Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Article

Measures and Methods for the Evaluation of ATO Algorithms

Patrick Bochmann [1], Birgit Jaekel [1,*]

[1] Technische Universität Dresden; vlp@tu-dresden.de
* Correspondence: birgit.jaekel@tu-dresden.de; Tel.: (0049) 351-463-36786

Abstract: There is increasing interest in automating train operations of mainline services, e.g. to increase network capacity. Automatic train operation (ATO) is already achieved by several pilot projects, but not implemented on a large scale. Before the general introduction of new or adapted technologies can have a transformative effect on the operation of such a complex system as train operation on mainlines, they have to pass functional, interoperability and performance tests. A virtual preliminary analysis is one way to ensure a smooth as well as safe introduction and implementation. This paper aims to present an approach that applies to the performance testing of ATO systems. Therefore, methods and test standards for technologies enabling automatic operation in other transport sectors are reviewed. The main findings have been adapted, transformed and combined to be used as a general strategy for virtual performance testing in the railway sector. Specifically, universal performance indicators, namely punctuality, accuracy, energy consumption, safety and comfort, are presented. A layer model for scenario description is adapted from the automotive sector, as well as the definition of different scenario types. Lastly, factors that can influence the performance of an ATO algorithm are identified. To demonstrate the developed approach, a straightforward investigation of a case study is conducted using a microscopic train simulator in combination with an ATO algorithm.

Keywords: ATO, Simulation, Performance Evaluation, Scenario-based Testing

1. Introduction

Automatic Train Operation (ATO) will enable more efficient train traffic by increasing track capacity, reducing travel time, and lowering energy consumption [1]. All these benefits will become crucial due to a steady increase of global passenger mobility and freight transportation demand and the resulting necessity of infrastructural development [2].

The development of theoretical and technical principles to automate railway operations is already well advanced. Similar trends can be found in other traffic and transportation sectors that aim to take advantage of the vehicle and operational automation to realize more efficient and safer transportation. Certain systems automating the operation of trains are already qualified for series production and, by this time, are being used in some special fields of the sector like metro systems. However general automation of mainline operation is still not achieved [3].

To ensure efficient implementation of ATO, careful testing and analysis of the technical components and software used is required and achieved by demonstrating functionality and performance. Due to the impact of the introduction of such a system, virtual validation and testing are vital prior to verification on tracks and under real-world conditions.

Proving the functionality of a system is the responsibility of the manufacturer. Customers (infrastructure managers and train operators) must decide which system complies with their requirements and can be purchased. To make this decision, testing the performance of available systems is an essential step. Virtual testing can solve this issue efficiently by utilizing a simulation environment. By focusing on the core software elements of ATO,

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

this paper wants to contribute a structured approach to enable the virtual performance evaluation of ATO algorithms.

The general scope of the approach is envisaged to allow a flexible performance analysis of ATO algorithms. The tests are specifically focused on determining whether an algorithm fulfils operational requirements. Runtime analysis and software specific questions are not addressed by the proposed approach since these requirements should be met in general.

In Europe, ATO in combination with the European train control system (ETCS) is the preferred way. This is referred to as *ATO over ETCS*. An ATO system shall enable the train to regulate traction and braking automatically. The *ASTRail Consortium*, part of the European *Shift2Rail* project, was aiming for the identification of general functionalities of an ATO system and describes in [4] the tasks of an ATO system as follows:

- retrieving necessary data for automatic operation,
- calculation of optimal speed profile,
- journey planning while including restrictions as well as provided trackside data and defining the resulting trajectory.

Furthermore, based on the systems Grade of Automation (GoA) the following tasks need to be carried out by the ATO software components:

- giving control to the human driver regularly or in case of an emergency or failure,
- opening and closing vehicle doors,
- establishing communication with trackside infrastructure, e. g. platform screen doors (PSD), passenger information systems, and send out commands.

The use of ATO in different settings leads to the need to incorporate multiple train characteristics while simultaneously taking different track layouts together with challenging sections due to gradient changes, varying speed limits, or complex routing into account. In combination with changing technical trackside equipment even within a country and differences in environmental conditions, all these characteristics contribute to a complex set of external factors. [1,5]

The performance of ATO is connected to the combination of algorithms for the two main tasks of ATO, namely calculating the optimal speed profile and controlling the movement of the train. Different algorithms of each category exist with individual advantages regarding either the computational time or accuracy of calculations for speed profile optimization as well as meeting other criteria like driving comfort and energy consumption in the case of vehicle control. One major condition must be met by all algorithms taking into account the real-world application. The computational time needs to be sufficiently short. In [6] one to two seconds is considered acceptable.

It is to be expected that the performance of ATO systems varies due to the use of different algorithms. Especially under demanding restrictions, differences can be noticeable. Therefore, the performance evaluation could provide useful information, e. g. for the question of which system satisfies the requirements of the customer.

The remainder of the paper, following this introductory section, is separated into four parts. After a comprehensive literature review of methods and standards for virtual testing of operation automating systems in other sectors an approach is designed which provides concepts for measuring the ATO algorithm performance, a structured design of test scenario specifications, and the extraction of relevant and challenging test scenarios. Finally, this approach is used in specific performance-relevant test cases, by using a simulator provided by *ProRail*.

2. literature review

2.1. Research methodology

Automation is to be found in all traffic modalities. Hereby software plays an important role in planning and controlling vehicle movements. Systems of other traffic modes and how they were developed and tested can hold methods and approaches, which might be useful to transfer and adapt when testing software with similar scope of duties in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

railway sector. Therefore, a literature review was conducted to identify standards and strategies used in other traffic domains Figure 1.

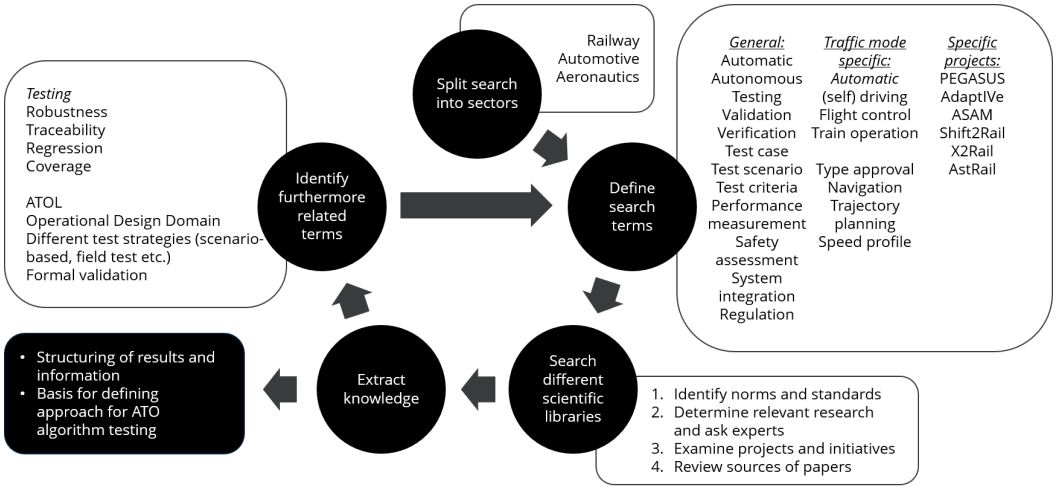


Figure 1. Visualization of the literature review process including examples for search terms and identified related terms.

Firstly, splitting the search into different traffic modalities allowed the identification of standards and norms that outline and give instructions on testing criteria and strategies for each sector. They form the basis for all subsequently used approaches. Finding these specific test strategies and methodologies was the main goal of the literature review process. The general terms *test strategies* and *methodologies* can hereby include in which form tests are performed, which criteria apply, how test scenarios were designed or test data was either extracted or generated.

2.2. Current state-of-the-art ATO test methods

One of the few resources of scientific research on ATO testing is based on the innovation initiative of *smartrail 4.0* and in connection to the European program *Shift2Rail*. The SBB and other Swiss railway companies and operators were testing ATO over ETCS on two different track sections. All specifications and results are explained in the reports [7] and [8] although a closer insight into the technology used is not possible.

Multiple test cases were executed testing the general functionality and in greater context the following characteristics:

- driving behaviour,
- punctuality,
- stopping position,
- energy saving,
- comfort for passengers,
- advantages for rail operation and train drivers,
- compliance of the TSI norm drafts with Swiss railway operation.

Already in 2019, ATO over ETCS was implemented in the core of London's rail network. The Thameslink program was aiming to enable a high-capacity passenger railway route by introducing ETCS Level 2 and ATO over ETCS functionalities. To introduce the system under real operational conditions several stages of testing were used. At first, a simulator was developed to showcase the technology and increase efficiency in the early development phase of the system. Furthermore, a system integration laboratory enabling a simulation of the Thameslink route was arranged. The next step was to conduct field tests at the ETCS National Integration Facility and finally operational tests on the Thameslink route itself. [9]

In [10] the authors are proposing an operational design domain (ODD) for a Chinese highspeed main line ATO system. The paper focuses on model validation and is including

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

a case study of the linkage scenario between train doors and platform doors to extract system requirements. In this context, the ODD is used for mapping components relevant to the case study. It is not demonstrated how the ODD contributes to the extraction of the interaction process the scenario model construction is based on.

