Scenario-Based Safety Assessment of Automated Driving Systems

TNO StreetWise

TNO Public › TNO 2024 R10983
29 July 2024

Mobility & Built Environment
www.tno.nl
+31 88 866 57 34
secretaryhelmond@tno.nl

29 July 2024

TNO StreetWise

Scenario-Based Safety Assessment of Automated Driving Systems

Author(s)	Erwin de Gelder, Olaf Op den Camp, Jeroen Broos, Jan-Pieter Paardekooper, Sjef van Montfort, Sytze Kalisvaart, Henk Goossens
Classification report	TNO Public
Number of pages	39 (excl. front and back cover)
Report number	TNO 2024 R10983

Contents

1 Introduction

Automated Driving Systems (ADSs) hold the potential to enhance traffic safety by eliminating human errors, providing more comfortable rides, and reducing traffic congestion [1]. Moreover, the introduction of higher levels of ADSs bears the important promise that it increases the accessibility of the mobility system for all individuals.

There are two significant developments in the evolution of ADSs. Firstly, while lower-level ADSs like Advanced Driver Assistance Systems (ADASs) are already prevalent [2], higher-level ADSs are currently being introduced or will soon be deployed on public roads. Secondly, the functionality of lower-level ADSs is expanding to provide sustained motion control of the vehicle, leading to so-called Driver Control Assistance Systems (DCASs) [3].

For lower-level ADSs up to SAE level 2, the driver is still responsible for intervening [4]. Consequently, drivers are accountable for vehicles equipped with such ADSs and are expected to intervene if the system encounters a situation it cannot handle safely. However, it is essential to ensure that such ADSs do not adversely compromise road safety [3]. Especially with the larger operational domain in which a DCAS is expected to be deployed, it is challenging to assess whether the DCAS does not adversely impact road safety.

For higher-level ADSs, i.e., from SAE level 3 onwards, the driver is neither required nor expected to intervene in case the ADS fails within its operating domain. As a result, for a higher-level ADS, it can no longer be assumed that the human takes over control in case the ADS fails to respond appropriately. In such cases, the responsibility for appropriate behaviour falls entirely on the ADS, even in rare scenarios. Therefore the safety assessment of higher-level ADSs poses significant challenges, as they must demonstrate the capability to handle a wide range of scenarios they may encounter during operation.

The expansion of the operational domain of DCASs and the advancement to higher levels of ADSs necessitate a paradigm shift in safety assessment methodologies. Unlike lower-level ADSs, identifying challenging scenarios for higher-level ADSs is not straightforward. Safety risks are determined by the frequency of occurrence and the severity of a scenario if mishandled. For safety assessment, we need a practical method that provides a complete overview of scenarios with their realistic variations that an ADS can encounter in the real world. The method should allow for statistical analyses providing information on the occurrence of a scenario and the completeness of the scenario set. TNO StreetWise offers a practical methodology that fulfils these requirements.

TNO's ambition is to facilitate the safe and responsible introduction of automated driving technology onto the public road, with the objective of making traffic safer, more efficient, more comfortable and more accessible. We achieve this goal by providing innovative solutions and technologies to enhance vehicle safety and to increase vehicle automation. Simultaneously, we develop methods to accelerate the implementation of such solutions into vehicles on the road. This report delves into the development and application of TNO StreetWise as a methodology for collecting real-world scenarios in a scenario database. It explains how the StreetWise scenario database is utilised to generate test scenarios for safety assessment and addresses the challenges associated with safety assessment of ADSs.

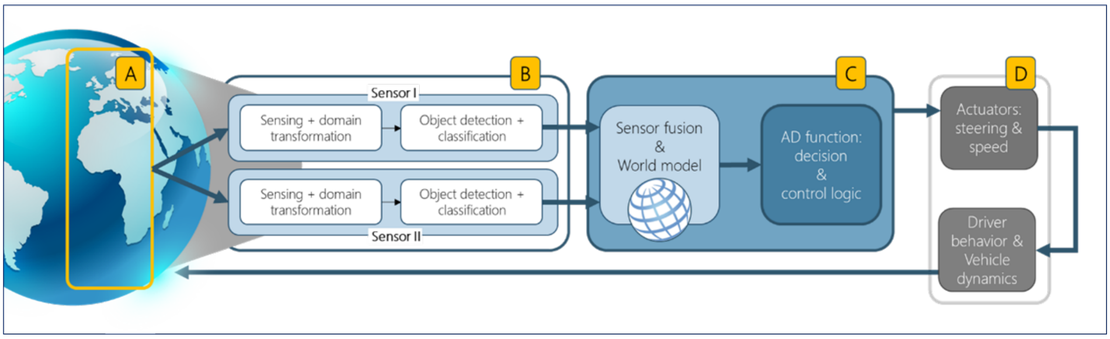

Figure 1.1: Schematic overview of ADS following a sense-think-act sequence.

1.1 StreetWise Vision

TNO plays a crucial role in supporting both the automotive industry and regulatory authorities in developing a scenario-based safety assessment methodology. This methodology is not limited to evaluating ADSs but extends to any automation system affecting the operational safety of motorised vehicles. The automotive industry requires relevant and realistic tests for vehicle system development and safety argumentation, while regulatory authorities bear the responsibility of ensuring that only safe vehicles are permitted on the road. Therefore, there is a mutual benefit in introducing a harmonised and standardised safety assessment method that is accepted globally. TNO StreetWise enables the wide application of scenario-based safety validation methods for ADSs.

As shown in Figure 1.1, an ADS typically consists of a sensor system (B) to perceive the environment (A), a think and plan system (C) to make an interpretation of the environment and plan the next manoeuvre, and a set of actuators (D) to actually follow the determined manoeuvre and control vehicle dynamics.

While type approval or certification tests typically evaluate the complete vehicle with the entire ADS, during the current phase of automated driving evolution, where applications are in testing rather than full product deployment, understanding the correct functioning of subsystems and components becomes crucial. Therefore, tests are often conducted to assess the performance of the subsystems. Perception tests evaluate the combination of sensor-, sensor fusion-, and world-modelling system, to assess the correct perception, identification, tracking and tracing of objects in the real world. Also, vehicle dynamics tests are performed to assess the capabilities and limitations of the actuators and the effect on the vehicle dynamics behaviour under different (road) conditions.

With StreetWise, TNO provides a methodology for the safety assessment of the complete vehicle, focusing on the control and decision logic of the ADS.

1.2 Scenarios and test scenarios

Scenarios lie at the heart of StreetWise, with the methodology distinguishing between scenarios and test scenarios. Given the significance of these concepts within StreetWise, it is essential to provide further clarification on each.

1.2.1 Scenario

In various scenario databases, scenarios are collected that are used to describe what ADS or Connected, Cooperative, and Automated Mobility (CCAM) system[1] may possibly encounter during its lifetime (or deployment time). The scenarios in these databases provide a structured view on what has happened on the road or what may realistically occur. Scenario databases are set up for various purposes, not only for safety assessment [5]. An accident database typically offers insights into road accidents, often including details such as the scenarios leading to the accidents and the associated conditions. Usually, accident databases provide information on injury severity resulting from the accidents. Other scenario databases identify, characterise, and record all scenarios from vehicles driving on the road — usually a limited number of vehicles driving for many kilometres (comparable to Field Operational Tests, or Naturalistic Driving Studies). They primarily rely on empirical data to compile and analyse scenarios. Alternatively, scenarios can also be formulated based on expert knowledge and experiences. These knowledge-based scenarios can be added to a scenario database.

> A scenario describes various elements of a road situation, including the intentions of the ego vehicle, the behaviours of other road users, road layouts, and environmental conditions such as weather and lighting. A drive on the road is considered a continuous sequence of scenarios that may overlap.

More formally [6], a scenario is a quantitative description of the relevant characteristics and activities and/or goals of the ego vehicle(s), the static environment, the dynamic environment, and all the events that are relevant to the ego vehicle(s) within the time interval between the first and the last relevant event [6]. Events mark specific moments in time, signifying mode transitions or system thresholds, influenced by both internal and external factors.

1.2.2 Test scenario

Test scenarios serve as the foundation for scenario-based safety assessments and are utilised in both virtual and track testing procedures. Following the IEEE definition of a test, which is clearly explained in [7], a test is an evaluation of:

- A statement on the system-under-test (*test criteria*; what is evaluated using the test?);
- under a set of specified conditions (*test scenario*; how are the test criteria evaluated?);
- using quantitative measures (*metrics*; how is the outcome of the test expressed quantitatively?);
- and possibly a reference of what would be the acceptable outcome (*reference*; when is the outcome acceptable?). Note: the reference may be formulated for each individual test, but also for a set of tests, e.g., in case of risk quantification.

Indeed, a test scenario may not perfectly mirror a scenario in a scenario database. Test scenarios may be derived through sampling or interpolation methods, resulting in scenarios that have not been precisely observed in the real world. A test scenario might not contain all information that is available from the scenario database, and a test scenario might be complemented with models or other information to be able to correctly perform a test. Additionally, a test scenario might be a simplification of a (real-world) scenario. For example, a test scenario of Euro NCAP

[1] The term Connected, Cooperative, and Automated Mobility (CCAM) systems is often used in Europe for higher levels of ADSs, often equipped with V2X communication functionality.

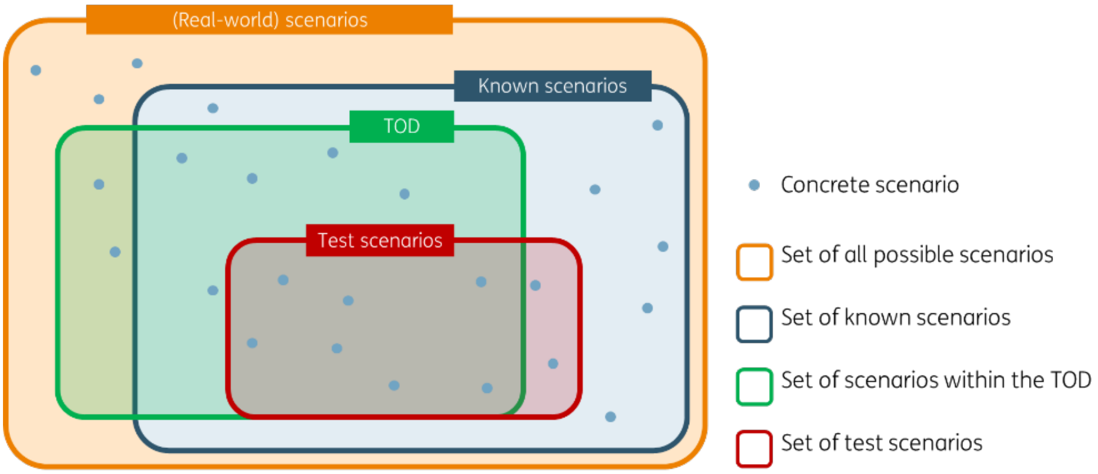

Figure 1.2: A schematic view on the relation between scenarios, the actual operating conditions of an ADS (coined as the Target Operational Domain (TOD), see Section 1.5), and test scenarios. Adapted from [8].

to test emergency functions such as Autonomous Emergency Braking (AEB) is usually defined with one vehicle and one collision partner. The real-world scenario on which such test scenario is based, may easily have contained more than one other road participant, and a different static environment than the environment represented on the test track. Assumptions are made in the translation from (real-world) scenarios to test scenarios. It is essential to evaluate the validity of these assumptions during an evaluation of the complete assessment process (e.g., by an audit).

Figure 1.2 visualises the difference between a scenario and a test scenario. Here, the set of known scenarios is a subset of all (real-world) scenarios. The actual operating conditions in which an ADS is operating is represented by the green box and referred to as the Target Operational Domain (TOD) (more on that in Section 1.5). The set of test scenarios, represented by the red box in Figure 1.2, is a subset of all known scenarios. The objective is to ensure that the set of test scenarios covers as much of the TOD as possible. However, this task is challenging due to the fact that the TOD typically contains scenarios that are simply unknown beforehand.

1.3 Data-driven, scenario-based assessment

In a scenario-based assessment of an ADS or CCAM system, the system is tested in many different individual (or concrete) driving scenarios. In this context, a scenario represents a description of the situation of a vehicle equipped with the ADS and its surroundings over a defined period (a more rigorous definition is presented in [6]). StreetWise employs a data-driven approach to gather and construct scenarios for assessment, utilising real-world driving data to collect the driving scenarios and subsequently generate test scenarios.

The objective of the data-driven, scenario-based assessment is to quantify the risk of a vehicle equipped with an ADS once the vehicle would be deployed on public roads. Figure 1.3 presents a schematic overview of scenario-based assessment methodology using real-world data, as followed by StreetWise and described in more detail in [9]. In the process, nine steps are distinguished:

1. Data collection, e.g., using a vehicle equipped with sensors [10].

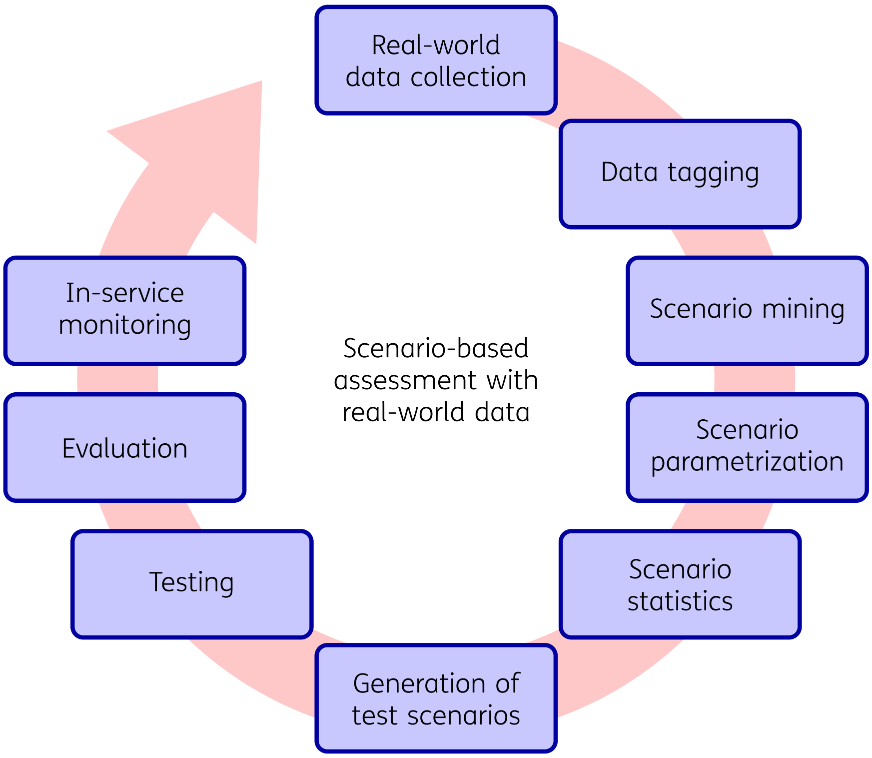

Figure 1.3: Schematic overview of the process of data-driven, scenario-based assessment. Figure based on [9].

