Towards «type approval» of automated vehicles: Means of safety validation and simulation-based methods, in particular

Wolfgang Kröger* and Ali Ayoub†

Summary: Within the "type approval" system", mandatory in Europe, a progressive approach is evolving allowing for virtual testing as part of a generic assessment method including the demonstration of the automated vehicles' capability to cope with the most critical scenarios. Besides questions on completeness, parameter space explosion emerges as a main problem. Advanced methods such as data-driven probabilistic frameworks based on PCE and statistical learning techniques are proposed. Finally, issues of complexity of fully automated and connected vehicles are addressed, calling for a "systems approach" and methods which capture the entire system and go beyond traditional methods of reliability theory.

Keywords: automated vehicles, complex system, safety validation, simulation-based methods

1 Introduction

Automated vehicles (AVs) and automated driving systems, respectively, rely on machine learning (ML) during different driving phases and for different functions. Therefore, while recognizing patterns and interpreting sensor and image data, the ML driving algorithms learn and get better. Hence, reliability and safety, which are driven by software, are not static but change, hopefully increasing continuously with modifications and updates during operation. However, the question remains when these systems are sufficiently safe and mature to be released to public use and how to guide and organize the safety validation and certification process.

Already existing methods of safety validation of automated vehicles can be systematically categorized into physical track and real-world testing, simulations, extreme value theory, formal verification, and scenario-based testing (Junietz et al., 2018). Each of these methods has its pros, cons, basic assumptions, and limitations.

* Swiss Federal Institute of Technology Zurich (ETH) & Swiss Academy of Engineering Sciences (SATW), email: kroeger@ethz.ch.

† Nuclear Science and Engineering Department, Massachusetts Institute of Technology (MIT), email: aliayoub@mit.edu.

2 Safety requirements and validation

The question "How safe is safe (or just good) enough?" is crucial and well known from other domains. Commonly agreed, even mandatory reliability and safety targets do not exist yet for (cooperative and) automated vehicles[1].

We distinguish three levels of abstraction and related targets. First, at the highest level, self-driving vehicles should better perform than vehicles driven by "attentive" humans with state-of-the-art assistant systems, which realistically appears hard to achieve. Targets or thresholds may remain in qualitative forms or exist in quantitative forms for which accident/collision-free driving per distance (km), time between crashes, or compliance with risk curves and associated tolerability lines are proposed as metrics/substitutes. The draft of the EC Implementing Regulation suggests as indicative target, *hazardous errors from the vehicle equipped with Automated Driving Systems (ADS) should be at most of the rate of 10^{-9} per hour, derived from the minimum endogenous mortality risk*[2]. The UN-ECE working party on functional requirements FRAV proposed 10^{-8} per hour for accidents with fatalities and 10^{-7} per hour for accidents with light or severe injuries, *aiming to achieve a neutral or positive risk balance compared to human driving* (FRAV, 2020).

Second, at the system level, the manufacturer has to demonstrate by a robust design and validation process that the system (here specified for Automated Lane Keeping System (ALKS)) is free of "unreasonable" risks for the driver, passengers and other road users and compliance with road traffic rules is ensured (UNECE, 2021).
Third, at components and subsystems level, safety must be ensured by compliance with standards, such as ISO 262 62: 2018 for development of safety-critical functions and devices and ISO PAS 21448:2019 for demonstrating safety of the intended functionality (SOTIF), geared to identify real world scenarios.

Test-driving in real-world traffic environment appears to be the most logical way to validate the safety of AVs and to evaluate and improve systems' performance while allowing to take the complexity of the entire system into account. However, this might be dangerous, not scalable, and turned out to be impossible/very inefficient proposition as the needed vehicle kilometers to be driven -- to meet most relevant traffic situations -- are huge (billions) and the time needed would be by far too long (see also Kalra and Paddock, 2016). Therefore, various ways out of this dilemma have been identified and pursued.

