United Nations ECE/TRANS/WP.29/GRVA/2022/2

Economic and Social Council

Distr.: General
15 November 2021

Original: English

Economic Commission for Europe

Inland Transport Committee

World Forum for Harmonization of Vehicle Regulations

Working Party on Automated/Autonomous and Connected Vehicles

Twelfth session
Geneva, 24-28 January 2022
Item 4(b) of the provisional agenda
Automated/Autonomous and Connected Vehicles:
Deliverables of the Informal Working Group on
Validation Methods for Automated Driving

Proposal for a second iteration of the New Assessment/Test Method for Automated Driving - Master Document

Submitted by Informal Working Group on Validation Methods for Automated Driving*

The text reproduced below was prepared by the the Informal Working Group (IWG) on Validation Methods for Automated Driving (VMAD). It proposes a second iteration of the Master Document on the New Assessment/Test Method for Automated Driving. It is based on ECE/TRANS/WP.29/2021/61.

* In accordance with the programme of work of the Inland Transport Committee for 2022 as outlined in proposed programme budget for 2022 (A/76/6 (Sect.20), para 20.76), the World Forum will develop, harmonize and update UN Regulations in order to enhance the performance of vehicles. The present document is submitted in conformity with that mandate.

I. Background

1. During the 178th session of the United Nations Economic Commission for Europe (UNECE)'s World Forum for Harmonization of Vehicle Regulations (WP.29), the Framework document on automated/autonomous vehicles (ECE/TRANS/WP.29/2019/34/Rev.2) was adopted and the Terms of Reference (ToR) (ECE/TRANS/WP.29/1147, Annex VI) for the Informal Working Group (IWG) on Validation Methods for Automated Driving (VMAD) were developed.

2. The Framework document included the action item of a 'New Assessment/Test Method for automated driving' (NATM) for consideration during the 183rd session of WP.29 (March 2021).

3. Consistent with the Framework document, the ToR outlines that VMAD's mandate under the Working Party on Automated/Autonomous and Connected Vehicles (GRVA) is to develop assessments methods, including scenarios, to validate the safety of automated systems based on a multi-pillar approach including audit, simulation/virtual testing, test track, and real-world testing.

4. During the development of this work, the ToR outlines that VMAD should:

(a) Pursue this work in line with the following principles/elements described in the WP.29 Framework Document on Autonomous Vehicles:

(i) Object event detection and response (assessment): The automated/autonomous vehicles shall be able to detect and respond to object/events that may be reasonably expected in the Operational Design Domain (ODD); and

(ii) Validation for system safety: vehicle manufacturers should demonstrate a robust design and validation process based on a system-engineering approach with the goal of designing Automated Driving Systems (ADS) free of unreasonable risks and ensuring compliance with road traffic regulation and the principles listed in this document. Design validation methods should include a hazard analysis and safety risk assessment for ADS, for the object event detection and response (OEDR), but also for the overall vehicle design into which it is being integrated and when applicable, for the broader transportation ecosystem. Design and validation methods should demonstrate the behavioural competencies an automated/autonomous vehicle would be expected to perform during a normal operation, the performance during crash avoidance situations and the performance of fall-back strategies. Test approaches may include a combination of simulation, test track and on road testing.

(b) Take account of the developments of other subsidiary Working Parties (GRs) of WP.29 and their IWGs and work in full cooperation with them; and,

(c) Consider existing data, research and technical standards, e.g. SAE International, International Standard Organization (ISO), available during the development of its action items.

(d) The first iteration of the Master Document was adopted as a Reference Document for the validation of ADS during the 183rd session of WP.29 in March 2021. At the same meeting, a second iteration was requested to be prepared for review at the March 2022 session of WP.29, to solve the outstanding issues and to include the output from the IWG on Functional Requirements for Automated Vehicles (FRAV), as far as available.

(e) Since then, VMAD and its subgroups have continued their work on this document. This version represents the status of this work at the end of October 2021. This is the basis for discussion at the January 2022 session of GRVA. In parallel, the work will be continued on a limited number of issues.

II. Purpose and scope

5. In order for the international community to maximize the potential safety benefits of ADS, a safety validation framework that can be adopted by Contracting Parties of both the

1958 and the 1998 UN vehicle regulations agreements must be established. The NATM developed by VMAD aims to provide clear direction for validating the safety of an ADS in a manner that is repeatable, objective and evidence-based, while remaining technology neutral and flexible enough to foster ongoing innovation by the automotive industry.

6. This document consolidates the work accomplished by VMAD to date to develop the NATM. It provides a clear overview of the NATM and its constituent pillars. This document also serves to promote coordination between VMAD and the work of the GRVA Informal Working Group on FRAV. This coordination will ensure that the NATM also addresses the validation of compliance of an ADS to common safety requirements to be developed by FRAV.

7. Given the substantial technical work that is still needed to operationalize the NATM in practice, this version of the Master Document provides a high-level framework for the NATM, outlining:

(a) Scope and general overviews of the scenario catalogue and each of the pillars (simulation/virtual testing, test track, and real-world testing, audit/assessment and in-use monitoring); and,

(b) Overall process of the NATM (e.g., how the components of the NATM (i.e., the scenarios catalogue and pillars) operate together, producing an efficient, comprehensive, and cohesive process).

8. Going forward, this document will be further developed and regularly updated and informed by the outcomes of future VMAD sessions.

9. As VMAD continues to develop the elements of the NATM and FRAV continues to develop safety requirements for ADS, this document will be updated to incorporate this work. Detailed technical documents will be outlined in an index of supporting reference materials, located at the end of this document, as these are developed by VMAD.

10. Subject to direction from GRVA and WP.29, once the NATM has reached a state of maturity to inform evaluation criteria (based on performance requirements specified by the IWG on FRAV), it is anticipated that this document (and any supporting resources developed by VMAD) will be used to help inform validation process guidelines and/or regulations/requirements that align with the needs of both 1958 and 1998 Agreement parties (subject to approval by WP.29).

III. Definitions

11. The introduction of ADS and related technologies has resulted in a proliferation of new technical terms and concepts. To ensure consistency, a glossary of terms and definitions used in the NATM Master Document is attached at Annex 1. This glossary will be further developed and updated on an ongoing basis as the Master Document and any supporting technical documents are developed. Where applicable, VMAD will ensure these terms are consistent with those adopted by WP.29, GRVA, and other GRVA Informal Working Groups, including definitions agreed upon by FRAV.

IV. Applying a Multi-pillar Approach to the NATM

12. The purpose of the NATM is to provide a framework for assessing an ADS and its ability to demonstrate safe behaviour when operating in the real world.

13. Validating these capabilities is a highly complex task which cannot be done comprehensively nor effectively through one validation methodology alone. As a result, VMAD has proposed that the NATM adopt a multi-pillar approach for the validation of ADS, composed of a scenarios catalogue and five validation methodologies (pillars) each of which is explored in greater detail in subsequent sections of this document:

(a) A scenario catalogue

14. It consists in descriptions of real-world driving situations that may occur during a given trip, will be a tool used by the NATM-pillars to validate the safety of an ADS;

(b) Simulation/virtual Testing

15. It uses different types of simulation toolchains to assess the compliance of an ADS with the safety requirements on a wide range of virtual scenarios including some which would be extremely difficult if not impossible to test in real-world settings. The aspect of credibility of simulation/virtual testing is included in this topic;

(c) Track testing

16. It uses a closed-access testing ground with various scenario elements to test the capabilities and functioning of an ADS;

(d) Real world testing

17. It uses public roads to test and evaluate the performance of ADS related to its capacity to drive in real traffic conditions;

(e) Audit/assessment procedures

18. They establish how manufacturers will be required to demonstrate to safety authorities using documentation, their simulation, test-track, and/or real-world testing of the capabilities of an ADS. The audit will validate that hazards and risks relevant for the system have been identified and that a consistent safety-by-design concept has been put in place. The audit will also verify that robust processes/mechanisms/strategies (i.e., safety management system) that are in place to ensure the ADS meets the relevant safety requirements throughout the vehicle lifecycle. It shall also assess the complementarity between the different pillars of the assessment and the overall scenario coverage; and finally,

(f) In-service monitoring and reporting

19. It addresses the in-service safety of the ADS after its placing on the market. It relies on the collection of fleet data in the field to assess whether the ADS continues to be safe when operated on the road. This data collection can also be used to fuel the common scenario database with new scenarios from the field and to allow the whole ADS community to learn from major ADS accidents/incidents.

V. Scenarios Catalogue

A. Why should scenario-based testing be included in the NATM?

20. In order to maximize the potential safety of Automated Vehicles (AVs), a robust safety validation framework should be established. Such a framework should provide clear direction for assessing safety requirements of AVs in a repeatable, objective, evidence-based and technology neutral manner.

21. At this relatively early stage in the development of AVs, much of the existing literature that assesses the current state of AV development uses metrics such as miles/kilometers travelled in real-world test situations with the absence of a collision, a legal infraction, or a disengagement by the vehicle’s ADS.

22. Simple metrics such as kilometers travelled without a collision, legal infraction, or disengagement can be helpful for informing public dialogue about the general progress being made to develop AVs. Such measurements on their own however, do not provide sufficient evidence to the international regulatory community that an AV will be able to safely navigate the vast array of different situations a vehicle could reasonably be expected to encounter.

23. In fact, some observers have suggested that an AV would have to drive billions of miles in the real-world to experience an adequate number of situations without an incident to prove that it has a significantly better safety performance than a human driver (Kalra & Paddock, 2016). Safety validation through such testing would not be cost and time effective,

nor would it be feasible to replicate the testing later on. As validation of AV in various traffic situations is needed, therefore different traffic scenarios should be considered.

24. A scenario-based approach helps to systematically organize safety validation activities in an efficient, objective, repeatable, and scalable manner and is a critical part of the NATM for ensuring a holistic and dense coverage of traffic situations.

25. Scenario~~s~~-based validation consists of reproducing specific real-world situations that exercise and challenge the capabilities of an ADS-equipped vehicle to operate safely.

B. What is a traffic scenario?

26. A scenario is a description of one or more real-world driving situations that may occur during a given trip. SG1 will design scenarios for use under the NATM pillars. A scenario can involve many elements, such as roadway layout, types of road users, objects exhibiting static or diverse dynamic behaviours, and diverse environmental conditions (among other factors).

Note: A trip is a traversal of an entire travel pathway by a vehicle from the point of origin to a destination.

27. As previously noted, the use of scenarios can be applied to different testing methodologies, such as virtual/simulation, test track, and real-world testing. Together these methodologies provide a multifaceted testing architecture, with each methodology possessing specific strengths and weaknesses. Therefore, some scenarios may be more appropriately tested using certain test methodologies over others.

28. Going forward, VMAD will establish a catalogue of scenarios that should be considered to validate, using the NATM pillars, each safety requirement – given by FRAV - for an ADS. While it is ideal that scenarios comprehensively reflect the subject traffic situation on world-wide public roads, in reality each scenario selected to test the ADS, will need to reflect the particular conditions (e.g., road configurations, direction of traffic in a given lane) relevant to the ODD in which the ADS is designed to operate. Scenarios will need to be appropriate for the ADS feature being validated. For example, an ADS feature intended only for highway use would not be subject to a scenario involving turns at intersections. In addition, because an ADS will need to be responsive to actions by other road users that may make a crash unavoidable, scenarios should not be limited to scenarios that are deemed preventable by the ADS. This work will be accomplished in consultation with VMAD subgroups.

29. If scenarios not covered by scenario catalogue are identified and deemed necessary, they should be included in the scenario catalogue.

30. It is envisaged that a scenario catalogue will have tags for all scenarios corresponding to their relevant ODD attributes (using a standardised ODD taxonomy) and behaviour competencies.

C. Identifying scenarios

31. Scenario-based validation methods must include an adequate representation/coverage of relevant, critical, and complex scenarios to effectively validate an ADS. There are a number of approaches for identifying scenarios to validate the safety of an AV. For example, scenarios can be identified based on:

(a) Analysing human driver behaviour, including evaluating naturalistic driving data;

(b) Analysing collision data, such as law enforcement and insurance companies' crash databases;

(c) Analysing traffic patterns in specific ODD (e.g., by recording and analysing road user behaviour at intersections);

(d) Analysing data collected from ADS’ sensors (e.g., accelerometer, camera, RADAR, and global positioning systems);

(e) Using specially configured measurement vehicle, onsite monitoring equipment, drone measurements, etc. for collecting various traffic data (including other road users);

(f) Knowledge/experience acquired during ADS development;

(g) Synthetically generated scenarios from key parameter variations; and

(h) Engineered scenarios based on functional safety requirements and safety of intended functionality.

32. “Coverage” of scenario catalogue, which means considered cases out of total cases, is an important aspect in order to estimate the effectiveness of the scenario catalogue. Thus, it is important to ensure that the scenario catalogue includes scenarios sufficient to address conditions in a wide range of ODDs covered (e.g. urban, highway, rural roadway configurations; various weather elements, etc.) from the viewpoint of completeness of the scenario catalogue, and that the scenarios relevant to the ODD of an ADS include a precise broad reflection of the ODD related real-world driving situations that can reasonably be expected to occur in the ODD from the viewpoint of credibility of the scenario based validation applied to the ADS.

33. Unsafe behaviours of other road users (e.g. wrong way driver, sudden crossing) —if reasonably foreseeable—should be included in the scenario catalogue. This does not mean that all collision will be avoided because the requirement for ADS depends on the situation and required level of safety.

34. Country specific scenarios should be respected and need to be covered in the scenario catalogue in the long term.

35. Continued collection of real-world data is important for identifying unexpected scenarios – scenarios that may be uniquely challenging to that vehicle’s specific ADS.

36. Once a wide range of scenarios has been identified, specific requirements can be tested and validated by virtual, test track, and real-world test validation methods. In order to identify relevant scenarios, ODD and behaviour competencies can be used.

D. Classifying scenarios

37. The amount of information that is included in a scenario can be extensive. For example, the description of a scenario could contain information specifying a wide range of different actions, characteristics and elements, such as objects (e.g., vehicles, pedestrians), roadways, and environments, as well as pre-planned courses of action and major events that should occur during the scenario. Therefore, it is critical that a standardized and structured language for describing scenarios is established so that AV stakeholders understand the intention of a scenario, each other’s objectives, and the capabilities of an ADS. One tool for establishing uniform language for describing a scenario is a template, which ensures that the information to be included in the scenario is consistent and minimizes the possibility of confusion in its interpretation.

38. While not constituting a template, one approach that researchers have established is to describe scenarios, by different levels of abstraction/detail, according to three categories: functional, logical, and concrete scenarios.

(a) Functional Scenario: Scenarios with the highest level of abstraction, outlining the core concept of the scenario, such as a basic description of the ego vehicle’s actions; the interactions of the ego vehicle with other road users and objects; roadway geometry; and other elements that compose the scenario (e.g. environmental conditions etc.). This approach uses accessible language to describe the situation and its corresponding elements. For the scenario catalogue, such an accessible (i.e., natural and non-technical) language needs to be standardised to ensure common understanding between different AV stakeholders about the scenarios.

(b) Logical Scenario: Building off the elements identified within the functional scenario, developers generate a logical scenario by selecting value ranges or probability distributions for each element within a scenario (e.g., the possible width of a lane in meters). The logical scenario description covers all elements and technical requirements necessary to implement a system that solves these scenarios.

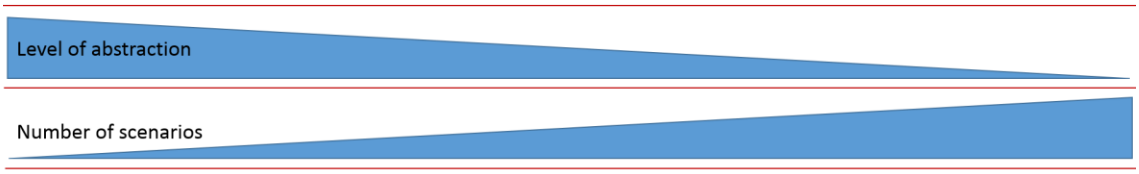
(c) Concrete Scenarios: Concrete scenarios are established by selecting specific values for each element. This step ensures that a specific test scenario is reproducible. In addition, for each logical scenario with continuous ranges, any number of concrete scenarios can be developed, helping to ensure a vehicle is exposed to a wide variety of situations.

(d) Refer to Figure 1 for examples of functional, logical and concrete scenarios.

Figure 1

Examples of a scenario during different stages of its development (Pegasus, 2018)

Functional scenarios	Logical scenarios	Concrete scenarios
Base road network: three-lane motorway in a curve, 100 km/h speed limit indicated by traffic signs	Base road network: Lane width [2.3..3.5] m Curve radius [0.6..0.9] km Position traffic sign [0..200] m	Base road network: Lane width [3.2] m Curve radius [0.7] km Position traffic sign [150] m
Stationary objects: -	Stationary objects: -	Stationary objects: -
Moveable objects: Ego vehicle, traffic jam; Interaction: Ego in maneuver „approaching" on the middle lane, traffic jam moves slowly	Moveable objects: End of traffic jam [10..200] m Traffic jam speed [0..30] km/h Ego distance [50..300] m Ego speed [80..130] km/h	Moveable objects: End of traffic jam 40 m Traffic jam speed 30 km/h Ego distance 200 m Ego speed 100 km/h
Environment: Summer, rain	Environment: Temperature [10..40] °C Droplet size [20..100] µm	Environment: Temperature 20 °C Droplet size 30 µm



E. Scenario elements

39. Traffic scenarios are derived by combining a number of relevant elements, taken from disjunct layers describing the scenario space systematically.

40. Functional scenarios[1] for divided highway application are described in Annex 2. This document should be regarded as "live document", meaning that the document should be updated based on the continuous discussion and the document is not the final version.

[1] After functional scenarios, it is natural to move down to a lower level of abstraction, and it is envisaged that some Logical scenarios and/or some possible ways of their description, as agreed in the continuous discussion, will also be included in Annex 2.

F. Scenario usage and testing related issues

41. Random sampling among scenarios relevant to a particular ADS and its ODD can be justified in order to avoid overfitting. Although more cases of random sampling are is preferable from a credibility perspective, the burden to manufacturers and authority (e.g. technical service) should be reasonably considered.

42. The scenario catalogue is not necessarily exhaustive and authorities may need to consider additional scenarios as necessary to support safety validation of an ADS feature.

VI. Simulation/Virtual Testing

A. Common terms

43. The following terms are used throughout this section.

(a) "*Abstraction*" is the process of selecting the essential aspects of a source system or referent system to be represented in a model or simulation, while ignoring those aspects not relevant. Any modelling abstraction carries with it the assumption that should not significantly affect the intended uses of the simulation tool.

(b) "*Closed Loop Testing"* means a virtual environment that does take the actions of the element-in-the loop into account. Simulated objects respond to the actions of the system (e.g. system interacting with a traffic model).

(c) "*Deterministic*" is a term describing a system whose time evolution can be predicted exactly and a given set of input stimuli will always produce the same output.

(d) "*Driver-In-the-Loop (DIL)*" is typically conducted in a driving simulator used for testing the human–automation interaction design. DIL has components for the driver to operate and communicate with the virtual environment.

(e) "*Hardware-In-the-Loop (HIL)*" involves the final hardware of a specific vehicle sub-system running the final software with input and output connected to a simulation environment to perform virtual testing. HIL testing provides a way of replicating sensors, actuators and mechanical components in a way that connects all the I/O of the Electronic Control Units (ECU) being tested, long before the final system is integrated.

(f) "*Model*" is a description or representation of a system, entity, phenomenon, or process.

(g) "*Model calibration*" is the process of adjusting numerical or modelling parameters in the model to improve agreement with a referent.

(h) "*Model Parameter*" are numerical values used to support characterizing a system functionality. A model parameter has a value that cannot be observed directly in the real world but that must be inferred from data collected in the real world (in the model calibration phase).

(i) "*Model-In-the-Loop (MIL)*" is an approach which allows quick algorithmic development without involving dedicated hardware. Usually, this level of development involves high-level abstraction software frameworks running on general-purpose computing systems.

(j) "*Open Loop Testing*" means a virtual environment that does not take the actions of the element-in-the loop into account (e.g. system interacting with a recorded traffic situation).

(k) "*Probabilistic*" is a term pertaining to non-deterministic events, the outcomes of which are described by a measure of likelihood.

(l) "*Proving Ground or test-track*" is a physical testing facility closed to the traffic where the performance of an ADS can be investigated on the real vehicle. Traffic agents can be introduced via sensor stimulation or via dummy devices positioned on the track.

(m) "*Sensor Stimulation*" is a technique whereby artificially generated signals are provided to the element under testing in order to trigger it to produce the result required for verification of the real world, training, maintenance, or for research and development.

(n) "*Simulation*" is the imitation of the operation of a real-world process or system over time.

(o) "*Simulation model*" is a model whose input variables vary over time.

(p) "*Simulation toolchain*" is a combination of simulation tools that are used to support the validation of an ADS.

(q) "*Software-In-the-Loop (SIL)*" is where the implementation of the developed model will be evaluated on general-purpose computing systems. This step can use a complete software implementation very close to the final one. SIL testing is used to describe a test methodology, where executable code such as algorithms (or even an entire controller strategy), is tested within a modelling environment that can help prove or test the software.

(r) "*Stochastic*" means a process involving or containing a random variable or variables. Pertaining to chance or probability.

(s) "*Validation of the simulation model*" is the process of determining the degree to which a simulation model is an accurate representation of the real world from the perspective of the intended uses of the tool.

(t) "*Vehicle-In-the-Loop (VIL)*" is a fusion environment of a real testing vehicle in the real-world and a virtual environment. It can reflect vehicle dynamics at the same level as the real-world and it can be operated on a vehicle test bed or on a test track.

(u) "*Verification of the simulation model*" is the process of determining the extent to which a simulation model or a virtual testing tool is compliant with its requirements and specifications as detailed in its conceptual models, mathematical models, or other constructs.

(v) "*Virtual testing*" is the process of testing a system using one or more simulation models.

B. Introduction

(a) Simulation provides powerful tools to assess the performance of an ADS under diverse and complex conditions which are prohibitive for conventional physical testing. Powered by simulation models, virtual testing plays a vital role in ensuring comprehensive assessment of an ADS. The major role virtual testing will play in the development and validation of ADS justifies its inclusion as a principal pillar of the NATM.

(b) While robust virtual test methods are available and widely used, the task of the NATM is to verify the possibility to produce reliable evidence of ADS safety performance in the physical world. Therefore, this section of the Master Document explains virtual testing tools and methods and complementarity of this pillar with the other testing methods.

C. Virtual testing and simulation in ADS development and validation

44. Virtual testing can be used in different phases of the ADS development and validation. Virtual testing can be used to explore in a comprehensive and cost-effective way an ADS (or of part of it) in a wide range of traffic scenarios across different ODDs and for a variety of additional purposes. Relying on simulation, virtual testing is particularly indicated to test the ADS under safety critical scenarios that would be difficult and/or unsafe to reproduce on test tracks or public roads.

45. Virtual testing includes replacing one or more physical elements characterized in a scenario-based test by a simulation model. The goal of such virtualization is to resemble, to a sufficient extent, the original physical elements. For automotive applications, virtual testing can be used to reproduce the driving environment and the objects operating therein that interact with either the entire system (e.g. a full vehicle with tires and ADS functions), a subsystem (e.g. an actuator or a hardware controller), or a component (e.g. a sensor).

46. Through this approach, an assessor can get confidence about the ADS based on the virtual tests and validation that was performed by the developer in an agile, controllable, predictable, repeatable, and efficient manner.

47. The simulation toolchain used for virtual testing may result in the combination of different approaches. In particular, tests can be performed:

(a) entirely inside a computer (referred to as Model or Software in the Loop testing, MIL/SIL), with the model of the elements involved (e.g. a simple representation of the control logic of an ADS) interacting in a simulated environment; and/or

(b) With a sensor, a subsystem, or the whole vehicle interacting with a virtual environment (Hardware or Vehicle in the Loop testing, HIL/VIL). For VIL testing, the vehicle can either be in:

(i) A laboratory where the vehicle would be standing still or moving on a chassis dynamometer or powertrain test bed and be connected to the environment model by wire or by direct stimulation of its sensors; or

(ii) A proving ground where the vehicle would be connected to an environment model and would interact with virtual objects by physically moving on the test-track.

(c) With a subsystem interacting with a real driver (DIL testing).

48. The interaction between the system under the test and the environment can either be an open- or closed-loop.

(a) Open-loop virtual tests (also referred to as software or hardware reprocessing, shadow mode, etc.) could be done through a variety of methods, such as the ADS interacting with virtual situations collected from the real world. In this case, virtual objects' actions are data-driven only and the information is not self-corrected based on feedback from the output. Because the open-loop controller may vary due to external disturbances without the ADS and/or the assessor being aware, the applicability of open-loop tests in the ADS validation may be limited.

(b) Closed-loop virtual tests includes a feedback loop that continuously sends information from the closed-loop controller to the ADS. Within these test systems, the behaviour of the digital objects could react in different ways depending on the action of the system under test.

