road safety along road links and intersections, incorporating both roadway data (such as link speed and vehicle density) and user-specific information (including vehicle type, age, and pavement condition). Supported by a qualitative review of previous studies, the index was validated as an effective measure for traffic safety. The model was tested on an actual road network, aiming to balance three main criteria in identifying the safest and shortest path: minimized travel time, a lower overall safety hazard index, and avoidance of the most hazardous location. The application of this safety index model successfully yielded safe routes that accommodated various user-specific conditions.

7.2 Limitations & Future Research

While this dissertation has yielded valuable insights, there are several limitations that should be addressed in future research.

One important direction is to improve the data collection and quality assurance processes. Researchers should explore the integration of additional sensor technologies, such as video cameras or LiDAR sensor, to validate and supplement the SPMD data. The use of these complementary data sources can help address the issues related to unreliable sensor measurements, leading to a more robust and comprehensive dataset. Developing methods to

enhance the accuracy and completeness of the data, including techniques to handle missing or erroneous information, will be crucial for providing a solid foundation for the safety analysis.

Expanding the scope of the safety measures and analysis is another key area for future research using the SPMD data. While the current studies mainly focused on rear-end and lane-departure crashes, future work should analyze other types of collision using other safety surrogate measures. Additionally, the network-level safety analysis should be extended beyond mid-block segments to include intersections, where complex interactions and conflicts often occur.

Enhancing the data processing algorithms and modeling approaches is another important direction. Researchers should refine the existing algorithms to better capture the nuances of real-world transportation scenarios, such as accurately distinguishing between segment-related and intersection-related conflicts. Furthermore, the application of machine learning and artificial intelligence techniques should be investigated, as these advanced methods have the potential to uncover safety-related patterns and insights from the rich connected vehicle data that may not be easily detected through traditional analytical approaches.

As the adoption of connected vehicle technologies continues to grow, future research should also focus on developing effective strategies to address the challenge of low connected vehicle penetration. This may involve the creation of algorithms that can estimate road safety conditions even when the connected vehicle sample is not representative of the entire driving population. Exploring ways to increase the penetration and utilization of connected vehicle technologies will be crucial for enhancing the reliability and applicability of the safety analysis.

Finally, future studies should adopt a network-level perspective, expanding the safety evaluation and route optimization methods to encompass entire transportation networks. This will enable

the integration of the safety index into comprehensive transportation planning and decision-making frameworks, ensuring that road safety considerations are systematically incorporated into infrastructure design, traffic management, and navigation systems.

By addressing these limitations and pursuing the future research directions, the field of safety surrogate measures in road safety studies can continue to evolve, providing increasingly robust and actionable insights to improve road traffic safety.

References

[1] A. Tarko, N. Saunier, and T. Sayed, “Surrogate Measures of Safety,” 2009.

[2] C. Hydén, “The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique,” *Bulletin Lund Institute of Technology, Department*, no. 70, 1987.

[3] Å. Svensson and C. Hydén, “Estimating the severity of safety related behaviour,” *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 379–385, Mar. 2006, doi: 10.1016/j.aap.2005.10.009.

[4] A. Laureshyn, Å. Svensson, and C. Hydén, “Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation,” *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1637–1646, Nov. 2010, doi: 10.1016/j.aap.2010.03.021.

[5] N. Saunier, T. Sayed, and K. Ismail, “Large-Scale Automated Analysis of Vehicle Interactions and Collisions,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2147, no. 1, pp. 42–50, Jan. 2010, doi: 10.3141/2147-06.

[6] A. P. Tarko, “Use of crash surrogates and exceedance statistics to estimate road safety,” *Accident Analysis & Prevention*, vol. 45, pp. 230–240, Mar. 2012, doi: 10.1016/j.aap.2011.07.008.

[7] L. Zheng, K. Ismail, and X. Meng, “Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes,” *Safety Science*, vol. 64, pp. 155–162, Apr. 2014, doi: 10.1016/j.ssci.2013.12.003.