As to the certification process two paradigms have been pursued in the past and become apparent today: on the one hand "self-certification" or "self-assessment" in the USA which encourages test-driving as much as possible and encourages industry to validate internally that designs meet best practice standards and a set of regulatory requirements, then leaving the remaining risk to automotive industry. The process is supported by a voluntary guidance document for industries and authorities with 10 priority safety design elements for consideration, such as *system safety,*

[1] RAND Corp. has carried out a comprehensive survey and developed *safety as a measurement, safety as a process and safety as a threshold* as categories of approaches for assessing AV safety (Blumenthal et al., 2020).

[2] Commission Implementing Regulation (EU) …/… on uniform procedures and technical specifications for the type-approval of motor vehicles with regards to their automated driving system (ADS), 2021, draft for discussions

operational design domain, object and event detection and response, fallback (minimal risk condition) validation methods, human machine interface, vehicle cybersecurity, crashworthiness, post-crash ADS behaviour and data recording, released by the Department of Transportation (USDOT/NHTSA, 2018).

On the other hand, the "type approval system" is mandatory for cars since 1998 in the European Union (EU) and adopted by other non-EU states. It operates with fully harmonized requirements, valid across borders, and heavily relies on international, notably UNECE[3] regulations for technical rules while single State authority will finally decide on the certification. An evolving "progressive approach" supports large-scale testing but does not solely rely on physical test-driving and allows for virtual testing methods. It encompasses guidelines under the EU exemption procedure (EC Directive, 2007 Art. 20) and provides a new legal framework (EU Regulation, 2019), applicable from July 2022. Following the expressed need, a new generic assessment method on automated driving has been developed which is based on three pillars: (1) audit of the manufacturer design/development process, (2) confirmation of the audit/minimum performance in normal and emergency conditions before market placement through testing and (3) confirmation of the audit after release through continual feedback from the operational experience. Testing includes the demonstration of basic driving capabilities (on public road, on test-track) and of the ability to cope with main critical scenarios (by desktop simulation/on test-track).

Critical scenarios are defined as a sequence or combination of situations for assessing the functional requirements for automated vehicles and involve a wide range of elements such as roadway layouts, interaction with different types of road users and objects as well as environmental conditions. They are classified as *logical, functional* and *concrete scenarios* with decreasing level of abstraction. Based on intensive use of available and specifically generated data as well as on results of theoretical studies, a large set of adequate and representative critical scenarios is fleshed out and agreed upon by key actors against which automated vehicles have to be tested. They should be transferred to a scenario database as a common framework for manufacturers and authorities, subject to continuous update (see Fig. 1 for illustration).

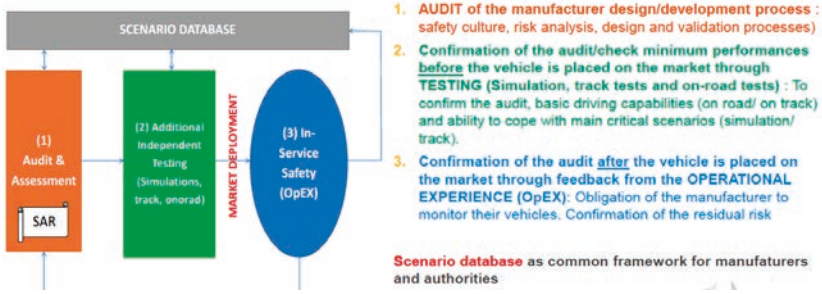

Fig. 1 Three pillars of the EU generic assessment method on automated driving (SAR stands for safety analysis report)