49. Selecting an open- or closed-loop test could depend on factors such as the objectives of the virtual testing activity and the status of development of the system under test. For ADS validation it is expected that mainly closed-loop virtual testing will be considered. Examples of virtual testing toolchains are reported in Annex III - Appendix 1 to the present document.

D. Strengths and weaknesses of the pillar

50. The flexibility of the pillar makes it a standard test method in vehicle design and validation in general. For ADS', given the impossibility to test the vehicle behaviour in real life in all possible situations and for any change in its driving logic, virtual testing becomes an indispensable tool to verify the capability of the automated system to deal with a wide variety of possible traffic scenarios. In addition, virtual testing can be extremely beneficial to replace real world and proving ground testing concerning safety-critical traffic scenarios.

51. Furthermore, virtual tests used for ADS validation can achieve different objectives, depending on the overall validation strategy and the accuracy of the underlying simulation models:

(a) Provide qualitative confidence in the safety of the full system.

(b) Contribute directly to statistical confidence in the safety of the full system (caveats apply).

(c) Provide qualitative or statistical confidence in the performance of specific subsystems or components.

(d) Discover challenging scenarios to test in the real world (e.g. real-world tests and track tests described in chapter 7 and 8 of this document).

52. In contrast to all its potential benefits, a limitation of this approach is in its intrinsic limited fidelity. As models can only provide a coarse representation of the reality, the suitability of a model to satisfactorily replace the real world for validating the safety of ADSs has to be carefully assessed. Therefore, the validation of the simulation models used in virtual testing is essential to determine the transferability and reliability of the results compared to real-world performance.

53. An approach for assessing the accuracy of a virtual testing toolchain is to compare the ADS' performance within a virtual test with its performance in the real world when executing the same scenario. Given the high number of scenarios that virtual testing can perform compared to track test, the validation will likely need to be performed on a smaller but still sufficiently representative subsection of the relevant scenarios in order to substantiate any extrapolation beyond the scenarios used for the validation.

54. Table 1 summarizes the main strengths and weaknesses of the virtual testing as part of the demonstration of the safety level by the manufacturer.

Table 1.
Strengths and Weaknesses of the Virtual Testing Pillar

Strengths	*Weaknesses*
• Controllability – Virtual testing affords an unmatched ability to control many aspects of a test. • Agility – Virtual tests allows for system changes to be revaluated immediately. • Efficiency – In MIL and SIL, virtual tests can be accelerated faster than real-time so that many tests can be run concurrently in a relatively short amount of time. • Cost effectiveness at test execution – In spite of the investments required to develop, validate and maintain a virtual testing toolchain, the running costs connected to its use are considerably lower than those required by physical testing. • Wide scenario coverage – Compared to other testing methods, virtual testing allows a wider exploration of safety-critical scenarios. By properly combining the experiments parameters it can for example reduce the space of the known unknowns and to the extent possible that of the unknown unknowns (including the effect of system failures). • Data gathering and analysis - Virtual testing offers a convenient and error-free platform for data gathering and analysis of the ADS performance. Once Qualified, that data can serve as a significant contribution for assessing the risk from the ADS. • Repeatability and replicability – Simulation affords the re-execution of the same virtual test without deviations due to stochastic phenomena. Faults in the functioning of the ADS can thus be identically replicated at any moment.	• Lower environmental fidelity/reliability – It is difficult, and likely impossible for models to completely reproduce the environment, responses, as well as the behaviour of the vehicle, other road users etc. in the real world. Also the validation process cannot prove the validity of the simulation across all possible scenarios. • Risk of over-reliance. Without proper consideration of models' intrinsic limitations, a risk exists to put too much emphasis on virtual testing results without sufficient proof of their validity by physical testing. • Expensive software life-cycle. The availability of a simulation model to execute virtual testing requires covering certain aspects of the software life-cycle which can be costly and time-consuming

E. Maturity of the pillar

55. Virtual testing is a constantly evolving test method. While it is in many ways mature and used commonly for design and development processes, the real-world reliability and validity of each embodiment of the tool still needs to be determined. Although virtual testing can be used both in the ADS development and validation process by the manufacturer and in the ADS certification process by the authority, it can be considered a mature option only as a tool used for the vehicle manufacturer. Further work is needed to define the requirements for using virtual testing in the certification process. Topics to be addressed are for instance the validation requirements. It needs to be proven, indeed, that the simulation toolchain used for virtual testing is an accurate representation of the real system for the purpose of the experimentation. For this reason, Annex III and related Appendixes to this document

describes the credibility assessment framework developed to prove the validity of a virtual testing toolchain to validate the safety requirements of ADSs.

56. Another area of research for the future application of virtual testing in ADS certification is the possibility for authorities to host and maintain a validated and standardized simulation environment where manufacturer can "plug" the system to be validated (either in the form of a physical system or of a model/software) to show its compliance to the safety requirements defined by the legislation.

57. Since this is currently the subject of research and standardization activities, in the short term virtual testing can only be allowed by simulation toolchains developed and maintained by vehicle manufacturers or ADS developers. Since their design depends on the validation and verification strategies implemented by the manufacturer, they should not be subject to regulation or standardization but rather explained and documented by the ADS/vehicle manufacturer and the basis for its verification and validation reviewed during the certification process. For this reason, it is envisioned that documentation and data provided by the manufacturer should be harmonized.

F. Interaction between Virtual Testing and the other Test Methods

58. Virtual testing will have strong relationships with all the pillars of the NATM. In particular:

(a) Virtual testing expands the scope of physical testing to account for the diversity of traffic. The strength of virtual testing lies in its capacity to cost-effectively assess performance across ranges of variables and arrays of scenarios. Virtual testing enables results of limited physical tests to be supplemented by verifiable data covering variations on the physical test scenario. Virtual testing enables coverage of safety-critical scenarios at their logical abstraction levels, confirming that an ADS will perform as intended across the parameter ranges. These advantages reduce the burden on physical tests (offsetting their weaknesses) to improve the efficiency of the overall assessment across the pillars. Virtual testing can also be effectively used to identify and cover edge cases and other low-probability scenarios to increase confidence on their performances.

(b) Virtual testing can play an important role in the development of performance requirements and traffic scenarios. Virtual testing also enables assessment of ADS performance boundaries, enabling precision of limits between collision avoidance and crash mitigation. Through methods of randomization and compositions, virtual testing enables the developer or the assessor to challenge the ADS with unexpected, unplanned scenarios, and thus increases the confidence in the performance of the ADS when challenged with low probability events.

(c) Virtual testing will be a key element in the audit assessment. Results of virtual testing carried out both during vehicle development and in the verification and validation phase will represent an important element to be subject to audit. Manufacturers will need to provide evidence and documentation about how the virtual testing is carried out and how the underlying simulation toolchain has been validated.

(d) Real-world tests can aid in the generation of realistic simulation models and in establishing their accuracy:

(i) Real-world data for vehicle and component model validation: vehicle data and data measured via vehicle sensors are important sources for quantifying and arguing model accuracy (e.g. vehicle dynamics or sensor models).

(ii) Real-world data for traffic modelling: the generation of novel scenarios requires realistic road user behaviour for the simulation environment to remain meaningful and representative.

(e) Virtual testing can play an important role in responding to concerns identified through in-use monitoring of ADS performance. Virtual testing provides speed and flexibility in analysing real-world events to verify ADS performance against such events and, if necessary, support modifications to improve performance. Scenario descriptions can be shared and integrated rapidly into virtual testing regimes worldwide. The various types of virtual testing, including HIL methods that come close to matching physical testing, ensure robust and rapid responses.

G. Use of the pillar to assess ADS safety requirements

59. Virtual testing using a validated simulation toolchain can be used to assess the ADS' compliance with the safety requirements. Considering the categories of safety requirements currently being considered, virtual testing seems particularly relevant for assessing requirements related to:

(a) ADS should drive safely, and ADS should manage safety critical situations. These are the requirements where virtual testing can play the most prominent role. MIL/SIL, HIL and VIL virtual testing can all be used to assess these requirements at different stages of vehicle verification and validation.

(b) ADS should interact safely with the user. DIL virtual testing can be helpful to support the assessment of this category of safety requirement by analysing the interaction between the driver and the ADS in a safe and controlled environment.

(c) ADS should safely manage failure modes and ADS should ensure a safe operational state. The use of virtual testing in these two categories is also very promising but would probably require further research work. SIL virtual testing could include simulated failures and maintenance requests. HIL and VIL virtual testing could be used to assess how the system would react to the occurrence of a real malfunctioning induced to the real system.

VII. Track Testing

A. Purpose

60. Track testing occurs on a closed-access testing ground that uses real obstacles and obstacle surrogates (e.g. vehicle crash targets, etc.) to assess the safety requirements of an ADS (e.g., human factors, safety system). This testing approach allows for the physical vehicles to be tested through a limited set of realistic scenarios (based on the test track's geometries, dimensions, size, and the ODD) to evaluate either sub-systems or the fully assembled system. These external inputs and conditions can be controlled or measured during a test.

61. In addition to this test method providing a higher level of environmental fidelity than simulation, it provides an opportunity to test the vehicle with less danger than what is likely posed within real-world tests. However, operating on test tracks can be resource-intensive, therefore testing on a test track will be based on selected known critical scenarios. Refer to Table 2. below for more information regarding the respective strengths and weaknesses of this testing methodology.

62. Track testing may be more suitable for assessing the ADS capabilities in a discrete number of nominal scenarios and critical scenarios. The same tests could be used to verify the performance of the vehicles regarding human factors or fallback in these scenarios.

Table 2
Strengths and Weaknesses of the Track Test Pillar

Strengths	*Weaknesses*
• Controllability – Track testing allows for control over many of the test elements, including certain aspects of the ODD. • Fidelity – Track testing involves functional, physical ADS-equipped vehicles and lifelike obstacles and environmental conditions.	• Significant time –Track testing can take a significant amount of time to set up and execute. • Costly – Track testing may require a substantial number of personnel and specialized test equipment (e.g., obstacle objects, measurement devices, safety driver).

Strengths	*Weaknesses*
• Reproducibility– Track testing scenarios can be replicated in different locations by different testing entities • Repeatability – Track testing allows for multiple iterations of tests to be run in the same fashion, with the same inputs and initial conditions. • Efficiency – Compared to real-world testing, closed-course testing can accelerate exposure to known rare events or safety critical scenarios by setting them up as explicitly designed test scenarios. Road testing by contrast could be an inefficient way to test less co manifesting by chance. • Track testing can be used to validate the quality of the simulation toolchain by comparing an ADS' performance within a simulation test with its performance on a test track when executing the same scenario.	• Limited variability – Track testing facility infrastructure and conditions may be difficult to modify to account for a wide variety of test elements (e.g., ODD conditions). They are restricted to their geometries, dimensions, size and ODD limitations such as weather conditions, time of day, number and type of other traffic agents. • Safety risks – Track testing with physical vehicles and real obstacles presents a potentially uncertain and hazardous environment for the test participants (e.g., safety driver and experiment observers). • Representativeness even with its increased fidelity. Whilst things like pedestrians can be included, these won't typically be real people due to safety reasons and the clutter or real-world environments cannot be replicated.

B. Why include this pillar in the NATM?

63. As per paragraph 7.3 as well as the strengths and limitation table, there are a number of reasons for including track testing in the NATM. For instance, track testing can be used to assess the performance of ADS in nominal and critical scenarios. Track testing can also provide a higher level of environmental fidelity than simulation. Unlike real-world testing, track testing can accelerate exposure to known rare events or safety critical scenarios.

C. Maturity of the pillar

64. Although track testing is a mature process which is used to assess safety requirements for some existing technologies, testing of ADS vehicles is fairly new and may need to be further refined. For instance, it may be difficult to develop specific ODD elements, such as rain, fog, and snow to reliably test how an ADS interacts with these environmental elements.

D. How the pillar interacts with other pillars?

65. The information generated during the track-test could also be used to validate the virtual tests by comparing an ADS' performance within a virtual test with its performance on a test track when executing the same scenario. For instance, track testing can be used to validate the quality/reliability of the virtual toolchain by comparing an ADS' performance within a virtual test with its performance on a test track when executing the same scenarios.

VIII. Real-world Testing

A. Purpose

66. Real-world testing uses public roads to test the capabilities and compliance with safety requirements (e.g., human factors, safety system) of a vehicle with an automated driving system (ADS) in real-world traffic.

67. This testing method can expose the ADS to a wide variety of real-world conditions related to an ODD. There are various approaches to real-world testing. For example, tests can be done within a specific ODD (e.g., highway driving) with a safety driver who is monitoring/ensuring the ADS is functioning safely.

68. Real world testing could be used to assess aspects of the ADS performance related to its capability to drive in real traffic conditions, e.g. smooth driving, capability to deal with dense traffic, interaction with other road users, maintaining flow of traffic, being considerate and courteous to other vehicles.

69. Real world testing could also be used to assess part of the ADS performance at some ODD boundaries (nominal and complex scenarios), i.e. is the system triggering transition demands to the driver when it is supposed to (e.g. end of the ODD, weather conditions). The same testing could be used to confirm the performances related to human factors under these conditions.

70. Finally, on road testing could be used to detect issues that may not be well captured by track tests and simulation, such as perception quality limitation (e.g. due to light conditions, rain, etc.).

71. Although it may not be possible to encounter all traffic scenarios during a real-world test, the likelihood of covering specific complex scenarios could be increased by selecting a specific type of ODD (e.g., highway) and examining when and where specific elements (e.g., high- or low-density traffic) typically occur.

72. Specific infractions identified during real-world testing may be later reviewed/assessed by evaluating the information/data using virtual, track and real-world testing. In addition, real-world testing data can be collected to identify and record new traffic scenarios and improve the environmental validity of track and virtual testing methodologies in the future.

73. Refer to Table 3 below for more information regarding the respective strengths and weaknesses of this testing method.

Table 3
Strengths and Weaknesses of the Real World Test Pillar

Strengths	*Weaknesses*
• High environmental validity – allows for validation of the vehicle in its intended ODD(s) and the diverse conditions these may present. • Can be used to test scenarios elements, such as weather and infrastructure (e.g., bridges, tunnels), that are unavailable through track testing • Real-world testing may be used to validate the simulation and track-testing by comparing an ADS' performance within a simulation and track test with its performance on in a real-world environment when executing the same scenario. • Can be used to assess aspects of the ADS performance related to its interaction with other road users, e.g. maintaining flow of traffic, being considerate and courteous to other vehicles. • Model, single software, and toolchain validation	• Restricted controllability – Public-road scenarios afford a limited amount of control over ODD conditions. • Restricted reproducibility – Public-road scenarios are difficult to replicate exactly in different locations. • Restricted repeatability – Public-road scenarios are difficult to repeat exactly over multiple iterations. • Limited scalability – Public-road scenarios may not scale up sufficiently. • Costly but not as costly as track testing – Requires a number of resources and is time-consuming • Potential impact on traffic and safety authorities • New competencies may need to be developed by authorities • Safety risks: on-road testing could subject test personnel and the public to significant risks of unsafe behavior.

B. Why include this pillar in the NATM?

74. Real world testing provides an opportunity to validate the safety of the ADS within its true operating environment, as set out in greater detail in paragraphs 8.3, 8.4, and 8.5.

C. Maturity of the pillar

75. Real-world testing is regularly conducted to assess the performance of human drivers. However, testing of ADS performance may pose some new challenges for this test methodology. Experiences could be drawn from other motor vehicle-related real-world testing schemes, such as real driving emissions (RDE) testing and market surveillance.

D. How the pillar interacts with other pillars

76. Real-world testing may be used to validate if portions of a virtual and/or track-testing environment were modelled properly by comparing an ADS' performance within a simulation and track test with its performance on in a real-world environment when executing the same test scenario.

77. It can also be used to identify new traffic scenarios for track and virtual testing, allowing for the identification of edge cases and other unknown hazard vulnerabilities that could challenge the ADS. The information gathered from real world testing can also be used in the hazard and risk analysis and design of the ADS systems.

IX. Audit

A. Purpose

78. The purpose of the audit pillar is to assess/demonstrate that the:

(a) Manufacturer has the right processes to ensure operational and functional safety during the vehicle lifecycle, and

(b) Vehicle design is safe by design and this design is sufficiently validated before market introduction. The validation should be confirmed by in use monitoring.

79. The manufacturer will be required to demonstrate that:

(a) Robust processes are in place to ensure safety throughout the vehicle lifecycle (development phase, production, but also operation on the road and decommissioning). It shall include taking the right measures to monitor the vehicle in the field and to take the right action when necessary;

(b) Hazard and risks relevant for the system have been identified and a consistent safety-by-design concept has been put in place to mitigate these risks; and

(c) The risk assessment and the safety- by-design concept have been validated by the manufacturer through testing showing before the vehicle is placed on the market that the vehicle meets the safety requirements and in particular is free of unreasonable safety risks to the broader transport ecosystem in particular the driver, passengers and other road users.

B. Why should the Audit pillar be included in the NATM?

80. On the basis of the evidence provided by the manufacturer and the targeted tests, authorities will be able to audit and check whether the processes, the risk assessment, the design and the validation of the manufacturer are robust enough with regard functional and operational safety.

81. As such, these elements: risk assessments, safety- by-design concept, and validation tests can be used to demonstrate the ADS' overall safety in a far more robust manner than a limited number of physical/virtual tests on their own.

C. Strengths and weaknesses of the audit pillar

82. Risk analysis, safety-by-design concepts as well verification/validation test methods are standard development methods used in the automotive industry for years to ensure functional safety of electronic system (fail safe). It is expected that similar methods will be followed by manufacturers to minimize unsafe and unknown scenarios for ADSs in a systematic manner (operational safety beyond failures).

83. Regarding the safety assessment, the tools under this pillar will provide a more robust demonstration on the ADS safety (coverage) than a few test runs. The manufacturer's safety case will be reinforced if it is assessed by an independent auditor and confirmed by targeted physical or virtual tests. Test runs will in particular be needed to demonstrate that the vehicle exhibits minimum performances for standard manoeuvres (e.g. normal lane keeping, lane change), key critical scenarios (e.g. emergency braking) and in traffic conditions (e.g. smooth integration in the traffic). It remains to be decided at this stage whether these tests shall be standardized across manufacturers for some defined situations or shall be tailored to the results of the risk assessment/design of the ADS or both.

D. Maturity of the Audit Pillar

84. The audit pillar is already used for a long time in UN Regulations, e.g. UN Regulation No. 79 on steering, UN Regulation No. 13 on braking and UN Regulation No. 152 on Advanced Emergency Braking System (AEBS). VMAD also proposed an updated audit pillar for the regulation on automated lane keeping systems in line with the concepts described above. The new UN Regulation No. 155 on cyber security and cyber security management system also uses audits.

85. Risk analysis, verification/validation and safety management systems is a well-established practice in the industry (ISO 26262 on Industry functional safety). There is on-going work to cover new risks raised by ADAS/ADS such as operational safety (ISO/PAS 21448, BSI PAS 1880:2020, BSI PAS 1881:2020 (https://www.bsigroup.com/en-GB/CAV/pas-1881/) and UL 4600). Similar standardization work exists for cybersecurity (ISO/SAE 21434).

86. It should be noted that the publication of voluntary safety assessment reports documented by the manufacturer are also currently encouraged by some Contracting Parties (e.g. in the United States of America and Canada).

87. The relevant issues for auditing the safety of the design concept are elaborated in Annex 4. The requirements for auditing the management system of the manufacturer are elaborated in Annex 5.

X. In-service monitoring and reporting

A. Purpose

88. The in-service monitoring and reporting pillar addresses the in-service safety of automated vehicles after market introduction. In practice, the application of the other pillars of the NATM will assess whether the ADS is reasonably safe for market introduction whereas the in-service monitoring and reporting will gather additional evidence from the field operation to demonstrate that that the ADS continues to be safe when operated on the road. This pillar addresses the dynamic nature of road transportation to ensure attention to and continuous improvement of road safety through the use of ADS.

89. The pillar consists in the collection of relevant data during AVs operation.

90. The three main purposes of in-service monitoring and reporting is to use retrospective analysis of data from manufacturers and other relevant sources to:

(a) Demonstrate that the initial safety assessment (residual risk) in the audit phase before the market introduction is confirmed in the field overtime (“safety confirmation”).

(b) To fuel the common scenario database with important new scenarios that may happen with automated vehicles in the field (“scenario generation”) and

(c) To derive safety recommendations for the whole community by sharing learnings derived from key safety accidents/ incidents to allow the whole community to learn from operational feedback, fostering continuous improvement of both technology and legislation (“safety recommendations”).

91. The obligation to have “real-time monitoring” (self-checks/ on board diagnostics) of the performance of ADs subsystems by the manufacturer is not part of this pillar but is part of the safety requirements. However, some reporting mechanisms on the performance of ADS subsystems overtime could be part of the first subpoint above, and contribute to the predictive monitoring of safety performance degradation.

92. The processes put in place by the manufacturer to manage safety during in use (e.g. to manage changes in the traffic rules and in the infrastructure) fall outside this pillar and are assessed with the audit pillar. This pillar focuses on the type of data to be monitored and reported.

B. Why should this pillar be included in the NATM?

93. Whatever a safety evaluation is done before market introduction, the actual level of safety will only be confirmed once a sufficient number of vehicles is in the field and once they are subjected to a sufficient range of traffic and environmental conditions. It is therefore essential that a feedback loop (fleet monitoring) is in place to confirm the safety by design concept and the validation carried out by the manufacturer before market introduction. The operational experience feedback from in-use monitoring will allow ex-post evaluation of regulatory requirements and validation methods, providing indications on gaps and needs for review.

94. New scenarios and new risks might be introduced by AVs on the market. Therefore, the In-Use Monitoring pillar could be used to generate new scenarios in the common scenario database to cover these new safety risks.

95. Finally, in the early phase of market introduction of ADS, it is essential that the whole community learns from crashes involving AVs in order to quickly react and lead to safety developments and subsequent prevention of that crash scenario for all other ADS.

C. Strengths and weaknesses of the pillar

96. Data from the field will be the most realistic way to assess the safety performance of an ADS over a wide range of real driving traffic and environmental conditions.

97. Data from the field are also instrumental to ensure that the scenario database is updated with the latest scenarios, in particular those deriving from the increasing use of ADS.

98. Regarding safety recommendations, learning from in-service data is a central component to the safety potential of ADSs. Lessons learned from a crash involving ADSs could lead to safety developments and subsequent prevention of that crash scenario in other ADS. Feedback from the operational experience is recognized as best practice for safety management in the automotive sector as well as in other transport sectors (e.g. already in place in aviation, railway and maritime sectors). Field operation data can also provide evidence of the positive impact of ADs on road safety.

99. Limitations might derive from the quantity of data to be handled (too much data is as problematic as too little data), availability of tools for automatic scenario generation, and identification of responsibility handlers. Therefore, the outcome shall be a proportionate, efficient and uniform system.

100. Methods to verify the reliability of collected data should be developed. The data collected should be comparable amongst manufacturers. It will create challenges on which data and how these data are collected and reported (definition of suitable reporting criteria). Timewise, another challenge is the development of the in-service safety monitoring framework in a timely manner in order to serve AVs market deployment. Data privacy should also be taken into account. A standardized format for communication of information will be needed to allow processing by authorities in a standard manner and that any outcomes are easily shareable or open for analysis by other authorities. Different type of data may be needed depending on the purpose of the data collection.

101. Processes for reporting the operational feedback from AVs should be developed for the automotive sector taking into account the higher number of monitored vehicles and events to be recorded.

D. Maturity of the pillar

102. In-service monitoring and reporting is standard practice in the industry to develop and improve driver assistant systems (see ISO 26262 and SOTIF[1]). It was introduced as part of the audit of the new UN Regulation No. 157 on Automated Lane Keeping System (ALKS). Starting from this requirement, additional elements should be developed in order to establish a more comprehensive approach for information sharing. In-service monitoring and reporting have already been implemented for many years as part of EU emissions regulations. In-use reporting was established in 1966 in the USA and formalized into a comprehensive safety reporting system under US law in 2000.

103. The development of new traffic scenarios on the basis of traffic data has already started from the manufacturers' side, through post-processing of recorded data elements and images (tools for complete automatic scenario generation are not available yet).

104. Operational Accident/incident analysis is a well-established practice in some vehicle safety regimes, e.g. through the analysis of data from event data recorders (EDRs) from conventional vehicles which are collecting relevant information in certain crash situations [2]. No standard data elements are currently defined for ADS crash/near-missed investigation: entities engaging in testing or deployment are encouraged to voluntarily collect data associated with crashes [3]. Because it is first time the concept of in-service-monitoring is introduced into the automotive safety sector and vehicles are usually used by normal citizens (different from air or rail sector), feasibility, such as how to collect data, which data (e.g. including or not if the ADs caused the circumstances that resulted in near-miss-crash), would be important view points and it should be well discussed.

