ORIGINAL ARTICLE

A Big Testing Framework for Automated Truck Driving

Mohamed Elgharbawy
Karlsruhe Institute of Technology (KIT), Germany

Abstract: Long-distance commercial vehicles are predestined for automated driving due to their high performance and long monotonous routes. Automation offers the prospect of improved road safety, increased fuel efficiency, optimised vehicle utilisation, higher driver productivity and lower freight costs. Even if the widespread use of full automation is not imminent, the vision of accident-free driving accelerates the further development of driver assistance functions to autonomous vehicle stages on the global market. The status quo evaluation refers to large-scale verification as one of the decisive challenges for the economical, reliable and safe use of automated driving functions in truck series development. In this scheme, the evaluation of software releases must be carried out in different phases up to the Start of Production (SoP) to provide an argument that the residual risk is below an acceptable level. In driving simulator tests, various system concepts of a truck series are first evaluated. The verification and validation strategy then performs X-in-the-Loop tests, proving grounds and long-term endurance tests. Finally, homologation meets the market-specific type-approval requirements based on the evidence collected during development. This paper summarises previous works dealing with the large-scale verification requirements and challenges of intelligent transportation systems. The basis of large-scale verification is presented, including the verification and validation procedures commonly used in large-scale verification schemes. The criteria of test completion are specified for assessing the performance of automated driving functions. The quality measures are presented to achieve sufficient reliability within the software quality management process. The several possible topics for future research are identified.

Keywords: Big data analysis, Large-scale verification, Start of Production, X-in-the-Loop tests, Homologation

1. Introduction

Long-distance commercial vehicles are vehicles designed to create economic value. They are highly specialised in fulfilling specific tasks and are primarily controlled by economic efficiency. Commercial vehicles are characterised by a large number of product ranges and models with tractors, semi-trailers, or trailer combinations. The legislators in large regions regulate the concepts and functions of commercial vehicles up to and including a particular truck system. The current challenges are to improve the use of existing infrastructure, enhance the utilisation and combination of assistance functions and make the truck driver profession more attractive[1].

1.1 Motivation

Freight traffic is growing worldwide and is the dominant means of transport. According to the traffic forecast for 2030, road freight transport performance in Germany will increase by 38% compared with the level of 2010[1]. Long-distance vehicle accidents often have serious consequences such as personal injury and death, as well as considerable financial costs and environmental risks. Therefore, the road safety of commercial vehicles is an essential aspect

doi: 10.18686/utc.v4i1.27

of society. The number of truck accidents involving seriously injured road users has fallen by more than 45.8% between 1992 and 2014. Although the volume of truck traffic in the same period has increased by 85.3%, the number of people who have died at these accidents has decreased by more than 59.7 %[2], as illustrated in **Figure 1**. The most frequent crash with heavy trucks is the rear-ended collision with a passenger car, in which the severity of the accident for the passenger car is considerably worse. For this reason, the European Commission has started to equip all common trucks with automatic emergency braking systems with a gross vehicle weight of more eight tonnes. The only exceptions are off-road vehicles, steel-sprung heavy-duty vehicles and trucks with more than three axles.

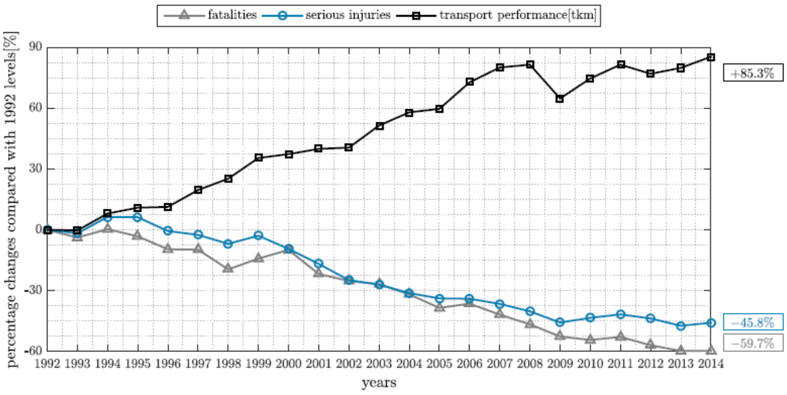


Figure 1: Fatalities and seriously injured persons in truck accidents on German roads compared to truck transport performance (1992 – 2014)

The development of automation in civil aviation to meet the increased safety require- ments can be seen as an indication of the challenges of the same expansion in trucking. Both sectors focus on the commercialisation of freight and passenger transport in a scal- able environment. Statistical studies on the number of accidents in civil aviation confirm a significant and sustained decline in accident rates worldwide, although the number of aircraft has increased. The study of statistics on the life cycle of each generation of jets shows that the lowest fatal accident rate of first-generation jets was around 3.0 accidents per million flights. For the second generation, it was about 0.7, which corresponds to an 80% reduction in fatal accidents. By comparison, third-generation jets achieve about 0.2 accidents per million flights. Finally, fourth generation jets have the lowest accident rate, with a stable average of about 0.1 fatal accidents per million flights, which represents a further 50% reduction compared to the third generation, as depicted in **Figure 2-Figure 3.**

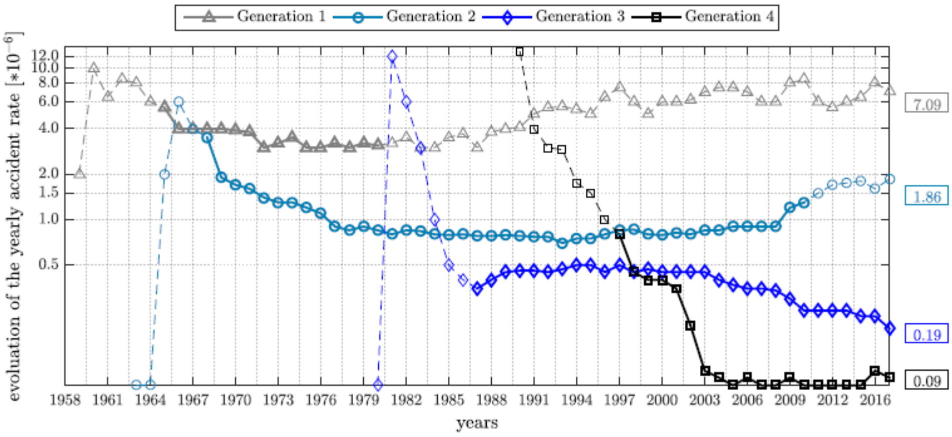


Figure 2: Statistical evaluation of accidents in commercial aviation(1958 – 2014)

[1] The 2030 Federal Transport Infrastructure Plan

[2] German federal statistics office, accidents and mileage

[3] Airbus- A Statistical Analysis of Commercial Aviation Accidents 1958-2016

1.1 Problem definition

The conversion from driver assistance systems of levels 0, 1 and 2 to higher levels of au- tomation in accordance with SAE J3016 represents a new challenge for the verification of autonomous trucks. The main difference is that driver assistance can intervene unintend- edly. The driver can override the side effects of the assistance system at any time if there are system limits[2]. Their functions are therefore designed to be controllable, but this can reduce their benefits. The controllability of system interventions and the effectiveness in the field with minimisation of undesired consequences are therefore decisive for the series development of these drive functions. Accordingly, Systems Engineering requires state-of-the-art evaluation procedures to verify and validate these systems. Long-term endurance tests are carried out to define thresholds for intervening systems based on the collected data. On the one hand, trigger algorithms can be optimised to minimise the frequency and impact of false-triggered interventions, and on the other hand to maximise the number of legitimate responses. Nevertheless, automated driving requires that a system exploits the limits of dynamic driving tasks as required and masters most environmental conditions controlled by a human driver. The new ISO26262 standard regulates the functional safety of electrical/electronic (E/E) systems in heavy commercial vehicles. However, the safety standard is limited to avoiding potentially safety-critical situations caused by undetected random hardware failures and dangerous systematic failures. Safety violations due to technological and system-technical deficiencies remain outside the scope of ISO26262(e.g. insufficient robustness to environmental conditions, inadequate training data with machine learning algorithms, uncertainty issues with perception sensors, etc.). In particular, automated driving without driver monitoring can also lead to potential safety-critical situations resulting from deficiencies in the estimation, interpretation and perception processes. The primary question is therefore how to verify automated driving efficiently and adequately for the required test completion criteria.