2. Data tagging based on, e.g., a particular lane change of a vehicle or a braking action of a vehicle.
3. Based on the detected activities in the previous step and other information, e.g., the road layout, scenarios are mined, i.e., identified and extracted, from the data.
4. Parameters are defined for characterising the mined scenarios, with values of the parameters computed for each scenario.
5. Based on the parametrised scenarios, statistics on the scenarios are computed, including the likelihood of encountering scenarios with certain parameter values.
6. Test scenarios are generated using the scenario statistics from the previous step.
7. Tests are performed to measure the response of the vehicle equipped with the ADS in the generated test scenarios, often using virtual simulations.
8. All test results are aggregated, typically using key performance indicators, to evaluate the ADS performance.
9. Upon approval, the ADS may be deployed on public roads. In-service monitoring may be required by regulatory authorities to continuously assess safety during deployment.

The data collected during in-service monitoring can be used to improve the generation of tests as it is possible that some scenarios have been overlooked during the initial assessment process. Additionally, as traffic conditions evolve due to factors like the introduction of new mobility systems, ongoing data collection is necessary to update the assessment process as indicated in Figure 1.3. The acquired data during in-service monitoring may be used for this purpose. The arrow in Figure 1.3 indicates the continuous repetition of the scenario-based assessment approach for the continuous improvement of the scenario-based assessment, accommodating both tests that have been overlooked and changes to the traffic system.

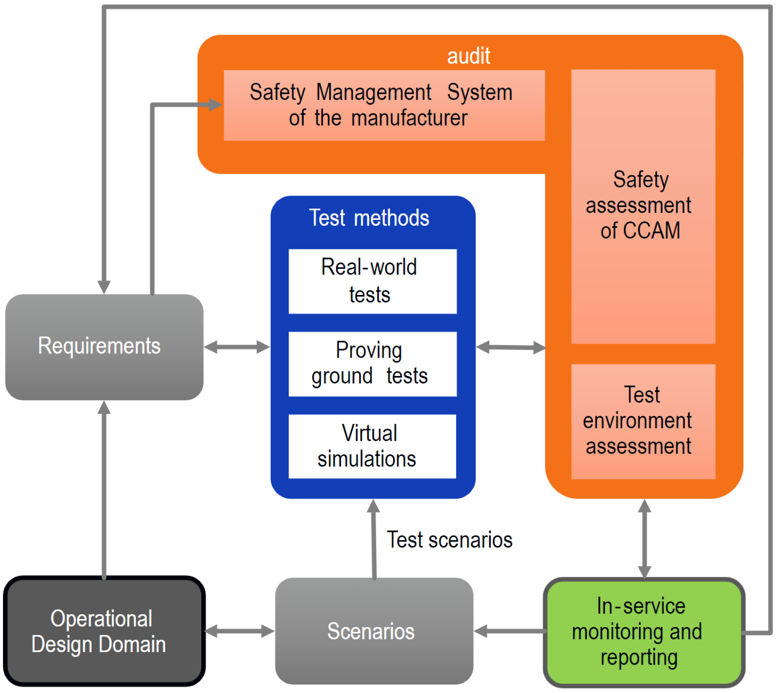

Figure 1.4: The multi-pillar approach. Adapted from [15].

1.4 Scenarios for compliance with regulations

To establish a legal framework for the deployment or testing of ADSs on public roads, regulations are being implemented by both the United Nations Economic Commission for Europe (UNECE), such as those for Automated Lane Keeping Systems (ALKSs) [11, 12], and the European Commission, specifically for ADSs in four designated use cases [13]. Following the formulation of the regulations for ALKSs, an ADS should be free of reasonably foreseeable and preventable safety risks. The challenge is how to quantify a safety risk, and what is considered to be "reasonably foreseeable" and "reasonably preventable", considering all plausible combinations of situations and conditions that an ADS might encounter on the road.

A solution lies in identifying and characterising real-world scenarios. StreetWise accomplishes this by collecting real-world scenarios from driving data, enabling not only the derivation of scenario categories exposure from the resulting scenario database but also the determination of Probability Density Functions (PDFs) for the characteristic parameters used to describe specific scenarios within a category.

1.4.1 The UNECE multi-pillar approach

The UNECE WP.29 Working Party on Automated/Autonomous and Connected Vehicles (GRVA) has developed the New Assessment/Test Methods (NATM) Master Document [14], which proposes a multi-pillar Safety Assessment Framework (SAF) for the type-approval process of CCAM systems (see Figure 1.4).

Despite the potential complexity of CCAM systems and the intricacies of their assessment procedures, the assessment results should be unambiguous, easily understood by experts in the field, and explainable to authorities and the public. This is why scenarios, serving as a structured means of describing the diverse array of situations and conditions that a CCAM system may encounter on the road, form the most important source of information for generating test scenarios across different testing methodologies: virtual testing using computer models and simulation tools, track testing under realistic and reproducible conditions, and real-world tests on the road, e.g., by means of field operational tests.

In addition to the three testing pillars, an In-Service Monitoring and Reporting (ISMR) pillar is distinguished. ISMR is utilised to examine the behaviour of the CCAM system during deployment or large-scale testing in real-world conditions. The monitoring process involves analysing vehicle-collected data to identify issues such as erratic behaviour, degrading performance aspects of the CCAM system, or scenarios that were previously unidentified. A reporting mechanism is incorporated to facilitate the sharing of scenarios and lessons learned among various stakeholders, including regulators and manufacturers.

A fifth and final pillar is established in the form of an audit. This pillar represents an analytical phase to verify the Safety Management System (SMS) of the CCAM manufacturer, the safety concept implemented in the CCAM design (verified in the safety assessment), and the tools and test environment used in the assessment. The definition of the audit process fall beyond the scope of this document.

1.4.2 Operationalising the multi-pillar approach

TNO has developed an SAF, depicted in Figure 1.5, to implement the multi-pillar approach as proposed by UNECE. This framework adopts a scenario-based approach, where tests (virtual testing, track testing and real-world testing) are designed around scenarios that take into account the Operational Design Domain (ODD) of the CCAM system or ADS. These tests assess the safety of the system prior to its deployment on public roads. The SAF includes, but is not limited to, processes to:

- *Generate the relevant test scenarios*: Based on a description of the ODD and the system requirements (incl. requirements resulting from a description of the dynamic driving task), the set of relevant test scenarios is determined. These test scenarios are rooted in real-world situations that fall within the ODD. Various scenario databases may be utilised for this purpose. In this document, we focus on TNO StreetWise, a scenario database derived from real-world driving data that identifies and characterises scenarios. Additionally, within the Horizon Europe project SUNRISE[2], a federation layer is under development to facilitate the querying of scenarios from different sources.
- *Allocate test scenarios to the different test methods*: Once the set of test scenarios has been selected or generated from a sampling procedure, test descriptions are provided by enhancing the test scenarios with appropriate test metrics and test objectives. Additionally, it is necessary to determine which test scenario is allocated to which test method. Alongside established the well-accepted testing methods in the real world (such as field operational tests) and on proving grounds (similar to Euro NCAP tests), a virtual simulation pillar is introduced. Virtual simulation is becoming increasingly important for enabling safety assessment based on a large number of test scenarios.
- *Execute the test scenarios*: All tests must be executed to provide test results for input into the safety assessment process. If a test cannot be performed using the allocated test method, it may need to be reassigned to another suitable method.

[2] `https://ccam-sunrise-project.eu/`

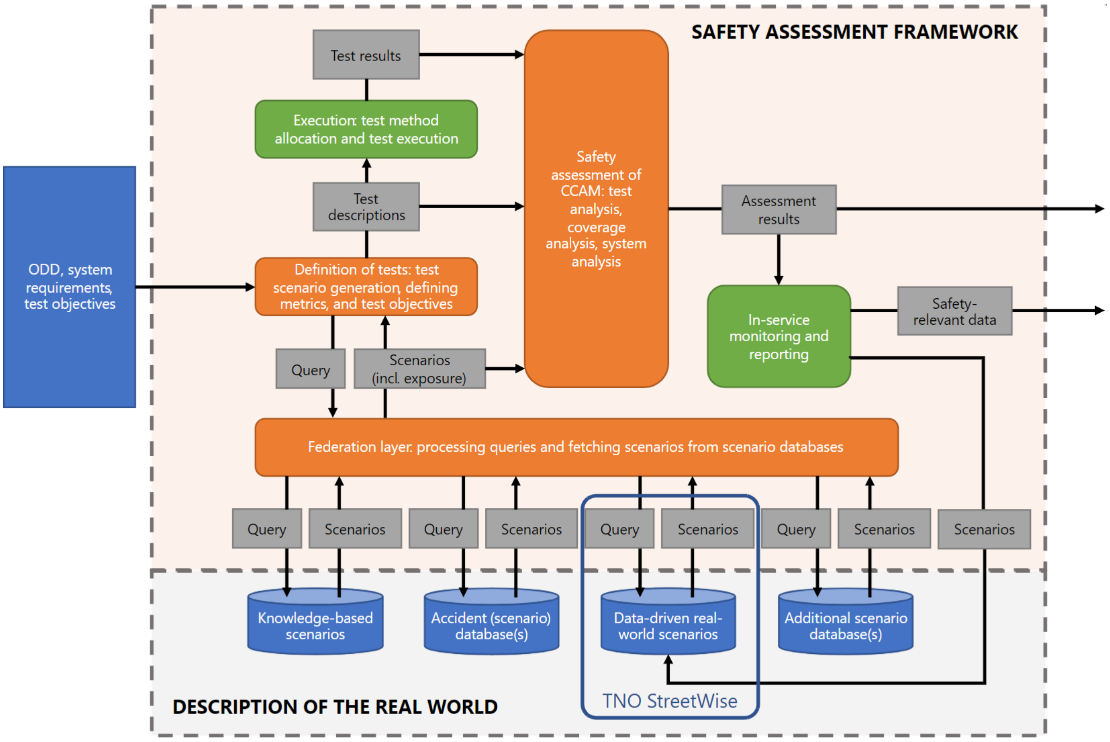

Figure 1.5: The scenario-based safety assessment framework that is in agreement with the NATM multi-pillar approach [15]. The position of the TNO StreetWise scenario database has been indicated along with other types of scenario databases.

The Horizon Europe project V4SAFETY[3] is developing a predictive assessment framework for road safety to provide prognoses on the effects of in-vehicle safety systems. A key element in the framework is the use of simulations, which can model future traffic scenarios and analyse, in depth, solutions not yet available on the market. The V4SAFETY framework is also designed to support the virtual testing of CCAM functions.

- *Analyse and assess the test results to reach a safety statement*: Utilising the test results, along with the test descriptions (including objectives) and the provided test scenarios with their exposure values, a safety assessment of the CCAM system is conducted. This assessment evaluates whether the system appropriately responds in all test scenarios, meeting the requirements and expectations for scenarios relevant to its ODD. The safety assessment includes verifying whether the ODD is adequately covered to make a statement about the system's safety within its defined domain. Based on the assessment results, authorities may provide approval for deploying the system on the public roads, e.g., as part of a large-scale test.
- It is typical for authorities to impose additional constraints on such deployments after the assessment and approval of the CCAM system. To address these constraints, the framework incorporates a process for in-service monitoring and reporting, enabling early detection of any potential compromises to the system's operational safety. Safety-critical data from the system during operation in the real world is shared with stakeholders, including relevant authorities.

In Chapter 2, we will delve into how StreetWise contributes to scenario-based safety assessment by providing essential inputs.

[3] `https://v4safetyproject.eu/`

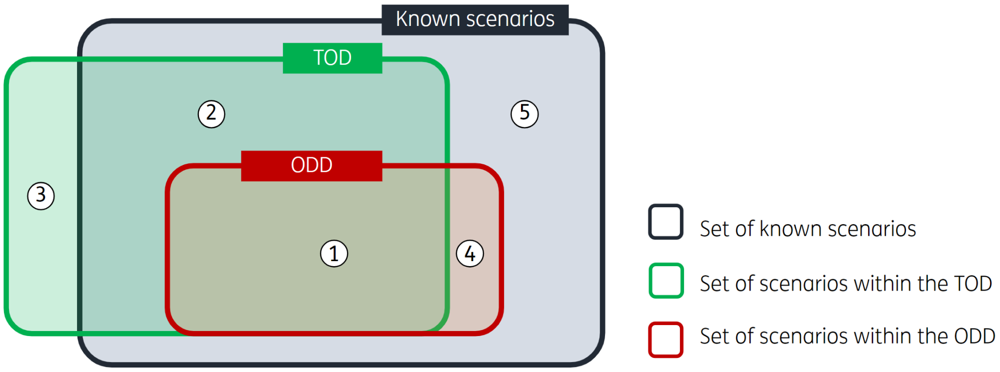

Figure 1.6: Schematic overview of the difference between the ODD and the TOD.

1.5 Scenarios for the description of the operational design domain

The ODD refers to the "operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics" [4, 16]. Essentially, the ODD describes the external environment and conditions for which an ADS is intended. To describe the actual operational domain in which an ADS is operating, the term Target Operational Domain (TOD) is introduced [16].

Ideally, the ODD and the TOD perfectly align, and the scenario database provides a sufficiently complete view on the scenarios that may occur within the TOD. However, in practice, this alignment is often imperfect. Figure 1.6 illustrates the disparity between the ODD and TOD, delineating five distinct areas:

1. The overlap between the ODD and the TOD. This area concerns the part of the TOD that is included in the design of the ADS.
2. The area of known scenarios that is only covered by the TOD. This area refers to the known external conditions that ADS can encounter while it has not been designed for it. Since the safety cannot be guaranteed for these conditions, it is of vital importance to minimise this area.
3. There might be a part of the TOD for which the scenario description is incomplete. These are scenarios that the ADS may encounter but has not been explicitly designed for. Minimising this area by expanding the scenario database is crucial as safety cannot be guaranteed for these scenarios.
4. The area that is only covered by the ODD. This area indicates the parts of the ODD that an ADS will not encounter in its actual operating area. While overengineering for safety is common, minimising this area is desirable to avoid resource wastage.
5. The area that is neither covered by the ODD nor the TOD. These scenarios are outside the scope of the ADS's intended functionality and are thus irrelevant for consideration within the ODD or TOD.

The importance of describing the ODD and TOD is underlined by various initiatives aimed at specifying an ODD, as evidenced by references [17, 18, 16]. It is noteworthy that these initiatives assume a knowledge-based approach: based on the targeted area of operation, the ODD is specified. Also, the environmental conditions that are part of the ODD are typically based on what the ADS can handle, rather than what can be encountered during operation.

