Virtual ITS European Congress, 9-10 November 2020

Paper number ITS-19700

Common Methodology for Data-Driven Scenario-Based Safety Assurance in the HEADSTART Project

Nicolas Wagener, M.Sc.[1*]; Dipl.-Ing. Patrick Weißensteiner, B.Sc.[2]; Jean-Baptiste Coget, M.Sc. [3]; Univ.-Prof. Dr.-Ing. Lutz Eckstein [1] Dr. Annie Bracquemond [3];

1. Institute for Automotive Engineering RWTH Aachen University, Germany

2. VIRTUAL VEHICLE Research GmbH, Austria

3. Institut VEDECOM, France

* Nicolas.Wagener@ika.rwth-aachen.de

Abstract

One objective of introducing Connected and Automated Driving (CAD) functions on the roads is to reduce the number of accidents due to human errors by reducing the tasks of the driver (partial automation) or removing it completely from the driving system. Building safe and reliable automated vehicles require specific testing methods that are adapted to higher levels of automation. Harmonised European Solutions for Testing Automated Road Transport (HEADSTART) is a research project funded by the European Union that aims to define testing and validation procedures for CAD functions. HEADSTART brings a methodology for testing and validating these functions with Data-Driven Scenario-Based Safety Assurance. The goal of this paper is first to present the overall HEADSTART methodology for validating CAD safety, and then further explain three aspects of it, such as: The database mechanics to extract and parametrize logical scenarios, the relevance metrics for the selection of scenarios, and the allocation of scenarios for test execution.

Keywords:

Safety Assurance, Automated Driving, Validation

Motivation

With autonomous driving on the rise for SAE Level 3+ systems being developed by the time this paper has been written, new methods for the validation and verification of these systems are needed. Due to the fact, that the driver will not necessarily control and monitor the vehicle at all times when using a SAE Level 3+ driving function, advanced methods for the safety assurance and the validation of the automated driving need to be developed. Situations which driving functions need to handle safely are sheer endless in their variability. Thus, classical approaches like [5] cannot be conducted due to the sheer number of driven kilometres which are necessary, shown by example calculations [6]. Putting a driving function on the road for millions of kilometres, possibly for each new update it receives, is not feasible.

There are several approaches and projects, like the German PEGASUS and VVMethods projects, the

French MOOVE project or the Japanese SAKURA project, on national and international levels to tackle this problem [1][7][8][14][9]. The shared idea of these initiatives is to build up a consolidated database of driving scenarios and test only relevant parts in simulations and on proving grounds. However, there are many different projects from different countries and stakeholders but the validation and verification is a problem, which needs to be solved on a global level. Therefore, the European Commission funded the HEADSTART project to identify and harmonize a methodology for the validation and verification of automated driving functions, taking the needs of various stakeholders into account.

The HEADSTART Project

The results and methods presented in this paper were created within the context of the European research project HEADSTART. The HEADSTART project is a European project to develop and define a harmonized validation methodology for connected and automated driving functions. There are several countries all over Europe by 17 partners involved, including Germany, Spain, France, Sweden, Belgium, Greece and the Netherlands. To find a harmonized European solution for testing automated road transport, the HEADSTART project will define testing and validation procedures of connected and automated driving functions including its key enabling technologies (i.e. communications, cyber-security, positioning) by cross-linking of all test instances such as simulation, proving ground and real world field tests to validate safety and security performance according to the needs of key user groups (technology developers, consumer testing groups and type approval authorities). This paper is based on the results of the workpackage on defining a common methodology. The methodology was compiled on the basis of a detailed state-of-the-art analysis. Moreover, a global network of experts and stakeholders is constantly involved in the development process by expert workshops and interviews.

Common Methodology

This chapter introduces the overall methodology, which was defined in the HEADSTART project. The methodology is based on knowledge and insights derived from a state-of-the-art and user needs analysis. As stated before, especially, the analysis of state-of-the-art methodologies from ongoing and finished projects including but not limited to the German PEGASUS project [8], the French MOOVE [14,15] project and the Dutch TNOStreetWise project [10] were taken into account. Finally, new concepts were developed and harmonized among the partners of the HEADSTART project.