1.2 Contribution

A statistical method for predicting the required test distance is being elaborated to demon- strate the safety of automated trucks on the basis of fatalities and injuries. Accident-based statistical safety assessment is an unresolved challenge for the developers of these tech- nologies when physical driving testing have to prove safety during the development phase. Despite the fact that the uncertainties of machine learning remain before au- tomated driving is released for widespread use, it is essential to develop innovative methods that complement physical driving experiments. This paper describes a modular framework that suggests the verification process of automated truck driving. In addition, the cluster-in-the-loop framework offers an evaluation of perception sensors, decision al- gorithms and functional robustness. The structure employed utilises a backend database that is filled with catalogs of relevant driving scenarios from various sources of field- based observations. In this scheme, the processing chain includes clustering multivariate time series data sets and locating critical driving situations to identify and assign the necessary test cases for different appropriate test environments. These new test cases then complement the existing test cases, which were developed from expert knowledge in adaptive test coverage manner. The platform-independent mechanism is intended to provide a consistent scenario description format for the various test environments. The proposed framework therefore contributes to a potential trade-off between the efficiency and effectiveness criteria of a scenario-based test concept[3].

1.3 Structure of the paper

The paper is structured as follows: Section 2 provides a brief overview of related work. Section 3 explains the proposed methodology and details of the framework implementa- tion. Finally, quantitative results are presented in section 4 and the performance of the overall system is evaluated. The paper concludes with a summary in Section 5.

2. Related work

Passive and active safety systems differ from their evaluation methods. A standard eval- uation approach has been developed for passive safety systems to verify their behaviour with an appropriate number of crash test cases under certain worst-case conditions. In contrast, the safeguarding of active safety systems poses a number of challenges with regard to the variety of relevant scenarios and environmental conditions, the complexity of the systems, the variability of driver behaviour and functional deficiencies.

2.1 Towards autonomous trucks

Due to the high mileage and long monotonous distances of long-haul trucks, various busi- ness cases for cooperative automated driving are demonstrated, such as truck platooning. Despite the fact that the widespread use of full automation is not imminent, the vision of accident-free driving expedites the further development of driver assistance functions to evolutionary autonomy stages on the global market. These stages are expected to overlap and are not sequentially available on the market. Despite strong support of industry and academia, questions about their business cases, ethical dilemmas, legal liability and safety frequently arise. For example, a further adaptation of the Vienna Convention on Road Traffic is necessary with regard to the provision of an automated steering sys-tem, which is prohibited in the UN-ECE R79 for use above 10 km/h[4]. The traditional method of sense-plan-act robot control provides a functional view of the data flow in the sensor and control system of an autonomous truck. The sensor system is responsible for understanding the current state of the environment[5]. While the planning part is responsible for finding out what the best next step is, the acting part is responsible for implementing the plan. A truck equipped with automated driving can be identified as a Cyber-Physical Vehicle System (CPVS) whose driving functions enable the intelligent handling of dynamic traffic situations in an extremely safety-critical environment, as illustrated in figure 3. In order to identify the challenges of proof of safety for automated driving functions, we propose four such evolutionary stages of automated truck driving adapted to the OICA/SAE standard J3016 automation levels[6]. Each of these stages is identified as follows:

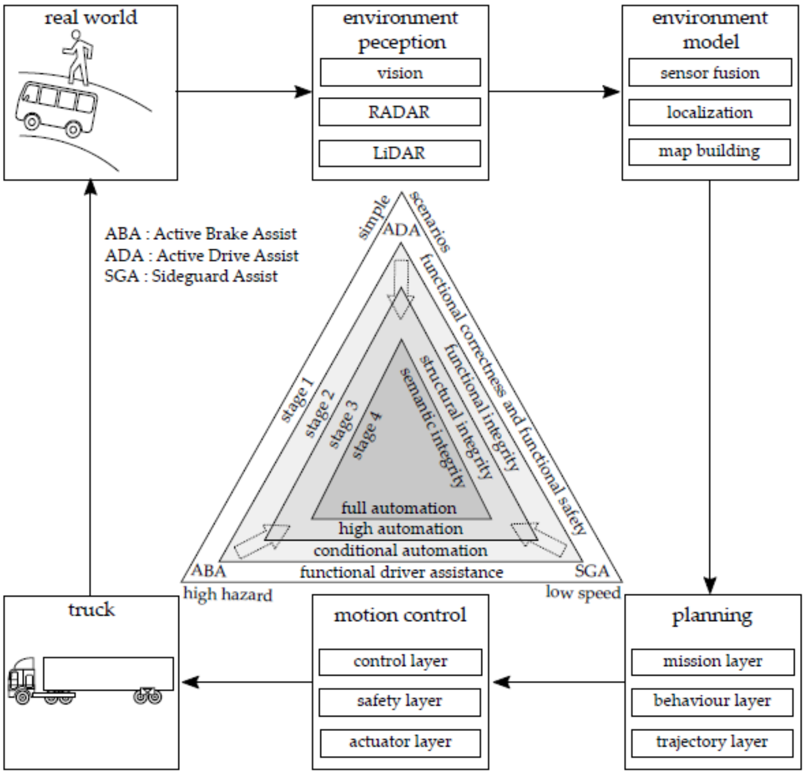


Figure 3: Evolutionary triangle of automated truck driving using the sense-plan-act control methodology

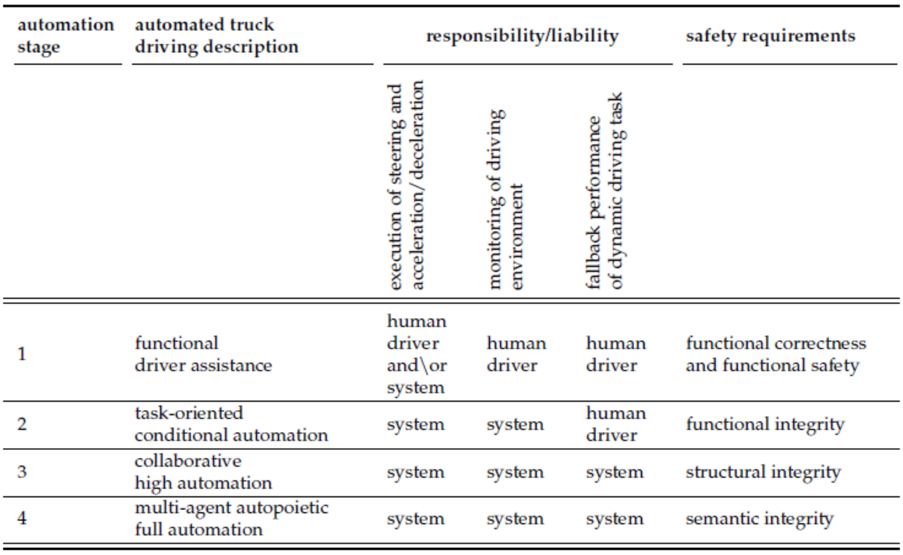
- The first stage can be divided into three sub-categories of driver assistance systems. Functional information and warning systems are the first sub-category in which the driver is fully engaged (e.g. traffic sign recognition, lane departure warning, sideguard assist, etc.). The second sub-category concentrates on functional intervening and continuously assisting driver assistance systems (e.g. active brake assist, adaptive cruise control etc.). The third sub-category concerns with functional combinations and multiple interacting driver assistance systems (e.g. active drive assist, highway assist, etc.). All of these systems do not learn during operation and perform limited tasks in a clearly defined context. Cooperation is therefore limited to the exchange of information on the system context. The verification and validation strategy aims to ensure the functional correctness and functional safety of the system. Critical driving situations due to E/E system failures can be addressed within the framework of ISO26262. The Safety of the Intended Function-ality (SOTIF) regulates the absence of unreasonable risks due to hazards arising from performance limitations and insufficient situational awareness.

- The second stage comprises task-oriented conditional automation systems (e.g. highway pilot) in which the system operates in a sequence of manageable situations. While the system does not learn during operation, it optimises its trajectories during the control process according to defined objectives such as time or other resources. The cooperation with other systems is therefore limited to the exchange of information about the system context and the system itself. Handling uncertain information in fail-operational mode is essential for predicting and interpreting situations because environmental awareness is not 100% reliable. On this basis, the functional integrity of the system must be ensured with the verification and validation strategy.

