CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

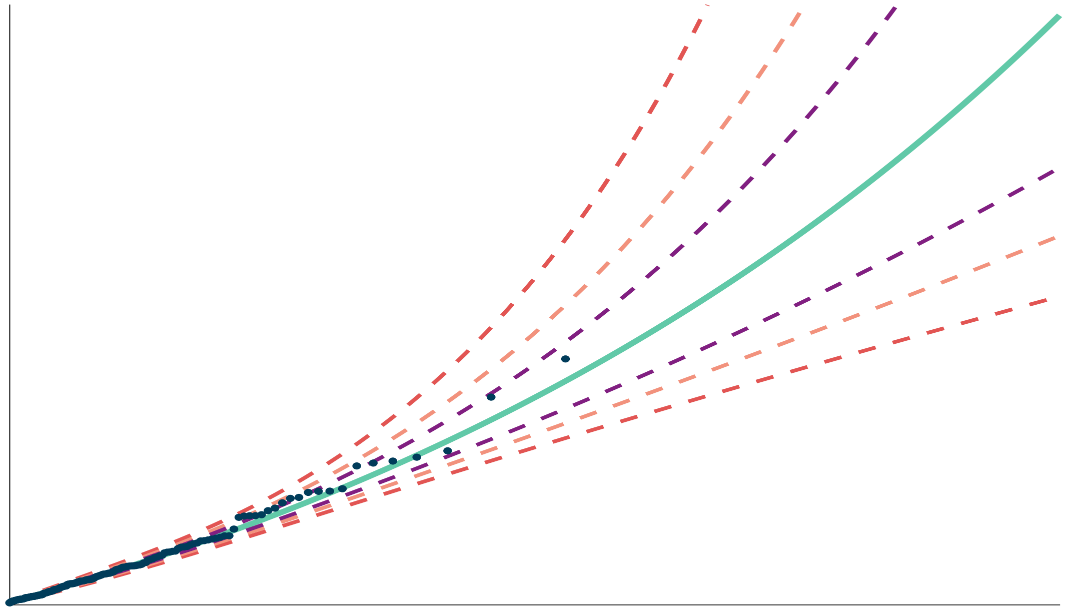
PHD THESIS

Every year more than a million lives are cut short due to traffic accidents. However, most traffic accidents are caused by human error and if these causes can be removed, many lives could be saved. Autonomous vehicles (AVs) will never be as tired or distracted as humans are and are expected to lead to a significantly safer traffic environment. Before AVs can be used by the public and enable a safer future, it needs to be shown that they are as safe as they should be. As a result, we need evidence that AVs have fewer accidents than human drivers in real traffic. This evidence is not simple to obtain since humans are, on average good drivers, and fatalities may occur less often than once every 100 million kilometers. Driving this distance to show the absence of accidents before every release does not scale well.

This thesis presents multiple approaches to creating this evidence more efficiently. The first method uses computer simulations of the actual vehicle to provide safety evidence of the software before it is used in an actual vehicle. Simulation efforts are also focused on the areas where it is believed to be closest to failure, which makes it more efficient. The result is a precise estimate of how often the AV software will fail and the specific scenarios where it will happen.

A second method is evaluating the safety of AVs in real traffic. It evaluates situations that were close to being accidents and uses them to estimate the frequency of actual accidents. The method makes it possible to show that the AVs are safe without experiencing any real accidents. In addition, the second method is also used to form a predictive safety monitor for a fleet of AVs. The results show that the predictive monitor significantly reduces the risk of deploying unsafe AVs.

DANIEL ÅSLJUNG • On Statistical Methods for Safety Validation of Automated Vehicles • 2022



On Statistical Methods for Safety Validation of Automated Vehicles

Using Threat Metrics to Accelerate Safety Evidence Generation

DANIEL ÅSLJUNG

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

On Statistical Methods for Safety Validation of Automated Vehicles

Using Threat Metrics to Accelerate Safety Evidence Generation

Daniel Åsljung

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2022

On Statistical Methods for Safety Validation of Automated Vehicles
Using Threat Metrics to Accelerate Safety Evidence Generation

Daniel Åsljung
ISBN 978-91-7905-757-2

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serier nr 5223
ISSN 0346-718X
Department of Electrical Engineering

Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000

Printed by Chalmers digitaltryck
Gothenburg, Sweden, November 2022

To Aron, Sigrid & Samuel.

Abstract

Automated vehicles (AVs) are expected to bring safer and more convenient transport in the future. Consequently, before introducing AVs at scale to the general public, the required levels of safety should be shown with evidence. However, statistical evidence generated by brute force testing using safety drivers in real traffic does not scale well. Therefore, more efficient methods are needed to evaluate if an AV exhibits acceptable levels of risk.

This thesis studies the use of two methods to evaluate the AV's safety performance efficiently. Both methods are based on assessing near-collision using threat metrics to estimate the frequency of actual collisions. The first method, called subset simulation, is here used to search the scenario parameter space in a simulation environment to estimate the probability of collision for an AV under development. More specifically, this thesis explores how the choice of threat metric, used to guide the search, affects the precision of the failure rate estimation. The result shows significant differences between the metrics and that some provide precise and accurate estimates.

The second method is based on Extreme Value Theory (EVT), which is used to model the behavior of rare events. In this thesis, near-collision scenarios are identified using threat metrics and then extrapolated to estimate the frequency of actual collisions. The collision frequency estimates from different types of threat metrics are assessed when used with EVT for AV safety validation. Results show that a metric relating to the point where a collision is unavoidable works best and provides credible estimates.

In addition, this thesis proposes how EVT and threat metrics can be used as a proactive safety monitor for AVs deployed in real traffic. The concept is evaluated in a fictive development case and compared to a reactive approach of counting the actual events. It is found that the risk exposure of releasing a non-safe function can be significantly reduced by applying the proposed EVT monitor.

Keywords: Automotive, automated driving systems, automated vehicles, extreme value theory, performance evaluation, simulation, validation, verification.

List of Publications

This thesis is based on the following publications:

[A] **Daniel Åsljung**, C. Zandén, J. Fredriksson, M. K. Vakilzadeh, "On Automated Vehicle Collision Risk Estimation using Threat Metrics in Subset Simulation". Published in IEEE International Intelligent Transportation Systems Conference (ITSC), 2021.

[B] **Daniel Åsljung**, J. Nilsson, J. Fredriksson, "Comparing Collision Threat Measures for Verification of Autonomous Vehicles using Extreme Value Theory". Published in 9th IFAC Symposium on Intelligent Autonomous Vehicles, 2016.

[C] **Daniel Åsljung**, J. Nilsson, J. Fredriksson, "Validation of Collision Frequency Estimation Using Extreme Value Theory". Published in IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017.

[D] **Daniel Åsljung**, J. Nilsson, J. Fredriksson, "Using Extreme Value Theory for Vehicle Level Safety Validation and Implications for Autonomous Vehicles". Published in IEEE Transactions on Intelligent Vehicles, Dec. 2017.

[E] **Daniel Åsljung**, C. Zandén, J. Fredriksson, "A Risk Reducing Fleet Monitor for Automated Vehicles Based on Extreme Value Theory". Submitted for publication in IEEE Transactions on Intelligent Transportation Systems.

[F] **Daniel Åsljung**, M. Westlund, J. Fredriksson, "A Probabilistic Framework for Collision Probability Estimation and an Analysis of the Discretization Precision". Published in IEEE Intelligent Vehicles Symposium (IV), 2019.

Other publications by the author, not included in this thesis, are:

[G] **Daniel Åsljung**, "On Safety Validation of Automated Driving Systems using Extreme Value Theory". Licentiate thesis, Chalmers University of Technology, Dec. 2017.

Acknowledgments

I want to start by giving a big thank you to my latest industrial supervisor at Zenseact, Dr. Carl Zandén. You have always provided good discussions and feedback, which have greatly improved my results. I also want to give a big thank you to my first industrial supervisor at Volvo Cars and Zenuity, Dr. Jonas Nilsson, for your part in my recruitment and your constant support in the first part of my research. You continuously provided me with good discussion, new ideas, and insights during the first part of my research. A big thank you also goes to Dr. Mattias Westlund who was my industrial supervisor during a period at Zenuity and helped me stay focused on what was important during that time.

I would also like to thank my previous manager at Zenseact, Mats Nordlund, who has taught me much about the importance of research communication and crisp research questions. A big thank you also goes to Carl Lindberg, my coach at Zenseact, for all your support during the later part of my studies. I also would like to thank my first manager at Volvo Cars, Jonas Ekmark, for hiring me and giving me this opportunity from the beginning.

I also want to express my gratitude to my academic supervisor at Chalmers, Prof. Jonas Fredriksson, for guiding me and always providing support when I need it during my studies. Furthermore, funding for this project has been received from Volvo Cars, Zenseact, and FFI under the Traffic Safety and Automated Vehicles program, for which I am grateful.

I want to thank all my colleagues at Zenseact and Chalmers for creating two very stimulating workplaces. Good discussions have always been going on that have made an inspiring environment. A special thank you goes to Dr. Majid K. Vakilzadeh, who participated in one of the papers and has provided me with many good thoughts and discussions.

Finally, my appreciation goes to my wife Lai Jee, for being supportive throughout the journey and always inspiring me to try new things.