The main idea of the project, based on state-of-the-art approaches, is to utilize a database of relevant scenarios as image of reality for testing automated driving functions. These scenarios can be rolled out to different test capabilities, enabling a safe evaluation and validation method for DIN SAE Level 3 to 5 driving functions. An overview on the overall methodology is given in Figure 1 and described in the following. The selection of relevant scenarios and the allocation of scenarios is presented in more detail in subsequent chapters.

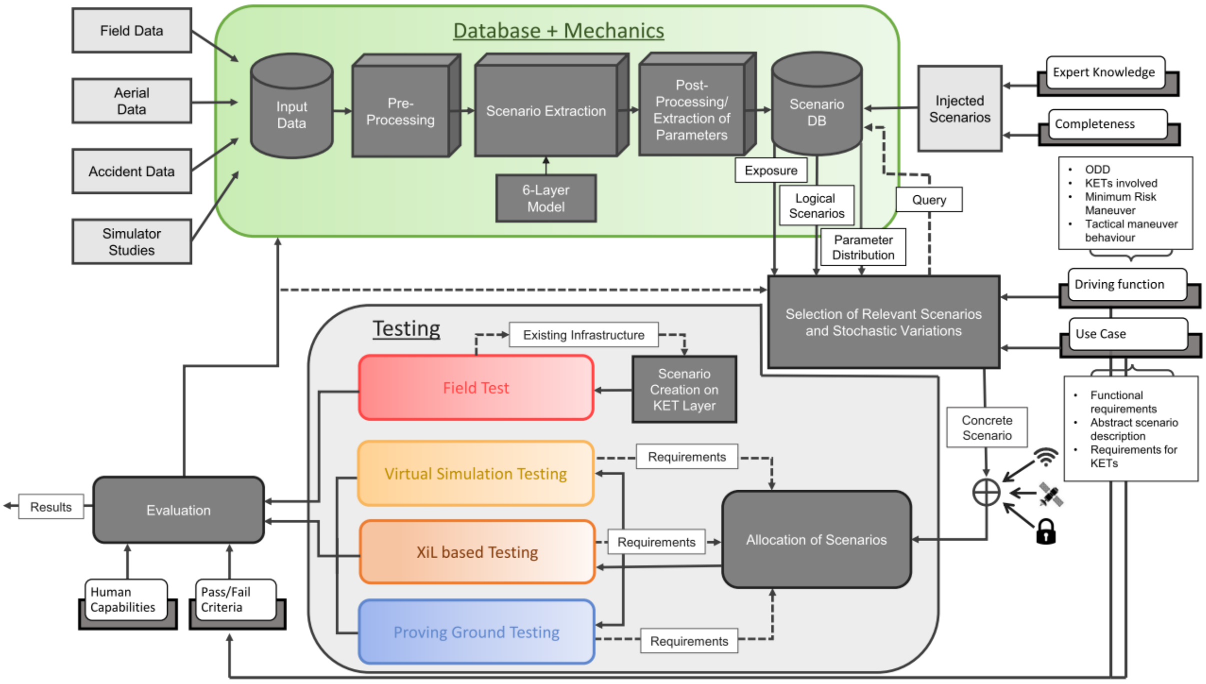

Figure 1 – Overview on the HEADSTART Methodology

The HEADSTART project is based on a data-driven scenario-based approach. Thus, input data is needed. To have a good representation of reality as data-basis, different sources are taken into account. This can be field data from field operational tests (FOT) or naturalistic driving studies (NDS). Moreover, aerial data from drones, such as the highD dataset [11], infrastructure data and accident data can be used as input. It was identified, that nearly all state-of-the-art approaches utilize similar database mechanics to extract scenarios from the data and maintain a scenario database [15]. Scenario databases emerging from large projects are currently on the market. The implementation of such database is out of scope of HEADSTART. Thus, only a general core functionality is described in this paper. Moreover, there are several scenario concepts, which scenario are derived and extracted by the respective database. For sake of brevity these concepts will not be included in this paper either.

