Transportation Research Part C 154 (2023) 104243

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

# Assessing safety functionalities in the design and validation of driving automation

Angelo Coppola *, Claudio D'Aniello, Luigi Pariota, Gennaro Nicola Bifulco

*Department of Civil, Environmental and Architectural Engineering, University of Naples Federico II, Via Claudio, 21, Napoli, NA 80125, Italy*

ARTICLE INFO

*Keywords:*
Automated vehicle system
Safety assessment
Testing of automated driving function
Vulnerable road users
Uncertainty analysis
Cooperative connected and automated mobility

ABSTRACT

This paper aims to contribute to the comprehensive and systematic safety assessment of Automated Driving Systems (ADSs) by identifying unknown hazardous areas of operation. The current methodologies employed in this domain typically involve estimating the distributions of situational variables based on human-centered field test, crash databases, or expert knowledge of critical values. However, due to the lack of a-priori knowledge regarding the influential factors, their critical ranges, and their distributions, these approaches may not be entirely suitable for the assessment of emerging automated driving technologies. To deal with this challenging problem, here we propose a testing methodology incorporating realistic yet unobserved driving conditions, distinguished by numerous situational variables, so to encompass unknown unsafe conditions comprehensively. Our methodology utilizes stochastic simulation and uncertainty modeling techniques to account for the variability of realistic driving conditions and their impact on ADSs' performances. By doing so, we aim to identify unsafe operational regions and triggering conditions that can lead to hazardous behaviors, thus improving the development and safety of automated driving functions. For our purposes, the Latin Hypercube Sampling technique and the recently proposed PAWN density-based sensitivity analysis method are employed. We apply this methodology for the first time in the specific field of ADSs design and validation, using an exemplificative use case. We discuss and compare the results obtained from our approach with those obtained from a traditional approach.

## 1. Introduction

The trajectory towards fully automated driving is currently in progress and can reshape future mobility. The spread of vehicles equipped with advanced automated driving capabilities that can accomplish the entire range of Dynamic Driving Tasks (DDTs) is anticipated to significantly enhance road safety (Singh, 2015), optimize energy consumption (Wadud et al., 2016), and alleviate traffic congestion (Bengler et al., 2014). Nonetheless, before market deployment, newly developed or emerging Automated Driving Systems (ADSs) must undergo thorough testing to guarantee acceptable levels of reliability and safety, which are primary objectives to accomplish. Consequently, the assessment of the safety performance of ADSs is a crucial and increasingly important matter in the automotive industry (Varga et al., 2023).

The assessment of innovative driving solutions is carried out at different stages of the development process in the automotive field,

* Corresponding author.
*E-mail addresses:* angelo.coppola@unina.it (A. Coppola), claudio.daniello@unina.it (C. D'Aniello), luigi.pariota@unina.it (L. Pariota), gnbifulc@unina.it (G.N. Bifulco).

https://doi.org/10.1016/j.trc.2023.104243
Received 6 March 2023; Received in revised form 15 June 2023; Accepted 5 July 2023
Available online 8 July 2023

from design to pre-sale prototypes and Field Operational Tests (FOTs). Distance-based methods were initially used, but they resulted to be inefficient due to the large number of testing kilometers required to assess safety and reliability performance of ADSs (Kalra and Paddock, 2016). Consequently, scenario-based testing has become the primary adopted methodology. The scenario-based approach focuses on intentionally varying and analyzing the operating scenarios of ADSs, emphasizing the more challenging conditions a vehicle may encounter. The SOTIF (Safety of The Intended Functionality) paradigm, based on the ISO standard (ISO 21448:2022, 2022), guides the verification/validation process for ADSs. This standard extends the functional safety model used for conventional vehicles (International Organization for Standardization, 2011), with a focus on hazardous events resulting from insufficiencies of the intended functionality. ISO 21448 demands the identification of unknown unsafe regions of operation in as much detail as possible. The ADSs' Operational Design Domain (ODD) establishes the boundaries of the entire *operation region*, identifying the conditions under which a given ADS is designed to operate. In any given ODD, a comprehensive and systematic safety assessment plays a vital role in the verification/validation process to ensure ADSs meet the requirements of ISO 21448.

In this context, we propose a methodology to *analyze influencing factors*, i.e., the parameters of an ADS and/or of the related ODD that may influence the system's performance. However, due to the limited or absent experience with innovative ADSs, a-priori knowledge of the critical range and distributions of these parameters is missing. Therefore, testing must also ensure sufficient *coverage* of scenarios, which involves exploring the scenario space to ensure that all factors that could affect the System Under Test (SUT) are adequately explored.