[3] United Nations Economic Commission for Europe

The application should start with “easier cases”, focussed on Level 3 up to Level 4 automated vehicles, and advanced assistance systems in particular, as building blocks towards entire autonomous vehicles.
To provide technical regulations and uniform provisions concerning the approval of automated vehicles the UNECE has set up a working party on “Automated/autonomous and Connected Vehicles” (GRVA) with a set of informal working groups as the group on Functional Safety of Automated and Autonomous Vehicles (FRAV) and on Validation Method for Automated Driving (VMAD) and the task Force on Cyber Security (OTA) and Software Updates (VMAD). In accordance with the Framework Document on the Safety of Automated Vehicles[4], the GRVA has proposed a regulation concerning the approval of vehicles with regard to Automated Lane Keeping Systems (ALKS) as first regulatory step for an automated driving system (ADS) in traffic and innovative provision aimed at addressing the complexity related to the evaluation of the system safety (UNECE/TRANS/WP.29 2020/8). The original text limits the operational speed to a maximum of 60 km/h, and further to passenger cars and activation of the ALKS under certain conditions on roads where pedestrians and cyclists are prohibited and divided lanes are physically separated. The general requirements relate to the system safety and fail-safe response; the possibility for the driver to override the system must be ensured at any time.

To support the process and to provide input to EC and UNECE working groups, a cooperative project between the EC DG GROW and the EC Joint Research Centre has been established. Their achievements include the draft of a voluntary Safety Guide with format and content of the information document for AV type approval, to be submitted by manufacturers and general procedures for different kinds of testing as well as contributions to ALKS regulation. This regulation should be expanded to Automated Lane Changing Systems (ALCS), motorway applications beyond 60 km/h maximum speed, and from passenger cars to other applications (e.g., valet parking, robot taxi/shuttles). Ongoing project work is focussed on testing concepts of general validity and for specific use-cases (e.g. highway chauffeur), on (quantitative) safety targets/thresholds, on analytical safety envelopes to define preventable accidents in traffic scenarios and on operational feedback recording and storage systems is ongoing and results are coming next.

As mentioned before, the services of the European Commission have expressed their views in a document with detailed annexes on uniform procedures and technical specifications for the type-approval of motor vehicles with regards to their ADS (see footnote [2]). Nevertheless, the whole design-simulation-test-redesign-certification procedure is still not established, neither by industry nor the regulator. Most recently, the working group VMAD of the working party on automated and connected vehicles (GRVA), established by the UNECE in mid 2018, proposed a validation framework called New Assessment/Test Methods for AV (NATM) to foster ongoing innovation in the automotive industry. The framework (see Fig. 2) is based on several pillars and five validation methods including a catalogue of critical scenarios and simulations. The testing might follow a logical sequence from simulation (based on use of various simulation toolkits), to testing on dedicated tracks and then real-world testing.

[4] UNECE/TRANS/WP.29/2019/34

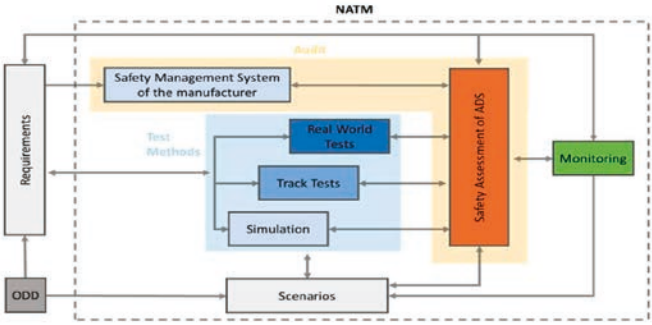

Fig. 2 Multi-pillar validation framework NATM and its integration with FRAV functional safety requirements (GRVA 2021), ODD stands for operational design domain

Harmonization of regulations is necessary and an ongoing effort at the UN level, in particular at the UNECE's Forum for Harmonization of Vehicle Regulation including Japan and China, among others.

3 Simulation-based methods and ways to reduce the parameter space

Simulation-based (or scenario-based) methods are a proposed alternative to the statistical approach of real-world testing and is favored as part of the type-approval process to validate automated driving functions. It is by far a much faster method of assessing advanced driver-assistance systems (ADAS) performance and safety where these systems have driven almost an order of magnitude more miles in simulation than in real-world testing. Based on the assumption that a large portion of road scenarios are uncritical in reality, it is proposed to identify "critical scenarios" out of a large set of developed scenarios and to expose single vehicles, equipped with automated systems, to exclusively critical scenarios to check their performance under "real world conditions" and on test benches in particular.