Since, data sources are finite, there might be important scenarios, which did not occur in recorded data. Moreover, there can be relevant scenarios with a low occurrence rate, which are, nevertheless, very important to be tested. An example for this could be a pedestrian on the highway. It is very unlikely that this scenario occurs, however, a driving function should be able to handle such scenario. Finally, there might be scenarios or situations, which do not occur currently on open roads, but will be introduced by automated driving functions. Thus, an injection of scenarios, created by experts or completeness methodologies, is necessary. The injection of scenarios is out of scope of the HEADSTART methodology and will therefore not be described in detail.

The output of such scenario database is, based on a query to select relevant scenarios. Of course, the selection of specific scenarios behoves to the OEM or developer of the driving function. Nevertheless, the HEADSTART approach introduces several characteristics for that process. The specific methodology is presented in subsequent sections. The output itself is a set of logical scenarios with their respective parameter distributions. Logical scenarios and their parameters are not specified within HEADSTART. There are several projects defining logical scenarios like [2]. Moreover, the exposure, severity and controllability, with focus on the exposure are a mandatory output of the database. The exposure is a necessary input, amongst others, for the determination of relevant scenarios and stochastic variations within the given parameter distributions. This shall be based on the driving function and the use-case. The selection of relevant scenarios and parameter variations is described in more detail in the next section.

The output of such selection and variation is a concrete, parametrized scenario, in HEADSTART described as OpenSCENARIO [12] and OpenDRIVE [13] files. HEADSTART tries to build on open and widespread standards to maintain a good compatibility for all users. Both standards are maintained by the ASAM consortium.

Within HEADSTART three key enabling technologies (KETs) are in focus. Namely these are communication, positioning and cyber-security. The general derivation of concrete scenarios and test cases is based on existing approaches and does not include explicitly these KETs. Therefore, scenarios are enriched by specific KET extensions. For sake of brevity, the HEADSTART KETs will not be described in detail in this paper.

The next step in the HEADSTART methodology is the allocation of scenarios. It needs to be decided where the test is carried out, depending on the scenario, the use-case, the driving function and other parameters. Thus, the allocation determines the optimal mapping of scenarios on different test capabilities. These are proving ground tests, virtual tests and X-in-the-loop tests, as a combination of proving grounds and virtual environments. Proving ground tests are expensive and cannot cover the wide variety of tests, which are possible to conduct in virtual environments. However, proving grounds offer the most reliable results in terms of realism. A combination of virtual and real tests are executed as X-in-the-loop tests, whereby certain elements are real, while other are simulated.

After the test was conducted, it needs to be evaluated. Inputs to the evaluation are, besides others, human capabilities and pass-fail criteria. Pass-fail criteria are derived from the driving function and the use-case. An easy to understand example would be, that the distance to other objects is greater than zero, respectively, there is no crash. These pass-fail criteria need to be described in machine readable and defined format. However, a detailed description of the evaluation is out of scope of this paper. Results from the evaluation can be redirected to the selection of relevant scenarios to test certain edge cases in a more detailed way. Finally, the result is given as output to the test engineer.

Relevance Metrics for the Selection of Scenarios

The process for selecting the relevant scenarios, including their respective stochastic parameter variations, is the first step of the proposed HEADSTART methodology, as seen in Figure 1. In general, this step of the methodology is using the available information from the driving function, by means of functional requirements. This information is used to select the applicable scenarios for the respective driving function out of the available scenarios from the database. The process can be divided into two main sections (see Figure 2):

Filtering of all relevant logical scenarios based on functional requirements:

The first step of this section is to define the functional requirements of the driving function from the vehicle under test (VUT), which should be safety assured. The provided input should include the following pieces of information:

- Operational Design Domain (ODD)
    - The ODD does reflect the limitation of the technological capability of automated driving systems (SAE J 3016, 2018). One example of dividing the ODD into different categories is given by Thorn et al. (2018). They list physical infrastructure (e.g., roadway types), operational constraints (e.g., speed limits), environmental constraints (e.g., weather and illumination) as well as connectivity as categories for defining an ODD.
- Object and Event Detection and Response
    - The monitoring of the driving environment via object and event detection, as well as classification, response preparation and execution, is collectively referred to as object and event detection and response (OEDR). Next to the longitudinal and lateral vehicle motion control, it is one of the subtasks of the dynamic driving task (DDT) (SAE J 3016, 2018). Thorn et al. (2018) map defined events to the respective response of the driving function, which can be used for checking the quality of the executed scenarios from the database.
- Tactical and Operational Maneuver Behavior
    - Michon (1985) divides the overall act of driving into three types of driver efforts: Strategic (e.g., trip planning), tactical (e.g., deciding on a takeover) and operational (e.g., executing actuator commands). The definition of DDT, provided by (SAE J 3016, 2018), includes tactical and operational effort and therefore defines the portion of driving that entails operating a vehicle in an active lane of traffic. By defining the tactical and operational maneuver behavior, it is possible to categorize the abilities of different driving functions.

With these requirements defined, a query to the scenario database can be created. The scenario database is composed as a list of different logical scenarios, including the ranges of their respective parameters for each applicable layer, based on the proposed layer-model of the PEGASUS (2019) project [4] derived from [3].

After the scenario extraction from the database, a check if the scenarios cover all events listed in the OEDR table, is conducted. If this is not the case (e.g., because of limited scenarios in the database), additional scenarios need to be injected at this stage. Such a situation could, for example, apply to the behavior outside the defined ODD, as the query excludes those scenarios. Additionally, feedback from the evaluation should be addressed at this stage.

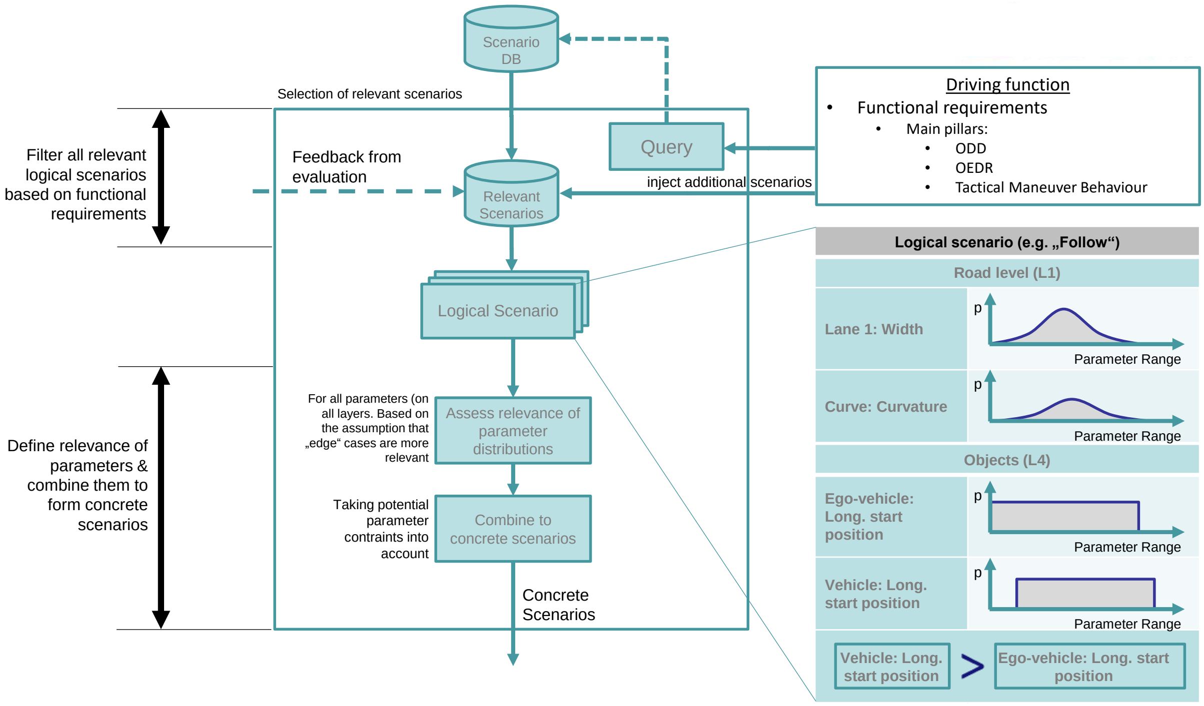

Figure 2 – Structure of the scenario selection process

Define the relevance of parameters and combine them to concrete scenarios:

After these initial steps, all the relevant scenarios, including their respective parameter distributions, are gathered. The next primary step is to define the relevance of the parameters and combine them to concrete scenarios. First, the relevance of the parameters is assessed using the associated distributions, for all parameters and layers. The underlying probability distribution of each scenario parameter is used to combine the overall occurrence probability for a specific scenario.