- The second stage comprises task-oriented conditional automation systems (e.g. highway pilot) in which the system operates in a sequence of manageable situations. While the system does not learn during operation, it optimises its trajectories during the control process according to defined objectives such as time or other resources. The cooperation with other systems is therefore limited to the exchange of information about the system context and the system itself. Handling uncertain information in fail-operational mode is essential for predicting and interpreting situations because environmental awareness is not 100% reliable. On this basis, the functional integrity of the system must be ensured with the verification and validation strategy.

-The third stage involves collaborative high automation systems that collaborate with other systems to perform their tasks (e.g. truck platooning, etc.). They negoti-ate their objectives, plans and actions with other systems and adapt their behaviour to the negotiated procedure. Since the system boundary changes dynamically due to the collaborative relationship, mechanisms for distributed planning and coordi-nation of interpretations are required to ensure safe system functionality. Therefore, their verification and validation strategy focuses on the structural integrity.

- The fourth stage includes multi-agent autopoietic full automation by autonomously expanding their perception, situational awareness and actions. The ability of unsupervised learning during operation at all levels of perception and action is the main feature of this system class. For this purpose, a concept of semantic integrity is necessary in order to safeguard any possible expansion.



automation stage	automated truck driving description	responsibility/liability			safety requirements
		execution of steering and acceleration/deceleration	monitoring of driving environment	fallback performance of dynamic driving task	
1	functional driver assistance	human driver and\or system	human driver	human driver	functional correctness and functional safety
2	task-oriented conditional automation	system	system	human driver	functional integrity
3	collaborative high automation	system	system	system	structural integrity
4	multi-agent autopoietic full automation	system	system	system	semantic integrity

Table 1: Safety integrity requirements for each automation stage, adapted from the SAE J3016 automation levels[7]

2.2 Deficiencies in environment sensing

The perception and situation prediction sensors have different measurement principles, which can be classified as follows:

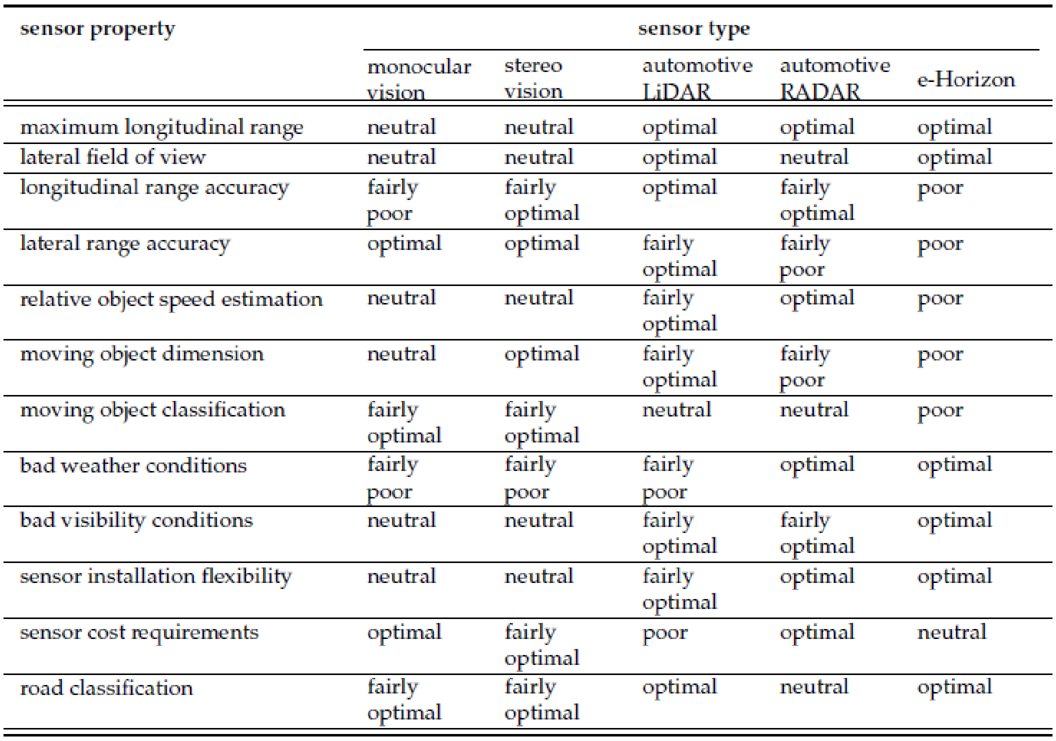
-Monocular vision sensors measure the incident light using an optical system. However, no depth or speed information can be measured directly, whereby the three-dimensional world is projected onto the two-dimensional image. The recog-nised object features are mapped to a vector representing an object hypothesis in the state space of the used classifier.

- Stereo vision sensors consist of an arrangement of two monocular cameras with a certain distance (base width) to each other and measure an environmental detail from different perspectives. The measurements are carried out synchronously, whereby a depth estimation is generated in the two images by comparing the displacement (disparity) of individual pixels or patterns. In addition, the distance accuracy is limited by the resolution, especially at long distances.

-Automotive RADAR (RAdio Detection And Ranging) sensors transmit and receive radio waves to determine the speed, range and angle of objects. Its strengths occupy in an extended longitudinal range, an optimal accuracy of the range rate with weather independence. Radar sensors, however, can poorly resolve closely spaced objects over long distances.

-Automotive LiDAR (LIght Detection And Ranging) sensors are based on an optical measurement principle to locate and measure the distance of objects in space. LiDAR sensors typically use the time of flight principle for distance measurement, where a laser pulse is emitted and the elapsed time is measured until the reflected signal is received again. The time delay between transmit and receive is directly proportional to the distance due to the proportionality between the time of flight and distance.

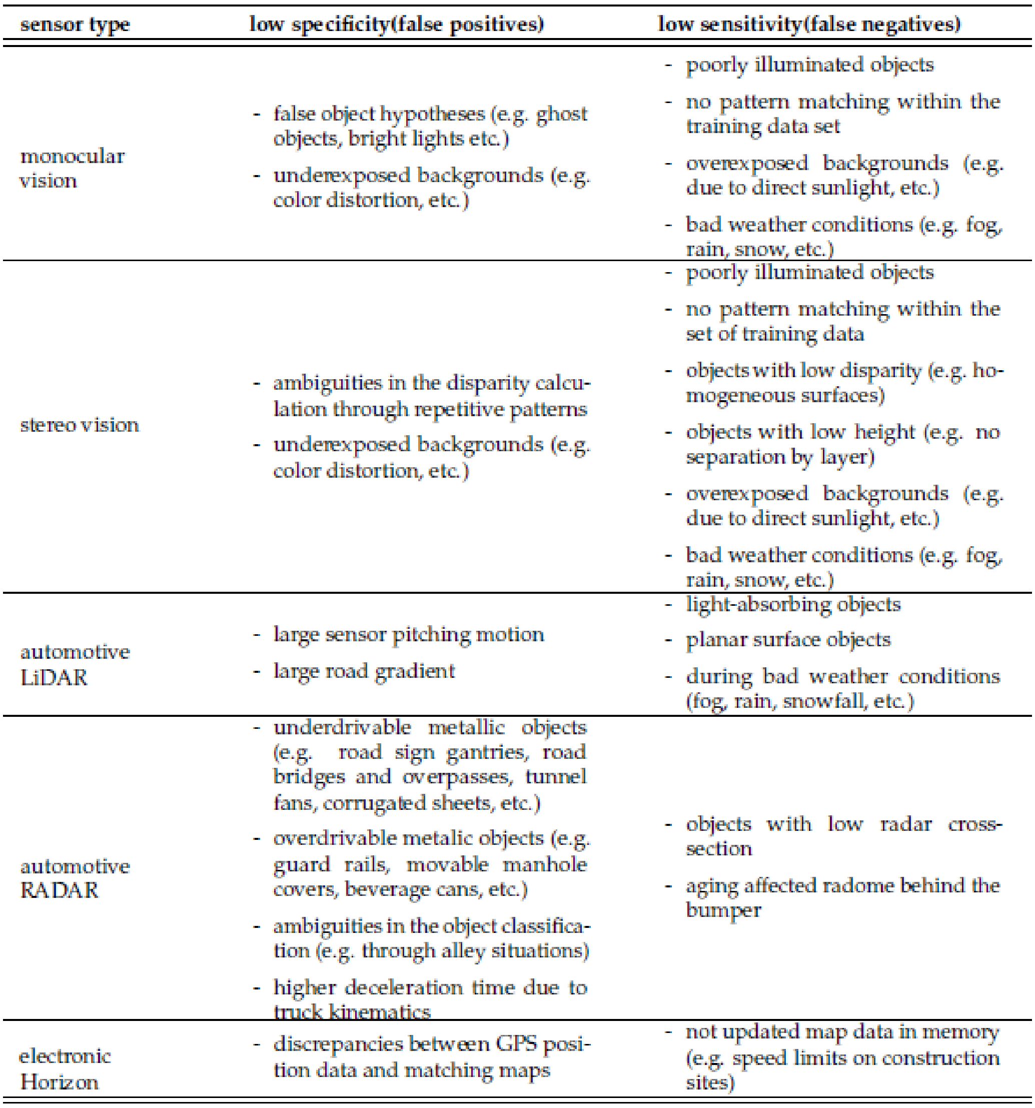
-e-Horizon (electronic Horizon) sensors employ the digital map data and GPS sensors to predict the driving route. The GPS sensor determines the vehicle position in world coordinates. The map matching transforms this place into map coordinates and assigns it to a specific road on the map. Subsequently, the most probable path extracts and processes the relevant map characteristics along the most likely future route using prediction algorithms. The e-Horizon data includes vehicle position data and road segment attributes such as road geometry, road class, number of lanes and speed limits.



