Operational Safety Assessment Methodology Framework:

An Approach to Quantifying Automated Vehicle Safety

by

Steven G. Como Jr.

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved September 2022 by the
Graduate Supervisory Committee:

Jeffrey Wishart, Co-Chair
Yezhou Yang, Co-Chair
Yan Chen
Francesca Favarò

ARIZONA STATE UNIVERSITY

December 2022

ABSTRACT

To date, there is not a standardized method for consistently quantifying the performance of an automated driving system (ADS)-equipped vehicle (AV). The purpose of this dissertation is to contribute to a framework for such an approach referred to throughout as the operational safety assessment (OSA) methodology. Through this research, safety metrics are identified, researched, and analyzed to capture aspects of the operational safety of AVs, interacting with other salient objects. This dissertation outlines the approach for developing this methodology through a series of key steps including: (1) comprehensive literature review; (2) research and refinement of OSA metrics; (3) generation of MATLAB script for metric calculations; (4) generation of simulated events for analysis; (5) collection of real-world data for analysis; (6) review of OSA methodology results; and (7) discussion of future work to expand complexity, fidelity, and relevance aspects of the OSA methodology.

The detailed literature review includes the identification of metrics historically used in both traditional and more recent evaluations of vehicle performance. Subsequently, the metric formulations are refined, and robust severity evaluations are proposed. A MATLAB script is then presented which was generated to calculate the metrics from any given source assuming proper formatting of the data. To further refine the formulations and the MATLAB script, a variety of simulated scenarios are discussed including car-following, intersection, and lane change situations. Additionally, a data collection activity is presented, leveraging the SMARTDRIVE testbed operated by Maricopa County Department of Transportation in Anthem, AZ to collect real-world data from an active intersection.

Lastly, the efficacy of the OSA methodology with respect to the evaluation of vehicle performance for a set of scenarios is evaluated utilizing both simulated and real-world data. This assessment provides a demonstration of the ability and robustness of this methodology to evaluate vehicle performance for a given scenario. At the conclusion of this dissertation, additional factors including fidelity, complexity, and relevance are explored to contribute to a more comprehensive evaluation.

DEDICATION

Working full-time while completing a PhD full-time sounded like a challenging task from the start; however, I made it only several weeks into the first semester before questioning whether it was even a possibility. The only way I was able to complete this PhD and maintain my sanity was the incredible support network including my family, friends, advisors, coworkers, professors, and fellow students. As many people as there were that played a fundamental role in helping me accomplish this feat, the one person that I absolutely could not have done this without is my incredible wife, Katie. Whether it was doing extra chores around the house when I was staying up late working on my dissertation or just pulling me away from my computer for an hour when I would get stressed, Katie deserves this PhD just as much as I do. I understand more than I could've imagined that getting a PhD is a team effort and for that I am so incredibly grateful to all of the people that continued to push and support me to accomplish my goals. While there are too many to list that deserve thanks, you know who you are.

AUTHOR CONTRIBUTIONS

Numerous students, professors, government employees, and industry experts have been involved in the research contributing to the information presented in this dissertation. As the PhD candidate for this program, my contributions are explicitly listed. From the beginning of my PhD program, I joined the Verification and Validation (V&V) task force of the SAE On-Road Automated Driving (ORAD) committee and have participated in discussions surrounding the taxonomy and classification of operational safety metrics, an extensive literature review for activities and entities related to AVs, and safety principles for AVs. I assisted in publishing the aforementioned literature review in the SAE International Journal of Connected and Automated Vehicles and have participated as a co-author for numerous papers submitted to SAE World Congress Experience (WCX) in addition to one that was submitted to IEEE International Conference on Robotics and Automation (ICRA). Further, I was the lead author on an SAE paper titled *Evaluating the Severity of Safety Envelope Violations in the Proposed Operational Safety Assessment (OSA) Methodology for Automated Vehicles* as well as the SAE paper currently in progress titled *Evaluating Automated Vehicle Scenario Navigation using the Operational Safety Assessment (OSA) Methodology.*

I have helped provide guidance for three students who contributed literature reviews to the safety assessments of other industries utilizing automation and the quantification of fidelity. Through my research, I helped expand and refine the metric formulations previously proposed in [1], established severity formulations for each of the proposed OSA metrics, and generated simulations for a variety of scenarios to better contextualize these metrics and severity formulations. I developed a MATLAB script capable of identifying

OSA metric violations and calculating the corresponding severities and refined the script to accept inputs from a variety of sources including simulations and real-world data. I led the generation of additional data through simulation in addition to assisting in multiple data collection activities to capture real-world data with an established ground truth baseline and processed these data for usage by various groups. Utilizing the script I generated for metrics calculations, I further evaluated the impact of accuracy and resolution on the metrics calculations for contextualization of data capture systems. Lastly, I proposed methods for incorporating the fidelity of a test methodology and relevance and complexity of a scenario to further refine the OSA methodology framework. The culmination of this dissertation establishes a robust framework with corresponding tools to evaluate scenarios with the flexibility and convenience of utilizing data from varying sources in addition to easily adaptable parameters for the metrics calculations.

A list of publications related to my dissertation work is included below:

- 2020. J. Wishart, **S. Como**, U. Forgione, J. Weast, L. Weston, A. Smart, G. Nicols, and Ramesh S., "Literature Review of Verification & Validation Activities of Automated Driving Systems," Accepted for publication in SAE International Journal of Connected and Automated Vehicles.
- 2020. J. Wishart, **S. Como**, M. Elli, B. Russo, N. Altekar, and J. Weast, "Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles," SAE WCX Conference, Paper 2020-01-1206, Detroit, USA.
- 2021. M. Elli, J. Wishart, **S. Como**, S. Dhakshinamoorthy, and J. Weast, "Evaluation of Operational Safety Assessment (OSA) Metrics for Automated

Vehicles in Simulation," SAE WCX Conference, Paper 2021-01-0868, Detroit, USA.

- 2021. N. Altekar, **S. Como**, D. Lu, J. Wishart, D. Bruyere, F. Saleem, and L. Head, "Infrastructure-Based Sensor Data Capture Systems for Measurement of Operational Safety Assessment (OSA) Metrics," SAE WCX Conference, Paper 2021-01-0175, Detroit, USA.

- 2021. D. Lu, V. Jammula, **S. Como**, J. Wishart, Y. Chen, and Y. Yang, "CAROM – Vehicle Localization and Traffic Scene Reconstruction from Monocular Cameras on Road Infrastructures," International Conference on Robotics and Automation, Xi'an, China.

- 2022. N. Kidambi, J. Wishart, M. Elli, and **S. Como**, "Sensitivity of Automated Vehicle Operational Safety Assessment (OSA) Metrics to Measurement and Parameter Uncertainty," SAE WCX Conference, Paper 2022-01-0815, Detroit, USA.

- 2022. **S. Como**, J. Wishart, M. Elli, and N. Kidambi, "Evaluating the Severity of Safety Envelope Violations in the Proposed Operational Safety Assessment (OSA) Methodology for Automated Vehicles," SAE WCX Conference, Paper 2022-01-0819, Detroit, USA.

- (Work in Progress). **S. Como**, J. Wishart, M. Elli, and N. Kidambi, "Evaluating Automated Vehicle Scenario Navigation using the Operational Safety Assessment (OSA) Methodology," Submitted to 2023 SAE WCX Conference.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

LIST OF ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
ABC	Achieved Behavioral Competency
ADS	Automated Driving System
ADSA	Automated Driving System Active
AI	Artificial Intelligence
AIS	Abbreviated Injury Scale
AR	Augmented Reality
AV	ADS-Equipped Vehicle
CARLA	Cars Learning to Act
CAV	Connected and Automated Vehicle
CDS	Crashworthiness Data System
CI	Collision Incident
CNN	Convoluted Neural Network
DDT	Dynamic Driving Task
DNN	Deep Neural Network
Dstop	Distance to Stop
EDR	Event Data Recorder
FARS	Fatality Analysis Reporting System
GES	General Estimates System
GST	Global Soft Target
HARA	Hazard Analysis and Risk Assessment
HTC	Human Traffic Controller
HTCDER	Human Traffic Control Detection Error Rate
HTCVR	Human Traffic Control Violation Rate
HVE	Human, Vehicle, Environment

IAM	Institute of Automated Mobility
IIHS	Insurance Institute for Highway Safety
MCDOT	Maricopa County Department of Transportation
ML	Machine Learning
MPrISM	Model Predictive Instantaneous Safety Metric
MRD	Minimum Required Deceleration
MSE	Minimum Safety Envelope
MSECE	Minimum Safety Envelope Calculation Error
MSEF	Minimum Safety Envelope Factor
MTTC	Modified Time-to-Collision
NASS	National Automotive Sampling System
NHTSA	National Highway Traffic Safety Administration
ODD	Operational Design Domain
OEF-CI	Other Entity Fault Collision Incident
OEM	Original Equipment Manufacturer
ORAD	On-Road Automated Driving
OSA	Operational Safety Assessment
PA	Predictable Acceleration
PDOF	Principal Direction of Force
PET	Post Encroachment Time
PR	Proper Response
RCRI	Rear-End Crash Risk Index
RRT	Rapidly Exploring Random Trees
RSS	Responsibility-Sensitive Safety
SDO	Standards Development Organization
STAMP	System-Theoretic Accident Model and Processes
SVF-CI	Subject Vehicle Fault Collision Incident
TET	Time-Exposed Time-to-Collision

THW	Time Headway
TIT	Time-Integrated Time-to-Collision
TLV	Traffic Law Violation
TRI	Transportation Research Institute
TTC	Time-to-Collision
USDOT	United States Department of Transportation
V&V	Verification and Validation
VRU	Vulnerable Road User
VSSA	Voluntary Safety Self-Assessment
VUT	Vehicle Under Test

PREFACE

One of the biggest challenges facing the advancement of AVs[1] today is the determination and validation of operational safety,[2] as well as agreement on the level of safety that AVs must exhibit. There are numerous possible ways to quantify safety ranging from comparison to the operational safety performance of a human-driven vehicle, to the evaluation of a given criteria for specific scenarios similar to the driving test humans are required to take before receiving a license. Unfortunately, the answer to this safety evaluation is not as straightforward as the latter situation since vehicles utilizing artificial intelligence (AI) could be trained to pass a regulated test without being capable of safely navigating scenarios for which the AI was not specifically trained. At the same time, unlimited time and resources are not available to complete the necessary amount of testing to consider the infinite number of scenarios an AV could face in the real world. Although there have been significant advancements as of late in AV development, there is not yet an agreed-upon standard that can be used to consistently and quantitatively measure the operational safety performance of an AV. To date, all guidelines proposed in the U.S. by regulators such as the National Highway Traffic Safety Administration (NHTSA), as well as standards development organizations (SDOs) such as SAE International, ISO, and IEEE are only voluntary, not regulatory, and most importantly, are incomplete. The operational safety assessment (OSA) methodology described here is thought to be a critical step toward

[1] Automated vehicles, abbreviated AVs throughout this paper will encompass all automated driving system (ADS)-equipped vehicles in addition to connected and automated vehicles (CAVs). AVs is used throughout for simplification.

[2] The term "operational" is used in describing the OSA methodology; however, it should be noted that there are operational aspects not included here. For instance, aspects of passenger securement, pickups/drop-offs, and other logistical operations are not the focus of this work. Throughout, "operational" describes the behavioral evaluation of the subject vehicle in conducting a given scenario.

quantifying vehicle performance to assist in the safe deployment and commercialization of AVs.

The Institute of Automated Mobility (IAM), created by Executive Order in Arizona in order to assist the AV industry through research into policy and technology, has initiated a multi-phase project to develop the OSA methodology as part of the overall safety assurance framework. The first phase of the project involved the development of the metrics to be used to quantify safety, including a literature review of existing metrics such as traditional surrogate safety measures used commonly in traffic engineering to analyze roadway use and other more recently proposed metrics that are more specific to AVs [1]. Next, an algorithm was developed to measure the parameters needed to quantify the proposed safety metrics from existing infrastructure in the form of traffic cameras [2]; temporary sensors were utilized to understand the need for different collection modalities. Data collection was performed to evaluate the performance of the algorithm and assess the results in the real physical environment. Parallel work was conducted to evaluate the use of these metrics in simulation utilizing various driving simulation software [3], [4]. The metrics project was expanded into additional phases which included the further refinement of these metrics and much of the efforts discussed throughout this dissertation.

In the context of this work, it is first important to understand the four key components of an overall safety case for the development of AVs: 1) Test Methods, 2) Metrics, 3) Evaluation Methodologies, and 4) Evaluation Criteria. Test methods are first needed to establish consistent testing for AVs, most likely including some combination of simulation, closed course testing, and public road testing with a proper balance of fidelity for validation and efficiency of testing. To optimize these testing efforts, a set of metrics should be established to measure important data during testing which are capable of providing insight into the successful operation of the vehicle. Once testing methods have been determined to collect the necessary data for a set of defined metrics, an evaluation methodology must be put in place to interpret the measured metrics' values and determine the level of success the vehicle has achieved for the designed scenario. The evaluation methodology piece will be the center of the proposed work, leveraging the previous work conducted by the IAM surrounding OSA metrics and test methods. Finally, the evaluation criteria component will rely on some level of best practices, standards, and eventually, legislation to truly answer the questions "how safe is the vehicle?" and "how safe is safe enough?".

Beyond the scope of this research, an overarching safety case could be constructed to demonstrate the safe operation of a vehicle prior to deployment on public roads. This safety case would extend beyond the scope of the OSA methodology by utilizing the described methodology to contribute to the display of safety, but also utilizing a combination of methods listed in Figure 1. While the OSA methodology quantifies safety for a single scenario, the safety case would require a demonstration of safety throughout the operational design domain (ODD) of the vehicle. The three legged stool in the below figure utilizes known design tools and strategies to methodically establish safety consisting of: 1) *Safety Management System and Safety Culture*; 2) *Scenario-Based Testing*; and 3) *Design and Development Methods* [5]. The scope of this research focuses on the quantification of the Scenario-Based Testing to construct the OSA methodology approach, although elements of the Design and Development and Safety Management System categories are important considerations to incorporate as the research progresses.

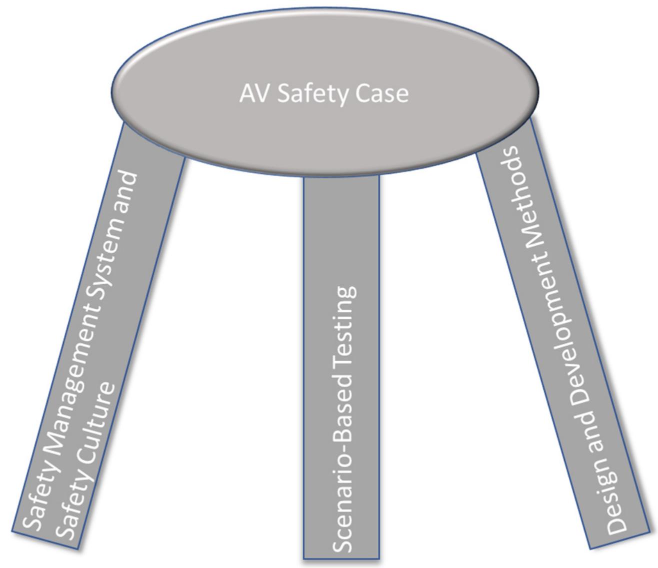


Figure 1. Possible Construction of a Safety Assurance Case [5]

The focus of the dissertation herein encompasses the research, development, and testing of the OSA methodology. Validation and refinement will be left for future work as these activities will require a significant amount of data for a variety of ODDs. The goal of this research is to provide the foundation for the framework of the OSA methodology that could be further developed into an evaluation tool for AV performance to facilitate original equipment manufacturers (OEMs) in test evaluations, assist the government and SDOs in the development of standards and regulations surrounding AV operations, and build trust with the public by increasing the transparency of AV performance prior to and during deployment on public roadways.

1. OVERVIEW OF AV EVALUATION TECHNIQUES

1.1 Review of Existing Safety Metrics for Vehicle Evaluations

In Phase I of the IAM metrics project, a list of proposed safety metrics was compiled based on the results of the review of over 50 published papers. The complete list of safety metrics identified in the literature review as well as a detailed formulation of the proposed metrics were published in [1]. As previously discussed, the safety evaluation of AVs will likely depend on a combination of simulation, closed-course, and on-road testing techniques; however, in order for this testing to be meaningful, operational safety must be quantified through metrics such as the proposed list, to evaluate vehicle performance in any given test scenario. This quantification also should be universal for comparison across AV platforms and even to human driven vehicles (for those applicable). As such, the OSA methodology seeks to establish a procedure for interpreting these metrics. A complete list of the OSA metric formulations from [1] is presented in Table 1.

Table 1. OSA Metric Formulations (Source: [1])

Minimum Safety Envelope (MSE)[3]	$d_{min}^{long} = \left[v_1^{long}\rho_1 + \frac{1}{2} a_{1,max,accel}^{long}\rho_1^2 + \frac{\left(v_1^{long} + \rho_1 a_{1,max,accel}^{long}\right)^2}{2a_{1,min,decel}^{long}} - \frac{\left(v_2^{long}\right)^2}{2a_{2,max,decel}^{long}} \right]$ (1) $d_{min}^{lat} = \mu + \left[\frac{2v_1^{lat} + \rho_1 a_{1,max,accel}^{lat}}{2}\rho_1 + \frac{\left(v_1^{lat} + \rho_1 a_{1,max,accel}^{lat}\right)^2}{2a_{1,min,decel}^{lat}} - \left(\frac{2v_2^{lat} - \rho_2 a_{2,max,accel}^{lat}}{2}\rho_2 - \frac{\left(v_2^{lat} - \rho_2 a_{2,max,accel}^{lat}\right)^2}{2a_{2,min,decel}^{lat}} \right) \right]$ (2)

[3] This metric was defined as the Minimum Safe Distance (MSD) in the original reference but has since been updated to the Minimum Safety Envelope (MSE) and other related metrics have been updated accordingly.

Proper Response (PR)	$PR = \begin{cases} 1 & if\ MSDV' = 1 \wedge \begin{pmatrix} a^{lat} \in [a^{lat}_{min,accel}, a^{lat}_{max,accel}] \vee \\ a^{long} \in [a^{long}_{min,accel}, a^{long}_{max,accel}] \end{pmatrix} \\ 0 & else \end{cases}$	(3)
Minimum Safety Envelope Factor (MSEF)	$MSEF^{lat} = \frac{d^{lat}}{d^{lat}_{min}}$ $MSEF^{long} = \frac{d^{long}}{d^{long}_{min}}$	(4) (5)
Minimum Safety Envelope Calculation Error (MSECE)	$MSECE_{long} = \frac{\lvert d^{long}_{gt,min} - d^{long}_{min} \rvert}{d^{long}_{gt,min}}$ $MSECE_{lat} = \frac{\lvert d^{lat}_{gt,min} - d^{lat}_{min} \rvert}{d^{lat}_{gt,min}}$ $MSECE = \sqrt{MSECE^2_{long} + MSECE^2_{lat}}$	(6) (7) (8)
Collision Incident (CI)	$CI = \begin{cases} 1 & if\ d^{lat} = 0 \wedge d^{long} = 0 \\ 0 & else \end{cases}$	(9)
Traffic Law Violation (TLV)[4]	$TLV = \begin{cases} 1 & if\ Law\ Violated \\ 0 & else \end{cases}$	(10)
Distance to Stop (Dstop)	$Dstop = \frac{v^{long}_1}{2a^{long}_{1,min,decel}}$	(11)
Time to Collision (TTC)	$TTC = \frac{X_2 - X_1}{v^{long}_1 - v^{long}_2}$ (12)	
Modified Time to Collision (MTTC)	$\mathrm{MTTC} = \frac{-\Delta\vec{V} \pm \sqrt{\Delta\vec{V}^2 + 2\Delta\vec{A}\vec{D}}}{\Delta\vec{A}}$ (13)	
Post Encroachment Time (PET)	$PET = t_2 - t_1$ (14)	
Time Headway (THW)	$THW = \frac{X_2 - X_1}{v^{long}_1}$	(15)
Human Traffic Control	$HTCDER = \frac{GTI - CDI}{GTI}$ (16)	

[4] This metric was previously defined in [1] as Rules of the Road Violation (RRV); but was updated to Traffic Law Violation (TLV) at a later date to remove ambiguity.

Detection Error Rate (HTCDER)	
Human Traffic Control Violation Rate (HTCVR)	$HTCVR = \frac{CDI-CCI}{CDI}$ (17)
Predictable Acceleration (PA)[5]	$PA = \begin{cases} 1 & a^{lat} \geq \lvert 0.47g \rvert \vee a^{long,} \geq \lvert 0.43g \rvert \vee a^{long,decel} \geq \lvert 0.61g \rvert \\ 0 & else \end{cases}$ (18)
Achieved Behavioral Competency (ABC)	$\begin{cases} 1 & if\ yes \\ 0 & else \end{cases}$ (19)
ADS Active (ADSA)	$\begin{cases} 1 & if\ yes \\ 0 & else \end{cases}$ (20)
Measured Variables	$v_1^{long} = Follow\ vehicle\ longitudinal\ velocity$ $v_2^{long} = Lead\ vehicle\ longitudinal\ velocity$ $d^{long} = Longitudinal\ relative\ distance\ between\ vehicles$ $d^{lat} = Lateral\ relative\ distance\ between\ vehicles$ $a^{long} = Subject\ vehicle\ longitudinal\ acceleration$ $a^{lat} = Subject\ vehicle\ lateral\ acceleration$ $d_{gt,min}^{long} = Longitudinal\ ground\ truth\ relative\ distance\ between\ vehicles$ $d_{gt,min}^{lat} = Lateral\ ground\ truth\ relative\ distance\ between\ vehicles$ $X_1 = Follow\ vehicle\ global\ position$ $X_2 = Lead\ vehicle\ global\ position$ $t_1 = Follow\ vehicle\ time\ to\ reach\ target\ position$ $t_2 = Lead\ vehicle\ time\ to\ reach\ target\ position$ $GTI = Number\ of\ ground\ truth\ instructions$ $CDI = Number\ of\ correctly\ detected\ instructions$ $CCI = Number\ of\ correctly\ complied\ instructions$
Assumed Parameters	$\rho_1 = Follow\ vehicle\ reaction\ time$ $a_{1,max,accel}^{long} = Follow\ vehicle\ maximum\ acceleration$ $a_{1,min,decel}^{long} = Follow\ vehicle\ minimum\ deceleration$ $a_{2,max,decel}^{long} = Lead\ vehicle\ maximum\ deceleration$ $\mu = Lateral\ fluctuation\ margin$

[5] This metric was previously defined in [1] as Aggressive Driving (AD); but was updated to Predictable Acceleration (PA) for a more concise description.

The SAE On-Road Automated Driving (ORAD) Verification and Validation (V&V) Task Force is in the process of generating a recommended practice compiling a list of OSA metrics to quantify vehicle performance [6]. While the industry and corresponding documents are constantly in flux as the technology evolves, one way to portray the taxonomy for these metrics is broken down into five data classes as shown in Figure 2. The data source level identifies the data which must be supplied in order to calculate the associated metrics. Class 0 references data related to static scenario elements such as roadway markings and signage as well as dynamic elements such as VRUs. Class 1 includes data related to the pose and/or motion of the subject vehicle. Class 2 involves data originating in the ADS of the subject vehicle related to its status. Class 3 and 4 relates to data originating within the ADS with Class 3 referencing processed data such as the output of the perception module while Class 4 includes unprocessed raw data. Black Box metrics may be calculated without any access to the ADS of the vehicle, allowing for increased versatility for independent calculations; Grey Box metrics require information in the form of messages from the ADS, while White Box metrics require processed data from the ADS; and Clear Box metrics require raw, unprocessed information from the ADS. Since an actual ADS was not available for evaluation, the black box metrics are explored in detail throughout this document. At the classification level, the metrics can be broken down into safety envelope, behavioral, component, and ADS module related metrics. The safety envelope metrics describe the safe boundary surrounding the vehicle which may be calculated without any access to the ADS including metrics such as the TTC, MSE, and PET. The behavioral, component, and ADS module level metrics require some amount of ADS data consisting of metrics such as HTCVR, ABC, and ADSA.

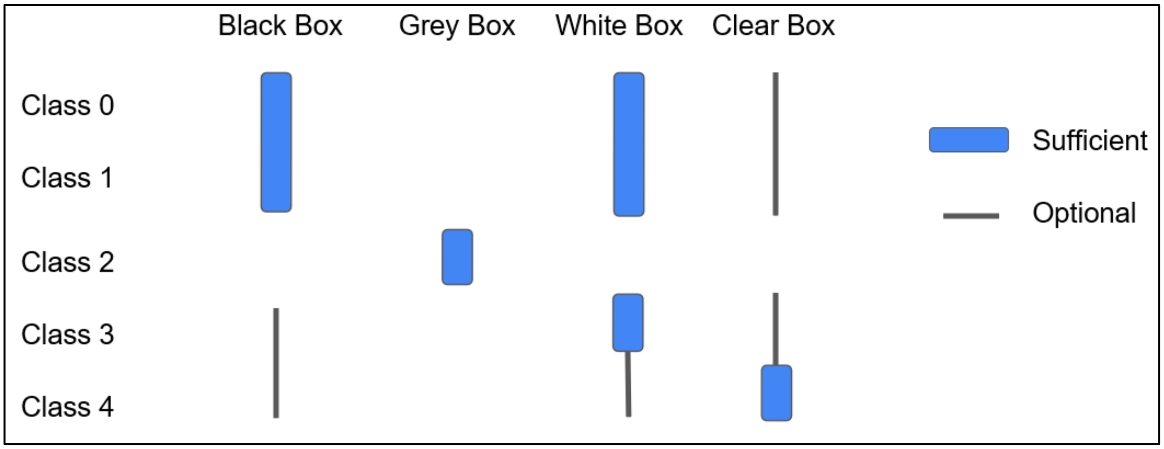


Figure 2. OSA Metrics Taxonomy (Source: [6])

The current list of OSA metrics being considered in this recommended practice are summarized in Table 2. This dissertation focuses on the Black Box metrics because an actual ADS was not available for evaluation. The Black Box metrics only require Class 1 and Class 2 data elements which were accessible for the evaluations presented herein.

Table 2. Current List of OSA Metrics Considered in Recommended Practice (Source: [6])

	Metric	**Class 0**	**Class 1**	**Class 2**	**Class 3**
Black Box	**Safety Envelope Metrics**	X	X		
	Collision Incident	X	X		
	Lane Stability	X	X		
	Traffic Law Violation	X	X		
	Predictable Acceleration		X		
Grey Box	**ADS DDT Execution**			X	
	Feature Level of Automation			X	
	Takeover Request to Fallback-Ready User			X	
White Box	**Achieved Behavioral Competency**	X	X		X
	Human Traffic Control Direction Identification Error Rate	X	X		X
	Compliance Error Rate to Human Traffic Controller Directions	X			X
	Safety Envelope Ratio	X	X		X
	ODD Recognition	X	X		X

An important aspect of the proposed safety metrics is the assumption of parameters needed for their calculation. To best characterize these parameters, naturalistic driving studies can be evaluated. Zhang *et al*. conducted a study which gathered naturalistic driving data by analyzing the effects of a dedicated AV lane while calculating surrogate safety measures such as time-to-collision (TTC), time-exposed TTC (TET), time-integrated TTC (TIT), rear-end crash risk index (RCRI), among others [7]. This study introduces an interesting approach to isolating AVs to operate within the same lane to enable the use of connectivity while understanding the impacts of a dedicated CAV lane on existing traffic patterns. Studies such as this one are important to help identify nominal driving performance parameters to be assumed in metrics calculations for AV safety evaluations. Additional studies have been conducted to calibrate the parameters of different metrics and even road designs using naturalistic driving data to better represent actual driving scenarios [8], [9], [10], [11]. By utilizing known driving data, important parameters for safety metric calculations can be better informed, helping to refine and improve the resolution of these calculations for specific scenarios. The more data which can be known in any given environment, the more accurate the assumptions can be surrounding the analyzed scenarios, thus providing higher fidelity when evaluating a vehicle using such data. Although the use of available data can help inform estimated driving behaviors, drivers do not always comply with "expected" driving performance and limitations in these predictions should be noted as such. Additionally, driving behaviors may vary drastically across ODDs and even in differing regions with similar ODDs. When utilizing naturalistic data to inform driving behavior estimations, it is important to consider all possible limitations.

In an approach attempting to characterize vehicle safety all within a single metric, Weng *et al*. proposed the Model Predictive Instantaneous Safety Metric (MPrISM) with the goal of measuring vehicle safety given any traffic snapshot in time considering the dynamic nature of vehicle safety [12]. This is an interesting approach which could be combined with the concepts presented in [13] in which a complete hazard analysis is conducted to assess the risks associated with an automated protected construction vehicle. Such an approach could be used to determine instantaneous safety of a given vehicle at any specific time on a system level. Although this approach could be implemented to help identify points of failure for the system, some level of scenario library generation must be utilized in the overall safety assessment in order to evaluate vehicle capability in the context different situations. As much of the literature repeats, the rarity of conflict events in general would require too much time and resources to validate the ability of an AV simply utilizing metrics such as the MPrISM, or any safety envelope related metric for that matter, for a given vehicle collecting miles; thus, relying on the supplementation of simulation for vehicle assessment.

Similar to the MPrISM, Mobileye has developed a principle of safety defined as Responsibility-Sensitive Safety (RSS) which seeks to computationally model safety for AVs based on the kinematic equations of motion [14]. In order to convey the safety model, Mobileye demonstrated how a vehicle following the principles of RSS would respond to scenarios contained within NHTSA's established list of 37 pre-crash scenarios. In a similar manner to that of the MPrISM, RSS attempts to formalize a model of safety by establishing safe following distances, avoiding dangerous situations, and enacting proper responses when necessary. Further studies have since been conducted to demonstrate the uses of RSS and refine the assumed parameters through the aforementioned technique of applying naturalistic driving data to calibrate the model in various situations and environments [11], [9], [10].

While many different approaches to the safety envelope-related metrics exist in the literature, [6] proposes characteristics of a good safety envelope metric. This list includes:

1. Definition needs to be transparent (i.e., not proprietary)
2. Allows for a good balance between safety and usefulness
3. Is comprehensive of the kinematic components of the dynamic driving task (DDT)
4. Is dynamic and adjusts based on the unfolding scenario
5. Useful in all scenarios within the ODD and in multiple ODDs
6. Needs to be implementable (i.e., doesn't require intensive compute, doesn't require proprietary IP or data (e.g., ADS data))
7. Minimizes False Positives (i.e., incorrectly indicates an unsafe situation) and False Negatives (i.e., fails to identify an unsafe situation)
8. Minimizes subjectivity

In 2019, Ren *et al*. explored another unique modeling approach which focuses on "empathy" for AVs. In this study, rather than using offline datasets to train vehicles to behave in a particular manner, this methodology introduces "social gracefulness" into the ego vehicle, allowing it to infer possible actions of surrounding vehicles [15]. One of the common critiques for AVs is that the social norms of driving are not easily achieved through the use of Deep Neural Networks (DNNs) and AI such as hand gesturing and drivers making eye contact to facilitate the expression of their intentions. This approach attempts to bridge the gap by integrating this "social gracefulness" into the driving model to better accommodate such driving behaviors. Although this concept is interesting, the approach is human factors driven and further exploration was outside of the scope of this dissertation.

In addition to the extensive literature that has been developed and reviewed surrounding operational safety metrics for AVs, the SAE International ORAD V&V task force is in the process of developing a recommended practice considering various OSA metrics. This task force consists of stakeholders from OEMs, government agencies, and academia and aims to establish a set of OSA metrics that may be used to understand the driving performance associated with an AV [6]. Among the discussions surrounding the OSA metrics, presentations and research have been shared highlighting key aspects such as risk models, good qualities for effective metrics, and an evaluation of different conflict indicators through naturalistic studies [16], [17], [18]. These presentations highlight important considerations for the OSA metrics which may be leveraged to comprehensively examine AVs and human driven vehicles alike, for any given scenario.

1.2 Testing Methodologies and OSA ODD

Once the review of OSA metrics and traditional surrogate safety measures was completed, another comprehensive literature review was conducted to understand the different methodologies being used to test operational safety in AVs. To accomplish this task, peer-reviewed articles, textbooks, and standards were reviewed pertaining to library scenario generation; safety assessments utilizing closed-course testing, public road testing, and simulation; industry standards; and even safety assessments from other industries (e.g., aerospace, locomotion, etc.). Exploring contributions in both the AV industry and other prominent industries requiring high levels of confidence such as aerospace provides context for work currently being performed but also stems insight into how the AV industry can utilize existing techniques to adequately solve the challenging issue of operational safety assurance. Examination of other industries provided useful insights for hazard approaches in other areas; however, all of the industries reviewed demonstrated much simpler ODDs than what could be faced by an AV. While others are considering methods for evaluating AVs, the OSA methodology is unique in that it is a comprehensive approach to establishing a unified assessment for AVs in any given scenario. The development of this methodology is a major step towards proving operational safety as AVs are developed and providing a groundwork to facilitate regulation across the industry. This work aims to bridge a major gap in the industry to connect the aspects of metrics measurements and interpretation of such metrics to quantify performance. Furthermore, other important aspects of the OSA methodology are introduced to connect the performance metrics to the fidelity of the testing by which these data were captured in addition to the relevance and complexity of the tested scenarios, thereby creating a complete assessment of a vehicle in a given scenario.

Comparable to the concept that an AV may only be designed for a specific ODD, certain metrics may only be useful for certain scenarios. Similarly, specific information may be unavailable for certain metrics calculations in some cases. These limitations should be considered in the selection of metrics for use in the OSA methodology as well as the feasibility of collecting data elements within a reasonable measurement uncertainty (i.e., a determination based on accuracy, precision, and resolution of the measurement method).

Fewer all-inclusive metrics capable of evaluating an AV in any given scenario will help to simplify the OSA formulation for a vehicle; however, sufficient metrics must be included to ensure completeness and robustness of the vehicle evaluation. Techniques reviewed in the literature associated with scenario library generation were applied to the assessment of the robustness of the OSA for the determination of scenarios; although, future validation efforts may introduce scenarios that have not been encompassed within the proposed methodology and it may be refined to expand the range for evaluation. Once the OSA methodology was formulated, it was tested and refined through similar methods to those explained in this section.

Feng *et al.*, explored a process for scenario library generation, using four key constituent components including: scenario description, metric design, library generation, and CAV evaluation tasks [19]. In a follow-on paper, Feng *et al.* applied the proposed scenario library generation methodology to specific case studies including a cut-in scenario, highway exit scenario, and car following scenario [20]. The difficulty in evaluating AVs from a scenario generation perspective is the challenge of capturing all possible scenarios which, as mentioned previously, is not feasible. An important factor when considering scenario generation is the evaluation of corner, edge, and long tail cases which are in the process of being defined by the SAE International taxonomy and definitions document for terms related to V&V of AVs [21]. Although it is important to ensure difficult scenarios are tested, Feng *et al.* discuss the consideration of not overweighting these instances to the point where AVs are always being designed for worst-case scenarios which can begin to degrade the comfort, drivability, and usefulness of the vehicle.

Fremont *et al.* took Feng's scenario library generation approach a step further by creating scenarios on a closed-course test track to evaluate AV performance. They utilized the GoMentum test track in California to conduct testing of scenarios which were also simulated [22]. This technique implemented an intriguing solution to what Ridemaier *et al*. consider to be one of the major gaps in the research which is that of the microscopic versus macroscopic assessments [23]. The project carried out by Fremont *et al*. involved generating a digital twin of the GoMentum facility so as to validate the fidelity of the simulation environment. Combining this approach with the statistical methodology of Riedmaier *et al*. may potentially bridge the gap between the microscopic scenario evaluation and macroscopic operational safety assessment. Riedmaier *et al*. explored a statistical higher-order model which considers the use of rapidly exploring random trees (RRTs) in order to assess scenarios with greater computational efficiency [23]. The combination of these approaches could facilitate the rapid generation of numerous scenarios that can be validated through real-world testing of a much smaller subset improving the efficiency of testing while maintaining the fidelity of the results. This approach is comparable to an automated iteration technique detailed in Chapter 6 in which the CARLA simulation was utilized to conduct a sensitivity study to changing variables from a baseline scenario. Saigol *et al*. took a similar approach in a project which utilized a digital twin of a proving ground to test sensors in conjunction with simulation of adaptive scenarios with pass/fail criteria [24]. Rather than requiring the high-fidelity environment needed to validate sensors within simulation, this project utilized physical testing for sensor validation while relying on simulation for efficiently varying parameters for rapid scenario generation. One of the difficulties with this method lies in the fact that the sensor validation and decision-making are not being performed at the same time since the simulation model does not rely on a photorealistic environment, thus perfect sensor function must be assumed for the simulation while perfect decision-making must be assumed for the physical testing of sensors.

Conversely, the method discussed by Pinter and Engelstein detailing the work of AImotive discusses how simulation can be used as a technique to evaluate AVs efficiently in many different scenarios. In this work, they stress the importance of photorealistic environmental data to adequately capture the detail required for a validated assessment [25]. Although it is important to simulate the vehicles and environment (including actor appearances and movements, lighting, signage and marking conspicuity, etc.), the use of previously mentioned techniques could potentially reduce the need for such a high-fidelity simulation environment which can be computationally expensive and laborious. In their discussions of photorealistic simulations, Pinter and Engelstein highlight the need to utilize lessons learned from other industries such as the advanced simulation techniques found in the aerospace industry. The importance of this research as it relates back to the OSA methodology work is in the quantification of fidelity. The highest fidelity is achieved through demonstration of the ability to replicate real-world scenarios with high accuracy; however, sensitivity to differences should be quantified in order to define the fidelity as it relates to the scenario outcome.

Although photorealistic simulation environments can be useful in improving the efficiency of testing through simulation, Tuncali *et al*. discuss the importance in validating simulation to ensure accuracy of assessments. This work utilized falsification and simulated annealing to efficiently detect issues within the simulation and locate corner cases [26]. Much of the current research is considering simulation approaches utilizing machine learning (ML) techniques; however, without proper validation, the results of simulation are meaningless. This concept ties in directly with the fidelity aspect of the OSA methodology discussed in Chapter 8. In a similar way which corner cases exist for vehicles, corner cases exist for simulation where ML struggles to accurately evaluate components such as sensors, perception systems, etc. As such, it is important to employ the strengths of simulation to increase the efficiency of testing while utilizing the ground truth, physical evaluations of hardware where necessary to validate the accuracy of the results. A combination of approaches can be achieved through augmented reality (AR) as is being done at Mcity's test track, yet the challenges of verifying and validating the test results remain important for such testing [27].

Earlier in 2020, Wen *et al*. proposed a unique methodology within the scenario library generation approach to incorporate convoluted neural networks (CNNs) to choose agents to interact with the ego vehicle and each other including other vehicles, pedestrians, and animals at scenario specific nodes [28]. By incorporating CNNs to choose agents based on positions and trajectories within the environment, this approach can efficiently simulate a wide variety of scenarios. The usefulness of this approach is depicted within the ability of AI to generate varying scenarios rather than the more tedious and expensive traditional technique of manually generating scenarios. This addition to the scenario generation pipeline could improve the efficiency of scenario generation, although it is important to consider the ability of AI to generate both relevant scenarios as well as corner and edge cases.

1.3 Safety Assessment Methodologies

Once the methodology for testing the operational safety of an AV is decided, whether that includes closed-course testing, public road testing, simulation, or likely, a combination of the three, the operational safety still must be quantified in a consistent manner. Kusano and Gabler discuss utilization of three major government crash databases for scenario library generation choosing from the Fatality Analysis Reporting System (FARS), General Estimates System (GES), and National Automotive Sampling System (NASS) Crashworthiness Data System (CDS) databases in which severity is denoted by the Abbreviated Injury Scale (AIS) [29]. The difficulty with evaluation of these databases is the subjectivity in how some quantifications such as change in velocity (delta-v), crush energy, injury potential, and the like are determined based on the amount of available data. When utilizing these types of sources for evaluations of severity, crash type, and other naturalistic data driven parameters, the subjective nature and reliance on expert analysis to interpret these data are important considerations. For this reason, Wardzinski proposed four risk levels attempting to unify this somewhat subjective process. These risk levels ranged from RL-A with one other vehicle within a specified distance to RL-D being an accident at the highest risk level [30]. Pending any type of federal standards or regulations, research continues to explore the most effective way to consistently quantify operational safety throughout the industry. Other studies have also utilized the NASS CDS database for Event Data Recorder (EDR) information providing context of vehicle dynamics information relating to actual collisions. Since crashes are infrequent events (and waiting for a crash to occur is not a suitable or ethical method of evaluating AVs) historical crash data can be utilized to evaluate vehicle performance in a similar scenario. Historical EDR data related to crash events have been utilized to evaluate crash severity and driver behavior [31], [32]. More recently, studies have employed EDR data to directly evaluate AVs and the metrics used to assess their performance. For example, Waymo considered numerous fatal collisions for which historical crash data existed and compared the actual outcome with that of a simulated Waymo vehicle, demonstrating the collisions that could have been avoided by employing their driving model [33]. Como *et. al.*, use a similar approach to examine the safety envelope violation severity for vehicles involved in historical crashes by considering available pre-crash data [34]. These types of approaches allow for the testing and contextualization of the evaluation of the OSA metrics using existing datasets; however, each individual study tends to focus on a single metric such as the collision incident metric for [33] and the MSE for [34]. The uniqueness of the proposed OSA methodology is the use of multiple metrics along with their interpretation in the form of violation severities and contextualization within the fidelity of the test method and relevance and complexity of the scenario to provide a comprehensive operational safety assessment rather than focusing on a single metric which may lead to an incomplete or inconclusive analysis.