StreetWise offers an alternative approach. Since scenarios are a description of the ADS-equipped vehicle's environment, the scenarios can be used to describe the ODD and TOD. Using data captured within the TOD, StreetWise's automatic scenario extraction capability facilitates the creation of a scenario catalogue that comprehensively describes the TOD. While acknowledging the effectiveness of a knowledge-based approach, we advocate for complementing it with a data-driven approach as proposed with StreetWise. The collected scenarios from real-world data can be used to minimise the part of the TOD that is not covered by the ODD, which is critical for guaranteeing the safe performance of the ADS in its entire TOD.

1.6 Scenarios and in-service monitoring

In Section 1.4, the concept of In-Service Monitoring and Reporting (ISMR) has been introduced as an important mechanism to provide information on the responses of an ADS during deployment, both to the ADS developers as to the authorities. This feedback loop enables continuous verification of whether the ADS responds appropriately in all encountered real-world scenarios.

There are several reasons that may lead to an inappropriate response of an ADS in practice. One possibility is that the TOD is not fully covered by the ODD. In other words, the ADS may encounter scenarios for which it is not designed or tested. A monitoring function can be used to determine the actual scenarios that are being encountered, and to compare those with the scenarios within the ODD. Encountered scenarios outside the ODD should be reported, e.g., to enhance and extend the current ODD description and extend the set of test scenarios for safety assessment accordingly.

Moreover, the traffic system evolves continuously, with the introduction of new mobility systems or technologies potentially leading to unforeseen scenarios. A monitoring function can detect when such long-term changes become significant enough to require updates to the ADS functionality.

Furthermore, the internal state of the ADS may change, leading to degradation in its original functionality, such as due to component failure or degradation. This could result in variations in the response to an already known scenario. Any such deviations in response should be reported by the monitoring function, as they could lead to unsafe behaviour.

In all of these cases, it is imported that not only the ADS response is reported, but also information on the scenario (or sequence of scenarios) at the time of the observed response. Only then, a proper analysis is possible on the implications of the observed changes in ADS behaviour, potentially resulting in additional requirements for the ADS and its safety assessment.

For effective in-service monitoring and reporting, an online functionality with efficient anomaly detection is required to promptly share safety-relevant data with the ADS developers and relevant authorities. Additionally, offline data analysis is foreseen to obtain detailed information about encountered scenarios. Continuing scenario identification and characterisation after the deployment of the ADS is essential not only for monitoring the behaviour of the ADS, but also to further enhance the representativeness and coverage of scenario databases for use in ADS development and safety assessment.

1.7 Scenarios for consumer testing

Consumer testing programs around the world serve as extensions of local legal requirements, often imposing stricter and more rigorous, albeit non-mandatory, standards. These assessments handle active, passive, and even post-crash safety. Currently, consumer organisations predominantly focus on assessing assisted driving features rather than fully automated driving.

Euro NCAP stands out among consumer testing organisations due to its challenging test and assessment requirements, reflecting the high standard of equipment in European vehicles. Within the Euro NCAP standard star rating, the focus is on the emergency ADAS, such as AEB and ALKS[4], where Euro NCAP also has an Assisted Driving grading[5], which addresses assisted systems with longitudinal and lateral support.

Traditionally, Euro NCAP and other consumer organisations have based their scenario selection heavily on accidentology to address the most impactful scenarios. In addition, testability in terms of the feasibility of the test itself and the costs and efforts to execute the tests, is taken into account to verify safety performance.

Increasingly, a scenario-based approach has been adopted by consumer testing programs worldwide, particularly in evaluating active safety performance. This allows addressing a much wider area of application to ensure more robust systems with a larger safety benefit.

In the coming years, Euro NCAP will further extend the assessment envelope by introducing a new test scenarios, such as encounters with powered two-wheelers, and incorporating more variations on existing scenarios. These variations will include different lighting and weather conditions and variations in the appearance of the other road users (wearing different clothing) and static environment (such as different lane markings). Additionally, the characteristics of the dynamic behaviour of various road users will be diversified beyond current combinations.

Euro NCAP is also considering incorporating on-road verification testing, where public road testing will be used to verify the expected performance. For the scenarios encountered during public road testing, expected behaviour will be compared with observed behaviour.

As Euro NCAP and other consumer organisations increasingly assess vehicles with higher levels of automation, both assisted and fully automated, the scenario-based approach becomes even more crucial in addressing the operational domain of these features and vehicles.

1.8 Scenarios for the verification and validation of data-driven AI systems

Automated driving functionality increasingly relies on machine learning (or data-driven AI) components. For camera-based object detection, essential for many driving tasks, the state of the art is deep neural networks that are trained on large sets of labelled images. Deep learning is also being deployed in the planning task, or in end-to-end learning of the driving task (from images directly to control signals). The use of these types of data-driven AI poses unique challenges to the Verification & Validation (V&V) process and safety assessment of the full ADS.

The V&V process aims to ensure that a system meets specified requirements. However, chal-

[4] Euro NCAP | The Ratings Explained

[5] Euro NCAP | Assisted Driving Gradings Explained

lenges arise with data-driven AI due to the mismatch between the statistical level at which data-driven AI algorithms work, and the symbolic level at which the ODD and the requirements are specified. For instance, while the requirement may be to detect all pedestrians within a certain range, deep neural networks operate at the pixel level and lack an understanding of symbolic concepts like "pedestrian." The question is then how to ensure that the system has learned all possible pixel arrangements that comprise a pedestrian in the specified range. Similarly, the ODD of the system is given in the form of a dataset of images. How can we ensure that these data encompass everything in the symbolic ODD description, and in the validation process, how can we ensure that this matches the TOD?

Deep neural networks are notoriously black box in nature, making it impossible to ascertain their performance through inspection of the algorithm itself. Testing becomes essential, but exhaustive testing of all possible inputs is impractical: it is impossible to test all the possible pixel arrangements in a camera image to verify the correctness of the algorithm in all possible inputs. Symbolic testing offers a solution: instead of testing all possible pixel configurations, you only test pixel configurations that are likely to occur in the real world. In case of the pedestrian detector, this would mean testing all possible ways a pedestrian could occur in the image. An additional advantage is that this test description provides a way to quantify performance at the symbolic level, thus bridging the gap with the requirement specification.

Scenarios provide a natural framework for describing and quantifying the ODD and TOD of data-driven AI systems. They allow analysis to ensure dataset compatibility with system specifications and absence of biases. Scenarios, including scene snapshots, play a crucial role in providing input for system tests. This step is integral to the V&V process of data-driven AI systems.

2 StreetWise method

StreetWise entails an ongoing continuous process of mining scenarios out of real-world data and storing them into a database. The various applications for which the scenarios can be used are described in the next chapter. This chapter focuses on the process of scenario mining. Initially, the chapter elucidates the distinction between the term “scenario” and its counterpart, the so-called “scenario category,” within the context of StreetWise. Subsequently, Section 2.2 delineates the types of data processed by StreetWise to extract scenarios. The identification and categorisation of the scenarios is detailed in Section 2.3, which also describes the current scope of StreetWise scenario mining process.

2.1 What are scenarios and scenario categories?

When discussing scenarios for the assessment of an Automated Driving System (ADS), typically a distinction is made between different abstraction levels of scenarios [19, 6, 20]. StreetWise adopts the terms “scenario” and “scenario category”, where the former is defined as follows [6]:

A scenario is a quantitative description of the relevant characteristics and activities and/or goals of the ego vehicle(s), the static environment, the dynamic environment, and all events that are relevant to the ego vehicle(s) within the time interval between the first and the last relevant event.

When scenarios are used for, e.g., scenario-based assessment or validation of the compliance with regulations, the practically infinite number of distinct scenarios makes these tasks challenging. To deal with this infinite number in a structured manner, we have proposed the use of scenario categories [6, 21].

A scenario category refers to — quite literally — a category of scenarios and can be regarded as an abstraction of a scenario. A scenario category comprises multiple scenarios. Note that scenario categories are not mutually exclusive, i.e., a scenario can be comprised by multiple scenario categories. Additionally, there can be distinct levels of abstraction among scenario categories, with some categories including others at a more specific level.

StreetWise formalises the relationship among scenario categories and between scenario categories and scenarios using tags. In simple terms, a scenario is comprised by a scenario category if the scenario contains all tags of the scenario category (more details on the categorisation are presented in Section 2.3). To accommodate the need for both generic scenario categories and more specific scenario categories, the tags are structured in hierarchical trees, where the different layers of the trees can be regarded as different abstraction levels. These trees of tags enable the specification of various scenarios with varying levels of detail. For example, a tree of tags might include distinctions like “standing still,” “driving forward,” and “reversing” within the category of longitudinal activity, with further specifications for each subcategory like “decelerating,” “keeping speed,” and “accelerating.” Examples of such tag trees are depicted in Figures 2.1 and 2.2, illustrating different levels of abstraction within the categorisation framework. Extensive examples are presented in [22] and [23].

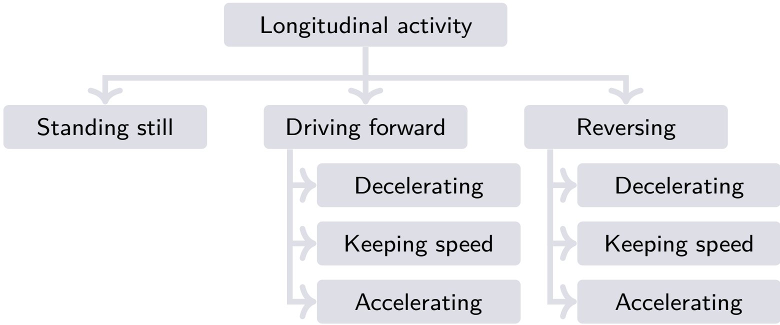

Figure 2.1: Tree of tags for longitudinal activity. Adapted from [23].

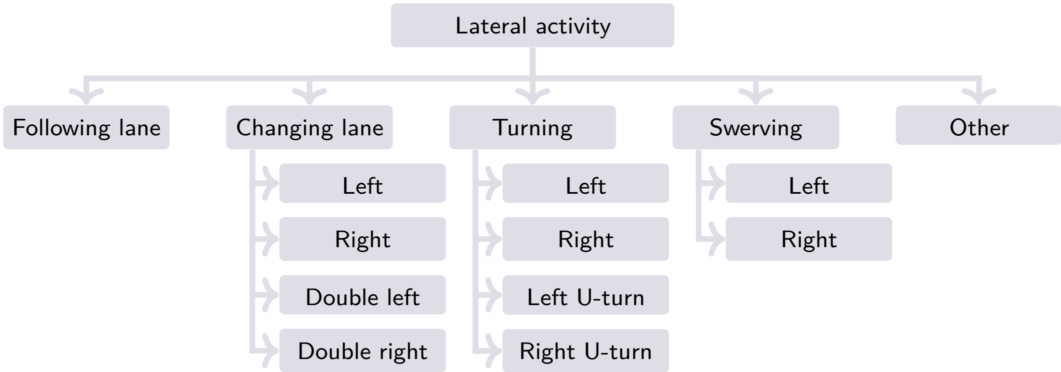

Figure 2.2: Tree of tags for lateral activity. Adapted from [23].

2.2 Data collection

As scenarios describe the dynamic manoeuvres of all relevant actors possibly interacting with an ego vehicle, the real-world scenarios can be generated from the sensor output of a vehicle, such as an accelerometer, camera, radar, and Global Navigational Satellite Systems (GNSSs). This is considered as a minimum set of information required to describe the dynamic part of a scenario. Further enriching a scenario is possible by incorporating additional on-board information from the CAN-bus, such as rain and light sensor output and windshield wiper system and direction indicator status.

For StreetWise, object-level data is required, instead of raw sensor data[6]. Object-level data, often produced by sensor fusion algorithms or on-board world modelling tools, assigns each detected object an ID, type (e.g., pedestrian, passenger car, truck, motorcycle, general object), and state variables like relative position, speed, and heading with respect to the ego vehicle, all tracked over time. While a 360-degree view of the environment is optimal, StreetWise can still function with object-level data from a forward-looking sensor set. As the most relevant activities are the ones of the ego vehicle and the other traffic participants in front of the ego vehicle, no strict requirements apply to the field-of-view of a sensor system for data collection. However, limited field-of-view might result in limitations regarding the scenario mining possibilities and

[6]There are several practical reasons to use object data as input to StreetWise instead of raw sensor data:

- The conversion of raw sensor data to object-level data is proprietary to the sensor manufacturers. By collecting object-level data, the method of analysing this data becomes independent of the type of sensor, the sensor technology and the sensor make;
- The storage size of object-level data is easily a factor of 100 smaller than that of raw data. Collecting object-level data limits the amount of data that needs to be transferred from data collection vehicles, and it limits storage size.

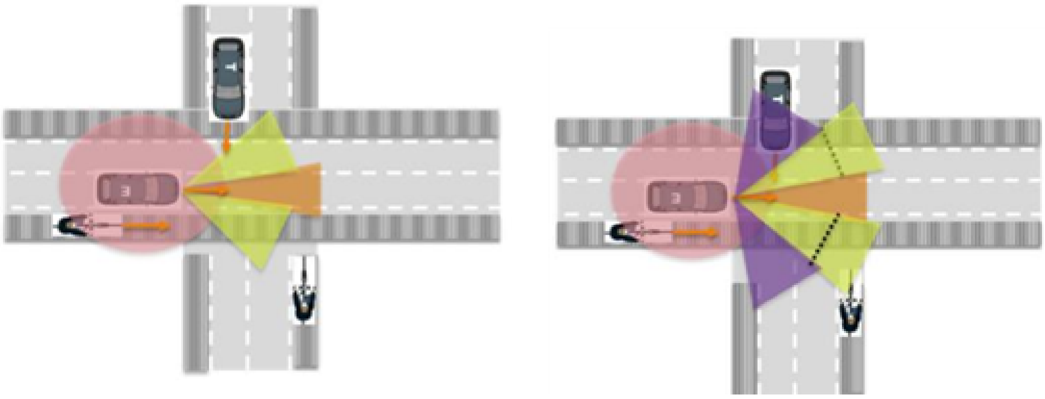

Figure 2.3: Example to illustrate how the field-of-view of the sensor set affects the scenario mining possibilities. A sensor set with limited field-of-view (left) does not perceive the crossing actors that are perceived by the sensor set with wider (recommended) field of view (right).

scenario description, illustrated in Figure 2.3. An example of a dataset with object-level data is published by Paardekooper et al. [10].

