mag. ing. mech. Tomislav(T.) Mihalj, Graz University of Technology, Graz, Austria
M. Tech Sajin(S.) Gopi, AVL List GmbH, Graz Austria
Ph.D. M.Sc. Philippe(P.) Nitsche, AVL List GmbH, Graz, Austria
Dr.-techn. Hans-Michael(H.-M.) Koegeler, AVL List GmbH, Graz, Austria
Assoc.Prof. Dipl.-Ing. Dr.techn Arno(A.) Eichberger, Graz University of Technology, Graz, Austria

Scenario-based simulation framework for validation of automated driving using next generation DoE

Szenariobasiertes Simulationsframework für das Testen von automatisiertem Fahren unter Verwendung neuer DoE Ansätze

Please keep the footers free from own writings!

Abstract

The future of mobility is being driven by automated driving, connected vehicles, electrification and mobility-as-a-service, which require enormous effort in testing and validation of those systems. Virtual test environments introduce a faster, safer and more economical approach than real field tests. They provide significant benefits in the establishment of product requirements during the early phases of product development. However, despite the virtual approach, an enormous number of scenarios and their variations are still required to achieve sufficient test coverage.
This paper proposes a methodology that significantly reduces the simulation effort by identifying relevant test cases. The methodology is based on Active Design of Experiments (Active DoE), an advanced and sequential process that establishes behaviour models of Key Performance Indicators (KPIs), and automatically adapts the test design to accurately predict simulation outcomes within the range of interest. The benefits are an overall reduction in the number of simulations in comparison to the one-shot designs such as full factorial, and reusable, seamless behavior models for optimization and search purposes.
This work presents a case study of an automated electric bus, that navigates through an urban environment with the task to identify safety-critical and uncomfortable operating ranges. Results demonstrate an excellent fit of KPI models within the range of interest with a more than 60% reduction in the number of simulations with respect to full factorial design.

Kurzfassung

Die Zukunft der Mobilität, die durch automatisiertes Fahren, vernetzte Fahrzeuge, Elektrifizierung und Mobilität-as-a-Service vorangetrieben wird, erfordert einen enormen Aufwand für die Prüfung und Validierung dieser Systeme. Virtuelle Testumgebungen bieten einen schnelleren, sichereren und wirtschaftlicheren Ansatz als reale Feldtests. Sie bieten erhebliche Vorteile bei der Festlegung der Produktanforderungen in den frühen Phasen der Produktentwicklung. Trotz des virtuellen Ansatzes ist jedoch immer noch eine große Anzahl von Szenarien und deren Variationen erforderlich, um eine ausreichende Testabdeckung zu erreichen.
In diesem Beitrag wird eine Methodik vorgeschlagen, die den Aufwand für umfangreiche Simulationen reduziert und relevante Testfälle identifiziert. Die Methode basiert auf dem so genannten Active Design of Experience (Active DoE), einem fortschrittlichen und sequenziellen Prozess, der Verhaltensmodelle von Key Performance Indicators (KPIs) erstellt und das Testdesign automatisch anpasst, um Simulationsergebnisse innerhalb des interessierenden Bereichs genau vorherzusagen. Vorteilhaft ist einerseits eine signifikante Verringerung der erforderlichen Simulationsdurchläufe im Vergleich zu vorab festgelegten Versuchsplänen, wie z. B. vollfaktoriellen Ansätzen. Anderseits können die gewonnen Verhaltensmodelle - lückenlos - für Optimierungs- und Suchzwecke genutzt werden.
In dieser Arbeit wird eine Fallstudie eines automatisierten Elektrobusses vorgestellt, der durch eine städtische Umgebung navigiert. Es werden sicherheitskritische und unangenehme Betriebsbereiche identifiziert. Die Ergebnisse zeigen eine hervorragende Anpassung der KPI-Modelle innerhalb des interessierenden Bereichs - wobei die Anzahl der Simulationsläufe um mehr als 60 % im Vergleich zum vollfaktoriellen Design reduziert wird.

1. Introduction

The release of safe and robust products on the market requires extensive testing within a product's development. Automated vehicles (AV) are complex systems that can lead to hazardous events caused by a system failure, or unforeseen events if insufficiently tested. Therefore, AVs are exposed to a vast array of testing including open-road tests, closed-track tests and virtual tests [1]. Simulations are a type of virtual testing that offers better controllability, repeatability, scalability and predictability at a lower cost than e.g. open-road tests, that require millions of kilometers of driving to statistically prove that they are as reliable as human drivers [2]. Due to their advantages, simulations represent a reasonable approach to support the development process of an AV and offer an alternative to on-road and track tests. That is especially the case in the early stages of development, when hardware may not even be available, and a relatively fast and approximated evaluation of the AV's behaviour is needed.

A common form of using simulations for the evaluation of AVs is scenario-based testing. Scenario-based testing is a process in which an AV's behaviour is validated within predefined scenarios with a specific environment, actors and manouvers. Scenarios are concretized to test cases through the fixing of input parameters within a desired operating space. The main idea is to identify under which conditions, relevant test cases are triggered. The relevance of test cases depends upon scenario requirements, therefore, relevant test cases can be defined in many ways, such as crash, no-crash, near-crash, energy inefficient, uncomfortable etc. However, the problem lies in the vast number of test cases required for each of the scenarios to cover whole operating space, especially as only a handful of them might be relevant. Therefore, the common N-wise sampling with all possible combinations of parameters is unrealistic.

Different techniques have been used in recent years to address this problem. Gelder [3] uses importance sampling to reduce the variance of the probability density of randomly generated test cases. Derived importance sampling density provides a way to generate more critical cases. However, the process still requires a huge number of simulations, and simulation reduction stays an open point. More related work to our study is work from Mullins [4], which uses sequential sampling and surrogate models to identify boundaries in the operating space, which as a consequence reduces the number of simulations. Boundaries are defined as part of the scenario space where small changes in inputs, cause large changes in outputs. The authors use Gaussian Process Regression (GPR) and k-nearest neighbor techniques to model output behaviour. Batsch in his work [5], also performs a sequential sampling which combines classification and regression based on the Gaussian Process (GP) to model Key Performance Indicators (KPIs). The selection of new candidates is done by a heuristic equation that contains predictions and targeted values of KPIs. In the studies mentioned above, a global surrogate model has been used that increases computation costs, while the performance of methodologies and algorithms is presented on a single KPI.

In our study, we propose an approach to identify relevant test cases based on the Active Design of Experiments (Active DoE). This method is available within the commercial software AVL CAMEO^{TM} developed for data-driven testing, modelling, and optimization. Active DoE, is a sequential process that uses split optimization. This allows localized modelling of KPIs that consequently reduces computation costs. This makes the methodology efficient, even when dealing with a larger number of input parameters and KPIs. Moreover, the search is focused on the custom output space known as the range of interest (RoI). The RoI is defined by the KPI thresholds in order to narrow the operating space to relevant test cases. Focused sampling within the RoI, increases the accuracy of

the behavior models of the KPIs, also known as surrogate models. This leads to overall simulation case reduction as the algorithm disregards most of the uninteresting test cases outside the RoI. Finally, derived surrogate models can be reused for optimization.
Several studies have already been written using Active DoE. Beglerovic [6] used Active DoE to identify near-crash scenarios in a Highway Pilot case study. The study also considers the Adaptive Cruise Control (ACC) controller parameters as inputs, which means the method is also useful for controller tuning/calibration. Yet, the approach is not limited to studies in the field of AV testing. Rainer and Koegeler [7] used the method for powertrain calibration to improve emissions.
In the scope of this study, we are using Active DoE to identify relevant test cases based on the case study of an automated electric bus that navigates through the urban environment. In particular, we are using a cut-in scenario to show the potential of the proposed methodology. The scenario considers the lead vehicle – referred to hereafter as target object front (TOF) – and the bus that represents the vehicle under test, which is referred to as the ego vehicle (EV). For the evaluation of EV behaviour, we are using a requirement-based approach in which the requirements are grouped into safety, comfort and legal criteria.
The remainder of the paper is outlined as follows. In Section 2 we are covering the background theory on Active DoE, and the model architecture to build KPI surrogates. Section 3 presents a co-simulation framework, a definition of evaluation requirements, the selected KPIs and the evaluation plan. Section 4 shows the evaluation process and results of the simulated cut-in scenario, by comparing full factorial simulation and Active DoE. Finally, the conclusion is provided in Section 5.

