EXAMINING THE EFFECTS OF NON-INFRASTRUCTURE VARIABLES ON THE SAFETY PERFORMANCE OF MIXED ENVIRONMENTS WITH DIFFERENT AUTOMATED VEHICLES MARKET PENETRATION RATES AT SIGNALIZED INTERSECTIONS

A Dissertation

by

SEYEDEH MARYAM MOUSAVI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Dominique Lord
Committee Members,	Karen K. Dixon
	Luca Quadrifoglio
	Wei Li
Head of Department,	Robin Autenrieth

December 2020

Major Subject: Civil Engineering

ABSTRACT

To accommodate autonomous vehicles (AVs) and make them perform at their capacity, the current roadway infrastructure should be upgraded. However, upgrading the entire infrastructure cannot be conducted at once since it is time-consuming and costly. Therefore, evaluating alternative solutions to enhance or maintain traffic safety is critical, especially when primarily employing AVs. Among the roadway network components, intersections are one of the critical locations since they are proportionally experiencing more crashes, and signalized ones result in the majority of the fatal intersection related crashes. Hence, this dissertation examines the safety effects of various non-infrastructure variables at a signalized intersection in mixed traffic environments. To this aim, different levels of signal cycle length, speed limit, and left-turn (LT) signal phasing were considered under seven AV market penetration rates (MPRs). In addition, the safety effects of AV size were analyzed. A micro-simulation program was employed to develop and run 3,850 simulation runs that were developed using a full factorial design. Eventually, the traffic safety of the scenarios was analyzed from various aspects, including 1) 15 longitudinal driving volatility measure, 2) 15 lateral driving volatility measure, and 3) three machine learning (ML) regression models using the percentage of jerks. The results showed that: 1) increasing the AV MPR improved the majority of the lateral volatility measures, 2) larger AV size is associated with higher longitudinal and lower lateral volatility measures, 3) each volatility model is different, and each model could be implemented according to the objective of a study, 4) the ML models for total jerks, AV jerk, and regular vehicles (RVs) jerk consistently indicated that higher speed limits and permitted LT phasing results in the

lower percentage of jerks, 5) increasing the AV size reduces the percentage of AV jerks and meanwhile has a negative effect on the RV jerks, but the benefit is higher, and 6) decreasing the cycle length reduces the number of jerky driving maneuvers for AVs. Eventually, the total and rear-end conflicts using the surrogate safety assessment model (SSAM) were compared to the number of jerks and indicated they could not be used interchangeably.

DEDICATION

To my parents and brother, without whom none of my success would be possible.

ACKNOWLEDGEMENTS

CONTRIBUTORS AND FUNDING SOURCES

Contributors

NOMENCLATURE

ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance System
ADS	Automated Driving Systems
AV	Autonomous Vehicle
C.V.	Coefficient of Variation
CACC	Cooperative Adaptive Cruise Control
CAV	Connected and Automated Vehicle
CV	Connected Vehicle
DR	Deceleration Rate
GLM	Generalized Linear Model
GT	Gap Time
HAV	High-Level AV
LAV	Low-Level AV
LT	Left-Turn
MAE	Mean Absolute Error
MBE	Mean Bias Error
ML	Machine Learning
MPR	Market Penetration Rate
MSE	Mean Square Error
MUTCD	Manual on Uniform Traffic Control Devices
NHTSA	National Highway Traffic Safety Administration

PET	Post Encroachment Time
PRT	Perception Reaction Time
PS.D.	Proportion of Stopping Distance
RT	Right-Turn
RV	Regular Vehicle
S.D.	Standard Deviation
SAE	Society of Automotive Engineers
SSAM	Surrogate Safety Assessment Model
SSM	Surrogate Safety Measure
THW	Time Headway
TSM	Traffic Safety-Based Maneuvering
TTC	Time-To-Collision
VSM	Vehicle Safety-based Maneuvering
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This chapter introduces a general overview of traffic safety and the development of autonomous vehicles, explains the problem statement, discusses the objectives of the dissertation, and lastly provides the dissertation outline.

1.1. Introduction

Cities have experienced several transformations and growth over time, and transportation systems have been an integral part of the transformations (*1*). The growth in population and commercial areas in urban environments leads to significant traffic congestion. Traffic congestion causes delays, air pollution, the stress of drivers, fuel consumption, and noise (*2*). Traffic congestion resulted in a $160 billion loss in the United States in 2014 (*3*). Also, evaluating 471 urban areas in the U.S. indicated that drivers lose 6.9 billion years and 3.1 billion gallons of fuels in traffic jam each year (*4*).

In addition to the foregoing drawbacks, the growth of cities and traffic congestion contributes to roadway crashes, injuries, and fatalities. Based on the World Health Organization (WHO), nearly 1.35 million people died in roadway crashes in 2018, while injury due to crashes is the 8th leading cause of death for the entire world. These statistics make the roadway safety a major concern (*5*). The United States statistics indicate that there were 37,133 and 40,000 fatalities due to the roadway crashes in 2017 and 2018, respectively (*6*, *7*).

Since the conventional solutions cannot fully solve the traffic safety concerns, due to the cost and space limitations, there is a need to implement an innovative solution to alleviate traffic congestion and overcome its consequences, including safety concerns, traffic delay, air pollution, fuel consumption, and noise level. As about 94% of all crashes in the U.S. involve human error, including fatigue, alcohol, or drug, eliminating or reducing the effect of the human factor can reduce the risk of roadway collisions (*2*, *8*, *9*).

Vehicle automation is one solution that is widely recognized for its expected ability to minimize reliance on the human element. Over the past few years, the development of technologies also affects the automobile industry and leads to growth by bringing in computerization (*10*). In fact, by integrating sensing technology and wireless communication in the traffic systems, autonomous vehicle technology has been developing significantly and is expected to revolutionize the vehicle industry (*1*, *11*, *12*). Autonomous vehicles (AVs), also known as automated or self-driving cars, can drive themselves on the roadway networks and navigate through various types of roadways and environments without any interference of human control (*10*). *Table 1* represents six levels of vehicle automation defined by the Society of Automotive Engineers (SAE) (*13*):

Table 1. SAE Levels of Vehicle Automation. Reprinted from *(13)*

SAE Level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human Driver Monitors the Driving Environment
0	No Automation	The full-time performance by the human driver of all aspects of the dynamic driving task, even when "enhanced by warning or intervention systems."	Human Driver	Human Driver	Human Driver	N/A
1	Driver Assistance	The driving mode-specific execution by a driver assistance system of "either steering or acceleration/deceleration." using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task	Human Driver And System	Human Driver	Human Driver	Some Driving Modes
2	Partial Automation	The driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration. using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task	System	Human Driver	Human Driver	Some Driving Modes
Automated Driving System Monitors the Driving Environment
3	Conditional Automation	The driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task. With the expectation that the human driver will respond appropriately to a request to intervene	System	System	Human Driver	Some Driving Modes
4	High Automation	The driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task. Even if a human driver does not respond appropriately to a request to intervene	System	System	System	Many Driving Modes
5	Full Automation	The driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task. Under all roadway and environmental conditions that can be managed by a human driver	System	System	System	All Driving Modes

AVs can address important concerns in transportation, including (*8*, *14–16*):

1) improving roadway safety by eliminating the human element;

2) providing mobility for everyone, including abled and disabled people, and all age groups who are not capable of driving;

3) saving time by reallocating commute time to another task;

4) shortening travel time by eliminating the required time to find a parking spot;

5) reducing parking cost;

6) increasing roadway capacity by reducing the headway/gap due to the improved safety features, *Figure 1*; and

7) saving fuel and decreasing emissions.
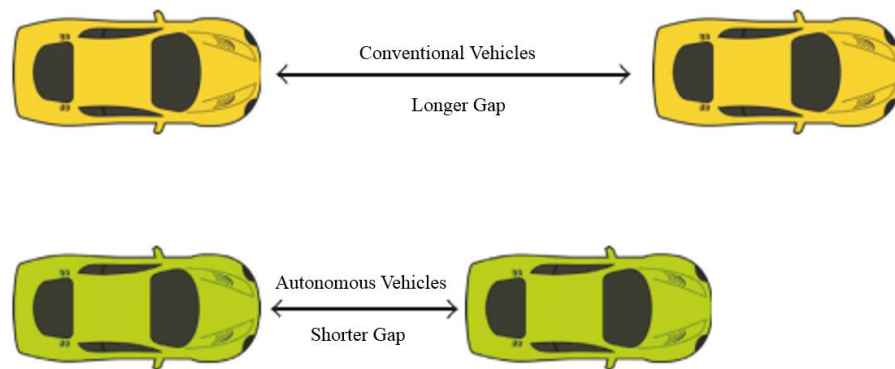
Figure 1. Gap in the AV Environment vs. Conventional Vehicle Environment.

Reprinted from (*8*)

The major social impacts of AVs are estimated to be $2,000 per vehicle per year, considering crash saving, fuel consumption, travel time reduction, and parking benefits. This cost reaches up to $4,000 when accounting for comprehensive crash costs (*10*).

Automated driving illustrates the potential to improve safety in addition to the traffic operation by gradually reducing the human-driven vehicles (*17*). AVs improve traffic safety by restraining the leading cause of the roadway crashes that are drivers' errors

(*18*). The safety of the AV transportation networks can be promoted due to the embedded sensors, cameras, lasers, and radars in the AV system, as depicted in *Figure 2*. This equipment is to monitor the vehicle and its surrounding environment and prompt either vehicle or driver to react to the risky situations to avoid a collision (*19*).
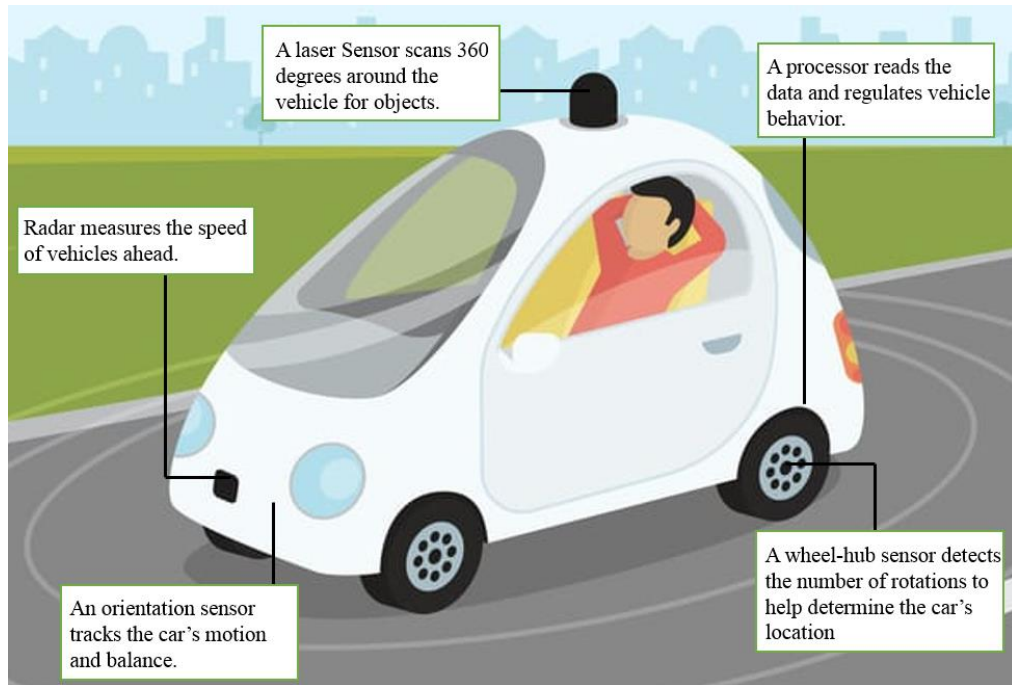
Figure 2. Concept of the Automated Cars. Adapted from (*17*, *20*)

However, there are also concerns associated with AVs, such as loss of awareness of the drivers and overreliance on automation that can cause critical issues in the case of complex traffic situations (*21*). Furthermore, due to the lack of real-world data, the exact impact of AVs cannot be determined and confirmed thoroughly (*22*). Therefore, evaluating the safety impacts of AVs is critical, but meanwhile, it is a challenge due to the unavailability of the real-world data. (*22*).

1.2. Problem Statement

As mentioned earlier, human factors and driving behaviors are known as the leading causes of traffic crashes (*23–26*). Therefore, replacing human-driven vehicles by AVs could promote both traffic safety and operation. But, due to the lack of data, very few research studies have been performed to examine the safety effects of AVs, especially in mixed traffic environments. In addition, different countries and cities provide different road infrastructures and regulations, which makes it challenging for AVs to perform before standardizing the infrastructures and regulations (*27*). However, with the rapid advancement of AV technology, AVs are expected to operate on the current roadway infrastructure in the near future. For AVs to operate properly, there are studies that indicate the necessity of providing smart roadways to communicate with vehicles and improve pavement structures (*28*, *29*). In fact, to achieve a smart city and develop autonomous intelligent vehicles, cooperating intelligent transportation systems and infrastructure should be further developed (*30*). Therefore, it is important to evaluate AVs' safety and consider promoting the existing infrastructure components to safely accommodate AVs. However, since it is costly and time-consuming to upgrade the entire infrastructure at once, it is beneficial to primarily evaluating and determining how altering non-infrastructure components can change traffic safety.

Although it is crucial to address the roadway crashes, injuries, and fatalities for the entire roadway network, network screening should be conducted to prioritize the locations based on their hazardousness. In addition, limitations in funding sources prevent the chance of promoting safety for the entire network at once. Among the roadway network components, signalized intersections are heavily traveled locations, complex due to having

several conflicts points, associated with more severe vehicular and pedestrian-involved crashes, and experience more crashes than segments given the length of the transportation network (*31–33*). More importantly, signalized intersection related crashes tend to be more severe compared to the other types of intersections. In fact, 30% of the intersection related fatal crashes occur at signalized intersections, while only 10% of the intersections are signalized (*33*). Also, analyzing the location of the crashes in the AV test conducted in California, *Figure 3*, indicated that 88% of the AV involved crashes happened at intersections, and 54% of the incidents occurred at the signalized intersections (*34*).

Therefore, it is critical to give the signalized intersection the priority for safety considerations. Therefore, traffic signal cycle length, left-turn (LT) signal phasing, and the speed limit are such non-infrastructure variables at signalized intersections that could be assessed to determine how they affect traffic safety in mixed traffic environments. In addition, the effect of AV size will be evaluated along with the abovementioned non-infrastructure variables.
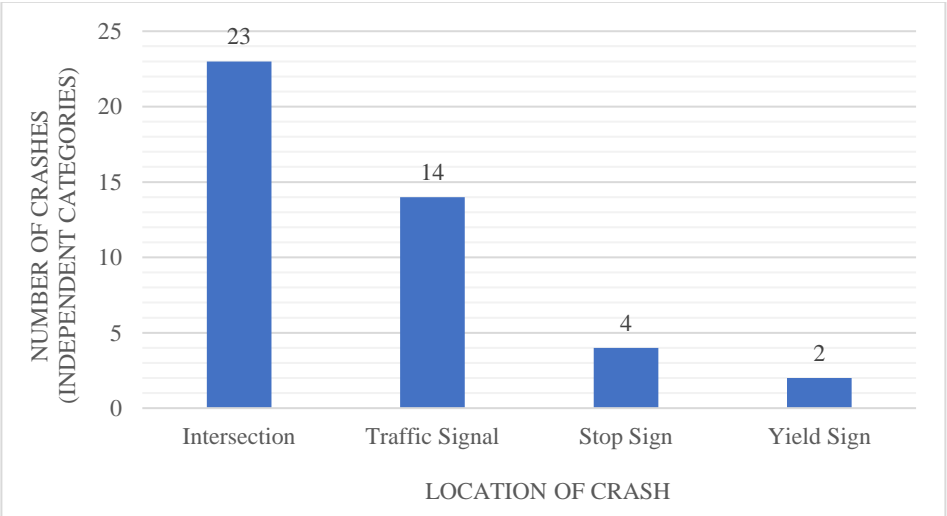
Figure 3. Location of the AV-Involved Crashes. Adapted from (*34*)

In summary, this dissertation focuses on evaluating the effects of non-infrastructure variables on the safety of signalized intersections under various AV MPRs. To accomplish this objective, the effects of various LT signal phasing, signal cycle length, speed limit, and vehicle size will be evaluated under various AV MPRs. The safety will be investigated through developing safety models based on the percentage of jerks for each vehicle type as well as determining how the study variables change different driving volatility measures.

1.3. Research Objectives

The primary purpose of this research is to evaluate the effects of various non-infrastructure variables on the safety of signalized intersections under various AV MPRs. The targeted objectives are as follows:

1) Assess the effects of AV size and various MPRs of AVs on the safety of signalized intersections;

2) Evaluate the impacts of different non-infrastructure variables, including signal cycle length, LT signal phasing, and the speed limit on the safety of signalized intersections;

3) Determine how the study variables affect various longitudinal and lateral driving volatility measures, and therefore, traffic safety;

4) Develop various machine learning models to estimate the safety of signalized intersections when implementing AVs at various levels of MPRs; and

5) Lastly, determine the relationship between the number of jerky driving maneuvers and the number of Surrogate Safety Assessment Model (SSAM) conflicts.

1.4. Research Contributions

The majority of the studies have considered an ideal roadway infrastructure condition while evaluating the operation and safety of AVs or mixed traffic environments. However, to be realistic, it is impractical, costly, and time-consuming to upgrade the entire infrastructure for AVs at once, especially for low AV MPRs. In addition, another concern is the lack of rich literature on the safety performance of AVs and/or mixed traffic environments. Hence, this dissertation, as a pioneer research study, evaluates the effects of economical and rapid non-infrastructure solutions, including adjusting signal cycle length, LT signal phasing, and the speed limit on traffic safety at signalized intersections under various AV MPRs. Also, the effect of AV size on traffic safety in mixed traffic environments is assessed to be used as a guideline for AV manufacturers to contribute to reducing the total number of crashes while employing AVs.

For the safety evaluation, this dissertation conducts comprehensive analyses to assess all the safety aspects of mixed traffic environments. The safety analyses are categorized as follows:

1) Analyzing longitudinal and lateral driving volatility measures: as a common practice, traffic safety has been assessed by analyzing only longitudinal movements of the vehicles, while the lateral movements have been neglected. However, to comprehensively measure the safety of the mixed traffic

environments, this dissertation, for the very first time, not only evaluates the safety of longitudinal movements but also thoroughly assesses the lateral safety through lateral driving volatility measures. In this dissertation, simulation outputs are utilized to, first, analyze various longitudinal and lateral driving volatility measures, and second, assess how all the study non-infrastructure variables influence each driving volatility measure. Hence, in addition to evaluating 30 different longitudinal and lateral driving volatility measures, the effects of four non-infrastructure variables under seven AV MPRs are also analyzed.

2) Developing machine learning algorithms to predict the percentage of the total, only AV, and only regular vehicle (RV) jerks: generally, to evaluate safety through simulation models, the SSAM software program is used. However, since this software program is controversial, this dissertation uses a novel safety approach for mixed traffic environments by studying driving jerk events. In this approach, a machine learning algorithm has been used to develop various safety prediction models for the percentage of total jerks, AV jerks, and RV jerks and determine how the study variables affect each type of jerk.

3) Comparing the calculated number of jerks and the number of conflicts from the SSAM: eventually, since SSAM is a controversial software program, this dissertation compares the number of calculated jerks and the number of total and rear-end SSAM conflicts to determine if they are correlated and could be used interchangeably.

In summary, this dissertation is the leading work that comprehensively evaluates traffic safety of mixed traffic environments from various aspects by adjusting non-infrastructure variables at a signalized intersection. This is a critical step to ensure traffic safety while employing AVs, especially for low MPRs.

1.5. Dissertation Outline

This dissertation is divided into six different chapters. The first chapter is the current chapter, and the following chapters are presented as follows:

1) Chapter two presents the literature review. The review covers the previous research studies to illustrate the safety effects of AVs and mixed traffic environments, determine how different surrogate safety measures (SSMs) have been used to evaluate the safety, and indicate if the safety effects of the study non-infrastructure variables have ever been analyzed.

2) Chapter three describes the methodology. This chapter comprehensively explains the development of the simulation models, calibration and validation process, creating AV behaviors, developing simulation scenarios, extracting data, processing data, calculating the required SSMs, assessing safety through analyzing both longitudinal and lateral driving volatility measures, and developing safety models using machine learning algorithms.

3) Chapter four contains primary statistical analyses of the study variables and the surrogate safety measures. The basic statistical analyses include descriptive statistics, explanatory analyses, and graphs.

4) Chapter five focuses on advanced statistical analyses and discussions, including the longitudinal driving volatility models, the lateral driving volatility models, and the machine learning models. This chapter describes the characteristics of advanced statistical models and the results of the modeling effort.

5) Chapter six is the last chapter and concludes the dissertation. The conclusion represents a brief summary of the work and provides recommendations for further research.

The next chapter covers the previous research studies that focused on the safety effects of AVs and various study variables as well as different methods of evaluating traffic safety.

CHAPTER II

LITERATURE REVIEW

This chapter aims to review, investigate, and summarize the previous research studies on the safety and operational effects of AVs in mixed traffic environments; the influence of LT signal phasing, signal cycle length, speed limit, and vehicle size on safety; and the methods for assessing traffic safety.

2.1. Background

Roadway crashes, fatalities, and injuries have been one of the main concerns of traffic engineers for decades. Drivers' errors are among the leading causes of roadway crashes (*2*, *8*, *9*). According to the World Health Organization (WHO) (*5*), roadway crashes are the 8th leading cause of deaths worldwide, and nearly 1.35 million people died in roadway crashes in 2018. In 2017 and 2018, the United States reported 37,133 and 40,000 roadway crash fatalities, respectively (*35*, *36*). Various studies evaluated traffic safety of roadway components (*37–42*). Among the various components of roadway networks, signalized intersections are known as high crash risk areas and contribute to 30% of the intersection related fatal crashes, even though only 10% of the intersections are signalized (*33*, *43*). "Vision Zero" that was initiated in Sweden in the 1990s has been focusing on increasing traffic safety by eliminating all traffic fatalities and severe injuries (*44*). Several cities in the United States have been implementing the Vision Zero approach to execute safety plans and enhance roadway safety. *Figure 4* depicts cities in the United States with the Vision Zero target (*45*).
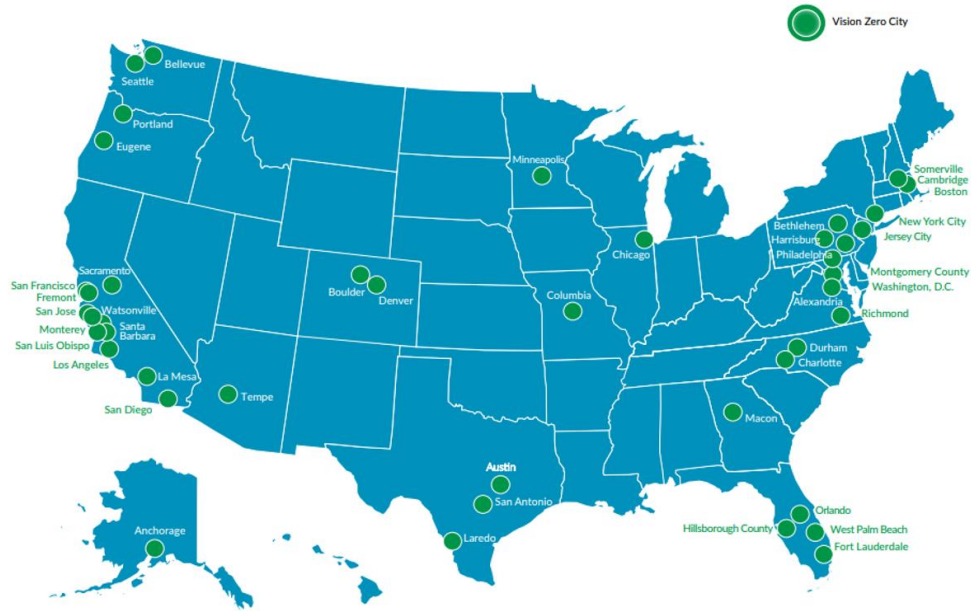
Figure 4. Vision Zero Cities in the U.S. Reprinted from (*46*)

Generally, there are two types of safety features in vehicles to improve roadway safety (*44*):

1. Passive safety features: these features are designed to protect victims of roadway crashes during and after a crash occurs when the crash is unavoidable.

2. Active safety features: these features are designed to prevent crashes or mitigate crash severities.

Eskandarian (*44*) indicated passive safety equipment reach a "point of diminishing returns"; however, on the other hand, the exploitation of the active safety features has been experiencing an upward trend.

If an active safety paradigm is fully implemented, instead of accepting crashes as given and considering the consequences, the pre-crash situation can be evaluated and addressed to prevent crashes (*44*).

14

AVs have been developed to implement active safety features to prevent roadway crashes and improve traffic safety. AVs, as evolving technology, also limit the human factor and behavior from driving in order to enhance the operation and safety of the roadways (*47*). Besides the benefits associated with AVs, they can be subjected to uncertainties such as unaffordable initial cost, state-level licensing, and standard testing (rather than national-level) that lead to inconsistencies, security concerns, and undefined liability details. In addition, the interaction between AVs and other transportation system components, especially in a mixed traffic environment, remains undefined. To address these issues, the federal government should develop a national framework for licensing, security, privacy, and standard for liability (*10*). The following section focus on various aspects of AVs.

2.2. Effects of AVs on Traffic Safety and Operation

This section expands on the influence of AVs on the safety and operation of the roadway traffic, respectively.

2.1.1. Influence of Vehicle Automation on Traffic Safety

Many people lose their lives or get injured in roadway crashes due to human errors or violations of traffic laws (*48*). Eliminating drivers' distraction, fatigue, inefficiency, health issues, and misjudgment leads to a 10-30% reduction in fatal crashes (*49*). Managing and minimizing the perception reaction time (PRT), as is the focused objective of vehicle automation, is a critical factor in reducing the number of fatal and severe crashes (*49*). Even after dismissing the human factor element from the AVs system, the AVs are

still vulnerable to the software and hardware hacking and glitches. During a course of

testing AVs, operators activated 2,700 disengagements, including bad weather, poor

pavement marking, potholes, and construction zones, in which the control of AVs was

taken over by humans (*50*). The following section presents various studies on the safety

effects of AVs on the roadway network and traffic.

By equipping vehicles to the partially-automated crash avoidance features,

including lane departure warning, blind-spot monitoring, and forward collision warning,

the frequency and severity of the roadway crashes can be reduced because of the

diminished effects of human errors and distracted driving. Assessing the costs and benefits

of deploying crash avoidance features within the U.S. light-duty vehicle fleet revealed that

the equipment could collectively prevent or reduce the severity of the crashes by 1.3

million per year in the United States. Harper et al. (*51*) considered two upper and lower

bounds of benefits, in which the upper bound is the case where all the crashes are

prevented, and the lower bound is based on observing the insurance data. The results show

$20 and $861 per vehicle net benefit for lower and upper bound, respectively. Even though

the evolution of technologies improves the performance of AVs and decrease the number

of roadway crashes and fatalities, the transition period to reach a 100% MPR takes time,

and the interaction between AVs and conventional vehicles/regular vehicles (RVs) remains

a challenge (*50*, *52*). During the course of transition, AVs are at risk of colliding with

human-driven vehicles and may even lead to a higher than anticipated rate of injury if the

occupants are not belted or out of position (reading, sleeping, conversing, etc.) (*52*). To

identify the risk associated with the failure of AVs in a mixed traffic environment, Bhavsar

et al. (*50*) developed a fault-tree analysis. After disassembling the AV system into two

components of vehicular and transportation infrastructure, two fault tree models were established for each of the components separately. The results indicated a failure probability of about 14% resulting from the failure of the AV components, while this value was 158 per one million miles of travel if considering the failure of the transportation infrastructure in the estimations.

Analyzing the real-world data from testing AVs in California indicated that among all the AV involved crashes, AVs were not at fault in any of the crashes, and the overall severity was lower than the conventional vehicle crashes (*53*). Although vehicle automation significantly reduces the overall number of crashes, it may cause new types of crashes at the same time. *Figure 5* depicts the effect of automation on collision (*49*).
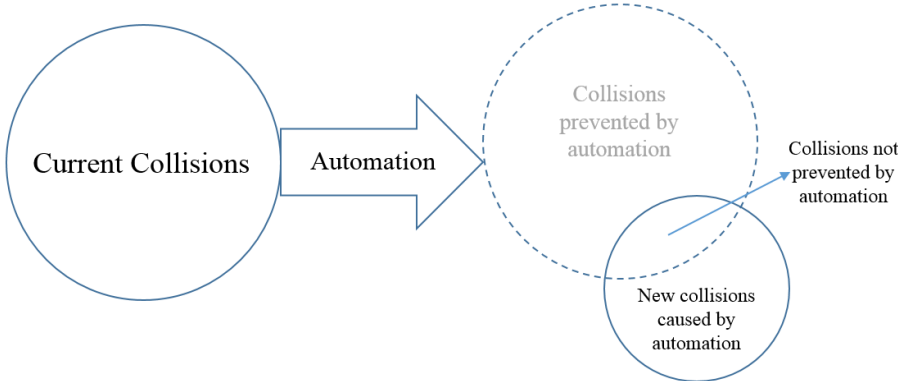
Figure 5. Effect of Vehicle Automation on Crashes. Adapted from (*49*)

The test that was conducted for AVs in California required the involved AV manufacturers to report the crashes that AVs were involved in during the test period. Among the 26 reported crashes, the majority of them were rear-end crashes where the AV was hit from the back. Moreover, 94% of human-driven vehicle crashes were related to human errors (*14*).

By implementing an Advanced Driver Assistance System (ADAS), the semi and fully AVs are growing. This advancement develops a mixed traffic condition in which equipped and unequipped vehicles use the same infrastructure (*11*). California PATH (*54*) evaluated and indicated the advantage of implementing platoons using Cooperative Adaptive Cruise Control (CACC) to reduce the time gap to 0.6- 1.5 sec and improve roadway capacity. Gouy et al. (*11*) evaluated the effect of AVs time headway (THW) on the THW of human-driven vehicles. As AVs are accompanied by small-time headways, the results indicated that manually-driven vehicles adjust the THW in accordance with the THW of their adjacent AV platoon. In fact, they spent most of the time keeping THW below the safety threshold that threatens their safety that indicates the negative impacts of providing a mixed traffic environment.

In addition, a few studies have implemented micro-simulation models for analyzing the safety and operational impacts of AVs on traffic. Morando et al. (*8*) evaluated the safety impacts of two different scenarios of signalized intersection and roundabout under various AV penetration rates using micro-simulation software. The results indicated that AVs improve traffic safety significantly under higher MPRs. The number of conflicts decreased by 20% to 65%, with a statistically significant p-value of less than 0.05, with the AV penetration rates of 50% to 100% for the signalized intersection. For the roundabout scenario, a 29% to 64% reduction was observed in the number of conflicts. Papadoulis et al. (*22*) confirmed the benefit of implementing connected and automated vehicles (CAVs) on the safety of the motorways, even at a low MPR. Mousavi et al. (*55*) compared the safety effect of a fully conventional vehicle environment vs. a fully AV environment near an unsignalized intersection. The results proved the significant benefit of AVs to traffic

safety. Fagnant and Kockelman (*10*) estimated that 90% AV MPR is associated with a saving of 4.2 million dollars due to crash reduction, and 21,700 dollars resulted from fatality reductions per year. In another study, Kockelman et al. (*56*) simulated four MPRs of 25%, 50%, 75%, and 100% for 4-leg intersections. The results of their study using Time-To-Collision (TTC) as the SSM indicated that by increasing the AV MPR, the number of conflicts decreased 4%, 31%, and 77% for three of the studied intersections; however, a 17% increase was observed for another intersection.

Arvin et al. (*57*) also evaluated an intersection located in Ann Arbor, Michigan, under various MPRs of level three and level four AVs, including 0%, 7%, 15%, 40%, 60%, 80%, and 100% (for MPR 100%, different combinations of AV level three and AV level five were used). This study demonstrated that by increasing the level three and level five AV MPR from 0% to 100%, the number of conflicts decreased from nine to zero. Moreover, by considering only level five AV and human-driven vehicles in the simulation environment, the number of conflicts increased from nine to 10 at low AV MPR, which is below 40%; however, by increasing the MPR to 100%, the number of conflicts reduced to zero. Another study by Arvin et al. (*58*) proved that the baseline intersection with fully RVs experienced an average of 9.43 conflicts. However, by increasing the MPR of low-level AV (LAV) and high-level AV (HAV), a 90.1% improvement was observed. Additionally, with having a fully HAV environment, the intersection became conflict-free.

Furthermore, Virdi et al. (*59*) examined the safety impacts of 11 CAV MPRs, starting from 0% until reaching to 100% for each 10% increment, at a signalized intersection in Australia. The results indicated that for the first 20% of the MPR, there was a 22% increase in conflicts, while there was a 48% reduction in the number of conflicts at

MPR of 90%. Another study, conducted by Jeong (*60*), developed a micro-simulation model to evaluate the effect of an optimization process, which minimizes the overall crash risk by focusing on vehicle maneuvering control parameters on a freeway traffic stream under different MPRs of automated driving systems (ADS). Two different approaches were considered: 1) vehicle safety-based maneuvering (VSM), and 2) traffic safety-based maneuvering (TSM). The former one considers the crash risk of an equipped vehicle and its surrounding vehicles, while the latter one examined the overall crash risk of the traffic stream. The results suggested significant reductions in the rear-end crash risks for both VSM and TSM (*60*).

In summary, this section has indicated that providing partially AVs can improve traffic safety. Moreover, the results of the majority of the simulation-based studies showed high AV/CAV MPR enhances the safety of the roadways significantly.

2.1.2.   *Influence of Vehicle Automation on Traffic Operation*

One of the expected operational effects of AVs is the increase in the capacity of the roadways as a result of shorter headways due to the enhanced safety features of the AVs (*8*). Aria et al. (*21*) assessed the impacts of AVs on the traffic network by developing two simulation scenarios of 100% AVs and 100% RVs. The results revealed that AVs positively impacts the network operation, especially when the network is congested (e.g., peak period). The AV scenario suggested improvements in the roadway capacity, average travel speed, and travel time of the network. Hoogendoorn et al. (*16*) analyzed the effect of vehicle automation on the traffic flow efficiency and indicated that although AVs influence traffic flow in the future, the effect of other roadway users should not be neglected.

Additionally, depending on the level of automation, the behavior of the drivers of both AVs and RVs in the vicinity of the AVs impact the performance of the AVs and as a consequence the traffic flow.

Considering the effect of vehicle automation on the capacity of the freeways, Karaaslan et al. (*61*) indicated that a substantial improvement could occur through implementing automated platooning, compare to 1,800-2,200 vphpl for manually driven vehicles. Vanderwerf et al. (*62*) also confirmed the effect of automation on a considerable capacity increase of the freeways. The capacity can be increased from 2,100 vphpl to 2,900 vphpl by implementing 20% Adaptive Cruise Control (ACC), 60% CACC, besides 20% manual driven vehicles. The effect of combining different MPRs of connected vehicles (CVs) and AVs with a flat 10% RVs on the fundamental diagram indicated that as long as the number of AVs is more than CVs, the scatter in a fundamental diagram is negligible. However, by increasing the number of connected vehicles, the scatter of the fundamental diagram increases dramatically (*63*, *64*).

A few studies evaluated the effect of dedicating a lane to CAVs. Ye and Yamamoto (*65*) assessed the impact of dedicating a lane to CAVs under different MPRs. The results indicated that allocating a lane for CAVs improves the traffic throughput only for the medium density range of traffic volume. Laan and Farokhi Sadabadi (*66*) also showed that in case of allocating a lane to AVs on the multilane freeways, the overall performance of the freeways improves when the MPR is increased to 30%, 40%, or 50% depending on the other AV behavioral factors and assumptions. Another study indicated that the MPR of AVs, as well as the lane policies on accommodating AVs, RVs, or both, affect the capacity of the roadways (*67*).

Nilsson et al. (*68*) focused on the lane change of AVs. They developed an algorithm for lane change maneuvers to find an appropriate inter-vehicle traffic gap and time instance for AVs. The results indicated the capability of the proposed algorithm to improve traffic operation. Sun et al. (*69*) proposed an intersection operation algorithm, called maximum capacity intersection operation scheme with signals, in which intersection capacity will be maximized by using all the available lanes simultaneously in addition to optimizing signal green time dynamically. The results indicated that this approach could double the capacity of an intersection. A study evaluated the effect of conventional, connected, and automated vehicles on the operation of an isolated signalized intersection. Various total flows, demand ratios, and MPRs were simulated, and the performances were compared to an actuated signal control algorithm. Eventually, an algorithm was proposed for connected vehicles. Based on the results, the total delay and number of stops showed an evident decrease and improved performance of the intersection (*70*). Mousavi et al. (*71*) also assessed the performance of AVs in the proximity of an unsignalized intersection and indicated that AVs have a superior capability in relieving traffic congestion significantly in the vicinity of a driveway in moderate LOS.

To improve the operation of automated vehicles at freeway merging areas, Letter and Elefteriadou (*72*) developed an algorithm to search for the best and optimized route for AVs to follow to perform the merging maneuver. The results of evaluating a simulated merging segment indicated that the algorithm enhances traffic, compared to RVs situation, by reducing travel time and increasing travel speed for AVs, and under the congested situation, it provides safe merging of the vehicles. The results of Talebpour and Mahmassani (*1*) revealed that CAVs improve string stability and increase throughput under certain MRRs.

Also, it is indicated that the vehicle automation performs better in preventing shockwave formation.

Although the AVs are known to be beneficial and influence daily travel significantly, many studies speculate that the system-level effects at low MPRs will be minimal. Talebpour et al. (*73*) explored the effect of reserving a lane for AVs on travel time reliability and traffic flow dynamics. Evaluation of three different scenarios, including 1) mandatory use of the reserved lane by AVs, 2) arbitrary use of the reserved lane by AVs, and 3) limiting AVs to operate autonomously in the reserved lane, indicated that the optional use of the reserved lane for AVs without imposing any limitation could improve the operation of the network and reduces traffic congestion.

In summary, the literature has explored the operational benefits of AVs from various aspects. The results indicated that AVs are capable of improving roadway capacity and throughput compared to the CVs, especially at higher AV MPRs and congested traffic conditions. However, the effects of other roadway users in mixed traffic environments should not be neglected.

2.3. Effects of LT Signal Phasing on Crash Risk

LT movements are known as the highest-risk movement at the intersections (*74*). The safety concerns arise from three different sources of potential conflicts, including opposing through, the same direction through, crossing vehicles, and pedestrians (*75*). In addition, performing a left-turning maneuver requires the evaluation of both available gaps and the speed of the opposing through vehicles. These complications make the left turn movements as one of the most critical maneuvers at signalized intersections that need

special attention and management (*76*). The Manual on Uniform Traffic Control Devices (MUTCD) (*77*) indicates that there are four designs for left-turning movements at a signalized intersection, including protected, permitted, protected/permitted, and variable mode. Determining an appropriate LT signal phasing is dependent upon various factors such as turning traffic volume, opposing through volume, number of lanes, approach speeds, pedestrian crossing, sight distance, and crash experience (*75*).

Amiridis et al. (*76*) indicated that even though the protected only left-turning signal provides the highest level of safety, it also can increase delay and congestion. Federal Highway Administration (FHWA) (*74*) conducted a before-and-after study for three signalized intersections by converting the permitted LTs to protected-permitted LTs. The results indicated that the total number of crashes decreased by 32 percent, while the head-on crashes declined by 84 percent. However, Chen et al. showed that converting permitted LT signal phasing to protected-permitted does not result in a significant reduction in the number of crashes, but protected LT signal phasing reduces LT crashes and pedestrian crashes significantly (*78*).

2.4. Effects of Speed Limit on Crash Risk

Speed plays an important role in traffic safety as it influences both crash occurrence and crash severity (*79*, *80*). The speed limit is known as one of the factors that are fed into a driver's speed choice (*81*). Previous studies focused on the impacts of increasing the speed limit on the crash severity and consistently reported that higher speed limits result in more severe crashes (*82*, *83*). In addition, Islam and El-Basyouny studied residential areas and reported that decreasing the speed limit improves traffic safety in these areas (*84*). In

general, there are not many studies that assessed the effects of the speed limit on the crash risk in urban areas, and specifically at signalized intersections.

2.5. Effects of Vehicle Size on Crash Risk

The effects of vehicle size on the safety of roadways have not been assessed extensively. National Highway Traffic Safety Administration (NHTSA) (*85*) report on the passenger vehicle occupant fatality rates indicated that by decreasing the vehicle size, the fatality rate increases. While compact cars have the highest fatality rate, large vans are associated with the lowest fatality rate. Kahane (*86*) reported consistent results and showed that a reduction in the weight and size of passenger cars leads to an increase in the fatality rate.

2.6. Effects of Cycle Length on Crash Risk

The effects of cycle length on the safety and number of crashes has not been studied in detail. Stevanovic et al. (*78*) suggested optimizing signal length in order to improve traffic safety. However, on the other hand, the impact of cycle length on the traffic operation, including delay and queue spillback, has been evaluated in several research studies (*87–89*). In this research, the effects of cycle length will be analyzed on the safety of conventional and AVs.

2.7. Detecting Safety-Critical Events in Simulation Environments

Generally, traffic simulation models do not result in any crashes. However, there are SSMs that could be implemented to determine the number of risky driving maneuvers in a

simulation environment, and hence assess the overall level of safety. The following two sections cover the two general categories of SSMs that could be used to assess the safety of the simulation scenarios.

