Technische Universität München

Fakultät für Maschinenwesen

Lehrstuhl für Fahrzeugtechnik

How to Define System-Specific Corner Cases for the Type Approval of Automated Vehicles

Thomas Ponn, M.Sc.

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der
Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Prof. Dr.-Ing. Boris Lohmann
Prüfer der Dissertation:	1. Prof. Dr.-Ing. Markus Lienkamp
	2. Prof. Dr.-Ing. Matthias Althoff

Die Dissertation wurde am 21. Oktober 2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 12. Januar 2021 angenommen.

We don’t claim that the cars are going to be perfect. Our goal is to beat human drivers.

Sergey Brin

Acknowledgments

This work was created during my work as a research assistant at the Institute of Automotive Technology at the Technical University of Munich in cooperation with TÜV SÜD Auto Service GmbH from 2017 to 2020.

First and foremost, I would like to thank my supervisor Prof. Dr.-Ing. Markus Lienkamp, who gave me the opportunity to do my doctorate. I would also like to thank him for the trust he has placed in me and the freedom to make decisions about the project. I would also like to thank my second examiner Prof. Dr.-Ing. Matthias Althoff for taking over the second assessment and for his valuable advice. I would also like to thank Prof. Dr.-Ing. Boris Lohmann for taking over the chairmanship of the examination.

I would like to pay my special regards to TÜV SÜD Auto Service GmbH for financing the research project. Special thanks goes to Christian Gnandt, Emmeram Klotz, Christoph Miethaner, Dieter Ludwig, Dirk Fratzke, Alexander Schwab, Benjamin Koller and Houssem Abdellatif. I would like to thank all of you for your continuous support and many enriching discussions.

I wish to express my sincere appreciation to my former FTM colleagues for an interesting, informative and unforgettable time. Thank you very much for your commitment, your support and the time spent together. I would especially like to thank Dr.-Ing. Frank Diermeyer who, as group leader at the chair, supported the development of the work and accompanied it with professional help. I would also like to thank my proofreaders Simon Hoffmann, Stefan Riedmaier and Tim Stahl.

I would like to pay my special regards to my students, who actively supported me in all phases of the project. I would also like to thank the RWTH Aachen University for providing the highD data set.

My family, especially my grandparents, parents, brother and sister as well as my girlfriend Franziska, deserve my deepest gratitude. You have encouraged and always supported me in all my plans. I thank you for your continued support.

Garching, October 2020

Thomas Ponn

Contents

List of Abbreviations

ACC	Adaptive Cruise Control
ACSF	Automatically Commanded Steering Function
ADAS	Advanced Driver Assistance Systems
AI	Artificial Intelligence
ALKS	Automated Lane Keeping System
AV	Automated Vehicle
DDT	Dynamic Driving Task
GA	Genetic Algorithm
GRVA	Working Party on Automated/Autonomous and Connected Vehicles
HAC	Hierarchical-Agglomerative Clustering
IDM	Intelligent Driver Model
KPI	Key Performance Indicator
LKA	Lane Keeping Assistant
MC	Monte Carlo
MOBIL	Minimizing Overall Braking Induced by Lane Changes
NGSIM	Next Generation SIMulation
NHTSA	National Highway Traffic Safety Administration
ODD	Operational Design Domain
OEDR	Object and Event Detection and Response
ROI	Region Of Interest
SAE	Society of Automotive Engineers
SBA	Scenario-based Approach
SHAP	SHapley Additive exPlanations
SNR	Signal-to-Noise Ratio
SOTIF	Safety Of The Intended Functionality
TP	Traffic Participant
TTB	Time-to-Brake
TTC	Time-to-Collision
UNECE	United Nations Economic Commission for Europe
VMAD	Validation Methods for Automated Driving
VUT	Vehicle Under Test
WHO	World Health Organization
XiL	X-in-the-Loop

Formula Symbols

Formula Symbols	Unit	Description
a	$\mathrm{m/s^2}$	Acceleration
a_{ego}	$\mathrm{m/s^2}$	Acceleration of ego-vehicle
$a_{\mathrm{follow,initial}}$	$\mathrm{m/s^2}$	Acceleration of following vehicle in initial ego lane
$a_{\mathrm{follow,target}}$	$\mathrm{m/s^2}$	Acceleration of following vehicle in target lane
a_{safe}	$\mathrm{m/s^2}$	Maximal deceleration that is considered as safe
$a^*_{\mathrm{x,ego}}$	$\mathrm{m/s^2}$	Desired Intelligent Driver Model (IDM) acceleration
$a_{\mathrm{x,max}}$	$\mathrm{m/s^2}$	Maximal allowed acceleration
$a_{\mathrm{x,max,dec}}$	$\mathrm{m/s^2}$	Maximal reachable deceleration
$A_{\mathrm{blindspot}}$	$\mathrm{m^2}$	Area of the blind spots
A_{ROI}	$\mathrm{m^2}$	Area of the Region Of Interest (ROI)
$b_{\mathrm{x,com}}$	$\mathrm{m/s^2}$	Comfortable deceleration
B_{n}	$\mathrm{1/s}$	Noise bandwidth
$\mathbf{c}_{\mathrm{factors}}$	-	Vector of the complexity factors
$\mathbf{c}_{\mathrm{weighting}}$	-	Vector of the complexity weightings for each factors
C_{GA}	-	Mean complexity of a scenario used for the Genetic Algorithm (GA)
C_{scenario}	-	Complexity of a scenario
C_{scene}	-	Complexity of a scene
d	m	Euclidean distance
d_{safety}	m	Legally specified and speed-dependent safety distance
$\Delta d_{\mathrm{overall}}$	m	Overall deviation of the prediction
f_{highD}	Hz	Frame rate of the highD data
$f(v_j)$	-	Heuristic evaluation function via node v_j
$g(v_j)$	-	Cost of the path from the start to node v_j
G_{e}	-	Emitter antenna gain factor
G_{r}	-	Receiver antenna gain factor

$h(v_j)$	-	Estimate of remaining cost from node v_j to final node
H	-	Test statistic
H_{crit}	-	Critical value of the test statistic
j_x	$\mathrm{m/s^3}$	Jerk in x-direction
j_y	$\mathrm{m/s^3}$	Jerk in y-direction
J_{GA}	-	Cost function of the GA
J_{Node}	-	Cost function of a node
J_{Path}	-	Cost function of the entire path
k	$\mathrm{J/K}$	Boltzmann constant
k_{J}	-	Weighting factor
l_{coll}	m	Length of potential collision area
l_{ego}	m	Length of ego-vehicle
L_{oa}	-	Overall damping coefficient
n_{comp}	-	Number of comparisons
$n_{\text{ele},i}$	-	Number of elements (ratings) of factor i
n_{experts}	-	Number of participating experts
n_{factors}	-	Number of complexity factors
n_{ff}	-	Factor for reducing the number of frames during optimization
n_{frames}	-	Number of frames of a scenario
$n_{\text{frames,opt}}$	-	Number of frames of a scenario used in optimization
$n_{\text{scenarios}}$	-	Number of scenarios
n_{TPs}	-	Number of traffic participants
n_{var}	-	Number of optimization variables
N_{total}	-	Total number of elements (ratings)
p	-	Politeness factor
P_{D}	-	Detection probability
P_{e}	W	Emitting power
P_{jerk}	-	Penalty regarding the jerk of vehicles
P_{n}	W	Power of the noise
P_{pos}	-	Penalty regarding the position of vehicles
P_{r}	W	Power of the received signal
P_{vel}	-	Penalty regarding the velocity of vehicles
P_{GA}	-	Penalty function of the GA

$Q_{0.05}$	-	Value of the standard normal distribution for adjusted $\frac{\alpha}{2}$
R	m	Range
R_{curve}	m	Curve radius
$R_{\text{factor},i}$	-	Rank sum of factor i
$R_{\text{scenario},i}$	-	Rank sum of scenario i
s	m	Distance between two vehicles
s^*	m	Desired distance between two vehicles
s_0	m	Linear jam distance
s_1	m	Non-linear jam distance
$S^{\sigma}_{c_{\text{factors},j}}$	-	Sigma-normalized derivative
$t_{\text{ch,predict}}$	s	Prediction time for challenger identification
$t_{\text{gap},i}$	s	Time gap
$t_{\text{laneMarking}}$	s	Time a vehicle drives over a lane marking
t_{predict}	s	Prediction time for complexity assessment
t_{ttb}	s	Time-to-brake
Δt_{highD}	s	Step size in highD data set
Δt_{opt}	s	Step size during optimization
T_{sys}	K	System noise temperature
T_g	s	Desired time gap
v	km/h	Velocity
v_{ego}	km/h	Velocity of ego-vehicle
v_j	-	Node j of the grid
v_{set}	km/h	Desired velocity
Δv	km/h	Velocity difference between two vehicle
v_0	km/h	Desired velocity
V	-	Variance
w_{coll}	m	Width of potential collision area
$\mathbf{x}_{\text{oap}}$	-	Optimal approaching path
$\mathbf{y}_{\text{opt}}$	m/s^2	Vector of the optimization variables
α	-	Significance level
α_{cronbach}	-	Cronbach's Alpha
δ	-	Acceleration exponent
δ_{bias}	-	Keep-right bias

δ_{th}	-	Switching threshold
λ	m	Wavelength of emitted signal
σ	m^2	Radar cross-section
$\sigma_{c_{\mathrm{factors},j}}$	-	Variance of input j
$\sigma_{C_{\mathrm{scenario}}}$	-	Variance of output
τ_k	-	Number of tied observation with rank k
ϕ_{s}	rad	Azimuth angle of a sensor
ψ	rad	Yaw angle
ψ_{s}	rad	Elevation angle of a sensor

1 Introduction

1.1 Motivation

According to the World Health Organization (WHO) [1], more than one million people died in road accidents in 2016. In addition to improving passive vehicle safety, Automated Vehicles (AVs) (Level 3 and higher according to Society of Automotive Engineers (SAE) [2]) should make a substantial contribution to significantly reduce this number. In order to achieve this goal, a considerable amount of resources has been invested in the implementation of such systems in recent years, so that various prototypes of these vehicles currently exist. Already with the start of test drives of these systems on public roads it has become clear that the safety of AVs is an extremely sensitive issue, which means that its consideration in research and industry is becoming increasingly important. The major difference to Level 2 Advanced Driver Assistance Systems (ADAS) currently available in mass production vehicles is that responsibility is transferred from the driver to the vehicle from Level 3 onwards. This means that the driver does not have to constantly monitor the system and intervene immediately in the event of a system fault, as is the case with Level 2. Consequently, the safety of AVs must be thoroughly tested before market launch, which is a demanding task [3–5].

When assessing the safety of AVs, various aspects must be taken into account. One of these is, for example, functional safety, where the effects of random hardware and software failures are investigated and procedures to minimize these failures are recommended [6]. Another aspect is the so-called Safety Of The Intended Functionality (SOTIF), where it is checked whether risks or hazards resulting from the AV are caused by functional insufficiencies of the intended functionality [7]. An alternative term for SOTIF is the so-called Object and Event Detection and Response (OEDR) according to the National Highway Traffic Safety Administration (NHTSA) [8, 9]. It examines whether the vehicle is able to correctly detect objects and events and to plan and execute an appropriate response, which in turn corresponds to the intended functionality. This must be checked for the entire Operational Design Domain (ODD) of the system. Since this aspect of safety is the central topic of this thesis and no further safety aspects are considered, the investigation of SOTIF / OEDR capabilities of the vehicle will be considered equivalent to the term safety assessment for the further course. It is explicitly emphasized that a vehicle that has been successfully validated with SOTIF / OEDR cannot be regarded as holistically safe, because further aspects like functional safety have to be considered. Furthermore, we only consider the ODD specified by the manufacturer. Therefore, tests whether the vehicle correctly detects an exit from the ODD are not covered by this thesis.

The biggest challenge in the safety assessment of AVs is that road traffic is an open parameter space in which an infinite number of different traffic situations can occur. A complete safety assessment is therefore not possible. For this reason, the so-called Scenario-based Approach (SBA) has been developed [10], which focuses on individual scenarios that are important for safety assessment. All situations that do not contribute to the safety demonstration, such as

driving on an empty straight motorway, are neglected, thus reducing the effort required. The question remains of how all the necessary scenarios for a sufficiently accurate safety assessment can be identified.

The safety of the system must not only be checked during development, but also before it can be launched on the market. For this purpose, it must receive a type approval from the legislator. The basis for a type approval are (for Europe) the regulations of the United Nations Economic Commission for Europe (UNECE). Tests included in these documents are carried out by an independent third party (technical service) and are intended to provide minimum safety requirements for all types of vehicles. For the type approval of AVs, new regulations and concepts must be developed analogous to the safety assessment accompanying the development. In this context, the technical service has an expanding responsibility for the definition of concrete test scenarios due to the increasingly generalized formulation of future regulations. One consequence of this is an enhanced integration of the approval process into the development process. Type approval by an independent technical service is the main application of the methodology developed in this thesis. Therefore, the type approval process as well as the specific requirements for type approval of AVs are explained in detail in Section 2.4.

In summary, the type approval of AVs is essential for their successful market launch, manufacturers, legislators and technical services have the responsibility to protect society from intolerable risks. For this purpose, an efficient methodology for technical services for the type approval of AVs must be developed and established.

1.2 Contributions

The basic objective of the present work is to contribute to an efficient safety assessment in the type approval process of AVs conducted by a independent third party (technical service), thus supporting the market introduction of AVs. The main contributions of this work are explained in more detail below.

1. **Taxonomy of the SBA:**
   There are many publications in the literature dealing with the safety assessment of AVs using the SBA. A classification of this literature into a generic overall concept is not possible because there is still a lack of a uniform understanding of the generic approach. Therefore, a taxonomy for the SBA is introduced in Section 2.3, which is used to classify the existing literature. Similarly, there is still a lack of an uniform definition and use of terms. Therefore, in Section 2.1 already established definitions will be taken up and these will be extended by definitions missing in the state of the art.
2. **Identification of special requirements for the type approval of AVs:**
   Almost all of the literature is focused on safety assessment during the development of the systems from the perspective of a vehicle manufacturer. Since the aim of this thesis is a methodology for technical services to efficiently conduct future type approval of AVs according to upcoming UNECE regulations [11], the special requirements from the point of view of an independent technical service are defined in Section 2.4 and 3.1.

3. **Development of a methodology to identify corner cases:**
   Chapter 3 introduces a novel methodology for the identification and definition of challenging scenarios that can support a technical service in carrying out the type approval. For this purpose, three different sub-methods are defined in order to address the overall system consisting of a sense-, plan- and act-module. The development process is based on the special requirements for the type approval of AVs.

4. **Implementation of the methodology:**
   Besides the development of the overall and sub-methods, the exemplary implementation and validation of the sub-methods is an important contribution of the present work. For this purpose, the sub-methods are implemented and their results are analyzed in Section 4.1 - 4.3. The focus is especially on the third sub-method, the development and validation of a complexity metric (Section 4.3), which allows an objective evaluation of the difficulty of test scenarios. The focus is on this sub-method, because the first two sub-methods are already published by the author.

1.3 Structure of the Work

An overview of the structure of the present work is shown in Figure 1.1. After a short introduction to the topic of AVs and their safety verification, the current state of the art is discussed in Chapter 2. First, important terms are defined and then an overview of concepts for validating the safety is given. The focus here is strongly emphasized on the SBA. Then the type approval process is described in detail and special requirements for type approval of AVs are derived. At

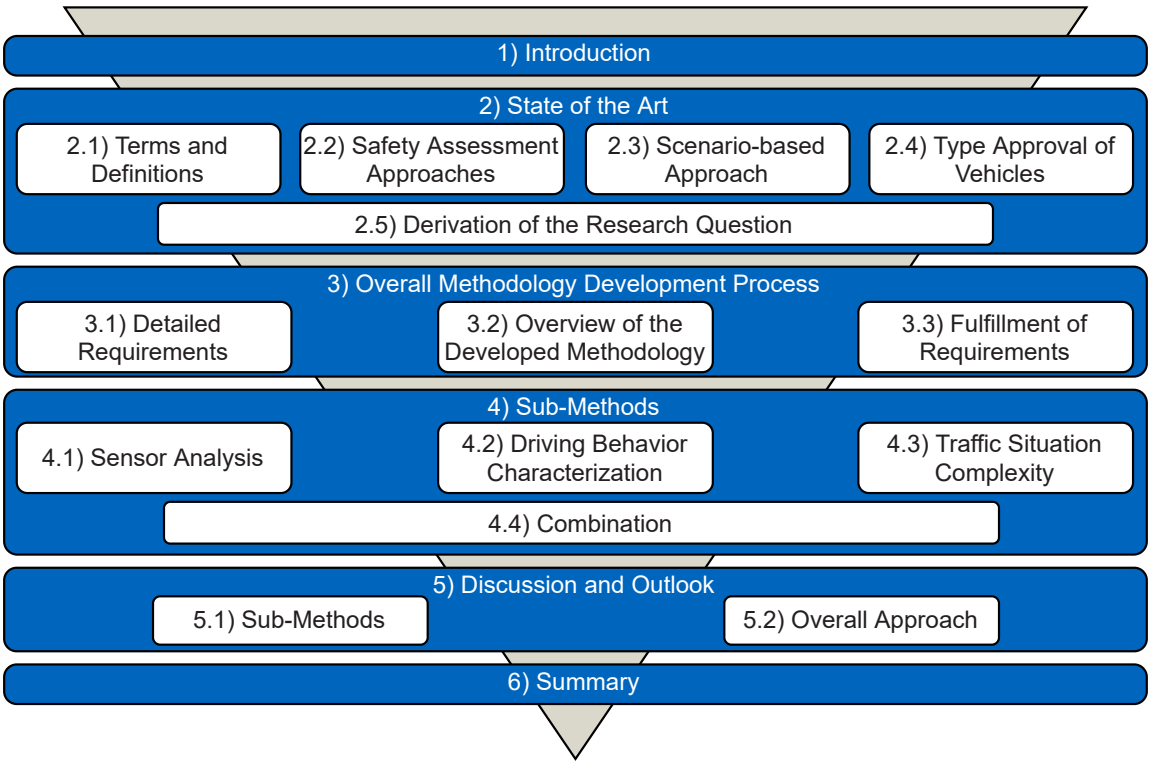


Figure 1.1: Structure of the thesis.

the end of Chapter 2, the research question of the present thesis is derived. Chapter 3 describes the procedure with which the approach was developed and also provides an overview of it. The defined sub-methods are described in Chapter 4 and their results are analyzed. Based on this, the results are discussed in Chapter 5 and an outlook on future research work is given. Finally, this work is summarized in Chapter 6.

2 State of the Art

This chapter first introduces important terms and defines their meaning. It then provides a comprehensive overview of the SBA. Building on this, the following section identifies specific requirements for the type approval of AVs. Finally, a research question is derived from the findings.

2.1 Terms and Definitions

For a better understanding, the most important terms are defined in a logical order so that previous definitions can be taken into account. The definitions are therefore not arranged alphabetically.

Automated Vehicle (AV): An AV is a vehicle with an automation level of at least Level 3 according to SAE [2]. If the AV is the vehicle to be tested, it is also called Vehicle Under Test (VUT) or ego-vehicle.

Dynamic Driving Task (DDT): According to SAE [2], DDT includes all of the real-time operational and tactical functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints. It includes longitudinal and lateral vehicle motion control and monitoring the driving environment via object and event detection, recognition, classification, and response preparation.

Operational Design Domain (ODD): According to SAE [2], the ODD is defined by the operating conditions for which the AV was developed. The ODD can be restricted, for example, by road classes (e. g. highway or city center) or environmental conditions (e. g. weather conditions).

Object and Event Detection and Response (OEDR): Detection of objects that are relevant to the immediate DDT, as well as the appropriate response to such circumstance. The AV is responsible for performing the OEDR while in its ODD. [2]

Safety Of The Intended Functionality (SOTIF): analyzes if hazards from functional insufficiencies of the intended functionality occur [12].

Scene: A scene describes the entire content of the environment at a specific time. All values of the describing parameters correspond to reality (ground truth). [13]

Situation: A situation is defined as the subjective perception of the scene from the perspective of an element therein (e. g. ego-vehicle). All information which is necessary to decide on the further behavior of the element is available in subjective form. [13]

Scenario: A scenario is a chronological sequence of individual scenes that depend on each other. Starting from a starting scene, actions and events occur through the participating elements. The aims and objectives of individual participants can also be included. [13]

Scenario description: Furthermore, Menzel et al. [14] distinguish between three descriptions of scenarios – functional, logical and concrete scenarios:

Functional Scenario: Functional scenarios describe scenarios on a semantic level. The entities and the relationships of these entities are described by a linguistic representation. The vocabulary used to describe functional scenarios is specific to the use case and can have different levels of detail. [14]

Logical Scenario: Based on the functional scenarios, the logical ones describe the scenario in the state space. They represent the entities and the relationships of these entities using parameter ranges in the state space. The parameter ranges can be specified using probability distributions. In addition, the parameter range relationships can be optionally specified using correlations or numerical conditions. A logical scenario thus contains a formal notation of the scenario. [14]

Concrete Scenario: When specific numerical values for each parameter are assigned to the logical scenarios, a concrete scenario is defined [14]. Theoretically, an infinite number of concrete scenarios can be derived from one logical scenario by means of an arbitrarily fine discretization of continuous parameters.

Semi-concrete Scenario: In the present work, we also define semi-concrete scenarios. This means that not all parameters of the logical scenario are exactly defined, but the parameter space is narrowed by reducing the parameter ranges or by defining a part of the parameters. All undefined parameters remain as parameter ranges or can be chosen to default values.

For the definition of logical and concrete scenarios all descriptive parameters are needed. In order to do this in a structured way, Bagschik et al. [15] introduce a five-layer model. The five layers are defined as below:

Layer 1: Road-level
Layer 2: Traffic infrastructure
Layer 3: Temporary manipulation of Layer 1 and Layer 2
Layer 4: Objects
Layer 5: Environment

Traffic Participant (TP): All kinds of movable objects (part of Layer 4) within a traffic situation are called TPs. Among others, this includes pedestrians, cyclists, motorcycles, passenger cars and trucks.

Key Performance Indicator (KPI): For the evaluation of AVs, KPIs are used. These consist of a metric and a criterion in which range the measured value is allowed to occur. An example of a metric is the Time-to-Collision (TTC) and an associated criterion may be that the measured value may not be less than a predefined value.

Test Case: A test case consists of a concrete scenario together with associated KPIs [12].

Types of concrete scenarios: Various adjectives are used in literature to describe concrete scenarios. Since these are not used uniformly, they are defined here and additional adjectives used in this thesis are introduced.

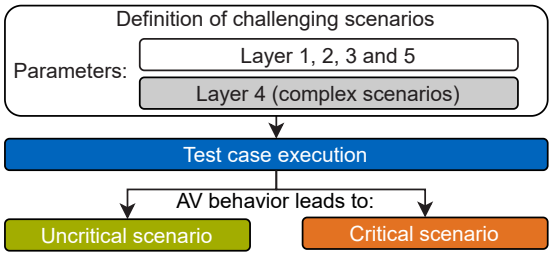
Relevant scenario: All scenarios that are relevant for the certification of AVs. These are all scenarios of the safety assessment and also scenarios that are not safety related, e. g. to check whether the ego-vehicle is capable to comply to all traffic rules.

Critical scenario: A critical scenario is a concrete scenario in which the safety performance of the ego-vehicle is insufficient. Therefore, it means an assessment of the ego-vehicle behavior. It is only determinable after test case execution and the behavior of different AV-functions lead to different criticality-results for the same concrete scenario. [16, 17]

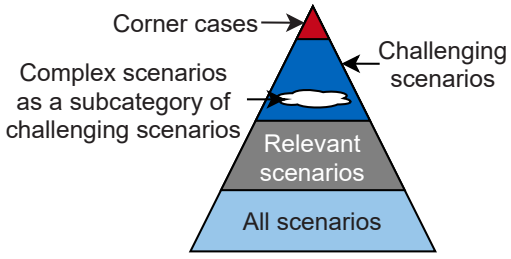
Challenging scenario: It means an assessment of a concrete scenario itself. It is determinable before test case execution and independent of the AV-performance. Whether a concrete scenario is challenging or not, depends on the chosen parameter values. Therefore, challenging can be seen as the difficulty for the AV to master the concrete scenario without the occurrence of a critical situation. [16, 17]

Complex scenario: The definition is the same as for challenging, but for complex scenarios the focus is especially on the objects in Layer 4 of the five-layer model [15]. Complex scenarios are therefore a subset of challenging scenarios where the difficulty results from the behavior of the participating road users [16]. The relationship between critical, challenging and complex scenarios is shown in Figure 2.1a.

Corner case scenario: Concrete scenario in which two or more parameter values are each within the capabilities of the system, but together constitute a rare condition that challenges its capabilities [18]. Corner cases can therefore be considered as the most challenging of the challenging scenarios. A summary of the different kinds of scenarios is given in Figure 2.1b.



(a) Difference between challenging, complex and critical scenarios. The figure is adapted from [16].



(b) Amount of possible scenarios before test case execution.

Figure 2.1: Classification of different kinds of scenarios. In Figure 2.1a, the focus is on the distinction between before and after test case execution. Complexity is a subcategory of challenging and means the difficulty of a scenario due to the behavior (trajectories) of the objects (TPs) that are part of Layer 4 of the five-layer model [15]. In Figure 2.1b the focus is on the number of definable concrete scenarios before test case execution. The pyramid-like shape symbolizes that starting from all scenarios up to the corner cases there are fewer and fewer concrete scenarios that can be identified.

Challenger: In general, the challenger is the most difficult aspect of a challenging scenario. The challenger can be, for example, another TP or difficult weather conditions.

Macroscopic assessment: For a socially accepted market introduction of AVs, it is crucial that they have a lower accident probability than human drivers [19]. To be able to make such a macroscopic (statistical) statement about the overall impact of AVs on traffic, a large amount of data must be available [20, p. VII].

Microscopic assessment: In contrast, particularly in scenario-based safety assessment, individual traffic situations (scenarios) are tested and evaluated. The evaluation of a single scenario is called microscopic assessment. The transition from a microscopic assessment of a single scenario to a macroscopic assessment of safety is one of the most important challenges of the SBA. [20, p. VII]

2.2 Overview of Safety Assessment Approaches

The main content of this section has already been published open access in the author's previous publication [17]. For a better understanding of the present thesis a summary of the already published work is given.

Currently, in literature, seven approaches are used for assessing the SOTIF and OEDR related capabilities and therefore the safety of AVs, which are shown in Figure 2.2. The SBA (highlighted in orange) is of major importance for the present work. In the following, all seven approaches are briefly introduced.

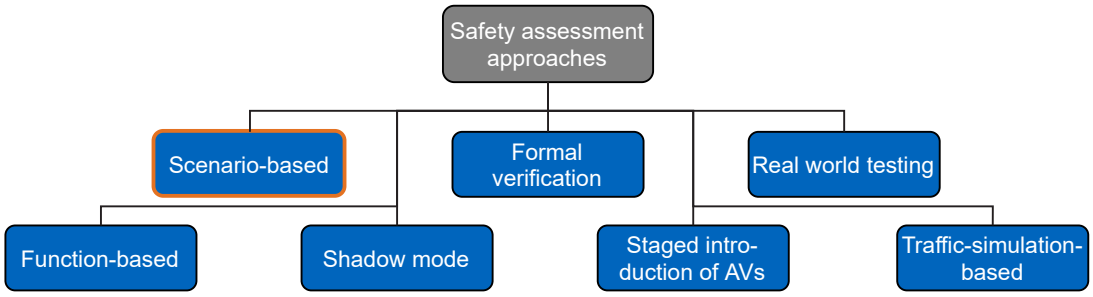


Figure 2.2: Overview of the seven existing approaches to SOTIF and OEDR related safety assessment of AVs. The Figure is adapted from [17].

Scenario-based Approach

The SBA is the most promising among the approaches currently available in literature because it has the potential to provide an efficient and reliable safety evaluation.The approach is described in detail in Section 2.3. Only a short introduction is given here in order to be able to make a more understandable distinction and differentiation from the other approaches.

By definition (Section 2.1) a scenario is a sequence of actions and events. If, for example, a typical highway ride is examined, there is a considerable period of time during which no actions and events occur. The SBA, which is also used in research projects (e. g. in Germany [10], Japan [21] and Singapore [22]), neglects the part without significant actions and thus reduces the test scope. In addition, frequent scenarios, such as cut-in situations with large relative distances and a higher speed of the leading vehicle, which do not provide relevant information for the safety assessment, can also be neglected. Nevertheless, the question remains unanswered which scenarios have to be considered in scenario-based tests and how they can be found in order to make a reliable statement about the vehicle's safety. With the SBA, individual scenarios are tested and microscopically evaluated. After a large number of these microscopic evaluations have been carried out, a transformation to a macroscopic safety statement has to be performed.

Formal Verification

Formal verification is a mathematical method by which the safety of systems can be formally demonstrated across the entire ODD. It is not a selection technique for dividing the parameter space into scenarios and is therefore not part of the SBA. Disadvantages of this approach are, on the one hand, that often simplifying assumptions have to be made and that "only" the trajectory planning module and not the overall vehicle is considered. For a separate validation of this module, formal verification can be a promising approach in the future. For detailed information it is referred to important publications in this area [23–44].

Real World Testing

An exclusively distance-based evaluation of safety by field tests is no longer economically feasible at higher levels of automation. In order to determine with sufficient certainty that AVs exceed the safety level of human drivers by a defined factor, 11 billion miles would have to be driven in the USA according to [45]. In this context, exceeding means that fewer fatal accidents occur. Analogous statistical investigations exist for Germany, where Wachenfeld and Winner [46] conclude that a highway chauffeur needs about 6.6 billion test kilometers. With a low degree of automation, testing in real world is the standard. But from Level 3 on, the required scope increases so much due to the transfer of responsibility that it is no longer economically feasible.

Function-based Approach

In function-based testing, the system's functions are checked on the basis of requirements in tests on the test track or in simulation. This is a widely used procedure for ADAS. Current ISO standards (e. g. ISO 15622 for Adaptive Cruise Control (ACC)) and UNECE regulations (e. g. UNECE R131 for Advanced Emergency Braking Systems) for ADAS follow this method and define a number of precisely specified tests for the individual systems, which check the basic functionality and thus ensure a minimum level of safety. This reduced testing effort is possible with ADAS on the one hand because of the reduced functional scope of the systems and on the other hand because the driver has to monitor the system permanently.

In order for the function-based approach to be applicable, the functionalities of the system must be defined. This is feasible for ADAS but difficult for AVs because it is impossible to specify the required functionality of AVs in all possible situations. In addition, future standards and regulations must not use a small selection of predefined standardized scenarios for testing AVs, because this leads to performance optimization in these test cases. As a result, the evaluation result will not reflect the actual driving behavior of the system in real road traffic.

The type approval process in general and the problems and challenges briefly outlined here are discussed in detail in Section 2.4.

Shadow Mode

Wang and Winner [47] present a method in which the automated driving function in production vehicles is passively executed, which is sometimes called shadow mode. The driving function is provided with the real inputs of the sensors, but cannot access the actuators of the vehicle. Simulation can be used to evaluate the decisions of the automated driving function and thus determine the safety level. The same approach is used by car manufacturers, e. g. Tesla[1], to test new systems or new versions of existing systems.

[1] Tesla Autonomy Day: https://www.youtube.com/watch?v=Ucp0TTmvqOE at 2:55:43

A major drawback, however, is that the behavior of the potential conflict partner (other TPs) in the simulation does not correspond to reality, since other road users also plan and execute their actions based on the actions of the AV. If the passive driving function in a situation decides differently than the actual (active) driving function in the vehicle or the human driver respectively, then perhaps another TP would have decided differently and the results of the simulation have only limited validity.

Staged Introduction of AVs

The idea of the staged introduction is to reduce the ODD of the vehicle and thus the number of traffic situations that occur, so that a safety assessment based on real world tests can be carried out in an economically viable way. A severely restricted ODD is, for example, driving on a certain section of a road for a few hundred meters or a few kilometers and only when visibility conditions are good. In addition, a trained safety driver is part of the safety concept, who can intervene immediately if the system makes wrong decisions. If the vehicle is assessed as safe in this ODD, the ODD can be gradually expanded and/or the safety driver can be omitted. Many system manufacturers apply this procedure, especially in China and the USA. The most recent example is Daimler and Bosch [48], who are testing their systems in San Jose on a precisely specified road section.

This approach can be promising for the introduction of Level 4 vehicles, e. g. in a selected city center. In practice, however, it is not suitable for the validation and approval of Level 5 systems, because by definition these systems have an unlimited ODD.

Traffic-simulation-based Approach

The concept of traffic simulation is to simulate not only a single scenario, but a whole road network with hundreds of TPs (so-called agents). This method is therefore particularly suitable for making a macroscopic statement about the safety of AVs. Here, the effect of AVs on human TPs can also be modeled and investigated. In addition, it can be determined how the probability of occurrence of scenarios changes due to the introduction of AVs or what influence an increasing proportion of AVs has on overall traffic.

The approach based on traffic simulation can be used to increase the efficiency of the staged introduction of AVs, since the entire ODD can be simulated in traffic simulation. As is also the case for the staged introduction, this approach is no longer possible for Level 5 systems. Further information on this concept can be found in [49–51].

2.3 Scenario-based Approach

After briefly introducing seven different approaches of how to assess safety of AVs in Section 2.2, the SBA is discussed in detail in this section.

2.3.1 Overall Process

As already mentioned in Section 2.2, the main challenge of the SBA is to answer the question of how to find a manageable set of representative scenarios for a reliable safety assessment and certification of AVs. A large amount of literature in recent years has been dealing with exactly

this question. The available literature covers various aspects of the SBA, all of which can be classified into a generic framework (Figure 2.3). This work flow is derived in the author's previous publication [17] on the basis of existing literature and major research projects [10, 21]. In general, the assignment of references to one of the steps within the framework is ambiguous in some cases. For this reason, an attempt is made to identify the main aspect of the reference and categorize it accordingly.

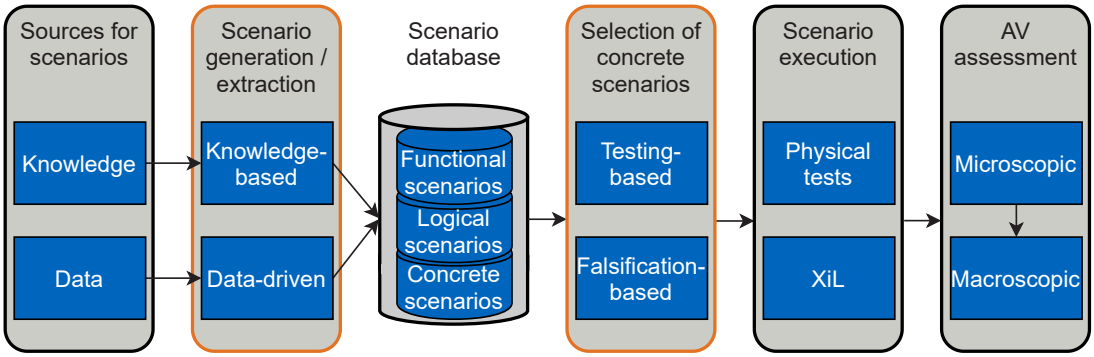


Figure 2.3: Generic framework of the SBA (adapted from [17]).

The central element of the SBA is the scenario database in which all scenarios are accumulated and stored. This does not necessarily have to be an actual database. Rather, the term is used in this thesis as a synonym for a storage container for scenarios. The framework can therefore be divided into a part to the left and a part to the right of the database. To the left of the database is the entire process for creating scenarios for the database. On the right side are all the procedures for the intelligent extraction of scenarios from the database as well as their execution and assessment. The individual blocks are described briefly below. The blocks of scenario generation and scenario selection highlighted in orange in Figure 2.3 are most relevant to the overall process. For this reason, these are discussed in more detail in the following.

2.3.2 Sources for Scenarios

In general, two types of information are available as sources for scenarios. On the one hand, this can be knowledge in various forms and on the other hand, data can be used as a source.

In the first case, the information can be in the form of abstract expert knowledge, standards and guidelines, such as the German guideline for the construction of motorways [52] and consumer tests, or in the form of accident data.

In the latter case, data from real traffic situations (e. g. field tests, etc.) are the source of information. A requirement for the generation of scenarios from driving data is a data set that is as representative as possible. Many organizations and enterprises have their own, non-accessible data sets, but in recent years, various institutions have made publicly accessible data sets available. An overview of available data sets is given in [53, 54]. Zhu et al. [55] also provide an overview of data sets and attempt to unify them.

Krajewski et al. [56] demonstrate a novel method for collecting real driving data. The traffic is logged with the use of a drone and the trajectories of the individual road users are then obtained from the images using computer vision. This method offers the major advantage that no costly and time-consuming test vehicles with extensive sensor technology need to be set up and that the traffic is not influenced by the recording. A drawback of the introduced procedure is the small

section of about 400 meters that can be captured by the drone, which is a limitation particularly in the acquisition of highway scenarios.

2.3.3 Scenario Generation and Extraction

In this paragraph, not all references are explained in detail, but rather the focus is on the description of the basic principles of different approaches. For a comprehensive explanation of individual references, the reader is referred to the author's previously published work [17].

Based on the two different sources for scenarios, a distinction is also made between two different approaches for generating and extracting scenarios. These are the knowledge-based as well as the data-driven scenario development. Both have in common that they subsequently store the scenarios identified in the database. However, the type of scenarios can differ. An overview of both methods can be seen in Figure 2.4.

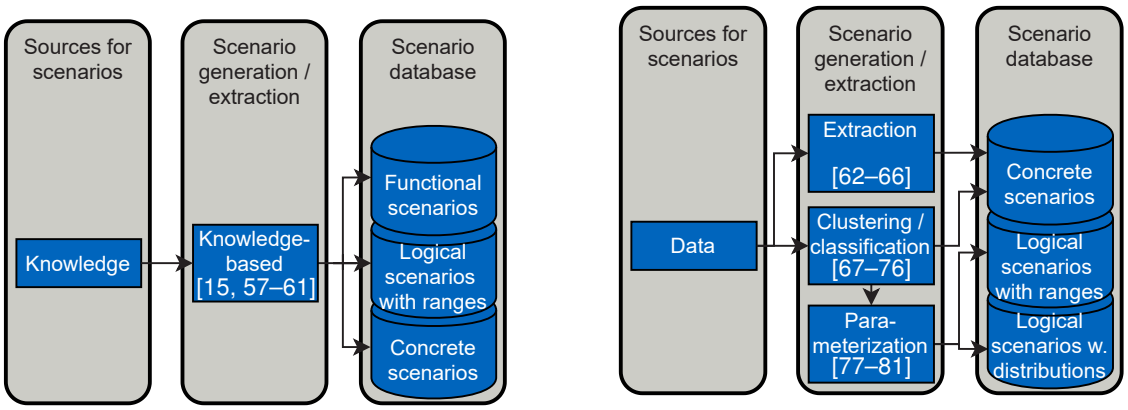


Figure 2.4: Knowledge-based and data-driven scenario generation and extraction methods including relevant literature (adapted from [17]).

Knowledge-based

The fundamental concept of knowledge-based scenario design is the structured transfer of knowledge into scenarios. In other words, abstract information is used to generate functional, logical (including parameter ranges) or directly concrete scenarios. Among other things, road traffic laws, regulations, accident data, consumer tests, ethics guidelines, safety analysis methods or expert knowledge can be used as knowledge base. Most often of the existing knowledge-based sources, expert knowledge is used to create scenarios. Within this method, ontologies are often used to store and structure expert knowledge [15, 57–61].

Data-driven

Three different approaches can be distinguished for the data-driven creation of scenarios (Figure 2.4b). These are extraction, clustering / classification as well as parameterization, whereby all of them typically use methods of machine learning or pattern recognition.

The basic idea of extraction methods is to extract concrete scenarios from measurement data without assigning them to a predefined group of logical scenarios. This approach is often used to find corner case scenarios. This can be done, for example, by finding a high level of uniqueness in the scenarios [66] or by triggering actions [62]. In the latter case, the trigger is initiated by a

particularly challenging prediction of the behavior of other road users. With the help of neural networks [63, 65] and Bayesian networks [64], this method can also be used to generate new concrete scenarios based on the measured data.

The common feature of scenario clustering and classification methods is that they can group individual scenarios from real data. The fundamental difference between the two approaches is that in clustering, the groups are not known in advance and are created during the process. In classification, on the other hand, the groups are defined in advance and the measurement data are assigned to the defined groups based on their properties. The former is usually done by means of unsupervised learning and the latter by supervised learning.

Clustering methods often use similarity measures [72, 73], hierarchical agglomerative clustering [75], or procedures based on Bayesian models [74]. In literature, there is a greater variety of approaches to scenario classification. For example, the predefined scenario groups can be based on the potential collision direction between the ego-vehicle and other TPs [76] or on the relative movement between these vehicles [70]. Furthermore, artificial intelligence in different variations (e. g. learned classification trees, neural networks and deep learning) is used to assign scenarios to defined groups [67–69, 71].

By definition (Section 2.1), logical scenarios are described by parameters and associated parameter ranges and/or parameter distributions. The groups of scenarios created by clustering and classification form the starting point for the description of logical scenarios by parameterization. This means that for each scenario group the parameters necessary for the description are determined from the measurement data. Then, for continuous parameters, their minimum and maximum values can be specified as parameter ranges. For discrete values, the individual discrete states can be determined. Based on this information, a parameter distribution can be specified for each parameter by examining the probability of occurrence of individual parameter values. Thereby, dependencies between the parameters must also be taken into account. One challenge is to obtain the most accurate conclusions about the parameter distributions that actually occur on the basis of a limited amount of real data [77, 78]. Further approaches of parameterization can be found in [79–81].

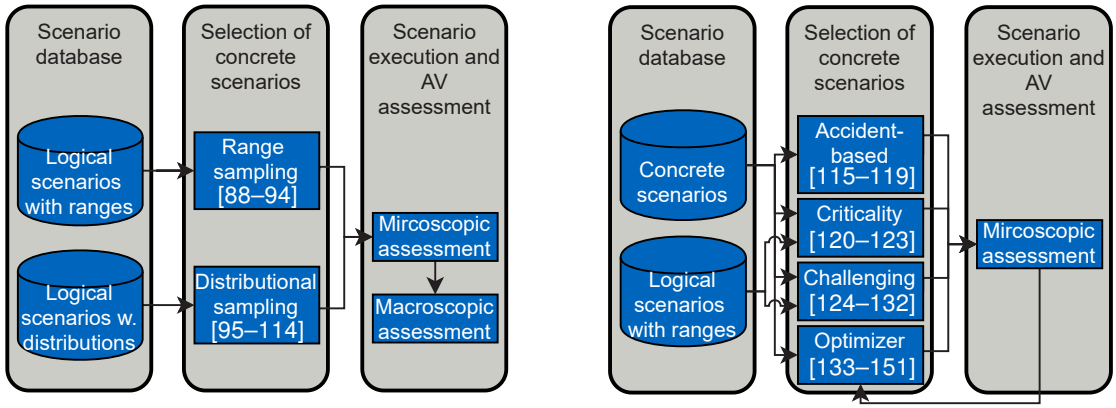
2.3.4 Scenario Database

For the SBA, a database with test scenarios is the core element. Due to the large number of scenarios and parameters, an efficient description and storage of the scenarios is indispensable. Within the PEGASUS project [10] a database with relevant scenarios for the ODD highway [82, 83] is established. The primary objective is to provide a standardized interface for reading different data sources and processing them into machine-readable formats. Another framework for developing a database, called Testing Scenario Library, is explained in [84, 85]. These authors also use the definitions for the different scenario types from the PEGASUS project, which are also introduced in this thesis in Section 2.1. Althoff et al. [86] present the Commonroad framework, which contains not only scenarios but also models and cost functions for the complete reproduction of the simulation-based assessment of trajectory planners. A further description of the procedure for building a scenario database can also be found in [87].

2.3.5 Selection of Concrete Scenarios

In this paragraph, not all references are explained in detail, but rather the focus is on the description of the basic principles of different approaches. For a comprehensive explanation of individual references, the reader is referred to the author's previously published work [17].