Another example for improving interoperability and additionally maintainability is the initiative of defining an open-source version of the ETCS called *openETCS*. Between 2012 and 2015 a comprehensive toolchain for requirements identification, system modelling, code development, and verification as well as testing and further related development steps was established ([11]). Performance testing was not the main goal of this initiative rather than providing a reference system of the ETCS onboard unit (European Vital Computer) which allows the validation of commercially developed implementations.

Throughout all of these projects, testing, both virtual and on the track, is necessary and is performed to prove the function of the system. On the other hand, comprehensive performance tests and the evaluation of different algorithms to determine variances in performance could not be identified during the literature research. This confirms the goal of this paper to provide an approach to evaluate and compare the performance of ATO algorithms in a virtual environment.

While some standards for virtual testing in the railway sector has been identified, it could not be verified whether the described principles are used for testing ATO systems in current projects. Generally, the standardization of test methods and especially the therefore used formats to allow transferability can contribute to an improvement of interoperability and comparability of technical systems in the railway sector.

2.3. Testing in other modes of transportation

Automatic systems in road vehicles are implemented in great variety and quantity. Therefore extensive testing and evaluation of such systems is part of the industry's daily work. According to [12] the most used development and testing strategy is the so-called V-model, which is also used in aviation ([13]). Being a typical approach for designing, developing, and testing software it introduces test phases corresponding to the individual design phases and therefore leads to a more efficient development and design process. In addition, this model refers to the requirements as defined in ISO 26262 [14] and the certification specifications *CS-25* released by the European Union Aviation Safety Agency.

Standards for evaluating automatic vehicles (AVs) and their components, for example in the context of the type approval process in Germany, are not generally available yet. There are two main ways to test and validate the safety and performance of electronic systems in vehicle automation. First, testing virtually, and second, testing components in the real world with a further distinction between field testing in closed-off areas or real traffic situations. A complete list of safety assessment approaches for the introduction of AVs is given in the dissertation of PONN [15]. The author lists the strategies scenario-based and traffic-simulation-based testing for virtual system tests. In the survey, [16] the authors propose the use of scenario-based testing for safety evaluations at the overall system level.

Simulation-based testing can be focused on validating the functionality of a system under test. This can include checking compliance with different requirements by validating several levels of the software. For more details, the reader is referred to [17]. Traffic-simulation-based testing can be useful to evaluate the performance and safety of AVs. As an example, in [18] traffic simulation is used to estimate the impact of automated systems on safety. The whitepaper [19] is suggesting some important considerations while testing software with regard to aviation standard *DO-178C*. The author is proposing techniques, like parallel test case definition during requirements specification or the implementation of testing standards to check all important qualifications like code coverage, robustness and traceability. He further advises testing worst-case scenarios to verify given performance requirements.

In the context of evaluating the performance of an ATO algorithm traffic-simulation-based testing might be applicable if a simulator, capable of replicating a railway network, is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

used. For this research, the provided simulator can reproduce a train drive on a microscopic level, but without the ability to replicate a complex railway network. Therefore, traffic-simulation-based testing can not be used with the available simulation environment.

Performance and safety assessment is often done by evaluating specific predefined test cases also referred to as scenario-based testing. These scenarios describe relevant situations within traffic during the operation of an AV. Only critical or challenging situations are selected. However, it is not possible to identify all potentially occurring traffic scenarios due to an indefinite combination of external influences like mentioned in [20]. Scenario-based testing shows to be a promising virtual test strategy.

In [21] authors from the industry and scientific research proposed a first taxonomy for testing Advanced Driver Assistance System (ADAS). To successfully carry out virtual validation of systems and software components, references based on collected data need to be defined. In [21] two ways to retrieve ground truth information are listed:

- Reference by measurement in real-world traffic,
- References based on simulation.

The authors conclude that both are necessary to validate technical and software-based components of ADAS. The *references* are used to verify the outcome of test *scenarios*, which can be defined based on test criteria. Therefore, *metrics* are necessary to quantify the results of the test scenarios. These three elements combined yield a taxonomy for testing ADAS. It is outlined that test scenarios can be defined as either knowledge-driven or data-driven. Secondly, the scope of the scenario is different, either being used to evaluate the functionality or safety for specific use-cases with case-by-case evaluation or traffic-based, providing an overview of the system's overall impact and revealing unwanted side-effects (see also [22]).

In the field of avionics, scenario testing is used during the verification and evaluation of subsystems. In [23] a vision-augmented landing technology for general aircraft is presented. Simulations were used for functional validation and performance testing of essential algorithms of the system within a self-developed simulation environment. The performance was checked for "integrity, accuracy, availability and continuity" to reach common aviation standards ([24]). In [25] an automatic flight path controller for general aircraft was developed and described. Besides the description of the model, the paper also discusses the various testing stages that were carried out. First, model-in-the-loop testing provided the necessary model and functional validation. Hardware-in-the-loop and aircraft-in-the-loop tests were used for functional and implementation validation of the system. Performance tests were performed again by flight tests, also evaluating turbulent situations. Unmanned aerial vehicles make also use of automatic flight functionalities that allow an autonomous operation. In [26] a combination of different simulation sources is proposed to provide a testing environment where seamless transitions are possible. Using live data, virtual and constructed entities enable multiple inputs that can be handled in the same way by the system under test without the need to change the simulation environment.

The extraction and generation of scenarios is a major research field. In [16] a more refined taxonomy of the scenario-based approach is proposed. The important module of this strategy is the creation of a scenario database filled with scenarios either generated or more explicitly extracted from knowledge or data. Three different types of scenarios are generated. *Functional scenarios* can be described as linguistically expressed descriptions of a scenario abstractly. Defining functional scenarios more detailed by using parameters and placing them in a physical state space they can be called *logical scenarios*. Finally, assigning one specific value to each parameter results in the definition of *concrete scenarios* ([27]). A list of safety metrics to evaluate scenarios can be found in [22].

Standardization of tests and even more important standardization of tools, data exchange formats is beneficial to ensure an efficient and improved test and validation phase. Two major standards, according to [16] already known to most researchers and manufacturers in the automotive field, are *OpenDRIVE* and *OpenSCENARIO*. The first one is used to define static elements in a simulation, the latter is used to describe dynamic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

elements, such as vehicle manoeuvres [28]. Furthermore, the standard *ASAM XIL* allows the communication between test automation tools and test benches as well as transferring tests between different test systems [29].

Next to the efforts in standardization by the Association for Standardization of Automation and Measuring Systems (*ASAM*) the German federal ministry for economic affairs and energy launched a project called *PEGASUS* to "deliver[s] the standards for automated driving", as stated on the website of the project [30]. The goal of this program was to develop a cross-manufacturer method for testing and evaluating automatic driving vehicles and their components. In [31] an overview of the established method is given.

Within the project, the so-called *6-layer-model* was introduced which shall enable a structured design of an ODD (defined also in SAE standard J3016) and a simpler extraction of test scenarios. By SCHOLTES et al. the model was then further developed and refined. In [32] it is described how the original model was designed and what changes were performed to further improve it. By using this model a clear structure of an operational design domain shall be feasible.

Another way to structure an ODD is given in [33]. The proposed framework for automatic driving system testable cases and scenarios is using a "ODD classification framework with top-level categories and immediate subcategories". A method for deriving test scenarios, namely critical corner cases, to falsify the functionality of an automatic vehicle in certain situations is also proposed by PONN in [15]. This approach is also applying the layer model to retrieve test cases and set parameters for each scenario.

It can be concluded that most research in the field of automotive and avionics is referring to safety assessment and thus proofing the safe functioning of a system controlling a means of transportation. Therefore it is important that a huge amount of critical scenarios and situations can get tested to evaluate the behaviour of the systems under test to ensure sufficient performance in all possible real-world situations. The ambitions towards interoperability of test tools in the automotive sector are notable. Comparable efforts in the railway sector may accelerate the implementation of ATO systems.

It is observable that especially in the automotive sector a wide range of strategies is exploited. In contrast, there seems to be a clear structure as to which test methodology is used in particular test phases in the aerospace sector. Like in the automotive sector coverage of all relevant test scenarios is important but not finally achieved. Therefore, formal validation is used to reduce the amount of testing. However, the concept of traceability of design requirements can add value to the testing of ATO algorithms. Not in order to be able to relate certain code areas directly to a given requirement but instead it would be useful to enable traceability of performance changes to the initiating change in the test situation.

As a promising approach, the scenario-based testing strategy was identified. This strategy will be the foundation of the approach developed within this paper. This includes the already mentioned *6-layer model* and the different scenario types. Both will be transferred to the railway sector to make use of the benefits of a structured way to describe and generate test scenarios. A definition of test scenarios in combination with the use of defined metrics to evaluate the simulation results based on the proposed strategies can be useful for testing ATO systems. Especially the 6-layer-model can contribute to a structured test case definition.

A clear definition for the term scenario in the context of simulating railway operation is beneficial. Therefore, the literature-based definitions for a scene and scenario in [34] can be transferred to the railway sector.