105. Mechanisms for operational feedback to improve common knowledge are already in place for decades in other transport sectors (see European Co-ordination centre for Accident and Incident Reporting Systems (ECCAIRS) portal, http://eccairsportal.jrc.ec.europa.eu/). Existing systems for reporting of safety concerns in the automotive sector have also been developed over decades of experience. A first step would be to investigate the suitability of such tools for ADs too. However, the main effort would still remain in defining common reporting criteria and developing a common repository. According to mechanisms already in place in other sectors (e.g. see Figure 2), in-service data recorded related to safety-relevant events (i.e. accidents, near-miss events, abnormal functioning etc.) are processed by manufacturers/operators and then an accident report (what happened) by manufacturers/operators is delivered to the National authority. National authorities are then responsible to perform the accident analysis (why did it happened), derive safety recommendations (how could this be avoided), and evaluate the possible impact on existing legislation. National information is then recorded into:

[1] Safety of the intended function: ISO/PAS 21448,

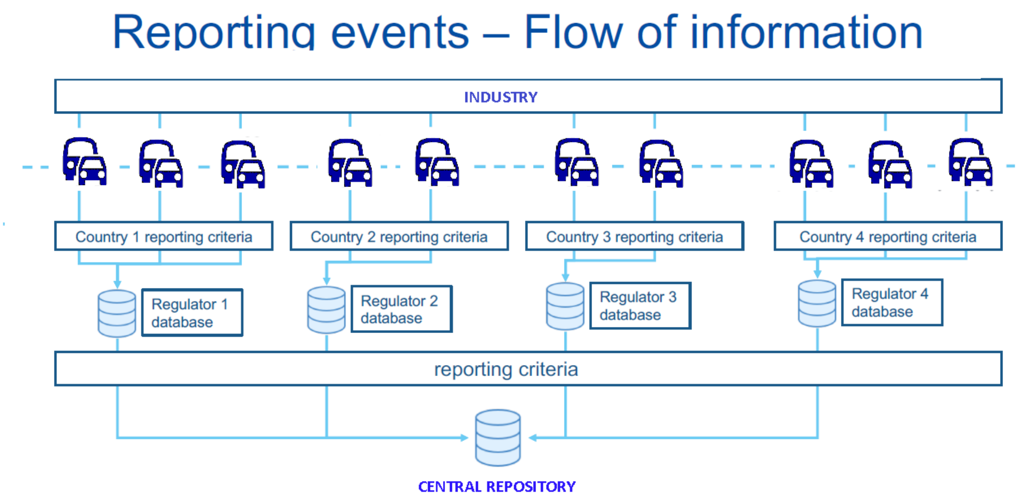
[2] See 49 CFR Part 563, Event Data Recorders. www.gpo.gov/fdsys/pkg/CFR-2016-title49-vol6/xml/CFR-2016-title49-vol6-part563.xml

[3] NHTSA Voluntary Guidance, https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

(a) Central Repository of Occurrences; and,

(b) a Central Repository of Safety Recommendations. Access to data recorded into the Central Repository is subject to strict rules and mainly limited to competent Authorities. Safety recommendations are shared internationally according to the guiding principle that transport safety is of global concern and its improvement should not be limited by geographical or organizational borders. Privacy is ensured at all levels. Another option could be for the measured data to be directly communicated to the authorities, who will then be in charge of collection, storage and post-processing of the information.

Figure 2
Reporting events – flow of information



106. There is some link with the Informal Working Group that is already working on data recording requirements for conventional and automated vehicles (IWG on DSSAD/EDR IWG[4]) in particular regarding accident analysis. However in-service monitoring as part of the ADS assessment method has a different purpose (i.e. confirming the safety assessment, fueling the scenario database, detailed analysis of accidents/incidents) than EDR/DSSAD (accident reconstruction and liability in case of road traffic offense).

107. The detailed requirements with regard to in-use monitoring are listed in Annex 6.

XI. NATM Pillars/Element Interaction

108. The goal of the NATM is to assess the safety of an ADS in a manner that is as repeatable, objective and evidence-based as possible, whilst remaining technology neutral and flexible enough to foster ongoing innovation in the automotive industry.

109. The overall purpose of the NATM is to assess, based on the safety requirements, whether the ADS is able to cope with the occurrences that may be encountered in the real world. In particular by looking at scenarios linked to road user behaviour/environmental conditions in traffic scenarios but also scenarios linked to driver behaviour (e.g. HMI) and ADS failures.

110. As previously noted, the multi-pillar approach recognizes that the safety of an ADS cannot be reliably assessed/validated using only one of the pillars. Each of the aforementioned testing methodologies possesses its own strengths and limitations, such as differing levels of environmental control, environmental fidelity, and scalability.

111. A single assessment or test method may not be enough to assess whether the ADS is able to cope with all occurrences that may be encountered in the real world.

[4] DSSAD/EDR https://wiki.unece.org/pages/viewpage.action?pageId=87621709

112. For instance, while real-world testing provides a high degree of environmental fidelity, a scenario-based testing methodology using only real-world testing could be costly, time-consuming, difficult to replicate, and pose safety risks. Consequently, track testing may be more appropriate methods to run higher risk scenarios without exposing other road users to potential harm. Further, test scenarios can also be more easily replicated in a closed track environment compared to the real-world. That said, test track scenarios can be potentially difficult to develop and implement, especially if there are numerous or complex scenarios, involving a variety of scenario elements.

113. Simulation/virtual testing, by contrast, can be more scalable, cost-effective, safe, and efficient compared to track or real-world testing, allowing a test administrator to safely and easily create a wide range of scenarios including complex scenarios where a diverse range of elements are examined. However, simulations may have lower fidelity than the other methodologies. Simulation software may also vary in quality and tests could be difficult to replicate across different simulation platforms.

114. In-service monitoring and reporting can confirm the pre-deployment safety assessment and fill the gaps between safety validation through virtual/physical testing and real-life conditions. Evaluation of in-service performance will also serve to update the scenario database with new scenarios deriving from increasing deployment of driving automation. Finally, the feedback from operational experience can support ex-post evaluation of regulatory requirements.

115. In addition to the respective strengths and weakness of each test pillar, the nature of the safety requirements being assessed will also inform what pillars are used.

116. For instance: the most appropriate method to assess an ADS's overall system safety prior to market introduction may be the audit pillar, using a systematic approach to perform a risk analysis. The audit could include information such as safety by design confirmed validation outputs as well as analysis of data collected in the field by the manufacturer.

117. Virtual testing may be more suitable when there is a need to vary test parameters and a large number of tests need to be carried out to support efficient scenario coverage (e.g., for path planning and control, or assessing perception quality with pre-recorded sensor data).

118. Track tests may be best suited for when the performance of an ADS can be assessed in a discrete number of physical tests, and the assessment would benefit from higher levels of fidelity (e.g., for HMI or fall back, critical traffic situations).

119. Real-world testing may be more suitable where the scenario may not be precisely represented virtually or on a test track (e.g., interactions with other road-users and perception quality may be assessed through real world evaluation).

120. In-service monitoring and reporting of field data represent the best way to confirm the safety performance of an ADS in the field after market introduction over a wide variety of real driving traffic and environmental conditions.

121. Given these considerations, the sequence and composition of test pillars used to assess each safety requirement may vary. While some testing might follow a logical sequence from simulation to track and then to real world testing, there may be deviations depending on the specific safety requirement being tested.

122. It is therefore necessary for the NATM pillars to be used together to produce an efficient, comprehensive, and cohesive process, considering their strengths and limitations. The methods should complement one another, avoiding excessive overlaps or redundancy to ensure an efficient and effective validation strategy.

123. As previously noted, the NATM pillars not only include the three aforementioned test methods but also an aggregated analysis (e.g., an audit/assessment /in service monitoring/reporting pillar). Whereas the test methods will assess the safety of the ADS, the audit/assessment pillar will serve to assess the safety of the ADS as well as the robustness of organizational processes/strategies. Elements of the audit are:

(a) Assessment of the robustness of safety management system,

(b) Assessment of the (identified) hazards and risks for the system,

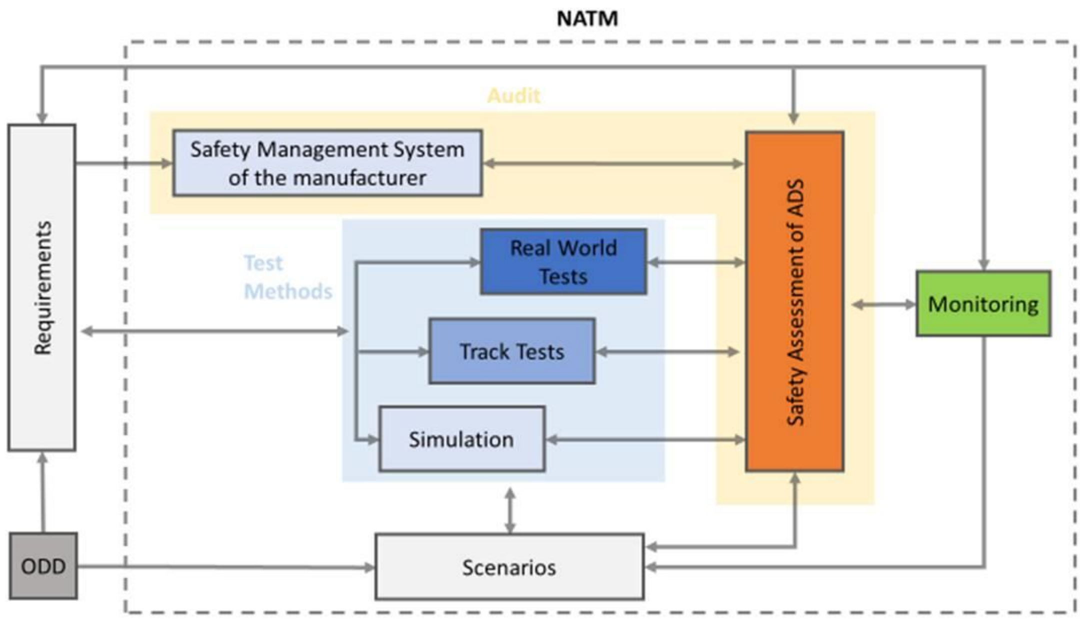
(c) Assessment of the Verification strategy (e.g. verification plan and matrix) that describe the validation strategy and the integrated use of the pillars to achieve the adequate coverage.

(d) Assessment of the level of compliance with requirements achieved through an integrated use of all pillars, including consistency between the outcomes of one pillar as input for another pillar (forward and backward) and adequate use of scenarios. This level of compliance concerns both new vehicles as vehicles in use.

(e) The audit/assessment phase also incorporate results from the Simulation, Track test and Real-World tests carried out by the manufacturer.

124. Figure 3 provides a diagram that outlines how the pillars, scenarios, and safety requirements (developed by FRAV) will interact. Further examination of each of these elements follows in the subsequent sections of this document.

Figure 3
Relationship between VMAD Pillars, Scenarios and FRAV Safety Requirements



XII. Integration

125. This document contains the description of a generic validation method. Likewise, the IWG on FRAV is developing generic requirements for the product to be validated. There is a clear relation between these two developments: functional requirements may affect the detailed validation requirement and vice versa, validation requirement may result in input for functional requirements. So far, FRAV has delivered a list of 28 high level functional requirements (FRAV-05). In detailing the functional requirements, the possible impact for validation methods will have to be checked. This process is managed by including representatives of both informal working groups in each other meetings.

126. As the safety requirements and technical aspects of each of the pillars are further developed, each of these sections will be updated to include additional detail. To provide further context, this section will also include examples of how the NATM pillars can be applied to certain functional capabilities of an ADS (e.g., highway driving) based on the established safety requirements.

Annex I

Glossary of terms and definitions (draft only)

"*Complex Scenarios*" means a traffic scenario containing one or more situations that involve a large number of other road users, unlikely road infrastructure, or abnormal geographic/environmental conditions.

"*Critical Scenarios*" means a traffic scenario containing a situation in which the ADS need to perform an emergency manoeuvre in order to avoid/mitigate a potential collision, or react to a system failure.

"*Edge Case*" is a rare situation that still requires specific design attention for it to be dealt with by the AV in a reasonable and safe way. The quantification of "rare" is relative, and generally refers to situations or conditions that will occur often enough in a full-scale deployed fleet to be a problem but may have not been captured in the design process. Edge cases can be individual unexpected events, such as the appearance of a unique road sign or an unexpected animal type on a highway

"*Nominal Scenarios*" means a traffic scenario containing situations that reflect regular and non-critical driving manoeuvres.

"*Test case specification*" are the detailed specifications of what must be done by the tester to prepare for the test.

"*Test methods*" is a structured approach to consistently derive knowledge about the ADS by means for executing tests, e.g. virtual testing in simulated environments, physical, structured testing in controlled test facility environments, and real world on-road conditions.

"*Traffic scenario*" (or scenario for short) is a sequence or combination of situations used to assess the safety requirements for an ADS. Scenarios include a DDT or sequence of DDTs. Scenarios can also involve a wide range of elements, such as some or all portions of the DDT; different roadway layouts; different types of road users and objects exhibiting static or diverse dynamic behaviours; and, diverse environmental conditions (among many other factors).

Annex II

Functional scenarios for divided highway application

Contents

I. Introduction

This text is a synthesis of different recent elaborations of traffic scenarios, with the designated purpose to create a functional scenario list for ADS in motorway use-case. It is envisaged that some Logical scenarios and/or some possible ways of their description, as agreed in the continuous discussion, will also be included in this text. ODD range: highways with up to 130 km/h and lane changes allowed.

II. Inputs to this proposal

The following input has been considered:

(a) UN Regulation No. 157 (ALKS);

(b) The Netherlands (Netherlands Organisation for Applied Scientific Research (TNO)) Scenario Categories V1.7;

(c) Securing America's Future Energy (SAFE) (Fortellix) scenario library;

(d) Japan Crash scenarios;

(e) China functional scenario proposal (China Automotive Technology and Research Center (CATARC));

(f) JRC own elaborations;

(g) Germany, Initiative for the Global Harmonization of Accident Data (IGLAD) catalogue of conflict types.

Inputs provided by Japan, the Netherlands, SAFE, China were submitted for consideration and discussion during the VMAD Subgroup 1 (VMAD SG1) meeting held on 10 December 2020, the proposal from Germany was submitted on 16 December 2020.

III. Building blocks of functional scenarios

Functional scenarios can cover several aspects (e.g. road geometry at different abstraction levels, ego-vehicle behaviour, moving/stable objects).

Additional aspects that are not covered by functional scenarios (e.g. speeds, accelerations, positions, environmental conditions, failures, miscommunications, road geometries at more detailed levels) should be covered by logical scenario.

Since classification of aspects to functional and logical scenarios (i.e. "which aspects should be considered in functional scenarios" and "which aspects should be considered in logical scenarios") has not yet been discussed and agreed, the classification in this document is initial version and will be updated through discussion.

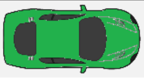
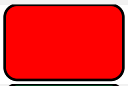
IV. Coverage

Since collisions always occur with other vehicles/objects (assuming that they can operate properly when there are no other vehicles/objects), and 24 functional scenarios in the figure described in "2. Interaction with other vehicles" can cover all interactions between other vehicles/objects and ego vehicle, the scenarios can cover collision with other vehicles/objects appropriately.

As described in paragraph 3., factors not covered in the proposed functional scenarios (e.g. initial speed of ego vehicle, size, initial position, initial speed, acceleration of other vehicles/objects), perception factor (e.g. weather, brightness, blind spot, false positive factor, blinkers of other vehicles) and vehicle stability factors (e.g. curve, slope, road surface μ, wind, etc.) can be described with parameters in logical scenarios.

Functional scenarios should be added anytime if SG1 and IWG on VMAD discussed and agreed.

V. Symbols used in this document

ICON	DESCRIPTION
	Ego vehicle
	Lead vehicle
	Other vehicles part of the scenario
	Impassable object on intended path
	Passable object on intended path

VI. A list of possible scenarios for level 3 highway chauffeur ADS

Input matrix from VMAD Subgroup 1 (SG1) participants:

Scenario family		*Sub-scenario*	*Japan crash scenarios*	*The Netherlands (TNO)*	*SAFE scenario library*	*China functional scenarios*	*Conflict Type*
A. Nominal driving	1. Perform lane keeping	a. Driving straight		X	X	X	X
		b. Manoeuvring a bend		X	X	X	X
B. Interaction with other vehicles/objects	1. Perform lane change	a. Ego vehicle performing lane change with vehicle behind	X	X			X
		b. Merging at highway entry	X		X	X	X
		c. Merging at lane end	X		X		X
		d. Merging into an occupied lane	X	X			X
	2..Critical (Emergency) braking scenarios during lane keeping	a. Impassable object on intended path	X	X	X		X
		b. Passable object on intended path	X	X		X	X
		c. Lead vehicle (LV) braking	X	X	X	X	X
		d. Approaching slower/stopped LV	X	X	X	X	X
		e. Cut-in in front of the ego vehicle	X	X	X	X	X
		f. Cut-out in front of the ego vehicle	X	X	X	X	X
		g. Detect and respond to swerving vehicles	X	X	X		X
C. Detect and response to traffic rules and road furniture		a. Speed limit sign			X	X	
		b. Signal lights				X	X
		c. Drive through tunnel				X	

Scenario family	*Sub-scenario*	*Japan crash scenarios*	*The Netherlands (TNO)*	*SAFE scenario library*	*China functional scenarios*	*Conflict Type*
	d. Toll				X	
	e. Conventional obstacles				X	X
D. Country specific road geometry	a. Interceptor			X		
E. Unusual situation	a. Wrong way driver (oncoming)			X		X

Notes to the inputs from VMAD SG1 members:

- China (CATARC): This is a list cut from a general catalogue describing different ODDs, like "General road", "City expressway" or "The highway" and their test items, like "speed limit sign", "lane line", "toll station", etc. The functional scenarios proposed below in this document are much more generic than the ones proposed by China, so they form a subset of this list. For example China proposal: "toll station" on the road or "conventional obstacles" can be in line with "impassable object on intended path" from this scenario list.
- The Netherlands (TNO): a very thorough scenario catalogue containing much more scenarios than needed for the highway use case. Terminology and descriptions worked out fully. Scenarios can be created using a combination of tags from the different layers.
- Japan: crash scenarios, scenarios only containing interaction with other vehicles. They describe different road geometries and possible other vehicle positions around ego. All other parameters considered as features (acceleration – deceleration, lane change – lane keeping, etc.).
- SAFE: a list of scenarios sometimes with very concrete examples, sometimes more generic approach. There is a different scenario for passing by slowly moving vehicles in the adjacent lane and a different one for passing by standing vehicles, but handles Lead Vehicle (LV) following as one scenario.
- Conflict Type: a list of "conflict types" used i.a. by accident investigators to sort scenarios, leading to accidents on road to different groups. These conflict types can be sorted into conflicts with or without influence of other road user. Uses different symbols than other documents for the description of a scenario or situation (mainly different kinds of arrows). Separates left and right hand traffic. Contains 251 scenario types, structured in seven larger types of conflicts, like: "longitudinal traffic" or "pedestrian crossing the road".

Note: "emphasized scenario parameters" and "tested parameters" in this paragraph are some examples of parameters. Other parameters may be essential for the validation testing.

A. Nominal driving (Perform lane keeping)

1. Nominal driving (Perform lane keeping)

Note: lane keeping is addressed in current UN Regulation No. 157 (ALKS) up to 60 km/h. As a functional scenario, lane keeping can be sorted into two groups depending on road geometry. It can also be sorted into more groups depends on the lane that the vehicle is in: center, side, middle, etc.

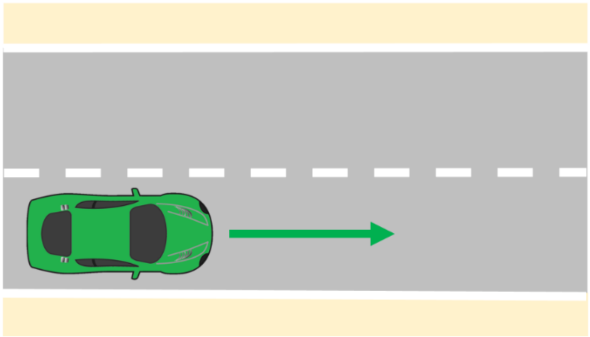
(a) Driving straight

(a) Without LV

(b) With LV

(c) With other vehicles in adjacent lanes (moving or stopped)

Figure 1
Schematic representation of driving straight



General description:

The ego vehicle is driving on a straight road. The aim of this scenario is to test the lane keeping ability of the vehicle under normal or demanding conditions and parameters [1,2,4].

Emphasized scenario parameters: ego speed demand (road rules), lane width, LV speed profile (if present), layout and speed profile of other vehicles (if present).

Tested parameters: deviation from lane centre (nominal value and distribution), deviation from desired speed, obeying to speed changes, temporal modifications, distance between ego and LV (if present), reaction to other vehicles…

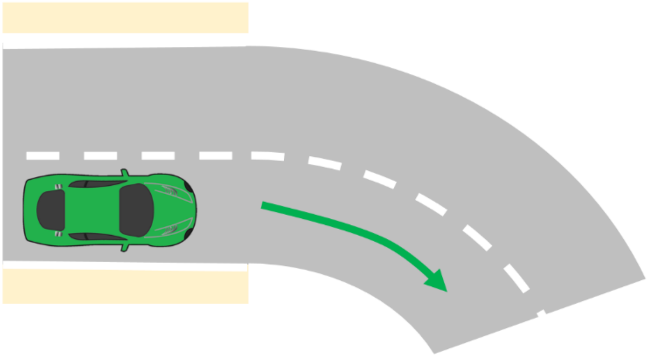
(b) Manoeuvring a bend (right curve and left curve)

(a) Without LV

(b) With LV

(c) With other vehicles in adjacent lanes (moving or stopped)

Figure 2
Schematic representation of manoeuvring a bend



General description:

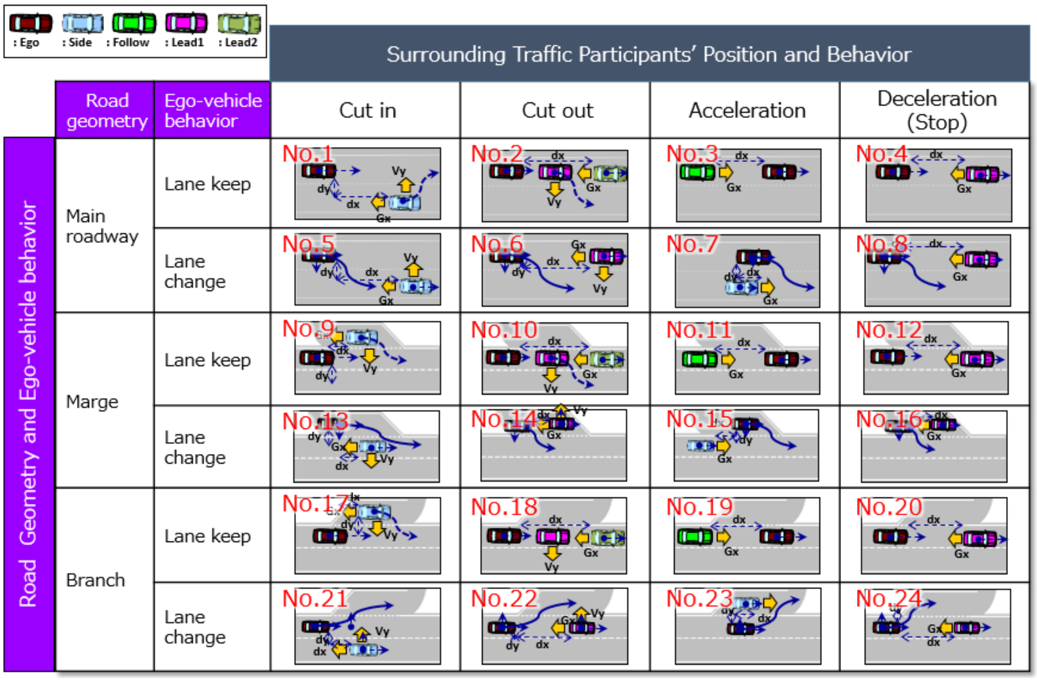
The ego vehicle is driving on a curved road. The aim of this scenario is to test if the vehicle is able to handle the road curvatures specified as part of the ODD [1], [2], [4].

Emphasized scenario parameters: ego speed demand (road rules), lane width, LV speed profile (if present), layout and speed profile of other vehicles (if present).

Tested parameters: deviation from lane centre (nominal value and distribution), deviation from desired speed, obeying to speed changes, temporal modifications, distance between ego and LV (if present), distance to other vehicles, etc.

B. Interaction with other vehicles/objects

The 24 scenarios below can cover the interaction with other vehicles driving in the same direction on the same or adjacent lanes.



In the 12 scenarios in which the ego vehicle performs lane change, the vehicle closest to the ego vehicle may not be necessarily in the same lane or an adjacent lane to the ego vehicle. It may be 2 lanes over from the ego vehicle, and even in such cases, the vehicle has to be detected by the ego vehicle because they can interact with one another if both change lanes. To describe these cases in the 12 scenarios properly, some parameters should be included such as "number of lanes", "lane of ego vehicle" and "relative position between ego and other vehicle". The examples of "main road case" are shown below. Other cases in "merged road" and "branched road" should be considered too.