The decisive process step in the SBA for the safety assessment of AVs is the selection of concrete test scenarios from the database which should be performed for the safety validation. As already described, there is an infinite number of concrete scenarios due to the open parameter space. In order to make the safety assessment practicable, a manageable subset must be selected. Ideally, a maximal reliable proof of safety is performed with the smallest possible number of tests. As already introduced in Figure 2.3, the selection of concrete scenarios is divided into a testing-based and a falsification-based approach. For both approaches, there are again sub-methods, which are shown in Figure 2.5. The basic difference between testing and falsifying is that testing is used to verify that all requirements for the AV are met in a subset of the infinite number of scenarios. Falsification, on the other hand, specifically targets to find scenarios in which the AV does not fulfill the requirements.



(a) Testing-based (b) Falsification-based

Figure 2.5: Testing-based and Falsification-based scenario selection methods including relevant literature (adapted from [17]).

Testing-based

Testing-based approaches select concrete scenarios and evaluate them microscopically after their execution (physically or simulation-based). The evaluation is based on defined KPIs, which can be accident-related, criticality related or based on guidelines or standards. Within the testing-based approach, the selection of concrete parameter values of the logical scenarios can be based on samples of the parameter range specified by minimum and maximum values or on samples of the parameter distribution (Figure 2.5a). The former is particularly suitable for determining the parameter space coverage, or for achieving a good parameter space coverage respectively. However, a transfer to a macroscopic safety statement is only possible to a limited extent due to the missing parameter distributions. A much more reliable macroscopic safety statement, e. g. concerning the number of accidents that occurred, can be achieved by sampling the parameter distributions, which contain the probability of occurrence (exposure) of the concrete scenarios.

The simplest method for sampling from parameter ranges is the so-called N-wise sampling. In this process, all continuous parameters are discretized and then all possible parameter combinations of the discrete values are created. Due to the number of required parameters and an adequate discretization of continuous parameters, this procedure is only applicable for simple driver assistance systems [92]. For the safety assessment of AVs, more intelligent procedures that make the process more efficient are inevitable. Examples are design of experiments [89], regression tests [93], scenario importance [90, 94, 152] or rapidly exploring random trees [88, 153].

The standard procedure for sampling from parameter distributions is the Monte Carlo (MC) method, which uses parameter values with high exposure frequently and parameter values with low exposure correspondingly less frequently. Since the vast majority of frequently occurring concrete scenarios are not critical for AVs, using the MC method is inefficient. Therefore, the available literature attempts to accelerate this process. The standard performance criterion for measuring improvement is usually the comparison with the MC method.

A frequently applied approach [97–99, 108, 110, 112–114] is the use of the so-called Extreme Value Theory in combination with the Importance Sampling Theory, with which an accelerated safety assessment can be performed. The basic idea is to increasingly take parameter values from the outer limits of the parameter distribution and to execute and microscopically evaluate the selected scenarios. The systematic examination of the outer areas results in more critical scenarios. Subsequently, the biased microscopic results obtained with the adjusted distributions are recalculated back to the original parameter distributions and transformed into an unbiased macroscopic safety statement. According to Zhao [112], this allows the process to be accelerated by a factor of up to 10^5.

Falsification-based

The falsification-based approaches attempt to find concrete scenarios within the ODD as efficiently as possible where the performance of the AV does not meet the specified requirements. Either concrete or logical scenarios with the corresponding parameter ranges can be used as a basis for this. In literature there are four different approaches to find the best potential scenarios that can be used as counter-examples to satisfy the safety requirement (Figure 2.5b). These are described in more detail hereinafter.

Potentially suitable candidates for counterexamples are scenarios from an accident database. This procedure is established for the evaluation of ADAS systems. ADAS are permanently monitored by the driver and must therefore be safe, especially in cases where the performance of the human driver is insufficient, which in turn corresponds to the accidents from current accident databases. A transfer of this procedure to Level 3 and higher can be seen in [116, 118, 119]. For a more general statement, [115, 117] vary the parameters of existing accidents. Nevertheless, the use of current accident data is not representative because AVs must be safe not only in situations where the human driver has experienced difficulties, but in all of them due to the lack of monitoring. Moreover, current accident databases almost exclusively contain accidents involving human drivers, which do not necessarily have to be challenging and critical for AVs as well.

Another possibility is to focus particularly on critical scenarios. This can be done by microscopic evaluation of real data [122] or by using an exemplary concrete scenario whose criticality is increased by optimization [120, 121]. Here, however, there is always an influence of the behavior of the AV function under consideration. In addition, when using real data based on human drivers, the problem is that critical scenarios that occur do not necessarily have to be critical for AVs.

As a third possibility, an approach independent of the performance of the VUT is presented. The basic idea here is to take concrete or logical scenarios and make them more challenging for AVs. This is done by adjusting the parameter values for concrete scenarios or limiting the parameter ranges for logical scenarios. By increasing the difficulty, it is assumed that the occurrence of insufficient system performance during the execution of the scenarios is more likely. Note that most of the following references use the term complex scenario instead of challenging scenario.

However, in this thesis, complex is considered a subset of challenging (Section 2.1). The most important references are briefly summarized below.

Several references [124, 130, 131] use what they call complexity index to describe particularly challenging concrete scenarios. This index is based on the Analytic Hierarchy Process and assigns a certain difficulty to parameter values. It is similarly performed in [129], in which the description of the challenging scenario is divided into a static and a dynamic component. Still others [132] consider it particularly challenging when the cognitive abilities of the AV are severely challenged. In particular, the focus on the influence of the behavior of other road users is emphasized in [128]. An overview of factors that influence the difficulty of a scenario can be found in [53, 125].

The fourth method is the simulation-based falsification. The central element of this approach is, according to Figure 2.5b, the optimizer. It uses the microscopic evaluation results and takes them into account when selecting scenarios in the following iteration. Due to different types of optimizer, one or more scenarios can be examined in parallel per iteration, as is the case with genetic algorithms where it is called population size. The initialization of the optimizer at the beginning of the procedure is done by the required number of concrete scenarios including the corresponding microscopic evaluation results. Another important element of the optimizer is the cost function used, which contains an evaluation variable assessing the vehicle safety (e. g. the TTC [135, 147]). On the basis of this, the optimizer selects the concrete scenarios, which are then executed using virtual simulation and evaluated microscopically. These results form the basis for the optimizer to select the next concrete scenarios and a new iteration begins. By minimizing the cost function, the optimizer can determine concrete scenarios for the VUT with each iteration that are closer to a violation of the requirements. Due to the large number of concrete scenarios to be executed, this approach can only be applied reasonably in a virtual simulation environment.

In literature there are many different approaches to the design of the optimizer, but the basic principle described in the previous paragraph does not change. In the following, the optimization approaches available in literature are listed:

- Reinforcement learning [136, 139, 141, 142]
- Differential evolution genetic optimization [134, 135]
- Particle swarm optimization [134, 135]
- Adaptive search-algorithms [143–145]
- Bayesian optimization [138]
- Random forest models [146]
- Simulated annealing [133, 147, 149–151]
- Gradient descent optimization [148]
- Forward / backward search [140]

The references differ not only in the type of the optimizer, but also in the type of use case applied. However, with all the existing references, the use case is only a simplified use case for the proof of concept. This can be, for example, the use of simplified driving functions [135] or simple scenarios such as the following drive scenario [140].

2.3.6 Scenario Execution

For the execution of the selected concrete scenarios, various test environments are available. Concrete scenarios can be carried out either in the real world via field or test site tests or with increasing degrees of virtualization via X-in-the-Loop (XiL) simulation [154]. Since simulation has many advantages in terms of costs, effort and safety risks, almost all references use it for their proof of concept. There are many commercial and free simulation tools on the market. Additionally, there are different frameworks available in literature [155, 156].

2.3.7 AV Assessment

Based on the definitions in Section 2.1, we differentiate between microscopic and macroscopic safety assessment in Figure 2.3. It is possible to determine the level of safety within a microscopic evaluation, using KPIs. It is advantageous to use criticality metrics as KPIs because accidents are rare events. The best known of these is the TTC [157], which also exists in many variations, e. g. in [158–160]. An overview of existing criticality metrics can be found in [161]. Hallerbach et al. [155] also distinguish four areas in the environment of the AV that are of interest for the evaluation of criticality. The number of critical situations and accidents that have been encountered in the microscopic assessment can be the basis for the transition to a macroscopic assessment. A more detailed investigation of the transition from many microscopic to one macroscopic assessment can be found in [20].

In the SBA, both, the testing- and the falsification-based scenario selection methods, microscopically assess the safety for each scenario. The testing-based methods focus more on covering the scenario space, while the falsification-based methods concentrate more on finding corner case scenarios. Although the selection of corner cases is efficient for finding counterexamples, it is less suitable for a macroscopic evaluation than the testing-based approach. Especially the distributional sampling approaches include a more comprehensive representation of real traffic behavior. They are therefore better at transforming the microscopic results into a statistical macroscopic statement using the parameter distributions / exposure of microscopically evaluated scenarios.

2.3.8 Conclusion

The generic framework for the scenario-based safety assessment of AVs from Figure 2.3 contains all high-level process steps within the approach. The central element is the database in which all scenarios are stored. The most important steps are scenario generation and the selection of the concrete scenarios to be used for the safety assessment. The former generates the scenarios (Subsection 2.3.3) and stores them in the database. The latter (Subsection 2.3.5) is the central aspect of the SBA. For both steps, a large number of references exist that deal with these topics, but without having shown a comprehensive applicability in practice. After selecting the concrete scenarios, they have to be executed. Due to the high number of concrete test cases, as many scenarios as possible have to be executed in simulation for both, time and financial reasons. Thereby, a big challenge still to be solved is the validation of the simulation models [154]. In the last step of the procedure, the microscopically evaluated scenarios still have to be aggregated to a macroscopic safety statement, which has not yet been conclusively clarified [20].

In summary, it can be concluded that there are promising papers for all individual steps of the SBA. However, these are limited to simple use cases and proof of concepts. Thus, one of the biggest challenges in future will be the large-scale implementation of existing methods.

2.4 Type Approval of Vehicles

This section explains the vehicle type approval process in more detail. First, the general process as applied to current series production vehicles is described (Subsection 2.4.1). Due to the increasing functional scope of automation and the associated imminent market introduction of Level 3+ systems, new requirements for type approval arise, which are explained in Subsection 2.4.2. Finally, the current status of the development of new standards and regulations for Level 3+ systems is shown in Subsection 2.4.3.

2.4.1 Process

After the development of vehicles and before they are introduced to the market, they must receive a type approval (also called certification or homologation) that confirms the absence of unreasonable risks. The type approval is thus at the end of the classic V-model and vehicles may not be sold without passing this certification. The basic procedure here is that the legislator specifies requirements that must be verified by defined tests.

In principle, type approval can be carried out in two different ways. These are self-certification and certification by an independent third party (test organization / technical service). The fundamental difference between the two procedures is by whom the required tests are carried out. In the case of self-certification, the manufacturer himself is responsible for carrying out the tests properly and checking that all requirements are met. In the case of certification by an independent test organization, this organization takes the manufacturer's vehicle, carries out the tests and checks whether all the requirements of the legislator are met. The former procedure is used in the USA, for example, and the latter in Europe. Since this thesis focuses specifically on type approval by an independent test organization, only this application will be discussed in the following.

The legal framework for the certification of vehicles in Europe is provided by the UNECE. Within the UNECE there are various working parties in the field of transport. Working Party 29 (WP.29) [162], called the World Forum for Harmonization of Vehicle Regulations, is most important for this work. In addition to this, there are also working groups within the so-called 'Global Forum for Road Traffic Safety (WP.1)' which deal with topics related to automated driving. Within WP.29, general regulations for type approval are developed and revised. These are addressed in various sub-working groups, arranged according to different subject areas (Figure 2.6). As of August 2020, there are 152 regulations that must be fulfilled before a vehicle can be released for sale. All these are part of the so-called 1958 Agreement [163], which 54 countries [164] have signed, mainly but not limited to Europe. In addition, there are 20 so-called UN Global Technical Regulations from the 1998 Agreement, which only 38 countries have signed, but among them, in addition to numerous large European countries, the USA and China.

As described in the previous paragraph, there have been standardized regulations for the certification of vehicles since 1958. At the beginning, all regulations were strongly focused on testing the hardware (e. g. brakes), because the proportion of software in the vehicle was either non-existent or low. All hardware-related tests are specified in fixed test specifications and can be performed reproducibly under standardized boundary conditions (e. g. temperature). With the increasing use of software in recent years, the demands on software testing have also increased. This aspect is discussed in detail in Subsection 2.4.2.

Since the introduction of the first regulations, these have not been in support of a comprehensive proof of safety, but rather confirm a minimum level of safety by spot-checks using predefined

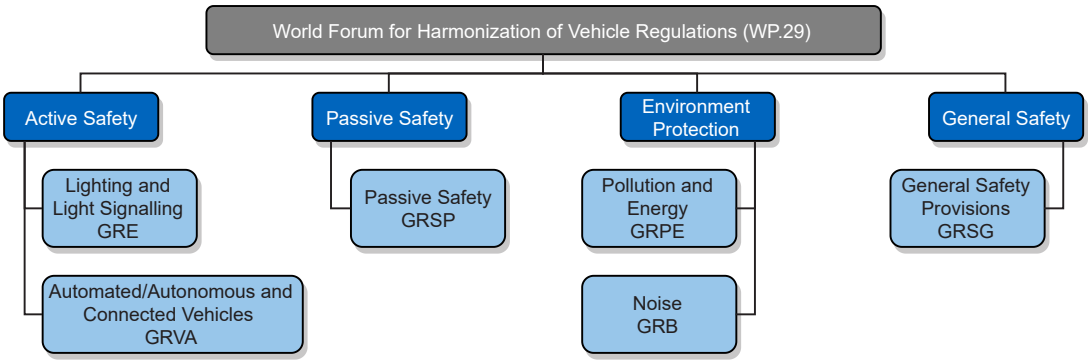


Figure 2.6: Structure of WP.29 as a part of the UNECE. Of particular interest for this work is the Sub-Working Group on Automated/Autonomous and Connected Vehicles (GRVA), which is introduced in more detail in Subsection 2.4.3.

tests. A comprehensive safety assessment, as the manufacturer must carry out for product liability, is therefore not the focus of the type approval process.

2.4.2 Requirements for AVs

As already introduced in Subsection 2.4.1, the origin of the UNECE regulations is hardware certification. As the digitalization of vehicles continues to progress, more and more software is being installed in vehicles. This software can be used in entertainment systems (no requirements for type approval), or can take over parts of the driving task. The latter has begun with the introduction of driver assistance systems, for which there are also regulations for their certification available. Examples are UNECE R130 for lane departure warning systems and UNECE R131 for advanced emergency braking systems.

There are therefore already regulations in place which include testing of software. However, these support the driver in the execution of the driving task, but do not take any responsibility. The driver must therefore permanently monitor the system and immediately override the system in the event of any errors. Due to this distribution of responsibility, it is sufficient in the certification of driver assistance systems to transfer the well-established procedure from the type approval of hardware. This means that it is sufficient to verify the basic functionality in predefined tests.

Compared to ADAS, there are two fundamental changes which result in two new requirements for AVs. These are:

1. Systems from Level 3 upwards take over the responsibility for the driving task. This leads to a higher number of required tests.
2. With predefined tests, there is a risk that manufacturers will prepare themselves specifically for these tests. A transfer of the behavior of the system in real road traffic is only conditionally given.

This two new requirements are explained in more detail in the following.

Higher test amount

By eliminating the monitoring of the system by the driver, any error in the driving function from Level 3 onwards can result in considerable risks and even fatal accidents. Even if the type approval does not have to fulfill the purpose of a comprehensive safety assessment, as it is necessary in the context of product safety, the scope of testing must be significantly increased

compared to existing regulations. Furthermore, a change from current simple tests to prove the basic function through function-based testing (Section 2.2) to more challenging, realistic tests through scenario-based testing must be implemented.

This is necessary because society has high demands on the safety of AVs, as demonstrated by the great media interest in the Uber accident in Arizona [165]. Reasons for this include, on the one hand, the fact that this is a new technology that is not yet established and, on the other hand, the fact that uninvolved persons can be injured by an error of a computer. The risk that society will accept when AVs are introduced has not yet been conclusively clarified [19]. Junietz et al. [19] assume that AVs must have a significantly lower accident rate than human drivers in order to be accepted by society. This also increases the pressure on legislators to test AVs more intensively than driver assistance systems before they are introduced to the market. Consequently, the scope of testing for type approval of L3+ vehicles will increase significantly, although the focus will remain on spot checks.

No predefined tests

The concern with pre-defined tests, which are currently the standard within the UNECE regulations, is that manufacturers may optimize the performance of the system to meet the requirements of these tests. As a result, the system behavior in real operation deviates from the behavior documented in the tests, respectively the system only achieves a lower performance in real operation. The reason for the precise definition of the tests as well as test conditions in current UNECE regulations is the reproducibility and comparability of the results, which are thereby ensured. A performance optimization by the manufacturer is therefore possible with the existing UNECE regulations, but due to the permanent monitoring of the systems by the driver, the impact is small. For the first time, problems with predefined tests in the type approval of vehicles have arisen with the discovery of the problems in emission testing. More detailed information on this topic can be found in our already published work [126, p. 6]. In summary, several manufacturers used software that detects when the vehicle is in the test cycle and then reduces emissions. As a result, new type approval guidelines have been drafted by UNECE which have a lower reproducibility and comparability, but which reflect the system behavior under real driving conditions [166].

Due to the high safety requirements that AVs have to meet, it is of great importance that the test results of the type approval also correspond to the performance of the vehicle in real road traffic. This is the only way to ensure that the requirements are met and that a socially accepted introduction of AVs is possible. The fact that standardized tests are not useful for type approval of AVs also reflects the opinion of the European Commission [167, Slide 5] and NHTSA [8, p. 77].

2.4.3 Development of New Regulations

This section briefly outlines the process of developing new regulations and then gives an overview of the future UNECE regulation for a Automated Lane Keeping System (ALKS) [11], the first regulation, for Level 3 systems.

Development process

The development process of new regulations is based on a bottom up principle. For this purpose, the relevant subgroup of the WP. 29, the Working Party on Automated/Autonomous

and Connected Vehicles (GRVA) [168] (Figure 2.6), is divided into further working groups. The working groups include participants from legislators, from research and from industry (both manufacturers and testing organizations). Among others, there are the working groups Validation Methods for Automated Driving (VMAD) and Automatically Commanded Steering Function (ACSF). The former deals more with the overall process of approval of AVs and has proposed that approval is a multi-pillar process consisting of virtual simulations, proving ground tests, field tests and an audit [169, 170]. The ACSF working group is involved in the concrete elaboration of regulations for AVs and developed a first draft for the regulations for the approval of ALKS systems. If all parties involved agree, this draft will be passed on step by step via the GRVA to WP.29. Currently (as of August 2020), WP.29 has already approved the draft [171] and the regulation [11] is expected to come into force in January 2021 and will be called R157 [172]. The following section provides a brief overview of the content of this future regulation.

Future ALKS regulation

The content of the future regulation [11] is of great importance for the present work, because the scope of the method to be developed is the certification of AVs and the regulation is the legal framework to be complied with. By definition, Level 3 systems have a restricted ODD, which is why the ODD is also clearly restricted in the future regulation. The regulation covers ALKS systems which operate up to a speed of $60\,\mathrm{km/h}$ on motorways with spatial separation of the two directional lanes (no pedestrians, cyclists and crossing traffic) and thereby take responsibility for the driving task during this time. By definition, the driver must take control again when the system requests it. The system does not change lanes, but can initiate evasive maneuvers in emergency situations. In addition to OEDR, the regulation also contains requirements and specifications for other areas (e. g. human-machine interface and safe takeover of the driving task). In the following, only the content of the regulation regarding OEDR will be highlighted, because it is the focus of this thesis.

In the following, the most important contents of the future regulation for the present work are summarized:

Technical service shall conduct spot checks: The testing organization should use targeted spot checks to verify that the requirements for the system are met and that the system has the functional capabilities specified by the manufacturer (Annex 4 Paragraph 1 and 4.1).

Technical service shall conduct tests that are critical for OEDR: In particular, tests that are critical / challenging for OEDR should be conducted (Annex 4 Paragraph 4.1.2.1). It is also mentioned that these so-called traffic critical scenarios can be generated using the parameters of road geometry and the behavior of surrounding TPs (Annex 4 Appendix 3 Paragraph 2).

Draft provides functional scenarios: The functional scenarios lane keeping, blocked lane, cut-in, cut-out and decelerating leading vehicle are defined (Annex 5 Paragraph 4).

Draft provides minimum set of parameters: Parameters are defined for the functional scenarios, whereby the number of parameters is not complete and only in rare cases a range of the parameters is specified. According to the definition in Section 2.1, this does not correspond exactly to logical scenarios, but because there is no standardized term for this form, these will be referred to as logical scenarios in the following.

Draft does not provide concrete scenarios: Throughout the document, no concrete scenarios are given and no specific statement is made about the number of scenarios required. There is also no discretization step size for the specified parameters of the logical scenarios provided.

Technical Service may define additional parameters: Essential parameters are defined, including road geometry, environmental conditions and position, speed and acceleration of the challenger. It is also emphasized that other parameters such as road curvature and lighting conditions can be added (Annex 4 Appendix 3 Paragraph 4).

Technical Service may define additional scenarios: The defined scenarios represent a minimum of tests and the technical service can define any test scenario within the defined system boundaries (Annex 5 Paragraph 1). In addition, areas in the parameter space can be identified by the technical service during the tests, which make further investigation through tests appropriate (Annex 5 Paragraph 5.4).

Specific test parameters shall be selected by the technical service: The technical service is responsible for the selection of parameter combinations (Annex 5 Paragraph 1) and can specify any combination as a test that it deems appropriate (Annex 5 Paragraph 3.2).

The manufacturer shall provide a documentation of the system: Even if the technical service does not have the software available as a white-box model in order to protect the manufacturer's intellectual property, the manufacturer still has to provide a documentation of the system. This includes, for example, the system design, control strategy, system layout, sensor setup and system boundaries (Annex 4 Paragraph 3 and Annex 4 Appendix 2).

Tests can be executed in virtual simulation: Simulation tools can be used especially for scenarios that are difficult to execute in real experiments, if they have been validated by physical tests (Annex 4 Paragraph 4.2).

Skilled human driver is performance reference: The microscopic evaluation of the scenarios is mainly carried out via the distinction accident / no accident. The reference for whether ALKS must be able to prevent an accident in a scenario is created using a model of an attentive human driver (Annex 4 Appendix 3 Paragraph 1).

In summary, the formulation of the regulation is not precise in many points and leaves the exact interpretation and application to the responsibility of the technical service in cooperation with the manufacturer. In addition, the technical service has a large degree of freedom in the design of the test scenarios in terms of their characteristics and scope. An efficient and comprehensible procedure for technical services for the homologation of AVs is therefore urgently needed.

2.5 Derivation of the Research Question

For an efficient certification of AVs performed by a technical service, methods from the safety assessment of AVs must be combined, integrated and adapted to the requirements of the given framework by the UNECE (i. e. upcoming Regulation [11]) (Figure 2.7a). This combination is not addressed in the current state of the art and a new methodology tailored to the individual

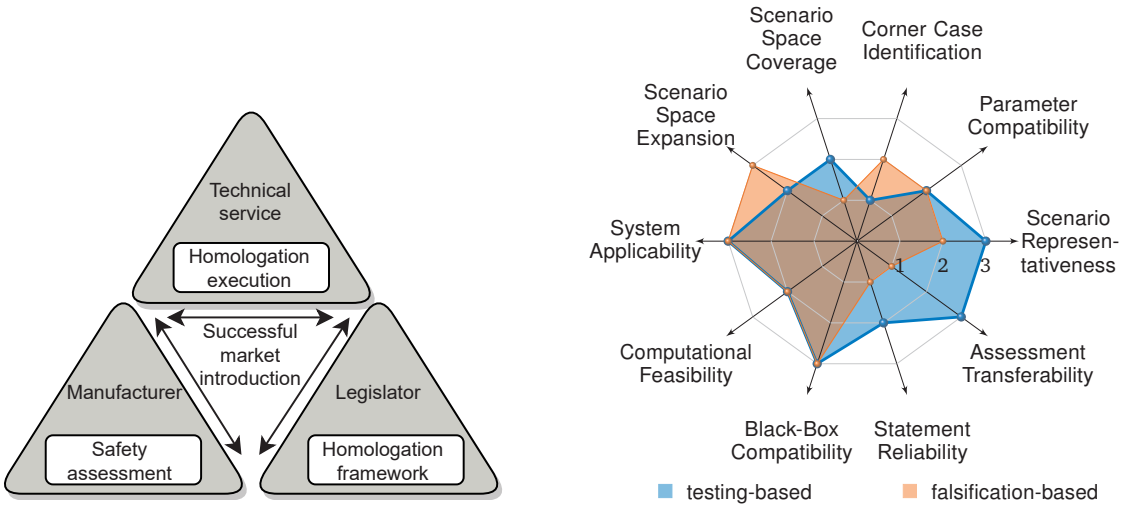
demands of a technical service is required. In the following, the interaction of the safety assessment and the given framework will be explained in more detail and the requirements from the viewpoint of a technical service will be derived.

The safety assessment of AVs by the manufacturer forms the methodical basis of the new method to be developed. Section 2.2 shows that the safety assessment of AVs is a problem that has not yet been solved and that different approaches exist for this purpose [17]. The SBA, which is explained in more detail in Section 2.3, is considered the most promising approach. The SBA (Figure 2.3) has various approaches for the individual steps, but it is not yet clear, how all relevant scenarios can be identified and selected.

Combining the approach of the SBA with the new UNECE framework (future regulation for ALKS from Section 2.4.2), the most important steps of the SBA for the technical service in the context of AV certification can be identified. If we look at the process of the SBA in Figure 2.3 and map the requirements of the future ALKS regulation to it, the steps left to the database can be assumed to be given by the specification of the logical scenarios. The final assessment of the AV can also be taken as given by the UNECE framework, because the human driver is to be used as a reference and further KPIs are also specified. According to the UNECE, the scenarios can be performed both, physically and virtually. For reasons of time and cost, virtual testing is applied as much as possible. For this purpose, the simulation models used must be valid. This is a large field of research and has not yet been conclusively investigated. Nevertheless, this aspect will not be considered in the present work. Therefore, the most important step is the selection of concrete scenarios. On the one hand, this aspect is still an open issue in AV safety assessment and, on the other hand, the UNECE does not make any precise specifications on this point and the technical service is commissioned to do so.

In the author's previous publication [17], the strengths and weaknesses of the two different scenario selection methods (testing-based and falsification-based) are discussed in detail. The resulting summary is shown in Figure 2.7b. An explanation of the individual categories can also be found in [17]. Mapping the requirements from the UNECE Framework (Section 2.4.2), it can be concluded that falsification has a more important role for the technical service. The main reason for this is the better rating in the identification of corner cases, which corresponds to the UNECE requirements for critical spot checks (Section 2.4.2).

Thus, according to Figure 2.5b, the four methods accident database, criticality, challenging and optimizer remain. Currently, there are no comprehensive accident databases of AVs, therefore this method is not applicable, but may become interesting for technical services after the market introduction of AVs. According to the definitions in Section 2.1, the assessment of whether a scenario is critical or not can only be made after the scenario has been carried out. As the technical service focuses on the definition phase of the scenarios, this approach is not applicable. To perform falsification using an optimizer, all simulation models must be available. However, the technical service only has documentation of the system available for homologation (Section 2.4.3), which means that this approach cannot be used either. What remains is the selection / definition of particularly challenging scenarios. This also makes sense for a technical service, because documentation of the system is sufficient, it can be defined before the execution of the scenarios and there is a higher probability that the challenging scenarios become critical during the execution. This then also fulfills the requirement for the selection of critical spot checks. However, none of the approaches available in the literature discusses homologation and the specific requirements from the point of view of a technical service. Therefore a new approach is needed here.



(a) Homologation (b) Comparison

Figure 2.7: a) Homologation as prerequisite for a successful market introduction of AVs, where three stakeholders are involved and a combination as well as integration of various requirements of the stakeholders is necessary. b) Comparison of scenario selection methods (adapted from [17]).

Even though more tests should be conducted during AV homologation (Section 2.4.2), the process must be as efficient as possible and should not be predictable / predefined (Section 2.4.2). In addition, the technical service has to test a wide variety of manufacturers with different system characteristics and features, which is why a system-specific selection of challenging scenarios is advisable. To ensure that the procedure is as efficient as possible, the most challenging scenarios should be identified, which according to the definitions (Section 2.1) are called corner cases. Recently, Batsch et al. [173, p. 16] also concluded that there is still a research gap to include knowledge about the system to identify failure regions in the parameter space to efficiently define corner cases.

From this, the research question of the present thesis can be derived:

How should a procedure for technical services be designed that efficiently identifies system-specific corner cases within the homologation of automated vehicles?

Based on Figure 2.7a, we take falsification as a method of safety assessment from the lower left triangle as a basis and adapt it to the homologation framework so that an appropriate procedure for technical services can be developed for the homologation of AVs. The following chapter describes in detail how this development is carried out in practice.

3 Overall Methodology Development Process

This chapter first refines and specifies the requirements and then provides an overview of the newly developed approach. Finally, it demonstrates how the requirements of the first part of the chapter are met.

3.1 Detailed Requirements

Section 2.4 introduces general requirements for AV homologation and presents the relevant content of the future UNECE framework [11]. From this, detailed requirements are derived in the following, which the methodology to be developed has to fulfill.

R1) On the one hand, the logical scenarios defined by the UNECE should be used as input for the method, on the other hand, the documentation of the system, which the manufacturer has to provide, can be used. According to section 2.4.3, this consists, among other things, of the system structure including all sensors.

R2) As output, the method should provide (semi) concrete scenarios for the spot checks during homologation. Semi-concrete scenarios are sufficient, because they allow to set priorities in the tests and in order to pre-define a larger subset of scenarios in an uncertain case. This conservative selection is intended to prevent scenarios that are relevant accidentally being excluded.

R3) After certification, the market launch of the vehicles can be started. The method must therefore be able to address the overall vehicle.

R4) The definition of the most important parameters of the future UNECE regulation must be covered. These are according to Section 2.4.3 the road geometry, environmental conditions and position, speed and acceleration of the challengers. This corresponds to the layers L1, L4 and L5 according to the five-layer model of Bagschik et al. [15].

R5) To ensure that the results also represent the vehicle behavior in real traffic situations, not only standardized and predefined tests should be used (Section 2.4.2).

R6) The technical services test a large number of different system versions during homologation. Therefore, the method must allow an efficient selection of corner cases for different system versions of different manufacturers.

R7) According to Section 2.4.3, the technical service can define any other scenarios it deems appropriate. However, these should be within the system specification. This

can be taken into account by the documentation provided. Scenarios outside the ODD are therefore not in focus.

These seven requirements are the most important from the technical service point of view and have to be taken into account during the development. How this is achieved is presented in remainder of this chapter.

3.2 Overview of the Developed Methodology

The logical scenarios defined by the upcoming UNECE regulation [11] are the starting point for the approach. The parameters of these logical scenarios should be specified or at least restricted, so that test cases are defined which are as meaningful and challenging as possible. In other words, as few scenarios as possible should be selected which have a high informative impact in terms of system performance. This means that (semi) concrete corner case scenarios will be obtained as output. A visualization of the procedure is shown in Figure 2.3. The overall method thus consists of three sub-methods and a subsequent combination of the parameter restrictions. Next, an overview of the three sub-methods is given.

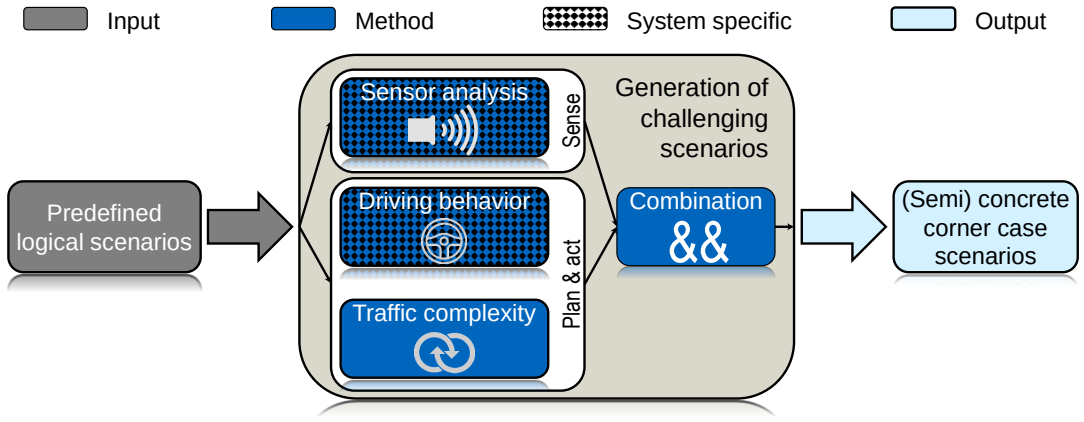


Figure 3.1: Overview of the developed methodology. Every aspect of a single block is challenging and according to the definitions in Section 2.1 these resulting scenarios are called a corner cases.

In all three sub-methods, the aim is to define challenging scenarios. The basic assumption here is analogue to the literature in Subsection 2.3.5 that challenging scenarios are more likely to lead to critical situations, and are therefore especially informative. This is illustrated in Figure 3.2 using a logical scenario with two parameters. In the unlimited parameter space, areas that are particularly challenging for the VUT are identified during scenario definition. It is assumed that critical situations will occur more often during the subsequent execution of test cases, turning the corresponding challenging scenarios into critical scenarios. Due to different system characteristics and implementations, the challenging scenarios are not identical for all systems, as shown in Figure 3.2 using two different systems as an example. Next, the three sub-methods are introduced shortly.

Sensor analysis

For the sake of simplicity, all sensors used for environmental perception are referred to as sensors in this thesis. Additionally, the term sensor analysis covers not only the sensors itself but

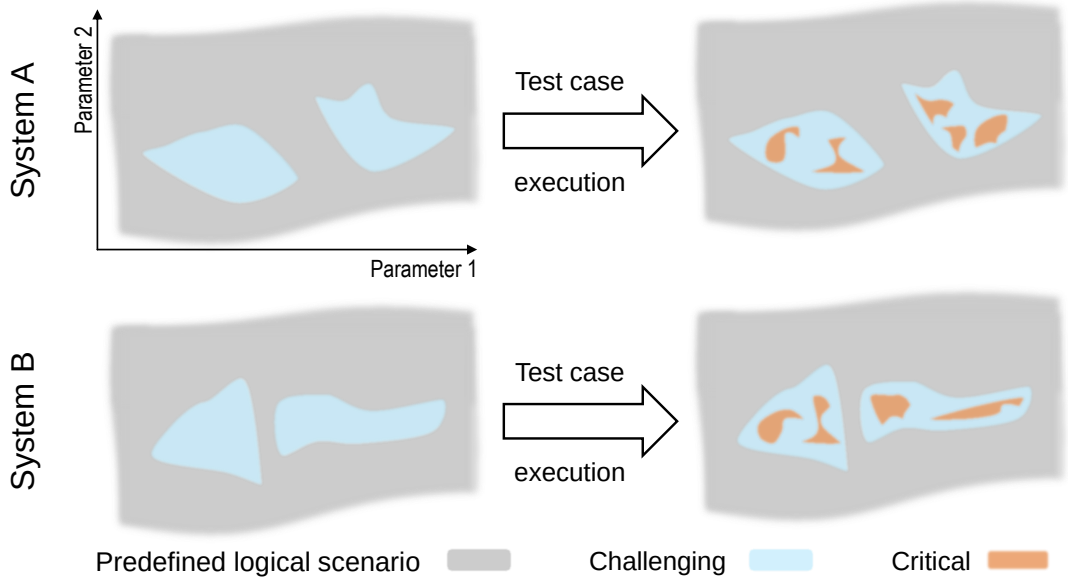


Figure 3.2: Simplified representation of a logical scenario with two parameters. The assumption is used that challenging scenarios lead more often to critical situations. The identification of challenging scenarios is individual for each AV. Due to bugs, critical situations can of course also occur in simple scenarios. This is not shown in this figure and will not be considered in the remainder of this thesis.

also the processing of the sensor data. AVs use different sensors like Lidar, Radar, ultrasonic and cameras for this task. These sensors have the responsibility to perceive the environment accurately. Via the involved data processing, all objects relevant for the driving task have to be detected and tracked. Only if this is ensured, a safe trajectory can be planned by the subsequent planning module.

In principle, each individual sensor type has certain strengths and weaknesses, which is why a combination of different sensor types is usually used and redundancy is established. More sensors increase the safety, but also the costs and therefore the manufacturers make a compromise here. As a result, each manufacturer has an individual sensor setup, which can be influenced by the cost as well as the package and the ODD of the vehicle. The goal of this sub-method is the structured analysis of the sensor setup as well as a derivation of weaknesses, which are then taken into account when defining the test scenarios. According to R1), this analysis has to work with the documentation of the system as input, which in this case means that the technical service has access to information about the sensors used, such as type and characteristics, opening angle, range, mounting position and orientation.

Based on this information and so-called phenomenological sensor models, a sensor coverage and a detection probability in three-dimensional space are calculated. This can be used to analyze the following aspects related to the sensor setup in use:

- Investigation of the immediate area around the vehicle for blind spots (independent of weather conditions, because the influence is negligible at short distances)
- Far-field analysis under different weather conditions to identify unfavorable environmental conditions.
- Optimization of an object's approaching path so that the object is seen as poorly as possible by the AV.

- Transforming the optimal approaching path into a road geometry for scenario execution.

The results of this sub-method are scenarios that are particularly challenging for the perception module and are derived individually based on the VUT's characteristics. A detailed description of the implementation of this sub-method is given in Section 4.1.

Driving behavior

In this sub-method, scenarios are identified that are particularly challenging when considering the driving behavior of the AV. In order to have information about the driving behavior of the AV, a small number of tests must be performed. This approach is therefore not completely applicable in advance, but is carried out during the certification process. The basic idea of this sub-method is that future test scenarios take the driving behavior of the AV into account (such as particularly defensive behavior, overtaking behavior, curve driving behavior, etc.) and that the scenarios are adapted in such a way that they are particularly challenging for the driving behavior of the AV under test, respectively are more likely to lead to critical situations.

For example, it can be investigated whether the vehicle tends to oscillate to the inside or outside of the curve during cornering. If it is determined that an AV tends to cut the curve, then this knowledge can be used for example to place (stationary) objects on the inside of the curve in future scenarios. If a vehicle behaves in exactly the opposite direction and drifts in the curve towards the outside of the curve, then objects can be placed accordingly on the outside of the curve. This leads to smaller distances to the objects, which increases criticality. This is not a guarantee that a critical situation will occur, but at least the AV must deviate from its usual control strategy, which is an increased challenge.

If no significant driving behavior can be identified for a vehicle, then this sub-method cannot be used to adapt future scenarios. Compared to the other two sub-methods, this sub-method has the smallest extent and is introduced in Section 4.2.

Traffic complexity

This sub-method (Section 4.3) examines the complexity of traffic situations. According to the definitions in Section 2.1, complex scenarios are a subcategory of challenging scenarios and address the behavior of surrounding TPs. The question that arises is: How should the surrounding TPs move so that the scenario is particularly complex respectively challenging for the AV? Hence the trajectories of the surrounding TPs are to be determined. Parameters like trajectories are particularly difficult to define because they are time-dependent and therefore not suitable for simple discretization, such as the scalar curve radius.

The core of this sub-method is the development of a metric to measure the complexity of traffic situations, and ultimately to derive complex scenarios from them. For this purpose, real data (highD data set [56]) are used to have realistic baseline data. These data of several minutes are clustered into short individual scenarios and these scenarios are grouped into different functional scenarios. In this way, the functional scenarios that are specified for certification by the UNECE can be considered in particular. In a next step, a newly developed complexity metric is applied to the scenarios and thus their complexity is determined. The metric consists of 13 different factors that are used to evaluate the scenario. For simplicity, each of these factors, such as the number of participating TPs, contributes linearly to the overall complexity. The factors are individually weighted according to their importance. Subsequently, the most complex scenarios can be

further used and in a next optional optimization step. An optimizer can be used to optimize the behavior of surrounding TPs in the most complex scenarios, thus further increasing the complexity. The scenarios with the highest complexity represent the greatest challenge for the AV and can be used for certification.

3.3 Fulfillment of Requirements

The following describes how the requirements (Section 3.1) are taken into account in the developed method (Section 3.2 and Figure 3.1).

Addressing R1) This requirement is strictly predetermined and does not represent an active implementation. According to Section 3.2, this requirement has been met by the methodology developed. For the sensor analysis sub-method, the system specification including the description of the sensors is required, which the manufacturer must provide to the technical service. Furthermore, for all three sub-methods the ODD of the vehicle, which the manufacturer must also provide, and the predefined logical scenarios of the UNECE are sufficient.

Addressing R2) This requirement is the result of the method and is taken into account by the sub-methods. Each of the three sub-methods has the objective of identifying challenging ranges for certain parameters. Thus, the output of a single sub-method can be regarded as challenging semi-concrete scenarios. By then simply combining these parameter ranges, according to the definitions in Section 2.1, the challenging scenarios become corner case scenarios.

Addressing R3) In order to be able to evaluate the entire vehicle, the AV is first divided into the sense, plan and act divisions, which are widely known from robotics (Figure 3.1 middle). At least one sub-method is defined for each of the three partial aspects. To further simplify the process, plan and act are considered in combination, because these two aspects strongly influence each other. Especially if the system is observed from the external perspective of a technical service who does not have the insight into the exact implementation.

Addressing R4) The most important parameters to be addressed by the method are L1, L4 and L5. The sensor analysis mainly addresses challenging environmental conditions (L5) for the system under test. In addition, the definition of the challenging road geometry for the sensor setup used also examines L1, i.e. the road geometry. The investigation of the driving behavior includes aspects of the road geometry (e. g. curve radii) as well as objects (L4). In the latter case, for example, the positioning of static objects is examined. The analysis of traffic complexity defines particularly challenging trajectories of surrounding TPs, which corresponds to the dynamic objects. Thus, the overall method includes both, static and dynamic objects in L4.

Addressing R5) The analysis of the sensor setup as well as the driving behavior are individually adapted to the VUT. Thus, for each vehicle individual test scenarios, which are not known to the manufacturer before, are derived and can be tested during homologation. By using real data for the extraction of particularly complex scenarios, the pre-defined logical scenarios, where only the challenger is defined, turn into realistic scenarios that can occur in this form on German motorways.

Addressing R6) This is also, where the system-specific consideration in the first two sub-methods comes into effect. In addition, the description of the ODD of the vehicle can be included in all three sub-methods. For example, the particularly challenging trajectories of traffic complexity can be filtered according to the velocities involved so that the ODD specified by the manufacturer is complied with.

Addressing R7) This requirement is met, as mentioned above, by the inclusion of the manufacturer‘s specified ODD in all three sub-methods. The more precisely the manufacturer specifies the ODD of the vehicle, the more specifically the scenarios can be selected.

According to Chapter 2, the scenario selection for the safety demonstration is still an unsolved problem. It is possible to consider any number of scenarios and any number of parameters, but especially for homologation, where only a limited scope of testing can be performed, only the most important parameters can be considered. This is taken into account by the overall method described in Section 3.2. Especially the reduction of continuous parameters to the most challenging regions, as it is done in the identification of the most challenging trajectories of the surrounding TPs based on traffic complexity, is valuable because these parameters are not well suited for a standard discretization.

In the further progress of the work, the development of the sub-methods is the main focus. The most comprehensive description is of the traffic situation complexity (Section 4.3), because the first two sub-methods in Section 4.1 and 4.2 were already published in previous works of the author. The respective implementation of the sub-methods will be prototypical / exemplary. The subsequent block of combination is a simple combination of the challenging scenarios or parameter ranges without considering a mutual influence of the sub-methods and is only considered theoretically.

4 Sub-Methods

This chapter represents the core of the work and describes the development and results of the three sub-methods of the overall method (Figure 3.1). At the end, the combination step is briefly described.

4.1 Sensor Analysis

The main content of this section has already been published in the author's previous publications [174–176]. For a better understanding of the present thesis a summary of the already published work is given.