A qualitative example of this can be seen in Figure 3. This categorization, which is taken from OICA (2019), can also be used as an essential distinguishing factor for the target groups defined in the HEADSTART project. Secondly, the parameters are combined to create concrete scenarios, taking potential parameter constraints into account. Also, potentially prohibited parameter combinations due to given constraints need to be considered as well. This can also be seen in Figure 2, with the logical scenario "*Follow*" as an example. In this scenario, where the ego vehicle should follow a lead vehicle, the start position of this exact lead vehicle needs to be in front of the ego vehicle, leading to the constraint that the longitudinal position of the lead vehicle needs to be greater than of the ego vehicle. This leads to concrete testable scenarios, which are in the next step of the overall methodology allocated to a specific test method.

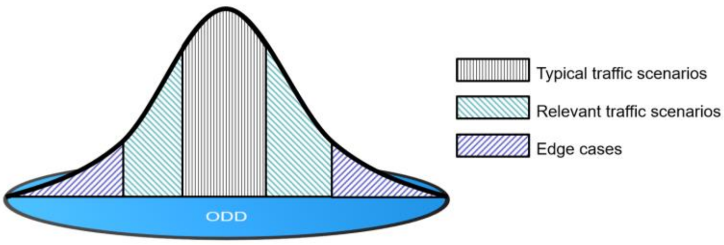

Figure 3 – Occurrence probability of traffic scenarios and their respective location within the ODD (qualitative representation)

Allocation of Scenarios for Test Execution

Once relevant scenarios are selected for the tests in the first step of the methodology, we need to define where to test them. To achieve this, we will allocate scenarios to one or several test instances. This is an essential step in the testing and validation methodology because each test instance has its advantages and restrictions. Some test instances are better to test some scenarios than others. For example, some accident scenarios can be too risky to test on proving ground because of safety concerns as well as high testing cost, therefore, these scenarios will be more suitable for virtual testing.

Previously selected logical and concrete scenarios are needed as input for this allocation process. Logical scenarios are used because they bring a well-defined layered approach. The scenario parameters are described with ranges and are not fixed, they allow more flexibility in the testing, especially for proving ground testing. In order to allocate scenarios to test instances, we need to define the capabilities of each test instance and map them to the selected scenarios. Capabilities can be divided into the categories “Sensor”, “Environment” and “Vehicle Dynamics”. Simulation models or the real world can be represented in these three categories. This framework, therefore, provides a possibility to categorize every possible testing method in terms of its capabilities. The testing methods targeted by HEADSTART for the scenario allocation process are:

- Proving ground testing
- Virtual testing / Simulation
- X in the Loop testing (XiL)

Figure 4 shows an example of this map of capabilities (MoC) for each testing method. In addition to the three main categories, there are also “resource-based” capabilities like time, costs and availability (e.g., available area of a proving ground). These resource-based capabilities will be further detailed in the next subsection.