Acronyms

ACC:	Adaptive Cruise Control
ADAS:	Advanced Driver Assistance Systems
ADS:	Automated Driving System
ALKS:	Automated Lane Keeping Systems
AV:	Automated Vehicle
BM:	Block Maxima
BTN:	Brake Threat Number
EVT:	Extreme Value Theory
GEV:	General Extreme Value
GP:	Generalized Pareto
ISO:	International Organization for Standardization

LKA:	Lane Keeping Aid
LSF:	Limit-State Function
MC:	Monte Carlo
MCMC:	Markov Chain Monte Carlo
MIL:	Model-In-the-Loop
ODD:	Operational Design Domain
PDF:	Probability Density Function
PET:	Post Encroachment Time
POT:	Peak Over Threshold
SuS:	Subset Simulation
SIL:	Software-In-the-Loop
SOTIF:	Safety Of The Intended Functionality
THW:	Time Headway
TTB:	Time To Brake
TTC:	Time To Collision
TTM:	Time To Maneuver

Contents

Part I

Overview

CHAPTER 1

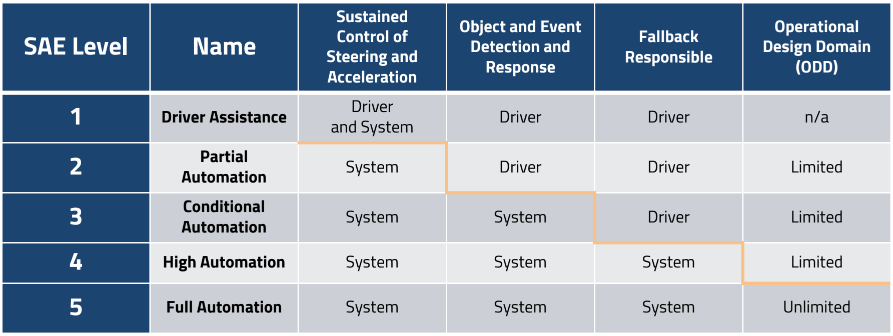
Introduction

Automated vehicles are expected to bring significant benefits to the traffic environment. The National Highway Traffic Safety Administration (NHTSA), [1], have shown that human factors are the cause of over 90% of traffic accidents. With the human removed from the equation, significantly reducing the number of casualties in traffic is possible. In addition, it enables the driver to do something else with the time in the vehicle, like reading an exciting book or thesis. Furthermore, the vehicles could also drive without passengers to allow relocation and delivery of goods. Currently, there is a lot of effort put into developing automated vehicles. Many actors promise to have vehicles with a higher level of autonomy available right now or during the coming years, e.g. [2]–[5].

The driver of an automated vehicle is put out of the loop and cannot be used as a fallback when things go wrong. Consequently, there will be very high dependability requirements connected to the safety of the vehicle and its functions. Moreover, it has to be understood what safe behavior means to know these requirements in practice. The vehicle must handle traffic laws, everyday driving, and rare road hazards that are hard to foresee. Then there must be a strategy to validate that the vehicle has reached the required level of safety. A considerable effort across many different domains has to be made to solve this problem, e.g., Safety Engineering, Legal, Testing, Security, and Computing Hardware [6].

1.1 Driving Automation

Advanced Driver Assistance Systems (ADAS) aims to support the driver by automating some mode of control in the vehicle. However, the driver is still responsible and has the possibility to override the function. The ADAS system also often has limited capabilities when it comes to more extreme maneuvers. Therefore, the driver must monitor the system for failures and act as a fallback. The basic type of assistance system relieves the driver of one specific driving task. These are referred to as Level 1 automation according to the SAE J3016 standard [7]. An overview of the different levels of automation can be seen in Figure 1.1. The first two levels are considered ADAS, while levels three and above are referred to as an Automated Driving System (ADS).



SAE Level	Name	Sustained Control of Steering and Acceleration	Object and Event Detection and Response	Fallback Responsible	Operational Design Domain (ODD)
1	**Driver Assistance**	Driver and System	Driver	Driver	n/a
2	**Partial Automation**	System	Driver	Driver	Limited
3	**Conditional Automation**	System	System	Driver	Limited
4	**High Automation**	System	System	System	Limited
5	**Full Automation**	System	System	System	Unlimited

Figure 1.1: Illustration of the five levels of automation from the SAE J3016 standard. The columns highlight where the responsibility lies within different areas for the respective level.

An example of a Level 1 system is Adaptive Cruise Control (ACC), which controls the acceleration and braking to maintain a set gap to the vehicle in front. If ACC is combined with Lane Keeping Assistance (LKA) into one function controlling acceleration, deceleration, and steering, it becomes a Level 2 system. There are systems of this type in production, e.g., Mercedes' Drive Pilot, Tesla's Autopilot, and Volvo's Pilot Assist. In these systems, the driver still needs to monitor the system and the environment. A detailed description of ACC, LKA, and other ADAS systems can be found in [8].

Unsupervised Automated Driving

By moving to Level 3 and higher, the driver's responsibility to monitor is removed, resulting in an unsupervised ADS. Consequently, it allows the driver to do other things while the car is driving. As a first step, the system could be limited to special conditions such as weather, traffic, and roadway characteristics. These types of operating conditions, which an ADS is specifically designed to function under, form an Operational Design Domain (ODD) [7]. An example of an ADS with a limited ODD is an Automated Lane Keeping System (ALKS), which is a system that takes over the driving task at speeds below 60 km/h [9]. The system then controls the lateral and longitudinal movements of the vehicle for these low-speed scenarios without any input from the driver. When the vehicle is about to exit the function's scope, it hands back the control to the driver. If the driver does not take over, the system needs to have a backup plan that it can execute to put the vehicle in a safe state. In the ALKS regulation, this is called Minimum Risk Manoeuvre (MRM) and consists of braking the vehicle to a standstill in the lane and activating the hazard lights. The ODD can be expanded to increase the system's capability and include more driving scenarios. Ultimately, the vehicle can be driven autonomously without a driver present, Level 4, and in all situations and conditions, Level 5.

Implication for System Design

In the case of an ADAS function such as ACC, the scope is limited. The function should keep a distance to a preceding vehicle, and if there is no vehicle in front, the system should act as regular cruise control, keeping a set speed. This function can be designed with a single radar sensor, measuring the position and speed of possible preceding vehicles. Out of the potential objects detected, a target vehicle has to be selected. Based on that, an action is taken to keep the set distance to that vehicle.

Suppose that a similar function with the same capabilities is to be developed, but now as an unsupervised function. The driver is no longer responsible for monitoring and is unavailable as a fallback option. Consequently, there would be much higher requirements on perception to detect all possible objects around the vehicle. That is because there is no longer a driver that monitors the road and can intervene if an object is missed. As a result, additional sensors for redundancy could be needed, which has to be handled by the perception layer. The higher requirements will also affect decision-making on interpreting the situation correctly and choosing the right target to follow. Ultimately, there will also be a requirement on vehicle control that guarantees the execution of a braking maneuver. The guarantee is needed for the decision-making to plan a safety margin to the target vehicle and be able to handle other possible events. To fulfill these requirements, adding a redundant braking system might be necessary.

When the function's scope expands towards unsupervised automated driving and a complete ODD, the function needs to handle many more situations than the ACC case. As a result, the environment that the system should be designed to act in will be much more complex. The implication for the perception block is that there will be high requirements to detect objects all around the vehicle and at long distances. Therefore, many more sensors must be added to give a full surrounding view of the environment. There will also be a need for redundant sensors in many places to reach the high level of robustness.

For decision-making, there will be many more scenarios that should be correctly interpreted and complex traffic scenarios with many different participants, which behavior needs to be predicted. There also needs to be decisions on multiple levels taking care of strategic and operational planning with logic determining what is currently the most important to reach the target safely.

For vehicle control, the scope now also includes steering, which probably needs to be redundant to guarantee high enough availability. In addition, the range of actions that should be possible to actuate has increased to include a large variety of highly dynamic maneuvers. The result is a highly complex system with very tough safety requirements that must be handled by every part of the system.

1.2 Safe System Design

To develop a complex system such as an ADS, one needs to define what needs to be developed, how it will be implemented, and ensure that the system is doing what it is supposed to. This process falls under an area called systems engineering, which deals with designing and managing this type of complex system [10].

The process usually contains the steps of refining requirements, functional allocation, and physical implementation. In addition, each of these steps has to be verified and validated against the top-level requirements. In the automotive industry, the process follows a framework called the V-model [11], which is also a part of the ISO 26262 standard for functional safety [12]. In this standard, safety goals are defined, forming the vehicle level safety requirements that should be met to ensure a safe function. For an ADS, the safety goals might be more general to cover all situations, but that leads to more abstract formulations that are more difficult to verify [13].

To ensure safe behavior of the system, possible failures have to be detected and mitigated. These failures include both hardware- and software-related faults, and it needs to be shown that these are sufficiently rare events. For an ADS, it is also vital to ensure that the nominal performance of the system is good enough to ensure a safe operation. Hence, the system must be designed to be safe when everything is working as intended.

Nominal Safety Performance

With the emergence of systems that take more control over the vehicle, new safety-related problems arise that the standard ISO 26262 does not address. ISO 26262 includes possible hardware or software failures that may lead to safety-critical situations. However, it does not explicitly describe how to handle the potential safety issues when there is no fault. With the emergence of ADAS and ADS, this will become more common.