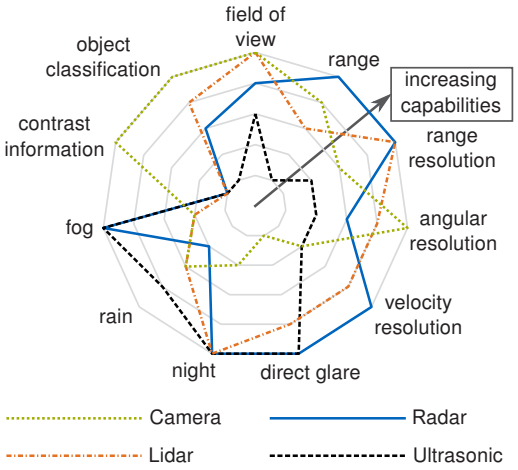
This section highlights the main aspects of the sensor analysis sub-method. First, the necessary basics are presented. Next, the approach is described and a summary of the results is given at the end.

4.1.1 Preliminaries

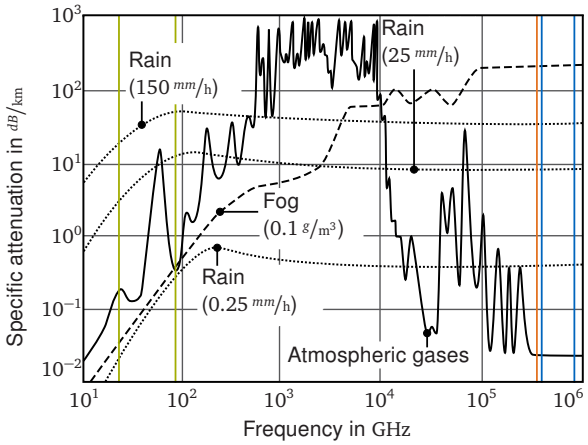
The aim of this sub-method is the analysis of the sensor setup and the derivation of weaknesses. According to [17], there are not many publications covering the sensor setup of AVs in the context of the safety assessment. A relevant example is [177], which models the occlusion by other road users in dependency of the traffic density. On the other hand, the work of Berk [178] must be highlighted, who investigates the reliability of individual sensors and entire sensor setups. In simple terms, he tries to answer the question how many sensors are needed to ensure that the occurrence of a fault is lower than a defined risk. In contrast, the method presented here tries to identify weak spots in the sensor setup and derives test scenarios from these weak spots.

The basis for this is provided in the already published work [174], where the necessary basics are described in more detail. For the analysis of the sensor setup, models of the sensors are needed. Based on [179–182], a classification of the models into ideal, phenomenological and physical sensor models is used. Ideal sensor models represent the ground truth. If an object is in the field of view of the sensor, it is detected, if not, then not. Phenomenological models can reproduce individual effects such as attenuation by fog without reproducing the physical relationship between input and output. Physical models include the physics of the sensor. For example, the propagation of rays can be calculated using the ray-tracing method, thereby simulating physical processes such as reflection and refraction of rays on objects. They have the highest accuracy, but also require by far the most computing capacity. Based on a compromise between accuracy and computing power, phenomenological sensor models are used in this thesis. Furthermore, high-precision physical sensor models do not add value because the inputs (e. g. weather conditions or material properties of the object to be detected) are provided in rather generic terms.

In the context of automated driving, ultrasonic, camera, Radar and Lidar sensors are used to perceive the environment. Since each type of sensor has certain strengths and weaknesses (Figure 4.1a) and is, for example, influenced differently by environmental conditions (Figure 4.1b), a combination of individual sensors is used and their detections are fused. The sensor types can be divided into active (ultrasonic, Radar, Lidar) and passive sensors (camera). Active sensors emit actively modulated radiation which is reflected by objects and absorbed by a receiver. Passive sensors (in this case cameras) do not emit radiation and instead a receiver absorbs secondary high-frequency radiation in the visible spectrum (light).



(a) Strength and weaknesses of the most common sensors for automated driving (adapted from [183] and [184]).



(b) Specific attenuation due to atmospheric gases and weather influences based on [185]. The frequencies of 24 GHz and 77 GHz Radar are depicted in green, the Lidar in red and the visible light spectrum in blue, respectively.

Figure 4.1: Preliminaries for the sensor analysis sub-method. Both figures are adapted from [174].

For the phenomenological modeling of active sensors, the Signal-to-Noise Ratio (SNR) is suitable. This equals the ratio of the power of the received signal P_r to the power of the noise P_n:

$$SNR = \frac{P_\mathrm{r}}{P_\mathrm{n}}. \tag{4.1}$$

As an example, the equation for the Radar from [186, chap. 2] and [187] is given as:

$$SNR_\mathrm{Radar}(R, \phi_\mathrm{s}, \psi_\mathrm{s}) = \frac{P_\mathrm{e} G_\mathrm{e}(\phi_\mathrm{s}, \psi_\mathrm{s}) G_\mathrm{r}(\phi_\mathrm{s}, \psi_\mathrm{s}) \sigma \lambda^2}{(4\pi)^3 R^4 L_\mathrm{oa}(R) k B_\mathrm{n} T_\mathrm{sys}}. \tag{4.2}$$

Due to the large number of formula symbols and because these are not discussed in detail, please refer to the list of formula symbols on page V for a description of them. Based on the SNR a detection probability can then be derived.

In the author's first publication [174], the passive camera is modeled with a surrogate model from the literature as an active sensor. However, since the algorithms for object detection have a considerable influence on the detection probability of passive sensors and especially of the camera, the author's publication [175] develops a model of camera-based object detection based on meta-information such as the distance to the object. The approaches are described in more detail in the following section.

4.1.2 Approach

The procedure is explained using Figure 4.2, which shows the entire sensor analysis approach. The method has three inputs with the sensor setup, the environmental conditions and the object to be detected. The former is available from the manufacturer through the system documentation provided. Input two and three depend on the ODD of the vehicle and this must also defined by the manufacturer and made available to the technical service. As output (block 'scenario transformation'), the procedure generates a cost-optimal road geometry for the given inputs. Optimal in this case means that it is most difficult for the test vehicle to detect the object correctly. In the following, the individual blocks are briefly described.

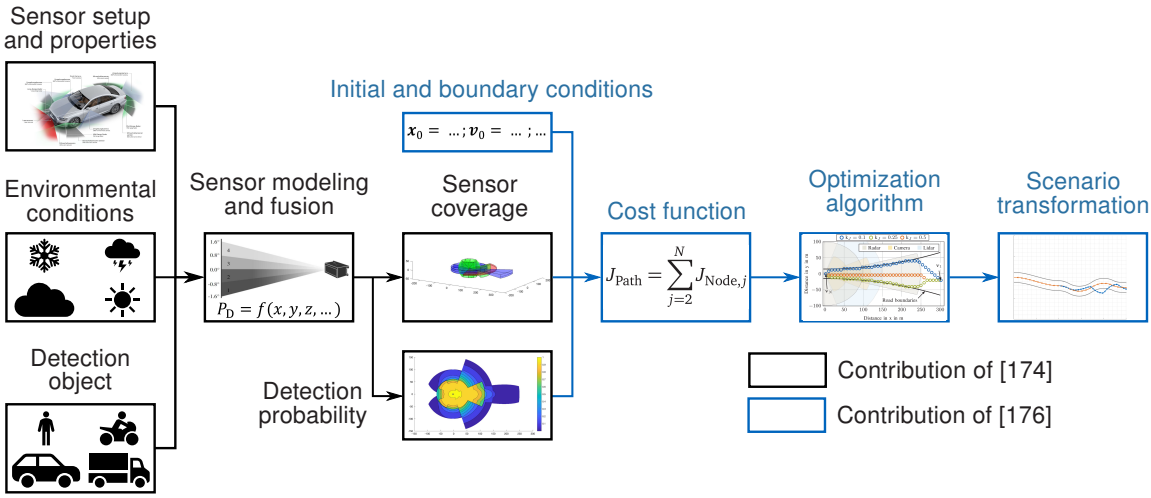


Figure 4.2: Summary of the sensor analysis approach (adapted from [176]).

Sensor modeling and fusion

The modeling of the sensors is performed in [174] using phenomenological models based on literature. The camera is also modeled as an active sensor with a surrogate model. An improved approach is presented in [175], in which the camera-based object detection is modeled on the basis of meta information. This model can replace the surrogate model from [174]. This improves the quality of the camera detection probability calculation but has no influence on the further steps of the methodology.

The properties of the sensors, such as the opening angle and the emitted power, are included in the calculation of the sensor models. In addition, the environmental conditions such as lighting conditions or rain as well as the type of object to be detected are also taken into account. The latter has an influence on the reflection properties of the radiation emitted by the sensor. For example, a truck has a much larger radar cross section than a pedestrian.

To develop a model for camera-based object detection, a publicly accessible data set (nuScenes data set [188]) is used in [175] and the performance of three state-of-the-art camera-based object detectors is investigated. For each object detection 22 defined meta-information are assigned to each object. This includes, for example, the relative position to the vehicle, the orientation to the vehicle, the velocities of the vehicle and the object, the degree of occlusion of the object and rain. Subsequently, the performance of the neural network based object detectors is imitated using simple random forest models. Thereby, the defined meta-information is used as input for the random forest. The random forest learns to predict, solely based on the meta information, whether the underlying object detector can correctly detect an object. With

the help of a subsequent examination using a explainable Artificial Intelligence (AI) algorithm called SHapley Additive exPlanations (SHAP) method [189], it can then be analyzed which meta-information has a large influence on the detection performance for both, in general and for specific detections. This represents the model that can be used to predict whether camera-based object recognition can correctly detect a specified object. In [175] it is shown that general statements can be derived, but for the application it is advantageous to apply this procedure again for the camera-based object detection of the VUT. As a result, suitable phenomenological models are now available for both, the active sensors and the passive camera sensor.

Sensor coverage

The sensor properties of all sensors used are sufficient for the sensor coverage, because in this step only the coverage is considered, which is equivalent to ideal sensor models. This can later be used to identify blind spots of the vehicle.

Detection probability

To calculate the detection probability, the existing and developed phenomenological sensor models are used to assign a detection probability to each point in a three-dimensional grid. This value indicates how likely it is that a pre-defined object at that location can be correctly detected by the vehicle under defined environmental conditions. This enables a more detailed investigation to be carried out to determine whether the sensor setup used by the manufacturer has weaknesses that can be addressed with test scenarios.

Cost function

Based on the detection probability grid, a cost-optimal approaching path via the nodes of the grid to the vehicle is calculated in [176], with which an object must approach the ego-vehicle in order to be seen as poorly as possible by the vehicle.

In a first step, a cost function must be defined. The cost function $J_{\mathrm{Node},j}$ of a node v_j of the grid is defined in [176] as

$$J_{\mathrm{Node},j} = \frac{k_{\mathrm{J}} + P_{\mathrm{D}}(v_j)}{k_{\mathrm{J}} + 1}\, d(v_j, v_{j-1}), \tag{4.3}$$

with the detection probability $P_{\mathrm{D}}(v_j)$ of the current node v_j and the Euclidean distance $d(v_j, v_{j-1})$ between the current v_j and the previous node v_{j-1}. In addition, a weighting factor k_{J} is introduced in order to weight $P_{\mathrm{D}}(v_j)$ and $d(v_j, v_{j-1})$ differently in relation to each other. A minimum weighting in the cost function of the distance of the approaching path is necessary to ensure a purposeful approach of the object towards the vehicle. More details can be found in [176]. The cost function of the entire path J_{Path} with N nodes yields to

$$J_{\mathrm{Path}} = \sum_{j=2}^{N} J_{\mathrm{Node},j}. \tag{4.4}$$

Optimization algorithm

The aim of this step is to calculate the optimal approaching path based on the detection probability grid using the previously defined cost function. To calculate the path from a given starting point to

the vehicle through the three-dimensional grid, the A-star (A*) algorithm is used. The A* algorithm is one of the heuristic search algorithms that guarantees to find the cost-optimal path between two nodes if it exists ([190, chap. 6.3.2]). The total cost of the path via node v_j is calculated in the A* algorithm using a heuristic evaluation function $f(v_j)$:

$$f(v_j) = g(v_j) + h(v_j). \tag{4.5}$$

Thereby, $g(v_j)$ is the cost of the path from the start node v_1 to the current node v_j and $h(v_j)$ is an estimate of the remaining cost from the current node v_j to the end node v_N. Both, v_1 and v_N can be defined by the user. The cost function from the previous section can be used directly as $g(v_j)$. When choosing $h(v_j)$, it must be noted that $h(v_j)$ must never overestimate the remaining costs. For this reason, the cost function with vanishing detection probability P_D is used for all nodes from v_{j+1} to the end node v_N. Thus the A* algorithm calculates in an iterative process the optimal approaching path $\mathbf{x}_\mathrm{oap}$ from a given start node v_1 to an end node v_N.

To limit the search space, the number of nodes is reduced by taking the ODD of the vehicle into account. This is done by considering the different design classes according to the German Motorway Construction Guideline [52]. For example, this specifies lane widths and minimum allowed curve radii. This restriction excludes all nodes in the calculation that are located outside a valid highway geometry.

Scenario transformation

In the previous step, an optimal relative approaching path of the object (challenger) to the ego-vehicle is calculated. This represents the static relative path that must be followed by the challenger in relation to the ego-vehicle. In simple terms, the challenger must perform this path when the ego-vehicle would be stationary. In a realistic test scenario, however, both vehicles move. Therefore, the relative approaching path is transformed into a dynamic scenario. Assuming a given relative speed and again taking into account the ODD, the German Motorway Construction Guideline [52] and the maximum possible vehicle dynamics, a road geometry is calculated that results in $\mathbf{x}_\mathrm{oap}$ and complies with the constraints of the German Motorway Construction Guideline [52]. For a summary of the required algorithms, please refer to [176].

The output is the road geometry as well as the trajectory of the challenger. This in turn is the input for test scenarios, with regard to the sensor setup used, which represent the highest possible challenge for the ego-vehicle.

4.1.3 Results and Validation

The results are divided into sections according to the blocks in Figure 4.2. As can be seen in the figure on the overall methodology in Figure 3.1, the results are system-specific. This means that for each AV, specially tailored results can be derived.

Sensor modeling and fusion

In the methodology of Section 4.1.2, the phenomenological sensor models are the starting point. The aim is not to develop these models, but to use them with information available in the literature. This is applicable for the active sensors (ultrasonic, Radar, Lidar), but not for the camera as a passive sensor. Therefore, the developed model of camera-based object detection [175] is briefly presented at the beginning of the results.

For the development of the model of camera-based object detection, the performance of object detectors is investigated first. Afterwards the behavior is imitated by a simple random forest and the performance is traced back to meta-information. Three object detectors are investigated, but they show similar results. Therefore, for simplicity only the result for the RetinaNet [191] is given. Figure 4.3 shows the 15 most important meta-information with the highest influence on the detection result. On the left side of the figure, it can be seen that the occlusion has the highest influence. This means that having information about the degree of occlusion of the object significantly assists in predicting whether the RetinaNet will detect the object correctly. On the right side of Figure 4.3 the influence of the values of the meta-information is shown. Each colored dot corresponds to a prediction made. Positive SHAP values mean an increased probability that the object will be correctly recognized and negative values indicate a decrease. It is clearly visible that high values for occlusion lead to negative SHAP values and thus reduce the probability of a correct object detection.

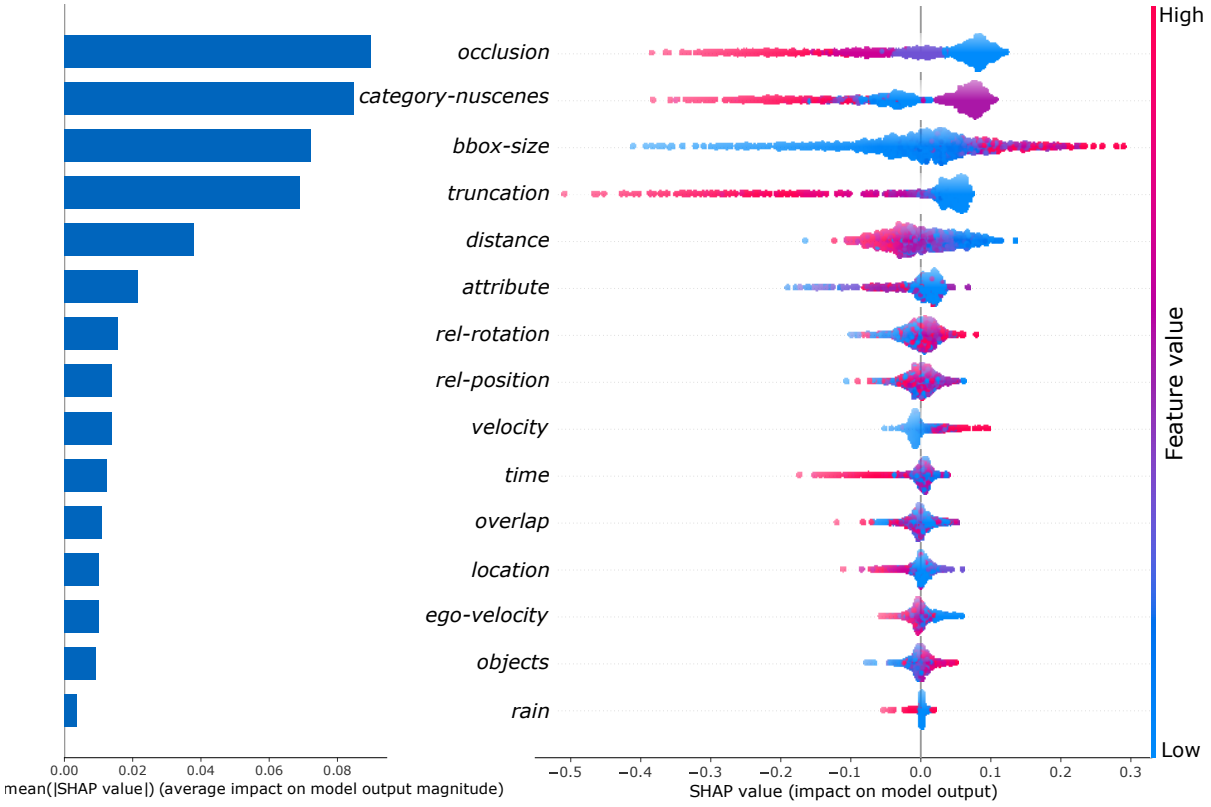


Figure 4.3: Left: Overview of the importance of the meta-information measured as the average impact of each meta-information on the detection score of the random forest modeling the detection performance of the RetinaNet. Right: Global impact of the meta-information on the prediction output of the random forest modeling the detection performance of the RetinaNet. Each dot represents an individual Shapley value. The corresponding value of the meta-information to the Shapley value is indicated by the coloring, where blue corresponds to low and red to high values. The figure is adapted from [175].

Figure 4.4 shows an example application of the model. The question arises whether the camera-based object detection can correctly detect the pedestrian framed in red. Below the image the SHAP values of the meta-information are shown and the model predicts with a probability of $60\,\%$ (marked in bold) that the RetinaNet correctly detects the pedestrian. The highest positive influence is that the object on the image is not truncated ($truncation = 0\,\%$). The validation shows that the RetinaNet actually detects the pedestrian correctly and that an overall accuracy of almost $85\,\%$ is achieved. Since the ground truth data of the meta-information must be available

for the application of the model, it can only be used for offline applications. However, this is no limitation for the definition of test scenarios.

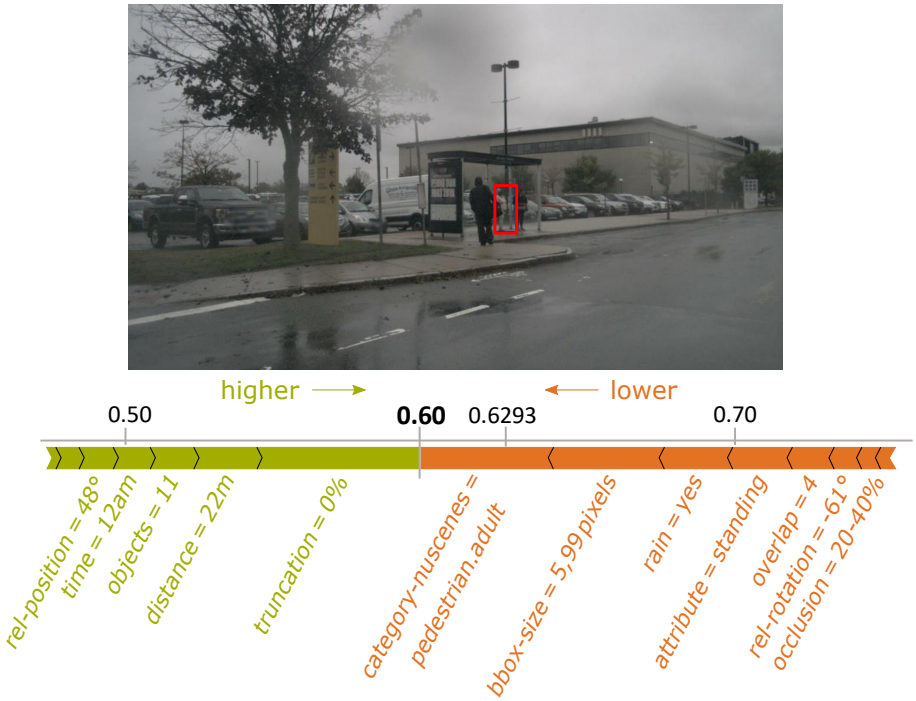
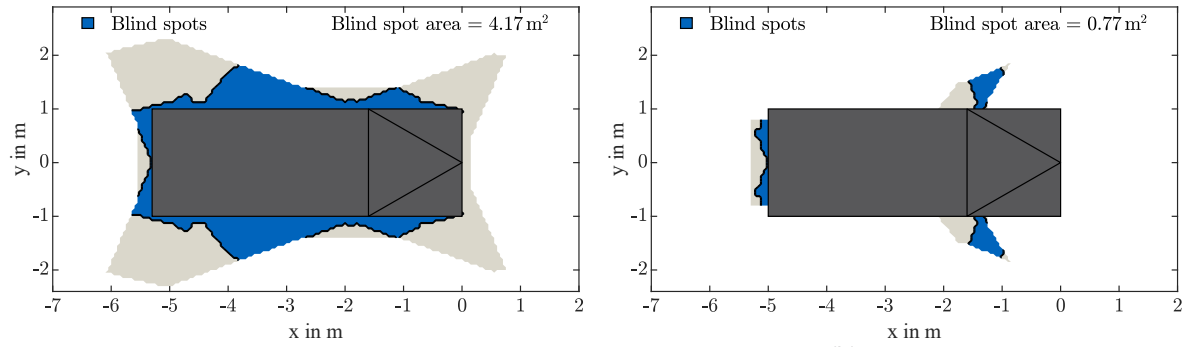


Figure 4.4: So-called force plot to visualize the estimated impact of each meta-information on the detection score of the random forest (0.60) for the marked pedestrian. The influence of each meta-information with respect to the baseline of 0.6293 is illustrated as an arrow. The baseline is the average of the output of the random forest over all leaves. Meta-information with positive Shapley values (green arrows) promote a correct detection, while meta-information with negative Shapley values (red arrow) reduce the probability of a correct detection. The length of each arrow indicates the height of its impact. The figure is adapted from [175].

Sensor coverage

The sensor coverage can be used, for example, to investigate whether there are blind spots in the manufacturer's sensor setup that can be specifically addressed when generating test scenarios. As already mentioned, the results are system-specific, which is why two systems are compared below. These two systems are exemplary systems and are intended to illustrate the system-specific aspect of the developed sub-method. They are based on (but are not an exact replica of) the sensor setup of commercially available cars from two different manufacturers and are generally referred to as system A and system B in the following.

When comparing the two vehicles in the near field around the vehicle in Figure 4.5, it is observed that the area of blind spots at manufacturer A is significantly larger. Blind spots are defined as areas that cannot be seen at any height by any sensor. In addition, Figure 4.5 also shows the area in gray that does not fall within the field of view of any sensor at a height of $0.1\,\mathrm{m}$. In this area, the detection of small objects is therefore not possible. Overall, it can be concluded that system A has a significantly higher risk of not detecting nearby objects. Therefore, special scenarios must be selected in which objects are located in the close vicinity. For the considered highway use case, this can be of particular importance during starting from standstill and slow driving in congestion. With system B this focus is not necessary due to the good coverage.

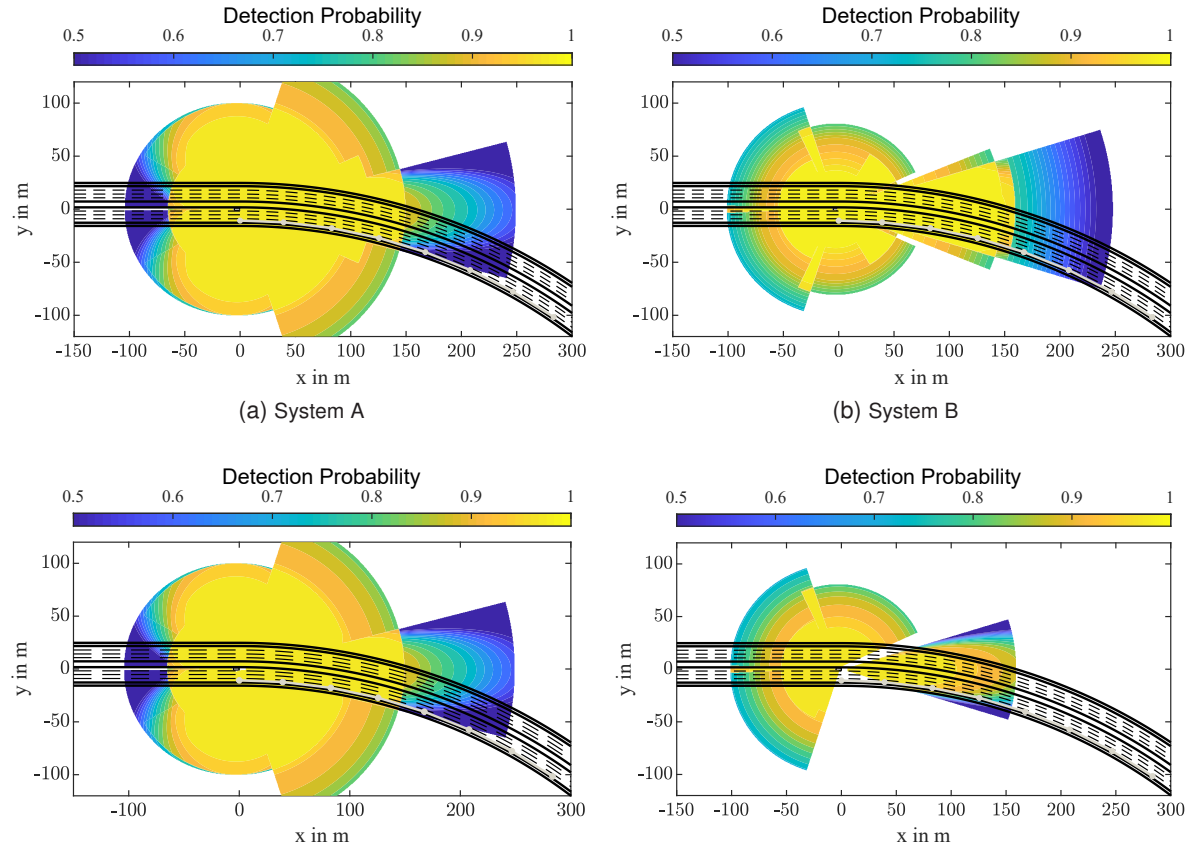


(a) System A (b) System B

Figure 4.5: Blind spots (blue) of both sensor configurations. For a description of the sensor setups see [176]. Blind spots are areas that do not fall within the detection range of a sensor at any height. The sum of all blind spots equals an area of $4.17\,\mathrm{m}^2$ for System A and $0.77\,\mathrm{m}^2$ for System B. In addition, the area that does not fall into any detection area at a height of $0.1\,\mathrm{m}$ is shown in gray.

Detection probability

Based on the detection probability, the far periphery of the vehicle can be examined more precisely, taking into account defined weather conditions. For this purpose, the two systems are compared again and two different weather conditions are considered (Figure 4.6). Under good



(a) System A (b) System B

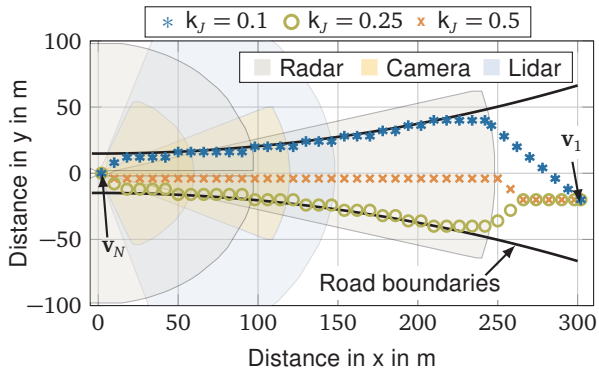
(c) System A (d) System B

Figure 4.6: Calculated detection probability of the ego-vehicle under normal (Figure (a) and (b)) weather conditions and with direct glare from the front right (Figure (c) and (d)). The ego-vehicle drives in the left of a four-lane highway approaching a curve of radius $r = 500\,\mathrm{m}$. Depicted in gray is the path of a passenger car driving in the right-hand lane. The path of the passenger car is considered only in the area in front of the automated vehicle. All four figures are adapted from [176].

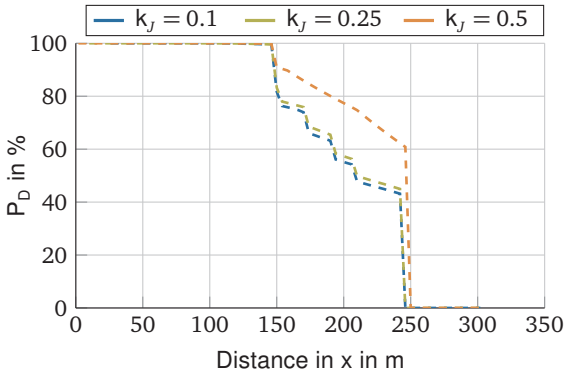
weather conditions (Figure 4.6a and 4.6b), both vehicles have good coverage of a four-lane highway. In direct glare from the front right, i.e. when the performance of the forward and right-facing cameras is significantly reduced, Figure 4.6c and 4.6d show that system B has significant shortcomings in the coverage and this must be taken into account in the tests, unless the manufacturer excludes these conditions from the ODD of the vehicle. How the object (challenger) has to approach optimally in the defined detection probability grid is explained below.

Optimization algorithm

In the following, the result of the path optimization using the A* algorithm is shown using system A under good weather conditions and a car as the object to be detected. The start node v_1 of the approaching path is located outside the field of view of the sensors (Figure 4.7a). In addition, the constraints (adherence to road boundaries) are shown. Although the optimization is performed in a three-dimensional grid, only the representation in the x-y-plane is discussed, because the effects in z-direction are small. A noteworthy influence in the z-direction only occurs with hilltops, which is addressed in the discussion in Subsection 5.1.1. It can be seen that as k_J increases, the focus is increasingly put on the shortest possible approach path. Figure 4.7b shows that the optimizer for decreasing k_J focuses more and more on exploiting areas with the lowest possible detection probability. This confirms the functionality of the optimizer.



(a) Optimal approaching path $\mathbf{x}_\mathrm{oap}$.



(b) Detection probability P_D over distance.

Figure 4.7: Results of the approaching path $\mathbf{x}_\mathrm{oap}$ optimization. Left: Optimal approaching path $\mathbf{x}_\mathrm{oap}$ with different weighting factors k_J. Right: Detection probability P_D with different weighting factors k_J. The high detection probabilities P_D of almost $100\,\%$ are due to multiple overlapping sensors and the assumed good weather conditions. This can change in adverse weather conditions [174]. Both figures are adapted from [176].

Scenario transformation

Using scenario transformation, the optimal approaching path $\mathbf{x}_\mathrm{oap}$ is transformed into a dynamic scenario and a corresponding road geometry (Figure 4.8). The ego-vehicle starts in v_N and the challenger in v_1, which corresponds to the starting position of the path optimization in Figure 4.7a. Then both vehicles move in positive x-direction and the relative approach of the two vehicles corresponds to the optimal approaching path $\mathbf{x}_\mathrm{oap}$. It can be seen that the two vehicles are not on a common valid road geometry at approximately $400\,\mathrm{m}$. The constraint is therefore violated in this area and must be considered by an improved implementation in future work.

Validation

The focus of the approach is not on the development of sensor models of active sensors, because these are well documented in the literature and are therefore assumed to be valid. The

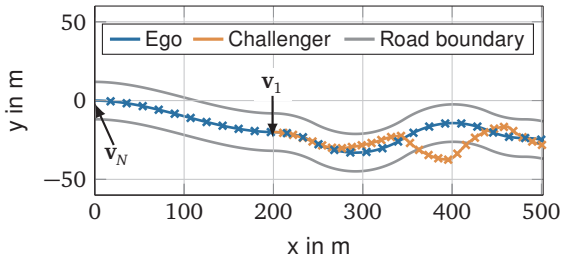


Figure 4.8: Calculated optimal road geometry to generate $\mathbf{x}_{\text{oap}}$ between the two vehicles. From $\mathbf{v}_1$, i.e. from $200\,\text{m}$, the road can theoretically also be defined around the challenger.

self-developed model for camera-based object detection can also be regarded as valid with an accuracy of almost $85\,\%$ which is in the range of state of the art object detectors [192]. The optimization based on these models uses the A* algorithm, an algorithm which, if implemented correctly, finds the optimal path as long as it exists. Based on the plausible results, a correct implementation and thus validity can be considered. The transformation into a dynamic scenario provides plausible results, but cannot yet meet all constraints, which must be addressed in future work. In summary, it can be concluded that the generated results are valid.

4.2 Driving Behavior Characterization

The main content of this section has already been published in the author's previous publications [127]. For a better understanding of the present thesis a summary of the already published work is given.

This section highlights the main aspects of the driving behavior characterization sub-method. First, the necessary basics are presented. Next, the approach is described and a summary of the results is given at the end.

4.2.1 Preliminaries

In order to characterize the driving behavior of AVs, on the one hand, KPIs are required to evaluate the driving behavior. On the other hand, meaningful traffic situations are necessary in which the driving behavior can be observed properly. These situations can be diverse. For example, a simple straight drive can be used to investigate whether the vehicle oscillates between the lane markings. However, situations that are more extensive may also be necessary, for example to characterize specifically the merging behavior on motorways. The basics for both, the meaningful traffic situations and the KPIs are briefly introduced below.

For the evaluation of the driving behavior, there is literature that studies the driving behavior of human drivers. For the present work, however, the publications on driving behavior of ADAS are of particular interest. Holzinger [193, 194], who is dealing with the evaluation of ACC and Lane Keeping Assistant (LKA), does important work in this field. Characteristic values of particularly aggressive and defensive driving behavior are investigated in [195–197]. It should be noted that the threshold values for these behaviors depend strongly on the driving situation [197].

The characteristic situations in which the driving behavior is evaluated are usually available in simple representations described in words, i. e. as functional scenarios. In some cases, parameters are given so that logical scenarios are already provided. An example of a functional scenario of longitudinal control for the investigation of reaction time is the sudden stopping of

the front vehicle [194]. Further situations can be taken from [198, 199]. Besides the technical literature, other sources can be used to define characteristic situations. These include, for example, the (German) Driving License Directive [200] or driving safety training courses. The former, together with the theoretical Driving License Questionnaire [201], contain a variety of suitable scenarios for characterizing driving behavior. In the driving license guideline, these are only tested in theory because their probability of occurrence is low. When testing AVs, such scenarios can be simulated using virtual simulation.

4.2.2 Approach

In the following, the procedure for characterizing the driving behavior for the adaptation of future scenarios is presented. The structure of the method can be seen in Figure 4.9. With the scenario definition, the scenario execution and the scenario evaluation three main process steps are introduced, whereby the methodological focus is on the former and the latter. All three steps are described below.

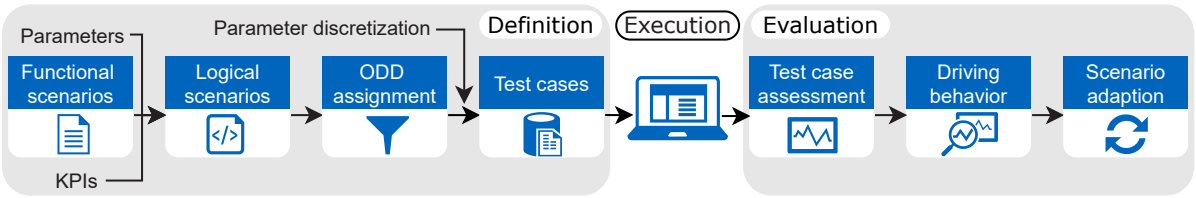


Figure 4.9: Overview of the driving behavior characterization sub-method (adapted from [127]).

Definition

The scenario definition is divided into four sub-steps as shown in Figure 4.9. First, functional scenarios are extracted from the sources described in Subsection 4.2.1. These are differentiated according to the distinction between lateral and longitudinal guidance known from vehicle dynamics in order to allow a more comprehensible structure. In addition, the scenarios are structured according to the primary type of acceleration. For example, when driving through a curve with constant velocity, a constant lateral acceleration occurs, which is why the functional scenario 'driving through a curve' is assigned to lateral guidance under constant acceleration. A summary of the number of derived functional scenarios is given in Table 4.1.

Table 4.1: Distinction of the scenarios into longitudinal and lateral control. Further subdivision according to the primary acceleration type occurring during the scenario. A distinction is made between no, a constant, a transient and a periodic acceleration. The table shows the number of identified scenarios for each category. A description of all scenarios can be found in [202].

	Number of scenarios			
	Null	Constant	Transient	Periodical
Longitudinal	18	29	5	1
Lateral	0	9	13	2

In order to turn the functional scenarios into logical ones, the necessary parameters and their ranges must be defined. Since all concrete scenarios derived from the same logical scenario that are executed later are assessed with the same KPIs in the evaluation, the KPIs are already defined here and assigned to the logical scenarios. An overview of the defined KPIs can be found

in Table 4.2. The five-layer model [15] is used for the structured definition of the parameters and a total of 34 parameters are defined for the derived functional scenarios. If a parameter is not defined for a scenario, it is assumed to have no influence and a default value can be used.

Table 4.2: The KPIs are divided into six groups according to the corresponding physical quantities. The table shows the number of defined KPIs per category and one corresponding KPI as an example.

Based on	Number of KPIs	Example
Distance	6	Distance to center of ego lane
Velocity	2	Maximum lateral velocity
Acceleration	10	Maximum lateral acceleration
Angle	3	Yaw angle
Time	3	Time-To-Collision (TTC)
Frequency	2	Oscillation around the center of the lane

Then the logical scenarios can be selected according to the ODD of the AV under test. The categories general, city center, country road and motorway are defined in Table 4.3 and the corresponding logical scenarios are assigned. The general category includes all scenarios that cannot be uniquely assigned to one of the other three categories. To derive test cases, i. e. concrete scenarios including KPIs, the existing parameter ranges must be discretized so that concrete numerical values for the parameters are available. Thereby, the real range of usage of the AV should be covered, because the goal is to identify characteristic driving behaviors that occur generally and not only under a certain parameter combination.

Table 4.3: The defined scenarios are subdivided according to four different ODDs. The table shows the number of scenarios per category.

Operational Design Domain (ODD)			
General	City center	Country road	Highway
38	26	3	10

Execution

The execution of the scenarios is not the focus of this thesis and is only mentioned for the sake of completeness. For the execution by means of virtual simulation, the concrete scenarios can be transferred into a machine-readable format (e. g. OpenDRIVE and OpenSCENARIO), or implemented and executed directly in a simulation tool such as IPG CarMaker®.

Evaluation

The concluding evaluation is divided into three sub-steps according to Figure 4.9. First, the previously defined KPIs are applied to the results of the virtual simulation. The KPIs can be used to identify characteristics in the driving behavior. The characteristics are defined manually based on expert knowledge. The evaluation of different parameter combinations ensures that the behavior is consistent and not a randomly occurring single event. Finally, this knowledge will be used to make future scenarios even more challenging for the AV under test. Subsequently, a proof of concept of this approach is shown using exemplary results.

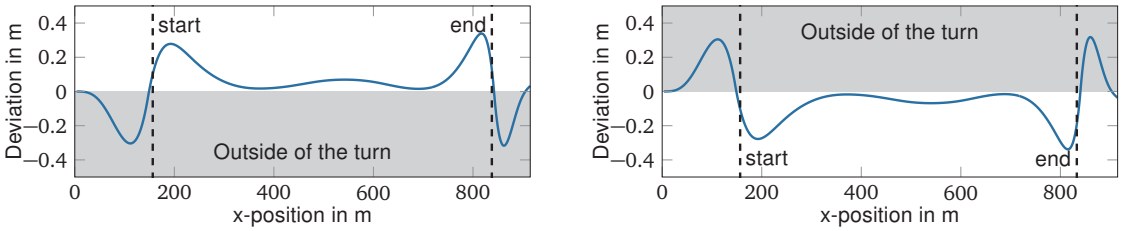
4.2.3 Results and Validation

This sub-method has the smallest scope of the three sub-methods and is only prototypically implemented. For this reason, the results and the validation are only briefly addressed here.

Results

The proof of concept is confirmed by exemplary results of the functional scenario 'driving through a curve'. The KPI is the deviation from the center of the vehicle to the center of the ego lane. The parameters defined are the curve radius R_{curve}, the desired velocity v_{set} and the direction of the curve. The curve radius R_{curve} is discretized in a range of $400\,\text{m}$ to $1000\,\text{m}$ in $200\,\text{m}$ steps. The speed in $20\,\text{km/h}$ steps from $60\,\text{km/h}$ to $100\,\text{km/h}$ and for the direction of the curve the values left and right exist. For the proof of concept, a simple driving function for lateral control was used and the scenarios were simulated with IPG CarMaker®.

Figure 4.10 shows the KPI under consideration, the deviation from the center of the vehicle from the center of the ego lane over the driven distance. The vertical lines mark the beginning and end of the curve, respectively. It is clearly visible that the vehicle first drives to the outside of the curve at the beginning of the curve and then cuts the curve towards the inside. At the end of the curve, the opposite behavior can be observed accordingly. In order to prove that the driving behavior of the ego-vehicle is characteristic for the whole parameter space, the maximum values of the considered KPI are plotted over the parameters curve radius and desired velocity in Figure 4.11. The figure confirms that this is a characteristic driving behavior of the AV.



(a) Left turn

(b) Right turn

Figure 4.10: Comparison of the system behavior between a left and a right curve with an identical radius of $R_{\text{curve}} = 800\,\text{m}$ at identical speed of $v_{\text{set}} = 70\,\text{km/h}$. The system shows the same driving behavior, a clearly visible curve cutting, in both left and right-hand curves. Due to the comparable behavior, only the left-hand curve will be considered in the following. The dashed vertical lines represent the beginning and the end of the curve respectively. Both figures are adapted from [127].

It is intended to adapt future scenarios to make them even more challenging for the AV under test. If, for example, scenarios with static obstacles are considered, then these can be placed in a precise position using the information obtained. In the case under consideration, before the curve entrance on the outside of the curve, or shortly after the curve entrance on the inside of the curve. These are more challenging because the AV will drive at these areas on the respective side of its own driving lane and thus the distance to the static object is smaller due to the driving behavior. Smaller distances in turn increase the probability of a critical situation occurring. Thus, the number of necessary parameter variations can be reduced with the described method. For example, if one assumes that the position of the static obstacle is discretized in ten steps, then the optimum placement of the obstacle can reduce the parameter combination to $10\,\%$ of the initial quantity. Even if no critical situation occurs when executing the scenario, this scenario is more difficult for the AV under consideration because in this case it had to change its standard control strategy.

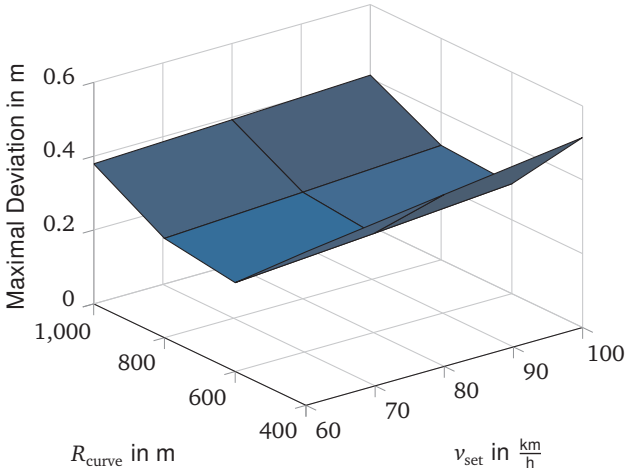


Figure 4.11: Maximal deviation of the center point of the vehicle to the center line of the ego-lane shortly after the entrance of a left turn. The AV shows an identical performance over the entire speed range and over all curve radii. In all concrete scenarios, there is a positive deviation, i. e. the AV drives much closer to the inside of the curve in all tested left-hand curves shortly after the curve entrance. The figure is adapted from [127].