2.7.1. Driving Volatility Measures

More recently, a new term has been introduced as the drivers' volatility that captures the aggressiveness of the drivers (*90–94*). Driving volatility captures the variation in instantaneous driving decisions on a roadway segment or at an intersection (*95*, *96*). Wali et al. (*91*) and Khattak and Wali (*93*) proved that volatility in instantaneous driving maneuvers represent unsafe traffic situations and could potentially result in incidents/crashes. Various studies introduced and evaluated different driving volatility measures, including standard deviation, coefficient of variation, percent of events over a threshold for speed, acceleration/deceleration, and positive/negative jerk events (*58*, *91*, *93*, *95*). Moreover, the volatility measures could be used for both longitudinal and lateral driving movements (*90–93*); however, most of the studies have more focused on the longitudinal movements of the vehicles, rather than the lateral driving volatilities. Previously conducted studies indicated that high driving volatility measures are associated with a lower level of traffic safety (*58*, *93*, *97*).

2.7.2. Driving Jerk

Recently, several studies have sought to use driver behavior on roadway segments as an explanatory variable for crash frequency, e.g. (*98*, *99*). Rates of change of deceleration (i.e., "jerks") are one of the potential SSMs to detect roadways black spots

(*100*, *101*). Previous studies indicated that the higher rate of jerk events are associated with a higher number of crashes on both segments and signalized intersections (*7*, *38*, *43*, *73*, *98*, *99*). Studying jerk is advantageous over acceleration/deceleration because incidents with milder braking reactions (low deceleration and high rate of change of deceleration) occur more frequently than the events with a high deceleration (zero or small jerk), which enables detecting crash-prone locations earlier with a higher level of accuracy (*101*).

Bagdadi and Varhelyi (*98*) used 10 Hz GPS data from 166 private vehicles to analyze jerks. By using a critical jerk threshold of -32.4 ft./ s³ (-9.9 m/s³), the results indicated that drivers with more jerk events above the threshold are more likely to have a history of self-reported crashes. Bagdadi conducted another study to develop a method for detecting crash-prone locations by using -1 g/s as the critical jerk threshold (*102*). Pande et al. (*100*) studied an uninterrupted traffic flow and proved that there is a high correlation between the number of jerky maneuvers and the total number of crashes on quarter-mile segments. In addition, Mousavi et al. (*43*, *101*) and Wolshon et al. (*97*) conducted a segment-based analysis on multiple drivers on the segments of interrupted traffic flow roadways. By using a lower jerk threshold of -2.5 $ft/_{s3}$ on 3 Hz GPS data, they indicated that segments with a higher number of jerks over the threshold were experiencing a higher crash rate as well.

2.7.3. Surrogate Safety Assessment Model (SSAM) as an SSM

The safety of any roadway network and facilities is often evaluated through tracking and analyzing the historical crash data. Due to the infrequency and random nature of crashes, it takes a longer time to obtain a reasonable sample size and find crash-prone

locations that need remedies (*101*, *103*). Therefore, there are various SSMs available to detect the potential conflicts of the vehicles on a roadway before numerous crashes, fatalities, and injuries occur. Time-to-collision (TTC), gap time (GT), deceleration rate (DR), the proportion of stopping distance (PS.D.), and post encroachment time (PET) are such variables (*8*).

TTC can be an indicator of the crash risk and is defined as the required time to collide if two vehicles continue moving on the same path at the same speed (*104*, *105*). Therefore, low TTC associates with a higher risk of crashes, and high TTC represents a lower crash risk (*105*). *Equation (1)* demonstrates how to calculate TTC mathematically based on *Figure 6* (*106*).

A vehicle on a section of a roadway only has one TTC value. However, the TTC on a junction is calculated based on one or more vehicles coming from the arms of the intersection (*105*).
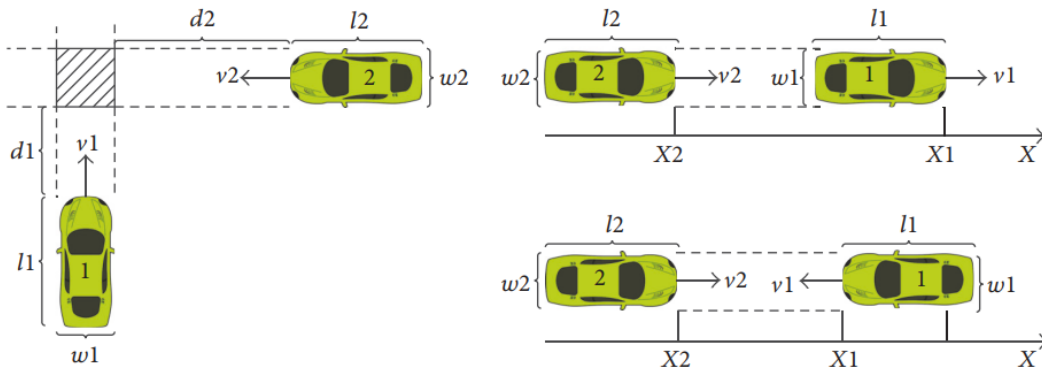
Figure 6. Time-To-Collision Definition and Variables. Reprinted from (*8*)

SSAM is a software that is used for analyzing SSMs obtained from various micro-simulation models to identify the number of potential conflicts (*107*, *108*). SSAM is

capable of detecting three different types of conflicts, including rear-end, lane-changing, and crossing. *Figure 7* depicts the angle diagram used by SSAM to distinguish different types of conflicts (*108*).

$$
TTC = \begin{cases}
\dfrac{d_2}{v_2} & \text{if } \dfrac{d_1}{v_1} < \dfrac{d_2}{v_2} < \dfrac{d_1 + l_1 + w_2}{v_1} \text{ (side)} \\[3ex]
\dfrac{d_1}{v_1} & \text{if } \dfrac{d_2}{v_2} < \dfrac{d_1}{v_1} < \dfrac{d_2 + l_2 + w_1}{v_2} \text{ (side)} \\[3ex]
\dfrac{X_1 - X_2 - l_1}{v_2 - v_1} & \text{if } v_2 > v_1 \text{ (rear-end)} \\[3ex]
\dfrac{X_1 - X_2}{v_2 + v_1} & \text{(head-on)}
\end{cases}
\qquad (1)
$$

Where:

v_1 and v_2: vehicle speeds;

l_1 and l_2: vehicle lengths;

w_1 and w_2: vehicles widths;

X_1 and X_2: vehicle positions;

d_1 and d_2: distance to conflict areas.
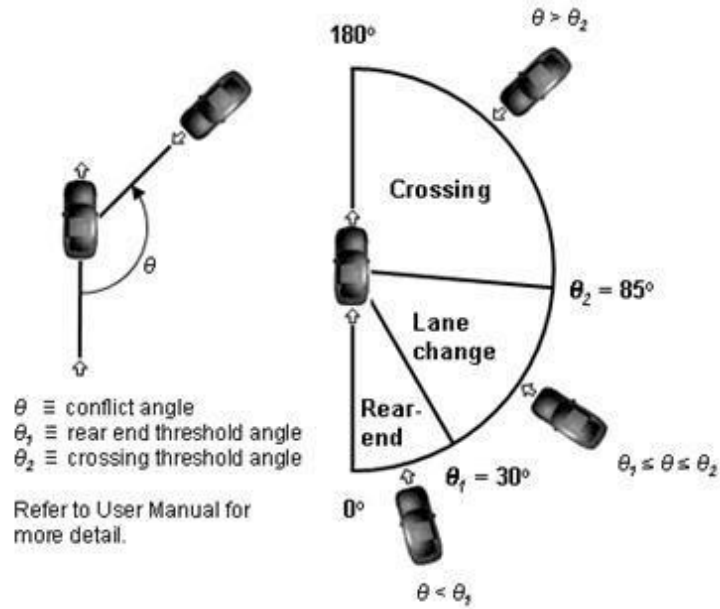
Figure 7. SSAM Conflict Angle Diagram. Reprinted from (*108*)

Several previous studies implemented SSAM to identify either the total number of conflicts or a specific type of conflict for traffic simulation models (*8*, *22*, *56*, *109–111*).

2.8. Chapter Summary

This chapter has summarized the literature on related research topics. It indicates that overall, AVs are capable of improving traffic safety and operation. However, not many research studies have extensively evaluated the traffic safety of mixed traffic environments.

Moreover, as indicated, there are very few studies that assessed the safety effects of the speed limit, LT signal phasing, and signal cycle length. Still, none of them considered AVs and mixed traffic environments. In addition, although the safety of vehicle size for

RVs has been evaluated, the safety effect of AV size has been neglected. In fact, AV size is a crucial factor to be evaluated before extensively employing AVs.

Lastly, various SSMs, including jerk, longitudinal driving volatility measures, lateral driving volatility measures, and SSAM, were explained to indicate how they have been implemented in the previous research studies. Although jerk and longitudinal driving volatility measures have been studies, lateral driving volatility measures need more investigations. In other words, none of the research studies determined how various non-infrastructure variables influence these SSMs. The next chapter describes the methodology used for this dissertation.

CHAPTER III

METHODOLOGY

This chapter describes the development of the simulation environment, designing various scenarios, processing simulation results, analyzing data, conducting safety analyses by assessing longitudinal and lateral driving volatility measures, and developing safety prediction models using machine learning. Lastly, the implementation of the SSAM is explained to find the number of conflicts, compare them with the number of jerks, and determine if there is an associative relationship.

3.1. Introduction

As mentioned in Chapter 1, analyzing the safety performance of mixed traffic environments at a signalized intersection before employing AVs in the real world is crucial. Since this analysis is infeasible due to the lack of real-world data, VISSIM 10.00 (*112*), as a micro-simulation software, was used to simulate and replicate an existing signalized intersection located in College Station, TX. In addition to assessing various AV MPRs, the effects of several other non-infrastructure variables, including speed limit, LT signal phasing, and signal cycle length, as well as AV size were evaluated. The following sections expand on the methodology, including developing and running the simulations scenarios, assessing the safety of each scenario using various driving volatility measures and jerks, and comparing the number of SSAM conflicts and jerks.

3.2. Traffic Simulation Design and Development

VISSIM 10.00 (*112*) was used as a micro-simulation platform to develop the simulation environments for assessing the objectives of the study. The followings provide details on the process of establishing the scenarios.

3.2.1. Base Simulation Environment

Since the purpose of the dissertation is to evaluate and compare the effectiveness of various non-infrastructure treatments, only one complicated intersection was selected for the analyses. *Figure 8* depicts the base signalized intersection scenario that was developed using VISSIM 10.00. This signalized intersection replicates an existing intersection in College Station, TX, and has been selected due to its complexity of being facilitated with exclusive right-turn (RT) and LT lanes for all the approaches. In addition, this intersection has various speed limits along each approach.

The simulated signalized intersection is a four-leg intersection. The major road has three through lanes in addition to a single exclusive LT lane and an exclusive RT lane for each approach. The minor road is facilitated with two through lanes as well as a dual LT lane and a single RT lane for each approach. The speed limit was set to 40 mph for the eastbound and westbound approaches of the intersection, 45 mph for the southbound leg, and 25 mph for the northbound leg.

Figure 8. Signalized Intersection Scenario in VISSIM

3.2.2. *Input Traffic Volume*

The next step in establishing the simulation environment is to input the traffic volume by each movement at the intersection. For the purpose of this dissertation, the peak hour traffic volume data were used. In fact, during the peak hour period, the highest number of risky driving maneuvers is expected because of the high level of traffic exposure. Due to the lack of traffic volume per movement for 2019, the collected data in 2015 were used to estimate the 2019 traffic volume data for each movement. To do so, the flat growth rate, presented in *Equation (2)*, was applied to the 2015 traffic data:

$$ADT_{Current} = ADT_{Previous}(1 + i)^n \qquad \textbf{(2)}$$

Where,

$ADT_{Current} =$ ADT for the year 2019

$ADT_{Previous} =$ ADT for a previous year (in this case, it is 2015)

$i =$ Growth rate (defined in *Equation* (**3**))

$n =$ Number of years

The value of the growth rate, i, was calculated from the previous years' overall traffic volume data, regardless of the per movement distribution, using *Equation (3)*.

$$Growth\ Rate = \frac{\frac{(ADT_n - ADT_{n-m})}{ADT_{(n-m)}} \times 100}{n - m} \tag{3}$$

Where,

$ADT_n =$ ADT for present

$ADT_{n-m} =$ ADT for past

Eventually, the traffic volume for every movement at the intersection in 2019 was calculated and used as the input volume data for the VISSIM scenario.

3.2.3. Driving Behaviors

VISSIM offers two car-following models of Wiedemann 74 and Wiedemann 99. Wiedemann 74 is implemented to replicate driving behaviors in urban environments and merging areas, while Wiedemann 99 is used for freeway segments with no merging area (*112*, *113*). Therefore, since the study intersection is located in an urban environment, the

behavior of RVs was modeled using the Wiedemann 74 car-following model with the default values for the predefined parameters.

VISSIM 10.00 does not offer AV as a predefined vehicle class. Therefore, different lateral and longitudinal controls of Wiedemann 74, including average standstill distance, additive part of safety distance, and multiplicative part of safety distance, were adjusted to replicate the behavior of the AVs (*112*). It is notable that due to the time constraints and since it was out of the dissertation's scope, a sensitivity analysis was not conducted to determine how the results would change by modifying the AV parameters. *Table 2* presents the driving behaviors that were adjusted to reflect AVs (*112*). It is notable that AVs in this study are only automated and not connected; therefore, there is no communication between the vehicles.

Table 2. Wiedemann 74 Driving Behavior Parameters

Driving Behavior	AV Value
Average Standstill Distance	1.0
Additive Part of Safety Distance	1.5
Multiplicative Part of Safety Distance	0.0

3.2.4. Calibrating and Validating the Base Scenario

Now, with having the input traffic volumes and driving behaviors, the base simulation scenario with fully RVs (i.e., AV MPR of 0%) should be calibrated and validated. The calibration and validation are required to ensure the acquisition of the best possible match between the simulation performance and the field measurements (*114*). The base scenario that replicates the real-world signalized intersection has protected LT signal

phasing, the cycle length of 135 seconds, AV MPR of 0%, and different speed limits for each approach, as indicated in section *3.2.1*.

The variable that was used to calibrate the model was travel time. As recommended by the VISSIM calibration guidelines (*115*), an acceptable error of 15% between the actual and simulated travel times was considered. To this aim, travel time data were collected at the study intersection for ten runs during the peak-hour period to calibrate the base scenario. Since the intersection is located nearby the campus of Texas A&M University (TAMU), College Station, TX, the class schedules affect the traffic pattern and peak-hour time. At TAMU, Tuesdays and Thursdays follow similar class schedules, and the rest of the weekdays follow another plan. Hence, to ensure accuracy, the travel time data were collected on a similar day of the week as the traffic volume data were collected.

Ultimately, the average collected travel time was used to calibrate and validate the based simulation model with AV MPR of 0%. It is worth mentioning that due to the unavailability of the crash data, the base model's number of SSMs (jerks, conflicts, etc.) could not be compared to the number of crashes to calibrate the model further. Eventually, the calibrated and validated model was further used to develop other scenarios to represent different LT signal phasing, signal cycle lengths, speed limits, and AV sizes. Each model was also considered at seven different AV MPRs, starting from 0%, representing an entirely conventional vehicle environment (base model), up to 100%, representing a fully AV environment. The next section provides various levels of the study variables that were simulated.

3.2.5. Simulated Variables and the Associated Levels

Abovementioned, in addition to the various AV MPRs, other variables, including different speed limits, LT signal phasing, signal cycle lengths, and AV sizes, were evaluated. It is worth mentioning that the speed limit is a surrogate for the operating speed, geometry of the intersection approaches, and the level of traffic congestion. *Table 3* presents the various levels of each simulated variable that were used to develop different simulation models.

Table 3. Defining Levels of the Simulated Variables

Variable	Levels	Details and Descriptions
Signal Cycle Length	Shorter	75 seconds
	Actual	135 seconds
	Longer	193 seconds
Speed Limit	Lower[1]	EB and WB: 35 mph SB: 40 mph NB: 20 mph
	Actual[2]	EB and WB: 40 mph SB: 45 mph NB: 25 mph
	Higher[3]	EB and WB: 45 mph SB: 50 mph NB: 30 mph
AV Size	Smaller	13.00 ft
	Larger	15.1 ft
LT Phasing	Protected (Actual)	All the LTs are protected
	Protected-Permitted	All the LTs are protected-permitted
	Permitted	All the LTs are permitted
AV MPR	0%	Fully RV environment
	20%	AV MPR of 20%
	40%	AV MPR of 40%
	60%	AV MPR of 60%
	80%	AV MPR of 80%

Table 3 Continued

Variable	Levels	Details and Descriptions
	90%	AV MPR of 90%
	100%	Fully AV environment
[1]In the graphs, this speed limit is indicated as 32.5 mph.		
[2]In the graphs, this speed limit is indicated as 37.5 mph.		
[3]In the graphs, this speed limit is indicated as 42.5 mph.		

3.2.6. *Number of Runs per Each Simulation Scenario*

One of the critical considerations in managing simulations runs is defining the minimum number of runs per each scenario. In fact, since simulations models are stochastic in nature, each scenario should be run more than once with various random seeds to ensure the accuracy and reliability of the outputs and results. To this aim, *Equation (4)* was used to determine the minimum required number of runs that should be conducted to reach to a 95% confidence interval as well as an acceptable error rate of 10% (*115*):

$$n = \left(\frac{S \times Z}{\mu \times \mathcal{E}} \right)^{2} \tag{4}$$

Where,

n = minimum sample size ;

S = standard deviation of a sample data;

$Z = \text{z} - \text{statistic, which is 1.96 for a 95\% confidence interval;}$

μ = mean of the sample data;

$\mathcal{E}$ = acceptable error rate.

To determine the number of runs, the variable travel time was used. The travel time data were measured from the point where vehicles entered the west-bound leg and went towards the west leg until exiting the network. Five sample runs were conducted on the calibrated and validated base case scenario with travel times of 57, 47, 37, 49, and 45 seconds; mean of 47 seconds; and a standard deviation of 7.21 seconds. The following shows the calculation to find the minimum required number of runs:

$$n = \left(\frac{S \times Z}{\mu \times \mathcal{E}} \right)^2 = \left(\frac{7.21 \times 1.96}{47 \times 0.1} \right)^2 = 9.04 \approx 10 \; runs$$

Therefore, ten iterations were required for each scenario to be able to reach a 95% confidence interval and an error rate of 10%. Also, each simulation iteration was run for 70 simulation minutes, considering the first 10 minutes as the warm-up period.

It is notable that the time step was set to 10 data points per second, i.e., 10 Hz, to capture the movements of the vehicles in more detail, especially for AVs that have shorter headways and quicker reaction times (8).

3.2.7. Full Factorial Design of the Simulation Scenarios

By defining all the levels of the study variables and the required number of runs per each scenario, the full factorial design was adopted to develop the VISSIM scenarios. By full factorial design, all the possible combinations of all the levels of the study variables

were investigated throughout a complete replicate (*116*). *Figure 9* depicts the full factorial design of the simulation scenarios.

As indicated in the figure, each and every level of the study variables was evaluated in combination with all possible levels of the other variables. Every developed combination was run at seven AV MPRs of 0%, 20%, 40%, 60%, 80%, 90%, and 100%. In addition, each scenario was run ten times to account for the stochastic nature of the scenarios. This full factorial design resulted in a total of 3,850 simulation runs. Eventually, the outputs were investigated and assembled for further analyses. The following section covers the data manipulation and preparation process.

3.3. Data Cleaning and Processing

After developing and running all the 3,850 simulation runs, the output data were cleaned and manipulated to acquire the final dataset for calculating the required variables as well as conducting statistical analyses.

As the first step, the VISSIM output files for each simulation run were merged to have vehicle type, location, and time information in one single file to calculate speed and acceleration/deceleration using *Equation (5)* and *Equation (6)*, respectively.

$$v = \frac{dx}{dt} \tag{5}$$

$$a = \frac{dv}{dt} \tag{6}$$

41
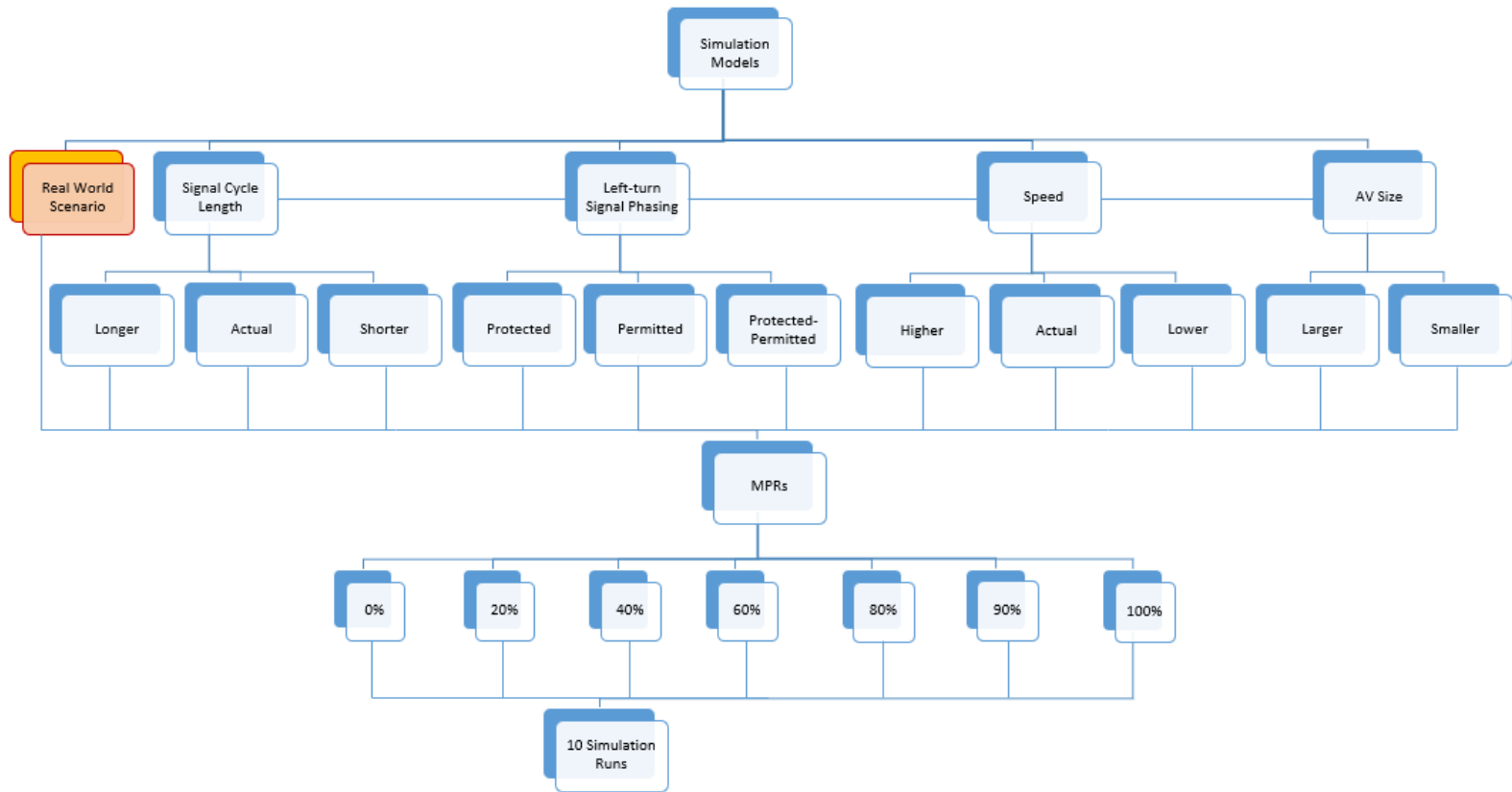
Figure 9. Overview of the Full Factorial Micro-Simulation Model

Where,

v = speed ($\frac{ft}{s}$);

dx = change in location between successive observations (ft);

dt = change in time between two successive observations (second);

a = acceleration/deceleration ($\frac{ft}{s2}$);

dv = change in speed between successive observations ($\frac{ft}{s}$);

The abovementioned variables were calculated for each 0.1 second time interval (i.e., data frequency of 10 Hz) and every single vehicle separately.

Since the study site was an intersection at which the vehicles were making turning maneuvers frequently, further data investigations were required to ensure the accuracy of the data. In other words, if a vehicle made any turning maneuvers, the calculated speed, acceleration, and deceleration values were erroneous due to the abrupt change in location coordinates. Therefore, the incorrect points were removed from the dataset. Each variable was monitored separately to be able to only keep the acceptable range of values.

By having the final dataset, the variable jerks and other driving volatility measures could be calculated to develop safety models. The following sections consider the calculations of the driving volatility measures and jerks as well as developing the safety models.

3.4. Driving Volatility Measures

Driving volatility measures are a comprehensive set of variables that thoroughly quantify traffic safety by considering various aspects. This section covers the methodology

that was implemented to calculate various driving volatility measures to be used for evaluating traffic safety as well as quantifying safety through developing models.

3.4.1. Calculating Driving Volatility Measures

To quantify the safety of the simulation scenarios, various measures of volatility were calculated. The driving volatility measures that were considered include standard deviation (S.D.), coefficient of variation (C.V.), and percent of values over a threshold for any variable.

Standard deviation represents variation in a dataset, and the coefficient of variation is a measure of dispersion. Also, the percent of values over a threshold shows the number of observations beyond a common threshold, which is calculated and defined based on each dataset and for each variable (*117*). Although these measures are frequently used variables, the implications in the field of traffic safety is not a common practice (*58*). In this dissertation, the driving volatility measures were calculated for speed, acceleration, deceleration, positive jerk, and negative jerk separately.

In general, vehicles have two types of movements, longitudinal and lateral. Since longitudinal movements are more critical, lateral movements have been neglected from traffic safety analyses. However, this dissertation is one of the leading works that considers both longitudinal and lateral movements simultaneously to evaluate traffic safety. Hence, all the driving volatility measures were calculated for both lateral and longitudinal speed, acceleration, deceleration, positive jerk, and negative jerk. Eventually, there were a total of 30 driving volatility measures for the final analyses.

Table 4 was used to calculate the measures of volatility for each variable and every single vehicle (*91*, *117*). The marked cells represent the driving volatility measures that were calculated and investigated.

After calculating the driving volatility measures for each vehicle, the network-level aggregated values for every simulation run was calculated and determined for the safety analyses, regardless of the driving behaviors of single vehicles. It is notable that since at signalized intersections vehicles come to a full stop frequently due to the presence of a red signal phase, the data with speed equal to zero were removed from the volatility measure analyses to avoid wrong results.

Table 4. Measures of Driving Volatility

Driving Volatility Measure	Formula	Performance Measure Variables for Longitudinal and Lateral Movements				
		Sp	Acc	Dec	Pos Jerk	Neg Jerk
Standard Deviation	$Sd = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{\mu})}{n}}$	×	×	×	×	×
Coefficient of Variation	$Cv = \dfrac{Sd}{\bar{\mu}} \times 100$	×	×	×	×	×
Percent of Outliers	$\%OL = \dfrac{Count(x_i > Threshold)}{n} \times 100$ $(Threshold = \bar{\mu} \pm 2Sd)$	×	×	×	×	×
Sp: Speed (ft/s); Acc: Acceleration (ft/s2); Dec: Deceleration (ft/s2); Pos Jerk: Positive Jerk (ft/s3); Neg Jerk: Negative Jerk (ft/s3)						

3.4.2. Developing Driving Volatility Measure Models

After calculating the aggregated network-level driving volatility measures by merging the measures over each simulation run, these measures were used to develop safety models. The Generalized Linear Model (GLM) regression was implemented to develop safety models and determine how the simulated variables affect each driving volatility measure. Also, *Equation (7)* was used to calculate the goodness of fit for each model (*118*, *119*).

$$Model\ Goodness\ of\ Fit = 1 - \frac{Residual\ Deviance}{Null\ Deviance} \tag{7}$$

Where,

$Null\ Deviance$

$= indicates\ how\ well\ the\ response\ variable\ is\ predicted\ by\ only\ using\ the$

$intercept;$

$Residual\ Deviance$

$= presents\ how\ well\ the\ response\ variable\ is\ predicted\ by\ including\ all\ the$

$predictors.$

In summary, these driving volatility measures cover a wide range of traffic safety aspects by considering both the longitudinal and lateral movements of the vehicles.

3.5. Developing Safety Models Based on Jerks Using ML Algorithm

To develop more advanced statistical models, jerk was used as the dependent variable in ML safety models. As mentioned earlier, the literature has indicated that the frequency of jerky driving maneuvers at a segment/location is highly correlated with its frequency of the crashes (*100*, *101*). The previous research studies also proved that jerk is a promising SSM in detecting crash-prone locations (*98*, *102*). Hence, the percentage of jerks could be used as an SSM to determine safety. The following section explains the process of calculating jerk.

3.5.1. Calculating Jerk

The final dataset containing speed, acceleration, and deceleration values was used to calculate jerk using *Equation (8)*.

$$j = \frac{da}{dt} \tag{8}$$

Where,

j = jerk, defined as the rate of change of acceleration/deceleration $\left(\frac{\text{ft}}{\text{s}^3}\right)$

da = change in acceleration between two successive observations $\left(\frac{\text{ft}}{\text{s}^2}\right)$

As for the speed, acceleration, and deceleration, the jerk was also calculated for every 0.1 second time interval and every single vehicle separately.

To implement the jerk as an SSM, a jerk threshold should be defined to distinguish between normal and jerky driving maneuvers. As the literature indicates, various studies implemented different jerk thresholds to detect risky driving maneuvers; however, the most similar research study with GPS data at a rate of 10 Hz, used $1\ ^g\!/_s\ (\approx 32.17\ ^{ft}\!/_{s^3})$ as the jerk threshold (*98*, *102*). Hence, due to the similarity in the frequency of the data, $32.17\ ^{ft}\!/_{s^3}$ was used as the jerk threshold to determine jerky driving maneuvers in this dissertation.

3.5.2. *Developing the ML Algorithm*

Eventually, by having the final dataset, including the jerk values, the ML algorithm could be applied using the GLM to develop safety models and evaluate the safety effects of the simulated variables. The percent of total jerky driving maneuvers, percentage of AV jerks, and percentage of RV jerk were considered as the dependent variables.

It is worth noting that the percentage of the total number of jerky driving maneuvers and the percentage of negative jerks over a threshold in the previous section are different. In fact, in this section, a global jerk threshold was used to determine jerky driving maneuvers for all the simulation runs; however, in the previous section, each simulation run had its jerk threshold according to its dataset using the equation represented in *Table 4*.

Figure 10 represents the ML approach to develop a model. The flow chart illustrates that the available data should be divided into two categories of training and test datasets.
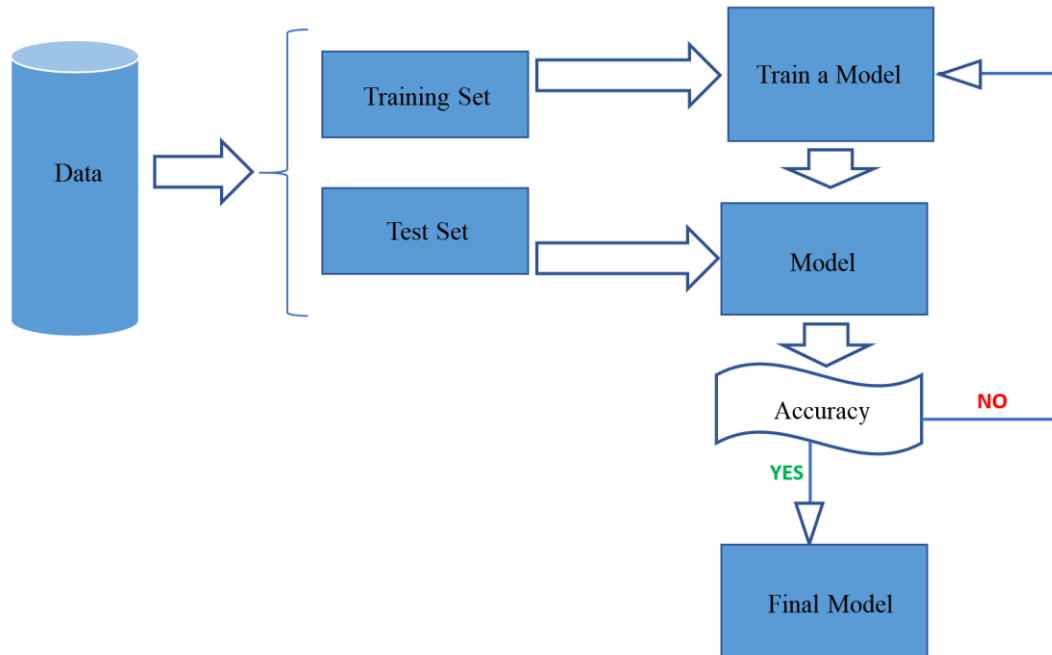
Figure 10. ML Methodology

The training dataset will be utilized to train the model, and the test dataset will be used for tuning the model and verifying its quality and accuracy by evaluating the error associated with it.

To determine the error term, the loss function is used. The overall objective is to minimize the error term to be able to predict any sets of data as accurately as possible. The general form of the loss function is shown in *Equation (9)*, in which the difference between the actual and estimated Y values is of interest:

$$L\left(Y - f_{\hat{w}}(x)\right) \tag{9}$$

For the regression model, there are three types of loss functions with different complexities, including:

a) Mean Square Error (MSE): This function only considers the average magnitude of the error irrespective of the direction, *Equation (10).*

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \qquad (10)$$

Where,

y_i = actual value;

$\hat{y}_i$ = predicted/estimated value;

n = number of data points.

b) Mean Absolute Error (MAE): This loss function also calculates the magnitude of the error and disregards the direction, *Equation (11)*. However, compared to the MSE, it needs linear programing to measure the error.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \qquad (11)$$

c) Mean Bias Error (MBE): The MBE is less common and the least accurate loss function. However, using this loss function can indicate the direction of the bias.

$$MBE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \qquad (12)$$

Eventually, the ML algorithm looks for the best fit with the minimum error that represents the final function. Lastly, the test set was used to evaluate the tuned model. The final algorithm to estimate the number of potentially risky driving events at the signalized intersection was presented in the form of:

$$y_i = \widehat{w_0} + \widehat{w_1}(x_1) + \widehat{w_2}(x_2) + \widehat{w_3}(x_3) + \cdots + \widehat{w_n}(x_n) + \widehat{w_{n+1}}(x_1)(x_2)$$
$$+ \cdots + \widehat{w_{n+2}}(x_n)(x_{n-1}) + \varepsilon$$

(13)

Where,

y_i = estimated y value (dependent variable);

$\widehat{w_0}, \widehat{w_1}, \widehat{w_2}, \dots, \widehat{w_n}$ = coefficients or weights;

$x_1, x_2, \dots, x_n$ = predictors (independent variables);

$(x_1)(x_2), \dots, (x_n)(x_{n-1})$ = interaction terms;

ε = error term.

Three different safety models were developed based on the vehicle type. The models considered the percentage of jerks for only RVs, the percentage of jerks for only AVs, and the percentage of the total jerks for both RVs and AVs, as the dependent variables, separately.

3.6. Number of Jerks vs. Number of Conflicts Using SSAM

SSAM has been implemented widely to evaluate the safety of simulation scenarios based on output trajectory files (*8*, *56*, *109*, *110*, *120*). However, the accuracy of this

software program has been controversial among researchers. The purpose of this section is to evaluate the correlation between the number of jerks and SSAM conflicts and determine if they are correlated and could be used interchangeably.

The following parts illustrate the process of finding the number of conflicts using SSAM and the correlation test.

3.6.1. Finding Total and Rear-End Conflicts Using SSAM

The software SSAM uses a user-defined TTC value to find the number of near-miss events. In other words, the input TTC is considered as the critical TTC to detect unsafe situations where the actual TTC of the vehicles goes below the critical value (*105*). Various studies implemented different TTC values. Archer (*121*) determined a TTC of 1.5 seconds as the critical value for conventional vehicles in urban areas; however, Horst (*122*) used 2.5 seconds. It is worth noting that the critical TTC threshold for the AVs is smaller due to their ability to react to different situations more abruptly compared to the human-driven vehicles. Hence, Morando et al. (*8*) implemented 1.0 second as the TTC for AVs. Mousavi et al. (*109*) defined the value of AV TTC as a percentage of RV TTC according to the VISSIM settings. In other words, as in this dissertation, since both the following distance and standstill distance for AVs were two-third of RVs (*112*), the implemented TTC for AVs was reduced to two-third of RVs (*109*). In this dissertation, the same approach as Mousavi et al.'s (*109*) was used, and 1.5 seconds and 1.0 second were used as the critical TTC values for RVs and AVs, respectively.

Since the TTC values are different for AVs and RVs, it is not practical to find an accurate number of conflicts for mixed traffic environments using SSAM. Also, the

majority of the previous research studies only used one single TTC value for mixed traffic environments. However, Mousavi et al. (*109*) came up with a solution to enable accurate safety analysis of mixed traffic environments by considering each conflict type separately and implement the correct corresponding TTC values for either AVs or RVs. For instance, for rear-end conflicts, if the second vehicle (which hits from the back and is at fault) is an RV, the TTC value of 1.5 seconds was used; and, if the second vehicle is an AV, 1.0 second was implemented. In this dissertation, the same method was used to evaluate rear-end conflicts, which is the most common type of conflict at signalized intersections.

Overall, to be comprehensive, this research used two approaches to find the number of conflicts in SSAM and compare them with the number of jerks. The followings introduce the methods for the analyses:

1) evaluating all the conflicts using 1.5 sec (commonly practiced value) as the TTC; and

2) assessing only rear-end conflicts while considering different TTCs for AVs and RVs.

To these aims, trajectory files that were obtained from the VISSIM outputs were used as inputs for SSAM. According to each method, a TTC value was determined to find the number of conflicts. *Figure 11* provides an SSAM interface at which the trajectory data should be analyzed by defining an appropriate TTC value.
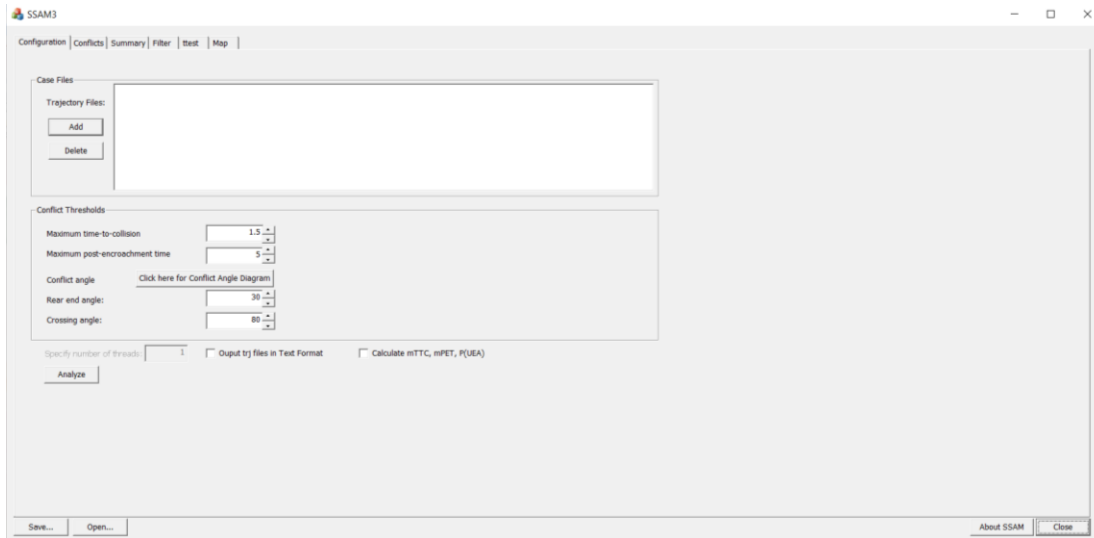
Figure 11. SSAM Interface

For the first approach, the TTC value of 1.5 seconds was directly used in the

SSAM, and the outputs were used for further analysis.

For the second approach, since the conflicts were dependent upon the type of the

second vehicle (vehicle in the back), more data analyses were required. Primarily, the TTC

value of 1.0 second was used to find the number of conflicts using SSAM. Since in the

second approach, only rear-end conflicts were of-interest, other types of conflicts were

excluded from further analysis. In the output files, SSAM only provides the ID of the

vehicles, and not the vehicle types, that were involved in conflicts. Therefore, the SSAM

outputs were joined to another VISSIM output in which both vehicle type and ID were

included. Eventually, by having the vehicle type in the SSAM outputs, the precise number

of rear-end conflicts based on the corresponding TTC values for each vehicle type was

determined.

3.6.2. Correlation Test between The Number of Jerks and Conflicts from

SSAM

Having the total number of jerks as well as the total number of conflicts and rear-

end conflicts from SSAM for each simulation run enabled conducting the correlation test

to determine if these two are correlated and could be used interchangeably.

For evaluating the correlation, Pearson's correlation test was used, which measures

the linear correlation between two variables, and is calculated according to *Equation (14)*

(*123*).

$$\rho\,(x,y) = \frac{cov\,(x,y)}{\sigma_x.\sigma_y} \qquad\qquad \textbf{(14)}$$

Where,

$cov\,(x,y) = $ covariance of x and y $= \dfrac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$; $n = $ sample size

$\sigma_x = $ standard deviation of x;

$\sigma_y = $ standard deviation of y.

Various correlation tests were conducted to determine if any correlation could be

observed between the variables. The results are presented in section *5.3*.

3.7. Chapter Summary

In summary, this chapter has explained the development of the simulation

environment and the calibration and validation process. For each study variable, including

signal cycle length, LT phasing, speed limit, AV size, and AV MPR, various levels were introduced. Eventually, the full factorial design was implemented to develop different simulation scenarios by considering all the combinations of all the levels of all the study variables. Moreover, the minimum required number of runs was calculated to reach an acceptable error rate and confidence interval.