The reason for this is that the nominal safety performance of the function may be inadequate to ensure safe operation. A function could, for example, be designed so that the host should always keep a minimum distance to the preceding vehicle. However, this distance could be insufficient in some situations to drive safely. Another critical area is the sensor performance, which includes, for example, technological limitations. Specifically, a vision sensor could be trained on a data set that does not contain a particular type of object and therefore fails to classify it. The same problem exists for actuators, where technical limitations prevent the vehicle from carrying out specific control requests safely. Unfortunately, the ISO 26262 standard does not explicitly describe extracting and verifying this type of requirement.

Due to the lack of addressing these issues, the standard ISO 21448 named Safety Of The Intended Functionality (SOTIF) [14] has emerged. The standard handles requirements based on functional insufficiencies of the system. These insufficiencies includes previously mentioned performance limitations or specifications that could lead to hazards. The standard also introduces the term acceptance criterion:

Definition 1 (Acceptance criterion)**:** *Criterion representing the absence of an unreasonable level of risk.*

For example, at the complete vehicle level, a criterion could be to have fewer accidents than once every million hours. In addition, a validation target is defined to provide evidence that the acceptance criterion is fulfilled.

Definition 2 (Validation target)**:** *Value to argue that the acceptance criterion (Definition 1) is met.*

For example, a validation target could be to let the system run with safety drivers for 3 million hours without accidents. Consequently, it is possible to show with 95 percent confidence that the criterion is met by assuming accidents are Poisson distributed [15]. These two standards should ensure that enough is done to ensure that an ADS system is safe before it is released.

There is also a section in SOTIF that deals with the operation phase. Since the environment constantly evolves during the system's lifetime, its safety has to be continuously monitored. It could also be that some assumptions are not valid or new functional insufficiencies are discovered. However, the standard does not explicitly suggest how this monitoring should be performed efficiently and safely.

1.3 Problem Formulation

The challenge of assuring safety for an ADS has given rise to the following questions: How to efficiently show that the acceptance criterion is met? How to monitor the safety of the system and make sure that the residual risk is acceptable? The first question addresses finding evidence that the acceptance criterion is met, i.e., a validation target and ensuring that the ADS is safe before it is released. It should also be possible to efficiently generate the evidence and avoid driving the 3 million hours needed in the example mentioned earlier. By addressing the second question, we can ensure that the system is safe after it is released and potentially catch any excessive residual risk as soon as possible.

1.4 Delimitation

This thesis considers only safety validation on the vehicle level. The acceptance criterion addressed in the problem formulation is in the form of a quantitative risk level of an ADS responsible for the driving task. It is here assumed that a verified ADS function already exists. The acceptance criterion is in this thesis delimited only to consider the situation of collision with other vehicles on the road. Data based on human drivers have been used to validate the method since no complete ADS is available for testing. For the simulation studies, the ADS implemented is an ACC function with emergency braking capability to have a transparent and simple function to illustrate the methods.

1.5 Contributions

This thesis presents two methods for accelerated failure frequency estimations of automated vehicles. The first method is called Subset Simulation (SuS) and is applied in a simulation environment, see Paper A. The second method uses Extreme Value Theory (EVT) and can be applied both in simulation and at the complete vehicle level, see Paper B, C, D and E. Common for both methods is that a metric of the closeness to a collision with another vehicle is required. In this thesis, different types of metrics are evaluated using the two methods with both simulated and recorded vehicle data, see Paper A, B, D and E. In addition, specific aspects concerning the metrics, such as multiple probable outcomes, see Paper F, and prediction, see Paper E, are also investigated.

The EVT method also generates confidence intervals that consider the uncertainty of the extrapolation, which can be used for safety validation purposes. Using data gathered from human drivers, the method is validated by comparing the results with data from crash statistics, see Paper C and D. The confidence interval can also be used to do statistical testing, see Paper E, where sequential statistical testing is used to create a safety monitor. Several methods for automatically applying the EVT model to the data have been evaluated in Paper D with further considerations related to conservative estimations in Paper E.

1.6 Outline

This thesis comprises two parts where Part I acts as an introduction to what is presented in Part II. In Part II, there are six scientific papers, which are the base of the thesis. Part I provides background information and puts the appended papers into context with the following structure. In Chapter 1, the setting of the thesis is introduced by first describing an unsupervised ADS. It is then explained what it takes to design this system to perform safely. This background is followed by a formulation of the problem that this thesis addresses and what delimitations have been made. Next, Chapter 2 describes different types of verification and validation methods. Chapter 3 provides an introduction to EVT and explains how it can be applied to traffic safety. This is followed in Chapter 4 by an introduction to the SuS method and how it can be used together with an ADS. In Chapter 5, the papers included in Part II are briefly summarized, and in Chapter 6, the thesis is concluded with suggestions for further research.

CHAPTER 2

Verification and Validation Methods

Chapter 1 describes how the complexity and increased responsibility of an ADS make it more difficult to ensure safety. There is a need to create evidence from verification and validation that the system does not impose unnecessary risk. The evidence can consist of both qualitative and quantitative arguments based on different models and assumptions. Qualitative methods aim to show that the system can handle a specific set or scope of scenarios. On the other hand, quantitative methods aim to measure the system's safety performance in a given environment.

This thesis considers two methods to address the presented research questions. The first approach is a simulation-based method called Subset Simulation (SuS), which is explained in more detail in Chapter 4. SuS is a method that can efficiently estimate a failure frequency and provide quantitative evidence during the development phase of a system. The second approach is based on Extreme Value Theory (EVT), a statistical method used to model rare events. EVT and how it can be used is explained in more detail in Chapter 3.

This chapter presents a selection of verification and validation approaches and describes their respective strengths and weaknesses. It is also argued why the selected methods for this thesis are chosen and what role the other approaches could serve in a safety case.

2.1 Simulation

Using simulation for verification aims to test the system in closed loop based on computer-generated inputs. Some parts of the system and the environment are modeled as close to the real experience as possible. One type of simulation is Model-In-the-Loop (MIL), where the whole system is a model. This type of simulation is used early in the development process before the actual code has been developed and can save time and cost before putting the code in a vehicle [16]. Another type of simulation is called Software-In-the-Loop (SIL), which uses the actual system implementation in the simulation. The results from SIL can then be compared to the MIL to see that the software is implemented according to the model. Examples of implementations of MIL and SIL can be found in [17]–[19]. In both these simulation types, virtually generated scenarios are used as input to the system. The scenarios can be created from the specifications to provide qualitative evidence but also based on experiences in real traffic for a quantitative approach, as seen in [20].

Generating these scenarios is an essential part of simulation and consists of two steps [21]. Firstly, the test space has to be modeled by, for example, observing real traffic and characterizing the scenarios. Secondly, the test space must be sampled by some principle to extract scenarios that the ADS should experience. The sampling can be done statistically, where one tries to find quantitative evidence that the failure rate is sufficiently low. However, with the low failure frequencies required by an ADS, more efficient sampling techniques such as important sampling [22]–[24] or SuS [25] might be needed.

Simulation has the benefit of being able to perform tests of scenarios much faster than in the real world. It can also test variations of scenarios that have not yet been seen. For this to be possible, validated models of the system and the environment are needed. In this thesis, the SuS method is chosen for its ability to provide quantitative evidence from much fewer simulations than a pure Monte Carlo approach.

2.2 Statistical Methods

To capture the stochastic behavior of the system due to the uncertainty in, e.g., sensor information and prediction models, one can use statistical verification methods. For estimating the frequency of failures, the system is often modeled as a Poisson process for the number of failures during a specific time. A confidence interval can be created to verify that the failure rate is lower than the requirement. This method is the basis for the proven-in-use argument in ISO 26262 [12]. It is also an option for a quantitative validation target for the acceptance criterion in the SOTIF standard [14]. An automated driving function has very tough requirements on failure rates, which leads to a large amount of driving data being needed for verification [26]. To get a representative sample of the driving, a real-world user profile is used as in [17], [27], where statistical methods are used to verify that the false positive rate is sufficiently low. A similar approach could be taken to verify false negatives for sensor detection in the case of missed objects. However, that would require a dependable reference sensor system or similar for comparison.

Statistical methods can also be used to monitor events in the real world by applying sequential testing [28], [29]. For instance, it has been applied for a long time in the medical field to monitor the testing of new treatments [30], [31]. Significant parameters are observed to be within a specific limit, and if that is not the case, the trial is stopped. Statistical monitoring has also more recently been applied to monitor parameters or safety indicators during run-time in the aviation industry [32], [33].

The main strength of statistical methods is the possibility of having high validity by testing the function in its natural environment. However, a drawback is that this requires a large amount of data for each new system version that needs to be verified. To address the issue, this thesis explores an EVT approach to get quantitative safety evidence from less amount of data.

2.3 Directed Testing

For testing the performance of ADAS, directed testing on test tracks has been used in [17], [27]. In directed testing, several scenarios based on real-world driving situations are tested. The tests are also done in several different weather and light conditions and variations of similar cases. A benefit of using directed testing is that the whole system, from sensors to actuators, is used as it is implemented. It is also possible to repeatedly test rare challenging scenarios, which is impossible in real traffic.

It is hard to recreate variations of situations realistically with directed testing at a test track. The worst-case scenarios are often tested when using directed testing on a test track for verification. An example of how worst-case scenarios can be defined for a collision avoidance system is found in [34]. It is in those situations where a system error is most likely and based on the results, it can be argued that the system can handle less challenging scenarios as well. However, for an automated vehicle, it is not evident in many situations what is the worst-case situation and how to argue that all other cases are handled.