Scene: A scene captures the current state of the environment and all dynamic elements as well as relationships between them for one specific point of time. This includes static objects (rail network, signs, stations, etc.) and dynamic elements (trains, passengers, environmental conditions, etc.). Interactions and relations between all elements are static within a scene.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Scenario: A scenario represents the temporal development of a sequence of scenes. The timespan of a scenario can vary. How the scenes and their elements change over time is influenced dynamically by external constraints.

3. Approach for ATO Testing

A survey [35] as part of the *X2Rail 1* (*Shift2Rail*) project showed that 75 % of the responding companies of the railway sector are in favour of harmonizing test strategies. This includes suppliers as well as customers, namely infrastructure management companies, and regulative authorities.

Defining a structured and general way for evaluating the performance of ATO algorithms is just one part of the overall necessity of harmonizing test strategies. In [36], HOFFMANN gives a detailed overview of the process of software testing, and based on that different test properties are outlined.

The envisaged test level of the approach is the system level. It can also be used for acceptance tests since one or multiple algorithms could be tested against certain data and criteria e. g. provided or set by the customer respectively.

Test criteria are defining the scope of the test, in other words, what is the aspect getting tested. This can either be a functional, operational, or temporal aspect [36]. The envisaged performance tests are part of operational and temporal testing. They are intended to give an impression of how well the system under test can handle certain test scenarios with increasing complexity.

HOFFMANN is listing black-box, white-box, and grey-box testing as different test methods. For the performance evaluations of ATO algorithms, black-box testing is the appropriate test method because normally the customer does neither have insight into the specifics of the algorithm nor its implementation. Thus, defining test cases based on the algorithm itself is not feasible.

These principles are the basis for defining a general test strategy. By incorporating the knowledge extracted during the literature review phase following major steps were identified:

1. Defining indicators for performance measurement,
2. Finding a strategy to enable structured scenario definition,
3. Identifying ATO performance relevant scenarios.

In subsequent research, it is necessary to find out how significant their influence is on the performance of an ATO algorithm.

3.1. Key performance indicators

Key performance indicators (KPI) are intended to provide a better understanding of whether an algorithm is able to fulfil the performance requirements under certain conditions in specific operating situations. They represent a universal and flexible metric, also allowing the comparison of multiple algorithms.

To gain a concise representation we suppose a multilevel structure, i. e., a top-level KPI is characterized by several indicators that can be measured or calculated and thus represent a quantitative value that allows further interpretation and comparison. These given KPIs are Punctuality, Energy and Safety and are commonly used in the railway sector when evaluating train operation. The same applies for the sub-level indicators of these KPIs. Within the literature of other projects in the context of testing ATO systems like [8], similar KPIs are used. Furthermore, accuracy, comfort and capacity were mentioned as relevant measures and are included in the presented approach. Table 1 is showing the complete list of the selected KPIs.

The calculation of the KPIs is performed for each track section, defined by a departure and arrival station (or point). This shall provide an additional grade of detail for the subsequent analysis of the measures. A more condensed evaluation is possible by determining each KPI for the complete journey of the train under test.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Table 1. List of KPIs for the evaluation of the performance of ATO algorithms.

KPI	Indicator	Unit
Punctuality	Departure deviation	$[s]$
	Passing point deviation	$[s]$
	Arrival deviation	$[s]$
Accuracy	Stopping position deviation at station	$[m]$
	Stopping position deviation at red signals	$[m]$
Energy	Energy consumption	$[kWh]$
Safety	Yellow signal approaches	-
	Yellow signal passages	-
	Red signal approaches	-
	Red signal stops	-
	Minimal differences to ETCS braking curves (vInd, vPerm, vWarn, vSBI, vEBI)	$[\frac{km}{h}]$
Comfort	Max. Acceleration	$[\frac{m}{s^2}]$
	Max. Jerk	$[\frac{m}{s^3}]$
Capacity	Meeting time windows	$Pass/Fail$
	Area enclosed by vPerm and actual speed curve	-

The KPI "Punctuality" seeks a detailed evaluation at multiple locations during a train journey. The comprehensive analysis of punctuality is of particular importance due to its significant operational impact. Besides the obvious evaluation points *departure* and *arrival*, passing points are part of the scenario definition. These are characterized by the fact that the train does not stop at such locations. For each point, a passing time is defined which allows an evaluation of the punctuality at this specific point of the journey. Emerging delays can be identified and possibly traced back to their origin. Moreover, it is possible to assess whether and how well the ATO algorithm can reduce given delays.

Passing points can also be *critical timing points* (CTP). CTPs are vital points within a timetable equipped with a time window that is characterized by the earliest and latest passing time. These time windows define a frame in which a train must pass the CTP in order to not negatively influence the timetable. Both delays and early arrivals are to be avoided. As a consequence, they can have a great influence on punctuality as well as on capacity as explained later.

The accuracy or precision of an ATO system is mainly defined by the accuracy of the stopping process and the resulting deviation of the specified and actual stopping position. A stop is not only necessary at stations but also in front of red signals. Thus, two indicators are introduced.

If using PSD in higher GoA, the acceptable range of deviation of the stopping position is significantly lower. In [4] "the need to the ATO function of achieving a centimetre-level accuracy for specific operations" is outlined. For stopping at stations with PSD a precision of less than 10 cm is seen as a prerequisite.

As ATO systems shall contribute to a more efficient drive, characterized by a reduced energy demand, measuring the total energy consumption is essential. The total consumption can be separated into different components such as traction and subsystem energy consumption as well as regenerative braking energy. The most important component of the overall energy use is the consumption of traction energy which is influenced by multiple parameters. Those influences can be traction resistance, air resistance, efficiency rates of system components and others.

In this paper, the energy consumption is calculated based on the data available within the simulation log files of the simulator. Therefore, the energy consumption is calculated per time step and summed up for each track section. In general, the following equations apply:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

$$F_{\text{trac}} = m \cdot (a - a_r), \tag{1}$$

$$E = \sum_{i=1}^{n} \frac{F_{\text{trac},i} \cdot v_i + F_{\text{trac},i+1} \cdot v_{i+1}}{2} \cdot \Delta t. \tag{2}$$

where F_{trac} is traction force in N, a_r is the coasting resistance $[\frac{m}{s^2}]$ determined by the simulator based on an empirical equation (*Davis equation*), a is the actual resulting acceleration $[\frac{m}{s^2}]$ and m is the train mass in kg. In Equation 2 E is the total energy consumption in $[kWh]$, v is the actual speed in $[\frac{m}{s}]$ and $\Delta t = t_{i+1} - t_i$ is the duration of the time step and n is the number of time steps.

Although ATO does not include the execution of safety-critical tasks, the performance of the connected algorithms can be evaluated by measuring safety-related indicators. These measures are well-accepted values and are commonly in use when examining rail operation.

- **Yellow Signal Approaches:** number of times driving up to a yellow signal,
- **Yellow Signal Passages:** number of times passing yellow signals,
- **Red Signal Approaches:** number of times driving up to a red signal,
- **Red Signal Stops:** number of times stopping at red signals,
- **Minimal Differences to ETCS braking curves:** minimal differences between the actual speed and the speed defined by the indication, permitted, warning, service brake indication and emergency brake indication speed curves of the ETCS on-board unit.

These values can be compared with a reference run on the evaluated route under standardized (optimal) conditions, allowing a statement to be made about the quality of the calculated speed profile. Furthermore, they can also serve as indicators for capacity analyses, whereby further interpretation is necessary.

The indicators describing the number of approaches of yellow and red signals as well as counting the passages of yellow signals and stops at red signals provide information on how unhindered a train can pass a track section. In addition, approaching a yellow or particularly a red signal is always a situation that involves a higher risk than passing a signal that indicates to proceed.

By calculating the minimal differences between the actual speed and the ETCS braking curves, information is obtained about how close the algorithm comes to forced braking or to what extent speed restrictions are observed. With a continuous evaluation of these values (for each time step), more in-depth analyses would be possible.

In addition to the aforementioned KPIs, customer satisfaction is also influenced by other factors related to comfort. One possibility to determine (parts of) this subjective feeling is to calculate the physical values of acceleration and jerk in order to get a statement about how comfortable or uncomfortable a drive is.

Based on relevant literature [37–39], ranges for the acceleration for the defined comfort levels are introduced (Table 2). The absolute value of jerk should not exceed $3\frac{m}{s^3}$. Braking and acceleration values shall remain equal to or below $1.5\frac{m}{s^2}$ to be considered comfortable. The proposed classification can serve as an indication of how to interpret the measured values. All defined values must be seen in the context of the concrete application. Note that comfort is only evaluated in terms of longitudinal motion.

Table 2. Comfort levels for absolute jerk and acceleration values based on the reviewed literature.