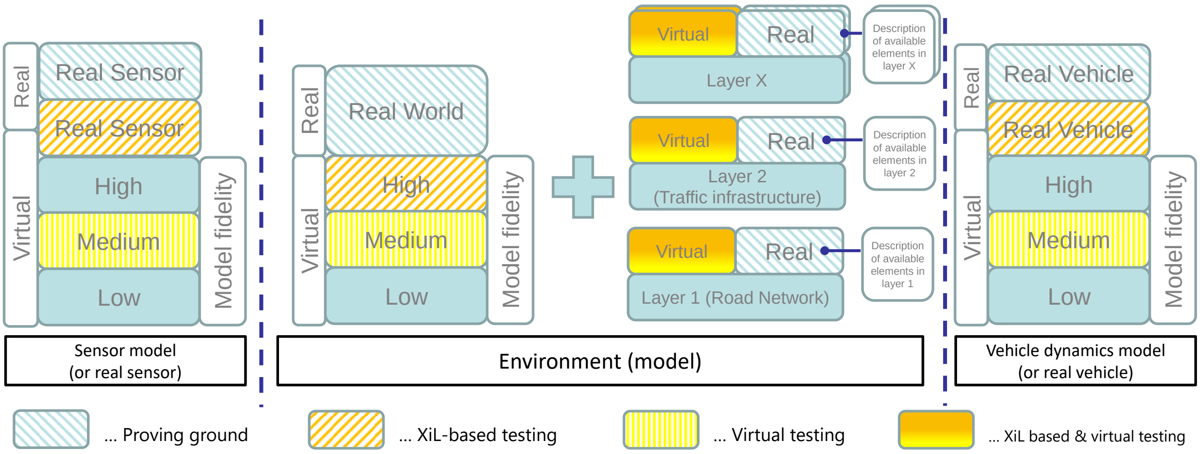

Figure 4 – Example of the developed map of capabilities for the different test methods

After defining the capabilities for each test method, the concrete scenarios will be matched according to that. To keep a structured format for the matching process, logical scenarios are used as a baseline for the allocation. From this starting point, the concrete scenarios are derived and finally allocated to the particular test method. An overview of this allocation process can be seen in Figure 5.

Subprocess for proving ground:

The first step for the proving ground subprocess consists of a comparison between the elements of the current logical scenario with the capabilities of the proving ground. If the proving ground can represent each needed element of the logical scenario, the capabilities are aligned with the specific parameter ranges. This concludes the initial matching process and allows the allocation of executable concrete scenarios.

Subprocess for virtual testing:

For virtual testing, the first step is based on the elements of the logical scenario. They need to be displayed accordingly in the simulation. Additionally, the OEDR response is used to derive potential additional requirements for the virtual elements. The requirements from the specific use case are a significant part, as they define potential minimum requirements for the simulation framework. These requirements need to be fulfilled, as otherwise, statistical coverage cannot be achieved.

Subprocess for XiL-based testing:

XiL-based testing is a mixture between the virtual testing and the proving ground, as it combines elements of both virtual and real elements. Therefore, the same step as explained for the proving ground, as well as for the virtual testing, apply to the XiL-based testing as well.