[8] L. Zheng and K. Ismail, “A generalized exponential link function to map a conflict indicator into severity index within safety continuum framework,” *Accident Analysis & Prevention*, vol. 102, pp. 23–30, May 2017, doi: 10.1016/j.aap.2017.02.013.

[9] H.-C. Chin and S.-T. Quek, “Measurement of traffic conflicts,” *Safety Science*, vol. 26, no. 3, pp. 169–185, 1997.

[10] L. N. Peesapati, M. P. Hunter, and M. O. Rodgers, “Evaluation of Postencroachment Time as Surrogate for Opposing Left-Turn Crashes,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2386, no. 1, pp. 42–51, Jan. 2013, doi: 10.3141/2386-06.

[11] L. Zheng, K. Ismail, and X. Meng, “Traffic conflict techniques for road safety analysis: open questions and some insights,” *Canadian Journal of Civil Engineering*, vol. 41, no. 7, pp. 633–641, Jul. 2014, doi: 10.1139/cjce-2013-0558.

[12] K. El-Basyouny and T. Sayed, “Safety performance functions using traffic conflicts,” *Safety Science*, vol. 51, no. 1, pp. 160–164, Jan. 2013, doi: 10.1016/j.ssci.2012.04.015.

[13] T. A. Dingus *et al.*, “The 100-Car Naturalistic Driving Study: Phase II - Results of the 100-Car Field Experiment,” *PsycEXTRA Dataset*. American Psychological Association (APA), 2006. doi: 10.1037/e624282011-001.

[14] F. Guo, S. G. Klauer, J. M. Hankey, and T. A. Dingus, “Near Crashes as Crash Surrogate for Naturalistic Driving Studies,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2147, no. 1, pp. 66–74, Jan. 2010, doi: 10.3141/2147-09.

[15] K.-F. Wu and P. P. Jovanis, “Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data,” *Accident Analysis & Prevention*, vol. 45, pp. 507–516, Mar. 2012, doi: 10.1016/j.aap.2011.09.002.

[16] A. Arun, M. M. Haque, A. Bhaskar, S. Washington, and T. Sayed, “A systematic mapping review of surrogate safety assessment using traffic conflict techniques,” *Accident Analysis & Prevention*, vol. 153, p. 106016, 2021.

[17] M. R. Parker Jr and C. V. Zegeer, “Traffic conflict techniques for safety and operations: Observers manual,” United States. Federal Highway Administration, 1989.

[18] S. R. Perkins and J. I. Harris, *Criteria for traffic conflict characteristics, signalized intersections*. Research Laboratories, General Motors Corporation, 1967.

[19] F. Orsini, G. Gecchele, R. Rossi, and M. Gastaldi, “A conflict-based approach for real-time road safety analysis: Comparative evaluation with crash-based models,” *Accident Analysis & Prevention*, vol. 161, p. 106382, Oct. 2021, doi: 10.1016/j.aap.2021.106382.

[20] M. G. Mohamed and N. Saunier, “Motion Prediction Methods for Surrogate Safety Analysis,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2386, no. 1, pp. 168–178, Jan. 2013, doi: 10.3141/2386-19.

[21] R. J. Kiefer, D. J. LeBlanc, and C. A. Flannagan, “Developing an inverse time-to-collision crash alert timing approach based on drivers’ last-second braking and steering judgments,” *Accident Analysis & Prevention*, vol. 37, no. 2, pp. 295–303, Mar. 2005, doi: 10.1016/j.aap.2004.09.003.

[22] P. St-Aubin, N. Saunier, and L. Miranda-Moreno, “Large-scale automated proactive road safety analysis using video data,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 363–379, Sep. 2015, doi: 10.1016/j.trc.2015.04.007.

[23] N. Saunier and T. Sayed, “Automated Analysis of Road Safety with Video Data,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2019, no. 1, pp. 57–64, Jan. 2007, doi: 10.3141/2019-08.

[24] J. C. Hayward, “Near miss determination through use of a scale of danger”.