The conventional assessment methodology, where a-priori information is available, is the so-called Test Matrix (TM), as described by Zhao et al. (2017) and Menzel et al. (2018). Under the assumption that all parameters in the ODD are discrete and have finite values, the tests cases are selected based on cluster analysis of historical datasets, such as Crash Imminent Test Scenarios for Integrated Vehicle-Based Safety Systems, as described by Najm and Smith (2007) and in the National Motor Vehicle Crash Causation Survey (Highway Traffic Safety Administration, 2008). Possible critical situations are identified based on expert knowledge or data analysis (e.g., observing the frequency of occurrence, the degree of risk or injury level of recorded accidents). Test repeatability is a significant advantage of TM methodology, which made it a standard in many FOTs, such as Autonomous Emergency Braking (AEB) protocol (Euro NCAP, 2021), APROSYS (APROSYS, Dissemination and Use Plan M1-M60 Deliverable D8.2.10", 2009) and ASSESS (Lemmen et al., 2012). For a complete overview of studies exploiting the TM methodology, see table 5 in Alghodhaifi and Lakshmanan (2021) and related references. Given its ease of use, the TM will probably remain the preferred methodology for safety performance assessment (Zhao et al., 2017) under simplified testing conditions. However, its suitability for validating autonomous driving solutions has been recently put at stake. A major limitation is that the exhaustive exploration of scenarios to achieve the coverage requirement is only applicable if the ODD is reasonably small (Sun et al., 2022; Zhang et al., 2021). Moreover, the unsafe tested scenarios in current TM-oriented FOTs are only a tiny subset of real-world hazardous conditions as accidents, fortunately, are rare events to observe. This limitation becomes particularly apparent when testing innovative systems, as the required observations for large datasets on which the TM methods are founded are not suitable. Indeed, the algorithms (or technologies) being evaluated may exhibit superior performances within a limited set of predefined conditions, thereby potentially undermining the overall effectiveness of the testing process (Zhao et al., 2017). For instance, in the EuroNCAP tests for AEBs, some situational variables are fixed, like the road alignment, the speed of surrounding vulnerable users (e.g., bicyclists and pedestrians), or the obstacle position. This work will prove the limitations of such an approach, which considers only a deterministically identified (small) set of conditions.

To deal with the problem of covering the vast sets of testing conditions due to the explosion of parameter combination (wide ODDs with many parameters), Combinatorial Testing (CT) has been widely adopted, as by Amersbach and Winner (2019). The main idea is to generate a minimum set of test cases satisfying the so-called *T-wise* coverage, that is all T-tuples of the considered parameter must be tested at least once. The methodology assumes that most system defects are due to a subset of factors and could be exploited by examining, according to Kuhn et al. (2010), the combinations of both scenario configuration parameters and SUT's input parameters. Since a failed scenario could emerge from a T-wise combination of configuration and input parameters, the number of tests could drastically increase again. To overcome the problem, some researchers recently combined CT with other methods, such as the Bayesian one to realize the black-box optimization (Gao et al., 2020) or backtracking algorithm and motion planner (Rocklage et al., 2018). Moreover, as for TM, the CT methodology assumes discrete parameters, whereas they are prevalently continuous in real-world scenarios. In any case, the T parameters should come from empirical evidence currently absent for testing innovative ADSs.

In summary, testing new or emerging technologies presents several significant challenges. One issue is the exploration of use cases that do not exist in currently available datasets. Additionally, new or improved technologies that empower existing driving functions can result in unexpected critical conditions or driving circumstances that are not present in currently adopted testing datasets. The number of unknown unsafe scenarios may be much greater than that of known dangerous scenarios, and it is questionable if the data from human-centered FOTs or crash databases are valid to obtain the empirical distributions of the influencing factors needed to test the performance of innovative ADSs (Batsch et al., 2021). Furthermore, concerning the ODD of innovative ADSs, it is unknown how new technologies interact with factors related to the heterogeneity of surrounding road users, their behaviors, the variability of road characteristics or weather conditions, and the highly non-linear interactions among all influencing factors. Therefore, there is a need to improve the assessment procedure of ADSs by considering realistic (but unobserved) driving conditions into a test to cover unknown unsafe scenarios.

A possible way to deal with this problem is to improve the representativeness of available datasets using Naturalistic Driving Studies (NDSs) (see Tables 3 and 4 in Alghodhaifi and Lakshmanan (2021)), where automated vehicles are tested under real-world driving conditions. Thus, also real-world dispersion of factors is experienced. However, this approach is challenging during the early stages of the development process due to the high number of kilometers of road tests required (Kalra and Paddock, 2016) since the rate of exposure to hazardous conditions is low (Favarò et al., 2017). As a result, assessment results are not robust enough.