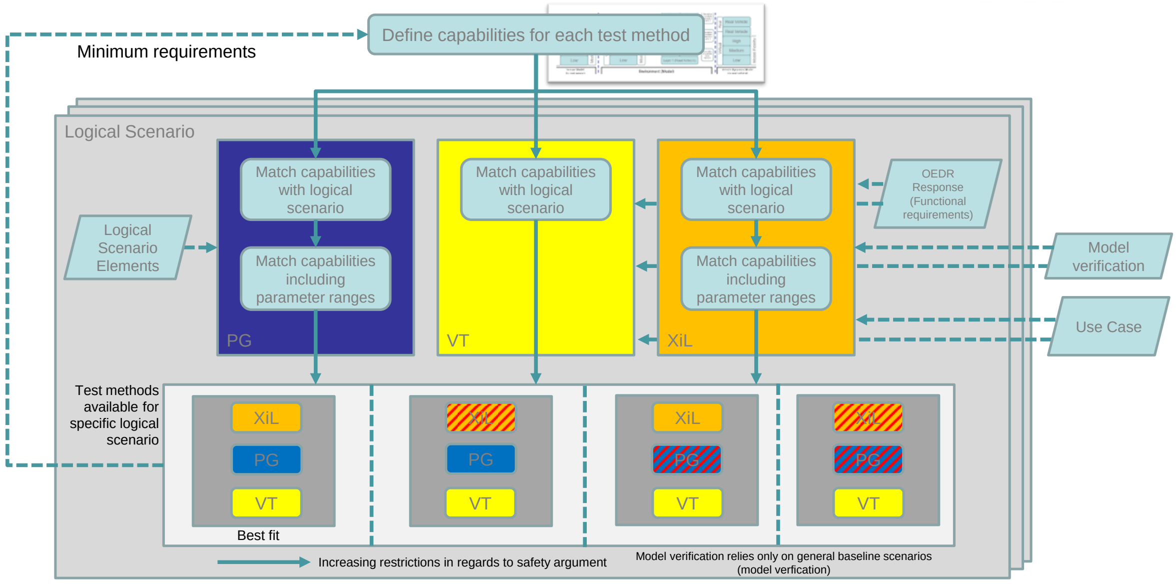

Figure 5 – Overview of the scenario allocation process

Safety argument and resource-based capabilities

At the end of the process, the requirements and capabilities will indicate which test methods can be used. In the best case, all three methods fulfill the requirements, which leads to a "best fit". The three other possible test method combinations have increasing restrictions regarding the safety argument and resource-based capabilities (from left to right). Test method combinations without the usage of virtual testing are prohibited, as the other test methods are not able to fulfill the needed statistical test coverage. Test method combinations without the usage of the proving ground are possible, but the model verification needs to rely on either the baseline scenarios or other similar logical scenarios. After determining the 'best fit' from a technical perspective, safety and resource-based capabilities are taken into account for the final selection of the test method. Resource based capabilities include: Time; Cost; Efficiency; Technical feasibility (mainly linked to the sensors of the tested vehicle); Availability of the testing tools, etc. They provide additional KPI to do the final selection of the test method. The method selected in the end will be a compromise between technical capabilities, safety and other resource-base capabilities.

Specificities of open road testing

Open road testing is also considered in the HEADSTART methodology for the testing and validation of CAD. However, this approach cannot be tackled the same way as the other three test instances presented in the previous section. There are several reasons for this limitation: First, the lack of controllability over the ODD of the function. There is no complete control of the ODD, the real situations encountered will depend on the interaction with other public road users. The second limitation is the lack of reproducibility and repeatability of the scenarios. Public-road scenarios are difficult to replicate exactly in different locations and over multiple iterations. Finally, the lack of scalability is also a limitation as public-road scenarios may not scale up well. Indeed, the connected and automated vehicle may require additional data, such as a priori digital maps, and infrastructure equipped for V2X communication technology.

The consequence of these limitations is that the scenario-based approach for testing Automated vehicles cannot be applied with only open road testing or Field Operational Tests (FOTs). These test methods can be used as a complement to the scenario-based approach mainly to test the ergonomics of the car, the driver behaviour and Human Machine Interactions (HMI). FOTs can also be used to test some conditions of the operational design domain of the vehicle and sensor perception, by testing the vehicle in extreme environmental condition.

Pertinence of test methods for the final users

The three final users considered by the HEADSTART project are : Development testing, Consumer testing and type approval testing. Each of these users have specific needs and objectives in terms of testing CAD functions, consequently, not all test methods presented earlier (proving ground testing, virtual testing, XiL and field testing) are relevant to them.

- **Development testing** regroups stakeholders such as OEMs and Tier 1 suppliers. Development testing often use all four test methods in order to test the safety of their vehicles.
- **Consumer testing** is done by non-profit organisations such as EuroNCAP to rate the safety of various models of cars. These organisations are only testing vehicles that are already commercially available. Traditionally, consumer testing is using proving ground testing, but with the introduction of ADAS systems and the first automated cars, simulation and virtual testing are becoming more relevant and will probably be used for testing level 3+ systems.
- Finally, **type approval testing** and homologation is done by public authorities to certify that a car meets a minimum set of regulatory, technical and safety requirements. Type approval testing is also done via proving ground testing. But other test methods are being considered for testing CAD functions such as simulations in cooperation with the OEMs as well as FOTs via a Digital Driving Licence.

Conclusion

Many projects are trying to create a methodology for the verification and validation of automated driving functions. Projects like PEGASUS, MOOVE [14][15], TNOStreetWise, VVMethods [16] or SAKURA are just some examples from the international horizon. However, the verification and validation of SAE Level 3+ driving functions is a global problem. Especially in the European Union, it is necessary to have a common methodology to be provided for all countries and to enable the cross-border functionality of driving functions. The HEADSTART project is a project to identify and analyse finished and ongoing projects in this context and to define a common and harmonized methodology based on this analysis. In this paper, a general overview of the defined HEADSTART methodology was given. This methodology will be exemplarily executed and tested in further work packages of the project. Moreover, international coordination by having workshops and an expert network is key to further developments of the HEADSTART methodology and implementations.