sensor property	sensor type				
	monocular vision	stereo vision	automotive LiDAR	automotive RADAR	e-Horizon
maximum longitudinal range	neutral	neutral	optimal	optimal	optimal
lateral field of view	neutral	neutral	optimal	neutral	optimal
longitudinal range accuracy	fairly poor	fairly optimal	optimal	fairly optimal	poor
lateral range accuracy	optimal	optimal	fairly optimal	fairly poor	poor
relative object speed estimation	neutral	neutral	fairly optimal	optimal	poor
moving object dimension	neutral	optimal	fairly optimal	fairly poor	poor
moving object classification	fairly optimal	fairly optimal	neutral	neutral	poor
bad weather conditions	fairly poor	fairly poor	fairly poor	optimal	optimal
bad visibility conditions	neutral	neutral	fairly optimal	fairly optimal	optimal
sensor installation flexibility	neutral	neutral	fairly optimal	optimal	optimal
sensor cost requirements	optimal	fairly optimal	poor	optimal	neutral
road classification	fairly optimal	fairly optimal	optimal	neutral	optimal

Table 2: Evaluation of environmental perception and situation prediction sensors

Since automated driving of trucks depends on environmental perception, safety vio-lations may be caused by limitations due to physical or technical constraints on the intended functioning of a system. In addition, object recognition and classification are primarily performed by machine learning techniques to extract relevant features in an unstructured operational context[8]. Although machine learning paradigms offer a promising perceptual performance, high values of false-negative and false-positive rates can have critical safety consequences within the overall system[9]. Therefore, the performance evaluation of the environment sensing should be established to provide a sufficiently safe level of residual risk associated with functional deficiencies in machine learning functions[10]. Accordingly, the various sensors must be verified not only in terms of failure rates, but also in terms of possible causes of technical deficiencies in machine learning. The evaluation criteria of S1 perception sensor contain false positive and false negative rates (P 1 and P 2 respectively) by implying some assumptions about the system context. The sensor detects pedestrian objects from a longitudinal distance of P 3 with a lateral distance P 4 from the longitudinal axis of the truck. The parameters P 1,P 2,P 3 and P 4 also refer to the system parameters (e.g. ego vehicle speed). Robust-ness in real traffic can be achieved by the intelligent fusion of sensor data and reasonable system design. The possible uncertainties for different perception sensors are classified as shown in **Table 3**.



sensor type	low specificity(false positives)	low sensitivity(false negatives)
monocular vision	- false object hypotheses (e.g. ghost objects, bright lights etc.) - underexposed backgrounds (e.g. color distortion, etc.)	- poorly illuminated objects - no pattern matching within the training data set - overexposed backgrounds (e.g. due to direct sunlight, etc.) - bad weather conditions (e.g. fog, rain, snow, etc.)
stereo vision	- ambiguities in the disparity calculation through repetitive patterns - underexposed backgrounds (e.g. color distortion, etc.)	- poorly illuminated objects - no pattern matching within the set of training data - objects with low disparity (e.g. homogeneous surfaces) - objects with low height (e.g. no separation by layer) - overexposed backgrounds (e.g. due to direct sunlight, etc.) - bad weather conditions (e.g. fog, rain, snow, etc.)
automotive LiDAR	- large sensor pitching motion - large road gradient	- light-absorbing objects - planar surface objects - during bad weather conditions (fog, rain, snowfall, etc.)
automotive RADAR	- underdrivable metallic objects (e.g. road sign gantries, road bridges and overpasses, tunnel fans, corrugated sheets, etc.) - overdrivable metalic objects (e.g. guard rails, movable manhole covers, beverage cans, etc.) - ambiguities in the object classification (e.g. through alley situations) - higher deceleration time due to truck kinematics	- objects with low radar cross-section - aging affected radome behind the bumper
electronic Horizon	- discrepancies between GPS position data and matching maps	- not updated map data in memory (e.g. speed limits on construction sites)

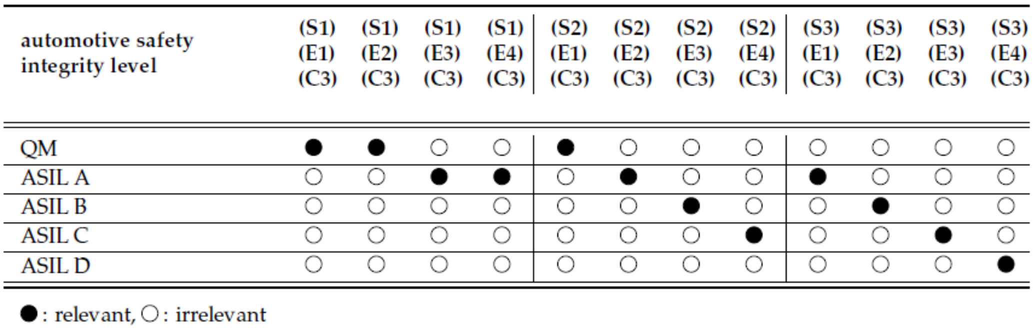
Table 3: Potential causes of uncertainty within environmental perception and situation prediction sensors

2.3 Safety of the Intended Functionality

ISO 26262 has been established as the state of the art in the development of safety-relevant systems in passenger cars[11]. ISO26262-3 includes hazard analysis and risk assessment to determine the required Automotive Safety Integrity Level (ASIL) and to assess the potential risks of E/E malfunctions that may violate the safety goals. For the analytical approach, a risk R can be described as a function F of three impact factors: severity S, the probability of exposure E and controllability C.

$$R = F\left(S, E, C\right) \tag{1}$$

The hazard classification defines each potentially hazardous driving situation according to the following classes: severity (S0-S3), exposure probability (E0-E4), controllability (C0-C3). This determination follows one of five categories to specify the risk and its risk reduction requirements, where ASIL D is the highest and Quality Management (QM) the lowest risk reduction class (ISO/WD 26262-1). For example, a system specified for the implementation a driver-assisted truck platooning may exhibit undesired behaviour due to misclassification of objects and require driver intervention. Controllability is therefore the probability that the driver can control driving situations, such as automated driving function, system limits and system failures. The processes and methods for assessing the controllability of unintended driver assistance reactions are specified in the Code of Practice. With automated driving without driver intervention, the ASIL determination can be assigned to the level [C3] of controllability[< 90%], where the intended function is difficult to control or uncontrollable, as illustrated in table 4[12].



automotive safety integrity level	(S1) (E1) (C3)	(S1) (E2) (C3)	(S1) (E3) (C3)	(S1) (E4) (C3)	(S2) (E1) (C3)	(S2) (E2) (C3)	(S2) (E3) (C3)	(S2) (E4) (C3)	(S3) (E1) (C3)	(S3) (E2) (C3)	(S3) (E3) (C3)	(S3) (E4) (C3)
QM	●	●	○	○	●	○	○	○	○	○	○	○
ASIL A	○	○	●	●	○	●	○	○	●	○	○	○
ASIL B	○	○	○	○	○	○	●	○	○	●	○	○
ASIL C	○	○	○	○	○	○	○	●	○	○	●	○
ASIL D	○	○	○	○	○	○	○	○	○	○	○	●

●: relevant, ○: irrelevant

Table 4: Automotive safety integrity level (ASIL) determination for automated driving without driver monitoring for uncontrollable level (C3) according to ISO 26262

Despite the updating of the scope of the ISO26262 standard for the inclusion of heavy-duty trucks in Edition 2, its safety goals mainly address undetected random hardware failures of the system components and systematic failures. Assuming that the E/E system malfunctions are managed using ISO26262, the safety violations, that may be caused by the environmental perception sensors remain outside the scope. Redundancy, diversity and functional restrictions can compensate system limitations. The Safety of Intended Functionality (SOTIF) ISO/WD PAS 21448 approach is currently under development and serves as an extension schema to specify the intended function in such a way that it is robust and safe enough to take into account the variations in sensor inputs and the different environmental conditions. Therefore, new verification and validation measures are needed to assess unintended system behaviour due to technological and systemic deficiencies. At the same time, SOTIF activities complement ISO 26262 with its focus on driver assistance rather than automated driving without driver engagement.