A different approach to explore potentially unsafe scenarios is employing stochastic simulations, such as Crude Monte Carlo (CMC). CMC involves sampling situational variables from estimated distributions, including those from existing databases and distributions based on NDSs data. For factors that are unknown or not covered by available data (or invalidated by innovation), uniform/normal distribution is assumed over a feasible range of variability. A significant advantage of the CMC method is the stochasticity of scenarios (de Nicolao et al., 2007; Eidehall and Petersson, 2008; Yang and Peng, 2010). For example, de Nicolao et al. (2007) developed a stochastic pedestrian model to design the Risk Assessment Unit of the PROTECTOR project properly. They evaluated the collision probability for each pedestrian-vehicle configuration by a Monte Carlo strategy. A comprehensive review of Monte Carlo simulation studies can be found in Alghodhaifi and Lakshmanan (2021), Batsch et al. (2021), and (Zhang et al., 2021). Since critical events are rare, CMC-based approaches are inefficient in finding critical scenarios. To this end, variance reduction techniques, such as Importance Sampling (IS) (Glynn and Iglehart, 1989), have been proposed (Xu et al., 2018). For instance, Zhao et al. (2017) exploited the IS approach to assess an automated vehicle in cut-in, lane-changing, and car-following scenarios, proving that the sample size can be reduced by 300 to 100,000 times w.r.t. the CMC method. However, IS can only be applied in low-dimensional situations (Au and Beck, 2003), and a minimal level of a-priory knowledge about the parameters' distributions is required (see Alghodhaifi and Lakshmanan (2021) and Batsch et al. (2021)).

The methodologies previously described rely on expert knowledge or available data to estimate situational variable distributions in ADS testing, which may not be suitable for new or emerging automated driving technologies due to the lack of reliable data. Additionally, these methods may not adequately cover unsafe or unknown conditions for an ADS. Therefore, a more unbiased approach that addresses real-world driving conditions is necessary for ADS testing.

To address the challenges related to testing ADSs under realistic and uncertain driving conditions, we propose a testing methodology that utilizes stochastic simulation and uncertainty modeling techniques. Specifically, it employs a quasi-random Monte Carlo technique, namely the Latin Hypercube Sampling (LHS), and the PAWN density-based sensitivity method. The primary aim is to quantify the impact of situation variability on the pass-fail rate of a driving automation use case. We will compare the results of our proposed approach with those obtained from a deterministic testing approach to evaluate its effectiveness.

As far as we know, the PAWN method (Pianosi and Wagener, 2018, 2015) has not yet been used to test ADSs. Pianosi et al. (2020) provide an in-depth overview of the current fields of application of the method. Combining the PWAN method with the LHS approach has been previously tested to address sampling issues for unknown distributions. The combination of these two methods ensures that