In addition to in-vehicle data (as described above), dynamic manoeuvres of interacting actors can be collected from other data sources like road-side cameras or drones. In the analysis of a dataset from such sources, all detected vehicles can take the role of ego vehicle and the scenarios are identified with respect to each of the detected vehicles separately [24]. This approach extends the field-of-view around the ego vehicle can be extended, which is especially useful for scenario mining in urban areas, e.g., at intersections and roundabouts, or specific road layouts such as motorway ramps. Notable examples of datasets captured by drones include the highD [25] and inD [26] datasets.

To complete the scenario description, data describing the static environment is required. This includes details on the road layout such as lane information, types of lane markers, road edges, and road furniture. TNO uses a GNSS signal with timestamp to get access to information regarding the static environment during data collection. By combining the GNSS signal with a map, the ego vehicle and surrounding objects can be accurately projected on a map. This approach enhances the scenario description by incorporating map information such as local road layout, intersections, sidewalks, cycle lanes, tunnels, lane counts, and speed limits. While freely available maps like OpenStreetMap[7] can be used, employing a high-definition map is recommended for enhanced accuracy, particularly for map matching of objects around the ego vehicle. The GNSS signal and timestamp also support the collection of data regarding weather and lighting conditions, which can be referenced against meteorological databases.

2.3 Scenario identification and categorisation

Several data processing steps are distinguished for scenario identification and categorisation.

2.3.1 Activity detection

An activity constitutes a quantitative description of the time evolution of one or more state variables of an actor between two events. For instance, an activity could describe the longitudinal

[7] www.openstreetmap.org

acceleration or speed during an acceleration or deceleration. Similarly, activities pertaining to the lateral position of a vehicle with respect to the centre of the corresponding lane might be categorised as “following lane” or “changing lane,” as depicted in Figure 2.2 Activities are the smallest building blocks to describe the dynamic and static environment in a scenario.

As a first step towards scenario identification and categorisation, TNO develops techniques and algorithms to automatically detect activities in collected real-world (microscopic) traffic data. These are hybrid techniques that combine physical/deterministic models with data analytics to detect activities hidden within terabytes of data. Through activity detection, the time history data is converted into discrete data.

This technique uses domain expertise of vehicle dynamics modelling as well as data analytics including artificial intelligence. The detection method provides an overview of the type and frequency of activities. Subsequently, during processing, activities within mined scenarios are stored including parameters describing their characteristics. For example, the maximum speed in lateral direction during a lane change serves as an indicator of the aggressiveness of the lane change.

Both longitudinal (e.g., keeping speed, accelerating, decelerating, standing still, reversing) and lateral (e.g., following lane, changing lane, turning) activities are used to describe the dynamic traffic. Furthermore, relative activities are determined to discretise the longitudinal and lateral relative position and velocity between actors. Likewise, activities can also be recognised in the description of the static environment, e.g., entering a tunnel, approaching an intersection with a view-blocking obstruction, or passing a billboard. Similarly, communication between traffic participants and infrastructure (V2X) can be described with activities (more details on incorporating V2X communication in scenarios is found in [27, 28]). Regarding lighting and weather conditions, activities are used to characterise scenarios, though they tend to be less dynamic compared to activities related to dynamic traffic. While dynamic traffic activities typically exhibit time constants in the order of seconds, changes in lighting and weather conditions are usually occur on a scale of minutes or even hours. In case of entering a tunnel or passing under an overpass, the change is drastic and lighting conditions for the ego vehicle will strongly vary within seconds.

2.3.2 Scenario mining

In scenario mining, independently identified activities for the ego vehicle, other traffic participants’ dynamic behaviour, the static environment, and environmental conditions are amalgamated to formulate a scenario [29]. An example is given in Figure 2.4, for a vehicle being overtaken while entering a tunnel.

The figure showcases blocks representing the identified activities for the scenario. It is evident that activities might overlap, yet exhibit a certain relationship. In this case, the lighting conditions for the ego vehicle are related to entering a tunnel (as part of the static environment). Similarly, the manoeuvres of the ego vehicle and the passing vehicle in the dynamic traffic description need to follow the road inclination into and out of the tunnel. Notably, that the example scenario of being overtaken by a vehicle while driving into a tunnel can be considered as the superposition of multiple scenarios overlapping in time:

- Cut-out at rear of ego vehicle
- Ego vehicle being overtaken
- Ego vehicle driving through a tunnel

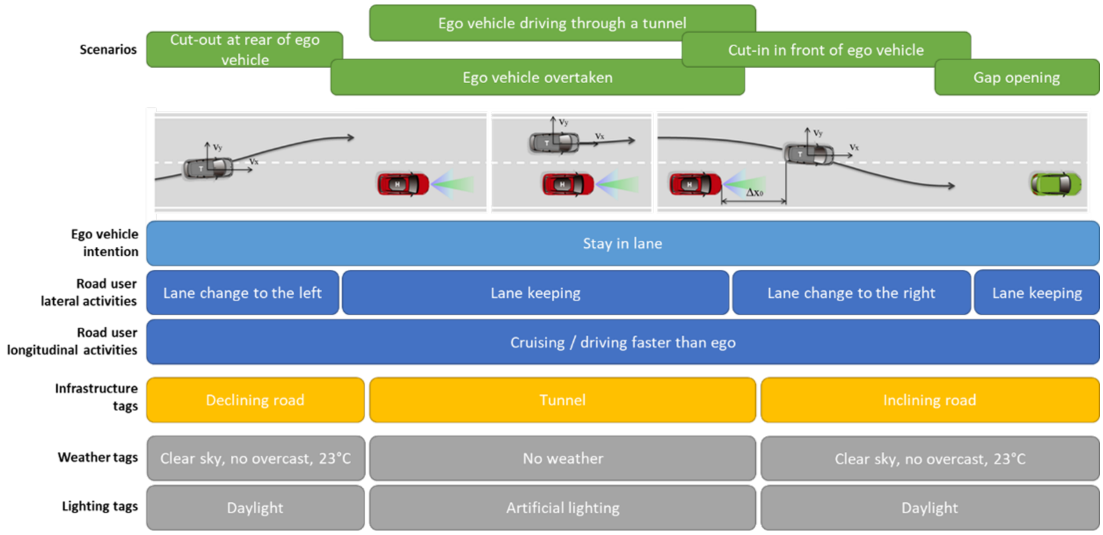

Figure 2.4: Activities and scenarios for an ego vehicle (marked red) that is being overtaken in a tunnel by another vehicle (marked grey).

- Cut-in in front of ego vehicle
- Gap opening

2.3.3 Scenario parametrisation and database

In the final processing stage, the mined scenarios are stored using the StreetWise domain model, an object-oriented framework for scenario and scenario category descriptions [6]. The StreetWise scenario database does not contain raw sensor data, but a parametrised model of the real world based on the sensor signals, thereby minimising dependence on the specific sensor set used during scenario recording.

Relevant time history signals of dynamic activities and the static environment within the mined scenarios are parametrised by using curve fitting to achieve a strong data reduction. Furthermore, based on the activities, tags are added to the scenario. Key parameters of a scenario are computed and added to the scenario description. Finally, the scenarios, as described in the StreetWise domain model, can be stored to a scenario database.

Characterising scenarios with key parameters facilitates the extraction of multidimensional Probability Density Functions (PDFs) for scenario categories. In this way, differences in scenario categories can be quantified. These parameter distributions will unveil typical behaviour within a category, as well as less common instances. The database encompasses not only critical scenarios but also everyday road behaviour and typical parameter ranges describing it. The mechanism of tagging and combining scenarios into scenario categories allows for investigation of how often certain scenarios occur on the road or in a specific Target Operational Domain (TOD). By selecting specific tags, analysts can scrutinise disparities between scenario categories and gauge the relevance of certain categories and parameter ranges.

The StreetWise platform offers an Application Programming Interface (API) that provides access to the database content, facilitating automatic test scenario generation. This API serves as the backbone for a web-based Graphical User Interface (GUI) that enables users to retrieve information and statistics about the scenarios stored in the database. Through the API, users can define parameter ranges for test scenario generation and subsequently download the generated test scenarios. Furthermore, individual parameter distributions can be downloaded. Various software tools, including test automation applications, can directly interact with the API. Within the API and GUI, scenario tags can be used to specify the scenario space for test scenario generation and tag-based visualisations to get more insight into scenario statistics for different Operational Design Domains (ODDs).

3 How to use StreetWise

While Chapter 1 presented multiple challenges that can be addressed with StreetWise, this chapter details how StreetWise can be used to actually address those challenges. First, Section 3.1 presents on a high level how StreetWise supports risk assessment according to the process shown in Section 1.3. Section 3.2 describes how StreetWise can be instrumental in demonstrating compliance with relevant regulations, as discussed in Section 1.4. In Section 1.5, we have explained how scenarios can be used to describe an Operational Design Domain (ODD); in Section 3.3 elaborates on how StreetWise contributes to this task. Section 3.4, which relates to Section 1.6, explains how StreetWise facilitates the in-service monitoring of Automated Driving Systems (ADSs), ensuring ongoing safety and performance. This chapter ends with a short note on StreetWise in relation to consumer testing in Section 3.5.

3.1 StreetWise for risk assessment

In the context of ADSs, ensuring traffic safety is paramount before deployment on public roads. Retrospective safety validation, relying solely on (test) drives with prototypes, is neither safe nor feasible due to the vast amount of driving required. Therefore, prospective safety validation is widely embraced, with scenario-based safety validation being a prevalent approach that is widely adopted in the automotive domain, by regulatory bodies, such as United Nations Economic Commission for Europe (UNECE) [11, 15] and European Commission (EC) [13], industry, research institutes, and academia [30, 31, 32, 33, 34].

StreetWise plays a crucial role in supporting data-driven, scenario-based assessment, as mentioned in Section 1.3. Terms like “reasonably foreseeable” and “reasonably preventable” are essential in this context, but they require quantification for practical application. Additionally, quantifying safety risk is imperative, along with determining the confidence level in the estimated risk.

3.1.1 Risk quantification

Figure 3.1 provides a schematic overview of the method for quantifying risk using StreetWise data. A summary of the risk quantification method is given here; more details are provided in [35]. The first step is to identify the scenarios that are part of the ODD of the ADS. To deal with the large variety of scenarios, the scenarios are grouped into scenario categories and the remainder of the risk analysis is performed for each scenario category. This categorisation can be based on observed data, and triggering conditions (factors causing hazardous behaviour) can be appended to scenarios at this stage, as per the ISO 21448 standard on Safety of the Intended Functionality (SOTIF) [36].

The next step is to calculate the exposure of the scenarios that belong to the specific scenario category. Following the scenario identification and categorisation as explained in Section 2.3, the exposure can be calculated from data analysis. If, however, the scenario occurrence is very rare, e.g., if rare triggering conditions are included, expert knowledge might be considered as part of the exposure estimation. We express the exposure as $\mathbb{E}[n_{\mathcal{C}}]$, where $n_{\mathcal{C}}$ is the number of occurrences of a scenario from scenario category $\mathcal{C}$ per hour of driving and $\mathbb{E}[\cdot]$ denotes the expectation.

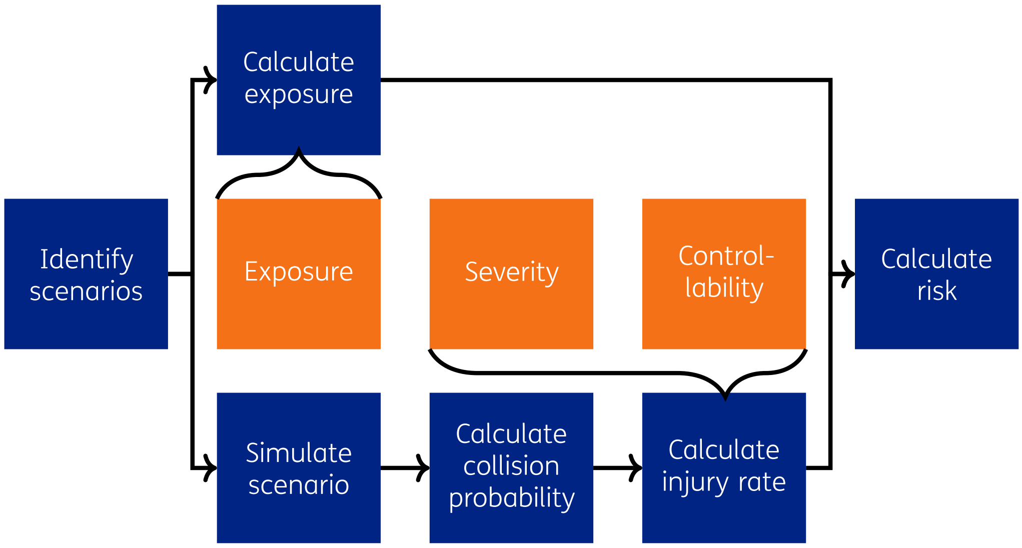

Figure 3.1: Schematic overview of the risk quantification using StreetWise data. The blue blocks represent the six steps of risk quantification. The orange blocks show the relation between the risk quantification method and the three aspects of risk as proposed in the ISO 26262 standard [37]. The figure has been adapted from [35].

The third step in the process is the simulation of the response of the ADS in a specific scenario. Let $R(S)$ denote the (potentially stochastic) result of the simulation of scenario S, where $R(S) = 1$ indicates a collision and $R(S) = 0$ otherwise. We assume that a tool is available that produces this simulation result $R(S)$. This may vary from a straightforward simulation tool to very complex simulations with some parts of the system under test integrated physically in the simulator. Physical simulation is feasible, although we expect that virtual simulations is inevitable due to the impracticality and expense of conducting a large number of tests physically.

Simulation results are used to calculate the collision probability, which is the fourth step. In addition, the probability of the scenario S is used. Or, more accurately, the probability density, since the scenario space is typically a continues space. We denote this probability density by $p(S|S \in \mathcal{C})$, i.e., the probability density of the scenario S, given that the S belongs to scenario category $\mathcal{C}$. Notably, this probability density must integrate to 1 when integrated over all scenarios that belong to scenario category, i.e., $\int_{S\in\mathcal{C}} p(S|S \in \mathcal{C})\,\mathrm{d}S = 1$. A practical approach for obtaining the Probability Density Function (PDF) $p(S|S \in \mathcal{C})$ is to parametrise the scenarios belonging to $\mathcal{C}$ and to use any numerical PDF estimation technique. Following the StreetWise paradigm, the scenarios collected from driving data could be used for estimating the PDF, e.g., see [35]. Using the probability density of S, the collision probability is computed by evaluating the following integral:

$$\int_{S\in\mathcal{C}} R(S)p(S|S \in \mathcal{C})\,\mathrm{d}S. \tag{3.1}$$

Notably the (numerical) evaluation of this integral is far from straightforward. Estimating the integral using the fewest evaluations of $R(S)$ possible presents a particular challenge, given that evaluating $R(S)$ is typically costly. That is why there is a significant amount of literature dedicated to this topic. Further elaboration on this subject is beyond this work's scope.