Afterward, extracting, processing, and analyzing data were illustrated thoroughly. For the safety analyses, three different approaches were defined:

1) Considering all the longitudinal driving volatility measures;

2) Analyzing the lateral driving volatility measures; and

3) Developing safety models using ML and jerks as the dependent variable;

Lastly, the process of calculating SSAM conflicts was explained to investigate if there is any correlation between the number of conflicts and jerks.

The following chapter presents the results of descriptive statistics and explanatory analyses of the data.

CHAPTER IV

RESULTS AND DISCUSSIONS: DESCRIPTIVE STATISTICS

This chapter describes the descriptive statistics of the simulated variables and the simulation outputs. Since traffic safety was analyzed from various aspects, the descriptive statistics were also followed a similar structure. Hence, this chapter is divided into two sections of driving volatility measures and ML models (which represents the jerk data).

4.1. Introduction

Each simulation scenario was run ten times to obtain the final data for conducting the statistical analyses. The results of the base scenario indicated that the 95% confidence interval for travel time is [62.51, 63.25] with a mean and standard deviation of 62.88 and 0.60 seconds, respectively.

With having the final dataset, the objectives of this study could be assessed to determine how changing the defined non-infrastructure variables affect traffic safety at the signalized intersection in mixed traffic environments. The following sections present preliminary data analyses by providing descriptive statistics of the data.

4.2. Descriptive Statistics for Driving Volatility Measures

As mentioned earlier, the driving volatility measures were calculated at the network level. In other words, the standard deviation, coefficient of variation, and percent of values over a threshold, as the measures of volatility, were calculated for speed, acceleration, deceleration, positive jerk, and negative jerk for each and every simulation run according

57

to *Table 4*. The following sections provide the descriptive statistics and explanatory

analysis of the longitudinal and lateral driving volatility measures, respectively, using the

R statistical package (*124*).

4.2.1. Longitudinal Driving Volatility

Table 5 presents the descriptive statistics for the longitudinal driving volatility

measures (S.D., C.V., percent of values over threshold) of the study variables (speed,

acceleration, deceleration, positive jerk, and negative jerk).

Table 5. Descriptive Statistics for Driving Volatility Measures

Measure	Mean	S.D.	C.V.	Percent Over Threshold
Longitudinal Speed (ft/s)				
Min	48.40	15.42	-2.13E+17	5.32
Max	56.97	19.34	8.88E+16	7.83
Median	51.57	17.66	500248.6728	6.84
Mean	51.95	17.40	-4.59E+13	6.73
Var	3.29	0.60	1.52E+31	0.21
S.D.	1.82	0.77	3.90E+15	0.45
C.V.	0.04	0.04	-84.95	0.07
Longitudinal Acceleration (ft/s^2)				
Min	0.87	1.91	141.40	5.57
Max	1.65	2.69	241.87	10.22
Median	1.24	2.27	181.04	8.01
Mean	1.24	2.27	186.90	7.95
Var	0.05	0.03	667.59	0.93
S.D.	0.22	0.17	25.84	0.96
C.V.	0.18	0.07	0.14	0.12
Longitudinal Deceleration (ft/s^2)				
Min	-2.73	2.43	-138.76	6.13
Max	-1.86	3.29	-109.44	11.68
Median	-2.37	2.78	-117.56	7.71
Mean	-2.33	2.78	-119.93	7.84
Var	0.04	0.02	45.21	1.13

Table 5 Continued

Measure	Mean	S.D.	C.V.	Percent Over Threshold
S.D.	0.21	0.16	6.72	1.06
C.V.	-0.09	0.06	-0.06	0.14
Longitudinal Positive Jerk ($ft/_{s^3}$)				
Min	3.93	7.16	93.89	2.16
Max	9.01	8.48	182.14	12.09
Median	6.28	8.19	130.52	3.33
Mean	6.28	8.02	134.05	4.95
Var	2.36	0.18	698.93	11.40
S.D.	1.53	0.42	26.44	3.38
C.V.	0.24	0.05	0.20	0.68
Longitudinal Negative Jerk ($ft/_{s^3}$)				
Min	-10.22	8.29	-84.18	4.15
Max	-9.88	8.37	-81.23	4.78
Median	-9.97	8.34	-83.60	4.62
Mean	-9.97	8.34	-83.60	4.60
Var	0.00	0.00	0.03	0.01
S.D.	0.03	0.01	0.17	0.10
C.V.	0.00	0.00	0.00	0.02

To explore more details about the effect of various AV MPR on the distribution of the driving volatility measures, boxplots for all the driving volatility measures are provided in the following section. As mentioned, the driving volatility measures include S.D., C.V., and percent of values above the threshold for speed, acceleration, deceleration, positive jerk, and negative jerk. Moreover, the effects of the study variables (speed limit, signal cycle length, AV size, and LT signal phasing) were also graphically investigated on each driving volatility measure to determine if any pattern could be detected. For the depictions, the x-axis represents one of the study variables (speed limit, signal cycle length, AV size, and LT signal phasing), the y-axis shows one of the driving volatility measures, and the color of the boxes presents the MPR.

As depicted in ***Error! Reference source not found.**,* the standard deviation of the speed indicates that by increasing the cycle length, the standard deviation of the speed increases. However, any increase in the AV MPRs over various cycle length decreases the discrepancy of the speed standard deviation. By increasing the speed limit, the standard deviation of the speed increases. Also, the larger AV size results in higher standard deviations, but increasing MPR decreases the volatility measure. Also, permitted LT phasing indicates to provide the lowest level of speed standard deviation since vehicles require to slow down and stop less frequently, compared to the protected-permitted and protected LT phasing. But it is noticeable that higher AV MPRs globally reduces the speed standard deviation, regardless of the LT signal phasing.

As depicted in the graphs, for some variables, the negative skewness increases by increasing the AV MPR. In other words, at low MPRs, the driving volatility measures do not experience high variations, while increasing the AV MPR results in a negative skewness. The increase in the skewness indicates that AVs are capable of providing safer traffic conditions for some scenarios by having smaller driving volatility measures. In other words, even though the median value might be higher for some variables at higher AV MPRs, the higher skewness indicates the capability of AVs to provide safer traffic environments compared to the RVs.

For some driving volatility measures of some study variables, bimodality or multimodality could be observed. This bimodality/multimodality is due to the effects of other non-infrastructure variables that could be captured simultaneously in the safety models presented in the next chapter.
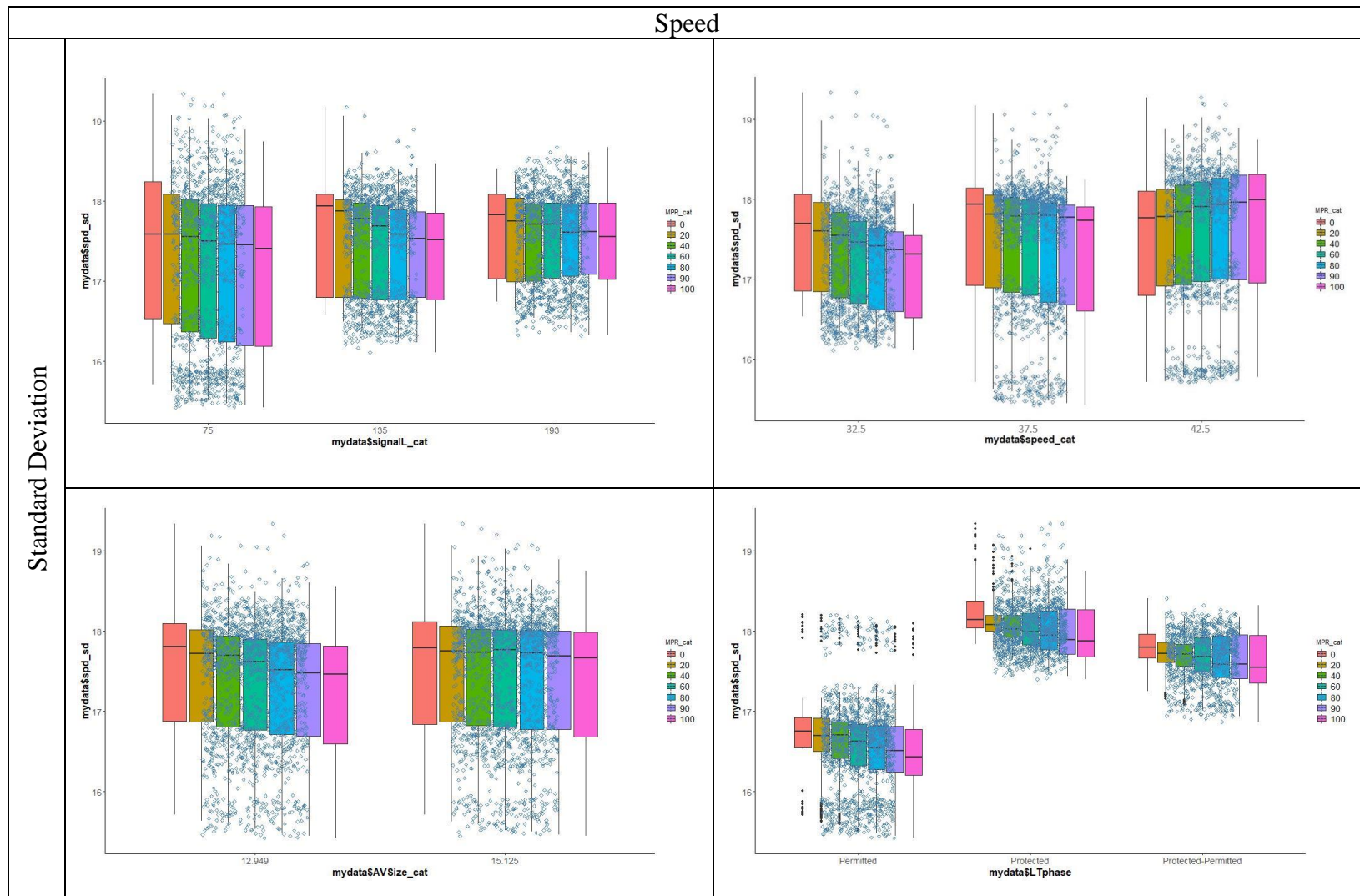
Figure 12. Boxplots of the Volatility Measures: Longitudinal Speed
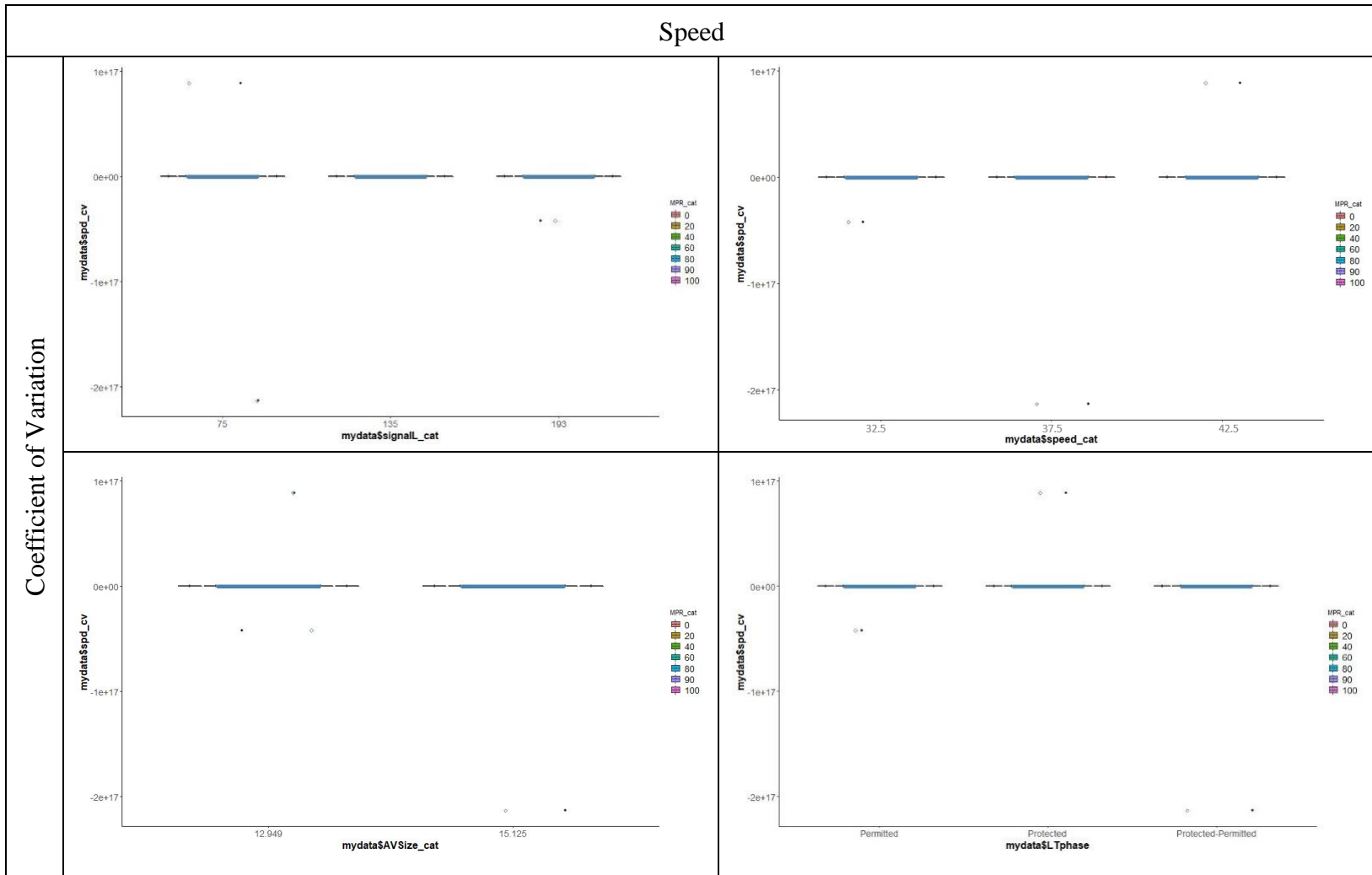
Figure 12 Continued
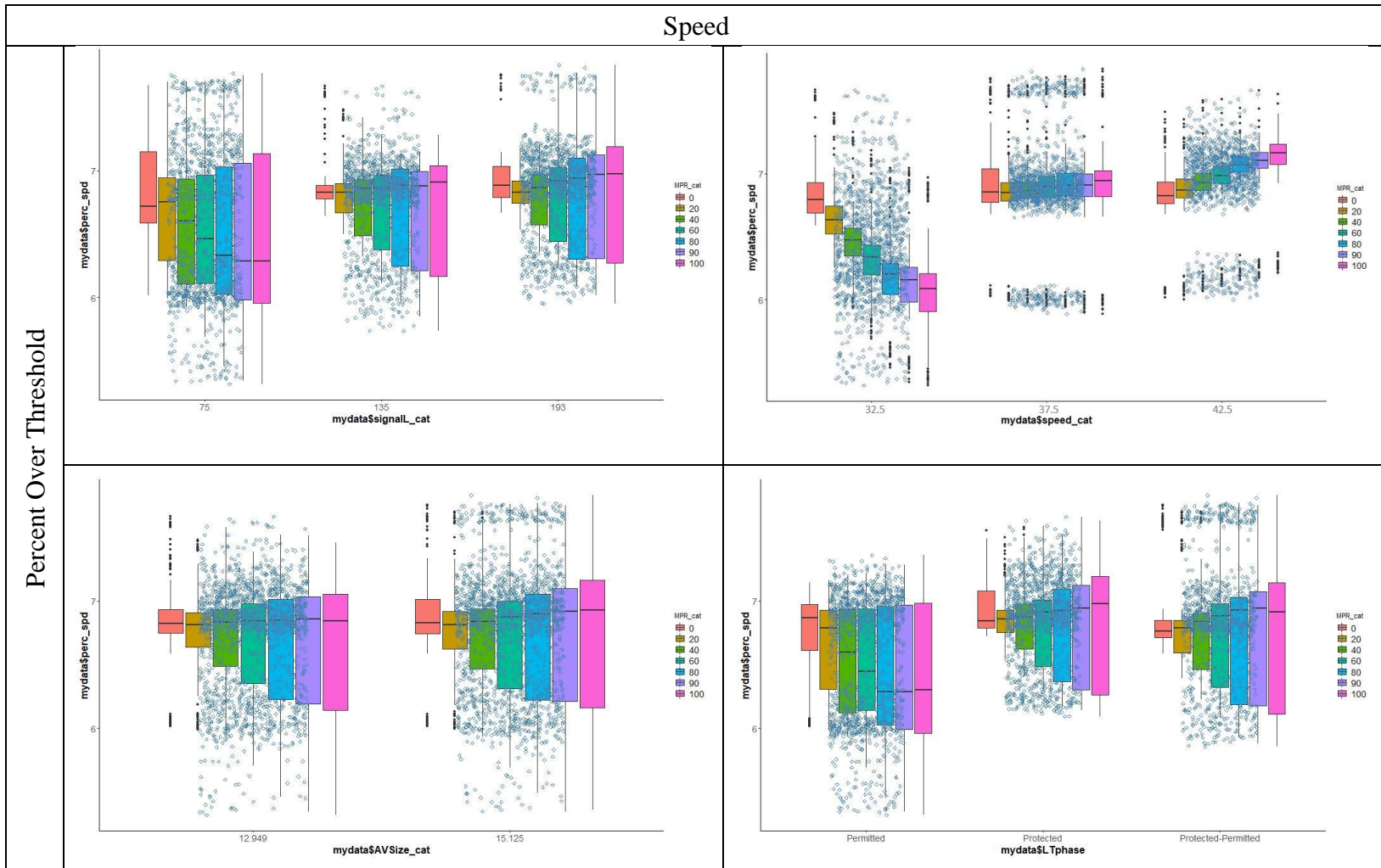
Figure 12 Continued

Moreover, the figures for the speed coefficient of variation do not provide detailed information. Visually comparing the effects of various study variables on the percent of speeding maneuvers indicate that increasing the cycle length increases the mean value of this driving volatility since the majority of the simulation time is allocated to the green signal timing and not stopped traffic. Increasing the speed limit and AV size result in higher speeding maneuvers. Also, protected LT signal phasing increases this driving volatility measure.

The next driving volatility measure focuses on longitudinal acceleration, for which the depictions are provided in ***Error! Reference source not found.***. Increasing the signal length decreases the standard deviation of acceleration and increases the acceleration C.V. Meanwhile, increasing the AV MPR increases the abovementioned volatility measures consistently. By providing higher posted speed limits, the standard deviation and coefficient of variation of the acceleration decreases. Also, any increases in the AV MPR results in an increase in the volatility measure. Enlarging AV size decreases the standard deviation and coefficient of variation of the acceleration globally. No consistent pattern could be detected for percent of the acceleration maneuvers over the threshold.
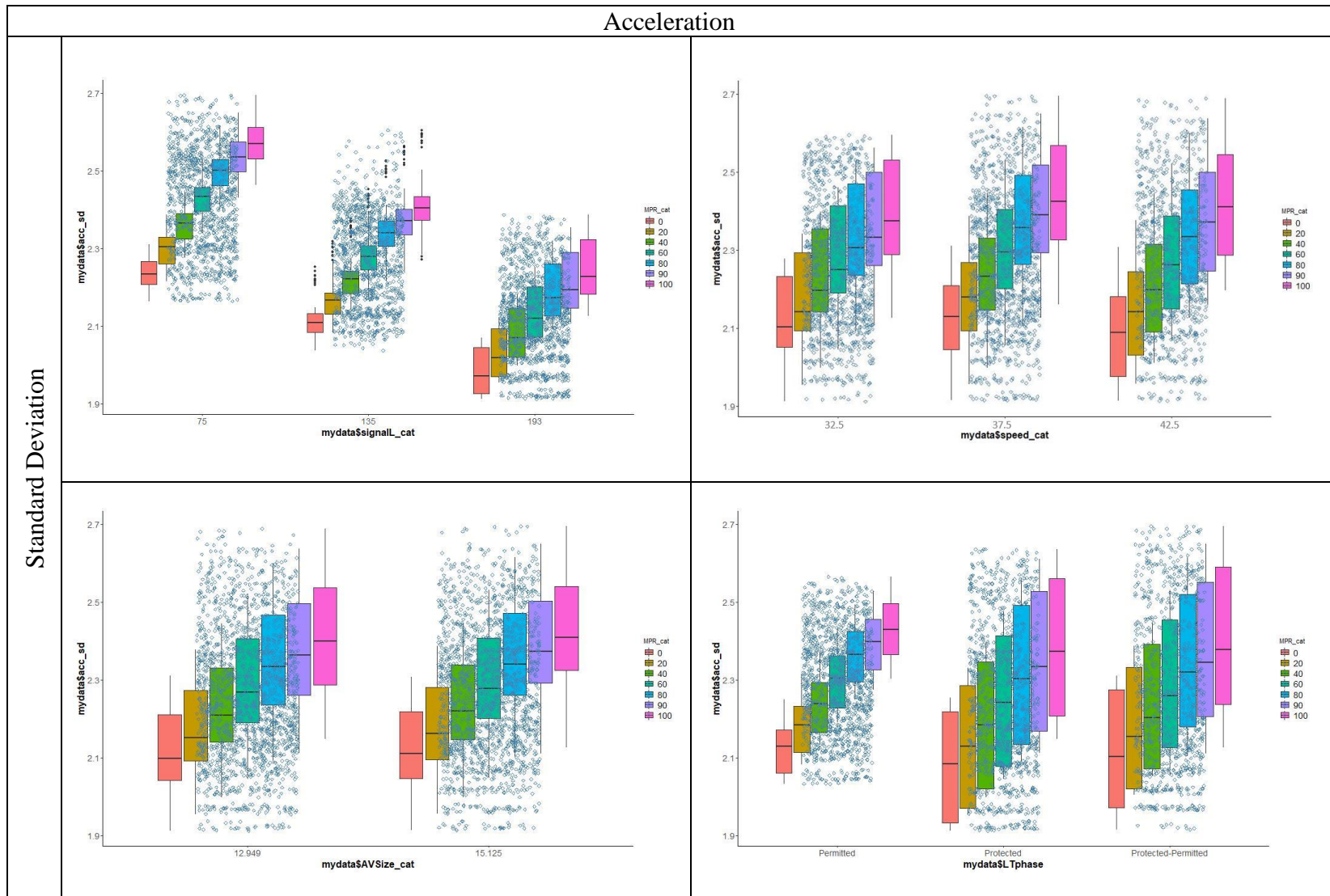
Figure 13. Boxplots of the Volatility Measures: Longitudinal Acceleration
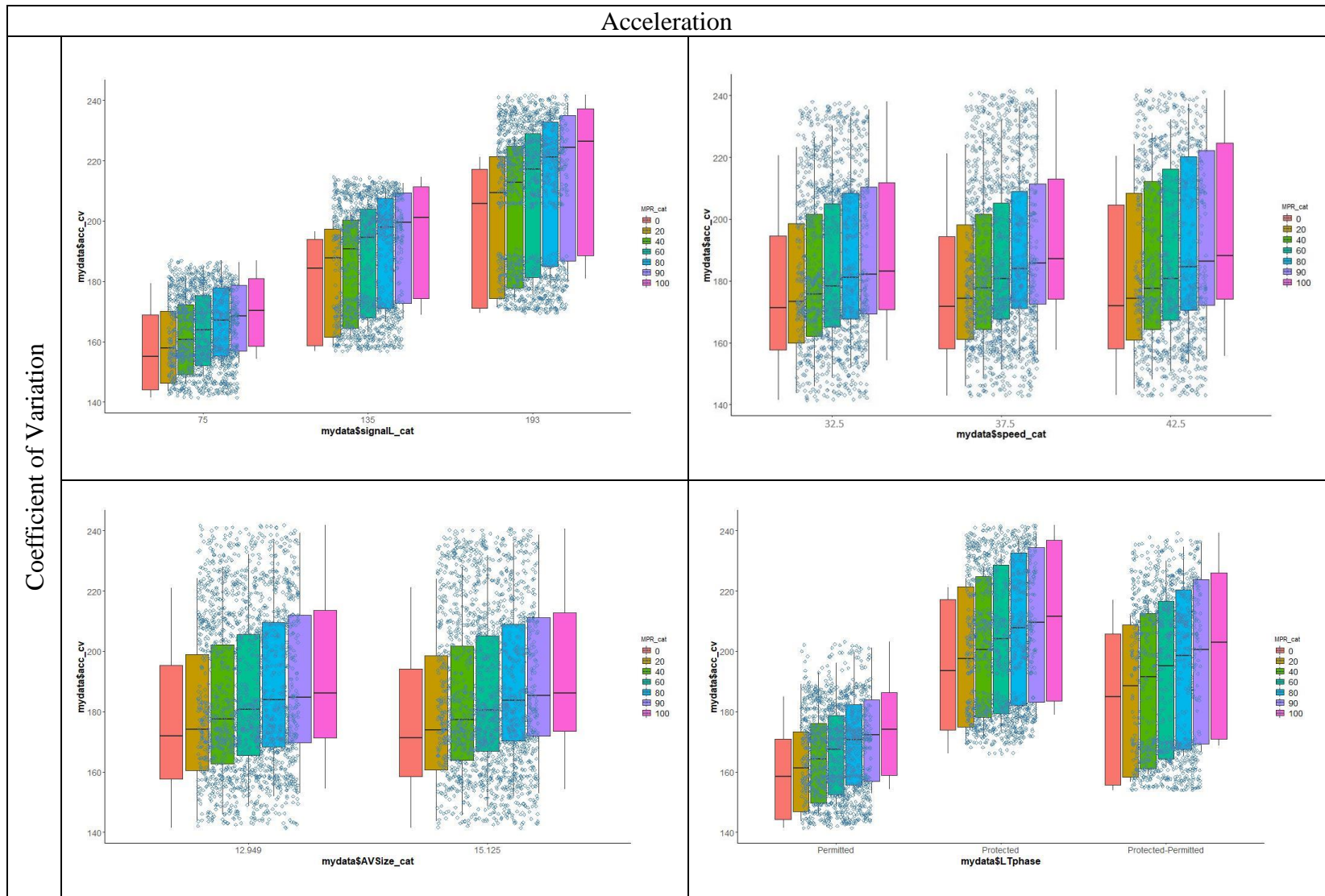
Figure 13 Continued

Figure 13 Continued

67
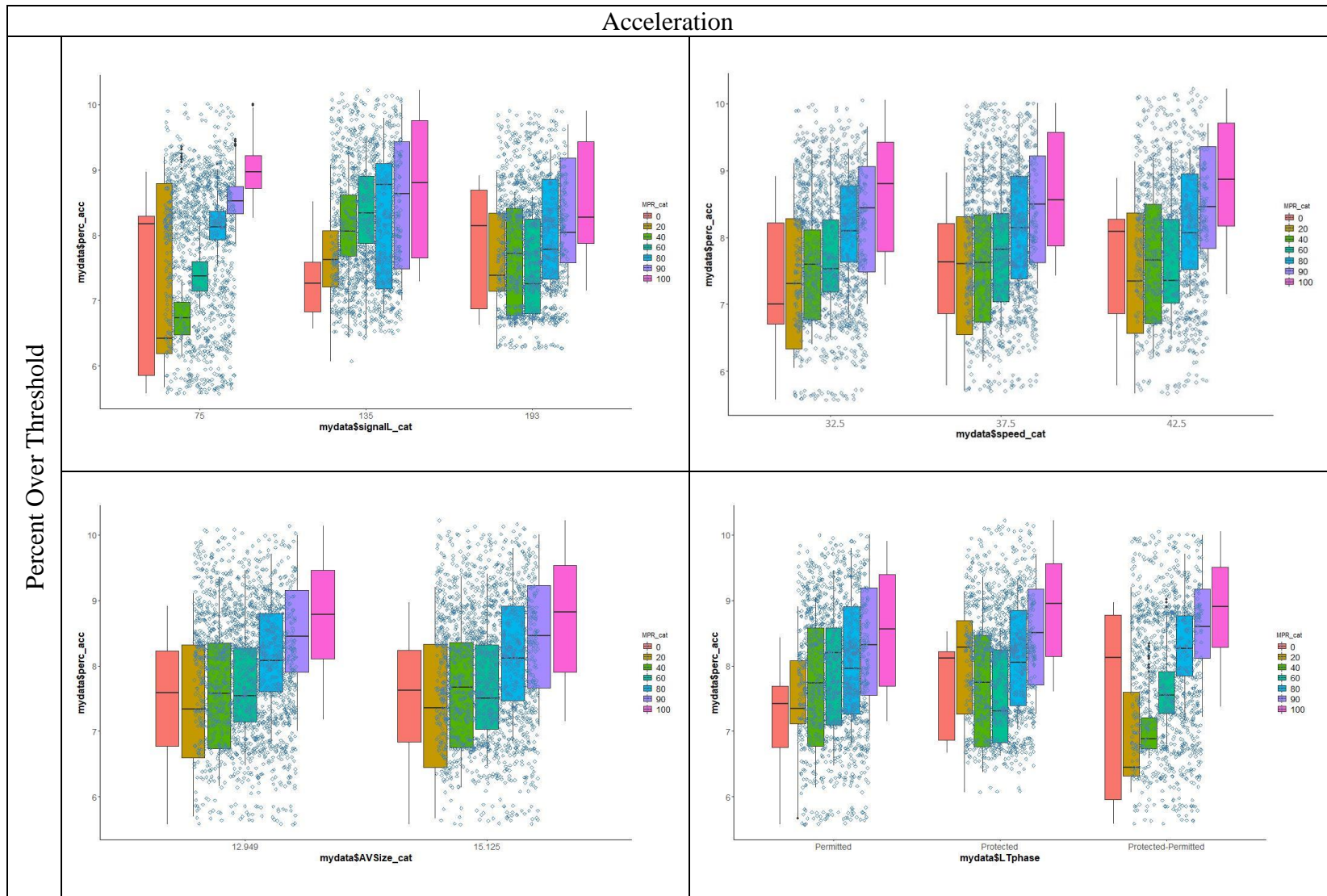
Error! Reference source not found. provides depictions for longitudinal deceleration. As indicated above, by increasing the cycle length, the S.D. of deceleration decreases and gets closer to zero. This is because the traffic experiences fewer red signals, which requires vehicles to decelerate and stop. Additionally, by increasing the AV MPR, the S.D. of the deceleration rate decreases. In other words, higher AV MPRs provide smoother traffic flow, and therefore, enhance traffic safety. Protected LT signal phasing results in the largest S.D. of deceleration due to the LT vehicles that need to come to a stop frequently until obtaining the right of way to perform their maneuver.

Inspecting the positive jerks in ***Error! Reference source not found.*** indicates that by increasing the signal length, the S.D. of positive jerk decreases. Additionally, protected, protected-permitted, and permitted LT signal phasing results in a higher S.D. of positive jerk values, respectively.

Moreover, by increasing the speed limit, the percent of values above the threshold does not change significantly, but there is a high fluctuation in the provided highest speed limit. The fluctuation could be observed in various LT signal phasing as well.

Last but not least, *Figure* explores longitudinal negative jerk. As indicated in *Figure*, for various cycle lengths, the standard deviation of negative jerk goes up at low AV MPRs. However, by increasing the AV MPR, the standard deviation starts decreasing. The same pattern could be observed on AV size and signal LT phasing. Increasing the AV MPR increases the percent of negative jerk values above the threshold at all the levels of cycle length, speed limit, and AV size.

Notably, the majority of the negative jerk graphs are normally distributed, and a noticeable skewness could not be observed.
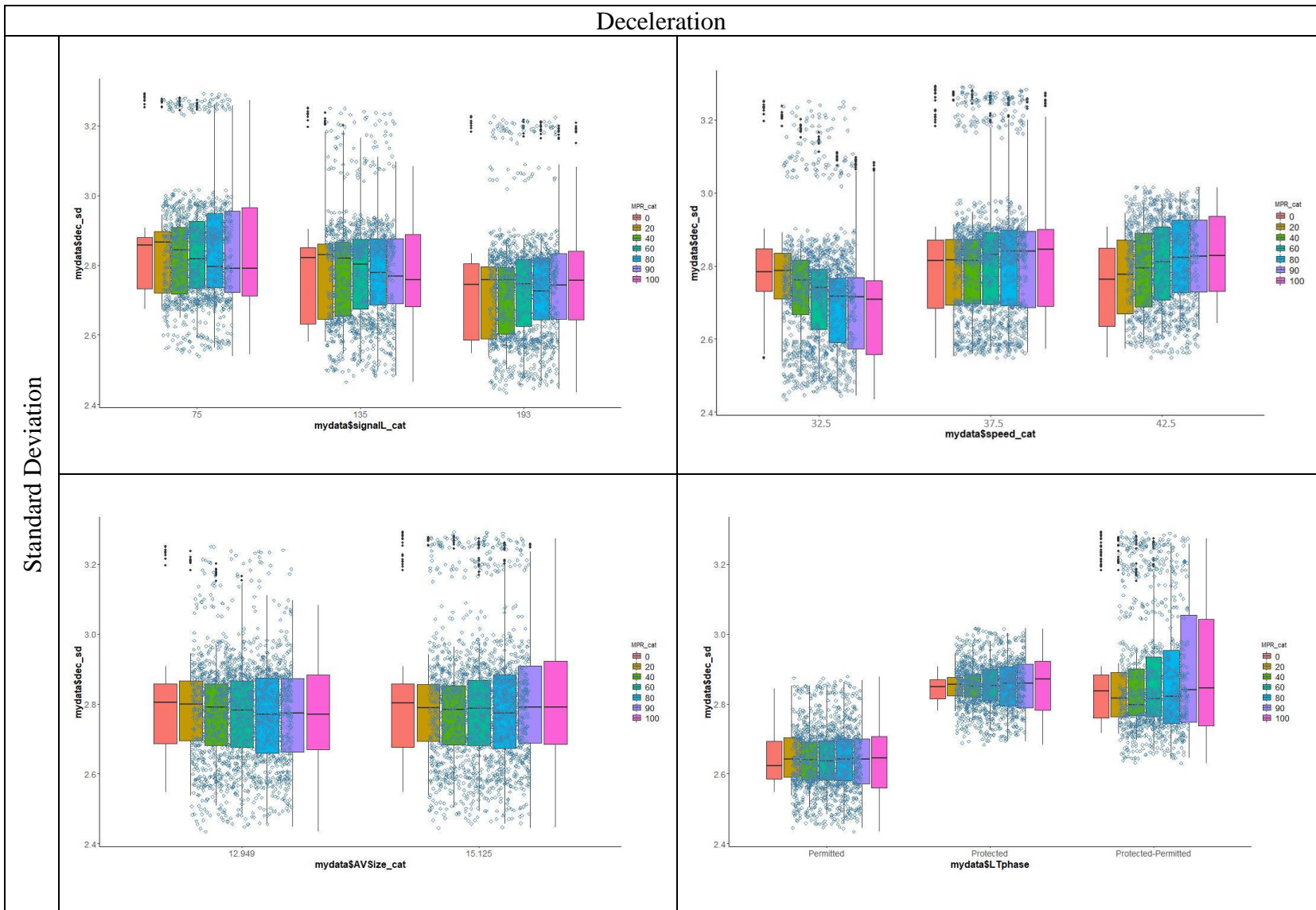
Figure 14. Boxplots of the Volatility Measures: Longitudinal Deceleration