2. Active DoE

The Active DoE is a data-driven sequential method that plans and executes test cases in such a way to cover predefined output space.
To gain knowledge about outputs without observed data, the Active DoE relays on the so-called Robust Neural Network (RNN), a neuro-fuzzy model architecture based on the evolving local model networks (ELMNs). In this section, we briefly discuss the RNN and strategy behind the Active DoE, for more details refer to [8] and [9].
The RNN splits the input space into smaller sets, with each set being modelled by local models ($\mathrm{LM}_1, \mathrm{LM}_2, \ldots, \mathrm{LM}_M$), see Figure 1. Each local model contains a local regression polynomial and validity function. Regression polynomial provides a local estimate combining the model parameter vector θ_i and a regression vector $x(k)$.

$$\hat{y}_i(k) = x^T(k)\theta_i \tag{1}$$

The validity function $\Phi_i(\tilde{x}(k))$ of predefine input set $\tilde{x}(k)$ defines the weight of the regression polynomial. The global model is then estimated by combining all local models:

$$\hat{y}(k) = \sum_{i=1}^{M} \Phi_i(k)\hat{y}_i(k) \tag{2}$$

The main challenge with training in a sequential process is to keep the computational effort low while rebuilding the model after each iteration. To support that premise, the RNN uses a discriminant tree structure to replace the worse local model with the new node that contains two additional local models. Such a concept allows split optimization of model parameters and validity function of the actual node only, while other local models remain unchanged.

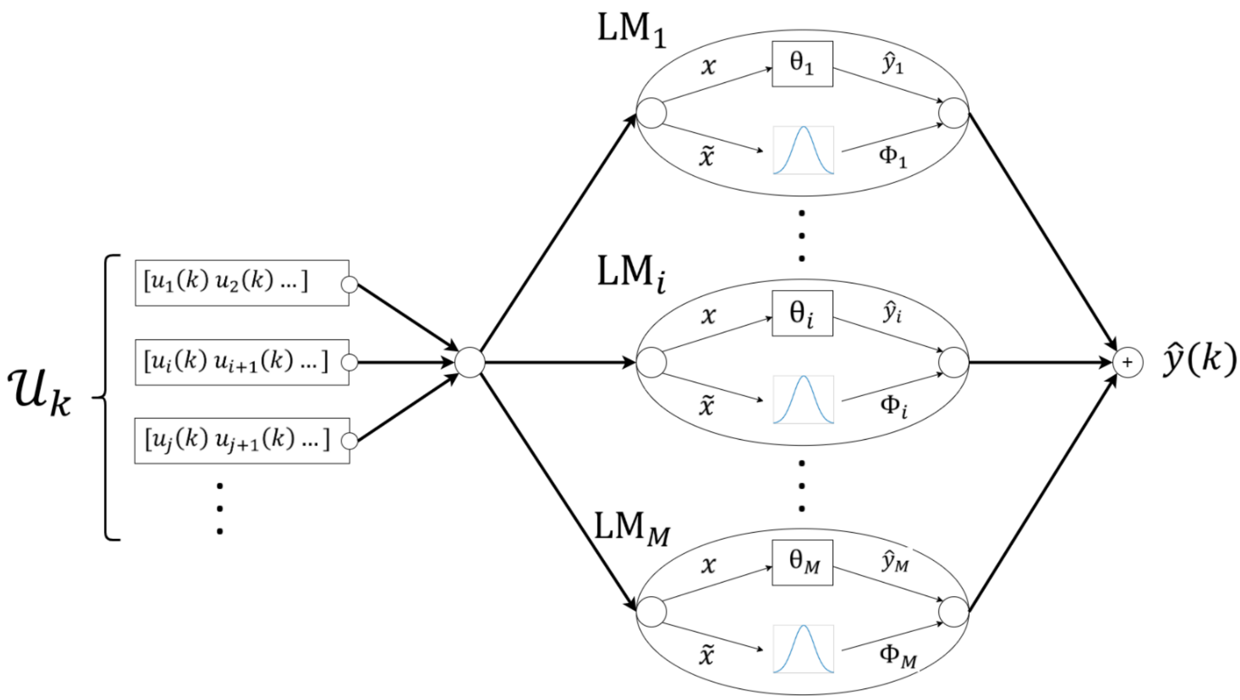

Figure 1 Architecture of local model network (LMN)

Considering those advantages of the RNN, the Active DoE can efficiently sample test cases until the required accuracy of the models is not reached. Since Active DoE explores the output space to find the next test case, the RoI should be predefined in advance bounded with lower (y_{min}) and upper limit (y_{max}). Therefore, the RoI can be written as follows:

$$R_y = [y_{min}, y_{max}] \quad (3)$$

The Active DoE works with two data sets collected from the input space. One data set is dedicated to observed data ($\mathcal{U}_k$) collected via simulations or measurements and the second set ($\mathcal{U}^C$) is reserved for potential candidates of the sequential process. Active DoE starts with the initial $\mathcal{U}_k$, for which methods like the Sobol design are used that equidistantly distribute samples within the design space by keeping discrepancy as low as possible considering the number of samples. The $\mathcal{U}^C$, on the other hand, is intended to fill the whole design space, therefore, the full-factorial design of q factors and p levels is used. After each iteration size of the $\mathcal{U}^C$ decrease while $\mathcal{U}_k$ increase as a new sample is moved from one set to another.

To select from output space, the Active DoE relays on the estimates of the RNN. The sequentially built RNN model is used to predict outputs on the $\mathcal{U}^C$ which makes the method aware of the outcomes. Selection is done using *maximin* distance on the output space, see Equation (4). The benefit of sampling from the output using the *maximin* approach is capturing the non-linear behaviour of the model, which is not the case if samples are only collected considering input space. This can be seen in Figure 2, where equidistant sampling from input space is compared to the Active DoE. It is shown that due to unawareness of non-linearities, the sampling from input space could perform poorly in preparing data for training the models. Moreover, knowing the RoI, the Active DoE focuses its sampling on that specific area which assures better accuracy of the RNN model, while at the same time, reducing the overall number of simulations or measurements in comparison to one-shot methods.

$$\boldsymbol{u}(k) = \arg\max\min|\hat{y}(u^C) - y|,\quad u^C \in \mathcal{U}_k^{R_y} \text{ and } y \in \mathcal{Y}_{k-1}. \tag{4}$$