One problem with simulation-based methods is undermining the complexity of the real world and its uncertainties. This casts questions on coverage and completeness of scenarios driven in simulation (Kröger and Ayoub, 2021). More concretely, a good question to be answered in this direction is: how many miles driven in simulation equate to a mile in the real world?

Another main problem which emerges in the scenario-based approach is the parameter-space explosion. More abstractly, if the driving scenario has n parameters: P_1, P_2,⋯,P_n , and each parameter has m_i assumed discrete realizations, then the overall number of possible parameter combinations (scenarios) grows exponentially as m^n. For example, the cut-in scenario alone results in an order of 10^{23} parameter combinations or concrete scenarios (Amersbach and

Wimmer, 2019). This clearly indicates that a brute force method of trying all the scenarios doesn't work and even focusing on critical scenarios results in a huge number which still seems to exceed the existing capacity so far.

Novel approaches have been investigated to reduce the parameter space of critical scenarios and thus the test coverage significantly. Various attempts/proposals deserve attention. (Koné et al., 2020) proposed a hazardous behavior criterion with five severity classes for evaluation of scenarios identified by assuming functional insufficiencies. (Weber et al., 2020) proposed a simulation-based statistical approach to derive concrete scenarios for highly automated driving functions (SAE level 3 and higher) with a takeover process prompted by the vehicle. The methodology extends the framework of (Hallerbach et al., 2018) for the derivation of logical scenarios and encompasses the statistical evaluation and discretization of influence parameters identified by the traffic simulation package, their application to functional layers of the decomposed automated driving function, and finally a deterministic variation of previously discretized parameters which define a concrete scenario. The application to *cut-in* and *traffic-jam dissolution* functional scenarios showed a significant reduction of the parameter space.

Within a student's projects at ETH Zurich, a data-driven probabilistic framework was proposed, based on Stochastic Spectral Methods, namely Polynomial Chaos Expansion (PCE). Relying on both virtual and physical simulations of the autonomous vehicle system (considered as a black box), a dataset of input/ output pairs is used to train a metamodel, which surrogates an unknown, generically nonlinear criticality function, mapping the input scenarios to a risk metric. This so-called criticality function quantifies the severity of the input scenarios by weighting various Safety Performance Indicators obtained through experiments, and ultimately can help in reducing the input space dimensionality by identifying the most influential input parameters.

Another proposed methodology -- within a collaborative student's project between ETH Zürich and American University Beirut (AUB) – was to use statistical learning techniques to reduce the dimensionality of the exploding parameter space. One prominent example is using Linear Discriminant Analysis (LDA) to project the initial huge input parameter space into a lower dimensional one, spanned by the most affecting parameters. The basic idea behind LDA is to find a linear combination of the input features which results in the maximal separation of different classes into distinct clusters. In this case, the input features are the parameters of the scenario, and the different classes represent varying levels of crash criticality (a binary class critical/non-critical is used for simplicity). Using this linear combination of features, the higher dimensional data can be projected into a much lower dimensional space without losing a lot of information. By looking at the Euclidean components of the projected vectors, one can determine which features have more weight affecting the output-label distribution, thus, giving an idea of which feature bears more importance regarding the criticality of the crash.