Deceleration

Coefficient of Variation
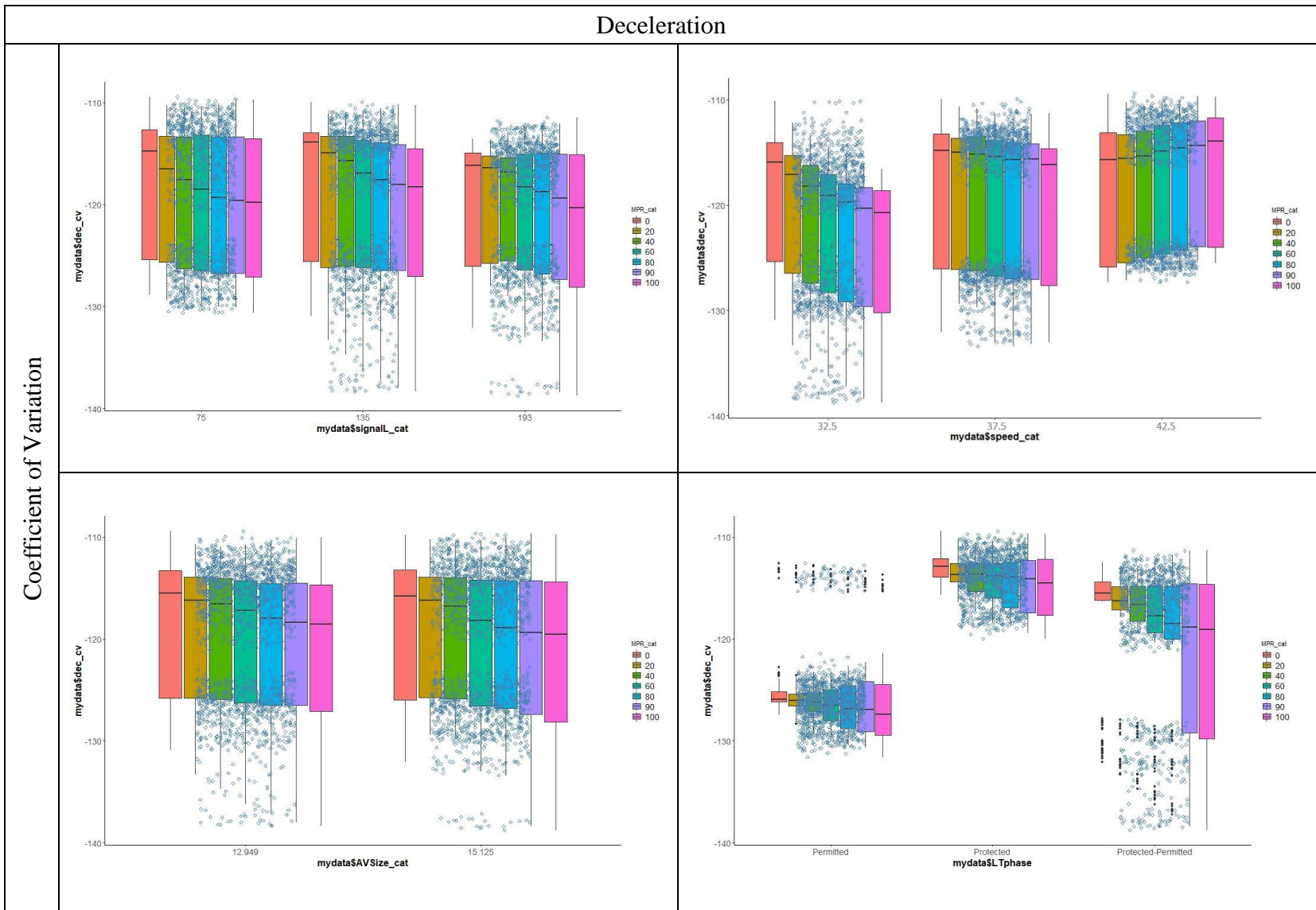
Figure 14 Continued
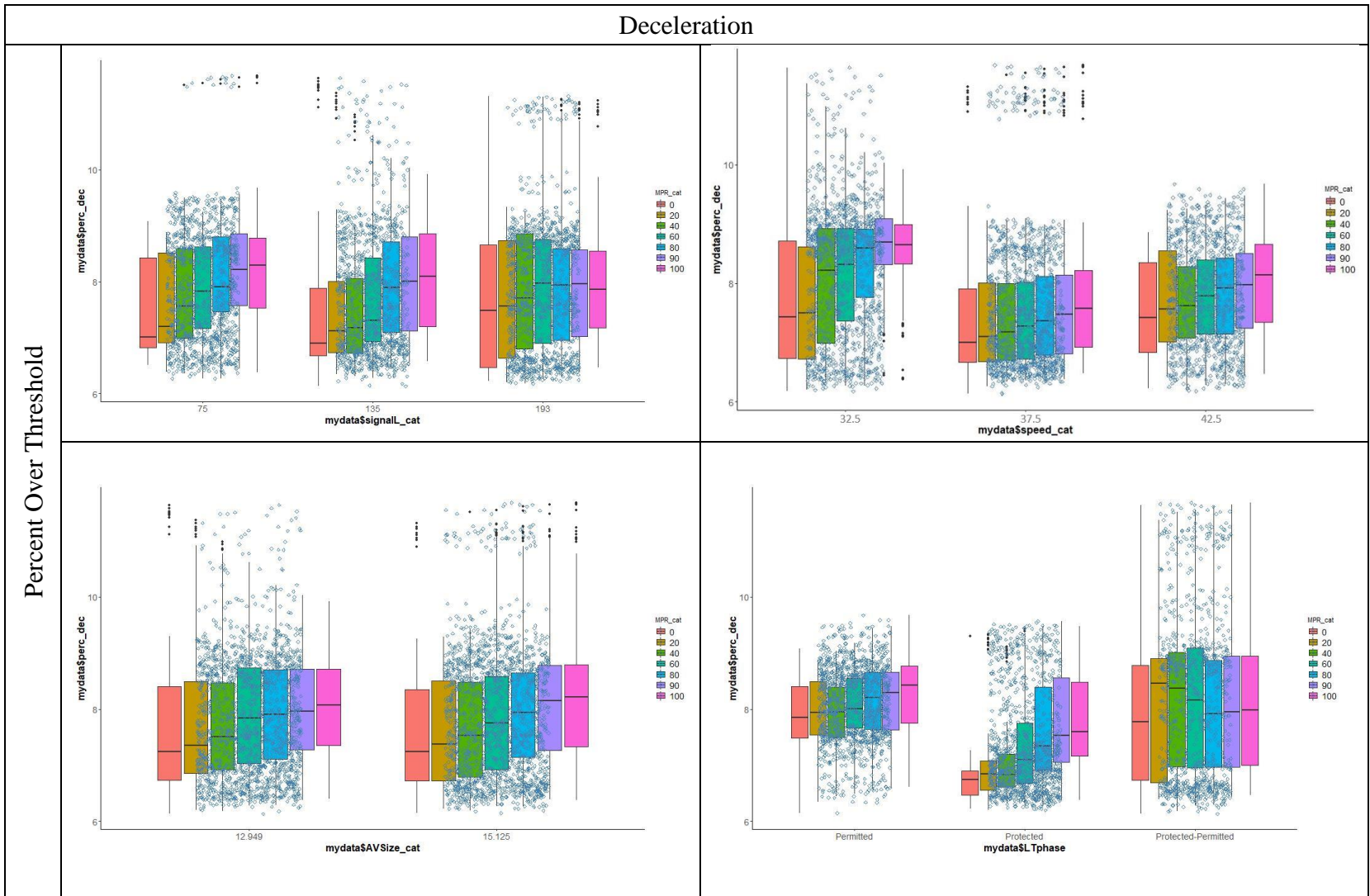
Figure 14 Continued
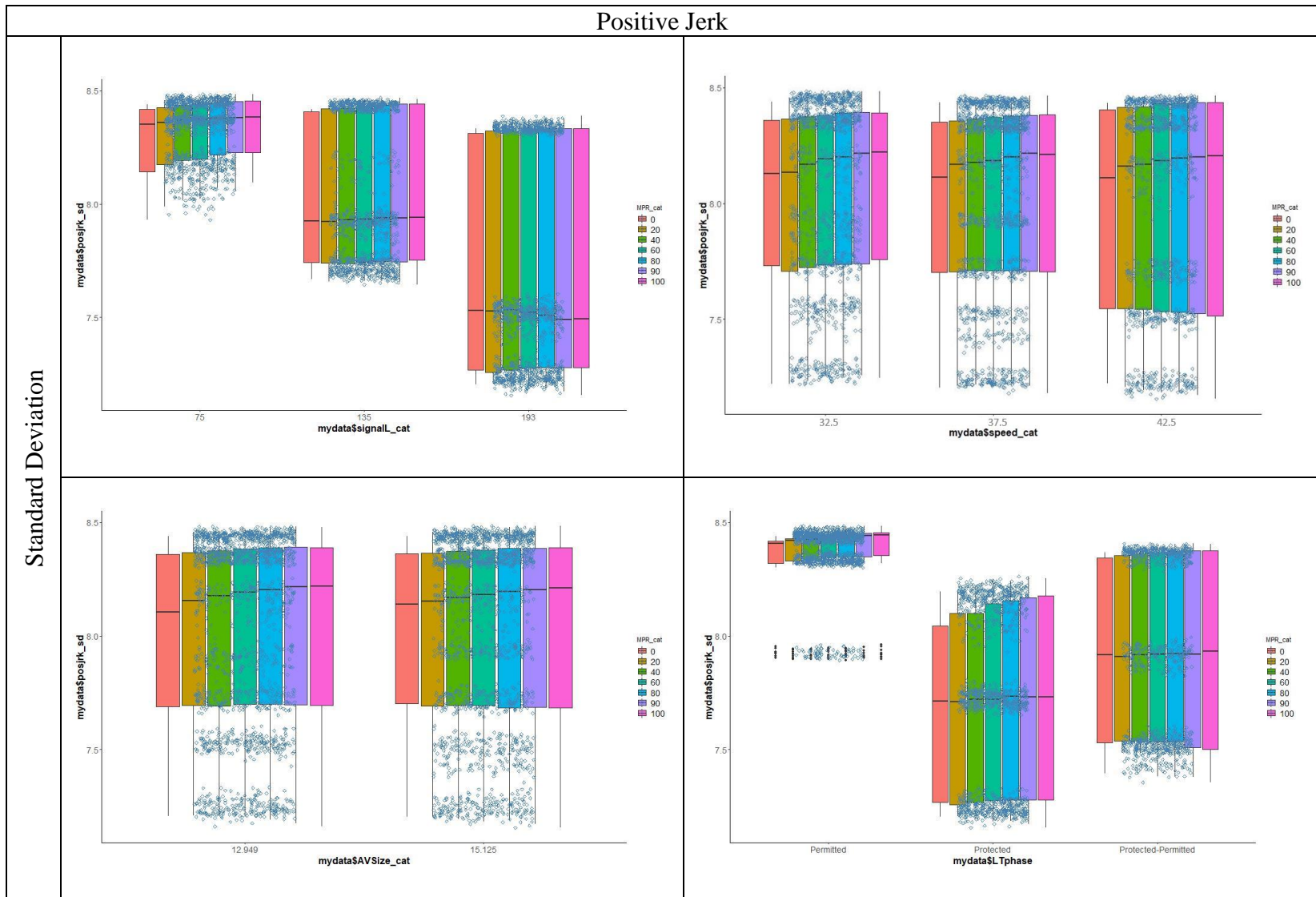
Figure 15. Boxplots of the Volatility Measures: Positive Jerk
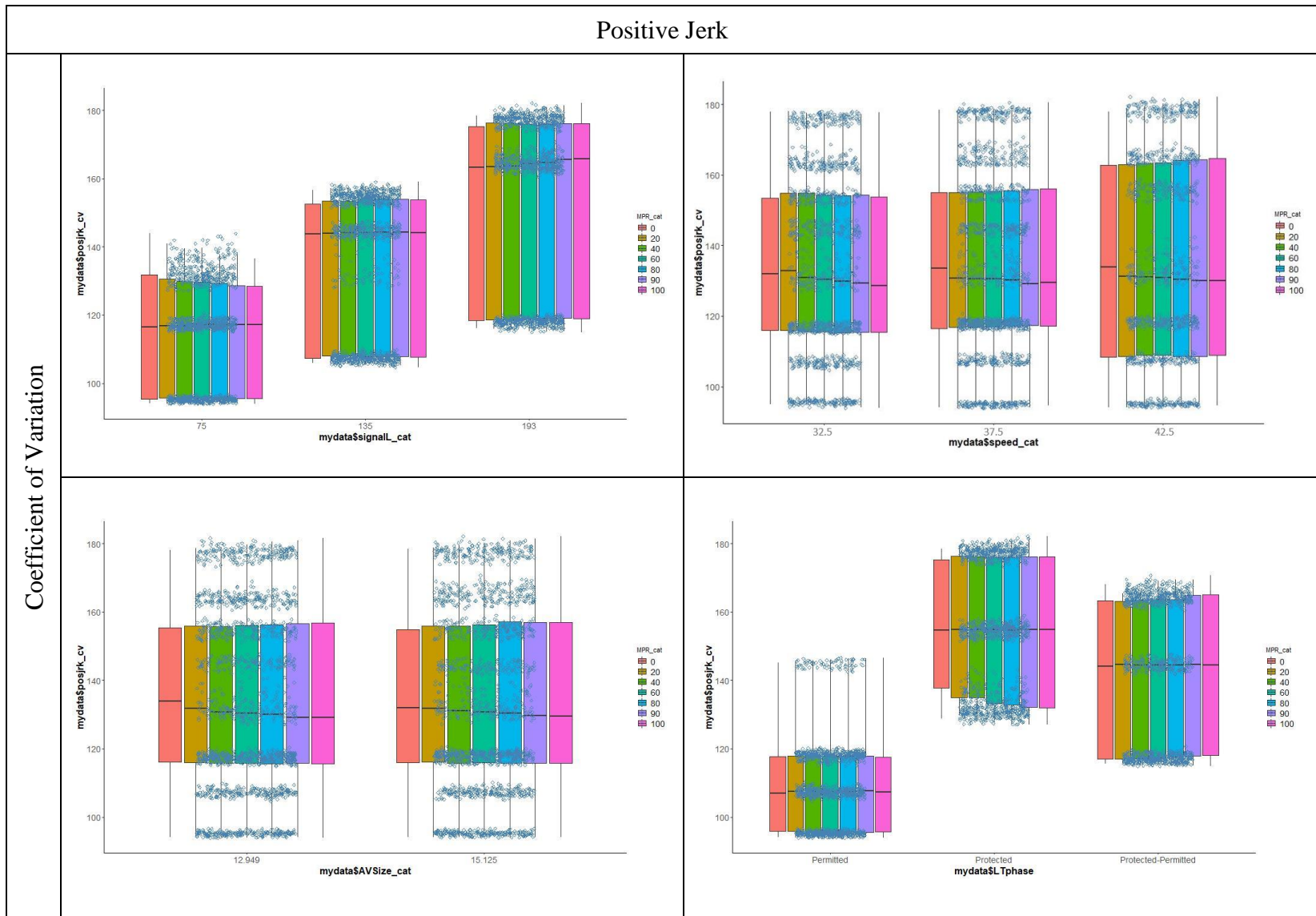
Figure 15 Continued
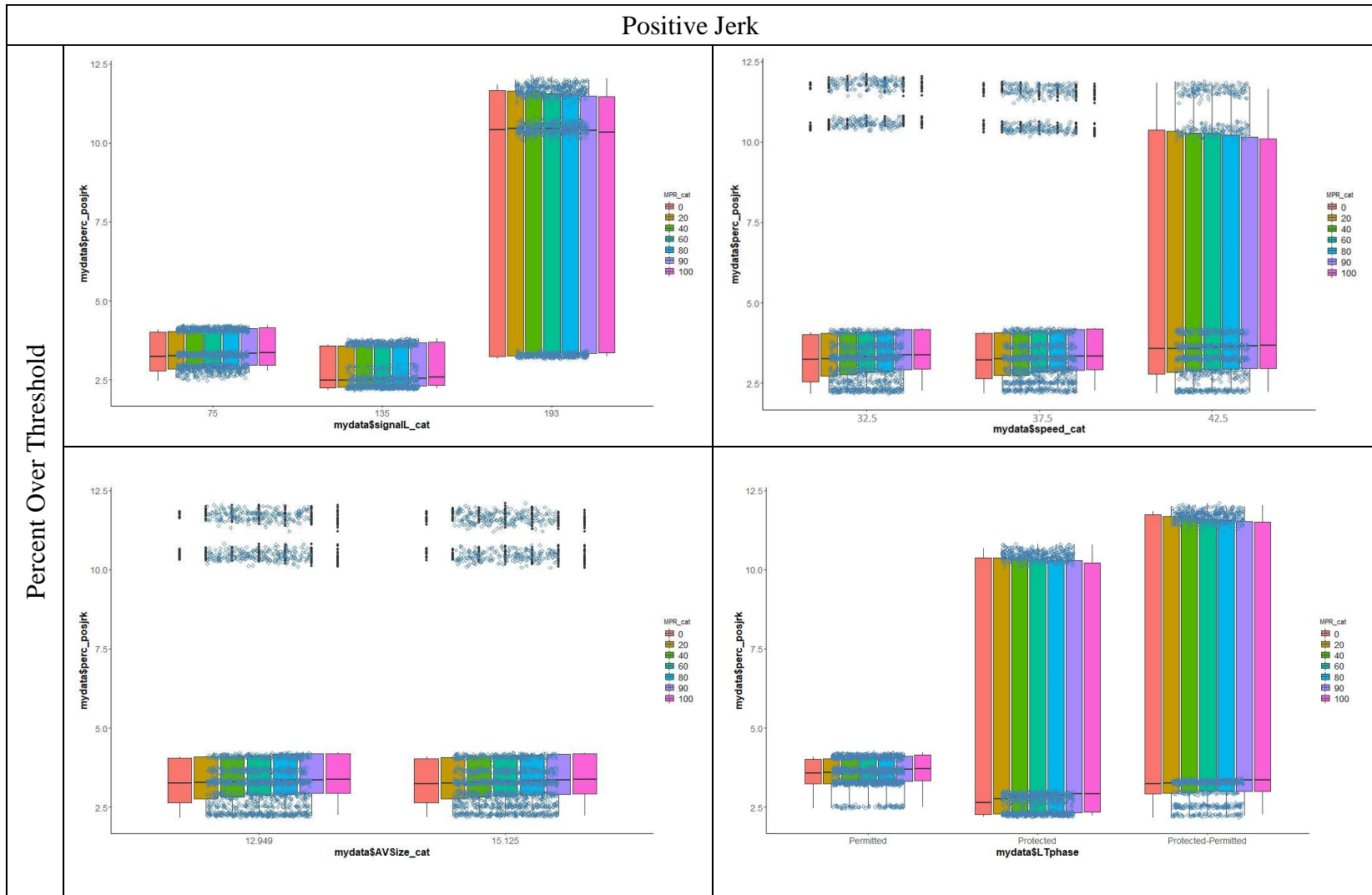
Figure 15 Continued
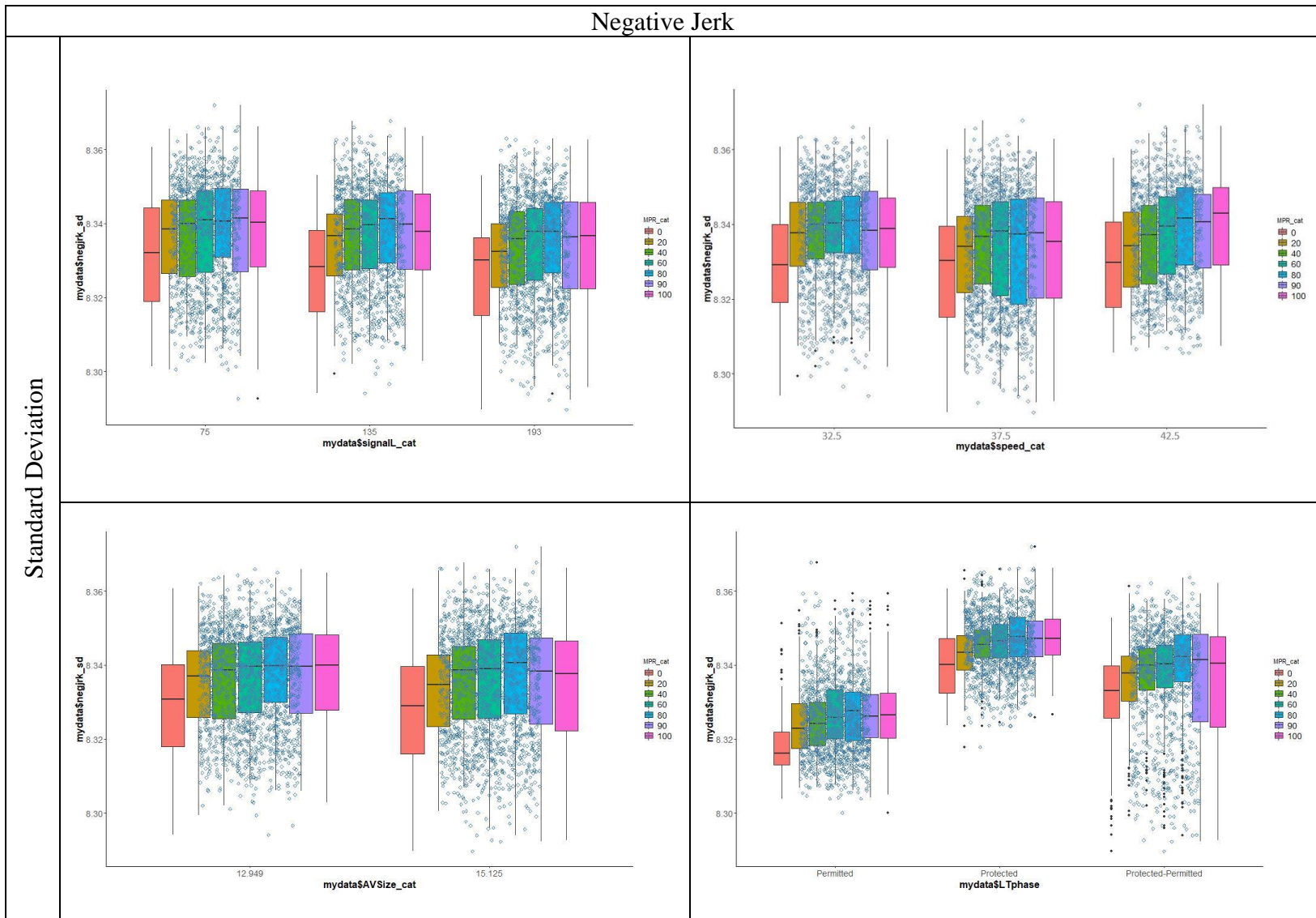
Figure 16. Boxplots of the Volatility Measures: Negative Jerk

Figure 16 Continued

Negative Jerk
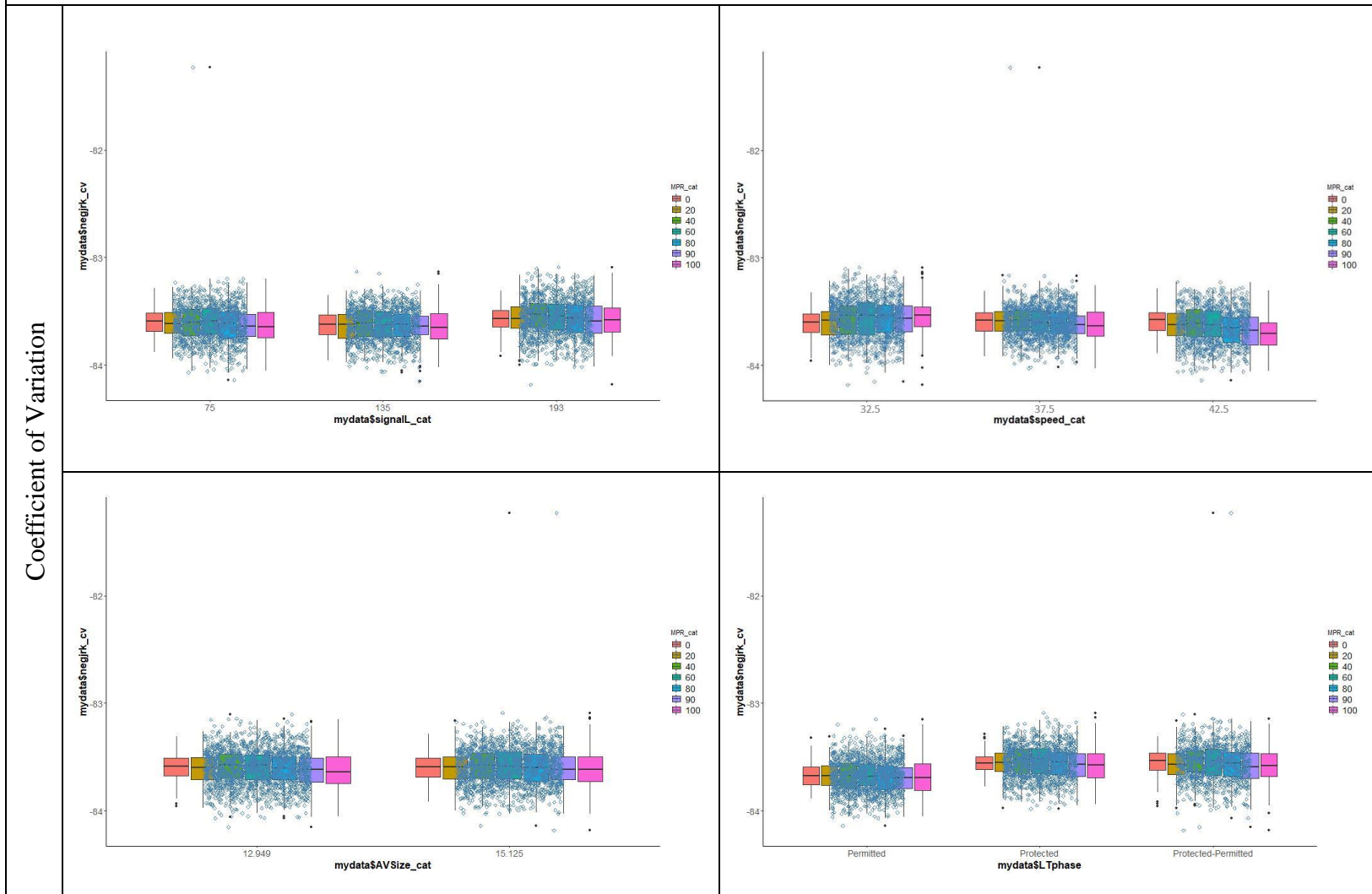
Percent Over Threshold
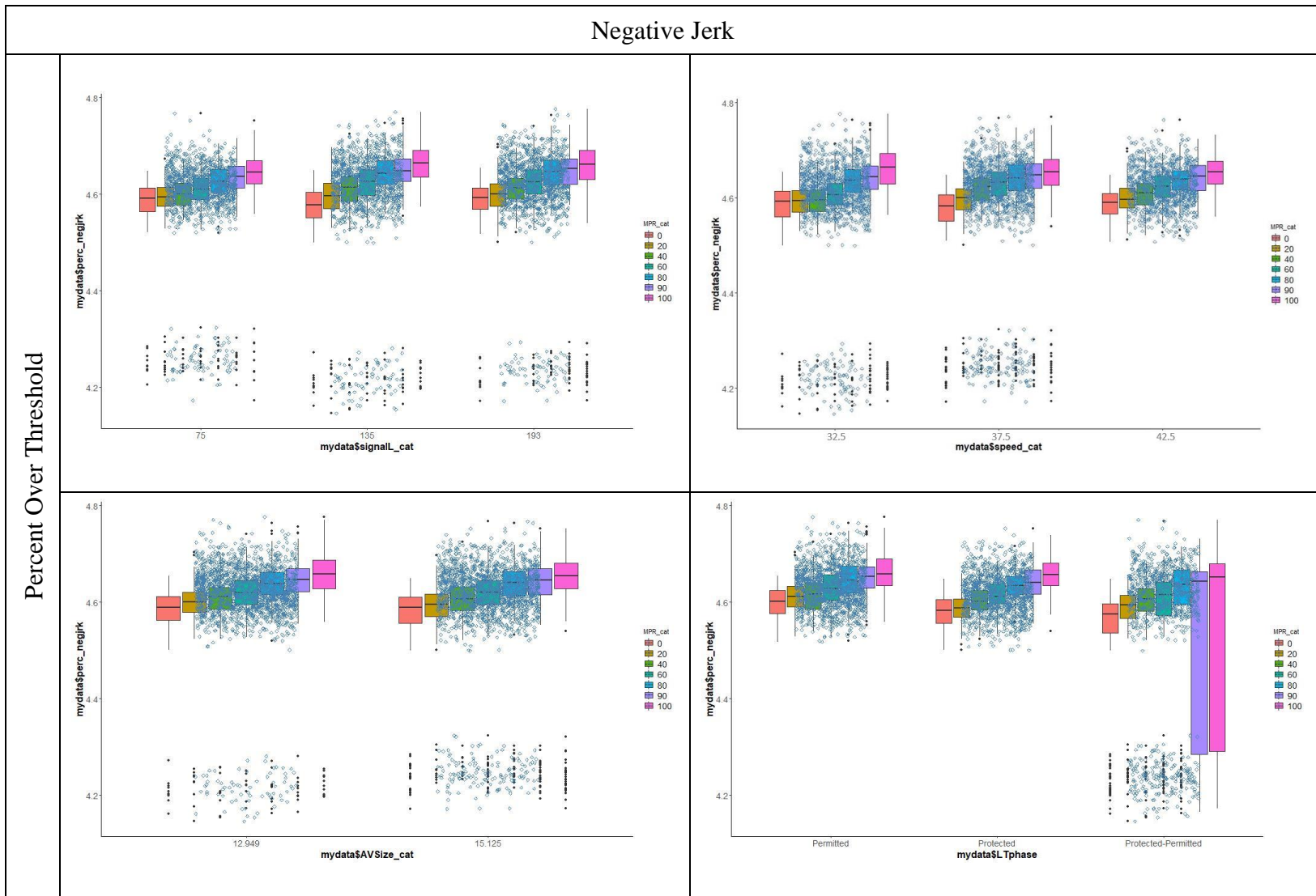
Figure 16 Continued

Lateral Driving Volatility

This section intends to assess the lateral driving volatility measures, which have not been studied extensively and thoroughly in the previous research studies.

Table 6 presents the descriptive statistics for each lateral driving volatility measure, including speed, acceleration, deceleration, positive jerk, and negative jerk. As indicated, the lateral driving volatility measures experience fewer fluctuations among the simulation runs of the scenarios, compared to the longitudinal driving volatilities that were more dispersed.

Table 6. Descriptive Statistics for Lateral Driving Volatility Measures

Measure	Mean	S.D.	C.V.	Percent Over Threshold
Lateral Speed ($ft/_s$)				
Min	-0.0002	0.23	-2.13E+17	0.06
Max	0.0002	0.39	8.88E+16	0.16
Median	0.00006	0.33	500,248.6728	0.11
Mean	0.00006	0.33	-4.59065E+13	0.11
Var	0.00000	0.0008	1.52E+31	0.0003
S.D.	0.00004	0.03	3.90011E+15	0.02
C.V.	0.73	0.09	-84.95	0.16
Lateral Acceleration ($ft/_{s^2}$)				
Min	0.06	2.37	2,198.96	0.06
Max	0.18	4.04	3,806.35	0.16
Median	0.13	3.36	2,643.33	0.11
Mean	0.13	3.37	2,658.79	0.1
Var	0.0005	0.08	66,087.45	0.000
S.D.	0.02	0.29	257.08	0.02
C.V.	0.17	0.09	0.10	0.17
Lateral Deceleration ($ft/_{s^2}$)				
Min	-7.62	24.36	-338.82	6.40
Max	-7.21	25.07	-325.78	6.80
Median	-7.39	24.56	-332.02	6.51
Mean	-7.40	24.57	-332.08	6.51
Var	0.004	0.006	7.60	0.002

Table 6 Continued

Measure	Mean	S.D.	C.V.	Percent Over Threshold
S.D.	0.07	0.08	2.76	0.04
C.V.	-0.009	0.003	-0.008	0.006
Lateral Positive Jerk ($ft/_{s3}$)				
Min	-0.04	9.72	-48,630,858.7	0.15
Max	0.003	12.65	33,788,442.40	0.83
Median	-0.02	11.39	-62,389.78	0.40
Mean	-0.02	11.27	-89,355.81	0.42
Var	.000	0.78	1.40E+12	0.02
S.D.	0.006	0.88	1,183,082.015	0.14
C.V.	-0.35	0.08	-13.24	0.32
Lateral Negative Jerk ($ft/_{s3}$)				
Min	-135.06	405.32	-320.36	8.16
Max	-129.53	417.53	-302.32	8.83
Median	-132.47	410.19	-309.63	8.49
Mean	-132.48	410.41	-309.79	8.49
Var	0.54	4.23	7.12	0.009
S.D.	0.73	2.06	2.67	0.09
C.V.	-0.006	0.005	-0.009	0.01

As in the longitudinal volatility measures, to visually inspect more details about the lateral driving volatility measures, ***Error! Reference source not found.*** to *Figure* are provided by having the x-axis representing the study variables, the y-axis showing the driving volatility measures, and the color of each box depicting AV MPRs.

As indicated in ***Error! Reference source not found.***, by increasing the cycle length, the S.D. of the lateral speed decreases because vehicles move through the intersection more uniformly with less mandatory stops. Hence, there is a smaller chance of vehicles being able to make discretionary lateral movements, e.g., lane changing maneuvers. Regardless of the cycle length, speed limit, AV size, and LT phasing, higher AV MPRs results in a lower S.D. of the lateral speed. The percent of lateral speed over the

threshold also indicates that higher cycle lengths and lower speed limits result in a smaller percent. For all the variables, regardless of the level, higher MPRs decrease the percent of lateral speed over the threshold almost consistently.

According to **Error! Reference source not found.**, the standard deviation of lateral acceleration decreases by increasing the cycle length and reducing the speed limit. Similarly, by increasing the AV MPR, S.D. decreases, regardless of the levels of the variables. In addition, the data dispersion increases by reducing the cycle length and increasing the speed limit.

Error! Reference source not found. depicts the lateral deceleration. As indicated, the deceleration S.D., as well as the percent of deceleration over the threshold, are not sensitive to any levels of the study variables. However, increasing the AV MPR increases lateral deceleration.

The graphs for the C.V. of all the study variables, except for the AV size, show a consistent pattern while increasing the AV MPR. In fact, by raising the AV MPR, the variations in the C.V. increases. Since the S.D. graphs do not present any skewness, this means that all the fluctuations are dependent upon the mean values.

Evaluating lateral positive jerk, presented in **Error! Reference source not found.**, indicates that increasing the cycle length decreases the S.D. of the lateral positive jerk noticeably. Moreover, the change in AV MPR does not impact the S.D. of lateral jerk noticeably. No significant variation could be detected in the C.V. Also, the percent of values over the threshold presents no uniform pattern for the variables.

Lastly, *Figure* shows that at all the levels of cycle length, speed limit, and LT phasing, any increase in the AV MPR increases the S.D. of the lateral negative jerk as well as the percent of values over the threshold.
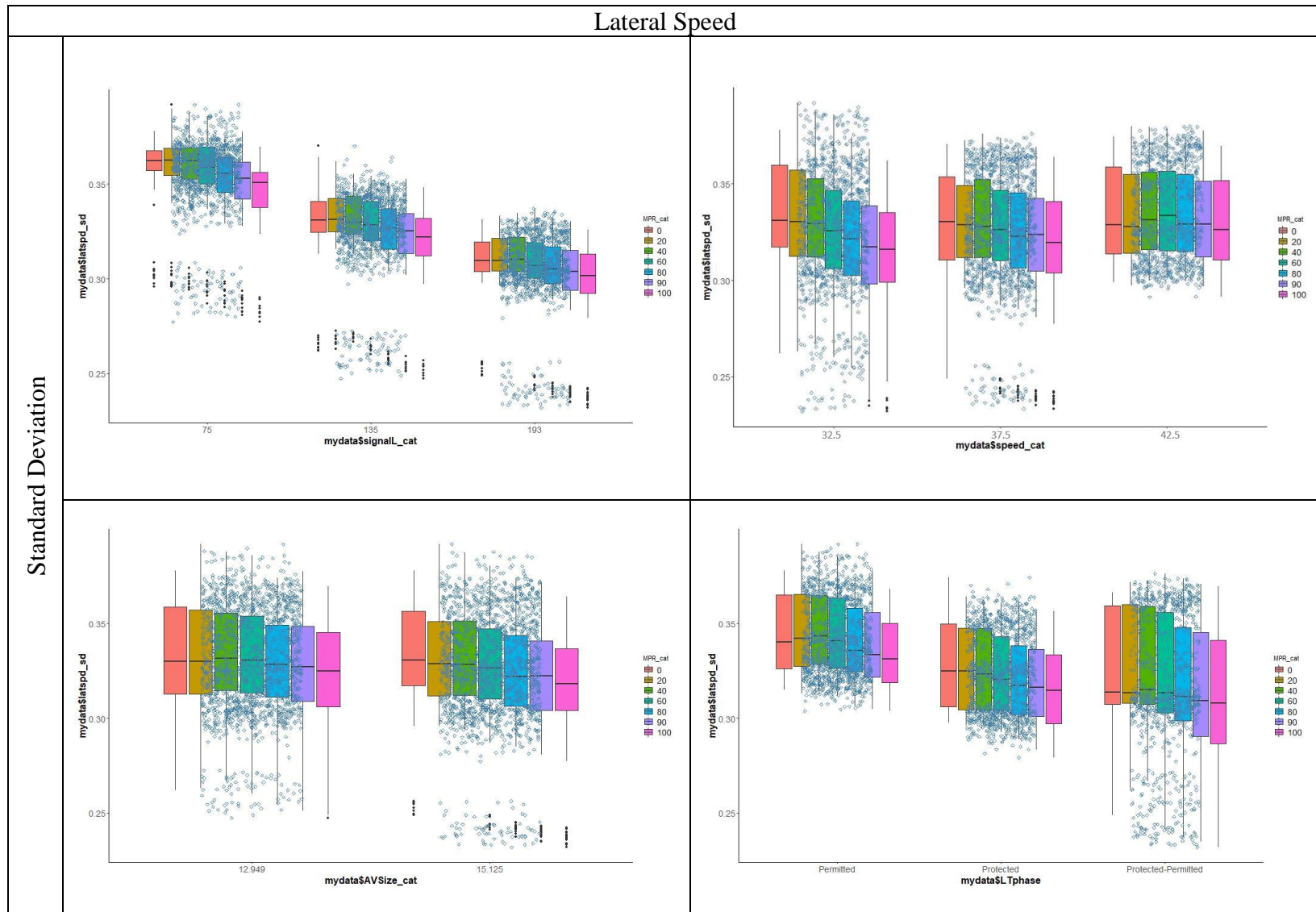
Figure 17. Boxplots of the Volatility Measures: Lateral Speed