[35], [13], and [36] all explore various risk assessment techniques as applied to vehicles ranging from level 0 to level 5 according to the SAE International definitions. Stolte *et al*. provides a detailed hazard analysis and risk assessment (HARA) conducted by industry experts evaluating an L4 unmanned protected vehicle used on the shoulders of German highways in accordance with ISO 26262 [13]. Although this is one of few published articles containing an in-depth HARA for an AV, the ODD for this vehicle is extremely limited making this approach much more manageable than what would be required for a typical

L4 or L5 AV operating in complicated ODDs. In a separate article, Stolte *et al*. discuss this challenge proposing a system-theoretic accident model and processes (STAMP) methodology to evaluate AVs on a system level [36]. This approach is unique in that rather than focusing on specific scenarios as many researchers are, this work discusses the importance of thinking on a system level, ensuring that the actuation and behavior of the vehicle is appropriate in any given situation. Rather than detailing the scenario for which the vehicle must correctly maneuver, this paper details the potential failures the vehicle may face.

In addition to all of the innovative research being conducted by OEMs, simulation software companies, and academia, government organizations such as NHTSA are leading AV studies in contribution to industry developments. [37] demonstrates a list of 37 pre-crash scenarios statistically evaluated by the United States Department of Transportation (USDOT). This list was compiled utilizing information from the GES crash database in which all accidents were categorized and statistically reviewed. As previously mentioned, it is not possible to evaluate an infinite number of scenarios for AVs and therefore, information pertaining to the relevance and severity of known scenarios provides great insight for this work. In a similar approach, NHTSA conducted severity analyses in a more detailed manner relating to the injury severity and corresponding injury mechanism for occupants in the context of the pre-crash scenario, category of impact, and restraint usage by the vehicle occupants [38]. In a more recent report, NHTSA evaluated the ways in which simulation will be a crucial aspect to the operational safety assessment of AVs in conjunction with physical testing [39]. This work included a literature review of existing projects which have previously used simulation, standards which have guided simulation implementation, and fidelity of simulation results. Although it is commonly agreed upon that some level of physical testing is required to prove safety for AVs, it has similarly been recognized that simulation will play a role in efficiently conducting a large variety of testing to help build a safety case in an assortment of scenarios.

Significant work has been conducted to date in efforts to test and prove operational safety of AVs as they continue to be deployed on public roads; however, there is not yet an established, unified methodology to measure and quantify vehicle operational safety for a given scenario. The scope of this dissertation is to propose such a methodology to ensure a consistent quantification of operational safety across vehicle platforms for any given scenario. This methodology focuses on metric violations and corresponding severity evaluations for interpretation of the metrics results. Additional discussion is incorporated to suggest ways in which the fidelity of testing as well as complexity and relevance of a scenario can be further explored to contribute a comprehensive assessment methodology.

2. OSA METHODOLOGY

2.1 OSA Methodology Motivation

While the concept of vehicle performance metrics is not novel, the interpretation of these metrics is less obvious. The incorporation of ML and AI in vehicles makes the automation of tasks possible, but also adds complexity when trying to test the performance of such a vehicle due to its ability to “learn” how to pass a standardized test. For this reason, the task of testing AVs prior to deployment on roadways is both important and challenging. The intent of the OSA methodology is to provide a solution to establishing a baseline approach to evaluating the performance of a vehicle, either human-driven or automated, across scenarios in a consistent manner allowing for direct comparison. With such information, regulatory and standards bodies may be given the insight needed to generate more standardized guidelines surrounding the implementation of AVs in the public domain; OEMs will receive feedback on the strengths and potential weaknesses of their vehicles; and the general public can rest assured that their safety is being considered whether they are riding in a robo-taxi or driving alongside one on public roads.

2.2 General Formulation

The proposed OSA methodology is composed of four primary components including the severity of a metric violation (if present), the complexity of the test scenario, the relevance of the scenario, and the fidelity of the test method. These four components are the foundation for the OSA methodology and are discussed at length throughout this dissertation. The general equation for the score of a vehicle in a given scenario generated by the OSA methodology is as follows [5]:

$$OSA\,Score = \left(Complexity * Relevance * Fidelity * \left(1 - \sum_{i} (Severity_i * Metric_i) \right) \right) x100\%$$

(1)

Where $OSA\,Score$ is the score of a vehicle evaluated through the OSA methodology for the specific scenario tested (0 to 100%), $Complexity$ is the level of difficulty of a given scenario attempted by the vehicle, $Relevance$ is the relative frequency which the vehicle is likely to experience the attempted scenario over the duration of its lifetime, $Fidelity$ is the confidence in the measurement of the outcome of a scenario, and $Severity_i$ is the criticality of a violation of $Metric_i$ for which a violation is present.

The process for calculating the *OSA Score* based on the proposed methodology is illustrated in Figure 3. The first level is comprised of collection of the necessary data elements to perform the metrics calculations. These data elements may be collected through a combination of closed-course, public road, and simulation-based testing. Based on the collection methods, measurement uncertainty for each data element should be considered within the calculation of the metrics. Upon calculation of the metrics for a scenario, violations may be identified, and a severity score may be assigned. At the scenario level, the complexity, relevance, and fidelity may be considered to generate a *Scenario Score*. The *Scenario Score* and metric violation results may then be combined according to the previously defined formula to generate an *OSA Score*.

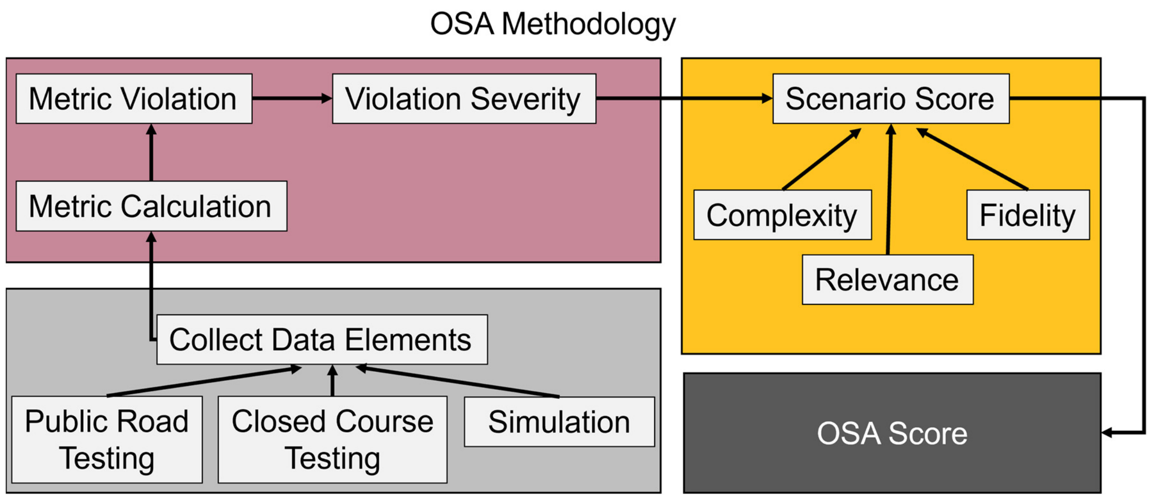


Figure 3. Workflow for Calculating OSA Methodology

2.3 Methodology for OSA Development

The components for the OSA formulation are discussed in detail throughout the remaining sections of this dissertation. First, the proposed OSA metrics are defined in detail in addition to the severity formulations. Next, the evaluation of the metrics is performed utilizing a combination of simulation data and real-world data. A discussion of the results obtained for the metrics calculations of these datasets is included in addition to sensitivity studies and comparative analyses. Lastly, the scenario level components of the fidelity, relevance, and complexity are reviewed and suggested future work is presented to improve the construction of the OSA methodology framework.

3. OSA METHODOLOGY METRIC VIOLATION

3.1 Proposed Metric Violation Definitions

The purpose of the metrics proposed through the OSA methodology is to quantify vehicle performance throughout different scenarios. These metrics have been assigned violation thresholds for the purposes of developing a formulation for the OSA score. While these thresholds have been developed based on research and literature, they may not meet the needs of specific evaluations for a vehicle. These thresholds are proposed in the current formulation of the OSA methodology and may be adapted over time as more testing is conducted and additional data are considered. One example of how the thresholds could be adapted is to identify specific categories of scenarios. For instance, the definition of a near-miss event could be used to formulate a minimum safety envelope trigger which identifies such an event.

As is highlighted in the background section of this document, there is a wide variety of metrics that have been researched and studied over the years throughout the traffic engineering industry; and, more recently, for AV applications. Many of these metrics provide similar evaluations with differing levels of detail based on the number of parameters used to define such metric equations. The RSS formulation of the safety envelope as researched and detailed thoroughly by Intel is one example of a safety envelope-related metric proposed in the OSA methodology based on the comprehensive consideration of the kinematic equations of motion. Variations of the safety envelope-related metrics include TTC, MTTC, PET, and THW amongst others. All of these safety envelope metrics rely on the fundamental equations of motion with varying levels of detail.

Additionally, the RSS based MSE considers a reaction time to account for realistic scenarios that do not follow idealized assumptions.

While it is important to identify which metrics should be considered when evaluating the performance of a vehicle, these metrics have little meaning unless a quantitative measure of success is established. This chapter will define what a violation means for each of the proposed metrics as a basis for the OSA methodology. The following sections break the metric classes down into two primary categories: (1) Behavioral Metrics and (2) Measurement Uncertainty Metrics.

3.2 Black Box Metric Definitions

The Black Box metrics can be evaluated without access to any information from the vehicle itself. These metrics evaluate the performance of the vehicle acting within a specified scenario including items such as applying the proper response to the encroachment of a safety envelope, driving in a predictable manner, and obeying traffic laws utilizing either onboard or offboard data sources. These metrics focus on the physical aspects of the vehicle rather than evaluating things like the hardware and software components responsible for vehicle perception for instance. As an example, consider a scenario in which a pedestrian crosses the street at a crosswalk in front of the subject vehicle. Assuming the subject vehicle obeys the relevant traffic laws but fails to identify the pedestrian and hits it, the Black Box metrics would identify violations such as a traffic law violation, proper response violation, and a collision incident; yet, these physical metrics would not indicate a failure of the perception system to identify the pedestrian in the first place which would require access to data contained within the ADS.

3.2.1 Safety Envelope-Related Metric Violations

The safety envelope metric for a vehicle can be expressed through a variety of traditional metrics such as TTC, MTTC, PET, and THW. These metrics are traditionally used in the traffic engineering industry as surrogate safety measures to evaluate roadway characteristics and promote safer designs. More recently, the MSE metric has been defined by Intel based on RSS as formulated by [40] for its comprehensive inclusion of the vehicle dynamics of both follow and lead vehicles in the determination of a safety envelope violation. The various safety envelope-related metrics are discussed in the following sections, including shortcomings and benefits of the different formulations.

3.2.1.1 MSE Violation

The MSE is one of the more comprehensive safety envelope-related metrics with a robust set of formulations for different scenarios as defined in [14] for NHTSA pre-crash scenarios. The MSE not only considers the current velocity of both vehicles but also assumes parameters for possible decelerations of the lead vehicle and a reaction time for the follow vehicle to respond to conservatively evaluate a given scenario. Additionally, the development of the MSE for a wide variety of scenarios ensures the metric can be applied to any situation while others are limited to specific events such as car-following scenarios. The disadvantage of this approach lies in the assumptions needed to calculate the MSE which may result in misinformation if the actual scenario does not follow the assumed parameters. The general formulations for the longitudinal and lateral components of the MSE violation (MSEV) are provided in Equations (2) and (3), respectively.

$$\boldsymbol{MSEV = 1\ if\ d_{closing,long} < d_{min}^{long} = \left[v_1^{long}\rho_1 + \frac{1}{2}a_{1,max,accel}^{long}\rho_1^2 + \frac{\left(v_1^{long}+\rho_1 a_{1,max,accel}^{long}\right)^2}{2a_{1,min,decel}^{long}} - \frac{\left(v_2^{long}\right)^2}{2a_{2,max,decel}^{long}}\right]} \quad (2)$$

$$\boldsymbol{and\ d_{closing,lat} < d_{min}^{lat} = \mu + \left[\frac{2v_1^{lat}+\rho_1 a_{1,max,accel}^{lat}}{2}\rho_1 + \frac{\left(v_1^{lat}+\rho_1 a_{1,max,accel}^{lat}\right)^2}{2a_{1,min,decel}^{lat}} - \left(\frac{2v_2^{lat}-\rho_2 a_{2,max,accel}^{lat}}{2}\rho_2 - \frac{\left(v_2^{lat}-\rho_2 a_{2,max,accel}^{lat}\right)^2}{2a_{2,min,decel}^{lat}}\right)\right]}$$

$$\boldsymbol{else\ MSEV = 0} \quad (3)$$

3.2.1.2 Proper Response (PR) Violation

The Proper Response metric evaluates the behavior of the vehicle when a safety envelope violation has occurred which may be the fault of the subject vehicle or another salient object. The safety envelope metrics consider the distance component of a safety envelope violation while the PR violation (PRV) is proposed to evaluate the reaction time of the subject vehicle. For consistency with the minimum safety enveloped-related metric violations, the same proper response assumed in the MSEV definition is proposed as the threshold for the PRV. Furthermore, a proper response does not just require the vehicle to apply a brake or steer maneuver; however, it requires a maneuver to be applied (to the best of the vehicle's ability) to reestablish the minimum safety envelope. As such, a successful PR is defined as the vehicle's application of a deceleration equal to or exceeding the Minimum Required Deceleration (MRD) for the subject vehicle or applying a steering input that results in reestablishing the MSE within one second of an MSEV taking place as defined in Equation (4).

$$\boldsymbol{if\ a_{subject}^{Long} < MRD\ and\ MSE = 1\ while\ MSEV_{Time} > \rho, PRV = 1}$$
$$\boldsymbol{else, PRV = 0}\ (4)$$

It should be noted that the MSEV may be the fault of the subject vehicle or the other salient object. In the case that the other salient object initiated the MSEV, a proper response negates the MSEV and there would not be a corresponding MSEV counted against the subject vehicle in the overall evaluation. In contrast, if the MSEV was the fault of the subject vehicle, a corresponding MSEV severity would be applied to the evaluation of the subject vehicle; however, a proper response would avoid an additional penalization of the PRV and corresponding severity assignment.

3.2.1.3 Collision Incident (CI) Violation

The Collision Incident (CI) metric simply identifies when a collision occurs with the subject vehicle. Similar to the MSEV, a collision may be the fault of the subject vehicle but could be that of another salient object. When evaluating an AV, one should differentiate these occurrences as it is possible, and even likely, that at times an AV will be in a situation where through no fault of its own, a collision will occur. It is expected that the available technology will help mitigate the severity of a collision in such case, but it may not always be possible to avoid (i.e., a driver cuts in front of the subject vehicle and brakes rapidly). While an AV may be designed to avoid a collision in reasonable and foreseeable scenarios, other salient objects do not always behave in a reasonable or foreseeable manner. When encountering such occurrences, the subject vehicle should not be penalized in the same manner as if it caused the collision (assuming the subject vehicle was otherwise behaving appropriately). Therefore, a CI violation (CIV) is segmented into two possible instances.

3.2.1.3.1 Subject Vehicle Fault Collision Incident

As the name implies, a Subject Vehicle Fault Collision Incident (SVF-CI) is a case in which the subject vehicle is responsible for the collision. Similar to traditional methods in which fault is designated for a collision, the fault responsibility will require additional evaluation and may not always be straightforward. A collision scenario may prompt investigation to decide the at fault party and may not be obvious in all circumstances. Accident reconstruction techniques may need to be employed to evaluate the collision and similar methodologies to the traditional sense may be used to resolve the collision incident typology. Regardless, a collision incident metric will require further evaluation in the context of the scenario; but, the indication of a violation is binary as shown in Equation (5).

$$\boldsymbol{CIV = 1\ if\ collision\ incident\ occurs\ and\ subject\ vehicle\ is\ at\ fault} \\ \boldsymbol{else, CIV = 0} \quad (5)$$

3.2.1.3.2 Other Entity-Fault Collision Incident

Conversely to the SVF-CI, the Other Entity Fault Collision Incident (OEF-CI) categorizes instances in which the subject vehicle experienced a collision; however, was deemed not at fault. For an OEF-CI to occur, another salient object needs to be determined “at fault” which could be the result of numerous factors such as violation of the traffic law metric.

An example of an OEF-CI would be a vehicle failing to stop at a red light resulting in the subject vehicle striking the other vehicle in an intersection and is represented as CIV* to denote the difference from a subject fault collision as shown in Equation (6).

$$\boldsymbol{CIV^{*} = 1\ if\ collision\ incident\ occurs\ and\ other\ entity\ is\ at\ fault}$$
$$\boldsymbol{else, CIV^{*} = 0} \quad (6)$$

In order to better visualize the relationship between the safety envelope metrics, proper response, and collision incident, the flow chart shown in Figure 4 was generated outlining the assignment of a violation and corresponding severity to be applied as defined in the next chapter. When the minimum safety envelope factor (MSEF) is greater than 1, there is no MSEV. Once an MSEV occurs, a determination must be made as to whether the violation was initiated by the subject vehicle or another salient object. If a PR is performed by the subject vehicle, the MSE is restored and only an MSEV is counted with a corresponding severity. When a PRV occurs, there is a possibility for a collision; however, it is also possible the subject vehicle has a delayed or interrupted response which results in a PRV but then reestablishes the MSE. If there is a delayed or interrupted response, the MSEV and PRV severities are attributed to the overall evaluation. The CI and associated severity would also be implemented if a collision occurs.

A separate path is defined for an MSEV which was initiated by another salient object, denoted as MSEV*. Similarly, the options exist for a PRV to occur or the subject vehicle to make a PR. The difference in the alternative branch comes at the next step for which there is a determination of whether or not the MSEV* presented an unrecoverable event. An example of a non-recoverable event is one in which the other salient object initiated an MSEV* with an MRD immediately exceeding the capability of the subject vehicle. In such an instance, an MSEV occurs, a PRV occurs, and a CIV occurs; however, the subject vehicle was not the initiator and was incapable of avoiding the collision from the start of the violation; thus, resulting in a non-foreseeable event. Conversely, if the event was recoverable, the flow chart follows the same progression as the subject induced MSEV. The primary difference of note between the paths is the lack of penalization for the scenarios in which the MSEV* leads to a CI* since the collision is not within the control of the subject vehicle.

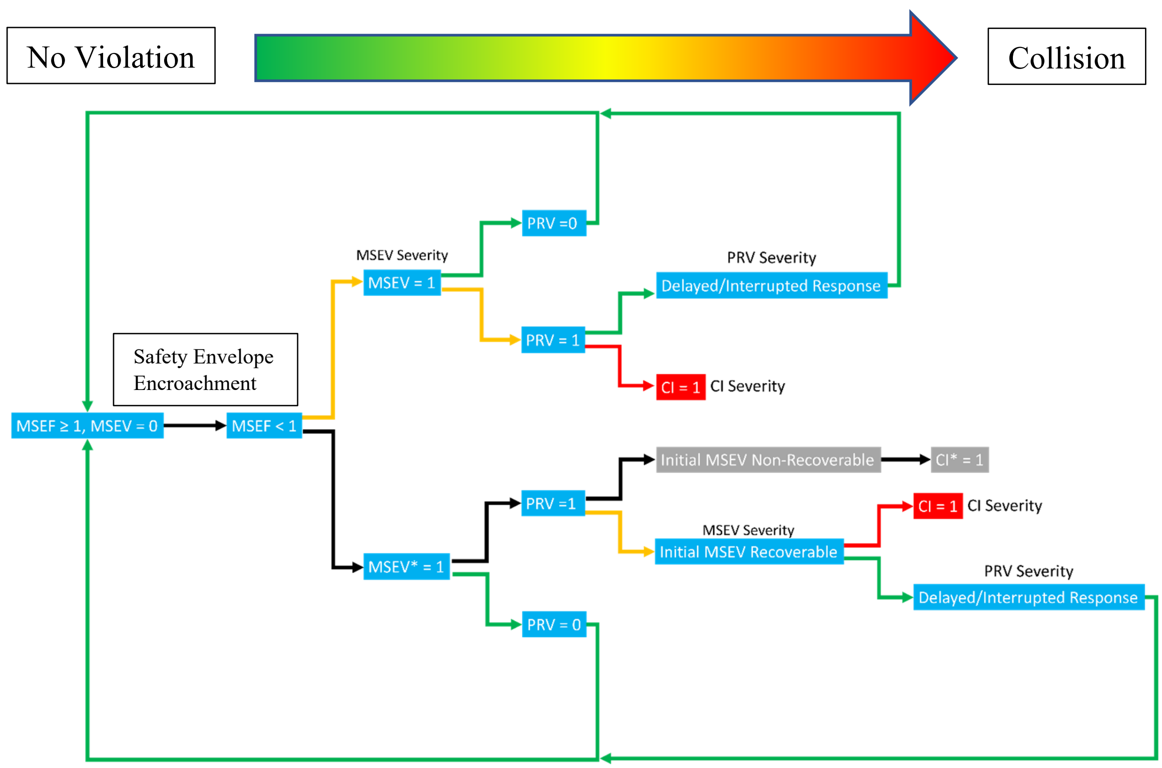


Figure 4. Flow Chart Defining Evaluation for an MSEV

3.2.1.4 Traffic Law Violation

Traffic laws help maintain order and structure on public roadways. There is a discrete number of traffic laws which are strictly defined for specific regions (i.e., ODDs); therefore, the traffic law violation metric has been defined to identify failure of the subject vehicle to obey established traffic laws. It should be noted that there are instances in which violating a traffic law is considered acceptable, and even required. An example scenario is one in which a construction zone requires a vehicle to cross over a solid yellow line and travel on the wrong side of the road. Just as human drivers are expected to obey all traffic laws, the same expectations will be established for AVs. These laws help to generate predictable behavior for drivers following a given set of rules and will help AVs predict the behavior of surrounding entities. Unless the subject vehicle is instructed to violate a traffic law (e.g., by a traffic cop), any other infraction would constitute a violation for this metric as expressed in Equation (7).

$$\boldsymbol{TLV = 1 \; if \; traffic \; law \; violated} \\ \boldsymbol{else, TLV = 0} \quad (7)$$

3.2.1.5 Predictable Acceleration (PA) Violation

While aggressive driving may not be considered “illegal”, it can lead to higher risk situations, relating to nominal driving performance characteristics of the vehicle. For example, a vehicle that accelerates and decelerates at high rates may be more susceptible to being rear-ended or rear-ending another vehicle. Thus, when the specified acceleration and deceleration (either longitudinal or lateral) thresholds are exceeded, a Predictable Acceleration Violation (PAV) is generated. The thresholds utilized in Equation (8) are examples of aggressive lateral and longitudinal accelerations as expressed in [41] and shown in Figure 5.

Event	Passenger Car (G)	Truck/Cube Van (G)	Heavy-Duty (G)
Harsh Braking	< −0.61	< −0.54	< −0.47
Hard Acceleration	> 0.43	> 0.34	> 0.29
Harsh Cornering	> 0.47 & < −0.47	> 0.4 & < −0.4	> 0.32 & < −0.32

Note: When viewing **Harsh Cornering**, negative values represent acceleration to the right and positive values represent acceleration to the left.

Figure 5. Example Reference for Harsh Acceleration Values [41]

The values displayed in Figure 5 are example values that could be employed for the OSA methodology framework; however, the exact values could be adapted. Varying thresholds could be used for the PA metric dependent on the vehicle type and specific use cases. For the purposes of this dissertation, these values were chosen to represent values above that of nominal driving and will be used for the evaluated scenarios in later sections.

$$\boldsymbol{PAV = 1\ if\ a_{Long} \geq 0.43g\ or\ a_{Long} \leq -0.61g\ or\ |a_{Lat}| > 0.47g}$$
$$\boldsymbol{else, PAV = 0}\ (8)$$

3.2.2 White Box Metric Definitions

The White Box metric definitions evaluate vehicle performance at a component level. Rather than evaluating a system level violation, the White Box metrics are related to the underlying components that may explain why a system level violation occurred. These metrics help define the resolution of the hardware and software components needed to ensure safety for the occupants and surrounding entities. In most cases, the White Box metrics will depend on the sensor technology employed within the AV. The limitations of the sensors should be considered through the evaluation of the White Box metrics to verify adequate safety factors are incorporated into the metrics calculations. Since these metrics are dependent on internal vehicle information, they require Class 3 data and cannot be achieved without some level of access to vehicle data.

3.2.2.1 Minimum Safety Envelope Calculation Error (MSECE) Violation

Rather than identifying whether or not a safety envelope violation occurred, the Minimum Safety Envelope Calculation Error (MSECE) evaluates the uncertainty in determining the MSE for the subject vehicle. For a high-level accuracy localization system, the vehicle will require less buffer to ensure avoidance of a conflict (e.g., a vehicle with 1cm error in the determination of the safety envelope needs a smaller buffer for uncertainty than a vehicle with 1m error). The measurement uncertainty can be thought of as a sphere around the subject vehicle in which it is possible that the vehicle inhabits at any given time. Unlike the behavioral metric definitions, specified thresholds are not identified for the measurement uncertainty metrics. The manufacturer should account for the uncertainty in their system through the implementation of their decision-making algorithms, although larger uncertainties may lead to less "useful" vehicles given larger required safety factors. A violation occurs when the actual distance (lateral or longitudinal) of the subject vehicle falls within the range of uncertainty for the MSE as follows in Equation (9):

$$\boldsymbol{MSECEV = 1, if\ d_{closing} \leq MSE + MSECE}$$
$$\boldsymbol{else, MSECEV = 0} \quad (9)$$

3.2.2.2 Achieved Behavioral Competency (ABC) Violation

The California PATH program identified 24 behavioral competencies for testing AVs to ensure they are capable of maneuvering in most generic situation types [42]. Additionally, the AVSC developed a best practice document for the evaluation of behavioral competencies for AVs [43]. These behavioral competency documents address the importance of establishing the ability of an AV to navigate common, generalized scenarios as a baseline. This metric focuses on intent as the subject vehicle may perfectly navigate a left turn through an intersection; however, if the instruction was given for the vehicle to make a right turn, the behavioral competency would be failed. Again, without knowing the intent of the vehicle, this metric cannot be evaluated and therefore requires data from the vehicle for the assessment. The achieved behavioral competency violation (ABCV) formulation is given in Equation (10).

$$\boldsymbol{ABCV = 1,}$$
$$\boldsymbol{if\ instructed\ behavior\ is\ not\ performed\ or\ performed\ incorrectly}$$
$$\boldsymbol{else, ABCV = 0} \quad (10)$$

3.2.2.3 Human Traffic Control Direction Identification Error Rate (HTCDIER) Violation

Human traffic controllers (HTCs) are commonly found in construction zones where lane closures demand the violation of a normal traffic law. Failure of an AV to identify and follow instruction from a HTC may result in undesirable behavior for the subject vehicle. HTCs are assumed to be synonymous to traffic laws as they serve the same purpose to ensure all road users are operating under the same set of rules. Failing to stop at a red light and failure to obey a traffic cop signaling the vehicle to stop would have similar results. The human traffic control direction identification error rate (HTCDIER) metric can be compared to the ABC metric with the behavioral competency being relayed by the HTC rather than the vehicle planning module. For evaluation purposes, it is necessary to understand whether the vehicle saw and understood the HTC for a given scenario or if it happened to be performing the maneuver signaled by the HTC but did not recognize the instruction, again requiring some level of access to the vehicle data. The HTCDIER violation (HTCDIERV) is expressed in Equation (11).

$$\boldsymbol{HTCDIERV = 1,}$$
$$\boldsymbol{if\ HTC\ not\ detected\ or\ HTC\ designated\ behavioral\ competency\ failed}$$
$$\boldsymbol{else, HTCDIERV = 0} \quad (11)$$

As demonstrated in this section, the behavioral metrics provide a means to evaluate the vehicle performance at a system level in the context of a given scenario, while the measurement uncertainty-related metrics focus on the ability of the hardware and software of the vehicle to perform at a component level. Both sets of metrics are important in establishing the overall performance of the vehicle; however, require different levels of data access. The behavioral metrics can be evaluated without having access to the ADS and as such, are the primary focus of this dissertation since an ADS was not accessible for evaluation. The measurement uncertainty metrics require varying levels of access to ADS information and are included for theoretical application.

4. METRIC VIOLATION SEVERITY

The proposed OSA metrics provide context to help identify behaviors by the subject vehicle that may be considered unsafe. Establishing a severity[6] for these violations goes a step further by better determining the extent of a hazard presented by such behaviors. More importantly, the severity formulation provides a more granular evaluation of performance across vehicle platforms for a given scenario. For example, for two vehicles experiencing an MSEV, a simple determination of a metric violation would indicate the same assumed performance for both vehicles. However, once a severity is assigned to the MSEV, a vehicle that exceeds the MSE by 1 m could be scored to perform better than a vehicle that exceeds the MSE by 10 m. In this chapter, the severity formulation for each metric violation will be detailed and explained.

4.1 MSEV Severity (Car-Following)

The MSE for a vehicle as referred to in this work is defined by RSS [40], [14]. When the MSE for a vehicle is encroached upon, a violation occurs. In order for an MSEV to occur, both the longitudinal and lateral components of the safety envelope must be violated. For instance, a vehicle driving parallel and adjacent to another vehicle would not necessarily experience an MSEV whereas a vehicle traveling within the longitudinal boundary that suddenly cuts into the path of the other vehicle would result in an MSEV. The violation severity has been formulated based on reasonable assumptions about the physical limitations of the involved vehicles. A reorganization of the MSE formulation and removal

[6] It should be noted that "severity" used throughout discusses the extent of a metric violation and does not necessarily correspond to injury risk or harm levels.

of the reaction time results in the minimum required deceleration (MRD) as defined in Equation (12):

$$\boldsymbol{MRD} = \left[\frac{\left(v_1^{long}\right)^2}{2*d_{current}^{long,same} + \frac{\left(v_2^{long}\right)^2}{2*n*a_{2,max,decel}^{long}}}\right] \quad (12)$$

Where v_1^{long} is the follow vehicle longitudinal velocity, $d_{current}^{long,same}$ is the minimum distance between the lead and follow vehicle, v_2^{long} is the lead vehicle longitudinal velocity, $a_{2,max,decel}^{long}$ is the maximum assumed longitudinal deceleration for the lead vehicle, and n is a percentage applied to the lead vehicle deceleration indicating a proportion of the maximum assumed deceleration for a given scenario (from 0 to 100%). This formulation allows for the consideration of not only the particular scenario being evaluated, but also more or less aggressive decelerations for the lead vehicle. The reason reaction time is not incorporated into the MRD formulation as it is in the MSE is because the MRD is the instantaneous quantification of the deceleration required to avoid a collision. If the reaction time of the vehicle was added to the formulation, the result would no longer be an instantaneous evaluation, but a projection based on assumed values. This robust formula may identify situations in which the follow vehicle would have avoided a collision regardless of the braking inputs for the lead vehicle versus cases in which a collision may have occurred had the lead vehicle braked harder. This removes the possibility of the vehicle under test (VUT) to be "lucky" in avoiding a collision simply because the other salient objects did not perform in a certain manner and facilitates the consideration of corner, edge, and long tail scenarios. The MSEV severity is based on the assumed physical limitations of the vehicle with 1 being the greatest severity and 0 being no violation as demonstrated by Equation (13). To normalize the various metric violation severities, 1 is the maximum value for a given severity. In this case, a severity of 1 corresponds to a scenario in which a collision will not be avoided unless an action is taken by the lead vehicle since the maximum capability of the following vehicle has already been reached.

$$\mathbf{MSEV_{SEV}} = \frac{\mathbf{Max(MRD)}}{\mathbf{decel_{capability}}}, \ \mathbf{Max(MSEV_{SEV}) = 1} \quad (13)$$

Although the MRD value is explicitly defined in Equation (12), the required braking levels have been broken down into four zones based on available literature as follows [34]:

- ***Low Braking Zone:*** 0 g – 0.34 g, corresponding to a range of deceleration values preceding the identified zones which are defined utilizing existing literature.
- ***Moderate Braking Zone:*** 0.35 g – 0.45 g, corresponding to a comfortable deceleration range at which drivers are capable of maintaining their lane of travel on wet roads according to the AASHTO “Green Book” [44].
- ***Reactionary Braking Zone:*** 0.46 g – 0.79 g, corresponding to a range at which most drivers decelerate when confronted with an unexpected obstacle (i.e., are able to react) [44].
- ***High Braking Zone:*** 0.80 and higher, corresponding to values inclusive and above 80% of the assumed maximum braking of the subject vehicle.

These braking zone definitions are utilized in later sections when illustrating the results of specific scenarios to serve as a visual aid.

4.2 MSEV Severity (Intersection)

While the fundamental concept of a MSEV for an intersection scenario is the same as a car-following scenario, an illustrative example is included in this section to convey the nuances of the implementation for this type of scenario. In the following examples, the trajectory overlap zone is defined as the intersecting region between the trajectories of two or more salient objects. If one of the salient objects never reaches the trajectory overlap zone, a collision will not occur. The various scenarios in which an intersection-based MSEV may occur are illustrated in Figure 6 through Figure 8 to explain this concept. In the first scenario, the subject vehicle (white) and other vehicle (red) are shown approaching an intersection at a perpendicular angle to one another. When the vehicle trajectories intersect and distance to trajectory overlap zone for both vehicles are exceeded, a MSEV has occurred. Since the avoidance of either of these two vehicles reaching the trajectory overlap zone will prevent a collision, the minimum required deceleration for the vehicles to stop prior to the trajectory overlap zone will be used to define the MSEV severity. Furthermore, if the longitudinal MSEV has not yet been reached, the vehicles are capable of safely stopping prior to reaching the trajectory overlap zone in the same manner as a car-following scenario and as such, would not result in a violation.

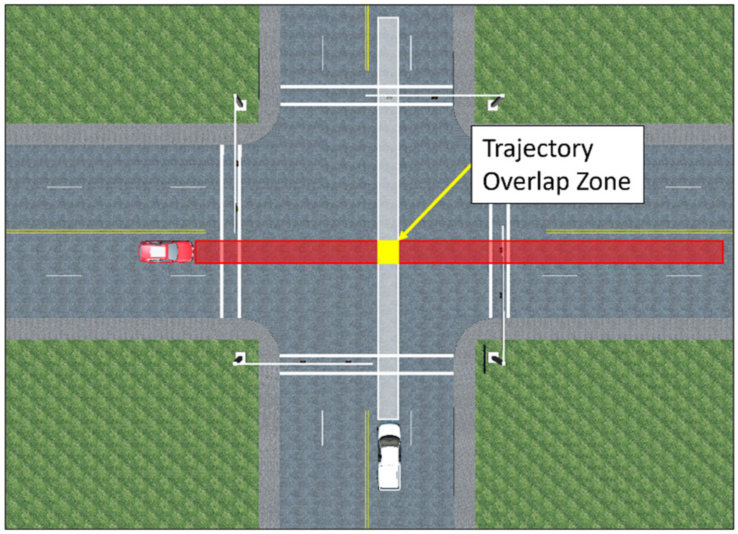


Figure 6. Intersection Scenario in Which the Vehicles Approach Perpendicular to One Another

Now consider the other vehicle initiates a left turn in front of the subject vehicle as depicted in Figure 7, or vice versa (Figure 8). In these scenario configurations, an extreme position may be reached at which the trajectory overlap zone will be defined as one vehicle or the other. As a result, the time to reach the trajectory overlap zone for that particular vehicle (i.e., the subject vehicle in Figure 7 and the other vehicle in Figure 8) is zero and the corresponding MRD goes to infinity.

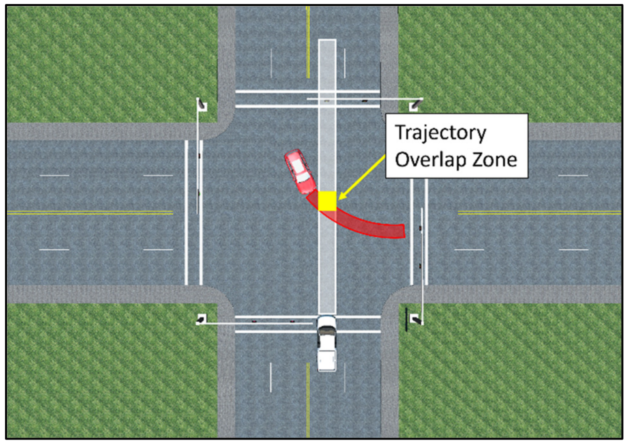


Figure 7. Intersection Scenario in Which the Other Vehicle (Red) Initiates a Left Turn in Front of the Subject Vehicle (White)

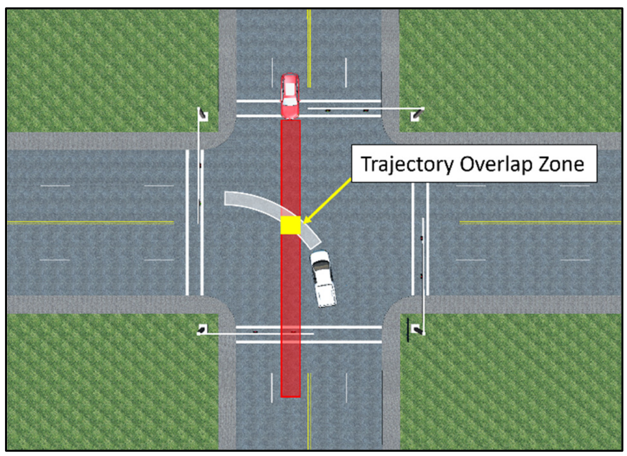


Figure 8. Intersection Scenario in Which the Subject Vehicle (White) Initiates a Left Turn in Front of the Other Vehicle (Red)

For this reason, the vehicle which has a greater separation distance to the trajectory overlap zone is considered for the MSEV severity formulation. Again, since the avoidance of the trajectory overlap zone for one of the vehicles will prevent a collision between the two vehicles and the MRD will be infinite when the trajectory overlap zone coincides with the boundary of one of the vehicles, using the MRD for the vehicle with the greater separation to the trajectory overlap zone is considered appropriate. Thus, the formulation for the MSEV severity for an intersection scenario is as follows:

$$MSEV_{SEV\_INT} = \frac{\frac{v_{long}^2}{2*d_{cz\_max}}}{decel_{capability}} \quad (14)$$

When considering the straight-line trajectory for a vehicle, an obvious limitation is the lack of attention to the lateral component of the vehicle dynamics as it relates to the future vehicle positions. Assuming straight line motion for either vehicle (especially for intersection-based scenarios) limits the ability to identify a potentially unsafe situation of one vehicle turning in front of another until the vehicles are close to impact. In order to accommodate such scenarios, the MSE was established by utilizing the lateral acceleration to project a more accurate trajectory for the vehicles.

4.3 PRV Severity

Once an MSEV occurs, a response is needed by the subject vehicle to avoid a PRV. The PRV severity is a function of the amount of time it takes for the subject vehicle to respond to an MSEV. While it is important for a vehicle to reestablish the safety envelope once a violation occurs, doing so too quickly may increase the likelihood of another collision due to unpredictable behaviors by the subject vehicle such as stopping suddenly or quickly changing lanes. Thus, the PRV severity is a function of the time it takes for the subject vehicle to respond to a safety envelope violation rather than the time it takes to regain the safety envelope. As discussed in the previous section, a PRV occurs when the subject vehicle does not achieve or exceed the necessary deceleration or steering within the assumed reaction time of the MSEV initiation. As was the case for an MSEV, a PRV severity of 1 or greater would imply an imminent collision unless an action is taken by the other vehicle to avoid the collision. This formulation is accomplished by taking the ratio of the actual response time of the subject vehicle to the current time to the trajectory overlap zone at the initiation of the PRV as defined in Equation (15):

$$\boldsymbol{PRV_{Sev}} = \frac{\textbf{\textit{Actual Response Time}}}{\textbf{Time to Collision Zone}}, \boldsymbol{if\,CI = 1, PRV_{Sev} = 1} \qquad (15)$$

4.4 CIV Severity

Collision severity has been a heavily studied topic in the fields of biomechanics and accident reconstruction for decades and is backed by numerous research studies. One study in particular was chosen to define the CIV severity due to its inclusion of many collisions which can be found in the NHTSA NASS CDS database, consideration of different levels of severity based on impact configuration (i.e., frontal, rear, or side impact), and generalized trend lines which can be applied to the OSA framework. As part of this study, the likelihood for a severe injury defined as a MAIS injury of 4+F, or severe to fatal injury was evaluated. The CIV severity is a function of delta-V, or change in velocity, for the involved vehicles. In cases where a large differential in mass ratio between the vehicles exists, the delta-V can be significantly higher for one vehicle over the other; therefore, to ensure the CIV metric accounts for any injury potentially resulting from a collision and not just those occupants within the subject vehicle, the more severe value is utilized.