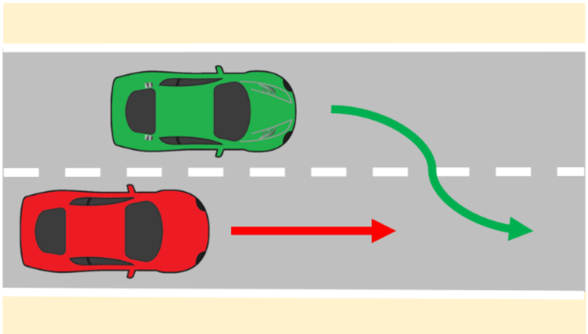
1. Perform lane change

Note: LC scenarios are complicated by the fact that the ADS cannot be forced to make a lane change. In addition, lane change functionality and principles shall be defined in a later stage (like technical requirements, definitions, activation criteria, indication of lane change, etc.).

Lane changes can be grouped based on the number of vehicles in the target lane. If there is enough space to execute the lane change, there is no need to cooperate with other vehicles. If the target lane is occupied by other traffic participants, than the ego vehicle has to adapt to the other participants and perform merging.

(a) Ego vehicle performing lane change with vehicle behind

Figure 3
Schematic representation of a lane change



General description

In an adjacent lane, another vehicle is driving in the same direction as the ego vehicle. The intention of the ego vehicle is, to perform a lane change to the lane in which the other player is driving [1], [3].

Emphasized scenario parameters: time of lane change, ego speed demand (road rules), lane width, LV speed profile (if present), layout and speed profile of other vehicles (if present).

Tested parameters: deviation from lane centres (nominal value, overshoot), time of lane change (lateral velocity of ego), distance between ego and LV (if present), distance to other vehicles, etc.

(b) Merging at highway entry

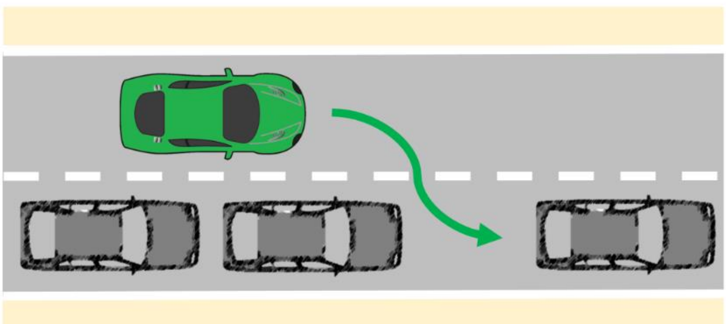
No description provided.

(c) Merging at lane end

No description provided.

(d) Merging into an occupied lane

Figure 4
Schematic representation of merging



General description

Other vehicles occupy the lane adjacent to the ego lane. The ego vehicle intends to perform a lane change to the lane in which the other vehicles are driving [1], [2], [3], [4]. According to road geometry, speed, number and layout of other vehicles, the difficulty of the scenario changes.

Emphasized scenario parameters: road layout, layout and speed profile of other vehicles (if present), ego speed (road rules), lane width etc.

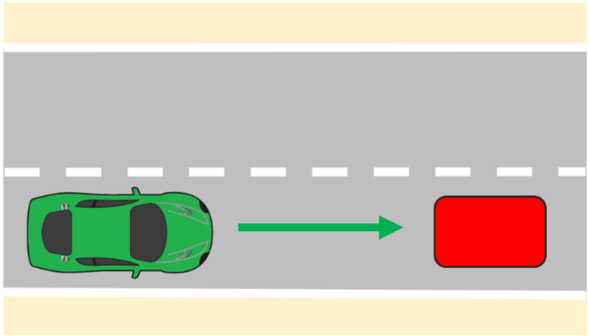
Tested parameters: distance to other vehicles, time of lane change (lateral velocity of ego),…

2. Critical (Emergency) braking scenarios during lane keeping

Note: In this family of scenarios a couple critical functional scenarios are present. It can be noticed in the input matrix of SG1 as well, these are scenarios that nearly every participant highlighted in the input documents.

(a) Impassable object on intended path (Including other cars and Vulnerable Road Users (VRUs))

Figure 5
Schematic representation of an impassable object



General description:

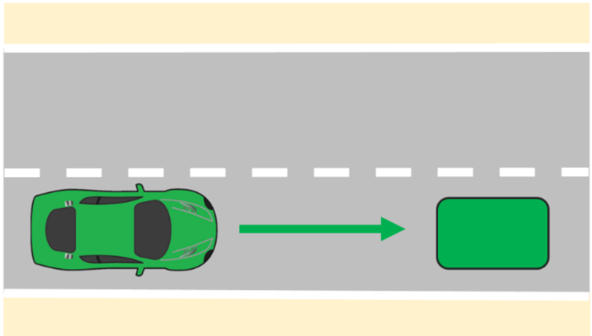
The ego vehicle is driving on a road with an impassable object in the ego lane. The objective of the ego vehicle is to continue driving straight. The ego vehicle needs to react [1], [2]. Depending on the velocity of the ego vehicle, the severity of the scenario is changing.

Emphasized scenario parameters: road layout (visibility of the object on the path), layout and speed profile of other vehicles (if present), ego velocity.

Tested parameters: reaction of ego (lane change/braking), distance to object, lateral velocity of ego (if changing lane), etc.

(b) Passable object on intended path (e.g. manhole lid)

Figure 6
Schematic representation of a passable object



General description:

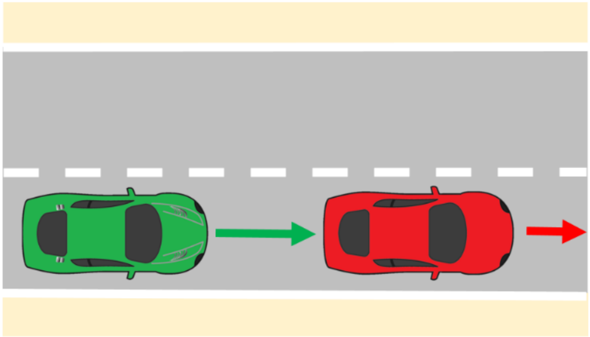
The ego vehicle is driving on a road with a passable object in the ego lane, e.g., a manhole lid or a small branch. The objective of the ego vehicle is to continue driving straight. The ego vehicle needs to react [1,4]. Depending on the velocity of the ego vehicle, the difficulty of the scenario is changing.

Emphasized scenario parameters: road layout (visibility of the object on the path), layout and speed profile of other vehicles (if present), ego velocity.

Tested parameters: reaction of ego (false positive, lane change/braking), distance to object, lateral velocity of ego (if changing lane), etc.

(c) Lead vehicle braking

Figure 7
Schematic representation of lead vehicle braking



General description:

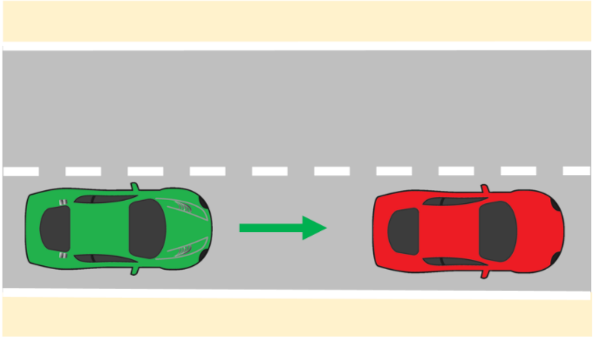
The ego vehicle is following a LV. The LV brakes, the ego vehicle has to adapt its speed in order to stay at a safe distance from the lead vehicle [1], [2], [3], [4].

Emphasized scenario parameters: ego velocity (road rules), LV speed profile (deceleration), layout and speed profile of other vehicles (if present).

Tested parameters: distance between ego and LV, reaction to other vehicles in adjacent lanes, etc.

(d) Approaching slower/stopped LV

Figure 8
Schematic representation of approaching stopped lead vehicle



General description:

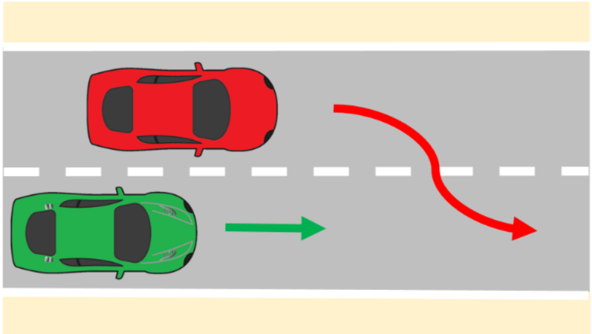
LV is driving in front of the ego vehicle at a slower speed. The ego vehicle might brake or perform a lane change to avoid a collision [1], [2], [3], [4]. According to the speed of the LV and ego vehicle, the severity of this scenario can be assessed.

Emphasized scenario parameters: ego velocity (road rules), LV speed profile (deceleration), layout and speed profile of other vehicles (if present).

Tested parameters: distance between ego and LV, reaction to other vehicles in adjacent lanes, etc.

(e) Cut-in in front of the ego vehicle

Figure 9
Schematic representation of cut-in



General description:

Another vehicle is driving in the same direction as the ego vehicle in an adjacent lane. The other vehicle makes a lane change, such that is becomes the LV from the ego vehicle's perspective [1-4]. Depending on the distance and lateral velocity of the LV, the severity of the cut-in manoeuvre changes.

Emphasized scenario parameters: LV lateral speed, distance to LV, ego velocity, lane width, layout and speed profile of other vehicles (if present).

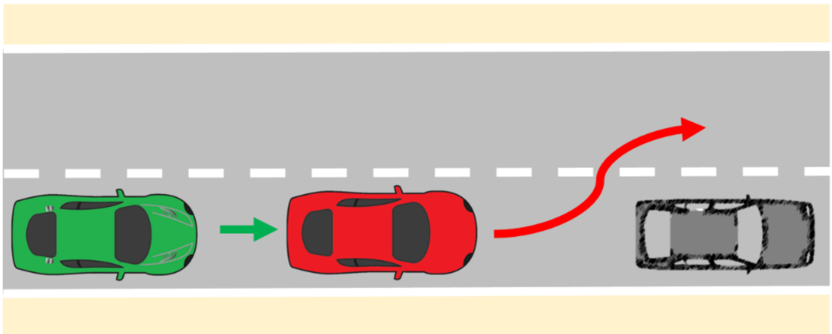
Tested parameters: distance between ego and LV, distance to other vehicles, etc.

(f) Cut-out in front of the ego vehicle

(a) Cut-out to highway exit

(b) Cut-out on highway lanes

Figure 10
Schematic representation of cut-out



General description:

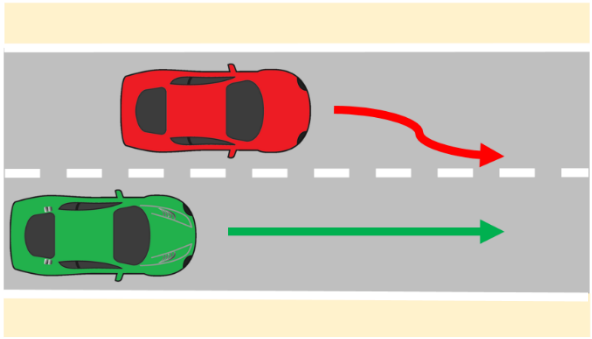
LV is driving in the same direction as the ego vehicle in front of the ego vehicle. The LV makes a lane change, such that it will no longer be the ego vehicle's LV [1], [2], [3], [4]. In order to test the behaviour of the ego vehicle, an obstacle is present in the ego lane in front of the ego vehicle. Depending on the velocity of the ego vehicle and the lateral velocity of the LV, the difficulty of this scenario changes.

Emphasized scenario parameters: LV lateral speed, distance to LV, ego velocity, lane width, layout and speed profile of other vehicles (if present).

Tested parameters: distance between ego and obstacle, distance to other vehicles, etc.

(g) Detect and respond to swerving vehicles

Figure 11
Schematic representation of a swerving vehicle



General description:

Another vehicle is driving in the same direction as the ego vehicle in an adjacent lane. The other vehicle swerves towards the ego vehicle's lane [1], [2], [3].

Emphasized scenario parameters: lateral speed of other vehicle, ego velocity, lane width, layout and speed profile of other vehicles (if present).

Tested parameters: distance between ego and swerving vehicle, distance to other vehicles…

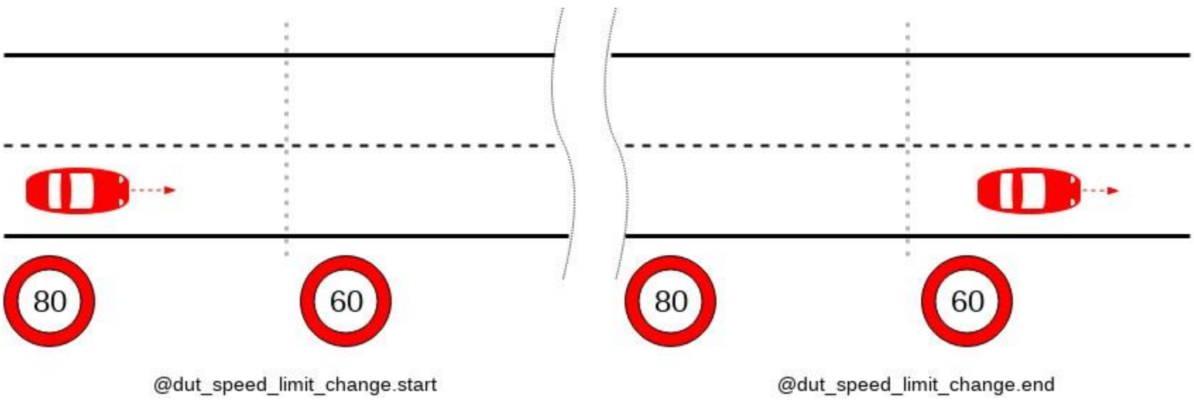
C. Detect and response to traffic rules and road furniture

Note: These scenarios are implicitly present in nearly every document, but sometimes are treated as special road furniture. It should be considered that these scenarios can be occurred simultaneously with other scenarios. It should be also noted that traffic rules are different from different countries or regions.

(a) Speed limit sign

This scenario challenges the ego vehicle to respond appropriately to speed limit changes by decelerating when entering a lower speed zone and accelerating when entering a higher speed zone. In the example shown below, the speed limit decreases from 80 km/h to 60 km/h.

Figure 12
Ego vehicle speed limit change scenario



Environmental requirements: A road that has at least one change in the speed limit.

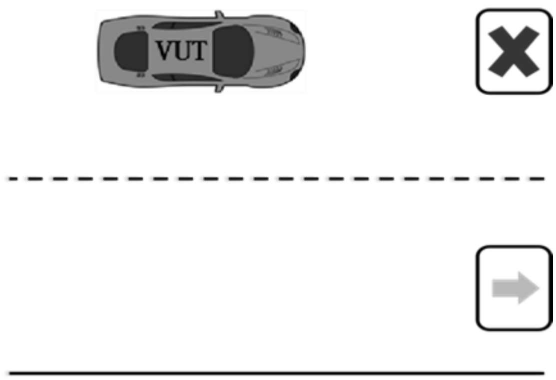
ego vehicle behaviour: The ego vehicle drives on the road, presumably adapting its speed to the changing limitations.

The ego vehicle merges at lane end.

(b) Signal lights

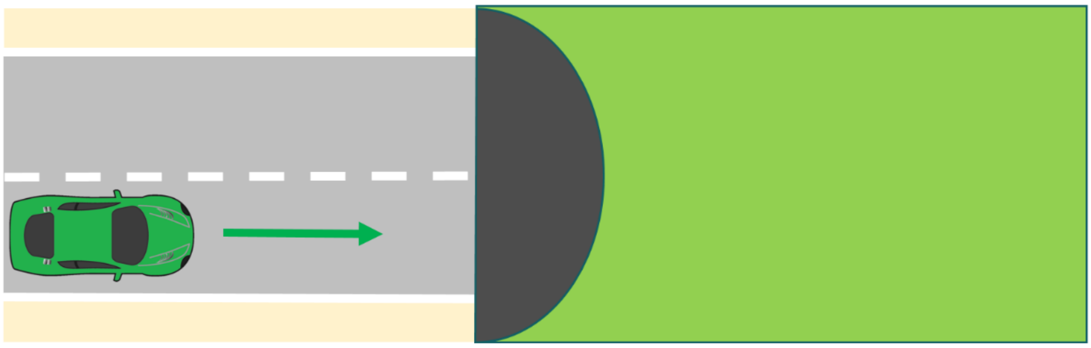
The test road consists of at least two lanes. The signal lights are set above the road, and the signal lights of adjacent lanes are kept in green state.

Figure 13
Testing scenario diagram for expressway signal lights



(c) Drive through tunnel

Figure 14
Schematic representation of driving through tunnel



General description:

The ego vehicle is driving through a tunnel (lack of GPS signals and natural light) [4]. The vehicle needs to adapt to the quickly changing light parameters and lack of global positioning. Depending on the speed of the ego vehicle, the difference between the light conditions outside and inside the tunnel and the length of the tunnel, the difficulty of the scenario is changing.

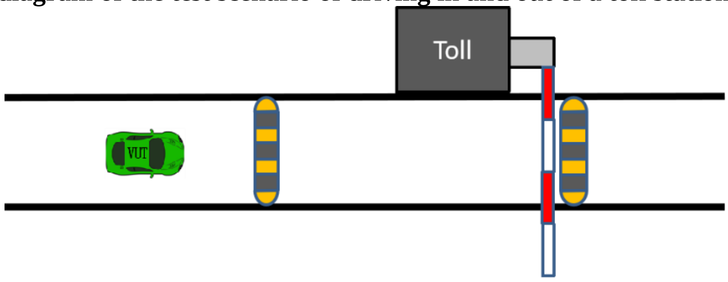
Emphasized scenario parameters: ego velocity, light conditions.

Tested parameters: ego lateral and longitudinal velocity, deviation to lane centre, etc.

(d) Toll

The test road is a long straight road with at least one lane. A toll station is set on this section, and toll station signs, speed limit signs and speed bumps are set in front of the toll station. This is shown in Figure 15.

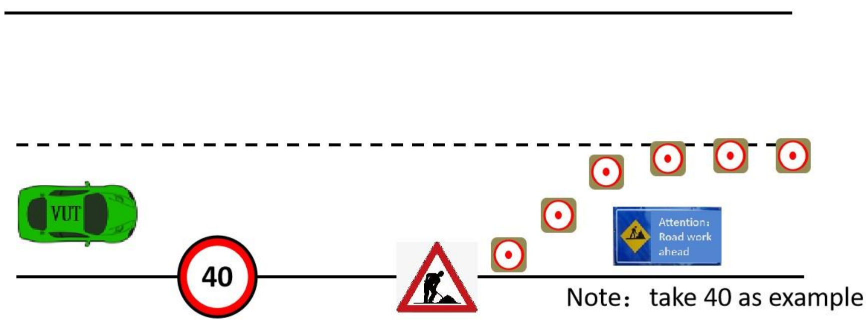
Figure 15
Schematic diagram of the test scenario of driving in and out of a toll station



(e) Conventional obstacles

The test road is a long straight road containing at least two lanes, and the middle lane line is a white dashed line. Within the lanes, conical traffic signs and traffic markings are placed according to the traffic control requirements of the road maintenance operation. This is shown in Figure 16.

Figure 16
Diagram of a conventional obstacle course.



D. Country specific road geometry

Note: This scenario is only applicable for limited countries or regions. Therefore, application of this scenario can be unnecessary depends on the target market of the ADS.

(a) Interceptor

For the ego vehicle, junctions present a challenge due to the increased likelihood of conflicts with other actors.

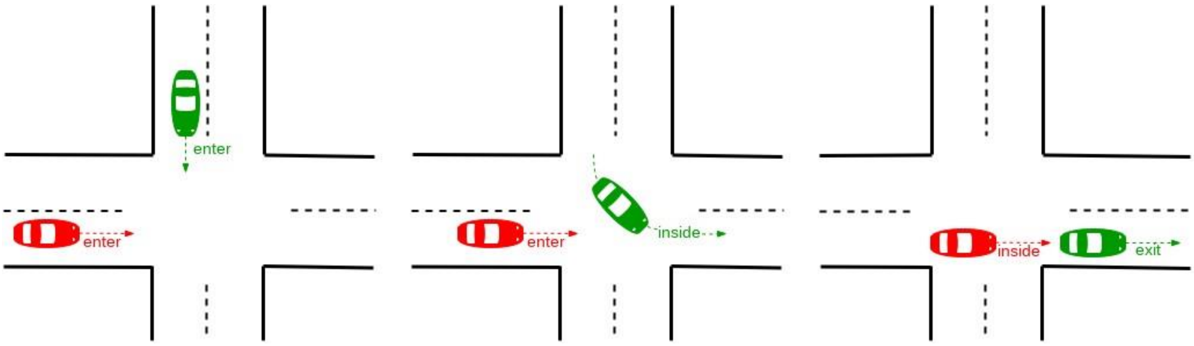
In this scenario, the ego vehicle traverses an intersection simultaneously with another car - the interceptor. This scenario tests the ego vehicle’s behaviour when on a collision course with another car in an intersection, possibly with signs, signals, or traffic lights. The ego vehicle should be able to safely manoeuvre through the intersection and avoid or mitigate a collision.

Environmental requirements: A junction with at least three ways. It may or may not be controlled (i.e. have yield sign, traffic lights, etc.).

Ego vehicle behaviour: The ego vehicle traverses the junction in any direction (left, right or straight).

Other actors' behaviour: Another car approaches the same junction, from a different direction and traverses the junction such that its trajectory intersects with the ego vehicle’s trajectory.

Figure 17
Interceptor scenario



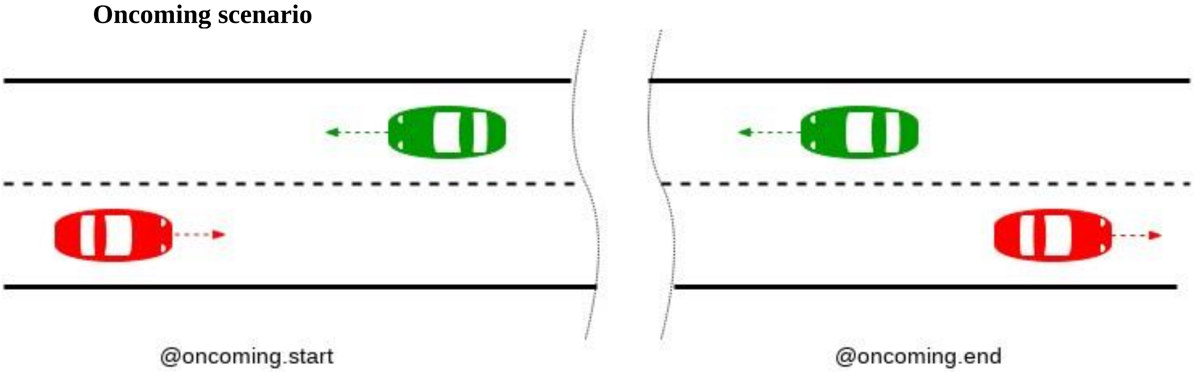
E. Unusual situation

Note: This scenario can happen in the real world. However, whether this kind of scenarios should be covered should be discussed in the appropriate group.

(a) Wrong way driver (oncoming)

Oncoming is a scenario in which a car approaches the DUT from the opposite direction and drives past the ego vehicle.

Figure 18
Oncoming scenario



Environmental requirements: A two-lane road with traffic moving in opposite directions.

Ego vehicle behaviour: The ego vehicle drives in a lane, presumably at a constant speed.

Other actors' behaviour: At the start of the scenario, another car is in the opposing lane, approaching the ego vehicle. At the end, the other car is still in the opposing lane, having passed the ego vehicle.

VII. References

1. UN Regulation No. 157 (Automated Lane Keeping System), Available online at https://undocs.org/ECE/TRANS/WP.29/2020/81 (original version) or https://unece.org/transport/documents/2021/03/standards/un-regulation-no-157-automated-lane-keeping-systems-alks

2. E. de Gelder, O. Op den Camp, N. de Boer, (The Netherlands): Scenario Categories for the Assessment of Automated Vehicles, Version 1.7, January 21, 2020.

3. SAFE (Foretellix) Highway and ADAS Traffic Scenario Library, Scenario Definitions at the functional Level, Version 1.0, November 2020.

4. Japan: Proposal of Traffic Scenarios for Highway Driving (Supplemental version for presentation), December 2020.

5. China (CATARC): Proposal about functional scenario from CATARC, December 2020.

6. European Commission - Joint Research Center. Speed profile for car-following tests. Available online at https://wiki.unece.org/download/attachments/92013066/ACSF-25-13%20%28EC%29%2020190121_TestSpecification_ALKS_JRC.pdf?api=v2

7. IGLAD 2019 Codebook, Conflict Types, 2019.

Annex III

Credibility assessment for using virtual toolchain in ADS validation

I. Introduction, motivation, and scope

The use of Modelling and Simulation (M&S) is becoming widespread thanks to the increasing computational capabilities, accuracy, usability, and availability of M&S software packages. M&S can be beneficial for ADS safety validation because it allows to overcome some real testing limitations and to increase the number of testing scenarios. Nonetheless, M&S can also lead to erroneous/seemingly correct results, especially in relation to complex simulations not adequately supported by robust practices addressing all M&S aspects beyond pure validation. Therefore, higher confidence in M&S credibility is needed to apply virtual testing instead of/in conjunction with the other NATM pillars. In other words, M&S can be used for virtual testing if an assessor is able to consider the simulation results *credible* enough to make sound decisions taking into account the potential uncertainties of M&S. The validation of M&S can be considered the hallmark of simulation credibility. However, the validation has some limitations, which include the limited scope of the validation tests and the difficulty in retrieving data supporting the validation procedures. The use of M&S requires more attention towards all factors influencing the quality and validity of M&S with aim at:

(a) Identifying a common framework to determine, justify, assess and report the overall credibility of the M&S;

(b) Indicating the levels of confidence in results from the validation phase.