3. Dilemmas of automated driving assessment

3.1 Statistical proof of safety

The validation process begins with the selection of a validation target, which is calculated by the system use case, the crash statistics and a safety margin. For a particular use case, human drivers experience an average number of kilometres between events and the system owner defines a certain safety margin. The stopping rule assumes that the failure rate has a binomial distribution. It can be shown that the system has a failure rate greater than or equal to the benchmark reference with a certain confidence level. Therefore, the validation distance required to provide statistical proof of the safety of an automated driving system can be calculated on the basis of a benchmark reference for the expected interval between accidents of certain severity. The total fatality rate in Germany caused by trucks in 2015 was 787 fatalities, totaling 58.934 billion kilometres. According to the binomial distribution, an autonomous vehicle with the reliability with m failures during the travel distance x with confidence level C is:

$$C_{(\kappa=m)} = 1 - \sum_{\kappa=0}^{m} \frac{x!}{\kappa!\,(x-\kappa)!} \nu^{\kappa} \left(1-\nu\right)^{x-\kappa} \qquad (2)$$

If the failure rate of a truck is ν, then the reliability γ is $(1-\nu)$ and can be interpreted as the probability that is no failure in the route driven. A hypothesis about the scenario ”no failures driving” can be used to estimate a lower limit for the number of failure-free kilometres n to determine the reliability of autonomous trucks with a confidence level C. This determines the reliability that can be claimed for a certain number of failure-free kilometres at a particular confidence level.

$$C_{(\kappa=0)} = 1 - \left(1-\nu\right)^{x} \qquad (3)$$

The required test distance x without failures is defined for a given confidence C and reliability γ, as represented in equation 4.

$$x = \frac{ln\left(1 - C_{(\kappa=0)}\right)}{ln\left(1-\nu\right)} \qquad (4)$$

substituting ν with $\frac{787}{58.934*10^{9}} = 1.34 * 10^{-8}$and confidence level C with 95%.

$$x = \frac{ln\left(1-0.95\right)}{ln\left(1-\left(1.34*10^{-8}\right)\right)} \approx 220 * 10^{6} \qquad (5)$$

The required test distance is approximately 220 million km. **Figure 4** represents the failure rate factor $\frac{\nu_{AD}}{\nu_{HD}}$, where $\nu_{AD}$is the failure rate of an automated driving system and ν_{HD} is the benchmark failure rate of human driver. For today’s trucks, there is no necessity for such long validation distances, at which the controllability of the driver provides the necessary proof of safety. Nevertheless, the 2 million kilometres used to validate current driver assistance systems are sufficient to prove a fatality rate of $(\varsigma_{(2*10^{6}\text{km})=25.5})$ times that of humans with 50% confidence, in case of a fully automated truck. In order to prove that an autonomous truck has a failure rate similar to that of humans in 2015 as a benchmark failure rate and assuming that the truck has no failure ($\kappa = 0$) during endurance testing, with 99% confidence is approximately 340 million kilometres are needed. This analysis applies to failure-free kilometres. For this reason, it is economically impossible to demonstrate the safety of automated driving systems with widespread usage statistically prior to introduction (approval trap).

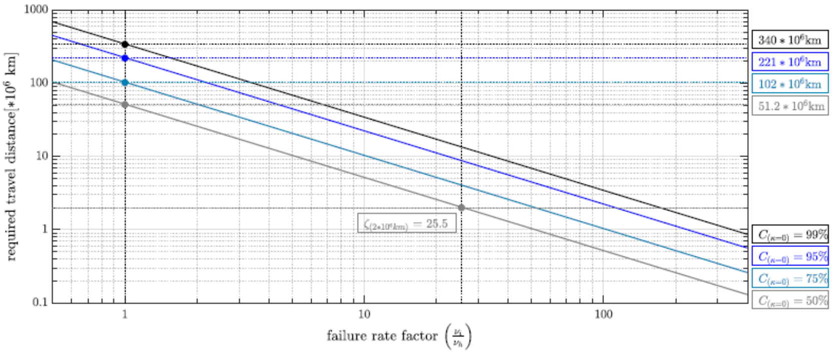


Figure 4: Failure free kilometers for a failure rate factor compared to human-driven truck fatality rate of year 2015

3.2 Distance between critical events

While critical traffic events are typically rare and not reproducible, early identification of functional deficiencies is essential for automated driving. Despite the difficulty of predicting a priori all possible operating scenarios, the coverage of critical driving scenarios needs to be adequately investigated. Recent research suggests the hypothesis of Poisson distribution to calculate the required validation distance with the following assumptions. On the one hand, the route used is representative; on the second hand, critical events occur independently of each other within a random process. In the equation 6, k corresponds to the number of accident events and λ is the predicted distance at which this event occurs at a given confidence level.

$$C_{(\kappa=m)} = \frac{\lambda^{\kappa}}{\kappa!}e^{-\lambda};\ \kappa = 0, 1, 2, \cdots, \infty \tag{6}$$

The mean time between failure (MTBF) can be determined at a given confidence level using the hypothesis of the Chi-square distribution according to ISO 26262. Accordingly, an exponential failure distribution with a constant failure rate is assumed. Regarding the safeguarding of driver assistance systems, there are no legal requirements for the validation distance. Since unintended reactions are rare events, a Chi-square distribution can be applied. If no critical event occurs at a sample distance with a required failure rate of one million kilometres each, the necessary validation requires around 3 million kilometres. In this case, no event should occur during the driven interval in order to argue the residual risk with a confidence level of 95%. The required mileage will increase if more events occur during validation (e.g. $x_{(\mathrm{k}=1)=}4.8 * 10^{6}$ km, $\mathrm{x}_{(\mathrm{k}=2)=}6.3 * 10^{6}$ km, $\mathrm{x}_{(\mathrm{k}=3)=}7.8 * 10^{6}$ km, etc

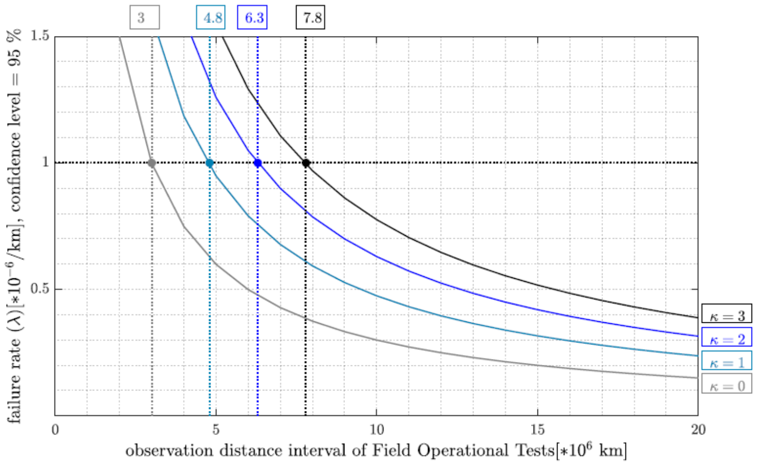


Figure 5: Required validation distance for various accident events using the chi-squared distribution with confidence level = 95%

In practice, the validation distance does not play the central role, but the variance of test conditions as much as possible (e.g. different weather conditions, time of day, road conditions, traffic conditions, pedestrian conditions, etc.) to cover rare operating situations. Therefore, route diversity in physical road tests is a significant measure of the probability distribution. Alternative safety assessment methods are therefore required, where the validation distance in long-term endurance tests will increase dramatically by using the current test concepts for automated driving without driver engagement.