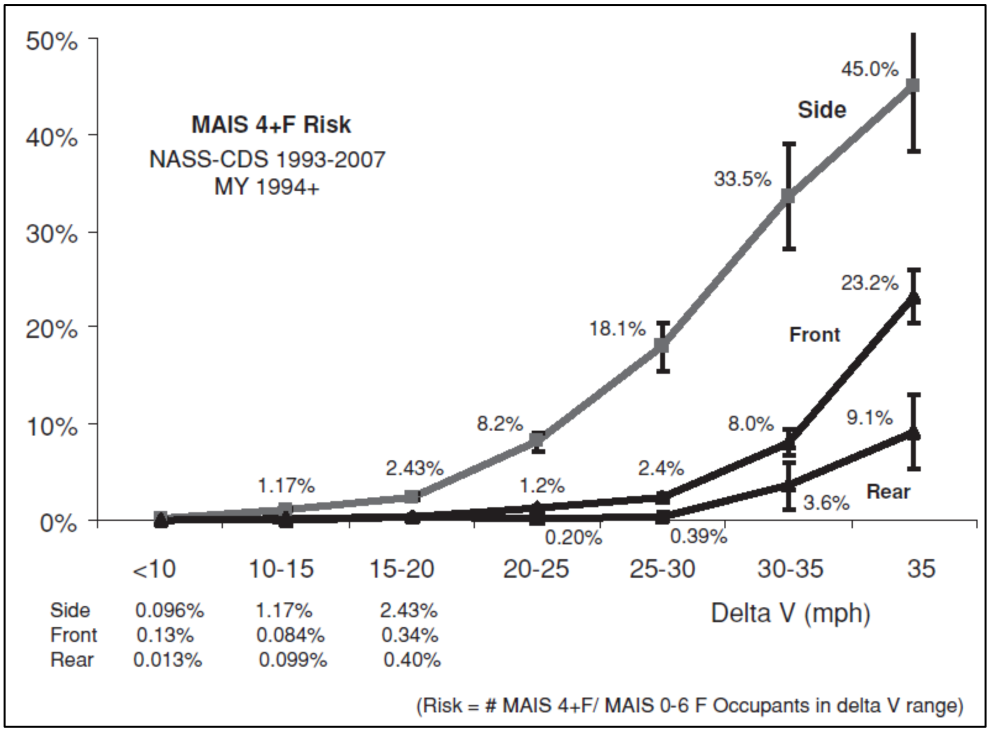


Figure 9. Risk of an MAIS 4+F Injury for NASS-CDS Study Conducted in [45]

Based on [45], the formulation for the CIV severities for side, frontal, and rear collisions are expressed in Equations (16), (17), and (18), respectively.

$$\textbf{Side Impact Severity}: \mathbf{CI_{Sev\_Side}} = \mathbf{0.1548e^{0.1784*DV}} \tag{16}$$

$$\textbf{Frontal Impact Severity}: \mathbf{CI_{Sev\_Front}} = \mathbf{0.0458e^{0.165*DV}} \tag{17}$$

$$\textbf{Rear Impact Severity}: \mathbf{CI_{Sev\_Rear}} = \mathbf{0.0137e^{0.1733*DV}} \tag{18}$$

These severity formulations represent trendlines generated from the referenced literature to predict an MAIS 4+F injury. The result of these equations gives the risk of an MAIS 4+F injury as a percentage.

This formulation of CIV severity does not explicitly account for a collision involving a vulnerable road user (VRU). As one would expect, the injury potential for a pedestrian or a cyclist as opposed to the occupant of another vehicle is significantly greater when considering similar delta-Vs. Given this increased injury potential for collisions involving VRUs, the severity for such a collision will be assigned the maximum severity of 1. As previously discussed, the severity formulations developed within this dissertation are example implementations and may be further refined or adapted in future work based on additional research and data. Furthermore, the formulations used here may not meet the needs of specific evaluations and could be adapted for other purposes.

4.5 Traffic Law Violation Severity

Traffic laws can vary depending on jurisdiction and the severity of a violation may vary drastically depending on the specific scenario. For instance, a vehicle rolling through a stop sign may be harmless when no other vehicles are around; however, if a pedestrian is crossing the road at the same time, the result could be fatal. Furthermore, traffic laws are in place to maintain order and structure on the roadways. Thus, a traffic law violation severity is not considered for the OSA methodology, rather the metric is proposed as a pass/fail situation in which case the vehicle receives a failing score for a traffic law violation similarly to how a student driver would fail a driving test for any traffic law violation. As previously discussed, there may be instances in which a traffic law violation is necessary (i.e., traffic cop directing vehicles to cross into an opposing lane for construction) and in such cases, there would be no metric violation severity since there would be no corresponding metric violation.

4.6 PAV Severity

As noted in the previous chapter, aggressive driving is not necessarily illegal, but may increase the risk of a collision. The PAV severity is defined as the percentage of time during the evaluation period for which the subject vehicle was driving aggressively according to the violation threshold. Additionally, the amount by which the violation threshold was exceeded is an important consideration for the severity. Minimally exceeding the threshold for the duration of the event may pose a lower risk than a vehicle pushing its physical limits even for a limited duration. Therefore, the severity of a violation is defined as the summation of violations for the duration of the event, normalized by the time interval for each evaluation period and multiplied by the ratio of the magnitude of the acceleration to the assumed vehicle limit. The longitudinal and lateral components of the PAV severity are shown in Equations (19) and (20), respectively.

$$\mathbf{PAV_{Long\_SEV}} = \sum_{i=1}^{n} \frac{\mathbf{PAV_{Long\_i}}}{\mathbf{Time}_{Interval}} * \frac{a_{\mathbf{Long}}}{a_{\mathbf{Long\_Limit}}} \quad (19)$$

$$\mathbf{PAV_{Lat\_SEV}} = \sum_{i=1}^{n} \frac{\mathbf{PAV_{Lat\_i}}}{\mathbf{Time}_{Interval}} * \frac{a_{\mathbf{Lat}}}{a_{\mathbf{Lat\_Limit}}} \quad (20)$$

4.7 Measurement Uncertainty Metric Violation Severity

A severity has not been defined for the measurement uncertainty metrics as alluded to in the introduction of the metric violations discussion. The reason for the lack of a severity formulation for these metrics is similar to that of the traffic law violation explanation in that a violation of these metrics is highly dependent on the specific scenario and difficult to generalize. The fundamental purpose for these metrics is to ensure a baseline behavior for the vehicle and as such, a violation triggered for one of these metrics is handled utilizing a pass/fail criteria. For instance, if an ABC of a vehicle making an unprotected left turn is failed, a severity of the ABCV would not be assigned. There may be instances where this failure would not be problematic (i.e., the subject vehicle failed to yield to an oncoming vehicle that stopped to avoid a collision) or it could result in other hazardous indicators (i.e., CIV). Therefore, a violation of a measurement uncertainty metric for the purposes of the OSA methodology proposed herein is denoted as a fail without a specific assignment of severity.

5. OSA METHODOLOGY EVALUATION SCRIPT

5.1 Overview

In order to evaluate the efficacy of the OSA methodology across a variety of scenarios, a MATLAB script was developed with the capability of calculating the metrics based on available data. White box metrics pertaining to vehicle automation systems such as whether the ADS was active for a scenario or measuring the performance of the perception module could not be evaluated without a physical or simulated ADS; therefore, the primary focus of this demonstration was on the previously defined behavioral metrics and does not include the measurement uncertainty metrics. The high-level goal of the MATLAB script generated for this evaluation was to develop a robust tool that would be capable of performing the necessary calculations to follow the OSA methodology, regardless of the data source. A basic intersection scenario is described throughout this section to illustrate the functionality of the MATLAB script.

5.2 Vehicle Polygons

The underlying observable variables for the majority of the behavioral metrics include position, velocity, and acceleration of the salient objects involved in the scenario. To accomplish this measurement, code was written to establish a polygon for each of the involved vehicles as depicted in Figure 10. As the vehicles move through the scenario, their positions are processed within the MATLAB script. While other VRUs such as pedestrians or cyclists are not shown, their implementation would be achieved simply by adjusting the extents of the polygon based on the length and width. Other adjustments would need to be incorporated based on the parameter assumptions for VRUs such that the velocity and acceleration limits of a cyclist vary drastically from that of a vehicle, and even further yet for a pedestrian. The velocities and accelerations of the involved vehicles are exported from the simulation to provide additional data related to the vehicle kinematics which could be measured for a real-world scenario through on-board or off-board sources as discussed in later sections. Additionally, the vehicle headings were exported and incorporated into the polygon calculation to establish the directionality of the vehicles which is an important factor in relating the context of various metrics as to the future path of the vehicles. This is represented in Figure 11 showing the two vehicles shortly before impact occurs within the intersection.

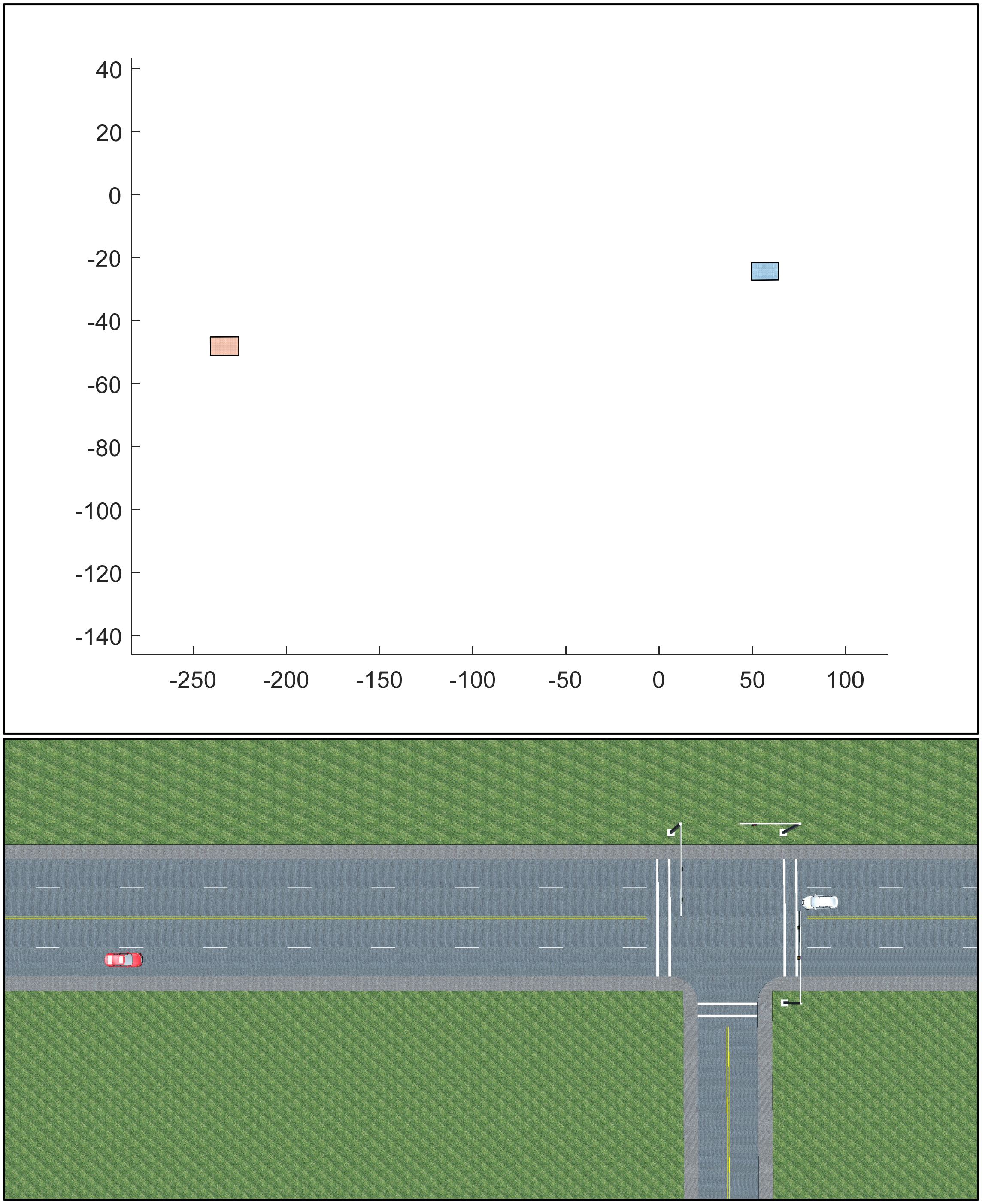
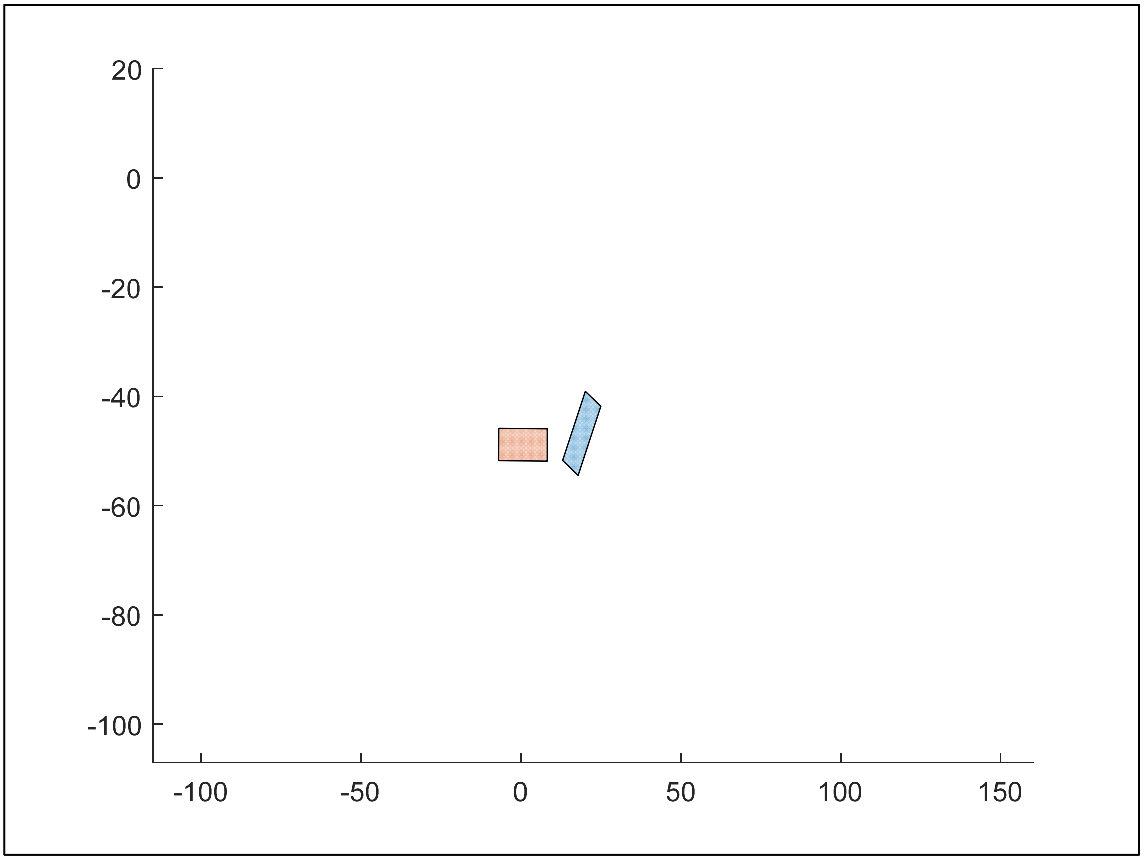


Figure 10. MATLAB-Generated Polygons Representing Vehicle Locations on 2D Map (Top) and HVE Simulation Depicting Corresponding Vehicle Locations (Bottom)



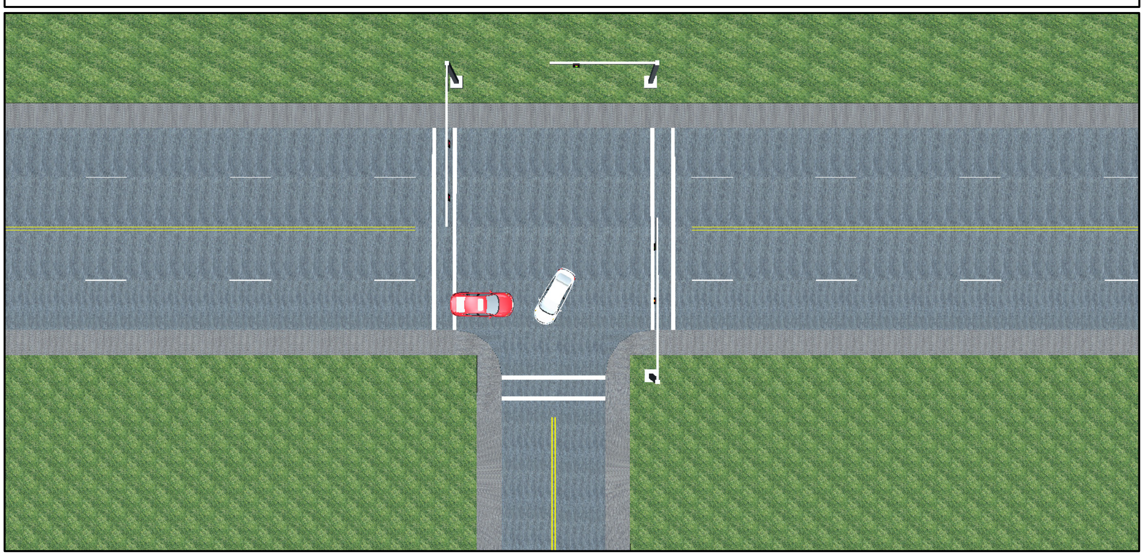


Figure 11. MATLAB-Generated Polygons Representing Vehicle Locations Just Before Collision (Top) and HVE Simulation Depicting Corresponding Vehicle Locations (Bottom)

5.3 Extended Vehicle Polygons

In addition to understanding the current position of the vehicles, many of the metrics require the context of the trajectory of the vehicles with respect to one another. For example, an MSEV is only applicable when both the longitudinal and lateral limits of the safety envelope are exceeded. In order to understand the intersection of vehicle trajectories, extended polygons were created as shown in Figure 12. To improve the model accuracy, curved extended polygons were generated based on the lateral acceleration of the vehicles for each timestamp within the scenario. When the vehicles are close to each other, the lateral component has minimal effect on evaluating the intersection point (i.e., the straight-line path of the vehicles intersect); however, the further away the vehicles are and the larger the lateral acceleration, the greater the effect on the trajectories with respect to one another. This phenomenon is illustrated in Figure 13.

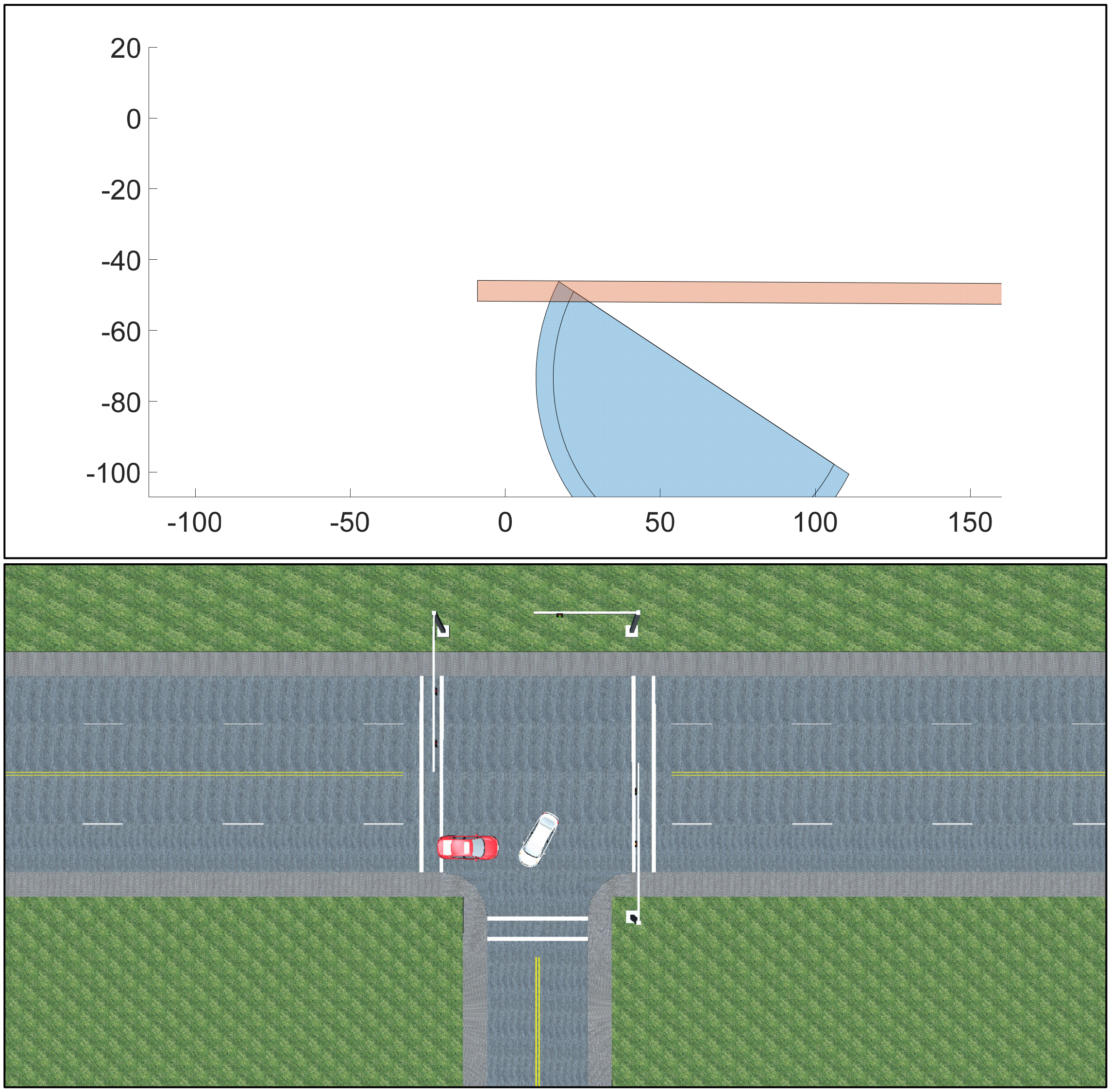


Figure 12. MATLAB-Generated, Extended Polygons Representing Vehicle Locations Just Before Collision (Top) and HVE Simulation Depicting Corresponding Vehicle Locations (Bottom)

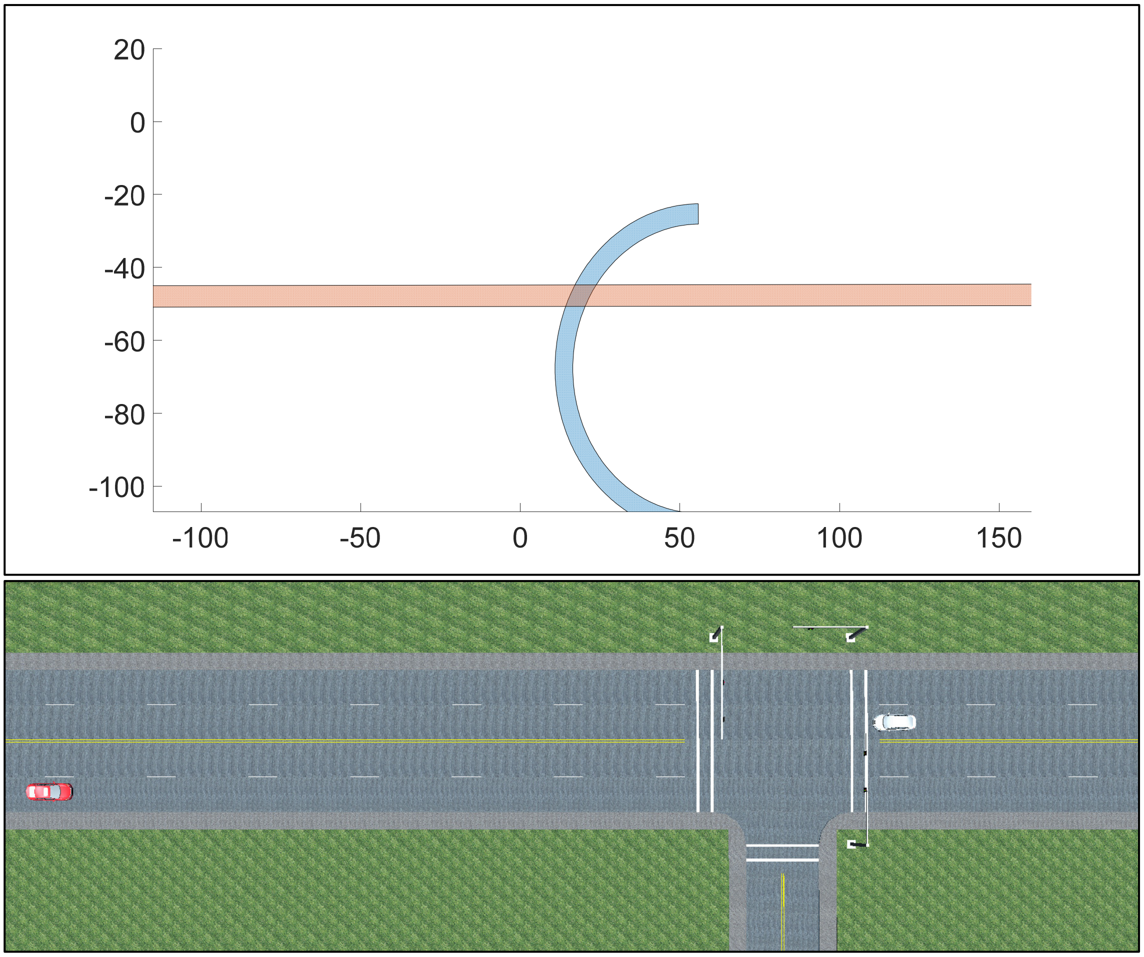


Figure 13. MATLAB-Generated Curved Extended Polygons Representing Vehicle Locations at Beginning of Lead Vehicle Left Turn (Top) and HVE Simulation Depicting Corresponding Vehicle Locations (Bottom)

5.4 Metric Calculations

Although the metric calculations are relatively straightforward as formulated in the previous sections, incorporating the logic into a robust code with the ability to evaluate a variety of scenarios was not so trivial. In this section, the logic is described for incorporating the OSA metric violation and severity evaluations into the MATLAB script. The primary focus in generating the script to evaluate the OSA metrics was to allow the inputs to be obtained from any given source including datasets such as real-world collected data, simulated events, or naturalistic driving databases. A generalized template was created to organize scenario inputs relating to the vehicle dynamics of an event which could be populated in a manual or automated manner from any given source.

5.4.1 Safety Envelope-Related Metrics

The safety envelope-related metrics are generally straightforward in their definitions, relying on varying levels of inputs, from the comprehensive equation of the MSE to the simplified formulation of TTC. The instantaneous vehicle dynamics for a given scenario are computed utilizing the aforementioned polygon function within MATLAB and the parameters for a safety envelope violation are determined based on the trajectory information defined by the extended polygons. To determine the minimum distance between the two polygons, the "Minimum distance between two polygons" MATLAB function was utilized [46]. By calculating the minimum distance between the polygons at any given time, the regions of the vehicle with the potential to first come into contact with one another provided a conservative estimate of a potential collision.

5.4.2 Proper Response Metric

The PR metric required implementation of one of the safety envelope-related metrics to trigger the requirement for a proper response. The MSE metric was chosen for its comprehensive formulation. Once a violation of the MSE occurs, the script calculates the timing for the subject vehicle to apply an appropriate level of braking or steering to avoid a collision based on the trajectory data. If the time required to complete the proper response exceeds the pre-defined threshold, or a collision occurs, the metric violation is triggered. The implementation of severity can then simply be determined by calculating the ratio of the timing of the actual proper response compared to the threshold value of a proper response. It should be noted that any of the safety envelope-related metrics could be used to define the PR metric by adjusting the corresponding lines of code to reference a different metric such as the TTC rather than the MSE.

5.4.3 Collision Incident Metric

Although a collision incident event is easily observable, the mathematical logic to define the severity of such an event is more complex. The initial observation of a collision was determined using the "Intersect" function within MATLAB for the defined vehicle polygons and the time of the collision was recorded. The severity was defined as a function of the maximum delta-V, or change in velocity, for the involved vehicles in addition to the generalized PDOF (i.e., frontal, rear, or side impact) according to the equations detailed in Section 4.4. To determine these values, the velocities of the vehicles at impact were recorded in addition to the separation velocities. In theory, these values are easy to record; however, in many cases, the vehicle polygons did not completely separate following the collision, resulting in no change of the "Intersect" function to indicate a separation velocity for a severity evaluation. To accommodate these scenarios, the separation velocity was defined as either the time at which the vehicle polygons no longer intersected, or the time at which the vehicle accelerations dropped below 1 g of acceleration (whichever occurred first) to generalize a point in time at which the crash pulse was ending. Additionally, the directionality of the collision was required for the severity calculation. To accomplish this, the velocity change direction was evaluated for each vehicle based on the lateral and longitudinal components of the velocities in the vehicle coordinate system. For example, if the magnitude of the lateral velocity change for the vehicle was greater than the longitudinal change in velocity, the collision was defined as a side impact.

5.4.4 Predictable Acceleration Metric

The PA metric was straightforward to implement in the script, evaluated based on the condition being achieved in which the longitudinal or lateral accelerations of the vehicles exceeded the pre-determined threshold. The summation of these events was accomplished by determining whether there was a violation at each timestamp and relating the corresponding lateral or longitudinal acceleration at that time compared to the assumed limit of the vehicle.

5.5 MATLAB Results

Developing the MATLAB script in a robust format that worked for any given scenario such as car-following, intersection, and lane change scenarios required numerous iterations as the logical implementation was not as straightforward as the formulations themselves. Once the evaluations were accomplished, several graphical outputs were automated to illustrate the results. These plots included the MSEV and PRVs with respect to the context of the scenario including vehicle velocities and relative positions, instantaneous MRD values, and a comparison of safety envelope-related metric violations. An indication of the occurrence of a collision was also added to the plots to provide context for rapidly and potentially unexpected changes in the vehicle dynamics resulting from a collision rather than driver inputs. Sample plots for the simulated scenarios are depicted in the following sections. Additionally, the severities of the various metric violations were calculated and stored in the MATLAB script to be reported for any given scenario.

6. EXAMPLE SCENARIOS WITH OSA SCORES

6.1 Defining Example Scenarios

This chapter provides examples of different scenario types with evaluations based on the previously defined OSA metric violations and severity measurements. To demonstrate the efficacy and robustness of the evaluation, a diverse set of scenarios have been selected as summarized in Table 3. This selection includes both successful and failed attempts to navigate scenarios included in NHTSA's typology of pre-crash scenarios [37] in addition to several scenarios identified in the PATH program's behavioral competencies [42]. The purpose for including both scenarios in which the vehicle successfully navigates the maneuver and failed attempts by the vehicle is to demonstrate the competency of the evaluation methodology with regards to identifying hazards of a given situation. This chapter will also discuss the methodology for the various testing techniques and the results obtained.

Table 3. Example Scenario Test Matrix

Scenario Name	Scenario Type	Scenario Description	Scenario Outcome
CF_LB_C	Car-Following	Traveling same speed, lead vehicle braking	Collision
CF_LB_NM	Car-Following		Near-Miss
CF_LB_NE	Car-Following		No Event
I_LT_C	Intersection	Left turn in front of oncoming traffic	Collision
I_LT_NM	Intersection		Near-Miss
I_LT_NE	Intersection		No Event
LC_CI_C	Lane Change	Vehicle cut-in at slower speed	Collision
LC_CI_NM	Lane Change		Near-Miss
LC_CI_NE	Lane Change		No Event

6.2 Example Scenarios in Simulation

First, the test matrix from Table 3 was evaluated utilizing a simulation program called Human, Vehicle, Environment (HVE). HVE is a physics-based program capable of simulating vehicle dynamics during collision events. In conjunction with the simulations performed in HVE, the CARLA software was utilized to iterate key parameters for a sensitivity analysis of the tested scenarios. The previously described MATLAB script was developed to automate the calculations for the safety metrics defined in the previous chapters based on the outputs from the simulation programs. As a result, the methodology can be applied for any dataset containing the proper variables regardless of the test type (e.g., simulation, closed-course testing, public-road testing).

6.2.1 Simulated Scenario Methodology

The vehicles chosen for the simulated scenarios are popular passenger vehicles on U.S. roadways to be representative of commonly driven vehicles. The lead vehicle for each simulation was a 2019 Ford Fiesta SE sedan and the follow vehicle was a 2020 Honda Civic EX-T sedan. The default dimensional and weight specifications for the Ford Fiesta and Honda Civic are shown in Figure 14 and Figure 15, respectively.

Sprung Mass Dimensional Data ---	
Overall Length (in):	174.00
Overall Width (in):	66.93
Overall Height (in):	58.42
Ground Clearance (in):	9.41
Wheelbase (in):	97.99
CG to Front Axle (in):	40.97
CG to Back Axle (in):	-57.02
CG Height (in):	21.35
Front Overhang (in):	34.14
Rear Overhang (in):	41.87
Sprung Mass Inertial Data ---	
Total Weight (lb):	2722.76
Sprung Weight (lb):	2427.71
Sprung Mass (lb-sec^2/in):	6.28
Sprg Mass Rot Inertia (lb-sec^2-in) - Roll:	3270.72
Pitch:	14400.01
Yaw:	14906.05
XZ Product:	0.00

Figure 14. Default HVE 2019 Ford Fiesta Specifications

Sprung Mass Dimensional Data ---	
Overall Length (in):	182.28
Overall Width (in):	70.84
Overall Height (in):	55.29
Ground Clearance (in):	6.94
Wheelbase (in):	106.20
CG to Front Axle (in):	43.91
CG to Back Axle (in):	-62.29
CG Height (in):	24.07
Front Overhang (in):	35.17
Rear Overhang (in):	40.91
Sprung Mass Inertial Data ---	
Total Weight (lb):	2879.99
Sprung Weight (lb):	2718.01
Sprung Mass (lb-sec^2/in):	7.03
Sprg Mass Rot Inertia (lb-sec^2-in) - Roll:	3598.05
Pitch:	17001.67
Yaw:	17493.87
XZ Product:	0.00

Figure 15. Default HVE 2020 Honda Civic Specifications

As discussed previously, a number of the variables utilized in calculating the OSA metrics are assumed values. These assumed parameters include components such as the subject vehicle reaction time, the maximum deceleration of the lead vehicle, and the maximum acceleration of the follow vehicle. A detailed validation activity was conducted specifically for the MSEV based on RSS in [11]. Important consideration should be applied when determining the values to be used for the assumed parameters in the OSA metrics equations, as the value chosen may significantly impact the outcome of the calculations. The scenarios presented throughout this paper utilize the following assumed parameters:

- Reaction Time: 1.0 second
- Maximum Longitudinal Deceleration: 1.0 g
- Maximum Longitudinal Acceleration: 0.05 g
- Minimum Longitudinal Deceleration: 0.46 g
- Maximum Lateral Acceleration: 0.7 g

The acceleration limits are utilized in the equations to ensure maximum acceleration of the follow vehicle for the duration of the assumed reaction time would provide sufficient time for the vehicle to avoid a collision based on the MSEV. The acceleration values used here are consistent with reasonable expected values based on the simulation environment. The minimum deceleration for the follow vehicle is consistent with typical deceleration levels applied by drivers due to an unexpected obstacle [44].

The values used for the assumed parameters may vary depending on the specific implementation of the OSA methodology and the accuracy of the metric calculations can be improved by optimizing these values. Naturalistic data are a valuable source of information to ensure the assumed parameters are as reasonable and accurate as possible; however, the limitations of these data should be understood. For instance, naturalistic data for highway driving in Los Angeles will not necessarily provide useful assumptions for parameters of vehicles located in rural Ohio. Depending on the compute power and sensing capabilities of the subject vehicle, some parameters may be obtained in real-time to improve the metric calculations. One example of this would be classifying the lead vehicle as a passenger vehicle versus a pickup towing a trailer, versus a heavy commercial vehicle. These different lead vehicles can have drastically differing deceleration capabilities and could be accommodated as such if classification of these objects occurs. While determining the specific make and model of the lead vehicle may provide even further in-depth details regarding the vehicle capabilities, a more generalized classification would likely be sufficient to improve the calculations.

In addition to the classification of specific vehicle types, vulnerable road users (VRUs) such as pedestrians, cyclists, and animals will have differing values for their assumed parameters. The scenarios evaluated throughout this work do not include VRUs due to simulation limitations; however, the metric calculations could easily be implemented for such scenarios as long as the appropriate assumed parameters are applied.

The differing scenario types were selected to ensure the MATLAB script performed accurate calculations in all cases. To accomplish this, it was necessary to expand the formulations to not just calculate the metrics for a given car-following scenario, but also incorporate the necessary constraints to determine the scenario type and calculate the metrics accordingly. In order to demonstrate the effects of varying levels of severity for violation of the OSA metrics, three different scenario outcomes were designed into the simulation definition including collision, near-miss, and no event. For the collision outcome, the scenario was defined in such a manner that the two vehicles collided with one another in an obvious manner. For the near-miss scenario, the same initial parameters were utilized; however, an avoidance maneuver was performed by the subject vehicle (and in some cases the aggressiveness of the other vehicle maneuver was reduced). The no event outcome was designed to border the edge of violations for the subject vehicle.

6.2.2 Simulated Scenario Results

Once the baseline HVE simulations were generated, the results were plotted and compared for each of the scenario outcomes of the same scenario description. Each OSA metric was evaluated for each of the scenarios to understand the impact of potentially hazardous scenarios with respect to the severity of a safety metric violation. Although a specific vehicle ADS was not evaluated, the purpose of this exercise was to evaluate the methodology and showcase the ability to quantify the performance of a given vehicle (human-drive and automated alike) leveraging observable variables that can be used to calculate metrics from a Black Box approach. Metrics requiring information directly from the vehicle ADS would require either a simulated or physical ADS for testing.

6.2.2.1 Car-Following Scenario

The first scenario evaluated was a basic car-following scenario in which both vehicles were initially driving at the same speed until the lead vehicle began braking. The timing and aggressiveness of the braking for the lead and follow vehicle were varied for the different scenario outcomes to alter the results between the collision, near-miss, and no event. The initial parameters for the lead and follow vehicles are summarized in Table 4. The initial setup for the scenario was then replicated in CARLA and the parameters were varied to generate additional scenarios. An illustration of the initial car-following scenario is depicted in Figure 16.

Table 4. Car-Following Scenario Initial Parameters

	Collision		**Near-Miss**		**No Event**	
Variable Name	**Follow Vehicle**	**Lead Vehicle**	**Follow Vehicle**	**Lead Vehicle**	**Follow Vehicle**	**Lead Vehicle**
Initial Position	(-200.0 ft, 61.0 ft)	(0.0 ft, 60.7 ft)	(-200.0 ft, 61.0 ft)	(0.0 ft, 60.7 ft)	(-200.0 ft, 61.0 ft)	(0.0 ft, 60.7 ft)
Initial Velocity	45 mph	45 mph	45 mph	45 mph	45 mph	45 mph
Time of Brake Initiation	8.5 seconds	5.0 seconds	8.5 seconds	5.0 seconds	7.3 seconds	5.0 seconds
Average Brake Magnitude	-0.75 g	-0.68 g	-0.76 g	-0.50 g	-0.76 g	-0.31 g

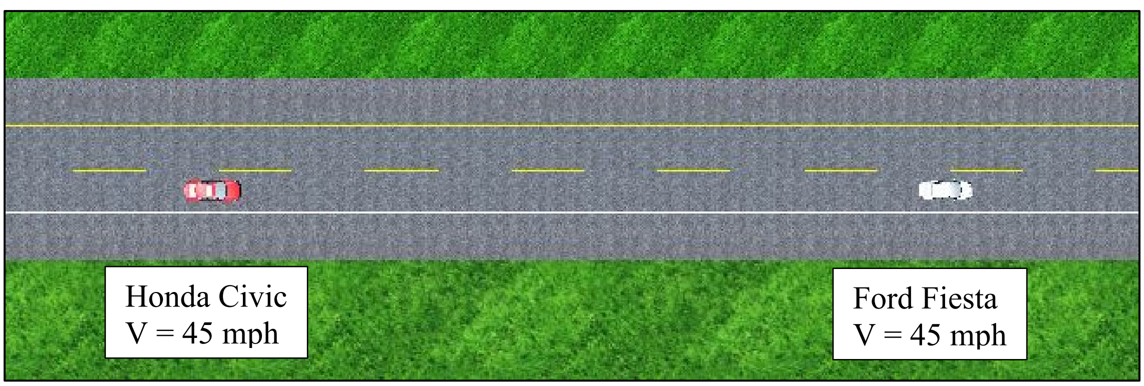


Figure 16. Car-Following Scenario Setup in HVE

6.2.2.1.1 Car-Following Scenario – MSE and PR Evaluation

The graph shown in Figure 17 depicts the initial conditions for the scenario including the velocities for both vehicles and the closing distance throughout the scenario. The decrease in the dashed green line illustrates the lead vehicle braking five seconds into the simulation, resulting in a reduction in the closing distance (solid blue). The MSEV and PRV are triggered when the longitudinal distance between the vehicles falls below the calculated MSE (dashed blue). In the simulated scenario, a collision occurs due to a delayed braking response by the follow vehicle with a resulting PRV severity of 1.0 given a sufficient braking response did not occur prior to the subject vehicle reaching the trajectory overlap zone.