Validation

The developed methodology cannot be completely automated and requires expert knowledge. For these reasons, the procedure has only been implemented as a prototype. The thereby performed proof of concept confirms the feasibility and delivers plausible results. Thus, plausibility is given but a comprehensive validation must be the subject of future work.

4.3 Traffic Situation Complexity

The concept of this section and first results have already been published in the author's previous publication [16]. In the following, this concept is taken up, explained in depth and the results are presented in detail.

The following section forms the core of this document and is explained in detail, which is why the extent is much larger compared to the first two sub-methods in Section 4.1 and 4.2. It explains a methodology for assessing traffic complexity in relation to AVs. It is important to note that complexity is defined in Section 2.1 as particularly challenging scenarios for the planning module of the AV in terms of the behavior of other TPs, i. e. Layer 4 of the five-layer model [15]. A second important aspect is that complexity refers to the AV. Therefore, the question is what is a complex traffic situation for AVs and not, what is a complex traffic situation for human drivers. Furthermore, the developed metric is intended to assess the scenario itself and not the behavior of the ego-vehicle. The metric is therefore (as far as possible) independent of the ego-vehicle behavior.

This section highlights the main aspects of the traffic situation complexity sub-method. First, the necessary basics are presented. Next, the approach is described and the results are discussed at the end.

4.3.1 Preliminaries

The methodology described in this section is intended to enable the assessment of the complexity of scenarios so that the most complex and thus most challenging ones can be identified and selected. For the complexity assessment, a newly developed metric is applied. In literature, related work exists in the context of automated driving, which will be introduced briefly below. Subsequently, necessary basics in relation to the sub-method are explained. It should be noted that the term complexity is often used differently in the relevant literature than in this thesis.

Related work

Wang et al. [129] present a two-part metric for assessing complexity. These are a road semantic complexity and a traffic element complexity. Both are weighted with a weighting factor and added linearly. For the present thesis, traffic element complexity is the more important part. However, only the distance and the angle to the surrounding TPs are considered in [129]. The evaluation of the total complexity is normalized to a range of values from zero to one and is divided into three equidistant ranges (general, medium and extreme).

Complexity metrics based on the analytic hierarchy process in combination with a relative importance index are developed by [124, 130, 131]. The resulting metric is called complexity index by the authors. The publications use a lane departure warning system as use case. Therefore, the focus is strongly on complex road geometry. Complex road design is not that important for the considered highway use case, because it is well standardized by the German Motorway Construction Guideline [52], but for city center and country roads this may be interesting in future work.

For a reinforcement learning process for trajectory generation, Gonzalez et al. [203] use a cost function, which can also be considered as a complexity function. It is based on a linear combination of weighted features. The features are divided into static and dynamic features. The dynamic features relevant for this thesis are limited to criticality measures like time headway.

Based on the trajectories that lead to inadequate ego-vehicle performance, Qi et al. [128] use the so-called scenario character parameter. By examining the scenario character parameter, scenario groups can be generated and then compressed into a challenging scenario.

A general description of the complexity of traffic situations for AVs based on eight influencing factors is presented by Schuldt [204, chap. 2.3.5] in his dissertation. These eight factors are:

1. **Number of Variables:** Describes how many elements (TPs) participate in a scenario.
2. **Connectivity:** Describes the mutual influence on several elements when the behavior of an element changes.
3. **Self-dynamic:** Describes whether the scenario develops itself without having received modified external impact.
4. **Non-transparency:** Describes whether elements are not visible or not known to the decision maker (AV under test).
5. **Multiple targets:** Describes whether elements pursue contradictory goals.
6. **Indistinction:** Describes whether the target situation of the scenario is clearly identifiable.

7. **Novelty:** Describes whether the scenario and the actions it contains are known to the decision maker.
8. **Number of states per element:** Describes the number of states that an element can occupy in the scenario.

Schuldt [204, chap. 2.3.5] applies these factors to example scenarios and evaluates them subjectively. His conclusion is that the method is applicable, but that an objectification of the factors must be implemented in future work. A validation of the influencing factors is not conducted.

A suitable metric for the evaluation of the trajectories of the surrounding TPs in order to assess the complexity is not available in literature, because many authors limit themselves to the road geometry, or the metric is only available in subjective form. Nevertheless, important information can be derived from the existing publications, which can be used for the development of the complexity metric:

- Use of a linear combination of influence factors and associated weighting factors [124, 129–131, 203].
- Normalization of the complexity values to a value range from zero to one and subsequent equidistant division of complexity into three classes [129].
- The factors considered by Schuldt [204, chap. 2.3.5] represent a good starting point for the influencing factors to be considered.

Fundamentals

This section briefly introduces necessary tools for the sub-method to be developed. These include the highD data set, the Genetic Algorithm (GA), and a simple model to mimic an AV.

Scenarios are needed so that the metric to be developed can be applied. In order to be able to use scenarios that are as realistic as possible, a freely available data set covering the considered highway use case is required. Popular data sets are for example the Next Generation SIMulation (NGSIM) [205] and the highD [56] data set. The latter is used in the presented thesis, because it is comprehensive and covers exactly the considered use case (German Autobahn). An overview of available data sets can be found in [53, 54]. The highD data set was recorded with the help of a drone with a frame rate f_{highD} of $25\,\mathrm{Hz}$ and contains the trajectories of all vehicles passing through a $400\,\mathrm{m}$ to $420\,\mathrm{m}$ long, straight section of a motorway. A summary of the data set can be seen in Table 4.4. The values are based on [56] and are all rounded.

Table 4.4: Summary of the highD data set based on [56]. The values given by [56] are rounded.

Recording duration	Lanes per direction	Recorded distance	Vehicles	Cars	Trucks	Driven distance	Driven time	Frame rate f_{highD}
16.5 h	2 to 3	400 m to 420 m	110,000	90,000	20,000	45.00 km	447 h	25 Hz

Within the developed method there is a step that optimizes the trajectories of the surrounding TPs. For this purpose, a GA, an optimization algorithm from the class of evolutionary algorithms, is used. In the following, the principle of the GA is briefly presented in Algorithm 1. For further information, the interested reader is referred to the detailed description in [206].

During optimization, the trajectories of the surrounding TPs from a scenario from the highD data set are adjusted in such a way that the considered scenarios become even more challenging.

Algorithm 1 Genetic algorithm

Require: Initial population, fitness function ▷ Initial population consists of individuals
Ensure: Optimal solution
1: Evaluate initial population ▷ Initial population is evaluated with fitness function
2: $stopCondition \leftarrow false$
3: **while** $\neg stopCondition$ **do**
4: Select best individuals ▷ Evaluated with fitness function
5: Reproduce individuals ▷ Apply crossover and mutation to children
6: Evaluate new individuals
7: **end while**
8: **return** Optimal solution

Thereby, the ego-vehicle controlled by a human driver in the highD data set is replaced by a simple and computationally cost efficient replica of an AV. This simplicity is for two reasons. First, the model has to be calculated often during optimization and second, the complexity metric is mainly independent of the ego-vehicle behavior. For the simulation of the longitudinal behavior the Intelligent Driver Model (IDM) [207] is used as a surrogate model according to Equation (4.6) and (4.7) with $a^*_{\mathrm{x,ego}}$ as the desired IDM acceleration in longitudinal direction, s the bumper to bumper distance between the leading vehicle x_{lead} and the ego-vehicle x_{ego} with length l_{ego}, v_{ego} as the actual ego velocity, Δv as the velocity difference of the leading vehicle v_{lead} to the ego-vehicle v_{ego}, $a_{\mathrm{x,max}}$ as the maximal allowed acceleration, v_0 as the desired velocity, δ as the acceleration exponent, s^* as the desired distance between both vehicles, s_0 as the linear jam distance, s_1 as the non-linear jam distance, T_{g} the desired time gap to the leading vehicle and $b_{\mathrm{x,com}}$ as the value of comfortable deceleration. The acceleration according to the IDM is calculated as

$$a^*_{\mathrm{x,ego}}(s, v_{\mathrm{ego}}, \Delta v) = a_{\mathrm{x,max}}\left[1 - \left(\frac{v_{\mathrm{ego}}}{v_0}\right)^{\delta} - \left(\frac{s^*(v_{\mathrm{ego}}, \Delta v)}{s}\right)^2\right] \tag{4.6}$$

with

$$\begin{aligned} s^*(v_{\mathrm{ego}}, \Delta v) &= s_0 + s_1\sqrt{\frac{v_{\mathrm{ego}}}{v_0}} + T_{\mathrm{g}} v_{\mathrm{ego}} + \frac{v_{\mathrm{ego}}\Delta v}{2\sqrt{a_{\mathrm{x,max}} b_{\mathrm{x,com}}}} \\ s &= x_{\mathrm{lead}} - x_{\mathrm{ego}} - l_{\mathrm{ego}} \\ \Delta v &= v_{\mathrm{lead}} - v_{\mathrm{ego}}. \end{aligned} \tag{4.7}$$

The lateral guidance is controlled by the so-called Minimizing Overall Braking Induced by Lane Changes (MOBIL) [208] approach. Here, lane changes are performed if the conditions from Equation (4.8) - (4.10) are fulfilled. A distinction is introduced between the symbols before ($\bullet$) and after ($\tilde{\bullet}$) the hypothetical lane change. The basic requirement for a lane change is that

$$|\tilde{a}_{\mathrm{follow,target}}| \leq |a_{\mathrm{safe}}|, \tag{4.8}$$

with the necessary deceleration of the following vehicle in the target lane $\tilde{a}_{\mathrm{follow,target}}$ and the safe deceleration a_{safe} which is defined as $8\,\mathrm{m/s^2}$.

If Equation (4.8) is fulfilled, the conditions in Equation (4.9) and (4.10) are checked and thus it is decided whether a lane change is performed. Therein, a_{ego} denotes the current ego acceleration, p a politeness factor that is set to 0.5 to balance egoistic and altruistic behavior of the ego-vehicle,

$a_{\text{follow,initial}}$ the acceleration of the following vehicle in the initial ego lane, δ_{th} a switching threshold which is set to $0.1\,\mathrm{m/s^2}$ and δ_{bias} the bias to account for the European keep-right directive which is set to $0.3\,\mathrm{m/s^2}$.

Condition 1: changing lane from left to the right:

$$a_{\text{ego}} + p(a_{\text{follow,initial}} + a_{\text{follow,target}}) > \tilde{a}_{\text{ego}} + p(\tilde{a}_{\text{follow,target}} + \tilde{a}_{\text{follow,initial}}) + \delta_{\text{th}} - \delta_{\text{bias}} \quad (4.9)$$

Condition 2: changing lane from right to the left:

$$\tilde{a}_{\text{ego}} + p(\tilde{a}_{\text{follow,target}} + \tilde{a}_{\text{follow,initial}}) > a_{\text{ego}} + p(a_{\text{follow,initial}} + a_{\text{follow,target}}) + \delta_{\text{th}} + \delta_{\text{bias}} \quad (4.10)$$

After this short introduction of the basics, the developed sub-method is explained in the next section.

4.3.2 Approach

This section first summaries the overall approach of the sub-method. Then, the individual method modules are outlined.

Overall approach

The primary goal of this sub-method is the development of a metric for evaluating the trajectories of the surrounding TPs in a scenario, so that the most complex scenarios can be extracted from real driving data. These are the most challenging and, according to the assumption in Section 3.2, lead more often to critical scenarios. The entire procedure is summarized in Figure 4.12. The part to the left of the database is used to develop and identify complex scenarios. The most important part is the complexity assessment, which is the focus of this sub-method. The part to the right of the database is used for the certification of different AVs and is not the focus of this work. However, this part is exemplary performed once for the validation of the complexity assessment method.

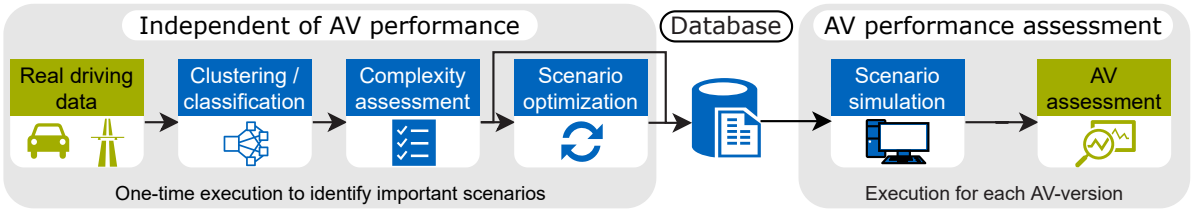
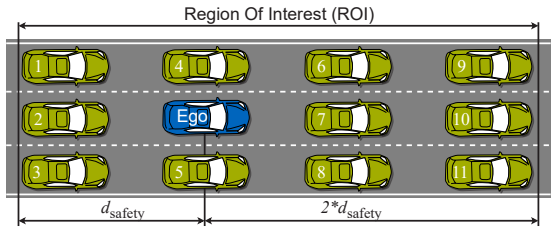


Figure 4.12: Overview of the traffic situation complexity sub-method (adapted from [16]).

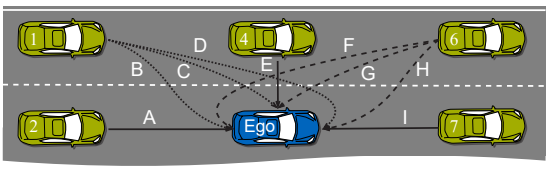
The scenario optimization step shown in Figure 4.12 is an optional step. Scenarios can be stored in the database both, directly after the complexity assessment and including the additional optimization. A short introduction to the overall approach is already given in Section 3.2. The sub-modules starting from the real driving data to the database are described in detail in the following subsections.

Before the individual modules are explained in more detail, two basic aspects are defined that affect the entire sub-method. This is the definition of an Region Of Interest (ROI), as well as the definition of the considered functional scenarios. The ROI describes the area around the ego-vehicle in which surrounding TPs must be located in order to be considered relevant for

the behavior of the ego-vehicle. For example, TPs that are too far away or on the structurally separated opposite roadway are not relevant. The definition of the ROI is based on [21] and is calculated in the x-direction using the legally specified and speed-dependent safety distance d_{safety}, which corresponds to a time gap of $1.8\,\text{s}$. In lateral direction, the lane of the ego-vehicle as well as an adjacent one to the left and to the right are included, if they exist. According to Figure 4.13a, if the safety distance is maintained and both adjacent lanes exist, a maximum of eleven TPs are within the ROI of the ego-vehicle.



(a) Definition of ROI for highway scenarios based on eleven surrounding TPs (adapted from [16, 21]) and the safety distance d_{safety} defined as a time gap of $1.8\,\text{s}$.



(b) Classification of functional scenarios based on the challenger trajectory. Only the left side is shown from the perspective of the ego-vehicle, because the trajectories and designations are symmetrical. The figure is based on [16, 76].

Figure 4.13: Definition of the ROI as well as the considered functional scenarios.

The given UNECE framework (Subsection 2.4.3, [11]) is structured according to functional highway scenarios. In order to address these scenarios, individual functional scenarios can be considered for each module of this sub-method. In this thesis, a classification of highway scenarios based on [76] as used in the PEGASUS project [10] is used. This classification has a higher information content than the functional scenarios of the UNECE framework. For example, the one cut-in scenario is subdivided into the functional scenarios with label D and H in Figure 4.13b.

Scenario clustering and classification

This module of the sub-method was supported by the student thesis of Breitfuß *[209].*

The individual steps of this module are shown in Figure 4.14. Starting from the continuous real driving data of the highD data set, the data is split into individual scenarios using hierarchical clustering. Subsequently, the maneuvers contained in the scenario are determined and checked whether a surrounding TP represents a challenger for the ego-vehicle. Finally, the scenarios containing a challenger are divided into the functional scenarios shown in Figure 4.13b according to the relative movement of the challenger. These scenarios can then be passed to the next module, where the complexity of the scenario is calculated.

Hierarchical clustering Before the actual clustering is started, the real data is unified. This means that one of the two driving directions is mirrored so that all vehicles drive in the same direction. For further standardization, all measurement recordings are shortened to $400\,\text{m}$ and the numbering of the lanes is standardized.

Hierarchical-Agglomerative Clustering (HAC) is a bottom-up method in which each data point first forms its own cluster. Using a distance measure, the similarity of individual clusters is calculated and all clusters that fall below a defined threshold of the distance measure are fused into one

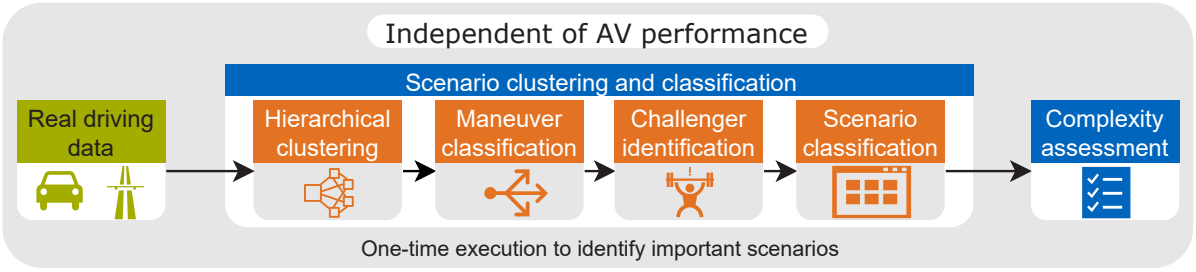


Figure 4.14: Overview of the scenario clustering and classification sub-sub-method (adapted from [16]).

cluster. An introduction to this technique can be found in [210]. For the clustering of the highD data, the HAC is adapted to the given problem. Thus, the distance between the vehicles is used as a distance measure and the limits of the ROI as a threshold value. According to Figure 4.13a a different threshold is thus chosen in front of and behind the vehicle. This is carried out for a considered vehicle (in the following called ego-vehicle) of the data set in each time step (frame) and all vehicles, which are part of the cluster of the ego-vehicle, are saved. Subsequently, all results of the individual frames are merged and this procedure is performed iteratively over all vehicles of the data set.

The result of this step is the assignment of the relevant TPs for a considered ego-vehicle. Thus, a maximum of as many scenarios can be extracted from the entire data set as the data set has vehicles. According to Table 4.4, this is approximately 110,000. If an ego-vehicle is alone in its cluster in all frames, then no other TP relevant to the ego-vehicle under consideration occurs. Consequently, the ego-vehicle is in an uninteresting free driving situation and the vehicle respectively scenario can be sorted out.

Maneuver classification A simple rule-based approach is used to determine the maneuvers because the driving situations on highways can still be handled with a relatively small number of rules. For a city-center application, this is difficult to perform and machine-learning methods should be preferred. The rule-based determination of maneuvers is done in two steps and is implemented in Stateflow from *The MathWorks*. First, the state between the ego-vehicle and the considered surrounding TP is determined in each frame. The eight basic states from the left hand side of Table 4.5 are used here. Furthermore, it is assigned to the states which of the two vehicles has the higher velocity and in which lane the TP is located in relation to the ego-vehicle.

In the second step, maneuvers are defined by the sequence of states respectively state transitions. This is done by searching the sequences of states according to previously defined patterns. The predetermined patterns thus represent the maneuvers. This is carried out with a trained decision tree. A total of 12 (right side of Table 4.5) high-level maneuver classes and 158 (Appendix Table A.1) maneuvers derived from them are defined. A sequence of maneuvers is also possible, which can be divided into separate maneuvers by a splitting step.

Challenger identification The next step is to examine each scenario to check whether it contains a challenger. Challenger in this context refers to a surrounding TP that requires a reaction from the ego-vehicle. If, for example, the ego-vehicle drives behind a TP and this TP starts to brake so that the ego-vehicle has to react within a defined period of time in order to prevent a potential collision, then this TP is classified as a challenger. However, if the TP drives

Table 4.5: Basic states (left) and maneuvers (right) between the ego-vehicle and a TP.

Name	Semantic description	Name	Semantic description
S1	TP in adjacent lane behind ego-vehicle	M1	Follow drive
S2	TP in adjacent lane next to ego-vehicle	M2	Passing
S3	TP in adjacent lane in front of ego-vehicle	M3	Parallel driving
S4	TP in same lane behind ego-vehicle	M4	Passing with cut-in in front of vehicle
S5	TP in same lane in front of ego-vehicle	M5	Follow drive with cut-out and passing vehicle
S6	Ongoing lane change of TP behind ego-vehicle	M6	Complete overtake maneuver
S7	Ongoing lane change of TP in front of ego-vehicle	M7	Passing slower vehicle and cut-in of this vehicle
S8	TP located in not relevant area (outside of ROI)	M8	Follow drive and cut-out of the leading vehicle
		M9	Passing of slower vehicle and cut-in behind this vehicle
		M10	Passing with lane change out of ROI
		M11	Lane change from out of ROI into ROI
		M12	Lane change of both vehicles

with constant velocity, or if the TP in front is even faster than the ego-vehicle, no reaction from the ego-vehicle is required and therefore the TP is not considered a challenger.

For this assessment, the movement of the ego-vehicle is predicted into the future using the current longitudinal and lateral acceleration. This is done for a defined period of time $t_{\mathrm{ch,predict}}$ in each time step of the scenario. This represents the behavior if the human driver respectively the AV does not perform any action. The predicted position of the ego-vehicle is compared with the actual position of the surrounding TPs at the predicted time. If a TP is located in the immediate vicinity of the ego-vehicle (Figure 4.15), it is considered a challenger.

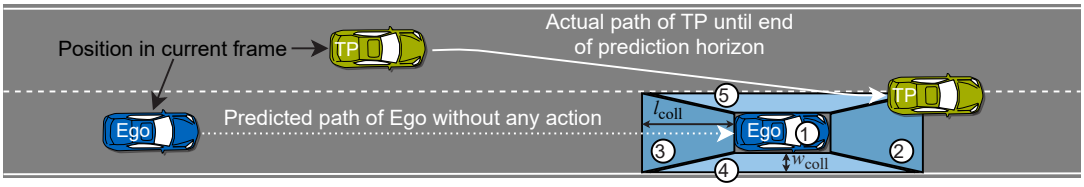


Figure 4.15: Prediction principle of the challenger identification. A TP is classified as challenger if the actual position of the TP from the highD data at the end of the prediction horizon is within an area of increased risk of collision (marked as 1 to 5) in the vicinity of the predicted position of the ego-vehicle at the end of the prediction horizon. The potential collision area is symmetrical and is described by their length l_{coll} and width w_{coll}.

In addition, the relative position of the challenger at the predicted time is stored based on five areas (Figure 4.15), such that this information can be used in the subsequent classification of scenarios based on challenger movement and collision potential. The size of the potential collision area, which is indicated symmetrically by the length l_{coll} and width w_{coll}, as well as the prediction horizon have a decisive influence on the number of challengers occurring. According to the German Motorway Construction Guideline [52], a lane width of $3.50\,\mathrm{m}$ is a frequently used standard width and is used here as a reference for determining w_{coll}. The collision potential area should fill the lane width when driving in the middle of its own lane. A vehicle width of $2.0\,\mathrm{m}$ is assumed, which results in $w_{\mathrm{coll}} = 0.75\,\mathrm{m}$. In contrast to w_{coll}, the length l_{coll} is not defined as a constant, but variable depending on the velocity of the ego-vehicle. As with ROI, the legally required safety distance is used as a reference value. The determination of a meaningful value for this parameter and the prediction duration is discussed in more detail in Subsection 4.3.3.

The decision of whether it is a challenger or not is determined based on the overlap of the TP with the potential collision area (Figure 4.15). This overlap represents a collision probability and the larger the area of overlap, the higher is the collision probability rated. As can be seen in Figure 4.15, a TP can overlap with several areas and in addition, it is possible that during the course of the scenario several different TPs are located within the potential collision area. Thus, two different strategies can be pursued to determine the challenger and the potential collision area. On the one hand, the first contact can be counted, i. e. the first TP that creates an overlap and the area with which it overlaps is the characterizing challenger. On the other hand, the maximum overlap can be used as a criterion. The difference between the two approaches is discussed in more detail in Subsection 4.3.3.

All scenarios that do not include a challenger do not require any action from the ego-vehicle, are thus uninteresting and can be sorted out. For all other scenarios that contain a challenger, the next step is to classify the scenarios.

Scenario classification In this step, information known from the previous steps is used to classify the challenger scenarios using the functional scenarios in Figure 4.13b. The first information required is the start frame of the prediction in which the characterizing challenger occurs. If the first contact principle is used, the start frame is relevant, in whose prediction an overlap of a TP with the potential collision area occurs for the first time. In addition, the localization of the overlap (one of the five defined areas from Figure 4.15) as well as the state (Table 4.5) of the start frame is required. Based on these three kinds of information, a simple decision tree can be created to classify the scenarios into the nine defined functional scenarios (Figure 4.13b). This means that the relative trajectory of the challenger is used to decide in which class a scenario is grouped. The movement and behavior of all other TPs of the scenario has no influence on the classification of the scenario.

The classified scenarios represent the output of the clustering and classification module. The next module calculates the complexity of the classified scenarios.

Complexity assessment

This module of the sub-method was supported by the student theses of Yu *[211, 212].*

This module determines the complexity of each challenger scenario. For this purpose, a complexity metric is developed that fulfills this task. Based on the related work in Subsection 4.3.1 the following basic aspects of complexity are defined.

- The complexity is an objective measure to calculate the difficulty of a scenario for the planning module of an AV (not for a human driver).
- Complexity assesses the trajectories / time dependent behavior of the surrounding TPs.
- There are different influencing factors / attributes that contribute to the complexity.
- Different factors have different importance.
- Coupling effects between factors are neglected.
- Every attribute has linear contribution to the overall complexity.

- Complexity evaluates the scenario (as far as possible) independently of the ego-vehicle behavior.
- All surrounding TPs within the ROI are relevant for the calculation of the complexity.
- The value range of the complexity should be normalized to values between zero and one.

The core of the method is the definition of the influencing factors that are used to objectively determine the complexity of a scenario. The basis for the influencing factors are the eight subjectively described and evaluated factors of Schuldt [204, chap. 2.3.5] from Subsection 4.3.1. Based on this, 13 factors were derived with the help of discussions with experts from industry and academia to describe the complexity of a scenario in relation to an AV objectively. These are summarized in Table 4.6 and are described in more detail below. A comprehensive validation of whether these factors are suitable and sufficient to describe the complexity is analyzed in Subsection 4.3.3 using an online expert questionnaire.

Table 4.6: Overview of the 13 defined factors influencing the complexity. Factor one to eleven can be calculated in every frame of the scenario. Factors twelve and 13 are only calculated for the whole scenario.

Nr.	Influence factor
1	Types of surrounding TPs
2	Number of surrounding TPs
3	Number of connections between all vehicles (Connectivity)
4	Dynamics of surrounding TPs
5	Variation of dynamical parameters of the surrounding TPs
6	Predictability (with a simple constant acceleration model) of future behavior of surrounding TPs
7	Number of possible actions of the ego-vehicle (due to action restriction by other TPs)
8	Number of possible actions of the surrounding TPs (due to action restriction by other TPs)
9	Time-gap between ego-vehicle and surrounding TPs
10	Time-to-Brake between ego-vehicle and surrounding TPs
11	Occluded blind-spot area from the ego-vehicle perspective
12	Number of actions of the ego-vehicle performed in the scenario
13	Number of actions of the surrounding TPs performed in the scenario

Factor 1: Types of surrounding TPs It is determined how many different vehicle types are within the ROI at each time step of the scenario. Vehicle types can be cars, trucks, motorcycles and similar. Each of these vehicle types has its own characteristic driving behavior and dynamics. For example, motorcycles are much more dynamic than trucks, which must be taken into account when planning a safe trajectory of the vehicle. The higher the number of different vehicle types, the more complex the scenario is. Because only cars and trucks appear in the utilized highD data set, the value two is used to normalize this factor.

Factor 2: Number of surrounding TPs It is determined how many TPs are within the ROI at each time step of the scenario. The ego-vehicle is not counted. The more vehicles are within the ROI, the more complex the scenario is considered. According to the definition of the ROI (Figure 4.13a) and in compliance with the required safety distance, a maximum of eleven vehicles can be positioned within the ROI in a frame, which is why this value is used to normalize this factor. However, it can be observed here that the normalization to a value between zero and one cannot be guaranteed. If TPs do not obey the safety distance, there can be more than eleven vehicles within the ROI. A physical maximum that can be used as a normalization factor

is therefore not available. This aspect occurs with several factors, which is why it cannot be guaranteed that the total complexity will not exceed the value 1 in rare cases.

Factor 3: Connectivity This factor is used to express the dependence of the behavior on neighboring vehicles. This means that the ROI is divided into twelve areas (eleven for the surrounding TPs and one area for the ego-vehicle), as shown in Figure 4.16 using the dashed white lines. The lateral boundaries of the individual areas are the lane markings. In the longitudinal direction, the boundaries are on the one hand the boundaries of the ROI and on the other hand, the additionally inserted white vertical dashed lines. The determination of the connections is determined by the occupied areas. Occupied means that a TP is present in the corresponding area. To calculate the connections, the number of arrows (Figure 4.16) for occupied areas is determined and normalized with the maximum number of connections (21).

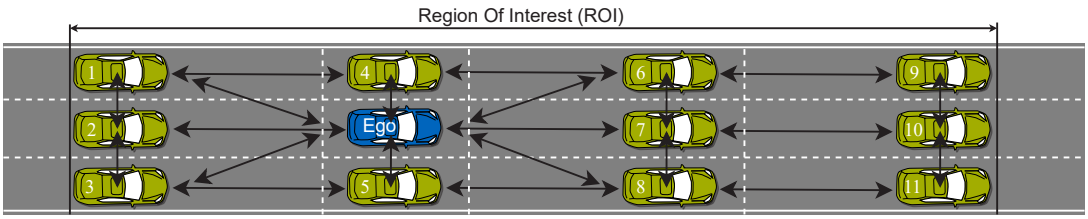


Figure 4.16: Number of connections between all vehicles. Every arrow represents one connection. The maximum number of connections is 21, which is also the number used for normalization.

This factor thus represents the mutual impact of the change in behavior of a TP. A scenario with three participating vehicles can be connected to different degrees depending on how these three vehicles are arranged in the scenario. If a scenario has many connections, on the one hand, there are action restrictions for other TPs and the ego-vehicle, and on the other hand the behavior of one TP affects more vehicles and chain reactions can occur. The more connections there are, the more complex it is assumed for the trajectory planning module to plan a safe trajectory. Since the focus is on the ego-vehicle, the diagonal connections are only counted for the ego-vehicle (Figure 4.16).

Factor 4: Dynamics of surrounding TPs With this factor, the dynamic values of the surrounding TPs are evaluated. The higher the occurring velocities and accelerations, the more complex the scenario is considered. Depending on the relative position of the TP to the ego-vehicle according to Figure 4.13a, the contributions of the TPs are weighted differently. Highly weighted are all vehicles that have the highest influence on planning a safe trajectory of the ego-vehicle. In the longitudinal direction, vehicles that are behind the ego-vehicle (positions 1 to 3) and have a higher velocity than the ego-vehicle are weighted high. Slower vehicles are weighted accordingly if they are in front of the ego-vehicle (positions 6 to 11). In the lateral direction, vehicles to the left of the ego-vehicle having a lateral speed to the right and vice versa are given a high weighting. Again, there is no physical maximum value for the normalization of this factor. Therefore, the following experience-based values derived from the highD data set are used: $dynamic_{\mathrm{v,x}} = 35\,\mathrm{m/s}$, $dynamic_{\mathrm{v,y}} = 0.65\,\mathrm{m/s}$, $dynamic_{\mathrm{a,x}} = 0.65\,\mathrm{m/s^2}$ and $dynamic_{\mathrm{a,y}} = 0.22\,\mathrm{m/s^2}$.

Factor 5: Variation of dynamical parameters of the surrounding TPs This factor considers the differences in the dynamic state of the surrounding TPs. It is assumed that it is more difficult for the ego-vehicle to plan a safe trajectory if the scenario includes slow as well as fast TPs and

the occurring accelerations vary as much as possible. Thus, a scenario in which all vehicles move at the same constant velocity is considered particularly simple. In the calculation, the lowest value that occurred in each frame of the scenario is subtracted from the highest value. For this factor, normalization is also based on the highD data set. The following values are used in this work: $variation_{\mathrm{v,x}} = 15\,\mathrm{m/s}$, $variation_{\mathrm{v,y}} = 6\,\mathrm{m/s}$, $variation_{\mathrm{a,x}} = 12\,\mathrm{m/s^2}$ and $variation_{\mathrm{a,y}} = 6\,\mathrm{m/s^2}$.

Factor 6: Predictability of future behavior of surrounding TPs To plan a safe trajectory, it is necessary to predict how the surrounding TPs will behave in the future. The better this is achieved, the easier it is to plan a safe trajectory. In other words, the more difficult it is to predict the behavior of a TP, the more difficult it is to plan a safe trajectory. Prediction can be made using simple models such as those based on the assumption of constant velocity or acceleration, or more sophisticated models based on machine learning. Based on literature [213, 214], the accuracy of prediction for highway scenarios using simple models is only slightly worse, so a simple constant acceleration model is used here. This model uses the current longitudinal and lateral acceleration to predict the position of the TP at a future time. The prediction time t_{predict} is chosen depending on the velocity of the ego-vehicle and is calculated as

$$t_{\mathrm{predict}} = \frac{v_{\mathrm{ego}}}{a_{\mathrm{x,max,dec}}}, \tag{4.11}$$

and represents the time the ego-vehicle needs to brake to a standstill with a maximum deceleration $a_{\mathrm{x,max,dec}}$ of $10\,\mathrm{m/s^2}$.

The predicted position of a TP in x- and y-direction can therefore be calculated using a simple point mass model via

$$\begin{bmatrix} x_{\mathrm{predict}} \\ y_{\mathrm{predict}} \end{bmatrix} = \begin{bmatrix} x_0 + v_{\mathrm{x},0} t_{\mathrm{predict}} + \frac{1}{2} a_{\mathrm{x},0} t_{\mathrm{predict}}^2 \\ y_0 + v_{\mathrm{y},0} t_{\mathrm{predict}} + \frac{1}{2} a_{\mathrm{y},0} t_{\mathrm{predict}}^2 \end{bmatrix}, \tag{4.12}$$

with the initial position (x_0, y_0), velocity v_0 and acceleration a_0 of the TP.

The overall deviation $\Delta d_{\mathrm{overall}}$ of the prediction is calculated as the euclidean distance of the predicted positions to the actual positions $(x_{\mathrm{highD}}, y_{\mathrm{highD}})$ of the highD data at the end of the prediction time as

$$\Delta d_{\mathrm{overall}} = \sqrt{\left(x_{\mathrm{predict}} - x_{\mathrm{highD}}\right)^2 + \left(y_{\mathrm{predict}} - y_{\mathrm{highD}}\right)^2}. \tag{4.13}$$

If there are several TPs, the average of the overall deviation $\Delta d_{\mathrm{overall}}$ over all TPs is calculated. Again, for this factor there is no physical upper limit that can be used for normalization. Therefore, a reference value of $1.4\,\mathrm{m}$ is used based on experience.

Factor 7 and 8: Possible actions of ego-vehicle / TPs The actions of both, the ego-vehicle and the surrounding TPs are are based on the same principle and therefore they are explained together. The basic principle of these factors is that a TP or the ego-vehicle can experience action restrictions due to the presence of other vehicles and therefore can only perform a certain number of actions. The possible actions are explained here from the perspective of the ego-vehicle and with regard to the definition of the ROI and its division into twelve areas from Figure 4.13a and 4.16. An action is the hypothetical change of the position of the ego-vehicle to one of the surrounding areas. Thus, only the eight surrounding areas of the ROI (areas one

to eight) are considered and therefore a maximum of eight actions are possible. For example, it is assumed that a lane change to the left is only possible if area four is not occupied. If it is determined which actions a surrounding TP can perform, it is temporarily considered as the ego-vehicle and the possible actions are determined from its point of view. An overview of the possible actions is given in Table 4.7.

Table 4.7: Overview of the possible actions. If an area is not occupied, this means that the vehicle with the corresponding number in Figure 4.13a and 4.16 is not present. A necessary criterion for lane changes is that the adjacent lane exists.

Nr.	Action	Occupancy of surrounding areas according to Figure 4.13a and 4.16
1	Deceleration	Assumed to be always possible
2	Acceleration	Area 7 is not occupied
3	Lane change to the left	Area 4 not occupied and vehicle 1 not faster than the ego-vehicle
4	Acceleration after lane change to the left	Action 3 possible and area 6 not occupied
5	Deceleration after lane change to the left	Action 3 possible and area 1 not occupied
6	Lane change to the right	Area 5 not occupied and vehicle 3 not faster than the ego-vehicle
7	Acceleration after lane change to the right	Action 6 possible and area 8 not occupied
8	Deceleration after lane change to the right	Action 6 possible and area 3 not occupied

Since the test scenarios to be defined are intended to test the capabilities of the ego-vehicle, it is assumed for the ego-vehicle that braking actions are always possible. This results in a smaller number of possible actions. In addition, the normalization is handled differently for these two factors. From the point of view of the ego-vehicle, it is assumed most difficult for trajectory planning if an intermediate number of actions is possible. If no action is possible, the ego-vehicle can only brake and the decision is easy. Likewise, if all actions are possible, then there are no surrounding TPs and it has no effect which action is chosen. The progression between the explained extreme values is assumed linear. For the surrounding TPs, the effect on the ego-vehicle is considered. The more actions are possible for a surrounding TP, the more difficult it is for the ego-vehicle to predict which action will be executed next by a TP. Therefore, the reference value eight is used for normalization. If several surrounding TPs exist, the average of all TPs is used.

Factor 9: Time-gap between ego-vehicle and surrounding TPs This factor represents the precision with which an action of the ego-vehicle must be performed. The smaller the time gap $t_{\mathrm{gap},i}$ to surrounding TPs, the more precisely an action of the ego-vehicle has to be planned and executed. This factor therefore addresses both, the planning and the execution module of the AV. The time gap $t_{\mathrm{gap},i}$ is calculated from the perspective of the ego-vehicle for a subset i of the surrounding TPs via

$$t_{\mathrm{gap},i} = \frac{d_i}{v_{\mathrm{ego}}} \qquad \text{with} \quad i \in \{1,3,6,7,8\}, \tag{4.14}$$

where d_i is the distance to traffic participant i (Figure 4.13a) and is defined as the distance between the front of the rear vehicle and the rear of the front vehicle. Here, only a relevant subgroup of the surrounding TPs is considered. These are the TPs driving directly in front of the ego-vehicle and the potential overtaking vehicles on the adjacent lanes. If several relevant TPs exist, the average value of these vehicles is calculated. The physical minimum value that can occur is zero. Therefore this value should result in the maximum value of one after normalization.

In order that small t_{gap} have a larger effect, no linearly decreasing course is used for the normalized factor $t_{\text{gap,normed}}$, but an exponentially decreasing course according to:

$$t_{\text{gap,normed}} = e^{-0.5 t_{\text{gap}}}. \tag{4.15}$$

The choice of the parameters of the e-function is based on experience.

Factor 10: Time-to-Brake between ego-vehicle and surrounding TPs This factor represents the required reaction time of the ego-vehicle. Even though AVs tend to have shorter reaction times than human drivers do, a certain amount of time is still required for the detection and tracking of objects as well as for planning a trajectory. The shorter the required reaction time, which is described by the Time-to-Brake (TTB), the more difficult it is for the ego-vehicle to calculate a safe trajectory. As a simplification, no steering intervention is considered. A meaningful calculation of the TTB exists only for a vehicle directly ahead, which is why only vehicle 7 from the definition of the ROI in Figure 4.13a is considered here. If the ego-vehicle is moving at a higher speed than vehicle 7, then the distance d_{dec}, which the ego-vehicle needs to decelerate with maximum deceleration $a_{\text{x,max,dec}}$ to the speed of vehicle 7 v_7 is defined as:

$$d_{\text{dec}} = \begin{cases} 0, & \text{for } v_7 \geq v_{\text{ego}} \\ \frac{v_{\text{ego}}^2 - v_7^2}{2 \cdot |a_{\text{x,max,dec}}|}, & \text{else.} \end{cases} \tag{4.16}$$

Then the required distance can be subtracted from the actual distance d_7 and the TTB t_{ttb} can be calculated as:

$$t_{\text{ttb}} = \frac{d_7 - d_{\text{dec}}}{v_{\text{ego}}}. \tag{4.17}$$

The value for t_{ttb} can be negative, which means that the ego-vehicle can no longer prevent a critical situation by a braking maneuver alone. For the normalization of this factor, reference is made to Junietz [20, p. 83]. In his dissertation, he also uses the TTB for the evaluation of scenarios for AVs, thereby using a reference value of $2\,s$ and a linearly decreasing progression. The reference value is determined on the basis of the traffic events of human drivers due to a lack of data for AVs. It is handled analogously in the present work, which yields to the normalized TTB $t_{\text{ttb,normed}}$:

$$t_{\text{ttb,normed}} = \begin{cases} 1, & \text{for } t_{\text{ttb}} \leq 0 \quad \text{and } v_7 \leq v_{\text{ego}} \\ 1 - \frac{t_{\text{ttb}}}{2} & \text{for } 0 < t_{\text{ttb}} < 2 \text{ and } v_7 \leq v_{\text{ego}} \\ 0, & \text{else.} \end{cases} \tag{4.18}$$

Factor 11: Occluded blind-spot area from the ego-vehicle perspective This factor describes the non-transparency of the scenario for the ego-vehicle, which is described by the size of the occluded area from the perspective of the ego-vehicle. These are blind spots that cannot be seen by the ego-vehicle's environmental sensors because they are occluded by other vehicles. This is illustrated by the black area in Figure 4.17. It is assumed that there is no common environment model of all TPs through vehicle-to-vehicle communication. This is realistic because the market penetration of the vehicles with which this is possible will be low in the first years after market launch. The larger the areas occluded by surrounding TPs, the greater

the uncertainty in planning a safe trajectory. This is seen as an increased difficulty because the probability of unexpected situations increases.

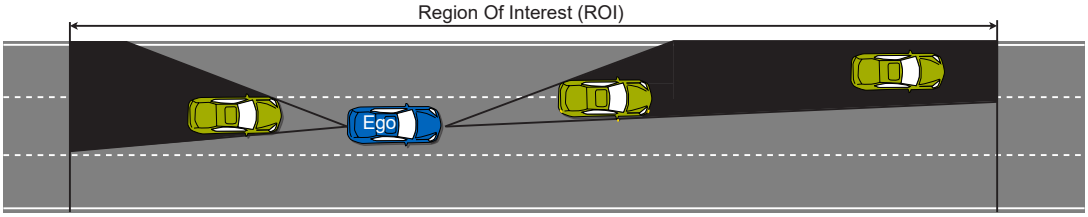


Figure 4.17: Visualization of the blind spot (marked in black) calculation. The blind spots are only considered within the ROI.

For a simplified and computationally efficient calculation of the blind spot, several assumptions are made. For example, the actual used sensor positions and orientations of the ego-vehicle under investigation are not used, but for simplification, it is assumed that the sensors are concentrated on two points (middle of the front and rear of the vehicle) according to Figure 4.17. Starting from these points, the blind spots are calculated based on beam propagation. Consequently, only the area within the ROI is considered. For the normalized size of the blind spot $A_{\text{blindspot,normed}}$, the calculated size of the blind spot $A_{\text{blindspot}}$ is divided by the size of the ROI A_{ROI} according to:

$$A_{\text{blindspot,normed}} = \frac{A_{\text{blindspot}}}{A_{\text{ROI}}}. \tag{4.19}$$

Factor 12 and 13: Number of actions of the ego-vehicle / TPs performed in the scenario
All factors introduced so far can be evaluated in each scene of a scenario. The last two factors, which are explained again together because of the same principle, can only be calculated at the end of the scenario for the overall scenario. After assessing with the factors 7 and 8 how many actions are possible in the current scene, the number of actions performed in a scenario is assessed here. The basic idea here is that the more actions are performed by surrounding TPs and the more actions are required by the ego-vehicle, the more difficult it is for the ego-vehicle to plan a safe trajectory. Thereby, it is distinguished between actions in longitudinal and lateral direction. An action in longitudinal direction is evaluated as a change in acceleration, unlike with factors 7 and 8. The concept is based on the work of [215], in which different actions are defined based on the intensity of the acceleration. In the present work, we use the states according to Table 4.8 and an action is defined as a change of the state.