This method is very effective in testing the system against extreme scenarios, which are often hard to experience in field tests. However, it is often difficult to define a complete set of test cases covering the ODD. In the case of providing evidence that the acceptance criterion is fulfilled, directed testing can be a complement to quantitative evidence. As an example, there might be corner cases that the ADS should be able to handle, but they are too rare to show up often enough in a quantitative method.

2.4 Formal Methods

Formal methods use mathematical models to verify that the system fulfills the requirements. They can be used in the entire development process, from requirements engineering to implementation [35]. In [36], formal methods are investigated in the scope of tactical planners, where they provide evidence that safe decisions are always made. At the implementation level, the software is connected to mathematical contracts between input and program variables. With these mathematical models present, the code can also be automatically generated. This method has been applied to verify the safety of ADAS and ADS, see e.g. [37]–[39].

The main benefit of these methods is that it is powerful to prove that requirements are always fulfilled mathematically. The drawback is that validated mathematical models of every part of the system are needed. There are, however, methods of automatically learning these models by observing an implemented software [40]. Moreover, formal methods could also complement quantitative methods to provide evidence on sub-parts of the system that it performs according to specification [41]. In [42], formal methods are used in the verification of decision and control logic. In particular, it is shown in a structured approach how that can be used as evidence in the safety argumentation for an ADS.

CHAPTER 3

Extreme Value Theory

Extreme Value Theory (EVT) is an area of statistics focusing on rare rather than frequent events. The theory was first applied in civil engineering to understand better the requirements for what structures need to be able to handle over a long period [43]. It provided a framework to describe the magnitude of expected forces based on historical data. The framework of EVT contains a set of models that enable the usage of observed levels of data and extrapolate that into estimates of unobserved levels. It has previously been used in, e.g., engineering, finance, and risk management [44]–[46].

3.1 Block Maxima

The statistical behavior modeled in the classical extreme value theory is the maximum, M_n, of a sequence of independent random variables.

$$M_n = \max\{X_1, ..., X_n\} \tag{3.1}$$

These measurements, $X_1, .., X_n$, could, for example, be a continuous stream of values, as visualized in Figure 4.1. The value M_n is the maximum of these values during a particular time. Therefore, the method is often referred to as the Block Maxima (BM) method.

If the cumulative distribution F of the maximum values in each block is known, it could be used to estimate the frequency of rare events. In practice, the distribution F is unknown but can be approximated to a set of models based only on extreme data [43]. It is similar to the normal approximation of sample means using the central limit theorem. The set of models can be represented by the Generalized Extreme Value (GEV) distribution, as

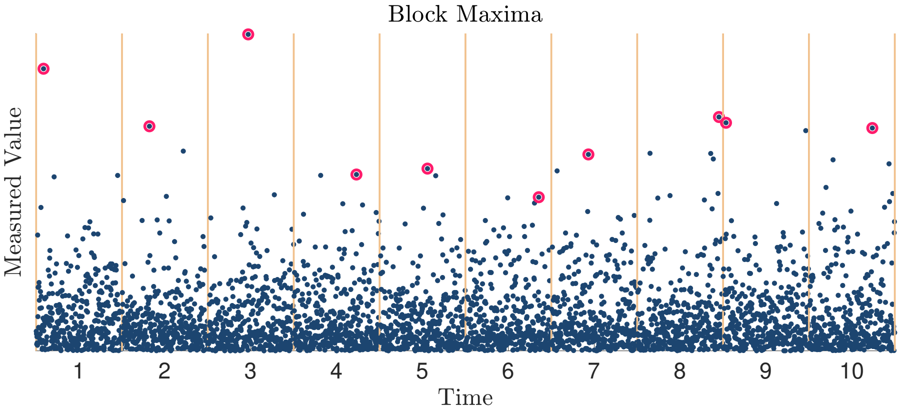


Figure 3.1: This figure illustrates how the block maxima values are selected from a continuous stream of values, shown as blue dots. The selected maximum values of each block are highlighted with a red ring. Adapted from [47]

illustrated in Figure 3.2.

The distribution consists of the three parameters location (μ), shape (ξ) and scale (σ) with the following probability density function:

$$f(x|\xi,\sigma,\mu) = \frac{1}{\sigma}\exp\Big(-(1+\xi\frac{(x-\mu)}{\sigma})^{-\frac{1}{\xi}}\Big)(1+\xi\frac{(x-\mu)}{\sigma})^{-1-\frac{1}{\xi}}. \tag{3.2}$$

If data is collected for multiple blocks, a series of maxima, $M_{n,1},...,M_{n,m}$, can be used to fit a GEV distribution. Then the probability that a yearly maximum exceeds a value x_p can be found using the inverse cumulative distribution function:

$$p = 1 - F(x_p). \tag{3.3}$$

When implementing this model on a data set, the choice of block size can significantly impact the result. Choosing a too small block size leads to bias in the estimation due to the poor approximation of the limit theorem. On the other hand, a large block size will instead lead to few maxima and thereby a sizable estimate variance. Another critical aspect of choosing the block size is that maxima must be equally distributed. Therefore, if seasonal differences exist in the measured variable, each block must have the same conditions. Using block maxima could mean that a large part of the available data is wasted, and it is especially true if many of the extreme events occur in the same block.

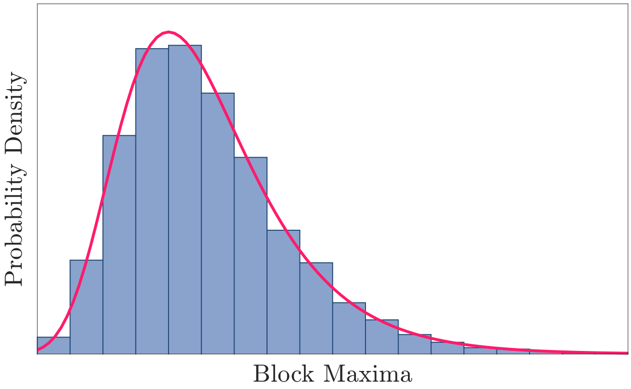


Figure 3.2: This figure illustrates how the GEV distribution is fitted to data. The probability density function for the distribution is shown as the solid red line. The values on the x-axis represent the maximum measurement from each block. Adapted from [47]

3.2 Peak Over Threshold

Peak Over Threshold (POT) is another method that avoids the blocking and only models the most extreme events that exceed some threshold, u, which is visualized in Figure 3.3. The k values that are exceeding the threshold, $x_i : x_i > u$, are called exceedances and are labeled $x_{(1)}, ..., x_{(k)}$.

These values then belong to a distribution family called the Generalized Pareto (GP) distribution, as shown in Figure 3.4. The GP distribution consists of similar parameters as the GEV distribution, with shape (ξ), scale (σ), and threshold (u), and it has the following probability density function:

$$f(x|\xi, \sigma, u) = \frac{1}{\sigma}\Big(1 + \xi\frac{x - u}{\sigma}\Big)^{-(1/\xi+1)}. \tag{3.4}$$

To avoid bias or high variance in the estimation, the threshold, u, is chosen as low as possible while still having a satisfying fit to the model [43]. The selection is often made by manually inspecting the shape parameter for different choices of thresholds. As long as the shape parameter is constant, the estimation is stable, indicating a good model fit. However, finding a suitable threshold in practice can be difficult and often relies on experience.

The probability that a specific value is exceeded can be calculated similarly to the block maxima method. Suppose that $\zeta_u = \Pr\{X > u\}$, then the probability, p, that the value x_p is exceeded is:

$$p = \zeta_u\left(1 - F(x_p)\right). \tag{3.5}$$

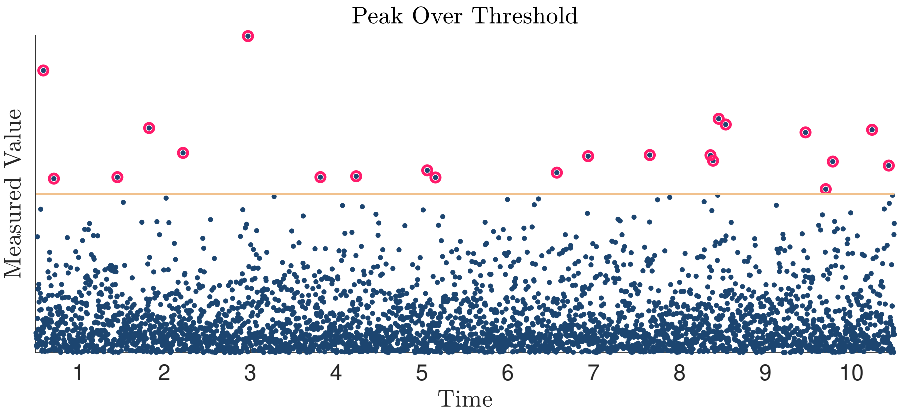


Figure 3.3: This figure illustrates how the exceedances are selected from the stream of values. The selected peak values that exceed the threshold are highlighted with a red circle. The threshold is represented with a horizontal yellow line. Adapted from [47]

3.3 Return Level

The probability, p, that is received for a certain value, x_p, can be used to find the average period between events that exceed this value. In EVT, this time is called the return period, and the corresponding value is called the return level. Given a probability, the return period, t_p, can be found using the following formula:

$$t_p = \frac{t_{tot}}{np}, \tag{3.6}$$

where t_{tot} is the total time of data gathering and n is the number of blocks for the BM method or the total number of measurements for the POT method.