3.3 Evidences of adaptive test coverage

User-oriented assessment procedures are the current de-facto standard for the validation of driver assistance functions. These methods provide system performance metrics, e.g. confusion matrices for all possible system reactions with classification as intended functional interventions or unintended side effects. In this scheme, the evaluation of software releases must be carried out in various phases up to the Start of Production (SoP). Initially by driving simulators, followed by X-in-the-Loop (XiL) technologies, proving grounds and long-term endurance tests. An optimised test strategy demands a selection of the necessary test method (simulation/laboratory, proving ground and field testing) for different scenarios and their interaction with other test methods[13]. Consequently, new innovative approaches need to be established, especially in simulation and laboratories. Efficient testing thus helps to achieve high development quality of new market-ready products in a short time span and at low cost. It will not be viable to prove safety of the required level of system performance through driving test hours alone during the development phase. Evidence should be provided by scenario coverage of the tests combined with statistical extrapolation techniques, field-based observations, component and integration tests including simulation as well as reasonable safety measures. The Goal Structuring Notation is used to highlight the verification methods for automated truck driving and explain main lines of the argumentation of adaptive test coverage, as illustrated in figure 6[14].

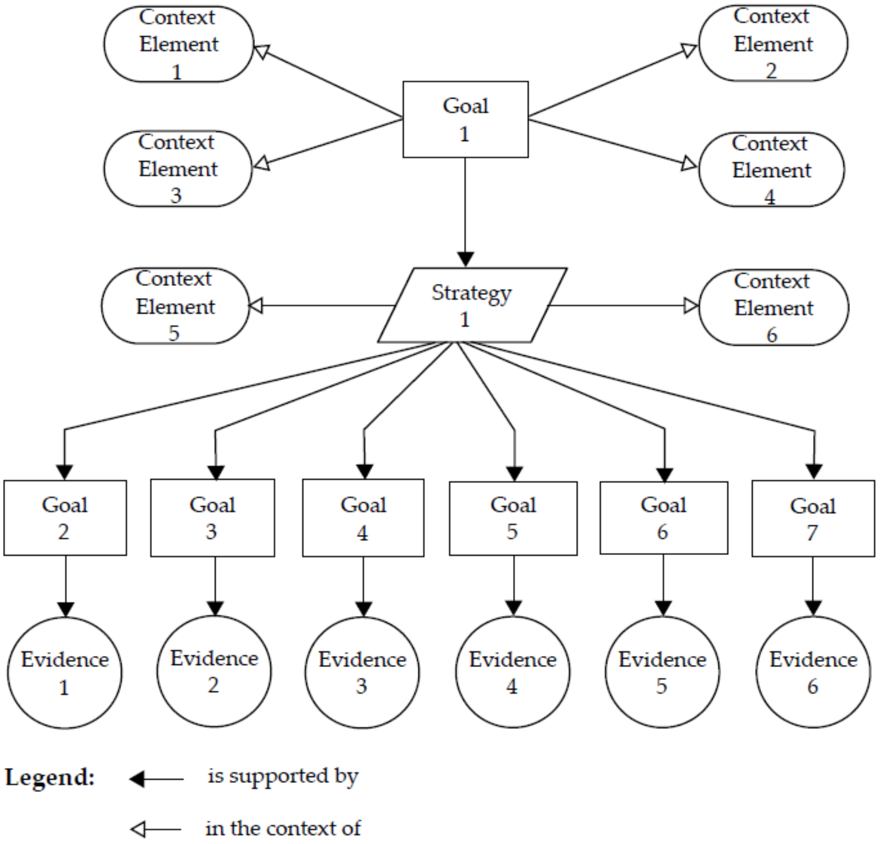


Figure 6: Adaptive verification case structure using Goal Structuring Notation

The following list contains typical context elements that are relevant for the adaptive verification case.

-Context Element 1 - definition of automated driving stages with regard to their required safety integrity:

-Context Element 2 - definition of functional requirements on the automated driving functions:

-Context Element 3 - definition of the system context including its quality gate within the series development:

-Context Element 4 - definition of the effectivity and efficiency criteria of the scenario-based test concept requirements:

-Context Element 5 - definition of field-based observations using clustering of multivariate time series analysis:

-Context Element 6 - definition of acceptable pass/fail criteria using criticality matrix:

The adaptive verification strategy can be described as follows:

Strategy 1 - argument on the required successful test completion through adaptive test coverage:

The Goal 1 forms the top-level claim and the sub-goals within the adaptive verification case are defined as follows:

-Goal 1 - residual risk associated with individual hazards in the automated driving is acceptable:

-Goal 2 - coverage of functional requirements:

-Goal 3 - coverage of algorithmic-based software structures:

-Goal 4 - coverage of system integration and variance:

-Goal 5 - coverage of software performance:

-Goal 6 - coverage of training dataset and uncertainties of machine learning pre-dictions: Autonomous vehicle technology typically involves some machine learn-ing, especially for object detection and classification. A driving function using non-deterministic algorithms (e.g. Bayesian estimators, neural networks, etc.) for object recognition is affected by failures of a different nature to those defined within IS26262:2018. The requirements are not in the typical V format of a set of functional requirements for the system itself, but preferably in the form of a set of training data or a plan to collect the set of training data due to the black swan problem. The gathered data is then annotated for specific features to be learned (e.g. road boundaries, pedestrians, cars). If the annotation is a manual process, a check of the annotations is required. The annotated data is then used to determine parameters through training. The training result is then verified using pass/fail criteria, such as acceptable false positive and false negative rates. If the self-verification fails, the process can be restarted after more data is obtained. Uncertainty quantification can provide information that is employed in object plausibility within sensor fusion algorithms. Two types of uncertainties can be distinguished. Aleatoric uncertainty covers noise that is inherent in the observation (e.g. sensor or motion noise). This uncertainty cannot be reduced by increasing training data. In contrast, epistemic uncertainty has the effect that for a given input class, the system performs inconsistently within a particular range of error. False negatives (not detected objects), false positives (ghost objects) and misclassification issues can be tuned using coverage of training dataset. The performance of machine learning algorithms relies on the amount of training dataset. The statistically relevant spread of operational situations can ensure an adequate coverage during training for environmental perception tasks.

- Goal 7 - coverage of critical driving scenarios:

Possible types of evidence that can be formulated to argue the required successful test completion, as follows:

-Evidence 1 - test results of Cluster-in-the-Loop tests and proving grounds: [15],[16], [17].

-Evidence 2 - test results of back-to-back tests:

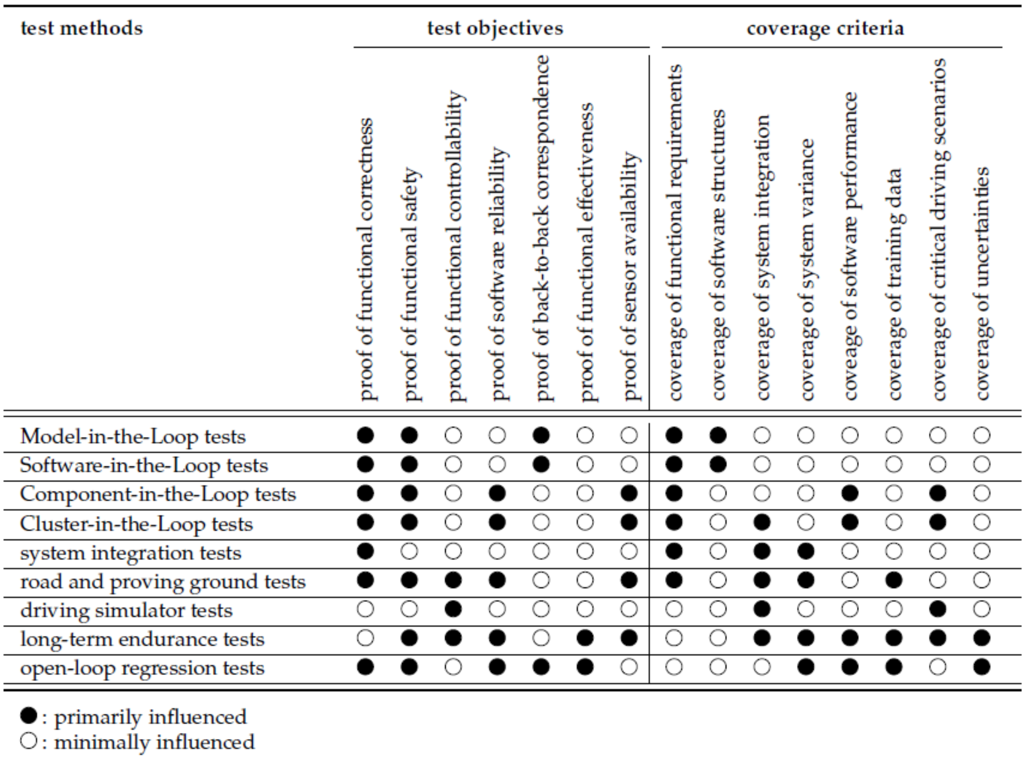
-Evidence 3 - test results of system integration tests and proving grounds:

-Evidence 4 - test results of fault injections tests: [18].