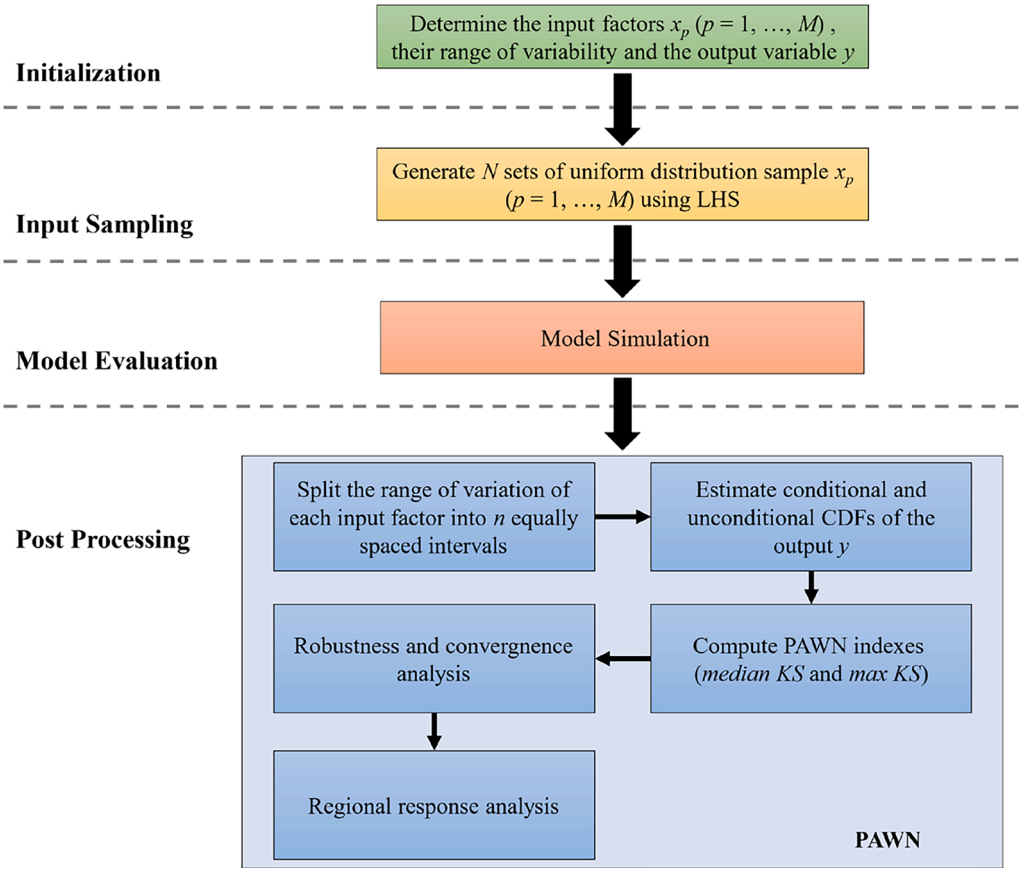


**Fig. 1.** Schematic representation of the methodological framework employed in this study.

the proposed approach meets the coverage requirements for effective scenario exploration. That is, the proposed approach satisfies the requirements of SOTIF for the reliable verification of ADSs. It is possible to identify safe and unsafe operational regions, as well as the triggering conditions leading to hazardous behaviors, to improve the development of automated driving functions and enhance the safety of ADSs.

The LHS technique (Helton and Davis, 2003) allows quasi-random sampling of situational variables to explore the full range of situations in a cost-effective manner. This avoids the creation of clusters or gaps that can distort statistical analyses of simulation results or increase computational effort. Situational variables are sampled from physically feasible ranges of variation rather than being inferred from real-world driving data. Rather, the study aims to discover unsafe unknown situations and unveil harmful conditions. As the method performs a full exploration of the space of situational variables, it is possible to encounter combinations of factors that are more critical than those of the testing scenarios based on known real-world data, which represent only a small subset of cases in deep uncertainty conditions. Thus, the proposed approach encompasses a Worst-Case Scenario Evaluation (WCSE) (Sun et al., 2022) by exploring the critical values of situational variables when there is no other information available on unknown unsafe conditions or when such information is deliberately ignored. Moreover, the PAWN density-based sensitivity analysis method allows identifying the situational variables that contribute the most to the pass-fail rate variability and the values that trigger failures. This information can help understand the sources of harmful conditions for a given use case and improve the testing protocol. The PAWN method is employed since: (*i*) it is easy applicable to nonlinear models; (*ii*) it is independent from the type of output distributions (symmetric, multimodal, or highly skewed); (*iii*) provide robust results for a relatively low sample size.

The proposed approach is tested for robustness on the EuroNCAP AEB Car-to-Bicyclist Nearside Adult 50% (CBNA-50) scenario. The approach evaluates the impact of several factors, including vehicle speed, bicycle speed and size, obstacle position, and road slope, on the pass-fail rate of the trial. Trials are generated through MATLAB/Simulink simulation environment, which emulates the motion of vehicles, the virtual road environment, and the AEB response to interactions with the surrounding environment. Although the automation functions of this use case are not highly complex, the study demonstrates that even moderately complex cases can fail to discover unknown unsafe testing conditions through deterministic approaches.

In this paper, we describe the proposed methodology in section 2. Section 3 describes the use case, the simulation environment, and the design of the experiment. Results are discussed in section 4, and conclusions and future research perspectives are drawn in Section 5.

## 2. Methodology

To achieve the objectives described in Section 1, we apply the methodological framework outlined in Fig. 1. The framework ground on the well-known modelling under uncertainty. Since engineering applications are pervaded by uncertainties, deriving from different and heterogeneous sources, retrieving information on the magnitude, composition, and most significant sources of uncertainty is essential for accurate decision-making process. These systems (or their models) are characterized by several parameters, which we refer to as *influencing factors* or simply *factors*. These factors include situational variables that describe the scenario and boundary conditions in which the system operates. The greater the number of factors, the higher the uncertainty of the *quantity of interest* (*output*), a summary scalar variable measuring the system's performances. The proposed framework employs a simulation model of an ADS, and then propagate the uncertainty of factors by means of simulations.