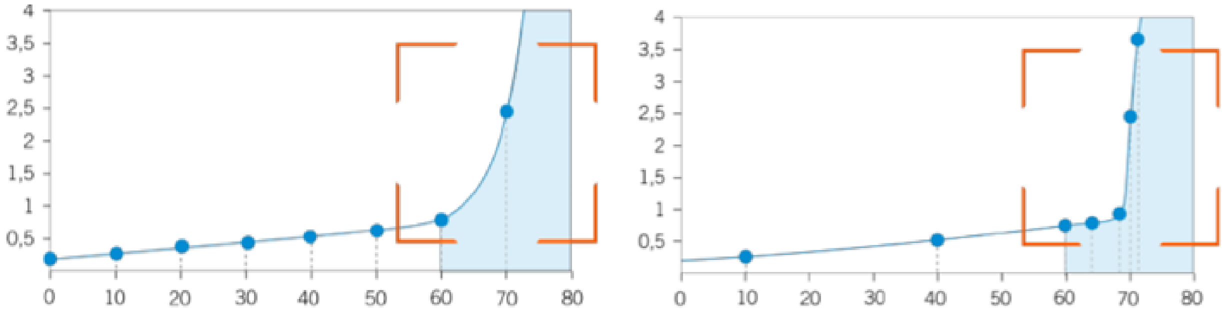

Figure 2 Comparison of one-shot sampling (standard DoE) and Active DoE on a 1D exponential function. The left image shows an example of the one-shot sampling method in which equidistantly spaced inputs are sampled. The right shows how the Active DoE places the sample regarding the output by shrinking unknown output space and covering non-linear space better. As consequence more accurate surrogate models are built. Images are taken from the AVL CAMEO™ online manual.

3. Method

3.1. Co-simulation Framework

Test cases are simulated with AVL Model.CONNECT™ software, a co-simulation platform that offers a vast number of interfaces to commonly used vehicle simulation software. The co-simulation integrates and synchronizes different simulation subsystems, dedicated to perform a specific task. Within Model.CONNECT, we have integrated AVL VSM™, the vehicle simulation software that uses a non-linear multi-body approach for precise representation of the vehicle dynamics. It has been used for the electric bus model, and exchanges signals with the automated driving controller (integrated as an FMU) and the environment simulation, performed with the software VTD (Virtual Test Drive). The automated driving controller combines lane keeping, adaptive cruise control and automated emergency braking, and is further denoted as ALKS (Automated Lane Keeping System). To perform the Active DoE study, AVL CAMEO™ is connected as the master, with Model.CONNECT. In this way, CAMEO handles parameter exchange and sequentially runs simulations to find relevant test cases. To prevent parameter explosion, the parameters for light and weather variation are excluded from the investigation. Thus, the usage of ideal sensors, i.e. geometrical models with object lists as output, in VTD is sufficient.

3.2. Evaluation requirements

To evaluate test cases, we defined requirements according to the following categories, safety, comfort, and legal criteria. Safety requirements are set to avoid hazardous events. Such requirements define the necessity to avoid a crash, keep a safe distance (even in the worst case), and timely adjustment of speed and distance to the TOF. Moreover, as we are dealing with a city bus, we must take into consideration every passenger. Free-standing passengers are the most exposed to injury due to hard braking, acceleration, or turning. Therefore, we have defined a safety requirement to keep acceleration and jerk within thresholds that prevent passengers' loss of balance. Due to the relatively short duration and quick changes of the longitudinal and lateral movements, frequency-based KPIs to assess driving comfort are not the most suitable. Instead, we use acceleration and jerk. The third criteria considers legal aspects, in this study, the TR-68 [10] has been accounted for, while a digital twin of the road represents the campus at the Nanyang Technical University in Singapore. The list of requirements is provided in Table 1.

Table 1 Requirements for a test case evaluation

Safety requirements	Description
[S1] keep safety distance (worst-case)	The ego should always keep a safe distance behind the TOF, which includes also the worst-case scenario when the TOF brakes with maximum possible deceleration.
[S2] adjust speed and distance	The ego should adjust its speed and distance regarding the TOF states.
[S3] keep deceleration and jerk within safety thresholds	The ego should always keep deceleration below levels at which standing passengers could lose their balance.
[S4] avoid a crash	The ego should avoid crashing with the TOF.
Comfort requirements	
[C1] keep deceleration within comfort thresholds	Ego should keep deceleration within comfortable thresholds, including for standing passengers.
[C2] keep jerks within comfort thresholds	Ego should keep jerks within comfortable thresholds, including for standing passengers.
[C3] keep a comfortable distance	Ego bus should keep such a distance behind the TOF, which even under the worst-case won’t require deceleration to exceed the comfortable threshold.
Legal requirements	
[L1] keep distance above 2m	[TR-68: 6.2] "…AVs should drive at a safe distance such that if the leading car came to a complete stop, the AV should also be able to safely come to a stop (2 m before the obstacle in front).
[L2] keep the time gap above 2s	[BTD 135] the "two-second" rule

3.3. Key Performance Indicators (KPI)

To each requirement, one or more KPIs are assigned, along with their thresholds. Thresholds represent ranges to assess requirements as fail or pass. The following text provides background on the selected KPIs, while Table 2 provides their summary and assignments. Thresholds provided in Table 2 indicate satisfied requirements.

Table 2 KPIs for evaluation requirements

Safety requirement	KPI	Threshold
S1	Remaining safety distance ($d_{\mathrm{rem,safe}}$)	$d_{\mathrm{rem,safe}} > 0$
S2	Time-to-collision (TTC)	TTC > 3s
S3	Longitudinal acceleration (a_{ego}) & Longitudinal jerk (j_{ego})	$-1.5\mathrm{m/s}^2 < a_{\mathrm{ego}} < 1.5\mathrm{m/s}^2$ & $\lvert j_{\mathrm{ego}} \rvert < 6\mathrm{m/s}^3$
S4	Crash	0 – no crash, 1 - crash
C1	Longitudinal acceleration	$-1.5\mathrm{m/s}^2 < a_{\mathrm{ego}} < 1.5\mathrm{m/s}^2$
C2	Jerk	$\lvert j_{\mathrm{ego}} \rvert < 0.75\mathrm{m/s}^3$
C3	Remaining comfortable distance ($d_{\mathrm{rem,comf}}$)	$d_{\mathrm{rem,comf}} > 0$
L1	Relative distance (d_{rel})	$d_{\mathrm{rel}} > 2\mathrm{m}$
L2	Time gap (t_{gap})	$t_{\mathrm{gap}} > 2\mathrm{s}$

a) **Time gap (t_{gap}):** This KPI defines the time passed between TOF's tail end and EV's front for the same spatial point. We use t_{gap} to evaluate the L2 requirement that refers to the widely known 2s time gap rule. To provide robustness towards possible short peaks in signal, the representative t_{gap} is calculated as the minimum value of the moving average with a 1s window over the scenario duration.

b) **Time-to-collision (TTC):** TTC is a well-known and widely used KPI, suitable for the safety assessment of rear-end collisions. The advantage of the TTC is in considering both distance and velocities. However, small relative velocity v_{rel} between EV and TOF, can lead to a non-critical assessment in case of small d_{rel}, which could be insufficient to avoid a crash in case of the braking event of the TOF. Therefore, if classical TTC is used it is good practice to use it along with others, e.g. distance KPIs. In this study, we use TTC to evaluate the S2 requirement, the EV's ability to adjust speed and distance to the TOF. An additional distance KPI is considered within the S1 requirement. TTC is defined as follows:

$$\mathrm{TTC} = \frac{d_{\mathrm{rel}}}{v_{\mathrm{ego}} - v_{\mathrm{tof}}} \qquad \forall\ v_{\mathrm{ego}} > v_{\mathrm{tof}} \tag{5}$$

where v_{ego} and v_{tof} denote the EV and TOF velocity, respectively. To distinguish safe and unsafe following maneuvers, a threshold of 3s is considered according to [11], [12] and [13]. In the same manner as t_{gap}, the representative TTC is calculated as the minimum value of the moving average with a 1s window.