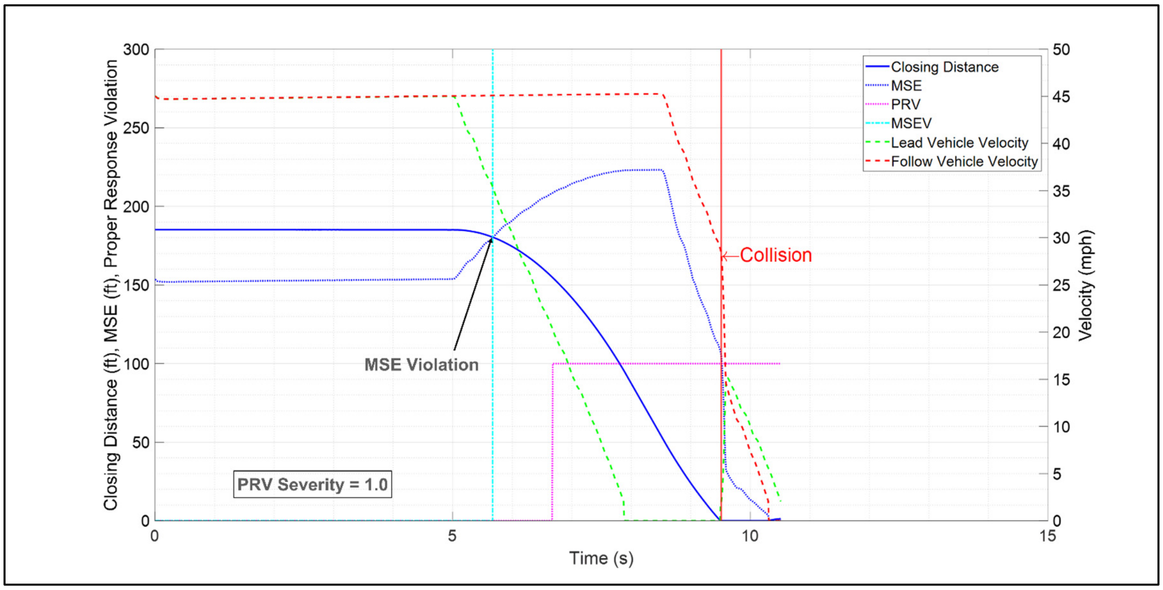


Figure 17. Scenario CF_LB_C Initial Conditions and MSEV Plot

Two additional scenarios were generated with varying parameters to lessen the severity of the event. Figure 18 was generated, illustrating the same general scenario characteristics resulting in the follow vehicle narrowly avoiding the lead vehicle by applying a braking maneuver. Similarly, the parameters were adapted further yet to generate the scenario depicted in Figure 19 resulting in a momentary proper response violation. The decreasing levels of PRV severity of 1.0, 0.70, and 0.004 correspond to the lessening of the event severity for the collision, near-miss, and no event scenarios, respectively.

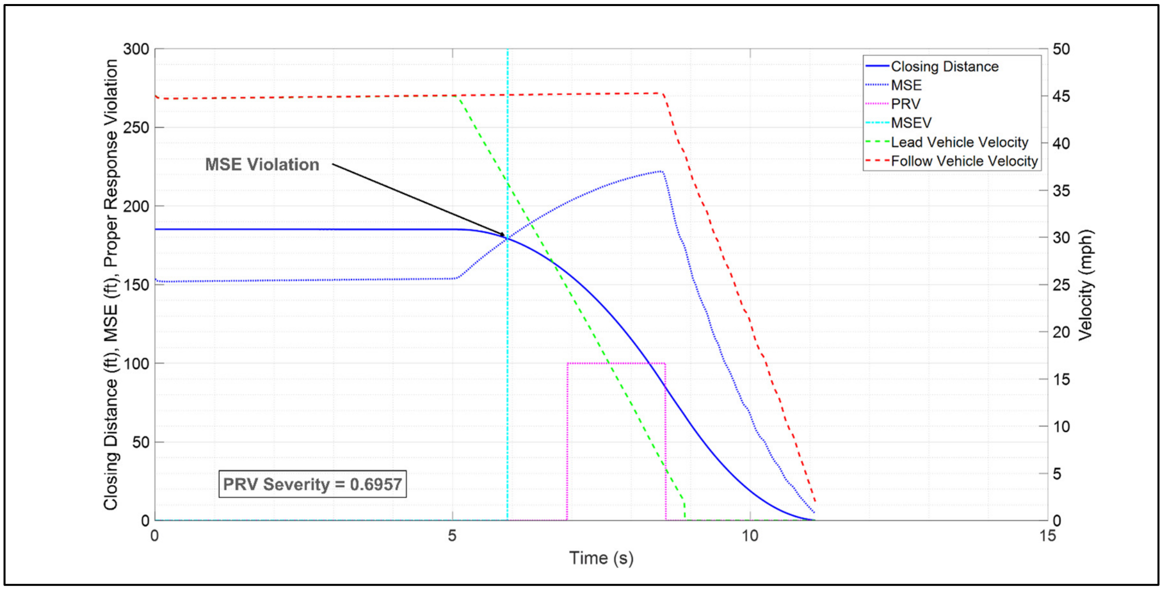


Figure 18. Scenario CF_LB_NM Initial Conditions and MSEV Plot

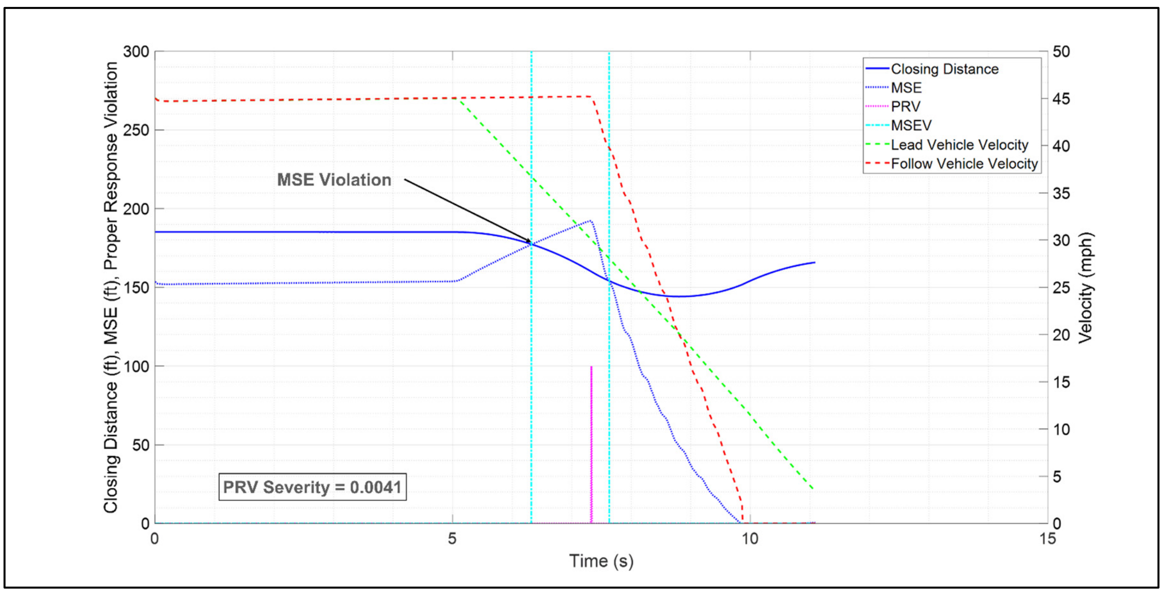


Figure 19. CF_LB_NE Initial Conditions and MSEV Plot

6.2.2.1.2 Car-Following Scenario – MRD Evaluation

As explained previously, the MRD acts as the severity evaluation of an MSEV. The MRD utilizes the current distance between the vehicles, current velocities, and an assumed level of braking for the lead vehicle to determine the deceleration required to avoid a collision. Ranging the lead vehicle deceleration from 10% to 100% of the vehicle's braking capability provides confidence in the results of the scenario for a variety of lead vehicle behaviors. As depicted in Figure 20 for the analyzed scenario, the MRD begins in the low braking zone as the follow vehicle maintains a safe distance from the lead vehicle. Upon the lead vehicle braking, the closing distance is reduced and the MRD correspondingly increases. Once the lead vehicle reaches a complete stop (at approximately 8 seconds), the varying MRD curves converge since additional braking will no longer impact the dynamics of the lead vehicle. As a result of the delayed reaction by the follow vehicle in applying the brakes, the MRD rapidly approaches the high braking zone and quickly exceeds the capabilities of the follow vehicle, thus resulting in a collision.

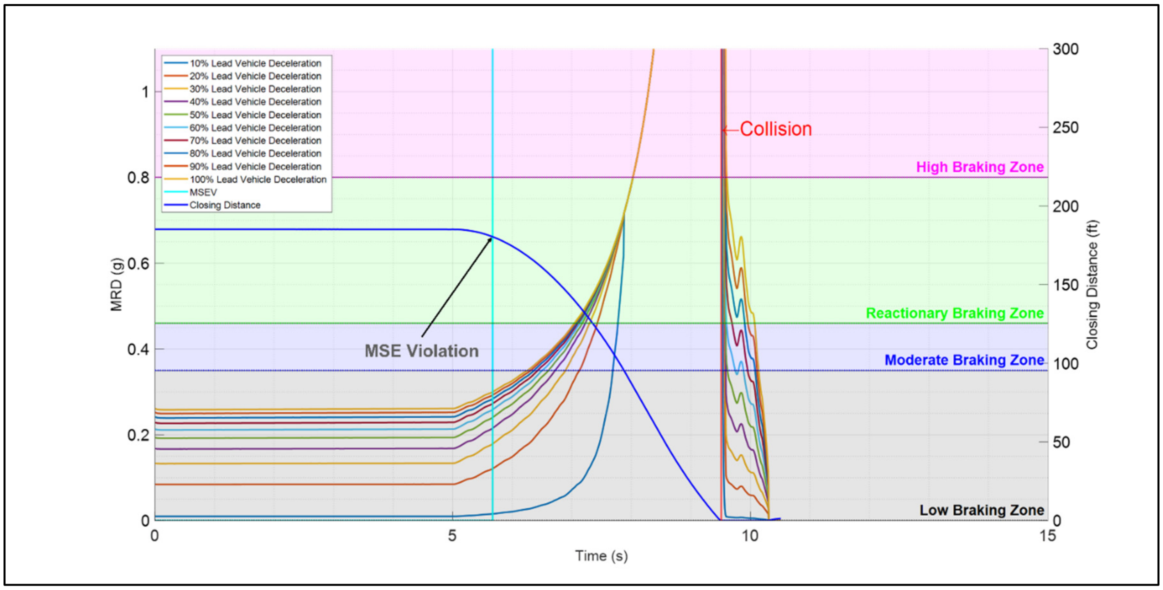


Figure 20. Scenario CF_LB_C MRD Evaluation

As expected, with the lessening severity of events for the near-miss and no event scenarios, the MRD is reduced, as illustrated in Figure 21 and Figure 22. For the collision scenario, the maximum braking capability of the follow vehicle was exceeded just over one second prior to the collision due to the delayed braking response. The near-miss scenario resulted in a maximum MRD of 0.9 g demonstrating the close proximity to a collision whereas the no event scenario MRD never exceeded the low braking zone. The varying scenarios presented here illustrate the formulation for the MSE as an MSEV occurs when the MRD approaches the moderate braking zone and will continue to increase until a proper response is applied to reestablish the MSE.

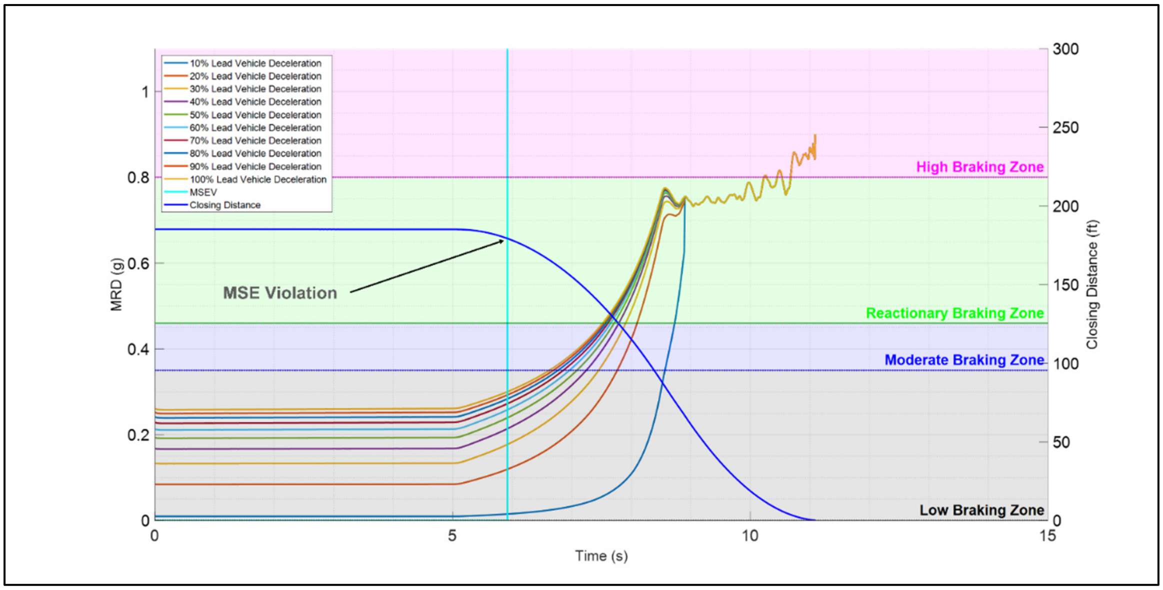


Figure 21. Scenario CF_LB_NM MRD Evaluation

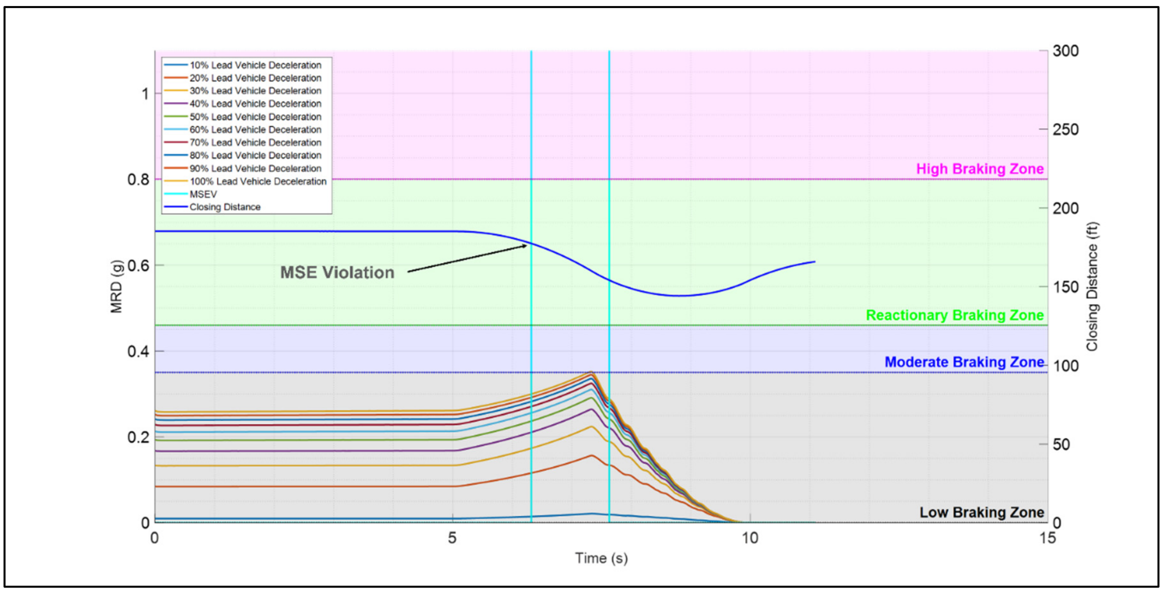


Figure 22. Scenario CF_LB_NE MRD Evaluation

6.2.2.1.3 Car-Following Scenario – Safety Envelope-Related Metric Violations

The reviewed literature revealed numerous safety envelope-related metrics which could be utilized to evaluate a vehicle for a given scenario. The various safety envelope-related metrics were calculated and plotted against one another for the simulated scenarios to compare their robustness, consistency, and overall ability to detect a potential event. As can be seen in Figure 23, a brief MTTC violation (MTTCV) occurs near the beginning of the scenario with no other violations occurring until the MSEV approximately four seconds prior to the collision. The remainder of the safety envelope-related metric violations follow the MSEV intermittently including the PET, TTC, MTTC, and THW. It should be noted that the other safety envelope-related metric violations are based on a threshold of 2.5 seconds. This threshold was chosen based on the simulation work conducted in [3] and affects the frequency as well as timing at which a violation will be triggered. These other safety envelope-related metrics will be included for demonstrative purposes throughout this section; however, this work identified a more consistent and comprehensive evaluation through the implementation of the MSE metric throughout the different scenarios presented. The consistency of the MSEV is shown throughout due to its comprehensive consideration of the vehicle kinematics.

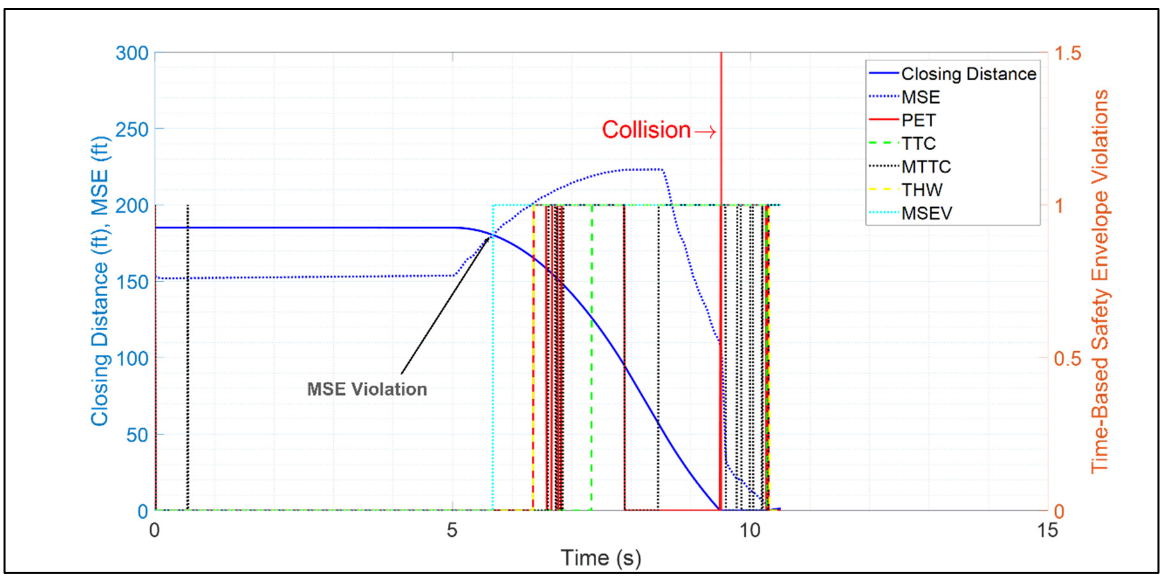


Figure 23. Scenario CF_LB_C Safety Envelope-Related Metric Violations

The MSEV occurs at the same time for each of the car-following scenarios because the dynamics leading up to the MSEV remained constant. This approach helped demonstrate the consistency of the MSEV in relation to the other safety envelope-related metrics which differ in timing and frequency for the near-miss and no event scenarios as depicted in Figure 24 and Figure 25, respectively.

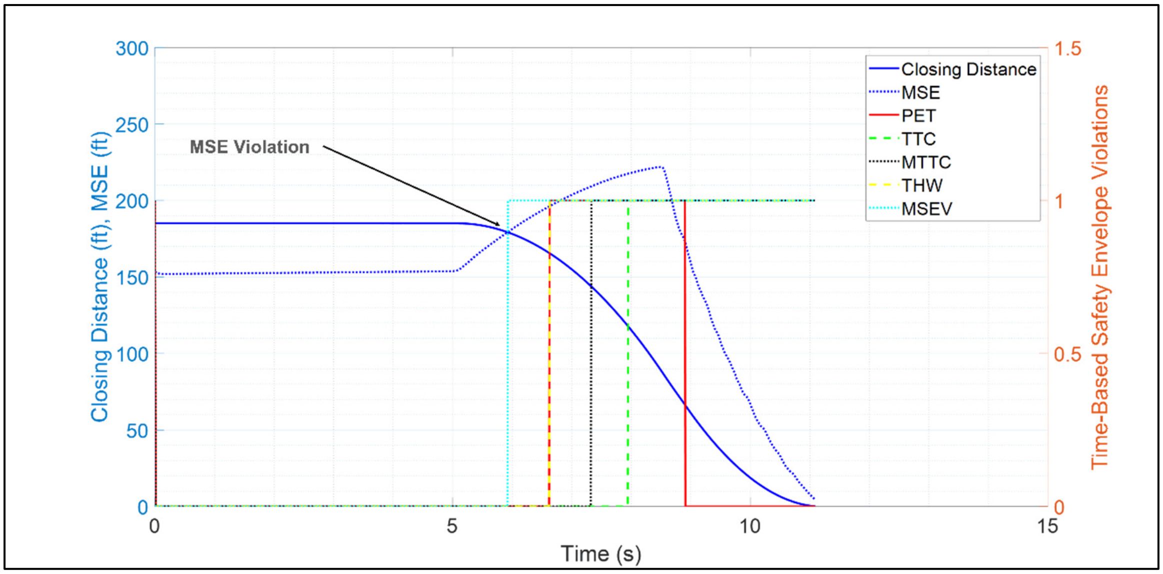


Figure 24. Scenario CF_LB_NM Safety Envelope-Related Metric Violations

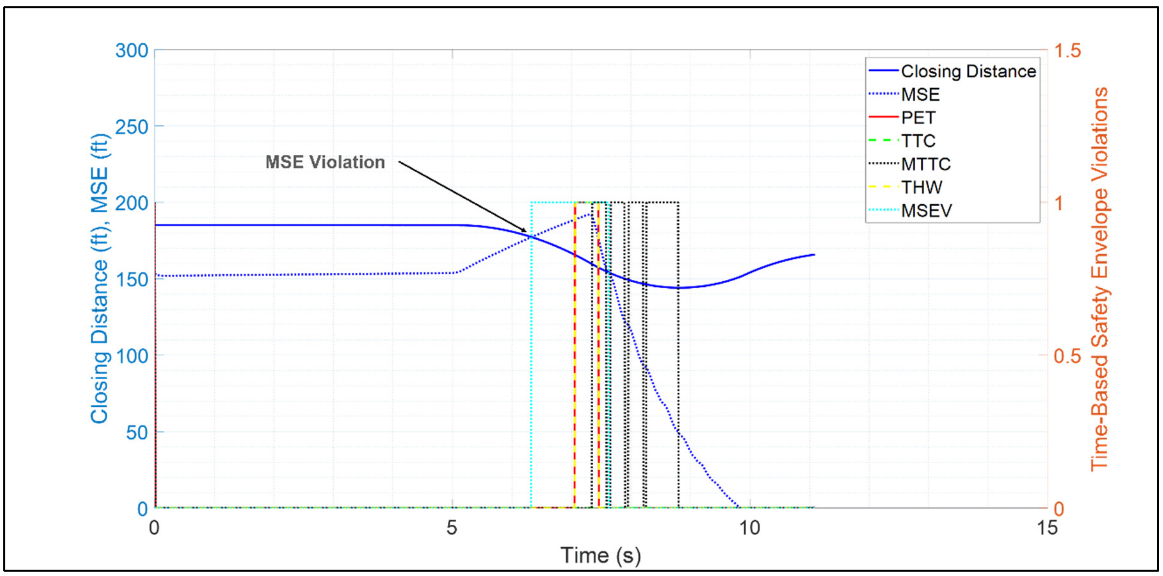


Figure 25. Scenario CF_LB_NE Safety Envelope-Related Metric Violations

6.2.2.1.4 Car-Following Scenario – CIV Severity Evaluation

In addition to the OSA metrics depicted in the above graphs, the CI and PA metrics were also evaluated for each of the given scenarios. A CIV severity only exists for scenarios in which a collision occurs and therefore is not compared across the scenarios presented for the car-following instances. Table 5 summarizes the CIV severity for the lead and follow vehicles for the collision event. Given the similar mass of the involved vehicles and the inline nature of the collision, the change in velocity, or delta-V, for both vehicles is approximately the same. As defined in the metric severity definition, the generalized formulation assumes rear impacts to be the least severe followed by frontal impacts, and side impacts being the most severe assuming the delta-V remains constant. It should be noted that this is a generalized formulation based on statistical data from [45] and exceptions may exist for specific cases. Given these assumptions, the follow vehicle experiences the greater CIV severity due to the similar delta-V because it experienced a frontal collision versus the lead vehicle experiencing a rear impact. The severity of this collision is relatively minor with a maximum CIV severity of 0.005.

Table 5. Scenario CF_LB_C CIV Severity

Lead Vehicle CIV Severity	**Follow Vehicle CIV Severity**
0.002	0.005

6.2.2.1.5 Car-Following Scenario – PAV Severity Evaluation

Unlike the CI metric, the PA metric can be applied to each of the car-following scenarios analyzed. Table 6 summarizes the PAV for both the lead and follow vehicles in each scenario. The PAV severity by itself does not necessarily provide a useful metric for evaluation of a vehicle as an initial review would indicate the follow vehicle was least aggressive in the collision scenario and therefore performed the best; however, we know that a higher deceleration in the initial scenario may have prevented the collision. All other metrics being equal, a lower PAV severity would indicate greater overall performance for the subject vehicle; yet, the context of the scenarios should be considered utilizing the other metrics to avoid misleading conclusions of increased performance for a more hazardous result. Although the subject vehicle is the vehicle under evaluation, understanding the PAV severity of the other salient object provides additional context as to the complexity of the scenario which will be discussed in detail later as another aspect of the OSA methodology. For example, if the lead vehicle is driving highly aggressively, the scenario may be more difficult for the subject vehicle to navigate just as it would be for a human driver.

Table 6. Car-Following Scenario PAV Severity

Scenario	**Lead Vehicle PAV Severity**	**Follow Vehicle PAV Severity**
CF_LB_C	0.1622	0.0645
CF_LB_NM	0.0	0.1511
CF_LB_NE	0.0	0.1454

6.2.2.2 Intersection Scenario

Next, a set of intersection scenarios were developed, again with a collision, near-miss, and no event scenario. A similar comparison was conducted to evaluate the same metrics from a different scenario type and prove the robustness of the formulations. The intersection used in this set was a T-intersection with two lanes in each direction of travel. The lead vehicle in this set of scenarios was positioned at the east side of the intersection making a left turn across traffic while the follow vehicle was traveling eastbound towards the turning vehicle in the number two (outside) travel lane. The initial parameters for each scenario are defined in Table 7 and a snapshot from the collision scenario with the lead vehicle turning left in front of the follow vehicle is depicted in Figure 26.

Table 7. Intersection Scenario Initial Parameters

	Collision		Near-Miss		No Event	
Variable Name	**Follow Vehicle**	**Lead Vehicle**	**Follow Vehicle**	**Lead Vehicle**	**Follow Vehicle**	**Lead Vehicle**
Initial Position	(-428.0 ft, 48.8 ft)	(57.0 ft, 24.4 ft)	(-428.0 ft, 48.8 ft)	(57.0 ft, 24.4 ft)	(-428.0 ft, 48.8 ft)	(57.0 ft, 24.4 ft)
Initial Velocity	45 mph	0 mph	45 mph	0 mph	45 mph	0 mph
Time of Brake Initiation	N/A	N/A	5.4 seconds	N/A	4.7 seconds	N/A
Average Brake Magnitude	N/A	N/A	-0.76 g	N/A	-0.77 g	N/A

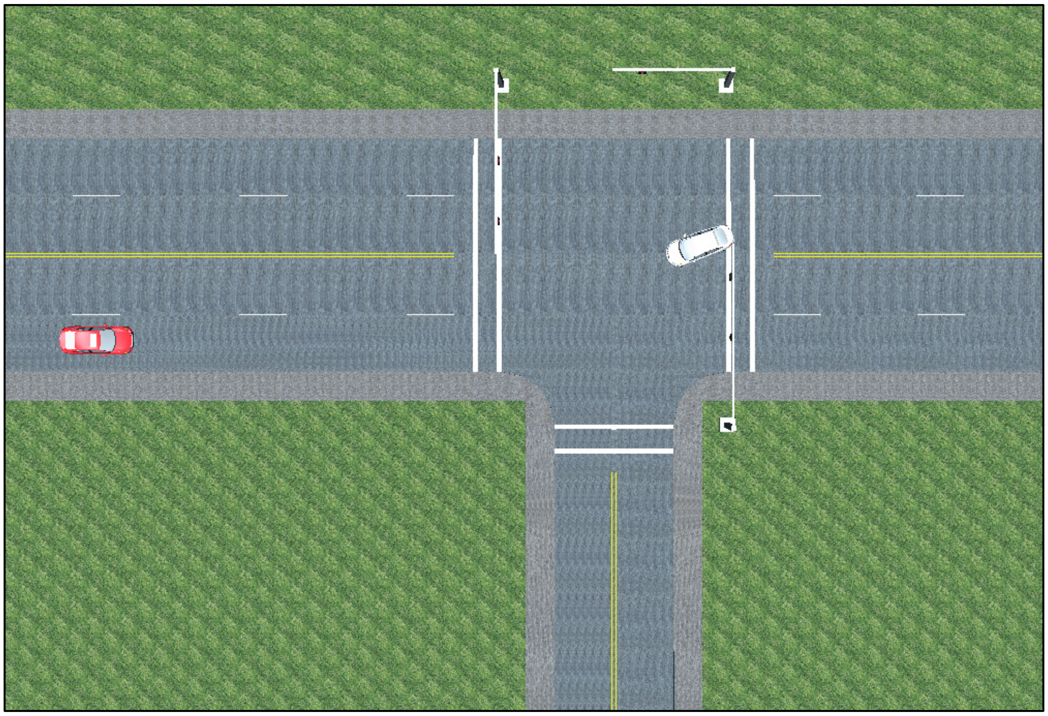


Figure 26. Intersection Scenario Setup in HVE Illustrating Lead Vehicle (White) Turning Left in Front of Follow Vehicle (Red)

6.2.2.2.1 Intersection Scenario – MSEV and PRV Evaluation

The intersection scenario was added to demonstrate the robustness of the evaluation methodology for differing types of events. In this scenario, the lead vehicle does not initially start in the path of the following vehicle; however, as the lead vehicle accelerates and steers through the intersection, its trajectory quickly intersects with that of the follow vehicle. In order to account for this mathematically, the future positions for each vehicle are determined based on the current position and the longitudinal and lateral velocities. This method is demonstrated in Figure 27 where the MSEV is shown to occur just prior to 4 seconds into the event, even though the lead vehicle does not physically enter the follow vehicle's lane of travel until just after 6 seconds into the simulation. Since the subject vehicle does not react to the vehicle turning in front, thus resulting in a collision, the PRV severity is shown to be 1.0.

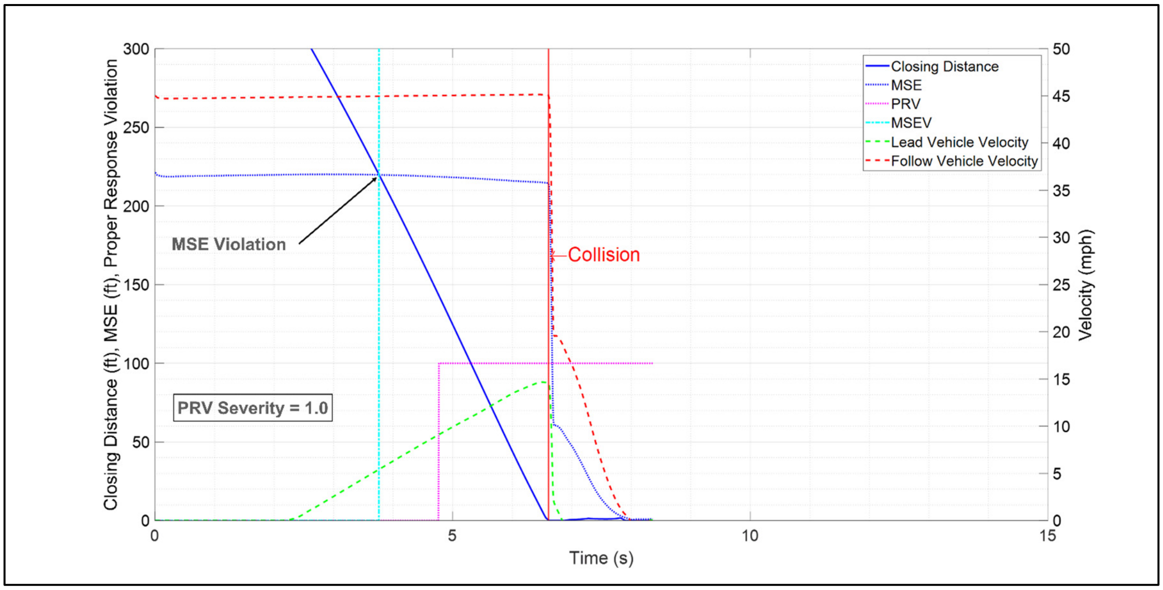


Figure 27. I_LT_C Initial Conditions and MSEV Plot

The subject vehicle was given a braking input in the near-miss and no event scenarios to avoid striking the lead vehicle with the same initial conditions. Since the conditions leading up to the MSEV are held constant, the violation occurs at the same point in time; although, the PRV severity is reduced to 0.3723 in the near-miss scenario and a PRV is altogether removed from the no event scenario as a result of the earlier brake application by the subject vehicle. Figure 28 and Figure 29 illustrate this reduction in severity for the alternate scenarios.

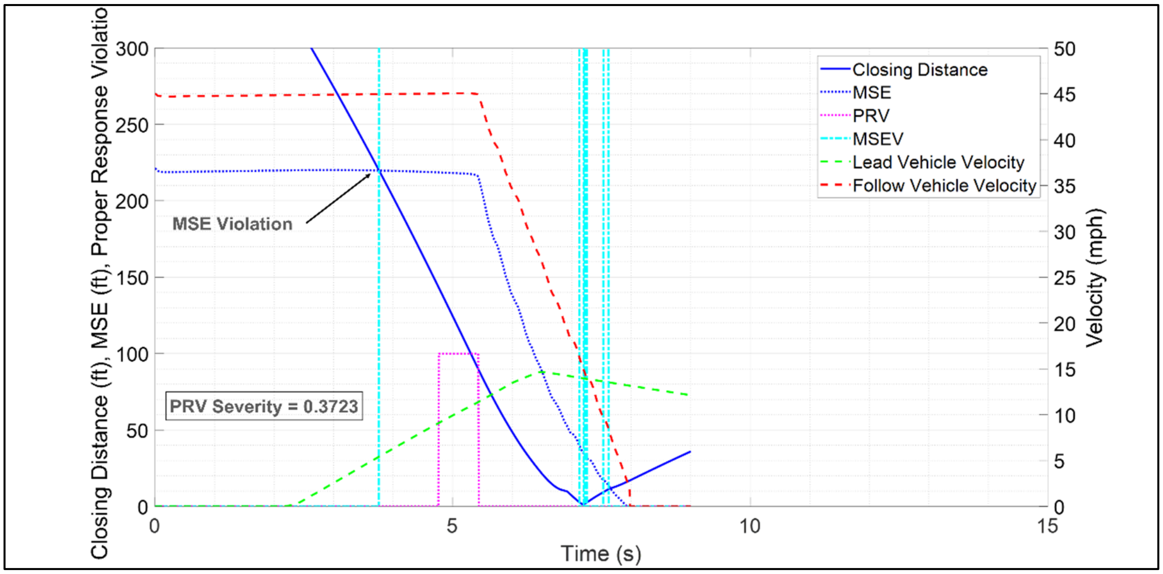


Figure 28. I_LT_NM Initial Conditions and MSEV Plot

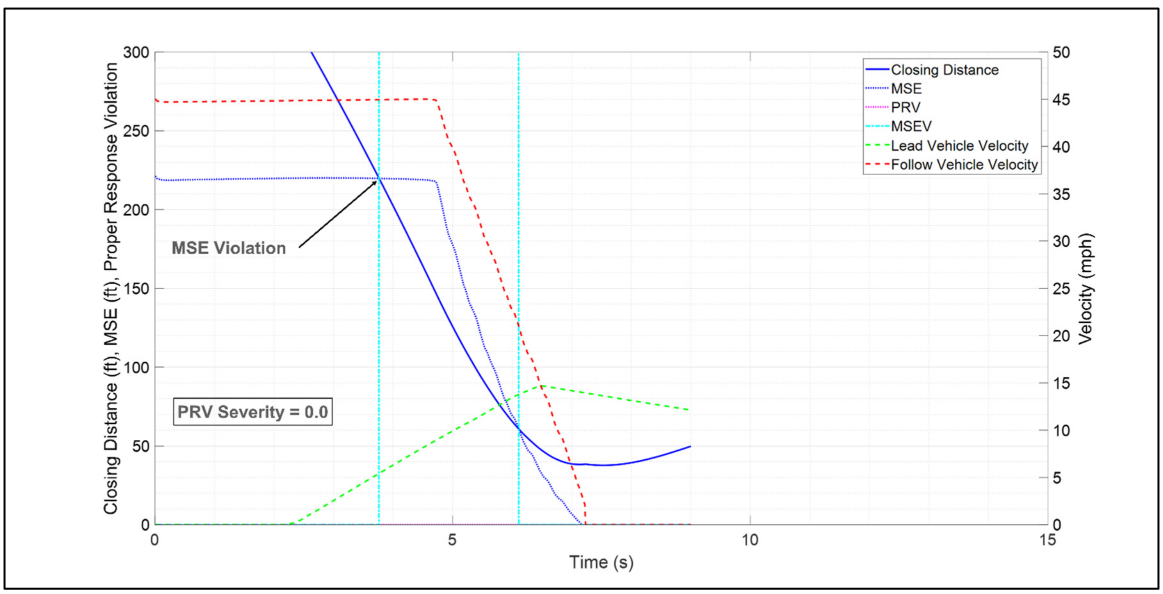


Figure 29. I_LT_NE Initial Conditions and MSEV Plot

6.2.2.2.2 Intersection Scenario – MRD Evaluation

As was the case with the car-following scenarios, the MRD decreases from the collision, to the near-miss, and finally, to the no event scenarios. Since the lead vehicle starts from a stop and accelerates rather than the previous set in which the lead vehicle started at a constant velocity and slowed to a stop, the varying levels of assumed braking for the lead vehicle cause the MRD curves to diverge as the lead vehicle accelerates. The MRD quickly enters the high braking zone and exceeds the vehicle braking capability as a result of the lack of response to the oncoming collision for the collision event scenario. Conversely, the MRD hardly surpasses the reactionary braking zone for the near-miss scenario and remains within the moderate braking zone for the no event scenario. The differing levels of MSEV severity as quantified by the MRD is shown in Figure 30 through Figure 32.