Figure 17 Continued
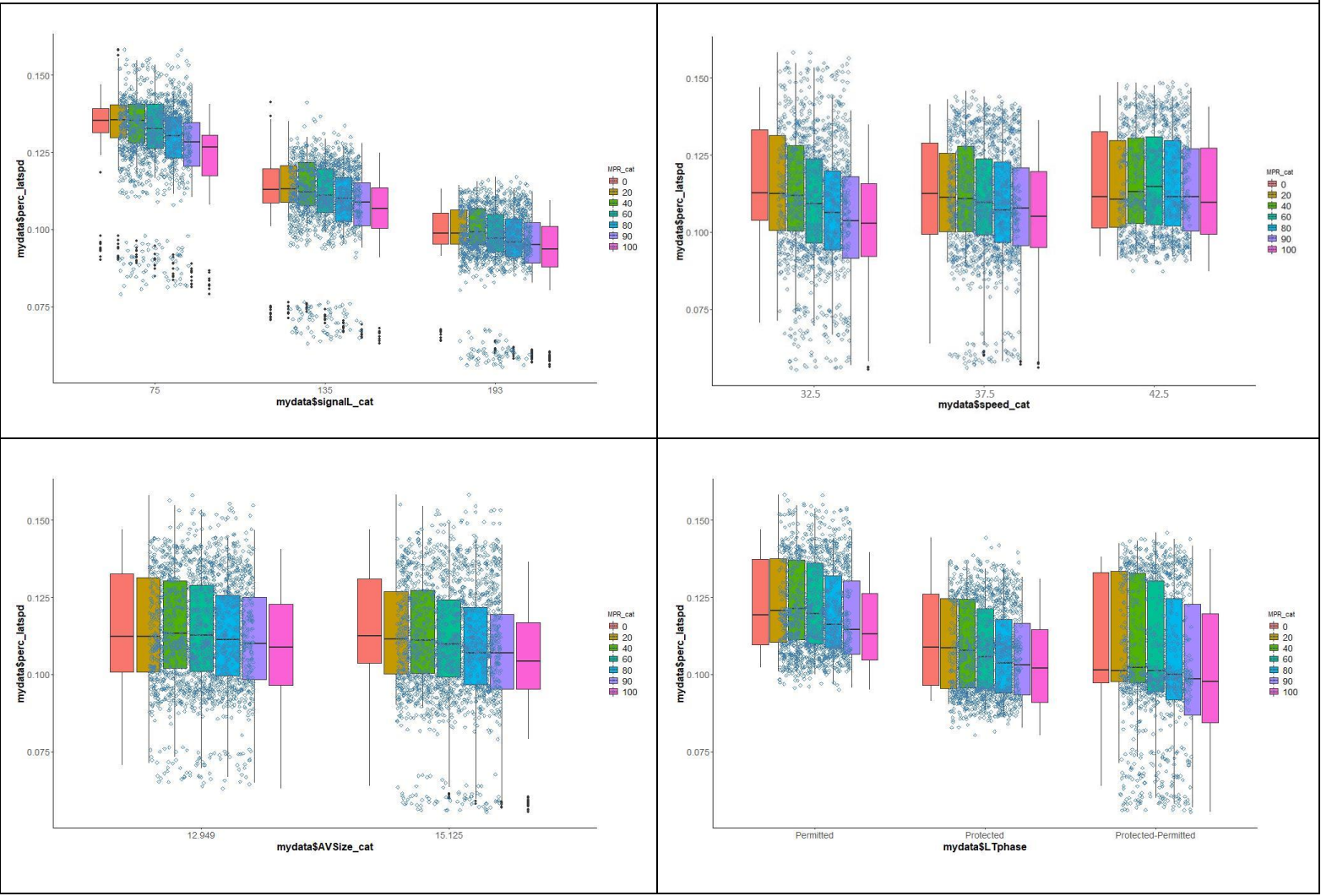
Figure 17 Continued
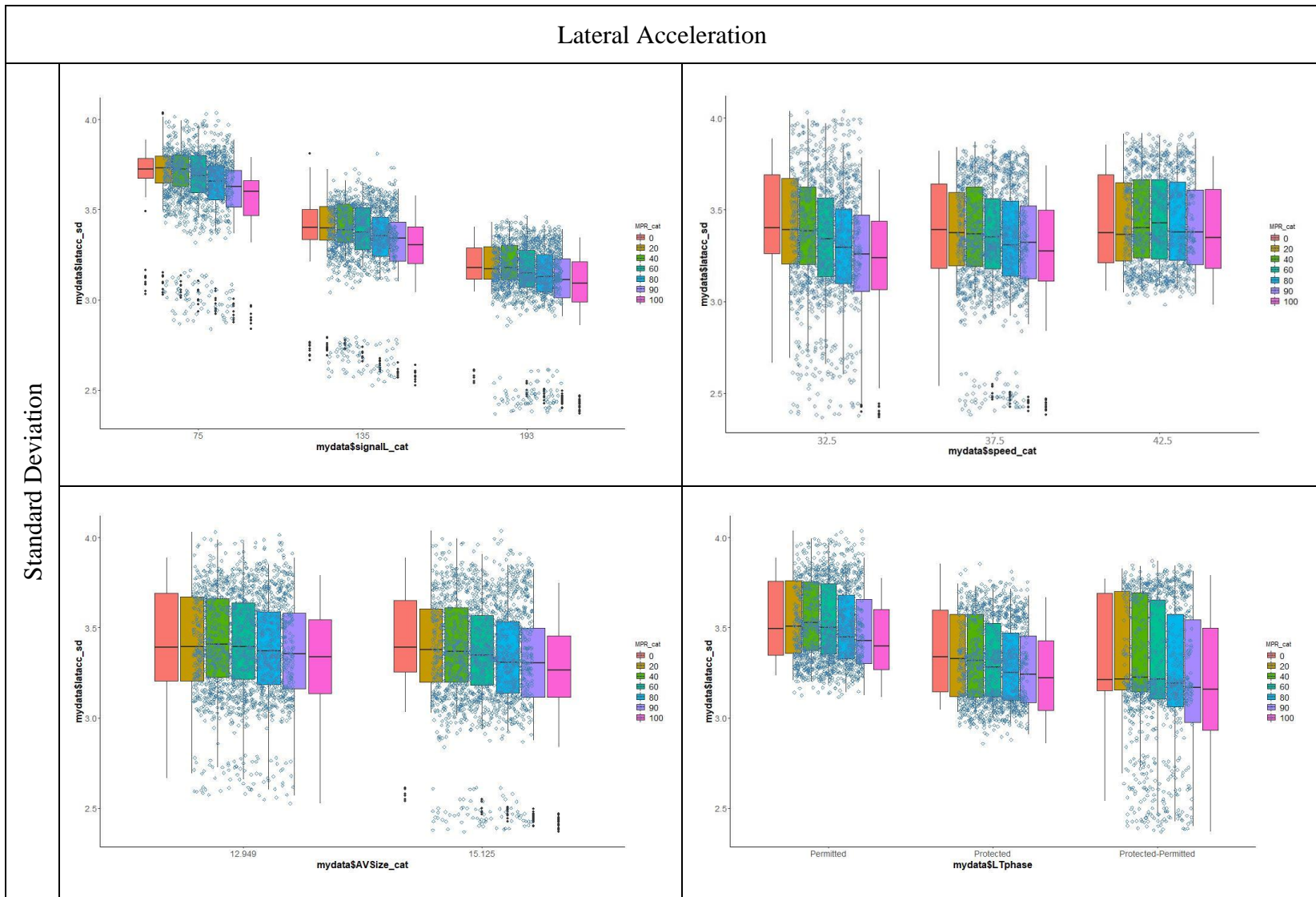
Figure 18. Boxplots of the Volatility Measures: Lateral Acceleration
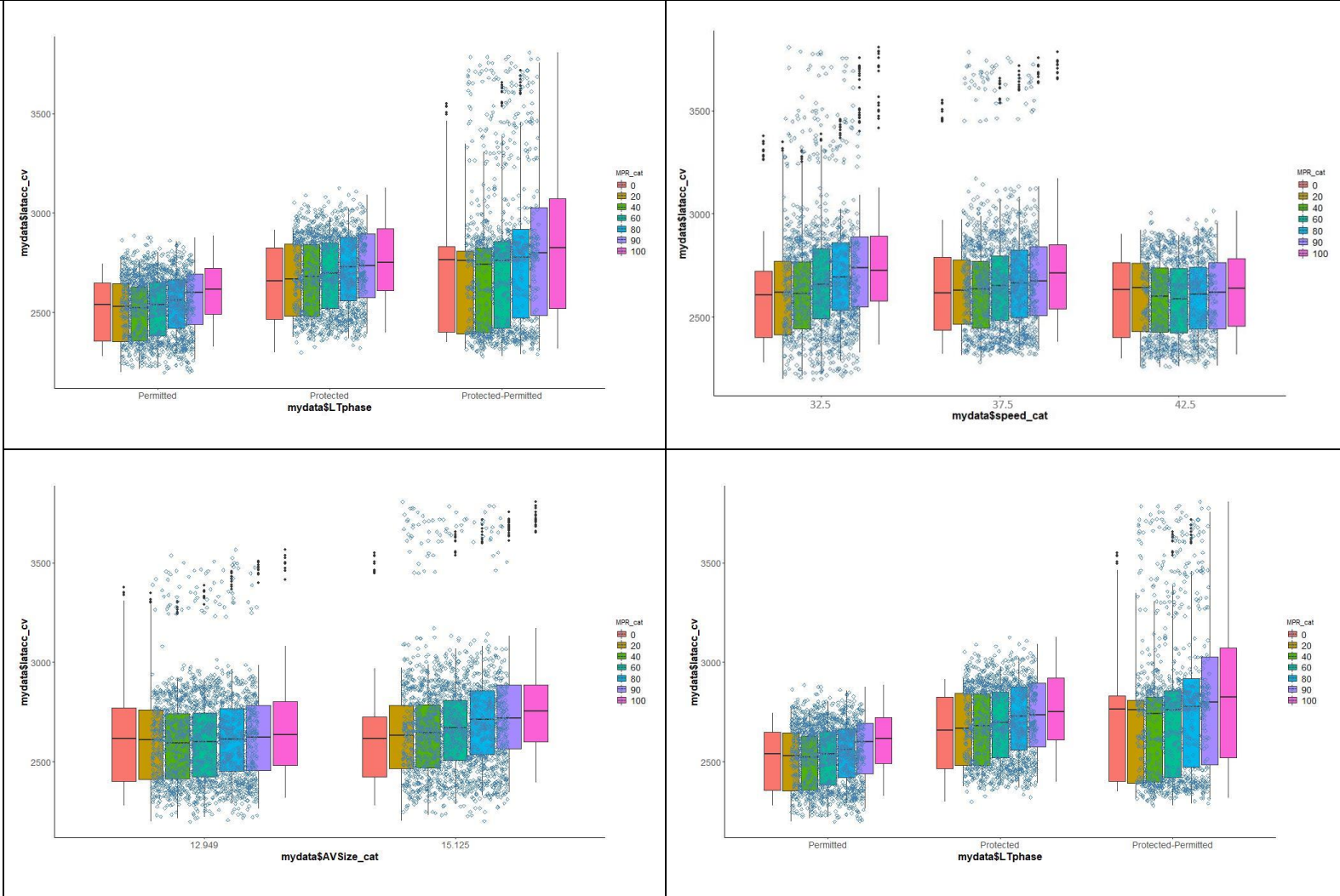
Figure 18 Continued
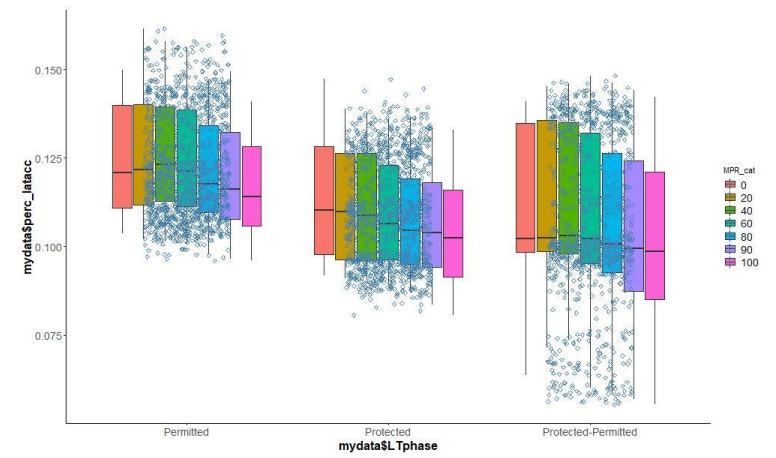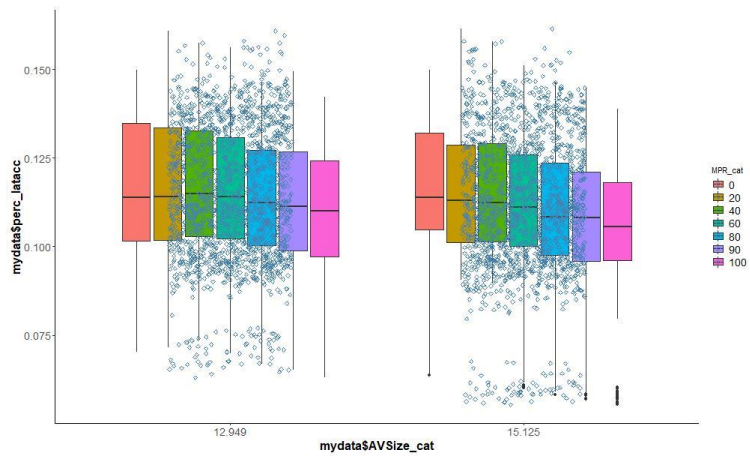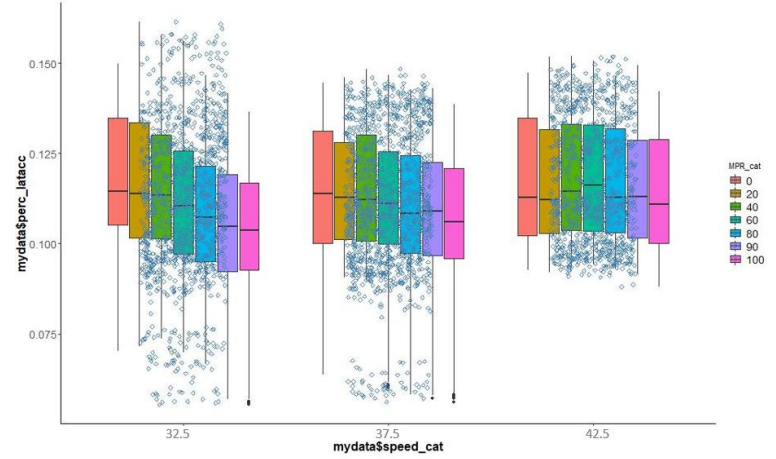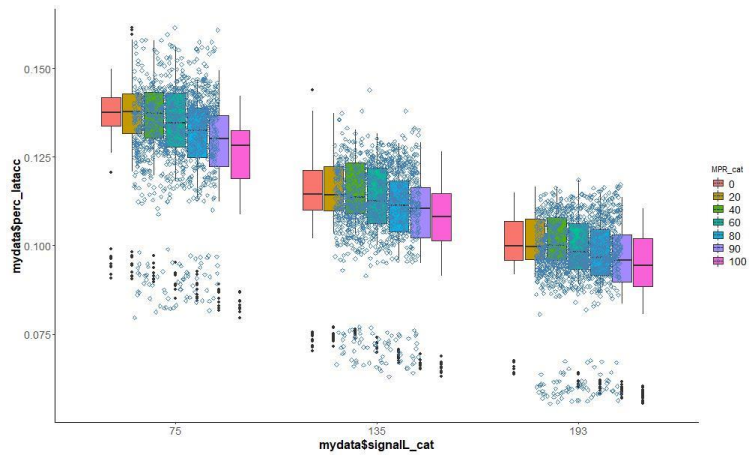
Figure 18 Continued
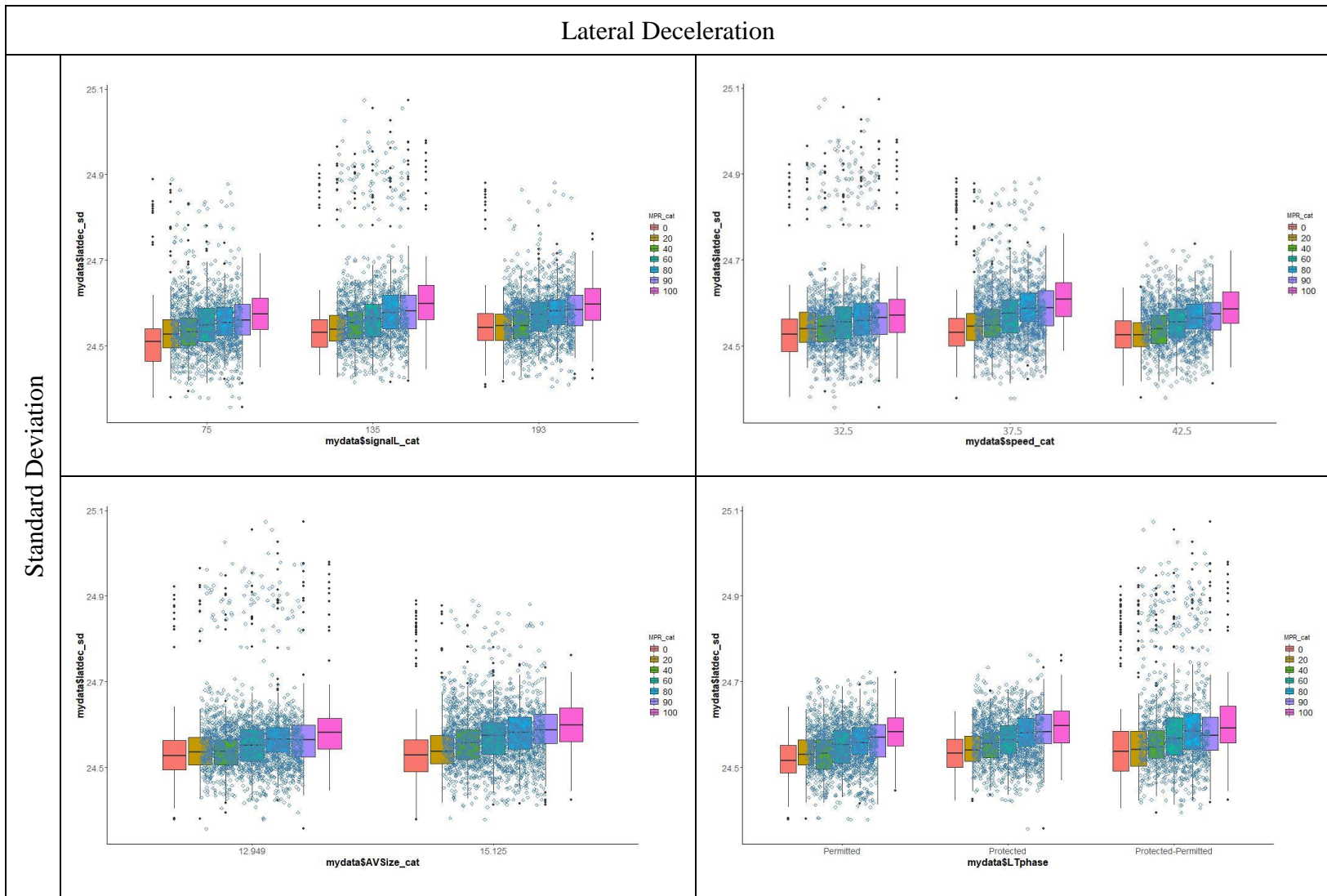
Figure 19. Boxplots of the Volatility Measures: Lateral Deceleration
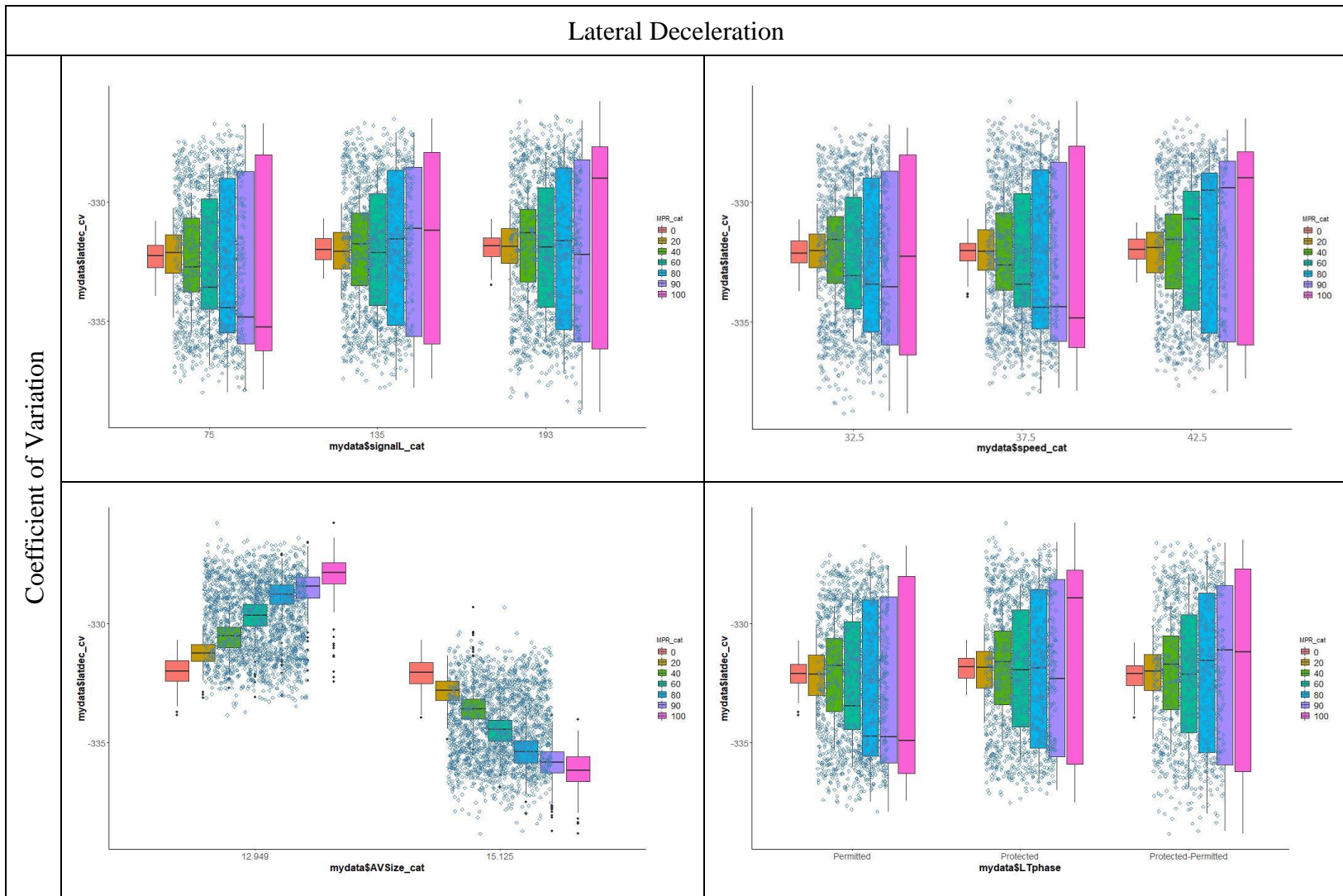
Figure 19 Continued
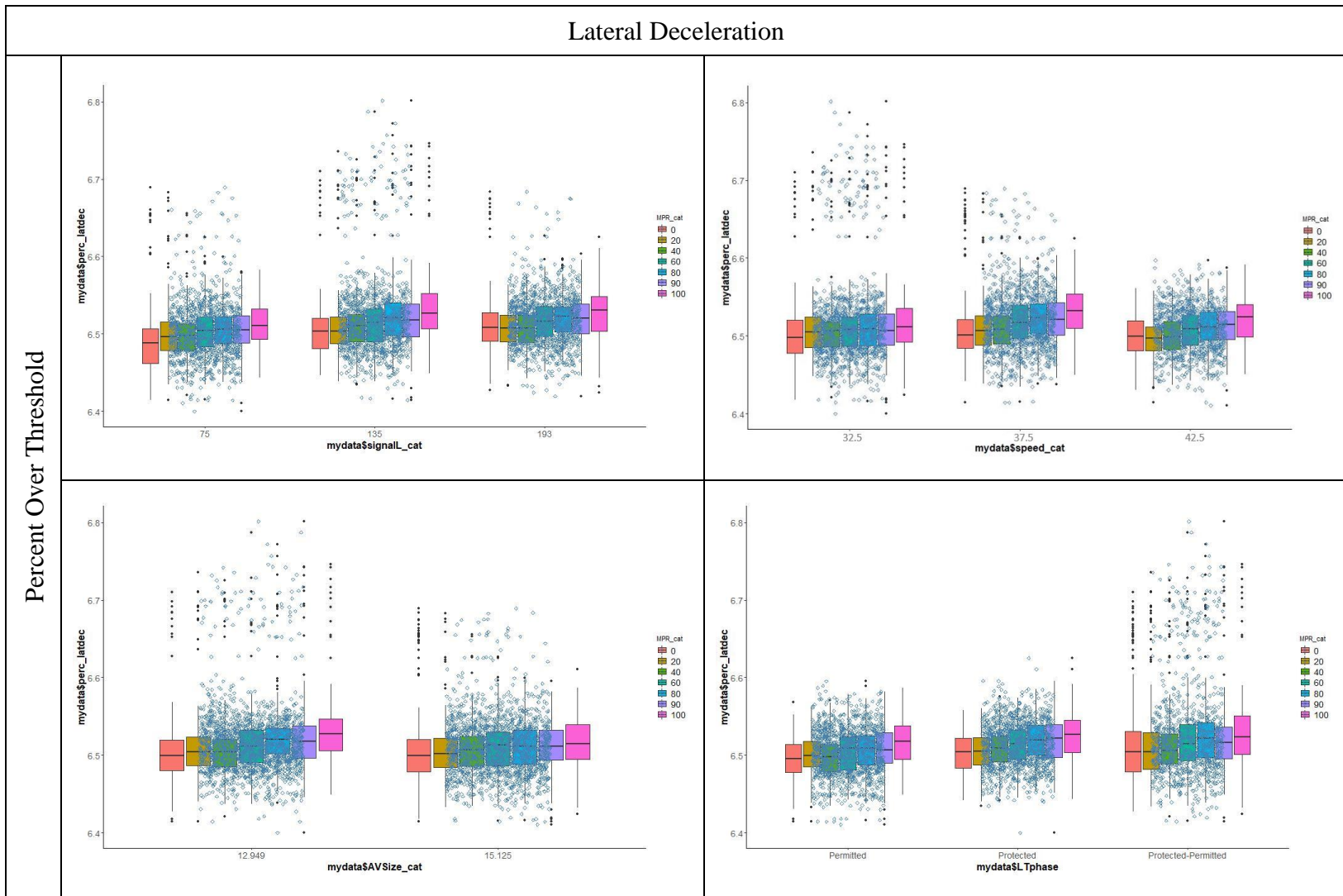
Figure 19 Continued
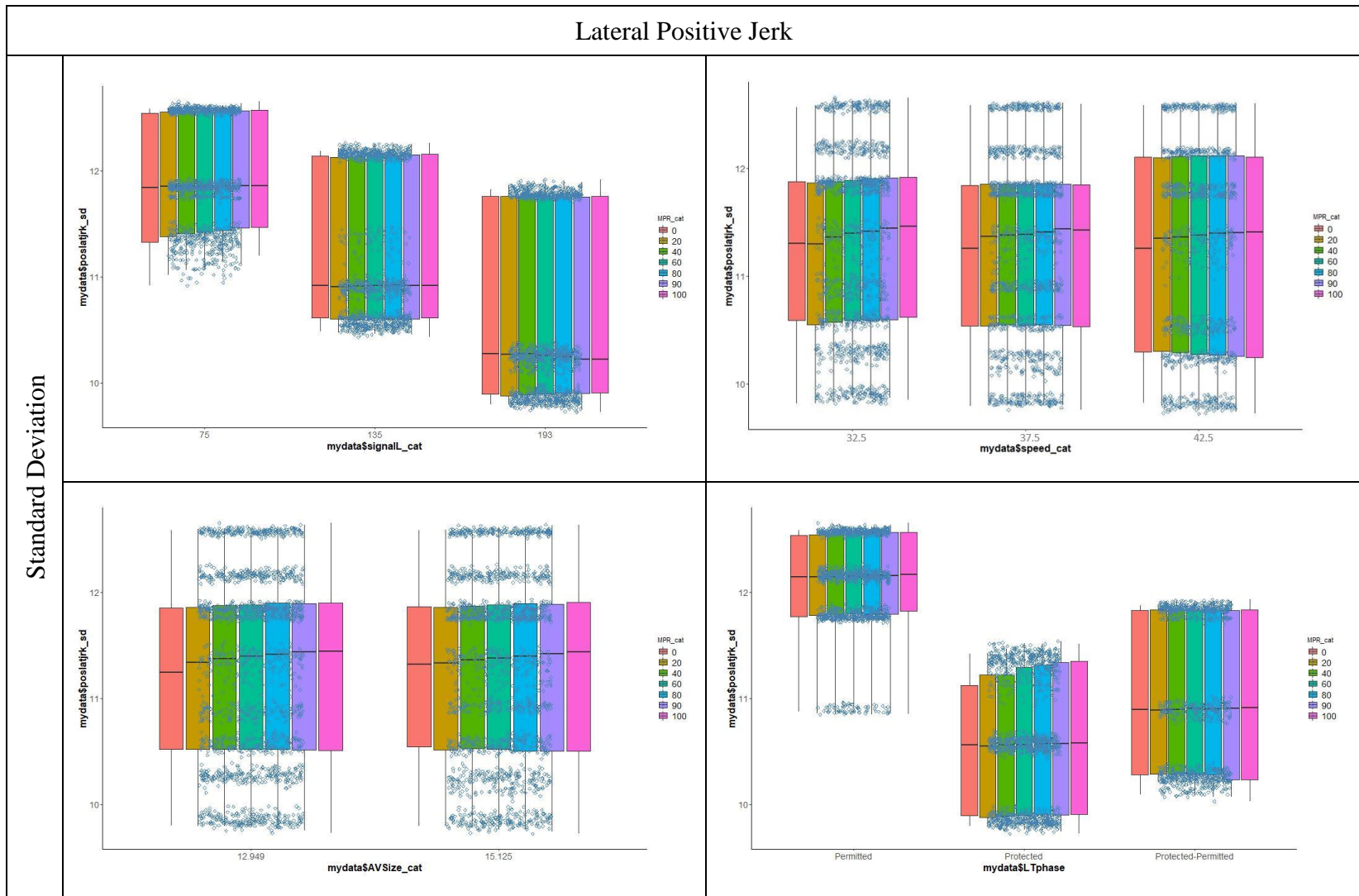
Figure 20. Boxplots of the Volatility Measures: Lateral Positive Jerk
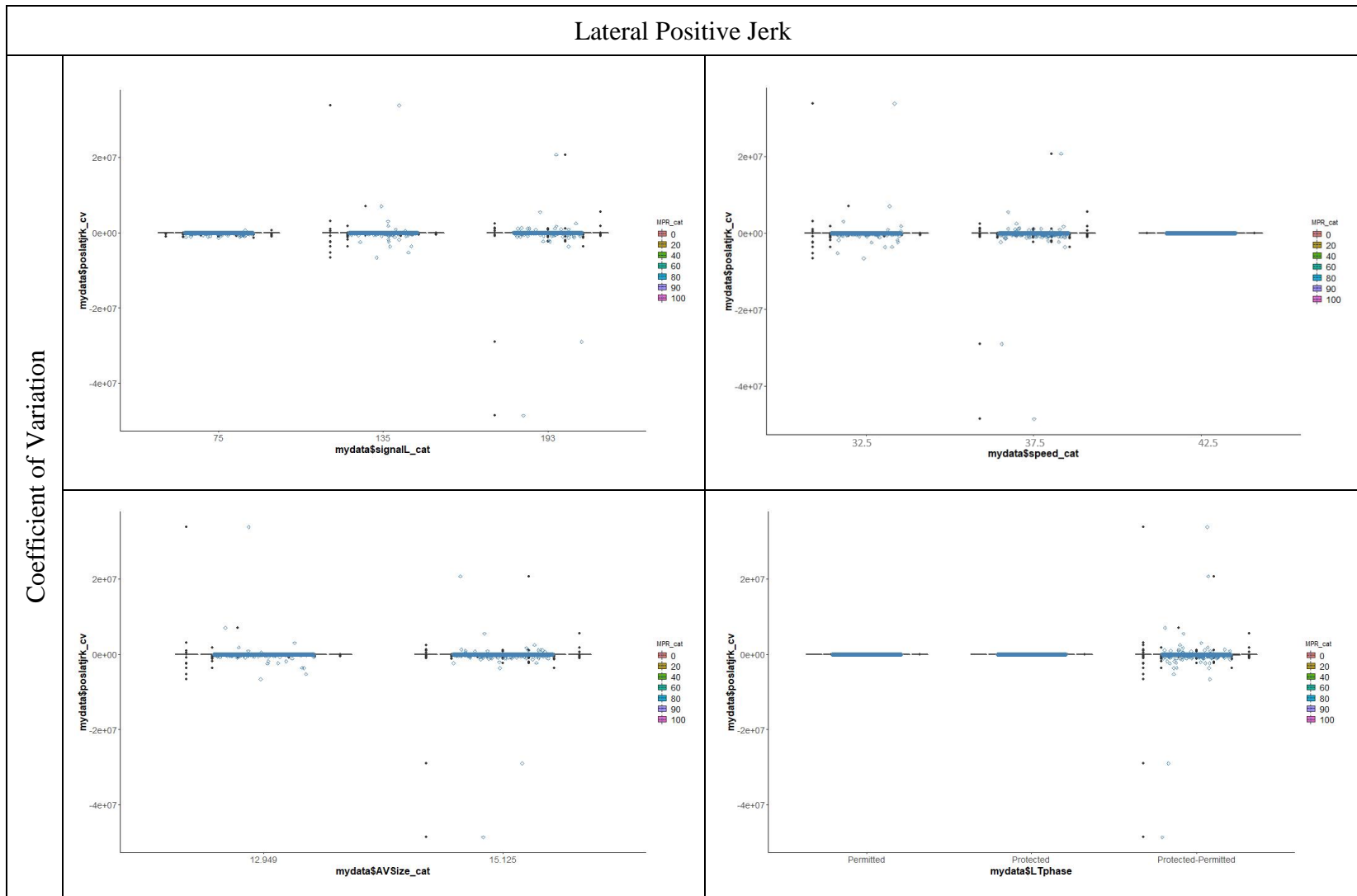
Figure 20 Continued
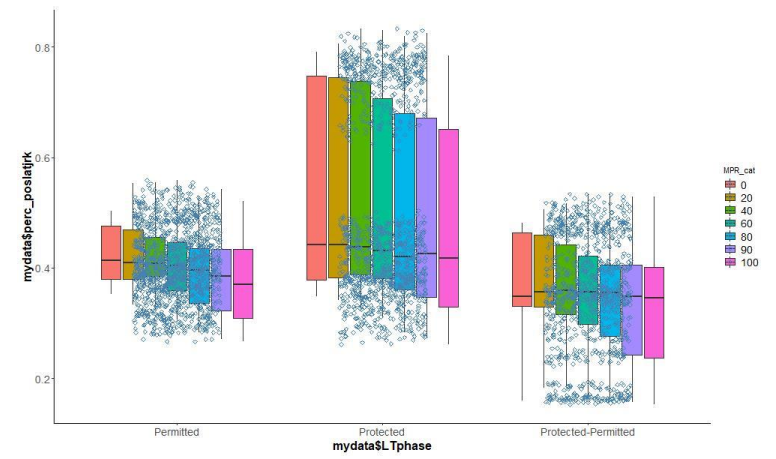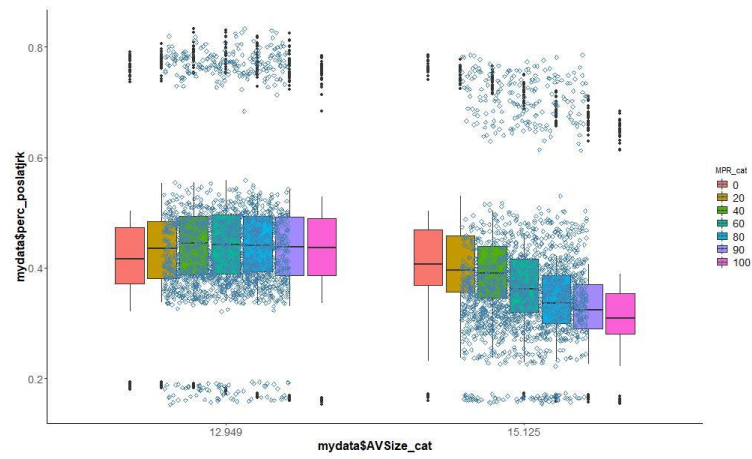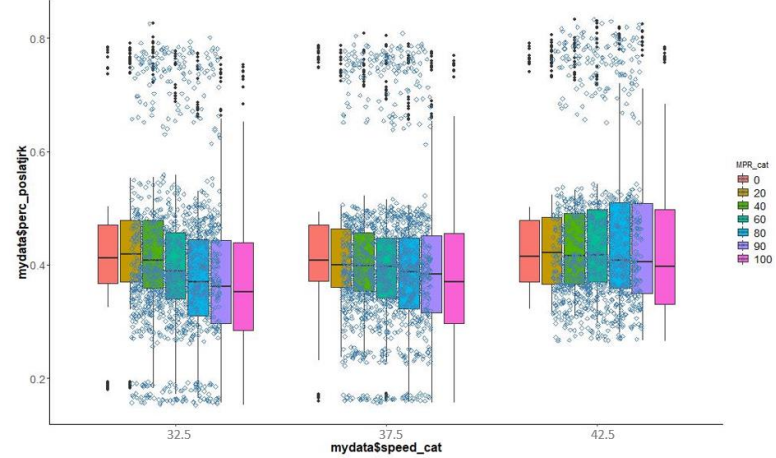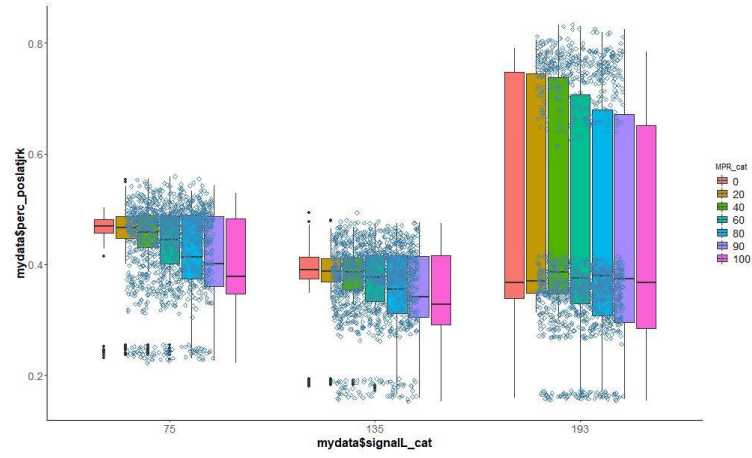
Figure 20 Continued
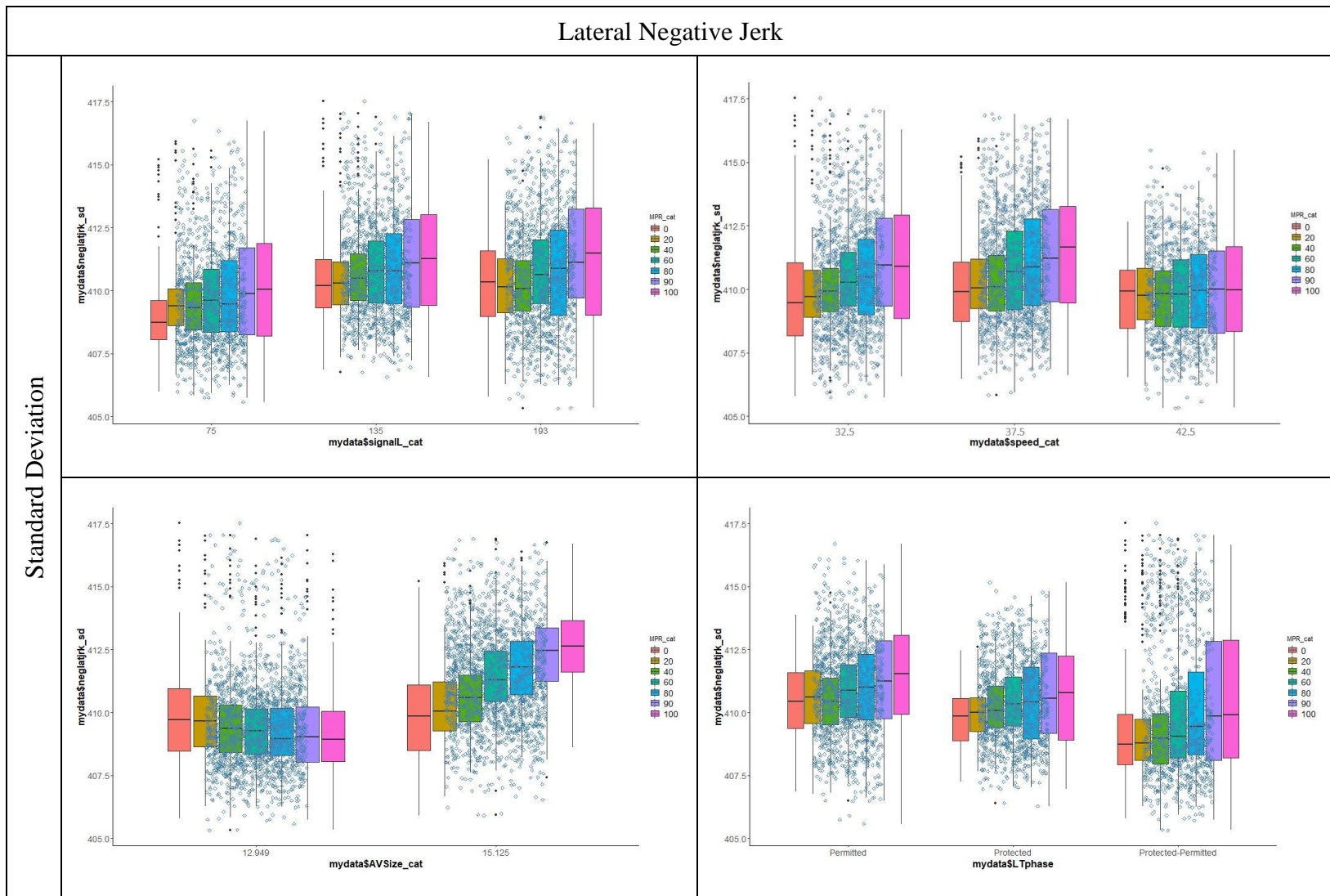
Figure 21. Boxplots of the Volatility Measures: Lateral Negative Jerk
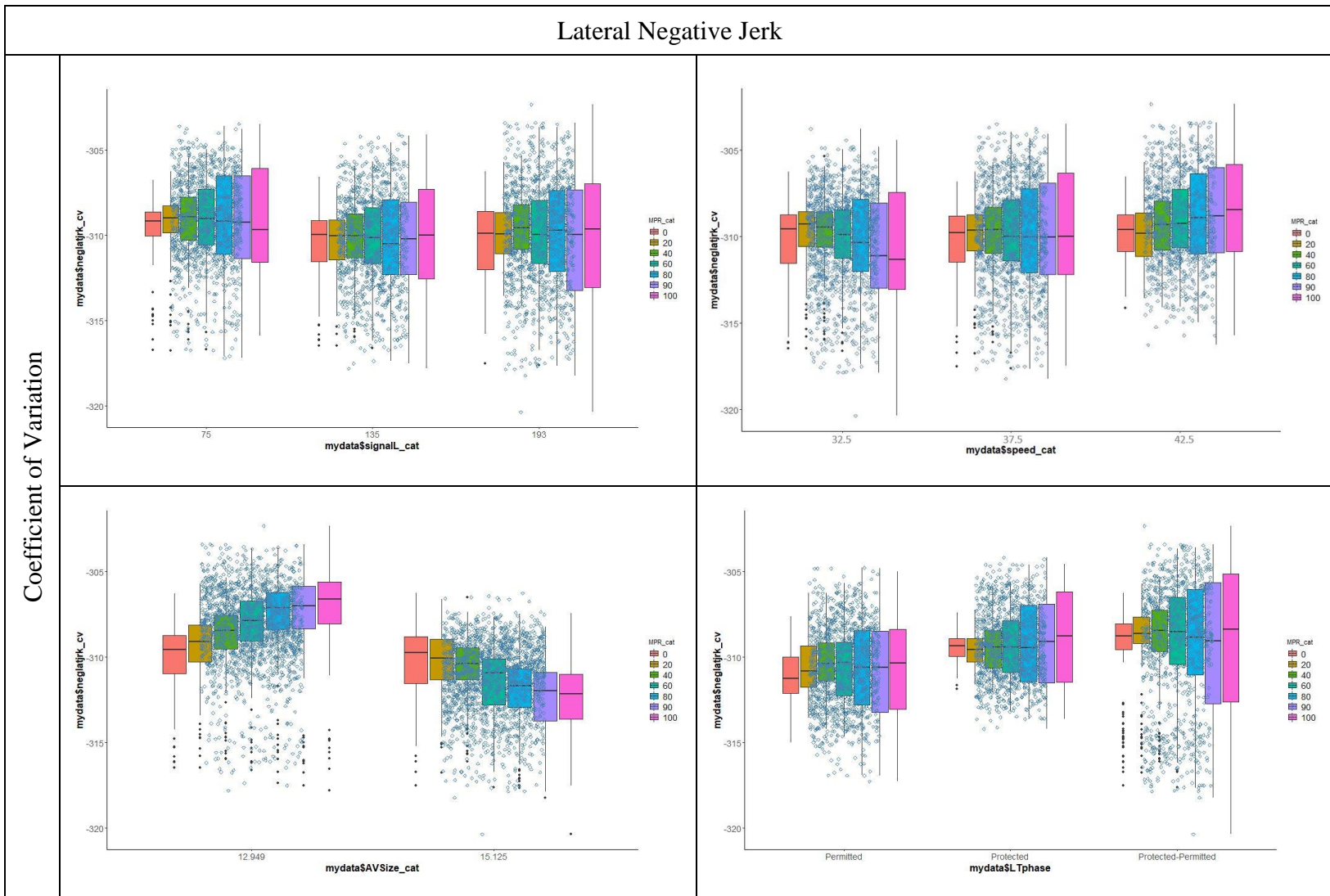
Figure 21 Continued
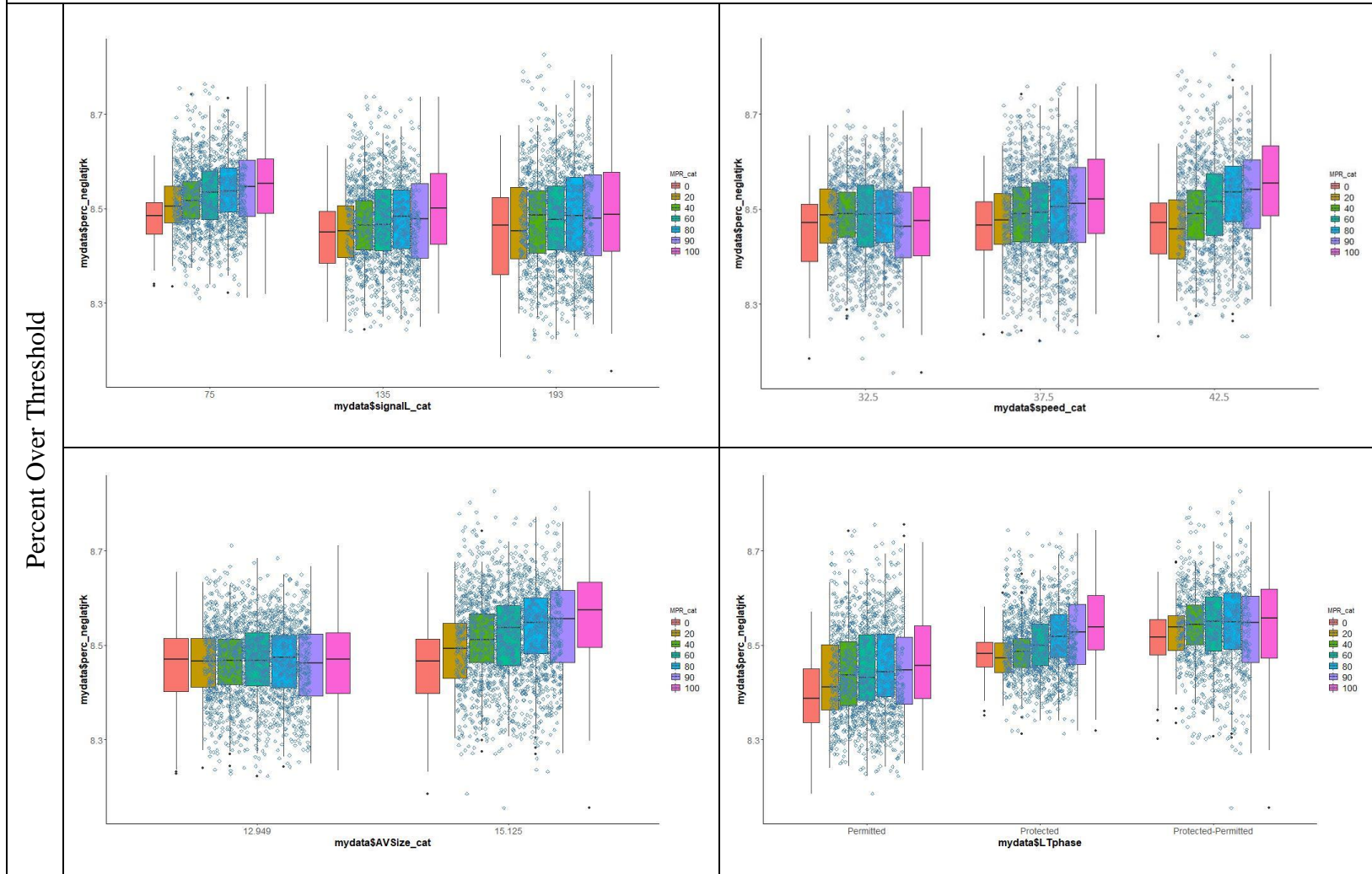
Figure 21 Continued

4.3. Descriptive Statistics for the ML Data

This section provides descriptive statistics of the final dataset that was implemented to develop the ML safety models, at which different jerks were considered as the dependent variables.

After using the defined jerk threshold to determine jerky driving maneuvers, the descriptive statistics of the jerks and the associated non-infrastructure variables of each simulation run were calculated. The outputs of the analysis were then combined so that each simulation run has a row representing its jerk information. Primarily, the merged dataset consisted of NA values that were produced from either the fully AV or RV scenarios, at which there were no RV jerks or AV jerks, respectively. *Table 7* summarizes the descriptive statistics of the simulated variables and their associated histograms, indicating their distributions before manipulating the NA values.