Table 4.8: Definition of states in dependency of the intensity of the longitudinal acceleration a_x. An action in longitudinal direction is defined as the transition between two states.

Range of a_x	Driving state
$a_x \leq -6.0\,\mathrm{m/s^2}$	Emergency brake
$-6.0\,\mathrm{m/s^2} < a_x \leq -3.0\,\mathrm{m/s^2}$	Strong deceleration
$-3.0\,\mathrm{m/s^2} < a_x \leq -0.2\,\mathrm{m/s^2}$	Normal deceleration
$-0.2\,\mathrm{m/s^2} < a_x \leq +0.2\,\mathrm{m/s^2}$	Constant driving
$+0.2\,\mathrm{m/s^2} < a_x \leq +2.0\,\mathrm{m/s^2}$	Normal acceleration
$a_x > +2.0\,\mathrm{m/s^2}$	Strong acceleration

The highD data used are real driving data. To compensate for measurement fluctuations and not to count short-term changes, an average of the longitudinal acceleration over ten time steps, which corresponds to a time of $0.4\,\mathrm{s}$, is used. Only when this average indicates a change in the

driving status, an action in the longitudinal direction is counted. The highD data set contains the trajectories of vehicles controlled by human drivers and thus includes small fluctuations in constant driving. To ensure that these slightly changing positive and negative accelerations a_{x} are not counted as actions, the range of constant driving is set to $\pm 0.2\,\mathrm{m/s^2}$.

In contrast to longitudinal control, the number of actions in the lateral direction is not determined by the intensity of the acceleration, but the number of actions is determined by the number of lane changes performed by the vehicle under consideration. This number can be determined simply by comparing the lane ID stored in the highD data. If this number changes, the vehicle has changed lanes and an action is counted.

There are no physical upper limits available for the normalization of these factors, thus values based on experience are used. As already mentioned, these two factors are not calculated for every scene of a scenario but can only be calculated for the entire scenario. For the longitudinal direction, a reference value is calculated based on the number of scenes. Thus, for a scenario duration per 50 scenes, which corresponds to $2.0\,\mathrm{s}$ for the highD data, one action is used as a reference value. No additional normalization is used for the lateral actions because not many lane changes occur due to the comparatively short scenario duration compared to the duration of a lane change. The mean value of the actions in longitudinal and lateral direction is then calculated and considered as the final value for the number of actions of the ego-vehicle. For the number of actions of the surrounding TPs, additionally the average over all existing TPs is calculated. For a scene-based consideration of complexity, the calculated values of these factors are applied to all scenes.

Complexity metric On the basis of the 13 defined factors, the developed metric is designed to objectively assess the complexity of scenarios. The aim is to make the evaluation as independent as possible of the behavior of the ego-vehicle. Because reactions to the actions of other road users are always carried out in the progress of a scenario, this is not fully possible, but the 13 defined factors enable this as far as possible. As mentioned at the beginning of this subsection, the complexity metric is defined as a linear combination of these factors which are aggregated in the vector $\mathbf{c}_{\mathrm{factors}}$. These should be weighted differently using the weighting factors $\mathbf{c}_{\mathrm{weighting}}$. The general calculation of the complexity of a scene $C_{\mathrm{scene},i}$ in a scenario with n_{frames} frames is given by

$$C_{\mathrm{scene},i} = \mathbf{c}_{\mathrm{weighting}}^{T} \cdot \mathbf{c}_{\mathrm{factors},i} \qquad \text{for} \quad i \in \{1, 2, ..., n_{\mathrm{frames}}\}, \tag{4.20}$$

whereby $\mathbf{c}_{\mathrm{weighting}}, \mathbf{c}_{\mathrm{factors},i} \in \mathbb{R}^{n_{\mathrm{factors}} \times 1}$ with $n_{\mathrm{factors}} = 13$.

In order that a structured determination of the weighting factors can be carried out in the further course of the work and in order to limit the values of the complexity of a scene $C_{\mathrm{scene},i}$, it is defined that the sum of all weighting factors $c_{\mathrm{weighting},j}$ results in

$$\sum_{j=1}^{n_{\mathrm{factors}}} c_{\mathrm{weighting},j} = 1. \tag{4.21}$$

In addition, the value range for the individual complexity factors $c_{\mathrm{factors},i,j}$ is normalized as

$$c_{\mathrm{factors},i,j} \in [0, 1], \tag{4.22}$$

even if compliance with the upper limit cannot be proven, which is already discussed in the introduction of the individual factors.

Because even a short difficult part of a scenario can cause a critical situation, the maximum complexity of a scene within each scenario is used as the descriptive complexity of a scenario C_{scenario} according to:

$$C_{\text{scenario}} = \max_i(C_{\text{scene},i}). \tag{4.23}$$

Complexity calculation For the complexity assessment, only all classified scenarios that contain a challenger are considered. For all scenarios, the complexity in each scene (frame) is calculated using the 13 defined and weighted factors and the result is stored in the scenario data. In addition, the number of the scene in which the highest complexity occurred is saved.

Scenario optimization

This module of the sub-method was supported by the student thesis of Lin *[216].*

This module represents an optional step of the tool chain. If this step is not performed, the scenarios with the highest complexity can be selected for the final database after calculating the complexity score. However, if this module is executed, the behavior of the surrounding TP is adjusted to maximize the calculated complexity of the scenario. A simplified visualization of the procedure is depicted in Figure 4.18.

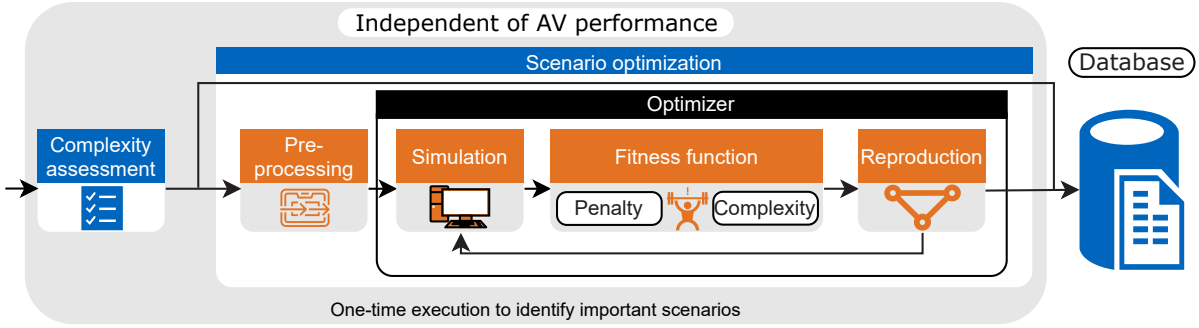


Figure 4.18: Visualization of the optional scenario optimization module. The core element of this module is the fitness function, which consists of a penalty function and the complexity metric described above.

The individual blocks of this module are described in the following. Thereby, the focus is on the fitness function, which is the core of this module.

Preprocessing The preprocessing step serves as preparation for the optimization. Thereby, any desired challenger scenario is loaded. This means that information about the number of frames of the scenario n_{frames}, the number of lanes, their width as well as the position, velocity, acceleration, orientation, width, height and vehicle type of the ego-vehicle and the surrounding TPs is provided. Only the starting state consisting of position, speed, acceleration and orientation of the ego-vehicle is used and the trajectory from the highD data, which was driven by a human driver, is replaced by the combination of IDM and MOBIL from Subsection 4.3.1.

This is necessary because a scenario always consists of actions and reactions of the participating vehicles and therefore the ego-vehicle must also be able to adapt its actions during the

optimization process. Furthermore, the IDM and MOBIL model eliminates the influence of the human driver. However, it cannot be guaranteed that these models behave identically to the human driver from the original highD data set. Since the complexity is defined largely independent of the behavior of the ego-vehicle, the complexity of the scenario changes only slightly due to a different ego-vehicle behavior, which is discussed in more detail in Subsection 4.3.3. Nevertheless, such a change can always occur when an identical scenario is tested for different AVs. Each of these AVs can behave differently and influence the outcome of the scenario. With the developed complexity metric, the difficulty of the scenario can be described independently of this behavior.

During the optimization process, the behavior of the surrounding TPs is adjusted to maximize the complexity of the scenario. The adaptation of the behavior is done by varying the accelerations of the TPs in x- and y-direction (a_x and a_y). These two variables are adjusted in each frame of the scenario for each TP. Thereby, lower and upper allowed bounds for the acceleration values are used. These are for passenger cars $-9\,\mathrm{m/s^2}$ and $+3\,\mathrm{m/s^2}$ in longitudinal and $\pm 3\,\mathrm{m/s^2}$ in lateral direction and for trucks $-7\,\mathrm{m/s^2}$ and $+1\,\mathrm{m/s^2}$ in longitudinal and $\pm 1\,\mathrm{m/s^2}$ in lateral direction. To reduce the computational effort, not every frame of the original highD data is used in the optimization, but only every n_{ff}-th frame. The size of the time steps in the optimization Δt_{opt} thus also increases by the factor n_{ff} compared to the time step $\Delta t_{\mathrm{highD}}$ of the highD data which is $0.04\,\mathrm{s}$. The number of optimization variables n_{var} finally results as:

$$n_{\mathrm{var}} = 2 n_{\mathrm{TPs}} n_{\mathrm{frames,opt}} \qquad \text{with} \quad n_{\mathrm{frames,opt}} = \left\lfloor \frac{n_{\mathrm{frames}}}{n_{\mathrm{ff}}} \right\rfloor, \tag{4.24}$$

where n_{TPs} is the number of traffic participants in the scenario and $n_{\mathrm{frames,opt}}$ is the number of frames used in the optimization of the scenario that has n_{frames} frames in the original highD data.

The vector of the optimization variable $\mathbf{y}_{\mathrm{opt}}$ contains all acceleration values in both directions, of all surrounding TPs and of all considered frames and is represented as:

$$\mathbf{y}_{\mathrm{opt}} = \begin{bmatrix} a_{\mathrm{x},1,1} & a_{\mathrm{y},1,1} & a_{\mathrm{x},2,1} & a_{\mathrm{y},2,1} & \cdots & a_{\mathrm{x},n_{\mathrm{TPs}},n_{\mathrm{frames,opt}}} & a_{\mathrm{y},n_{\mathrm{TPs}},n_{\mathrm{frames,opt}}} \end{bmatrix}^T \in \mathbb{R}^{n_{\mathrm{var}} \times 1}. \tag{4.25}$$

Thereby, each entry has three indices according to the scheme $a_{\mathrm{direction,TP\ number,frame\ number}}$.

For the initialization of the genetic optimization algorithm, the acceleration values from the highD data are used as initial population and forwarded to the optimizer. The optimization is performed using the genetic algorithm of the global optimization toolbox in Matlab [217] which was introduced in Subsection 4.3.1. The next section describes the simulation of the scenarios within the optimizer.

Simulation At the beginning of each optimization iteration the scenarios of the single individuals of the population are simulated. This means that each individual of the population corresponds to an acceleration vector according to Equation (4.25) and thus as many simulations have to be performed as there are individuals in a population. For each TP, the acceleration values must be transformed into a trajectory. Because this has to be done many times in the optimization, a computationally efficient extended nonlinear point mass model is used for each TP of the scenario, which is summarized as:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \qquad \text{with} \quad \mathbf{x} = \begin{bmatrix} x & y & v & \psi \end{bmatrix}^T \qquad \text{and} \quad \mathbf{u} = \begin{bmatrix} a_{\mathrm{x}} & a_{\mathrm{y}} \end{bmatrix}^T \tag{4.26}$$

as well as

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{v} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} v \cdot \cos(\psi) \\ v \cdot \sin(\psi) \\ a_{\mathrm{x}} \\ \frac{a_{\mathrm{y}}}{v} \end{bmatrix}. \tag{4.27}$$

The state vector $\mathbf{x}$ consists of the position (x, y) of the vehicle in global x- and y-direction, the velocity v and the yaw angle ψ of the TP. As already introduced in the previous section, the acceleration in x- and y-direction is used as input vector. For the simulation, the time-discrete form of this model is used:

$$\begin{bmatrix} x_t \\ y_t \\ v_t \\ \psi_t \end{bmatrix} = \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ v_{t-1} \\ \psi_{t-1} \end{bmatrix} + \begin{bmatrix} \cos(\psi_{t-1}) \cdot v_{t-1} \\ \sin(\psi_{t-1}) \cdot v_{t-1} \\ a_{\mathrm{x},t} \\ \frac{a_{\mathrm{y},t}}{v_{t-1}} \end{bmatrix} \cdot \Delta t_{\mathrm{opt}}. \tag{4.28}$$

Thus, the state vector at the current time t is calculated based on the values of the previous time step $t-1$ and the inputs $\mathbf{u}$ at the current time step t.

Based on the trajectories of the surrounding TPs, the trajectory of the ego-vehicle is calculated using the IDM and MOBIL model to calculate the longitudinal and lateral acceleration with a subsequent application of the extended point mass model from Equation (4.28). Thus, all trajectories are available for all scenarios (individuals) and next the fitness function can be calculated for all individuals. The next paragraph explains how this function is defined.

Fitness function The fitness function J_{GA} is the function that the optimizer tries to minimize. It consists of the complexity metric C_{GA} and a penalty function P_{GA}. The latter is used to prevent certain resulting behaviors, such as accidents between surrounding TPs. Since an optimization problem is formulated by default as a minimization problem and in the present case P_{GA} is to be minimized and C_{GA} maximized, the fitness function J_{GA} results in:

$$J_{\mathrm{GA}} = P_{\mathrm{GA}} - C_{\mathrm{GA}}. \tag{4.29}$$

In Equation (4.29), not the maximum occurring complexity C_{scenario} is used as descriptive complexity, but the average complexity of all scenes of the scenario C_{GA} according to:

$$C_{\mathrm{GA}} = \frac{1}{n_{\mathrm{frames,opt}}} \sum_{i=1}^{n_{\mathrm{frames,opt}}} C_{\mathrm{scene},i}. \tag{4.30}$$

The mean value of complexity over all scenes of the scenario is used in J_{GA} in order to increase the difficulty of the whole scenario and not only of one time step. The reason for the change from maximum to mean complexity is that the application has shown that more realistic scenarios can be achieved using mean complexity, since the focus is not only on one situation within the scenario. The mean complexity is only used within the optimizer. As with all scenarios from the highD data set in the previous complexity assessment step, the maximum complexity is used as the descriptive characteristic for evaluating the optimized scenarios.

The penalty function P_{GA} is used to ensure that defined boundary conditions for the behavior of the surrounding TPs are met during optimization. These boundary conditions are divided according to the descriptive variable into the three areas jerk P_{jerk}, velocity P_{vel} and position P_{pos}. Thereby, P_{pos} is again divided into three different aspects. The entries of the penalty function are calculated in each frame and summed up over all frames $n_{\mathrm{frames,opt}}$ of the scenario according to:

$$P_{\mathrm{GA}} = \sum_{i=1}^{n_{\mathrm{frames,opt}}} \left(P_{\mathrm{jerk},i} + P_{\mathrm{vel},i} + \sum_{j=1}^{3} P_{\mathrm{pos},j,i} \right). \tag{4.31}$$

Thus, a total of five different aspects are defined, causing an increase of the penalty. These are explained in more detail below and summarized in Table 4.9.

Table 4.9: Overview of the five parts of the penalty function P_{GA}.

Aspect	Penalty if	Description
Jerk	$j_x > 10\,\mathrm{m/s^3}$; $j_y > 10\,\mathrm{m/s^3}$	Physically motivated limitation of jerk
Vel	$v < 0\,\mathrm{m/s}$	Reverse driving is not allowed
Pos 1	(x, y) outside road boundaries	Prevent vehicles from leaving the road
Pos 2	$t_{\mathrm{laneMarking}} > 3\,\mathrm{s}$	Driving between lanes is not allowed
Pos 3	Vehicles overlapping	Prevent collisions between vehicles

The penalty in relation to the jerk of the surrounding TPs is physically motivated. Driving dynamics does not allow the acceleration to change too much between two time steps. In order to prevent this, a penalty is added if the jerk is too high and the fitness function J_{GA} therefore deteriorates. For simplicity, the jerk limit is set identically for all vehicle types.

If the optimizer determines permanently negative values for the longitudinal acceleration of a TP, then its velocity can become negative and the corresponding TP drives in the opposite direction. To prevent this, a penalty is inserted which takes effect as soon as the velocity of a TP is negative.

Regarding the position of the surrounding TPs there are three aspects that are considered in the penalty function. First, the surrounding TPs are not allowed to leave the road. Second, the TPs should drive within one lane and not permanently between two lanes. This is determined by the time a TP drives over a lane marking. Since lane changes are still allowed and also encouraged, the penalty is introduced from a time of $3\,\mathrm{s}$ upwards. The value of $3\,\mathrm{s}$ is based on [218], whose work investigates the duration of lane changes on motorways. The value is based on the median in free-flowing traffic, and the duration of a lane change is simplified to be equal to the time the vehicle is located over the lane marking. As the last part of the penalty function, accidents between the surrounding TPs are penalized. This is calculated based on an overlap of the vehicle shapes, which are assumed to be rectangles. In addition, collisions between the ego-vehicle and surrounding TPs are also penalized if the TP under consideration drives with higher speed into the rear of the ego-vehicle. This can happen because the surrounding TPs have no intelligence and do not brake due to other objects. If accidents occur in which the ego-vehicle crashes into a TP because it has made a wrong decision or cannot brake in time, no penalty is added. In other words, in a simplified form, the blame for the accident is determined and if the TP is responsible, a penalty is introduced. If the ego-vehicle is to blame, then the difficulty of the scenario was too high for the ego-vehicle under consideration (in this case the simple IDM / MOBIL model) and this is exactly what should be achieved with challenging scenarios.

If all constraints are met, $P_{\mathrm{GA}} = 0$ and the fitness function becomes the negative value of the complexity, which is then minimized by the optimizer, thus maximizing the complexity of the scenario. In the next step the reproduction of the individuals can be performed.

Reproduction Each individual in the population is assigned a fitness function value, which describes how well the individual fulfills the given problem formulation. The reproduction is performed internally in the Matlab pre-implemented GA algorithm and does not have to be implemented manually. The algorithm bases the reproduction on a small percentage of the best individuals (elite) and combines them to new individuals. Additionally, mutations are inserted, which represent a random change of individual acceleration values. The most unsuitable individuals are sorted out so that the population size remains identical in each iteration.

After reproduction, all individuals are simulated again, and a new iteration starts. This is repeated according to Algorithm 1 as long as defined end conditions are not met. In the optimal case, this is the convergence of the solution to the optimal solution $\mathbf{y}^*$

$$\mathbf{y}^*(\mathbf{x}(t), \mathbf{u}(t), t_0) = \underset{\mathbf{y}(\cdot)}{\operatorname{argmin}}\, J_{\mathrm{GA}}(\mathbf{x}(t), \mathbf{u}(t), t_0) = \underset{\mathbf{y}(\cdot)}{\operatorname{argmin}}\, (P_{\mathrm{GA}} - C_{\mathrm{GA}}), \tag{4.32}$$

otherwise it can be aborted based on other conditions, such as reaching a maximum number of iterations.

With this approach, the scenarios extracted from the highD data set can be made even more challenging and these scenarios are a good choice for the certification of AVs as they are challenging in terms of the behavior of surrounding TPs.

Regardless of whether the most complex scenarios from the highD data set are stored directly in the database, or the scenarios are subject to the optimization process just described, the reduced number represents a selection of good test scenarios for all AV systems. Scenarios that do not correspond to the ODD of the ego-vehicle to be tested can also be sorted out here. For example, if an AV can only be activated and used up to a defined maximum velocity, all scenarios with higher ego-vehicle velocities can be sorted out.

In the previous sections, the methodology on the left of the database from Figure 4.12 has been introduced. In the next section, the scenarios extracted using this method are examined in more detail and the procedure is validated.

4.3.3 Results and Validation

Analogous to the approach, the results are divided into the three steps to the left of the database from Figure 4.12. Afterwards, the method is validated. The validation focuses on the core element of the approach, the complexity metric, which is validated by an expert survey and an exemplary execution of the part to the right of the database (scenario simulation and AV assessment).

Scenario clustering and classification results

The results of the scenario clustering and classification depend mainly on the three parameters $t_{\mathrm{ch,predict}}$, l_{coll} and the type of the challenger determination strategy, as already mentioned in Subsection 4.3.2. In order to determine their influence and to keep the computational costs low, the first ten of the 60 measurements of the highD data set are used. In these ten measurements $11,370$ vehicles are recorded, which leads to a maximum possible number of as many concrete

scenarios. First, the time $t_{\text{ch,predict}}$ is examined, which describes how long the trajectory of the ego-vehicle is predicted into the future (Figure 4.15). Thereby, the length of the collision area l_{coll} is chosen constant to $0.5 \cdot d_{\text{safety}}$ and the challenger is determined according to the first contact strategy. A summary of the share of scenarios containing a challenger is shown in Table 4.10.

Table 4.10: Share of challenger scenarios as a function of the prediction time $t_{\text{ch,predict}}$. The values shown have been calculated based on the first ten of the 60 highD measurements, which include a total of $11,370$ vehicles. For the collision length l_{coll}, the value $0.5 \cdot d_{\text{safety}}$ is used and the challenger is determined according to the first contact strategy.

$t_{\text{ch,predict}}$ in s	1	2	3	4	5
Challenger scenarios in %	29.2	29.9	32.8	38.2	44.3

As expected, the share of challenging scenarios with longer prediction times increases as more potential collisions occur. A maximum prediction time of $5\,s$ is chosen. Longer prediction times do not make sense from the author's point of view because the deviation from the actual trajectory increases with higher prediction times, especially since only a simple constant acceleration model is used. The percentage of challenger scenarios that already become challenger scenarios at low prediction times are the most interesting ones, because they require a reaction of the ego-vehicle within a short time. A reasonable value for $t_{\text{ch,predict}}$ in Table 4.10 is $2\,s$, because this is comparable to the reaction time of a human driver. Next, we investigate how the share of challenger scenarios changes depending on the collision length l_{coll}. A prediction time $t_{\text{ch,predict}}$ of $2\,s$ and the first contact strategy is used. A summary of the percentage of scenarios containing a challenger is given in Table 4.11.

Table 4.11: Share of challenger scenarios in relation to the collision length l_{coll}. The values shown have been calculated based on the first ten of the 60 highD measurements, which include a total of $11,370$ vehicles. For the prediction time $t_{\text{ch,predict}}$, the value $2\,s$ is used and the challenger is determined according to the first contact strategy.

l_{coll} in m	$0.1 \cdot d_{\text{safety}}$	$0.2 \cdot d_{\text{safety}}$	$0.3 \cdot d_{\text{safety}}$	$0.4 \cdot d_{\text{safety}}$	$0.5 \cdot d_{\text{safety}}$
Challenger scenarios in %	1.6	5.1	12.2	21.6	29.9

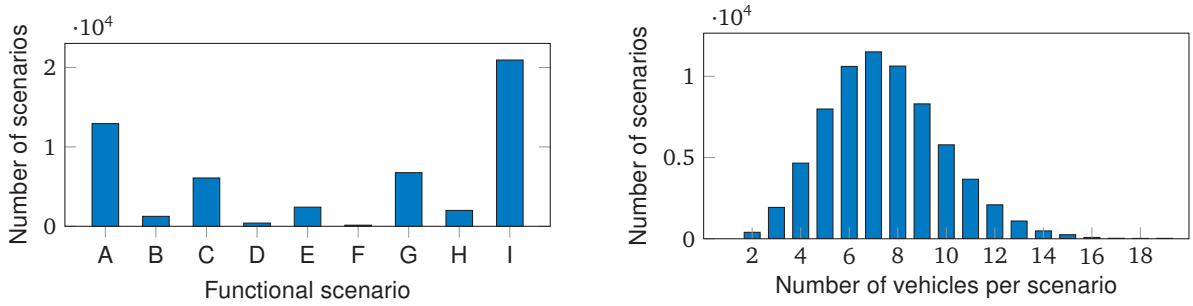
As expected, the share of challenger scenarios increases with increasing l_{coll}. The upper limit is $0.5 \cdot d_{\text{safety}}$, because this corresponds to the distance from which the fines become higher and so-called "Punkte" (English: points) are assigned in Germany[2]. Scenarios which become challenging at small collision lengths are more interesting because this corresponds to a high intensity of the required action (e. g. strong deceleration required). A reasonable value for l_{coll} in Table 4.11 is $0.5 \cdot d_{\text{safety}}$, because even at this length a comprehensive reaction is required to avoid any points.

There is no difference in the number of challenger scenarios between the two different strategies (first contact and maximum overlap) for challenger determination. What may change is the classification of the scenario into one of the nine defined functional scenarios because the TP that is considered as the challenger may change. However, $54.2\,\%$ of the considered scenarios have only one challenger. In theory, a shift can also occur in scenarios that only have one challenger. This is the case if the collision area (Figure 4.15) at first contact is different from that at maximum contact. Due to the short duration of the scenarios and the fact that the first challenger already requires an action, the first contact strategy is used in the further course of the work.

[2] https://www.bussgeldkatalog.org/abstand/ → If a driver has received more than eight points, there is the risk of a (temporary) withdrawal of the driving license.

Consideration of the entire data set Although a prediction time $t_{\text{ch,predict}}$ of $2\,\text{s}$ is defined as suitable, the entire data set is calculated with $t_{\text{ch,predict}} = 5\,\text{s}$. This is done in order to have a larger number of scenarios for the classification and the complexity assessment. Otherwise, only a small number is available for rarely occurring functional scenarios such as class F (Figure 4.13b), with which an evaluation of the occurring complexities is not reasonably feasible. If a more comprehensive data set than the highD data set is available, it makes more sense to use $t_{\text{ch,predict}} = 2\,\text{s}$. The collision length l_{coll} is chosen as discussed above to $0.5 \cdot d_{\text{safety}}$ and the first contact strategy to determine the challenger is used.

In Figure 4.19a the distribution of the 9 functional scenarios is visualized. The most common scenarios are the classes A and I, which refer to a challenger in the same lane as the ego-vehicle. The classic cut-in scenario, as defined in the future UNECE regulation [11], corresponds to the functional scenario H and occurs comparatively rarely. However, the functional scenarios D and G can also be considered as cut-in scenarios, which increases the number of cut-in situations. A total of 69,584 challenger scenarios occur in the entire data set with the defined parameter values. All maneuvers of the surrounding TPs according to the procedure in Subsection 4.3.2 can be determined for 52,918 (76.05 %) scenarios. For the remaining scenarios, undefined maneuvers occur that can be integrated in future work.



(a) Number of concrete scenarios in the highD data set which include a challenger. 52.918 scenarios are assigned to the nine defined functional scenarios from Figure 4.13b.

(b) Distribution of the number of involved vehicles per scenario, whereby the ego-vehicle is also counted. The median is seven and the maximum (without traffic jam situations) is 19 vehicles.

Figure 4.19: Results of the clustering and classification of the highD data set.

One motivation for the approach of extracting challenging scenarios based on real world data is the representation of actually occurring situations in real road traffic. In the future UNECE regulation [11], only the framework for the scenarios under consideration (e. g. cut-in) will be specified. This means that only a simple situation consisting of one challenger and the ego-vehicle will be defined. According to Figure 4.19b, which shows the number of vehicles involved in a scenario including ego-vehicle, this combination rarely occurs in real traffic. Only 397 of 69,584 challenger scenarios, which corresponds to a share of 0.57 %, are scenarios where only a challenger and the ego-vehicle occur. Considering the required transferability of the results to the behavior of the vehicles in real traffic, the definition of challenging scenarios with several TPs is therefore of essential importance.

According to the definition of the ROI in Figure 4.13a, the maximum of surrounding TPs is eleven. As can be seen in Figure 4.19b, scenarios with more than eleven TPs also occur. This can be caused by either not maintaining the safety distance of surrounding TPs, or by faster / slower TPs leaving the ROI or new TPs entering the ROI during the scenario. In the depiction in Figure 4.19b, traffic jam situations are excluded because it can happen that an adjacent lane has

not yet come to a standstill and therefore many vehicles pass through the ROI of the ego-vehicle under consideration. However, this does not provide any added value in terms of information.

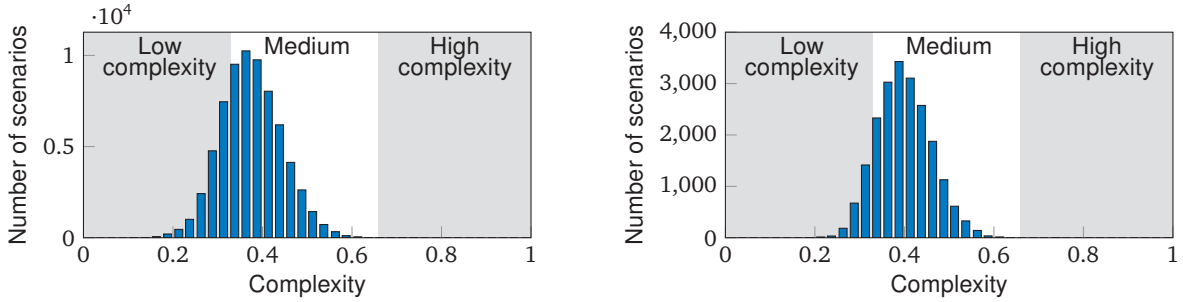
Complexity assessment results

In this section, the complexity of all challenger scenarios is calculated according to Equation (4.20) - (4.23). Since the factors influencing complexity are defined mostly independent of the behavior of the ego-vehicle, the highD data set can be used here, although the data set consists purely of vehicles controlled by human drivers. To calculate the complexity, the vector with the weighting factors $\mathbf{c}_{\text{weighting}}$ of the individual factors must first be defined. In the present work the weighting vector

$$\mathbf{c}_{\text{weighting}} = [0.01, 0.087, 0.087, 0.1, 0.087, 0.077, 0.087, 0.087, 0.087, 0.087, 0.1, 0.02, 0.084] \quad (4.33)$$

is used. How the values for the weighting vector $\mathbf{c}_{\text{weighting}}$ are determined is explained in the subsequent validation subsection.

Figure 4.20a shows the distribution of the complexity of all challenger scenarios independent of the assignment to the functional scenarios. An equidistant division of complexity based on [129] into low, medium and high complexity is used. With the defined weightings, only eight of the 69,584 challenger scenarios are assigned a high complexity. This corresponds to a share of $0.0115\,\%$. The course of example scenarios from a top view is illustrated in the appendix in Figure B.1.



(a) Complexity distribution of all challenger scenarios. Only 8 out of the 69,584 challenger scenarios are classified with high complexity.

(b) Distribution of the scenario complexity of the 20,939 concrete scenarios of functional scenario I which has the highest occurrence frequency of the nine functional scenarios.

Figure 4.20: Complexity distribution of the extracted highD scenarios using an equidistant classification into low, medium and high complexity based on [129]. Each scenario class shows a similar distribution compared to the distribution of all scenarios.

The different functional scenarios all show similar distributions compared to the distribution of all scenarios. As an example, this is shown for the most frequently occurring functional scenario I in Figure 4.20b. From this it can be concluded that the type of functional scenario does not allow any conclusions regarding its complexity. The most complex scenarios from the highD data set can be used for the application or the entry into the database. Optionally, the scenarios can even be improved in terms of complexity using the optimization process.

Scenario optimization results

Theoretically, all scenarios can be subjected to the optimization process. It is also possible to use scenarios that do not contain a challenger and to generate this challenger during the optimization process. To prove the applicability and functionality of the optimizer, a small number

of exemplary scenarios is used. The selection is based on the number of surrounding TPs that a scenario contains. For each number, the scenario with the highest complexity is selected from the highD data. We use these scenarios because as an approximation, each scenario with an identical number of TPs can be transformed into each other. Only the initial states, which are taken from the highD data, can still differ. As already described in Figure 4.19b, traffic jam situations are excluded and therefore the maximum number of participating vehicles including the ego-vehicle is 19, i. e. a total of 18 scenarios (1 surrounding TP to 18 surrounding TPs) are selected, each of which has the highest complexity value. This is shown in the Kiviat diagram in Figure 4.21 in orange.

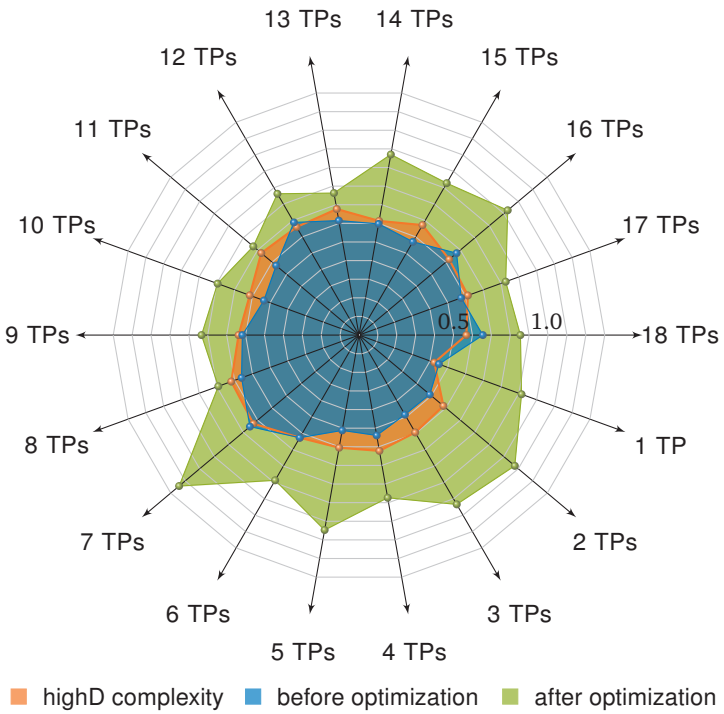


Figure 4.21: Complexity values of the scenarios for the proof of concept of the optimization process. The 'highD complexity' is the complexity of the original highD scenario. 'Before optimization' is the complexity of the scenarios where the human driver of the ego-vehicle is replaced by the approximate implementation of a AV by the IDM and MOBIL model. Additionally, in 'after optimization' the behavior of the surrounding TPs is optimized by the GA algorithm.

It can be seen that the complexity does not only depend on the number of surrounding TPs. The highest values are obtained with an intermediate number of surrounding TPs. This means that the highest value occurs with $C_{\text{scenario}} = 0.73$ with 7 TPs, followed by $C_{\text{scenario}} = 0.72$ with 8 surrounding TPs. The lowest value is obtained with 1 TP with $C_{\text{scenario}} = 0.42$, followed by $C_{\text{scenario}} = 0.57$ with 18 TPs. An overview of all scenarios can be found in the appendix in Table C.3. Figure 4.21 also shows that the developed metric is mostly independent of the behavior of the ego-vehicle. This can be observed using the values shown in blue. As explained in Subsection 4.3.2, during optimization the ego-vehicle is replaced by the IDM and MOBIL model, thus emulating a simple automated system. The values shown in blue represent the complexity values when the surrounding TPs behave exactly as in the highD data and the ego-vehicle is replaced by the model. It can be seen that there are only minor differences. There is also a slight tendency that the original highD scenarios with the human driver as ego-vehicle show slightly higher values.

The complexity values after optimization are shown in green in Figure 4.21. The results show that the GA algorithm achieves an improvement of the values in all scenarios. The functionality

of the algorithm is thus proven. That more complex scenarios generally lead more often to critical situations is shown at the end of this subsection. However, values > 1 occur in some scenarios, which is due to the normalization already mentioned in Subsection 4.3.2, for which a physical upper limit cannot always be found. Even if this does not have a negative influence on the results obtained, a further well-founded investigation of the normalization values of the individual complexity factors can be carried out in further work. The parameters of the GA algorithm used are summarized in the appendix in Table C.2. An overview of the most important information on the selected scenarios for optimization including the number of optimization variables and the required duration can be found in the appendix in Table C.3. The number of optimization variables tends to increase with the number of surrounding TPs and thus also the required calculation time of the optimization.

As introduced in Subsection 4.3.2, the trajectories of the TPs from the original highD data are used as the initial solution for the optimization. In combination with the replacement of the ego-vehicle controlled by a human driver using the IDM and MOBIL model, collisions between a TP and the 'new' ego-vehicle may occur for which the TP is to blame. In most cases, the TP collides with the rear of the ego-vehicle because the ego-vehicle adjusts a larger distance to the front vehicle compared to the human driver from the original highD data. The reason for this is the frequent non-compliance with the required safety distance by human drivers, especially at higher velocities [219]. The other aspects of the penalty function from Equation (4.31) do not raise problems when defining the initial solution. To improve the optimization results, an adaptive adjustment of the start solution can be investigated so that no collisions occur.

Regardless of whether scenarios are used directly from real data or via the optional optimization module, after this step a database with the most challenging scenarios in terms of the behavior of surrounding TPs is available.

Validation

In the previous sections, a methodology and the corresponding results are explained to identify the most challenging scenarios with respect to the behavior of surrounding TPs. The core of this methodology is a complexity metric that allows to determine the difficulty of the scenario mostly independent of the behavior of the ego-vehicle under test. The validation of this metric is an important process that is accomplished with the following two-step approach (Figure 4.22). Next, the two steps are explained in more detail.

Step 1 of the validation process was supported by the student thesis of Yu *[212].*

Step 1: Online expert survey A significant impact factor in the complexity metric is the relative weighting of the individual factors $c_{\text{weighting},j}$ to each other. Without additional information, only an equal weighting of all 13 factors with $1/13$ is reasonable. One approach to determine a suitable weighting vector $\mathbf{c}_{\text{weighting}}$ is an online expert survey, with which $\mathbf{c}_{\text{weighting}}$ can be determined based on expert knowledge. In addition to querying standard information such as age and duration of experience in the field of safety assessment of AVs, the designed questionnaire consists of two parts.

In the first part, the importance of each of the 13 factors is determined by the experts. They can rate each factor on a scale from 1 (not important) to 5 (very important). Here, it is additionally asked whether, from the experts' point of view, other factors must be considered in the complexity metric. This can be used to check the completeness of the metric.

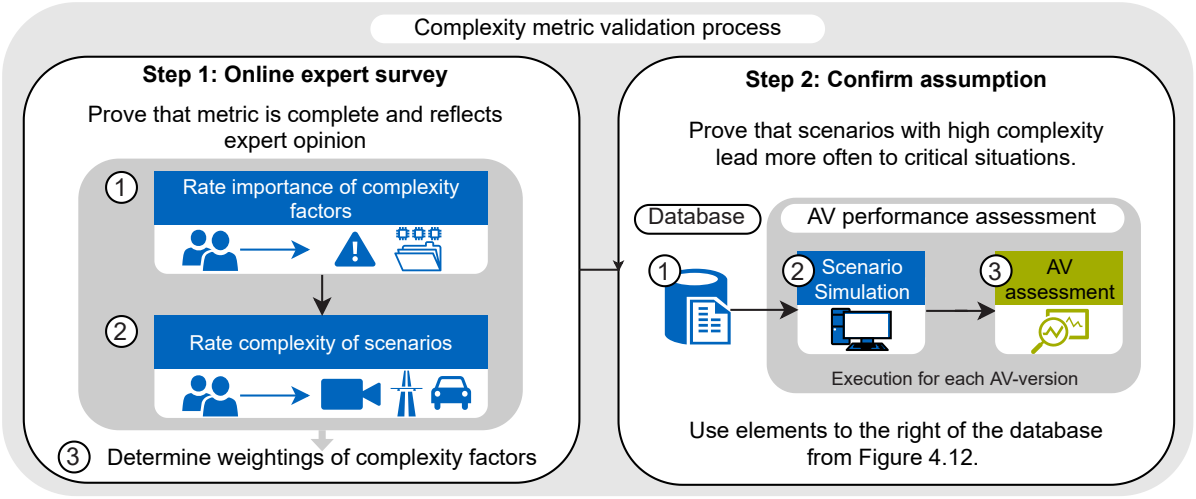


Figure 4.22: Overview of the process to validate the complexity metric. The first step is a online expert survey to determine the weighting factors $c_{\text{weighting},j}$. The second step is to prove the assumption from Section 3.2 that scenarios with high complexity lead more often to critical situation when they are executed. For this, the modules to the right of the database according to Figure 4.12 are used.

In the second part of the survey, the experts will be provided with videos[3] of 20 different highD scenarios, which the experts will also rate on a scale of 1 (not complex) to 5 (very complex). Both parts of the questionnaire are therefore ordinally scaled and will then be used to determine the weighting vector $\mathbf{c}_{\text{weighting}}$. This division is used to first focus on the individual factors and then to compare the overall result of the metric with the expert judgment. This is to prevent that a certain factor is classified by the experts as very important, but that scenarios where this factor is very prominent are only evaluated as not complex. Since this conflict can occur, a compromise must be found in the weighting determination process.

Step 1.1 Factor importance rating: In the first part of the survey, 25 participants ($n_{\text{experts}} = 25$) from industry and academia from the automotive sector and mainly from the safety assessment of AVs responded, with about three-quarters of them being from Germany. The remaining persons are distributed among China, Singapore, Czech Republic, Slovakia and Croatia. The average age of the participants is almost 32 years. The Code of Practice for the Design and Evaluation of ADAS [220, p. 15] states in the context of user experiments a number of at least 20 participants for an indication of the validity of the results. Thus, a sufficient number of participants can be assumed in the conducted online questionnaire.

As already briefly introduced, in the first part of the questionnaire the 13 complexity factors are rated by the participants on a scale of 1 (not important) to 5 (very important). The results of all participants for all factors are summarized in Table D.1 in the appendix. A reliability analysis based on Cronbach's Alpha α_{cronbach} [221]:

$$\alpha_{\text{cronbach}} = \frac{n_{\text{factors}}}{n_{\text{factors}} - 1}\left(1 - \frac{\sum_{i=1}^{n_{\text{factors}}} V_i}{V_t}\right) = 0.8216 \tag{4.34}$$

is used to determine whether the generated results are also reliable. With $n_{\text{factors}} = 13$ and the variances V_i of the individual 13 factors as well as the variance from the total score V_t from Table D.1, the value is 0.8216. A result of more than 0.8 indicates a 'good' reliability. Furthermore,

[3] The videos are available via https://www.youtube.com/channel/UC3IV32GfmVKXouqvF74jMFg/videos?view=0&sort=dd&shelfid=0

the experts did not name any additional factor when querying for influencing factors not taken into account, which can be seen as an indication of the completeness of the metric. All comments of the participants can be found in the appendix in Section D.2.

To calculate the order of the ordinally scaled data, the sum of the ranks of the individual factors $R_{\text{factor},i}$ and their mean rank $\bar{R}_{\text{factor},i}$ are calculated. The rank is the numbering of the order of the ratings by the experts, sorted by size. Since only five different values can be assigned, there are often identical values. This is called ties and is taken into account in the calculation of ranks by averaging. A summary of the rank sums $R_{\text{factor},i}$ as well as mean rank $\bar{R}_{\text{factor},i}$ for all 13 factors can be found in Table D.2 in the appendix. A visualization of the mean rank of all 13 factors is shown in Figure 4.23. The identical order of the factors as in Table 4.6 is used in Figure 4.23.

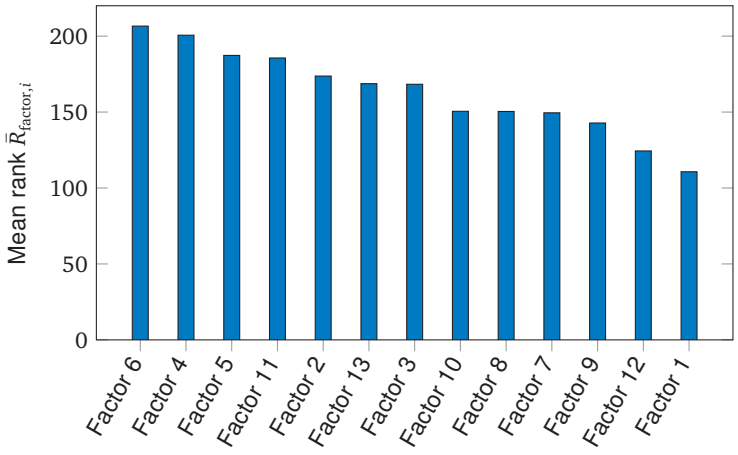


Figure 4.23: Evaluation of the importance of the individual factors by 25 experts. The higher the mean rank $\bar{R}_{\text{factor},i}$ the more important the respective factor i is rated. The exact values of each factor are summarized in the appendix in Table D.2. A description of the factors can be found in Table 4.6.