When the return level is plotted against different return periods, the result can be seen in Figure 3.5. Confidence intervals of these estimates that consider the uncertainty of more extreme return levels that have not yet occurred can also be constructed.

If one is interested in how often a particular value is exceeded, the answer would be the corresponding return period. For example, that could be interesting in evaluating a certain height's effectiveness for a seawall. The return period would then correspond to how often the barrier is expected to be flooded.

3.4 Application to Vehicle Safety

Extreme value methods can estimate the frequency of events that have not yet occurred by extrapolating from the models fitted to the rare data that has been recorded. Examples of how EVT has been applied in the automotive safety setting can be found in [48]–[55].

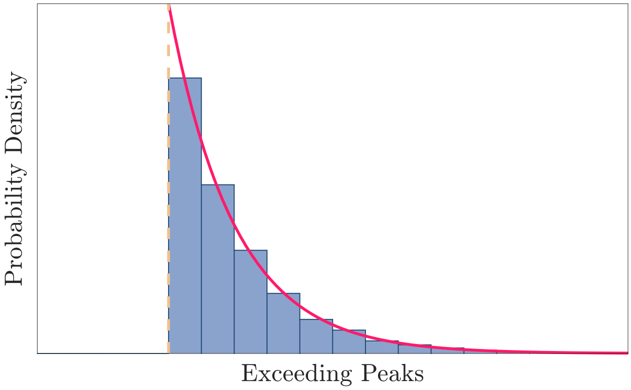


Figure 3.4: This figure illustrates how the GP distribution is fitted to all values exceeding a certain threshold. The threshold is represented by the dashed yellow line, and the probability density function by the solid red line. Adapted from [47]

For vehicle safety related to ADAS and ADS, it is reasonable to assume that there will be a lot of data available about how the system performs. The data could also be generated continuously to create a stream of values that can be used for EVT. In that case, it is better to use the POT method than the BM method since that would enable more data to be used [43]. For this to be possible, there is a need for a metric that reflects the closeness to an accident. The metric also needs a definite value where a collision happens or is unavoidable.

Such metrics have been developed in the active safety area for avoiding, for example, rear-end collisions with an auto-braking system. These metrics are called threat assessment metrics since they are used to decide if the situation is threatening enough for the collision avoidance system to activate. The main differences between the presented threat assessment methods are the model used for the host vehicle, the objects around it, and how their future actions and motions are predicted [56].

Deterministic Threat Metrics

Generally, a vehicle can avoid a collision in many different ways, for example, by steering, braking, or accelerating, and there are a lot of combinations of these actions. Therefore, threat assessment is often simplified for computational reasons. Deterministic threat metrics assume a given model, which offers one prediction that results in one specific value of the threat for a given time instant. These predictions are often made for one of the vehicle's possible actions at a time. Below is a description of some common deterministic threat assessment methods.

One of the most straightforward metrics is the distance to an obstacle ahead in the host's path. This metric is called headway, p_{HW}, and for a straight road, it is equal to the distance

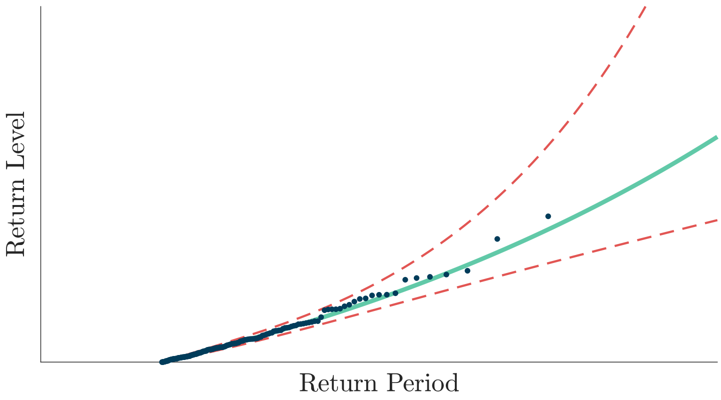


Figure 3.5: The figure illustrates how EVT can estimate the value that is expected to be exceeded once in a specific return period. The solid green line represents the most likely estimate, while the red dashed lines correspond to this estimate's confidence interval. The blue dots correspond to the measurements used to fit the EVT model, plotted along the estimate to indicate how well the model fits the data. Adapted from [47]

between the host bumper and the rear of the obstacle [56]. In the case of a curved road, it is the distance traveled along the middle of the road to reach the object. This metric can also be expressed in time headway (THW), t_{HW}, which is the time it takes for the host to reach the object's position. If the host's acceleration is zero, then:

$$t_{HW} = \frac{p_{HW}}{v_{0,host}}, \tag{3.7}$$

where $v_{0,host}$ is the initial speed of the host vehicle. A variant of this metric is the post-encroachment time (PET), commonly used in a retrospective analysis of conflict scenarios [57]. The PET metric is defined as the time between the obstacle leaving a conflict area and the ego vehicle entering the same area. In the case of the ego vehicle following the obstacle in the same lane, the PET is identical to THW.

The headway metric relates to the exposure to a hazardous situation, i.e., how sensitive the host vehicle is to sudden events. However, the metric does not predict the future motions of the object, which becomes a problem if there is a high relative speed. A metric that handles this is the time to collision (TTC), t_{TTC}. It is often assumed that the acceleration of the host and the object is constant [56]. With that assumption, the t_{TTC} is found by solving:

$$0 = p_{x,0} + v_{x,0}t_{TTC} + \frac{a_{x,0}t_{TTC}^2}{2}, \tag{3.8}$$

where $p_{x,0}$, $v_{x,0}$ and $a_{x,0}$ are the initial relative position, velocity, and acceleration, respectively. The correct t_{TTC} is the lowest positive solution found. This metric is directly related to the point of a collision. There are also metrics such as required longitudinal acceleration, a_x, reflecting how much effort is needed to avoid a collision. The required acceleration can also be combined with the braking capacity and create a ratio of the needed braking to avoid a collision, called Brake Threat Number (BTN) [58]. With the assumption of constant acceleration for both the host vehicle and the object, the required acceleration can be found by solving the following system of equations:

$$\begin{cases} 0 = v_{x,0} + a_x t, \\ 0 = p_{x,0} + v_{x,0} t + \frac{a_x t^2}{2}. \end{cases} \tag{3.9}$$

There is a difference between the metrics presented here in what aspect they relate to a possible threatening situation. The metric of TTC reflects the closeness in time of a predicted collision. Time headway does not predict a crash but instead relates to an obstacle-free distance, to some extent a conservative metric of the closeness to a collision. These metrics are the same in the case of a standstill object or an object that stops instantly. The required acceleration metric is different from the other two metrics since it does not relate to a collision event. Instead, it measures the action needed to avoid a collision and hence the closeness to the point where a collision is practically unavoidable. Required acceleration, therefore, gives an earlier indication of when a crash is happening compared to the other two metrics. A variant of the TTC metric that addresses this issue is the time to maneuver (TTM) or time to brake (TTB) metric [57]. These metrics relate to this point of an unavoidable collision with time as a unit.

Advanced Threat Metrics

The mentioned threat assessment methods can also be extended to include more detailed models for the actuation of actions, such as braking, to make them more realistic. The simple models presented here only consider one target at a time, which sometimes underestimates the threat since other objects might block some paths. Including multiple objects in the threat assessment can mitigate this at the cost of increased complexity. There are also a lot of uncertainties in state measurement and prediction. The uncertainties can be countered by introducing safety margins in the deterministic models or using stochastic models instead.

Stochastic models of uncertainties can give a more realistic measurement of the current risk. The models can include both measurement uncertainties and consider multiple future trajectories. Furthermore, stochastic models can be applied to the metrics presented in Section 3.4. For TTC, that would mean that the result will be a distribution of values instead of a single one, as seen in [56]. In addition, stochastic models can also estimate the probability of collision for each given instance, as shown in [59], [60]. By creating stochastic models of the future paths, it is possible to calculate the risk that an object will occupy the same place as the ego vehicle at the same time in the future. We can see this in, e.g., [61], where uncertainties of the measurements are modeled together with the other traffic participants. Then stochastic reachable sets can be used to predict the probability of collision for a particular path of the ego vehicle.

CHAPTER 4

Subset Simulation

Subset Simulation (SuS) is a method used to estimate very low frequencies and, therefore, expensive to sample uniformly. The technique has been applied to estimate very low failure frequencies in engineering systems such as building constructions, mechanical components, and automated vehicles [25], [62], [63]. The principle behind the method is to iteratively explore a parameter space and direct the simulations toward where it is believed to be closest to having a failure or an incident. The parameters resulting in simulations closest to a failure are chosen as a subset in each iteration. The selected scenarios are then used as the starting point for exploring in the next iteration. Consequently, there is a need for a metric that guides the search toward failure. This metric should measure the closeness to failure to rank them and determine if a failure has occurred.

4.1 Performance Metrics

A performance evaluation function is needed to guide the search of SuS and evaluate which parameter values result in simulations closer to failure. This function is defined as $g(\theta)$ and is used to create a Limit-State Function (LSF). The LSF is designed so that a lower value is closer to a failure, and a value below zero is considered a failure. Therefore, the performance function, $g(\theta)$, needs to represent the closeness to a failure and have a clearly defined limit where a failure has occurred. The function is transformed to match the LSF requirement of decreasing values toward its zero failure limit. For each iteration of SuS, an intermediate failure region is defined as:

$$LSF(\boldsymbol{\theta}) \leq y^*, \tag{4.1}$$

where y^* is defined as the value that makes a certain share, p_0, of simulations fulfill this inequality.