[25] K. Ozbay, H. Yang, B. Bartin, and S. Mudigonda, “Derivation and Validation of New Simulation-Based Surrogate Safety Measure,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2083, no. 1, pp. 105–113, Jan. 2008, doi: 10.3141/2083-12.

[26] M. M. Minderhoud and P. H. L. Bovy, “Extended time-to-collision measures for road traffic safety assessment,” *Accident Analysis & Prevention*, vol. 33, no. 1, pp. 89–97, Jan. 2001, doi: 10.1016/s0001-4575(00)00019-1.

[27] B. L. Allen, B. T. Shin, and P. J. Cooper, “Analysis of traffic conflicts and collisions,” 1978.

[28] Y. Kuang, X. Qu, J. Weng, and A. Etemad-Shahidi, “How Does the Driver’s Perception Reaction Time Affect the Performances of Crash Surrogate Measures?,” *PLoS One*, vol. 10, no. 9, pp. e0138617–e0138617, Sep. 2015, doi: 10.1371/journal.pone.0138617.

[29] F. Cunto and F. F. Saccomanno, “Calibration and validation of simulated vehicle safety performance at signalized intersections,” *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1171–1179, May 2008, doi: 10.1016/j.aap.2008.01.003.

[30] S. M. S. Mahmud, L. Ferreira, Md. S. Hoque, and A. Tavassoli, “Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs,” *IATSS Research*, vol. 41, no. 4, pp. 153–163, Dec. 2017, doi: 10.1016/j.iatssr.2017.02.001.

[31] D. J. Gabauer and H. C. Gabler, “Comparison of roadside crash injury metrics using event data recorders,” *Accident Analysis & Prevention*, vol. 40, no. 2, pp. 548–558, Mar. 2008, doi: 10.1016/j.aap.2007.08.011.

[32] S. G. Shelby, “Delta-V as a measure of traffic conflict severity,” in *3rd International Conference on Road Safety and Simulati. September*, 2011, pp. 14–16.

[33] A. Laureshyn, M. de Goede, N. Saunier, and A. Fyhri, “Cross-comparison of three surrogate safety methods to diagnose cyclist safety problems at intersections in Norway,” *Accident Analysis & Prevention*, vol. 105, pp. 11–20, Aug. 2017, doi: 10.1016/j.aap.2016.04.035.

[34] Y. Kuang, X. Qu, and S. Wang, “A tree-structured crash surrogate measure for freeways,” *Accident Analysis & Prevention*, vol. 77, pp. 137–148, 2015.

[35] K. Ismail, T. Sayed, and N. Saunier, “Methodologies for Aggregating Indicators of Traffic Conflict,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2237, no. 1, pp. 10–19, Jan. 2011, doi: 10.3141/2237-02.

[36] D. Gettman, L. Pu, T. Sayed, S. G. Shelby, and S. Energy, “Surrogate safety assessment model and validation,” Turner-Fairbank Highway Research Center, 2008.

[37] C. Wang and N. Stamatiadis, “Evaluation of a simulation-based surrogate safety metric,” *Accident Analysis & Prevention*, vol. 71, pp. 82–92, Oct. 2014, doi: 10.1016/j.aap.2014.05.004.

[38] K. Ismail, T. Sayed, N. Saunier, and C. Lim, “Automated Analysis of Pedestrian–Vehicle Conflicts Using Video Data,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2140, no. 1, pp. 44–54, Jan. 2009, doi: 10.3141/2140-05.

[39] K. Ismail, T. Sayed, and N. Saunier, “Automated Analysis of Pedestrian–Vehicle Conflicts,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2198, no. 1, pp. 52–64, Jan. 2010, doi: 10.3141/2198-07.

[40] N. Saunier and M. G. Mohamed, “Clustering surrogate safety indicators to understand collision processes,” in *Transportation Research Board, 93rd meeting*, 2014, pp. 14–2380.

[41] L. Zheng, T. Sayed, and F. Mannering, “Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions,” *Analytic methods in accident research*, vol. 29, p. 100142, 2021.