Jerk	Acceleration	Comfort level
$\lvert\Delta a\rvert \leq 1.5\frac{m}{s^3}$	$\lvert a\rvert \leq 0.7\frac{m}{s^2}$	Very comfortable
$1.5\frac{m}{s^3} < \lvert\Delta a\rvert \leq 3\frac{m}{s^3}$	$0.7\frac{m}{s^2} < \lvert a\rvert \leq 1.5\frac{m}{s^2}$	General comfortable
$3\frac{m}{s^3} < \lvert\Delta a\rvert$	$1.5\frac{m}{s^2} < \lvert a\rvert$	Uncomfortable

Mainly of importance to infrastructure managers is the capacity. To measure such a complex train operation performance indicator, it is necessary to define what capacity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

means in the context of the conducted investigations. In [40], capacity effects related to the introduction of ETCS are linked to a reduction of the train following times. This reduction increases in the number of trains that can travel on one route within a given period, e. g. one hour.

Because this approach is testing an ATO algorithm within a defined scenario and not by simulating a comprehensive railway network and determining its effects on the capacity within this network an evaluation of simpler indicators is chosen like the compliance with given CTPs. When a train passes a CTP within its time window, the full (or intended) capacity is reached under the condition that the time window can be respected with a time-optimal drive.

By comparing the actual and permitted speed (ETCS permitted speed curve), a potential to increase the speed driven on this section and thus to increase capacity can be identified. This way of identifying capacity potential was mentioned in [8] (Tests of ATO over ETCS by SBB), but not fully evaluated.

The difference between both speed curves can be examined by calculating the area enclosed by them. Therefore, Equation 3 can be utilized to stepwise approximate this area. This value is then comparable to a reference value, e. g. from a drive under optimal conditions.

$$A = \sum_{i=1}^{n} \begin{cases} (v_{\text{perm},i} - v_i) \cdot \Delta t & (v_{\text{perm},i} - v_i) \geq 0, \\ 0 & (v_{\text{perm},i} - v_i) < 0. \end{cases} \tag{3}$$

3.2. Layer Model

The identified key principle *Structured Scenario Definition* will be implemented in this approach by transferring the *6-layer model* (6LM) by SCHOLTES et al. from the automotive domain to the railway sector. To find comprehensive documentation of the original 6LM, the transferred model is based on, the reader is referred to [32]. A visualization of the proposed layer model for the railway sector in the context of the evaluation of ATO algorithms is presented in Figure 2.

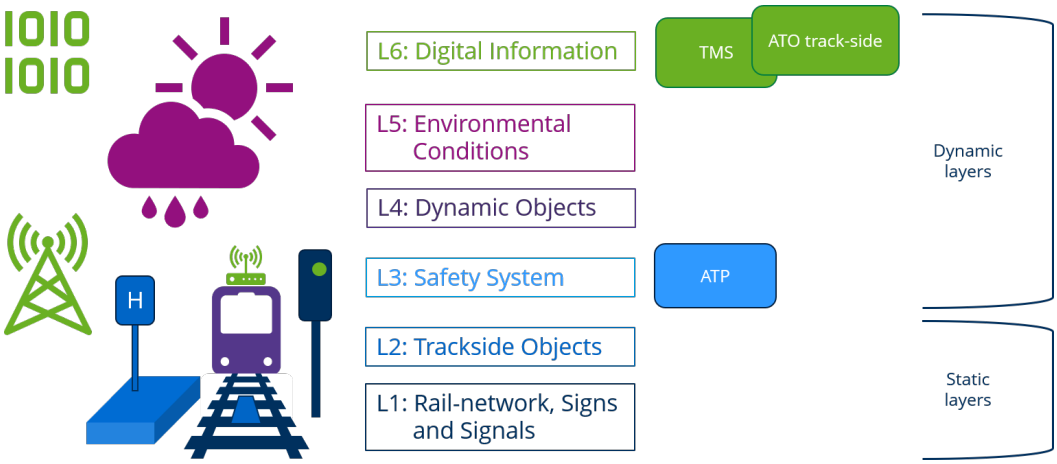


Figure 2. Layer model for the railway sector derived from the 6LM in [32].

All layers do contain entities, which are furthermore characterized by their properties. One entity can have a variety of properties and apart from that entities could include a definition of relationships to another or multiple entities as defined in the 6LM ([32]).

To improve readability, the following more detailed description refers to the layers of the 6LM using numerical phrases (one, two, etc.) and layers of the transferred model are addressed using cyphers (1,2, etc.).

Layer one, two and three in the 6LM are describing the static elements of a test case, thereby including spatial information of all important structures and elements which describe the static scenario components. Equal to these layers, layers 1 and 2 of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

transferred model were defined, whereby track elements, signs and signals (without the information, if it is not static) are subject to layer 1 and trackside structures like platforms, tunnels, bridges, PSD and balises are elements of layer 2.

Layer three in the 6LM describes temporal changes of layers one and two if they are consistently existing during the whole scenario. This is reasonable for road traffic because changes to the road marking, lane guidance and connected traffic rule changes (e. g. signs and traffic lights) can occur frequently. However, changes to the track position of rail lines during construction are very rare and, if present, are known in advance. The installation of fences and warning lights can serve as an example for temporal modifications on layer 2. Such changes will be subject to layer 2 for simplicity reasons, due to their limited occurrences and often long-term installation.

From layer three upwards dynamic elements of a scenario are implemented which do or can contain temporal information. Their state does not necessarily need to change during the scenario but changes can occur, e. g. the sun is shining from the beginning of a scenario till the end, or clouds are slowly covering the sun and it starts to rain.

The automatic train protection (ATP) is a vital function in train operation because it influences all operational phases. Layer 3, which was originally containing temporal changes to layers one and two in the 6LM, is now reserved for the ATP system, e. g. ETCS, its components and possible interactions. When testing ATO functionalities, influences by the ATP need to be evaluated. These influences can range from providing new movement authorities (MA) to intervening when the actual speed is exceeding the ETCS emergency brake indication speed curve or maximum track speed. As a result, the train might have to brake to a full stop before continuing its drive.

Layer 4 is similar to layer four in the 6LM. SCHOLTES et al. are referring to it as the "traffic layer" because it includes "movable objects whose movement could evolve over time". Dynamic objects in the railway sector are passengers (on a platform), other trains, cars driving on level crossings, or other vehicles and objects connected to railway operation like cranes for loading and unloading freight trains. They might not exist in such a great amount as in road traffic. Nevertheless, they can play a significant role within a scenario.

Layer 5 copies the equivalent layer of the 6LM. All environment conditions are included in this layer. In [10], proposing an ODD for ATO systems of Chinese highspeed trains the following environmental parameters are conducted: humidity, wind speed and temperature. Furthermore, the amount of precipitation, as well as the range of vision due to fog or rainfall, can also be added to the list of environmental parameters.

On Layer 6 all digital information is incorporated, like in the original 6LM. Besides the signal state, all information received from the ATO trackside component is subject to this layer. Examples of such information are journey profile updates with timetable changes, temporary speed restrictions, disruptions, rerouting, and more. Generally, all transferred information is modelled by this layer including connections of onboard systems.

Within the railway sector, several simulation environments are available either including their own scenario description language and format or making use of standardized formats in the sector. The latter can also be highly relevant when aiming for a universal scenario description. Potentially, the layer model can provide a basic structure for a new scenario description format.

3.3. Extraction of challenging scenarios for performance evaluation

A third step to evaluate the performance of an ATO algorithm is the extraction of challenging scenarios. This would mean generating a set of situations that could reveal limits of the algorithm under test or confirm the satisfaction of the given requirements.

Other transportation sectors use scenarios with a small time scope in the context of safety assessment. Scenarios for ATO algorithm performance evaluation need to be much longer than just a critical situation to allow the identification of performance changes. The envisaged scenario-based performance test approach evaluates comprehensive driving situations, e. g. a train route between two or more stations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Two ways to generate or define scenarios exist: data-driven and knowledge-based. The data-driven approach requires the existence of a sufficiently large database consisting of recorded traffic situations and drives under a wide range of conditions. Since such comprehensive datasets are not available in the rail sector to the author's knowledge, data-driven scenario definition is not explored further, but may be an interesting research direction.

The proposed approach is therefore employing the knowledge-based concept. Several scenario types were developed by BAGSCHIK et al. and described in [27]. They are utilized within this approach as well and consist of the following steps:

- Identification of functional scenarios,
- Extracting corresponding logical scenarios,
- Further development of logical scenarios into concrete scenarios with little effort.

3.3.1. Functional scenarios

According to the definition in [27], functional scenarios are describing a test case linguistically in a general way. The following text passage can be seen as an exemplary definition of a functional scenario for a train journey.

> The track (layer 1) includes three stations: A, B, C. Train X is starting its journey at station A, passing station B and arriving at station C. The journey is not interrupted by any halt signal. Train X does not stop in station B. (layer 6) Different weather conditions prevail (layer 5) and the movement authority is not provided in time. (layer 6)

Furthermore, for the concrete application in the railway sector functional scenarios can include many more information and entities such as general speed limits, the presence of static and dynamic objects and a general description of the environmental conditions, like the season and the presence of precipitation. To specify all those information, the layer model for the railway sector can be used. In the example, the layer the information is related to is given in brackets.

3.3.2. Base scenario

Before logical scenarios are defined allowing the variation of parameter values, a minimal scenario definition is a crucial requirement for the lengthy scenarios necessary for testing the performance of ATO algorithms. This minimal definition is introduced in this paper and will be referred to as *base scenario*. It can be provided either by designing a track from the scratch or by making use of already existing reference tracks.