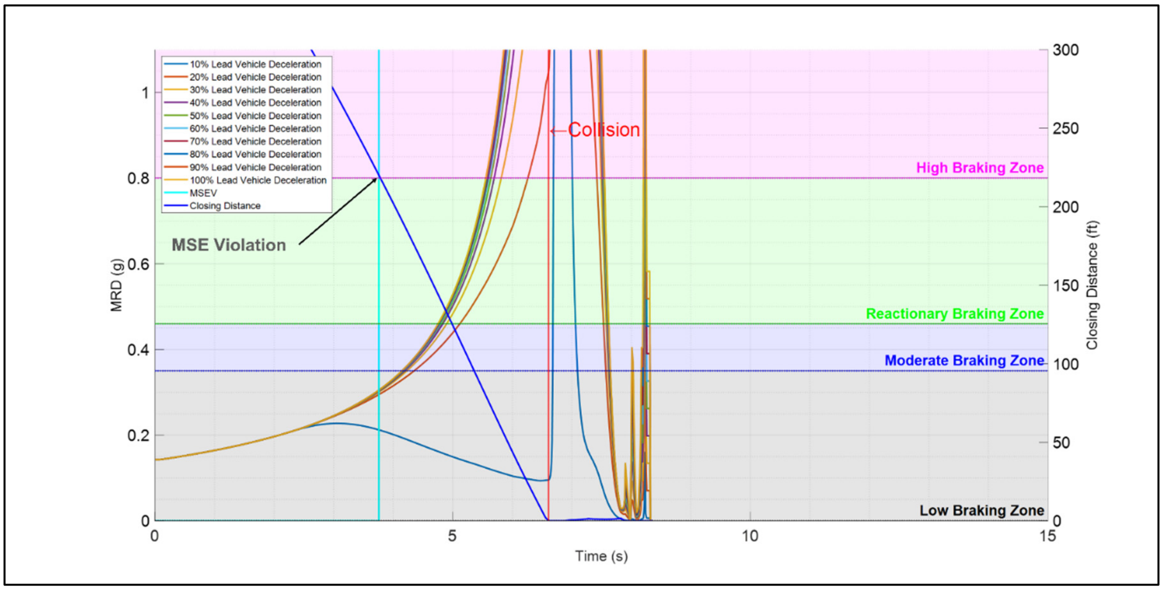


Figure 30. Scenario I_LT_C MRD Evaluation

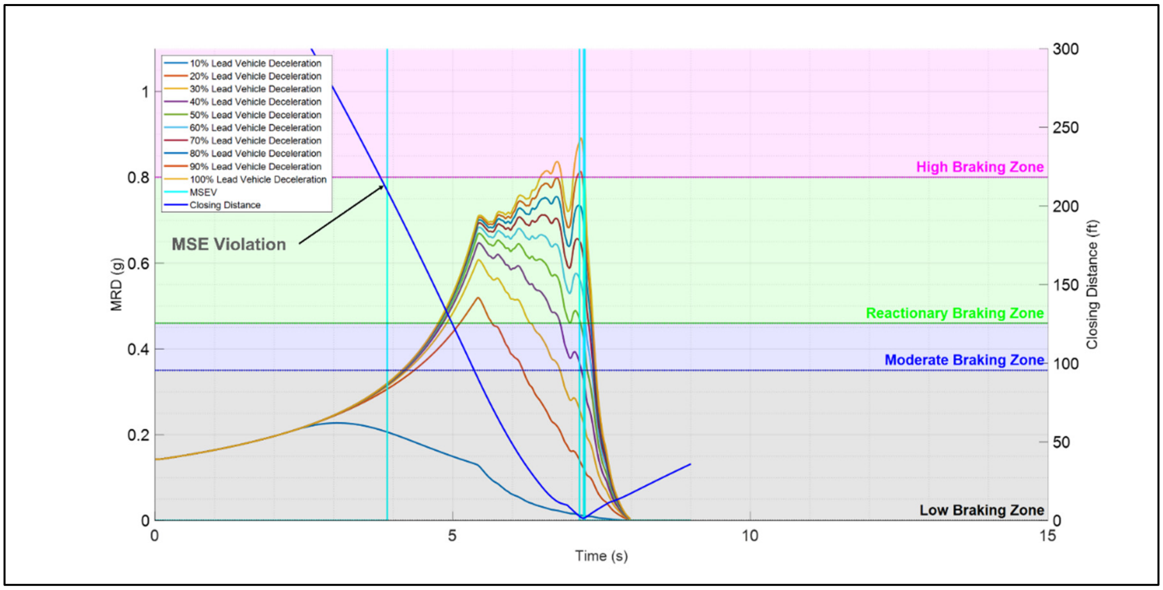


Figure 31. Scenario I_LT_NM MRD Evaluation

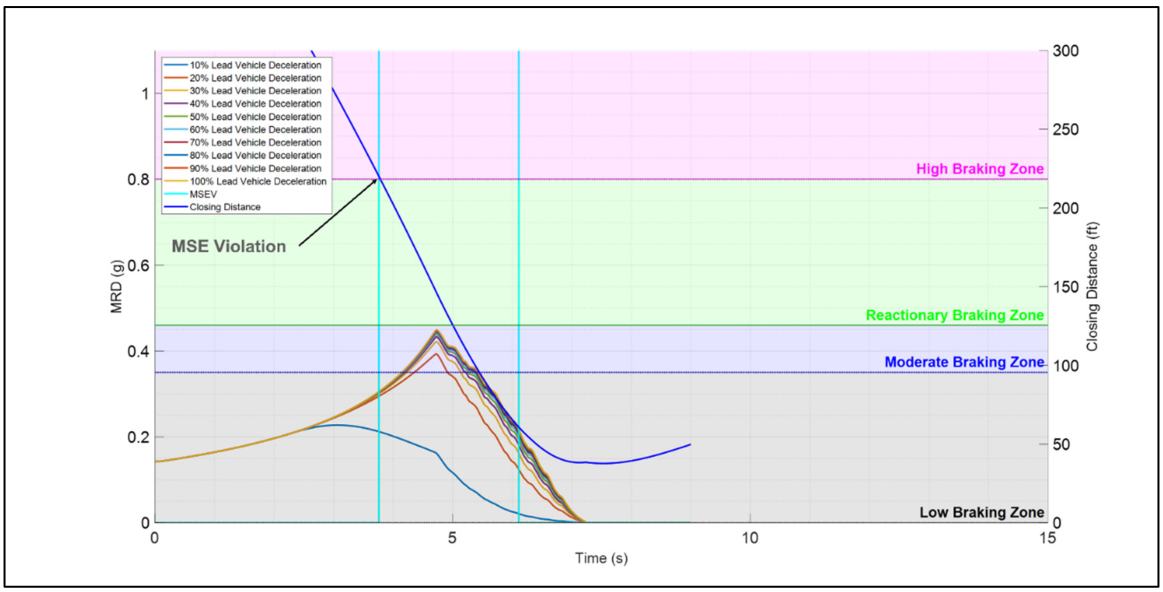


Figure 32. Scenario I_LT_NE MRD Evaluation

Intersection Scenario – Safety Envelope-Related Metric Violations

As was demonstrated in Figure 27 through Figure 29, the MSEV occurs at the same point in time for each scenario with differing levels of severity based on the calculated MRD, resulting from differences in the brake initiation time for the subject vehicle. As was the case with the car-following scenarios, the other safety envelope-related metrics displayed some variation between the differing scenarios in both occurrence time and frequency, shown in Figure 33 through Figure 35. In all cases, the other safety envelope-related metrics triggered after the MSEV; however, as explained previously, the trigger time is dependent on the threshold set for the violations. Again, the MSEV appears the most consistent and comprehensive across scenarios by incorporating the most information from the involved vehicles and basing the violation severity on the assumed physical limitations of the vehicles.

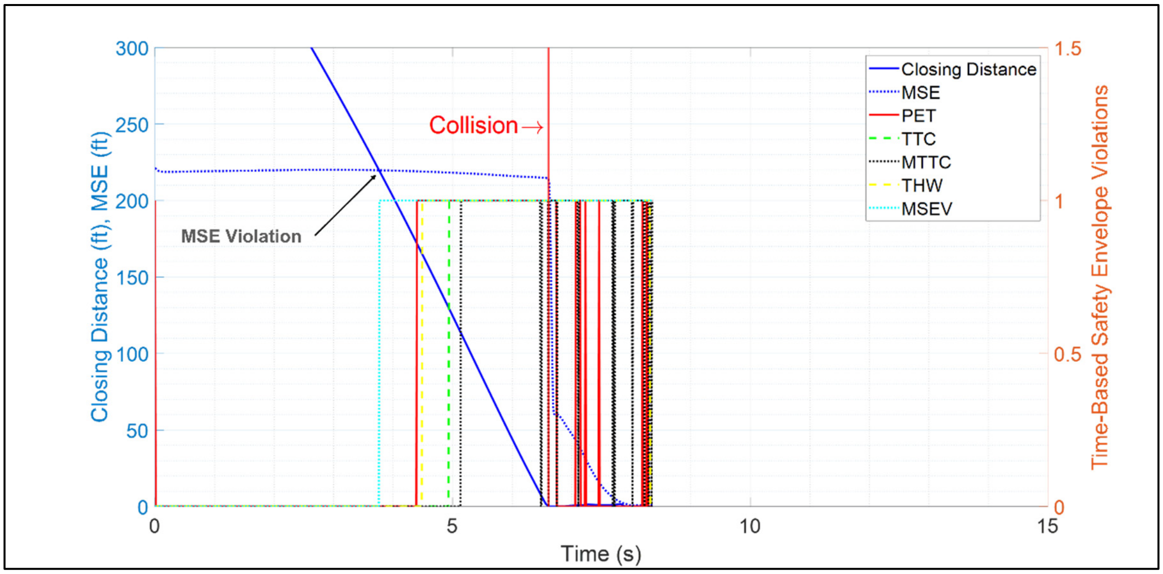


Figure 33. Scenario I_LT_C Safety Envelope-Related Metric Violations

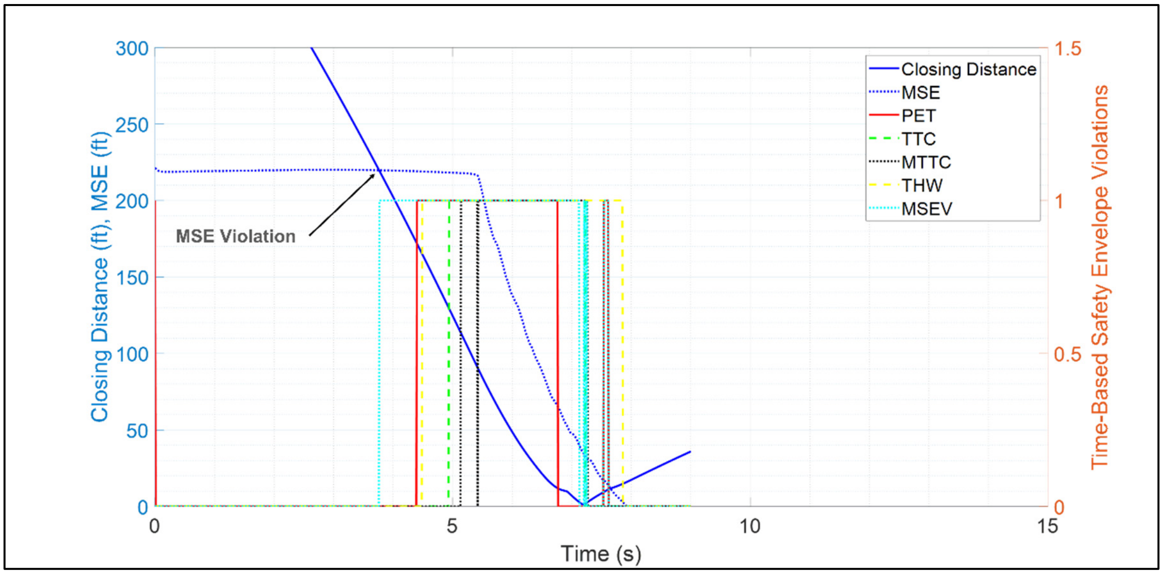


Figure 34. Scenario I_LT_NM Safety Envelope-Related Metric Violations

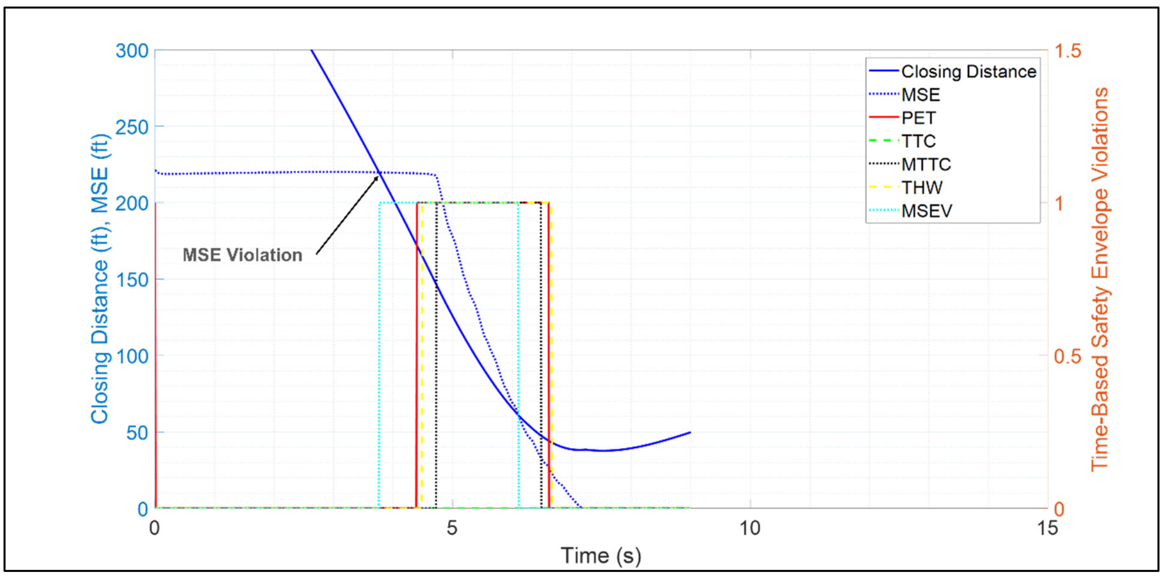


Figure 35. Scenario I_LT_NE Safety Envelope-Related Metric Violations

6.2.2.2.3 Intersection Scenario – CIV Severity Evaluation

The collision event for the intersection scenario resulted in a substantially higher severity value than that of the car-following scenario. The primary factor attributing to the increase in severity is the impact modality, introducing a side impact rather than the frontal and rear collisions that resulted from the car-following collision event. Additionally, the delta-Vs for the vehicles in this scenario were greater than those resulting from the car-following scenario, further increasing the severity and producing the 16.2% risk of an MAIS 4+F for the lead vehicle occupants in this scenario in addition to a 3.1% risk for the subject vehicle.

Table 8. Scenario I_LT_C CIV Severity

Lead Vehicle CIV Severity	Follow Vehicle CIV Severity
0.162	0.031

6.2.2.2.4 Intersection Scenario – PAV Severity Evaluation

Similar to the car-following collision scenario, the intersection collision scenario represented the lowest PAV severity since the subject vehicle did not apply any braking to avoid the collision. The similar magnitude of braking for the near-miss and no event scenarios resulted in a similar PAV severity of approximately 0.19. The lead vehicle in this case did not perform any maneuvers resulting in the vehicle dynamics exceeding the threshold of predictable acceleration and therefore has a PAV severity of 0.0; however, the failure to yield to the subject vehicle in the case of the collision would represent a traffic law violation and could be accounted for in the complexity of the scenario as there would be a limit to which the lead vehicle could turn in front of the subject vehicle without leaving enough time for the subject vehicle to avoid a collision. Conversely, it is possible the turning vehicle had the right-of-way based on the traffic signal phase and timing, requiring further investigation to determine the at fault vehicle which would follow the flow chart previously depicted in Figure 4.

Table 9. Intersection Scenario PAV Severity

Scenario	Lead Vehicle PAV	Follow Vehicle PAV
I_LT_C	0.0	0.0
I_LT_NM	0.0	0.1900
I_LT_NE	0.0	0.1878

6.2.2.3 Cut-In Scenario

To further test the robustness of the metric violations and severity formulations, a cut-in scenario was simulated in which the lead vehicle initially traveled in the lane adjacent to the subject vehicle, then cut into the lane at a slower speed than the subject vehicle necessitating a response. The initial longitudinal position and speed of the subject vehicle was varied for each alternate scenario to generate less severe situations with the same cut-in maneuver by the lead vehicle. Table 10 summarizes the initial parameters for the various cut-in scenarios and Figure 36 depicts the lead vehicle cutting into the subject vehicle's lane during the collision scenario in HVE.

Table 10. Cut-In Scenario Initial Parameters

	Collision		Near-Miss		No Event	
Variable Name	**Follow Vehicle**	**Lead Vehicle**	**Follow Vehicle**	**Lead Vehicle**	**Follow Vehicle**	**Lead Vehicle**
Initial Position	(-340.0 ft, 53.0 ft)	(-100.0 ft, 66.0 ft)	(-440.0 ft, 53.0 ft)	(-100.0 ft, 66.0 ft)	(-350.0 ft, 53.0 ft)	(-100.0 ft, 66.0 ft)
Initial Velocity	65 mph	45 mph	65 mph	45 mph	55 mph	45 mph
Time of Brake Initiation	6.7 seconds	5.5 seconds	6.7 seconds	5.5 seconds	5.5 seconds	5.5 seconds
Average Brake Magnitude	-0.81 g	-0.66 g	-0.77 g	-0.67 g	-0.78 g	-0.67 g

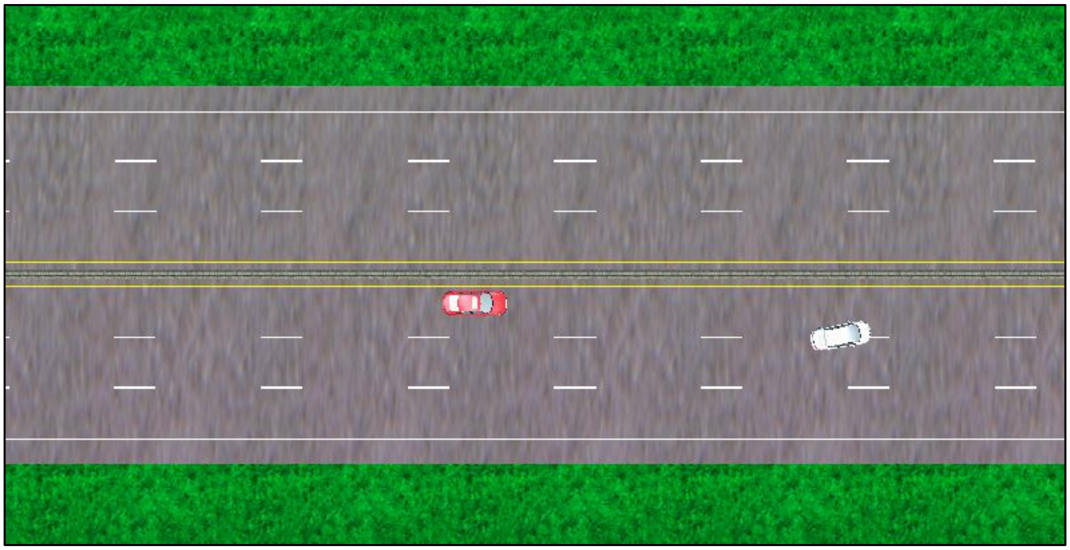


Figure 36. Cut-In Scenario Setup in HVE Illustrating Lead Vehicle (White) Changing Lanes in Front of Follow Vehicle (Red)

6.2.2.3.1 Cut-In Scenario – MSE Violation and PR Violation Evaluation

The cut-in scenarios were simulated with differing levels of severity by maintaining a constant lane change maneuver by the lead vehicle and varying the initial position and speed of the subject vehicle. The MSEV is less intuitive for the cut-in scenario because the longitudinal MSE is violated from the beginning of the simulation as depicted in Figure 37 through Figure 39. The MSEV is much less stable for the lane change scenarios. The reason for the instability of the violation measure is the sensitivity to slight heading changes during the scenario. Since the longitudinal MSE is violated from the beginning of each scenario, a slight yaw for either the lead or follow vehicle will result in an intersection of the paths creating a lateral violation as well. Although this results in a high frequency of MSEVs, the severity of the violation is still a result of the calculated MRD. Additionally, because the MSEVs are intermittent, a PRV does not occur for the brief violation measurements. The near-miss scenario results in a lower PRV severity as a result of the greater longitudinal distance at the time of the lane change, allowing sufficient time for the subject vehicle to stop prior to impacting the slower lead vehicle. The PRV then goes away entirely for the no event scenario due to the lower closing speed between the two vehicles and the earlier braking response by the subject vehicle.

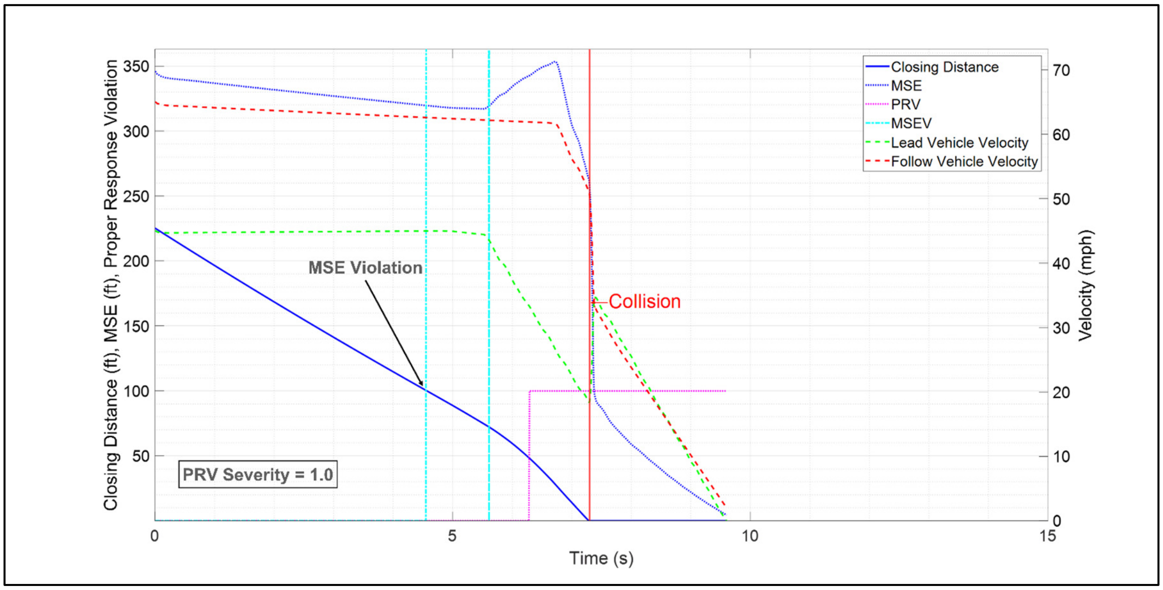


Figure 37. LC_CI_C Initial Conditions and MSEV Plot

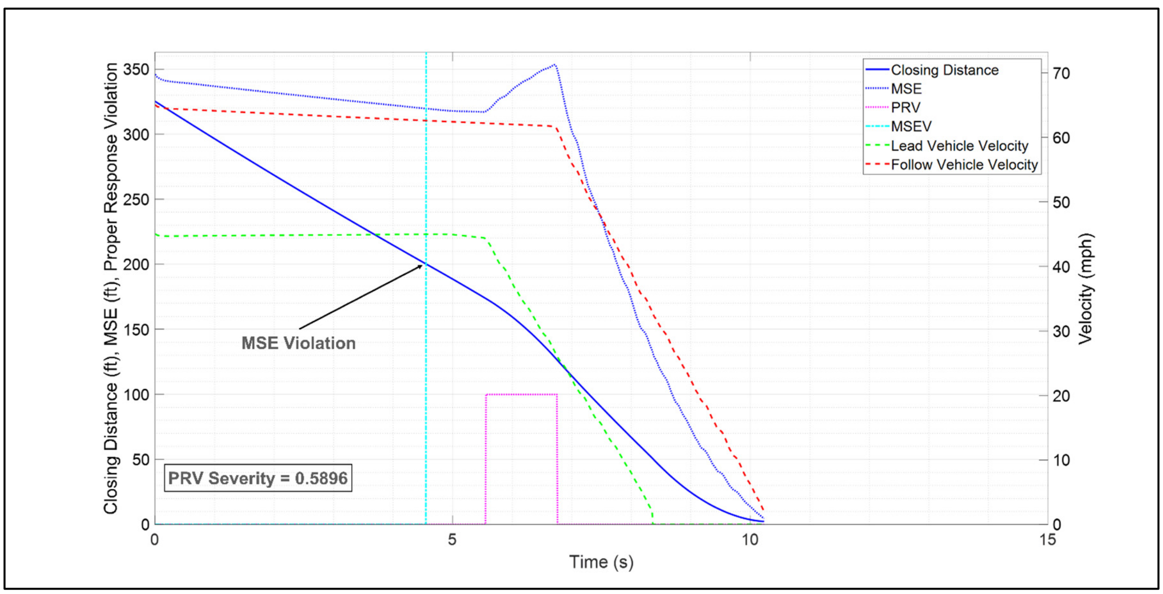


Figure 38. LC_CI_NM Initial Conditions and MSEV Plot

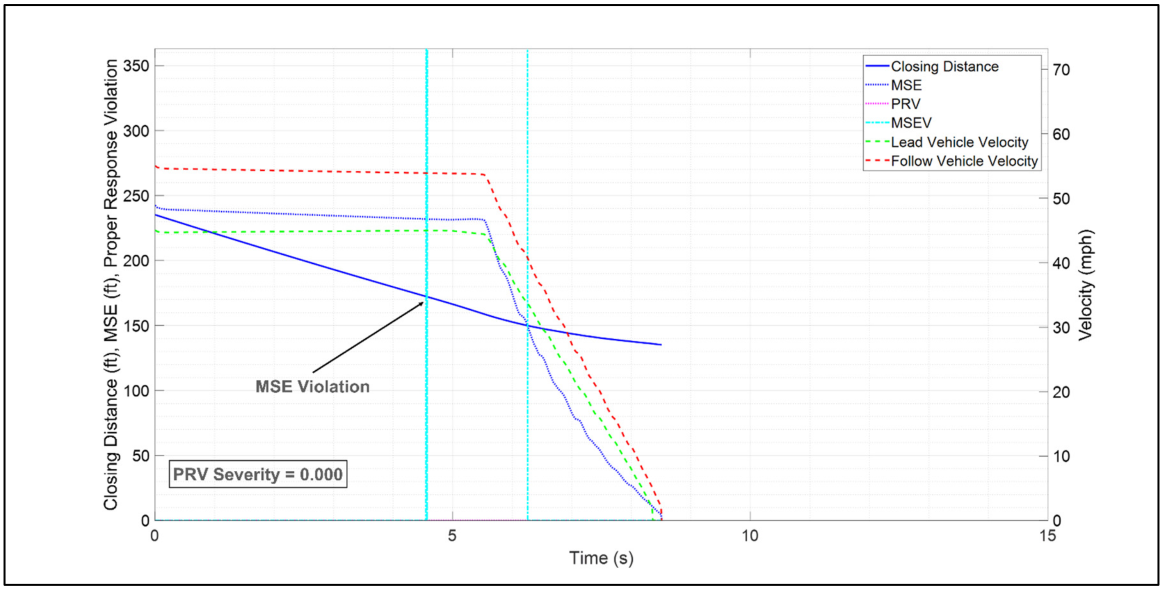


Figure 39. LC_CI_NE Initial Conditions and MSEV Plot

6.2.2.3.2 Cut-In Scenario – MRD Evaluation

Since the cut-in scenario occurred when the longitudinal MSE had already been compromised, the MRD begins near the high braking zone by the time the MSEV triggers (almost 5 seconds into the simulation) as shown in Figure 40. As was the case for the intersection scenario, the near-miss and no event scenarios resulted in an MRD reaching the reactionary and moderate braking zones, respectively, as illustrated in Figure 41 and Figure 42. The cut-in collision scenario represents a situation where the MSEV is initiated by the other salient object and creates an almost unavoidable impact assuming a 1 second reaction time for the subject vehicle to achieve the MRD. Had the lane change by the lead vehicle been initiated slightly later, the collision would have been unavoidable from the start and the subject vehicle would not be penalized for the collision based on the proposed evaluation methodology for an OEF-CI.

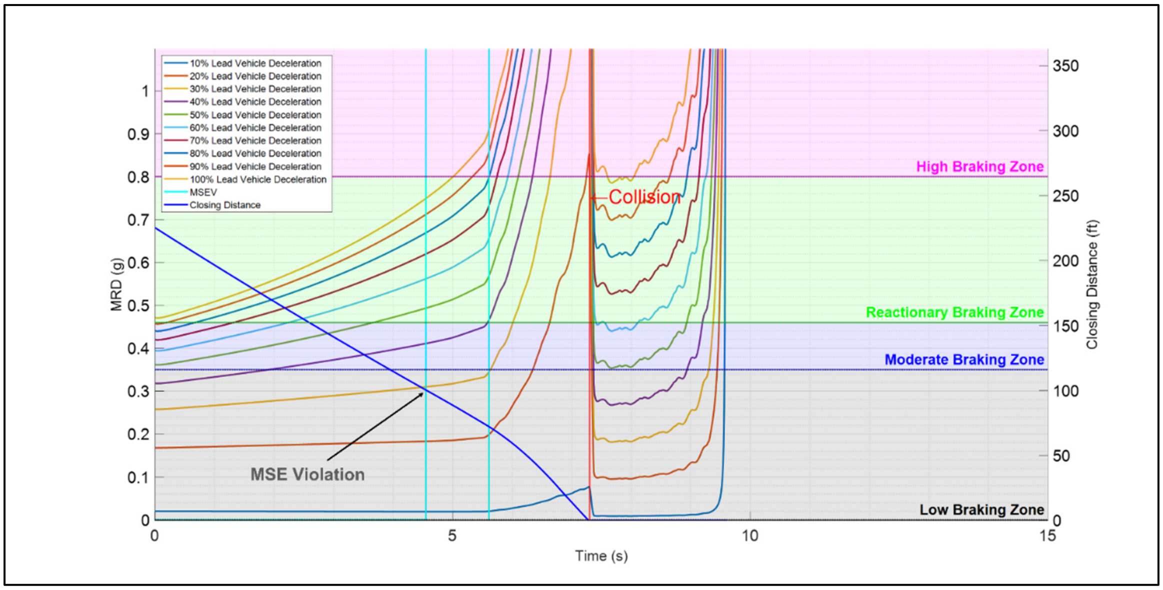


Figure 40. Scenario LC_CI_C MRD Evaluation

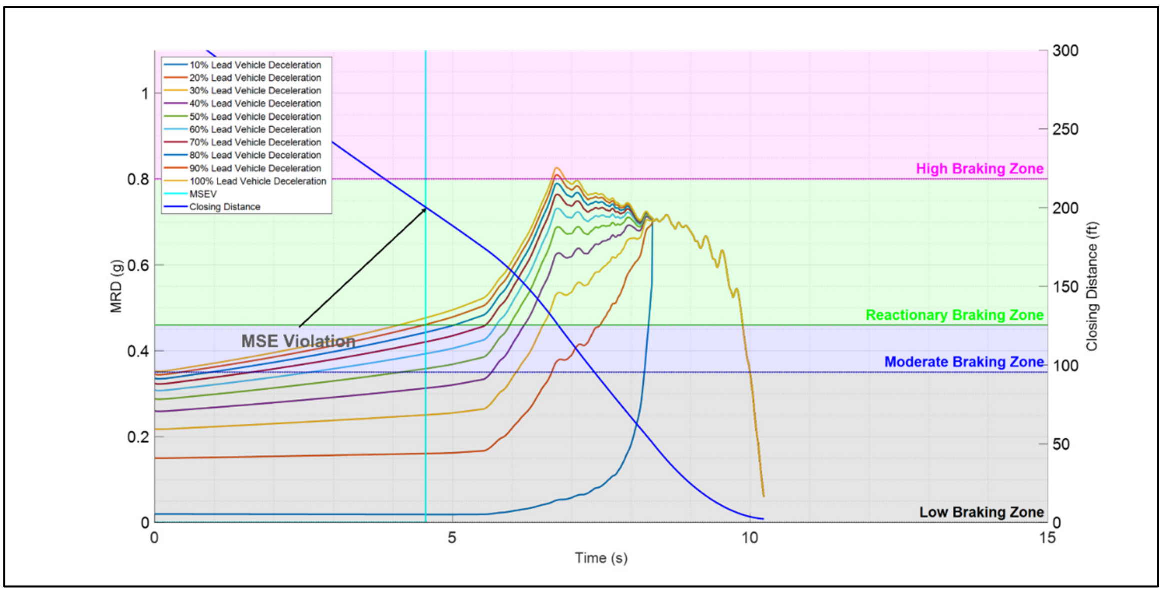


Figure 41. Scenario LC_CI_NM MRD Evaluation

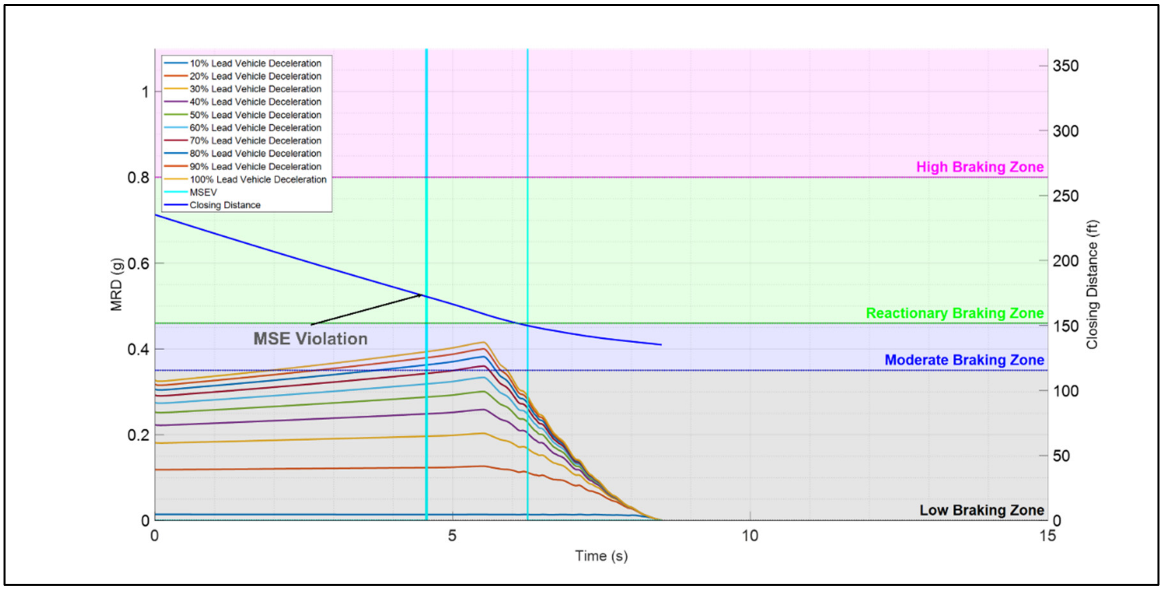


Figure 42. Scenario LC_CI_NE MRD Evaluation

6.2.2.3.3 Cut-In Scenario – Safety Envelope-Related Metric Violations

As was the case with the previously described scenarios, the other safety envelope-related metric violations vary with regards to timing and frequency while the MSEV occurs at the same time for each alternate simulation as depicted in Figure 43 through Figure 45. The THW violation (THWV) was the only metric violation that was triggered prior to the MSEV in each of the simulated scenarios; however, it should be noted that many of the various safety envelope-related metrics do not consider the lateral component of the vehicles demonstrating an additional limitation in that the TTC, MTTC, and THW would be incapable of differentiating the follow vehicle safely traveling parallel to the lead vehicle in the adjacent lane versus traveling towards the lead vehicle in the same lane.

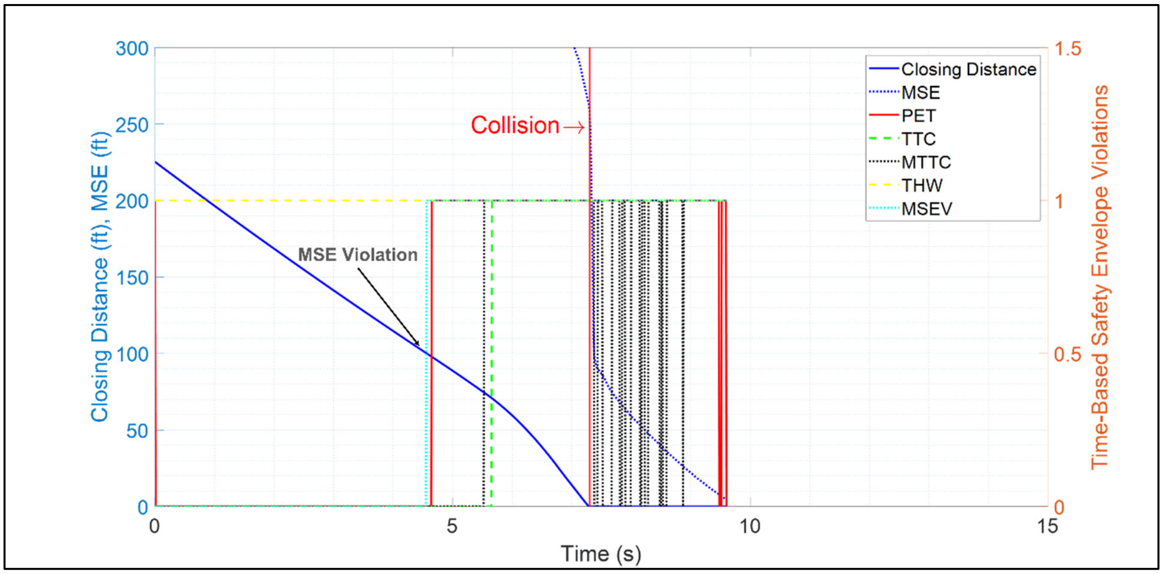


Figure 43. Scenario LC_CI_C Safety Envelope-Related Metric Violations

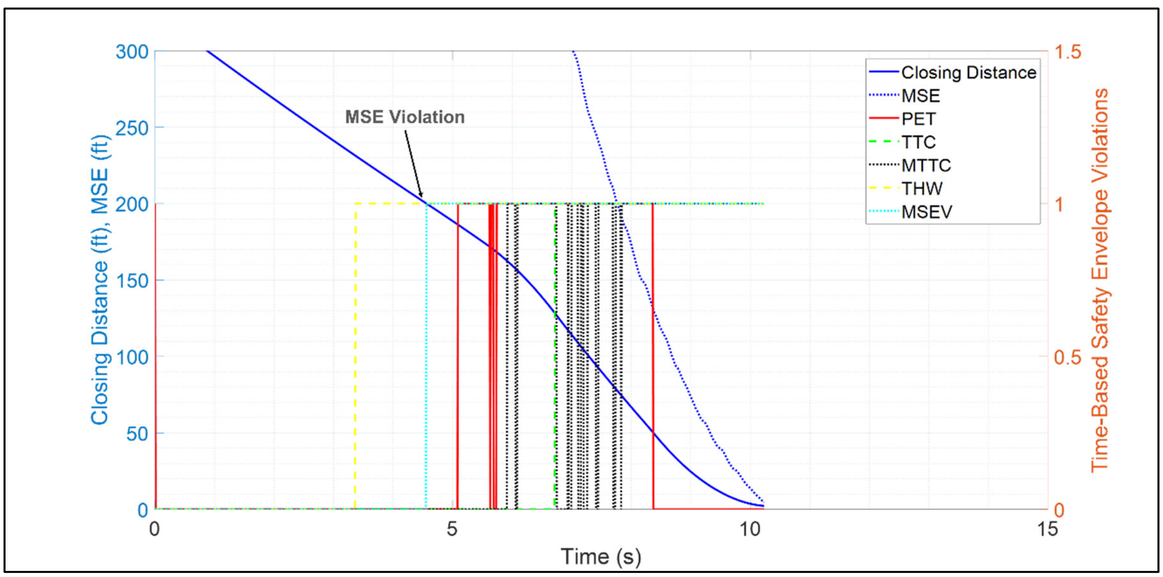


Figure 44. Scenario LC_CI_NM Safety Envelope-Related Metric Violations

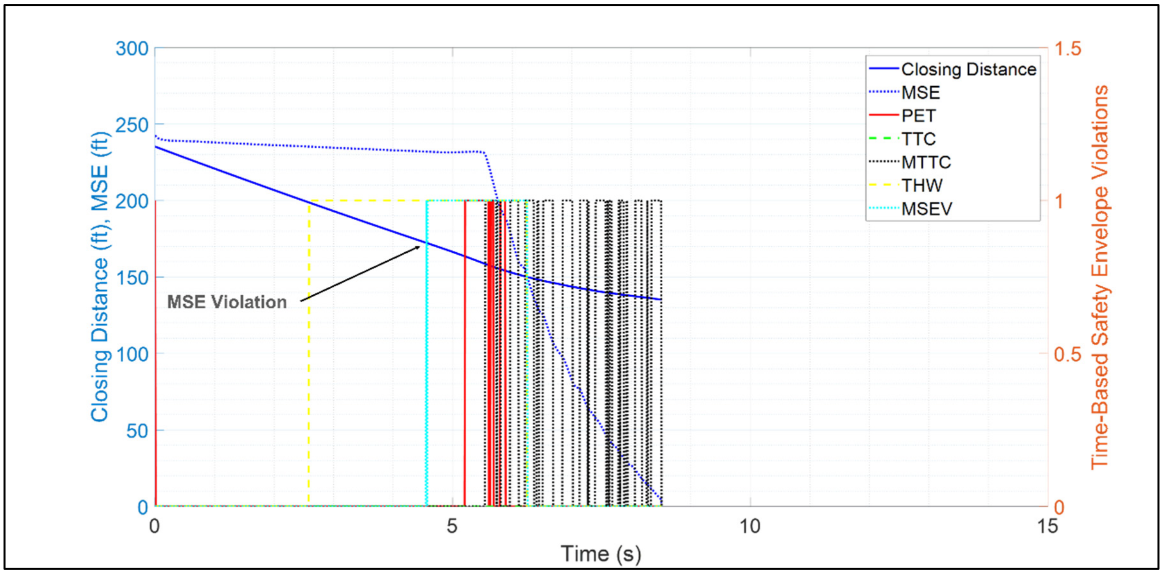


Figure 45. Scenario LC_CI_NE Safety Envelope-Related Metric Violations

6.2.2.3.4   Cut-In Scenario – CIV Severity Evaluation

The CIV severity is more comparable to that of the initial car-following scenario set since the impact modality is the same even though the initial parameters were different. The delta-V resulting from the simulated collision was slightly higher than that of the initial car-following collision scenario, therefore yielding a slightly higher CIV severity of 0.010 shown in Table 11.