c) **Relative distance (d_{rel}):** d_{rel} is the distance between the TOF's tail end and the EV's front. The TR-68 [10] defines 2m as the minimum allowed distance that the EV should keep even when the TOF is at standstill. Therefore, we are using this measure to evaluate the L1 requirement.

d) **Remaining safety distance ($\boldsymbol{d_{\mathrm{rem,safe}}}$):** $d_{\mathrm{rem,safe}}$ is the difference between relative distance and minimum safety distance $d_{\mathrm{min,safe}}$, see (6). $d_{\mathrm{min,safe}}$, is calculated according to [14], and it defines the worst-case scenario in which the TOF brakes with maximum available braking (b_{tof}). Equations take into account the tire-road friction coefficient (μ) through braking, and the response time (ρ) needed for the EV to detect and react to the event.

$$d_{\mathrm{rem,safe}} = d_{\mathrm{rel}} - d_{\mathrm{min,safe}} \tag{6}$$

$$d_{\mathrm{min,safe}} = v_{\mathrm{ego}}\rho + \frac{1}{2}a_{\mathrm{ego}}\rho^2 + \frac{(v_{\mathrm{ego}} + \rho a_{\mathrm{ego}})^2}{2b_{\mathrm{ego}}} - \frac{v_{\mathrm{tof}}^2}{2b_{\mathrm{tof}}} \tag{7}$$

In (7), a_{ego} is EV acceleration, ρ EV response time, while b_{ego} and b_{tof} are the same and denote the maximum available braking. For simplicity, ρ is taken to be 1s according to [15]. When $d_{\mathrm{rem,safe}}$ is bigger than 0, it means that d_{rel} is bigger than $d_{\mathrm{min,safe}}$ and crash is avoidable, in other words, d_{rel} is sufficient to avoid a crash with the TOF under maximum available braking. This KPI is also calculated as the minimum moving average over the 1s window.

e) **Remaining comfort distance ($\mathbf{d_{rem,conf}}$):** $d_{\mathrm{rem,conf}}$ is defined similarly as the $d_{\mathrm{rem,safe}}$ except we replaced $d_{\mathrm{min,safe}}$ with the minimum comfortable distance ($d_{\mathrm{min,conf}}$). In contras to $d_{\mathrm{min,safe}}$, the $d_{\mathrm{min,conf}}$ uses the comfortable longitudinal deceleration threshold for free-standing passengers, -1.5m/s^2 [16], as the EV braking deceleration (b_{ego}).

f) **Longitudinal acceleration ($\boldsymbol{a_{\mathrm{ego}}}$):** a_{ego}, used alone is an indicator of an uncomfortable ride. However, if coupled with a jerk metric, can be used to assess potentially safety-critical scenarios in which a free-standing passenger in a bus can lose their balance and injure themselves [17]. To distinguish a comfortable from an uncomfortable ride, the threshold value of -1.5 m/s^2 is considered. The same threshold is used to evaluate the loss of balance requirement, only in that case, a_{ego} must be considered along with longitudinal jerk (j_{ego}). To consider acceleration as KPI, it should be averaged over a 2s time interval [17]. Therefore, to extract representative values from the acceleration time series we calculate it as the minimum moving average with a 2s window.

g) **Longitudinal jerk ($\boldsymbol{j_{\mathrm{ego}}}$):** j_{ego}, similar to a_{ego}, is another indicator of discomfort. However, when coupled with acceleration, it can be used to evaluate a safety-critical requirement in which free-standing passengers in a bus can lose their balance. As a measure of comfort, jerks should be within the range 0.75m/s^3 to -0.75m/s^3 according to [18] and [19]. Yet, when used to assess safety-critical scenarios, the range is between 6m/s^3 and -6m/s^3 along with accelerations that do not exceed the range between 1.5m/s^2 and -1.5m/s^2 [17]. As a representative value, a simple peak value can be considered only, without applying a moving average. This assumption is acceptable since we are dealing with the simulations. If the measurement were the subject of an investigation, other approaches, such as deriving the jerk from the smooth acceleration signal, would be more appropriate.

h) **Crash:** To evaluate crashes we used a binary approach where 0 indicates no crash and 1 a crash.

3.4. Evaluation plan

In this section, we are introducing the example scenario, the test plan and the evaluation process. To demonstrate methodology, we selected the cut-in scenario in which the EV ALKS functionality [20] can be tested. In the scope of this study, we define the relevant test cases as cases in which a crash did not happen but the d_{rel} is not sufficient to avoid collision in case of abrupt braking of the TOF.

a) **Scenario definition:** At the start of a simulation, the TOF is located in front of the EV in an adjacent lane. As the EV reaches the cut-in triggering distance, also called the initial distance d_0, the TOF cuts into the EV lane. The EV should recognize the cut-in maneuver and if possible, safely adjust its speed and distance accordingly. A schematic representation of the scenario is shown in Figure 3.

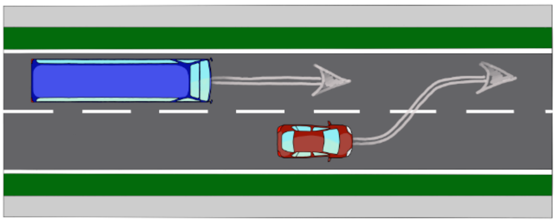

Figure 3 Schematic representation of the Cut-In scenario

b) **Test plan:** Four parameters, namely tire-road friction coefficient (μ), TOF velocity (v_{tof}), initial distance (d_0), and lane change time (t_{LC}) are subject to variation. Road, TOF type, and EV velocity (v_{ego}) are kept constant. Those parameters are independent except for the lane change time in which partial dependency is built. The boundaries of lane change time are calculated as a function of v_{tof} and d_0, and act as a preprocessing filter to keep the TOF in the EV's field of view during all variations. Other than that, t_{LC} is selected independently between a minimum and maximum boundary, which are limited to 3 and 10s, respectively. Those limits match driving simulator studies [21] and naturalistic driving data [22]. Due to its variable character, the t_{LC} is normalized over test cases and ranged down between 0 – 1, where 0 = 3s and 1 = 10s. A detailed test plan is given in Table 3. Defined steps indicated with superscript [a] refer to the fixed step size for the full factorial design.

Table 3 Test plan for test case simulations

Static Environment	
Road	Nanyang Ave (from the Cleantech View intersection to the lane merging at Nanyang Ave)
TOF type	Car
Tire-road friction coefficient (μ)	0.6 : 1 (step[a]: 0.2)
Dynamic Environment	
EV velocity (v_{ego})	20 km/h

TOF velocity (v_{tof})	5 : 18 km/h (step[a]: 1)
Initial distance (d_0)	5 : 30 m (step[a]: 5)
Lane change time (t_{LC})	3 : 10 s (step[a]: 0.2 – step is used when lane change time is normalized)

[a] Applicable only for the full factorial design

c) **Evaluation process:** The evaluation process is conducted through a full factorial test design, representing the traditional approach, and the Active DoE to show reduction potential.