Table 7. Descriptive Statistics of the Simulated Variables

Variable	Missing Values	Mean	S.D.	Min	P 25th	Median	P 75th	Max	Histogram
Total Number of Jerks for AVs	0	1216.85	759.98	0	511	1374	1883	2701	
Total Number of Jerks for AVs (*100)	0	12.17	7.60	0	5.11	13.74	18.83	27.01	
Percent of AV Jerks	3	58.19	34.49	0	24.50	65.10	91.07	100	
Average AV Jerk Value	523	-50.29	1.08	-53.93	-50.98	-50.33	-49.63	-45.51	
AV Jerk S.D.	523	20.43	1.99	11.79	19.14	20.33	21.69	27.42	
Total Number of Jerks for RVs	0	823.60	672.79	0	188	749	1332	2677	
Total Number of Jerks for RVs (*100)	0	8.24	6.73	0	1.88	7.49	13.32	26.77	
Percent of RV Jerks	3	41.45	34.36	0	8.57	34.48	75.10	99.94	

Table 7 Continued

Variable	Missing Values	Mean	S.D.	Min	P 25th	Median	P 75th	Max	Histogram
Average RV Jerk Value	523	-50.38	1.62	-61.49	-51.22	-50.27	-49.41	-43.72	
RV Jerk S.D.	523	21.34	2.65	8.70	19.72	21.24	23.00	32.88	
Total Number of Jerks (*100)	0	20.40	3.00	0	17.56	21.12	22.69	28.48	
MPR	3	0.56	0.35	0	0.2	0.6	0.9	1	
AV Size (ft)	3	14.03	1.09	12.95	12.95	12.95	15.125	15.125	
Speed Limit (mph)	0	37.32	4.04	32.5	32.5	37.5	42.5	42.5	
Signal Length (min)	0	2.24	0.82	1.25	1.25	2.25	3.22	3.22	

To develop the ML regression model, missing values should be removed or imputed. One of the most common practices is to use the average value of a column to replace the missing values from the same column (*125*). Moreover, the nearest neighbor is another method that could be implemented. However, for the purpose of this dissertation, if either the mean or the nearest neighbor is utilized for the missing jerk values, wrong information will be produced. For instance, in a fully AV environment, the jerk for RVs is nonexistent and cannot be replaced by any other values rather than zero. It is the same situation for a fully RV environment. Therefore, all the NA values should be replaced by zero to demonstrate that there is no jerk associated with that specific vehicle type in those scenarios. *Table 8* presents the descriptive statistics after replacing the NAs with zero.

As indicated in *Table 8*, AVs are committed to a higher number of jerky events compared to RVs, means of 1,217 and 824 for AVs and RVs, respectively. One reason for the higher number of AV jerks is the distribution of the AV MPRs. In other words, since AV MPRs of 0%, 20%, 40%, 60%, 80%, 90%, and 100% were examined, on average, 55.71 of the vehicles for all the scenarios were AV and 44.29 were RVs.

Table 8. Descriptive Statistics of the Simulated Variables after Replacing NAs

Variable	Missing Values	Mean	S.D.	Min	P 25th	Median	P 75th	Max	Histogram
Total Number of Jerks for AVs	0	1216.85	759.98	0	511.00	1374.00	1883.00	2701.00	
Total Number of Jerks for AVs (*100)	0	12.17	7.60	0	5.11	13.74	18.83	27.01	
Percent of AV Jerks	0	58.14	34.52	0	24.48	65.09	91.06	100	
Average AV Jerk Value	0	-43.10	17.64	-53.93	-50.87	-50.12	-49.09	0	
AV Jerk S.D.	0	17.51	7.39	0	18.24	19.99	21.44	27.42	
Total Number of Jerks for RVs	0	823.60	672.79	0	188	749	1332	2677	
Total Number of Jerks for RVs (*100)	0	8.24	6.73	0	1.88	7.49	13.32	26.77	
Percent of RV Jerks	0	41.42	34.37	0	8.55	34.48	75.10	99.94	
Average RV Jerk Value	0	-43.17	17.71	-61.49	-51.03	-49.99	-48.73	0	
RV Jerk S.D.	0	18.28	7.86	0	18.55	20.74	22.64	32.88	
Total Number of Jerks (*100)	0	20.40	3.00	0	17.56	21.12	22.69	28.48	
MPR	0	0.56	0.35	0	0.2	0.6	0.9	1	
AV Size (ft)	0	14.02	1.16	0	12.95	12.95	15.125	15.125	
Speed Limit (mph)	0	37.32	4.04	32.5	32.5	37.5	42.5	42.5	
Signal Length (min)	0	2.24	0.82	1.25	1.25	2.25	3.22	3.22	

In addition, *Table 9* confirms that there is a significant difference between the number of jerks for the AVs and RVs in each scenario.

Table 9. T-Test for Comparing the Mean Number of Jerks for AVs and RVs

	Df	Sum Sq	Mean Sq	F-Value	Pr(>F)
Vehicle Type	1	28,263	28,263	548.7	*<0.0001*
Residuals	7,308	376,443	52		

Figure 22 graphically depicts the total number of jerks (*100) by each MPR. As indicated, by increasing the AV MPR, the percentage of jerks rises primarily, but it starts declining at AV MPR of 90%.
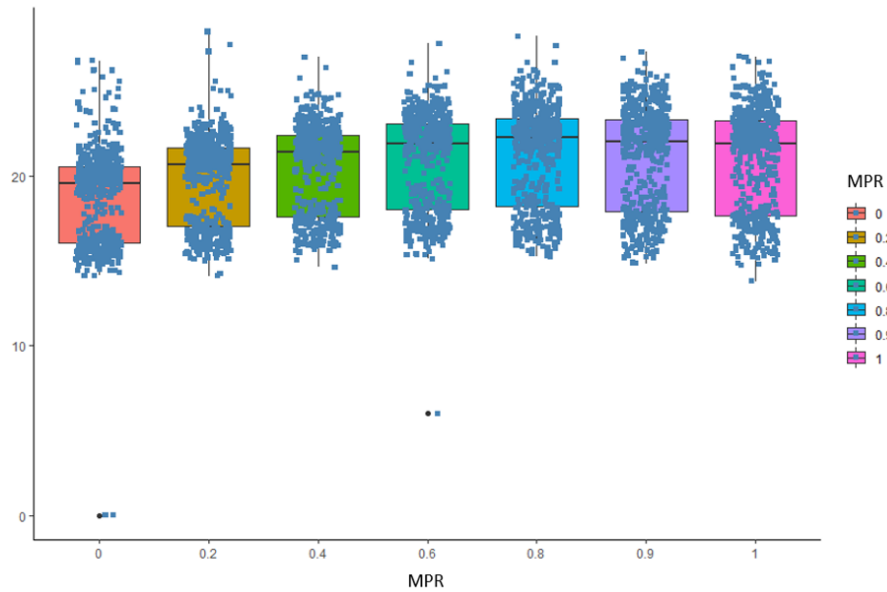
Figure 22. Total Number of Jerks (*100) by AV MPR

To better explore jerks for each vehicle type, the average jerk values were compared for AVs and RVs. According to *Table 8*, RVs execute more abrupt braking maneuvers compared to AVs and, consequently, involve in more hazardous situations; the minimum jerk value for RVs is -61.5 with a standard deviation of 17.8 as for the AVs, these values are -53.9 and 17.7, respectively. Also, as indicated in *Figure 23*, jerk values for RVs experience more fluctuations and sometimes reach to approximately -60 $ft/_{s^3}$ at AV MPR of 90%; however, based on *Figure 24*, the AVs jerks are more concentrated around a point with relatively smaller absolute values, which results in a safer traffic situation.
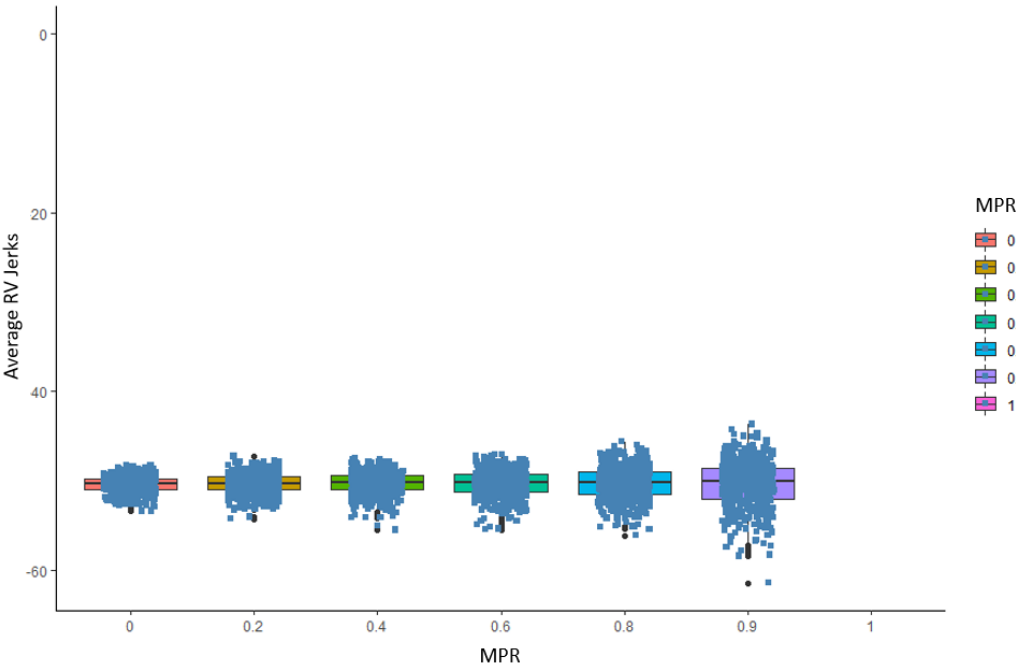
Figure 23. Average RV Jerks by AV MPR
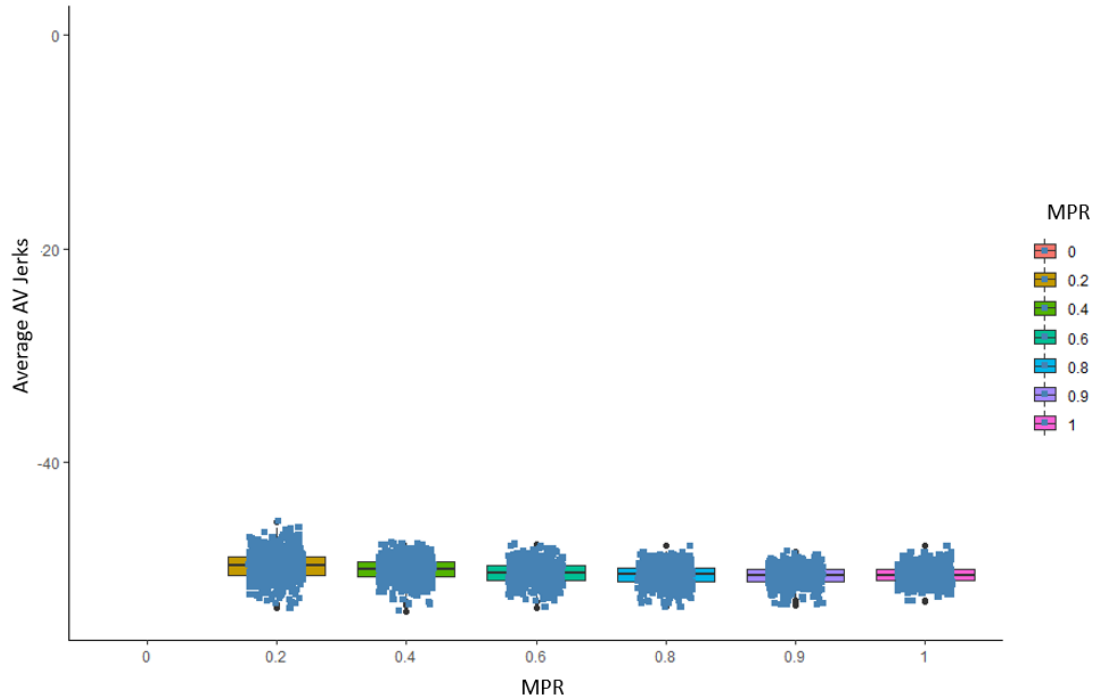
Figure 24. Average AV Jerks by AV MPR

By statistically comparing the average AV jerks and average RV jerks, the results indicate that there is no statistically significant difference between the mean jerk values for AVs and RVs, as shown in *Table 10*.

Figure 25 provides the distribution of the number of jerky maneuvers for AVs, RVs, and the sum of AVs and RVs. The depictions indicate that the number of jerks for AVs in the simulation runs tends to be normally distributed; however, the number of jerks for the RVs is more skewed to the right. It is worth noting that the scales of the graphs are different.

Table 10. T-Test for Comparing the Average Jerk Values for AVs and RVs

	Df	Sum Sq	Mean Sq	F-Value	Pr(>F)
Vehicle Type	1	9	9.23	0.03	0.863
Residuals	7,308	2,282,800	312.37		
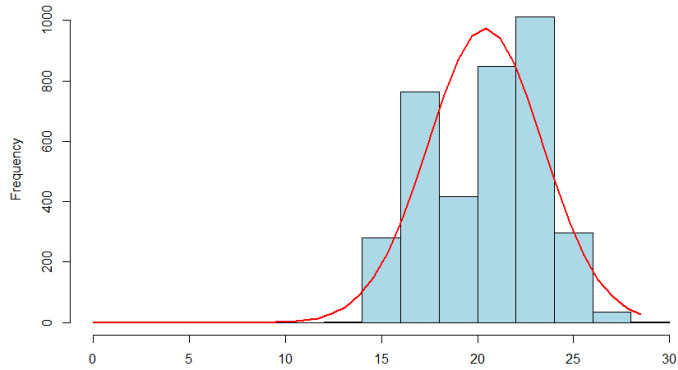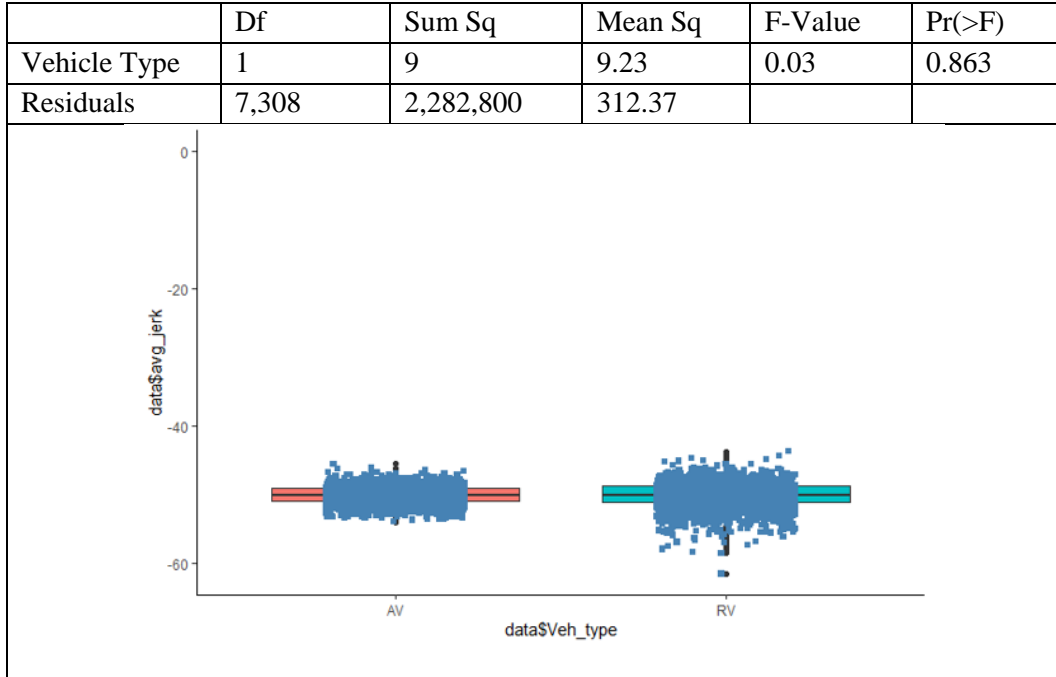
a. Distribution of the Total Number of Jerks/100
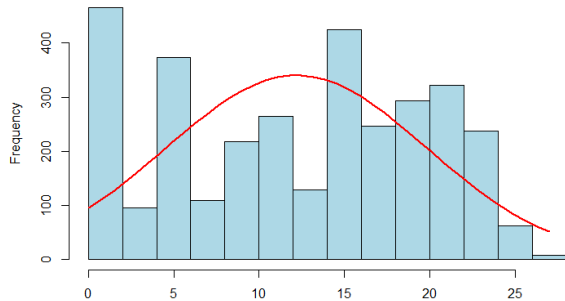
b. Distribution of AV Jerks/100
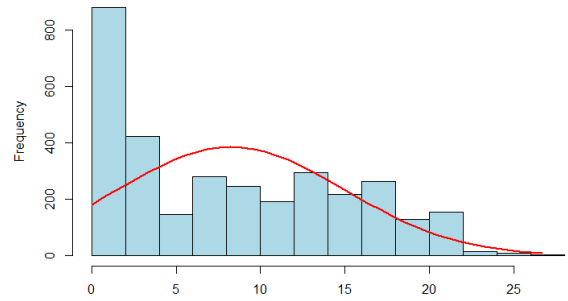
c. Distribution of RV Jerks/100

Figure 25. Distribution of the Jerk Frequency per Simulation Run for AVs vs. RVs

Ultimately, *Figure 26* depicts the distribution of the simulated variables among the simulation runs. As indicated, the variables are reasonably distributed to be able to explain their effects on the jerk, as the response variable.
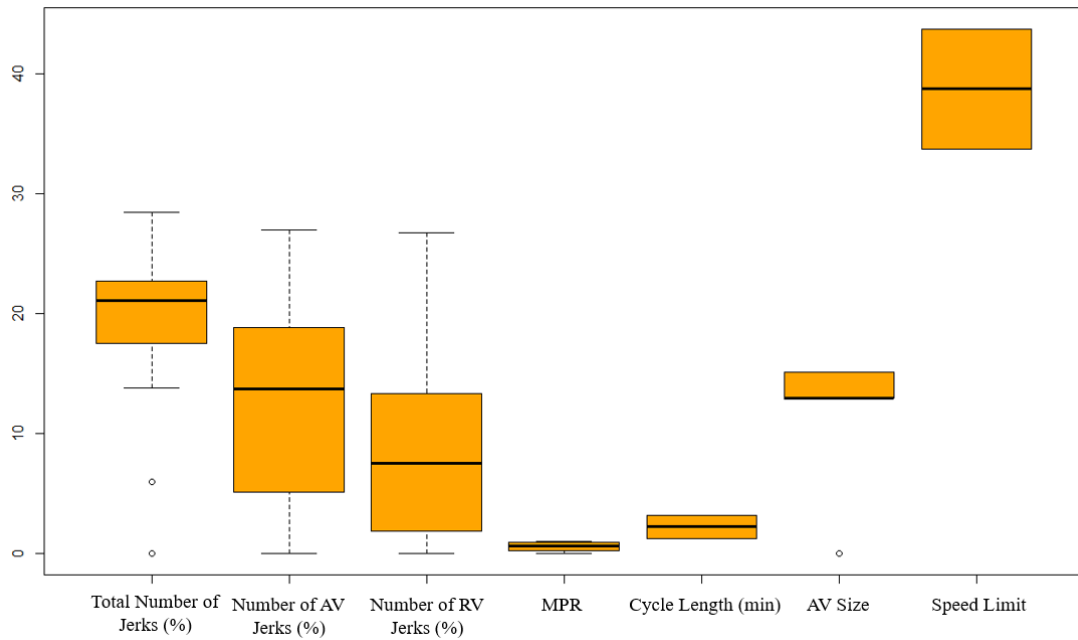
Figure 26. Boxplot for the Simulated Variables

4.4. Chapter Summary

This chapter includes primary statistical analyses of the driving volatility measures, jerk data, and the studied non-infrastructure variables. The variables have been explored by providing descriptive statistics, explanatory analyses, and plots.

Thirty different driving volatility measures for the four studied variables under seven AV MPRs were assessed. Overall, the results indicated that longitudinal driving volatility measures are more sensitive compared to the lateral measurements. However, the

outputs cannot be summarized in detail, and each driving volatility measure should be considered separately.

Analyzing the jerks indicated that by increasing the AV MPR, the percentage of the total jerks increases primarily, but it starts declining at higher AV MPRs. In addition, overall, AVs are involved in a higher number of jerks compared to RVs, but considering the absolute negative jerk values indicated that RVs have larger jerky driving maneuvers, i.e., RVs contribute to more dangerous situations than AVs. The following chapter presents advanced statistical analyses.

CHAPTER V

RESULTS AND DISCUSSIONS: ADVANCED STATISTICAL

ANALYSES

After conducting the descriptive statistics and explanatory analyses in Chapter 4, this chapter focuses on advanced statistical analyses. It describes the results of the data analyses by presenting how traffic safety would be affected by changing the non-infrastructure study variables. Traffic safety was evaluated from different aspects through developing different safety models for longitudinal driving volatility measures, lateral driving volatilities, and the percentage of jerks. Ultimately, correlation tests were conducted to compare the number of SSAM conflicts and the number of jerks.

5.1. Driving Volatility

The following sections present the statistical safety models for both longitudinal and lateral driving volatility measures, respectively, based on the non-infrastructure variables and AV MPR.

5.1.1. *Longitudinal Driving Volatility Safety Models*

By conducting a preliminary analysis of the driving volatility measures, each volatility measure could be statistically evaluated to determine how the study variables influence the volatility measures, and hence traffic safety. The results of conducting a correlation test before developing safety models indicated that the variables are not correlated. To assess the effects of the study variables on the driving volatility measures,

various GLMs were developed. *Table 11* provides a summary of the GLMs and their

associated goodness of fit values.

Table 11. Regression Models for the Longitudinal Driving Volatility Measures

	Estimate	Std. Error	t value	Pr(>|t|)
Speed S.D.				
(Intercept)	14.275	0.102	140.148	<0.0001
MPR	-0.002	0.000	-11.808	<0.0001
Signal Cycle Length	0.003	0.000	22.634	<0.0001
speed	0.030	0.002	19.476	<0.0001
AV Size	0.062	0.006	11.169	<0.0001
LT Protected	1.482	0.014	103.336	<0.0001
LT Protected-Permitted	1.142	0.015	75.626	<0.0001
Model Fit: 0.7786				
Percent of Speed Over Threshold				
(Intercept)	3.672	0.097	37.928	<0.0001
MPR	-0.002	0.000	-10.327	<0.0001
Signal Cycle Length	0.002	0.000	17.657	<0.0001
speed	0.057	0.001	38.838	<0.0001
AV Size	0.035	0.005	6.548	<0.0001
LT Protected	0.303	0.014	22.220	<0.0001
LT Protected-Permitted	0.341	0.014	23.742	<0.0001
Model Fit: 0.419				
Acceleration S.D.				
(Intercept)	2.355	0.009	258.114	<0.0001
MPR	0.003	0.000	118.271	<0.0001
Signal Cycle Length	-0.002	0.000	-137.461	<0.0001
speed	0.003	0.000	12.126	<0.0001
LT Protected	-0.043	0.002	-20.606	<0.0001
LT Protected-Permitted	-0.012	0.002	-5.525	<0.0001
Model Fit: 0.902				

Table 11 Continued

	Estimate	Std. Error	t value	Pr(>|t|)
Acceleration C.V.				
(Intercept)	90.567	2.129	42.542	<0.0001
MPR	0.163	0.004	44.624	<0.0001
Signal Cycle Length	0.375	0.003	143.524	<0.0001
speed	0.114	0.032	3.555	0.00038
AV Size	0.888	0.116	7.641	<0.0001
LT Protected	35.358	0.300	117.967	<0.0001
LT Protected-Permitted	25.894	0.316	82.007	<0.0001
Model Fit: 0.9136				
Percent of Acceleration Over Threshold				
(Intercept)	6.6215	0.1481	44.7134	<0.0001
MPR	0.0124	0.0004	30.0882	<0.0001
Signal Cycle Length	0.0008	0.0003	2.6607	0.008
speed	0.0120	0.0036	3.3252	0.0009
LT Protected	0.1824	0.0338	5.3929	<0.0001
LT Protected-Permitted	-0.0018	0.0356	-0.0515	0.9590
Model Fit: 0.2109				
Deceleration S.D.				
(Intercept)	2.194	0.027	81.167	<0.0001
Signal Cycle Length	-0.001	0.000	-27.175	<0.0001
speed	0.010	0.000	24.254	<0.0001
AV Size	0.013	0.001	8.852	<0.0001
LT Protected	0.213	0.004	55.818	<0.0001
LT Protected-Permitted	0.250	0.004	62.061	<0.0001
Model Fit: 0.6188				
Deceleration C.V.				
(Intercept)	-140.645	0.695	-202.401	<0.0001
MPR	-0.022	0.002	-11.261	<0.0001
Signal Cycle Length	-0.013	0.001	-9.237	<0.0001
speed	0.457	0.017	26.939	<0.0001

Table 11 Continued

	Estimate	Std. Error	t value	Pr(>|t|)
LT Protected	11.840	0.159	74.608	<0.0001
LT Protected-Permitted	6.927	0.167	41.473	<0.0001
Model Fit: 0.642				
Percent of Deceleration Over Threshold				
(Intercept)	9.286	0.162	57.298	<0.0001
MPR	0.006	0.000	12.845	<0.0001
speed	-0.040	0.004	-9.834	<0.0001
LT Protected	-0.730	0.038	-19.166	<0.0001
LT Protected-Permitted	0.054	0.040	1.340	0.18
Model Fit: 0.6188				
Positive Jerk S.D.				
(Intercept)	9.407	0.050	189.350	<0.0001
MPR	0.000	0.000	3.154	0.0016
Signal Cycle Length	-0.005	0.000	-85.186	<0.0001
speed	-0.003	0.001	-4.079	<0.0001
AV Size	-0.016	0.003	-5.909	<0.0001
LT Protected	-0.656	0.007	-93.808	<0.0001
LT Protected-Permitted	-0.455	0.007	-61.700	<0.0001
Model Fit: 0.823				
Positive Jerk C.V.				
(Intercept)	45.172	2.132	21.188	<0.0001
Signal Cycle Length	0.323	0.003	123.044	<0.0001
speed	0.136	0.032	4.206	<0.0001
AV Size	1.061	0.117	9.078	<0.0001
LT Protected	44.981	0.301	149.199	<0.0001
LT Protected-Permitted	33.067	0.318	104.112	<0.0001
Model Fit: 0.4925				
Negative Jerk S.D.				
(Intercept)	8.333	0.002	3546.206	<0.0001
MPR	0.0001	0.00001	12.456	<0.0001

Table 11 Continued

	Estimate	Std. Error	t value	Pr(>|t|)
Signal Cycle Length	-0.00004	0.000004	-10.694	<0.0001
AV Size	-0.0005	0.0002	-2.841	0.0045
LT Protected	0.020	0.0004	49.498	<0.0001
LT Protected-Permitted	0.010	0.0004	24.110	<0.0001
Model Fit: 0.4284				
Negative Jerk C.V.				
(Intercept)	-83.564	0.042	-1977.960	<0.0001
MPR	-0.0002	0.0001	-3.368	0.0008
Signal Cycle Length	0.0004	0.0001	8.197	<0.0001
speed	-0.007	0.001	-10.266	<0.0001
AV Size	0.007	0.002	2.961	0.0031
LT Protected	0.124	0.006	20.854	<0.0001
LT Protected-Permitted	0.113	0.006	17.971	<0.0001
Model Fit: 0.1675				
Percent of Negative Jerk Over Threshold				
(Intercept)	4.629	0.025	181.864	<0.0001
MPR	0.001	0.00004	12.974	<0.0001
Signal Cycle Length	0.0001	0.00003	2.087	0.0369
speed	0.002	0.0004	4.741	<0.0001
AV Size	-0.008	0.001	-5.579	<0.0001
LT Protected	-0.015	0.004	-4.061	<0.0001
LT Protected-Permitted	-0.083	0.004	-21.971	<0.0001
Model Fit: 0.1762				

As presented in the previous table, not all the models provide a good fit to be implemented to estimate traffic safety. Also, it is notable that a few driving volatility measures were not affected by any of the study variables and were excluded from the table.

In general, increasing the AV MPR significantly decreases the longitudinal speed S.D., the percent of speed over the threshold, deceleration C.V., and negative jerk C.V. On

the other hand, the variable MPR has a positive relationship with the rest of the driving volatility measures. In other words, increasing the AV MPR reduces some of the longitudinal volatility measures and enhances traffic safety, and at the same time, increases the other longitudinal volatility measures.

The majority of the models, which provide a good fit, indicate that there is a positive relationship between the AV size and the longitudinal driving volatility measures, including speed S.D., acceleration C.V., deceleration S.D., and positive jerk S.D.

Longer signal cycle lengths significantly reduce acceleration S.D., deceleration S.D., deceleration C.V., and negative jerk S.D. The reductions in the S.D.s indicate that the traffic flow is smoother as the vehicles do not need to come to a full stop frequently (and conduct abrupt driving maneuvers) due to the red signal. The results also suggest that the average speed limit of the legs of the intersection significantly influences the driving volatility measures. For instance, increasing the average speed limit results in a significant reduction in the percent of deceleration above the threshold, positive jerk S.D., and negative jerk C.V. On the other hand, increasing the speed limit raises the other driving volatility measures, e.g., the speed S.D. increases when the speed limit increases, which is expected since the difference between the highest speed of the vehicles and the vehicles that slow down at the signalized intersection increases.

Eventually, LT signal phasing is another significant variable in the driving volatility models. In the models, the permitted LT signal phasing was considered as the base scenario, and protected-permitted and protected LT phasing were compared to the base condition. For instance, the speed S.D. model indicates that the speed S.D. increases by 1.48 and 1.14 when providing protected LT phasing and protected-permitted LT

111

phasing compared to the permitted LT phasing, respectively, if keeping the rest of the variables constant.

In summary, since various safety models for the longitudinal driving volatility measures were developed, the studied variables presented different effects on each volatility measure. Hence, each given model could be implemented according to the purpose of the safety analysis to predict each volatility measure based on the given non-infrastructure variables.

5.1.2. Lateral Driving Volatility Safety Models

This section intends to assess the lateral driving volatility measures, which have not been studied extensively and thoroughly in the previous research studies.

By conducting a preliminary analysis of the lateral driving volatility measures, statistical modeling could be conducted to determine how the non-infrastructure variables influence each lateral driving volatility measure. To do so, various GLMs were developed. *Table 12* summarizes the results of GLMs with their goodness of fit provided alongside each model to indicate how well the independent variables predict each lateral driving volatility measure.

Table 12. Regression Models for the Lateral Driving Volatility Measures

	Estimate	Std. Error	t value	Pr(>|t|)
Speed S.D.				
(Intercept)	0.411	0.004	95.383	<0.0001
MPR	-0.0001	0.00001	-17.146	<0.0001
Signal Cycle Length	-0.0004	0.00001	-78.119	<0.0001
speed	0.001	0.0001	13.098	<0.0001

Table 12 Continued

	Estimate	Std. Error	t value	Pr(>|t|)
AV Size	-0.003	0.0002	-11.868	<0.0001
LT Protected	-0.018	0.001	-29.735	<0.0001
LT Protected-Permitted	-0.023	0.001	-35.770	<0.0001
Model Fit: 0.6942				
Percent of Speed Over Threshold				
(Intercept)	0.167	0.003	61.051	<0.0001
MPR	-0.0001	0.000005	-17.975	<0.0001
Signal Cycle Length	-0.0003	0.000003	-83.258	<0.0001
speed	0.001	0.0000	13.293	<0.0001
AV Size	-0.002	0.0001	-12.195	<0.0001
LT Protected	-0.012	0.0004	-31.755	<0.0001
LT Protected-Permitted	-0.015	0.0004	-35.955	<0.0001
Model Fit: 0.7171				
Acceleration S.D.				
(Intercept)	4.238	0.045	94.756	<0.0001
MPR	-0.001	0.0001	-17.557	<0.0001
Signal Cycle Length	-0.004	0.0001	-78.655	<0.0001
speed	0.009	0.001	13.125	<0.0001
AV Size	-0.029	0.002	-12.060	<0.0001
LT Protected	-0.189	0.006	-29.974	<0.0001
LT Protected-Permitted	-0.239	0.007	-36.105	<0.0001
Model Fit: 0.6975				
Acceleration C.V.				
(Intercept)	1757.914	44.980	39.082	<0.0001
MPR	1.215	0.077	15.726	<0.0001
Signal Cycle Length	3.429	0.055	62.087	<0.0001
speed	-8.059	0.677	-11.904	<0.0001
AV Size	40.225	2.455	16.382	<0.0001
LT Protected	147.617	6.333	23.310	<0.0001
LT Protected-Permitted	222.177	6.672	33.302	<0.0001

Table 12 Continued

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Model Fit: 0.6103				
Percent of Acceleration Over Threshold				
(Intercept)	0.167	0.003	61.051	<0.0001
MPR	-0.0001	0.000005	-17.975	<0.0001
Signal Cycle Length	-0.0003	0.000003	-83.258	<0.0001
speed	0.001	0.00004	13.293	<0.0001
AV Size	-0.002	0.0001	-12.195	<0.0001
LT Protected	-0.012	0.0004	-31.755	<0.0001
LT Protected-Permitted	-0.015	0.0004	-35.955	<0.0001
Model Fit: 0.7171				
Deceleration S.D.				
(Intercept)	24.4605	0.0202	1212.8152	0
MPR	0.0005	0.00003	15.3654	<0.0001
Signal Cycle Length	0.0002	0.00002	7.0521	<0.0001
speed	-0.0009	0.0003	-2.8039	0.0051
AV Size	0.0047	0.0011	4.2978	<0.0001
LT Protected	0.0171	0.0028	6.0182	<0.0001
LT Protected-Permitted	0.0422	0.0030	14.1205	<0.0001
Model Fit: 0.1234				
Deceleration C.V.				
(Intercept)	-303.803	0.438	-693.207	<0.0001
Signal Cycle Length	0.002	0.001	4.034	<0.0001
speed	0.011	0.007	1.700	0.0891
AV Size	-2.073	0.024	-86.262	<0.0001
LT Protected	0.264	0.062	4.256	<0.0001
LT Protected-Permitted	-0.038	0.065	-0.579	0.563
Model Fit: 0.6755				
Percent of Deceleration Over Threshold				
(Intercept)	6.537	0.011	587.563	<0.0001
MPR	0.0002	0.00002	9.304	<0.0001

Table 12 Continued

	Estimate	Std. Error	t value	Pr(>|t|)
Signal Cycle Length	0.0001	0.00001	7.814	<0.0001
speed	-0.0005	0.0002	-2.832	0.0046
AV Size	-0.003	0.001	-4.547	<0.0001
LT Protected	0.009	0.002	5.641	<0.0001
LT Protected-Permitted	0.021	0.002	12.836	<0.0001
Model Fit: 0.0884				
Positive Jerk S.D.				
(Intercept)	14.230	0.071	200.632	<0.0001
MPR	0.0003	0.0001	2.620	0.0088
Signal Cycle Length	-0.011	0.0001	-124.270	<0.0001
speed	-0.005	0.001	-4.515	<0.0001
AV Size	-0.036	0.004	-9.187	<0.0001
LT Protected	-1.494	0.010	-149.591	<0.0001
LT Protected-Permitted	-1.106	0.011	-105.145	<0.0001
Model Fit: 0.9174				
Negative Jerk S.D.				
(Intercept)	8.333	0.002	3546.206	<0.0001
MPR	0.0001	0.00001	12.456	<0.0001
Signal Cycle Length	-0.00004	0.000004	-10.694	<0.0001
AV Size	-0.0005	0.0002	-2.841	0.0045
LT Protected	0.020	0.0004	49.498	<0.0001
LT Protected-Permitted	0.010	0.0004	24.110	<0.0001
Model Fit: 0.4284				
Percent of Negative Jerk Over Threshold				
(Intercept)	7.947	0.021	384.077	<0.0001
MPR	0.001	0.00004	14.281	<0.0001
Signal Cycle Length	-0.0004	0.00003	-16.376	<0.0001
speed	0.004	0.0003	12.793	<0.0001
AV Size	0.026	0.001	23.118	<0.0001
LT Protected	0.065	0.003	22.230	<0.0001

Table 12 Continued

	Estimate	Std. Error	t value	Pr(>|t|)
LT Protected-Permitted	0.101	0.003	32.813	<0.0001
Model Fit: 0.3738				

As for the longitudinal volatility measures, not all the models in *Table 12* for lateral volatility measures provide a good fit. However, the outputs could be used to determine the overall statistical influences of the variables on each lateral volatility measure. The models for speed S.D., percent of speed over the threshold, acceleration S.D., acceleration C.V., percent of acceleration over the threshold, deceleration C.V., and positive jerk S.D. have a good model fit and could be implemented to evaluate the lateral safety of the mixed traffic environments accurately, given the independent variables.

According to the models, increasing the AV MPR reduces the lateral speed S.D., percent of speed over threshold, acceleration S.D., and the percent of acceleration over threshold significantly. The majority of the models with an acceptable goodness of fit value show a negative relationship between the cycle length and the lateral driving volatility measures. In fact, increasing the cycle length reduces the lateral driving volatility measures. Moreover, increasing the AV size reduces the overall lateral driving volatility measures significantly. Lastly, the average speed limit shows different behaviors on each driving volatility measure.

5.2. Evaluating Traffic Safety through ML Regression Models

After evaluating safety models for 30 lateral and longitudinal driving volatility measures, this section develops and assesses three SSM models in more detail by conducting ML analyses.

The analyses include data processing for ML as well as developing the ML safety models using different jerks as the response variables.

5.2.1. *Preprocessing Data to Develop ML Algorithms*

The first step to develop an ML algorithm is to check the structure of the variables. ML regression does not correctly analyze non-ordinal categorical variables; therefore, any non-ordinal categorical variables should be converted to several binary variables according to the number of classes of the categorical variables.

One-hot encoding (dummy variable) can be used to convert categorical variables into a series of binary variables (*126*). It is notable that just replacing the categories of a categorical variable with numbers may not result in a meaningful output, especially if the variable is not ordinal. Hence, variable LT signal phasing, which was a categorical variable, was converted to a series of dummy variables for further analysis using one-hot encoding.

A correlation test was done as the next step before conducting the algorithm development. *Figure 27* depicts the correlation test to examine the correlation between the predictors.
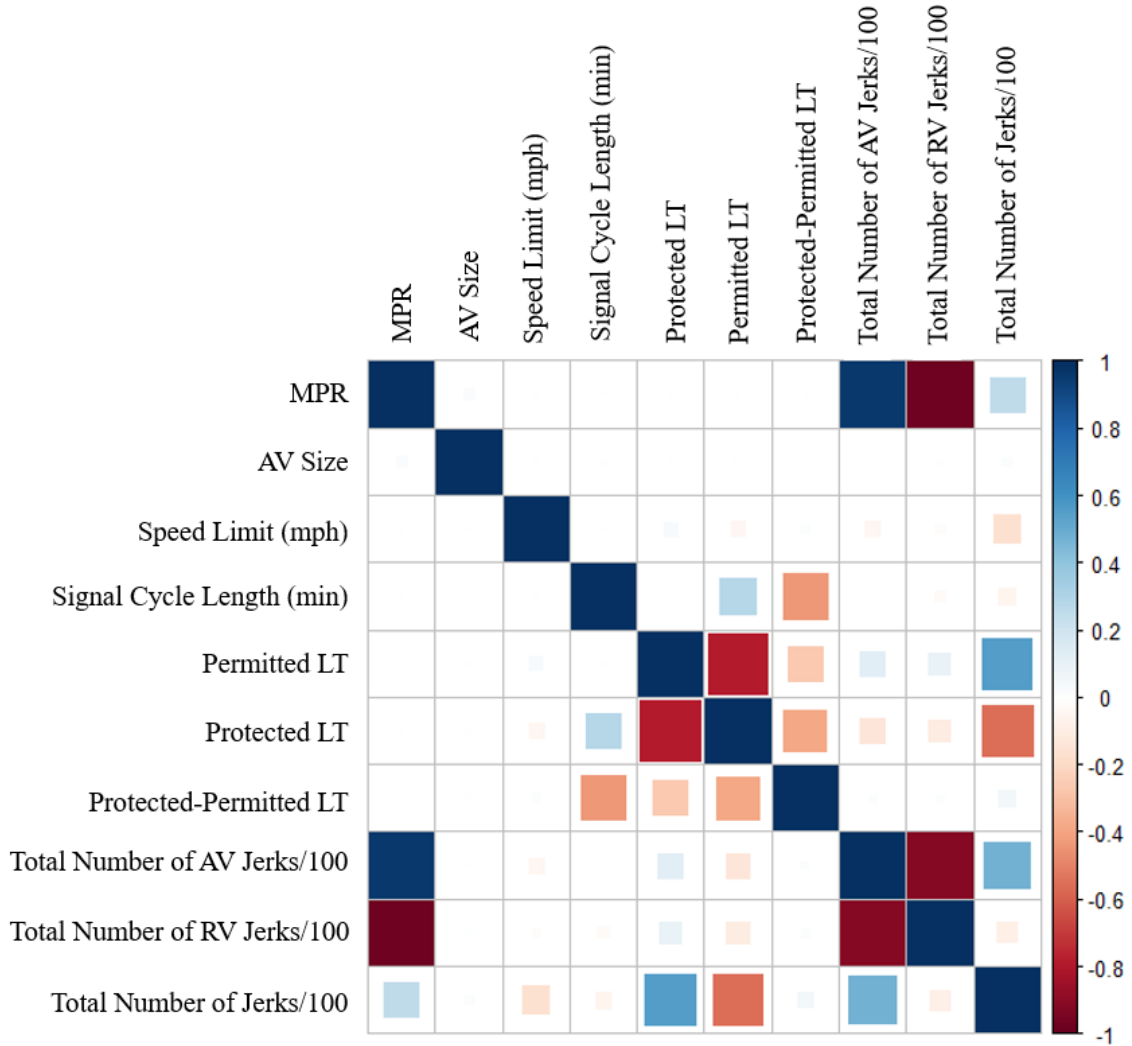
Figure 27. Correlation Test for the Predictors of the Regression Models

As indicated, for the predictors, only "LT protected" and "LT permitted" are highly correlated, with a Pearson's correlation value of -0.79. Therefore, all variables could be included in the ML models simultaneously without dealing with any interaction terms. The following section expands on the development of the ML algorithm.