[42] J. Autey, T. Sayed, and M. H. Zaki, “Safety evaluation of right-turn smart channels using automated traffic conflict analysis,” *Accident Analysis & Prevention*, vol. 45, pp. 120–130, Mar. 2012, doi: 10.1016/j.aap.2011.11.015.

[43] T. Sayed, M. H. Zaki, and J. Autey, “Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis,” *Safety Science*, vol. 59, pp. 163–172, Nov. 2013, doi: 10.1016/j.ssci.2013.05.009.

[44] M. H. Zaki, T. Sayed, and K. Shaaban, “Use of Drivers’ Jerk Profiles in Computer Vision–Based Traffic Safety Evaluations,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2434, no. 1, pp. 103–112, Jan. 2014, doi: 10.3141/2434-13.

[45] A. Sobhani, W. Young, and M. Sarvi, “A simulation based approach to assess the safety performance of road locations,” *Transportation Research Part C: Emerging Technologies*, vol. 32, pp. 144–158, Jul. 2013, doi: 10.1016/j.trc.2012.10.001.

[46] O. Bagdadi, “Assessing safety critical braking events in naturalistic driving studies,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 16, pp. 117–126, Jan. 2013, doi: 10.1016/j.trf.2012.08.006.

[47] J. K. Jonasson and H. Rootzén, “Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach,” *Accident Analysis & Prevention*, vol. 62, pp. 102–109, Jan. 2014, doi: 10.1016/j.aap.2013.09.013.

[48] K.-F. Wu and P. P. Jovanis, “Defining and screening crash surrogate events using naturalistic driving data,” *Accident Analysis & Prevention*, vol. 61, pp. 10–22, Dec. 2013, doi: 10.1016/j.aap.2012.10.004.

[49] L. Chong, M. M. Abbas, A. Medina Flintsch, and B. Higgs, “A rule-based neural network approach to model driver naturalistic behavior in traffic,” *Transportation Research Part C: Emerging Technologies*, vol. 32, pp. 207–223, Jul. 2013, doi: 10.1016/j.trc.2012.09.011.

[50] F. Huang, P. Liu, H. Yu, and W. Wang, “Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections,” *Accident Analysis & Prevention*, vol. 50, pp. 1014–1024, Jan. 2013, doi: 10.1016/j.aap.2012.08.018.

[51] W. Young, A. Sobhani, M. G. Lenné, and M. Sarvi, “Simulation of safety: A review of the state of the art in road safety simulation modelling,” *Accident Analysis & Prevention*, vol. 66, pp. 89–103, May 2014, doi: 10.1016/j.aap.2014.01.008.

[52] D. Gettman and L. Head, “Surrogate Safety Measures from Traffic Simulation Models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1840, no. 1, pp. 104–115, Jan. 2003, doi: 10.3141/1840-12.

[53] E. Hauer and P. Garder, “Research into the validity of the traffic conflicts technique,” *Accident Analysis & Prevention*, vol. 18, no. 6, pp. 471–481, Dec. 1986, doi: 10.1016/0001-4575(86)90020-5.

[54] T. Sayed and S. Zein, “Traffic conflict standards for intersections,” *Transportation Planning and Technology*, vol. 22, no. 4, pp. 309–323, Aug. 1999, doi: 10.1080/03081069908717634.

[55] D. Lord and F. Mannering, “The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives,” *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, Jun. 2010, doi: 10.1016/j.tra.2010.02.001.

[56] Y. Ali, M. M. Haque, and F. Mannering, “Assessing traffic conflict/crash relationships with extreme value theory: Recent developments and future directions for connected and autonomous vehicle and highway safety research,” *Analytic methods in accident research*, p. 100276, 2023.

[57] P. Songchitruksa and A. P. Tarko, “The extreme value theory approach to safety estimation,” *Accident Analysis & Prevention*, vol. 38, no. 4, pp. 811–822, Jul. 2006, doi: 10.1016/j.aap.2006.02.003.

[58] L. Zheng, K. Ismail, and X. Meng, “Freeway safety estimation using extreme value theory approaches: A comparative study,” *Accident Analysis & Prevention*, vol. 62, pp. 32–41, Jan. 2014, doi: 10.1016/j.aap.2013.09.006.