- *Full factorial simulations*: Within the full factorial test designs 1512 test cases according to the test plan defined in Table 3, are simulated. The full factorial design is deterministically defined and sampled stepwise within the parameter ranges. This simulation set represents the base case for comparison purposes. After test completion, KPIs derived from the simulation results are processed in CAMEO for further analysis and comparison to Active DoE.
- *Preparing Active DoE:* The first step in the preparation of Active DoE is to define the input space and the RoI that is constrained with the relevant KPIs. In this study, we focus our search on no-crash test cases that fail S2 and partially fail S1 requirements. The available input space is restricted to the test plan, see Table 3.
- *Active DoE:* Active DoE is initialized with a start design of 30 simulations defined by the Sobol design, this initial design is used by Active DoE to scan the operating space. We have limited the iterative simulation process of the Active DoE to 450 simulations without considering the start design. CAMEO offers repetition points that are scheduled after a certain number of tests, overall, 14 such test cases are simulated. They have the same input parameters and they aim to detect variation in output. This is more relevant in physical testing. However, it is also useful in simulations to verify repeatability. Overall 485 simulations with Active DoE are performed, which represent roughly 1/3 of the number required for the full factorial design, and ensure high model quality and good visualization possibility for comparison.
- *Distribution in RoI:* The Active DoE and full factorial results are compared to examine the benefits of the Active DoE approach, with regards to placement of test cases within the RoI. Placing samples in the RoI solves the problem of removing irrelevant test cases and directly leads to a reduction of the overall number of necessary tests. The second benefit of using such an approach is obtained in terms of surrogate modelling - more points are placed within the RoI. Thus, quality of the surrogate models is improved in this range of interest - even compared to behavior models gained out of full factorial sampling. This is achieved by dynamic adjustment of point density based on model quality. If data is lacking, then more points are sampled, until the required model quality is reached, meaning no unnecessary simulations are performed.

- *Fitting simulated data and model quality:* To fit the model to simulated data we chose the RNN from the palette of built-in fitting algorithms in CAMEO. The model quality is assessed through the goodness of fit and comparison with cross-validation points to check the model against lack of fit. For the goodness of fit assessment, we are going to show the results using statistical metrics like R^2, $\mathrm{R}^2_{\mathrm{adj}}$ and RMSE. To check against the lack of fit, including overfitting, we divided the simulation set into two randomly selected subsets, one containing 80% and the other 20% of simulations. The smaller subset is used as verification points. To assess the lack of fit, the ratio between RMSE^* (test data set) and RMSE (training data set) has been calculated and compared to p-values to verify if there is a significant difference between training and verification samples, see Figure 4.

p-value	RMSE* / RMSE		
	q < 0.3	0.3 < q > 2	q > 2
0 - 0.1	Poor	Very Good	Overfit
0.1 - 0.9	Very Good	Very Good	Very Good
0.9 - 0.975	Good	Very Good	Good
0.975 - 1	Poor	Very Good	Poor

Figure 4 Threshold values for assessing lack of fit (taken from the AVL CAMEO™ online manual)

- *Optimization:* Obtained surrogate models are reused for optimization purposes to identify the corner test cases regarding S1, S2 and S3 requirements. The AVL CAMEO™ offers different optimization algorithms. We have selected a multiobjective genetic algorithm that is based on Non-dominated Sorting Genetic Algorithm II (NSGA-II). This algorithm shows some favorable characteristics such as a lower computational complexity and elitism which makes execution faster. Furthermore, it is found that NSGA-II outperforms other popular multiobjective evaluation algorithms regarding finding the true Pareto-optimal set [23].

4. Evaluation and Results

a) **Range of interest (RoI) and overall assessment:** Overall 1512 test cases based on the full factorial design and 485 with Active DoE have been simulated. For Active DoE we defined the RoI by constraining TTC in the range between 0 and 3s, d_{rel} from 0 to 12m (one length of the bus) and no crash cases. The reason why we use d_{rel} instead $d_{\mathrm{rem,safe}}$ is to capture a slightly larger area around the safety-critical and uncomfortable region.

b) *Table 4* shows the overall assessment and demonstrates the efficiency of the Active DoE algorithm compared to the full factorial test design. Efficiency in this case means a higher percentage of test cases that failed requirements. While crash test cases are not of particular interest to this study, the overall evaluation summary provided in

c) *Table 4* is based on 944 no-crash test cases of the full factorial design, and 443 sampled with Active DoE. Therefore, it can be concluded that among 1512 simulations of full factorial design, 568 (37.6%) are crash test cases, while in Active DoE only 42 (8.7%). This result is even lower, at approximately 2%, when the repetition points are removed, as they happened to be crash cases.

d) *Table 4* presents the fail and pass criteria for each requirement. In general, we can see that the Active DoE obtains more failed test cases. This is expected, as the RoI has been partially defined on the safety criteria. However, due to this partial coverage, the results given in

e) Table *4* do not represent the performance of the Active DoE directly, as RoI defines a slightly wider space than space bounded with safety thresholds. Thus we should look at how many points are within the RoI. The 263 out of 485 test cases have met that criteria, which is 54.2%, in comparison to full factorial where only 177 (11.7%) out of 1512 are sampled within the RoI. Here we can see that using the Active DoE reduces the overall simulation number by an approximate factor of 3 and more relevant cases are collected in comparison to the full factorial design. Moreover, a higher number of test cases within the RoI, give us confidence that surrogate models can be well-fitted.

Table 4 Overall evaluation summary of no crash test cases

Requirement	Full factorial **Fail | Pass**	Active DoE **Fail | Pass**
[S1] keep safety distance	55 | 889 (94.2%)	163 | 280 (63.2%)
[S2] adjust speed to TOF	177 | 767 (81.3%)	263 | 180 (40.6%)
[S3] keep deceleration and jerk within safety thresholds	158 | 786 (83.3%)	144 | 299 (67.5%)
[S4] avoid a crash	0 | 944 (100%)	0 | 443 (100%)
[C1] keep deceleration within comfort thresholds	369 | 575 (60.9%)	276 | 167 (37.7%)
[C2] keep jerks within comfort thresholds	917 | 27 (2.9%)	442 | 1 (0.2%)
[C3] keep a comfortable distance	155 | 789 (83.5%)	304 | 139 (31.4%)
[L1] keep distance above 2m	41 | 903 (95.7%)	82 | 361 (81.5%)
[L2] keep the time gap above 2s	316 | 628 (66.5%)	395 | 48 (10.8%)

f) **Model Quality:** Along with the reduction of simulation effort and extracting relevant scenarios, it is also important to ensure good model quality of the surrogates. Table 5 shows the goodness of fit by examining the R^2, $\mathrm{R}^2_{\mathrm{adj}}$ and RMSE. It can be observed that the predictions fit very well with the simulated results.

Table 5 Goodness of fit summary

KPI	$\mathbf{R^2}$	$\mathbf{R^2_{adj}}$	**RMSE**

a_{ego}	0.97	0.97	0.067
j_{ego}	0.88	0.88	0.579
d_{rel}	0.99	0.99	0.167
$d_{\mathrm{rem,safe}}$	0.99	0.99	0.203
$d_{\mathrm{rem,comf}}$	0.96	0.96	0.503
t_{gap}	0.98	0.98	0.069
TTC	0.98	0.98	0.207

To check against the lack of fit, including overfit, we used an 80/20 split of simulated test cases. Results are presented in Table 6, which can be directly compared to Figure 4. We can note that no lack of fit has been observed. In addition to model quality, we analyzed the significance of variations, see Table 7, to inspect the significance of the input parameters on the KPIs. As expected, the initial cut-in distance and TOF velocity play the most significant roles for most of the KPIs. However, it is observed that for comfort-related KPIs like a_{ego}, j_{ego} and $d_{\mathrm{rem,comf}}$, the μ shows high influence.