With specific regard to the field of traffic simulation, the methodology has been applied in the study of car-following (Punzo et al., 2021), charging strategy for electric buses (Liu et al., 2021), demand flow (Dantsuji et al., 2022) and assessment of connected vehicles on the performance of traffic networks (Coppola et al., 2022). To the best of our knowledge, the methodology has never been applied to investigate the performances of ADSs.

Given the use case and the model of the SUT, the methodology consists of four consecutive steps. The first one involves the identification and characterization of the different source of uncertainty that could affect the performance of the SUT (Roy and Oberkampf, 2011). Factors assumed as non-influential or whose variation is not of interest for the experiment are keep fixed. Conversely, the input factors to be analyzed are randomly chosen from all the levels of interest. Hence, it is necessary to characterize a feasible variability space of each input factors according to its physical meaning. Moreover, if the input factors are considered stochastic variables, their Probability Density Functions (PDFs) must be defined. If the PDFs are unknown and/or the comparison of simulated occurrence with real-world ones is not of interest, a common approach is to assume them as uniformly and independently distributed within their range of variability. It is important to note that a trade-off between the number of factors considered and their range of variability and the need to limit the output variance is necessary to obtain robust results. In this study, the uncertain inputs are assumed to be uniformly and independently distributed with unknown PDFs.

The second step of the methodology entails selecting a quantity of interest (output), which is an output variable that serves as an indicator of the *degree* of pass-fail of the SUT in the given use case. This variable measures the extent to which an ADS has accomplished or failed to complete the designated driving task.

The third step involves mapping the PDFs of factors into the output distribution. To achieve this, a simulation model emulating the SUT behavior and the surrounding environment in used. Then, several input sets are randomly generated in the entire input feasibility space to estimate the empirical CDF of the output. Each set of the sample consists of a tuple containing a random value of each uncertain input factor. Latin Hypercube Sampling (LHS) approach is employed to generate a quasi-random sample in the multi-dimensional space of input factors, so to ensure an exhaustive input mapping by evenly exploring the whole input space. Then, the simulation model is run for each set of the sample, and the quantity of interest (output) is calculated. It is worth noting that the sample size significantly affects the computational effort, and the optimal sample size may vary depending on the number of input factors and

the complexity of the model's response. Thus, a trade-off between sample size and reliable results is required.

The final step is to perform *a sensitivity analysis* (Razavi et al., 2021) to quantify the relative contribution of each uncertain input factor to the variability of the output. The selection of the specific sensitivity method to employ depends on the purpose of the study, available computing resources, and specific problem characteristics, such as the linearity or non-linearity of the input–output relationship and the characteristics of the output distribution.

The following section introduces and explain the sensitivity analysis method employed in this study, which differs from those previously used in the transport sector.

### 2.1. PAWN sensitivity analysis

Let us consider a generic input–output relationship

$$y = f(\mathbf{x}) \tag{1}$$

where $\times = [x_1, x_2, \ldots, x_p, \ldots, x_m] \in \mathscr{H} \subseteq \mathbb{R}^M$ is the input vector of $M$ factors, while $y \in \mathbb{R}$ is a scalar output variable.

Global Sensitivity Analysis (GSA) approaches consider the variation of each input factor within its space of variability to identify the interaction with the output $y$. According to Saltelli et al. (2008), GSA approaches have three primary purposes: (*i*) *Factor Prioritization* (FP) or *Ranking*, i.e., ranking the inputs factors based on their relative contribution to output uncertainty; (*ii*) *Factor Fixing* (FF) or *Screening*, i.e., determining the inputs that do not give any contribution to output uncertainty; (*iii*) *Factor Mapping* (FM), i.e., determining the regions in the inputs space that produce specific output values. To determine the relative impact of each input factor on the output variability, sensitivity indexes are computed. These indexes range from 0 to 1, with higher values indicating greater influence and a value of zero indicating non-influence. This allows us to achieve the aforementioned objectives.

The technical literature commonly employs variance-based methods to determine the relative contribution of input factors to the output variability. These methods assume that the second-order moment (variance) is an appropriate indicator of output uncertainty. However, for output distributions that are multi-modal or highly skewed, this assumption can be unsuitable and yield contradictory outcomes (Borgonovo et al., 2011).