Finally, (Zanella, Shehab and Ayoub, 2021) proposed a practical algorithmic framework combining surrogate modeling and importance analysis in series. It starts with the identification of the input parameters (scenarios) along with their probability distributions based on historical data and experts' knowledge. From the constructed scenario space, a set of input scenarios is created using Latin hypercube sampling, which are then fed to a self-driving simulator, CARLA (Dosovitskiy et al., 2017), outputting various metrics of criticality. At this stage, an importance analysis method, based on Linear Discriminant Analysis (LDA) or Morris elementary effects, is

implemented on this input-output mapping dataset to determine which parameters are the most influential. The original input dataset is then refined by refining the selected most influential parameters, resulting in an expanded dataset that now includes scenarios sensitive to the most important parameters, which are then fed again into CARLA to calculate their criticality values. The researchers then propose to use this refined input-output mapping to surrogate the criticality function using PCE to reduce the computational burden of evaluating new input scenarios. After the surrogate model is generated, an optimization routine onto the fully differentiable surrogate criticality surface finds the most severe regions (critical regions) as the portions of the hypersurface where the criticality is above a certain threshold, called the criticality limit – defined by the user. The critical scenarios are defined then as the n-dimensional algebraic vectors containing the coordinates on the criticality surface identifying the parameter combination where the criticality function is higher than the criticality limit. Finally, they proposed to introduce a re-sampling strategy that refines the already built surrogate by increasing the chance that critical scenarios are sampled during the training phase by performing an informed sampling circumscribed within the already identified critical regions.

4 Complex system-of-systems and challenges to methods

Some expect highly and, notably, fully automated (SAE level 4 and level 5) vehicles, connected to other vehicles (V2V) and infrastructure (V2X) to evolve into a “complex system” or even into a “system-of-systems” rather than just into a “complicated system”. This distinction, with associated elements and attributes as well as challenges to methods, is considered worth to be carved out.

The term “complexity” is not well defined. However, it is commonly agreed that complexity is something with parts interacting with each other’s’ in multiple ways, culminating in a higher order of emergence greater than the sum of its parts. According to (Aven et al., 2015), complexity is when “it is not possible to establish an accurate prediction model of system behavior based on knowing the specific functions and states of its individual components”.

Characteristics of complex systems versus complicated systems are highlighted as follows, see (Kröger and Nan, 2019) for more details:

- Both system types entail a huge number of highly connected components, for complicated systems event frequency-consequence curves tend to follow a normal distribution while such curves for complex systems tend to have fat tails and follow a power law distribution.
- Rules of interaction between the components of complex systems may change over time and may not be well understood, while components of complicated systems have well-defined roles and are governed by prescribed interactions.
- Complex systems are more open, respond to external conditions and evolve, interact with their environment; structures do not remain closed and stable over time and the range of responses to changes in their environment is not limited, all in contrast to complicated systems.
- Complex systems tend to show high dynamic, emerging, and non-linear behavior, as well as sudden regime shifts; behaviors are not fully predictable, opposite to complicated systems.

- The overall behavior of complex systems cannot be described in terms of building blocks or by the sum of its parts as in complicated systems.

Despite the early stage of development, we conclude courageously that attributes and behaviors of complex systems may fully apply to envisaged coordinated, highly, or notably fully automated vehicles and associated mobility concepts. In this respect, they are considered similar to large cyber-physical networks of critical infrastructure systems such as the power grid. Thus, methods are needed for the proof of reliability and safety which are capable of mapping and analyzing the system as a whole entity, calling for a "systems approach". Some adapted traditional methods based on "reliability theory" and thus on decomposition, like deductive Fault Tree Analysis (FTA), and causal chains, like inductive Failure Mode and Effects Analysis (FMEA) or Event Tree Analysis (ETA), which are successfully applied in other industrial sectors, may prove insufficient alone for safety validation of coordinated autonomous vehicles.

The traditional methods based on reliability theory have also been criticized because they focus on hardware component failures and do not sufficiently consider software failures and human interactions, in addition to not considering the system as a whole entity. STPA (Systems-Theoretic Process Analysis) has been developed (Leveson, 2011) to overcome these limitations in terms of identifying design errors, flawed requirements, human factors implications, software failures and unsafe and unintended component interaction failures. STPA uses a "feedback loop safety control structure" to identify unsafe scenarios and develops a detailed set of safety constraints/requirements and has been applied to various development aspects of autonomous vehicles. Various attempts have been made to compare STPA results with other methods (Kröger and Ayoub, 2021), as well as to combine them.