Not all collisions are equal. That is why it is crucial to consider not only the probability of collision but also the potential harm inflicted in each collision. Let $p(I(S))$ indicate the probability that scenario S leads to a to-be-specified harm. For example, this specified harm could be a moderate injury or worse, corresponding to a Maximum Abbreviated Injury Scale (MAIS) [38] level of 2 or higher, but it could also be a fatality. We assume that there is a method available to determine $p(I(S))$. Typically, phenomenological injury risk functions or in-crash simulations for which the earlier mentioned pre-crash simulation provide the initial and boundary conditions, are used. Similar to the collision probability, the injury rate or fatality rate is computed by evaluating the following integral:

$$\int_{S\in\mathcal{C}} p(I(S))R(S)p(S|S\in\mathcal{C})\,\mathrm{d}S. \tag{3.2}$$

The risk associated with scenario category $\mathcal{C}$ can be defined as the combination of the probability of occurrence of a scenario of $\mathcal{C}$ and the probability of an injury or fatality given such a scenario, thus combining the exposure $\mathbb{E}[n_\mathcal{C}]$ with Eq. (3.2):

$$\mathsf{Risk}(\mathcal{C}) = \mathbb{E}[n_\mathcal{C}] \cdot \int_{S\in\mathcal{C}} p(I(S))R(S)p(S|S\in\mathcal{C})\,\mathrm{d}S. \tag{3.3}$$

To estimate the risk of an ADS, we are not only interested in the risk for one scenario category, but for all scenario categories. Therefore, the resulting risk associated with individual scenario categories need to be combined. The risks of two scenario categories can be combined as follows:

$$\mathsf{Risk}(\mathcal{C}_1 \cup \mathcal{C}_2) = \mathsf{Risk}(\mathcal{C}_1) + \mathsf{Risk}(\mathcal{C}_2) - \mathsf{Risk}(\mathcal{C}_1 \cap \mathcal{C}_2).$$

In general, it is sufficient to estimate the upper bound of the risk, so in case it is practically difficult to evaluate $\mathsf{Risk}(\mathcal{C}_1 \cap \mathcal{C}_2)$, one can use:

$$\mathsf{Risk}(\mathcal{C}_1 \cup \mathcal{C}_2) \leq \mathsf{Risk}(\mathcal{C}_1) + \mathsf{Risk}(\mathcal{C}_2)$$

where equality applies if two scenario categories, $\mathcal{C}_1$ and $\mathcal{C}_2$, do not overlap.

As shown in Figure 3.1, the risk as calculated using Eq. (3.3) is related to the aspects of risk as formulated in the ISO 26262 standard [37]: exposure, severity, and controllability. In [35], we have shown the $\mathsf{Risk}(\mathcal{C})$ can be expressed as the product of exposure, severity, and controllability, which is consistent with the risk estimation as proposed in the ISO 26262 standard.

3.1.2 Uncertainties in risk quantification

For using the estimated risk as outlined in Section 3.1.1, it is important to have enough confidence in the quantification of the risk. Uncertainties may arise for various reasons, for example due to:

- Lack of data: There is a limitation in the amount of data that can be captured and processed to estimate the exposure $\mathbb{E}[n_\mathcal{C}]$ and the probability density functions of the scenarios $p(S|S\in\mathcal{C})$.
- Inaccuracies in the data: Scenarios may have been missed in the data (false negatives) or the estimated parameters may be inaccurate.
- The potentially limited number of simulations: Even if virtual simulations require less (physical) effort, it will still be infeasible to simulate each and every scenario.
- Inaccuracies in the simulation of the scenarios: Since the simulations use models of the reality, inaccuracies are inevitable. As a result, the results of the simulations may also be inaccurate.

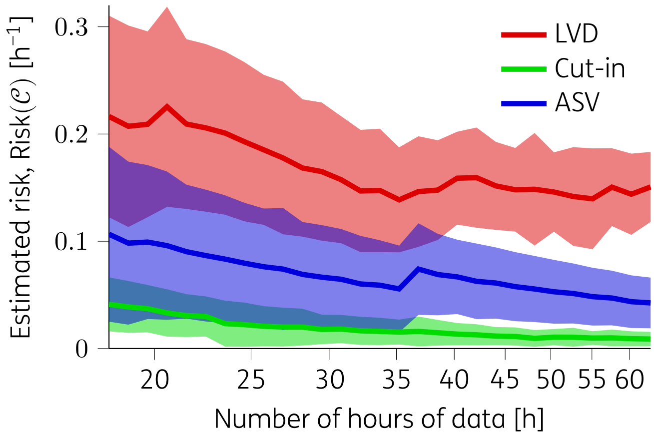

Figure 3.2: Estimated collision risk considering different amount of data for three scenario categories: Leading Vehicle Decelerating (LVD), cut-in, and Approaching Slower Vehicle (ASV). The coloured area marks the estimated uncertainty.

- Simplifications of the scenarios: By assuming, for example, that the scenario of a specific scenario category can be described by a finite number of parameters.
- Misspecification of the ODD of the ADS: If the actual ODD of an ADS is substantially different from the operating conditions under which the data were acquired, the data may not be representative of the Target Operational Domain (TOD). This could lead to inaccuracies of the estimated exposure and the estimated PDF of the scenario parameters.

Please note that this list is not complete. It is important to determine the impact of each of these uncertainties on the confidence of the estimated risk.

In [39], a method is provided that estimates the uncertainty of the risk as a result of the limited data and the limited number of simulations. First, based on the variations of how often a scenario is seen per hour of driving, the uncertainty of the exposure $\mathbb{E}[n_{\mathcal{C}}]$ is estimated. Second, the uncertainty of the crash probability of Eq. (3.1) is estimated using bootstrapping of the original data. Using importance sampling, it is possible to estimate the collision probability repeatedly with different input data without the need to redo the simulations. Third, the uncertainty of the collision probability as a result of the limited number of simulations is estimated using a standard formula used for importance sampling with Monte Carlo simulations. Finally, the three uncertainties are combined to calculate the uncertainty of the estimated risk.

Figure 3.2 shows an illustration of the estimated risk uncertainty for an adaptive cruise control in three different scenario categories. The solid lines represent the estimated crash probabilities according to Eq. (3.1). The coloured areas mark the estimated standard deviations of the estimated risk as a result of the limited data. Figure 3.2 shows that when more data are used, the estimated uncertainty drops.

3.2 StreetWise for compliance with regulations

This section discusses several application areas of StreetWise in relation to the existing regulations for ADSs. First, Section 3.2.1 discusses how StreetWise aids in the quantification of determining what “reasonably foreseeable and preventable” means. Section 3.2.2 demonstrates how StreetWise can be used for the validation methodologies (pillars) as mentioned in the New Assessment/Test Methods (NATM) [14], which are virtual testing, track testing, real-world testing, In-Service Monitoring and Reporting (ISMR), and auditing.

3.2.1 Determining what "reasonably foreseeable and preventable" means

As mentioned in Section 1.4 and stated in the regulations for Automated Lane Keeping System (ALKS) [11, 13], an activated ALKS "shall not cause any collisions that are reasonably foreseeable and preventable". The phrasing of "reasonably foreseeable and preventable" leaves room for different interpretations and, therefore, this might result in disagreements among developers and approval authorities. With StreetWise, it is possible to quantify what is "reasonably foreseeable and preventable" using real-world driving data [40]. By utilising data, determining what collisions are reasonably foreseeable and preventable relies less on expert judgement, thus rendering it more objective.

For determining reasonably foreseeable collisions, StreetWise can be used to identify scenarios that fall within this category. Ensuring that an ALKS can safely deal with those scenarios, it is assured that the ALKS avoids reasonably foreseeable collisions. A threshold ϵ_{F}, which has units "per hour", is used to quantify the range of "reasonably foreseeable" scenarios. The proposed method consists of three steps:

1. Identify the scenarios that are part of the ODD of the ADS.
2. Determine the exposure of these scenarios, i.e., the expected number of occurrences per hour of driving.
3. Determine the probability of encountering scenarios within a specified parameter range, denoted by $[x_{\mathsf{L}}, x_{\mathsf{U}}]$.

The first two steps are already explained in Section 3.1.1, where the exposure of scenario category $\mathcal{C}$ is expressed using $\mathbb{E}[n_{\mathcal{C}}]$. For the third step, it is assumed that the scenarios from scenario category $\mathcal{C}$ can be parametrised using the vector x. Two different approaches to determine the probability that $x \in [x_{\mathsf{L}}, x_{\mathsf{U}}]$ are proposed in [40].

The first approach to determine the probability that $x \in [x_{\mathsf{L}}, x_{\mathsf{U}}]$ uses a PDF that is estimated based on the available data. One common approach is the use of Kernel Density Estimation (KDE) [41, 42]. Let $f_{\mathcal{C}}(x)$ denote the probability density of x for scenario category $\mathcal{C}$. Based on the PDF, the Cumulative Distribution Function (CDF) $F_{\mathcal{C}}(\cdot)$ can be obtained. The CDF $F_{\mathcal{C}}(u)$ denotes the probability that[8] $x \leq u$, and is obtained as follows:

$$F_{\mathcal{C}}(u) = \int_{-\infty}^{(u)_1} \cdots \int_{-\infty}^{(u)_d} f_{\mathcal{C}}(x) \, \mathsf{d}(x)_1 \cdots \mathsf{d}(x)_d,$$

where $(u)_j$ and $(x)_j$ denote the j-th entry of the d-dimensional vectors u and x, respectively. The probability $p(x \in [x_{\mathsf{L}}, x_{\mathsf{U}}])$, i.e., x is within the range $[x_{\mathsf{L}}, x_{\mathsf{U}}]$, is estimated using the evaluation of the estimated CDF, $F_{\mathcal{C}}(\cdot)$, at the vertices of the hyperrectangle spanned by x_{L} and x_{U}. In case of a one-dimensional parameter vector ($d = 1$), this would simply be:

$$p(x \in [x_{\mathsf{L}}, x_{\mathsf{U}}]) = F_{\mathcal{C}}(x_{\mathsf{U}}) - F_{\mathcal{C}}(x_{\mathsf{L}}).$$

By solving the following equation with respect to x_{L} and x_{U}, the range of parameters that are reasonably foreseeable can be determined:

$$\mathbb{E}[n_{\mathcal{C}}] \cdot (1 - p(x \in [x_{\mathsf{L}}, x_{\mathsf{U}}])) = \epsilon_{\mathsf{F}}.$$

The second approach to determine the probability that $x \in [x_{\mathsf{L}}, x_{\mathsf{U}}]$ applies the Extreme Value Theory (EVT). Based on this theory, the parameter values that are above a certain threshold are

[8] In case of higher dimensions, $x \leq u$ indicates that each entry of x is not larger than the corresponding entry of u.

distributed according to the Generalised Pareto Distribution (GPD) [43]. The idea is to choose a certain threshold u and only utilise the observed scenarios for which the parameters are larger than this threshold u. Then, as justified by the EVT, the distribution of the parameters can be estimated using the GPD. For the mathematics behind this method, we refer to [40, page 4].

For determining what collisions are reasonably preventable, it is important to realise that not all drivers are equal. Therefore, it is not the question of whether a collision is preventable or not, but rather to which extent the collision is preventable. There are two different ways to address this question. One way is to look at all scenarios that belong to a specific scenario category and determine the collision probability considering all these scenarios. This can be done as outlined in Section 3.1.1 by evaluating Eq. (3.1). The second way to approach this question is to study the extent to which a collision is preventable for each individual scenario. This approach involves comparing how an ADS can safely handle each scenario that a human driver can manage safely. This might be closer to the intuition that ADS must prevent all collisions if it is reasonable to assume that a human driver can prevent these collisions. As noted by Kusano et al. [44], both approaches may complement each other, and a combination is also possible.

Notably, for the first approach, it matters how likely a certain scenario is. Scenarios that are more likely are weighted more, as illustrated in Eq. (3.1). For the second approach, the likeliness of a scenario is not considered — only the performance of the ADS compared to that of a (competent and attentive) human driver.

3.2.2 Using StreetWise for each pillar of the NATM

As outlined in Section 1.4.1, the UNECE is proposing five pillars for the type-approval process of Connected, Cooperative, and Automated Mobility (CCAM) systems: virtual testing, track testing, real-world testing, ISMR, and auditing. StreetWise can provide valuable input to each of these five pillars. This section explains in more detail how this is achieved. Note that next to these five pillars, the NATM proposal by UNECE [14] also marks the scenario catalogue as imperative for the type approval of CCAM systems and ADSs. How StreetWise contributes to the scenario catalogue is described in Section 3.3.

The first two pillars, virtual testing and track testing, consider testing of scenarios in a controlled environment. One of the main challenges for these two pillars is determining which test scenarios should be considered. Scenarios collected from real-world data can provide a valuable source of information for determining the test scenarios. Thus, scenarios collected using StreetWise can be used. Note that these scenarios can be directly translated into test scenarios, but since not all scenarios are equally challenging, it might be more useful to use a different strategy, e.g., see [45]. In any case, the parameter distributions that result from the collected real-world scenarios can be used in combination with the test results to quantify the (absence of) risk (see Section 3.1).

For the third pillar, real-world testing, the environment of the system under test is less controlled. Here, StreetWise can be used to analyse the real-world test. This includes the analysis of the scenarios that have been encountered during the real-world test. The encountered scenarios can be used to determine whether the ODD has been sufficiently covered. Additionally, StreetWise's scenario identification enables comparison of the system's performance within specific scenarios with its performance during virtual and track testing. While StreetWise is valuable for supplying test scenarios for the first two pillars, it contributes to the third pillar by analysing the results.

StreetWise can contribute to the ISMR pillar in a manner similar to real-world testing. That is, the StreetWise scenario identification algorithms can aid the analysis of the system's performance during ISMR. In addition, StreetWise can be used to develop so-called Surrogate Safety Measures (SSMs) that can monitor the safety performance of the system during its operation; more information on the use of StreetWise for the development of SSMs can be found in Section 3.4.

The audit pillar includes the assessment of whether the manufacturer has followed the right processes to ensure operational and functional safety. Since StreetWise can be a part of the process to ensure safety of the CCAM system or ADS, the use of StreetWise may be part of the audit as well. One useful aspect of StreetWise is that it provides measures to quantify the effort that has been put into identifying all situations that the assessed system may encounter during its lifetime. This can be valuable for determining whether sufficient effort has been made to minimise the so-called unknown unknowns, which is one of the goals of the ISO 21448 standard on SOTIF [36]. Additionally, StreetWise can contribute to the structured argumentation regarding the selection of test scenarios, both those that are considered and those that are not, for the first two pillars.