At the same time, this framework should be general enough to be used for different M&S types and applications. However, the goal is complicated by the broad differences across ADS features and the variety of M&S types and applications. These considerations lead to introduce a (risk-based/informed) credibility assessment framework relevant and appropriate to all M&S applications.

The proposed credibility assessment framework provides a general description of the main aspects considered for assessing the credibility of an M&S solution together with guidelines of the role played by third parties assessors[5] in the validation process with respect to credibility. Concerning the latter point, the assessor should investigate the produced documentation supporting credibility at the audit phase, whereas the actual validation tests occur once the ADS manufacturer has developed the integrated simulation systems.

Ultimately, the outcome of the current credibility assessment should define the envelope in which the virtual tool can be used to support the ADS assessment.

II. Components of the credibility assessment framework.

M&S can be used for virtual testing if its credibility is established by evaluating the fitness of M&S for the intended purpose. The credibility can be achieved by investigating and assessing five M&S properties:

(a) Capability – what the M&S can do, and what are the risks associated;

(b) Accuracy – how well M&S does reproduce the target data;

(c) Correctness – how sound & robust are M&S data and algorithms;

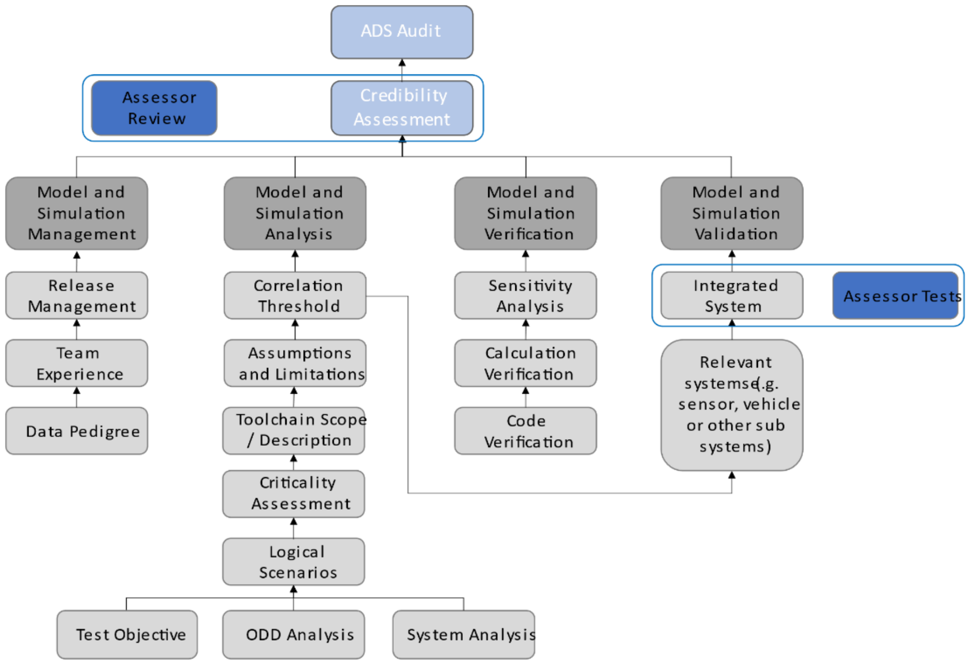
[5] For type-approval the manufacturer produces the whole documentation upfront with the authority requested to study it and provide its assessment. In the self-certification the same may be done by the assessor during market surveillance

(d) Usability – what training and experience is needed and what quality of the process applied to it.

(e) Fit for Purpose – how suitable the M&S is for the ODD and ADS assessment.

Therefore, credibility requires a unified method to investigate these properties and get confidence in the M&S results. The Credibility Assessment framework introduces a way to assess and report the credibility of M&S based on quality assurance criteria that allow indicating the levels of confidence in results. In other words, the credibility is established by evaluating the following M&S influencing factors that are considered as main contributors for M&S properties and therefore for the overall M&S credibility: M&S management, team's experience and expertise, M&S analysis and description, data/input pedigree, verification, validation, uncertainty characterization. Each of these factors indicates the level of quality achieved by M&S, and the comparison between the obtained levels and the required levels leads to consider the M&S credible and fitness to use for virtual testing. A graphical representation of the relationship among the components of the credibility assessment framework is reported in Figure 1.

Figure III.1.
Graphical representation of the relationships between the components of the credibility assessment framework



A. Models and Simulation Management.

The M&S lifecycle is a dynamic process with frequent releases that should be monitored and documented. Management activities should be established to support the M&S in a work product management fashion. Relevant information on the following aspects should be included in this section.

1. M&S management process

This part should:

(a) Describe the modifications within the releases,

(b) Designate the corresponding software (e.g., specific software product and version) and hardware arrangement (e.g., XiL configuration),

(c) Record the internal review processes that accepted the new releases,

(d) Be supported throughout the full duration of the virtual model utilization

2. Releases management

Any M&S toolchain's version used to release data for certification purposes should be stored. The virtual models constituting the testing toolchain should be documented in terms of the corresponding validation methods and acceptance thresholds to support the overall credibility of the toolchain. The developer should enforce a method to trace generated data to the corresponding M&S version.

Quality check of virtual data. Data completeness, accuracy, and consistency should be ensured throughout the releases and lifetime of an M&S toolchain to support the verification and validation procedures.

3. Team's Experience and Expertise.

Even though Experience and Expertise (E&E) are already covered in a general sense within organization, it is important to establish the basis for confidence on the specific experience and expertise for M&S activities.

In fact, the credibility of M&S depends not only on the quality of the simulation models but also on the E&E of the personnel involved in the validation and usage of the M&S. For instance, a proper understanding of the limitations and validation domain will prevent from possible misuse of M&S or from misinterpretation of its results.

In this perspective, it is important to establish the basis for the ADS manufacturer's confidence on the experience and expertise of:

(a) The Teams that will validate the simulation toolchain and,

(b) The Teams that will use the validated simulation for the execution of virtual testing with the purpose of validating the ADS.

Thus, Team's E&E increase the level of confidence on the credibility of M&S and its outcomes by ensuring that the human factors behind the M&S are taken into consideration and any possible human component risk is controlled as expected in any suitable Management System.

If the ADS manufacturer's tool chain incorporates or relies upon inputs from organizations or products outside of the manufacturer's own team, the ADS manufacturer will include an explanation of measures it has taken to support its confidence in the quality and integrity of those inputs.

Team's Experience and Expertise include two levels:

(a) Organizational level:

The credibility is established by setting up processes and procedures to identify and maintain skills, knowledge, and experience to perform M&S activities. The following processes should be established, maintained and documented:

(i) Process to identify and evaluate the individual's competence and skills;

(ii) Process for training competent personnel to perform M&S-related duties

(b) Team level:

Once a M&S has been finalized, its credibility is mainly dictated by the skills and knowledge of the individual/team that will validate the M&S Toolchain and will use the M&S for the validation of ADS. The credibility is established by documenting that these Teams have received adequate training to fulfil their duties.

The ADS manufacturer should then:

(i) Provide the basis for the ADS manufacturer's confidence in the Experience and Expertise of the individual/team that validates the M&S Toolchain.

(ii) Provide the basis for the ADS manufacturer's confidence in the Experience and Expertise of the individual/team that uses the simulation to execute virtual testing with the purpose of validating the ADS.

The ADS manufacturer's demonstration of how it applies the principles of ISO 9001 or a similar best practice or standard with regard to the competence of its M+S organization and the individuals in that organization will provide the necessary basis for this determination. The assessor may not substitute its judgment for that of the ADS manufacturer with regard to the experience and expertise of the organization or its members.

4. Data/Input pedigree

The data/input pedigree contains a record of traceability from the ADS manufacturer's data used in the validation of the M&S.

(a) Description of the data used for the M&S

(a) The ADS manufacturer should document the data used to validate the model and note important quality characteristics;

(b) The ADS manufacturer should provide documentation showing that the data used to validate the models covers the intended functionalities the toolchain aims at virtualizing;

(c) The ADS manufacturer should document the calibration procedures employed to fit the virtual models' parameters on the collected input data.

(b) Effect of the data quality (e.g. data coverage, signal to noise ratio, and sensors' uncertainty/bias/sampling rate) on model parameters uncertainty

The quality of the data used to develop the model will have an impact on model parameters' estimation and calibration. Uncertainty in model parameters will be another important aspect in the final uncertainty analysis.

5. Data/Output pedigree

The data/output pedigree contains a record of the signals selection that the M&S allows investigating.

(a) Description of the data generated by the M&S

(a) The ADS manufacturer should provide [information on] any data and scenarios used for virtual testing toolchain validation.

(b) The ADS manufacturer should document the exported data and note important quality characteristics e.g. using the correlation methodologies as defined Annex II.

(c) The ADS manufacturer should trace a M&S output to the corresponding simulation setup:

(i) Effect of the data quality M&S credibility

(a) The M&S output data should be sufficiently wide to ensure the correct execution of the validation computation. The data should sufficiently reflect the ODD relevant to the virtual assessment of the ADS.

(b) The output data should allow consistency/sanity check of the virtual models via possibly exploiting redundant information

(ii) Managing stochastic models

(a) Stochastic models should be characterized in terms of their variance

(b) Stochastic models should be ensured the possibility of deterministic re-execution

B. M&S Analysis and description

The M&S analysis and description aim to define the whole M&S and identify the parameter space that can be assessed via virtual testing. It defines the scope and limitations of the models and toolchain and the uncertainty sources that can affect its results.

(a) General description

(a) ADS manufacturer should provide a description of the complete toolchain along with how the simulation data will be used to support the ADS validation strategy.

(b) The ADS manufacturer should provide a clear description of the test objective.

(b) Assumptions, known limitations and uncertainty sources:

(a) The ADS manufacturer should motivate the modelling assumptions which guided the design of the M&S toolchain

(b) The ADS manufacturer should provide evidence on:

(i) How the manufacturer-defined assumptions play a role in defining the limitations of the toolchain;

(ii) The level of fidelity required for the simulation models.

(c) The ADS manufacturer should provide justification that the tolerance for sim-real correlation is acceptable for the test objective

(d) Finally, this section should include information about the sources of uncertainty in the model. This will represent an important input to final uncertainty analysis, which will define how the model outputs can be affected by the different sources of uncertainty of the model used.

(c) Scope (what is the model for?). It defines how the M&S is used in the ADS validation.

(a) The credibility of virtual tool should be enforced by a clearly defined scope of utilization the developed models.

(b) The matured M&S should allow a virtualization of the physical phenomena to a degree of accuracy which matches the fidelity level required for certification. Thus, the M&S will act as a “virtual proving ground” for ADS testing.

(c) Simulation models need dedicated scenarios and metrics for validation. The scenario selection used for validation should be sufficient such that there is confidence that the toolchain will perform in the same manner in those scenarios outside of the validation scope.

(d) ADS manufacturers should provide a list of validation scenarios together with the corresponding parameters’ limitation.

(e) ODD analysis is a crucial input to derive requirements, scope, effects that the M&S must consider in order to support ADS validation.

(f) Parameters generated for the scenarios will define extrinsic and intrinsic data for the toolchain and the simulation models.

(d) Criticality assessment

The simulation models and the simulation tools used in the overall tool-chain should be investigated in terms of their responsibility in case of a safety error in the final product. The proposed approach for criticality analysis is derived from ISO 26262, which requires qualification for some of the tools used in the development process. In order to derive how critical the simulated data is, the criticality assessment considers the following parameters:

(a) The consequences on human safety e.g. severity classes in ISO 26262.

(b) The degree in which the simulated results influence’s the ADS.

The table below provides a sample criticality assessment matrix to demonstrate this analysis. ADS manufacturers may adjust this matrix to their particular use case.

Table III.1
Criticality assessment matrix

Influence on ADS	Significant	N/A	Perform degraded mode within reduced system constraints	Create a collision free and lawful driving plan	Correctly execute and actuate the driving plan
	Moderate		Determine its location	Predict the future behaviour of other actors	Perceive relevant static and dynamic objects in the proximity of the ADS
	Minor	Strategic control of the ADS by the User	Communicate and interact with other road users	Safe management of transitions of control	Determine if specified nominal performance is not achieved
	Negligible	User interaction with HMI	User informed about operational status	N/A	
		Negligible	Minor	Moderate	Significant
		Decision consequence			

From the perspective of the criticality assessment, the three possible cases for assessment are:

(a) Those models or tools that fall within the red boxes are clear candidates for fully following the credibility assessment;

(b) Those models or tools that fall within the yellow boxes may or may not be candidates for fully following the credibility assessment at the discretion of the assessor;

(c) Those models or tools that fall within the green boxes are not required to follow the credibility assessment.

C. Verification

The verification of an M&S deals with the analysis of the correct implementation of the conceptual/mathematical models building up the M&S toolchain. The verification contributes to the M&S's credibility via providing assurance that the M&S will not exhibit unrealistic behavior for a set of input which cannot be tested. The procedure is grounded on a multi-step approach which includes code verification, calculation verification and sensitivity analysis.

1. Code verification

Code verification is concerned with the execution of test demonstrating that no numerical/logical flaws affect the virtual models

(a) The ADS manufacturer should document the execution of proper code verification techniques, e.g. static/dynamic code verification, convergence analysis and comparison with exact solutions if applicable6

(b) The ADS manufacturer should provide documentation showing that the exploration in the domain of the input parameters was sufficiently wide to identify parameters' combination for which the M&S shows unstable or unrealistic behavior. Coverage metrics of parameters combinations may be used to demonstrate the required exploration of the models behaviours.

(c) The ADS manufacturer should adopt sanity/consistency checking procedures whenever data allows

2. Calculation verification

Calculation verification deals with the estimation of numerical errors affecting the M&S:

(a) The ADS manufacturer should document numerical error estimates (e.g. discretization error, rounding error, iterative procedures convergence);

(b) The numerical errors should be kept sufficiently bounded to not affect validation.

3. Sensitivity analysis

Sensitivity analysis aims at quantifying how model output values are affected by changes in the model input values and thus pointing out the parameters having the greatest impact on the simulation model results. The sensitivity study also affords determining the extent to which the simulation model satisfies the validation thresholds when it is subjected to small variations of the parameters, thus it plays a fundamental role to support the credibility of the simulation results.

(a) The ADS manufacturer should provide supporting documentation demonstrating that the most critical parameters influencing the simulation output have been identified by means of sensitivity analysis techniques such as by applying a perturbation of the model's parameters;

(b) The ADS manufacturer should demonstrate that robust calibration procedures have been adopted while identifying and calibrating the most critical parameters to the end of increasing the credibility of the developed toolchain.

(c) Ultimately, the sensitivity analysis results will also help defining the inputs and parameters whose uncertainty characterization needs particular attention in order to properly define the uncertainty of the simulation results.

4. Validation

The quantitative process of determining the degree to which a model or a simulation is an accurate representation of the real world from the perspective of the intended uses of the M&S. Examples of virtual toolchain validation are reported in Annex 3 - Appendix 3 to the present document.

(a) Measures of Performance (metrics)

(a) The performance metrics are defined during the M&S analysis.

(b) Metrics for validation may include:

(i) Discrete value analysis e.g. detection rate, firing rate;

(ii) Time evolution e.g. positions, speeds, acceleration;

(iii) Flow of actions based analysis e.g. distance/speed calculations, TTC calculation, brake initiation.

---

[6] Roy, C. J. (2005). Review of code and solution verification procedures for computational simulation. *Journal of Computational Physics*, *205*(1), 131-156.

(b) *Goodness of Fit measures*

(a) The analytical frameworks used to compare real world and simulation metrics. They are generally Key Performance Indicators (KPIs) indicating the statistical comparability between two sets of data. Examples of goodness of fit measures and correlation methodologies that can be used in the validation process are reported in Annex 3 - Appendix 2 to the present document.

(b) The validation should show that these KPIs are met.

(c) *Validation methodology*

(a) The ADS manufacturer should define the logical scenarios used for virtual testing toolchain validation. They should be able to cover to the maximum possible extent the ODD of virtual testing for ADS validation.

(b) The exact methodology depends on the structure and purpose of the toolchain. The validation may consist of one or more of the following:

(i) Validate Subsystem models e.g. environment model (road network, weather conditions, road user interaction), sensor models (Radio Detection And Ranging (RADAR), Light Detection And Ranging (LiDARs), Camera), vehicle model (steering, braking, powertrain);

(ii) Validate vehicle system (vehicle dynamics model together with the environment model);

(iii) Validate sensor system (sensor model together with the environment model);

(iv) Validate integrated system (sensor model + environment model with influences form vehicle model).

(d) *Accuracy requirement*

Requirement for the correlation threshold is defined during the M&S analysis. The validation should show that these KPIs are met. e.g. using the correlation methodologies as defined in Annex 2.

(e) *Validation scope (what part of the toolchain to be validated)*

A toolchain consists of multiple tools, and each tool will use a number of models. The validation scope includes all tools and their relevant models.

(f) *Internal validation results*

(a) The documentation should not only provide evidence of the simulation model validation but also used to obtain sufficient information related to the processes and products that provide overall credibility of the toolchain used.

(b) Documentation/results may be carried over from previous credibility assessments.

(g) *Independent Validation of Results*

The assessor should audit the documentation provided by the manufacturer and may carrying out physical tests of the complete integrated tool

(h) *Uncertainty characterisation*

This section is concerned with characterizing the expected variability of the virtual toolchain results. The assessment should be made up of two phases. In a first phase the information collected the M&S Analysis and Description section and the Data/Input Pedigree are used to characterise the uncertainty in the input data, in the model parameters and in the modelling structure. Then, by propagating all the uncertainties through the virtual toolchain, the uncertainty in the model results is quantified. Depending on the uncertainty in the model results, proper safety margins will need to be introduced by the ADS manufacturer in the use of virtual testing of ADS validation.

(i) *Characterization of the uncertainty in the input data*

The ADS manufacturer should demonstrate to have opportunely estimated critical model's inputs by means of robust techniques such as providing multiple repetitions for the assessment of the quantity;

(j) *Characterization of the uncertainty in the model parameters (following calibration)*

The ADS manufacturer should demonstrate that critical model's parameters that cannot be estimated identically are characterized by means of a distribution and/or confidence intervals;

(k) *Characterization of the uncertainty in the M&S structure*

The ADS manufacturer should provide evidence that the modeling assumptions are given a quantitative characterization of the generated uncertainty (e.g. comparing the output of different modeling approaches whenever possible);

(l) *Characterization of aleatory vs. epistemic uncertainty:*

The ADS manufacturer should aim to distinguish between the aleatory component of the uncertainty (which can only be estimated but not reduced) and the epistemic uncertainty deriving from the lack of knowledge in the virtualization of the process.

III. Documentation structure

This section will define how the aforementioned information will be collected and organized in the documentation provided by the ADS manufacturer to the relevant authority:

(a) The ADS manufacturer should produce a document (a "simulation handbook") structured after the present outline providing evidence for the topics presented;

(b) The documentation should be delivered together with the corresponding release of the M&S and related produced data;

(c) The ADS manufacturer should provide clear reference that allows tracing the documentation to the corresponding M&S/data;

(d) The documentation should be maintained throughout the whole lifecycle of the M&S utilization. The assessor may audit the ADS manufacturer through assessment of their documentation and/or by conducting physical tests.

IV. Interdependences with VMAD Subgroups 1 and 3

The scenarios developed by the IWG on VMAD Subgroup are the input of the M&S toolchain.

The credibility analysis can be exploited to support industry audit's procedures established in VMAD Subgroup 3.

Annex III - Appendix 1

Virtual Testing Toolchain Example

Virtual Testing is introduced to reduce the burden of physical tests and effectively provide evidence on the ADS performance across the operational domain. However, no one simulation tool can be used to test all aspects of the ADS software, this is why manufacturers will exploit the attributes of various simulation tools to develop confidence in the safety of the full system.

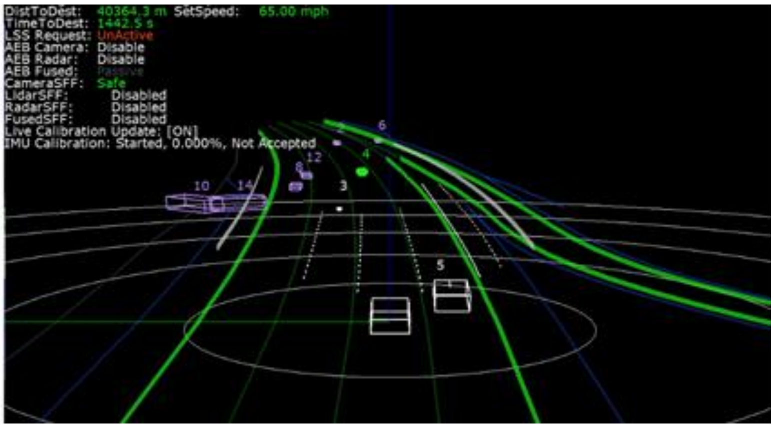
Each virtual testing tool will have its own strengths and weakness based on the speed and cost of execution and the level of fidelity achieved. Typically, lower fidelity tools are used to cover a vast number of scenarios to obtain a general understanding of the systems performance. Then it is possible to increase the level of fidelity within a subset of scenarios to validate the performance of the ADS in a statistically relevant number of realistic scenarios. A manufacturer's virtual testing toolchain may consist of the following tools:

I. Perception simulation

Perception simulation can be used to train and validate the perception algorithms of the ADS software with physical accurate sensor models in combination with ground-truth data. This can be done in open-loop since the planning and control algorithms are bypassed.

II. Planning & Control (P&C) simulation

P&C simulation can be used to validate the control algorithms of the ADS software with basic sensor models. This can be done faster than real time so is an effective way to test the control system over a vast number of scenarios.



III. Full AV Stack simulation (MIL, SIL or HIL)

Full AV Stack simulation can accurately render sensor data streams that represent a wide range of environments and scenarios. The ADS software processes the simulated data as if it were coming from the sensors of a vehicle actually driving on the road and sends actuation commands back to the simulator. This allows engineers to test rare conditions, such as rainstorms, snowstorms, or sharp glare at different times of the day and night. Each scenario can be tested repeatedly, adjusting multiple variables such as road surfaces and surroundings, weather conditions, other traffic, and time of day.

HIL can be used to test the entire hardware component or ECU before the real vehicle is available and to test the interactions/ networks of the components within the virtual prototype e.g. conduct E/E failure test of hardware components.

IV. Vehicle in the Loop (VIL)

VIL provides a validation environment for ready-to-drive vehicles in combination with a virtual environment simulation. It allows to execute complex and safety critical scenarios on vehicle level.

A. VIL on Test Beds

VIL on test beds combines this with the advantages of a lab and focuses on flexibility in scenario generation and reproducibility of scenario execution. It allows additionally to test the real sensors and perception in the loop.

VIL on Test bed may consist of the following elements:

(a) Longitudinal dynamics: The longitudinal dynamics are emulated by the test bed. This can either be a chassis dynamometer or a wheel hub / powertrain test bed. High dynamic dynamometers in combination with a vehicle dynamics simulation allow the execution of various maneuvers and scenarios including high dynamic maneuvers at the limits (realistic wheel slip, etc.);

(b) Lateral dynamics: In case of lateral dynamics, including the steering is required, test beds can be extended by additional devices to allow steering. Ideally steering is not only allowed but also the resulting reaction forces are emulated properly to avoid error states and to ensure a proper operation together with the AV function;

(c) Interface virtual environment simulation: Depending on the use case and the requirements, there are different possibilities: Object list injection (no sensor, no perception in the loop), raw data injection (no sensor but perception in the loop), over-the-air stimulation of the sensor (sensor and perception in the loop). Using the over-the-air stimulation, there are no modification on the vehicle required. Also, a mixed operation is possible.

B. VIL on proving grounds

VIL on proving grounds focuses more on the interaction between the driver/passenger and the vehicle. In this configuration the real acceleration (longitudinal and lateral) of the vehicle can be experienced by the driver/passenger (difference to Vehicle-in-the-Loop at test beds). A judgment and rating by the real driver are possible.