Table 11. Scenario LC_CI_C CIV Severity

Lead Vehicle CIV Severity	Follow Vehicle CIV Severity
0.002	0.010

6.2.2.3.5   Cut-In Scenario – PAV Severity Evaluation

The cut-in scenario set provides yet another example of the PAV severity giving misleading results without considering the full context of the scenario as summarized in Table 12. The lead vehicle behaves least aggressively during the impact scenario to most aggressively in the no event scenario with a corresponding increase in severity for the follow vehicle PAV. While the results shown here appear counterintuitive, a simple example demonstrating the reason for the PAV severity would be an identical scenario to the no event simulation in which the follow vehicle maintains a larger initial longitudinal distance from the lead vehicle in the adjacent lane, corresponding to a lighter brake application to avoid a collision. The reduction in braking would reduce the PAV severity for the follow vehicle to indicate the less aggressive driving behavior.

Table 12. Cut-In Scenario PAV Severity

Scenario	**Lead Vehicle PAV Severity**	**Follow Vehicle PAV Severity**
LC_CI_C	0.1388	0.0584
LC_CI_NM	0.1614	0.2355
LC_CI_NE	0.1940	0.2395

6.2.3 Scenario Testing Summary Discussion

The previously shown examples help to characterize the robustness and limitations of the various proposed metrics to be used in quantifying the safety of a vehicle for a given scenario. As previously discussed, the OSA methodology evaluates vehicle performance first by determining whether a metric violation exists, then applying a normalized weighting factor to the violation based on severity and the values are summed for the complete scenario. The weighted aggregate scores for the scenarios described in detail throughout this chapter were calculated and compiled for comparison. Table 13 summarizes the scores calculated for each of the simulated scenarios. It should be noted that the calculated scores only include metrics that could be measured for the simulated scenarios (i.e., Black Box metrics). Metrics related to the subject vehicle perception system and other subsystems that would include Grey Box, White Box, and Clear Box metrics could not be included in this evaluation since an actual vehicle ADS was not tested. There were no TLVs throughout the tested scenarios because an actual ADS was not being evaluated; therefore, any TLV would have been arbitrarily created. It is not unreasonable to assume a TLV should be an uncommon violation if the ADS is designed to conform with traffic laws as is the case in the evaluated scenarios.

Table 13. Summary Table of OSA Scores for Simulated Scenarios

Scenario Type	Description	MSEV	PRV	CIV	PAV	TLV	Score
Car-Following	CF_LB_C	1.000	1.000	0.005	0.065	0.000	**58.6%**
	CF_LB_NM	0.900	0.583	0.000	0.151	0.000	**67.3%**
	CF_LB_NE	0.352	0.004	0.000	0.145	0.000	**90.0%**
Intersection	I_LT_C	1.000	1.000	0.162	0.000	0.000	**56.8%**
	I_LT_NM	0.891	0.372	0.000	0.190	0.000	**70.9%**
	I_LT_NE	0.450	0.000	0.000	0.188	0.000	**87.3%**
Lane Change	LC_CI_C	1.000	1.000	0.010	0.058	0.000	**58.6%**
	LC_CI_NM	0.826	0.572	0.000	0.236	0.000	**67.3%**
	LC_CI_NE	0.416	0.000	0.000	0.24	0.000	**86.9%**

For illustrative purposes, consider a case in which the car-following, no-event scenario resulted in a TLV due to the follow vehicle exceeding the speed limit. Assuming the same evaluated kinematics with the addition of the TLV, the overall score would drop from 90.0% to 70.0%. This significant drop in the overall score demonstrates the negative impact of a TLV on the vehicle evaluation.

It should be noted that the scope of the OSA methodology defined within this paper extends beyond the analysis conducted within this section. The major limitation of the simulation used to define these scenarios lies in the lack of an actual ADS to be evaluated; therefore, removing the ability to analyze ADS-specific metric violations. Although the ADS-related metrics could not be evaluated within these scenarios, the framework has been laid out and the script has been designed for ease of implementation of these metrics in future work. Despite the lack of ADS-related metrics in the simulated scenarios, numerous interesting observations could be made from the Black Box metrics focused on the vehicle dynamics of the scenario. One such observation is that based on the current formulation, the resulting scores for the collision, near-miss, and no event scenarios followed the anticipated outcome of the highest score for no event (corresponding to the lowest cumulative violation severities) decreasing to the lowest score for the collision events regardless of scenario type.

Another interesting observation resulting from the scenarios is the increase in severity for the PAVs for the events avoiding a collision due to high deceleration levels enacted to avoid the collision. All other metric violations being equal, the ability to avoid a collision or even to minimize a safety envelope violation with lower deceleration levels would correspond to a lower severity. Although avoidance maneuvers resulted in higher severities of the PA metric, these increased severities did not cause the overall score to reach the same level as the collision events. The formulations described throughout this paper are just one example of possible thresholds and severity assessments for such metrics. Possible methods to overcome the penalization of a vehicle exceeding the acceleration thresholds as necessary to navigate a scenario may include options such as negating the PAV if a proper response is being performed by the subject vehicle as a result of another salient object's actions or establishing weighting factors for metrics relating to varying levels of safety.

It is important to consider all relevant metrics when evaluating the performance of a vehicle for a given scenario to provide as much context as possible; however, it may be determined that categorized scores are more useful than providing a single score for vehicle performance. Such a methodology would be similar to the Insurance Institute for Highway Safety (IIHS) safety ratings for vehicles broken down by various categories such as *Crashworthiness, Crash Avoidance & Mitigation,* and *Seat Belts & Child Restraints* [47]. Although the vehicle receives an overall performance rating in this structure, individual categories of performance can be more easily compared using this breakdown. This type of rating system could allow for a vehicle to have a pass/fail criterion for more critical metrics such as a collision incident while a metric like predictable acceleration that may be considered less safety critical than a collision itself may not carry the same implications as an independent rating. Such a pass/fail type rating could be further applied by establishing required thresholds for different metric categories. For example, a CIV or PRV may be considered a failure regardless of severity whereas PAVs and MSEVs may be acceptable up to a specified threshold, although their ratings may be lower for greater observed violation severities. An approach for evaluating the vehicle for different categories is shown in Table 14. This approach directly compares the severity for each metric violation across the scenarios.

Table 14. Summary Table of Independent Metric Scores for Simulated Scenarios

Scenario Type	Description	MSEV	PRV	CIV (Impact Mode, Severity)	PAV	TLV
Car-Following	CF_LB_C	0%	0%	Fail (Frontal, 0.005)	94%	100%
	CF_LB_NM	10%	42%	Pass (N/A, N/A)	85%	100%
	CF_LB_NE	65%	100%	Pass (N/A, N/A)	85%	100%
Intersection	I_LT_C	0%	0%	Fail (Side, 0.162)	100%	100%
	I_LT_NM	11%	63%	Pass (N/A, N/A)	81%	100%
	I_LT_NE	55%	100%	Pass (N/A, N/A)	81%	100%
Lane Change	LC_CI_C	0%	0%	Fail (Frontal, 0.010)	94%	100%
	LC_CI_NM	17%	43%	Pass (N/A, N/A)	76%	100%
	LC_CI_NE	58%	100%	Pass (N/A, N/A)	76%	100%

Yet another possible approach combines the results summarized in Table 13 and Table 14 by establishing category scores. This approach combines the relevant metrics to establish a Nominal Driving Score, Near-Miss Score, and Collision Score. For the data presented here, the proposed Nominal Driving Score is based on the average of the PAV severity and TLV severity; the MSEV and PRV severities are averaged to provide the Near-Miss Score; and the CIV severity establishes the Collision Score as shown in Table 15. This categorization provides the clearest picture with regards to relevant comparisons. For example, the PAV is no longer factored into the occurrence of a collision and the presence of a collision can be isolated from scenarios in which a collision does not occur. As in the previous example, consider again the car-following, no-event scenario resulting in a TLV.

While the near-miss and collision scores would not be impacted by the TLV, the nominal driving score would be reduced from 93% to 43%, again illustrating the negative impact of a TLV on the nominal driving-related vehicle performance.

Table 15. Summary Table of Categorized Metric Scores for Simulated Scenarios

Scenario Type	Description	Nominal Driving Score	Near-Miss Score	Collision Score
Car-Following	CF_LB_C	97%	0%	99%
	CF_LB_NM	92%	26%	100%
	CF_LB_NE	93%	82%	100%
Intersection	I_LT_C	100%	0%	84%
	I_LT_NM	91%	37%	100%
	I_LT_NE	91%	78%	100%
Lane Change	LC_CI_C	97%	0%	99%
	LC_CI_NM	88%	30%	100%
	LC_CI_NE	88%	79%	100%

6.3 CARLA Simulation Methodology

As discussed in the previous section, HVE was utilized to simulate the baseline scenarios defined in the test matrix in Table 3. An additional simulation was then created in CARLA to iterate different variables for the subject and other vehicle in each scenario, providing results for almost 500 scenarios with similar conditions. This was accomplished utilizing an automated script to iterate variables at a predetermined increment over a range of values [48]. The value of CARLA is demonstrated through this task in the ability to automate a large number of scenarios to conduct a sensitivity analysis with respect to the metric calculation results which would take substantial time and effort to compile in HVE. Although CARLA is unable to consider the severity of a collision event since the simulation model is not equipped to handle the dynamics of the collision itself, it can still be used to identify whether a collision occurs and HVE could be leveraged to further explore specific CIV severities. An evaluation of the CARLA scenarios was conducted to consider the effects of the applied variable ranges for the various scenarios. One important aspect of the metrics analysis is the measurement uncertainty for the collection of the observable variables in real-world environments. Simulated data provide the ability to measure exact values; however, when these metrics are considered for real world evaluation, different sensors have varying capabilities with regards to the accuracy and resolution of measurements which will have an impact on the overall metrics calculations. The iteration through variable ranges during this analysis provides additional context to understand the boundaries of a possible collision in addition to the variation of the other metric violations and severities based on differing levels of measurement uncertainty. Furthermore, since measurement uncertainties can result in the variation of actual data (i.e., the measured result may vary above and below the observed value), evaluation of these uncertainties requires more than simply considering the maximum or minimum realistic value. By iterating the variables and contrasting the results for different permutations of uncertainties, a comprehensive sensitivity analysis can be performed. Real-world evaluations utilizing the presented OSA methodology could incorporate known measurement uncertainties to bound performance evaluations and ensure the confidence of the assessment. Additionally, this approach could be utilized to assist in the identification of edge, corner, and long tail cases. These cases highlight the boundaries of a vehicle capabilities, or in some cases, may highlight the boundaries of a metric to accurately evaluate a scenario (i.e., measurement uncertainty limits of a metric). Table 16 summarizes the variables, ranges, and increments for the CARLA simulation iterations.

Table 16. Variable Iteration Table for CARLA Simulations

Variable Name	Variable Range	Iteration Increment
Velocity	±5 mph	1.25 mph
Initial Separation Distance (Longitudinal)	±16.4 ft	3.28 ft
Time of Braking	±2.5 sec	1.0 sec

6.4 CARLA Simulation Results

The MATLAB script was modified for compatibility with the CARLA results for the hundreds of scenarios generated with varied parameters. This automated process provided the capability of undertaking a sensitivity study to understand how the metrics calculations were affected by variations within the initial scenario parameters. Overall, the scenarios resulted in a range of outcomes from avoiding an MSEV altogether, to collision events.

6.4.1 CARLA Scenario Setup

The advantage to utilizing CARLA is demonstrated in this section providing a range of metrics calculations for the specified scenario. First, a baseline scenario was generated with the CARLA simulation tool and was replicated in HVE to establish the general repeatability of results. It should be noted that the fidelity of the vehicle models in CARLA is lower than that of the vehicle models in HVE, as the HVE models include validation efforts with access to far more parameters than the more simplified CARLA vehicle models. The tradeoffs are demonstrated as CARLA features ease of automation of scenarios while HVE provides heightened fidelity. A simple car-following scenario was chosen for this example to illustrate the potential benefits which could be applied to similar scenario types to those generated in the previous HVE simulations.

6.4.2 CARLA Baseline Result Repeatability

The same baseline scenario was generated in HVE to demonstrate the general repeatability of results between the two simulation software. The vehicle parameter inputs were incorporated to the CARLA simulation to achieve similar results to the HVE scenario and mimic the results for the metric violations. Table 17 reports the key metrics violation severities and the deviation between the two simulations to compare the results. The results of the metrics calculations performed for both simulations demonstrate similar results with minimal standard deviations. The collision incident severity was not compared because the baseline comparison involved a near-miss scenario in which the vehicles did not collide. Furthermore, since CARLA is not capable of handling the collision dynamics between two vehicles, the CIV results would not be expected to demonstrate consistency had a collision scenario been analyzed. The slightly higher deviations observed in the PRV and the PAV for the lead vehicle are due to the less stable vehicle model in CARLA with greater variation of the vehicle accelerations. In general, the results between the baseline CARLA and HVE scenarios demonstrated repeatability, allowing for the next step of proceeding with the iterations for the sensitivity analysis.

Table 17. Comparison of CARLA and HVE Baseline Scenario Results

Simulation	MSEV	PRV	CIV	PAV_Lead	PAV_Follow
CARLA	0.8989	0.8133	N/A	0.1780	0.1647
HVE	0.9004	0.6957	N/A	0	0.1511
Std Dev.	0.0008	0.0588	N/A	0.0890	0.0068

6.4.3 CARLA Iteration Results

Once the repeatability was verified between the simulation programs, the initial parameters for the baseline CARLA simulation were iterated according to Table 16. Utilizing the automated script for scenario iteration, 460 simulations were generated from the initial baseline scenario [48]. The resulting metrics scores ranged from 0 to 1 for both the MSEV and PRV severities. The CI metric was only evaluated as a binary metric violation since CARLA is not capable of handling the post-collision dynamics necessary to evaluate the CIV severity. Scatter plots for the results of the metrics calculations for the generated scenarios are included in Figure 46 to Figure 48 to illustrate the range of values measured for the iterated scenarios.

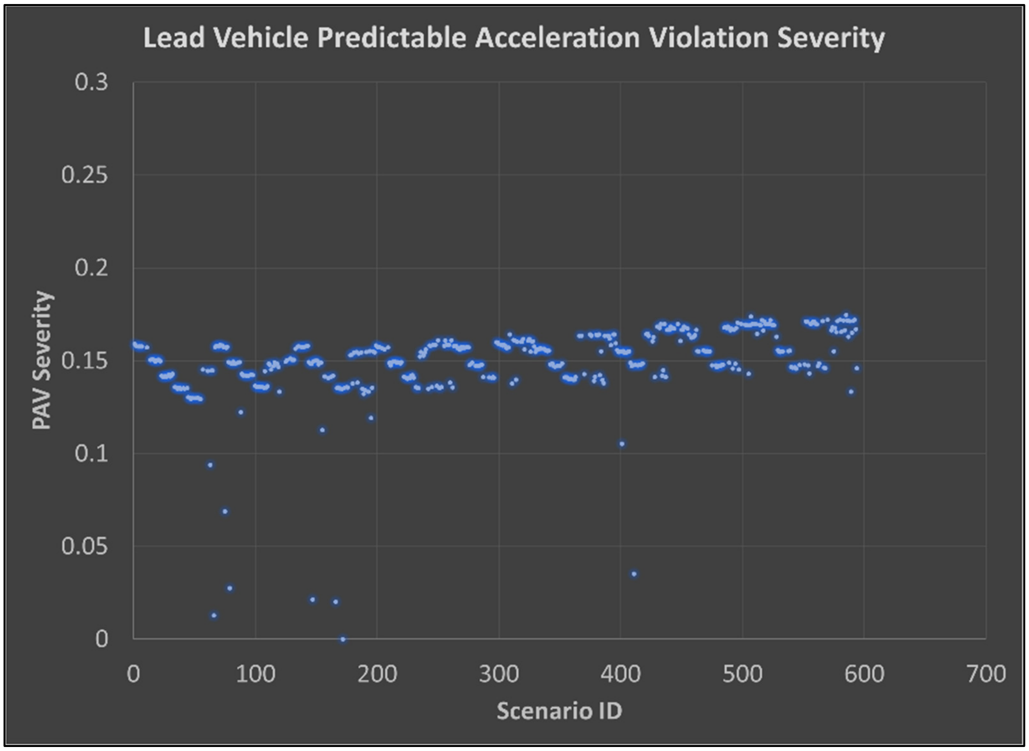


Figure 46. Lead Vehicle PAV Severity for CARLA Scenarios

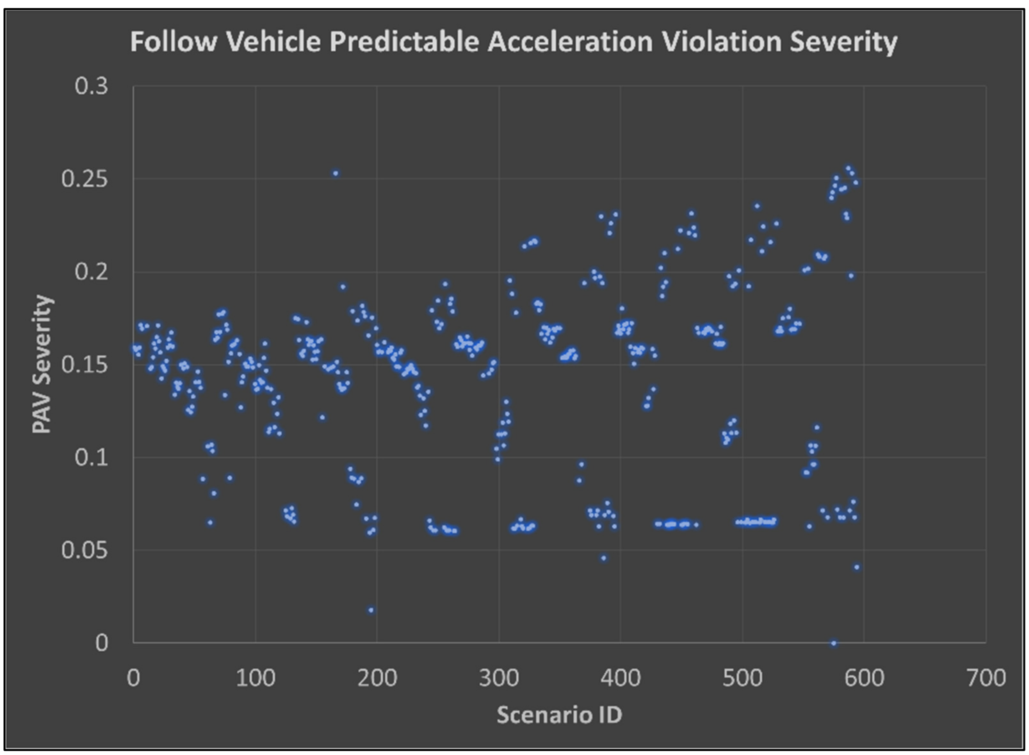


Figure 47. Follow Vehicle PAV Severity for CARLA Scenarios

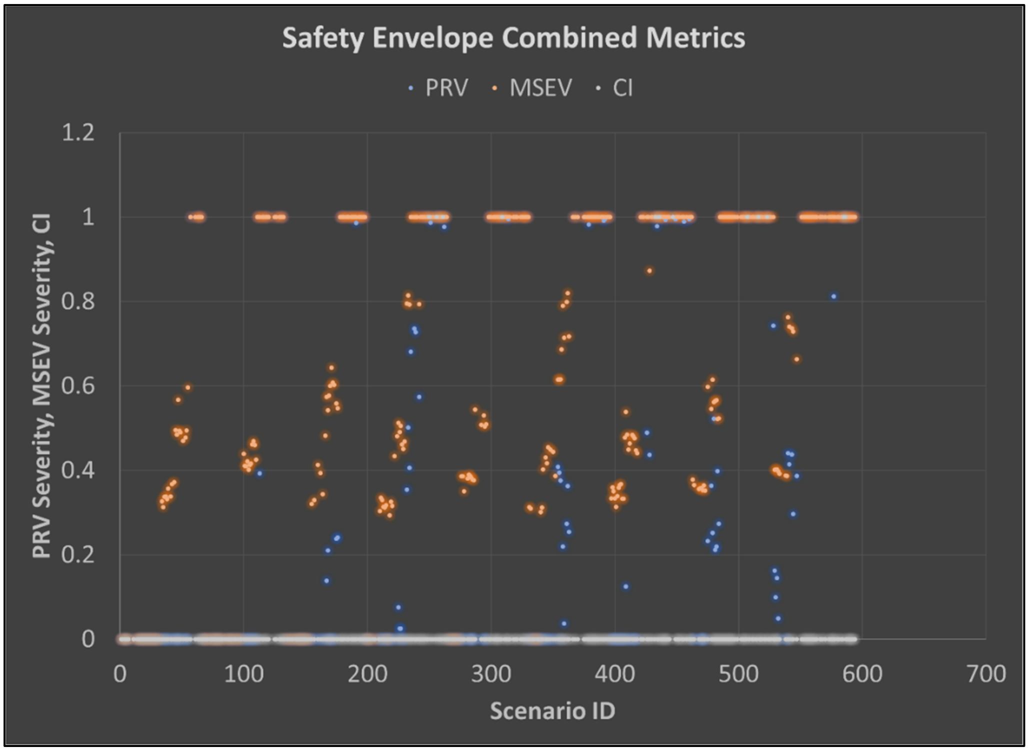


Figure 48. PRV Severity, MSEV Severity, and CI Violation Instances for CARLA Scenarios

This sensitivity analysis demonstrates the importance of considering measurement uncertainty within the OSA methodology and incorporating such analyses in the overall evaluation of a scenario. Given known measurement uncertainty parameters for a sensor modality validated through ground truth testing, this same process could be conducted. As was shown in this section, the iteration parameters resulted in a wide range of scenarios from not even experiencing an MSEV to the occurrence of a collision event. In most real-world systems, it would be expected that the measurement uncertainty would be lower than the values considered here for example purposes, corresponding to less variability in the metrics calculations.

6.5 Example Scenarios in Public Road Data Collection

In addition to the simulated scenarios analyzed, data were collected for real-world scenarios at the intersection of Daisy Mountain Drive and Gavilan Peak Parkway in Anthem, Arizona. This intersection is part of the SMARTDrive Program^{SM} Test Bed developed by the University of Arizona Transportation Research Institute (TRI) in cooperation with the Maricopa County Department of Transportation (MCDOT) [49]. Data were collected using a variety of sensor modalities including LIDAR, drone camera, infrastructure-based cameras, and RT differential GPS. The drone data were processed based on methodologies detailed in [2] to measure the necessary parameters for calculation of metrics violations and severities similar to the simulated scenarios. The data were then post-processed into a format that could be interpreted by the same MATLAB script used to calculate the metrics in the previous sections. A set of car-following scenarios were selected from the captured data to represent varying following distances as outlined in Table 18.

Table 18. Summary of Real-World Scenarios Measured During Public Road Data Collection

Scenario Name	Scenario Type	Scenario Description
CF_1	Car-Following	Traveling through intersection longest following distance
CF_2	Car-Following	Traveling through intersection, normal following distance
CF_3	Car-Following	Traveling through intersection, shorter following distance
CF_4	Car-Following	Traveling through intersection, shortest following distance

6.5.1 Public Road Data Collection Methodology

The real-world data were utilized in a similar manner to the simulated data to evaluate the OSA metrics for a variety of scenarios. Although any vehicles captured in the drone footage during the data collection could have been utilized for this analysis, the initial focus was to utilize scenarios involving the test vehicle to have a ground truth data comparison. The RT differential GPS installed in the Jeep Wrangler test vehicle was capable of recording position with an accuracy of 10 cm for comparison with the tracking algorithm applied to the drone data. Additionally, the test vehicle was scanned utilizing a 3D laser scanner to document the actual vehicle length and width, in addition to providing an accurate measurement of the location of the differential GPS within the vehicle as shown in Figure 49.

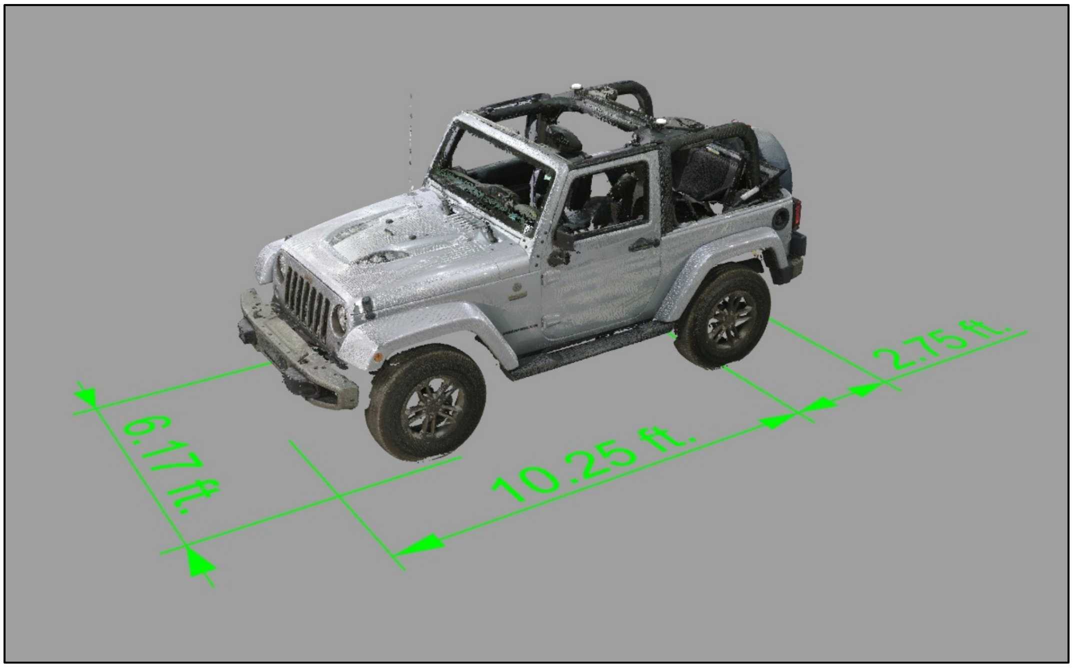


Figure 49. Jeep Wrangler Test Vehicle Laser Scan Data

The purpose of this section is to demonstrate the robustness of the MATLAB script and proposed OSA methodology to evaluate vehicle performance regardless of the data source and even applied to human-driven vehicles. The tracking algorithm detailed in [2] was applied to the collected drone data and the results were compared to the differential GPS data to understand the level of measurement uncertainty introduced by the tracking algorithm. The data were then processed using the MATLAB script to generate the metrics calculations for the algorithm results. An example frame of the drone data with both the Jeep Wrangler test vehicle and ASU Trident 1 test vehicle is depicted in Figure 50 and the tracking algorithm applied to the same frame is shown in Figure 51.

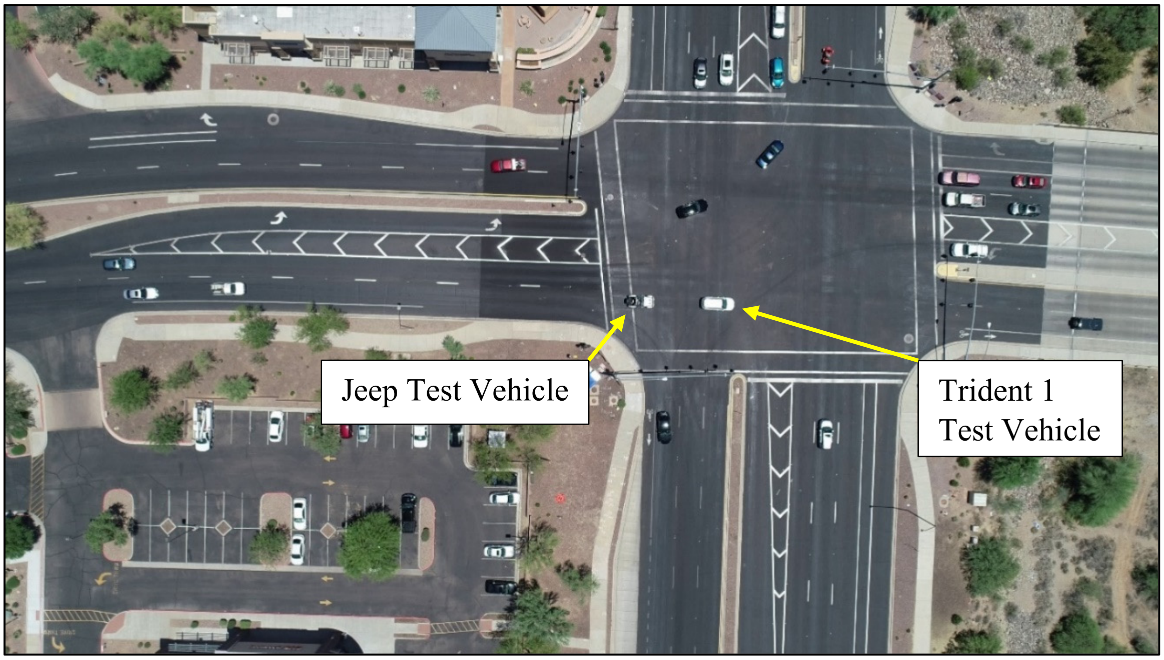


Figure 50. Drone Data Annotated with Jeep Wrangler Test Vehicle and ASU Trident 1 Test Vehicle

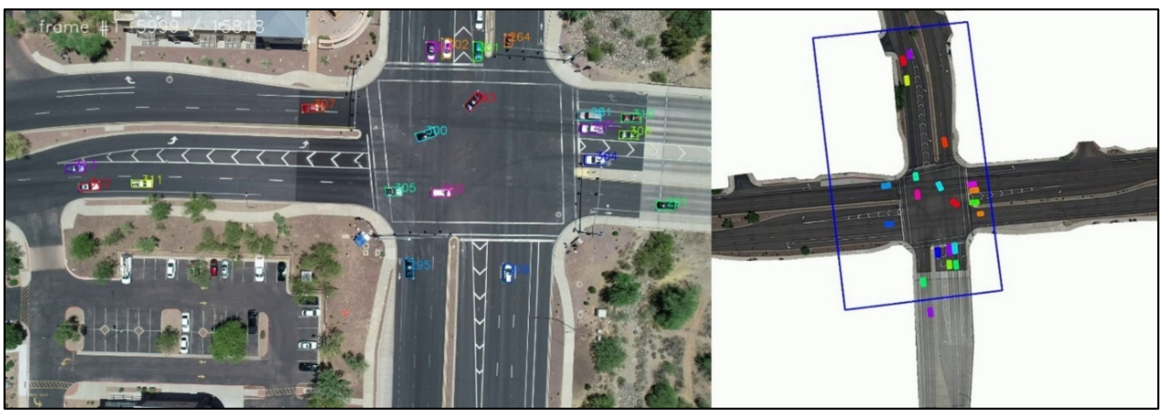


Figure 51. Tracking Algorithm Applied to Collected Drone Data [2]

6.5.2 Public Road Data Collection Results

This section demonstrates yet another use case for the MATLAB script generated throughout this work to showcase the robustness of this tool to calculate the OSA metrics and corresponding severities for a given scenario, regardless of the data source. The ground truth measurements for the test Jeep Wrangler in each of the analyzed real-world scenarios were compared to those estimated by the tracking algorithm and the resulting deviations for the observed variables necessary for the OSA metrics calculations are depicted in Table 19.

Table 19. Comparison of Ground Truth Measurements to Tracking Algorithm

Source		X Pos. (ft)	Y Pos. (ft)	Long Vel (mph)	Long Acc (g)	Length (ft)	Width (ft)
CF_1_Anthem	Avg. Deviation	0.09	0.36	0.47	0.04	1.23	1.05
CF_2_Anthem	Avg. Deviation	0.36	0.66	0.98	0.08	1.20	0.89
CF_3_Anthem	Avg. Deviation	0.13	0.58	0.70	0.12	0.93	0.88
CF_4_Anthem	Avg. Deviation	0.14	0.61	0.88	0.05	2.03	0.95

The metrics violations and resulting severities for the scenarios outlined in Table 18 were calculated utilizing the MATLAB script and are depicted in this section. In contrast to the simulated data, real-world data collection does not conform to idealized measurements obtained from simulated environments as was expressed in Table 19. This is an important consideration when evaluating such scenarios to incorporate uncertainty measurements and conduct sensitivity studies to understand the overall impact on the metrics calculations such as that demonstrated in the previous section utilizing simulation methods to iterate parameter definitions. The four outlined scenarios in the real-world dataset incorporated various following distances between the vehicles to illustrate the corresponding metrics calculations results. Graphs depicting the timing of the MSEVs and PRVs are shown in Figure 52 through Figure 55. For each of these scenarios, a MSEV was observed with a corresponding PRV since both vehicles avoided accelerating or decelerating during these scenarios.

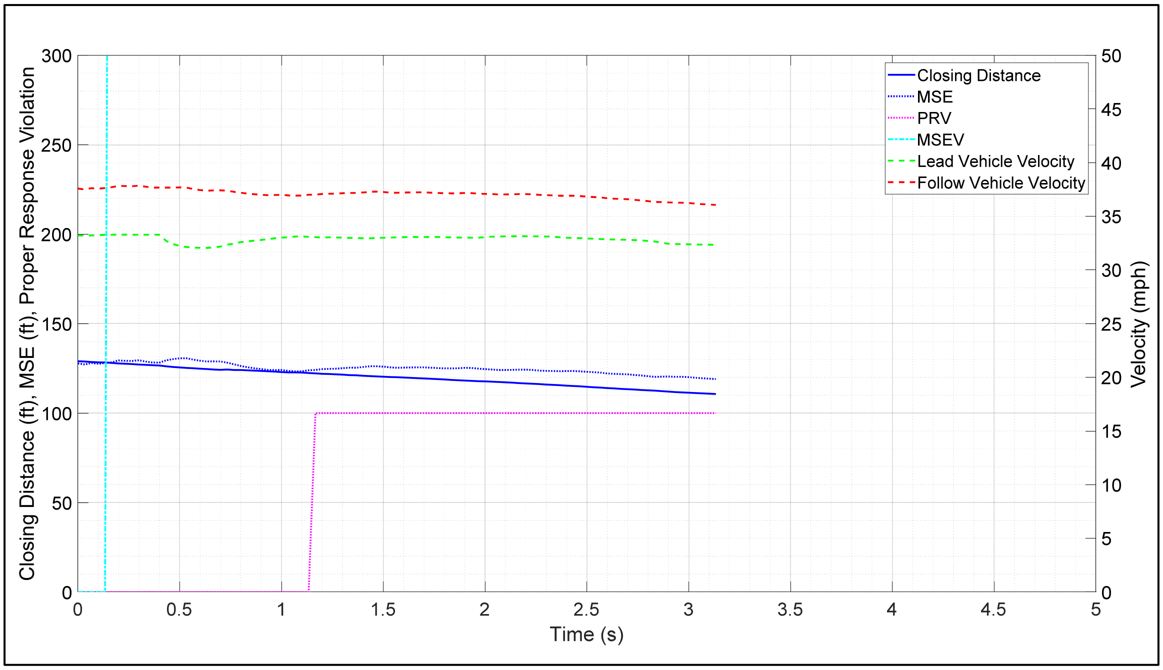


Figure 52. Scenario CF_1_Anthem Initial Conditions and MSEV Plot

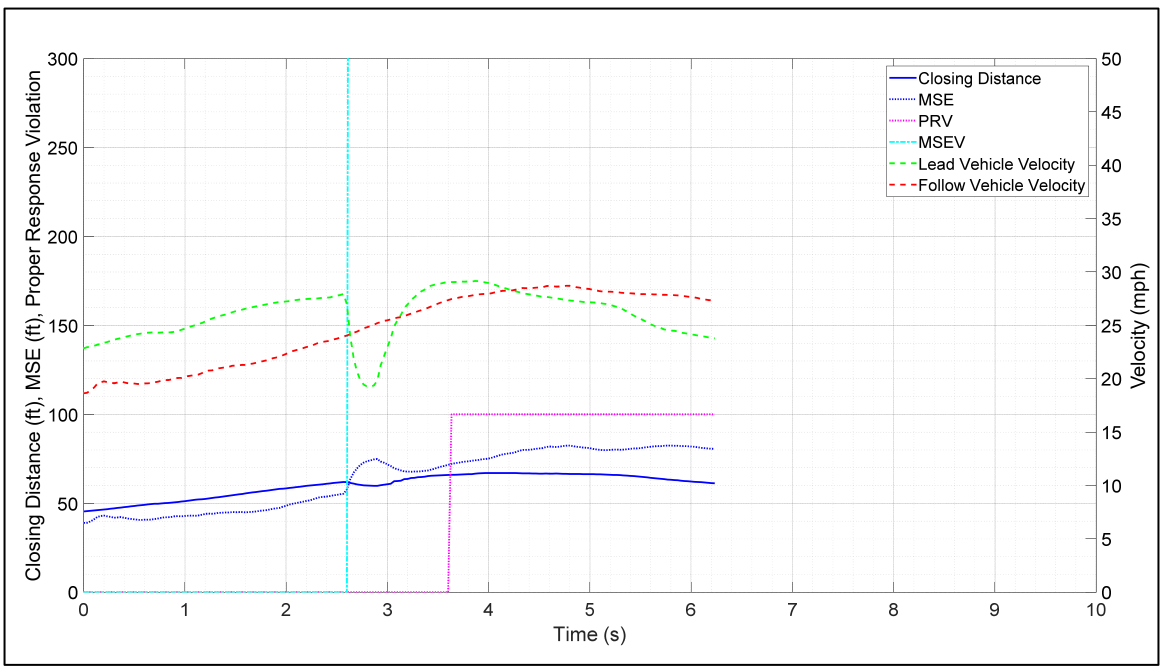


Figure 53. Scenario CF_2_Anthem Initial Conditions and MSEV Plot

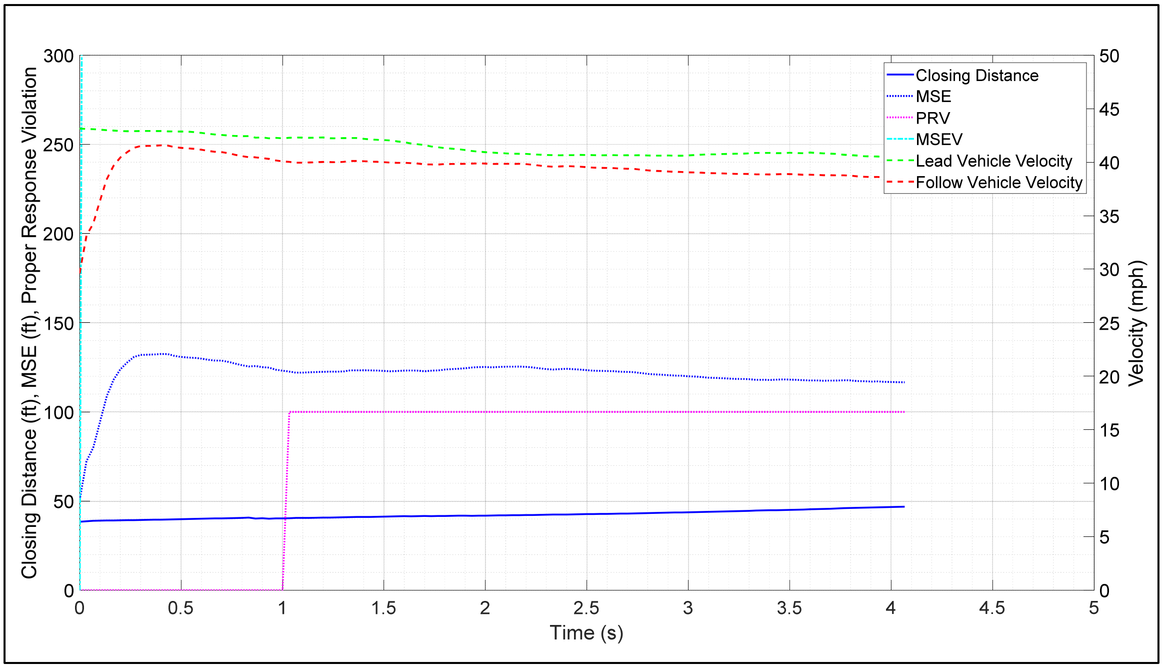


Figure 54. Scenario CF_3_Anthem Initial Conditions and MSEV Plot

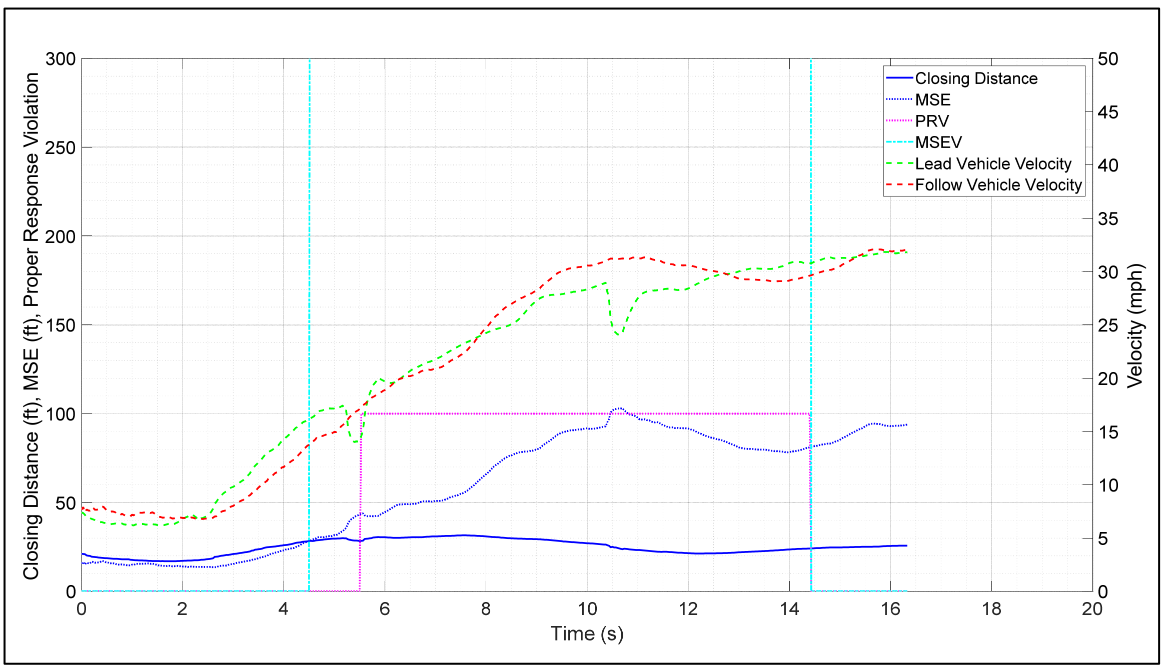


Figure 55. Scenario CF_4_Anthem Initial Conditions and MSEV Plot

The corresponding violation severity of the safety envelope was then plotted to characterize the differences for varying following distances and speeds. As can be seen in Figure 56 through Figure 59, the MSEV severities increased from scenario CF_1 to CF_4 due to the decrease in following distance. The MRD at the beginning of scenario CF_4_Anthem was low, due to the low speeds of the vehicles as they began from a stop at the intersection. As the vehicle speeds increase, so too does the MRD, resulting in the highest overall severity across the scenarios.