3.3 StreetWise for description of ODD and TOD

There is no single best way to describe an ODD, or a TOD. Several initiatives, such as ASAM OpenODD [18] and the ISO 34503 standard [16], aim to harmonise the description of a TOD. Although these initiatives are certainly useful towards the unambiguous description of a TOD, we believe that for the purpose of safety assessment, it is more useful to think of a TOD as a catalogue of scenarios that an ADS may encounter in its lifetime. Given that this scenario catalogue completely covers the TOD, the validation of the ADS can be done based on the scenario catalogue. This is in line with VMAD's proposal to establish a scenario catalogue that, using the NATM pillars, should validate each safety requirement.

Identifying the scenarios belonging to the scenario catalogue is far from straightforward. At TNO, we believe that a combination of methods is necessary, rather than relying solely on one approach, to develop a comprehensive scenario catalogue.

One method involves employing a data-driven approach to identify different scenario categories and the scenarios, including their parameter values, belonging to those scenario categories. This approach leads to an understanding of the possible variations of scenarios within a scenario category and allows us to quantify the possible parameter ranges (Section 3.2.1). The method to extract the scenarios from the data is already explained in Chapter 2 and, therefore, not further detailed here.

Another approach is to analyse driving behaviour data, including naturalistic driving data, data from real-world testing, and in-service monitoring data. New (types of) scenarios could be identified from the data. It is also possible to establish scenarios based on the data, even though those scenarios have not been observed directly. For example, certain artefacts observed in one context could prove relevant in other realistic scenarios. In [44], based on the analysis of driving data from instrumented Waymo vehicles and a set of so-called core scenarios identified earlier, new scenarios are created. Even though these new scenarios may not have been observed in the driving data, they are still deemed to be realistic and therefore relevant for the safety evaluation.

The TNO StreetWise approach is to use the data in combination with knowledge to come to a scenario catalogue. One of the advantages of using data for developing the scenario catalogue is that the scenario categories represent situations that an ADS can encounter; after all, the scenarios have been observed in the data. Furthermore, the scenario catalogue expands gradually with the collection of more data. It also accommodates changes in scenario types over time, such as those introduced by developments in the traffic system, like the increased presence of ADSs on the road. Combining knowledge with data enables a more efficient use of the available data. For instance, it allows for expanding the scenario catalogue by integrating specific aspects from various observed scenarios.

The TNO StreetWise approach has been used several times to come to a scenario catalogue for a to-be-developed ADS. One example is a set of scenario categories that represent scenarios that an ADS may encounter on Singaporean roads [8].

3.4 StreetWise and in-service monitoring

Before deploying an ADS on public roads, it is imperative to verify that overall traffic safety is not compromised with the ADS's introduction. Despite all validation efforts, it is expected that the actual level of safety still needs to be confirmed in the field, once the ADSs are subjected to a large variety of traffic situations and environmental conditions. Furthermore, it might well be that changes are introduced — either internally, e.g., due to system degradation, or externally, e.g., due to changes in the way other traffic participants respond to the ADS — so the safety of the ADS must be continuously monitored. Therefore, so-called in-service monitoring is required.

During in-service monitoring of an ADS, safety-relevant data is gathered. This includes the occurrence of crashes, as the number of crashes is a clear measure of safety. Given the anticipated low crash rate, it is beneficial to also examine "surrogate safety measures" (SSMs), which are used to express road safety in terms of the safety risk in traffic conflicts. The goal is to establish SSMs that effectively mirror the likelihood of crashes and/or harm; in essence, the higher the criticality of a specific SSM, the greater the likelihood of an impending crash. If an SSM is indeed a good proxy, it can be used to identify a safety issue before the actual occurrence of a crash.

SSMs typically rely on assumptions of how the system controlling the vehicle and the future environment will develop. For example, Time-To-Collision (TTC) [46], the ratio of the distance toward and the speed difference with an approaching object, is computed by assuming a constant relative velocity. As a result of these assumptions, SSMs are only applicable in certain types of scenarios. For example, TTC is only applicable when approaching an object. There is a plethora of SSMs, including SSMs considering complex models for predicting the future ambient traffic state and SSMs applicable for other types of scenarios. For an overview, see [47, 48].

The StreetWise data can be used to develop SSMs that consider the specific capabilities of the system controlling the vehicle as well as the local context for predicting the future of the vehicle's environment. The data-driven Probabilistic RISk Measure derivAtion (PRISMA) method is used to derive SSMs that can be used to calculate in real time the probability of a specific event (e.g., a crash). The PRISMA method [49] has the following characteristics:

- The derived SSMs give a probability that a specified event, e.g., a crash or a near miss, will happen in the near future, e.g., within the next 10 s, given an initial state and the foreseen evolutions of traffic participant trajectories. A probability is easier to interpret than, e.g., a value ranging from 0 to infinity such as the TTC.
- It is possible to use a model of an ADS, in such a way that the derived SSM estimates the safety risk if this ADS controls the vehicle.
- As a data-driven approach is embraced, the derived SSM adjusts to the recorded data. This flexibility allows for the adaptation of the SSM to local traffic behaviours, provided that these behaviours are reflected in the recorded data.

- The PRISMA method is not limited to one type of scenario.

The PRISMA method consists of four steps. First, the “initial situation” and the possible “future situations” are parametrised. Using the “initial situation” and the possible “future situations”, the type of scenario for which the SSM is to be developed is defined. Second, given the initial situation, the probability (density) for the potential future situations is estimated. This could be done using the scenario data obtained using the StreetWise method explained in Chapter 2. Third, using simulations, the probability of the specified event (e.g., crash, near miss) based on the initial and the future situations is estimated in a similar fashion as outlined in Section 3.1.1. The last step is to employ local regression to speed up the calculations and to enable real-time use of the SSM.

3.5 StreetWise for consumer testing

Consumer testing organisations need to understand the characteristics of real-world scenarios to define the specifications of the assessment envelope accurately. This understanding enables them to create a realistic representation of the scenarios encountered in everyday driving conditions. StreetWise provides the exposure of the scenarios including its parameters with ranges, distributions, and interrelations. Risk information based on this exposure information from StreetWise can be used to weigh the different scenarios and parameters if the consumer organisation aims to introduce new scenarios or prioritise specific cases.

4 StreetWise roadmap

StreetWise is still a work in progress. There remain unanswered questions to address and opportunities to explore with StreetWise. This chapter provides a preview on topics that are to be addressed in the future. The first topic that will be discussed in this chapter is the heavy-tail problem, i.e., the fact that many important scenarios occur infrequently. Section 4.2 elaborates on the expansion of the set of scenario categories. In Section 4.3, the use of StreetWise to define competent driving and good roadmanship is addressed.

4.1 Facing the heavy-tail problem

The safe and responsible introduction of Automated Driving Systems (ADSs) onto public roads requires extensive testing against realistic scenarios. In this work, a data-driven scenario-based approach has been proposed for such safety assessment. This approach starts with a description of the wide variety of traffic situations by means of scenarios stored in a scenario database. These scenarios are used to describe the Operational Design Domain (ODD) of ADSs and to generate test scenarios by sampling scenarios based on scenario statistics (scenario parameter distributions).

We have shown how the StreetWise approach is used in gathering scenarios by identifying and characterising them from object-level sensor data obtained from vehicles covering extensive distances on public roads. The collection of scenarios must encompass the diverse array of situations an ADS may encounter in real traffic throughout its operational lifespan. As a result, usually many different scenarios need to be considered to achieve a complete safety assessment. Since the number of variations in scenarios on the road is practically infinite, despite the fact that most variations are easily recognised, the exposure of scenarios is best described as a heavy-tail distribution. Even after driving millions of kilometres, it is not uncommon to encounter a scenario for the first time [50]. It is for this reason that it takes a lot of effort to come to a reasonably complete collection of scenarios, even for a fairly simple ODD. Now, when the ODD is large, e.g., covering multiple countries in the EU, it is sheer impossible for a single organisation, to reach a sufficient level of completeness.

The first step towards enhancing the coverage of scenario databases involves devising a proper quantification method for coverage, as detailed in [51, 52]. Subsequent sections outline various approaches to augment the completeness of scenario databases and the set of test scenarios for safety assessment within a given ODD.

4.1.1 Sharing scenarios

A first possibility to increase completeness of a scenario database is found in cooperation between stakeholders in data collection. Car manufacturers and automotive suppliers typically have multiple test vehicles with state-of-the-art data collection equipment for testing all types of functions on public roads. There is a large potential for increasing the geographical coverage of scenario collection when industry partners would share the scenarios they encountered during testing.

Though the data that is being collected by the test vehicles is highly confidential, the scenarios that can be identified and characterised from the sensor data only provide a description of the situations that the vehicle encountered on the road during testing. No sensitive or confidential information regarding the ego vehicle (functionality or equipment) or privacy sensitive information regarding the surrounding traffic participants is contained in the identified concrete scenarios, and consequently, this information is not contained in the individual partner's scenario database. The scenario mining process not only identifies and characterises scenarios but also anonymises them before storing them in a scenario database. This enables the sharing of scenarios between different organisations without any confidentiality breach.

Figure 1.5 illustrates how information from various scenario databases can be utilised for generating a set of test scenarios. To allow such querying of multiple scenario databases, the project SUNRISE intends to develop a federated approach. This approach involves the development of a federation layer, which acts as a smart interface connected to the various scenario databases, enabling seamless information exchange between them. As host of the StreetWise scenario database, TNO is leading the activities in SUNRISE to establish the requirements for such a federation layer. Moreover, TNO will establish the required interface(s) with StreetWise for flawless communication with the SUNRISE federation layer.

4.1.2 Combining different data sources

To capture the large variety of scenarios with a single source of data, would require large amounts of data unless the ODD is very restricted. Therefore, it is tempting to explore multiple data sources, leveraging the strengths of each to compensate for the weaknesses of others. Different data sources include, but are not limited to:

- Data from different sources: For example, data might be recorded from sensor-equipped vehicles with different field-of-views. Also, it might be possible to use road-side units or camera-equipped drones to collect the data in addition to vehicle data.
- Data from different operating conditions: It may be useful to dedicate parts of the data campaign to certain operating conditions. For example, collecting data during rainy conditions or collecting data at specific (complex) intersections.
- Continuous versus event-driven data: When collecting a continuous stream of data, the required storage capacity quickly becomes demanding. Therefore, vehicles can be equipped with so-called Event Data Recorders (EDRs) that only store data when the EDR is triggered, e.g., when the deceleration surpasses an *a priori* prescribed threshold.

As explained earlier in this work, StreetWise uses the collected scenarios to estimate Probability Density Functions (PDFs) to enable, among others, the quantification of risk (Section 3.1). The objective of using different data sources is to obtain more accurate PDFs of the scenario statistics, allowing for more accurate risk estimation. Note, however, that simply combining data from various sources would result in inaccurate PDF estimations, because the underlying distributions of the different data sources do not match. Therefore, the use of combined data from various sources requires additional efforts to estimate unbiased PDFs.

4.2 Expanding the scenario category library

At present, StreetWise is able to automatically identify scenarios from dozens of distinct scenario categories. These include scenarios that are commonly encountered on motorways as well as the more common scenarios in urban domains. As the ODDs of ADSs expand over time, for instance due to increased system functionality, the aim is to also broaden the scenario category library used in the StreetWise scenario database. This expansion facilitates the identification and characterisation of a wider range of scenario types.

The number of scenario categories encompassed within StreetWise depends on several factors. To illustrate, numerous scenario categories are accompanied by descriptive documentation; however, automated identification from driving data has yet to be implemented for a substantial portion of these scenario categories. To address these divers factors, we introduce the Scenario Category Maturity Levels (SCMLs) in Section 4.2.1. A crucial aspect of expanding the StreetWise scenario category library entails determining the appropriate scenario category for inclusion and increasing its maturity level. This determination depends on the context, and various perspectives are detailed in Section 4.2.2. We conclude in Section 4.2.3 with (research) directions related to the expansion of the StreetWise scenario category library.

4.2.1 Scenario Category Maturity Levels (SCMLs)

In total, five different SCMLs are considered from two distinctive perspectives, illustrated in Figure 4.1. The first perspective recognises the maturity with respect to the data and the second perspective comprehends the knowledge-driven development of the scenario category.

The first SCML refers to a scenario category that has been conceived and possibly discussed within the StreetWise team. While scenarios from this category may have been observed informally, e.g., while commuting, no official documentation exists yet. Therefore, the extent of categories achieving this SCML remains unknown.

A scenario category becomes more concrete when it achieves the second SCML. At this stage, at least one video example is available, and the category has been described with a schematic overview (e.g., see [8] for a description of 67 scenario categories). Though formal documentation is lacking, it is recommended to adopt a clear description understood by experts, possibly referencing the ISO 34504 standard [23]. A first validation, indicated with the orange arrow, can be performed by comparing the video footage with the description.

The next milestone, SCML 3, is reached when formal mining rules are established for detecting scenarios. These mining rules are implemented and used to detect scenarios belonging to the considered scenario category from driving data. Validation is conducted to ensure the accuracy of the automated detection process.

Once the fourth SCML has been reached, many automatically detected scenarios are available. In addition, the parameter set for all scenarios belonging to the scenario category have been defined and the values are automatically extracted from the identified scenarios. Given that many scenarios have been extracted from driving data, some initial statistics on the scenario exposure are available. In addition, a validation of the automatic scenario mining is performed, possibly using spot checking. Initial statistics on scenario exposure and mining performance, including precision, recall, and F1 score, are compiled.

Finally, SCML 5 signifies a stable scenario definition and implementation for a substantial time. Many more scenarios are automatically identified, which results in accurate statistics regarding the scenario exposure, the scenario parameter values, and the scenario mining performance.