-Evidence 5 - run-time plausibility checks of open-loop regression tests: [19].

-Evidence 6 - field-based observations of triggered-based field operational tests:

Table 5 highlights the required systems engineering and verification techniques. The possible types of evidence that can be assigned by multiple test methods to achieve their test objectives and coverage criteria.



test methods	test objectives							coverage criteria							
	proof of functional correctness	proof of functional safety	proof of functional controllability	proof of software reliability	proof of back-to-back correspondence	proof of functional effectiveness	proof of sensor availability	coverage of functional requirements	coverage of software structures	coverage of system integration	coverage of system variance	coveage of software performance	coverage of training data	coverage of critical driving scenarios	coverage of uncertainties
Model-in-the-Loop tests	●	●	○	○	●	○	○	●	●	○	○	○	○	○	○
Software-in-the-Loop tests	●	●	○	○	●	○	○	●	●	○	○	○	○	○	○
Component-in-the-Loop tests	●	●	○	●	○	○	●	●	○	○	○	●	○	●	○
Cluster-in-the-Loop tests	●	●	○	●	○	○	●	●	○	●	○	●	○	●	○
system integration tests	●	○	○	○	○	○	○	●	○	●	●	○	○	○	○
road and proving ground tests	●	●	●	●	○	○	●	●	○	●	●	○	●	○	○
driving simulator tests	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○
long-term endurance tests	○	●	●	●	○	●	●	○	○	●	●	●	●	●	●
open-loop regression tests	●	●	○	●	●	●	○	○	○	○	●	●	●	○	●

●: primarily influenced
○: minimally influenced

Table 5: Assignment of test methods used for development of driver assistance systems to their test objectives and coverage criteria

4. Big testing process

The big testing process provides a set of scenarios from field-based observations as a complementary approach to the existing functional specifications. Therefore, the adaptive learning process identifies critical scenarios in large data sets to learn from the system experience in the field. As a result, the adaptive test coverage contributes a potential trade-off between efficiency and effectiveness for a scenario-based test concept.

4.1 Verification of automated driving functions

Due to the interaction of automated driving functions with the surrounding environment, testing becomes highly complex and verification cannot be realised using a single testing approach[20]. **Figure 7** shows the differences between structure-, situation-based open-loop and scenario-based closed-loop testing over time. In the beginning, the scene block represents a snapshot of the environment, including the scenery, the dynamic elements and all actors and their relationships. Consequently, the situation block provides a selection of a particular behaviour pattern for an individual triggering event. As a result, the scenario block offers a description of the temporal evolution between several scenes in a sequence of scenes. The test case contains a logical scenario with a set of parameters that are applied to its pass/fail criteria as to whether a system is operating according to its intended functionality. Structure tests are executed to test the structure coverage of algorithmic-based software components, like particular software functions and parts of the code. As a next step, situation-based open-loop testing generates driving situations from the required behaviour and evaluates the behavioural response without feeding it back into future situations. If all test cases are passed successfully, scenario-based testing is used to test the behaviour in a closed-loop setup. Scenario-based closed-loop testing specifies an entire scenario in a test case. This includes a sequence of scenes, actions, events and goals for the driving function.

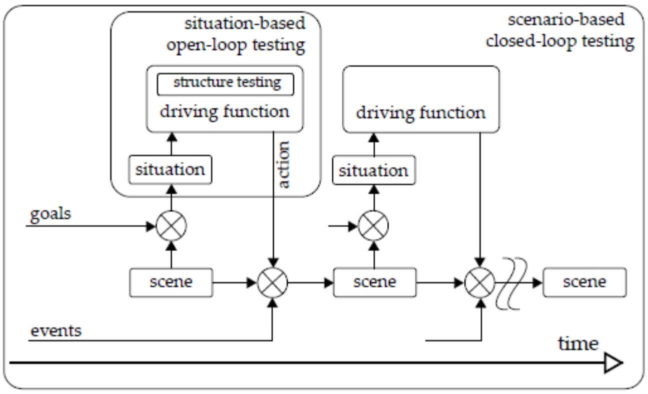


Figure 7: Illustration of the differences between structure-, open-loop and closed-loop testing

4.2 Big testing database

Big Data tools and techniques provide the measurement system and database infras-tructure needed to record, store and access data for re-simulation capabilities and root cause analysis. Besides, triggering events are utilised as a complementary information source for the discovery of critical driving scenarios experienced during long-term en-durance tests. The triggering event is a driving scenario with specific conditions, which serves as the initiator for a subsequent system reaction. For example, when an automatic emergency braking system incorrectly identifies a traffic sign as a preceding vehicle on a highway and leads to unintended braking. Therefore, the first step is logging and transmitting time series recordings of triggering events from the truck with its appropriate measurement system to the big data server. The second step is extracting and clustering of multivariate time-series data to provide them as a complementary information source to different suitable test environments. The root cause analysis can be divided into three subcategories(clustering, regression and classification). Initially, the clustering splits triggering events into one of several categorical clusters. Then, the regression represents each group with its corresponding signal prototype. Also, the open-loop prototypes can be converted to closed-loop control data to synthesise various driving scenarios applied to the Cluster-in-the-Loop test bench. In case of new triggering events, the classification maps them into one of the predefined categorical classes. Data from various sources (such as long-term endurance tests, accident databases, etc.) are formatted into a standard form to apply a typical processing chain[21]. Based on these steps, performance indicators can be measured for scenario group characterisation can be measured to derive complementary test specifications, as shown in figure 8. The knowledge discovery process is applied to identify critical scenarios from field-based observations and verify automated driving functions effectively and efficiently.

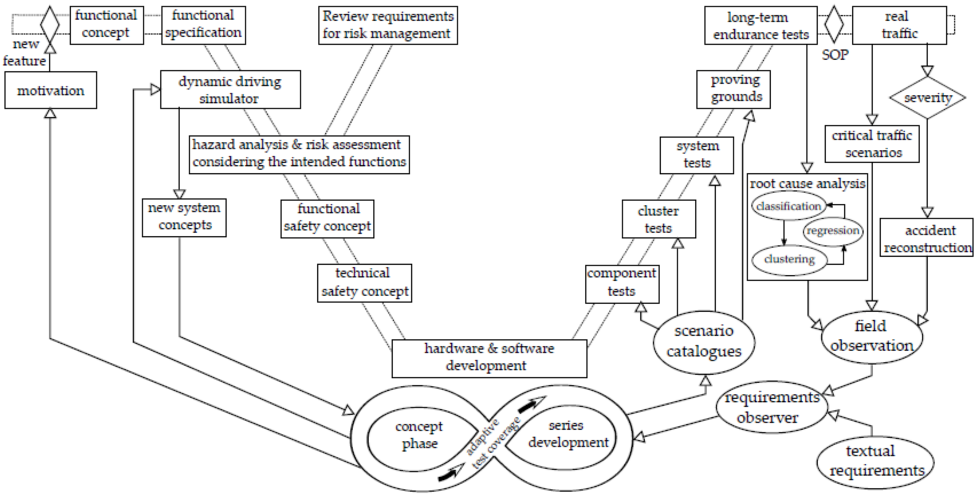


Figure 8: Adaptive learning process using the V-model and safety evaluation methods using big testing database

The knowledge discovery process can be summarised with the following steps to identify the scenarios from field-based observations.