[59] Z. Geng, X. Ji, R. Cao, M. Lu, and W. Qin, “A conflict measures-based extreme value theory approach to predicting truck collisions and identifying high-risk scenes on two-lane rural highways,” *Sustainability*, vol. 14, no. 18, p. 11212, 2022.

[60] Y. Guo, M. Essa, T. Sayed, M. M. Haque, and S. Washington, “A comparison between simulated and field-measured conflicts for safety assessment of signalized intersections in Australia,” *Transportation research part C: emerging technologies*, vol. 101, pp. 96–110, 2019.

[61] C. Fu and T. Sayed, “A multivariate method for evaluating safety from conflict extremes in real time,” *Analytic Methods in Accident Research*, vol. 36, p. 100244, Dec. 2022, doi: 10.1016/j.amar.2022.100244.

[62] G. A. Davis, J. Hourdos, H. Xiong, and I. Chatterjee, “Outline for a causal model of traffic conflicts and crashes,” *Accident Analysis & Prevention*, vol. 43, no. 6, pp. 1907–1919, Nov. 2011, doi: 10.1016/j.aap.2011.05.001.

[63] A. P. Tarko, “Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution–A theoretical and numerical exploration,” *Accident Analysis & Prevention*, vol. 113, pp. 63–73, 2018.

[64] P. Zhao and C. Lee, “Assessing rear-end collision risk of cars and heavy vehicles on freeways using a surrogate safety measure,” *Accident Analysis & Prevention*, vol. 113, pp. 149–158, 2018.

[65] D. Henclewood, M. Abramovich, and B. Yelchuru, “Safety pilot model deployment–one day sample data environment data handbook,” *Research and Technology Innovation Administration. Research and Technology Innovation Administration, US Department of Transportation, McLean, VA*, 2014.

[66] J. Liu and A. J. Khattak, “Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles,” *Transportation research part C: emerging technologies*, vol. 68, pp. 83–100, 2016.

[67] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*, vol. 208. Springer, 2001.

[68] Q. Meng and J. Weng, “Evaluation of rear-end crash risk at work zone using work zone traffic data,” *Accident Analysis & Prevention*, vol. 43, no. 4, pp. 1291–1300, Jul. 2011, doi: 10.1016/j.aap.2011.01.011.

[69] S. Chandra, “Safety-based path finding in urban areas for older drivers and bicyclists,” *Transportation research part C: emerging technologies*, vol. 48, pp. 143–157, 2014.

[70] R. C. Larson and A. R. Odoni, *Urban operations research*, no. Monograph. 1981.

[71] A. Y. Abul-Magd, “Modeling highway-traffic headway distributions using superstatistics,” *Physical Review E*, vol. 76, no. 5, p. 057101, 2007.

[72] M. Dozza, “What factors influence drivers’ response time for evasive maneuvers in real traffic?,” *Accident Analysis & Prevention*, vol. 58, pp. 299–308, 2013.

[73] M. W. Hancock and B. Wright, “A policy on geometric design of highways and streets,” *American Association of State Highway and Transportation Officials: Washington, DC, USA*, vol. 3, 2013.

[74] F. F. Saccomanno, F. Cunto, G. Guido, and A. Vitale, “Comparing safety at signalized intersections and roundabouts using simulated rear-end conflicts,” *Transportation Research Record*, vol. 2078, no. 1, pp. 90–95, 2008.

[75] D. Lord, A. Manar, and A. Vizioli, “Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments,” *Accident Analysis & Prevention*, vol. 37, no. 1, pp. 185–199, 2005.

[76] L. Aarts and I. Van Schagen, “Driving speed and the risk of road crashes: A review,” *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.

[77] Y.-J. Kweon, “Development of crash prediction models with individual vehicular data,” *Transportation research part C: emerging technologies*, vol. 19, no. 6, pp. 1353–1363, 2011.

[78] A. Hillier, “Manual for working with ArcGIS 10,” *University of Pennsylvania*, 2011.