4.2.2 Which scenario categories should be added?

Merely augmenting the scenario category library with additional entries or advancing the SCML of the existing scenario categories may not yield significant benefits. Therefore, the important question is: which scenario categories should be prioritised for being added to the existing library? Given the contextual nuances, there is no single best answer. Consequently,

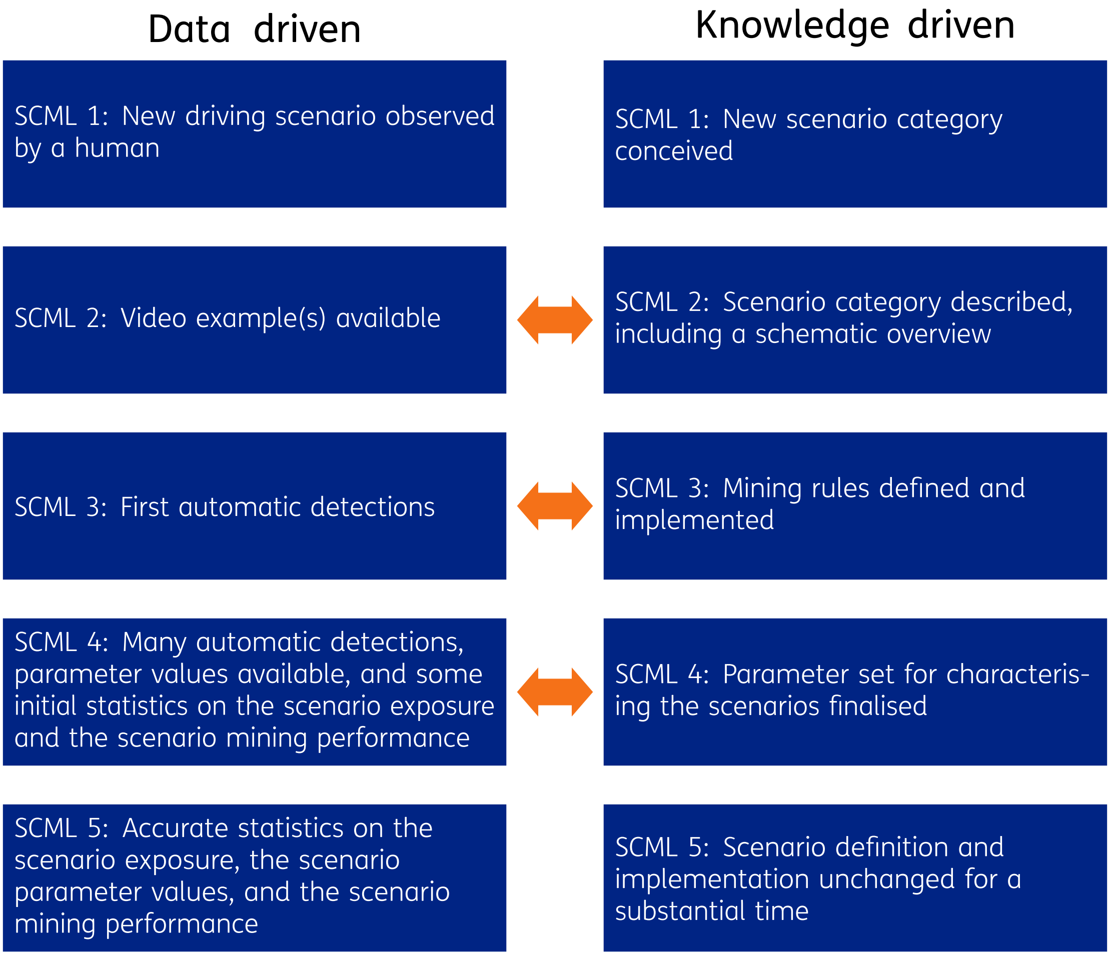

Figure 4.1: The definition of the different Scenario Category Maturity Levels (SCMLs) from a data-driven and knowledge-driven perspective.

no prescriptive method for determining the ideal set of scenario categories is provided here. Instead, we present a list of considerations:

- The set of required scenario categories depends on the ODD that is the subject of the analysis. For example, if only motorway driving is considered, there is no need to consider non-motorway scenarios[9].
- Scenario categories could be added to the scenario category library in order to improve the coverage of the ODD. TNO StreetWise includes coverage metrics that quantify the extent to which the ODD is covered by the identified scenarios [52]. The metrics may be used to decide which scenario category to add or mature.
- Given a data set with driving data, an objective could be to cover the data with scenarios. As with the ODD coverage, TNO StreetWise includes coverage metrics to quantify the extent to which the driving data has been covered by the identified scenarios [52].
- A delicate balance exists between defining broad scenario categories that contain many scenarios and more specific scenario categories that contain fewer scenarios. It could be

[9] This assumes that the ODD can be enforced by the ADS. If this is not the case, then also scenarios that are outside the ODD could be of interest. Also, if the enforcement of the ODD itself is subject of the analysis, scenarios outside the ODD should be considered.

decided to define a scenario category that is fully included in another scenario category. For example, in case the statistics for cut-ins from the left are significantly different from cut-ins from the right, it may be decided to consider both scenario categories “cut-in from left” and “cut-in from right”.

- Another reason to consider a more specific scenario category (of an already existing scenario category) is if diverse behaviour is anticipated from the system under test in the more detailed category. For example, a vehicle may not need to respond to a cut-in if the cutting-in vehicle is driving substantially faster or if the distance is large enough. However, in case a (braking) response is expected from a vehicle after a cut-in, this may be treated as a different scenario category.
- There might be a reason to combine two scenario categories if their respective scenarios occur consecutively, either due to their frequent recurrence or the need for particular attention during safety assessments. For example, a cut-in followed by a leading vehicle decelerating scenario may be considered as a separate scenario category.
- Another reason for combining scenario categories is if their scenarios happen simultaneously and if this concurrent occurrence is relevant for the safety assessment. For example, in case the ego vehicle is approaching a slower vehicle, the ego vehicle may decide to overtake this slower vehicle. However, this may not be possible because the adjacent lane is blocked by another vehicle that is overtaking the ego vehicle. Therefore, the scenario categories “approaching slower vehicle” and “vehicle overtaking ego vehicle” may be combined into one (more specific) scenario category.

4.2.3 Future directions for expanding the scenario category library

TNO aims to expand the StreetWise scenario category library to address a higher variety of scenarios and to cover a larger set of operating conditions. As a result, the StreetWise scenario database will be applicable for more use cases. Within the process of expanding the scenario category library, however, there are some research questions that should be addressed:

- StreetWise aims to achieve a high coverage. As presented in [52], there are multiple views on coverage, such as coverage of the ODD, coverage of all time instants in the data, coverage of all relevant actors in the data, etc. In [52], multiple coverage metrics are proposed. These coverage metrics depend on parameters, such as the tags that are deemed relevant or the actors that should be part as a main actor of a scenario. In an experiment, it is demonstrated that a 100 % coverage can be achieved with some choices of those parameters, and in case a 100 % coverage has not been achieved, the metrics identify which data and scenarios could be added to enhance the coverage. Since achieving 100 % coverage may prove challenging under certain conditions, further research is necessary to ascertain suitable values of coverage depending on the context and selected parameters.
- To get an understanding how to use the SCML that has been presented in Section 4.2.1 and what SCML is feasible for which scenario category, more research is required. For example, it might require a large effort to achieve SCML 5 for a particular scenario category whereas SCML 4 might be sufficient. How the SCML influences the confidence in the assessment results needs to be studied in practical applications.
- The StreetWise scenario categories with SCML 3 or higher currently cover most scenarios that are encountered on the motorway. There are already some scenario categories with automatic detections available for urban environments, but more scenario categories with automatic detections are required for a better coverage of urban areas. Future work is required to increment the SCMLs for scenario categories related to urban areas.

- Additionally, StreetWise has successfully been applied in confined areas, such as airside traffic [53], but this resulted mostly in scenario categories with SCML 2. It is the ambition to increase the SCMLs of the identified scenario categories for confined areas, to enhance confidence in the provided test scenarios for the various automated applications in this domain.
- As mentioned earlier, one of the objectives of expanding the StreetWise scenario category library is to achieve a better coverage of an ODD. It should be noted, however, that for the safe deployment of an ADS, the ADS must be capable of dealing with the Target Operational Domain (TOD) (see Section 1.5). Future research should address how to measure the extent to which the scenario category library covers the TOD.

4.3 Quantifying ‘responsible safety’

The process for assessing the safety of an ADS has been described in Figure 1.5. After an analysis of the tests themselves and the coverage that the tests have been able to achieve, it considers the analysis of the system (and its requirements) and the analysis of the test results to come to a conclusion whether or not the ADS is safe to be deployed in its TOD. The questions are “How safe is safe enough?” [50], and “How to quantify the level of safety?”

The European Commission (EC) Implementing Regulation for the type approval of the ADS of fully automated vehicles [13] requires an ADS to be “free of unreasonable safety risks”. Acceptance criteria to support such a qualification shall be derived considering (among other things) data on performances from competently and carefully driving human drivers [13].

In [11] on the UN Regulation for the introduction of Automated Lane Keeping Systems (ALKSs), a similar formulation is used. The ADS should be free of “unreasonable risk”, which is explained as “the performance of the system shall be ensured at least to the level at which a competent and careful human driver could minimise the risks” [12]. Moreover, the activated system shall not cause any collisions that are reasonably foreseeable and preventable. Both regulations take a competent and careful human driver as a reference, but it is not indicated how the behaviour of such a human driver is quantified.

Since StreetWise uses a large amount of driving data to determine a complete overview of scenarios and the corresponding PDFs of the describing parameters, statistical analyses can be performed regarding the exposure of scenarios. In this way, it can be shown what scenarios are reasonably foreseeable (high exposure) and what driving behaviour (as shown in the manoeuvres of the other traffic participants in the collected scenarios) is most common. To allow such statistical analysis, TNO identifies all driving scenarios in the continuous stream of data that is collected from each drive from departure at the starting point, to arrival at the destination. TNO checks for any gaps or discrepancies in the data stream before mining scenarios out of the data. The data and scenario analysis is finalised by verifying that each data point in the data stream corresponds to at least one scenario. It is important to note that in our scenario definition, scenarios can overlap, allowing a data point to be part of multiple scenarios simultaneously.

Where TNO StreetWise focuses on the description of scenarios and establishing a scenario database to support the generation of test scenarios and the description of ODDs, the TNO StreetProof team focuses on the quantification of competent driving and good roadmanship. StreetProof develops a safety envelope concept, prescribing ranges of acceptable behaviour with respect to other road users, infrastructure, and obstacles [54, 55]. Through scenario analysis, StreetWise provides a wealth of information on driving behaviour. However, for use in research towards competent driving and good roadmanship, a method for tagging scenarios with the quality of the observed driving behaviour is needed. Research is performed by both TNO StreetWise and StreetProof on the use of Surrogate Safety Measures (SSMs) to this end [49].

5 Conclusions

StreetWise is a data-driven methodology that has been developed by TNO to provide real-world scenarios for the development and safety assessment of Advanced Driver Assistance Systems (ADASs) and (Connected) Automated Driving Systems (ADSs). The StreetWise methodology has the following characteristics:

- StreetWise provides a structured approach to identify and characterise scenarios to describe the operational domain of an ADS and to provide appropriate test scenarios for the safety assessment of such system. The data-driven collection of scenarios, that provide a description of what an ADS may encounter on the road during its lifetime, is generic and independent of the function or the technology integrated into the system, such as AI or V2X.
- The way scenarios are collected and stored in the StreetWise scenario database allows for statistical analyses using the Probability Density Functions (PDFs) of the scenario parameters that are stored with the scenarios. Such statistical analyses can be used to determine the level of coverage of the database (How effectively do the scenarios in the database encompass all conceivable road scenarios in the real world?) or for instance the level of coverage of a given Operational Design Domain (ODD) (How effectively do the selected scenarios cover the ODD?) Moreover, the PDFs indicate the exposure of scenarios in the real world, which is an important factor in the estimation of safety risk [39].
- As TNO uses its scenario mining algorithms to extract scenarios and scenario parameters from object-level sensor data, the scenarios in the StreetWise database are automatically anonymised. This way of working is a giant leap forward in pre-competitive cooperation in reusing and sharing valuable scenario information collected in test drives, without running into showstopper discussions on intellectual property or confidentiality issues associated with the original raw data. It enables participants in StreetWise to study differences in scenario characteristics for different cities, countries and continents, without the need to run their own data collection campaigns in each of these areas, and without the need to share raw or object-level data.
- The StreetWise scenario database developed by TNO aligns closely with the requirements for scenarios used in safety assessment and assurance according to the United Nations Economic Commission for Europe (UNECE) multi-pillar approach as shown in Figure 1.5. Test scenarios can be ported as input to any (virtual) simulation tool chain using the OpenDRIVE and OpenSCENARIO XML format, which are recognised open standards. Also other output formats for StreetWise are possible upon request. As the underlying PDFs of the scenario parameters are stored with the scenarios, not only the exposure of scenarios can be quantified, but also the confidence with which the PDFs are estimated. As [39] shows, this approach provides all information to quantify the confidence in the scenario-based assessment results. This provides strong arguments for a well-founded safety assurance claim.
- For the safety of an ADS, it is of paramount importance that the ODD contains the actual operating conditions that it will encounter during deployment, i.e., the ODD should cover the Target Operational Domain (TOD) as much as possible. At TNO, we believe that a TOD can be regarded as a catalogue of scenarios that the ADS can encounter during deployment. StreetWise offers a data-driven approach to create and append such a scenario catalogue. Hence, StreetWise contributes to a complete description of the ODD. In addition, StreetWise enables the detection of changes in the operating conditions as a result of new technologies

or new modes of mobility entering the market, changes in traffic rules, the deployment of new infrastructure solutions, etc.

- StreetWise offers the opportunity to continuously assess the performance of an ADS using in-service monitoring. One way StreetWise contributes to in-service monitoring is through the definition of metrics that estimate the collision risk in real-time. Another approach involves detecting scenarios encountered while the ADS is in service, followed by comparing the ADS's performance in those scenarios with a baseline, such as its performance during pre-deployment tests.

The use of StreetWise has been demonstrated in multiple projects and publications, but that does not mean that StreetWise is a finished product. Instead, the development of StreetWise and expansion of its use is an ongoing process. One aspect that requires future research is how to cope with the challenge of the virtually infinite number of distinct scenarios that might need to be considered for a proper safety assessment of an ADS. We believe that achieving this goal necessitates the sharing of scenarios among various organisations and the integration of diverse data sources. However, further efforts are required to enable secure scenario sharing while safeguarding proprietary and sensitive data. Additionally, more research is needed to conduct statistical analysis on scenario data originating from different sources.

Another important topic for further research is the definition of the right set of scenario categories. Getting this right is a trade-off: on the one hand, the number of scenario categories must be limited in order to keep the safety assessment manageable. On the other hand, it is important that the scenario categories cover the complete ODD and contain all details that are relevant for the performance of the ADS. TNO will focus on, among other tasks, complementing the automatic scenario identification with additional scenario categories. To indicate the maturity of the scenario category description and the automatic identification of scenarios belonging to a scenario category, TNO uses the Scenario Category Maturity Levels (SCMLs).

This document has described how the scenarios obtained using StreetWise can be used for defining the tests needed for the safety assessment of an ADS, where safety is generally defined as the absence of unreasonable risk. Further research is needed to investigate how the scenarios — including the responses of humanly-driven vehicles to traffic conflicts — can be utilised to quantify competent driving and good roadmanship, thereby providing a benchmark for future ADSs.

References

[1] C.-Y. Chan. "Advancements, Prospects, and Impacts of Automated Driving Systems". In: *International Journal of Transportation Science and Technology* 6.3 (2017), pp. 208–216. DOI: 10.1016/j.ijtst.2017.07.008.

[2] Euro NCAP. *Assisted Driving — Highway & Interurban Assist Systems — Test & Assessment Protocol, Implementation 2024, version 2.0.* Tech. rep. 2024. URL: https://www.euroncap.com/media/79863/euro-ncap-ad-test-and-assessment-protocol-v20.pdf.