4.2 Sampling

In the first iteration of SuS, a Monte Carlo sampling of the parameter space is performed. Next, the samples resulting in simulations closer to a failure are selected to form an intermediate failure region. Finally, Markov Chain Monte Carlo simulations are performed based on the selected samples. The result is new samples that further explore the parameter space, and a new intermediate failure region can be formed.

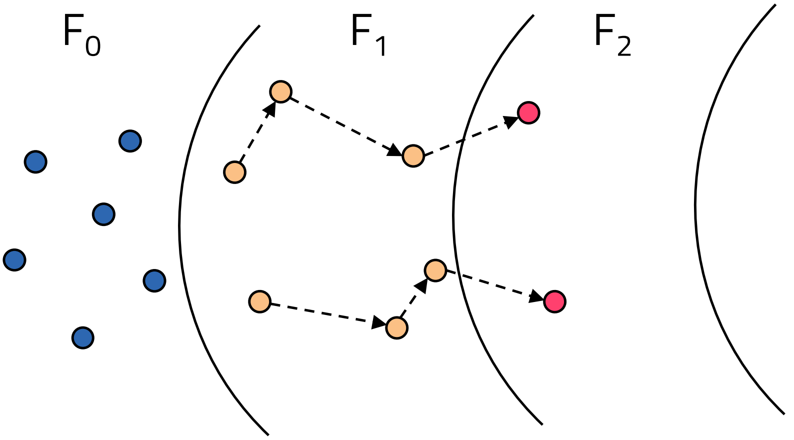


Figure 4.1: Illustration of the first two iterations of the SuS. A new intermediate failure region is created for each iteration, and MCMC sampling is used to generate new samples in that region. Each time, the simulations get closer to the failure region (F), and when enough samples end up in that region, the process is stopped.

This process is repeated until a significant share of the samples are actual failures. The result is a chain of intermediate failure regions ($F_1 \supset F_2 \cdots \supset F_n$) that can be used to calculate the probability of failure:

$$Pr(F) = Pr\left(\bigcap_{i=1}^{n} F_i\right) = \prod_{i=1}^{n} Pr(F_i|F_{i-1}). \tag{4.2}$$

The idea behind SuS is to make these intermediate probabilities large enough so that it is relatively easy to sample [64]. As a result, the original problem of sampling a low probability is transformed into a chain of larger conditional probabilities.

The MCMC sampling at each level i uses the states from $\boldsymbol{\theta}_{i-1}$ that end up in F_i as the starting point for a Markov chain. From that starting state a candidate state $\tilde{\boldsymbol{\theta}}$ is generated from a given PDF $\varphi_n(\boldsymbol{\theta}|F_i)$. If the $\tilde{\boldsymbol{\theta}}$ generates an LSF-value that is lower than the intermediate limit, y_i^*, and thereby makes $\tilde{\boldsymbol{\theta}} \in F_i$, the candidate is selected as the next state ($\boldsymbol{\theta}_{k+1} = \tilde{\boldsymbol{\theta}}$). Otherwise, the current state $\boldsymbol{\theta}_k$ is used as the next state $\boldsymbol{\theta}_{k+1}$. This is continued for each starting state to create multiple chains of states that, in the end, result in N new states in level i ($\boldsymbol{\theta}_i$).

Subset Simulation Algorithm

Here is a description of the subset simulation algorithm as described in [62].

1. Generate N samples $\{\boldsymbol{\theta}_0^{(k)} : k = 1, ..., N\}$ from the PDF $\varphi_n(\boldsymbol{\theta})$
2. Order the samples after the corresponding LSF value $\{LSF(\boldsymbol{\theta}_0^{(k)}) : k = 1, ..., N\}$ in ascending order and find y_1^* as the p_0-percentile.
3. Set $F_1 = \{\boldsymbol{\theta} \in \mathbb{R}^n : LSF(\boldsymbol{\theta}) \leq y_1^*\}$
4. Set $j = 1$ and repeat while $y_j^* > 0$
   a) Start from the N_S samples where $\boldsymbol{\theta}_{j-1}^{(k)} \in F_i$ where $N_S = p_0 N$.
   b) For each $\boldsymbol{\theta}$ of the N_S samples, generate $\frac{1}{p_0} - 1$ new states of a Markov chain from the PDF $\varphi_n(\boldsymbol{\theta}|F_i)$ using MCMC sampling and let the resulting N states be $\boldsymbol{\theta}_j$.
   c) Set $F_{j+1} = \{\boldsymbol{\theta} \in \mathbb{R}^n : LSF(\boldsymbol{\theta}) \leq y_{j+1}^*\}$ where y_{j+1}^* is the p_0-percentile of the ascending LSF-values.
   d) Set $j = j + 1$
5. Count the number of failures (N_f) of the last level where $\boldsymbol{\theta}_{\boldsymbol{j-1}} \in F$
6. Calculate the failure probability $\hat{P}_f = p_0^{j-1} \frac{N_f}{N}$

4.3 Application to Vehicle Safety

The SuS method requires a metric for closeness to failure similar to what is needed for EVT. Therefore, the metrics presented in Section 3.4 can be used for vehicle safety estimations. The difference to EVT is that it is more the ordering of the scenarios that are important than the specific value of the metric. The same scenarios will be selected regardless of the associated metric value, while for EVT, the values affect the shape of the distribution. However, it still applies that different scenario types using the same metric must be comparable in values. Otherwise, the SuS will be biased toward specific scenarios, and some parts of the parameter space will not be explored as much.

CHAPTER 5

Summary of included papers

This chapter provides a summary of the included papers. The included papers provides methods for accelerated testing of an ADS safety performance, both before and after release of the system. Paper A covers a method used during the development phase to validate the safety in simulation. The following papers B, C and D covers a method for statistical safety validation at vehicle level to be used before the release of an ADS software. Paper E introduces a monitor that is used after the launch of the system, which aborts operation if the safety performance is not according to expectations. Lastly, Paper F explores a probabilistic threat metric that can be applied in the previous methods in a more general context.

5.1 Paper A

> **Daniel Åsljung**, C. Zandén, J. Fredriksson, M. K. Vakilzadeh
> On Automated Vehicle Collision Risk Estimation using Threat Metrics in Subset Simulation
> *Published in IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 58-63, Indianapolis, IN, USA.
> ©2021 IEEE DOI: 10.1109/ITSC48978.2021.9564695 .

The Subset Simulation (SuS) method uses a metric to guide the simulations toward failure. Such a metric needs to relate correctly to the closeness of failure between different scenarios. Therefore, it is necessary to investigate how the choice of metric affects the failure rate estimates and how critical this choice is. In Paper A, a selection of different threat metrics is evaluated. The SuS is applied for an ACC function faced with simulated cut-in scenarios. All metrics gave results relatively close to the actual failure rate, and metrics relating to a state where failure could not be avoided proved a little better. This result is in line with previous results for EVT in Paper D and B. The thesis author was responsible for the problem formulation, implementation, analysis, and writing the paper.

5.2 Paper B

Daniel Åsljung, J. Nilsson, J. Fredriksson
Comparing Collision Threat Measures for Verification of Autonomous Vehicles using Extreme Value Theory
Published in 9th IFAC Symposium on Intelligent Autonomous Vehicles,
2016, pp. 57-62, Leipzig, Germany.
©2016 IFAC DOI: 10.1016/J.IFACOL.2016.07.709 .

As described in Chapter 3, there is a need for a measure that reflects the closeness to a collision to use EVT to estimate the collision frequency. The measure needs to be able to continuously show the closeness to a collision and be comparable between different situations. This paper investigates how different threat measures affect the inferences drawn from EVT. Two different types of threat measures are compared and a subset of a larger field test is used as input data, where the vehicles are driven by humans. The results show a clear difference between the two types, especially when looking at the estimated collision frequency. The measure that reflects the closeness to the point where a collision is unavoidable looks much more promising. The thesis author was responsible for the implementation, analysis, and writing the paper.

5.3 Paper C

Daniel Åsljung, J. Nilsson, J. Fredriksson
Validation of Collision Frequency Estimation Using Extreme Value Theory
Published in IEEE 20th International Conference on Intelligent Transportation Systems (ITSC),
2017, pp. 1-6, Yokohama, Japan.
©2017 IEEE DOI: 10.1109/ITSC.2017.8317596 .

In Paper B it was shown that one type of metric showed greater promise of being able to estimate the collision frequency using EVT. In order to be used as a validation method for safety requirements, as described in Chapter 2, the method needs to be shown to correctly estimate the collision frequency. To address this, the metric that was more promising is investigated more in this paper. To validate the correctness of the estimation using EVT, it is compared to an estimate from crash statistics. For the comparison to be valid, the data used for the EVT estimate is from a larger field test made up of 250 000 km driven by humans. The results confirmed the initial conclusions from Paper B that this metric gives credible results. It was also found that the EVT model could be fitted in two different ways resulting in some differences in the inferences drawn. By fitting the model to a few of the most extreme events, the drivers' performance was significantly better than the average human. The conclusion is that this is what can be expected from data based on trained test drivers. The thesis author was responsible for the implementation, analysis, and writing the paper.