A base scenario consists of a minimal set of scenario components. These requirements must allow a logical scenario to function in principle. Therefore, these minimal defined scenario components are:

- correctly designed track layout including track characteristics,
- presence of critical infrastructure elements,
- accurate use and positioning of safety system components,
- definition of a (standard) timetable,
- comprehensive train model and characteristics.

Bringing back the layer model for the railways, elements of the static layers one and two as well as parts of layers three and six need to be defined as a base scenario before logical scenarios can be set up. According to the simulator used and which default values need to be defined, these minimal requirements can be more comprehensive.

3.3.3. Logical scenarios

Functional scenarios introduce entities. If the description of those entities is appended with parameter ranges in the physical state space, one speaks of logical scenarios. They also can describe relationships between multiple entities. [27] By using the layer model for the railway sector a structured definition of logical scenarios is possible.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Following the example from above, by applying parameter ranges for the entities defined in the functional scenario, a corresponding logical scenario can be characterized as listed in Table 3.

Logical scenarios can introduce additional entities with parameter ranges for their properties. They could represent challenging factors whose influence on the performance behaviour of the ATO algorithm is to be investigated. The exemplary logical scenario can thus be titled: "Influence of a movement authority delay under certain weather conditions".

Table 3. Definition of entities and their properties of an exemplary logical scenario for a train journey based on the description of a simple functional scenario. Elements marked with an asterisk (*) do already have finalized values and are part of the base scenario, showcasing the difference between a base and a logical scenario.

Layer	Entity	Property	Parameter range
1	* Station A * Station B * Station C	Location Location Location	$0m$ $4800m$ $8000m$
2	* Platform Station A * Platform Station C	Length Length	$170m$ $350m$
3	-	-	-
4	-	-	-
5	Weather	Temperature Precipitation	$[10..30]°C$ $[5..8]l/m^2$
6	* Timetable Movement Authority	Departure A Arrival C Delay	$08:00h$ $10:00h$ $[0..240]s$

3.3.4. Concrete scenarios

Finally concrete scenarios need to be extracted. A concrete scenario is characterized by giving each parameter a fixed value from the parameter range previously defined in the logical scenario. Besides a fixed value, the change of a value over time is also an option, if the time and rate of change are fixed. Following the example from the other sections two possible concrete scenarios with the following specifications would be:

1. MA delay: $60s$, Temperature: $10°C$, Precipitation: $5l/m^2$,
2. MA delay: $90s$, Temperature: $10°C$, Precipitation: $5l/m^2$.

3.3.5. Identification of potential performance influencing factors

The structured identification of factors possibly influencing the performance of an ATO-system is based on the specific system characteristics of railways as presented in [41] and [42]. An extensive overview of resulting influencing factors is presented in section A. Please refer to the corresponding section at the end of the paper to learn how to retrieve this information.

For the general functional scenario *train journey* the subdivisions *Start*, *Drive* and *End* were introduced. The factors are therefore linked to the operational phase in which they might be relevant and sorted by the corresponding layers of the layer model. They are characterized by varying degrees of detail, due to the complexity of the topic they address. For example, an intervention of the safety system (e. g. because overshooting the emergency braking indication curve) is considered an influence on the ATO system the same way as the general timetable is treated as an influencing factor. Important to note, that it has not been proven that these factors do have a significant influence on the performance of an ATO algorithm. With this list, an overview of potential factors is created. To determine the effects and impact they provoke, a separate investigation is necessary.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

It can also be useful to choose a finer structure of the individual factors. For example, the aforementioned influences of the timetable could be separated. The current grade of detail for each factor was selected to give a first overview of possible influencing factors and to provide a structure by applying the layer model. Certainly, an extension of this list is possible, especially with country- and system-specific additions. Moreover, it is also relevant which train control system is used, which can lead to a multitude of further influencing factors or a reduction of the aforementioned.

3.4. Scenario evaluation and comparison

It is either possible to determine performance changes between multiple scenario variations or analyze the performance of multiple ATO algorithms. Figure 3 visualizes these two objectives.

If the performance of one algorithm is tested for several (challenging) situations, then the evaluation of scenario variations is necessary. Scenario variations are achieved by extracting a set of concrete scenarios from one logical scenario as described. In other words, the concrete values of one or several properties are varied. All other values stay as defined in the base scenario. By comparing the KPIs for those scenario variations it shall be possible to diagnose how extensive the influence of one (or multiple) changing parameters is on the performance of an ATO algorithm.

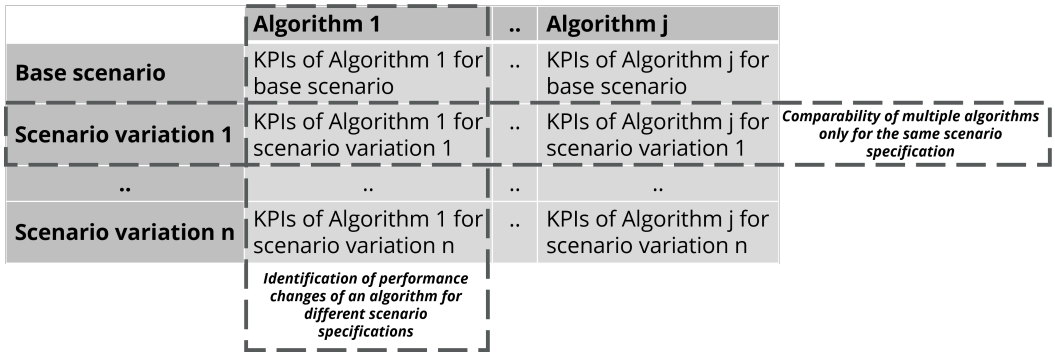


Figure 3. Two different scopes and the according ways to evaluate the results of concrete scenario simulations.

If the performance analysis is supposed to compare different ATO algorithms, then consequentially it is not useful to conduct such a comparison across different scenarios. To investigate the differences in the performance of two or more algorithms, the same scenario must be taken to ensure a level playing field. When testing the algorithms against the same scenario, it is possible to discover deviations in the handling of challenging situations. For this purpose, either the performance for the base scenario or the performance for the same scenario variation is comparable as illustrated in Figure 3.

4. Demonstration of Approach

This section describes a brief demonstration of the outlined approach. The simulation framework that has been used for an exemplary ATO algorithm evaluation consists of two components:

1. Driver Advisory System (DAS) named BEAOnline by TU Dresden (used as ATO algorithm),
2. Train simulator named NEO by ProRail.

BEAOnline was originally developed as a driver advice system and based on the offline functionality of BEA (Basic Energy Analysis), which can be used for energy consumption estimation [43]. The main functionality of an ATO and DAS algorithm, namely the calculation of the optimal speed curve, coincide. With BEAOnline, a DAS algorithm was refined for application as part of an ATO software.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

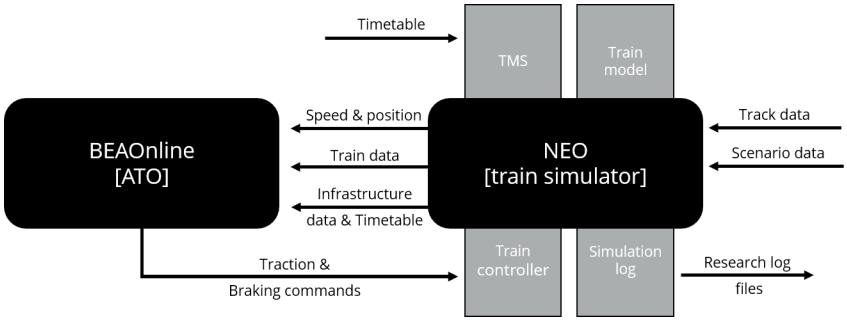


Figure 4. Visualization of the simulation framework consisting of two major components (train simulator and ATO algorithm), their subparts and the data exchange between them.

The NEO simulator by ProRail allows the microscopic simulation of one (or multiple) train drives. Thereby, train models and detailed virtual track images are necessary. The system is based on the original version named *MATRICS* following the concept of gaming simulation. It was used for several projects within ProRail either for demonstration or research purposes. [44] Furthermore, NEO is capable of generating comprehensive logging data, which is necessary to calculate the KPIs and thus evaluate the performance of the algorithm under test. By reading and analyzing all simulation log files, the KPIs for ATO algorithms are calculated and can be compared with other tests.

4.1. Test scenario definition

To demonstrate the use of the approach a functional scenario has been defined. The following influencing factor for ATO systems is the central element of the scenario:

> Different timetables (Layer 6) yield a challenging set of available running times including additional restrictions by CTPs

The resulting functional scenario can be described as follows:

Different timetables The train equipped with the system under test (BEAOnline ATO algorithm) is driving from a station on a reference line to a target destination and needs to stop at one additional station along the line. For this journey, different timetables are provided, which are characterized by the available running time. The timetable can include time windows for critical timing points. All other necessary definitions are given by the base scenario.

The case study is based on an existing reference route available for the NEO simulator. It represents the railway line from Schiphol to Zwolle in the Netherlands. Train operation is organized and constrained by timetables. Their change causes the algorithm to adapt the calculations and thus also the optimal speed profile to the newly arising constraints. Depending on whether the new limitations are tighter or less strict, the performance of the ATO algorithm may change. It is intended to investigate to what extent the algorithm can cope with varying running times set by the timetable.