Figure 4.23 shows that the Factor 6 (predictability) and Factor 4 (dynamic) are rated most important by the experts. The least important factors are therefore Factor 1 (types of surrounding TPs) followed by Factor 12 (number of actions of ego-vehicle).

Next, the Kruskal-Wallis test [222, p. 587] is used to determine whether there are significant differences in the evaluation of the individual factors. If this is not the case, then it cannot be statistically proven that one factor is rated by the experts as more important or less important than another factor. This is checked using the test statistic H:

$$H = \frac{\frac{12}{N_{\text{total}}(N_{\text{total}}+1)} \sum_{i=1}^{n_{\text{factors}}} \frac{R_{\text{factor},i}^2}{n_{\text{ele},i}} - 3(N_{\text{total}}+1)}{1 - \frac{1}{N_{\text{total}}^3 - N_{\text{total}}} \sum_{k=1}^{l} \tau_k^3 - \tau_k} = 29.67, \tag{4.35}$$

where N_{total} is the total number of expert ratings, which is calculated based on the number of factors n_{factors} rated and the number of elements (ratings) per factor $n_{\text{experts}} = n_{\text{ele},i}$ to $N_{\text{total}} = 13 \cdot 25 = 325$. In our case, the number of elements per factor is identical for all 13 factors. $R_{\text{factor},i}$ is the rank sum for all factors i listed in the appendix in Table D.2. The denominator in Equation (4.35) is a correction term for tied ranks and is calculated from the number of tied observations τ with rank k. With the values given, the test statistic H is 29.67.

The Kruskal-Wallis test can be used to check whether the null hypothesis (there is no difference between the factors) can be rejected. Because there are more than five groups (complexity factors), a Chi-square distributed test statistic can be assumed. To determine whether there

is a significant difference between the factors, H is compared with the critical value H_{crit} of the chi-square distribution determined by the degrees of freedom df. The number of df is given by $n_{\text{factors}} - 1 = 12$. Additionally, the significance level $\alpha = 0.05$ is predetermined. If H exceeds the critical value H_{crit}, it can be concluded that p-value$< \alpha$ and thus the null hypothesis can be rejected. The value of H_{crit} can be read from tables [223]. In the considered case, $H = 29.67 > H_{\text{crit}} = 21.03$ and thus it follows that there are differences in the importance of the factors. However, it is not possible to determine between which factors a significant difference has occurred. In the next step, the so-called Dunn's test [224] is used for this purpose.

It can be examined for each combination whether the importance of the two factors considered differs significantly. For the 13 factors, $n_{\text{factors}}(n_{\text{factors}} - 1) \cdot 0.5 = 78$ pairwise comparisons must be made. The null hypothesis for each comparison is that there is no difference in the importance of the two factors considered. The null hypothesis can be rejected if the inequality

$$Q_{0.05} < \frac{|\bar{R}_{\text{factor},i} - \bar{R}_{\text{factor},j}|}{\sqrt{\left(\frac{N_{\text{total}}(N_{\text{total}}+1)}{12} - \frac{\sum_{k=1}^{l} \tau_k^3 - \tau_k}{12(N_{\text{total}}-1)}\right)\left(\frac{1}{n_{\text{ele},i}} + \frac{1}{n_{\text{ele},j}}\right)}} \tag{4.36}$$

is fulfilled.

Thereby, $\bar{R}_{\text{factor},i}$ denotes the middle rank of factor i and $\bar{R}_{\text{factor},j}$ that of factor j that are compared with each other. The number of elements (ratings) $n_{\text{ele}} = n_{\text{experts}}$ is identical for all factors, so that $n_{\text{ele}} = 25 \; \forall \, i, j$. $Q_{0.05}$ equals the z-value of a standard normal distribution from the so called z-table [225] for adjusted $1 - \frac{\alpha}{2 \cdot n_{\text{comp}}}$ level depending on the number of comparisons n_{comp} that is 78 in our case. With $\alpha = 0.05$, this results in a z value of 0.99968 and therefore, $Q_{0.05} = 3.41$ according to the z-table [225].

In the appendix in Table D.3, the right side of the inequality from Equation (4.36) is listed for all 78 pairwise comparisons of the factors. The reference value for each comparison is $Q_{0.05} = 3.41$. The results show that there are only significant differences for the pairs factor 6 to 1 and factor 4 to 1. These are marked in gray in Table D.3. For all other combinations, some tendencies can be identified which factors are of higher importance (Figure 4.23), but these are not high enough to be verifiable. These findings are taken into account when determining the weighting vector $\mathbf{c}_{\text{weighting}}$. First, however, the second part of the online questionnaire is evaluated.

Step 1.2 Scenario complexity rating: In the second part of the questionnaire, the participants are provided with videos of 20 scenarios ($n_{\text{scenarios}} = 20$), which are again rated on a scale of 1 (not complex) to 5 (very complex). This part of the questionnaire has been completed by 20 participants, so that from now on $n_{\text{experts}} = 20$. Again, according to [220, p. 15] this is enough for valid conclusions. For the analysis of the results, a three-step procedure consisting of Cronbach's Alpha, Kruskal-Wallis test and Dunn's test is carried out analogous to the first part of the survey. The results are summarized in Table 4.12. Furthermore, the mean rank of the scenarios is shown in Figure 4.24. The links to the videos of each scenario are summarized in the appendix in Section D.3.

Table 4.13 contains all combinations for which the null hypothesis (there is no difference in the complexity of the two scenarios considered) is rejected in Dunn's test. In addition, the values to be compared with $Q_{0.05}$ are listed. Since the null hypothesis is rejected for all pairings, all values are $> Q_{0.05}$. This means, for example, that according to expert opinion there is a significant difference in complexity from scenario 15 to scenario 17. In comparison to the evaluation of the influencing factors, significant differences occur more often here. This indicates that the order based on the mean rank can be used as an appropriate order scheme and the complexity metric

Table 4.12: Summary of the results of the second part of the online survey. When using the equations, the sub-script $(\bullet)_{\text{factor}}$ must be replaced by $(\bullet)_{\text{scenario}}$. The number of elements (ratings) per scenario n_{ele} is 20 for all scenarios and the total number of ratings is $N_{\text{total}} = 400$.

Name of test	Value	Description
Cronbach's Alpha	$\alpha_{\text{cronbach}} = 0.86$	Calculated via Equation (4.34) and $\alpha_{\text{cronbach}} > 0.8$ which means that a 'good' reliability of the results is given.
Kruskal-Wallis test	$H = 156.2$ $H_{\text{crit}} = 30.14$	Calculated via Equation (4.35) and $df = 19$ and $\alpha = 0.05$. There is a significant difference in the complexity of the scenarios because $H > H_{\text{crit}}$.
Dunn's test	$Q_{0.05} = 3.64$	In total, 190 pairwise comparisons have to be made. The combinations with significant difference are summarized in Table 4.13

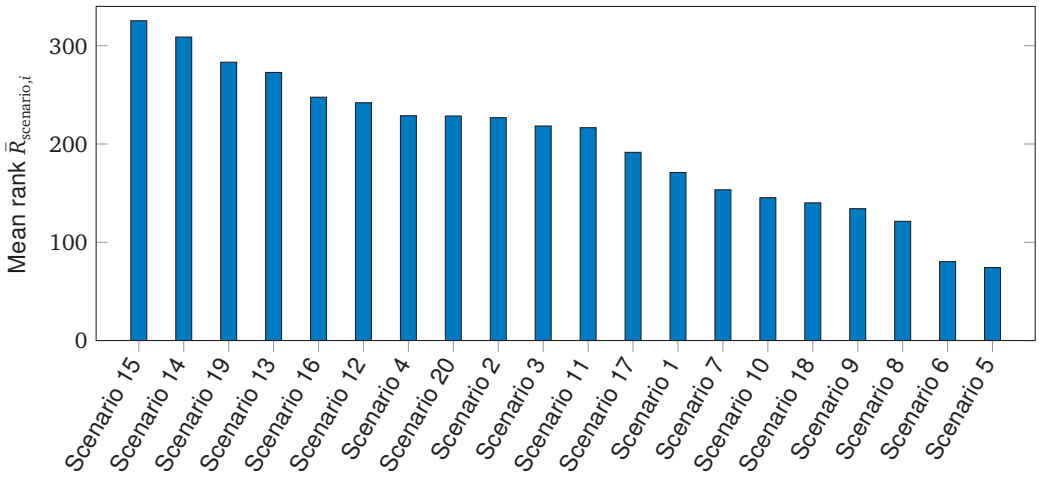


Figure 4.24: Evaluation of the complexity of the individual factors by 20 experts. The higher the mean rank $\bar{R}_{\text{scenario},i}$ the more complex the respective scenario i is rated by the experts. The exact values of each scenario are summarized in the appendix in Table D.2. The links to the videos can be found in the appendix in Section D.3.

Table 4.13: Summary of the results of the Dunn's test of the second part of the online survey. The experts rated the scenarios from each column significantly more complex than the scenarios in each row (if there is a value entered in the corresponding cell).

		Scenario										
		15	14	19	13	16	12	4	20	2	3	11
Scenario	5	7.12	6.65	5.92	5.63	4.92	4.76	4.38	4.37	4.32	4.08	4.04
	6	6.95	6.48	5.76	5.46	4.75	4.59	4.21	4.21	4.16	3.91	3.87
	8	5.79	5.32	4.59	4.30							
	9	5.42	4.95	4.23	3.93							
	18	5.26	4.78	4.06	3.76							
	10	5.11	4.63	3.91								
	7	4.88	4.41	3.68								
	1	4.38	3.91									
	17	3.80										

should create an order that is as identical as possible. In the next step, the findings from the first and second part of the survey are used to determine the weighting vector $\mathbf{c}_{\text{weighting}}$.

Step 1.3 Factor weightings: It is intended to take into account both, the assessment of the importance of the factors and the complexity assessment of the scenarios by the experts when determining the weighting vector $\mathbf{c}_{\text{weighting}}$. The basis is an identical weighting of all factors $c_{\text{weighting},j} = \frac{1}{n_{\text{factors}}}$. The importance of the factors in the complexity metric is determined by the

simple but effective sigma-normalized derivative [226, p.15]:

$$S^{\sigma}_{c_{\text{factors},j}} = \frac{\sigma_{c_{\text{factors},j}\,\partial\, C_{\text{scenario}}}}{\sigma_{C_{\text{scenario}}}\,\partial\, c_{\text{factors},j}} = \frac{\sigma_{c_{\text{factors},j}}}{\sigma_{C_{\text{scenario}}}} c_{\text{weighting},j}. \tag{4.37}$$

This is particularly suitable for linear models, as the complexity metric is. Here, the partial derivatives (which correspond to the weighting factors $c_{\text{weighting},j}$ in the linear model) are normalized with the ratio of input and output variance. This results in a measure with which it can be concluded which input (complexity factor $c_{\text{factors},j}$) is the most influential on the output (total complexity of scenario C_{scenario}). In Equation (4.37), $S^{\sigma}_{c_{\text{factors},j}}$ denominates the sigma-normalized derivative of input $c_{\text{factors},j}$ which is one of the complexity factors, $\sigma_{c_{\text{factors},j}}$ denominates the variance of the considered complexity factor and $\sigma_{C_{\text{scenario}}}$ the variance of the output which is the result of the complexity metric.

The complexity assessment of the scenarios by the experts is included by comparing the mean rank and the calculated complexity through the metric. The mean rank is also used here because the data is strictly ordinally scaled. The correlation between expert opinion and complexity metrics is determined using linear regression and least squares.

This is examined more closely in [212] and a conflict of targets arises because a higher weighting of the factors assessed by the experts as important leads to a deterioration of the least-squares error in the agreement of the assessed scenarios. Thus a compromise is obtained in the weighting vector, which is already introduced in Equation (4.33). Using these values, a level of influence of the influencing factors according to Figure 4.25 results, which is calculated using the sigma-normalized derivative. A value of 1 means that this is the most influential factor and a value of 13 that it is the factor with the least influence. In [212] it is found that the order between the individual measurement recordings (tracks) of the highD data set changes only slightly. Therefore, only two tracks are shown here. In addition, the order of importance of the factors determined by the experts is shown (analog to Figure 4.23).

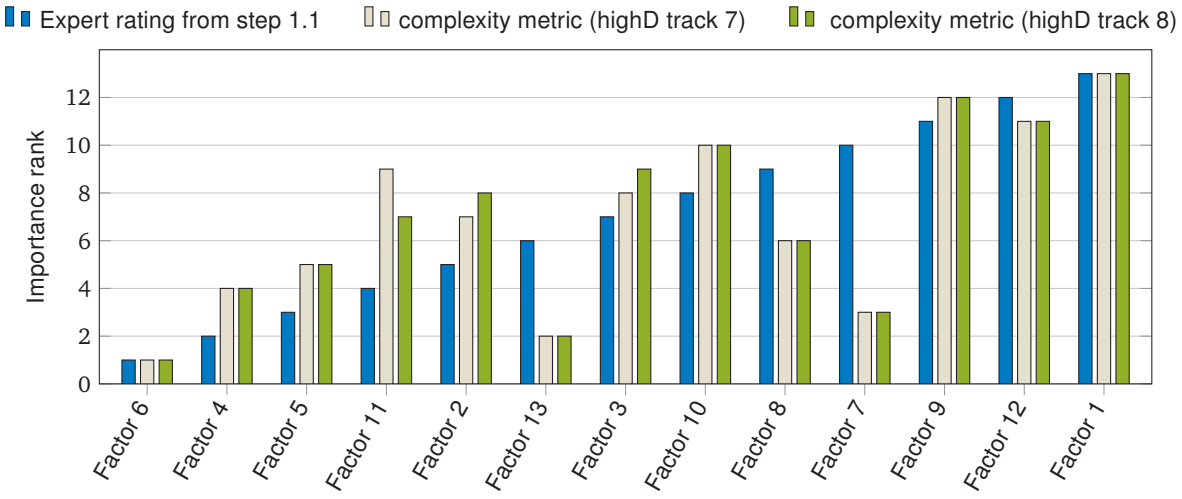


Figure 4.25: Order of importance of the factors influencing the result of the overall complexity of a scenario. The metric should reflect the expert opinion as far as possible. However, in the expert opinion there are only significant differences between Factor 6 and Factor 1 and between Factor 4 and 1, which the metric also reflects.

Figure 4.25 shows a close correlation between the actual importance of the influencing factors in the highD data set and the expert opinion. There are larger differences, for example, in the case of Factor 7, which is considered more intensively in the complexity metric than the

experts' opinion. However, it must be taken into account that no significant differences occurred in the expert opinion for Factor 7 compared to the other factors. Therefore, the order should not be overestimated. It is important that the significant differences that have occurred between Factor 6 and Factor 1 and between Factor 4 and 1 are included. Since the complexity metric takes Factor 6 into account most intensively, Factor 4 also very intensively and Factor 1 as the least intensively, this requirement is met by the metric.

Next, for the 20 scenarios, the correspondence of the calculated complexity based on the metric with the experts' assessment is checked. As already mentioned, the complexity rating of the expert opinion is represented by the mean rank $\bar{R}_{\text{scenario},i}$ from Figure 4.24. In Figure 4.26, this is plotted over the complexity C_{scenario} of the metric.

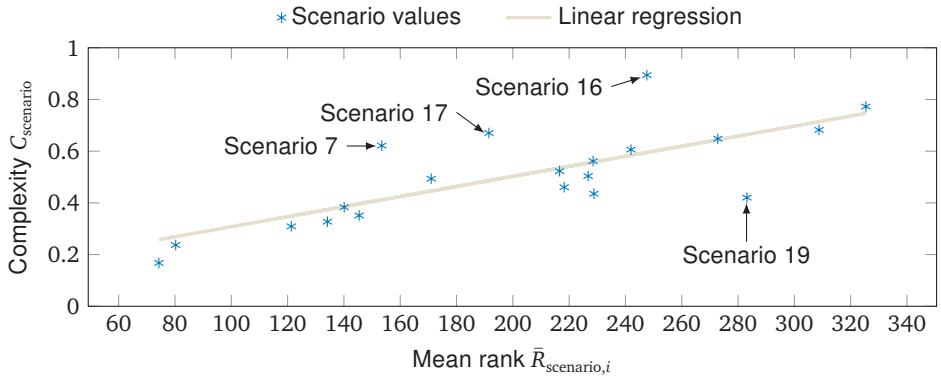


Figure 4.26: Correlation of the scenario assessment of the experts (mean rank of the scenarios $\bar{R}_{\text{scenario},i}$ from step 1.2) and of the complexity metric C_{scenario} with weightings according to Equation (4.33) based on the 20 selected scenarios and a linear regression. The metric reflects the experts' assessment well.

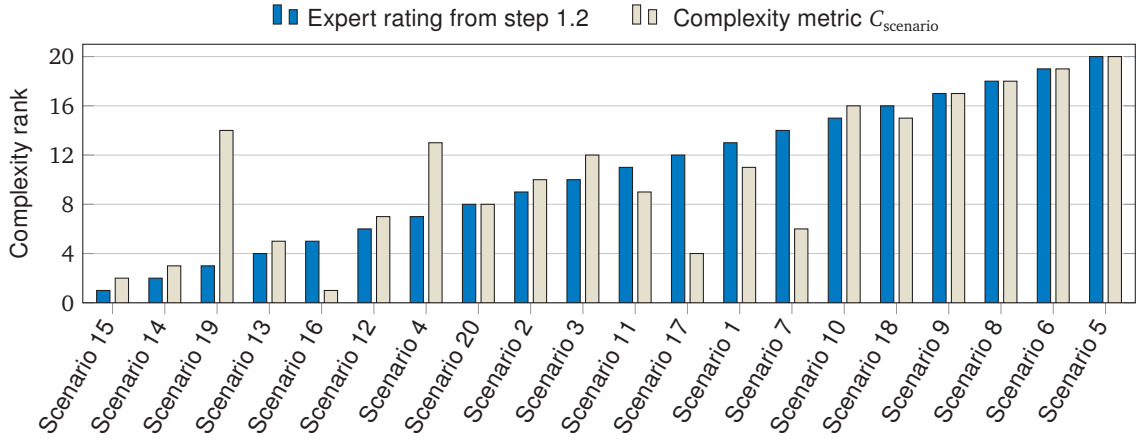


Figure 4.27: Order of the complexity values. The value 1 means the highest and 20 the lowest complexity of all scenarios. The metric with weightings according to Equation (4.33) should reflect the expert opinion as far as possible. However, the metric does not reflect one of the 43 significant differences according to Table 4.13, namely scenario 19 is not rated with higher complexity than scenario 7.

The linear regression in Figure 4.26 shows that the selected weighting vector from Equation (4.33) achieves a good agreement of the complexity metric with the expert opinion. Scenario 16, which is calculated with too high complexity, and scenario 19, which is calculated with too low complexity, are conspicuously higher and lower than the experts' assessment. Here again, the order of the complexity values of the scenarios compared to the experts' assessment (Figure 4.27) of complexity can be used as a check in combination with the table of significant complexity

differences from Table 4.13. On the one hand, this shows that the complexity metric also reflects the order of the scenarios well. However, it also shows that only 42 of the 43 significant differences from Table 4.13 are met. The only pairing not reflected by the metric is that scenario 19, according to the experts, is more complex than scenario 7. Even if scenario 16 is clearly overrated by the metric, the value does not violate any significant difference of the expert assessment to another scenario.

The previous paragraphs show that the developed metric mostly reflects the judgment of experts. An exact agreement and validation is not possible due to the described conflict of objectives. One reason for this is that expert opinions are assessments by humans, which can contradict each other. Therefore, a second validation step will check whether the developed metric can be used to confirm the assumption from Section 3.2 that challenging (complex) scenarios more often lead to critical situations. The procedure for this is explained in the next paragraph.

Step 2: Confirm assumption If the assumption is correct, scenarios with high complexity lead more often to critical situations. Whether the developed complexity metric can confirm this assumption is the objective of this second validation step. For this purpose, three sub-steps are performed as shown in Figure 4.22. First, suitable scenarios are selected from the scenario database. Secondly, these scenarios are implemented in a simulation tool and simulated with an exemplary implementation of an AV. Finally, in the third step, the scenarios are evaluated microscopically and the criticality of the simulations performed is examined in order to confirm or reject the assumption. The three steps are explained in detail below.

Step 2.1 Selection of Scenarios: When selecting the scenarios, the scenario database previously derived from the highD data is used. A distinction between the individual functional scenarios is not considered here. The distribution of the complexity of all scenarios in the database is shown in Figure 4.20a. Since the equidistant distribution of complexity results in too few scenarios with high complexity, a different approach is used when selecting the scenarios. Three groups are defined, which are located at the left and right boundary area as well as in the middle of the distribution. In this way, a total of almost 2000 scenarios are selected, so that there are 650 scenarios in each of the three groups. The group ’lowest complexity’ consists of the 650 scenarios with the lowest complexity values, i. e. the left end of the distribution from Figure 4.20a. The group ’average complexity’ consists of the 650 scenarios that are closest to the average value of the distribution of 0.38. The group ’highest complexity’ consists of the 650 scenarios with the highest complexity values, i. e. the right end of the distribution. These three categories can be compared later in the criticality evaluation.

The choice of the number of scenarios considered represents a compromise, so that enough scenarios are used for a reliable statement, but at the same time the differences in complexity are as large as possible. This allows to identify differences in criticality of the individual complexity classes best. The average complexity of the three groups increases from 0.20 for ’lowest complexity’ to 0.38 and 0.57 for ’highest complexity’.

Step 2.2 Scenario simulation: The selected scenarios are executed using Matlab/Simulink version R2019b by *The Mathworks*. The Automated Driving Toolbox [227] is used for this. First, the scenario information (trajectories of the surrounding TPs, their dimensions, the number and width of lanes and the initial state of the ego-vehicle in the scenario designer are loaded from the highD data. Thereby, the scenarios start at the frame with highest complexity, in order to reduce the influence of the ego-vehicle until the high complex scene is reached. Otherwise, it can occur that the AV drives slower than the original ego-vehicle of the highD data in the first ’boring’

part of the scenario and the complex situation does not even occur. For the implementation of the ego-vehicle, existing examples [228] of the Automated Driving Toolbox are used. To be able to reproduce the automated behavior for the considered highway scenarios, a combination of the existing examples of an model predictive control based ACC [229], a LKA [230] and an emergency brake assistant [231] is used. Since there is no pre-implemented function for changing lanes, a simple system is implemented which changes lanes when the adjacent lane is free and also complies with the German keep right directive. The vehicle dynamics are modeled using a bicycle model from the Vehicle Dynamics Blockset [232]. The sensor setup is defined by idealized camera and Radar sensors with $360°$ surround view. Starting from the initial state at time $t = 0\,s$ this simplified implementation of an AV takes over the calculation and execution of the trajectory of the ego-vehicle. After the simulation has been carried out, the trajectories are saved and the evaluation can begin.

Step 2.3 Criticality assessment: The microscopic evaluation of the simulated scenarios on the basis of criticality is an examination of the relative differences between the three defined complexity groups. Due to the simplified implementation of the AV, a macroscopic statement about the performance of the considered AV is not meaningful. The criticality evaluation is performed via the TTC [157] of the AV to the front vehicle. The minimum TTC value of the scenario is used as a comparison value.

Each of the three complexity groups contains 650 scenarios, whereby accidents occur for which the AV is not to blame. As with the optimization of complexity, these result almost exclusively from the more defensive driving behavior of the AV compared to the original ego-vehicle of this scenario from the highD data. This has the effect that TPs driving behind the AV will drive into the rear of the AV because the trajectories of the surrounding TPs are fixed and they have no intelligence. Because these kinds of critical situations and accidents have no relevance to the problem under investigation, these scenarios are sorted out. Therefore, a smaller number of scenarios are considered for the final evaluation according to Table 4.14.

Table 4.14: Results of the Matlab/Simulink simulation. The number of scenarios is not identical because all accidents that are not caused by the AV are sorted out. The critical TTC value is $1.5\,s$ and is based on [233].

	Complexity class		
	Lowest	Average	Highest
Average complexity	0.20	0.38	0.57
Number of scenarios	645	595	416
Scenarios below critical TTC	2	7	9
Scenarios below critical TTC at $t = 0\,s$	0	1	2
Number of accidents	2	13	22
Number of accidents with critical TTC at $t = 0\,s$	0	1	3

Table 4.14 shows that the number of cases sorted out increases with increasing complexity. Even if the complexity does not correlate directly with the number of TPs involved (Figure 4.21), there is still a tendency for low complexity scenarios to contain less TPs on average. As the number of TPs increases, the probability of irrelevant accidents increases as well. In addition, Table 4.14 shows that scenarios with high complexity are more likely to result in critical situations (TTC$<1.5\,s$). This is illustrated in Figure 4.28 using the cumulative distribution functions of the three complexity classes. Table 4.14 also shows that some scenarios and accidents already start with a critical situation (TTC$< 1.5\,s$). In order to be able to evaluate the influence of the start situation, Figure 4.29a shows the cumulative distribution function of the TTC in the start situation ($t = 0\,s$). From this it follows that the start situation has an influence, but it is only minor.

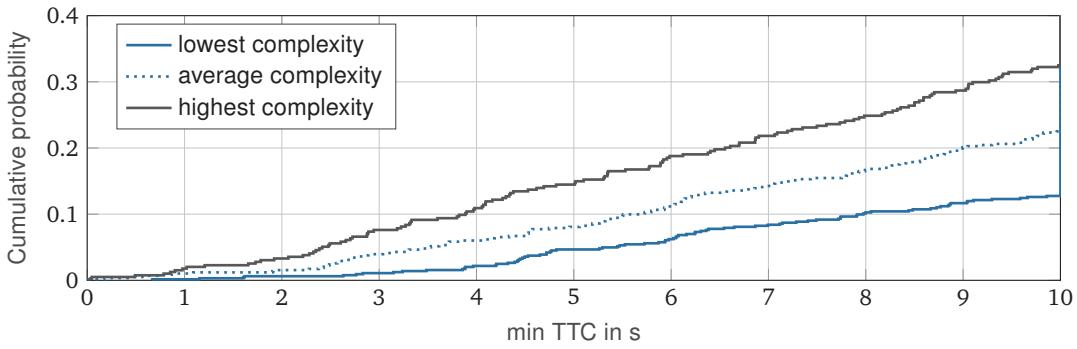
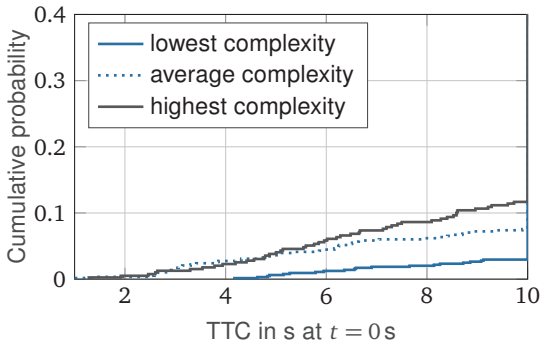
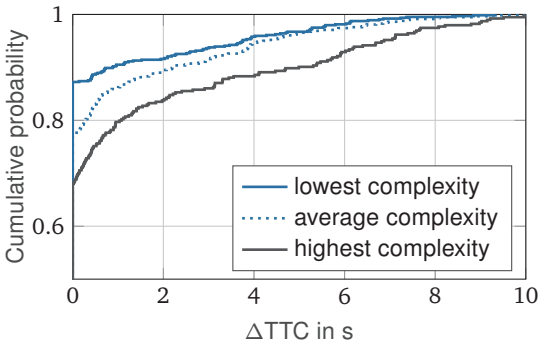


Figure 4.28: Cumulative distribution functions of the TTC of the lowest, average and highest rated scenarios. The three categories each contain approximately one percent of the scenarios shown in Figure 4.20a. It can be seen that scenarios with higher complexity more often show small TTC values. This means that more complex scenarios are on average more critical.



(a) Cumulative distribution functions of the TTC at $t = 0$ s of the lowest, average and highest rated scenarios. It can be observed that scenarios with higher complexity tend to have a slightly higher criticality at the beginning of the scenario than scenarios with lower complexity.



(b) Cumulative distribution functions of ΔTTC of the lowest, average and highest rated scenarios with ΔTTC as the difference between TTC at $t = 0$ s and the minimal TTC of the scenario. The higher ΔTTC, the more informative the scenarios are.

Figure 4.29: Analysis of simulated scenarios. The scenarios are the same as in Figure 4.28.

Figure 4.29b shows the difference in TTC called ΔTTC between the initial scene at $t = 0$ s and the minimum occurring TTC in the respective scenario. As more scenarios show a high ΔTTC, i. e. the lower right the course of the cumulative probability distribution, the more interesting the scenarios are. This in turn means that criticality increases over the course of the scenario due to decisions of the ego-vehicle. It is exactly this type of scenario that is therefore meaningful for the safety verification. Figure 4.29b shows that the most complex scenarios lead to higher ΔTTC and are therefore better suited for the homologation of AVs.

According to Table 4.14, the number of accidents in which AVs are at fault also increases with increasing complexity values. Since with increasing complexity values both, the number of scenarios with minimum TTC values below the critical limit of 1.5 s occur and the number of accidents caused by AVs increases, the assumption can be confirmed that more complex scenarios lead more often to critical situations (even if there is a small bias of the initial scene) and thus the complexity metric can be confirmed as valid.

In summary, this section develops and validates a complexity metric that can be used to assess the difficulty of concrete scenarios in relation to the trajectories of surrounding TPs (Layer 4 of five-layer model of Bagschik et al. [15]). Starting from the highD data set, concrete scenarios are extracted by means of clustering and classification and their complexity is subsequently

determined by a complexity metric based on expert knowledge and literature. The complexity metric consists of 13 individual factors, which are determined using a weighted linear combination to form a scalar complexity value for the scenario. The completeness of the metric and the weighting factors (Equation (4.33)) for the linear combination of the individual factors are examined and determined using an online expert survey. In addition, an optimization procedure is implemented to increase the complexity of selected scenarios. Finally, simulation executions are used to show that scenarios with high complexity are more likely to lead to critical situations, which proves the validity of the metric to identify challenging scenarios before their execution.

4.4 Combination

In this chapter the three elementary sub-methods of the overall procedure with which the homologation of a AV can be carried out at the overall vehicle level have been presented in detail so far. Each sub-method presented focuses on specific modules of the AV. Thus, a modular concept is available. The development and exemplary application of these sub-methods is the focus of the present work. The combination step introduced here is considered purely theoretically.

The three sub-methods developed must be applied for the homologation at overall vehicle level. In the simplest case, this means that all three sub-methods are executed one after the other in order to identify and thus define particularly challenging scenarios. The user (technical service) is thus provided with a tool which is able to address all relevant parameters according to the future UNECE regulation [11].

The execution of the sub-methods is partially automated as described in the previous sections. The importance of the individual sub-methods and their test scope cannot be determined in general, but depends, among other things, on the ODD of the AV to be tested as defined by the manufacturer. For example, if the manufacturer specifies that the system can only be activated under good weather conditions, then the sensor setup analysis in Section 4.1 is simplified.

Besides the step-by-step execution of the three sub-methods for the identification of challenging scenarios, the three methods can be combined. Here, the user executes the combination manually. For example, a particularly complex cut-in scenario from Section 4.3 can be defined with a curve with a static object positioned according to Section 4.2. This combination of challenging scenarios results in corner cases according to the definitions (Section 2.1). Since not all existing parameters are defined concretely, these scenarios are called semi-concrete in this work.

In summary, the overall results of this work are a simple and manual combination of the individual results from Section 4.1 - 4.3. The combination of these individual results provides technical services with a methodology to determine all the given parameters by the UNECE for the pre-defined functional scenarios, resulting in system-specific corner cases. Hereby, the individual parameters are considered and determined separately. An improvement of the performance of the presented method can be achieved by a simultaneous determination of all predefined parameters by means of a global optimization, as it is considered again in the discussion in Section 5.2. The main challenges here are the extensive automation of all sub-methods as well as the consideration of all occurring interactions between the sub-methods.

Alternatively, individual modules can also be used to test an AV at system or component level. Even if this is not sufficient to certify a vehicle according to UNECE regulations, it may be of

interest to suppliers, for instance. For example, a sensor manufacturer may develop a sensor setup and have it tested by a technical service using the sub-method described in Section 4.1. If all relevant objects are successfully detected in the selected scenarios, the manufacturer can advertise that his sensor setup is suitable for the homologation of AVs.

5 Discussion and Outlook

This chapter critically reviews the developed methodology and at the same time highlights possible improvements for future work. First the sub-methods are considered and then the overall method is discussed.

5.1 Sub-Methods

Of the three sub-methods from Section 4.1 - 4.3, the focus here is on the development of the complexity metric, because the other two sub-methods are already discussed in the previous publications and only a short summary is given here.

5.1.1 Sensor Analysis

The discussion of this methodology is also part of the author's previous publications [174–176]. For a better understanding of the present thesis a summary with extensions of the already published work is given.

The basis for the analysis of the sensor setup are the phenomenological sensor models. For the active sensors (Radar, Lidar and ultrasonic) these are taken from literature. These are validated in literature, but depend on the available information of the sensors used. According to the future UNECE regulation [11], the manufacturer has to provide this information to the technical service, but it is still unclear to what extent the quality of the provided information about the sensors affects the validity of the models, which has to be investigated in further work. For the camera as a passive sensor, a separate model has been developed [176] that describes the performance including object classification based on meta-information. The selection of the meta-information used is mainly motivated by the availability in the data set used (nuScenes) and is by no means complete. For example, glare caused by low sunlight has not been investigated and modeled because there are too few images available in the nuScenes data for this particular purpose. With the increasing number of freely available data sets there is a chance that further meta-information, such as the mentioned glare from low sunlight, can be considered in the future. Even better results can be achieved if the recorded data set is recorded with the identical camera at the identical mounting position and orientation, as is the case in the AV under test.

Based on the calculated sensor coverage, an optimal approaching path is calculated in [175] so that the AV can see the object (challenger) as poorly as possible. The optimal path is calculated in a three-dimensional grid. It is shown that the effect in z-direction is negligible with the sensor models used and the calculation can be reduced to a two-dimensional x-y grid. Only the influence of hilltops on the maximum range of the sensors must be considered. For example, the maximum range is significantly limited for the smallest permissible hilltop radius on German motorways [52]. Especially for low mounted sensors (e. g. front Radar), the visibility for small objects (height

of $0.5\,\mathrm{m}$) decreases to about $100\,\mathrm{m}$ during the crossing of the hilltop. Therefore, a reduction of the dimension is possible when choosing the grid. In contrast, when considering the challenger, an extension is necessary. The challenger is currently regarded as a point without dimensions, which can at least be replaced by modeling and considering the width of the object.

Improvement potential exists in the transformation of the approaching path into a dynamic scenario. Thereby a short deviation from the permitted road geometry occurs in Figure 4.8. This means that no road geometry could be determined that both establishes the optimal approaching path and complies with the road geometry permitted in Germany [52]. One reason for this is the occurrence of oscillations. A possible solution, which can be investigated in further work, is the reverse simulation in time starting from the end of the scenario. This can improve the results, because the most complex compliance with the road geometry at the time-related end of the scenario is then calculated at the beginning, where no oscillations have yet occurred. Another possibility is the consideration of maximum allowed curvature values when calculating the cost-optimal approaching path using the A* algorithm.

An extension of the approach can be implemented by an automated and comprehensive optimization of the choice of environmental conditions and the starting point of the challenger. Both are currently selected manually by the user. Especially the environmental conditions depend on how the ODD of the manufacturer is defined. If the system cannot be activated in a foggy environment, this weather condition does not have to be considered for certification. Due to the small number of different weather conditions, automated consideration is not as crucial, but still desirable. It is different when choosing the starting point of the challenger. For this, there is a variety of possibilities which must be examined more closely.

Subsection 4.1.3 argues by means of a mathematical validation. For the development of the methodology this is sufficient, but in a future application an additional validation of the concept is required. For the validation and also already for the implementation of the concept during homologation, simulations of the identified scenarios have to be performed. Because the calculated road geometries are usually not available on test sites, the only reasonable execution is by means of virtual simulation. Because the sensors are in the focus here, high-precision physical sensor models must be used that are capable of realistically representing a wide range of environmental influences. These are still topic of current research but are absolutely necessary for the simulation-based homologation of AVs at vehicle level. If the manufacturer cannot provide them, then the methodology from Section 4.1 can neither be validated nor reasonably applied and a simulation-based homologation of the overall vehicle for the considered AV is not possible.

5.1.2 Driving Behavior Characterization

The discussion of this methodology is also part of the author's previous publication [127]. For a better understanding of the present Thesis a summary with extensions of the already published work is given.

The main focus of this section has been the development of the methodology. Up to now, implementation has only been performed as proof of concept, which has been carried out successfully. For a comprehensive implementation, a high-performance automated driving function must be available in the simulation in order to execute the derived scenarios in a meaningful way. The implementation of the derived functional scenarios in a simulation tool involves an increased amount of work, because they were derived on the basis of knowledge and therefore cannot be automatically transferred from measurement data into the simulation. In addition, it can happen that no characteristics occur in the examined AV and therefore none can

be identified. In these cases, the execution of the scenarios would not generate any additional value.

In general, due to the limited implementation, there is a lack of experience as to which identified functional scenarios are particularly well suited to efficiently determine characteristic driving behavior. Similarly, there is a lack of experience on how many parameter variations of the functional scenarios should be executed as concrete scenarios. In the proof of concept under consideration, 24 variations of the functional scenario 'driving through a curve' are executed. To what extent this number can be transferred to other functional scenarios cannot be stated.

A side effect of the developed methodology is that the consideration of the theoretical Driving License Questionnaire [201] creates the basis for a simulation-based driving license test of AVs. Due to their rareness, many important situations, whose probability of occurrence is low, are not tested in the practical driving license test [200], but in the theoretical driving license questionnaire [201]. With the derived functional scenarios, parameters and KPIs it is possible for AVs to transfer this into virtual simulation and thus to perform a sort of simulation-based driving license test for AVs which has a strong symbolic effect. Of course, this is only applicable for questions about driving behavior and not for questions like the correct adjustment of the rear mirrors.

The conclusion is that this sub-methodology is applicable, but a lot of time must be invested to have the required scenarios, parameters (including discretization) and KPIs available in virtual simulation. Furthermore, even after extensive application, it is not ensured whether information can be derived from it. But if all scenarios and KPIs are implemented, an efficient and fast execution of the method for different AV versions is possible and therefore this sub-method can also contribute to an efficient certification of AVs and is an important tool for technical services.

5.1.3 Traffic Situation Complexity

The discussion of the traffic situation complexity is based on the order of the approach shown in Figure 4.12.

Scenario clustering and classification For the classification, the maneuvers performed by the surrounding TPs are determined using a rule-based approach. This approach is currently not yet able to correctly identify all maneuvers. Investigations in [209, chap. 4.2-4.3] show that 24.7 % of the challenger scenarios contain a non-assignable maneuver. These scenarios can still be taken into account in the database, only an assignment to the nine defined scenarios is not possible. In addition, it was shown that 94.8 % of the specified maneuvers are correctly assigned. These values are sufficient for the development of the methodology, but can be improved in future work. For example, it is possible to switch from a simple rule-based approach to methods of machine learning.

Currently, the extracted scenarios are divided into the nine functional scenarios shown in Figure 4.13b. This can be extended by information on action restrictions, so that, for example, cut-in scenarios can be tested specifically if the adjacent lane is occupied at a certain position. In addition, in congestion situations, the number of vehicles involved can be characterized by the maximum number of vehicles simultaneously within the ROI. Currently, as shown in Figure 4.19b, a maximum number of 19 involved vehicles (i. e. different 18 TPs are spread over the whole scenario and at least for one time step within the ROI) is considered. In congestion situations, it can happen that many vehicles in a faster moving adjacent lane drive through the ROI of the

ego-vehicle, but this does not provide any additional information. A description of the scenario using the maximum simultaneously present TPs can be more precise here.

Complexity assessment The complexity metric represents the core of this sub-method and is therefore of particular importance. One aspect of the metric is the normalization of the individual complexity factors so that the value of one is not exceeded. As already described in Subsection 4.3.2, this is not possible for all complexity factors because there is no physical upper limit. The execution of the optimization (Figure 4.21) shows that complexity values of more than one occur. Even if this circumstance does not reduce the quality of the metric, an adjustment of the normalization values can be performed so that the occurring complexity values are between 0 and 1. Because the importance of the individual factors changes according to Equation (4.37), a new investigation of the weighting factors is necessary afterwards.

For the development of the metric, it is assumed for simplicity that the overall complexity of the scenario consists of a linear combination of the individual complexity factors and without couplings between them. The validation of the metric has shown that these simplifications are acceptable and that valid results are achieved with them. In further work it can be investigated whether the performance of the metric can be improved by considering nonlinearities as well as couplings between the factors.

Scenario optimization There is always the potential for performance improvement during optimization. This can be achieved by adjusting the parameters of the optimization algorithm used or by choosing another algorithm that is more suitable for the given problem. Both are investigated in more detail in [216], first by studying different algorithms (simulated annealing, pattern search, particle-swarm and genetic algorithm) and then by studying the parameters for the best performing genetic algorithm. Due to the long optimization times (Table C.3) caused by the high number of variables, further optimization of the parameters can reduce computation time.

It can also be examined whether a simpler optimization with fewer optimization variables also achieves the desired result. This means that the maneuvers (i. e. the profile of the accelerations) of the surrounding TPs are not changed, but only the start positions and velocities. Depending on the duration of a scenario, this reduces the number of optimization variables from several thousands to a low two-digit number. Furthermore, in future work, additional constraints should be added so that the type of functional scenario cannot be changed or even specified in advance. In the current version, a scenario from class X may be a scenario from class Y after optimization.

During optimization, the penalty function punishes undesirable behavior of the surrounding TPs and thus prevents it. This is to ensure that as a result of the optimization, scenarios are obtained that are representative of the traffic situation on German highways and that the scenarios are therefore physically reasonable. This is only visually checked and found plausible in the present work. In future work, this can be investigated in detail on the basis of objective indicators.

Scenario simulation When executing the extracted scenarios using the simplified simulation model, it became apparent that testing exactly what should be tested with the considered scenario is difficult. The reason for this is that each AV can behave differently and thus the desired situation in the scenario does not necessarily have to occur when the surrounding TPs have fixed trajectories. This becomes more likely the longer the scenarios last. Therefore, the

scenarios used for the validation of the metric are started at the time step when the complex situation is already present. Here, it can also be examined whether it is better to let the scenarios start shortly before the most complex situation, so that the AV to be tested still has a chance to mitigate the situation. Thereby, the question arises how long before the most complex situation the scenario should be started. Ensuring that the desired situations occur during the execution of the scenarios can also be achieved by defining all trajectories relative to the ego-vehicle. However, this requires that the simulation tool used can handle this type of definition.

In addition, the executed scenarios often cause the problem that a TP driving behind the ego-vehicle drives into the rear of the AV, because the AV sets a larger safety distance to its front vehicle than the human driver in the highD data. Because this kind of collision has no relevance for the safety assessment of the considered AV, these scenarios have to be sorted out during the evaluation of the simulations. This problem also occurs during optimization, which leads to high penalty values in the initial solution, which the optimizer has to reduce in the iterations. In this case, adapting the start solution can improve the performance of the algorithm.

Overall complexity approach This sub-method focuses exclusively on the behavior of the surrounding TPs, which corresponds to Layer 4 according to the five-layer model of [15], and therefore considers only highway scenarios on straight roads. This is identical to the highD data set used, which also only considers scenarios on straight sections of the highway. However, the methodology is not limited to this special case, but can also be carried out with other data sets that include curves, for example. It is only important that the data is available for all vehicles within the ROI.