ML Algorithms

To evaluate safety at signalized intersections under various AV MPRs and the study variables, three different backward elimination ML models were developed considering the following variables as the dependent variables in each model, separately:

a. Percentage of the total jerks;

b. Percentage of the AV jerks; and

c. Percentage of the RV jerks.

In each model, the primary dataset was split into a training set and a test set based on the dependent variable. The training set, consisting of 75% of the original data, was used to train the model. Also, the test set, the remaining 25% of the original dataset, was used to evaluate the performance of the model. AV size, signal cycle length, speed limit, LT signal phasing, and AV MPR were used as predictors.

It is worth mentioning that even though the scenarios replicated the peak-hour traffic condition, the speed limit was still an influencing factor on the speed distribution of the vehicles, as indicated in *Table 13*. As depicted, by increasing the speed limit along the legs of the intersection, the mean operating speed of the vehicles increased as well. Statistical comparison was also conducted and confirmed that the means of the operating speeds of the vehicles were statistically different for the different speed limits. Therefore, the speed limit could be considered in the safety models to determine if it played a role in causing jerky driving maneuvers.

Table 13. T-Test for Comparing the Operating Speed by Speed Limit

	Df	Sum Sq	Mean Sq	F-Value	Pr(>F)
Vehicle Type	2	1.749×10^7	8,743,552	29,697	*<0.0001*
Residuals	7,483,644	2.203×10^9	294		
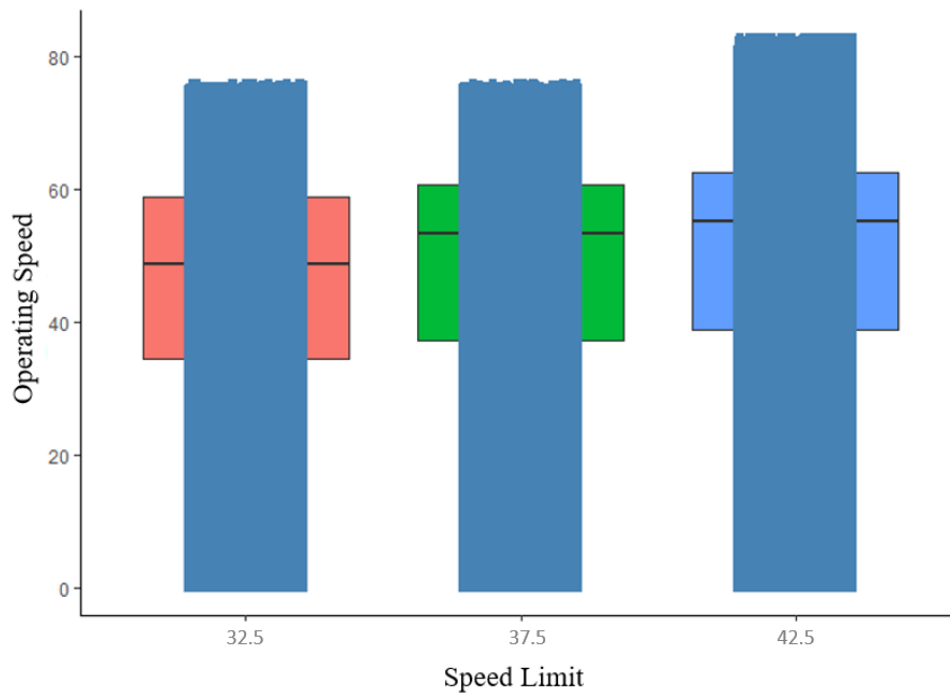
The x-axis of the graph:
- Speed Limit 32.5 mph → EB and WB: 35 mph, SB: 40 mph, NB: 20 mph
- Speed Limit 37.5 mph → EB and WB: 40 mph, SB: 45 mph, NB: 25 mph
- Speed Limit 42.5 mph → EB and WB: 45 mph, SB: 50 mph, NB: 30 mph

By having all the study variables, models for each dependent variable could be developed. The following sections cover the models.

5.2.2.1. ML Model with Percentage of Total Jerks

In the first model, the percent of the total jerks (considering both AVs and RVs) was set as the dependent variable, and the other variables were included as the regressors in a GLM. *Table 14* provides the results of the ML model.

Table 14. ML Algorithm for the Percentage of the Total Jerks

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	22.635	0.413	54.850	<0.0001
Speed Limit	-0.131	0.010	-12.484	<0.0001
MPR	2.173	0.122	17.843	<0.0001
LT Protected	3.926	0.092	42.857	<0.0001
LT Protected-Permitted	2.122	0.138	15.381	<0.0001
AIC= 12,129, Model Fit= 0.459				

As indicated, the overall percentage of jerky driving maneuvers increases by increasing the AV MPR. As depicted in *Figure 22*, the total number of jerks increases for AV MPRs of up to 80% due to the interactions between the AVs and RVs; however, this percentage starts declining from AV MPR of 90%. Moreover, the results of the descriptive statistics indicated that although the number of jerky driving maneuvers is higher for the AVs, the absolute jerk values are smaller than RVs. It should be noted that jerky driving maneuvers not necessarily result in a crash, but they could be a good representative of risky traffic situations.

For LT signal phasing, the permitted LT phasing was considered as the base scenario, and the safety of protected and protected-permitted LT signals was compared to the base condition. Hence, as presented, permitted LT signal phasing results in the lowest number of jerky driving maneuvers. It is worth noting that this does not consider the number of conflicts between left-turning vehicles and opposing through vehicles, but it discusses jerky driving maneuvers. In fact, although protected LT signal phasing makes the vehicles to come to a full stop and results in a higher number of jerks, it should enhance the overall safety by reducing the chance of left-turning and opposing through conflicts. To compare the total number of conflicts for different LT signal phasing, SSAM was used

according to section *3.6.1*, and the results are given in *Table 15*. As indicated, even though protected LT signal phasing results in the highest number of jerks, it is associated with the lowest number of SSAM conflicts.

Table 15. Effects of LT Signal Phasing on the Number of SSAM Conflicts

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
LT Signal Phasing	2	17,309,394	8,654,697	461.1	*<0.0001*
Residuals	3,536	66,376,013	18,771		
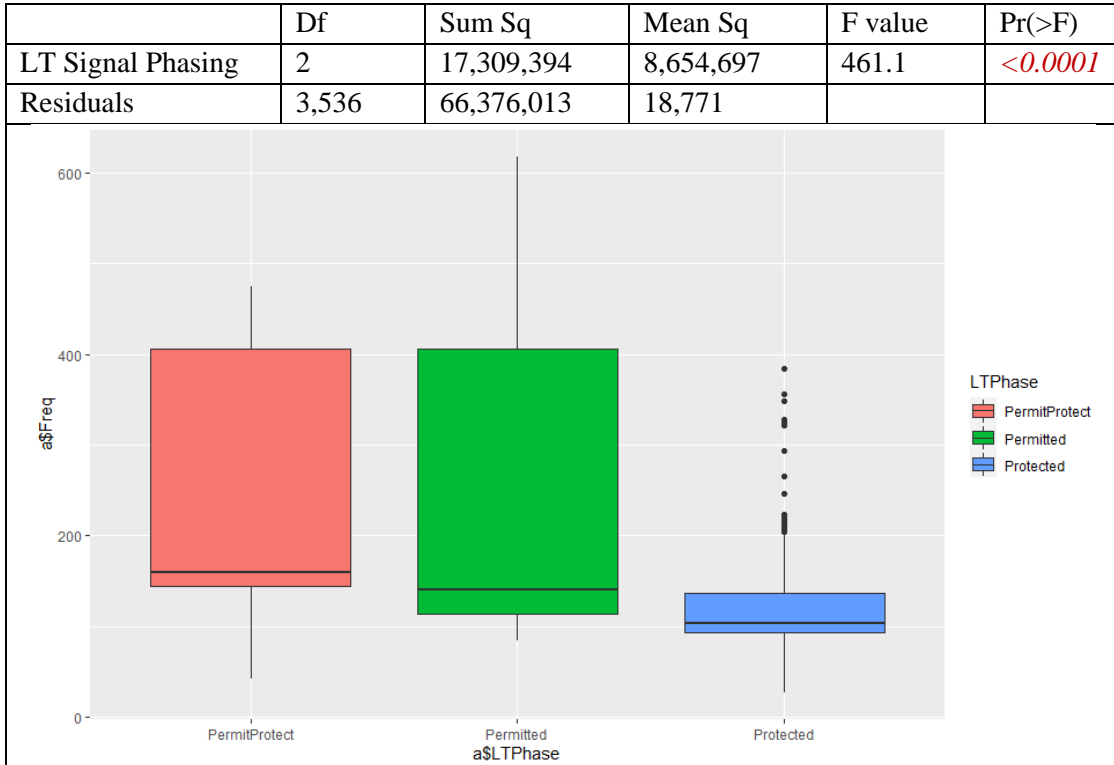
It is worth noting that the cycle length had a significant p-value in *Table 14*, but conducting a sensitivity analysis indicated that this variable does not influence the overall performance of the model significantly, and hence excluded from the model to satisfy the simplicity. Also, the Akaike Information Criterion (AIC) of the model is 12,1229, which is an in-sample prediction error estimator (*127*, *128*).

Lastly, *Equation (15)* represents the final form of the regression model, with a root mean square error (RMSE) of 2.329 for the test dataset. This equation could be used to

estimate the percentage of the total number of jerky driving maneuvers, given the study

variables.

$$Percentage \; of \; the \; Total \; Jerks$$
$$= \; 22.635 - 0.131(SL) + 2.173(MPR) + 3.926(LT_{Pro}) \qquad \textbf{(15)}$$
$$+ \; 2.122(LT_{Pro-Per})$$

Where,

SL= Speed limit (mph);

LT_{Pro}= Binary variable for protected LT phasing (value of 1 if protected LT, 0

otherwise)

$LT_{Pro-Per}$= Binary variable for protected-permitted LT phasing (value of 1 if

protected-permitted LT, 0 otherwise)

5.2.2.2. ML Model with Percentage of AV Jerks

The next ML regression model that was developed considers the percentage of AV

jerks as the dependent variable using the ML. The results of the regression model are given

in *Table 16*.

As presented, increasing the AV size and speed limit improves traffic safety by

decreasing the percentage of jerky driving maneuvers for AVs. Analyzing the lane

changing maneuvers in *Table 17* indicates that by increasing the AV size, the number of

lane-changing maneuvers decreases significantly. In other words, for the same scenarios

with only having different AV sizes, the scenario with larger AVs experiences a fewer

123

number of lane-changing maneuvers. In the scenario with the smaller AV size, vehicles

can find adequate gaps easier to make lane-changing maneuvers. Therefore, this results in

a higher number of jerky driving maneuvers since the following vehicles might require to

decelerate abruptly to keep an appropriate heading and prevent any conflict and/or crash.

Table 16. ML Algorithm for the Percentage of the AV Jerks

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.271	0.523	8.169	<0.0001
AV Size	-0.111	0.028	-3.946	<0.0001
Signal Cycle Length (min)	0.236	0.046	5.149	<0.0001
Speed Limit	-0.099	0.008	-11.919	<0.0001
MPR	21.134	0.096	219.277	<0.0001
LT Protected	2.401	0.074	32.594	<0.0001
LT Protected-Permitted	1.454	0.120	12.119	<0.0001
AIC= 10,856, Model Fit= 0.947				

It is worth mentioning that the reasons that this is not in line with the previous

model in section *5.1.1* are: a) this model only considers AV jerks while the previous model

included both AV and RV jerks, b) the previous model just considered the actual volatility

values while the current model evaluates the percentage of jerks, and lastly, c) the model

on the frequency of negative jerk values over a threshold, in the previous section, did not

provide a good fit to be reliable for implementation.

As before, protected LT signal phasing is the most dangerous LT signal phasing,

even for only AVs that are machine-driven vehicles. Increasing the cycle length also

increases the jerky maneuvers. Although higher cycle lengths provide smoother traffic

flow, it increases the LT cycles simultaneously, which results in more risky maneuvers.

Moreover, as expected, raising the AV MPR increases the frequency of the AV jerks but

meanwhile decreases the average jerk values, as indicated in *Figure 24*. Although higher

AV MPRs increase the AV jerks, that does not necessarily cause a crash as AVs are

capable of reacting to the situations abruptly. Moreover, AV jerks are not as harsh as RV

jerks, as indicated in *Table 8*.

Table 17. T-Test for the Mean Number of Lane Changing Maneuvers for

Different AV Sizes

	Df	Sum Sq	Mean Sq	F-Value	Pr(>F)
AV Size	1	3,290,900	3,290,900	23.83	*<0.0001*
Residuals	3,650	504,015,078	138,086		
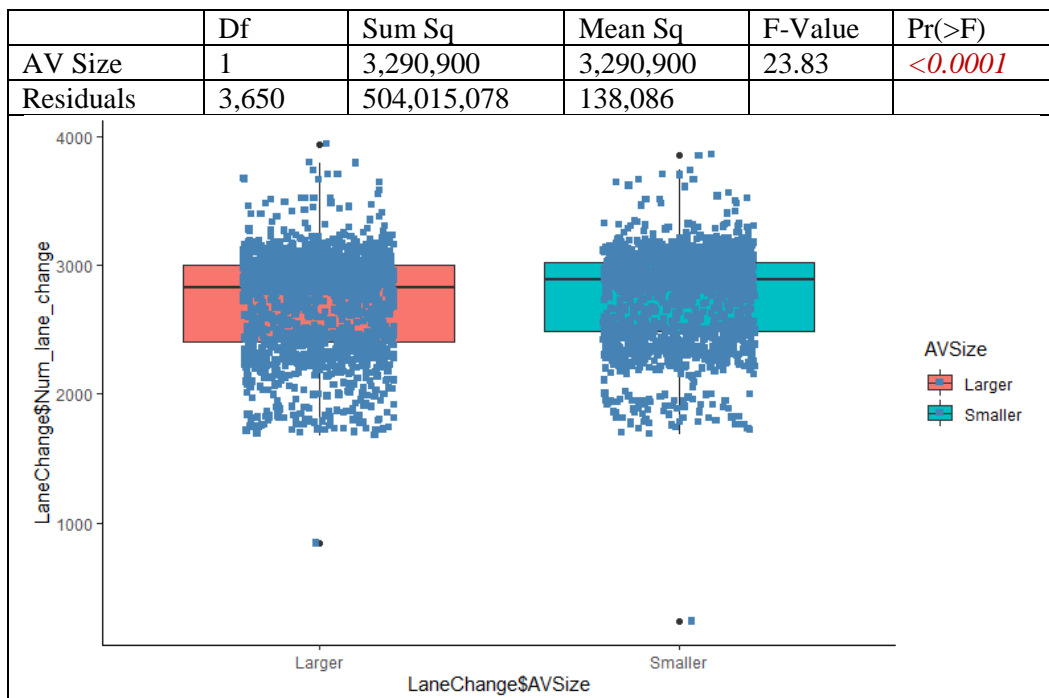
The results of testing the developed model on the test set indicated that the RMSE

is 1.714. Lastly, *Equation (16)* represents the model for estimating the percentage of AV

jerks.

$$Percentage\ of\ the\ Total\ AV\ Jerks$$

$$= 4.271 - 0.111(AV\ Size) + 0.236(L_{Cycle}) - 0.099(SL) \qquad \textbf{(16)}$$

$$+ 21.134(MPR) + 2.401(LT_{Pro}) + 1.454(LT_{Pro-Per})$$

Where,

$AV\ Size$ = Length of AVs (ft)Length of AVs (ft)

$L_{Cycle} = Cycle\ Length$ (min)

5.2.2.3. ML Model with Percentage of RV Jerks

Another ML regression model was developed to predict the percentage of RV jerks, and the results are given in *Table 18*.

Table 18. ML Algorithm for the Percentage of the RV Jerks

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	18.583	0.400	46.447	<0.0001
AV Size	0.073	0.023	3.251	0.0012
Speed Limit	-0.038	0.006	-6.007	<0.0001
MPR	-18.970	0.073	-259.716	<0.0001
LT Protected	1.556	0.055	28.248	<0.0001
LT Protected-Permitted	0.792	0.082	9.674	<0.0001
AIC= 9,328.0, Model Fit= 0.962				

The sign of the variable speed limit is consistent with the previously developed models and indicates that the percentage of RV jerks decreases by increasing the speed limit. As expected, the number of RV jerks declines by increasing the AV MPR, but the median value of the RV jerks increases, as depicted in *Figure 23*. The effect of LT signal phasing is in line with the previous models and indicates that the permitted LT phasing is

126

associated with the highest level of safety, while protected LT phasing results in the highest number of RV jerks.

Ultimately, the test dataset was used to determine how well the model could perform. The RMSE resulted from applying the developed model to the test dataset was 1.483. Therefore, this model is well representing and estimating the percentage of RV jerks. *Equation (17)* could be used to calculate the percentage of RV jerks, given the study variables.

$$Percentage\ of\ RV\ Jerks$$
$$= 18.583 + 0.073(AV\ Size) - 0.038(SL) - 18.97(MPR) \qquad \textbf{(17)}$$
$$+ 1.556(LT_{Pro}) + 0.792(LT_{Pro-Per})$$

5.3. Comparing Jerk and SSAM Outputs

As mentioned earlier, SSAM is another SSM that is implemented to assess the safety of micro-simulation scenarios using TTC. In this section, the relationship between the number of jerky driving maneuvers and the number of conflicts using SSAM will be calculated to determine if they are correlated and could be used interchangeably. The following sections consider the total number of conflicts and only rear-end conflicts vs. number of jerks, respectively.

5.3.1. Total Number of SSAM Conflicts vs. Number of Jerks

As mentioned in the methodology chapter, the total number of SSAM conflicts was calculated using the TTC value of 1.5 seconds. The total number of conflicts was then

compared to the total number of jerky driving maneuvers, the number of AV jerky driving

maneuvers, and the number of RV jerks to determine if there is any correlation between the

variables.

Pearson's correlation test was conducted in R, and the results are presented in

Table 19. As indicated, the correlation value is -36.58% for the total number of jerks,

which means that the total number of conflicts and the total number of jerks are not highly

correlated. This value is -11.45% and -34.9% for only AV jerks and only RV jerks,

respectively.

Table 19- Pearson's Correlation Coefficient Test: Total Number of SSAM Conflicts

vs. Number of Jerks

Variable 1	Variable 2	Pearson Correlation Value
Total number of jerks	Total Conflicts from SSAM	-0.3658
Number of AV Jerks	Total Conflicts from SSAM	-0.1145
Number of RV Jerks	Total Conflicts from SSAM	-0.349

In the next section, the correlation test will be conducted on the number of jerks

and only rear-end conflicts from SSAM.

5.3.2. *Number of Rear-End SSAM Conflicts vs. Number of Jerks*

As mentioned earlier, for mixed traffic environments, there is not any single value

for TTC to consider both RVs and AVs at the same time. In other words, if a TTC is

selected according to the RVs, the number of AV near-miss events would be overestimated

by increasing the AV MPR. Similarly, if the critical TTC is chosen based on the performance of AVs, the number of conflicts would be underestimated for RVs.

Hence, to overcome this issue, only rear-end conflicts were calculated and compared with the number of jerks. In fact, for the rear-end conflicts based on the current traffic law, the vehicle which hits from the back is at fault. Therefore, if the faulty vehicle was an RV, 1.5 seconds was used as the TTC; otherwise, 1.0 second was implemented.

The extracted number of rear-end near-miss events from SSAM was compared to the number of jerk events using Pearson's correlation test. *Table 20* summarizes the results of the Pearson's correlation coefficient tests. As indicated, compared to the total number of SSAM conflicts, the correlation values have been improved for AV jerks and RV jerks; however, the performance of Pearson's correlation test for the total number of jerks vs. rear-end conflicts deteriorates.

Table 20- Pearson's Correlation Coefficient Test: Number of SSAM Rear-end Conflicts vs. Number of Jerks

Variable 1	Variable 2	Pearson Correlation Value
Total number of jerks	Rear-end Conflicts from SSAM	-0.1411
Number of AV Jerks	Rear-end Conflicts from SSAM	-0.3243
Number of RV Jerks	Rear-end Conflicts from SSAM	-0.4295

As indicated, the number of jerks and SSAM total conflicts or rear-end conflicts were not highly correlated to be used interchangeably. However, this does not imply that one is superior over the other. Each SSM has advantages and disadvantages and could be selected for implementation accordingly. For instance, although TTC (through SSAM) is objective and easy to implement for simulation outputs, it fails to distinguish responders

from initiators, and it does not serve as a proxy for all types of crashes (*129*). On the other hand, although implementing jerks for simulation outputs is complicated and time-consuming, and there is no software program available to calculate these events automatically, they are known as a promising SSM to detect crash-prone locations accurately (*13*, *38*, *43*, *98*, *101*, *102*). Hence, according to the objective of a study and the available resources, different SSMs could be implemented.

5.4. Chapter Summary

To conclude, this chapter has comprehensively evaluated the safety of a signalized intersection considering various SSMs and using different non-infrastructure variables in mixed traffic environments. Traffic safety was analyzed from different aspects, including longitudinal driving volatility measures, lateral driving volatility measures, and jerks.

Considering both longitudinal and lateral driving volatility measures, a total of 30 models were developed to explore how the study variables affect each volatility measure, and hence traffic safety. The results indicated that longitudinal driving volatility measures are more sensitive to the changes of the study variables. Overall, depending on the purpose of the study, each longitudinal and lateral model could be implemented to evaluate traffic safety.

For the ML models using jerks, three dependent variables were considered separately, including the percentage of the total jerks, the percentage of AV jerks, and the percentage of RV jerks. For all the models, MPR, speed limit, and LT signal phasing were significant variables, while in the AV model and RV model, other variables were also significantly affecting the performance of the models. Increasing the speed limit and

providing permitted LT phasing resulted in the lowest percentage of jerks for all the models. It is notable that although protected LT phasing is associated with the highest number of jerky driving maneuvers, it results in the lowest number of SSAM conflicts. Therefore, LT-opposing should be evaluated separately and thoroughly using the number of conflicts, especially at low AV MPRs, where the majority of the vehicles are RVs and controlled by human judgments.

Ultimately, the number of jerks were compared with the total and only rear-end SSAM conflicts. The results indicated that there is no correlation between the variables, and they could not be used interchangeably.

The next and the last chapter provides a summary of the work and concludes the dissertation.

CHAPTER VI

SUMMARY AND CONCLUSION

This chapter wraps up the dissertation by providing a summary of the dissertation, conclusion, and further research.

6.1. Summary of Work

Transportation networks, as an integral part of urban areas, have been experiencing several transformations and growths over time (*1*). The growth in transportation networks results in significant traffic congestion, air pollution, fuel consumption, traffic delay, and, most importantly, deterioration in traffic safety (*2*). Based on the WHO, roadway crashes contributed to nearly 1.35 million fatalities worldwide in 2018 (*5*). The U.S. also reported 400,000 fatalities in 2018 (*6*).

The roadway crashes, fatalities, and injuries make public and private agencies to explore various solutions to enhance traffic safety. However, conventional solutions (including widening the roadways, widening shoulders, etc.) are not capable of fully addressing the concerns anymore, due to space and funding limitations. Hence, transportation agencies have been focusing on innovative technologies instead. AVs are one of the recent technologies that are expected to solve traffic safety concerns significantly as they are eliminating human factors from the driving task, which are contributing to 94% of the crashes (*2*, *8*, *9*). However, it should be noted that for AVs to reach their full capacity, upgrading the current infrastructure is required. But, upgrading the entire infrastructure at once is time-consuming and costly. Therefore, this research

aimed to determine how changing the non-infrastructure variables can influence traffic safety in mixed traffic environments.

Meanwhile, intersections proportionally experience more crashes than segments, and the signalized ones result in the highest number of intersection-related fatality crashes. Therefore, this dissertation focused on signalized intersections. Thus, speed limit, signal cycle length, LT signal phasing, and AV size were considered as non-infrastructure variables at signalized intersections, and their effects on traffic safety were analyzed under various AV MPRs. The results could be used to enhance safety by adjusting the non-infrastructure variables without imposing any extra costs, especially at low AV MPRs.

A full factorial design was implemented to combine all the levels of all the study variables to develop the simulation scenarios. Eventually, 3,850 simulation runs were set and run. The output data were then extracted, merged, and cleaned for further statistical analyses.

Traffic safety analyses of the simulation runs and the effects of the study variables were investigated from various aspects:

1) Evaluating various longitudinal and lateral driving volatility measures, including S.D., C.V., and percentage of values over a threshold for speed, acceleration, deceleration, positive jerk, and negative jerk:

   - For the first time, this study extensively explored both lateral and longitudinal driving volatility measures to evaluate safety. The lateral and longitudinal driving volatility measures were analyzed to determine how the non-infrastructure study variables influence each of them.

- The results indicated that the longitudinal driving volatility measures are more sensitive than lateral movements. By increasing the MPR, some of the longitudinal measures decreased while the others increased. However, the majority of the lateral driving volatility measures improved by increasing the AV MPR.

- Increasing the AV size increased the majority of the longitudinal driving volatility measures while decreased most of the lateral volatility measures.

- Higher signal cycle length improved most of the lateral volatility measures.

- Depending on the driving volatility measures, the speed limit and LT signal phasing showed different effects on both lateral and longitudinal measures.

- Since 30 different longitudinal and lateral driving volatility measures were considered, the results cannot be summarized thoroughly. Hence, according to the objective of a study, an appropriate model could be implemented.

2) Developing safety models using ML and considering the percentage of jerky driving maneuvers for AVs, RVs, and both AVs and RVs as the dependent variables:

- The results of the analyses using the ML algorithm indicated that in mixed traffic environments, the percentage of the total number of jerks could be reduced by increasing the speed limit. In fact, by increasing the speed limit, more vehicles can pass through the intersection without experiencing red signals, which requires them to come to a full stop and sometimes brake abruptly. The effects of increasing the speed limit on the AV jerks and RV jerk presented the same results.

- All the models indicated that protected LT signal phasing is associated with the highest percentage of the total jerks, AV jerks, and RV jerks because the vehicles require to come to a full stop more frequently. But meanwhile, considering the SSAM conflicts indicated that the total number of conflicts decreases while providing protected LT signal phasing. In fact, protected LT signal phasing avoids/significantly reduces the left-turning and opposing through conflicts.

- Increasing the size of AVs also illustrated that the larger the size, the fewer the number of AV jerks. The reason is that larger AVs have fewer opportunities to conduct lane-changing maneuvers because of the difficulty in finding gaps with adequate size.

Eventually, the number of jerks were compared to the SSAM conflicts through conducting multiple Pearson's correlation coefficient tests. Both total SSAM conflicts and only rear-end conflicts were examined. The results indicated that the jerks and SSAM total/rear-end conflicts are not correlated and cannot be used interchangeably. However, this does not imply that one outperforms the other. Depending on the purpose of a study and the availability of the resources, an appropriate SSM could be implemented.

6.2. Conclusions

In summary, it is not practical to upgrade the entire infrastructure at once before implementing AVs, especially at low AV MPRs. Therefore, this dissertation evaluated the effects of altering the non-infrastructure variables on the safety of mixed traffic environments at a signalized intersection.

Quantifying the safety of non-infrastructure variables was conducted on various surrogate safety measures, including different longitudinal and lateral driving volatility measures and the percentage of the jerks.

Each longitudinal and lateral driving volatility measure was influenced by a different combination of the non-infrastructure variables. Hence, each model could be evaluated and implemented individually according to the purpose of a study. For instance, the signal cycle length had a negative relationship with the majority of longitudinal and lateral driving volatility measure, most of the longitudinal driving volatility measures increased by increasing the speed limit, AV size had a negative relationship with the majority of the longitudinal measures and positive relationship with most of the lateral measures, and the effects of LT signal phasing were different on each model.

By only considering the percentage of jerks, the results indicated that the percentage of the total jerk increased by increasing the AV MPR. It is notable that although the percentage of jerky driving maneuvers increased by raising the AV MPR, the absolute jerk values decreased. In addition, providing permitted LT signal phasing could reduce the percentage of total jerks, the percentage of AV jerks, and the percentage of RV jerks, but meanwhile increases the total number of SSAM conflicts. Therefore, LT and opposing through conflicts should be evaluated in further detail. Although increasing the AV size reduced the percentage of AV jerks and increased the percentage of RV jerks, the benefit outweighed the loss (by having a larger coefficient in the AV model). Therefore, by increasing the AV size, the number of AV jerks could be reduced significantly.

Moreover, increasing the approach speed limit consistently reduced the percentage of the total jerks, percentage of AV jerks, and percentage of RV jerks. It is notable that the

speed limit is a surrogate for the operating speed of the network, the geometry of the

intersection legs, and the traffic congestion level. Hence, where the geometry of the legs of

an intersection allows, the approach speed limits could be increased. However, the impacts

of increasing the speed limit on pedestrians and bicyclists should not be neglected. In fact,

although increasing the speed limit significantly reduces the various percentage of jerks, it

could result in more severe pedestrian/bicyclists-involved crashes. Therefore, where

pedestrians and bicyclists are actively involved, the approach speed limits should be

increased cautiously (if not, putting a ceiling on

the limit based on the characteristics of the land use or policy used for setting speed limits

by the transportation agency). Moreover, by increasing the AV MPR, the effects of

increasing the speed limit become less perceptible since AVs follow the posted speed

limits and result in a more uniform speed distribution.

In general, when it is impractical to upgrade the infrastructure to safely

accommodate AVs, especially at the first deployment of AVs, the best practice is to adjust

the speed limit. To reach a higher level of safety, the approach speed limits could be

increased according to the geometry of the roadways and activities of

pedestrians/bicyclists. Although the positive effect of increasing the approach speed limits

is more visible in an RV dominant environment, it could result in a more vulnerable

environment for pedestrians and bicyclists as RVs have a larger reaction time and tend to

drive over the speed limits. In addition, permitted LT signal phasing results in the lowest

percentage of jerky driving maneuvers and the highest number of SSAM conflicts.

Therefore, for low AV MPRs, permitted LT signal phasing is not the safest choice.

Similar to low AV MPRs, for the medium AV MPRs, the best practice is to adjust the approach speed limits and LT signal phasing. Moreover, to reach a safer environment, the signal cycle length could be modified according to the type of LT signal phasing to control the jerky driving maneuvers of AVs.

When reaching a fully AV environment, if the infrastructure is still not designed properly to safely accommodate AVs, the safety could be enhanced by adjusting the signal cycle length, speed limit, LT signal phasing, and AV size. Decreasing the cycle length, increasing the speed limit (with constraints), and providing permitted LT signal phasing could improve traffic safety significantly at high AV MPRs. It is worth noting that in a fully AV environment, despite the RV environments, the permitted LT signal phasing could be the safest option since AVs are machine-driven vehicles and capable of detecting their environments with minimal, if not without any, errors as well as reacting to the situations abruptly. Moreover, the benefit of increasing the speed limit at high AV MPRs is not as noticeable as for the low MPRs; but, the negative impacts on the safety of pedestrians and bicyclists are lower as AVs analyze and react to the situations abruptly. Lastly, if manufacturers provide AVs in larger sizes, the number of jerky driving maneuvers will be reduced significantly at high AV MPRs.

6.3. Recommendations for Further Research

Traffic safety of mixed environments has not been explored extensively, and yet there are many aspects that could be examined and analyzed before employing AVs.

This dissertation defined various levels/categories for the studied non-infrastructure variables (speed limit, cycle length, and AV size) to be simulated and evaluated. For future

research, more levels could be introduced and assessed. Moreover, other easy-to-implement infrastructure variables, such as the density of driveways nearby a signalized intersection, could be examined to determine how the safety of mixed traffic environments would be affected.

In each simulation scenario of the dissertation, all the AVs had the same size. For future research, the safety effects of various combinations of different AV sizes could be explored.

This dissertation increased or decreased the current and actual speed limits of the intersection legs with the same rate for all the legs. However, future studies can evaluate the safety effects of uniform speed limits along all the legs.

While analyzing the safety of the simulation scenarios could be conducted using SSMs, implementing the real-world crash data is worth investigating. Where available, the real-world crash data could be compared to the results of jerks and also SSAM conflicts to determine how they are related. To this aim, other models, such as negative binomial and Poisson models, could also be implemented and assessed.

It is worth investigating the turning maneuvers at signalized intersections in more detail. SSAM could be implemented to find and evaluate turning conflicts since this software program is capable of providing the number of rear-end, lane-changing, and crossing conflicts. However, as mentioned earlier, the TTC should be selected cautiously to avoid overestimating/ underestimating the number of conflicts.

Lastly, this dissertation assessed 34 different SSMs to evaluate the safety of mixed traffic environments from various aspects. Although the majority of the SSMs are hard to implement on the simulation outputs, the investigation is worthy.

REFERENCES

1. Talebpour, A., and H. S. Mahmassani. Influence of Connected and Autonomous Vehicles on Traffic Flow Stability and Throughput. *Transportation Research Part C: Emerging Technologies*, Vol. 71, 2016, pp. 143–163. https://doi.org/10.1016/j.trc.2016.07.007.

2. Gora, P., and I. Rüb. Traffic Models for Self-Driving Connected Cars. No. 14, 2016, pp. 2207–2216.

3. American Society of Civil Engineers (ASCE). *Infrastructure Report Card: A Comprehensive Assessment of America's Infrastructure*. 2017.

4. Schrank, D., B. Eisele, T. Lomax, and J. Bak. 2015 Urban Mobility Scorecard. 2015.

5. World Health Organization (WHO). *Global Status Report on Road Safety 2018*. 2018.

6. National Safety Council. Vehicle Deaths Estimated at 40,000 for Third Straight Year. National Safety Council. https://www.nsc.org/road-safety/safety-topics/fatality-estimates. Accessed Jul. 25, 2019.

7. (NHTSA), N. H. T. S. A. *2017 Fatal Motor Vehicle Crashes: Overview*. 2018.

8. Morando, M. M., Q. Tian, L. T. Truong, and H. L. Vu. Studying the Safety Impact of Autonomous Vehicles Using Simulation-Based Surrogate Safety Measures. *Journal of Advanced Transportation*, 2018. https://doi.org/10.1155/2018/6135183.

9. Wayomo. *On the Road to Fully Self-Driving*. 2018.

10. Fagnant, D. J., and K. Kockelman. Preparing a Nation for Autonomous Vehicles:

Opportunities, Barriers and Policy Recommendations. *Transportation Research Part A: Policy and Practice*, Vol. 77, 2015, pp. 167–181. https://doi.org/10.1016/j.tra.2015.04.003.

11. Gouy, M. *Behavioural Adaptation of Drivers of Unequipped Vehicles to Short Time Headways Observed in a Vehicle Platoon*. 2013.

12. Bansal, P., and K. M. Kockelman. Forecasting Americans' Long-Term Adoption of Connected and Autonomous Vehicle Technologies. *Transportation Research Part A: Policy and Practice*, Vol. 95, 2017, pp. 49–63. https://doi.org/10.1016/j.tra.2016.10.013.

13. Autonomous Car. *Wikipedia*. https://en.wikipedia.org/wiki/Autonomous_car. Accessed Apr. 8, 2018.

14. Favarò, F. M., N. Nader, S. O. Eurich, M. Tripp, and N. Varadaraju. Examining Accident Reports Involving Autonomous Vehicles in California. *PLOS ONE*, Vol. 12, No. 9, 2017, p. e0184952. https://doi.org/10.1371/journal.pone.0184952.

15. Truong, L. T., C. De Gruyter, G. Currie, and A. Delbosc. Estimating the Trip Generation Impacts of Autonomous Vehicles on Car Travel in Victoria, Australia. *Transportation*, Vol. 44, No. 6, 2017, pp. 1279–1292. https://doi.org/10.1007/s11116-017-9802-2.

16. Hoogendoorn, R., B. van Arerm, and S. Hoogendoom. Automated Driving, Traffic Flow Efficiency, and Human Factors. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2422, No. 1, 2014, pp. 113–120. https://doi.org/10.3141/2422-13.

17. Milakis, D., B. van Arem, and B. van Wee. Policy and Society Related Implications

of Automated Driving: A Review of Literature and Directions for Future Research. *Journal of Intelligent Transportation Systems*, Vol. 21, No. 4, 2017, pp. 324–348. https://doi.org/10.1080/15472450.2017.1291351.

18. Highway Traffic Safety Administration, N., and U. Department of Transportation. *Stats Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey.* 2015.

19. Dhar, V. Equity, Safety, and Privacy in the Autonomous Vehicle Era. *Computer*, Vol. 49, No. 11, 2016, pp. 80–83. https://doi.org/10.1109/MC.2016.326.

20. Stopping the Driverless Car from Committing a Hit and Run. https://www.girardikeese.com/blog/2017/10/stopping-the-driverless-car-from-committing-a-hit-and-run.shtml. Accessed Nov. 9, 2018.

21. Aria, E., J. Olstam, and C. Schwietering. Investigation of Automated Vehicle Effects on Driver's Behavior and Traffic Performance Enhancing Highway Performance Selection and Peer-Review under Responsibility of the Scientific Programme Committee of ISEHP 2016. *Transportation Research Procedia*, Vol. 15, 2016, pp. 761–770. https://doi.org/10.1016/j.trpro.2016.06.063.

22. Papadoulis, A., M. Quddus, and M. Imprialou. Evaluating the Safety Impact of Connected and Autonomous Vehicles on Motorways. *Accident Analysis and Prevention*, Vol. 124, 2019, pp. 12–22. https://doi.org/10.1016/j.aap.2018.12.019.

23. Akamatsu, M., Y. Sakaguchi, and M. Okuwa. Modeling of Driving Behavior When Approaching an Intersection Based on Measured Behavioral Data on An Actual Road. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 47, No. 16, 2003, pp. 1895–1899.

https://doi.org/10.1177/154193120304701613.

24.   Zimmerman, K., and J. A. Bonneson. Intersection Safety at High-Speed Signalized
       Intersections: Number of Vehicles in Dilemma Zone as Potential Measure.
       Transportation Research Record: Journal of the Transportation Research Board,
       Vol. 1897, No. 1, 2004, pp. 126–133. https://doi.org/10.3141/1897-16.

25.   Khoda Bakhshi, A., and M. M. Ahmed. Utilizing Black-Box Visualization Tools to
       Interpret Non-Parametric Real-Time Risk Assessment Models. *Transportmetrica A:*
       Transport Science, 2020, pp. 1–27.
       https://doi.org/10.1080/23249935.2020.1810169.

26.   Khoda Bakhshi, A., and M. M. Ahmed. Practical Advantage of Crossed Random
       Intercepts under Bayesian Hierarchical Modeling to Tackle Unobserved
       Heterogeneity in Clustering Critical versus Non-Critical Crashes. *Accident Analysis*
       and Prevention, 2020.

27.   Oliver, N., K. Potočnik, and T. Calvard. To Make Self-Driving Cars Safe, We Also
       Need Better Roads and Infrastructure. *Harvard Business Review*.

28.   Axelrod, C. W. Integrating In-Vehicle, Vehicle-To-Vehicle, and Intelligent
       Roadway Systems. *International Journal of Design and Nature and Ecodynamics*,
       Vol. 13, No. 1, 2018, pp. 23–38. https://doi.org/10.2495/DNE-V13-N1-23-38.

29.   Steyn, W. J. v. M., and J. W. Maina. Guidelines for the Use of Accelerated
       Pavement Testing Data in Autonomous Vehicle Infrastructure Research. *Journal of*
       Traffic and Transportation Engineering (English Edition). 3. Volume 6, 273–281.

30.   Tokody, D., A. Albini, L. Ady, Z. Rajnai, and F. Pongrácz. Safety and Security
       Through the Design of Autonomous Intelligent Vehicle Systems and Intelligent

Infrastructure in the Smart City. *Interdisciplinary Description of Complex Systems*, Vol. 16, No. A, 2018, pp. 384–396. https://doi.org/10.7906/indecs.16.3.11.

31. Antonucci, N. D., K. K. Hardy, K. L. Slack, R. Pfefer, and T. R. Neuman. *NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan. Volume 12: A Guide for Reducing Collisions at Signalized Intersections*. 2004.

32. Polders, E., S. Daniels, E. Hermans, T. Brijs, and G. Wets. Crash Patterns at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2514, No. 1, 2015, pp. 105–116. https://doi.org/10.3141/2514-12.

33. Guo, F., X. Wang, and M. A. Abdel-Aty. Modeling Signalized Intersection Safety with Corridor-Level Spatial Correlations. *Accident Analysis and Prevention*, Vol. 42, No. 1, 2010, pp. 84–92. https://doi.org/10.1016/j.aap.2009.07.005.

34. Favarò, F. M., N. Nader, S. O. Eurich, M. Tripp, and N. Varadaraju. Examining Accident Reports Involving Autonomous Vehicles in California. *PLOS ONE*, Vol. 12, No. 9, 2017, p. e0184952. https://doi.org/10.1371/journal.pone.0184952.

35. U.S. Department of Transportation-NHTSA. Fatal Motor Vehicle Crashes 2017: Overview. *Dot Hs 812603*, No. October, 2018, pp. 1–9.

36. Highway Traffic Safety Administration, N., and U. Department of Transportation. *Research Note: 2017 Fatal Motor Vehicle Crashes: Overview*. 2017.