To overcome the above-described limitation, we employ the recently developed PAWN method, as described in Pianosi and Wagener (2015) and Pianosi and Wagener (2018), to quantitatively assess the influence of input factors on the output. It has been widely applied in climate or atmospheric sciences for assessing complex environmental models, as in Chaturvedi and Rajasekar (2022), Koo et al. (2020) and Yang et al. (2019). The employed method belongs to the density-based category, which use sensitivity indexes that are independent of the type of output distributions (symmetric, multimodal, or highly skewed) since the entire probability distribution is considered. These methods measure sensitivity by estimating the variations induced in the output distribution when removing the uncertainty about one (or more) input. Most existing methods estimate the sensitivity index of an input $x_p$ by measuring the distance between the unconditional PDF of the output $y$ (all inputs vary simultaneously) and the conditional distributions obtained by changing all inputs but $x_p$. However, PDFs are typically unknown, and their estimation (e.g., via kernel density estimation methods) could require a significant computational effort. Unlike other density-based methods, the PAWN one characterizes the conditional and unconditional distributions by their Cumulative Distribution Functions (CDFs). To measure the distance between these CDFs, the Kolmogorov-Smirnov (*KS*) statistic is exploited:

$$KS(x_p) = \max_y \left| F_y(y) - F_{y|x_p}(y|x_p) \right| \tag{2}$$

where $F_y(y)$ is the unconditional CDF of the output $y$ and $F_{y|xp}(y|x_p)$ is the conditional CDF of the output $y$ when the input factor $x_p$ ($p = 1, \ldots, M$) is kept fix. Since the $KS(x_p)$ statistic is a function of $x_p$, the related PAWN index $S_p$ considers a statistic (maximum, median or mean) over all possible value that $x_p$ can assume:

$$S_p = \operatorname*{stat}_{x_p} \max_y \left| F_y(y) - F_{y|x_p}(y|x_p) \right| \tag{3}$$

Considering Eq. (2), the previous equation can be also stated as

$$S_p = \operatorname*{stat}_{x_p} \left[ KS(x_p) \right] \tag{4}$$

Like other sensitivity indexes, the PAWN one assumes values in the range [0, 1] and the higher is its value, the more influential is $x_p$. Therefore, it can be used for both FP and FF purposes.

Due to the complexity of the input–output relationship, the computation of PAWN indexes needs to be approximated numerically. This also opens the possibility of exploiting datasets generated by a generic sampling strategy, such as the LHS employed in this work. The estimation procedure foresees dividing the range of variation of each input factor $x_p$ into $n$ equally spaced intervals. The generic interval is denoted as $\mathfrak{I}_k$ and is adopted to define the conditional sample $YU_{pk}$ ($\forall p \in M$). The unconditional sample $YU$ corresponds with the entire sample $Y$ or a subsample (this will be clearer once the regional response is introduced). Accordingly, the estimated PAWN sensitivity indexes can be computed as follows:

$$\widehat{S}_p = \operatorname*{stat}_{k=1,\cdots,n} \max_y \left| \widehat{F}_y(y) - \widehat{F}_{y|x_p}(y|x_p \in \mathscr{I}_k) \right| \tag{5}$$

where $F_y(y)$ is the empirical distribution of the unconditional sample $YU$, obtained by varying all factors at once; $F_{y|x}(y|x_p \in \mathfrak{I}_k)$ is

the observed distribution of a conditional sample $YU_{pk}$ obtained by varying all factors at once but $x_p$ ($\forall\, p \in M$), which is set to the k-th conditioning value $\bar{x}_p^{(k)}$ ($k = 1, \ldots, n$).

Since $KS(\mathscr{I}_k) = \max_y \left| \widehat{F}_y(y) - \widehat{F}_{y|x_p}(y|x_p \in \mathscr{I}_k) \right|$, Eq. (5) can be recast in a more compact way

$$\widehat{S}_p = \operatorname*{stat}_{k=1,\cdots,n} KS(\mathscr{I}_k) \tag{6}$$

The estimation procedure involves defining two tuning parameters, i.e., the sample size $N$ and the number $n$ of conditioning points. Regarding $N$, the more complex the model of the SUT, the higher the number of uncertainty factors, and the more extended the range for each of those, the larger the sample size. However, performing a large number of trials is the most computationally demanding step of the whole procedure. Hence, a trade-off between available computing resources and model complexity is necessary. More detailed suggestions about selecting the value of $N$ can be found in Pianosi and Wagener (2018), where authors provide two case studies considering models with different levels of complexity. As a rule of thumb, $N > 500$ is suggested. Regarding the number of conditioning points $n$, the proposed *rule of thumb* is $n \geq 10$.