5.4 Paper D

Daniel Åsljung, J. Nilsson, J. Fredriksson
Using Extreme Value Theory for Vehicle Level Safety Validation and Implications for Autonomous Vehicles
Published in IEEE Transactions on Intelligent Vehicles,
vol. 2, no. 4, pp. 288-297, Dec. 2017.
©2017 IEEE DOI: 10.1109/TIV.2017.2768219 .

The analysis of different types of threat metrics made in Paper B was done on a limited amount of data, making the results preliminary. In Paper C it was shown that depending on what threshold is used for the EVT model, the inferences drawn could differ. As described in Chapter 3, this process is often performed manually by visual inspection. To efficiently use EVT for validation of safety requirements, this has to be done automatically. In Paper D, the same more extensive field test as in Paper C is used to verify the result received from Paper B. The result from this larger field test is very similar to what was found in Paper B, which further strengthens the conclusion that a metric that reflects the closeness to a point where a collision is unavoidable is the better choice. The Paper also includes an evaluation of three different methods of automatically choosing a threshold for the EVT model. All methods choose a probable threshold for both metrics, suggesting that the whole process can be automatically performed. The thesis author was responsible for the implementation, analysis, and writing the paper.

5.5 Paper E

Daniel Åsljung, C. Zandén, J. Fredriksson
A Risk Reducing Fleet Monitor for Automated Vehicles Based on Extreme Value Theory
Submitted for publication in IEEE Transactions on Intelligent Transportation Systems.

The methods presented in Papers A,B,C & D aims to give evidence of a safe system before launch. However, even if the result is positive with high confidence, there is a residual risk that the system is not performing according to the safety target. It is therefore suggested to monitor the safety of the system after launch. However, a statistical monitor based on observing potential failures poses critical consequences if the system is not performing well enough. We do not want to see a lot of accidents before a sub-performing system is stopped. In Paper E, the strength of EVT shown in Paper D is used to create a safety monitor with predictive abilities. The result from a simulation study, similar to what is used in Paper A, shows a significant risk reduction compared to just observing failures. The risk reduction is achieved by stopping functions not fulfilling the requirement much sooner due to the predictive ability of EVT. Also, a new retrospective version of the BTN metric is introduced that achieves similar risk reduction as the predictive metric while being less conservative. The thesis author was responsible for the problem formulation, concept generation, implementation, analysis, and writing the paper.

5.6 Paper F

Daniel Åsljung, M. Westlund, J. Fredriksson
A Probabilistic Framework for Collision Probability Estimation and an Analysis of the Discretization Precision
Published in IEEE Intelligent Vehicles Symposium (IV),
2019, pp. 52-57, Paris, France.
©2019 IEEE DOI: 10.1109/IVS.2019.8813853 .

The different types of threat measures used in Paper B, D are only considering a single outcome of the future. Consequently, the metrics cannot account for possible future outcomes that could be more serious. In Chapter 3, these are referred to as deterministic threat assessment metrics in the way it treats the future as deterministic. There have also been attempts at constructing a metric stochastic based on stochastic models. In Paper F, a stochastic framework for calculating the probability of collision is presented. It is based on a discrete Markov Chain model populated by the same large data set used in C and D. The focus is on evaluating the precision in the metric and how the discretization could be improved. The thesis author was responsible for the problem formulation, implementation, analysis, and writing the paper.

CHAPTER 6

Concluding Remarks and Future Work

The attached papers present two quantitative methods to validate an ADS's safety performance. First, SuS have been used in simulation to give statistical safety estimates during the development. Secondly, it has been shown how EVT can validate safety on a vehicle level by extrapolating the distribution of near-failures. Consequently, providing evidence that an acceptance criterion is met at the launch. These methods address the first research question by providing evidence during different parts of the development process. Lastly, an approach to using EVT for predictive monitoring of safety after launch is also presented. The monitor thus addresses the second research question and is shown to reduce the risk during the operation significantly.

Common for all methods is the usage of threat metrics, or metrics of closeness to failure, to accelerate the testing. It has been shown in the attached papers that the choice of this metric can have significant effects on the result and the inferences drawn from them. For example, it has been shown that a metric relating to a state where failure is unavoidable, such as BTN, is better for extrapolation using EVT. For SuS, the same metric also showed better results, but the difference was not as significant. The choice of metric for SuS does not seem as critical as for EVT. A possible reason for this is that the specific value of the metric is not crucial in SuS, just the relative order of the scenarios.

It is also found that the constant acceleration prediction in the metric results in a very conservative estimate for the cut-in scenarios in the simulations. A metric calculated retrospectively without any prediction has four times fewer failures than the predictive metric. These results highlight an essential aspect to consider when choosing a metric, and it might differ between different situations how conservative the metric is. Another significant factor for both methods is designing a metric that gives a non-zero threat value to cases with a potential risk of failure. The reason is to have an indication of being close to failure from all causes and types of situations. By having more data points, the confidence in the estimates using EVT can also be increased. It also reduces the risk of missing critical scenarios during the search in SuS and getting a bias in the estimation.

The papers included in this thesis only consider collisions with other vehicles on the road. However, the method can be applied to other types of failures by using an appropriate threat metric. Therefore, there is a need for a metric or a group of metrics that captures a complete set of possible failures. It is preferable to have as few metrics as possible or metrics with comparable values to maximize the number of data points when using EVT. The same applies to SuS to reduce the number of parallel searches needed to validate the full scope of the function.

One paper in the thesis investigated the possibility of a probabilistic metric that could be extended to cover different types of collisions more generally. The metric could capture the possibility of different outcomes and more accurately reflect the closeness to a crash. However, the explored framework requires much data about the possible behavior of other traffic participants, especially in critical scenarios, which is not easily obtained. Compared to the other threat metrics used in the thesis, a probabilistic framework might not be needed to produce stable results in the presented methods. Instead, simple metrics relating to the system's capability could be enough to reflect the closeness of a failure for many types of scenarios.

There are some possible future directions to expand on the results of this thesis. Firstly, the EVT method should be applied to a more extensive field test. Consequently, it can be shown that the failure frequency estimations are credible, and the EVT monitor performs similarly to the simulation environment.

Secondly, it is desired to automate as much of this process as possible. For example, threshold selection is crucial in the EVT process, and improvements should be made in automatically evaluating different thresholds to select the appropriate one.

Lastly, new metrics that consider the severity of the failure should be evaluated. The possibility of measuring the higher severity failures is especially important for an ADS since the acceptance criterion for these will be of a very low frequency.

References

[1] S. Singh, "Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety Facts Crash • Stats. Report No. DOT HS 812 115)," National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. February, 2015.

[2] Volvo Car Group, *Volvo Cars' unsupervised autonomous driving feature Ride Pilot to debut in California - Volvo Cars Global Media Newsroom*, Jan. 2022.

[3] Waymo, *Waymo is opening its fully driverless service to the general public in Phoenix*, Oct. 2020.

[4] Mercedes-Benz Group, *First internationally valid system approval for conditionally automated driving*, Dec. 2021.

[5] BMW Group, *BMW Group, Qualcomm and Arriver to form long-lasting strategic cooperation for joint development of Automated Driving software solutions*, Mar. 2022.

[6] P. Koopman and M. Wagner, "Autonomous Vehicle Safety: An Interdisciplinary Challenge," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 1, pp. 90–96, 2017.

[7] SAE International Surface Vehicle Recommended Practice, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, Apr. 2021. SAE Standard J3016.

[8] H. Winner, S. Hakuli, F. Lotz, and C. Singer, *Handbook of Driver Assistance Systems*. Springer, 2014, pp. 1–30, ISBN: 978-3-319-09840-1.

[9] United Nations, *UN Regulation No. 157: Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems*, 2020.

[10] H. Winner, G. Prokop, and M. Maurer, *Automotive systems engineering. II.* Springer, 2018, ISBN: 978-3-319-61605-6.

[11] J. Schaeuffele and T. Zurawka, *Automotive Software Engineering, Second Edition.* SAE International, Sep. 2016.

[12] *ISO 26262-8:2018: Road vehicles — Functional safety.* International Organization for Standardization. Geneva, Switzerland, 2018.

[13] C. Bergenhem, R. Johansson, A. Söderberg, *et al.*, "How to Reach Complete Safety Requirement Refinement for Autonomous Vehicles," in *CARS - Critical Automotive applications: Robustness & Safety*, 2015.

[14] *ISO 21448:2022: Road vehicles — Safety of the intended functionality.* International Organization for Standardization. Geneva, Switzerland, 2022.

[15] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, *Autonomous Driving.* Springer-Verlag, Berlin, Germany, ISBN: 9783662488454.

[16] D. Bruggner, A. Hegde, F. S. Acerbo, D. Gulati, and T. D. Son, "Model in the Loop Testing and Validation of Embedded Autonomous Driving Algorithms," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, 2021, pp. 136–141.

[17] E. Coelingh, H. Lind, W. Birk, and D. Wetterberg, "Collision Warning with Auto Brake," in *FISITA World Congress*, Yokohama, Japan, 2006.

[18] J. Hillenbrand and K. Kroschel, "A Study on the Performance of Uncooperative Collision Mitigation Systems at Intersection-like Traffic Situations," in *IEEE Conference on Cybernetics and Intelligent Systems*, IEEE, Jun. 2006, pp. 1–6, ISBN: 1-4244-0022-8.