VIL Test bed may consist of the following elements:

(a) Longitudinal dynamics: The real longitudinal dynamics are available

(b) Lateral dynamics: The real lateral dynamics are available

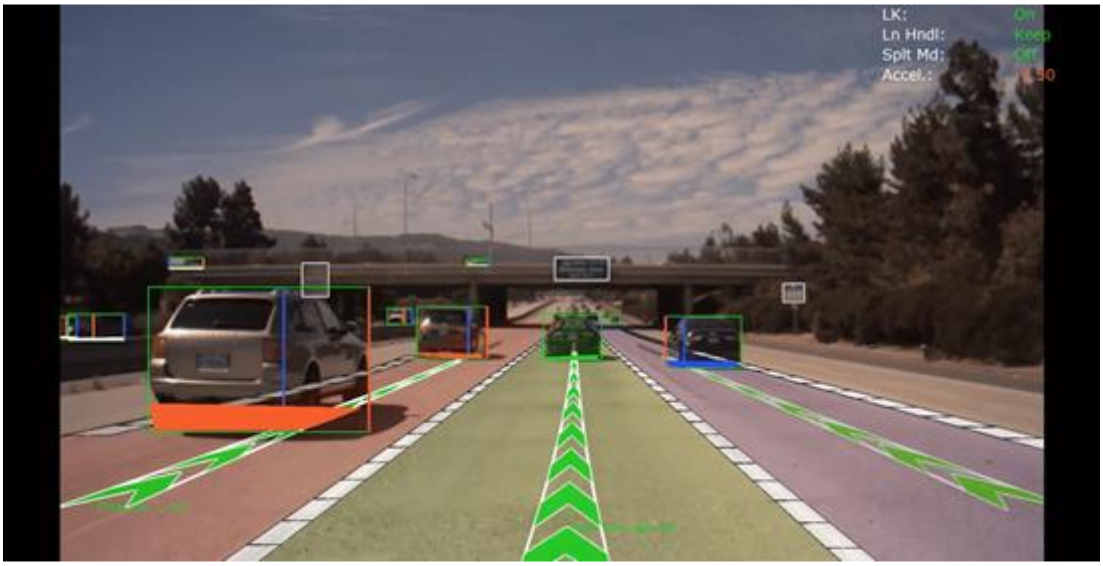
(c) Interface virtual environment simulation: Typically, the interface between the vehicle and the virtual environment is done via object list injection. Also, raw data injection is possible. Real sensors cannot be considered (with a few exceptions for very simple sensors like ultrasonic).

V. Driver in the Loop (DIL)

DIL virtual testing can be helpful to support the assessment of this category of functional requirement by analysing the interaction between the driver and the ADS in a safe and controlled environment.

VI. Software Reprocessing (SwR)

SwR involves playing back previously recorded sensor data, rather than synthetic data, to the ADS software to accurately assess the perception performance in an open loop system.



Considering the categories of functional requirements currently being considered, virtual testing seems particularly relevant for assessing requirements related to:

(a) ADS should drive safely and ADS should manage safety critical situations. These are the requirements where virtual testing can play the most prominent role. MIL/SIL, HIL and VIL virtual testing can all be used to assess these requirements at different stages of vehicle verification and validation.

(b) ADS should interact safely with the user. DIL virtual testing can be helpful to support the assessment of this category of functional requirement by analyzing the interaction between the driver and the ADS in a safe and controlled environment.

(c) ADS should safely manage failure modes and ADS should ensure a safe operational state. The use of virtual testing in these two categories is also very promising but would probably require further research work. SIL virtual testing could include simulated failures and maintenance requests. HIL and VIL virtual testing could be used to assess how the system would react to the occurrence of a real malfunctioning induced to the real system.

Functional Requirement	*SIL*	*HIL*	*VIL*	*DIL*	*SwR*
ADS should drive safely	Y	Y	Y	-	Y
ADS should interact safely with the user	Y	Y	Y	Y	-
ADS should manage safety-critical situations	Y	Y	Y	-	Y
ADS should safely manage failure modes	Y	Y	Y	--	-

The table below describes all available test environments. The main difference in these test environments is in the application of virtual and real stimuli and in the items being tested.

Virtual Testing Tool	*Software*	*Hardware*	*Vehicle*	*Driver*	*Environment*
Perception	Real	Virtual	Virtual	Virtual	Virtual
Planning & Control	Real	Virtual	Virtual	Virtual	Virtual
Full AV Stack (SIL)	Real	Virtual	Virtual	Virtual	Virtual
Full AV Stack (HIL)	Real	Real	Virtual	Virtual	Virtual
Vehicle in the Loop	Real	Real	Real	Virtual	Virtual
Driver in the Loop	Virtual	Virtual	Virtual	Real	Virtual
Software Reprocessing	Real	Virtual	None	None	Real
Proving Ground	Real	Real	Real	Real	None
Real World Test	Real	Real	Real	Real	Real

Annex III - Appendix 2

Example of correlation methodologies

The validation of a virtual testing toolchain shall be based on the quantitative evaluation of a set of KPIs with respect to the real-world data. The assessment returns a measure of correlation which has to be checked against a prescribed correlation threshold. It is recognized that no method for correlating sim-real data is suitable for all virtual testing tools, it is therefore the responsibility of the ADS manufacturer to justify the chosen correlation methodologies.

The computation of the correlation is carried out comparing either time-series or probability distributions depending on the data availability and the virtual testing setup. Deterministic virtual testing environments such as MIL and SIL will originate deterministic results with no possibility of assessing the confidence intervals. Similarly, real-world testing leveraging on a single execution per each test does not allow assessing confidence intervals. Thus, when a MIL testing environment is compared to a single execution for validation purposes, only time-series comparison analysis is possible.

On the other side, a HIL or VIL testing environment is subject to a certain degree of stochasticity, which implies that multiple repetitions will originate a statistical distribution of the results. An analogous result is obtained via the execution of several repetitions for a given proving ground scenario. This way of proceeding allows carrying out statistical testing on the collected data distributions.

I. Graphical comparison

Graphical comparisons provide a first validation step which displays the goodness of the simulation model. Nonetheless, the subjectivity inherent to the qualitative nature of the assessment implies that graphical comparisons are only suitable to support the credibility of the developed toolchain. A proper validation methodology shall be based on the quantitative methods described below.

II. Scalar data comparison

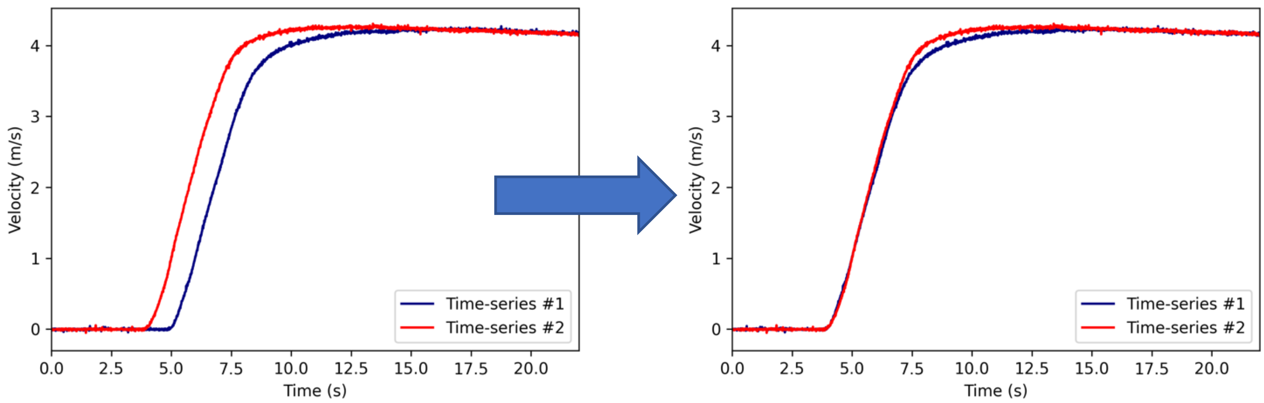
Scalar data comparisons are useful tools to compare significant values of a signal. When only the pick values of a signal is relevant (e.g. the maximum yaw-rate during an emergency obstacle avoidance maneuver) for the sake of validation, the Relative Error Criterion (REC) [1] difference amplitude criterion is a suitable metrics

$$\frac{|peak_{real} - peak_{sim}|}{peak_{real}} * 100.$$

III. Time-series comparison

Despite being the first step into the quantitative evaluation, scalar data provide limited information about the agreement of the signals. The study of time-series affords to investigate the correlation of the simulation-generated evidence with the real-world data to a greater extent.

Several tools exist to quantify the distance between time-series. Before any attempt of comparison can be pursued, the time-series have to be synchronized and resampled based on the lowest frequency between real-world and the simulated data. A widespread solution for the synchronization is to adopt the Time-of-Arrival (ToA) criterion. ToA implies the definition of a reference starting time for the signals which is derived from the first time the signal reached a pre-defined amplitude.



Once opportunely synchronized and resampled, the time series can be analyzed according to a distance function. Distance estimation is typically carried out by applying some norm function to the vector of residuals. For instance, the L2 norm (Euclidean distance) reads as:

$$\sqrt{\sum_{i}^{N}\left(y_{sim,i}-y_{real,i}\right)^{2}},$$

where N is the total list of samples. The normalization of the L_2 norm over the total number of samples yields the Root Mean Square Error (RMSE):

$$\sqrt{\frac{1}{N}\sum_{i}^{N}\left(y_{sim,i}-y_{real,i}\right)^{2}}.$$

Alternative norms can be used to quantify the discrepancies between the time-series, which are susceptible to different features or error signals. For instance, the L_∞ norm returns the maximum absolute value of the error

$$max_i\ \left(\left|y_{sim,i}-y_{real,i}\right|\right).$$

Recently developed metrics allow separating the contribution of phase error (thus the shape of the time-series) to the contribution of the magnitude error between the signals, thus providing more insights on possible inconsistencies affecting the model. A recent report published by Sandia [2] investigates such techniques. In particular, the Sprague-Geers [3] metric is presented therein. The same criterion is also adopted to validate virtual models for seats within the field of aviation [4]. The metric is based on establishing the integral distance between the signals

$$d_M=\sqrt{\frac{\sum_i^N y_{sim,i}^2}{\sum_i^N y_{real,i}^2}}-1,$$

and the phase difference

$$d_P=\frac{1}{\pi}\ \cos^{-1}\left(\frac{\sum_i^N y_{sim,i}*y_{pg,i}}{\sqrt{\sum_i^N y_{sim,i}^2*\sum_i^N y_{real,i}^2}}\right)$$

combined into the total error

$$d_{SG} = \sqrt{d_M^2 + d_P^2}.$$

An alternative analysis that can be carried is establishing the correlation between the signals. Several tools to calculate the correlation have been proposed in the literature [5]. Among them, a commonly adopted tool is the Pearson correlation

$$r_{sim,real} = \frac{\left|\sum_{i=1}^{N} \left(y_{sim,i} - \bar{y}_{sim}\right)\left(y_{real,i} - \bar{y}_{real}\right)\right|}{\sqrt{\sum_{i=1}^{m} \left(y_{sim,i} - \bar{y}_{sim}\right)^2 \sum_{i=1}^{m} \left(y_{real,i} - \bar{y}_{real}\right)^2}} .$$

Values of $r_{sim,real}$ close 1 suggest good agreement between the signals, whereas correlation degrades approaching 0.

IV. Statistical testing

Statistical testing is concerned with verifying whether the null hypothesis, i.e.,: “the model is an accurate representation of the real-world phenomena,” cannot be rejected given the evidence generated by the simulation. Statistical testing is particularly useful when dealing with non-deterministic virtual testing environments or multiple repetitions of the same driving scenario on the proving ground.

A common statistical test is the well-known T-test which analyzes whether two distributions have a significatively different mean. T-test can be performed on both one-sample or two-sample datasets. A one-sample case study involves determining whether the mean of a population ($\bar{x}$) is statistically different from a given reference mean μ_0. The “*t*”-value can be calculated as

$$t = \frac{(\bar{x} - \mu_0)\sqrt{N}}{s},$$

where *S* is the standard deviation of the sample. One can reject the null-hypothesis is the *t* value exceeds the critical value resulting from the sample size *N* and significance level.

A typical example for the one-sample T-test is investigating whether the experimental mean of a quantity differs significatively from the distribution of the same quantity deriving from multiple repetitions on a HIL/VIL setup. Similarly, comparing multiple repetitions on a proving ground with the evidence derived from a deterministic environment originates a one-sample exercise. Conversely, two-sample T-test is found when two distributions are compared. The comparison of more than two distributions can be carried out by exploiting ANOVA.

While the T-test is mainly concerned with studying the mean of distributions, alternative tests exist which do not make assumptions on input data normality. For instance, the Kolmogorov-Smirnov test evaluates the maximum vertical distance in the Cumulative Distribution Functions (CDFs) of the input distributions.

Annex III - Appendix 3

Validation examples

This section presents modeling and validation approaches for three models' classes: lane and camera, RADARs and LiDARs, and vehicle dynamical models. The first two paragraphs are concerned with describing the realization of virtual perception modules, which, together with the virtual vehicle models described in the last paragraph, enable interfacing the ADS with the simulation environment. The contribution includes examples of metrics and relevant KPIs which afford the determination of the fidelity level returned by the virtual solution.

The focus of the current discussion is on the simulation models per se (intrinsic properties). Nonetheless, a complete assessment of the fidelity level provided by the integrated virtual testing toolchain might also include investigating the sensor-grade realism offered by the virtual environment. That is how the simulation engine is capable of faithfully rendering real-world characteristics that are relevant for the perception systems but that might not necessarily match human vision peculiarities.

I. Lane Model Validation

Lane model validation is considered to provide a practical example on how the validation is performed as a part of the credibility assessment. Accurate representation of lane models are required for perception algorithm used for most lateral support systems e.g. lane keep assist, lane centering, lane change assist etc. In order to demonstrate that the lane models are fit for purpose we have used the processes defined in the credibility assessment. Vehicle dynamics is not considered during this process because the chassis dynamic will have negligible impact on the ability to detect the lane markings. The process consists of the following elements:

(a) Subsystem – camera model

(b) Sensor System – camera model with virtual lane markings.

(c) Integrated System – Lane detection algorithms

A. Camera Model Validation

Simulation needs to provide accurate image (intrinsic property) from the correct position (extrinsic property) for all cameras for a given scene. Specific intrinsic camera-related phenomena that should be considered during the validation include:

(a) Lens distortion: optical aberration due to projection;

(b) Vignette: darkening of the screen border;

(c) Grain jitter: white noise injection;

(d) Bloom: presence of fringes around bright areas;

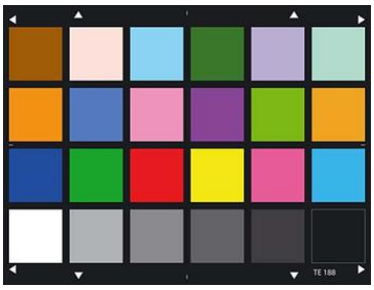
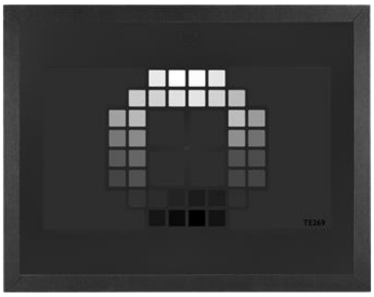
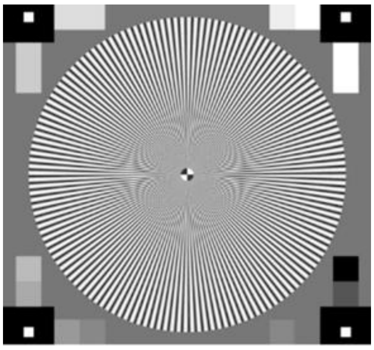
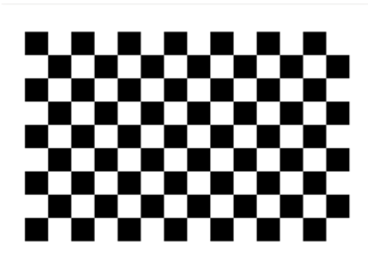
(e) Auto exposure: image gamma adaption to darker or brighter areas;

(f) Lens flares: reflection of bright objects on the lens;

(g) Depth of field: blurring of objects near or very far away of the camera;

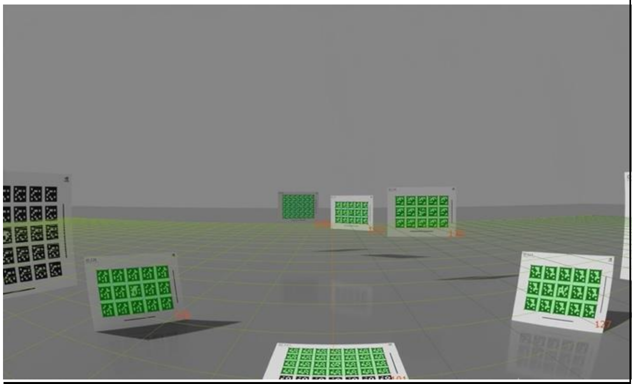
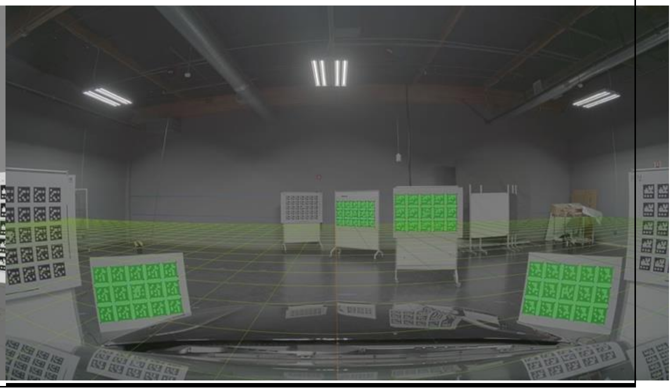
(h) Exposure time: shutter opening duration.

Below is a non-exhaustive list of tools that can be used to support the camera model validation.

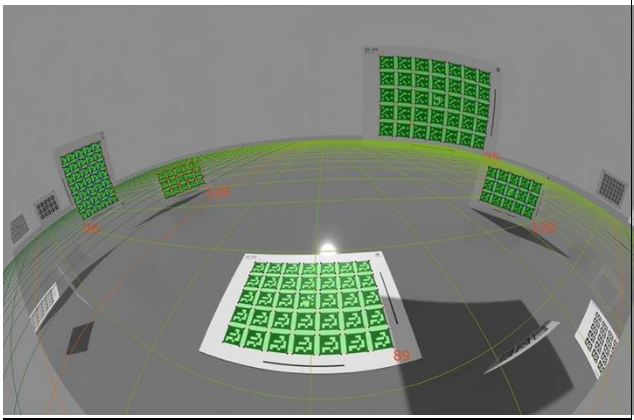
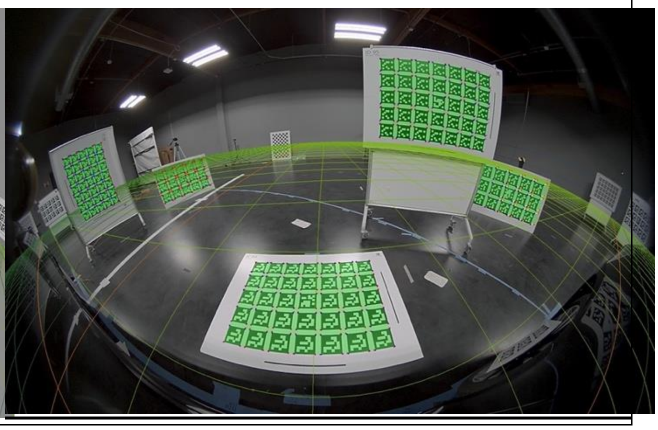
Tool	Image	Purpose
Macbeth Color chart Test		• To determine the camera color space of the camera • To determine the parameters for camera noise modelling • To learn about the exposure characteristics
OECF chart Tests		• Is designed for evaluating the opto-electronic-conversion-function of a camera.
SFR Chart		• To measure sharpness, contrast and lens effects
Lens Flare Characterization		• To differentiate the static and the dynamic components (dark shot noise) a video has to be recoded • To determine the lens characteristic for lens flares and ghosting artifacts
FTheta Calibration		• At every position, tilt the checkerboard target both horizontally and vertically up to 45 degrees • To determine the ftheta polynomial and to compare it with a more precise lens measurement

April Tags is a visual fiducial system, useful tool to supporting the validation of the extrinsic camera related properties. The tags provide a means of identification and 3D positioning, even in low visibility conditions. The tags act like barcodes, storing a small amount of information (tag ID), while also enabling simple and accurate 6D (x, y, z, roll, pitch, yaw) pose estimation of the tag.

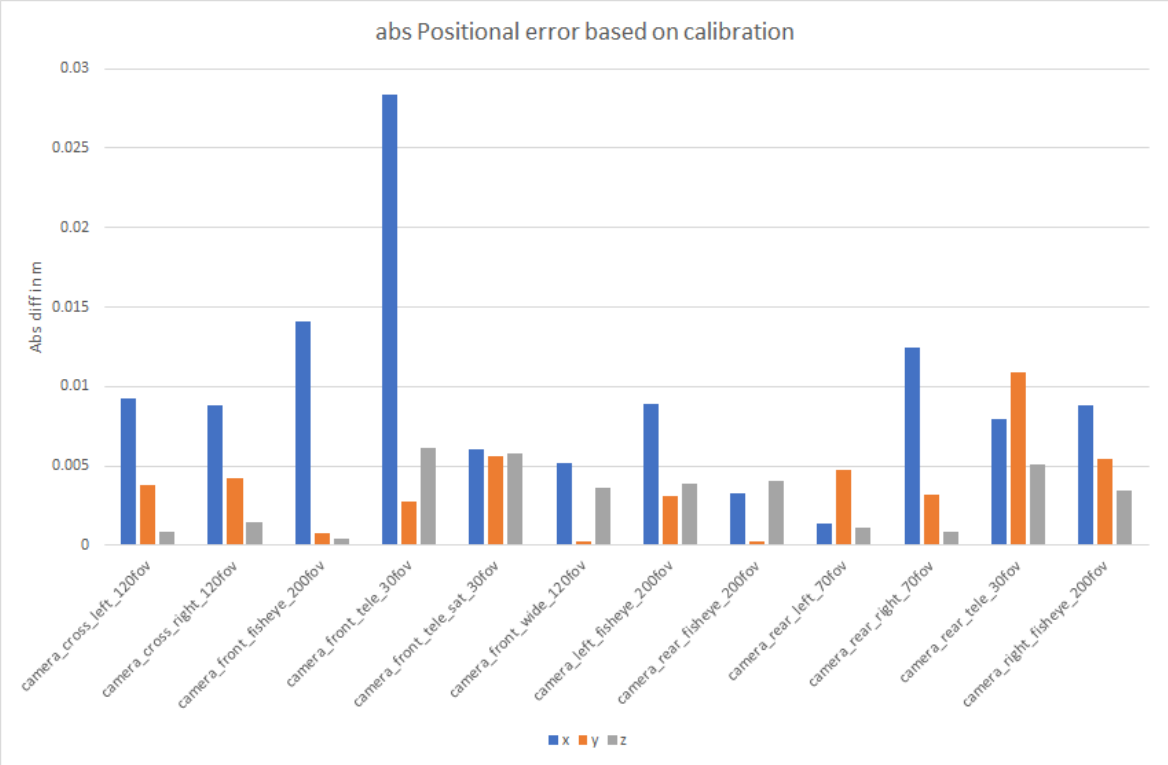
Camera front wide 120fov:

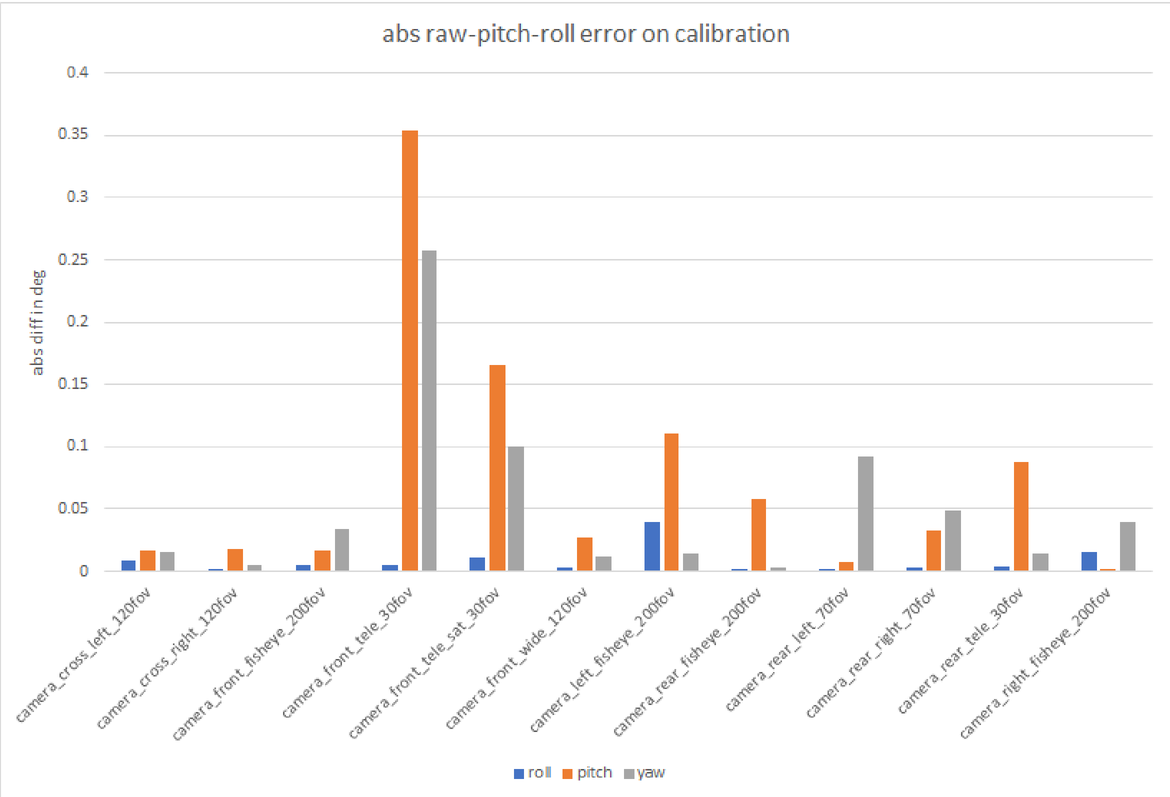
Simulation	Real
	

Camera_left_fisheye_200fov

Simulation	Real
	

The April tag chart positions and orientations are well constrained in the scene as they are visible from multiple cameras. Thresholds may be set on error derived from difference in absolute position / angle of the April Tags.