References

1. N.N. *PEGASUS Method for Assessment of Highly Automated Driving Function* PEGASUS Project https://www.pegasusprojekt.de/en/pegasus-method (accessed at 02.08.2019) 2nd PEGASUS Symposium Ehra-Lessien, Germany, (14.05.2019).
2. Weber, H.; Bock, J.; Klimke, J.; Roesener, C.; Hiller, J., Krajewski, R.; Zlocki, A.; Eckstein, L. *A Framework for Definition of Logical Scenarios for Safety Assurance of Automated Driving.* In: Traffic Injury Prevention. Special Edition ESV (2019) DOI:10.1080/15389588.2019.1630827.
3. Bagschik, G.; Menzel, T.; Körner, C; Maurer M. *Wissensbasierte Szenariengenerierung für Betriebsszenarien auf deutschen Autobahnen* In: 12. Workshop Fahrerassistenzsysteme und automatisiertes Fahren. Walting, Germany, (2018) UNI-DAS e.V., pages 1 - 14.
4. Bock, J., Krajewski, R., Eckstein, L., Klimke, J., Sauerbier, J., Zlocki, A. *Data Basis for Scenario-Based Validation of HAD on Highways* 27th Aachen Colloquium Automobile and Engine Technology 2018, Aachen, Germany, (2018).
5. Knapp A., Neumann M., Brockmann M., Walz R., Winkle T. *Code of practice for the design and evaluation of ADAS* Preventive and Active Safety Applications, eSafety for road and air transport, European Commission Project, Brussels .
6. Mustermann, Max (2000). *SRC_WINNER*, Veröffentlichung.
7. Wagener N., Zlocki A., Weber H., Bock J., Eckstein L. (2019). *State of Research on Data-Driven Safety Assurance*, 28th Aachen Colloquium Automobile and Engine Technology.
8. Winner H., Lemmer K., Form T., Mazzega J. (2019). *PEGASUS—First Steps for the Safe Introduction of Automated Driving*, Meyer G., Beiker S. (eds) Road Vehicle Automation 5. Lecture Notes in Mobility.
9. Antona-Makoshi J. (2019). *Towards global AD safety assurance*, AVS2019, San Francisco.
10. Elrofai, H., Paardekooper, J.-P., de Gelder, E., Kalisvaart, S., Op den Camp, O. (2018). *StreetWise Scenario-Based Safety Validation of Connected and Automated Driving*, White paper. http://publications.tno.nl/publication/34626550/AyT8Zc/TNO-2018-streetwise.pdf.
11. Krajewski R., Bock J., Kloeker L., Eckstein L. (2018). *The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems*, 21st International Conference on Intelligent Transportation Systems (ITSC).
12. Jullien J., Martel C., Vignollet L., Wentland M. (2009). *OpenScenario: A Flexible Integrated Environment to Develop Educational Activities Based on Pedagogical Scenarios*, Ninth IEEE International Conference on Advanced Learning Technologies.
13. Dupuis, M., Grezlikowski, H., (2006). OpenDRIVE – An Open Standard for the Description of Roads in Driving Simulations, Proceedings of the Driving Simulation Conference DSC Europe 2006, 25-35.
14. Bracquemond, A. *Status of the MOOVE project*, Automated Vehicle Symposium 2019 (AVS2019) Orlando, 2019

15. Bracquemond, A., Thiolon G., *MOOVE Project : Recognition of Road Scenes by the Data Collected at the Output of the Sensors of the Autonomous Vehicle*, 28th Aachen Colloquium Automobile and Engine Technology 2019.
16. Galbas, R.; 2018. *Projekt „VV-Methoden“ - Validierung und Verifikation für hochautomatisiertes Fahren.* In: Symposium „Testen - Automatisiertes und Vernetztes Fahren“. Braunschweig, Germany. (2018)