The main goal of this demonstration is to show how the developed approach can be applied and whether the defined KPIs allow concluding about the performance of the algorithm in practice.

Based on the functional scenario two examples for logical scenarios are defined. In Table 4 and Table 5 a description of these logical scenarios is given. For both logical scenarios, one part of the reference line functions as the corresponding base scenario respectively including all the specific characteristics of the section. The two logical scenarios differ in terms of the concepts used to vary the timetables. As the section from Zwolle to Almere Centrum includes two CTPs before the stop at Lelystad Centrum, logical scenario 2 incorporates these additional constraints. In contrast to that, no CTPs are defined for the section of the reference line between Schiphol Airport and Almere assigned to logical scenario 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Table 4. Description of the entities and their parameters relevant for testing the influence of different timetables on the ATO performance. Only showing relevant definitions differing from the base scenario.

Layer	Entity	Property	Parameter range
6	Timetable	Reduction of available running time	$[0..6]min$

Table 5. Description of the entities and their parameters relevant for testing the influence of different timetables and the variation of TWs of CTPs on the ATO performance. Only showing relevant definitions differing from the base scenario.

Layer	Entity	Property	Parameter range
6	Timetable	Reduction of available running time	$[0..4]min$
		Moving TW	$[False, True]$
		Varying TW length	$[False, True]$

Regarding the defined parameter ranges of the timetable properties, concrete timetables were defined. They are changing the original reference timetable of the base scenario in such a way, that the different property parameter values within the defined ranges are achieved. For example, for a concrete scenario derived from logical scenario one, this could result in a two-minute reduction in available running time and for a second concrete scenario, a four-minute reduction. A reasonable gradation of the parameters needs to be selected according to the evaluation objectives.

Overall, in comparison with a reference drive under base scenario conditions with the standard timetable, performance changes shall be identifiable if existing.

4.2. Simulation results

For demonstration purposes and to evaluate the use of the introduced approach, nine simulation runs as part of two logical scenarios have been performed. Their results have been analyzed accordingly.

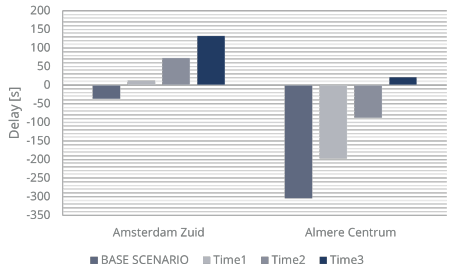
As intended, the KPIs were calculated based on the logging data of the simulation environment. In addition to the effects that could be expected from the introduction of restrictive boundary conditions in the timetable, important insights were gained about the simulation environment and the ATO algorithm itself.

Within Figure 5 the KPI values for different concrete scenarios of logical scenario one are shown exemplarily. Logical and expected results, e. g. as for punctuality, can be identified. Other KPIs need to be evaluated in a greater context. For example, the values of the accuracy KPI vary greatly from one concrete scenario to the other. However, this can be attributed to inaccuracies in the simulation environment, which resulted in a non-optimal stopping process. Regarding maximum acceleration and jerk levels, measured values for the available scenarios are within the defined limits. Energy consumption falls within the expected ranges as well.

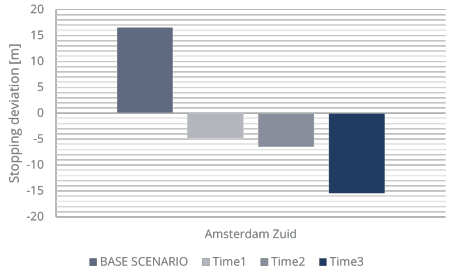
In the case of the capacity evaluation, a measurement area from the beginning of the track up to a speed restriction area before the station of Amsterdam Zuid was selected. This measurement allows determining how the algorithm tries to reduce the travel time to a minimum by following the allowed maximum speed as closely as possible. Within the base scenario, enough running time is available. Thus, it is not necessary to drive at full speed leading to a greater area enclosed by the permitted and actual speed curve. When the algorithm tries to achieve a time-optimal drive, the area enclosed by the two curves should be minimal, since it tries to maintain the maximum permissible speed as closely as possible.

In Figure 6 a visualization of the speed profile including relevant restrictions like the maximum allowed track speed and ETCS emergency braking curve is displayed.

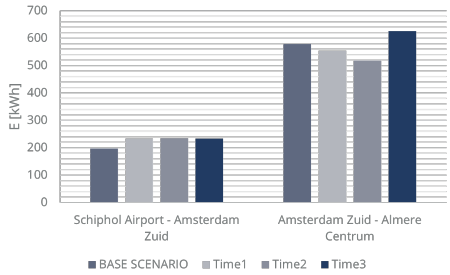
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1



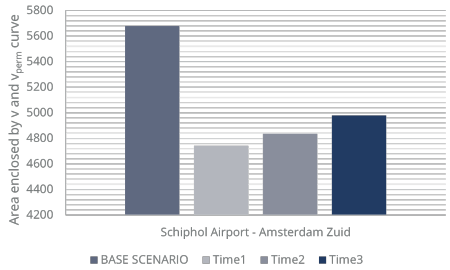
(a) Arrival delay at stations



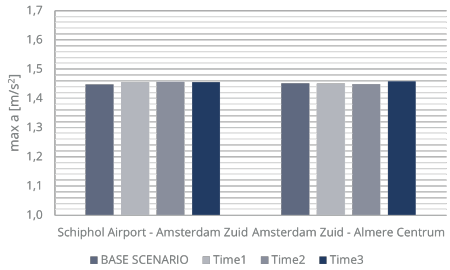
(b) Stopping accuracy at stations



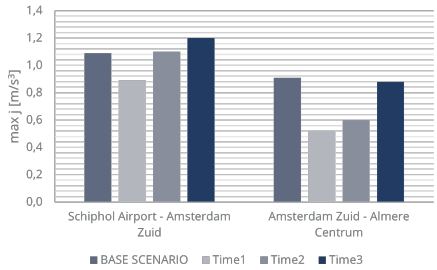
(c) Energy consumption per track part



(d) Capacity score



(e) Maximum acceleration per track part



(f) Maximum jerk per track part

Figure 5. Selection of plots comparing indicator values for concrete scenarios of logical scenario one.

Additionally, track data like station locations and height are given. This shall allow a more convenient and easier evaluation of the KPIs by providing context to the plain numbers.

The scenarios are illustrating how the algorithm is reacting to a set of challenging timetables. This is observable by the change of the KPI values. Additionally the speed plots provide viable information to further interpret the KPIs for example for a more in-depth analysis of weaknesses of the algorithm under test or to understand better the reason for certain specific behaviours.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

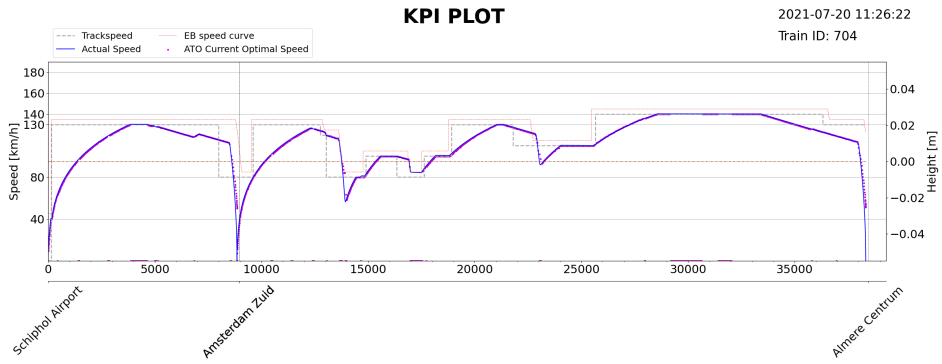


Figure 6. Additional plots generated by the evaluation tool for the original base scenario simulation of logical scenario one.

5. Conclusion

With the approach presented in this paper, two set goals are achieved, namely identifying a way to efficiently test ATO algorithms in a virtual environment and enabling the comparison of multiple algorithms based on logical measures. Its application may allow train operators and infrastructure managers to select the best suiting ATO systems according to their requirements.

The approach consists of universal KPIs to measure the performance of ATO algorithms, a framework for the structured definition of test scenarios taking advantage of the layer model for the railway sector as well as different scenario types, and the identification of performance relevant factors.

The KPIs introduced for this approach can provide valuable information about the performance of an ATO algorithm in specific situations and scenarios. It proved difficult to strike a balance between the need for further interpretation of the values and providing quantified metrics that are comparable when scenario definitions change. Calculating the KPIs for each departure and arrival pair results in a more or less extensive list of values, depending on the length of the scenario. If the KPIs were determined only for the entire run, a simpler but less detailed overview of the performance for a given scenario would be possible.

Furthermore, it is necessary to provide a range or overview of reference values for some indicators to be able to classify the measured quantitative values. This proved to be particularly difficult for the comfort indicator since the classification of these values is not commonly realized.