B. Sensor System Validation

The purpose of the sensor system validation is to demonstrate that camera models provide accurate results in the virtual environment which the system under test will be operated in. Pre-defined KPIs can be used to determine performance of the virtual sensor system. For the purpose of lane models the contrast ratio between lane marking and road surface is used to demonstrate the performance of the sensor system in both physical and virtual environments.

A simple framework for dividing sensor performance into several equivalence classes is shown here, as an example. The method relies on efficiently dividing the equivalence classes of the conditions that have a significant effect on a sensor performance metric, in this case, the brightness contrast ratio between the lane marking and the road surface. A requirement pattern can be formed that combines performance achievement with certain environmental or scenario-specific conditions.

A generic requirement pattern can be considered, as follows:

The {KPI} shall be {greater than} {KPI Threshold} if {Conditions Exist}.

The requirement pattern can be repeated with different conditions, as needed, in order to 1) fully cover all external conditions, including the extreme ends, and 2) define the boundary values at which performance requirements may change depending on the conditions, for example, relaxing the false-positive detection rate of a lane boundary if it is snowing. If this requirement pattern is well-defined across all possible conditions, independently verified, and has commitment from the developers to fulfill the requirements, then the problem of "functional insufficiencies" in sensor performance will likely be reduced or eliminated altogether.

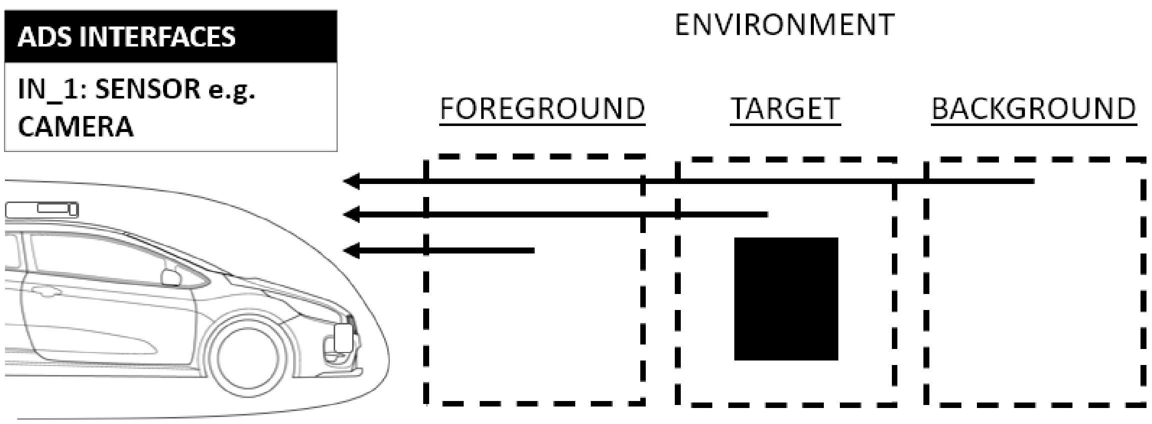
A method of division of the conditional classes follows this simple structure, as an example:

(a) Class 1: Nominal conditions - These are the ideal, best-case conditions.

(b) Class 2: Average conditions - These are the expected, real-world conditions that likely require significant development effort compared to Class 1 conditions, e.g. inner quartile range.

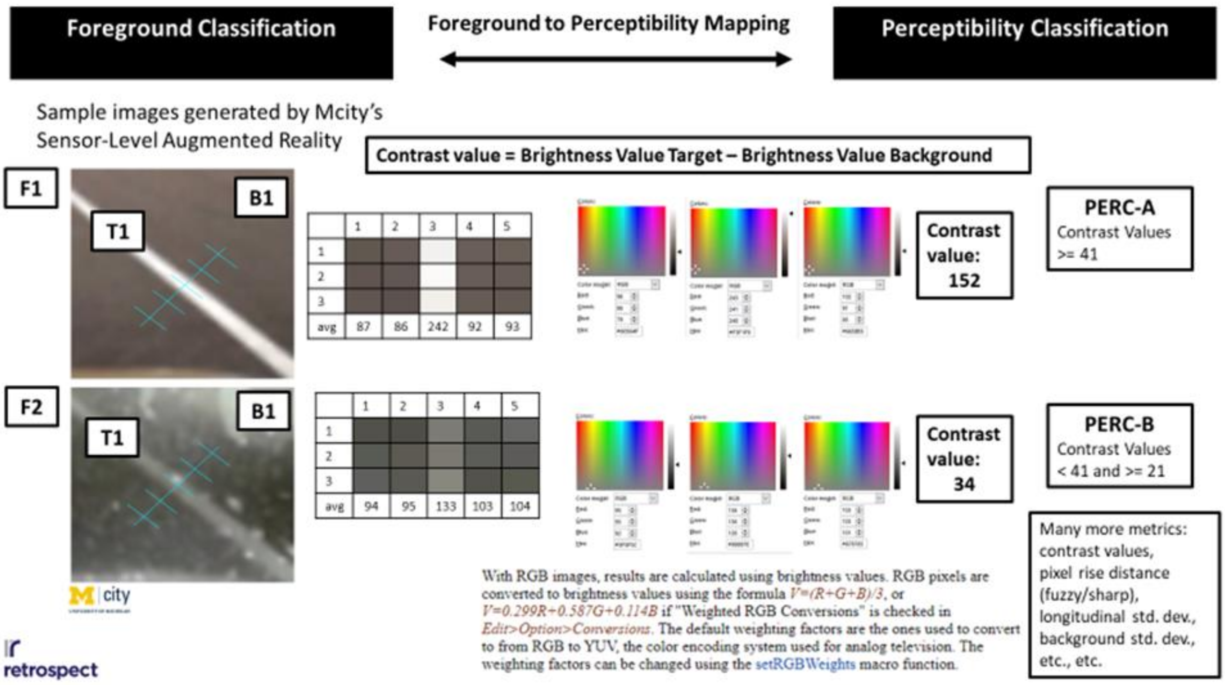
(c) Class 3: Worst-acceptable conditions - These are the worst conditions in which some level of performance will be guaranteed, e.g. 95'th percentile range. These likely require a tradeoff between the minimum required performance level and the remaining development effort. Beyond this class, no performance requirements are obligated. (Note, this can be tailored, as needed)

Finally, the conditions themselves may need to be separated into parameters according to the dependence or independence from each other into a minimal parameter set which adequately captures the environmental and scenario-specific conditions. For each sensing modality that has been considered so far, which includes: camera (visible light), RADAR, LiDAR, ultrasonic, and infrared cameras, the following generic sensor model has shown to be repeatable and useful in analyzing all environmental conditions for all sensing modes. It is broken down into three distinct parameters: Foreground, Target, and Background.



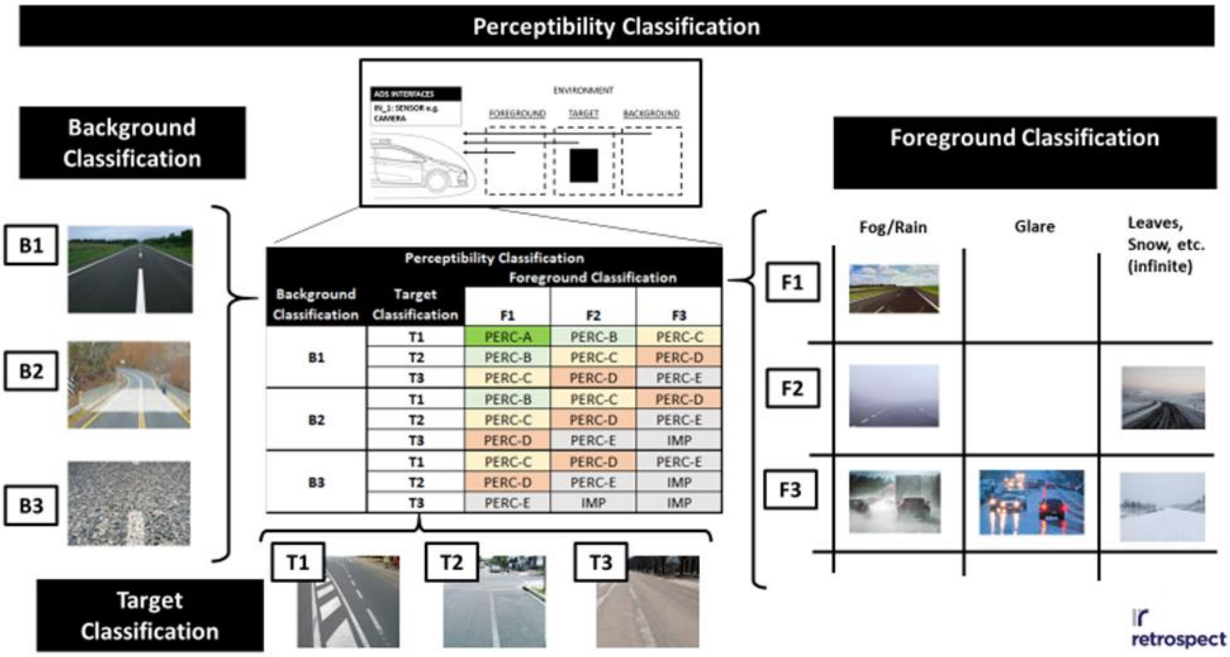
In the example of a camera-based system doing lane detection, the "Target" in this case would be the lane, itself. Many attributes may need to be developed to fully capture all the desired attributes of the Target, such as color, position, curvature, dash type, sharpness (or blurriness), etc. In this case, the attribute of interest is the contrast ratio of the lane with respect to the road surface. The road surface would be the "Background" in the sensing model. The “Foreground” could be anything between the sensor and the Target, such as fog, rain, or clear air, as well as, debris, objects, accumulated snow, etc.

In the following figure, a sample image from a sensor-level augmented reality demonstration is used to show how the Foreground, in this case snow, can alter the contrast value of the lane (Target) with respect to the road (Background). This highlights the usefulness of the conditional classification, as it can provide a consistent interface point between the infinite variations and combinations of environmental conditions and the finite set of the performance requirements that the developers must commit to deliver.



Internal or external classifications of the overall performance capability may (or may not) be useful to consider. For lane models, each parameter and their associated range of variables (Target, Foreground, and Background) should be considered during this phase. The sensor system should be qualified using known KPIs, such as very deterministic static scenarios, at first. This will allow the system to be validated against a measurable KPI. After that it can be extended to varying weather conditions etc. A test matrix can then be established that considers the variation of input parameters. Large variation in real and simulated results provide evidence where there may be limitations in the tool. Any sensor performance limitations should be noted during the assessment to put restrictions on what data can be generated to support the assessment of the ADS.

As performance limitations are encountered due to uncontrollable environmental conditions, the designers may either make reductions to a minimum performance level, as discussed above, or they may be able to strategically shift the “Target” in order to detect the environmental conditions, themselves. Considering the snowy example above, in light snow conditions the “Target” could be the lane, itself. However, in heavy snow conditions, the “Target” may be the heavy snow, itself, that is to say, the lane detection camera must detect heavy snow.



Background Classification	Target Classification	F1	F2	F3
B1	T1	PERC-A	PERC-B	PERC-C
	T2	PERC-B	PERC-C	PERC-D
	T3	PERC-C	PERC-D	PERC-E
B2	T1	PERC-B	PERC-C	PERC-D
	T2	PERC-C	PERC-D	PERC-E
	T3	PERC-D	PERC-E	IMP
B3	T1	PERC-C	PERC-D	PERC-E
	T2	PERC-D	PERC-E	IMP
	T3	PERC-E	IMP	IMP

It is up to the designers to determine what is useful for the overall system goals and the given technical capabilities, but the intent with the sensor system validation approach is to show how the designers and testers may fully specify performance and safety requirements in their development contracts, and show evidence of the fulfillment of their contracts. This can be done with little risk of finding out late in development that they are unable to meet the expected performance. Even if this should happen, the easiest solution may simply be to renegotiate the performance levels and communicate those to all stakeholders. The greatest source of risk would be in failing to write the requirements, in the first place, and plan for a "wait and see" approach with respect to sensor performance.

C. Integrated System Validation

Finally, the complete integrated system is tested. This includes the sensor system with the integrated perception algorithms. Simulated and real-world data are collected from the same environment and synchronized. State changes perception algorithms can then be compared to check if the simulated results match the real-world performance. The correlation threshold would determine if lane detection algorithms are used to support: LDW, LKAS or ADS.

Simulation	Real
	

After demonstrating that the lane model is accurate enough, the virtual testing tool can be used to support the assessment of lane detection algorithms. Virtual tests can be used to dramatically speed up the validation process and provide enough evidence that the system works as expected across the ODD. Once a base line correlation of the models and tool chain is achieved, the virtual testing tool can used to validate a large span of behaviors and confirm safe responses to unexpected situations. By applying variations and randomization of the different inputs, the system response is being tested across a wide range of scenarios and stimulus, and more confidence in its performance is gathered. The confidence can be reflected by coverage metrics (measured on the input data and/or ODD ranges), where higher measured coverage correlates to higher confidence in the system performance, as it was tested over wider set situations.

II. LiDAR/RADAR Model Validation

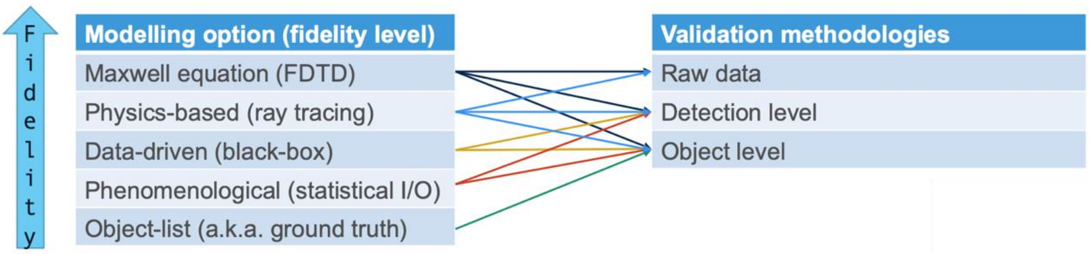
A. Modelling approaches

The LiDARs/RADARs modelling approaches can informally be divided into fidelity levels depending on the target application for the M&S. In particular, three reference classes [6] can be derived:

(a) “Low” fidelity models: retrieve the traffic objects’ list and status directly from the virtual environment ground-truth. This modelling paradigm does not afford statistical aspects related to the perception, such as false positives/negatives rate. Low fidelity models might however include basic sensor modelling such as accounting for the sensor’s Field of View (FoV) and occlusions to filter the whole object list;

(b) “Medium” fidelity model: similarly, to the low fidelity, medium fidelity models retrieve the objects’ status from the virtual environment kernel. Nonetheless, medium fidelity sensors introduce detection probability (false positive and false negative), the effect of objects’ shape and material on the detection, and environmental effects such as atmospheric degradation;

(c) “High” fidelity model: take advantage of advanced and computationally expensive rendering techniques to model physical processes happening in the real sensor. High fidelity sensors take as input the simulation rendered 3D environment following ray-tracing/rasterization. These sensor models are then allowed to operate with a similar input with respect to their physical counterparts.



Each fidelity level can be associated with a corresponding validation procedure. For instance, only “high” and “medium”-fidelity levels provide simulated raw-data that can be investigated against the real-world recording. Conversely, “low” fidelity model can only deliver information related to the object/detection level. Hence any validation procedure requiring raw data as an input cannot be embraced.

B. Metrics and KPIs for explicit LiDAR/RADAR Model Validation

The validation of a sensor model is concerned with establishing whether the developed sensor model is a viable solution for the purpose of performing ADS certification via virtual testing. “Explicit” validation techniques directly compare the direct output of the virtual model with respect to the real counterpart for the same set on input when applicable.

The ADS validation shall rely on the highest fidelity modelling approaches in virtual tests where the perception system plays a critical role. Hence, the annex is mainly concerned with the validation of “medium” and “high”-fidelity LiDAR/RADAR models. Such models are typically validated by exploiting the generated “point-clouds” (PC) or at the “Occupancy-Grid” (OG) level.

OGs are derived from the PCs where a cell (c_i) is assumed to be free ($c_i = 0$) or occupied ($c_i = 1$) if the probability of detecting an obstacle in the cell is greater than 0.5.

OGs deriving from simulation tests and real-world tests can be compared exploiting one of the following methods:

(a) OGs pixel-loss:

$$\sum_{x_c=0}^{\text{Width}} \sum_{y_c=0}^{\text{height}} \left| \text{sim}_{\text{grid}}(x_c, y_c) - \text{real}_{\text{grid}}(x_c, y_c) \right|;$$

(b) OGs Pearson correlation:

$$\frac{\left|\sum_{i=1}^{n_c}(c_{i,sim}-\bar{c}_{sim})(c_{i,real}-\bar{c}_{real})\right|}{\sqrt{\sum_{i=1}^{n_c}(c_{i,sim}-\bar{c}_{sim})^2 \sum_{i=1}^{m}(c_{i,real}-\bar{c}_{real})^2}};$$

(c) OGs ratio:

$$\frac{\sum_i^{N\ cell\ sim} c_j}{\sum_j^{N\ cell\ real} c_i};$$

As an alternative validation procedure, the virtual and real Point Clouds (PC) can be characterized taking advantage of a distance function, such as:

(a) PCs Euclidean distance:

$$D'_{pp} = \frac{1}{M}\sum_{m=1}^{M} \min_{1\le n\le N} \|p_{\text{sim}} - p_{\text{real}}\|;$$

(b) PCs Pearson correlation:

$$\frac{\left|\sum_{i=1}^{m}(x_{i,j}-\bar{x}_j)(y_i-\bar{y})\right|}{\sqrt{\sum_{i=1}^{m}(x_{i,j}-\bar{x}_j)^2 \sum_{i=1}^{m}(y_i-\bar{y})^2}}.$$

Based on the evidence provided in [7], [8], the following correlation thresholds have been reported in literature:

Metric	*Literature correlation*	*Optimum*
OG Pearson	0.59 – 0.76	1
OG ratio	0.2 – 0.5	1
PC Pearson	0.57 – 0.59	1

C. Implicit LiDAR/RADAR Model Validation

The perception system of an ADS is the element which acts as an interface between the simulation environment and the actual ADS. Thus, any information retrieved by the sensors is forwarded to the ADS. The validation of a sensor model shall then not disregard the impact that even small discrepancies between the real and virtual model can have on a complex system such as the ADS.

"Implicit" validation techniques establish the validity of the sensor model by including the perception algorithms [9] in the validation chain. The comparison is then carried out by establishing the difference between the simulation derived and real-world detected/tracked traffic objects.

The evaluation of implicit metrics can be carried out by directly compare the distance between the $(x,y)_{obj,sim}$ and $(x,y)_{obj,\quad real}$ coordinates of the tracked obstacles over the duration of the experiment using the techniques highlighted in Annex II. Alternatively, the Intersection-over-Union (IoU) metric

$$J(bb_{sim}, bb_{real}) = \frac{|bb_{sim} \cap bb_{real}|}{|bb_{sim} \cup bb_{real}|}$$

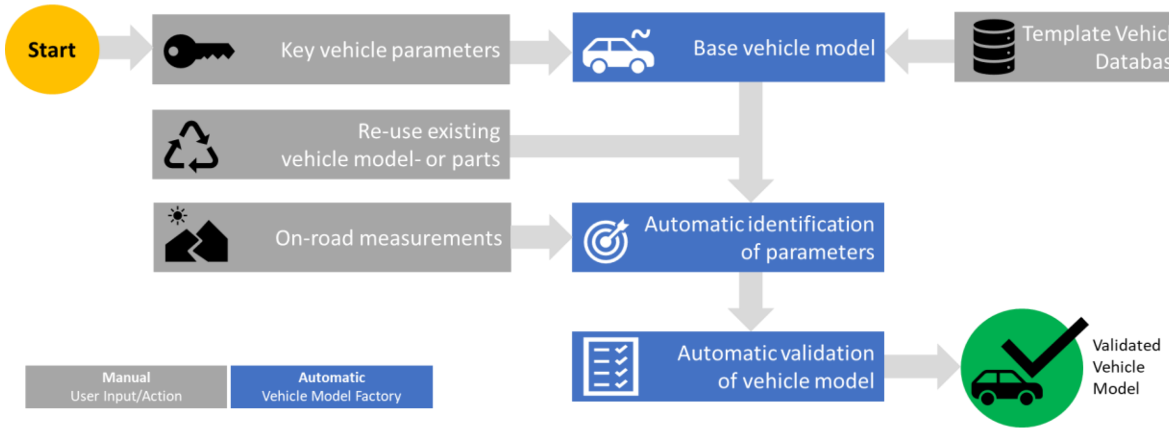
can be computed in case the detection layer returns bounding-boxes.

III. Vehicle Dynamics Model Validation

Beyond environment and sensor simulation described in the previous sections, vehicle dynamics simulation plays a major role in the virtual testing toolchain for certification and type approval. Building high fidelity vehicle dynamical models is a time- and cost-intensive activity since it requires precise component measurements (e.g. tire slips) to achieve the required accuracy quality on vehicle-level.

A. Vehicle Model Factory (VMF) Approach [10]

To reduce time, effort, and previous knowledge necessary to build up and validate a vehicle model, software tools can be used which guide the crafting high-fidelity models. For instance, the Vehicle Model Factory (VMF) approach requires 3 main inputs. Firstly, a minimal amount of key vehicle parameters, which can be obtained by vehicle-datasheets or workshop measurements (e.g.: vehicle corner weights, wheelbase, track-width, tire dimensions, etc.). Secondly, the test vehicle must be equipped with a minimum set of measurement- and recording equipment (CAN-access, inertial measurement unit, accelerometer, and GPS). The required instrumentation does not require any structural modifications on the vehicle. Finally, a set of predefined maneuvers must be performed on a test track.



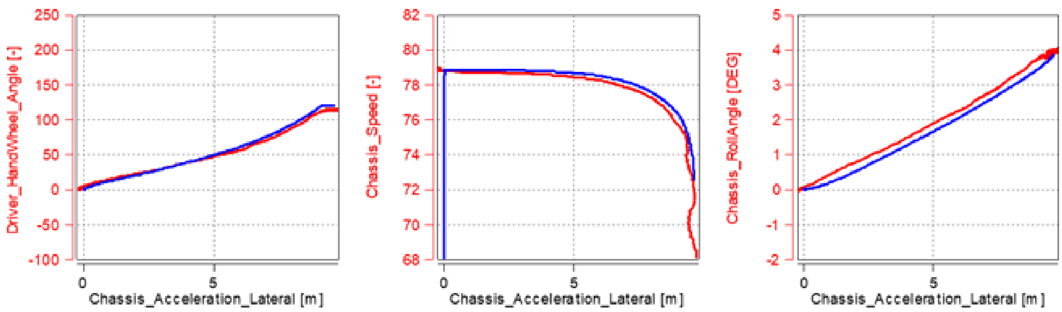
For the parameter identification task, groups are defined. For example:

(a) Driving resistance (Driving Resistance Coefficients),

(b) Weight distribution (Center-of-gravity position vertically and horizontally),

(c) Suspension (dynamic roll- and pitch behavior),

(d) Powertrain (motor/engine torque- and pedal-maps, gearbox- and total ratios).



Slowly Increased Steer (80kph)	Max. lateral acceleration [m/s²]	Roll Angle Gradient [deg/(m/s²)]	Pitch Angle Gradient [deg/(m/s²)] *Acceleration*	*Braking*
Road measurements	9.801	0.348	8.33	8.12
AVL VSM™	9.6	0.332	8.32	8.25
Accuracy	**97.9%**	**95.4%**	**98.8%**	**98.3%**

The software tool allows automatic identification of the parameters based on the collected dataset and on the chosen template model. After the parameters are identified, they are inserted into the vehicle model and a simulation for validation is performed automatically.