A disadvantage of the highD data set used is that the extracted scenarios are rather short for highway scenarios with a duration of slightly more than $10\,\mathrm{s}$. This means that, for example, complete overtaking maneuvers are almost impossible to find because their duration or distance is longer than that recorded in the data set. Furthermore, the maneuvers do not start at the beginning of the recording, but occur at arbitrary positions. To ensure that even long maneuvers are completely present and to extract the interesting part of a scenario, much longer scenarios are necessary. In the highD data set a distance of about $400\,\mathrm{m}$ is recorded and based on the experience gained a length of $1\,\mathrm{km}$ is desirable from the author's point of view.

Further investigations with regard to the overall methodology are necessary in the area of challenger determination for the classification of the scenarios into functional scenarios (Figure 4.13b). The evaluation of the highD data set showed that $45.8\,\%$ of the challenger scenarios have more than one challenger. This also raises the question whether a scenario in which the second challenger is a cut-in situation still belongs to the cut-in scenario according to the future UNECE regulation [11].

Even if the entire methodology is applied offline, future work may focus on how to reduce the required computing time. In particular, the extraction of the scenarios and the complexity assessment must be carried out for a large number of scenarios and therefore their efficiency is decisive for the entire sub-method.

The validity of the metric is determined using a multi-stage process consisting of an expert questionnaire and the evaluation of simulations of scenarios with different complexity values. This confirms the validity of the metric, which can be assumed to be reliable due to the multi-stage process. Nevertheless, a further validation of the metric can be carried out in the future using the simulation with different AV versions. Thus, it can be proven that the developed metric is not only valid for the exemplary AV used in this thesis.

5.2 Overall Approach

The safety assessment of AVs is still a relatively new research domain. Currently, many terms are not yet used consistently and there is no widely accepted definition for many of them. This is also affecting the present work, because terms such as challenging scenarios, corner cases or semi-concrete scenarios are used here as well. Because the work considers different aspects of challenging scenarios, the overall result is called corner cases. Based on the upcoming UNECE regulation [11] only the most important parameters are addressed and therefore not all parameters are defined. For this reason the term semi-concrete is used in this thesis for the identified scenarios. The corresponding definitions can be found in Section 2.1.

For the safety assessment and certification of AVs, it is always better if more parameters and scenarios are considered. But especially for homologation a reduced and efficient procedure is necessary to ensure a minimum level of safety based on the given framework of the UNECE. Therefore, the focus during the development of the methodology is on all parameters that are explicitly mentioned in the future UNECE regulation [11]. All these parameters are addressed by the developed overall method and thus the most important requirements of the UNECE are fulfilled, even if no objective evaluation can be made whether a sufficient number of tests have been performed.

The focus in this work is also on developing the method so that it is available to the technical service as an efficient tool for the certification of AVs. This is achieved and the exemplary implementation and validation of the sub-methods ensures the functionality of the overall method. A well founded validation of the overall method will only be possible after the introduction of the future UNECE regulation [11] into practice. Nevertheless, the method can be used for the homologation of the first systems by the technical service.

In a first step of implementation, the developed method has been implemented as far as possible in an automated way. However, it is currently not possible to completely replace and automate the expert judgment required in some places. However, this can also be seen as a positive aspect, because each assessor of the technical service can make a slightly different selection and therefore an additional variation is included in the test scenarios. This reduces the danger from Subsection 2.4.2 that the manufacturer of the vehicles prepares specifically for the tests to be carried out during homologation.

As already mentioned, highly accurate sensor models are required for the simulation-based certification of AVs. Because these are expensive to calculate, the simulation currently achieves only slightly higher simulation speeds than real time. This means that the execution of the scenarios takes a relatively long time and any reduction in the number of test scenarios, as achieved by the present methodology, is valuable. Even though the simulations can be parallelized well, and therefore the time required for execution can be reduced, each simulation execution costs money. This is examined in more detail in the author's already published work [92] and it is shown that it is not economically feasible to perform simulation-based homologation without reducing the number of test cases.

The developed method can also be interesting for system manufacturers during the development of the driving function in the context of optimization-based falsification. The developed method can be used to select the start scenario for the optimization process. The better this is chosen, the more efficiently the procedure can be carried out, thus saving time and costs.

The aim of this thesis is to develop a methodology for the use case of highways. In principle, the method is also suitable for other use cases (e. g. country road or city center), but special

adaptations have to be made. For example, the focus on complicated road design and the often unpredictable behavior of pedestrians must be addressed, which is not necessary for the highway use case. In this context, the requirements of the UNECE in the regulations that are intended for this purpose must be taken into account. Since there are no drafts for these regulations yet, it is difficult to estimate the future requirements.

The biggest potential for improvement of the overall method exists in the substitution of the simple combination block of the three sub-methods by a global optimization that takes all three parts into account. For this purpose, the sub-methods must be automated as far as possible and an iterative inclusion of all aspects must be carried out. For example, the challenger from the sensor analysis can be included as a constraint in the complexity optimization and it can be additionally specified that this challenger results in a cut-in situation. There are many more couplings between the three sub-methods, so that the implementation of this global approach represents a considerable workload for future work.

The developed methodology is an important tool for technical services in the homologation of AVs, but the user must be aware that this is only a small aspect of the overall AV-related certification. In addition, requirements for human-machine interface, functional safety or system self checks, for example, must also be considered. Even when selecting the scenarios, in addition to testing the particularly challenging corner cases, a certain coverage of the parameter space must be ensured. This is methodically not comprehensive and can be achieved by simple equidistant discretization, which is why it is not considered more closely in the context of this thesis. The only problem here is the question of the step size of the discretization, which must be specified by the technical service. An exemplary visualization, reduced to two parameters, of the interaction between the developed method for the system-specific definition of corner cases and a simple coverage-based approach, which was introduced in Subsection 2.3.5, is shown in Figure 5.1.

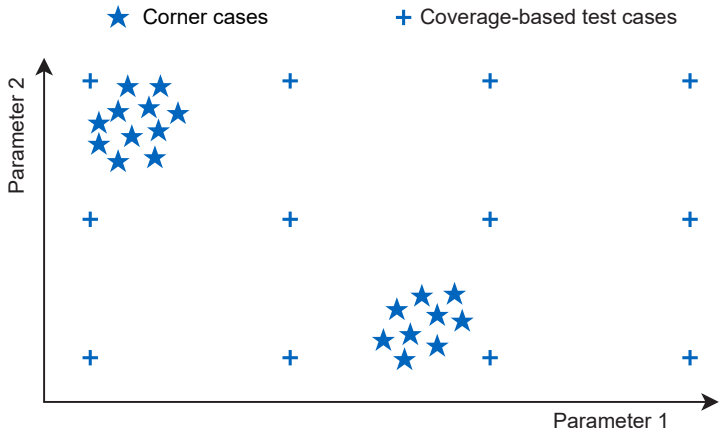


Figure 5.1: Simplified illustration of a combination of the developed corner case method with a simple coverage-based approach using only two parameters. The coverage-based approach can also be used for homologation in order to obtain information about the performance over the entire operating range of the AV.

6 Summary

The homologation of Automated Vehicles (AVs) is essential for the market introduction of these vehicles. This poses new challenges not only for vehicle manufacturers and the legislator, but also for technical services which are responsible in Europe for checking compliance with the requirements specified by the legislator. For this purpose, the United Nations Economic Commission for Europe (UNECE) has developed a framework [11] that represents the first regulation for AVs of Level 3 according to Society of Automotive Engineers (SAE) and is to come into force from 2021. This future regulation essentially sets the requirements for this work. The focus here is on the use case of the German highway and the examination of the correct Object and Event Detection and Response (OEDR).

Since the homologation covers a reduced safety proof, the work starts with a definition of the most important terms (Section 2.1) and a comprehensive presentation of the state of the art in the field of safety assessment of AVs (Section 2.2 and 2.3). The elaboration of this section is based on an already published work [17] of the author. Together with the special requirements for the type approval of AVs, the concrete question of the work is derived in Section 2.5. From the state of the art it can be concluded that the type approval represents a reduced proof of safety with special requirements. From the point of view of a technical service, however, this is not yet considered in the literature. Therefore the following research question is derived:

How should a procedure for technical services be designed that efficiently identifies system-specific corner cases within the homologation of automated vehicles?

On the basis of the derived requirements by the future UNECE regulation [11], an overall methodology is derived in Chapter 3 and it is explained how the requirements are addressed. Essentially, the overall method consists of three sub-methods, whose development and exemplary implementation is the core of the present work. A detailed description of the three sub-methods is given in Chapter 4.

First, a procedure is developed and implemented in Section 4.1, which deals in detail with the used sensor setup of the AV to be tested. Central for this section are the already published papers [174–176] of the author. Based on information about the sensors used and on phenomenological sensor models, quality of the sensor coverage is modeled and analyzed. An optimal approaching path to the AV is calculated in a three-dimensional grid using an A* algorithm. This path must be followed by an object (challenger) to approach the vehicle so that it can be detected as poorly as possible by the Vehicle Under Test (VUT). Thereby, different weather conditions can be considered. Finally, the static approaching path is transformed into a dynamic scenario and an optimal road geometry for the selected environmental conditions is derived for the scenario that establishes the calculated optimal approaching path. These scenarios thus represent a particular challenge for the perception of the vehicle.

The second sub-method in Section 4.2 examines the driving behavior of AVs and the question of how information can be derived from it to define particularly challenging scenarios specifically

for the AV under consideration. The basis for this is an already published work [127] by the author. Functional scenarios and objective Key Performance Indicators (KPIs) are derived from technical literature and from guidelines for driving license tests [200, 201], which can be used to characterize the driving behavior. The implementation of this sub-method is exemplified by the functional scenario of driving through a curve. It is shown how stationary obstacles can be optimally placed to provide a challenging situation for the AV based on an investigation of the cornering behavior.

Finally, in Section 4.3 the third sub-method is introduced, which deals specifically with the trajectories of surrounding Traffic Participants (TPs), so that it is particularly challenging for the AV to plan a safe trajectory. The term complex scenarios is used for this type of scenario. The basic concept has already been published in [16] by the author. The central element of this sub-method is a newly developed complexity metric that allows the difficulty (complexity) of scenarios to be assessed independently of the performance of the AV. The metric is based on 13 factors that are weighted differently, combined with each other and thus yield a scalar value for the complexity of a scenario. The validation of this metric is performed by a two-step process consisting of an online expert survey and the simulation of scenarios extracted from the highD data set. The former is used to determine the weightings of the individual factors. The latter to investigate the assumption that more complex scenarios lead more often to critical situations. The obtained results confirm this assumption and the developed metric can be considered valid. Chapter 4 ends with a brief description of how these three sub-methods can be combined by the user (technical service) in the execution of the type approval.

In the subsequent discussion and outlook (Chapter 5), the three sub-methods are first examined in more detail before the overall method is again critically reviewed. In the case of the sub-methods, special emphasis is placed on the development of the complexity metric, because the other two sub-methods have already been discussed in the author's previous publications [127, 174–176]. For the complexity metric, the biggest potential for improvement is a review of whether an improvement can be achieved by extending the metric from its current linear combination of individual factors to include non-linear elements. It can also be examined whether the consideration of couplings between the factors improves performance. For the overall method, it is advisable in future to use a global optimization that considers all three sub-methods simultaneously instead of the simple combination. Only then mutual interactions between the individual sub-methods can be efficiently taken into account.

The conclusion of the present work is that a method is developed which is specifically designed to meet the requirements of the type approval of AVs. The method is available to technical services as a tool to address the most important parameters of the future UNECE regulation [11] and to allow an efficient and system specific identification of corner cases, thus preparing the technical services for the introduction of the regulation. During the development of the method a lot of experience was gained, many questions were answered from the point of view of a technical service, but at least as many new research questions were raised in Chapter 5, to which a solution must be found in future research work.

List of Figures

List of Tables

Bibliography

[1] World Health Organization. „*Number of road traffic deaths*," 2016. Available: https://www.who.int/gho/road_safety/mortality/traffic_deaths_number/en/ [visited on 10/06/2020].

[2] SAE J3016. „*Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*," 2018.

[3] P. Koopman and M. Wagner, „Challenges in Autonomous Vehicle Testing and Validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, 2016.

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

[5] P. Koopman and M. Wagner, „Toward a Framework for Highly Automated Vehicle Safety Validation," in *WCX World Congress Experience*, 2018. Available: https://doi.org/10.4271/2018-01-1071.

[6] International Organization for Standardization. „*ISO 26262: Road vehicles — Functional safety*," 2018.

[7] International Organization for Standardization. „*ISO/PAS 21448: Road vehicles — Safety of the intended functionality*," 2019.

[8] NHTSA. „*Federal Automated Vehicles Policy: Accelerating the Next Revolution In Roadway Safety: September 2016*," 2016. Available: https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf [visited on 10/06/2020].

[9] NHTSA. „*A Framework for Automated Driving System Testable Cases and Scenarios*," 2018. Available: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automateddrivingsystems_092618_v1a_tag.pdf [visited on 10/06/2020].

[10] German Aerospace Center. „*PEGASUS-project*," 2019. Available: https://www.pegasusprojekt.de/en/home [visited on 10/06/2020].

[11] UNECE. „Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping System," 2020. [Online]. Available: http://www.unece.org/fileadmin/DAM/trans/doc/2020/wp29/ECE-TRANS-WP29-2020-081e.pdf [visited on 10/06/2020].

[12] International Organization for Standardization. „*Road vehicles — Safety of the intended functionality*," 2019.

[13] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt and M. Maurer, „Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 982–988.

[14] T. Menzel, G. Bagschik and M. Maurer, „Scenarios for Development, Test and Validation of Automated Vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1821–1827.

[15] G. Bagschik, T. Menzel and M. Maurer, „Ontology based Scene Creation for the Development of Automated Vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1813–1820, doi: 10.1109/IVS.2018.8500632.

[16] T. Ponn, M. Breitfuß, X. Yu and F. Diermeyer, „Identification of Challenging Highway-Scenarios for the Safety Validation of Automated Vehicles Based on Real Driving Data," (accepted), in *Fifteenth International Conference on Ecological Vehicles and Renewable Energies (EVER2020)*, 2020.

[17] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick and F. Diermeyer, „Survey on Scenario-Based Safety Assessment of Automated Vehicles," (shared co-first authorship), *IEEE Access*, vol. 8, pp. 87456–87477, 2020.

[18] DIN SAE SPEC 91381. „*Terms and Definitions Related to Testing of Automated Vehicle Technologies*," 2019.

[19] P. Junietz, U. Steininger and H. Winner, „Macroscopic safety requirements for highly automated driving," *Transportation research record*, vol. 2673, no. 3, pp. 1–10, 2019.

[20] P. M. Junietz, „Microscopic and Macroscopic Risk Metrics for the Safety Validation of Automated Driving," Dissertation, Technical University of Darmstadt, Darmstadt, 2019, doi: 10.25534/tuprints-00009282.

[21] J. Antona-Makoshi, N. Uchida, K. Yamaaki, K. Ozawa, E. Kitahara and S. Taniguchi, „Development of a Safety Assurance Process for Autonomous Vehicles in Japan," in *26th International Technical Conference and exhibition on the Enhanced Safety of Vehicles (ESV)*, 2019.

[22] CETRAN. „*Centre of Excellence for Testing & Research of Autonomous Vehicles*," Available: http://cetran.sg/ [visited on 10/06/2020].

[23] H. Abbas, M. E. O'Kelly, A. Rodionova and R. Mangharam, „A Driver's License Test for Driverless Vehicles," *Mechanical Engineering*, vol. 139, no. 12, S13, 2017.

[24] M. Althoff, O. Stursberg and M. Buss, „Reachability analysis of linear systems with uncertain parameters and inputs," in *2007 46th IEEE Conference on Decision and Control*, 2007, pp. 726–732, isbn: 978-1-4244-1497-0.

[25] M. Althoff, „Reachability Analysis and its Application to the Safety Assessment of Autonomous Cars," PhD thesis, Technical University of Munich, Munich, 2010.

[26] M. Althoff and J. M. Dolan, „Set-based computation of vehicle behaviors for the online verification of autonomous vehicles," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011, pp. 1162–1167, isbn: 978-1-4577-2197-7.

[27] 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, 2014.

[28] N. Aréchiga, S. M. Loos, A. Platzer and B. H. Krogh, „Using theorem provers to guarantee closed-loop system properties," in *2012 American Control Conference (ACC)*, 2012, pp. 3573–3580, isbn: 978-1-4577-1096-4.

[29] N. Aréchiga, „Specifying Safety of Autonomous Vehicles in Signal Temporal Logic," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 58–63.

[30] F. Gruber and M. Althoff, „Anytime Safety Verification of Autonomous Vehicles," in *2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 1708–1714.

[31] B. Johnson, F. Havlak, H. Kress-Gazit and M. Campbell, „Experimental Evaluation and Formal Analysis of High-Level Tasks with Dynamic Obstacle Anticipation on a Full-Sized Autonomous Vehicle," *Journal of Field Robotics*, vol. 34, no. 5, pp. 897–911, 2017.

[32] S. M. Loos, A. Platzer and L. Nistor, „Adaptive Cruise Control: Hybrid, Distributed, and Now Formally Verified," in *FM 2011*, M. Butler and W. Schulte, ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, isbn: 978-3-642-21436-3.

[33] P. Nilsson, O. Hussien, A. Balkan, Y. Chen, A. D. Ames, J. W. Grizzle, N. Ozay, H. Peng and P. Tabuada, „Correct-by-Construction Adaptive Cruise Control: Two Approaches," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1294–1307, 2016.

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

[35] M. E. O'Kelly, H. Abbas, S. Gao, S. Kato, S. Shiraishi and R. Mangharam, „APEX: Autonomous Vehicle Plan Verification and Execution," in *SAE 2016 World Congress and Exhibition*, 2016.

[36] M. O'Kelly, H. Abbas and R. Mangharam, „Computer-aided design for safe autonomous vehicles," in *2017 Resilience Week (RWS)*, 2017, pp. 90–96, isbn: 978-1-5090-6055-9.

[37] C. Pek, P. Zahn and M. Althoff, „Verifying the safety of lane change maneuvers of self-driving vehicles based on formalized traffic rules," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1477–1483, isbn: 978-1-5090-4804-5.

[38] A. Rizaldi and M. Althoff, „Formalising Traffic Rules for Accountability of Autonomous Vehicles," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 1658–1665, isbn: 978-1-4673-6596-3.

[39] B. Schürmann, D. Heß, J. Eilbrecht, O. Stursberg, F. Koster and M. Althoff, „Ensuring drivability of planned motions using formal methods," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–8, isbn: 978-1-5386-1526-3.

[40] S. Shalev-Shwartz, S. Shammah and A. Shashua, „On a Formal Model of Safe and Scalable Self-driving Cars," *arXiv preprint arXiv:1708.06374*, 2017. Available: http://arxiv.org/abs/1708.06374 [visited on 10/06/2020].

[41] H. Täubig, U. Frese, C. Hertzberg, C. Lüth, S. Mohr, E. Vorobev and D. Walter, „Guaranteeing functional safety: Design for provability and computer-aided verification," *Autonomous Robots*, vol. 32, no. 3, pp. 303–331, 2012.

[42] C. E. Tuncali, J. Kapinski, H. Ito and J. V. Deshmukh, „Reasoning about Safety of Learning-Enabled Components in Autonomous Cyber-physical Systems," in *2018 Design Automation Conference (DAC)*, 2018.

[43] B. Vanholme, D. Gruyer, B. Lusetti, S. Glaser and S. Mammar, „Highly Automated Driving on Highways Based on Legal Safety," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 333–347, 2013.

[44] T. Wongpiromsarn, U. Topcu and R. M. Murray, „Receding Horizon Temporal Logic Planning," *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2817–2830, 2012.

[45] N. Kalra and S. M. Paddock, „Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016. Available: http://www.sciencedirect.com/science/article/pii/S0965856416302129.

[46] W. Wachenfeld and H. Winner, „The Release of Autonomous Vehicles," in *Autonomous Driving: Technical, Legal and Social Aspects*, M. Maurer, J. C. Gerdes, B. Lenz and H. Winner, ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 425–449, isbn: 978-3-662-48847-8. Available: https://doi.org/10.1007/978-3-662-48847-8_21.

[47] C. Wang and H. Winner, „Overcoming Challenges of Validation Automated Driving and Identification of Critical Scenarios," in *IEEE Intelligent Transportation Systems Conference - ITSC*, 2019, pp. 2639–2644, isbn: 978-1-5386-7024-8.

[48] Daimler. „*Daimler and Bosch*," 2019. Available: https://www.daimler.com/innovation/case/autonomous/pilot-city-san-jose.html [visited on 10/06/2020].

[49] S. Kitajima, K. Shimono, J. Tajima, J. Antona-Makoshi and N. Uchida, „Multi-agent traffic simulations to estimate the impact of automated technologies on safety," *Traffic injury prevention*, vol. 20, no. sup1, pp. 58–64, 2019.

[50] C. Roesener, M. Harth, H. Weber, J. Josten and L. Eckstein, „Modelling Human Driver Performance for Safety Assessment of Road Vehicle Automation," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 735–741.

[51] M. Saraoglu, A. Morozov and K. Janschek, „MOBATSim: MOdel-Based Autonomous Traffic Simulation Framework for Fault-Error-Failure Chain Analysis," *IFAC-PapersOnLine*, vol. 52, no. 8, pp. 239–244, 2019. Available: http://www.sciencedirect.com/science/article/pii/S2405896319304100.

[52] Forschungsgesellschaft für Straßen- und Verkehrswesen. „*Richtlinien für die Anlage von Autobahnen*," 2008.

[53] J. Guo, U. Kurup and M. Shah, „Is It Safe to Drive? An Overview of Factors, Metrics, and Datasets for Driveability Assessment in Autonomous Driving," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–17, 2019.

[54] Y. Kang, H. Yin and C. Berger, „Test Your Self-Driving Algorithm: An Overview of Publicly Available Driving Datasets and Virtual Testing Environments," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 171–185, 2019.

[55] J. Zhu, W. Wang and D. Zhao, „A Tempt to Unify Heterogeneous Driving Databases using Traffic Primitives," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2052–2057.

[56] R. Krajewski, J. Bock, L. Kloeker and L. Eckstein, „The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2118–2125.

[57] W. Chen and L. Kloul, „An Ontology-based Approach to Generate the Advanced Driver Assistance Use Cases of Highway Traffic," in *10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2018. Available: https://hal.archives-ouvertes.fr/hal-01835139.

[58] S. Geyer, M. Baltzer, B. Franz, S. Hakuli, M. Kauer, M. Kienle, S. Meier, T. Weißgerber, K. Bengler, R. Bruder, F. Flemisch and H. Winner, „Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance," *IET Intelligent Transport Systems*, vol. 8, no. 3, pp. 183–189, 2014.

[59] F. Klueck, Y. Li, M. Nica, J. Tao and F. Wotawa, „Using Ontologies for Test Suites Generation for Automated and Autonomous Driving Functions," in *2018 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2018, pp. 118–123.

[60] F. Wotawa and Y. Li, „From Ontologies to Input Models for Combinatorial Testing," in *Testing Software and Systems*, 2018, pp. 155–170, isbn: 978-3-319-99927-2.

[61] Yihao Li, Jianbo Tao and Franz Wotawa, „Ontology-based test generation for automated and autonomous driving functions," *Information and Software Technology*, vol. 117, p. 106200, 2020. Available: http://www.sciencedirect.com/science/article/pii/S0950584918302271.

[62] J.-A. Bolte, A. Bär, D. Lipinski and T. Fingscheidt, „Towards Corner Case Detection for Autonomous Driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 438–445.

[63] I. R. Jenkins, L. O. Gee, A. Knauss, H. Yinz and J. Schroeder, „Accident Scenario Generation with Recurrent Neural Networks," in *2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 3340–3345.

[64] S. Jesenski, J. E. Stellet, F. Schiegg and J. M. Zöllner, „Generation of Scenes in Intersections for the Validation of Highly Automated Driving Functions," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 502–509.

[65] R. Krajewski, T. Moers, D. Nerger and L. Eckstein, „Data-Driven Maneuver Modeling using Generative Adversarial Networks and Variational Autoencoders for Safety Validation of Highly Automated Vehicles," in *2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2383–2390.

[66] J. Langner, J. Bach, L. Ries, S. Otten, M. Holzäpfel and E. Sax, „Estimating the Uniqueness of Test Scenarios derived from Recorded Real-World-Driving-Data using Autoencoders," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018.

[67] J. Bach, J. Langner, S. Otten, E. Sax and M. Holzäpfel, „Test scenario selection for system-level verification and validation of geolocation-dependent automotive control systems," in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2017, pp. 203–210, isbn: 978-1-5386-0774-9.

[68] H. Beglerovic, T. Schloemicher, S. Metzner and M. Horn, „Deep Learning Applied to Scenario Classification for Lane-Keep-Assist Systems," *Applied Sciences*, vol. 8, no. 12, p. 2590, 2018.

[69] B. Dávid, G. Láncz and Hunyady Gergely, „Highway Situation Analysis with Scenario Classification and Neural Network based Risk Estimation for Autonomous Vehicles," in *2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2019, pp. 375–380.

[70] A. Erdogan, B. Ugranli, E. Adali, A. Sentas, E. Mungan, E. Kaplan and A. Leitner, „Real-World Maneuver Extraction for Autonomous Vehicle Validation: A Comparative Study," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 267–272.

[71] R. Gruner, P. Henzler, G. Hinz, C. Eckstein and A. Knoll, „Spatiotemporal representation of driving scenarios and classification using neural networks," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1782–1788, isbn: 978-1-5090-4804-5.

[72] F. Kruber, J. Wurst, E. S. Morales, S. Chakraborty and M. Botsch, „Unsupervised and Supervised Learning with the Random Forest Algorithm for Traffic Scenario Clustering and Classification," in *30th IEEE Intelligent Vehicles Symposium*, 2019, pp. 2463–2470, isbn: 978-1-7281-0560-4.

[73] F. Kruber, J. Wurst and M. Botsch, „An Unsupervised Random Forest Clustering Technique for Automatic Traffic Scenario Categorization," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2811–2818.

[74] W. Wang and D. Zhao, „Extracting Traffic Primitives Directly From Naturalistically Logged Data for Self-Driving Applications," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1223–1229, 2018.

[75] H. Watanabe, T. Maly, J. Wallner, T. Dirndorfer, M. Mai and G. Prokop, „Methodology of Scenario Clustering for Predictive Safety Functions," in *9. Tagung Automatisiertes Fahren*, 2019.

[76] H. Weber, J. Bock, J. Klimke, C. Roesener, J. Hiller, R. Krajewski, A. Zlocki and L. Eckstein, „A framework for definition of logical scenarios for safety assurance of automated driving," *Traffic injury prevention*, vol. 20, no. sup1, S65–S70, 2019.

[77] E. de Gelder and J.-P. Paardekooper, „Assessment of Automated Driving Systems using real-life scenarios," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 589–594.

[78] E. de Gelder, J.-P. Paardekooper, O. Op den Camp and B. de Schutter, „Safety assessment of automated vehicles: how to determine whether we have collected enough field data?," *Traffic injury prevention*, vol. 20, no. sup1, S162–S170, 2019.

[79] L. Hartjen, R. Philipp, F. Schuldt, F. Howar and B. Friedrich, „Classification of Driving Maneuvers in Urban Traffic for Parametrization of Test Scenarios," in *9. Tagung Automatisiertes Fahren*, 2019.

[80] J. Zhou and L. del Re, „Identification of critical cases of ADAS safety by FOT based parameterization of a catalogue," in *2017 11th Asian Control Conference (ASCC)*, 2017, pp. 453–458, isbn: 978-1-5090-1573-3.

[81] M. R. Zofka, F. Kuhnt, R. Kohlhaas, C. Rist, T. Schamm and J. M. Zöllner, „Data-driven simulation and parametrization of traffic scenarios for the development of advanced driver assistance systems," in *18th International Conference on Information Fusion*, 2015.

[82] A. Pütz, A. Zlocki, J. Bock and L. Eckstein, „System validation of highly automated vehicles with a database of relevant traffic scenarios," in *12th ITS European Congress*, 2017.

[83] A. Pütz, A. Zlocki, J. Küfen, J. Bock and L. Eckstein, „Database Approach for the Sign-Off Process of Highly Automated Vehicles," in *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*, 2017.

[84] S. Feng, Y. Feng, C. Yu, Y. Zhang and H. X. Liu, „Testing Scenario Library Generation for Connected and Automated Vehicles, Part I: Methodology," *arXiv preprint arXiv:1905.03419*, 2019.

[85] S. Feng, Y. Feng, H. Sun, S. Bao, A. Misra, Y. Zhang and H. X. Liu, „Testing Scenario Library Generation for Connected and Automated Vehicles, Part II: Case Studies," *arXiv preprint arXiv:1905.03428*, 2019.

[86] M. Althoff, M. Koschi and S. Manzinger, „CommonRoad: Composable benchmarks for motion planning on roads," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 719–726, isbn: 978-1-5090-4804-5.

[87] D. Zhao, Y. Guo and Y. J. Jia, „TrafficNet: An open naturalistic driving scenario library," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–8.

[88] M. Althoff and J. M. Dolan, „Reachability computation of low-order models for the safety verification of high-order road vehicle models," in *2012 American Control Conference (ACC)*, 2012, pp. 3559–3566, isbn: 978-1-4577-1096-4.

[89] H. Beglerovic, A. Ravi, N. Wikström, H.-M. Koegeler, A. Leitner and J. Holzinger, „Model-based safety validation of the automated driving function highway pilot," in *8th International Munich Chassis Symposium 2017* (Proceedings), P. P. E. Pfeffer, ed. Wiesbaden: Springer Fachmedien Wiesbaden, 2017, pp. 309–329, isbn: 978-3-658-18458-2.

[90] L. Huang, Q. Xia, F. Xie, H.-L. Xiu and H. Shu, „Study on the Test Scenarios of Level 2 Automated Vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 49–54.

[91] B. Kim, A. Jarandikar, J. Shum, S. Shiraishi and M. Yamaura, „The SMT-based automatic road network generation in vehicle simulation environment," in *Proceedings of the 13th International Conference on Embedded Software - EMSOFT '16*, 2016, pp. 1–10, isbn: 9781450344852.

[92] T. Ponn, D. Fratzke, C. Gnandt and M. Lienkamp, „Towards Certification of Autonomous Driving: Systematic Test Case Generation for a Comprehensive but Economically-Feasible Assessment of Lane Keeping Assist Algorithms," in *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2019)*, 2019, pp. 333–342.

[93] E. Rocklage, H. Kraft, A. Karatas and J. Seewig, „Automated scenario generation for regression testing of autonomous vehicles," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 476–483.

[94] F. Xie, T. Chen, Q. Xia, L. Huang and H. Shu, „Study on the Controlled Field Test Scenarios of Automated Vehicles," in *SAE Technical Paper*, 2018. Available: https://doi.org/10.4271/2018-01-1633.

[95] Y. Akagi, R. Kato, S. Kitajima, J. Antona-Makoshi and N. Uchida, „A Risk-index based Sampling Method to Generate Scenarios for the Evaluation of Automated Driving Vehicle Safety *," in *IEEE Intelligent Transportation Systems Conference - ITSC*, 2019, pp. 667–672, isbn: 978-1-5386-7024-8.

[96] M. Arief, P. Glynn and D. Zhao, „An Accelerated Approach to Safely and Efficiently Test Pre-Production Autonomous Vehicles on Public Streets," in *2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2006–2011.

[97] D. Åsljung, J. Nilsson and J. Fredriksson, „Comparing Collision Threat Measures for Verification of Autonomous Vehicles using Extreme Value Theory," *IFAC-PapersOnLine*, vol. 49, no. 15, pp. 57–62, 2016. Available: http://www.sciencedirect.com/science/article/pii/S2405896316309855.

[98] D. Åsljung, J. Nilsson and J. Fredriksson, „Using Extreme Value Theory for Vehicle Level Safety Validation and Implications for Autonomous Vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 4, pp. 288–297, 2017.

[99] D. Åsljung, M. Westlund and J. Fredriksson, „A Probabilistic Framework for Collision Probability Estimation and an Analysis of the Discretization Precision," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 52–57.

[100] Z. Huang, D. Zhao, H. Lam, D. J. LeBlanc and H. Peng, „Evaluation of automated vehicles in the frontal cut-in scenario — An enhanced approach using piecewise mixture models," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 197–202.

[101] Z. Huang, H. Lam and D. Zhao, „Sequential experimentation to efficiently test automated vehicles," in *2017 Winter Simulation Conference (WSC)*, 2017, pp. 3078–3089.

[102] Z. Huang, H. Lam and D. Zhao, „Towards affordable on-track testing for autonomous vehicle — A Kriging-based statistical approach," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–6.

[103] Z. Huang, Y. Guo, M. Arief, H. Lam and D. Zhao, „A Versatile Approach to Evaluating and Testing Automated Vehicles based on Kernel Methods," in *2018 Annual American Control Conference (ACC)*, 2018, pp. 4796–4802.

[104] Z. Huang, H. Lam, D. J. LeBlanc and D. Zhao, „Accelerated Evaluation of Automated Vehicles Using Piecewise Mixture Models," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2017.

[105] Z. Huang, H. Lam and D. Zhao, „An accelerated testing approach for automated vehicles with background traffic described by joint distributions," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 933–938, isbn: 978-1-5386-1526-3.

[106] Z. Huang, M. Arief, H. Lam and D. Zhao, „Synthesis of Different Autonomous Vehicles Test Approaches," in *2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2000–2005.

[107] Z. Huang, M. Arief, H. Lam and D. Zhao, „Evaluation Uncertainty in Data-Driven Self-Driving Testing," in *2019 IEEE 22th International Conference on Intelligent Transportation Systems (ITSC)*, 2019.

[108] M. O'Kelly, A. Sinha, H. Namkoong, R. Tedrake and J. C. Duchi, „Scalable end-to-end autonomous vehicle testing via rare-event simulation," in *Advances in Neural Information Processing Systems*, 2018, pp. 9827–9838.

[109] S. P. Olivares, N. Rebernik, A. Eichberger and E. Stadlober, „Virtual stochastic testing of advanced driver assistance systems," in *Advanced Microsystems for Automotive Applications 2015* Springer, 2016, pp. 25–35.

[110] X. Wang, H. Peng and D. Zhao, eds. „*Combining Reachability Analysis and Importance Sampling for Accelerated Evaluation of Highly Automated Vehicles at Pedestrian Crossing*," vol. 3. Dynamic Systems and Control Conference. 2019.

[111] S. Zhang, H. Peng, D. Zhao and E. Tseng, „Accelerated Evaluation of Autonomous Vehicles in the Lane Change Scenario Based on Subset Simulation Technique," in *2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 3935–3940.

[112] D. Zhao, „Accelerated Evaluation of Automated Vehicles," PhD thesis, University of Michigan, Michigan, 2016.

[113] D. Zhao, H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa and C. S. Pan, „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.

[114] 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.

[115] F. Fahrenkrog, L. Wang, T. Platzer, A. Fries, F. Raisch and K. Kompaß, „Prospective Effectiveness Safety Assessment of Automated Driving Functions – From The Method to the Results," in *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2019.

[116] P. Feig, V. Labenski, T. Leonhardt and J. Schatz, „Assessment of Technical Requirements for Level 3 and Beyond Automated Driving Systems Based on Naturalistic Driving and Accident Data Analysis," in *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2019.

[117] J. So, I. Park, J. Wee, S. Park and I. Yun, „Generating Traffic Safety Test Scenarios for Automated Vehicles using a Big Data Technique," *KSCE Journal of Civil Engineering*, 2019. Available: https://doi.org/10.1007/s12205-019-1287-4.

[118] L. Stark, M. Düring, S. Schoenawa, J. E. Maschke and C. M. Do, „Quantifying Vision Zero: Crash avoidance in rural and motorway accident scenarios by combination of ACC, AEB, and LKS projected to German accident occurrence," *Traffic injury prevention*, vol. 20, no. sup1, S126–S132, 2019.

[119] L. Stark, S. Obst, S. Schoenawa and M. Düring, „Towards Vision Zero: Addressing White Spots by Accident Data based ADAS Design and Evaluation," in *2019 IEEE International Conference of Vehicular Electronics and Safety (ICVES)*, 2019, pp. 1–6.

[120] M. Althoff and S. Lutz, „Automatic Generation of Safety-Critical Test Scenarios for Collision Avoidance of Road Vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018.

[121] M. Klischat and M. Althoff, „Generating Critical Test Scenarios for Automated Vehicles with Evolutionary Algorithms," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 2352–2358.

[122] A. Pierson, W. Schwarting, S. Karaman and D. Rus, „Learning Risk Level Set Parameters from Data Sets for Safer Driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 273–280.

[123] D. Stumper and K. Dietmayer, „Towards Criticality Characterization of Situational Space," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 3378–3382, isbn: 978-1-7281-0321-1. doi: 10.1109/ITSC.2018.8569505.

[124] F. Gao, J. Duan, Y. He and Z. Wang, „A test scenario automatic generation strategy for intelligent driving systems," *Mathematical Problems in Engineering*, vol. 2019, 2019.

[125] P. Koopman and F. Fratrik, „How Many Operational Design Domains, Objects, and Events?," in *SafeAI*, 2019.

[126] T. Ponn, C. Gnandt and F. Diermeyer, „An Optimization-Based Method to Identify Relevant Scenarios for Type Approval of Automated Vehicles," in *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2019.

[127] T. Ponn, A. Schwab, F. Diermeyer, C. Gnandt and J. Záhorský, „A Method for the Selection of Challenging Driving Scenarios for Automated Vehicles Based on an Objective Characterization of the Driving Behavior," in *9. Tagung Automatisiertes Fahren*, 2019.

[128] Y. Qi, Y. Luo, K. Li, W. Kong and Y. Wang, „A Trajectory-Based Method for Scenario Analysis and Test Effort Reduction for Highly Automated Vehicle," in *SAE Technical Paper Series*, 2019.

[129] J. Wang, C. Zhang, Y. Liu and Q. Zhang, „Traffic Sensory Data Classification by Quantifying Scenario Complexity," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1543–1548.

[130] Q. Xia, J. Duan, F. Gao, T. Chen and C. Yang, „Automatic Generation Method of Test Scenario for ADAS Based on Complexity," in *SAE Technical Paper*, 2017. Available: https://doi.org/10.4271/2017-01-1992.

[131] Q. Xia, J. Duan, F. Gao, Q. Hu and Y. He, „Test Scenario Design for Intelligent Driving System Ensuring Coverage and Effectiveness," *International Journal of Automotive Technology*, vol. 19, no. 4, pp. 751–758, 2018.

[132] C. Zhang, Y. Liu, Q. Zhang and Wang Le, „A Graded Offline Evaluation Framework for Intelligent Vehicle's Cognitive Ability," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 320–325.

[133] H. Abbas, M. O'Kelly, A. Rodionova and R. Mangharam, „Safe At Any Speed: A Simulation-Based Test Harness for Autonomous Vehicles," in *Seventh Workshop on Design, Modeling and Evaluation of Cyber Physical Systems (CyPhy'17)*, 2017.

[134] R. B. Abdessalem, S. Nejati, L. C. Briand and T. Stifter, „Testing advanced driver assistance systems using multi-objective search and neural networks," in *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering - ASE 2016*, 2016, pp. 63–74, isbn: 9781450338455.

[135] H. Beglerovic, M. Stolz and M. Horn, „Testing of autonomous vehicles using surrogate models and stochastic optimization," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–6, isbn: 978-1-5386-1526-3.

[136] A. Corso, Du Peter, K. Driggs-Campbell and M. J. Kochenderfer, „Adaptive Stress Testing with Reward Augmentation for Autonomous Vehicle Validation," in *2019 IEEE 22th International Conference on Intelligent Transportation Systems (ITSC)*, 2019, pp. 163–168.

[137] H. Felbinger, F. Klück and M. Zimmermann, „Comparing two systematic approaches for testing automated driving functions," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE) Proceedings*, 2019.

[138] B. Gangopadhyay, S. Khastgir, S. Dey, P. Dasgupta, G. Montana and P. Jennings, „Identification of Test Cases for Automated Driving Systems Using Bayesian Optimization," in *2019 IEEE 22th International Conference on Intelligent Transportation Systems (ITSC)*, 2019.

[139] M. Koren, S. Alsaif, R. Lee and M. J. Kochenderfer, „Adaptive Stress Testing for Autonomous Vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1898–1904.

[140] M. Koschi, C. Pek, S. Maierhofer and M. Althoff, „Computationally Efficient Safety Falsification of Adaptive Cruise Control Systems," in *IEEE Intelligent Transportation Systems Conference - ITSC*, 2019, pp. 2879–2886, isbn: 978-1-5386-7024-8.

[141] R. Lee, M. J. Kochenderfer, O. J. Mengshoel, G. P. Brat and M. P. Owen, „Adaptive stress testing of airborne collision avoidance systems," in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, 2015, pp. 1–24, isbn: 978-1-4799-8940-9.

[142] R. Lee, O. J. Mengshoel, A. Saksena, R. Gardner, D. Genin, J. Brush and M. J. Kochenderfer, „Differential adaptive stress testing of airborne collision avoidance systems," in *2018 AIAA Modeling and Simulation Technologies Conference*, 2018.

[143] G. E. Mullins, P. G. Stankiewicz and S. K. Gupta, „Automated generation of diverse and challenging scenarios for test and evaluation of autonomous vehicles," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1443–1450.

[144] G. E. Mullins, P. G. Stankiewicz, R. C. Hawthorne and S. K. Gupta, „Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles," *Journal of Systems and Software*, vol. 137, pp. 197–215, 2018.

[145] G. E. Mullins, „Adaptive Sampling Methods for Testing Autonomous Systems," PhD thesis, Department of Mechanical Engineering, University of Maryland, College Park, 2018.

[146] M. Nabhan, M. Schoenauer, Y. Tourbier and H. Hage, „Optimizing coverage of simulated driving scenarios for the autonomous vehicle," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE) Proceedings*, 2019.

[147] C. E. Tuncali, T. P. Pavlic and G. Fainekos, „Utilizing S-TaLiRo as an Automatic Test Generation Framework for Autonomous Vehicles," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 1470–1475, isbn: 978-1-5090-1889-5.

[148] C. E. Tuncali, S. Yaghoubi, T. P. Pavlic and G. Fainekos, „Functional gradient descent optimization for automatic test case generation for vehicle controllers," in *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*, 2017, pp. 1059–1064, isbn: 978-1-5090-6781-7.

[149] C. Tuncali, G. Fainekos, H. Ito and J. Kapinski, „Simulation-based Adversarial Test Generation for Autonomous Vehicles with Machine Learning Components," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018.

[150] C. E. Tuncali, G. Fainekos, D. Prokhorov, H. Ito and J. Kapinski, „Requirements-driven Test Generation for Autonomous Vehicles with Machine Learning Components," *IEEE Transactions on Intelligent Vehicles*, 2019.

[151] C. E. Tuncali, „Search-based Test Generation for Automated Driving Systems: From Perception to Control Logic," PhD thesis, Arizona State University, Tempe, 2019.

[152] J. Zhou and L. d. Re, „Reduced Complexity Safety Testing for ADAS & ADF," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5985–5990, 2017. Available: http://www.sciencedirect.com/science/article/pii/S2405896317317755.

[153] C. E. Tuncali and G. Fainekos, „Rapidly-exploring Random Trees for Testing Automated Vehicles," in *2019 IEEE 22th International Conference on Intelligent Transportation Systems (ITSC)*, 2019.

[154] S. Riedmaier, J. Nesensohn, C. Gutenkunst, T. Düser, B. Schick and H. Abdellatif, „Validation of X-in-the-Loop Approaches for Virtual Homologation of Automated Driving Functions," in *11th Graz Symposium Virtual Vehicle (GSVF)*, 2018.

[155] S. Hallerbach, Y. Xia, U. Eberle and F. Koester, „Simulation-Based Identification of Critical Scenarios for Cooperative and Automated Vehicles," *SAE International Journal of Connected and Automated Vehicles*, vol. 1, no. 2, pp. 93–106, 2018.

[156] J. C. Kirchhof, E. Kusmenko, B. Rumpe and H. Zhang, „Simulation as a Service for Cooperative Vehicles," in *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*, 2019, pp. 28–37.

[157] J. C. Hayward, „Near-Miss Determination Through Use of a Scale of Danger," *Highway Research Record*, no. 384, 1972.