[19] D. Gruyer, S. Choi, C. Boussard, and B. D'Andrea-Novel, "From virtual to reality, how to prototype, test and evaluate new ADAS: Application to automatic car parking," *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 261–267, 2014.

[20] J. E. Stellet, M. R. Zofka, J. Schumacher, T. Schamm, F. Niewels, and J. M. Zöllner, "Testing of advanced driver assistance towards automated driving : A survey and taxonomy on existing approaches and open questions," *IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 1455–1462, 2015.

[21] S. Jesenski, J. E. Stellet, W. Branz, and J. M. Zöllner, "Simulation-Based Methods for Validation of Automated Driving: A Model-Based Analysis and an Overview about Methods for Implementation," *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, pp. 1914–1921, 2019.

[22] D. Zhao, H. Lam, H. Peng, *et al.*, "Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 595–607, 2017, ISSN: 15249050.

[23] D. Zhao, X. Huang, H. Peng, H. Lam, and D. J. Leblanc, "Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 733–744, 2018, ISSN: 15249050.

[24] 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 Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1573–1582, Mar. 2021, ISSN: 15580016.

[25] S. Zhang, H. Peng, D. Zhao, and H. E. Tseng, "Accelerated Evaluation of Autonomous Vehicles in the Lane Change Scenario Based on Subset Simulation Technique," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2018-Novem, pp. 3935–3940, 2018.

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

[27] M. Distner, M. Bengtsson, T. Broberg, and L. Jakobsson, "City Safety-A System Addressing Rear-End Collisions At Low Speeds," in *21st Enhanced Safety Vehicles Conference*, Stuttgart, Germany, 2009.

[28] P. Armitage, C. K. McPherson, and B. C. Rowe, "Repeated Significance Tests on Accumulating Data," *Journal of the Royal Statistical Society*, vol. 132, no. 2, pp. 235–244, 1969.

[29] A. Wald, "Sequential Tests of Statistical Hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[30] S. Pocock, "Group Sequential Methods in the Design and Analysis of Clinical Trials," *Biometrika*, vol. 64, no. 2, pp. 191–199, 1977.

[31] P. C. O'Brien and T. R. Fleming, "A multiple testing procedure for clinical trials.," *Biometrics*, vol. 35, no. 3, pp. 549–56, Sep. 1979.

[32] E. Denney, G. Pai, and I. Habli, "Dynamic Safety Cases for Through-Life Safety Assurance," *Proceedings - International Conference on Software Engineering*, vol. 2, pp. 587–590, Aug. 2015, ISSN: 02705257.

[33] E. Asaadi, E. Denney, J. Menzies, G. J. Pai, and D. Petroff, "Dynamic Assurance Cases: A Pathway to Trusted Autonomy," *Computer*, vol. 53, no. 12, pp. 35–46, Dec. 2020, ISSN: 15580814.

[34] J. Nilsson, A. C. E. Ödblom, and J. Fredriksson, "Worst-Case Analysis of Automotive Collision Avoidance Systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 1899–1911, Apr. 2016, ISSN: 0018-9545.

[35] J. Woodcock, P. G. Larsen, J. Bicarregui, and J. Fitzgerald, "Formal methods," *ACM Computing Surveys*, vol. 41, no. 4, pp. 1–36, Oct. 2009, ISSN: 03600300.

[36] J. Krook, "Formal Methods and Safety for Automated Vehicles," Ph.D. dissertation, Chalmers University of Technology, Gothenburg, 2022, ISBN: 9789179057312.

[37] J. Nilsson, J. Fredriksson, and A. Odblom, "Verification of Collision Avoidance Systems using Reachability Analysis," in *IFAC 19th World Congress*, Cape Town, South Africa, 2014.

[38] M. Althoff and J. M. Dolan, "Online Verification of Automated Road Vehicles Using Reachability Analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, Aug. 2014.

[39] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a Formal Model of Safe and Scalable Self-driving Cars," Aug. 2017.

[40] Y. Selvaraj, A. Farooqui, G. Panahandeh, W. Ahrendt, and M. Fabian, "Automatically Learning Formal Models from Autonomous Driving Software," *Electronics 2022, Vol. 11, Page 643*, vol. 11, no. 4, p. 643, Feb. 2022, ISSN: 2079-9292.

[41] J. Krook, L. Svensson, Y. Li, L. Feng, and M. Fabian, "Design and formal verification of a safe stop supervisor for an automated vehicle," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2019-May, pp. 5607–5613, May 2019, ISSN: 10504729.

[42] Y. Selvaraj, "Safety Proofs for Automated Driving using Formal Methods," Ph.D. dissertation, Chalmers University of Technology, Gothenburg, 2022, ISBN: 9789179057381.

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

[44] E. Castillo, *Extreme Value Theory in Engineering.* Elsevier, 1988, ISBN: 9780121634759.

[45] M. Gilli and E. Këllezi, "An Application of Extreme Value Theory for Measuring Financial Risk," *Computational Economics 2006 27:2*, vol. 27, no. 2, pp. 207–228, May 2006, ISSN: 1572-9974.

[46] P. Embrechts, S. I. Resnick, and G. Samorodnitsky, "Extreme Value Theory as a Risk Management Tool," *North American Actuarial Journal*, vol. 3, no. 2, pp. 30–41, Apr. 2013, ISSN: 10920277.

[47] D. Åsljung, *On Safety Validation of Automated Driving Systems using Extreme Value Theory*, Licentiate thesis. Chalmers University of Technology, Gothenburg, Dec. 2017.

[48] P. Songchitruksa and A. P. Tarko, "The extreme value theory approach to safety estimation," *Accident Analysis and Prevention*, vol. 38, no. 4, pp. 811–822, Jul. 2006.

[49] J. K. Jonasson and H. Rootzén, "Internal validation of near-crashes in naturalistic driving studies: A continuous and multivariate approach.," *Accident Analysis and Prevention*, vol. 62C, pp. 102–109, Sep. 2013, ISSN: 1879-2057.

[50] H. Farah and C. L. Azevedo, "Safety analysis of passing maneuvers using extreme value theory," *IATSS Research*, vol. 41, no. 1, pp. 12–21, 2017, ISSN: 03861112.

[51] L. Zheng and T. Sayed, "Application of Extreme Value Theory for Before-After Road Safety Analysis," *Transportation Research Record*, vol. 2673, no. 4, pp. 1001–1010, 2019.

[52] G. Gecchele, F. Orsini, M. Gastaldi, and R. Rossi, "Freeway rear-end collision risk estimation with extreme value theory approach. A case study," *Transportation Research Procedia*, vol. 37, pp. 195–202, 2019, ISSN: 23521465.

[53] F. Orsini, G. Gecchele, M. Gastaldi, and R. Rossi, "Large-scale road safety evaluation using extreme value theory," *IET Intelligent Transport Systems*, vol. 14, no. 9, pp. 1004–1012, Sep. 2020, ISSN: 1751956X.

[54] L. Zheng, T. Sayed, and M. Essa, "Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators," *Accident Analysis and Prevention*, vol. 123, no. December 2018, pp. 314–323, 2019, ISSN: 00014575.

[55] J. Cavadas, C. L. Azevedo, H. Farah, and A. Ferreira, "Road safety of passing maneuvers: A bivariate extreme value theory approach under non-stationary conditions," *Accident Analysis and Prevention*, vol. 134, Jan. 2020, ISSN: 00014575.

[56] J. Jansson, "Collision Avoidance Theory with Application to Automotive Collision Mitigation," Ph.D. dissertation, Linköping University, 2005, p. 188, ISBN: 9185299456.

[57] L. Westhofen, · . C. Neurohr, · . T. Koopmann, *et al.*, "Criticality Metrics for Automated Driving: A Review and Suitability Analysis of the State of the Art," *Archives of Computational Methods in Engineering 2022*, vol. 1, pp. 1–35, Aug. 2022, ISSN: 1886-1784.

[58] M. Brännström, J. Sjöberg, and E. Coelingh, "A situation and threat assessment algorithm for a rear-end collision avoidance system," in *IEEE Intelligent Vehicles Symposium*, Eindhoven, The Netherlands: IEEE, 2008, ISBN: 978-1-4244-2568-6.

[59] A. Eidehall and L. Petersson, "Statistical Threat Assessment for General Road Scenes Using Monte Carlo Sampling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 137–147, 2008, ISSN: 1524-9050.

[60] D. Greene, J. Liu, J. Reich, *et al.*, "An efficient computational architecture for a collision early-warning system for vehicles, pedestrians, and bicyclists," *IEEE Transactions on Intelligent Transportation Systems*, 2011, ISSN: 15249050.

[61] M. Althoff, O. Stursberg, and M. Buss, “Model-Based Probabilistic Collision Detection in Autonomous Driving,” *IEEE Transactions on Intelligent Transportation Systems*, Jun. 2009.

[62] S. K. Au and J. L. Beck, “Estimation of small failure probabilities in high dimensions by subset simulation,” *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263–277, 2001, ISSN: 02668920.

[63] H. S. Li, S. Xia, and D. M. Luo, “A probabilistic analysis for pin joint bearing strength in composite laminates using Subset Simulation,” *Composites Part B: Engineering*, vol. 56, pp. 780–789, Jan. 2014, ISSN: 1359-8368.

[64] I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub, “MCMC algorithms for Subset Simulation,” *Probabilistic Engineering Mechanics*, vol. 41, pp. 89–103, 2015, ISSN: 18784275.