Table 6 Lack of fit summary

KPI	p-value	RMSE	$\mathbf{RMSE^*}$	$\mathbf{RMSE^*/RMSE}$
a_{ego}	1	0.060	0.121	2
j_{ego}	0.504	0.523	0.536	1
d_{rel}	0.459	0.124	0.132	1.1
$d_{\mathrm{rem,safe}}$	0.976	0.146	0.192	1.3
$d_{\mathrm{rem,comf}}$	0.802	0.450	0.508	1.1
t_{gap}	0.516	0.051	0.054	1.1
TTC	0.843	0.186	0.213	1.1

Table 7 Significance of variations

KPI	d_0	t_{LC}	μ	v_{tof}
a_{ego}	0.120	0.032	0.696	0.15
j_{ego}	0.042	0.03	0.871	0.057
d_{rel}	0.323	0.113	0.145	0.419
$d_{\mathrm{rem,safe}}$	0.503	0.12	0.108	0.269
$d_{\mathrm{rem,comf}}$	0.494	0.133	0.228	0.146
t_{gap}	0.475	0.131	0.113	0.281
TTC	0.32	0.061	0.104	0.515

g) **Safety evaluation:** After achieving good quality surrogates, we perform an analysis of the safety-critical regions with respect to S1, S2 and S3 requirements. Figure 5 Left image: Safety-critical test cases regarding S1 and S2 with respect to $d_{rem,safe}$and TTC. Right image: Safety-critical test cases with respect to to d_0 and to v_{tof}. The highlighted area in light blue represents safety-critical test cases that failed to satisfy the S1 and S2 requirements. Blue points are the test cases generated with Active DoE, while green triangles are generated based onthe full factorial design. Gray points in the background are random samples generated with surrogate models. highlights the safety-critical cases (light blue points) that fall below the threshold of selected KPIs for S1 and S2. Those relevant cases show somewhat of a linear behaviour to v_{tof} and d_0. The highlighted test cases show higher density compared to the full factorial design (green triangles) which has an equidistant distribution over the whole operating space.

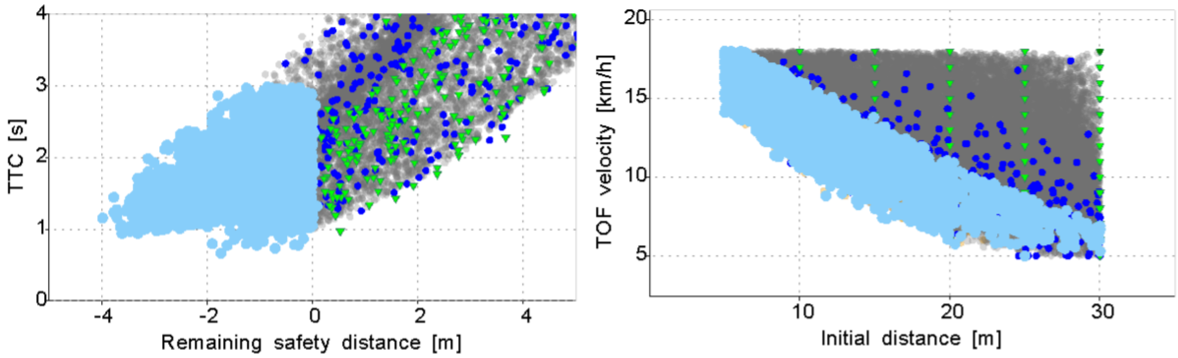

Figure 5 Left image: Safety-critical test cases regarding S1 and S2 with respect to $d_{rem,safe}$and TTC. Right image: Safety-critical test cases with respect to to d_0 and to v_{tof}. The highlighted area in light blue represents safety-critical test cases that failed to satisfy the S1 and S2 requirements. Blue points are the test cases generated with Active DoE, while green triangles are generated based onthe full factorial design. Gray points in the background are random samples generated with surrogate models.

Similarly, Figure 6 Left image: Safety-critical test cases regarding S3 with respect to a_{ego} and j_{ego}. Right image: Safety-critical test cases with respect to μ and to v_{tof}. The highlighted area in light blue represents safety-critical test cases that failed to satisfy the S3 requirement. Blue points are the test cases generated with Active DoE, while green triangles are generated based on the full factorial design. Gray points in the background are random samples generated with surrogate models. shows the overall evaluation of S3. It can be observed that the safety-critical region of a_{ego} and j_{ego} appear at higher values of μ. To identify the corner test cases related to S1, S2 and S3, we performed optimization using a multiobjective genetic algorithm.

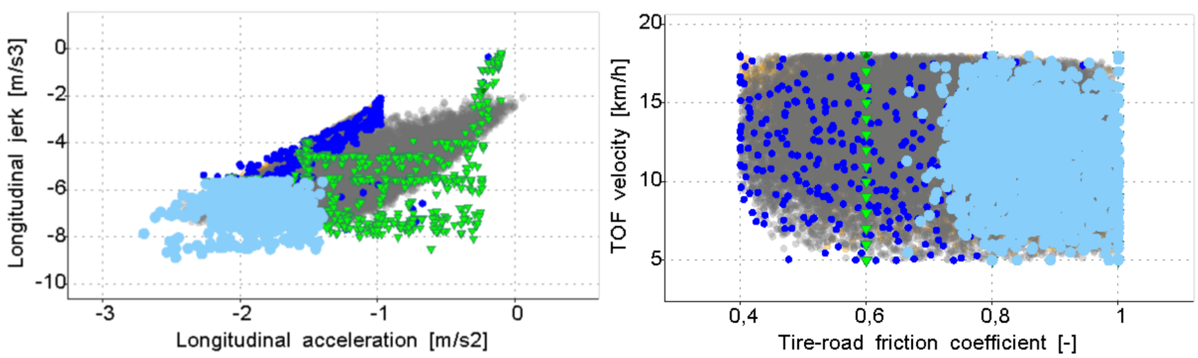

Figure 6 Left image: Safety-critical test cases regarding S3 with respect to a_{ego} and j_{ego}. Right image: Safety-critical test cases with respect to μ and to v_{tof}. The highlighted area in light blue represents safety-critical test cases that failed to satisfy the S3 requirement. Blue points are the test cases generated with Active DoE, while green triangles are generated based on the full factorial design. Gray points in the background are random samples generated with surrogate models.