4.3 Adaptive scenario-based test concept

The resulting control and behavioral response are used to influence future scenes and by this implicitly future situations, as well. The ontology-based method is identified to extract a category of adequate and relevant scenarios for existing Field Operational Tests. It presents a concept for semantic representation of worst-case scenarios. These are to be obtained using data-mining techniques and consequently systematically transformed into requirement test coverage. The proposed concept aims to bridge the gap between knowledge- and data-driven approaches to enable continuous extensibility of experience in an adaptive test coverage manner, as demonstrated in **Figure 9**.

5. Conclusions and future work

Since it is not possible to guarantee absolute safety for automated trucks, one of the biggest challenge in automated truck driving is to argue for a reasonably low residual risk resulting from imperfections of the environmental perception sensors. Such arguments are not currently supported by the relevant safety norms. This paper proposed applying an adaptive verification approach to determine how such an argument could be formed by decomposing the goals. The adaptive verification would be completed by providing systematically diverse evidence to support the claim that required successful test comple-tion through adaptive test coverage. The structure presented in this paper raises several issues that require substantial future research activities. Further research will also include the application of clustering of multivariate time-series data. These activities have to be integrated into a system engineering approach that supports the structure of the adaptive verification. This technical research work needs to be complemented by activities within industry to form a consensus on risk evaluation and acceptable argumentation structures that would feed into future standards and code of practice guidelines.

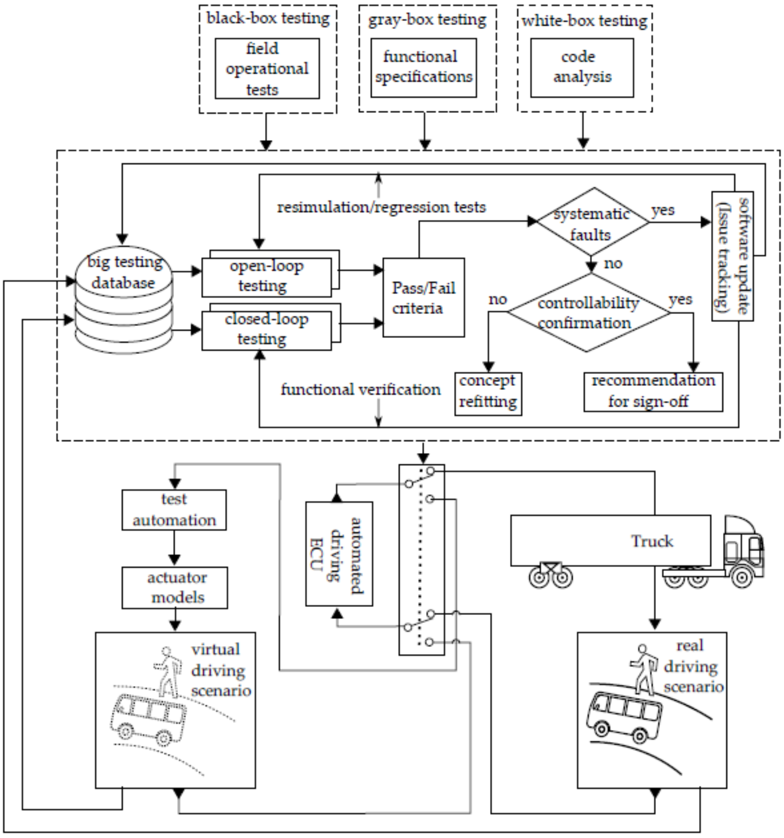


Figure 9: adaptive scenario based test concept based on field observations for sign-off recommendations

References

1. M. Kirschbaum, Highly automated driving for commercial vehicles, in: 6th International Munich Chassis Symposium 2015, Springer, pp. 5–15.
2. H. Winner, W. Wachenfeld, P. Junietz, Validation and introduction of automated driving, in: Automotive Systems Engineering II, Springer, 2018, pp. 177–196.
3. M. Kienle, B. Franz, H. Winner, K. Bengler, M. Baltzer, F. Flemisch, M. Kauer, T. Wei ßgerber, S. Geyer, R.Bruder, S. Hakuli, S. Meier, Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance, IET Intelligent Transport Systems 8 (2014) 183–189.
4. I. M. Sinclair, The Vienna Convention on the law of treaties, Manchester University Press, 1984.
5. B. W. Smith, Automated driving and product liability, Mich. St. L. Rev. (2017) 1.
6. W. Damm, P. Heidl, Safetrans working group highly automated systems:test, safety, and development processes, Recommendations on Actions and Research Challenges (2017).
7. S. O.-R. A. V. S. Committee, *et al.*, Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems, SAE International (2014).
8. K. R. Varshney, H. Alemzadeh, On the safety of machine learning: Cyber-physical systems, decision sciences, and data products, Big data 5 (2017) 246–255.
9. H.-P. Scho ¨ner, Challenges and approaches for testing of highly automated vehicles, in: J. Langheim (Ed.), Energy Consumption and Autonomous Driving, Lecture Notes in Mobility, Springer, 2016, pp. 101–109.
10. S. Burton, L. Gauerhof, C. Heinzemann, Making the case for safety of machine learning in highly automated driving, in: International Conference on Computer Safety, Reliability, and Security, Springer, pp. 5–16.
11. K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, H. Winner, Three decades of driver assistance systems: Review and future perspectives, IEEE Intelligent Transportation Systems Magazine 6 (2014) 6–22.
12. P. Wagner, Challenges in autonomous vehicle testing and validation, in: 2016 SAE World Congress.
13. A. Knauss, C. Berger, H. Eriksson, N. Lundin, J. Schro ¨der, H. Preenja, M. Ali, *et al.*, Proving ground support for automation of testing of active safety systems and automated vehicles (2017).
14. D. Kritzinger, 2 - safety assessment strategy (with goal structuring notation), in: D. Kritzinger (Ed.), Aircraft System Safety, Woodhead Publishing, 2017, pp. 23 – 35.
15. M. Elgharbawy, A. Schwarzhaupt, M. Frey, F. Gauterin, A real-time multisensor fusion verification framework for advanced driver assistance systems, Transportation Research Part F: Traffic Psychology and Behaviour (2017).

16. M. Elgharbawy, R. Schwarzhaupt, A. Arenskrieger, H. Elsayed, M. Frey, F. Gauterin, A testing frame- work for predictive driving features with an electronic horizon, Transportation Research Part F: Traffic Psychology and Behaviour (2018).
17. M. Feilhauer, J. Haering, S. Wyatt, Current approaches in hil-based adas testing, SAE International Journal of Commercial Vehicles 9 (2016) 63–69.
18. M. Elgharbawy, A. Schwarzhaupt, G. Scheike, M. Frey, F. Gauterin, A generic architecture of adas sensor fault injection for virtual tests, in: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–719. M. Elgharbawy, B. Bernier, M. Frey, F. Gauterin, An agile verification framework for traffic sign classi- fication algorithms in heavy vehicles, in: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–8.
20. S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, M. Maurer, Defining and substantiating the terms scene, situation, and scenario for automated driving, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 982–988.
21. A. Pu ̈tz, A. Zlocki, J. Bock, L. Eckstein, System validation of highly automated vehicles with a database of relevant traffic scenarios, 12th ITS European Congress 1 (2017) E5.
22. T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. Perez, J. Hankey, D. Ramsey,R.Gupta, *et al.*, The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment, Technical Report, 2006.
23. J. Dokic, B. Mu ̈ller, G. Meyer, European roadmap smart systems for automated driving, European Technology Platform on Smart Systems Integration (2015).
24. M. Miegler, R. Schieber, A. Kern, T. Ganslmeier, M. Nentwig, Hardware-in-the-loop test of advanced driver assistance systems, ATZelektronik worldwide 4 (2009) 4–9.
25. A. Pu ̈tz, A. Zlocki, J. Ku ̈fen, J. Bock, L. Eckstein, Database approach for the sign-off process of highly automated vehicles, in: 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.
26. V. Sasse, Autonomous driving from individualism towards collectivism, ATZelektronik worldwide 12 (2017) 72–72.
27. J. E. Stellet, M. R. Zofka, J. Schumacher, T. Schamm, F. Niewels, J. M. Zollner, Testing of advanced driver assistance towards automated driving: A survey and taxonomy on existing approaches and open questions, in: IEEE 18th International Conference on Intelligent Transportation Systems, IEEE, Piscataway, NJ, 2015, pp. 1455–1462.