5 Short outlook

Vehicles of different degree of automation are under massive development and testing or even close to deployment. Certification requirements and rules are in the process of being structured at international and national level, with validation of sufficient functional and operational vehicle safety as well as the elimination of unreasonable risks as key elements. Adequate modelling and testing methods for different phases of development and safety validation are advancing and under early case applications, currently focused on advanced assistance systems. These efforts seem to be lagging the development of adequate methods for reliability/safety validation, at least for highly to fully automated cars.

Note: The first author is engaged in corresponding type-approval process activities at European level and ongoing methodological developments at academic level; progress achieved, and further results gained will be presented at the workshop.

References

[1] C. Amersbach, H. Winner, Functional decomposition – A contribution to overcome parameter space explosion during interactions of highly automated driving, Traffic Injury Prevention, volume 20 (1), 2019

[2] T. Aven, Y. Ben-Haim, H. B. Andersen, T. Cox, E. L. Droguett, M. Greenberg, S. Guikema, W. Kröger, O. Renn, K. M. Thompson, SRA Glossary, Council of the Society of Risk Analysis (SRA), 2015

[3] M. S. Blumenthal, L. Fraade-Blanar, R. Best, J. L. Irwin, Safe enough, RAND Corporation, 2020

[4] DIRECTIVE 2007/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Establishing a framework for the approval of motor vehicles and their trailers, and of systems, component and separate technical units intended for such vehicles, Sept. 2007

[5] FRAV, Common safety requirements of autonomous vehicles, FRAV-02-05/Rev. 2, Jan. 2020

[6] GRVA, New assessment/test method for automated driving (NATM), WP.29-183-05, March 2021.

[7] A. Dosovitskiy, G. Ros, F. Codevilla, A. López and V. Koltun, "CARLA: An open urban driving simulator", Conference on Robot Learning (CoRL), 2017.

[8] S. Hallerbach, Y. Xia, U. Eberle, F. Köster, Simulation-based identification of critical scenarios for cooperative and automated vehicles, SAE Technical Papers, April 2018

[9] P. Junietz, W. Wachenfeld, K. Klonecki, H. Winner, Evaluation of different approaches to address safety validation of automated driving, 21st International Conference on Intelligent Transportation Systems (ITSC), 2018

[10] N. Kalra, S. M. Paddock, How many miles of driving would it take to demonstrate autonomous vehicle reliability? RAND Corporation, 2016

[11] T. Koné, E. Levrat, E. Bonjour, F. Mayer, S. Géronimi, Safety assessment of scenarios for the simulation-based validation process of AV with regards to its functional insufficiencies, Proc. of ESREL2020-PSAM15 conference, Venice, 2020

[12] W. Kröger, C. Nan, Power systems in transition – Dealing with complexity, in C. Büscher, J. Schippl, P. Sumpf (editors), Energy as a Sociotechnical Problem, Routledge, 2019

[13] W. Kröger, A. Ayoub, Autonomous driving: A survey with focus on reliability and risk issues, Environment Systems and Decisions, 2021 (invited article, under review)

[14] N. G. Leveson, Engineering a safer world: Systems thinking applied to safety (engineering systems). MIT Press Cambridge, 2011

[15] REGULATION (EU) 2019/2144 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, applicable from July 2022

[16] UNECE, Regulation No. 157 on Uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping Systems, March 2021.
[17] USDOT/NHTSA, Automated driving systems 2.0, 2018.
[18] N. Weber, D. Frerichs, U. Eberle, A simulation-based, statistical approach for the derivation of concrete scenarios for the release of highly automated driving functions, AmE, GMM-Fachbericht, VDE, 2020.
[19] M. Zanella, M. L. Shehab, A. Ayoub, Autonomous driving safety validation: A method to overcome parameter space explosion using Importance Analysis, Surrogate Modeling, and informed sampling, Technical paper, SATW Document (under review), 2021