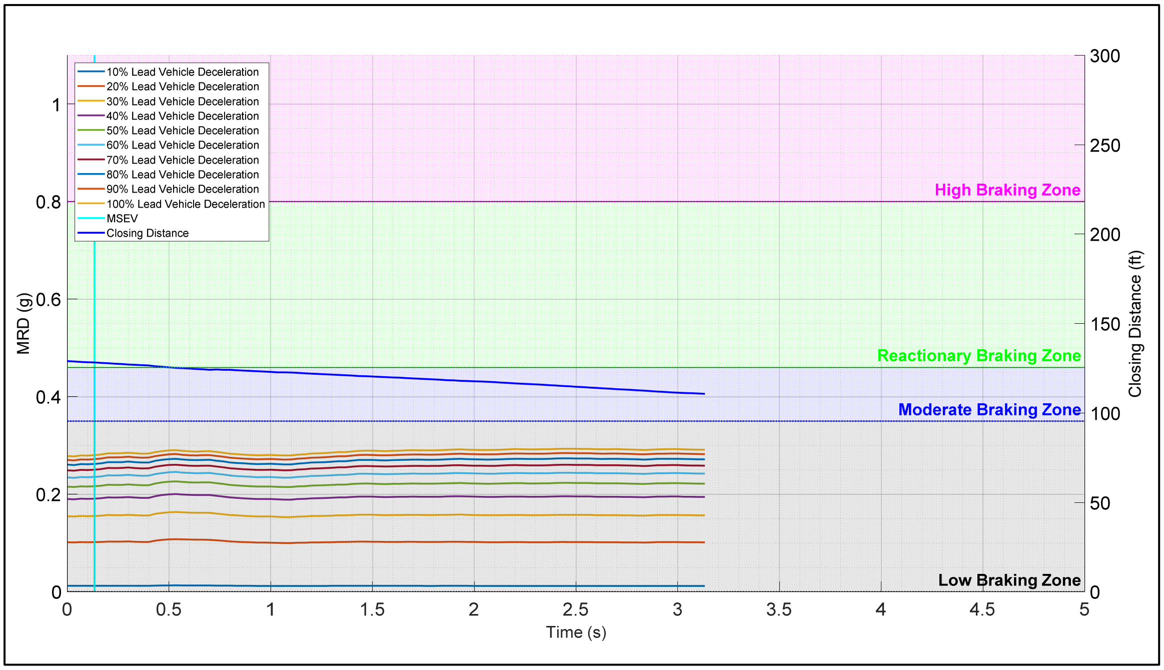


Figure 56. Scenario CF_1_Anthem MRD Evaluation

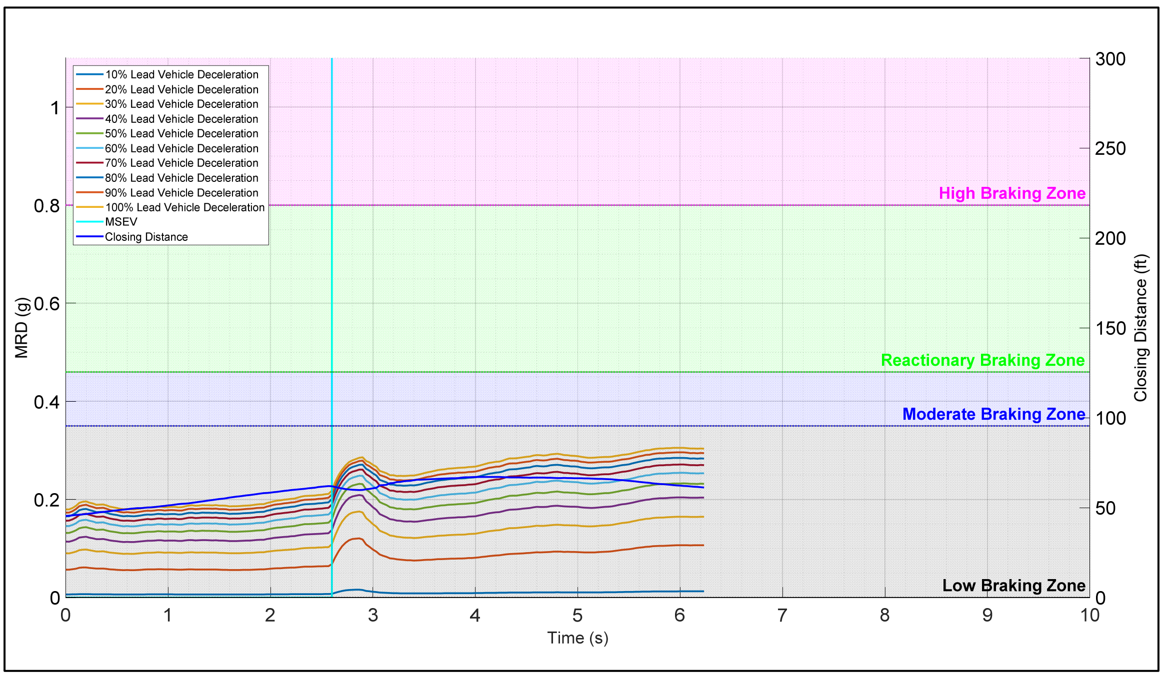


Figure 57. Scenario CF_2_Anthem MRD Evaluation

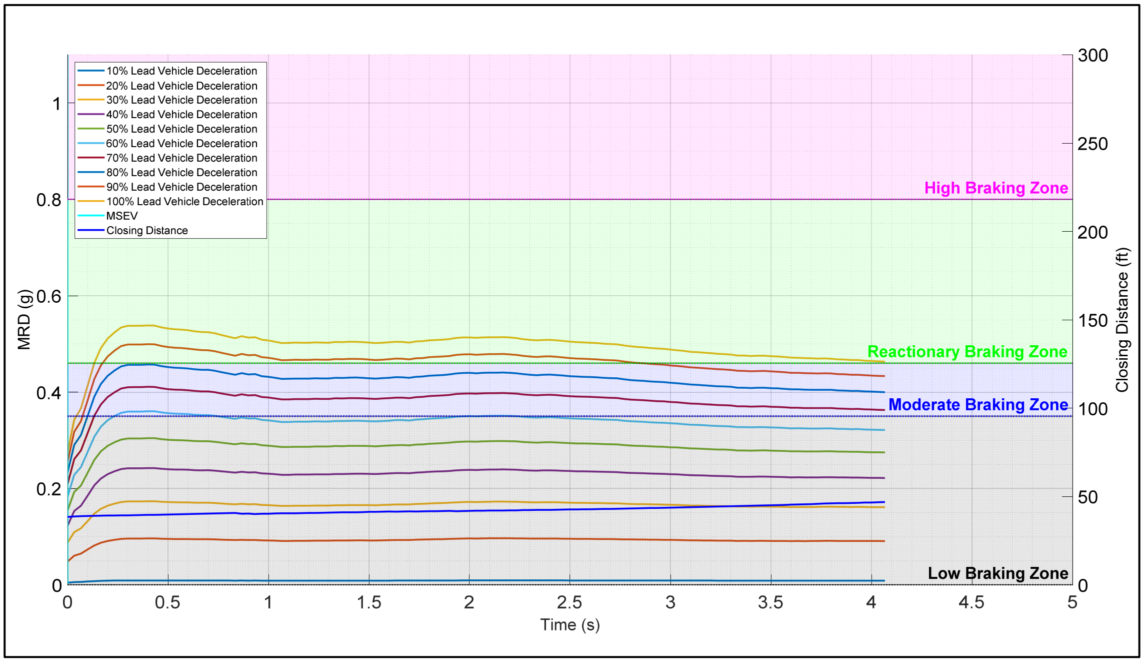


Figure 58. Scenario CF_3_Anthem MRD Evaluation

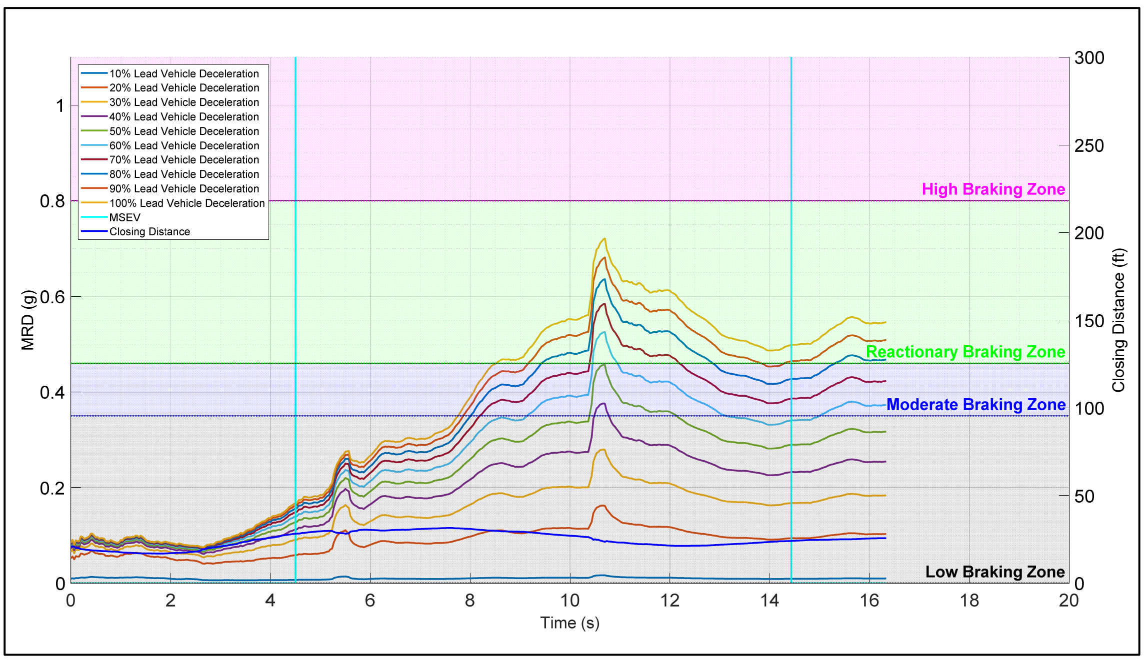


Figure 59. Scenario CF_4_Anthem MRD Evaluation

Similar to the previous section, the overall violation severities were calculated and compared for the four real-world scenarios. Obviously, collision events were not observed as the vehicles were being operated in communication with one another to travel at distances that could be safely maintained without a concern for a near-miss or a possible collision. Similarly, no PAVs were observed because high accelerations were avoided for safety reasons. As a result, the MSEV severity was the only varied scoring parameter, contributing to the overall scores observed in Table 20. Although the nominal driving score and collision score will be 100% for all of the real-world scenarios, considering there were no collisions, TLVs, or PAVs, the near-miss score was relatively poor as an MSEV occurred with no proper response in each of the scenarios summarized in Table 21.

Table 20. Overall Metric Violation Scores for Real-World Scenarios

Description	MSEV	PRV	CIV	PAV	TLV	Score
CF_1_Anthem	0.294	1.000	0.000	0.000	0.000	**74.1%**
CF_2_Anthem	0.305	1.000	0.000	0.000	0.000	**73.9%**
CF_3_Anthem	0.539	1.000	0.000	0.000	0.000	**69.2%**
CF_4_Anthem	0.721	1.000	0.000	0.000	0.000	**65.6%**

Table 21. Categorized Metric Scores for Real-World Scenarios

Description	Nominal Driving Score	Near-Miss Score	Collision Score
CF_1_Anthem	100%	35%	100%
CF_2_Anthem	100%	35%	100%
CF_3_Anthem	100%	23%	100%
CF_4_Anthem	100%	14%	100%

Screenshots from the drone footage captured for each of the four real-world scenarios are depicted in Figure 60 through Figure 63. Consistent with the trajectory estimation data and corresponding MSEV severities, the images show a decrease in closing distance from scenario CF_1 to CF_4. Since only the MSEV severity varied throughout these scenarios, the scores based only on the MSE metric violations are presented in Table 22.

Table 22. MSE Metric Score for Real-World Scenarios

Description	*MSEV*
CF_1_Anthem	71%
CF_2_Anthem	69%
CF_3_Anthem	46%
CF_4_Anthem	28%

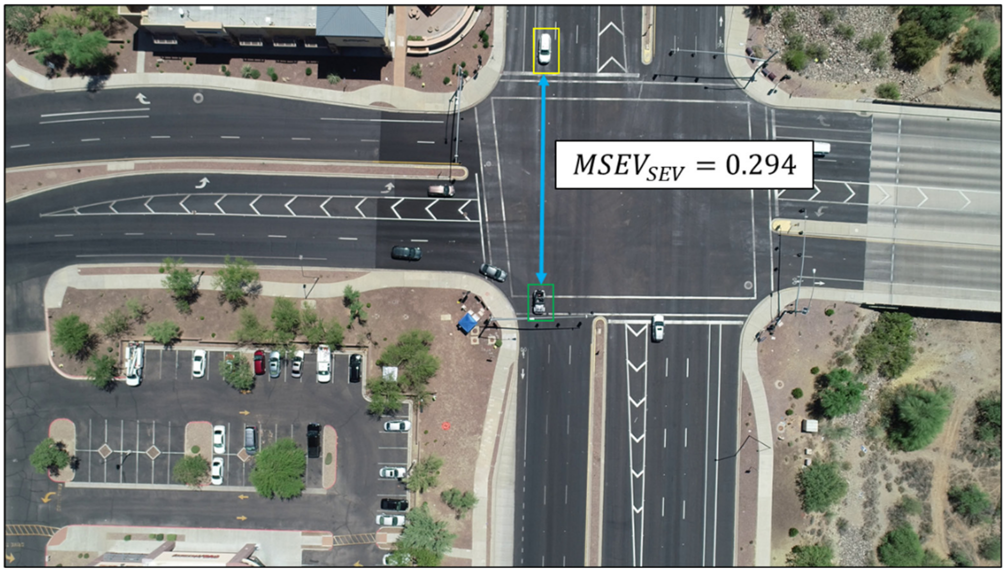


Figure 60. Scenario CF_1_Anthem Drone Image

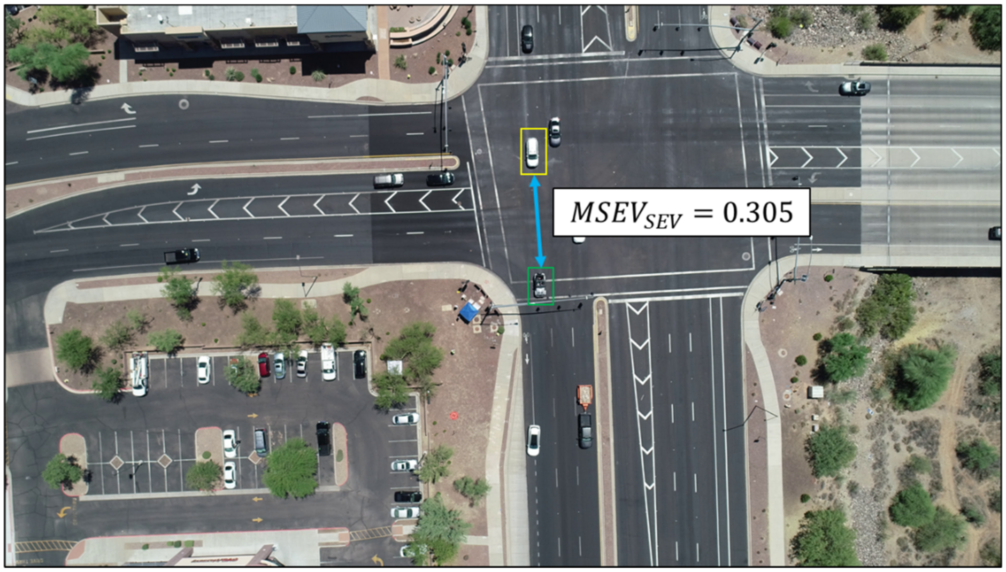


Figure 61. Scenario CF_2_Anthem Drone Image

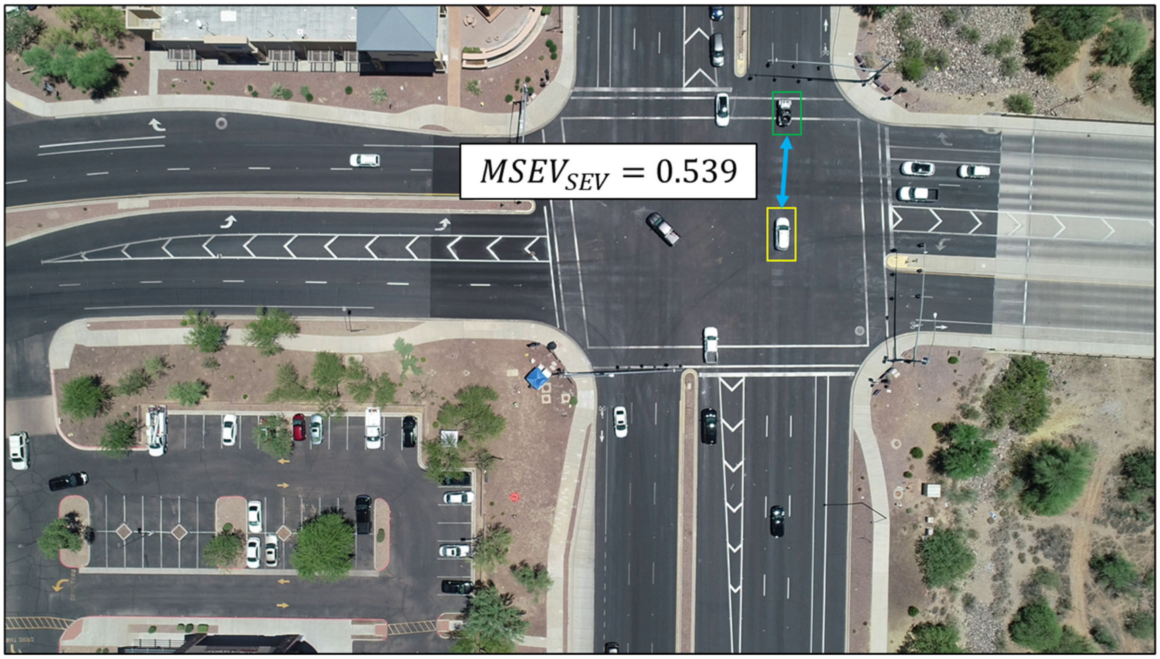


Figure 62. Scenario CF_3_Anthem Drone Image

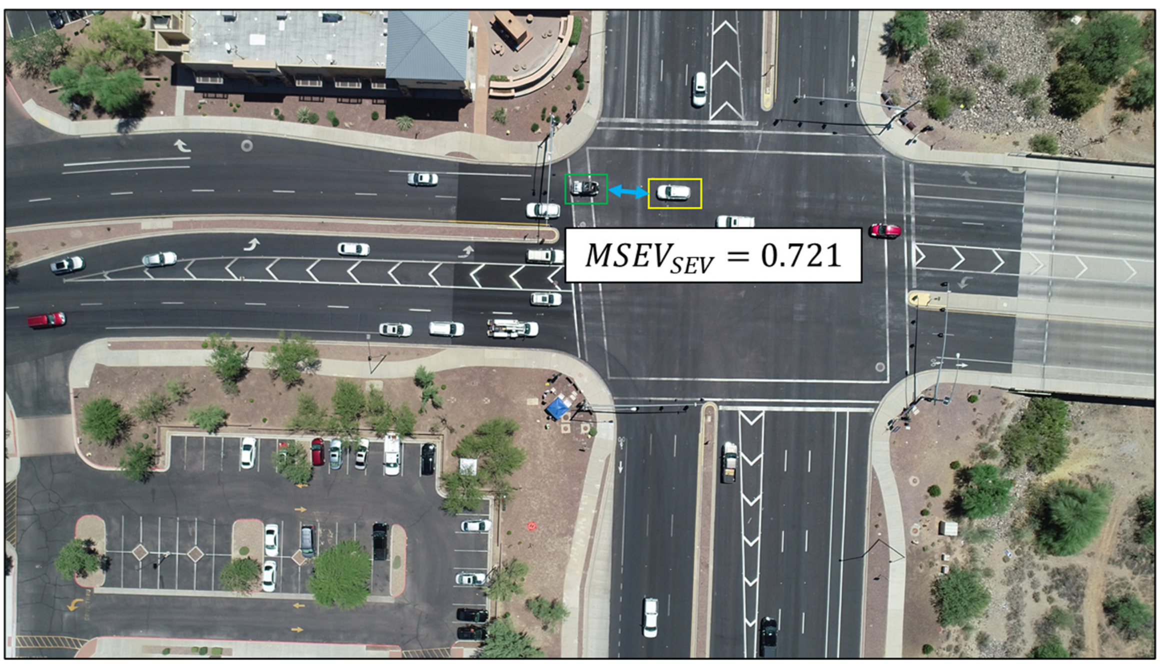


Figure 63. Scenario CF_4_Anthem Drone Image

While the real-world scenarios presented here focus on the MSEV, the usefulness of the MATLAB script is demonstrated in the ability to calculate such metric violations and corresponding severities from any dataset. Real-world testing could be designed (likely in a controlled environment such as a closed course track) to further exercise the other metric violations and corresponding severities. The methods described in these sections could be implemented for a vehicle utilizing the designed script to conduct physical testing, evaluate measurement uncertainty, iterate parameter definitions to consider the sensitivity to parameter changes within the scenario, and generate overall performance scores for various scenarios.

6.6 Comparison of Evaluation Methods

Although HVE and CARLA were the two simulation platforms utilized in this analysis, many simulation software exist to conduct such evaluations. HVE was leveraged for the high-fidelity vehicle models and capability of accurately modeling post-impact kinematics for the collision incident severity calculations. While these benefits led to employing HVE for the majority of the simulation work, CARLA was utilized for its automation capabilities to conduct the aforementioned sensitivity analysis. It is possible that others could utilize a single simulation platform with all of the benefits provided by both CARLA and HVE, in which case the same MATLAB script could be used to generate the metric calculations.

There are advantages and disadvantages to the different test modalities and as a result, a combination of methods will likely be employed in the evaluation of AVs. This section demonstrated the ability of the generated tool to calculate the OSA metrics results regardless of the testing modality. Additionally, this process considers measurement uncertainty to establish confidence in the results. Simulations produce exact values for the involved vehicles; however, they may produce lower fidelity results depending on the complexity and accuracy of the vehicle and environment models. On the contrary, real-world data collection can produce actual results for a given scenario; however, limitations in the data collection methodology may reduce the accuracy of the observable variables. In either situation, the usefulness of the MATLAB script showcased in this chapter is demonstrated by providing the ability to evaluate a scenario and apply parameter iterations to determine the range of metrics violations and severities.

7. COMPLEXITY OF TEST SCENARIO

The previously discussed metric violations and corresponding severities demonstrate an approach to quantify the performance of a vehicle for a given scenario. While this methodology provides a comparable score for vehicles acting within the same scenario, it does not account for a comprehensive evaluation of the vehicles across scenarios. For instance, while it may be easier to design an ADS capable of driving on a straight, flat roadway with no obstacles in ideal weather conditions, there are far more difficult scenarios that must be considered for the ODD of AVs being deployed across the country and around the world. This chapter will discuss the proposed formulation for the complexity factor assigned to a scenario based on factors such as the number of neighboring salient objects, the predictability of those salient objects, and driving conditions that may increase the difficulty of a scenario. Additionally, examples will be provided for each of the following sections to illustrate the purpose of the assigned complexity factor. Each of these factors will be normalized and have a range of 0 to 1, corresponding to the same process which was applied to the OSA metric violation severities. Applying a complexity factor for the OSA score accounts for challenges that may be experienced by an AV and could even be utilized to define the ODD for which the AV may be deployed.

7.1 Number of Salient Objects

According to SAE J3208, a salient object is defined as "Any object (excluding the subject vehicle), dynamic or static, that may have relevance for the safety performance of the DDT [21]." A set of safety metrics, and potential safety metric violations with corresponding severities, exists for each salient object neighboring the subject vehicle. This exposure to additional salient objects creates an opportunity for the same set of potential violations with each object; thus, a multiplier is proposed for the complexity factor of a scenario based on the number of surrounding salient objects. For normalizing the complexity factor, the number of salient objects will have a denominator of 10 (i.e., one salient object is equivalent to a factor of 0.1 while eight salient objects would be equivalent to 0.8 with an overall range from 0 to 1).

7.1.1 Salient Object Example Application

For this example, consider two scenarios for a vehicle approaching an intersection. This first scenario is depicted in Figure 64 in which the subject vehicle (silver Jeep) approaches a signalized intersection to make a right turn while a pedestrian crosses the street. One would expect the subject vehicle to approach the intersection, stop at the stop bar, allow the pedestrian to cross, and proceed through the intersection. Successful completion of this scenario would be communicated by a lack of violations for the metrics summarized in Table 23 and mapped to corresponding actions for the vehicle. The corresponding salient object factor for this scenario is 0.1.

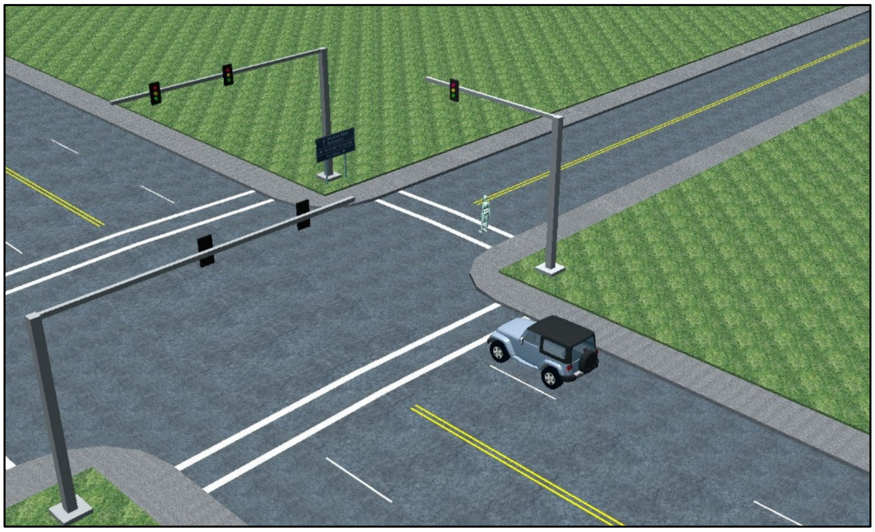


Figure 64. Intersection Scenario Involving One Salient Object

Table 23. OSA Metrics Related to Example Simple Scenario

Expected Actions	Corresponding Metric
Stop at stop bar	Traffic Law
Yield right-of-way to pedestrian	Traffic Law, Minimum Safety Envelope
Complete right turn when clear	Achieved behavioral competency

Next, consider a similar scenario in which the subject vehicle approaches the same intersection with a pedestrian crossing; however, there is now a vehicle located on the opposing side of the intersection intending to turn left across the intersection in addition to a vehicle traveling straight through the intersection on a green signal in front of the subject Jeep as shown in Figure 65. While the actions of the vehicle and corresponding metrics for analysis are generally the same, the increased number of salient objects creates more possibilities for a metric violation to occur. Now, the subject Jeep still must stop for the signal and crossing pedestrian; however, now it also must maintain a minimum safety envelope with the vehicle in front. Additionally, the vehicle intending to turn left across the intersection creates another possibility for a safety envelope violation. The increased complexity of this scenario is summarized within Table 24. In this scenario, the complexity is increased because there is greater exposure to have a metric violation with respect to one of the three salient objects, resulting in a salient object factor of 0.3. In a similar scenario where the lead vehicle was also turning right in front of the subject Jeep, the pedestrian would no longer be counted as a salient object for the subject Jeep because the pedestrian would have already crossed before the lead vehicle was able to turn. In that alternate scenario, the Jeep would no longer be yielding to the pedestrian but waiting behind the subject vehicle; thus, the complexity factor would only account for two other salient objects. These options for varying complexity based on minor variations of the described scenario indicates just how much the complexity can change due to the number of salient objects.

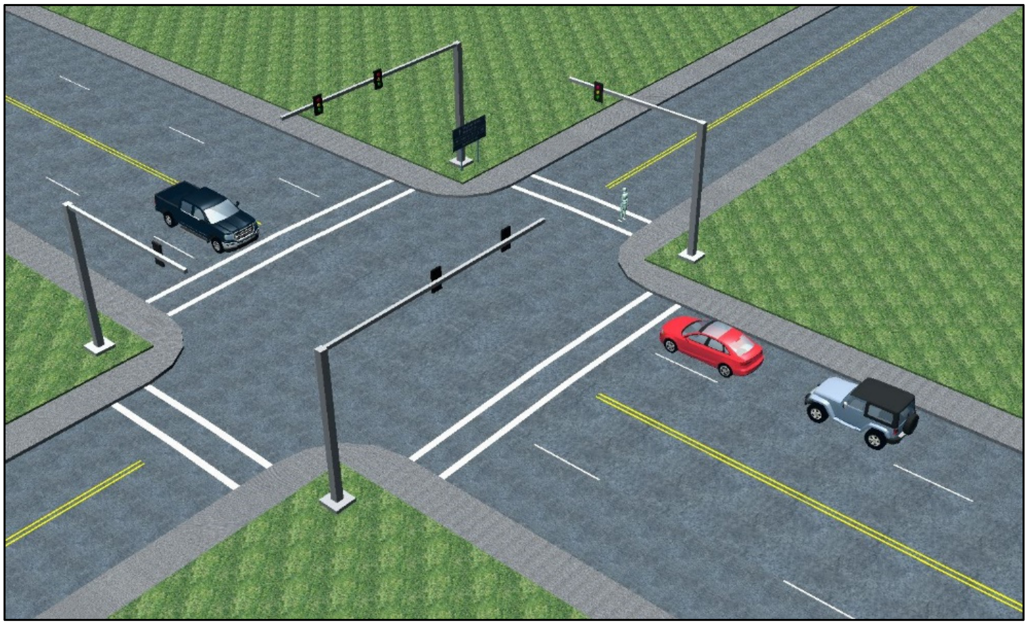


Figure 65. Intersection Scenario Involving Three Salient Objects

Table 24. OSA Metrics Related to Higher Complexity Scenario

Expected Actions	Corresponding Metric
Maintain safe distance behind lead vehicle	Traffic Law, Minimum Safety Envelope (with respect to lead vehicle)
Stop at stop bar	Traffic Law
Yield to pedestrian crossing	Traffic Law, Minimum Safety Envelope (with respect to pedestrian)
Complete right turn when clear	Achieved behavioral competency, Minimum Safety Envelope (with respect to pedestrian and vehicle intending to turn left)

7.2 Predictability Factor

The predictable acceleration metric is one of the proposed OSA metrics for the subject vehicle described earlier to evaluate the level of risk introduced due to increased accelerations. Greater longitudinal and lateral accelerations may reduce the amount of time in which a driver (or ADS) must respond to a given situation. Similarly, the aggressivity of the salient objects surrounding the subject vehicle may increase the difficulty of the test scenario. The proposed formulation for applying such a factor is accomplished in the same method for which the PAV severity was determined in Section 4.6. As such, the longitudinal and lateral components of the predictability factor for the complexity are defined in Equations (21) and (22), respectively. The predictability factors are summed for each salient object in a given scenario.

$$\textbf{Predictability Factor}_{\textbf{Long}} = \sum_{i=1}^{n} \frac{\textbf{PAV}_{\textbf{Long\_i}}}{\textbf{Time}_{Interval}} * \frac{a_{\textbf{Long}}}{a_{\textbf{Long\_Limit}}} \quad (21)$$

$$\textbf{Predictability Factor}_{\textbf{Lat}} = \sum_{i=1}^{n} \frac{\textbf{PAV}_{\textbf{Lat\_i}}}{\textbf{Time}_{Interval}} * \frac{a_{\textbf{Lat}}}{a_{\textbf{Lat\_Limit}}} \quad (22)$$

7.2.1 Predictability Factor Example Application

To review the predictability factor, consider two example scenarios. In the first scenario, the lead vehicle applies light braking to achieve a controlled stop for a traffic signal as shown in Figure 66. In this scenario, the lead vehicle never violates the PA metric; therefore, the predictability factor for the scenario is 0.

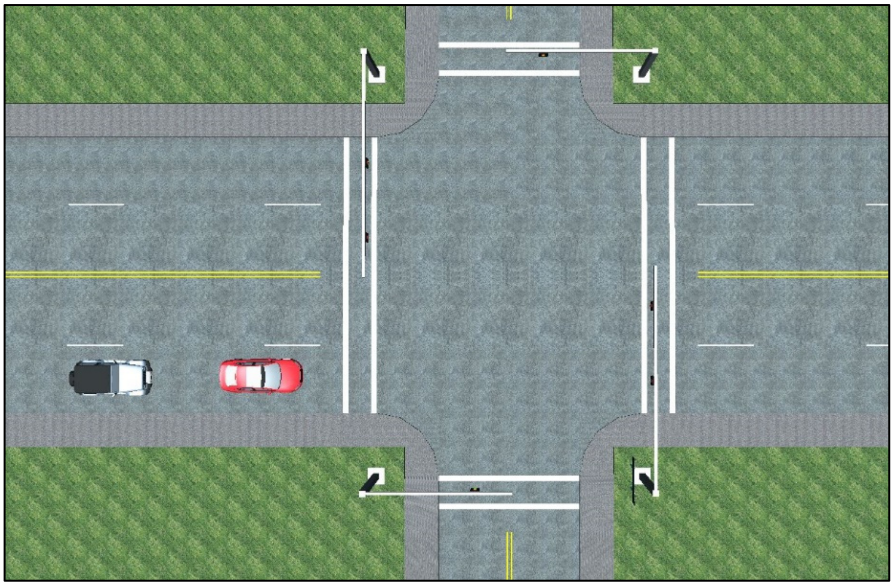


Figure 66. Intersection Scenario with Lead Vehicle Applying Light Braking at Red Traffic Signal

An alternate scenario is depicted in Figure 67 in which the lead vehicle applies aggressive braking of 0.8 g for 1 second to avoid a collision with a left turning vehicle that failed to yield the right-of-way. Given the proposed formulation, and assuming a scenario duration of 5 seconds and a longitudinal deceleration limit of 1.0 g, the resulting predictability factor for this scenario would be 0.16. Intuitively, as the accelerations during the scenario increase in duration and/or magnitude, the corresponding predictability factor (and complexity of the scenario) would increase.

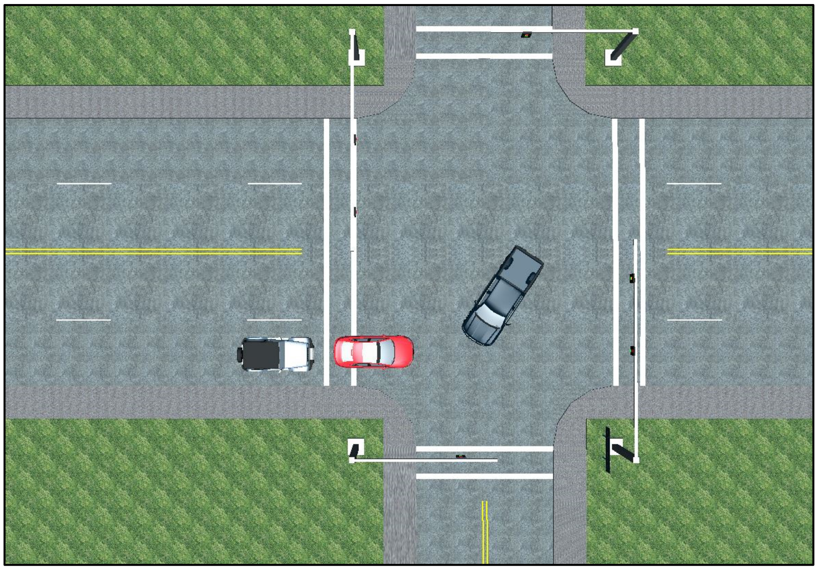


Figure 67. Intersection Scenario with Lead Vehicle Applying Aggressive Braking Due to Left Turning Vehicle

7.3 Surface Factor

The condition of the roadway is another factor which can increase the complexity of a given scenario and can vary considerably based on the ODD of the subject vehicle. Under ideal conditions, the subject vehicle (and other road users) would easily maintain control with corresponding predictable behaviors but as the friction of the road surface is reduced, the difficulty of navigating a scenario is increased. Several references for typical tire-to-surface friction values are included in Figure 68 and Figure 69.

TIRE-ROADWAY FRICTION VALUES

DESCRIPTION OF ROAD SURFACE	DRY UNDER 30 MPH	DRY OVER 30 MPH	WET UNDER 30 MPH	WET OVER 30 MPH
PORTLAND CEMENT				
New, Sharp	.80 - 1.20	.70 - 1.00	.50 - .80	.40 - .75
Travelled	.60 - .80	.60 - .75	.45 - .70	.45 - .65
Traffic Polished	.55 - .75	.50 - .65	.45 - .65	.45 - .60
ASPHALT, TAR				
New, Sharp	.80 - 1.20	.65 - 1.00	.50 - .80	.45 - .75
Travelled	.60 - .80	.55 - .70	.45 - .70	.40 - .65
Traffic Polished	.55 - .75	.45 - .65	.45 - .65	.40 - .60
Excess Tar	.50 - .60	.35 - .60	.30 - .60	.25 - .55
GRAVEL				
Packed, Oiled	.55 - .85	.50 - .80	.40 - .80	.40 - .60
Loose	.40 - .70	.40 - .70	.45 - .75	.45 - .75
CINDERS				
Packed	.50 - .70	.50 - .70	.65 - .75	.65 - .75
ROCK				
Crushed	.55 - .75	.55 - .75	.55 - .75	.55 - .75
ICE				
Smooth	.10 - .25	.07 - .20	.05 - .10	.05 - .10
SNOW				
Packed	.30 - .55	.35 - .55	.30 - .60	.30 - .60
Loose	.10 - .25	.10 - .20	.30 - .60	.30 - .60

Figure 68. Tire-to-Roadway Friction Values [50]

DESCRIPTION OF ROAD SURFACE	AUTOMOBILE TIRE	TRUCK TIRE
DRY CONCRETE	.85	.65
DRY ASPHALT	.80	.60
WET CONCRETE	.70-.80	.50
WET ASPHALT	.45-.80	.30
PACKED SNOW	.15	.15
ICE	.05	.11 (DRY) .07 (WET)
DRY DIRT	.65	--
MUD	.40-.50	--
GRAVEL or SAND	.55	--
WET, OILY, SMOOTH CONCRETE	--	.25
HARD-PACKED SNOW W/ CHAINS	--	.60
DRY ICE W/ CHAINS	--	.25

Figure 69. Passenger Vehicle and Heavy Truck Roadway Friction [51]

Road friction values can depend on a variety of factors ranging from weather to the actions of vehicles operating on the roadway. Figure 70 demonstrates the reduction in friction coefficient values as a vehicle's speed increases. These references provide examples of estimating the friction coefficient for a given surface which is dependent upon the ODD of the vehicle.