How the capacity KPI is evaluated in this approach might not provide all information important for capacity determination. Because the scope of the simulator is limited to a microscopic assessment, the capacity evaluation is adapted to this characteristic. If analysis on a network scale was possible, more general statements potentially of greater relevance and significance may be drawn.

Within the approach, different scenario types are introduced. These are helpful to infer from general test cases to exactly defined, complete, and thus also simulatable scenarios in a structured manner. However, the proposed approach only abstractly defines those scenario types. This decision was made due to the specific demands of each simulation software concerning the scenario definition. If more standardization for design and exchange of scenarios is achieved, an improved and standardized concept for delineating the different types of scenarios will also be beneficial and easier to establish. With the layer model, a foundation is laid, based on the definition of entities and their properties.

Another critical aspect of the approach is the identification of factors potentially having an impact on the ATO performance. Their extraction from the literature leads to an unequal grade of detail. What is meant by this is that one factor can have a variety of impacts on the performance of ATO systems due to the different effects it induces. Influences of several causative constituents of railway operation were partially combined, or the causing entity

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

itself was included as a factor. This was done with the intention of achieving conciseness of the overview but apparently resulted in redundancies. The list can certainly be extended and umbrella terms might be divided into more specific impacts on ATO.

During the identification of the factors having a potential impact on the ATO algorithm performance it was ascertained, that the layer model is applicable for defining the simulation scenario, the entities within those scenarios, and also including the definition of potential influencing factors of these entities. In contrast to that, it does not provide a possibility to allocate factors directly related to the train under test, the so-called *ego train*. More precisely, taking the impact of a reduced traction or braking power of the train probably not known to the ATO algorithm as an example can lead to an unexpected behaviour also provoking an unsafe situation. This factor needs to be evaluated as well and therefore is appending the list of influencing factors (Table A1) by a new row. This row is not assigned to a layer of the layer model and thus resembles an additional category called *Ego Train*.

Besides the identification of many potential performance influencing factors, a brief evaluation of the impact was only done for different timetable definitions including CTP variations as part of the demonstration of the approach. More profound analyses are necessary.

The approach focused on providing universal measures for evaluating and comparing ATO algorithms. However, it showed to be interesting to analyze the information about the calculations of the algorithm which are directly provided by the algorithm itself within a log file.

More importantly, the proposed algorithm evaluation strategy is useful to test the performance of ATO algorithms, especially due to the structured scenario design framework, allowing an organized investigation of the potential performance influencing factors. Furthermore, the KPIs seemed to be selected appropriately for the small case study conducted. To fully determine if this is the case when aiming for the comprehensive evaluation of an ATO algorithm and the comparison of multiple algorithms, additional explorations are crucial.

6. Future Research Perspectives

In this work, an approach was presented which emphasizes the importance of testing ATO algorithms for their performance behaviour in different challenging situations.

First, an improvement of the approach, as well as the other components of the framework based on the shortcomings described in the previous section, is desirable. This also includes the detailed testing of the different influencing factors to evaluate their actual impact on the performance.

The more specific definition of standards can be seen as a further opportunity for improvement. Here, the definition of a uniform scenario format, for example, based on the layer model, or the consideration of already existing formats is possible.

In general, the question remains open which level of performance can and needs to be expected under varying operating conditions. In this context, the KPIs can provide an important indication of the values to be achieved, but the definition of these performance criteria seems only to make sense with simultaneous consideration of the real routes existing in the network and their characteristics.

While for the indicators of comfort, stopping accuracy, and possibly punctuality, generally applicable values can be defined and have been already partially provided in the scope of this approach, an evaluation of energy demand, capacity and the individual indicators of the safety KPI is route-dependent and will have to be substantiated by empirical values and feasible targets.

Lastly, if the specifications for ATO over ETCS are finally realized a further analysis of the interfaces defined within those specifications can be worthwhile. These interfaces have to certainly include messages, status updates and variables which need to be sent out by the ATO onboard or trackside unit. Because of the therefore necessary standardization, useful

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

data might be available generally. This data could allow more extensive ATO performance analyses but is limited to systems under the ATO over ETCS concept.

Author Contributions: Conceptualization, P.B.; methodology, P.B.; software, P.B; validation, P.B.; investigation, P.B.; writing—original draft preparation, B.J. and P.B.; writing—review and editing, B.J. and P.B.; visualization, P.B.; supervision, B.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank Jelle van Luipen from ProRail, the railway infrastructure manager in The Netherlands, for supporting this work through thorough discussions and the internship opportunity with ProRail. It was a great chance for learning and professional development.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ATO	Automatic Train Operation
ETCS	European Train Control System
GoA	Grade of Automation
PSD	platform screen doors
SBB	Schweizerische Bundesbahnen
TSI	Technical Specifications for Interoperability
ODD	Operational Design Domain
ISO	International Standardization Organization
AV	Automatic vehicle
ADAS	Advanced Driver Assistance System
ASAM	Association for Standardization of Automation and Measuring Systems
SAE	Society of Automotive Engineers
KPI	Key performance indicator
CTP	critical timing point
6LM	6-layer model
ATP	automatic train protection
MA	Movement Authority
DAS	Driver Advisory System
BEA	Basic Energy Analysis

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Appendix A List of influencing factors

As described in the paper, based on general literature and knowledge about railway operation, the list (Table A1) of potential influencing factors on the performance of ATO algorithms and systems was created. This section is providing a few explanations of the table and its structure.

All factors are assigned to the respective layer of the layer model in which they occur and are relevant. The grade of detail varies, whereby it was the goal to give a comprehensive overview of possible factors. Thus, the individual factors might represent the result of multiple different reasons, e. g. the reason for speed restrictions can be manifold as shown in the table. On the other hand, the influence of tunnels and bridges is not further separated, because their impact on the performance needs to be evaluated. They might be relevant due to changing wind characteristics and light conditions.

The last column highlights the one or multiple KPIs that could be influenced. Following abbreviation apply:

Punctuality: P | **Accuracy:** A | **Energy:** E | **Safety:** S | **Comfort:** Co | **Capacity:** Ca

It is clear, that most of the factors do also have an indirect influence on most of the KPIs. Thus, only the most relevant KPIs according to the author are listed. This is based on an assumption, which itself is supported by the relevance of the individual factors for railway operations. Some factors are subsequently self-explanatory and their influence on the performance of ATO algorithms is logical and obvious, while others need to be evaluated first, to determine, which impact they can have.

Table A1. Overview of all factors capable of influencing the performance of an ATO system and especially the ATO algorithm. Structured by the layers of the layer model of the railway sector providing information about the operational phase they are relevant to.

Layer	Factor	Departure (Start)	Drive	Arrival (Stop)	Relevant KPIs
1	Speed restrictions - Constant - Temporary - Curves - Switches - Slopes	X	X	X	P,E,Ca
	Static friction coefficient (adhesion)	X	X	X	A,S,Co,E
	Gradients (changes)	X	X	X	E,P
	Curve radius	-	X	-	Co,P
2	Balises - Functionality - Position	X	X	X	A,S,Co,Ca
	PSD - Functionality - Connection	X	-	X	P,Ca,S
	Signals - Malfunction	X	X	X	S,P,Ca
	Tunnel	(X)	X	(X)	E,P,S
	Bridge	-	X	-	E,P,S
	Catenary (Energy supply)	X	X	X	E,P,Ca

(To be continued)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

Layer	Factor	Departure (Start)	Drive	Arrival (Stop)	Relevant KPIs
3	ATP intervention (EB)	X	X	X	S,P,Ca,Co,E
	Release speed	X	X	X	S,Ca,P
	Level changes (ETCS)	X	X	X	S,Ca,P
	Connection loss of ERTMS	X	X	X	S,Ca,P
4	Preceding train	X	X	X	P,Ca
	Following train	X	X	X	Ca,E,
	Passengers (on platform)	X	-	X	P
	Vehicles at level crossings	-	X	-	S,P,Ca
	Objects on track	X	X	X	S,A
5	Wind	X	X	X	E,P,Ca
	Temperature	X	X	X	E,A,P
	Humidity	X	X	X	E,A,P
	Precipitation	X	X	X	E,A,P
	Fog (sight limits)	X	X	X	S,A,P
	Time of day	X	X	X	P,Ca
	Lighting condition	X	X	X	S
6	Change of train number Timetable	X X	X X	X X	P,Ca P,Ca,E
	Timetable update - Delayed departure - Earlier arrival - New CTPs - Changing CTP TWs - Less available rt - More available rt	X	X	X	P,Ca,E
	Movement authority - Missing - Shortening - Extension	X	X	X	P,Ca
	Journey profile - Update - Wrong - Missing	X	X	X	P,Ca,E
	Segment profile - Update - Wrong - Missing	X	X	X	P,Ca,E
	Deviations in expected and actual train characteristics	X	X	X	S,E,P,Ca
Ego train	Maximum traction	X	X	-	S,P,E,Ca
	Maximum braking	-	X	X	S,P,E,Ca