B. AEB Simulation example

For a virtual certification of Advanced Driver Assistant System (ADAS)/AD functions, validated vehicle dynamics models are crucial to achieve a maximum correlation between the virtual and the real-world system behavior. For instance, accurate tire and brake models enable realistic Automated Emergency Brake (AEB) testing in terms of deceleration and braking distance results. Validated suspension models will affect the virtual sensor output, such as RADAR, LiDAR or Camera, in a realistic manner, including e.g. pitching and rolling motion of the chassis. In general, a correlation analysis as part of creating a digital vehicle twin for virtual ADAS testing should be conducted at several levels:

(a) Vehicle dynamics behaviour;

(b) Vehicle geometrics (digital 3D twin);

(c) Environment sensor output (raw data);

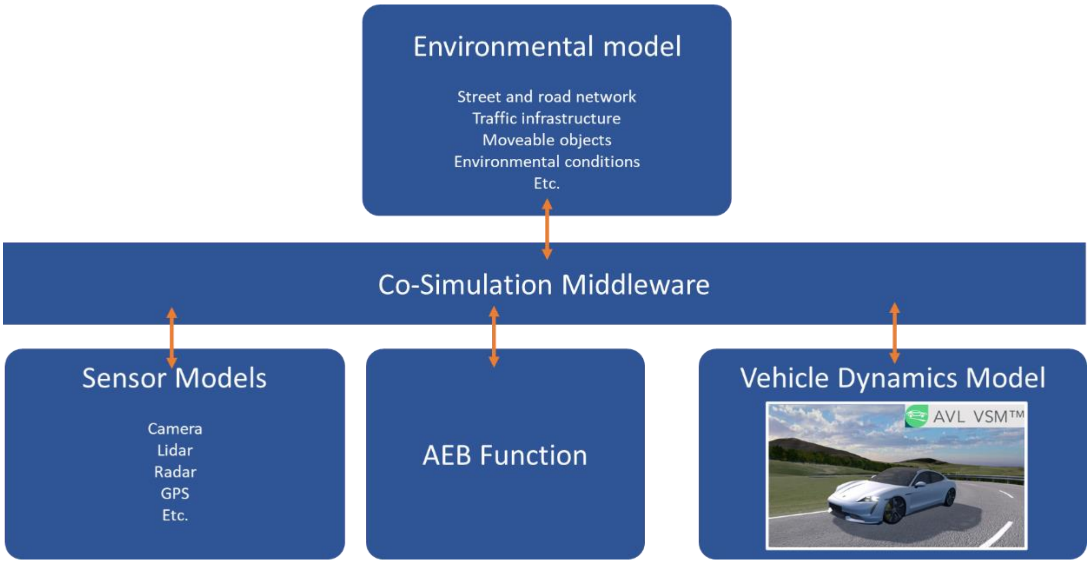
(d) Perception output (detected and classified objects);

(e) ADAS/AD controller behavior.

To showcase the importance of accurate vehicle dynamics models for ADAS testing, a simulation of an AEB function was conducted, and the results of two different virtual vehicle configurations of a limousine passenger car analyzed. Thereby configuration 1 featured summer tires and configuration 2 winter tires.

As part of a larger General Safety Regulation (GSR) scenario database, a simple AEB scenario was simulated, where the vehicle under test approaches a stationary Target Object Front (TOF) at varying velocities, namely 20 km/h, 40 km/h and 60 km/h. The AEB controller is based on Time to Collision (TTC) thresholds that initiate the different braking modes. The emergency braking is actuated as soon as the TTC exceeds the critical threshold.

In total, six simulations were conducted. Three for the configuration with summer tires, and three using winter tires, having a reduced longitudinal tire grip. The results show that winter tires lead to a substantially longer braking distance. For 60 km/h, the braking distance increases from 14.77 m to 17.51 m. The 20 km/h case resulted in a 2.2 m increase in braking distance when changing the tires.



Contrarily, the maximum longitudinal deceleration declines when changing from summer to winter tires. The different cases show a reduction in the range between 0.6 (for 20 km/h) and 1.8 m/s² (for 60 km/h). Although these results are no surprise and can be expected in real-world tests, the simulations demonstrate the effect of slight model variations on the AEB performance. In simulations with low-fidelity dynamics models, this effect might be overlooked, and critical scenarios or collisions might remain undetected.

IV. References

1. “proposed_cm-s-014_modelling_simulation_-_for_consultation.pdf.” Accessed: Jul. 30, 2021. [Online]. Available: https://www.easa.europa.eu/sites/default/files/dfu/proposed_cm-s-014_modelling_simulation_-_for_consultation.pdf

2. K. A. Maupin and L. P. Swiler, “Validation Metrics for Deterministic and Probabilistic Data,” p. 52.

3. M. A. Sprague and T. L. Geers, “A spectral-element method for modelling cavitation in transient fluid–structure interaction,” Int. J. Numer. Methods Eng., vol. 60, no. 15, pp. 2467–2499, Aug. 2004, doi: 10.1002/nme.1054.

4. “ARP5765B: Analytical Methods for Aircraft Seat Design and Evaluation - SAE International.” https://www.sae.org/standards/content/arp5765b/ (accessed Jul. 29, 2021).

5. J. P. C. Kleijnen, “1999: VALIDATION OF MODELS: STATISTICAL TECHNIQUES AND DATA AVAILABILITY,” p. 8.

6. B. Schlager et al., “State-of-the-Art Sensor Models for Virtual Testing of Advanced Driver Assistance Systems/ Autonomous Driving Functions,” vol. 3, no. 3, p. 30, 2020.

7. A. Schaermann, A. Rauch, N. Hirsenkorn, T. Hanke, R. Rasshofer, and E. Biebl, “Validation of vehicle environment sensor models,” in 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, Jun. 2017, pp. 405–411. doi: 10.1109/IVS.2017.7995752.

8. T. Hanke et al., “Generation and validation of virtual point cloud data for automated driving systems,” in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Oct. 2017, pp. 1–6. doi: 10.1109/ITSC.2017.8317864.

9. A. Ngo, M. P. Bauer, and M. Resch, “A Multi-Layered Approach for Measuring the Simulation-to-Reality Gap of Radar Perception for Autonomous Driving,” ArXiv210608372 Cs Eess, Jun. 2021, Accessed: Jun. 25, 2021. [Online]. Available: http://arxiv.org/abs/2106.08372

10. M. Oswald, “Vehicle Model Factory - Automatic Generation of Validated Virtual Prototypes,” presented at the Internationales Symposium für Entwicklungsmethodik, Wiesbaden, Nov. 2021.

Annex IV

Requirements for the audit of the safety design concept of the Automated Driving System (ADS)

General

The purpose of the audit of the safety by design concept of the ADS is to demonstrate that hazard and risks relevant for the ADS have been identified by the manufacturer and a consistent safety-by-design concept has been put in place to mitigate these risks. In addition, it should demonstrate that the risk assessment and the safety- by-design concept have been validated by the manufacturer through testing showing before the vehicle is placed on the market that the vehicle meets the safety requirements and in particular is free of unreasonable safety risks to the broader transport ecosystem in particular the driver, passengers and other road users.

I. ADS general description

A description should be provided which gives a simple explanation of the operational characteristics of the ADS and ADS feature:

(a) Operational Design Domain (Speed, road type, country, Environment, Road conditions, etc)/ Boundary conditions;

(b) Basic Performance (e.g. Object and Event Detection and Response (OEDR) etc.);

(c) Interaction with other road users;

(d) Main conditions for Minimum risk manoeuvres;

(e) Interaction concept with the driver (if relevant);

(f) Supervision center (if relevant));

(g) The means to activate, override or deactivate the ADS by the driver (if relevant) or the human supervision center (if relevant), passengers (if relevant) or other road users (if relevant).

II. Description of the functions of the ADS

A description should be provided which gives a simple explanation of all the functions including control strategies of "The ADS" and the methods employed to perform the dynamic driving tasks within the ODD and the boundaries under which the ADS is designed to operate, including a statement of the mechanism(s) by which control is exercised.

A list of all input and sensed variables should be provided and the working range of these defined, along with a description of how each variable affects system behaviour.

A list of all output variables which are controlled by "The ADS" should be provided and an explanation given, in each case, of whether the control is direct or via another vehicle system. The range of control (paragraph 2.7.) exercised on each such variable should be defined.

III. ADS layout and schematics

A. Inventory of components.

A list should be provided, collating all the units of "The ADS" and mentioning the other vehicle systems which are needed to achieve the control function in question.

An outline schematic showing these units in combination, should be provided with both the equipment distribution and the interconnections made clear.

This outline should include:

(a) Perception and objects detection including mapping and positioning;

(b) Characterization of Decision-making;

(c) Remote supervision and remote monitoring by a remote supervision centre (if applicable);

(d) information display / user interface;

(e) The data storage system (DSSAD).

B. Functions of the units

The function of each unit of "The ADS" should be outlined and the signals linking it with other units or with other vehicle systems should be shown. This may be provided by a labelled block diagram or other schematic, or by a description aided by such a diagram.

Interconnections within "The ADS" should be shown by a circuit diagram for the electric transmission links, by a piping diagram for pneumatic or hydraulic transmission equipment and by a simplified diagrammatic layout for mechanical linkages. The transmission links both to and from other systems should also be shown.

There should be a clear correspondence between transmission links and the signals carried between Units. Priorities of signals on multiplexed data paths should be stated wherever priority may be an issue affecting performance or safety.

C. Identification of units

Each unit should be clearly and unambiguously identifiable (e.g. by marking for hardware, and by marking or software output for software content) to provide corresponding hardware and documentation association. Where software version can be changed without requiring replacement of the marking or component, the software identification must be by software output only.

Where functions are combined within a single unit or indeed within a single computer, but shown in multiple blocks in the block diagram for clarity and ease of explanation, only a single hardware identification marking should be used. The manufacturer should, by the use of this identification, affirm that the equipment supplied conforms to the corresponding document.

The identification defines the hardware and software version and, where the latter changes such as to alter the function of the unit as far as this Regulation is concerned, this identification should also be changed.

D. Installation of sensing system components

The manufacturer should provide information regarding the installation options that will be employed for the individual components that comprise the sensing system. These options should include, but are not limited to, the location of the component in/on the vehicle, the material(s) surrounding the component, the dimensioning and geometry of the material surrounding the component, and the surface finish of the materials surrounding the component, once installed in the vehicle. The information should also include installation specifications that are critical to the ADS's performance, e.g. tolerances on installation angle.

Changes to the individual components of the sensing system, or the installation options, should be updated in the documentation.

E. ADS specifications

Description of ADS specifications in Normal and Emergency Conditions, the acceptability criteria and the demonstration of compliance with those criteria.

List of applied regulations, codes and standards.

F. Safety Concept and validation of the safety concept by the manufacturer

The manufacturer should provide a statement which affirms that the "The ADS" is free from unreasonable risks for the driver (if applicable), passengers and other road users.

In respect of software employed in "The ADS", the outline architecture should be explained and the design methods and tools used should be identified (see 3.5.1). The manufacturer should show evidence of the means by which they determined the realization of the ADS logic, during the design and development process.

The manufacturer should provide an explanation of the design provisions built into "The ADS" so as to ensure functional and operational safety. Possible design provisions in "The ADS" are for example:

(a) Fall-back to operation using a partial system;

(b) Redundancy with a separate system;

(c) Removal of the automated driving function(s).

If the chosen provision selects a partial performance mode of operation under certain fault conditions (e.g. in case of severe failures), then these conditions should be stated (e.g. type of severe failure) and the resulting limits of effectiveness defined (e.g. initiation of a minimum risk manoeuvre immediately) as well as the warning strategy to the driver/remote supervision center (if applicable).

If the chosen provision selects a second (back-up) means to realize the performance of the dynamic driving task, the principles of the change-over mechanism, the logic and level of redundancy and any built in back-up checking features should be explained and the resulting limits of back-up effectiveness defined.

If the chosen provision selects the removal of the automated driving function, this should be done in compliance with the relevant provisions of this Regulation. All the corresponding output control signals associated with this function should be inhibited.

The documentation should be supported, by an analysis which shows, in overall terms, how the ADS will behave to mitigate or avoid hazards which can have a bearing on the safety of the driver (if applicable), passengers and other road users. It should show how unknown hazardous scenarios will be managed by the manufacturer in order to keep the residual level or risk under control.

The chosen analytical approach(es) should be established by the manufacturer and made available to the relevant authority should before market introduction.

The auditor should perform an assessment of the application of the analytical approach(es):

(a) Inspection of the safety approach at the concept (vehicle) level. This approach should be based on a Hazard / Risk analysis appropriate to system safety.

(b) Inspection of the safety approach at the ADS level including a top down (from possible hazard to design) and bottom up approach (from design to possible hazards). The safety approach may be based on a Failure Mode and Effect Analysis (FMEA), a Fault Tree Analysis (FTA) and a System-Theoretic Process Analysis (STPA) or any similar process appropriate to system functional and operational safety.

(c) The documentation should demonstrate the validation/verification plans and results including appropriate acceptance criteria. This should include validation testing appropriate for validation, for example, Hardware in the Loop (HIL) testing, vehicle on-road operational testing, testing with real end users, or any other testing appropriate for validation/verification.

Results of validation and verification may be assessed by analyzing coverage of the different tests and setting coverage minimal thresholds for various metrics.

The documentation should confirm that at least each of the following items is covered where applicable under (a)-(c):

(i) Issues linked to interactions with other vehicle systems (e.g. braking, steering);

(ii) Failures of the automated driving system and system risk mitigation reactions;

(iii) Situations within the ODD when a system may create unreasonable safety risks for the driver (if applicable), passengers and other road users due to operational disturbances (e.g. lack of or wrong comprehension of the vehicle environment, lack of understanding of the reaction from the driver(if applicable), passenger or other road users, inadequate control, challenging scenarios)

(iv) Identification of the relevant scenarios within the boundary conditions and management method used to select scenarios and validation tool chosen.

(v) Decision making process resulting in the performance of the dynamic driving tasks (e.g. emergency manoeuvres), for the interaction with other road users and in compliance with traffic rules

(vi) Cyber-attacks having an impact on the safety of the vehicle.

(vii) Reasonably foreseeable misuse by the driver (if applicable) (e.g. driver availability recognition system and an explanation on how the availability criteria were established), mistakes or misunderstanding by the driver if applicable (e.g. unintentional override) and intentional tampering of the ADS.

The documentation should establish that argumentation supporting the safety concept is understandable and logical and implemented in the different functions of the ADS.

The documentation should also demonstrate that validation plans are robust enough to demonstrate safety (e.g. reasonable coverage of chosen scenarios testing by the validation tool chosen) and have been completed.

The documentation should demonstrate that the vehicle is free from unreasonable risks for the driver (if applicable); vehicle occupants and other road users in the operational design domain and the method, i.e. through:

(a) An overall validation target (i.e., validation acceptance criteria) supported by validation results, demonstrating that the entry into service of the automated driving system will overall not increase the level of risk for the driver (if applicable), vehicle occupants, and other road users compared to a manually driven vehicles; and

(b) A scenario specific approach showing that the ADS will overall not increase the level of risk for the driver (if applicable), passengers and other road users compared to a manually driven vehicles for each of the safety relevant scenarios;

The documentation should allow the certification Authority to tests to verify the safety concept.

The documentation should itemize the parameters being monitored and should set out, for each failure condition of the type defined in paragraph 3.4.4. of this annex, the warning signal to be given to the driver (if applicable) /vehicle occupants/other road users and/or to service/technical inspection personnel.

This documentation should also describe the measures in place to ensure the "The ADS" is free from unreasonable risks for the driver (if applicable), vehicle occupants, and other road users when the performance of "The ADS" is affected by environmental conditions e.g. climatic, temperature, dust ingress, water ingress, ice packing.

G. Data Storage System

The documentation should describe:

(a) Storage location and crash survivability;

(b) Data recorded during vehicle operation and occurrences;

(c) Data security and protection against unauthorized access or use;

(d) Means and tools to carry out authorized access to data.

H. Cyber security

The documentation should describe:

(a) Cyber security and software update management;

(b) Identification of risks, mitigation measures;

(c) Secondary risks and assessment of residual risks;

(d) Software update procedure and management put in place to comply with legislative requirements.

[I. Information provisions to users

The documentation should describe:

(a) Model of the information provided to users (including expected driver’s tasks within the ODD and when going out of the ODD;

(b) Extract of the relevant part of the owner`s manual.]

J. Safety management system

The manufacturer should have a valid Safety Management System relevant to the ADS concerned and should inform of any change that will affect the relevance of the safety management system for the ADS concerned.

K. Type of documentation to be provided

The manufacturer should provide a documentation package which gives access to the basic design of "ADS" and the means by which it is linked to other vehicle systems or by which it directly controls output variables.

The function(s) of "ADS", including the control strategies, and the safety concept, as laid down by the manufacturer, should be explained.

Documentation should be brief, yet provide evidence that the design and development has had the benefit of expertise from all the ADS fields which are involved.

For periodic technical inspections, the documentation should describe how the current operational status of "The ADS" can be checked.

Information about how the software version(s) and the failure warning signal status can be readable in a standardized way via the use of an electronic communication interface, at least be the standard interface (OBD port).

The documentation package should show that the "ADS":

(a) Is designed and was developed to operate in such a way that it is free from unreasonable risks for the driver (if applicable), passengers and other road users within the declared ODD and boundaries;

(b) Respects, under the performance requirements specified elsewhere by FRAV;

(c) Was developed according to the development process/method declared by the manufacturer.

Documentation should be made available in three parts:

(a) An information document which is submitted to the authority should contain brief information on the items.

(b) The formal documentation package annexed to the information document, which should be supplied to the certification Authority for the purpose of conducting the safety assessment.

(c) Additional confidential material and analysis data (intellectual property) which should be retained by the manufacturer, but made open for inspection (e.g. on-site in the engineering facilities of the manufacturer) at the time of the product assessment / process audit. The manufacturer should ensure that this material and analysis data remains available for a period of 10 years counted from the time when production of the ADS is definitely discontinued.

Any changes to ADS safety design should be communicated as required to the relevant authority.

Annex V

Requirements for the Audit of the manufacturer safety management system for Automated Driving Systems (ADS)

General

The purpose of the audit of the safety management system of the manufacturer is to demonstrate that the manufacturer has robust processes to manage safety risks and to ensure safety throughout the ADS lifecycle (development phase, production, but also operation on the road and decommissioning). It should include taking the right measures to monitor the vehicle in the field and to take the right action when necessary.

I. Safety management system

The documentation provided by a manufacturer should demonstrate that a safety management system that provides effective processes, methodologies and tools is in place, is up to date, and is being followed within the organization to manage safety and continued compliance throughout the product lifecycle (design, development, production, operation including respect of traffic rules, and decommissioning).

Guidelines:

Note: Safety risk management is a core activity that supports the safety management system and also contributes to the effectiveness of other organizational processes. The term safety risk management, as opposed to the more generic term risk management, is meant to restricts itself to the management of safety risks.(e.g. without considering financial risk, legal risk, economic risk and so forth)

The control of risk can be achieved addressing 3 critical dimensions:

(a) Human component thanks to people with appropriate skills, training and motivation;

(b) Organisational component consisting of procedures and methods defining the relationship of tasks; and

(c) Technical component by using appropriate tools and equipment.

The establishment of an adequate SMS serves to monitor and improve all three dimensions and control the relevant risks. The Safety Management System (SMS) evaluation is based on automotive engineering standards, guidebooks and best practice documents relevant to safety.

The product operational risks should be specifically addressed in the Design and Development chapter and implemented in the product assessment. Thus, this section should show the link between the overall risk management process (as per this chapter) and product operational risks.

Examples of processes and aspects to be documented:

(a) Risk identification (in line with ISO 3100 para. 6.4.2 or equivalent standard);

(b) Risk analysis (in line with ISO 3100 para. 6.4.3 or equivalent standard);

(c) Risk evaluation (in line with ISO 3100 para. 6.4.4 or equivalent standard);

(d) Risk treatment (in line with ISO 3100 para. 6.4.5 or equivalent standard), including:

(e) Processes used for keeping the risk assessments as current as possible;

(f) Safety performance of the organization and effectiveness of safety risk controls.

Examples of processes and aspects to be documented:

(a) Safety governance:

(i) Safety policies and principles (in line with the concept stated in ISO 21434 5.4.1 and ISO 9001 Automotive 5.2, but from safety perspective)

(ii) Management commitment (in line with the concept stated in ISO 21434 para. 5.4.1 and ISO 9001 Automotive 5.1, but from safety perspective)

(iii) Roles and responsibilities (ISO 26262-2 para. 6.4.2, this relates to the organizational as well as to the project dependent activities)

(b) Safety culture (ISO 26262-2 para. 5.4.2.);

(c) (Periodic) internal and external audit to ensure that all SMS processes are implemented consistently (UN R.157 para. 3.5.5, ISO 26262-2 para. 6.4.11.);

(d) Effective communications within the organization (ISO 26262-2 para. 5.4.2.3.);

(e) Information sharing outside of the organization (in line with the concept stated in ISO 21434 para. 5.4.5. and ISO 9001, but from safety perspective);

(f) Quality management system (e.g. IATF 16949) to support safety engineering, including change management, configuration management, requirement management, tool management etc.

Examples of processes and aspects to be documented to ensure the robustness of the production phase:

(a) Quality Management System accreditation (e.g. as per IATF 16949 or ISO 9001 or equivalent);

(b) A general description of the way in which the organization performs all the production functions including management of working conditions and the environment and equipment and tools.

Examples of processes and aspects to be documented to assure robustness of distributed production:

(a) Liaison between the vehicle manufacturer and all other organizations (partners or subcontractors) involved in the production of the system/vehicle;

(b) Criteria for the for the acceptability of “subsystem/components” manufactured by other partners or subcontractors. (i.e. deployment of production assurance requirements to supply chain).

The design and development process should be established and documented including risk management, requirements management, requirements’ implementation, testing, failure tracking, remedy and release

Guidelines:

Examples of processes and aspects to be documented to ensure the robustness of the design and development phase:

(a) A general description of the way in which the organization performs all the design and development activities;

(b) Vehicle\system development, integration and implementation:

(i) Requirements management (e.g. Requirement capture and validation)

(ii) Validation strategies, including but not limited to:

- Credibility assessment for simulation (link to Subgroup 2);
- System Integration level;
- Software level;
- Hardware level.

(iii) Management of functional Safety and SOTIF, including the continuing evaluation and update of risk assessments and relationship with In-Service Safety.

(c) Management of design changes and changes to design and development processes.

The manufacturer should institute and maintain effective communication channels between manufacturer departments responsible for functional/operational safety, cybersecurity and any other relevant disciplines related to the achievement of vehicle safety.

Guidelines:

Examples of processes and aspects to be documented to assure that responsibilities are properly discharged:

(a) Roles and responsibilities during the design and development;

(b) Qualifications and experience of persons responsible for making decisions affecting safety;

(c) Coordination between design and production.

The manufacturer should have processes to monitor safety-relevant incidents/ crashes/collisions caused by the engaged automated driving system and a process to manage potential safety-relevant gaps post-registration (closed loop of field monitoring) and to update the vehicles. They should have processes to report critical incidents (e.g. collision with another road users and potential safety-relevant gaps) to the relevant Authority when critical incidents occur.

Guidelines: Link with the in-service monitoring/reporting pillar. The manufacturers should set up process for the operational phase for confirmation of compliance to the safety requirements in the field, early detection of new unknown scenarios (in line with SOTIF safety development goal to minimize the unknown scenarios area), event Investigation, to share learnings derived from incidents and near-miss analysis to allow the whole community to learn from operational feedback and to contribute to the continuous improvement of Automotive Safety

Example of guiding principles: Is there a document describing the appropriate procedure of reporting incidents to the management? Is there evidence that the company is complying with that procedure? Is there a document describing the appropriate procedure of investigation and documentation of incidents? Is there evidence that the company is complying with that procedure?

The manufacturer should demonstrate that periodic independent internal process audits are carried out to ensure that the processes established according with paragraphs 1. to 4. are implemented consistently.

Guidelines:

Examples of processes and aspects to be documented to assure independent design audit and assessment:

(a) Assurance that all practices and procedure to be applied during the vehicle\system development are followed. (process assurance);

(b) Assurance an independent checking for the compliance with the applicable requirements and regulations. (Independent assessment from person not creating the compliance data);

(c) Process to assure the continuing evaluation of the Safety management system in order to ensure that it remains effective. (system audit that can be undertaken by the existing Quality Management System).

Manufacturers should put in place suitable arrangements (e.g. contractual arrangements, clear interfaces, quality management system) with suppliers to ensure that the supplier safety management system comply with the requirements of paragraphs.1. (except for vehicle related aspects like "operation" and "decommissioning"), 2,.3 and 5

Guidelines:

Examples of processes and aspects to be documented:

(a) Organizational policy for supply chain;

(b) Incorporation of risks originating from supply chain;

(c) Evaluation of supplier SMS capability and corresponding audits;

(d) Processes to establish contracts, agreements for ensuring safety across the phases of development, production and post-production;

(e) Processes for distributed safety activities.

Expiration/renewal of the SMS

Documentation should be regularly updated in line with any relevant changes to the SMS processes. Any changes to SMS documentation should be communicated as required to the relevant authority.

Annex VI

Draft requirements for In-Service Monitoring and Reporting for the Automated Driving System (ADS)

(forthcoming)