The results of optimizations are shown in Figure 7 Upper left: Pareto front on the borderline with respect to $d_{rem,safe}$; Lower left: Pareto front regarding d_0 and to v_{tof}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals (shaded green area, boudn by dashed green lines). The selected point is the light blue point selected on the pareto front on the diagrams on the left. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

and Figure 8 Upper left: Pareto front on the borderline with respect to a_{ego}; Lower left: Pareto front in μ vs v_{ego}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals and showing the selected light blue point on the pareto front on the left diagrams. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

. Figure 7 Upper left: Pareto front on the borderline with respect to $d_{rem,safe}$; Lower left: Pareto front regarding d_0 and to v_{tof}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals (shaded green area, boudn by dashed green lines). The selected point is the light blue point selected on the pareto front on the diagrams on the left. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

shows the corner cases where TTC is allowed to go below 3s, which means that the S2 requirement is not fulfilled, but $d_{\mathrm{rem,safe}}$ is kept on the border line. This can be seen at the upper left plot where the Pareto front (dark blue diamond shapes) is distributed along $d_{\mathrm{rem,safe}} \approx 0$. The corner cases identify the boundary between relevant and non-relevant test cases. The lower left plot shows that those test cases within the Pareto front are distributed along the d_0 and v_{tof}. The right plot shows the influence of the input parameters on the KPIs. As expected, with a decrease of d_0 or increase of v_{tof} the $d_{\mathrm{rem,safe}}$ becomes more critical. However, the prediction shows that increasing t_{LC} increases criticality also. This could indicate that the EV does not realize the lane change in time, which could come either from an insufficient field of view of the sensor or a suboptimal algorithm decision. In the same manner, Figure 8 Upper left: Pareto front on the borderline with respect to a_{ego}; Lower left: Pareto front in μ vs v_{ego}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals and showing the selected light blue point on the pareto front on the left diagrams. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

shows corner cases with respect to a_{ego}, while j_{ego} is allowed to go below the safety threshold with respect to S3. The upper left plot shows the Pareto front (dark blue diamond shapes) with respect to a_{ego} and j_{ego}. The lower left plot shows the Pareto front distribution in μ vs v_{tof}. The right plot shows the influence of the input parameters on the

KPIs. We can observe the trade-off behaviour of a_{ego} and j_{ego} in dependency to μ. Higher μ lowers j_{ego} in approximately linear manner, but lowest a_{ego} are predicted at $\mu \approx 0.8$.

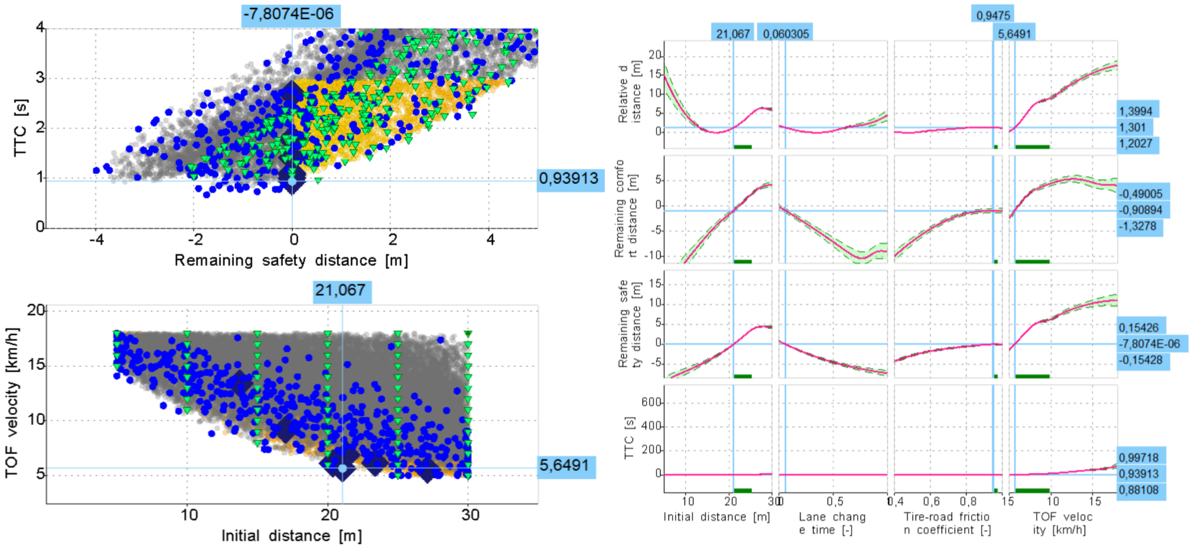

Figure 7 Upper left: Pareto front on the borderline with respect to $d_{rem,safe}$; Lower left: Pareto front regarding d_0 and to v_{tof}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals (shaded green area, boudn by dashed green lines). The selected point is the light blue point selected on the pareto front on the diagrams on the left. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

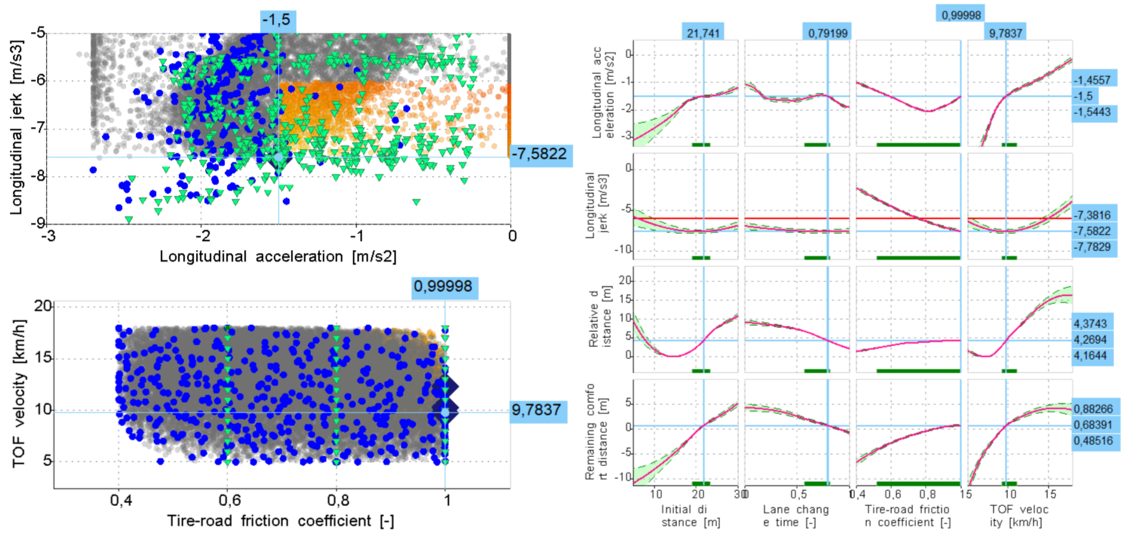

Figure 8 Upper left: Pareto front on the borderline with respect to a_{ego}; Lower left: Pareto front in μ vs v_{ego}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals and showing the selected light blue point on the pareto front on the left diagrams. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

In addition, Figure 7 Upper left: Pareto front on the borderline with respect to $d_{rem,safe}$; Lower left: Pareto front regarding d_0 and to v_{tof}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals (shaded green area, boudn by dashed green lines). The selected point is the light blue point selected on the pareto front on the diagrams on the left. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

and Figure 8 Upper left: Pareto front on the borderline with respect to a_{ego}; Lower left: Pareto front in μ vs v_{ego}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals and showing the selected light blue point on the pareto front on the left diagrams. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

also show the comparison of coverage between test cases generated with Active DoE and those based on the full factorial design. We can observe, especially in Figure 7 Upper left: Pareto front on the borderline with respect to $d_{rem,safe}$; Lower left: Pareto front regarding d_0 and to v_{tof}; Right: Prediction of influence of the input parameters on the KPIs with confidence intervals (shaded green area, boudn by dashed green lines). The selected point is the light blue point selected on the pareto front on the diagrams on the left. Dark blue diamonds – Pareto front, Blue circles – test cases generated with Active DoE, Green triangles – test cases generated based on the full factorial design, Gray points – random samples that do not satisfy optimization constraints, Yellow points – random samples that satisfy optimization constraints.

, that full factorial does not cover some critical ranges.