[3] ECE/TRANS/WP.29/GRVA/2023/20. *Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to Driver Control Assistance Systems.* Standard. World Forum for Harmonization of Vehicle Regulations, 2023. URL: https://unece.org/sites/default/files/2023-07/ECE-TRANS-WP.29-GRVA-2023-20e.pdf.

[4] SAE J3016. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.* Tech. rep. SAE International, 2021. DOI: 10.4271/J3016_202104.

[5] A. Thorsén et al. *SUNRISE D3.1: Report on Baseline Analysis of Existing Methodology.* Tech. rep. European Union, 2023.

[6] E. de Gelder, J.-P. Paardekooper, A. Khabbaz Saberi, H. Elrofai, O. Op den Camp, S. Kraines, J. Ploeg, and B. De Schutter. "Towards an Ontology for Scenario Definition for the Assessment of Automated Vehicles: An Object-Oriented Framework". In: *IEEE Transactions on Intelligent Vehicles* 7.2 (2022), pp. 300–314. DOI: 10.1109/TIV.2022.3144803.

[7] J. E. Stellet, M. R. Zofka, J. Schumacher, T. Schamm, F. Niewels, and J. M. Zöllner. "Testing of Advanced Driver Assistance Towards Automated Driving: A Survey and Taxonomy on Existing Approaches and Open Questions". In: *IEEE 18th International Conference on Intelligent Transportation Systems.* 2015, pp. 1455–1462. DOI: 10.1109/ITSC.2015.236.

[8] E. de Gelder, O. Op den Camp, and N. de Boer. *Scenario Categories for the Assessment of Automated Vehicles.* Tech. rep. Version 1.7. CETRAN, 2020. URL: http://cetran.sg/wp-content/uploads/2020/01/REP200121_Scenario_Categories_v1.7.pdf.

[9] E. de Gelder. "Safety Assessment of Automated Vehicles Using Real-World Driving Scenarios". PhD thesis. Delft University of Technology, 2022. DOI: 10.4233/uuid:bed383ed-c091-4c2a-a530-aada366283d8.

[10] J.-P. Paardekooper, S. Montfort, J. Manders, J. Goos, E. de Gelder, O. Op den Camp, A. Bracquemond, and G. Thiolon. "Automatic Identification of Critical Scenarios in a Public Dataset of 6000 km of Public-Road Driving". In: *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV).* 2019. URL: https://www-esv.nhtsa.dot.gov/Proceedings/26/26ESV-000255.pdf.

[11] E/ECE/TRANS/505/Rev.3/Add.156. *Uniform Provisions Concerning the Approval of Vehicles with Regard to Automated Lane Keeping Systems.* Standard. World Forum for Harmonization of Vehicle Regulations, 2021. URL: https://unece.org/sites/default/files/2021-03/R157e.pdf.

[12] ECE/TRANS/WP.29/2022/59/Rev.1. *Proposal for the 01 series of amendments to UN Regulation No. 157 (Automated Lane Keeping Systems).* Standard. World Forum for Harmonization of Vehicle Regulations, 2022. URL: https://unece.org/sites/default/files/2022-05/ECE-TRANS-WP.29-2022-59r1e.pdf.

[13] European Commission. "Commission Implementing Regulation (EU) 2022/1426 for the Application of Regulation (EU) 2019/2144 as Regards Uniform Procedures and Technical Specifications for the Type-Approval of the Automated Driving System (ADS) of Fully Automated Vehicles". In: *Official Journal of the European Union* L221 (2022), pp. 1–64. URL: `https://eur-lex.europa.eu/eli/reg_impl/2022/1426/oj`.

[14] ECE/TRANS/WP.29/2021/61. *New Assessment/Test Method for Automated Driving (NATM) – Master Document*. Tech. rep. World Forum for Harmonization of Vehicle Regulations, 2021. URL: `https://unece.org/sites/default/files/2021-04/ECE-TRANS-WP29-2021-61e.pdf`.

[15] R. Donà, B. Ciuffo, A. Tsakalidis, L. Di Cesare, C. Sollima, M. Sangiorgi, and M. C. Galassi. "Recent Advancements in Automated Vehicle Certification: How the Experience from the Nuclear Sector Contributed to Making Them a Reality". In: *Energies* 15.20 (2022), p. 7704. DOI: `10.3390/en15207704`.

[16] ISO 34503. *Road Vehicles – Test Scenarios for Automated Driving Systems – Taxonomy for Operational Design Domain for Automated Driving Systems*. Standard. International Organization for Standardization, 2023. URL: `https://www.iso.org/standard/78952.html`.

[17] PAS 1883:2020. *Operational Design Domain (ODD) Taxonomy for an Automated Driving System (ADS) — Specification*. Standard. The Britisch Standards Institution, 2020.

[18] Association for Standardization of Automation and Measuring Systems (ASAM). *ASAM OpenODD*. Accessed May, 2024. 2021. URL: `https://www.asam.net/standards/detail/openodd/`.

[19] C. Neurohr, L. Westhofen, M. Butz, M. Bollmann, U. Eberle, and R. Galbas. "Criticality Analysis for the Verification and Validation of Automated Vehicles". In: *IEEE Access* 9 (2021), pp. 18016–18041. DOI: `10.1109/ACCESS.2021.3053159`.

[20] ISO 34501. *Road Vehicles – Test Scenarios for Automated Driving Systems – Vocabulary*. Standard. International Organization for Standardization, 2022. URL: `https://www.iso.org/standard/78950.html`.

[21] H. Elrofai, J.-P. Paardekooper, E. de Gelder, S. Kalisvaart, and O. Op den Camp. *Scenario-Based Safety Validation of Connected and Automated Driving*. Tech. rep. Netherlands Organization for Applied Scientific Research, TNO, 2018. URL: `http://publications.tno.nl/publication/34626550/AyT8Zc/TNO-2018-streetwise.pdf`.

[22] E. de Gelder and O. Op den Camp. "Tagging Real-World Scenarios for the Assessment of Autonomous Vehicles". In: *38th FISITA World Congress*. F2020-PIF-048. 2020. URL: `https://arxiv.org/abs/2012.01081`.

[23] ISO 34504. *Road Vehicles – Test Scenarios for Automated Driving Systems – Scenario categorization*. Standard. International Organization for Standardization, 2024. URL: `https://www.iso.org/standard/78953.html`.

[24] O. Op den Camp, E. de Gelder, J. Uittenbogaard, and J. Broos. "StreetWise: Scenario-Based Approach to Describe Real-World Traffic". In: *JSAE Annual Congress (Spring)*. 2022.

[25] 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: *IEEE 21st International Conference on Intelligent Transportations Systems (ITSC)*. 2018, pp. 2118–2125. DOI: `10.1109/ITSC.2018.8569552`.

[26] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein. "The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections". In: *IEEE Intelligent Vehicles Symposium (IV)*. 2020, pp. 1929–1934. DOI: `10.1109/IV47402.2020.9304839`.

[27] O. Op den Camp, J. van de Sluis, E. de Gelder, and I. Yalcinkaya. "Generation of Tests for Safety Assessment of V2V Platooning Trucks". In: *27th ITS World Congress*. 70. 2021. DOI: `10.5281/zenodo.5040362`.

[28] J. van de Sluis, O. Op den Camp, J. Broos, I. Yalcinkaya, and E. de Gelder. "Describing I2V Communication in Scenarios for Simulation-Based Safety Assessment of Truck Platooning". In: *Electronics* 10.19 (2021), pp. 1–18. DOI: 10.3390/electronics10192362.

[29] E. de Gelder, J. Manders, C. Grappiolo, J.-P. Paardekooper, O. Op den Camp, and B. De Schutter. "Real-World Scenario Mining for the Assessment of Automated Vehicles". In: *IEEE International Transportation Systems Conference (ITSC)*. 2020, pp. 1073–1080. DOI: 10.1109/ITSC45102.2020.9294652.

[30] C. Roesener, J. Sauerbier, A. Zlocki, F. Fahrenkrog, L. Wang, A. Várhelyi, E. de Gelder, J. Dufils, S. Breunig, P. Mejuto, F. Tango, and J. Lanati. "A Comprehensive Evaluation Approach for Highly Automated Driving". In: *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. 2017. URL: https://www-esv.nhtsa.dot.gov/Proceedings/25/25ESV-000259.pdf.

[31] 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, pp. 1–8. URL: https://www.pegasusprojekt.de/files/tmpl/pdf/12th%20ITS%20European%20Congress_Folien.pdf.

[32] M. Althoff, M. Koschi, and S. Manzinger. "CommonRoad: Composable Benchmarks for Motion Planning on Roads". In: *IEEE Intelligent Vehicles Symposium (IV)*. 2017, pp. 719–726. DOI: 10.1109/IVS.2017.7995802.

[33] J. Antona-Makoshi, N. Uchida, K. Yamazaki, K. Ozawa, E. Kitahara, and S. Taniguchi. "Development of a Safety Assurance Process for Autonomous Vehicles in Japan". In: *26th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. 2019, pp. 1–18. URL: https://www-esv.nhtsa.dot.gov/Proceedings/26/26ESV-000286.pdf.

[34] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick, and F. Diermeyer. "Survey on Scenario-Based Safety Assessment of Automated Vehicles". In: *IEEE Access* 8 (2020), pp. 87456–87477. DOI: 10.1109/ACCESS.2020.2993730.

[35] E. de Gelder, H. Elrofai, A. Khabbaz Saberi, O. Op den Camp, J.-P. Paardekooper, and B. De Schutter. "Risk Quantification for Automated Driving Systems in Real-World Driving Scenarios". In: *IEEE Access* 9 (2021), pp. 168953–168970. DOI: 10.1109/ACCESS.2021.3136585.

[36] ISO 21448:2022. *Road Vehicles – Safety of the intended functionality*. Standard. International Organization for Standardization, 2022. URL: https://www.iso.org/standard/77490.html.

[37] ISO 26262. *Road Vehicles – Functional Safety*. Standard. International Organization for Standardization, 2018. URL: https://www.iso.org/standard/68383.html.

[38] E. Petrucelli, J. D. States, and L. N. Hames. "The Abbreviated Injury Scale: Evolution, Usage and Future Adaptability". In: *Accident Analysis & Prevention* 13.1 (1981), pp. 29–35. DOI: 10.1016/0001-4575(81)90040-3.

[39] E. de Gelder and O. Op den Camp. "How Certain Are We That Our Automated Driving System Is Safe?" In: *Traffic Injury Prevention* 24.sup1 (2023), S131–S140. DOI: 10.1080/15389588.2023.2186733.

[40] E. de Gelder and O. Op den Camp. "A Quantitative Method to Determine What Collisions Are Reasonably Foreseeable and Preventable". In: *Safety Science* 167 (2023), p. 106233. DOI: 10.1016/j.ssci.2023.106233.

[41] M. Rosenblatt. "Remarks on Some Nonparametric Estimates of a Density Function". In: *The Annals of Mathematical Statistics* 27.3 (1956), pp. 832–837. DOI: 10.1214/aoms/1177728190.

[42] E. Parzen. "On Estimation of a Probability Density Function and Mode". In: *The Annals of Mathematical Statistics* 33.3 (1962), pp. 1065–1076. DOI: 10.1214/aoms/1177704472.

[43] J. Franke, W. K. Härdle, and C. M. Hafner. *Statistics of Financial Markets*. Vol. 2. Springer, 2004. DOI: `10.1007/978-3-030-13751-9`.

[44] K. D. Kusano, K. Beatty, S. Schnelle, F. Favaro, C. Crary, and T. Victor. "Collision Avoidance Testing of the Waymo Automated Driving System". In: *arXiv preprint arXiv:2212.08148* (2022). URL: `https://arxiv.org/abs/2212.08148`.

[45] E. de Gelder and J.-P. Paardekooper. "Assessment of Automated Driving Systems Using Real-Life Scenarios". In: *IEEE Intelligent Vehicles Symposium (IV)*. 2017, pp. 589–594. DOI: `10.1109/ivs.2017.7995782`.

[46] J. C. Hayward. *Near Miss Determination Through Use of a Scale of Danger*. Tech. rep. TTSC-7115. Pennsylvania State University, 1972. URL: `https://onlinepubs.trb.org/Onlinepubs/hrr/1972/384/384-004.pdf`.

[47] A. Arun, M. M. Haque, A. Bhaskar, S. Washington, and T. Sayed. "A Systematic Mapping Review of Surrogate Safety Assessment Using Traffic Conflict Techniques". In: *Accident Analysis & Prevention* 153.106016 (2021). DOI: `10.1016/j.aap.2021.106016`.

[48] C. Wang, Y. Xie, H. Huang, and P. Liu. "A Review of Surrogate Safety Measures and Their Applications in Connected and Automated Vehicles Safety Modeling". In: *Accident Analysis & Prevention* 157 (2021), p. 106157. DOI: `10.1016/j.aap.2021.106157`.

[49] E. de Gelder, K. Adjenughwure, J. Manders, R. Snijders, J.-P. Paardekooper, O. Op den Camp, A. Tejada, and B. De Schutter. "PRISMA: A Novel Approach for Deriving Probabilistic Surrogate Safety Measures for Risk Evaluation". In: *Accident Analysis & Prevention* 192 (2023), p. 107273. DOI: `10.1016/j.aap.2023.107273`.

[50] P. Koopman. *How Safe Is Safe Enough? Measuring and Predicting Autonomous Vehicle Safety*. Pittsburgh, PA: Carnegie Mellon University, 2022.

[51] 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?" In: *Traffic Injury Prevention* 20.S1 (2019), pp. 162–170. DOI: `10.1080/15389588.2019.1602727`.

[52] E. de Gelder, M. Buermann, and O. Op den Camp. "Coverage Metrics for a Scenario Database for the Scenario-Based Assessment of Automated Driving Systems". In: *In review* (2024).

[53] O. Op den Camp, E. de Gelder, A. Kalose, and S. Gaggar. "Scenario Categories for Autonomous Baggage Transport at Amsterdam Schiphol Airport". In: *JSAE Annual Congress (Spring)*. 2024.

[54] A. Tejada, M. J. E. Legius, A. R. Kalose, P. F. V. Oliveira, E. van Dam, and J. H. Hogema. "Towards a Practical Methodology for Defining Competent Driving". In: *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. 2023, pp. 569–575. DOI: `10.1109/ITSC57777.2023.10422458`.

[55] A. Tejada, J. H. Hogema, E. Van Dam, J. L. Souman, and E. Silvas. "Streetproof: Monitored Deployment for Safe & Socal Automated Driving". In: *JSAE Annual Congress (Spring)*. 2023.

Mobility & Built Environment

Automotive Campus 30
5708 JZ Helmond
www.tno.nl

innovation
for life