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

References

1. Tasler, G.; Knollmann, V. The introduction of highly automatic operation - towards fully automatic train operation. *SIGNAL + DRAHT* **2018**, *110*, 6–14.
2. International Energy Agency. The future of Rail. Technical report, IEA, 2019.
3. Akbari, M.; Hoogewoonink, B.; Godziejewsk, B.; Rajabalinejad, M. Expediency of ATO in heavy rail: A survey for the Dutch Railways. *MATEC Web of Conferences* **2020**, *314*. doi:10.1051/matecconf/202031401005.
4. Sirti. D3.2 - Automatic Train Operations: implementation, operation characteristics and technologies for the Railway field. Technical report, ASTRail Consortium, 2019.
5. Kessell, C. Main Line ATO Evaluated. *Rail Engineer* **2017**, *150*, 20–24.
6. Yin, J.; Tang, T.; Yang, L.; Xun, J.; Huang, Y.; Gao, Z. Research and development of automatic train operation for railway transportation systems: A survey. *Transportation Research Part C: Emerging Technologies* **2017**, *85*, 548–572. doi:10.1016/j.trc.2017.09.009.
7. Nolte, J.; Wanner, F.; Matthias, M.; Kyburz, M. ATO 2 Basic Phase 1 Abschlussbericht. Technical report, SBB, 2019.
8. Nolte, J. ATO2Basic Phase 2 Abschlussbericht. Technical report, SBB, 2020.
9. Hartwell, G. A Review of the Thameslink Programme. Technical report, Institution of Railway Signal Engineers Australasia, 2015.
10. Meng, Z.; Tang, T.; Wei, G.; Yuan, L. Analysis of ATO System Operation Scenarios Based on UPPAAL and the Operational Design Domain. *Electronics* **2021**, *10*. doi:10.3390/electronics10040503.
11. Karg, S.; Raschke, A.; Tichy, M.; Liebel, G. Model-driven software engineering in the openETCS project. Proceedings of the ACMIEEE 19th International Conference on Model Driven Engineering Languages and Systems; Baudry, B., Ed. ACM, 2016, pp. 238–248. doi:10.1145/2976767.2976811.
12. Börcsök, J. *Funktionale Sicherheit*, 4 ed.; VDE Verlag GmbH, 2015.
13. Moy, Y.; Ledinot, E.; Delseny, H.; Wiels, V.; Monate, B. Testing or Formal Verification: DO-178C Alternatives and Industrial Experience. *IEEE Software* **2013**, *30*, 50–57. doi:10.1109/MS.2013.43.
14. International Standardization Organization. Road Vehicles. Technical Report ISO 26262, ISO, 2018.
15. Ponn, T. How to Define System-Specific Corner Cases for the Type Approval of Automated Vehicles. PhD thesis, Technische Universität München, 2021.
16. Riedmaier, S.; Ponn, T.; Ludwig, D.; Schick, B.; Diermeyer, F. Survey on Scenario-Based Safety Assessment of Automated Vehicles. *IEEE Access* **2020**, *8*, 87456–87477. doi:10.1109/ACCESS.2020.2993730.
17. Huang, W.; Wang, K.; Lv, Y.; Zhu, F. Autonomous vehicles testing methods review. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2016, pp. 163–168. doi:10.1109/ITSC.2016.7795548.
18. Kitajima, S.; Shimono, K.; Tajima, J.; Antona-Makoshi, J.; Uchida, N. Multi-agent traffic simulations to estimate the impact of automated technologies on safety. *Traffic injury prevention* **2019**, *20*, 58–64, [31381431]. doi:10.1080/15389588.2019.1625335.
19. Hilderman, V. DO-178C BEST PRACTICES. Technical report, AFuzion, 2020.
20. Schuldt, F.; Saust, F.; Lichte, B.; Maurer, M.; Scholz, S. Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen. AAET2013 - Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel. Institut für Regelungstechnik, TU Braunschweig, 2013.
21. Stellet, J.E.; Zofka, M.R.; Schumacher, J.; Schamm, T.; Niewels, F.; Zollner, J.M. Testing of Advanced Driver Assistance Towards Automated Driving: A Survey and Taxonomy on Existing Approaches and Open Questions. 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 1455–1462. doi:10.1109/ITSC.2015.236.
22. Nalic, D.; Mihalj, T.; Bäumler, M.; Lehmann, M.; Bernsteiner, S. Sceneario based testing of automated driving systems: A literature survey. Proceedings of the FISITA Web Congress 2020. FISITA, 2020.
23. Kügler, M.E.; Mumm, N.C.; Holzapfel, F.; Schwithal, A.; Angermann, M. Vision-Augmented Automatic Landing of a General Aviation Fly-by-Wire Demonstrator. AIAA Scitech 2019 Forum. American Institute of Aeronautics and Astronautics, 2019. doi:10.2514/6.2019-1641.
24. Angermann, M.; Wolkow, S.; Schwithal, A.; Tonhäuser, C.; Hecker, P. High Precision Approaches Enabled by an Optical-Based Navigation System. Proceedings of the ION 2015 Pacific PNT Meeting, 2015, pp. 694–701.
25. Karlsson, E.; Schatz, S.P.; Baier, T.; Dörhöfer, C.; Gabrys, A.; Krause, C.; Hochstrasser, M.; Lauffs, P.J.; Mumm, N.C.; Nürnberger, K.; et al. Development of an Automatic Flight Path Controller for a DA42 General Aviation Aircraft. In *Advances in Aerospace Guidance, Navigation and Control*; Dołęga, B.; Głębocki, R.; Kordos, D.; Żugaj, M., Eds.; Springer International Publishing, 2018; pp. 121–139. doi:10.1007/978-3-319-65283-2_7.
26. Theunissen, E.; Kotegawa, T. Applying LVC to testing and evaluation of DAA systems. 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC). IEEE, 2017, pp. 1–7. doi:10.1109/DASC.2017.8102089.
27. Bagschik, G.; Menzel, T.; Reschka, A.; Maurer, M. Szenarien für Entwicklung, Absicherung und Test von automatisierten Fahrzeugen. 11. Workshop Fahrerassistenzsysteme und automatisiertes Fahren, 2017, pp. 125–135.
28. ASAM group. ASAM openSCENARIO. Technical report, ASAM e.V., 2021.
29. ASAM group. ASAM XIL. Technical report, ASAM e.V., 2020.
30. PEGASUS Project Office. About PEGASUS. https://www.pegasusprojekt.de/en/about-PEGASUS, accessed on 2021-04-29.
31. PEGASUS Project Office. PEGASUS Method. Technical report, PEGASUS Project, 2019.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0239.v1

32. Scholtes, M.; Westhofen, L.; Turner, L.R.; Lotto, K.; Schuldes, M.; Weber, H.; Wagener, N.; Neurohr, C.; Bollmann, M.H.; Kortke, F.; et al. 6-Layer Model for a Structured Description and Categorization of Urban Traffic and Environment. *IEEE Access* **2021**, *9*, 59131–59147. doi:10.1109/ACCESS.2021.3072739.
33. Thorn, E.; Kimmel, S.; Chaka, M. A Framework for Automated Driving System Testable Cases and Scenarios. Technical report, National Highway Traffic Safety Administration, 2018.
34. Ulbrich, S.; Menzel, T.; Reschka, A.; Schuldt, F.; Maurer, M. Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving. 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 982–988. doi:10.1109/ITSC.2015.164.
35. X2Rail1 Consortium. D6.1 Current test condition and benchmarking report. Technical report, X2Rail1 Consortium, 2018.
36. Hoffmann, D.W. *Software-Qualität*; Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-35700-8.
37. Powell, J.P.; Palacín, R. Passenger Stability Within Moving Railway Vehicles: Limits on Maximum Longitudinal Acceleration. *Urban Rail Transit* **2015**, *1*, 95–103. doi:10.1007/s40864-015-0012-y.
38. Liu, K.w.; Wang, X.C.; Qu, Z.h. Research on Multi-Objective Optimization and Control Algorithms for Automatic Train Operation. *Energies* **2019**, *12*. doi:10.3390/en12203842.
39. Hoberock, L.L. A Survey of Longitudinal Acceleration Comfort Studies in Ground Transportation Vehicles. *Journal of Dynamic Systems, Measurement, and Control* **1977**, *99*, 76–84. doi:10.1115/1.3427093.
40. Stanley, P. *ETCS for engineers*, 1 ed.; DVV Media Group, Eurail Press: Hamburg, 2011.
41. Pachl, J. *Systemtechnik des Schienenverkehrs*, 10 ed.; Springer Fachmedien Wiesbaden and Springer Vieweg, 2021.
42. Lübke, D.; Hecht, M. *Das System Bahn*, 1 ed.; DVV Media Group (Eurailpress), 2008.
43. Albrecht, T.; Gassel, C.; Binder, A.; van Luipen, J. Dealing with operational constraints in energy efficient driving. IET Conference on Railway Traction Systems (RTS 2010). IET, 2010, pp. 1–7. doi:10.1049/ic.2010.0028.
44. van Luipen, J.; Meijer, S. Uploading to the MATRICS: Combining simulation and serious gaming in railway simulators. In *Rail human factors around the world*; Wilson, J.R.; Mills, A.; Clarke, T.; Rajan, J.; Dadashi, N., Eds.; CRC Press, 2012; pp. 165–177.