5. Conclusion

This study proposes a methodology to identify relevant test cases and reduce overall simulation effort in comparison to a full-factorial test design. The methodology is based on the Active DoE approach available in AVL CAMEO™. All required workflow steps and advanced algorithms are available in easy to use wizards that allow for preparing data, executing tests, modelling surrogates, analysis and optimization, with easy interfacing to many industry standard and custom toolchains. To perform KPI prediction with surrogate models, we use the built-in RNN architecture, which involves split optimization and localized behavior modelling, to make the model computation efficient and to close the gap to existing approaches.

The test cases are simulated using the co-simulation framework AVL Model.CONNECT™, which synchronizes time steps and manages signal exchange beween various specialized software like AVL VSM™, VTD and others. The benefit of using co-simulation is in getting higher simulation model quality for each simulation aspect, such as vehicle dynamics or environment.

To define the evaluation plan, we applied a requirements-based approach defined through safety, comfort and legal aspects as evaluation criteria. Firstly we performed 1512 simulations based on the full-factorial design to derive a baseline. The results are compared to the Active DoE approach, which uses a sequential and optimized approach to sample test cases within a predefined RoI. With a 68% reduction in simulation test cases, with respect to the full-factorial design, high-quality surrogate models are derived. All KPI models, except the j_{ego}, achieve an accuracy of 0.96 or more in terms of R^2. Even higher reduction is possible to still achieve good model quality, yet our focus within this study was to model high-precision surrogates. It is important to highlight that more than half of the sampled test cases are within the RoI, which demonstrates the efficiency of the search for the relevant test case. This is a considerably better result than in the full-factorial design where only 11.7% are in the RoI.

It is important to note here that it is not possible to reduce much further how many points fall outside of the RoI, as there is a requirement to ensure that no unexplored RoI areas remain. There is no requirement for the model quality outside of the RoI areas to be particularly good, however the point density must be such that we can be sure that no areas that meet the criteria for RoI are missed.

It has been observed that, in general, initial cut-in distance and TOF velocity play a significant role in the majority of KPIs. However, the exceptions are longitudinal acceleration and jerk, which show strong dependency on the tire-road friction coefficient. To identify the corner test cases for safety requirements, we performed multiobjective generic optimization using derived surrogates.

This study has been based on the relatively simple cut-in scenario where environmental influences like lighting and weather were neglected. However, the methodology is capable of dealing with much higher numbers of parameters. This methodology could therefore be applied to more challenging scenarios with a higher number of input parameters and KPIs. The platform is already cloud capable, as parallelization in the cloud is anticipated as a requirement for input variation numbers over 15.

Acknowledgment

The authors gratefully acknowledge Mr. Leo Xenakis (AVL List GmbH) for the review.

References

[1] E. Thorn, S. C. Kimmel, M. Chaka and B. A. Hamilton, “A framework for automated driving system testable cases and scenarios,” Department of Transportation. National Highway Traffic Safety Administration, United States, 2018.

[2] N. Kalra and S. M. Paddock, “Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?,” *Transportation Research Part A: Policy and Practice,* vol. 94, pp. 182-193, 2016.

[3] E. de Gelder and J.-P. Paardekooper, “Assessment of automated driving systems using real-life scenarios,” *2017 ieee intelligent vehicles symposium (iv),* pp. 589-594, 2017.

[4] G. E. Mullins, P. G. Stankiewicz and S. K. Gupta, “Automated generation of diverse and challenging scenarios for test and evaluation of autonomous vehicles,” *2017 IEEE international conference on robotics and automation (ICRA),* pp. 1443-1450, 2019.

[5] F. Batsch, A. Daneshkhah, V. Palade and M. Cheah, “Scenario optimisation and sensitivity analysis for safe automated driving using gaussian processes,” *Applied Sciences,* vol. 11, no. 2, p. 775, 2021.

[6] H. Beglerovic, A. Ravi, N. Wikström, H.-M. Koegeler, A. Leitner and J. Holzinger, “Model-based safety validation of the automated driving functio highway pilot,” *8th International Munich Chassis Symposium 2017,* pp. 309-329, 2017.

[7] A. Rainer and H.-M. Koegeler, “Iterative DoE-improved emission models and better optimisation results within a shortened measurement time,” *International Journal of Powertrains,* vol. 6, no. 2, pp. 109-124, 2017.

[8] C. Hametner and S. Jakubek, “Local model network identification for online engine modelling,” *Information Sciences,* vol. 220, pp. 210-225, 2013.

[9] M. Deregnaucourt, M. Stadlbauer, C. Hametner, S. Jakubek and H.-M. Koegeler, “Evolving model architecture for custom output range exploration,” *Mathematical and Computer Modelling of Dynamical Systems,* vol. 21, no. 1, pp. 1-22, 2015.

[10] “Technical reference - Autonomous vehicles - Part 1: Basic behaviour,” Singapore Standards Council, 2019.

[11] M. M. Minderhoud and P. H. Bovy, “Extended time-to-collision measures for road traffic safety assessment,” *Accident Analysis & Prevention,* vol. 33, no. 1, pp. 89-97, 2001.

[12] F. Bella and R. Russo, “A collision warning system for rear-end collision: a driving simulator study,” *Procedia-social and behavioral sciences,* vol. 20, pp. 676-686, 2011.

[13] X. Qu, Y. Kuang, E. Oh and S. Jin, “Safety evaluation for expressways: a comparative study for macroscopic and microscopic indicators,” *Traffic injury prevention,* vol. 15, no. 1, pp. 89-93, 2014.

[14] S. Shalev-Shwartz, S. Shammah and A. Shashua, “On a formal model of safe and scalable self-driving cars,” *arXiv preprint arXiv:1708.0637,* 2017.

[15] W. Hulshof, I. Knight, A. Edwards, M. Avery and C. Grover, “Autonomous emergency braking test results,” *Proceedings of the 23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV),* pp. 1-13, 2013.

[16] L. L. Hoberock, “A survey of longitudinal acceleration comfort studies in ground transportation vehicles,” Council for Advanced Transportation Studies, 1976.

[17] S. Krašna, A. Keller, A. Linder, A. P. Silvano, J.-C. Xu, R. Thomson and C. Klug, “Human response to longitudinal perturbations of standing passengers on public

transport during regular operation," *Frontiers in bioengineering and biotechnology, Frontiers Media SA,* vol. 9, 2021.

[18] T. Nguyen, N. NguyenDinh, B. Lechner and Y. D. Wong, "Insight into the lateral ride discomfort thresholds of young-adult bus passengers at multiple postures: Case of Singapore," *Case Studies on Transport Policy,* vol. 7, no. 3, pp. 617-627, 2019.

[19] I. Bae, J. Moon and J. Seo, "Toward a comfortable driving experience for a self-driving shuttle bus," *Electronics,* vol. 8, no. 9, p. 943, 2019.

[20] U. N. E. C. for Europe (UNECE), "Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with," 2020.

[21] B. Rogic, D. Nalic, A. Eichberger and S. Bernsteiner, "A novel approach to integrate human-in-the-loop testing in the development chain of automated driving: The example of automated lane change," *IFAC-PapersOnLine,* vol. 53, no. 2, pp. 10188-10195, 2020.

[22] M. Yang, X. Wang and M. Quddus, "Examining lane change gap acceptance, duration and impact using naturalistic driving data," *Transportation research part C: emerging technologies,* vol. 104, pp. 317-331, 2019.

[23] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation,* vol. 6, no. 2, pp. 182-197, 2002.