[158] P. Junietz, F. Bonakdar, B. Klamann and H. Winner, „Criticality Metric for the Safety Validation of Automated Driving using Model Predictive Trajectory Optimization," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 60–65.

[159] M. Schreier, V. Willert and J. Adamy, „An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2751–2766, 2016.

[160] W. Wachenfeld, P. Junietz, R. Wenzel and H. Winner, „The worst-time-to-collision metric for situation identification," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 729–734.

[161] S. M. S. Mahmud, L. Ferreira, M. S. Hoque and A. Tavassoli, „Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS Research*, vol. 41, no. 4, pp. 153–163, 2017. Available: http://www.sciencedirect.com/science/article/pii/S0386111217300286.

[162] UNECE. „*Vehicle Regulations*," Available: https://www.unece.org/trans/main/welcwp29.html [visited on 10/06/2020].

[163] UNECE. „*Text of the 1958 Agreement*," Available: http://www.unece.org/trans/main/wp29/wp29regs.html [visited on 10/06/2020].

[164] UNECE. „*UN Transport Agreements and Conventions*," Available: http://www.unece.org/trans/maps/un-transport-agreements-and-conventions-18.html [visited on 10/06/2020].

[165] NTSB. „*Docket And Docket Items*," Available: https://dms.ntsb.gov/pubdms/search/hitlist.cfm?docketID=62978&CurrentPage=1&EndRow=15&StartRow=1&order=1&sort=0&TXTSEARCHT= [visited on 10/06/2020].

[166] European Commission. „*New and improved car emissions tests become mandatory on 1 September*," 2017. Available: https://ec.europa.eu/commission/presscorner/detail/en/IP_17_2822 [visited on 10/06/2020].

[167] European Commission. „*The European Commission'sscienceandknowledgeservice*," 2019. Available: https://wiki.unece.org/download/attachments/81888491/VMAD-03-06%20JRC%20Safety%20Assessment%20.pdf?api=v2 [visited on 10/06/2020].

[168] UNECE. „*Introduction*," Available: https://www.unece.org/trans/main/wp29/meeting_docs_grva.html [visited on 10/06/2020].

[169] VMAD. „New Assessment/Test Method for Automated Driving," 2020. [Online]. Available: https://wiki.unece.org/download/attachments/94046591/VMAD-06-03%20NATM%20Master%20Document.docx?api=v2 [visited on 10/06/2020].

[170] VMAD. „New Assessment/Test Method," 2020. [Online]. Available: https://wiki.unece.org/download/attachments/101555220/VMAD-07-15%20%20NATM%20presentation.pptx?api=v2 [visited on 10/06/2020].

[171] UNECE. „UN Regulation on Automated Lane Keeping Systems is milestone for safe introduction of automated vehicles in traffic," 2020. [Online]. Available: https://www.unece.org/?id=54669 [visited on 10/06/2020].

[172] UNECE. „Adopted Proposals," 2020. [Online]. Available: http://www.unece.org/fileadmin/DAM/trans/doc/2020/wp29/June_2020-Adopted-proposals-_-Situation-of-EIF.pdf [visited on 10/06/2020].

[173] F. Batsch, S. Kanarachos, M. Cheah, R. Ponticelli and M. Blundell, „A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles," *Journal of Intelligent Transportation Systems*, vol. 9, no. 6, pp. 1–20, 2020, doi: 10.1080/15472450.2020.1738231.

[174] T. Ponn, F. Müller and F. Diermeyer, „Systematic Analysis of the Sensor Coverage of Automated Vehicles Using Phenomenological Sensor Models," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 1000–1006.

[175] T. Ponn, T. Kröger and F. Diermeyer, „Performance Analysis of Camera-based Object Detection for Automated Vehicles," (shared co-first authorship), *Sensors*, vol. 20, no. 13, 2020, doi: 10.3390/s20133699.

[176] T. Ponn, T. Lanz and F. Diermeyer, „Automatic Generation of Road Geometries to Create Challenging Scenarios for Automated Vehicles Based on the Sensor Setup," (accepted), in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020.

[177] F. Schiegg, I. Llatser and T. Michalke, „Object Detection Probability for Highly Automated Vehicles: An Analytical Sensor Model," in *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems*, 2019, pp. 223–231, isbn: 978-989-758-374-2. doi: 10.5220/0007767602230231.

[178] M. Berk, „Safety Assessment of Environment Perception in Automated Driving Vehicles," Dissertation, Technische Universität München, München, 2019.

[179] S. Bernsteiner, Z. Magosi, D. Lindvai-Soos and A. Eichberger, „Radar Sensor Model for the Virtual Development Process," *ATZelektronik worldwide*, vol. 10, no. 2, pp. 46–52, 2015, doi: 10.1007/s38314-015-0521-1.

[180] P. Cao, „Modeling Active Perception Sensors for Real-Time Virtual Validation of Automated Driving Systems," Dissertation, Fachbereich Maschinenbau, Technische Universität Darmstadt, Darmstadt, 2017.

[181] M. Feilhauer and J. Häring, „A real-time capable multi-sensor model to validate ADAS in a virtual environment," in *3. Internationale ATZ Tagung: Automatisiertes Fahren*, 2017.

[182] R. Schubert, N. Mattern and R. Bours, „Simulation of Sensor Models for the Evaluation of Advanced Driver Assistance Systems," *ATZelektronik worldwide*, no. 03/2014, pp. 26–29, 2014.

[183] J. Steinbaeck, C. Steger, G. Holweg and N. Druml, „Next generation radar sensors in automotive sensor fusion systems," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2017, pp. 1–6, doi: 10.1109/SDF.2017.8126389.

[184] J. Schrepfer, V. Picron, J. Mathes and H. Barth, „Automated Driving and its Sensors under Test," *ATZ worldwide*, vol. 120, no. 1, pp. 28–35, 2018, doi: 10.1007/s38311-017-0160-7.

[185] S. Hasirlioglu and A. Riener, „Introduction to rain and fog attenuation on automotive surround sensors," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, doi: 10.1109/ITSC.2017.8317823.

[186] A. Ludloff, *Praxiswissen Radar und Radarsignalverarbeitung*, Vieweg + Teubner, 1998, isbn: 978-3-663-12326-2.

[187] J. A. Scheer, „The Radar Range Equation," in *Principles of modern radar*, W. A. Holm, M. A. Richards and J. A. Scheer, ed. Raleigh, NC: SciTech Publ, 2015, pp. 59–86, isbn: 9781891121524.

[188] NUSCENES. „NuScenes," 2020. [Online]. Available: https://www.nuscenes.org/ [visited on 10/06/2020].

[189] S. M. Lundberg and S.-I. Lee, „A Unified Approach to Interpreting Model Predictions," *31st Conference on Neural Information Processing Systems (NIPS 2017)*, pp. 4768–4777, 2017. Available: https://arxiv.org/pdf/1705.07874.pdf.

[190] W. Ertel, *Introduction to Artificial Intelligence*, (Undergraduate Topics in Computer Science), 2nd ed. 2017, Cham, Springer International Publishing, 2017, isbn: 978-3-319-58487-4. doi: 10.1007/978-3-319-58487-4.

[191] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, „Focal Loss for Dense Object Detection," *2020 IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020, doi: 10.1109/TPAMI.2018.2858826. Available: https://arxiv.org/pdf/1708.02002.pdf.

[192] Z. Zou, Shi, Zhenwei, Y. Guo and J. Ye, „Object Detection in 20 Years: A Survey," *ArXiv:1905.05055*, 2019. Available: https://arxiv.org/pdf/1905.05055.pdf.

[193] J. Holzinger, P. Schöggl, M. Schrauf and E. Bogner, „Objective Assessment of Driveability while Automated Driving," *ATZ worldwide*, vol. 116, no. 12, pp. 24–29, 2014, doi: 10.1007/s38311-014-0250-8.

[194] J. Holzinger and E. Bogner, „Objective Assessment of Advanced Driver Assistance Systems," *ATZ worldwide*, vol. 119, no. 9, pp. 16–19, 2017, doi: 10.1007/s38311-017-0089-x.

[195] G. Büyükyildiz, O. Pion, C. Hildebrandt, M. Sedlmayr, R. Henze and F. Küçükay, „Identification of the driving style for the adaptation of assistance systems," *International Journal of Vehicle Autonomous Systems*, vol. 13, no. 3, pp. 244–260, 2017, doi: 10.1504/IJVAS.2017.083515.

[196] J. Karjanto, N. Md. Yusof, J. Terken, F. Delbressine, M. Z. Hassan, M. Rauterberg, S. A. Che Ghani, W. A. Wan Hamzah and A. Alias, „Simulating autonomous driving styles: Accelerations for three road profiles," *MATEC Web of Conferences*, vol. 90, no. 2, pp. 1–16, 2017, doi: 10.1051/matecconf/20179001005.

[197] C. M. Martinez, M. Heucke, F.-Y. Wang, B. Gao and D. Cao, „Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 666–676, 2018, doi: 10.1109/TITS.2017.2706978.

[198] H.-H. Nagel, „A vision of 'vision and language' comprises action: An example from road traffic," *Artificial Intelligence Review*, vol. 8, no. 2, pp. 189–214, 1994, doi: 10.1007/BF00849074. Available: https://doi.org/10.1007/BF00849074.

[199] H. .-H. Nagel, W. Enkelmann and G. Struck, „FhG-Co-driver: From map-guided automatic driving by machine vision to a cooperative driver support," *Mathematical and Computer Modelling*, vol. 22, no. 4, pp. 185–212, 1995, doi: 10.1016/0895-7177(95)00133-M. Available: http://www.sciencedirect.com/science/article/pii/089571779500133M.

[200] Bundesministerium für Verkehr und digitale Infrastruktur. „*Richtlinie für die Prüfung der Bewerber um eine Erlaubnis zum Führen von Kraftfahrzeugen*," 2014.

[201] Bundesministerium für Verkehr und digitale Infrastruktur. „*Fragenkatalog für die theoretische Fahrerlaubnisprüfung*," 2019.

[202] A. Schwab, „Eine Methode zur Auswahl kritischer Fahrszenarien für automatisierte Fahrzeuge anhand einer objektiven Charakterisierung des Fahrverhaltens," Master's Thesis, Technical University of Munich, Munich, 2019.

[203] D. S. Gonzalez, J. S. Dibangoye and C. Laugier, „High-speed highway scene prediction based on driver models learned from demonstrations," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, 2016, pp. 149–155, isbn: 978-1-5090-1889-5. doi: 10.1109/ITSC.2016.7795546.

[204] F. Schuldt, „Ein Beitrag für den methodischen Test von automatisierten Fahrfunktionen mit Hilfe von virtuellen Umgebungen," Dissertation, Technische Universität Braunschweig, Braunschweig, 2017, doi: 10.24355/dbbs.084-201704241210. Available: https://publikationsserver.tu-braunschweig.de/receive/dbbs_mods_00064747.

[205] Federal Highway Administration. „Next Generation Simulation," 2020. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm [visited on 10/06/2020].

[206] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*, Berlin, Heidelberg, Springer-Verlag Berlin Heidelberg, 2008, isbn: 978-3-540-73190-0. doi: 10.1007/978-3-540-73190-0.

[207] M. Treiber, A. Hennecke and D. Helbing, „Congested traffic states in empirical observations and microscopic simulations," *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, vol. 62, no. 2 Pt A, pp. 1805–1824, 2000, doi: 10.1103/physreve.62.1805.

[208] A. Kesting, M. Treiber and D. Helbing, „General Lane-Changing Model MOBIL for Car-Following Models," *Transportation research record*, vol. 1999, no. 1, pp. 86–94, 2007, doi: 10.3141/1999-10.

[209] M. Breitfuß, „Extraction and classification of real traffic scenarios for validation of autonomous vehicles," Semesterarbeit, Technical University of Munich, Munich, 2020.

[210] B. Everitt, *Cluster analysis*, (Wiley series in probability and statistics), 5. ed., Chichester, Wiley, 2011, isbn: 978-0-470-97780-4. doi: 10.1002/9780470977811. Available: http://lib.myilibrary.com/detail.asp?id=367872.

[211] X. Yu, „Method for quantitative evaluation of traffic complexity on the highway," Semesterarbeit, Technical University of Munich, Munich, 2019.

[212] X. Yu, „Improvement and validation of a highway traffic complexity metric for test scenarios of automated vehicles," Master's Thesis, Technical University of Munich, Munich, 2020.

[213] F. Diehl, T. Brunner, M. T. Le and A. Knoll, „Graph Neural Networks for Modelling Traffic Participant Interaction," in *30th IEEE Intelligent Vehicles Symposium : 9-12 June 2019, Paris*, Paris, France, 2019, pp. 695–701, isbn: 978-1-7281-0560-4. doi: 10.1109/IVS.2019.8814066.

[214] R. Schubert, E. Richter and G. Wanielik, „Comparison and evaluation of advanced motion models for vehicle tracking," in *2008 11th International Conference on Information Fusion*, 2008, pp. 1–6.

[215] P. Schörner, L. Töttel, J. Doll and J. M. Zöllner, „Predictive Trajectory Planning in Situations with Hidden Road Users Using Partially Observable Markov Decision Processes," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 2299–2306.

[216] Y. Lin, „Creation of Complex Test Scenarios for Automated Vehicles by Means of Evolutionary Algorithm," Semesterarbeit, Technical University of Munich, Munich, 2020.

[217] The MathWorks. „Global Optimization Toolbox," 2020. [Online]. Available: https://mathworks.com/products/global-optimization.html [visited on 10/06/2020].

[218] T. Toledo and D. Zohar, „Modeling Duration of Lane Changes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, no. 1, pp. 71–78, 2007, doi: 10.3141/1999-08.

[219] A. Loulizi, Y. Bichiou and H. Rakha, „Steady-State Car-Following Time Gaps: An Empirical Study Using Naturalistic Driving Data," *Journal of Advanced Transportation*, vol. 2019, no. 3, pp. 1–9, 2019, doi: 10.1155/2019/7659496.

[220] A. Knapp, M. Neumann, M. Brockmann, R. Walz and T. Winkle. „*Code of Practice for the Design and Evaluation of ADAS*," 2009. Available: https://www.acea.be/uploads/publications/20090831_Code_of_Practice_ADAS.pdf [visited on 10/06/2020].

[221] L. J. Cronbach, „Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951, doi: 10.1007/BF02310555.

[222] William H. Kruskal and W. Allen Wallis, „Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952. Available: http://www.jstor.org/stable/2280779.

[223] MEDCALC. „*Values of the Chi-squared distribution*," Available: https://www.medcalc.org/manual/chi-square-table.php [visited on 10/06/2020].

[224] O. J. Dunn, „Multiple Comparisons Using Rank Sums," *Technometrics*, vol. 6, no. 3, p. 241, 1964, doi: 10.2307/1266041.

[225] The University of Arizona. „*Standard normal distribution*," Available: https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf [visited on 10/06/2020].

[226] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, *Global Sensitivity Analysis. The Primer*, Chichester, UK, John Wiley & Sons, Ltd, 2007, isbn: 9780470725184. doi: 10.1002/9780470725184.

[227] The MathWorks. „Automated Driving Toolbox," 2020. [Online]. Available: https://de.mathworks.com/help/releases/R2019b/driving/index.html [visited on 10/06/2020].

[228] The MathWorks. „Getting Started with Automated Driving Toolbox," 2020. [Online]. Available: https://de.mathworks.com/help/releases/R2019b/driving/getting-started-with-automated-driving-toolbox.html [visited on 10/06/2020].

[229] The MathWorks. „Adaptive Cruise Control with Sensor Fusion," 2020. [Online]. Available: https://de.mathworks.com/help/releases/R2019b/driving/examples/adaptive-cruise-control-with-sensor-fusion.html [visited on 10/06/2020].

[230] The MathWorks. „Lane Keeping Assist with Lane Detection," 2020. [Online]. Available: https://de.mathworks.com/help/releases/R2019b/driving/examples/lane-keeping-assist-with-lane-detection.html [visited on 10/06/2020].

[231] The MathWorks. „Autonomous Emergency Braking with Sensor Fusion,“ 2020. [Online]. Available: https://de.mathworks.com/help/releases/R2019b/driving/examples/autonomous-emergency-braking-with-sensor-fusion.html [visited on 10/06/2020].

[232] The MathWorks. „Vehicle Dynamics Blockset,“ 2020. [Online]. Available: https://de.mathworks.com/help/releases/R2019b/vdynblks/index.html [visited on 10/06/2020].

[233] R. van der Horst and J. Hogema, „Time-to-collision and collision avoidance systems,“ in *6th ICTCT Workshop Safety Evaluation of Traffic Systems: Traffic Conflicts and Other Measures* 1993, pp. 109–121.

Prior Publications

During the development of this dissertation, publications and student theses were written in which partial aspects of this work were presented.

Journals; Scopus/Web of Science listed (peer-reviewed)

[17] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick and F. Diermeyer, „Survey on Scenario-Based Safety Assessment of Automated Vehicles," (shared co-first authorship), *IEEE Access*, vol. 8, pp. 87456–87477, 2020.

[175] T. Ponn, T. Kröger and F. Diermeyer, „Performance Analysis of Camera-based Object Detection for Automated Vehicles," (shared co-first authorship), *Sensors*, vol. 20, no. 13, 2020, doi: 10.3390/s20133699.

Conferences; Scopus/Web of Science listed (peer-reviewed)

[16] T. Ponn, M. Breitfuß, X. Yu and F. Diermeyer, „Identification of Challenging Highway-Scenarios for the Safety Validation of Automated Vehicles Based on Real Driving Data," (accepted), in *Fifteenth International Conference on Ecological Vehicles and Renewable Energies (EVER2020)*, 2020.

[92] T. Ponn, D. Fratzke, C. Gnandt and M. Lienkamp, „Towards Certification of Autonomous Driving: Systematic Test Case Generation for a Comprehensive but Economically-Feasible Assessment of Lane Keeping Assist Algorithms," in *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems (VE-HITS 2019)*, 2019, pp. 333–342.

[174] T. Ponn, F. Müller and F. Diermeyer, „Systematic Analysis of the Sensor Coverage of Automated Vehicles Using Phenomenological Sensor Models," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 1000–1006.

[176] T. Ponn, T. Lanz and F. Diermeyer, „Automatic Generation of Road Geometries to Create Challenging Scenarios for Automated Vehicles Based on the Sensor Setup," (accepted), in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020.

Journals, Conferences, Periodicals, Reports, Conference Proceedings and Poster, etc.; not Scopus/Web of Science listed

[126] T. Ponn, C. Gnandt and F. Diermeyer, „An Optimization-Based Method to Identify Relevant Scenarios for Type Approval of Automated Vehicles," in *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 2019.

[127] T. Ponn, A. Schwab, F. Diermeyer, C. Gnandt and J. Záhorský, „A Method for the Selection of Challenging Driving Scenarios for Automated Vehicles Based on an Objective Characterization of the Driving Behavior," in *9. Tagung Automatisiertes Fahren*, 2019.

Non-thesis-relevant publications; Scopus/Web of Science listed (peer-reviewed)

O. Hofmann, T. Ponn, R. Buchmann and D. Rixen, „Nonlinear Predictive Control of Combustion and Emissions in Direct Injection Engines with Nozzle Aging," in *2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2018, isbn: 9781538618547. doi: 10.1109/aim.2018.8452324.

Supervised Student's Thesis

The following student theses were written within the framework of the dissertation under the supervision of the author in terms of content, technical and scientific support as well as under relevant guidance of the author. In the following, the bachelor, semester and master theses relevant and related to this dissertation are listed. Many thanks to the authors of these theses for their extensive support within the framework of this research project.

[202] A. Schwab, „Eine Methode zur Auswahl kritischer Fahrszenarien für automatisierte Fahrzeuge anhand einer objektiven Charakterisierung des Fahrverhaltens," Master's Thesis, Technical University of Munich, Munich, 2019.

[209] M. Breitfuß, „Extraction and classification of real traffic scenarios for validation of autonomous vehicles," Semesterarbeit, Technical University of Munich, Munich, 2020.

[211] X. Yu, „Method for quantitative evaluation of traffic complexity on the highway," Semesterarbeit, Technical University of Munich, Munich, 2019.

[212] X. Yu, „Improvement and validation of a highway traffic complexity metric for test scenarios of automated vehicles," Master's Thesis, Technical University of Munich, Munich, 2020.

[216] Y. Lin, „Creation of Complex Test Scenarios for Automated Vehicles by Means of Evolutionary Algorithm," Semesterarbeit, Technical University of Munich, Munich, 2020.

T. Kröger, „Analysis of influence factors on the prediction performance of AI-based object detection algorithms for autonomous driving," Master's Thesis, Technical University of Munich, Munich, 2020.

T. Lanz, „Automatische Generierung relevanter Szenarien für den Test automatisierter Fahrzeuge unter Berücksichtigung der Sensorabdeckung," Master's Thesis, Technical University of Munich, Munich, 2019.

C. Miethaner, „Bestimmung relevanter Szenarien für die Freigabe hochautomatisierter Fahrzeuge Masterarbeit," Master's Thesis, Technical University of Munich, Munich, 2018.

M. Moslemi, „Autonomous Cars & ADAS: Complex Scenario Generation, Simulation and Evaluation of Collision Avoidance Systems," Master's Thesis, Politecnico di Torino, Torino, 2019.

F. Müller, „Modellierung der Sensorabdeckung autonomer Fahrzeuge zur Berechnung optimaler Annäherungspfade," Master's Thesis, Technical University of Munich, Munich, 2018.

M. Riedel, „Entwicklung eines automatisierten Frameworks zur Durchführung und Auswertung von Simulationen automatisierter Fahrzeuge," Semesterarbeit, Technical University of Munich, Munich, 2020.

L. Figueira de Oliveira Schmidt, „Development of a test catalogue for the approval of automated vehicles based on the activities of the UNECE," Bachelor's Thesis, Technical University of Munich, Munich, 2019.

J. Steck, „Methodological Approach to Identify Automation Risks of Highly Automated Vehicles Using STPA,“ Master’s Thesis, Technical University of Munich, Munich, 2018.

Appendix

A Maneuver Definitions

Table A.1: Definition of 158 maneuvers. The surrounding TP is denoted as surrounding vehicle (SV).

Name	Description
FFA	FollowDrive Front Approach
FFI	FollowDrive Front Identical
FFB	FollowDrive Front Backdrop
FBA	FollowDrive Back Approach
FBI	FollowDrive Back Identical
FBB	FollowDrive Back Backdrop
FFD	FollowDrive Front Diverse
FBD	FollowDrive Back Diverse
SPB	SV Passing Behind
SPM	SV Passing Middle
SPF	SV Passing Front
SPBM	SV Passing Behind Middle
SPMF	SV Passing Middle Front
SPC	SV Passing Complete
EPB	Ego Passing Behind
EPM	Ego Passing Middle
EPF	Ego Passing Front
EPBM	Ego Passing Behind Middle
EPMF	Ego Passing Middle Front
EPC	Ego Passing Completely
PSB	ParallelDrive SV Behind
PSM	ParallelDrive SV Middle
PSF	ParallelDrive SV Front
SPCCIFFA	SV Passing Complete Cut-In FollowDrive Front Approach
SPCCIFFI	SV Passing Complete Cut-In FollowDrive Front Identical
SPCCIFFB	SV Passing Complete Cut-In FollowDrive Front Backdrop
SPMFCIFFA	SV Passing Middle/Front Cut-In FollowDrive Front Approach
SPMFCIFFI	SV Passing Middle/Front Cut-In FollowDrive Front Identical
SPMFCIFFB	SV Passing Middle/Front Cut-In FollowDrive Front Backdrop
SPFCIFFA	SV Passing Front Cut-In FollowDrive Front Approach
SPFCIFFI	SV Passing Front Cut-In FollowDrive Front Identical
SPFCIFFB	SV Passing Front Cut-In FollowDrive Front Backdrop
SCIFFA	SV Cut-In FollowDrive Front Approach
SCIFFI	SV Cut-In FollowDrive Front Identical
SCIFFB	SV Cut-In FollowDrive Front Backdrop
SPCCI	SV Passing Complete Cut-In
SPMFCI	SV Passing Middle/Front Cut-In
SPFCI	SV Passing Front Cut-In
EPCCIFBA	Ego Passing Complete Cut-In FollowDrive Back Approach
EPCCIFBI	Ego Passing Complete Cut-In FollowDrive Back Identical
EPCCIFBB	Ego Passing Complete Cut-In FollowDrive Back Backdrop
EPMFCIFBA	Ego Passing Middle/Front Cut-In FollowDrive Back Approach
EPMFCIFBI	Ego Passing Middle/Front Cut-In FollowDrive Back Identical
EPMFCIFBB	Ego Passing Middle/Front Cut-In FollowDrive Back Backdrop

Continued on next page

Table A.1 – continued from previous page

Name	Description
EPFCIFBA	Ego Passing Front Cut-In FollowDrive Back Approach
EPFCIFBI	Ego Passing Front Cut-In FollowDrive Back Identical
EPFCIFBB	Ego Passing Front Cut-In FollowDrive Back Backdrop
ECIFBA	Ego Cut-In FollowDrive Back Approach
ECIFBI	Ego Cut-In FollowDrive Back Identical
ECIFBB	Ego Cut-In FollowDrive Back Backdrop
EPCCI	Ego Passing Complete Cut-In
EPMFCI	Ego Passing Middle/Front Cut-In
EPFCI	Ego Passing Front Cut-In
SFBCOPC	SV FollowDrive Behind Cut-Out Passing Complete
SFBCOPBM	SV FollowDrive Behind Cut-Out Passing Behind/Middle
SFBCOPB	SV FollowDrive Behind Cut-Out Passing Behind
SFBCOPSB	SV FollowDrive Behind Cut-Out ParallelDrive SV Behind
SFBCOEPF	SV FollowDrive Behind Cut-Out Ego Passing Front
SFBCO	SV FollowDrive Behind Cut-Out
SCOPC	SV Cut-Out Passing Complete
SCOPBM	SV Cut-Out Passing Behind/Middle
SCOPB	SV Cut-Out Passing Behind
SCOPSB	SV Cut-Out ParallelDrive SV Behind
SCOEPF	SV Cut-Out Ego Passing Front
EFBCOPC	Ego FollowDrive Behind Cut-Out Passing Complete
EFBCOPBM	Ego FollowDrive Behind Cut-Out Passing Behind/Middle
EFBCOPB	Ego FollowDrive Behind Cut-Out Passing Behind
EFBCOPSF	Ego FollowDrive Behind Cut-Out ParallelDrive SV Front
EFBCOSPF	Ego FollowDrive Behind Cut-Out SV Passing Front
EFBCO	Ego FollowDrive Behind Cut-Out
ECOPC	Ego Cut-Out Ego Passing Complete
ECOPBM	Ego Cut-Out Passing Behind/Middle
ECOPB	Ego Cut-Out Passing Behind
ECOPSF	Ego Cut-Out ParallelDrive SV Front
ECOSPF	Ego Cut-Out SV Passing Front
SLCB	SV Lane Change Behind
SLCF	SV Lane Change Front
ELCB	Ego Lane Change Behind
ELCF	Ego Lane Change Front
SFBOFFA	SV FollowDrive Behind Overtake FollowDrive Front Approach
SFBOFFI	SV FollowDrive Behind Overtake FollowDrive Front Identical
SFBOFFB	SV FollowDrive Behind Overtake FollowDrive Front Backdrop
SOFFA	SV Overtake FollowDrive Front Approach
SOFFI	SV Overtake FollowDrive Front Identical
SOFFB	SV Overtake FollowDrive Front Backdrop
SFBO	SV FollowDrive Behind Overtake
EFBOFBA	Ego FollowDrive Behind Overtake (SV) FollowDrive Back Approach
EFBOFBI	Ego FollowDrive Behind Overtake (SV) FollowDrive Back Identical
EFBOFBB	Ego FollowDrive Behind Overtake (SV) FollowDrive Back Back-drop
EOFBA	Ego Overtake (SV) FollowDrive Back Approach
EOFBI	Ego Overtake (SV) FollowDrive Back Identical
EOFBB	Ego Overtake (SV) FollowDrive Back Backdrop
EFBO	Ego FollowDrive Behind Overtake
EPBSCIFFA	Ego Passing Behind and SV Cut-In with FollowDrive Front Approach
EPBSCIFFI	Ego Passing Behind and SV Cut-In with FollowDrive Front Identical
EPBSCIFFB	Ego Passing Behind and SV Cut-In with FollowDrive Front Backdrop
EPBMSCIFFA	Ego Passing Behind/Middle and SV Cut-In with FollowDrive Front Approach
EPBMSCIFFI	Ego Passing Behind/Middle and SV Cut-In with FollowDrive Front Identical
EPBMSCIFFB	Ego Passing Middle and SV Cut-In with FollowDrive Front Backdrop
EPMSCIFFA	Ego Passing Middle and SV Cut-In with FollowDrive Front Approach

Continued on next page

Table A.1 – continued from previous page

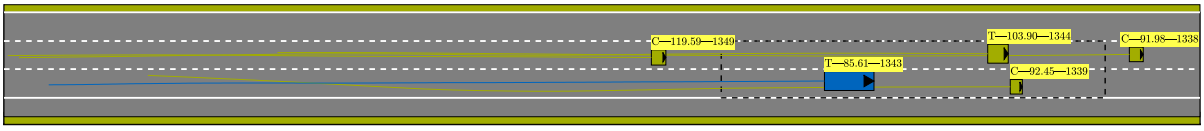
Name	Description
EPMSCIFFI	Ego Passing Middle and SV Cut-In with FollowDrive Front Identical
EPMSCIFFB	Ego Passing Middle and SV Cut-In with FollowDrive Front Backdrop
EPBSCI	Ego Passing Behind and SV Cut-In
EPBMSCI	Ego Passing Behind/Middle and SV Cut-In
EPMSCI	Ego Passing Middle and SV Cut-In
SPBECIFBA	SV Passing Behind and Ego Cut-In with FollowDrive Back Approach
SPBECIFBI	SV Passing Behind and Ego Cut-In with FollowDrive Back Identical
SPBECIFBB	SV Passing Behind and Ego Cut-In with FollowDrive Back Backdrop
SPBMECIFBA	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Approach
SPBMECIFBI	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Identical
SPBMECIFBB	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Backdrop
EPBSCI	Ego Passing Behind and SV Cut-In
EPBMSCI	Ego Passing Behind/Middle and SV Cut-In
EPMSCI	Ego Passing Middle and SV Cut-In
SPBECIFBA	SV Passing Behind and Ego Cut-In with FollowDrive Back Approach
SPBECIFBI	SV Passing Behind and Ego Cut-In with FollowDrive Back Identical
SPBECIFBB	SV Passing Behind and Ego Cut-In with FollowDrive Back Backdrop
SPBMECIFBA	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Approach
SPBMECIFBI	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Identical
SPBMECIFBB	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Backdrop
SPMECIFBA	SV Passing Middle and Ego Cut-In with FollowDrive Back Approach
SPMECIFBI	SV Passing Middle and Ego Cut-In with FollowDrive Back Identical
SPMECIFBB	SV Passing Middle and Ego Cut-In with FollowDrive Back Backdrop
SPBECI	SV Passing Behind and Ego Cut-In
SPBMECI	SV Passing Behind/Middle and Ego Cut-In
SPMECI	SV Passing Middle and Ego Cut-In
SFBECOSPC	SV FollowDrive Behind with Ego Cut-Out and SV Passing Com-plete
SFBECOSPBM	SV FollowDrive Behind with Ego Cut-Out and SV Passing Behind/Middle
SFBECOSPB	SV FollowDrive Behind with Ego Cut-Out and SV Passing Behind
SFBECOPSB	SV FollowDrive Behind with Ego Cut-Out and ParallelDrive SV Behind
SFBECOPF	SV FollowDrive Behind with Ego Cut-Out and (Ego) Passing Front
SFBECO	SV FollowDrive Behind with Ego Cut-Out
ECOSPC	Ego Cut-Out and SV Passing Complete
ECOSPBM	Ego Cut-Out and SV Passing Behind/Middle
ECOSPB	Ego Cut-Out and SV Passing Behind
ECOPSB	Ego Cut-Out and ParallelDrive SV Behind
ECOPF	Ego Cut-Out and (Ego) Passing Front
EFBSCOEPC	Ego FollowDrive Behind with SV Cut-Out and Ego Passing Complete
EFBSCOEPBM	Ego FollowDrive Behind with SV Cut-Out and Ego Passing Behind/Middle
EFBSCOEPB	Ego FollowDrive Behind with SV Cut-Out and Ego Passing Behind
EFBSCOPSF	Ego FollowDrive Behind with SV Cut-Out and ParallelDrive SV Front
EFBSCOPF	Ego FollowDrive Behind with SV Cut-Out and (SV) Passing Front
EFBSCO	Ego FollowDrive Behind with SV Cut-Out
SCOEPC	SV Cut-Out and Ego Passing Complete
SCOEPBM	SV Cut-Out and Ego Passing Behind/Middle
SCOEPB	SV Cut-Out and Ego Passing Behind
SCOPSF	SV Cut-Out and ParallelDrive SV Front
SCOPF	SV Cut-Out and (SV) Passing Front
EPBCIFFA	Ego Passing Behind Cut-In FollowDrive Front Approach
EPBCIFFI	Ego Passing Behind Cut-In FollowDrive Front Identical
EPBCIFFB	Ego Passing Behind Cut-In FollowDrive Front Backdrop
PSFECIFFA	ParallelDrive SV Front Ego Cut-In FollowDrive Front Approach
PSFECIFFI	ParallelDrive SV Front Ego Cut-In FollowDrive Front Identical
PSFECIFFB	ParallelDrive SV Front Ego Cut-In FollowDrive Front Backdrop
SPFECIFFA	SV Passing Front Ego Cut-In FollowDrive Front Approach
SPFECIFFI	SV Passing Front Ego Cut-In FollowDrive Front Identical

Continued on next page

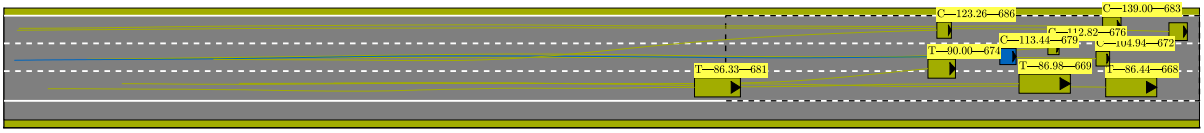
Table A.1 – continued from previous page

Name	Description
SPFECIFFB	SV Passing Front Ego Cut-In FollowDrive Front Backdrop
SPBCIFBA	SV Passing Behind Cut-In FollowDrive Back Approach
SPBCIFBI	SV Passing Behind Cut-In FollowDrive Back Identical
SPBCIFBB	SV Passing Behind Cut-In FollowDrive Back Backdrop
PSBCIFBA	ParallelDrive SV Behind Cut-In FollowDrive Back Approach
PSBCIFBI	ParallelDrive SV Behind Cut-In FollowDrive Back Identical
PSBCIFBB	ParallelDrive SV Behind Cut-In FollowDrive Back Backdrop
EPFSCIFBA	Ego Passing Front SV Cut-In FollowDrive Back Approach
EPFSCIFBI	Ego Passing Front SV Cut-In FollowDrive Back Identical
EPFSCIFBB	Ego Passing Front SV Cut-In FollowDrive Back Backdrop

B Exemplary Complex Scenarios



(a) Scenario from highD-track 34 with low complexity $C_{\text{scenario}} = 0.32$.



(b) Scenario from highD-track 10 with high complexity $C_{\text{scenario}} = 0.66$.

Figure B.1: Comparison of scenarios with low and high complexity. The ego-vehicle including the driven path is shown in blue and all surrounding TPs are shown in green. The ROI of the ego-vehicle at the depicted time step is marked by a dashed black rectangle. Each vehicle has a three-part identification consisting of the class (car or truck), the current speed in $^{\text{km}}/_{\text{h}}$ and the highD vehicle ID. The figures are adapted from [16]

C Scenario Optimization

C.1 IDM Parameters

Table C.1: List of the parameters used for IDM model.

Parameter	Symbol	value
Desired velocity	v_0	scenario-specific
Time-gap headway	T_g	1.2 s based on [219]
Maximum acceleration	$a_{x,max}$	3 m/s^2
Comfortable deceleration	$b_{x,com}$	4 m/s^2
Acceleration exponent	δ	4
Length of ego	l_{ego}	scenario-specific
Linear jam distance	s_0	2 m
Non-linear jam distance	s_1	3 m

C.2 GA Parameters

Table C.2: Parameters of the GA.

Parameter	Value
Population size	500
Maximal number of generations	500
Crossover fraction	80 %
Elite ratio	5 %
Selection type	Stochastic uniform
Crossover type	Two point
Mutation function	Adaptive feasible
Initial population matrix	Original highD

C.3 Optimization Results

Table C.3: Overview of the optimization results. Thereby, highD is the complexity of the original highD scenario, before means the complexity of the original highD date where the ego-vehicle is replaced by the IDM and MOBIL model, after means the complexity of optimized scenario, track nr. is the track number of the highD data set, ego ID is the number of the ego-vehicle of the highD data set, time is the optimization time in seconds for the 500 iterations on a 32 core E5-2670 Intel Xeon with $2.6\,\mathrm{GHz}$ and n_{var} is the number of optimization variables.

highD	before	after	Track Nr.	Ego ID	Time in s	n_{var}
0.424	0.450	0.915	56	1278	1348	124
0.583	0.489	1.076	3	178	2026	220
0.595	0.486	1.034	53	1108	3276	390
0.622	0.539	0.874	52	529	4164	512
0.605	0.512	1.047	59	140	5742	740
0.627	0.627	0.887	40	2194	5068	828
0.729	0.753	1.242	47	1135	5493	1064
0.719	0.661	0.792	36	2326	8707	1328
0.637	0.617	0.833	31	72	11165	1692
0.615	0.535	0.796	40	1291	8213	1420
0.675	0.572	0.730	10	676	14949	1804
0.660	0.688	0.863	10	679	12718	1920
0.677	0.615	0.763	40	1288	19020	2262
0.613	0.597	0.970	30	1709	24613	2744
0.672	0.570	0.928	41	254	25406	2730
0.623	0.674	1.027	41	253	29559	2912
0.613	0.575	0.826	41	544	23976	2822
0.568	0.655	0.854	44	2349	35438	3492

D Complexity Metric Validation

D.1 Complexity Factors Results

Table D.1: Rating of the 13 influencing factors by 25 experts via a scale from 1 (not important) to 5 (very important).

Participant	Influence factor 1	2	3	4	5	6	7	8	9	10	11	12	13	sum
1	1	5	3	4	3	5	4	4	2	2	5	2	2	42
2	3	4	3	5	5	5	5	4	2	5	5	3	5	54
3	2	4	4	4	3	4	3	4	2	2	3	2	3	40
4	2	5	5	4	1	5	5	5	5	5	5	1	1	49
5	3	3	4	4	4	3	3	3	3	3	2	3	3	41
6	3	2	4	4	3	2	2	2	4	4	4	3	3	40
7	2	4	4	4	5	4	3	4	5	4	3	4	4	50
8	3	4	2	3	4	5	2	4	2	2	4	3	4	42
9	2	4	3	5	5	5	5	4	2	5	5	4	5	54
10	2	4	4	5	5	5	5	4	2	5	5	4	5	55
11	5	5	5	5	5	5	5	5	5	5	5	5	5	65
12	4	2	3	4	5	4	2	2	2	1	3	1	5	38
13	3	4	4	5	1	5	4	2	5	2	3	1	5	44
14	4	3	1	1	5	5	1	3	3	3	5	4	4	42
15	4	4	4	5	5	5	3	3	5	5	5	3	3	54
16	4	4	5	5	5	4	3	4	5	3	4	4	4	54
17	2	3	3	4	3	4	3	2	4	4	3	2	2	39
18	5	5	5	5	5	4	5	5	5	5	5	4	4	62
19	4	5	5	5	4	5	5	5	4	4	4	5	5	60
20	4	5	5	4	4	4	4	3	4	4	3	4	3	51
21	5	5	5	5	5	4	4	4	5	5	5	4	4	60
22	2	4	3	2	4	5	5	5	3	3	5	5	5	51
23	3	2	4	4	3	3	2	4	1	2	2	2	4	36
24	3	4	4	5	5	2	3	3	3	3	4	5	5	49
25	2	4	4	4	3	5	3	3	3	3	4	2	2	42
Variance	1.24	0.91	1.06	1.00	1.50	0.88	1.51	0.99	1.76	1.59	1.04	1.67	1.42	68.51

Table D.2: Mean rank of each influence factor. The number of elements is the number of participating experts who rated the factors.

Factor	Number of elements	Rank sum $R_{\text{factor},i}$	Mean rank $\bar{R}_{\text{factor},i}$
1	25	2766.50	110.66
2	25	4342	173.68
3	25	4207	168.28
4	25	5016	200.64
5	25	4683.50	187.34
6	25	5165.50	206.62
7	25	3736.50	149.46
8	25	3760.50	150.42
9	25	3570	142.80
10	25	3762	150.48
11	25	4639.50	185.58
12	25	3109	124.36
13	25	4217	168.68

Table D.3: Right side of the inequality from Equation (4.36) of the Dunn's test for all influence factor comparisons. Factors with significant difference are marked in gray. The reference value for each comparison is $Q_{0.05} = 3.41$.

		Mean rank difference to factor											
		1	**12**	**9**	**7**	**8**	**10**	**3**	**13**	**2**	**11**	**5**	**4**
Factor	**6**	3.75	3.21	2.49	2.23	2.19	2.19	1.50	1.48	1.29	0.82	0.75	0.23
	4	3.51	2.98	2.26	2.00	1.96	1.96	1.26	1.25	1.05	0.59	0.52	
	5	2.99	2.46	1.74	1.48	1.44	1.44	0.74	0.73	0.53	0.07		
	11	2.93	2.39	1.67	1.41	1.37	1.37	0.68	0.66	0.46			
	2	2.46	1.93	1.21	0.95	0.91	0.91	0.21	0.20				
	13	2.27	1.73	1.01	0.75	0.71	0.71	0.02					
	3	2.25	1.71	0.99	0.73	0.70	0.70						
	10	1.55	1.02	0.30	0.04	0.00							
	8	1.55	1.02	0.30	0.04								
	7	1.51	0.98	0.26									
	9	1.26	0.72										
	12	0.53											
	1												

D.2 Complexity Factors Additional Factors Comments

- The distinction between actors can be an additional factor: Small angles between vehicles, Small distances, similar look and behavior might challenge the algorithm. So the AV doesn't know whether there are two objects or one object.
- Factors like size and shape of surrounding objects will have an Impact as well. Influence: object recognition will be challenged
- Factors for the shape and size of surrounding vehicles will have an impact: Influence: Object recognition will be challenged
- Speed difference of the vehicles (might be similar to factor 1)
- constellation of position and speed of the vehicles (difficult to evaluate)
- road obstacles (e.g. entry lane)
- I think the parameterization is already pretty detailed.

D.3 Links to Scenario Videos Used for Online Survey

1. https://www.youtube.com/watch?v=9ojVyRYyiPk
2. https://www.youtube.com/watch?v=w4S5mX6DXis
3. https://www.youtube.com/watch?v=xnmLJZqJTag
4. https://www.youtube.com/watch?v=f-d44ulvFMo
5. https://www.youtube.com/watch?v=K2engJ8GETs
6. https://www.youtube.com/watch?v=3uV1l8jdsXY
7. https://www.youtube.com/watch?v=TQXLO8SpihE
8. https://www.youtube.com/watch?v=ibNUY904Cdk
9. https://www.youtube.com/watch?v=zm9Uj428oeU
10. https://www.youtube.com/watch?v=xGOdYsfN5hA
11. https://www.youtube.com/watch?v=ZnhlrH6uVHk
12. https://www.youtube.com/watch?v=WkWs-1os60s
13. https://www.youtube.com/watch?v=2V8Wb8wAzp0
14. https://www.youtube.com/watch?v=XfLf5S1gr_Q
15. https://www.youtube.com/watch?v=8ifA54kV7Bo
16. https://www.youtube.com/watch?v=Y1GkeFIcACk
17. https://www.youtube.com/watch?v=k26SAxLzXAM
18. https://www.youtube.com/watch?v=GDx7PADYuhQ
19. https://www.youtube.com/watch?v=3zDyRoRxtMk
20. https://www.youtube.com/watch?v=UjIoIs7uH8M