A crucial step is to evaluate the robustness of the PAWN indexes to the chosen sample size. To this end, sensitivity computation is repeated using different bootstrap resamples (Efron and Tibshirani, 1994) of the input–output dataset to obtain a distribution of PAWN indexes, whose mean value is used as a robust estimation. Moreover, confidence intervals are derived to correctly infer whether differences in sensitivity indexes are significant enough to discriminate between relevant and non-relevant input factors (Sarrazin et al., 2016). In principle, the sensitivity index of a non-influential parameter should have a value of zero. However, due to estimation errors, non-zero indexes can be obtained also for the non-influential parameters. Accordingly, an artificial input factor (the dummy parameter labeled with $p = m + 1$) is introduced to identify influential and non-influential factors. The dummy parameter does not affect the output distribution, hence $F_{y|xp}(y|x_{m+1}) = F_y(y)$ for all conditioning points in Eq. (2) and $S_{dummy} = 0$. Again, due to estimation errors, it can assume non-zero values ($S_{dummy} > 0$) when using Eq. (5). If the sensitivity index of a factor is significantly larger than the one of the dummy, it is considered influential. Otherwise, nothing can be concluded about this input factor because its non-zero sensitivity may be due to an actual effect of the input on the output or approximation errors. The dummy parameter can be considered a measure of the accuracy in estimating estimated PAWN indices by exploiting a limited sample size. The procedure to prove the robustness of the sensitivity values, as well as to derive the PAWN index of the dummy parameter, includes: *i*) deriving the unconditional sample $YU$ by randomly extracting a subsample of size $N_c$ ($N_c = N/n$ is the size of the conditional sample) and repeat the subsampling for a prescribed number of times; *ii*) performing bootstrap without replacement from $YU$ since $N_c < N$; *iii*) deriving a distribution of estimated PAWN sensitivity indexes and confidence intervals by applying Eq. (6) for each bootstrap resample of $YU$; *iv*) computing the PAWN sensitivity of the dummy parameter and comparing it with the ones of the $M$ input factors.

The PAWN sensitivity of the dummy parameters can be estimated as follows:

$$\widehat{S}_{dummy} = \operatorname*{stat}_{k=1,\cdots,n} \max_y \left| \widehat{F}_y^{(k)}(y) - \widehat{F}_{y|x_{m+1}}(y) \right| \tag{7}$$

where $F_y^{(k)}(y)$ is the empirical distribution of the *k-th* resample of the unconditional output sample $YU$; $F_{y|xp}(y|x_{m+1})$ is the conditional distribution of the dummy parameter, equal to $F_y(y)$ since it does not affect the output distribution.

#### 2.1.1. Regional response sensitivity analysis

The PAWN sensitivity index $S_p$ can also focus on a specific sub-range of the output distribution rather than the entire range. By way of example, assume to estimate PAWN indexes over the range $y > y_t$, being $y_t$ a predefined minimum value of the output, Eq. (2) can be rewritten as follows:

$$\widehat{KS}(x_p) = \max_{y > y_t} \left| \widehat{F}_y(y) - \widehat{F}_{y|x_p}(y) \right| \tag{8}$$

The computation of PAWN indexes for different sub-ranges is easy since it does not require any new runs of the simulation but just the application of Eq. (5) and Eq. (8).

In technical literature, this technique for SA, able to focus on specific regions of the response surface, is known as Regional Response Probabilistic Sensitivity Analysis (RRPSA) and can be used for FM.

## 3. Experimental setup

As previously mentioned in Section 2, the implementation of the proposed methodology requires the propagation of the PDF of the input factors into the output distribution, which is achieved through simulation models. Here we shall elucidate the implementation of the methodology, delineating its four-step process. Prior to that, we shall introduce the use case along with the SUT and the (related) vehicle dynamics model. It is worth noting that, in what follows, the vehicle that is equipped with the SUT will be referred to as the Ego Vehicle.

It is worth to recall that this work does not aim to evaluate a specific driving assistance solution. Rather, it aims to develop a procedure to identify the factors that influence the output in the testing setup. This application effectively handles the uncertainty arising from various parameters that define operations in terms of initial and boundary conditions. As a result, it has the potential to be applied to numerous driving automation solutions and use cases.

Finally, we remark that the proposed testing methodology is, globally, implemented in the MATLAB/Simulink environment. In

what follows, for the sake of completeness, the simulation environment and tools used for each component will be defined.