VELOCITY DECREMENT VALUES

INDICATED SPEED, MPH	PERCENT REDUCTION OF FRICTION COEFFICIENT
40	3
50	7
60	9
70	11
80	14
90	18

REF. (3), p. 178

PARAMETERS FOR $\mu = \mu_0 - \nu V$

	DRY PEAK	DRY SLIDING	WET PEAK	WET SLIDING
μ_0	0.95	0.85	0.75	0.70
ν	0.0017	0.0025	0.0033	0.0050

Figure 70. Friction Coefficient Reduction Based on Increased Velocity [52]

As shown in Figure 68, a surface friction coefficient of approximately 1.0 is on the upper end of what a typical tire is capable of achieving on a given roadway. As is demonstrated by the references provided, this value can decrease due to differing conditions. Therefore, the surface factor for the scenario is defined in Equation (23):

$$\boldsymbol{Surface\ Factor = 1\ - Assumed\ Friction} \quad (23)$$

If the coefficient of friction achieved is 1.0 or greater, there is no added complexity due to the surface factor for the given scenario. As the coefficient of friction is reduced, the surface factor increases with a corresponding increase in scenario complexity.

7.3.1 Surface Factor Example Application

The surface factor for implementation into the scenario complexity is straightforward. Assume an example scenario where a vehicle is driving in Phoenix, Arizona during the winter where it is sunny and dry. Most of the time, the surface factor in this case will be at or near 0. Conversely consider a scenario in Ann Arbor, Michigan during the winter as the subject vehicle encounters a patch of black ice with a tire-to-roadway coefficient of friction of approximately 0.2. This would result in a surface factor of 0.8, again demonstrating a higher complexity value for a scenario resulting in decreased vehicle control.

7.4 Visibility Factor

In a similar manner, reduced visibility will increase the complexity of a scenario; however, in many cases, AVs will rely on sensors that improve visibility as compared to a human driver (i.e., radar, lidar, thermal). Not all AVs are being equipped with the same sensors and as such, the proposed visibility factor is not dependent on a sensor suite. The visible distance is proposed to be defined as the distance based on the ability of a human with 20/20 vision to identify all salient objects and traffic markings. The visibility factor is then calculated based on the visible distance and the distance it would take for the subject vehicle to stop at the posted speed limit assuming a moderate deceleration of 0.46 g which is the deceleration at which most vehicles brake when confronted with an unexpected obstacle [44] as formulated in Equation (24):

$$\boldsymbol{Visibility\ Factor} = \frac{\boldsymbol{Speed\ Limit^2}}{\boldsymbol{2*0.46g}} / \boldsymbol{Visible\ Distance} \qquad (24)$$

The formulation of the visibility factor is based on the speed limit and human visible distance so as to avoid defining the complexity of a scenario according to the design of a specific AV. To date, manufacturers are utilizing differing sensor suites and varying decision-making algorithms for the navigation of a given scenario. If the complexity of the scenario were based on the visible distance of a particular AV, the vehicle would be penalized by having a lower complexity value due to superior sensing technology compared to a vehicle performing the same scenario lacking access to the same sensors. The speed limit is also incorporated to account for the increased complexity of stopping at a higher speed with reduced visibility to avoid a potential hazard. Furthermore, by incorporating the speed limit, the visibility is normalized by the baseline vehicle stopping distance. The numerator expresses the distance required to stop with a predetermined deceleration of 0.46 g while the denominator is the visible distance as defined previously. A ratio of greater than 1.0 would indicate a scenario in which a human driver would be incapable of seeing the point at which they would be capable of stopping (i.e., creating a scenario of a potential hazard being invisible to the driver until it may be too late to stop), while a ratio below 1.0 would suggest the driver is capable of seeing beyond the point by which they are capable of stopping.

7.4.1 Visibility Factor Example Application

Although Phoenix, Arizona does not typically experience conditions that will increase the surface factor dramatically as described in the previous section, the weather can contribute to a wide range of visibility factors. Consider one example of an AV traveling on a flat, straight highway in Phoenix on a clear day with a visible distance of 1 mile and a speed limit of 70 mph. The resulting visibility factor would be 0.07, indicating low complexity resulting from the scenario visibility. During monsoon season, these weather conditions can change in an instant with dust storms reducing the visibility to blizzard-like conditions. In an alternate scenario, assume a dust storm rolls in resulting in a visible distance of 100 feet. In this scenario, the visibility factor would increase to more than 1.0; however, all of the complexity factors are limited to the range of 0 to 1.0 for normalization. Although the complexity of the scenario increases drastically for this example, a well-equipped sensor suite and robust decision-making algorithms to reduce travel speed as a human driver would likely do in such a scenario would contribute to preventing metric violations in this scenario.

7.5 Behavioral Competency Complexity Factor

When considering the complexity factor for a scenario, the difficulty of the behavioral competency should be evaluated. For instance, the act of stopping at a stop sign is significantly different than performing a lane change on the highway. Although the complexity of the various behavioral competencies will vary, the level of difficulty may also vary depending on the design of the ADS and the sensors employed. Compilation of data through AV testing will likely be required to establish parameters for the behavioral competency complexity factor and could be a focus of future work.

7.6 Complexity Factor Summary

As illustrated in the preceding examples, the complexity factor for a given scenario accounts for the difficulty in navigating a scenario to factor into the overarching OSA methodology. These factors facilitate in normalizing metric violations for scenarios posing substantially greater challenges. The proposed formulation of the total complexity factor is expressed in Equation (25).

$$\boldsymbol{Complexity\ Factor} = \sum \frac{\boldsymbol{Salient\ Object, Predictability, Surface, Visibility, Behavioral\ Competency\ Complexity}}{\mathbf{5}} \quad (25)$$

As explained in the prior sections, the complexity factor was carefully formulated to focus on scenario characteristics contributing to complexity while avoiding reliance on specific AV functionality. This approach avoids providing advantages or disadvantages to a vehicle based on design and bases the complexity directly on the parameters of a given scenario in an objective and discrete manner.

8. FIDELITY OF A TEST METHOD

The fidelity of a test method relates to the ability of an approach to replicate results through a test scenario that would be achieved by the vehicle in a real-world scenario. For example, evaluation of a vehicle on a public road is a high fidelity approach because it is as close as you can get to the real thing, whereas an overly simplified simulation that does not consider variables such as environmental factors, vehicle sensors, and abnormalities in other salient object behavior will have a lower fidelity as it is less representative of what the vehicle would actually experience in a given scenario. Significant research has been conducted utilizing high fidelity simulation [22], [25]; closed-course testing; injection of augmented reality into closed course testing [27]; in addition to other testing modalities. Unlike the complexity factor, the fidelity cannot be discretely quantified in the same manner. This section will introduce and explain aspects that would affect the results of different test modalities, although a single factor will not be proposed for the fidelity as was the case in the other sections. While it may seem obvious that the highest fidelity would be desired for every tested scenario, there are tradeoffs between the resources required to complete a sufficient amount of testing and the accuracy of the results. Furthermore, methods exist to combine test methods such as the augmented reality based closed-course testing described in [27] which leverages the advantages of both simulation and closed-course testing.

Table 25. Advantages and Disadvantages of Various Test Modalities

Fidelity Factors	Simulation	Closed-Course	Public Road
Realistic Environment	○	◑	●
Resource Efficient	●	◑	○
Realistic Vehicle Behavior	○	●	●
Accurate Sensor Representation	○	●	●

○ Weak Performance ◑ Nominal Performance ● Strong Performance

It should be noted that the table above is a general characterization of these test modalities and should not be considered a global truth. As an example, each individual test modality contains a wide array of fidelity. While there are simulation programs in existence that contain over-simplifications of vehicle performance models and low-level modeling of a simulated environment (e.g., a rectangular box moving at constant speed across a flat plane) other simulation programs utilize digital twin models with photorealistic environments in addition to vehicle digital twins with highly accurate operational parameters to attain incredibly high levels of fidelity. As previously stated, this wide variety of existing, and constantly expanding tools, makes establishing a quantified fidelity for various test methods difficult.

8.1 Simulation Environments

As implied in the prior table, a key benefit of simulation is the ability to test a large number of scenarios with minimal resources. While the accuracy of the simulation may be reduced from that of real-world testing, many iterations can be achieved quickly in a simulated environment and sensitivity studies can be leveraged to accommodate for the variation in a given scenario. Beyond the difference in capabilities of varying simulation tools, simulation software may be capable of providing a high-fidelity simulation for one scenario where it may be very low for another. For example, a simulation environment may not be properly defined to accurately simulate wheel friction for wet surfaces and therefore would have a lower fidelity for a scenario involving rainy weather, but it may have a highly accurate representation for a dry scenario. In a similar manner, the vehicle model in the scenario environment could be an exact representation of the subject vehicle but if the visibility of salient objects is not accurate, the perception system for the vehicle in simulation may produce completely different results from physical testing. Just as an AV has a specified ODD, a simulation program may have a limited ODD of validated test scenarios. It is important to understand the limitations of the simulation software being utilized and ensure scenarios within the ODD of the simulation are being considered. Further, validation experiments should be performed to prove the efficacy of the simulated test as it compares to available physical test data.

8.2 Closed-Course Test Approaches

Closed-course testing provides a controllable environment to evaluate scenarios that may pose dangerous situations on public roads such as near-misses and even collisions. The disadvantage of utilizing closed course testing is the amount of time that may be required to set up and test different scenarios. Although it is easier to test specific scenarios in closed-course testing rather than waiting for these instances to occur in public road testing, setting up a test can take substantial time and resources, and iterating for differing variables may be infeasible. The development of new tools to test AVs has improved the efficiency for completing such tests over more traditional crash test methods. One example of such technology is the global soft target (GST) which allows for automated programming of the vehicle paths, easy reassembly when a collision occurs, and repeatability for iterations of various scenario parameters. As previously discussed, augmented reality is another tool that can be utilized in closed-course testing to leverage the benefits of simulation by quickly iterating test parameters and reducing overall test setup; however, this method also carries some of the limitations of simulation such as assuming the perception system performs perfectly as the augmented reality informs the scenario by directly injecting information related an obstacle's presence without the need to have a physical object; thereby, surpassing the vehicle's perception module altogether.

8.3 Public Road Testing Considerations

While public road testing provides the highest fidelity because it represents the actual scenarios faced by a vehicle, significant time and resources would be required to ensure the test vehicle will experience a sufficient array of scenarios to properly test its capabilities. According to a study conducted by RAND, hundreds of millions and possibly even hundreds of billions of miles would need to be driven in order to establish safety of any given AV [53]. Furthermore, without first verifying the capability of an AV, it would be unsafe and unethical to deploy such a vehicle on a public road where other road users are present.

8.4 Test Modality Parameter Error Factor

When setting up a test, whether physical or simulated, there is a level of accuracy that may be achieved based on the tools and equipment being utilized. One aspect of a high-fidelity test will be one that is capable of closely representing the desired test parameters. Some of the designed test parameters may include vehicle speed, starting position, steer angle and rate, among others. There are multiple reasons for the importance of minimizing the test parameter error including repeatability of the test for consistent evaluation and ensuring the intent of the test scenario is properly evaluated (e.g., capturing the bounds of a long tail and edge cases without exceeding them). One such measure of fidelity for a test could incorporate the summation of error for all such test parameters as formulated in Equation (26):

$$\boldsymbol{Test\ Modality\ Parameter\ Error\ Factor} = \mathbf{1} - \frac{\Sigma\left|\frac{\boldsymbol{Measured\ Test\ Parameter} - \boldsymbol{Actual\ Test\ Parameter}}{\boldsymbol{Actual\ Test\ Parameter}}\right|}{\boldsymbol{Number\ of\ Test\ Parameters}} \tag{26}$$

One of the major challenges with this formulation is quantifying every test parameter error for an accurate measurement. For example, the position, velocity, and heading errors for a vehicle in simulation compared to a physical test vehicle may be easy to characterize; however, it could be exceedingly challenging to quantify other errors such as sensor accuracy. Furthermore, the amount of parameters that could be considered may render such an analysis infeasible for any given scenario. By the time a thorough validation is performed for a given parameter set, the resources required to physically test such a scenario may be equivalent.

8.5 Environmental Factors

Digital twins have been implemented to create a photorealistic replica of specific environments in simulation for testing various scenarios [22]. However, when simulation environments are utilized for testing, the fidelity relates to more than just the accuracy of the roadway geometry or topography. Additional considerations include simulation of proper vehicle dynamics, weather conditions, along with other aspects that may be difficult to replicate in a scenario environment. The visibility factor attributed to the complexity of the scenario could be very difficult to establish in a simulated environment while a surface factor could be relatively easily attained with an adequate tire model. While there are limitations in simulation based on the amount of physical data available for implementation and compute power, there are also significant advantages in the ability to test a large range of scenarios very quickly as compared to physical testing. The more detail provided to define the simulation environment and the more closely the simulation environment is able to replicate a real-world scenario, the higher fidelity the simulated test.

8.6 Vehicle Dynamics

Similar to the digital twins that have been utilized for replicating a physical environment, digital twin vehicle models are being used in both the automotive and the aviation industry for a variety of applications. A vehicle digital twin is the same concept as that of an environmental digital twin in that the physical representation of the vehicle is mimicked in simulation from the dimensional components to the functionality and performance. While these digital twins can be used for aspects such as manufacturing and training, they can also serve as vital components for operational safety testing in different scenarios. Without these high-fidelity digital vehicle models, variations in the vehicle dynamics of a simulated vehicle as compared to an actual vehicle can produce results differing from physical scenarios even if the test parameters and environmental factors are accurately represented.

8.7 Salient Object Appearance and Behavior

Like the dynamics of the subject vehicle, the appearance and behavior of the other salient objects in a given scenario are essential to the fidelity of a test. This factor expands beyond just simulation as tools utilized in closed-course testing such as pedestrian and vehicle targets have varying levels of fidelity. Such targets have made major advancements in recent years, representing even accurate radar signatures; however, the standard NCAP pedestrian target wears a black shirt and blue pants. With the practically infinite combination of fabrics, colors, and reflectivity of clothing; varying skin tones; and the varying shapes and sizes of all pedestrians, even physical testing may not be representative of the different scenarios potentially faced by an AV. This is only one example of the many factors that must be considered to accurately represent salient objects in any type of environment, from accurate depiction in simulation to sufficient representation during physical testing.

8.8 Testing Fidelity Summary

This section should have made the difficulty in quantifying the fidelity of a test method apparent. The numerous factors involved in representing an accurate test poses challenges to ensuring simulation can be conducted with adequate confidence levels; however, the infinite number of possible scenarios a vehicle could face establishes the need for utilizing expedited techniques such as simulation to ensure a sufficient set of scenarios are tested. One proposed solution to this dilemma is a certification process of a digital twin for both the environment and the vehicle model. Substantial validation efforts utilizing physical testing would likely be required for such a certification process. Regardless of the process, validation techniques should be incorporated to ensure confidence in the combination of testing conducted to evaluate vehicle performance. Significant efforts are currently underway by standards organizations such as the simulation task force under the SAE ORAD V&V committee to incorporate best practices and lessons learned into the development of simulation approaches.

9. RELEVANCE OF A TEST SCENARIO

The relevance of a test scenario is directly related to the ODD for a given test vehicle. If a given scenario falls outside the realm of the vehicle ODD, that particular scenario could be potentially disregarded for the evaluation of the vehicle since the ADS should never be responsible for the DDT under such conditions. For example, if the ODD specifies dry conditions and rainy conditions fall outside the ODD, any scenarios involving rain would be considered outside of the ODD and would be irrelevant for that particular vehicle; although a manufacturer may still decide to test varying conditions for future applications or expansions of an ODD. This chapter will discuss a proposed method for determining the relevance of a test scenario for consideration in the OSA methodology framework.

9.1 Operational Design Domain Applications

SAE J3016 defines an ODD as:

Operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics [54].

Examples of a specific ODD restrictions include snowy conditions, driving at speeds below 45 mph on rural roadways, or even operation within a particular region. One of the driving factors in determining relevance of a scenario is first considering whether it falls within the ODD of the subject vehicle for evaluation. The normalization convention utilized for the other factors of the OSA methodology framework will be once again utilized in this chapter in formulating the relevance factor. Previously a larger number was consistent with a higher severity of a violation and a greater complexity factor for a scenario. The proposed relevance factor will then utilize a 1.0 for maximum relevance and a 0.0 for a scenario falling outside of the ODD, thus rendering it irrelevant and not counting for or against the vehicle in a given evaluation. For this initial determination based on the subject vehicle ODD, the relevance will receive an indicator of either relevant or irrelevant with more granular considerations to be described in the following sections.

9.2 Behavioral Competency Requirements

While ABC is considered within the set of the OSA metrics, this is also an important component of scenario relevance. As discussed earlier, numerous lists have been established such as the California PATH program required behavioral competencies list [42] in addition to the AVSC best practice document [43]. In contrast to the ODD designations of relevant versus irrelevant, the premise of the required behavioral competencies for a vehicle is to prove the AV is capable of navigating at least a baseline set of scenarios that are considered to be essential for on-road operations. Therefore, in a similar manner to the ODD binary application of relevant or irrelevant, any scenario falling within the category of required behavioral competency could be assigned a 1.0 for the relevancy factor. The ODD consideration takes priority over any of the considerations discussed in these preceding sections since the ADS is not designed to operate outside of its ODD and should not be expected to do so.

9.3 Frequency of Occurrence Estimations

On a more granular level than the required behavioral competencies, an estimation of the frequency of occurrence can be considered for the relevance factor. This analysis is dependent on naturalistic driving datasets and may vary drastically from region to region. For instance, parallel parking is a frequent encounter for drivers living in crowded cities while someone living in a rural area may never have a need to parallel park. An additional complication in the implementation of a frequency factor is the unknown surrounding the activity of any given vehicle. A particular AV make and model sold within the United States will likely have a single defined ODD and will not change based on where it is sold within the United States. However, the frequency of events may change drastically across states based on factors such as weather conditions, population density, and infrastructure. Furthermore, even if the frequency of occurrence was adapted based on location, people relocate frequently and as such, there is no way to predict where a vehicle may operate with respect to region, climate, or even rural versus city settings.

To counteract the uncertainty and variability amongst the aforementioned factors, a proposed implementation of the frequency of occurrence is usage of data defining the entirety of a potential operating region. For instance, an AV designed to be sold in the United States could leverage studies such as the pre-crash scenario typology published by NHTSA [37]. The NHTSA pre-crash scenario typology leverages synthesized real-world data to assign relative frequencies to 37 scenarios accounting for all light-vehicle crashes as shown in Table 26. Similar analyses have been completed within the same study to map a concise list of scenarios to more granular datasets such as all single light-vehicle crashes; two-vehicle, light-vehicle crashes; and multi-vehicle, light-vehicle crashes.

Table 26. NHTSA Defined Pre-Crash Scenarios with Corresponding Frequency Analysis [37]

No.	Scenario	1-Frequency	Frequency	Rel. Freq.
1	Lead Vehicle Stopped	974,855	975,000	16.41%
2	Control Loss Without Prior Vehicle Action	528,930	529,000	8.90%
3	Vehicle(s) Turning at Non-Signalized Junctions	434,892	435,000	7.32%
4	Lead Vehicle Decelerating	428,067	428,000	7.20%
5	Road Edge Departure Without Prior Vehicle Maneuver	333,706	334,000	5.62%
6	Vehicle(s) Changing Lanes – Same Direction	338,309	338,000	5.69%
7	Animal Crash Without Prior Vehicle Maneuver	305,102	305,000	5.13%
8	Straight Crossing Paths at Non-Signalized Junctions	263,840	264,000	4.44%
9	Running Red Light	253,618	254,000	4.27%
10	Vehicle(s) Turning – Same Direction	221,791	222,000	3.73%
11	LTAP/OD at Signalized Junctions	220,206	220,000	3.71%
12	Lead Vehicle Moving at Lower Constant Speed	209,610	210,000	3.53%
13	LTAP/OD at Non-Signalized Junctions	189,816	190,000	3.19%
14	Backing Up Into Another Vehicle	130,701	131,000	2.20%
15	Vehicle(s) Not Making a Maneuver – Opposite Direction	123,699	124,000	2.08%
16	Control Loss With Prior Vehicle Action	102,617	103,000	1.73%
17	Vehicle(s) Drifting – Same Direction	97,973	98,000	1.65%
18	Following Vehicle Making a Maneuver	85,373	85,000	1.44%
19	Road Edge Departure With Prior Vehicle Maneuver	67,528	68,000	1.14%
20	Road Edge Departure While Backing Up	65,809	66,000	1.11%
21	Object Crash Without Prior Vehicle Maneuver	54,526	55,000	0.92%
22	Evasive Action Without Prior Vehicle Maneuver	56,199	56,000	0.95%
23	Vehicle(s) Parking – Same Direction	48,138	48,000	0.81%
24	Running Stop Sign	48,296	48,000	0.81%
25	Non-Collision Incident	45,910	46,000	0.77%
26	Vehicle Failure	42,147	42,000	0.71%
27	Pedestrian Crash Without Prior Vehicle Maneuver	39,324	39,000	0.66%
28	Vehicle Turning Right at Signalized Junctions	34,951	35,000	0.59%
29	Object Crash With Prior Vehicle Maneuver	30,301	30,000	0.51%
30	Pedalcyclist Crash Without Prior Vehicle Maneuver	24,071	24,000	0.41%
31	Animal Crash With Prior Vehicle Maneuver	23,322	23,000	0.39%
32	Pedalcyclist Crash With Prior Vehicle Maneuver	18,325	18,000	0.31%
33	Pedestrian Crash With Prior Vehicle Maneuver	17,118	17,000	0.29%
34	Lead Vehicle Accelerating	18,722	19,000	0.32%
35	Vehicle(s) Making a Maneuver – Opposite Direction	15,472	15,000	0.26%
36	Evasive Action With Prior Vehicle Maneuver	13,120	13,000	0.22%
37	Other	35,859	36,000	0.60%

It should be noted that the 37 pre-crash scenarios listed in Table 26 are based on a statistical analysis of real-world collisions involving human drivers across the United States. While this analysis provides a starting point for determining relevancy of potentially hazardous scenarios, additional data will be important to consider as AV-specific data may indicate differences between high percentage collision scenarios for human drivers as compared to AVs.

The 37 pre-crash scenarios highlighted by NHTSA do not necessarily correspond to the most frequently encountered scenarios, but the most frequent collision scenarios. Rather than proposing the relevance factor based on the expected occurrence of a given scenario, the relevance factor here is proposed to consider the most frequent collision scenarios. This methodology reduces the impact of variance for a given scenario occurring based on geographic location, daily driving route, etc. while highlighting the importance of the evaluation of scenarios that are known to create the most potential hazards. As previously noted, these statistics should be constantly monitored, especially as more AVs are deployed and additional crash statistics are available; however, they serve as a promising starting point for determining scenario relevance.

9.4 AV Use-Case Considerations

To date, AVs are being deployed for a variety of applications, including but not limited to, robo-taxis, delivery services, long-haul trucking, and shuttles. These different applications will likely have differing ODDs and should be considered in determining the relevance of a scenario. For instance, a long-haul trucking AV will spend a large majority of time on freeways in the specified region of operation, whereas a delivery AV may be restricted to roadways with speed limits of 35mph or less and likely won't be designed to transport people (e.g., Nuro). Additionally, it makes little sense to assign relevancy of a scenario for an automated heavy truck based on the NHTSA pre-crash scenario typology for light-vehicle collisions. Therefore, appropriate datasets should be considered when mapping relevant scenarios for a specific AV application for incorporation into the OSA methodology.

9.5 Scenario Relevance Summary

As discussed, the scenario relevance is a challenging topic given the wide range of AV applications and array of potential scenarios faced even for vehicles designed for the same ODD. As such, the scenario relevance factor should be determined with the consideration of all impacting factors. First, the specific ODD for the subject vehicle must be determined to account for relevant versus irrelevant scenarios. Next, an appropriate dataset should be chosen based on the specific AV application. The proposed scenario relevance factor is generalized in Equation (27) to weigh scenarios of greater potential crash frequency although the sources for determining the exact values may vary.

$$\boldsymbol{if\ Scenario\ is\ within\ ODD, Relevance\ Factor = Relative\ Frequency\ of\ Crash\ Occurence}$$

$$\boldsymbol{else, Relevance\ Factor = 0} \quad (27)$$

Statistical approaches can be implemented to establish a scenario relevance or existing literature could be leveraged to identify existing studies providing the necessary data to calculate relevancy. Although the scenario relevance could vary between AVs, it remains an essential element to the overall OSA methodology framework to ensure scenarios are being evaluated in a proper manner.

10. OSA METHODOLOGY REVIEW

The OSA methodology is a general approach that has been described throughout this dissertation as a possible solution to evaluating the performance of a vehicle for a given scenario, whether that vehicle is controlled by an ADS or a human driver. One of the major takeaways from this work is the ability to quantify vehicle performance can occur in numerous forms and there is not a single correct evaluation method. The approaches discussed here are example implementations that show promising results for comparing the performance of vehicles across a variety of scenarios and test modalities including real-world data collection and simulation. One of the current frameworks in use for assessing vehicles prior to public deployment is the voluntary safety self-assessment (VSSA) established by NHTSA in [55]. While the VSSA offers a framework for manufacturers to highlight vehicle performance and improve transparency with the public, it lacks consistency and the granular detail required to evaluate the technical aspects of many AVs. The OSA methodology proposed within this work offers a solution to establishing a more consistent framework for quantifying vehicle performance within a given scenario through the utilization of a set of OSA metrics in addition to assigned metric violation severity, complexity, and relevance factors. As noted in earlier sections, the fidelity of the test modality is also an important consideration; however, this aspect will likely require some form of validation for use in evaluating scenarios.

While this document proposes formulations for the various factors impacting an overall OSA score, this framework could easily be adapted to accommodate additional OSA metrics, refined formulations for specific factors, and incorporation of additional factors, all of which could be examined in future work. One Black Box metric that will be incorporated in future work is the lane stability metric which evaluates nominal driving performance of a vehicle by evaluating the ability of the subject vehicle to maintain its lane. The tools developed in this dissertation including a generalized MATLAB script for calculation of the OSA metrics and corresponding violations from any given dataset may serve as key tools for further refining the concepts described throughout. Additionally, the processes developed throughout the completion of this work involving simulation testing, rapid iteration of scenarios through automated processes, and real-world data collection activities demonstrate effective methods that could also be used to expand upon this methodology in future work.

The methodology presented here could be implemented through real-time calculation of the metrics provided sufficient compute power is available. Through real-time implementation of the metrics, information regarding a violation could be relayed to the vehicle as they occur, providing a potential for corrective actions utilizing the OSA metrics for guidance. Conversely, this methodology could be leveraged to generate a score for the subject vehicle at the conclusion of a scenario. This approach would not require the same computational workload as the data elements necessary for the calculations discussed previously could simply be recorded and post-processed for the metric evaluations.

The necessity of real-time computation versus post-processing of the metrics calculations is dependent on the use-case for the evaluation. Possible use-cases of the OSA methodology may range from a design tool by an OEM to consider performance of their vehicle based on the proposed metrics to a third-party evaluation process that could be implemented to certify a vehicle based on completion of a set of scenarios with inputs from some regulatory body such as NHTSA. In the event that the OSA methodology is being utilized to offer feedback to the subject vehicle to facilitate the communication of warnings and provide data to dictate vehicle actuation, real-time computation of the metrics would be necessary and the latency of the system should be incorporated into the measurement uncertainty of the collected data elements. For instances which the OSA methodology is used purely as an evaluation tool to determine how the subject vehicle performed in a given scenario, post-processing of the metrics calculations would likely suffice, utilizing the data elements collected during the vehicle maneuver. Given the appropriate compute resources for data element collection and metrics calculations, either approach could be employed.

While much of the focus of this work revolved around the formulation of OSA metric violations and severities, the scenario complexity and relevance as well as the test fidelity are important considerations that could be expanded upon in future efforts. Again, the framework has been laid out to accommodate the simple refinement of additions, subtractions, or adaptations of the discussed elements. As AVs continue to be deployed, additional datasets may lead to verification and validation efforts of the proposed OSA methodology and provide useful insights into important considerations that may not have been implemented at the time of this work. While the question remains "How safe is safe enough?", a crucial step towards measuring safety-related performance of AVs has been proposed in this work, providing a potential framework for helping answer this question and hopefully contributing to safer roadways around the world.

REFERENCES

[1] J. Wishart, S. Como, M. Elli and e. al., "Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles," *SAE International,* Vols. 2020-01-1206, 2020.

[2] D. Lu, V. C. Jammula, S. Como, J. Wishart, M. Elli, Y. Chen and Y. Yang, *CAROM - Vehicle Localization and Traffic Scene Reconstruction from Monocular Cameras on Road Infrastructures,* International Conference on Robotics and Automation (ICRA), (In Development).

[3] M. Elli, S. Como, J. Wishart and J. Weast, "Evaluation of Operational Safety Assessment (OSA) Metrics for Automated Vehicles in Simulation," in *SAE World Congress (WCX)*, Detroit, 2021.

[4] V. Chandra Jammula, J. Wishart and Y. Yang, "Evaluation of Operational Safety Assessment (OSA) Metrics for Automated Vehicles Using Real-World Data," SAE Technical Paper 2022-01-0062, 2022.

[5] J. Wishart, *Personal communication,* 2022.

[6] SAE International, "Recommended Practice J3237: Operational Safety Assessment (OSA) Metrics for Verification and Validation (V&V) of Automated Driving Systems (ADS)," SAE International, In Progress.

[7] J. Zhang, 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,* Wiley, 2020.

[8] M. Hussein, T. Sayed, K. Ismail and A. Van Espen, "Calibrating Road Design Guides Using Risk-Based Reliability Analysis," *Journal of Transportation Engineering,* vol. 140, no. 9, pp. 04014041-1-04014041-6, 2014.

[9] S. Liu, X. Wang, O. Hassanin, X. Xu, M. Yang, D. Hurwitz and X. Wu, "Calibration and Evaluation of Responsibility Sensitive Safety (RSS) in Automated Vehicle Performance During Cut-in Scenarios," *Transportation Research Part C: Emerging Technologies,* vol. 125, no. https://doi.org/10.1016/j.trc.2021.103037, pp. 1-15, 2021.

[10] X. Xu, X. Wang, X. Wu, O. Hassanin and C. Chai, "Calibration and Evaluation of the Responsibility Sensitive Safety Model of Autonomous Car-Following Maneuvers Using Naturalistic Driving Study Data," *Transportation Research Part C: Emerging Technologies,* vol. 123, no. https://doi.org/10.1016/j.trc.2021.102988, pp. 1-17, 2021.

[11] Y. Huang, M. S. Elli, J. Weast, Y. Lou and Y. Chen, "RSS Model Calibration and Evaluation for AV Driving Safety based on Naturalistic Driving Data," *TBD,* vol. TBD, no. TBD, p. TBD, In Progress.

[12] B. Weng, S. J. Rao, E. Deosthale, S. Schnelle and F. Barickman, *Model Predictive Instantaneous Safety Metric for Evaluation of Automated Driving Systems,* 2020.

[13] T. Stolte, A. Reschka, G. Bagschik and M. Maurer, *Hazard Analysis and Risk Assessment for an Automated Unmanned Protective Vehicle,* ResearchGate, 2017.

[14] Mobileye, *Implementing the RSS Model on NHTSA Pre-Crash Scenarios.*

[15] Y. Ren, S. Elliott, Y. Wang, Y. Yang and W. Zhang, *How Shall I Drive? Interaction Modeling and Motion Planning Towards Empathetic and Socially-Graceful Driving,* 2019.

[16] A. Bin-Nun, "What are the Properties of a Good Metric?," in *SAE ORAD V&V TF - J3237 Subcommittee*, 2021.

[17] F. M. Favarò, "Hazard Level for Safety Monitoring and Informed Interventions," in *SAE ORAD V&V TF - J3237 Subcommittee*, 2021.

[18] T. Sayed, "Use of Surrogate Measures in Road Safety Analysis," in *SAE ORAD V&V TF - J3237 Subcommittee*, 2021.

[19] S. Feng, Y. Feng, C. Yu, Y. Zhang and H. X. Liu, *Testing Scenario Library Generation for Connected and Automated Vehicles, Part I: Methodology,* IEEE, 2020.

[20] S. Feng, F. Yiheng, H. Sun, S. Bao, Y. Zhang and H. X. Liu, *Testing Scenario Library Generation for Connected and Automated Vehicles, Part II: Case Studies,* IEEE, 2019.

[21] SAE International, "Information Report J3208: Taxonomy and Definitions for Terms Related to Verification & Validation of Automated Driving Systems (ADS)," SAE International, In Progress.

[22] D. J. Fremont, E. Kim, Y. V. Pant and S. A. Seshia, *Formal Scenario-Based Testing of Autonomous Vehicles: From Simulation to the Real World.*

[23] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick and F. Diermeyer, *Survey on Scenario-Based Safety Assessment of Automated Vehicles,* IEEE, 2020.

[24] Z. Saigol, A. Peters, M. Barton and M. Taylor, *Regulating and Accelerating Development of Highly Automated and Autonomous Vehicles through Simulation and Modelling,* 2018.

[25] F. Pinter and A. Engelstein, *Automating Simulation for Safer Self-Driving.*

[26] C. E. Tuncali, G. Fainekos, H. Ito and J. Kapinski, *Simulation-based Adversarial Test Generation for Autonomous Vehicles with Machine Learning Components,* 2019.

[27] H. Liu and Y. Feng, "Real World Meets Virtual World: Augmented Reality Makes Driverless Vehicle Testing Faster, Safer, and Cheaper," *Mcity: University of Michigan,* pp. 1-8, 2018.

[28] M. Wen, J. Park and K. Cho, *A Scenario Generation Pipeline for Autonomous Vehicle Simulators,* SpringerOpen, 2020.

[29] K. D. Kusano and H. C. Gabler, *Pre-Crash Scenarios for Determining Target Populations of Active Safety Systems,* Blacksburg, Virginia: Virginia Tech.

[30] A. Wardzinski, *Dynamic Risk Assessment in Autonomous Vehicles Motion Planning,* Gdansk University of Technology: Research Gate, 2015.

[31] T. Toledo, O. Musicant and T. Lotan, "In-Vehicle Data Recorders for Monitoring and Feedback on Drivers' Behavior," *Transportation Research Part C,* vol. 16, no. doi:10.1016/j.trc.2008.01.001, pp. 320-331, 2008.

[32] J. M. Scanlon, R. Sherony and H. C. Gabler, "Models of Driver Acceleration Behavior Prior to Real-World Intersection Crashes," *IEEE Transactions on Intelligent Transportation Systems,* vol. 19, no. 3, pp. 774-786, 2018.

[33] J. M. Scanlon, K. D. Kusano, T. Daniel, C. Alderson, A. Ogle and T. Victor, "Waymo Simulated Driving Behavior in Reconstructed Fatal Crashes within an Autonomous Vehicle Operating Domain," Waymo LLC, 2021.

[34] S. Como, J. Wishart, M. S. Elli and N. Kidambi, "Operational Safety Assessment (OSA) Methodology: A Proposed Approach to Quantifying the Safety Envelope for Automated Vehicle Safety," in *SAE World Congress Experience*, Detroit, In Progress.

[35] S. Demmel, D. Gruyer, J.-M. Burkhardt, S. Glaser, G. Larue, O. Orfila and A. Rakotonirainy, *Global Risk Assessment in an Autonomous Driving Context: Impact on Both the Car and the Driver,* Elsevier Ltd, 2019.

[36] T. Stolte, R. S. Hosse, U. Becker and M. Maurer, *On Functional Safety of Vehicle Actuation Systems in the Context of Automated Driving,* Elsevier Ltd, 2016.

[37] W. G. Najm, J. D. Smith and M. Yanagisawa, *Pre-Crash Scenario Typology for Crash Avoidance Research,* Washington, DC: DOT-VNTSC-NHTSA-06-02, 2007.

[38] A. M. Eigen and W. G. Najm, *Problem Definition for Pre-Crash Sensing Advanced Restraints,* Cambridge: DOT HS 811 114, 2009.

[39] S. Schnelle, K. Salaani, S. J. Rao, F. S. Barickman and D. Elasasser, *Review of Simulation Frameworks and Standards Related to Driving Scenarios,* Washington, DC: DOT HS 812 815, 2019.

[40] S. Shalev-Shwartz, S. Shammah and A. Shashua, "On a Formal Model of Safe and Scalable Self-Driving Cars," arXiv:1708.06374v6 [cs.RO], 2017.

[41] Geotab, "Product Guide," Geotab, 2021. [Online]. Available: https://community.geotab.com/s/product-guide?language=en_US. [Accessed 16 July 2022].

[42] University of California PATH Program, "Peer Review of Behavioral Competencies for AVs," February 2016. [Online]. Available: https://www.nspe.org/sites/default/files/resources/pdfs/Peer-Review-Report-IntgratedV2.pdf. [Accessed 16 April 2022].

[43] Automated Vehicle Safety Consortium, "AVSC Best Practice for Evaluation of Behavioral Competencies for Automated Driving System Dedicated Vehicles (ADS-DVs)," SAE ITC, 2021.

[44] American Association of State Highway and Transportation Officials, A Policy on Geometric Design of Highways and Streets, Washington, DC: American Association of State Highway and Transportation Officials, 2018.

[45] D. C. Viano and C. S. Parenteau, "Severe Injury to Near- and Far-Seated Occupants in Side Impacts by Crash Severity and Belt Use," *Traffic Injury Prevention,* vol. 11, pp. 69-78, 2010.

[46] sjye, "Minimum distance between two polygons," 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/29719-minimum-distance-between-two-polygons. [Accessed 25 June 2022].

[47] IIHSA, "Vehicle Ratings," 2022. [Online]. Available: https://www.iihs.org/ratings. [Accessed 25 June 2022].

[48] S. Rahimi, *CARLA Automated Iteration Script,* 2021.

[49] N. Altekar, S. Como, D. Lu, J. Wishart, D. Bruyere, F. Saleem and L. K. Head, *Infrastructure-based Sensor Data Capture Systems for Measurement of Operational Safety Assessment (OSA) Metrics,* SAE World Congress, (in development).

[50] L. Fricke, Traffic Crash Reconstruction, Chicago: Northwestern University Center of Public Safety, 2010.

[51] J. Collins, Accident Reconstruction, Springfield, 1979, pp. 171, 187.

[52] R. Limpert, Motor Vehicle Accident Reconstruction and Cause Analysis, Charlottesville: Michie Co., 1978.

[53] N. Kalra and S. M. Paddock, "Driving to Safety: How Many Miles of Driving Would it Take to Demonstrate Autonomous Vehicle Reliability?," RAND Corporation, 2016.

[54] SAE International, "J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE International, 2018.

[55] U.S. Department of Transportation, "Automated Driving Systems 2.0: A Vision for Safety," NHTSA, Washington D.C., 2017.

[56] J. G. Fadaie, *The State of Modeling, Simulation, and Data Utilization within Industry: An Autonomous Vehicles Perspective,* The Boeing Company, 2018.

[57] D. Gettman and L. Head, "Surrogate Safety Measures from Traffic Simulation Models," *Transportation Research Record, 1840,* pp. 104-115, 2003.

[58] J. Wishart, S. Como, L. Head, Y. Chen and N. Altekar, "The Development of Automated Driving System-Equipped Vehicle Safety Metrics," in *SAE WCX World Congress*, Detroit, 2020.

[59] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. d. La Fortelle and M. Tomizuka, *INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps,* arXiv:1910.03088v1 [cs.RO], 2019.

[60] Underwriters Laboratories Inc, *UL 4600: The First Comprehensive Safety Standard for Autonomous Products,* 2019.

[61] V. Jammula, M. Elli, S. Como and J. Wishart, "Evaluation of Operational Safety Assessment (OSA) Metrics for Automated Vehicles Using Real-World Data," in *SAE International WCX*, 2022 in Preparation.

[62] N. Rossi, "EGR 598 Connected and Automated Vehicles," Arizona State University, 2020.

[63] N. Kidambi, J. Wishart, M. Elli and S. Como, "Effects of Measurement and Model Uncertainty on Operational Safety Assessment (OSA) Metrics," in *SAE World Congress*, Detroit, In Progress.