37. Mousavi, S. M. *Identifying High Crash Risk Roadways through Jerk-Cluster Analysis*. 2015.

38. Mousavi, S. M., H. Marzoughi, S. A. Parr, B. Wolshon, A. Pande, and A. Professor.

A Mixed Crash Frequency Estimation Model for Interrupted Flow Segments. 2019.

39.   Polders, E., S. Daniels, E. Hermans, T. Brijs, and G. Wets. Crash Patterns at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research*, Vol. 2514, 2015, pp. 105–116. https://doi.org/10.3141/2514-12.

40.   Mokhtarimousavi, S., J. C. Anderson, A. Azizinamini, and M. Hadi. Factors Affecting Injury Severity in Vehicle-Pedestrian Crashes: A Day-of-Week Analysis Using Random Parameter Ordered Response Models and Artificial Neural Networks. *International Journal of Transportation Science and Technology*, Vol. 9, No. 2, 2020, pp. 100–115. https://doi.org/10.1016/j.ijtst.2020.01.001.

41.   Dixon, K. K., K. Fitzpatrick, S. M. Mousavi, I. B. Potts, D. W. Harwood, J. Grotheer, and J. Ronchetto. *Unsignalized Full Median Openings in Close Proximity to Signalized Intersections*. Washington D.C., 2020.

42.   Eftekharzadeh, S. F., and A. Khodabakhshi. *Safety Evaluation of Highway Geometric Design Criteria in Horizontal Curves at Downgrades*. International Journal of Civil Engineering, 2014.

43.   Mousavi, S. M., Z. Zhang, S. A. Parr, A. Pande, and B. Wolshon. Identifying High Crash Risk Highway Segments Using Jerk-Cluster Analysis. 2019.

44.   Eskandarian, A. *Handbook of Intelligent Vehicles*. Springer, 2012.

45.   Vision Zero Communities Map. https://visionzeronetwork.org/wp-content/uploads/2019/04/vision-zero-cities-feb-2019.pdf. Accessed Jul. 5, 2020.

46.   Vision Zero Communities Map. https://visionzeronetwork.org/wp-content/uploads/2019/04/vision-zero-cities-feb-2019.pdf.

47.   Shladover, S. E. Connected and Automated Vehicle Systems: Introduction and Overview. *Journal of Intelligent Transportation Systems*, Vol. 22, No. 3, 2018, pp. 190–200. https://doi.org/10.1080/15472450.2017.1336053.

48.   Cui, J., L. S. Liew, G. Sabaliauskaite, and F. Zhou. A Review on Safety Failures, Security Attacks, and Available Countermeasures for Autonomous Vehicles. *Ad Hoc Networks*, Vol. 90, 2019, p. 101823. https://doi.org/10.1016/j.adhoc.2018.12.006.

49.   Malek, F. A. Autonomous Vehicles: Safety, Sustainability, and Fuel Efficiency. *i-manager's Journal on Future Engineering & Technology*, 2017.

50.   Bhavsar, P., P. Das, M. Paugh, K. Dey, and M. Chowdhury. Risk Analysis of Autonomous Vehicles in Mixed Traffic Streams. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2625, No. 1, 2017, pp. 51–61. https://doi.org/10.3141/2625-06.

51.   Harper, C. D., C. T. Hendrickson, and C. Samaras. Cost and Benefit Estimates of Partially-Automated Vehicle Collision Avoidance Technologies. *Accident Analysis and Prevention*, Vol. 95, 2016, pp. 104–115. https://doi.org/10.1016/j.aap.2016.06.017.

52.   McMurry, T. L., G. S. Poplin, G. Shaw, and M. B. Panzer. Crash Safety Concerns for Out-of-Position Occupant Postures: A Look toward Safety in Highly Automated Vehicles. *Traffic Injury Prevention*, Vol. 19, No. 6, 2018, pp. 582–587. https://doi.org/10.1080/15389588.2018.1458306.

53.   Schoettle, B., and M. Sivak. *A PRELIMINARY ANALYSIS OF REAL-WORLD CRASHES INVOLVING SELF-DRIVING VEHICLES*. 2015.

54. Bu, F., H. S. Tan, and J. Huang. Design and Field Testing of a Cooperative Adaptive Cruise Control System. 2010.

55. Maryam Mousavi, S., D. Lord, B. Dadashova, and S. Reza Mousavi. Can Autonomous Vehicles Enhance Traffic Safety at Unsignalized Intersections? 2020.

56. Kockelman, K., P. Avery, P. Bansal, S. D. Boyles, P. Bujanovic, T. Choudhary, L. Clements, G. Domnenko, D. Fagnant, J. Helsel, R. Hutchinson, M. Levin, J. Li, T. Li, L. Loftus-Otway, A. Nichols, M. Simoni, and D. Stewart. *Implications of Connected and Automated Vehicles on the Safety and Operations of Roadway Networks: A Final Report*. 2016.

57. Arvin, R., M. Kamrani, A. Khattak, and J. Rios-Torres. Safety Impacts of Automated Vehicles in Mixed Traffic. 2018.

58. Arvin, R., A. Khattak, and J. Rios Torres. Evaluating Safety with Automated Vehicles at Signalized Intersections: Application of Adaptive Cruise Control in Mixed Traffic. 2019.

59. Virdi, N., H. Grzybowska, S. T. Waller, and V. Dixit. A Safety Assessment of Mixed Fleets with Connected and Autonomous Vehicles Using the Surrogate Safety Assessment Module. *Accident Analysis and Prevention*, Vol. 131, 2019, pp. 95–111. https://doi.org/10.1016/j.aap.2019.06.001.

60. Jeong, E., C. Oh, and S. Lee. Is Vehicle Automation Enough to Prevent Crashes? Role of Traffic Operations in Automated Driving Environments for Traffic Safety. *Accident Analysis and Prevention*, Vol. 104, 2017, pp. 115–124. https://doi.org/10.1016/j.aap.2017.05.002.

61. Karaaslan, U., P. Varaiya, and J. Walrand. Two Proposals to Improve Freeway

Traffic Flow. 1991.

62. VanderWerf, Joel; Shladover, Steven; Miller, M. A. *Conceptual Development and Performance Assessment for the Deployment Staging of Advanced Vehicle Control and Safety Systems*. 2044.

63. Mahmassani, H. S. Autonomous Vehicles and Connected Vehicle Systems: Flow and Operations Considerations. *Transportation Science*, 2016, pp. 1140–1162.

64. Mahdinia, I., R. Arvin, A. J. Khattak, and A. Ghiasi. Safety, Energy, and Emissions Impacts of Adaptive Cruise Control and Cooperative Adaptive Cruise Control. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2674, No. 6, 2020, pp. 253–267. https://doi.org/10.1177/0361198120918572.

65. Ye, L., and T. Yamamoto. Impact of Dedicated Lanes for Connected and Autonomous Vehicle on Traffic Flow Throughput. *Physica A: Statistical Mechanics and its Applications*, Vol. 512, 2018, pp. 588–597. https://doi.org/10.1016/j.physa.2018.08.083.

66. Vander Laan, Z., and K. F. Sadabadi. Operational Performance of a Congested Corridor with Lanes Dedicated to Autonomous Vehicle Traffic. *International Journal of Transportation Science and Technology*, Vol. 6, No. 1, 2017, pp. 42–52. https://doi.org/10.1016/j.ijtst.2017.05.006.

67. Chen, D., S. Ahn, M. Chitturi, and D. A. Noyce. Towards Vehicle Automation: Roadway Capacity Formulation for Traffic Mixed with Regular and Automated Vehicles. *Transportation Research Part B: Methodological*, Vol. 100, 2017, pp. 196–221. https://doi.org/10.1016/j.trb.2017.01.017.

68. Nilsson, J., M. Brannstrom, E. Coelingh, and J. Fredriksson. Lane Change

Maneuvers for Automated Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 5, 2017, pp. 1087–1096. https://doi.org/10.1109/TITS.2016.2597966.

69.   Sun, W., J. Zheng, and H. X. Liu. A Capacity Maximization Scheme for Intersection Management with Automated Vehicles. No. 23, 2017, pp. 121–136.

70.   Yang, K., S. I. Guler, and M. Menendez. Isolated Intersection Control for Various Levels of Vehicle Technology: Conventional, Connected, and Automated Vehicles. *Transportation Research Part C: Emerging Technologies*, Vol. 72, 2016, pp. 109–129. https://doi.org/10.1016/j.trc.2016.08.009.

71.   Mousavi, S. M., O. A. Osman, and D. Lord. Impact of Urban Arterial Traffic LOS on the Vehicle Density of Different Lanes of the Arterial in Proximity of an Unsignalized Intersection for Autonomous Vehicle vs. Conventional Vehicle Environments. 2019.

72.   Letter, C., and L. Elefteriadou. Efficient Control of Fully Automated Connected Vehicles at Freeway Merge Segments. *Transportation Research Part C: Emerging Technologies*, Vol. 80, 2017, pp. 190–205. https://doi.org/10.1016/j.trc.2017.04.015.

73.   Talebpour, A., H. S. Mahmassani, and A. Elfar. Investigating the Effects of Reserved Lanes for Autonomous Vehicles on Congestion and Travel Time Reliability. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2622, No. 1, 2017, pp. 1–12. https://doi.org/10.3141/2622-01.

74.   Federal Highway Administration (FHWA). Permissive/Protected Left Turn Phasing - Safety | Federal Highway Administration. https://safety.fhwa.dot.gov/intersection/conventional/signalized/case_studies/fhwasa

09015/. Accessed Aug. 7, 2020.

75.   Antonucci, Nicholas D.; Hardy, Kelly K.; Slack, Kevin L.; Pfefer, Ronald; Neuman,
        T. R. *NCHRP 500: A Guide for Reducing Collisions at Signalized Intersections*.
        Transportation Research Board, 2004.

76.   Amiridis, K.; Stamatiadis, N., Kirk, A. Left-Turn Phasing Decisions Utilizing
        Simulated Traffic Conflicts and Historical Crashes. 71–82.
        http://www.atsinternationaljournal.com/index.php/2018-issues/special-issue-2018-
        vol1/979-left-turn-phasing-decisions-utilizing-simulated-traffic-conflicts-and-
        historical-crashes. Accessed Aug. 7, 2020.

77.   Federal Highway Administration (FHWA). *Manual on Uniform Traffic Control
        Devices (MUTCD)*. U.S. Department of Transportation, 2009.

78.   Chen, L., C. Chen, and R. Ewing. Left-Turn Phase: Permissive, Protected, or Both?
        A Quasi-Experimental Design in New York City. *Accident Analysis and Prevention*,
        Vol. 76, 2015, pp. 102–109. https://doi.org/10.1016/j.aap.2014.12.019.

79.   Elvik;, R., and T. Vaa; *The Handbook of Road Safety Measures: Second Edition*.
        Elsevier, 2004.

80.   De Pauw, E., S. Daniels, M. Thierie, and T. Brijs. Safety Effects of Reducing the
        Speed Limit from 90 Km/h to 70 Km/H. *Accident Analysis and Prevention*, Vol. 62,
        2014, pp. 426–431. https://doi.org/10.1016/j.aap.2013.05.003.

81.   McCarthy, P. Effect of Speed Limits on Speed Distributions and Highway Safety: A
        Survey of Recent Literature. *Transport Reviews*, Vol. 21, No. 1, 2001, pp. 31–50.
        https://doi.org/10.1080/014416400750059275.

82.   Shinar, D. A Controversial Topic and an Elusive Relationship. 1998.

83. Joksch, H. C. Velocity Change and Fatality Risk in a Crash. *Accident Analysis & Prevention*, Vol. 25, 1993.

84. Islam, M. T., and K. El-Basyouny. Full Bayesian Evaluation of the Safety Effects of Reducing the Posted Speed Limit in Urban Residential Area. *Accident Analysis and Prevention*, Vol. 80, 2015, pp. 18–25. https://doi.org/10.1016/j.aap.2015.02.026.

85. Subramanian, R. *Passenger Vehicle Occupant Fatality Rates by Type and Size of Vehicle*.

86. Kahane, C. J. *Relationships between Vehicle Size and Fatality Risk in Model Year 1985-93 Passenger Cars and Light Trucks | National Highway Traffic Safety Administration (NHTSA)*. Washington D.C., 1997.

87. Li, X. The Symmetric Intersection Design and Traffic Control Optimization. *Transportation Research Part C: Emerging Technologies*, Vol. 92, 2018, pp. 176–190. https://doi.org/10.1016/j.trc.2018.04.023.

88. Tang, K., K. Dong, and E. Chung. Queue Discharge Patterns at Signalized Intersections with Green Signal Countdown Device and Long Cycle Length. *Journal of Advanced Transportation*, Vol. 50, No. 8, 2016, pp. 2100–2115. https://doi.org/10.1002/atr.1448.

89. Zakariya, A. Y., and S. I. Rabia. Estimating the Minimum Delay Optimal Cycle Length Based on a Time-Dependent Delay Formula. *Alexandria Engineering Journal*, Vol. 55, 2016, pp. 2509–2514. https://doi.org/10.1016/j.aej.2016.07.029.

90. Kamrani, M., R. Arvin, and A. J. Khattak. Extracting Useful Information from Basic Safety Message Data: An Empirical Study of Driving Volatility Measures and Crash Frequency at Intersections. *Transportation Research Record: Journal of the*

Transportation Research Board, Vol. 2672, No. 38, 2018, pp. 290–301. https://doi.org/10.1177/0361198118773869.

91. Wali, B., A. J. Khattak, H. Bozdogan, and M. Kamrani. How Is Driving Volatility Related to Intersection Safety? A Bayesian Heterogeneity-Based Analysis of Instrumented Vehicles Data. *Transportation Research Part C: Emerging Technologies*, Vol. 92, 2018, pp. 504–524. https://doi.org/10.1016/j.trc.2018.05.017.

92. Wali, B., A. J. Khattak, and T. Karnowski. Exploring Microscopic Driving Volatility in Naturalistic Driving Environment Prior to Involvement in Safety Critical Events—Concept of Event-Based Driving Volatility. *Accident Analysis and Prevention*, Vol. 132, 2019, p. 105277. https://doi.org/10.1016/j.aap.2019.105277.

93. Khattak, A. J., and B. Wali. Analysis of Volatility in Driving Regimes Extracted from Basic Safety Messages Transmitted between Connected Vehicles. *Transportation Research Part C: Emerging Technologies*, Vol. 84, 2017, pp. 48–73. https://doi.org/10.1016/j.trc.2017.08.004.

94. Khoda Bakhshi, A. ., and M. M. Ahmed. Real-Time Crash Prediction for a Long Low-Traffic Volume Corridor Using Corrected-Impurity Importance and Semi-Parametric Generalized Additive Model. *Journal of Transportation Safety & Security (Unpublished results)*.

95. Wang, X., A. J. Khattak, J. Liu, G. Masghati-Amoli, and S. Son. What Is the Level of Volatility in Instantaneous Driving Decisions? *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 413–427. https://doi.org/10.1016/j.trc.2014.12.014.

96. Liu, J., 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, 2016, pp. 83–100. https://doi.org/10.1016/j.trc.2016.03.009.

97. Wolshon, B., S. A. Parr, and S. M. Mousavi. *Identifying High-Risk Roadways for Infrastructure Investment Using Naturalistic Driving Data*. 2015.

98. Bagdadi, O., and A. Várhelyi. Jerky Driving - An Indicator of Accident Proneness? *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011, pp. 1359–1363. https://doi.org/10.1016/j.aap.2011.02.009.

99. Bagdadi, O., and A. Várhelyi. Jerky Driving - An Indicator of Accident Proneness? *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011, pp. 1359–1363. https://doi.org/10.1016/j.aap.2011.02.009.

100. Pande, A., J. Loy, V. Dixit, K. Spansel, B. Wolshon, and J. D. Kent. Exploration of Naturalistic Driving Data for Identifying High Crash Risk Highway Locations. 2014.

101. Mousavi, S. M. Identifying High Crash Risk Roadways through Jerk-Cluster Analysis. *LSU Master's Theses*, 2015.

102. Bagdadi, O., and A. Várhelyi. Development of a Method for Detecting Jerks in Safety Critical Events. *Accident Analysis and Prevention*, Vol. 50, 2013, pp. 83–91. https://doi.org/10.1016/j.aap.2012.03.032.

103. Federal Highway Administration (FHWA). Surrogate Safety Assessment Model (SSAM) - FHWA-HRT-08-049. https://www.fhwa.dot.gov/publications/research/safety/08049/. Accessed Aug. 7, 2020.

104. Hayward, J. C. NEAR-MISS DETERMINATION THROUGH USE OF A SCALE OF DANGER. *Highway Research Record*, No. 384, 1972.

105. Dijkstra, A., P. Marchesini, F. Bijleveld, V. Kars, H. Drolenga, M. Van Maarseveen, A. Dijkstra, P. Marchesini, F. Bijleveld, and V. Kars. Do Calculated Conflicts in Microsimulation Model Predict Number of Crashes? *Transportation Research Record: Journal of the Transportation Research*, Vol. 2147, 2010, pp. 105–112. https://doi.org/10.3141/2147-13.

106. Saunier, N. *Surrogate Safety Analysis*. 2010.

107. Gettman, Douglas; Pu, Lili; Sayed, Tarek; Shelby, S. *Surrogate Safety Assessment Model and Validation: Final*. 2008.

108. U.S. Department of Transportation. *Surrogate Safety Assessment Model (SSAM)– SOFTWARE USER MANUAL - FHWA-HRT-08-050*. 2008.

109. Mousavi, S. M., O. A. Osman, D. Lord, K. K. Dixon, and B. Dadashova. *Investigating the Safety and Operational Benefits of Mixed Traffic Environments in the Proximity of a Driveway on an Urban Arterial*. 2020.

110. Katrakazas, C., M. Quddus, and W. H. Chen. A New Integrated Collision Risk Assessment Methodology for Autonomous Vehicles. *Accident Analysis and Prevention*, Vol. 127, 2019, pp. 61–79. https://doi.org/10.1016/j.aap.2019.01.029.

111. Zhang, J., K. Wu, M. Cheng, M. Yang, Y. Cheng, and S. Li. Safety Evaluation for Connected and Autonomous Vehicles' Exclusive Lanes Considering Penetrate Ratios and Impact of Trucks Using Surrogate Safety Measures. *Journal of Advanced Transportation*, 2020.

112. PTV VISSIM. *PTV VISSIM 10 User Manual*. 2018.

113. Chowdhury, M., N. Derov, P. Tan, and A. Sadek. Prohibiting Left-Turn Movements at Mid-Block Unsignalized Driveways: Simulation Analysis. *Journal of Transportation Engineering*, Vol. 131, No. 4, 2005, pp. 279–285. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:4(279).

114. Crossing, C. R. *VISSIM Calibration and Validation*. 2006.

115. Dowling, R., A. Skabardonis, J. Halkias, G. McHale, and G. Zammit. Guidelines for Calibration of Microsimulation Models: Framework and Applications. Transportation Research Board 96th Annual Meeting, National Research Council, 1876, 1, Jan 01, 2004, pp. 1–9.

116. Montgomery, D. C. *Design and Analysis of Experiments*. John Wiley & Sons, Inc., Danver, 2012.

117. Kamrani, M., R. Arvin, and A. J. Khattak. Extracting Useful Information from Basic Safety Message Data: An Empirical Study of Driving Volatility Measures and Crash Frequency at Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2672, No. 38, 2018, pp. 290–301. https://doi.org/10.1177/0361198118773869.

118. Stanford. 1 Dispersion and Deviance Residuals. https://statweb.stanford.edu/~jtaylo/courses/stats306b/restricted/notebooks/quasilikelihood.pdf. Accessed Sep. 24, 2020.

119. Lillis, D. Generalized Linear Models in R, Part 2: Understanding Model Fit in Logistic Regression Output. https://www.theanalysisfactor.com/r-glm-model-fit/#:~:text=R reports two forms of,the intercept (grand mean).&text=The Residual Deviance has reduced,of two degrees of freedom. Accessed Sep. 24, 2020.

120. Deluka Tibljaš, A., T. Giuffrè, S. Surdonja, and S. Trubia. Introduction of Autonomous Vehicles: Roundabouts Design and Safety Performance Evaluation. *Sustainability*, Vol. 10, No. 4, 2018, p. 1060. https://doi.org/10.3390/su10041060.

121. Archer, J. *Indicators for Traffic Safety Assessment and Prediction and Their Application in Micro-Simulation Modelling: A Study of Urban and Suburban Intersections The Greater Stockholm Area Defined by Accident Occurrence (from STRADA)*. 2005.

122. Van der Horst, A. R. A. *A Time-Based Analysis of Road User Behaviour in Normal and Critical Encounters*. Royal Institute of Technology, 1990.

123. Wikipedia. Pearson Correlation Coefficient. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient. Accessed Aug. 9, 2020.

124. R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed Sep. 26, 2020.

125. How to Handle Missing Data in Machine Learning: 5 Techniques. https://dev.acquia.com/blog/how-to-handle-missing-data-in-machine-learning-5-techniques/09/07/2018/19651. Accessed Aug. 30, 2020.

126. One_hot Function | R Documentation. https://www.rdocumentation.org/packages/mltools/versions/0.3.5/topics/one_hot. Accessed Aug. 30, 2020.

127. Akaike Information Criterion. *Wikipedia*. https://en.wikipedia.org/wiki/Akaike_information_criterion. Accessed Sep. 21, 2020.

128. Takane, Y., and H. Bozdogan. Akaike Information Criterion (AIC): Introduction. *Psychometrika*, Vol. 52, No. 3, 1987.

129. Fraade-Blanar, L., M. Blumenthal, J. Anderson, and N. Kalra. *Measuring Automated Vehicle Safety: Forging a Framework*. RAND Corporation, 2018.

APPENDIX A

CODING IN R

A. Calculating Longitudinal and Lateral Driving Volatility Measures

```
#install.packages('readr')

library(readr)


fzp <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzp" , pattern = ".txt")

fhz <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhz" , pattern = ".txt")


setwd('D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzp')
```

```r
for(i in fzp) {

  setwd('D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzp')

  # skip: skips the rows of the txt file- sep: shows what has separated the columns

  NewTable <- read.delim (file =  i, header = T, sep = ";" , skip = 17)

  setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzpDone")

  write.csv(NewTable, file = paste(i, ".csv"))

}


setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhz")


for(i in fhz) {

  setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhz")

  NewTable1 <- read.delim (file =  i, header = T, sep = ";" , skip = 8)

  setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhzDone")

  write.csv(NewTable1, file = paste(i, ".csv"))

}


#install.packages('dplyr')

library(dplyr)


main_df_fzp <- list.files(path = "D:/Dissertation/VISSIM/Larger vehicle Size/3/fzpDone" ,

pattern = ".csv")
```

```r
Veh_Type_Table <- list.files(path = "D:/Dissertation/VISSIM/Larger vehicle

Size/3/fhzDone" , pattern = ".csv")


setwd('D:/Dissertation/VISSIM/Larger vehicle Size/3/Final')


#if(

#  length(setdiff(main_df_fzp, Veh_Type_Table))>0)

#  stop("Actually, the two directories do not have the same files")


for(

  file in main_df_fzp) {

  varname <- substr(file, start=1, stop=nchar(file)-4)

  main <- read.csv(file.path("D:/Dissertation/VISSIM/Larger vehicle Size/3/fzpDone",

file))

  vehType <- read.csv(file.path("D:/Dissertation/VISSIM/Larger vehicle Size/3/fhzDone",

file))


  #rename just a single column to match to the other table to be able to  left_join

  names(main)[names(main)=="NO"] <- "VehNo"


  new_table <- left_join(main, vehType, by=c("VehNo"))


  #sort by time and veh #
```

```r
new_table <- new_table[with(new_table, order(VehNo , X.VEHICLE.SIMSEC)),]


new_table$pos_mi <- new_table$POS / 5280

new_table$time_hr <- new_table$X.VEHICLE.SIMSEC / 3600


#calculating speed by group
new_table$speed <- (ave(new_table$POS, factor(new_table$VehNo), FUN=function(x)
c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC, factor(new_table$VehNo),
FUN=function(x) c(NA,diff(x))))


new_table$acc <- (ave(new_table$speed, factor(new_table$VehNo), FUN=function(x)
c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC, factor(new_table$VehNo),
FUN=function(x) c(NA,diff(x))))


new_table$jrk <- (ave(new_table$acc, factor(new_table$VehNo), FUN=function(x)
c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC, factor(new_table$VehNo),
FUN=function(x) c(NA,diff(x))))


# add lateral changes


new_table$speed_Lat <- (ave(new_table$POSLAT, factor(new_table$VehNo),
FUN=function(x) c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC,
factor(new_table$VehNo), FUN=function(x) c(NA,diff(x))))
```

160

```
new_table$acc_Lat <- (ave(new_table$speed_Lat, factor(new_table$VehNo),

FUN=function(x) c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC,

factor(new_table$VehNo), FUN=function(x) c(NA,diff(x))))


new_table$jrk_Lat <- (ave(new_table$acc_Lat, factor(new_table$VehNo),

FUN=function(x) c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC,

factor(new_table$VehNo), FUN=function(x) c(NA,diff(x))))


new_table <- new_table[ , -c(1,  8:11 , 13:17)]


write.csv(new_table, file = paste(file, "-edit.csv"))
}
```

B. Calculating Network-Level Longitudinal and Lateral Volatility Measures

```
setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/19/Final")


filenames <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle Size/19/Final",

pattern = ".csv",

                full.names = TRUE)


for (file in filenames){


 mydata_1 <- read.csv(file)


 mydata_1 <- na.omit(mydata_1)


 non_zero_spd <- mydata_1[mydata_1$speed > 0,]

 spd_avg <- mean(non_zero_spd$speed)
```

```
spd_sd <- sd(non_zero_spd$speed)

spd_cv <- (spd_sd) * 100/spd_avg


up_thresh <- spd_avg + 2*(spd_sd)

dn_thresh <- spd_avg - 2*(spd_sd)

perc_spd <- nrow(subset(non_zero_spd , speed < dn_thresh | speed > up_thresh))*100 /

nrow(non_zero_spd)


acc <- mydata_1[mydata_1$acc >= 0,]

acc_avg <- mean(acc$acc)

acc_sd <- sd(acc$acc)

acc_cv <- (acc_sd) * 100/acc_avg


up_thresh <- acc_avg + 2*(acc_sd)

dn_thresh <- acc_avg - 2*(acc_sd)

perc_acc <- nrow(subset(acc , acc < dn_thresh | acc > up_thresh))*100 / nrow(acc)


dec <- mydata_1[mydata_1$acc < 0,]

dec_avg <- mean(dec$acc)

dec_sd <- sd(dec$acc)

dec_cv <- (dec_sd) * 100/dec_avg


up_thresh <- dec_avg + 2*(dec_sd)
```

```
dn_thresh <- dec_avg - 2*(dec_sd)

perc_dec <- nrow(subset(dec , acc < dn_thresh | acc > up_thresh))*100 / nrow(dec)


posjrk <- mydata_1[mydata_1$jrk >= 0,]

posjrk_avg <- mean(posjrk$jrk)

posjrk_sd <- sd(posjrk$jrk)

posjrk_cv <- (posjrk_sd) * 100/posjrk_avg


up_thresh <- posjrk_avg + 2*(posjrk_sd)

dn_thresh <- posjrk_avg - 2*(posjrk_sd)

perc_posjrk <- nrow(subset(posjrk , jrk < dn_thresh | jrk > up_thresh))*100 /

nrow(posjrk)


negjrk <- mydata_1[mydata_1$jrk < 0,]

negjrk_avg <- mean(negjrk$jrk)

negjrk_sd <- sd(negjrk$jrk)

negjrk_cv <- (negjrk_sd) * 100/negjrk_avg


up_thresh <- negjrk_avg + 2*(negjrk_sd)

dn_thresh <- negjrk_avg - 2*(negjrk_sd)

perc_negjrk <- nrow(subset(negjrk , jrk < dn_thresh | jrk > up_thresh))*100 /

nrow(negjrk)
```

```
#lateral


latspd_avg <- mean(mydata_1$speed_Lat)

latspd_sd <- sd(mydata_1$speed_Lat)

spd_cv <- (latspd_sd) * 100/latspd_avg


up_thresh <- latspd_avg + 2*(latspd_sd)

dn_thresh <- latspd_avg - 2*(latspd_sd)

perc_latspd <- nrow(subset(mydata_1 , speed_Lat < dn_thresh | speed_Lat >

up_thresh))*100 / nrow(mydata_1)


latacc <- mydata_1[mydata_1$acc_Lat >= 0,]

latacc_avg <- mean(latacc$acc_Lat)

latacc_sd <- sd(latacc$acc_Lat)

latacc_cv <- (latacc_sd) * 100/latacc_avg


up_thresh <- latacc_avg + 2*(latacc_sd)

dn_thresh <- latacc_avg - 2*(latacc_sd)

perc_latacc <- nrow(subset(latacc , acc_Lat < dn_thresh | acc_Lat > up_thresh))*100 /

nrow(latacc)


latdec <- mydata_1[mydata_1$acc_Lat < 0,]

latdec_avg <- mean(latdec$acc_Lat)
```

```r
latdec_sd <- sd(latdec$acc_Lat)

latdec_cv <- (latdec_sd) * 100/latdec_avg


up_thresh <- latdec_avg + 2*(latdec_sd)

dn_thresh <- latdec_avg - 2*(latdec_sd)

perc_latdec <- nrow(subset(latdec , acc_Lat < dn_thresh | acc_Lat > up_thresh))*100 /
nrow(latdec)


poslatjrk <- mydata_1[mydata_1$jrk_Lat >= 0,]

poslatjrk_avg <- mean(poslatjrk$jrk)

poslatjrk_sd <- sd(poslatjrk$jrk)

poslatjrk_cv <- (poslatjrk_sd) * 100/poslatjrk_avg


up_thresh <- poslatjrk_avg + 2*(poslatjrk_sd)

dn_thresh <- poslatjrk_avg - 2*(poslatjrk_sd)

perc_poslatjrk <- nrow(subset(poslatjrk , jrk_Lat < dn_thresh | jrk_Lat > up_thresh))*100
/ nrow(poslatjrk)


neglatjrk <- mydata_1[mydata_1$jrk_Lat < 0,]

neglatjrk_avg <- mean(neglatjrk$jrk_Lat)

neglatjrk_sd <- sd(neglatjrk$jrk_Lat)

neglatjrk_cv <- (neglatjrk_sd) * 100/neglatjrk_avg
```

```r
  up_thresh <- neglatjrk_avg + 2*(neglatjrk_sd)

  dn_thresh <- neglatjrk_avg - 2*(neglatjrk_sd)

  perc_neglatjrk <- nrow(subset(neglatjrk , jrk_Lat < dn_thresh | jrk_Lat > up_thresh))*100
/ nrow(neglatjrk)


  df <- NULL

  df <- rbind(df, data.frame( spd_avg , spd_sd , spd_cv , perc_spd,

                 acc_avg , acc_sd , acc_cv , perc_acc,

                 dec_avg , dec_sd , dec_cv , perc_dec,

                 posjrk_avg ,posjrk_sd , posjrk_cv , perc_posjrk,

                 negjrk_avg , negjrk_sd ,negjrk_cv , perc_negjrk,

                 latspd_avg , latspd_sd , spd_cv , perc_latspd,

                 latacc_avg , latacc_sd , latacc_cv , perc_latacc,

                 latdec_avg ,  latdec_sd , latdec_cv , perc_latdec,

                 poslatjrk_avg , poslatjrk_sd , poslatjrk_cv , perc_poslatjrk,

                 neglatjrk_avg , neglatjrk_sd ,  neglatjrk_cv , perc_neglatjrk
  ))


  write.csv( df , file = paste0( file , "_Volatility.csv"), row.names = FALSE)


filenames <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle
Size/19/Final/Volatility",

                 full.names = TRUE)
```

```r
my.df <- do.call(rbind,

          lapply(filenames, function(x)

            cbind(read.csv(x),

                name = tools::file_path_sans_ext(basename(x)))))


my.df$MPR <- ifelse(grepl("_20AV", my.df$name) , "20",

                  ifelse(grepl("_40AV", my.df$name) , "40" ,

                      ifelse(grepl( "_60AV", my.df$name) , "60" ,

                          ifelse(grepl( "_80AV", my.df$name) , "80" ,

                              ifelse(grepl( "_90AV", my.df$name) , "90" ,

                                  ifelse(grepl("_AV", my.df$name) , "100" , "00"))))))


my.df$filelocation <- "Smaller Vehicle Size/19/Final"


setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/19/Final/Volatility")


write.csv(my.df , "Larger vehicle Size_18.csv")
```

C. Merging Driving Volatility Measures and Adding the Studied Variables

```r
filenames <- list.files(path = "D:/Dissertation/VISSIM/VolatilityAll/Renamed",
                full.names = TRUE)


import_files <- lapply(filenames,read.csv,stringsAsFactors =FALSE)


my.df <- do.call(rbind,
         lapply(filenames, function(x)
           cbind(read.csv(x))))


my.df$signalL <- ifelse(grepl("Longer", my.df$Scenario) , 193,
                   ifelse(grepl("Shorter", my.df$Scenario) , 75 , 135))


my.df$speed <- ifelse(grepl("Lower", my.df$Scenario) , 32.5,
                  ifelse(grepl("Higher", my.df$Scenario) , 42.5 , 37.5))
```

my.df$AVSize <- ifelse(grepl("Larger", my.df$filelocation) , 15.125, 12.949)

my.df$LTphase <- ifelse(grepl("PermitProtect", my.df$Scenario) , "Protected-Permitted",

ifelse(grepl("Permit", my.df$Scenario) , "Permitted" , "Protected"))

write.csv(my.df , "Merged_Volatility.csv")

D. Calculating Jerks

```
#install.packages('readr')
library(readr)


fzp <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzp" , pattern =
".txt")
fhz <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhz" , pattern =
".txt")


setwd('D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzp')


for(i in fzp) {
 setwd('D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzp')
 # skip: skips the rows of the txt file- sep: shows what has separated the columns
 NewTable <- read.delim (file =  i, header = T, sep = ";" , skip = 17)
 setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzpDone")
```

```r
  write.csv(NewTable, file = paste(i, ".csv"))

}


setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhz")


for(i in fhz) {

  setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhz")

  NewTable1 <- read.delim (file =  i, header = T, sep = ";" , skip = 8)

  setwd("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fhzDone")

  write.csv(NewTable1, file = paste(i, ".csv"))

}


#install.packages('dplyr')

library(dplyr)


main_df_fzp <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle

Size/3/fzpDone" , pattern = ".csv")


Veh_Type_Table <- list.files(path = "D:/Dissertation/VISSIM/Smaller Vehicle

Size/3/fhzDone" , pattern = ".csv")


setwd('D:/Dissertation/VISSIM/Smaller Vehicle Size/3/Final')
```

```r
#if(
#  length(setdiff(main_df_fzp, Veh_Type_Table))>0)
#  stop("Actually, the two directories do not have the same files")


for(
 file in main_df_fzp) {
 varname <- substr(file, start=1, stop=nchar(file)-4)
 main <- read.csv(file.path("D:/Dissertation/VISSIM/Smaller Vehicle Size/3/fzpDone",
file))
 vehType <- read.csv(file.path("D:/Dissertation/VISSIM/Smaller Vehicle
Size/3/fhzDone", file))


 #rename just a single column to match to the other table to be able to  left_join
 names(main)[names(main)=="NO"] <- "VehNo"


 new_table <- left_join(main, vehType, by=c("VehNo"))


 #sort by time and veh #
 new_table <- new_table[with(new_table, order(VehNo , X.VEHICLE.SIMSEC)),]


 new_table$pos_mi <- new_table$POS / 5280
 new_table$time_hr <- new_table$X.VEHICLE.SIMSEC / 3600
```

```r
#calculating speed, acc, and jerk by group

new_table$speed <- (ave(new_table$POS, factor(new_table$VehNo), FUN=function(x)

c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC, factor(new_table$VehNo),

FUN=function(x) c(NA,diff(x))))


new_table$acc <- (ave(new_table$speed, factor(new_table$VehNo), FUN=function(x)

c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC, factor(new_table$VehNo),

FUN=function(x) c(NA,diff(x))))


new_table$jrk <- (ave(new_table$acc, factor(new_table$VehNo), FUN=function(x)

c(NA,diff(x))))/(ave(new_table$X.VEHICLE.SIMSEC, factor(new_table$VehNo),

FUN=function(x) c(NA,diff(x))))


new_table <- new_table[ , -c(1, 4:5, 7:11 , 13:15)]

#match(Table$SecondVID, Veh_Type_Table$Veh, incomparables = F)


#new_table <- new_table[ , c(1:43 , 49)]


write.csv(new_table, file = paste(file, "-edit.csv"))

}
```

E. Regression Using Machine Learning

#https://www.machinelearningplus.com/machine-learning/caret-

package/#35howtopreprocesstotransformthedata

#install.packages("lattice")

library(lattice)

#install.packages("ggplot2")

library(ggplot2)

#install.packages("caret")

library(caret)

#install.packages("MASS")

library(MASS)

#install.packages("ggplot2")

library(ggplot2)

```
setwd("C:/Users/maryam.mousavi/Google Drive/Dissertation

(11.30.19)/Analysis/MachineLearning")


# 1. Loading data


raw_data <- read.csv("MachineLearning(01.20.20)_csv.csv")


#since changing NAs for 0 for avg_jrk does not make any difference, we make NAs zero

to calculate the avg jrk

raw_data[is.na(raw_data)] <- 0


#for converting values a specific value in a column with another value

##raw_data$LTPhasing[raw_data$LTPhasing == "Actual"] <- "Protected"


#sum(is.na(raw_data$Avg_AV_Jrk))

#sum(is.na(raw_data$Avg_CV_Jrk))


raw_data$total_jrk <- raw_data$AV_cnt + raw_data$CV_cnt

raw_data$avg_jrk <- (raw_data$Avg_AV_Jrk + raw_data$Avg_CV_Jrk)/2

raw_data$avg_pr <- (raw_data$AV_pr + raw_data$CV_pr)/2


# 2. getting descriptive stats for variables
```

```r
#install.packages('skimr')

library(skimr)


write.csv(skim(raw_data , SignalLength_num , AVSize_num, AvgSpeedLimit_num, MPR

, AV_cnt, CV_cnt ,

        Avg_AV_Jrk, Avg_CV_Jrk, avg_jrk, avg_pr , avg_pr, total_jrk) , "123.csv")


#making training dataset just considering AVs

##inTrain <- createDataPartition(y=raw_data$AV_pr , p=0.75, list = FALSE)


##training <- raw_data[inTrain, ]

##testing <- raw_data[-inTrain, ]


#making training dataset considering both AVs and RVs (we can use either #of jerks or

#% of jerk (because we made NAs as zero))


# 3. making training and test sets


names(raw_data)

raw_data <- raw_data[ , c('avg_jrk' , 'AvgSpeedLimit_num' , 'SignalLength_num' ,

'AVSize_num' , 'MPR' , 'LTPhasing') ]
```

set.seed(1111)

inTrain <- createDataPartition(y=raw_data$avg_jrk , p=0.75, list = FALSE)

training <- raw_data[inTrain,]

testing <- raw_data[-inTrain,]

4. store x and y for later use

x <- training[,c('AvgSpeedLimit_num' , 'SignalLength_num' , 'AVSize_num' , 'MPR' ,

'LTPhasing')]

y <- training$avg_jrk

5. One-Hot Encoding

Creating dummy variables is converting a categorical variable to as many binary

variables as here are categories.

we just do it for training dataset

dummies_model <- dummyVars(avg_jrk ~ ., data=training)

Create the dummy variables using predict. The Y variable (Purchase) will not be present

in trainData_mat.

```r
training_mat <- predict(dummies_model, newdata = training)


# # Convert to dataframe

training <- data.frame(training_mat)


# # See the structure of the new dataset

str(training)


#now we should append the y variable

training$avg_jrk <- y


# 6. drawing featureplots


featurePlot(x= training[ ,c("AvgSpeedLimit_num" , "MPR" , "AVSize_num" ,

"SignalLength_num" ,

                "LTPhasing.Permitted" , "LTPhasing.Protected" ,

                "LTPhasing.ProtectedPermitted")], y = training$avg_jrk , plot= 'pairs')


# hist of the dependent variable
x <- training$avg_jrk
h<- hist(x, breaks=10, col="light blue", xlab="Average Jerk",

   main="Average Jerk Histogram with Normal Curve")
```

```
xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit*diff(h$mids[1:2])*length(x)

lines(xfit, yfit, col="red", lwd=2)



# 7.Correlateion test


install.packages("corrplot")

library(corrplot)



correlations <- cor(training[,1:8])

# create correlation plot

corrplot(correlations, method="circle")



# we also can do scatterplot matrix

pairs(training)



# 8. train model using training data


modelFit <- train(avg_jrk ~ AvgSpeedLimit_num + SignalLength_num + AVSize_num +

MPR + LTPhasing.Permitted +

        LTPhasing.Protected + LTPhasing.ProtectedPermitted ,
```

```
           data=training, method='glm')


write.csv(coef(summary(modelFit)) , "GLMresults.csv")


# 9. now, we should impute the testing data for prediction


x <- testing[,c('AvgSpeedLimit_num' , 'SignalLength_num' , 'AVSize_num' , 'MPR' ,

'LTPhasing')]

y <- testing$avg_jrk


dummies_model1  <- dummyVars( avg_jrk ~ ., data=testing)


# Create the dummy variables using predict. The Y variable (Purchase) will not be present

in trainData_mat.

testing_mat <- predict(dummies_model1, newdata = testing)


# # Convert to dataframe

testing <- data.frame(testing_mat)


# # See the structure of the new dataset

str(testing)


#now we should append the y variable
```

```
testing$avg_jrk <- y


# 10. now, predict based on test data


predicted <- predict(modelFit, newdata = testing)

head(predicted)


confusionMatrix(reference = testing$avg_jrk, data = predicted, mode='everything',

positive='MM')
```