### 3.1. Use-Case

We examinate the effectiveness of the proposed testing approach in a specific use case. The testing conditions we consider are the one of the CBNA-50 scenario (Euro NCAP Test Protocol-AEB/LSS VRU systems, 2021), which is an officially recognized EuroNCAP test. In this use case, the Ego Vehicle travels along a straight road at a constant speed towards a bicyclist crossing its path from the nearside. The Ego Vehicle is equipped with an AEB system (that is the SUT). The bicycle travels at a constant speed and approaches from the right side. Additionally, there is an obstacle on the right side of the road, blocking the Ego Vehicle's radar's field of view. The initial speed and position of both actors are set up such that without braking, the bicycle and the Ego Vehicle would collide at a predetermined point location (Fig. 2). Although the situation may seem straightforward, many parameters could impact the system's functionality and introduce significant uncertainty.

All the actors, their behavior (behavior of the bicyclist) and objects in the scenario are implemented in Simulink.

The official EuroNCAP test protocol specifies most of the testing variables as follows:

1. dry, uniform, solid-paved surface with a consistent slope between level and 1%;
2. daytime testing in dry conditions (no precipitation);
3. wind speeds below 5 [m/s];
4. Ego Vehicle speed in the range 20–60 [km/h], with 5 [km/h] incremental steps;
5. bicycle speed fixed at 15 [km/h];
6. position of the obstacle fixed at 3.55 [m] and 15 [m].

We use the official EuroNCAP test protocol as baseline for comparison purposes. The Ego Vehicle, with nine speed values ranging from 20 [km/h] to 60 [km/h], and the position of the obstacle, with two possible positions, are the only variables subjected to testing variability. For the sake of comparison, we also acknowledge the potential variation in road slope, which can assume values of $-1$, 0, or 1 [deg]. As for the wind speed, a constant value of 0 [m/s] is assumed. Taking into account all the feasible combinations of these discrete input parameters, the total number of tests amounts to 54. Here we dispute the approach adopted by the EuroNCAP test, assumed as an example of a testing procedure biased by limited a priori judgment on the effecting variables.

### 3.2. Ego Vehicle modelling

The longitudinal motion of the Ego Vehicle is here developed by the application of the Newton's laws of motion. More in detail, the vehicle state can be described by the set of forces acting along the direction of the velocity on the vehicle body (Coppola et al., 2022):

$$\begin{aligned} \dot{p}(t) &= v(t) \\ \dot{v}(t) &= u(t-\tau_a) - g\sin(\theta) - g\cos(\theta)\frac{C_r}{1000}(c_1 \cdot v(t-\tau_a) + c_2) - \frac{\rho_{air}}{2 \cdot m} C_D A_f v^2(t-\tau_a) \end{aligned} \tag{9}$$

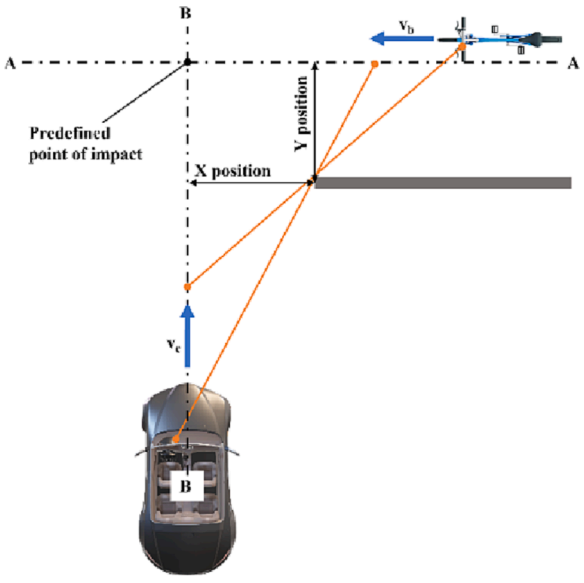


**Fig. 2.** Car-to-cyclist testing scenario, with specific measures to quantify the position and severity of a view-blocking obstruction.

where $\tau_a = 0.1$ [s] is the actuator delay; $p(t)$ [m] and $v(t)$ [m/s] are the position and the speed of the Ego Vehicle; $u(t)$ [m/s$^2$] is the driving/braking control acceleration; $m = 1430$ [kg] is the Ego Vehicle mass; $C_r = 1.75$, $c_1 = 0.0328$, $c_2 = 4.575$ are the rolling resistance parameters; $\rho_{air} = 1.2256$ [kg/m$^3$] is the air density; $C_D = 0.29$ is the air-drag resistance coefficient; $A_f = 2.46$ [m$^2$] is the frontal area of the Ego Vehicle; $g = 9.81$ [m/s$^2$] is the gravitational acceleration; $\theta$ [rad] is the road slope.

The Ego Vehicle is here assumed to be equipped with a GPS to measure the absolute position, an Inertial Sensor to measure the acceleration in the direction of the motion and a Wheel Speed Sensor to measure the velocity in the direction of the motion (Bifulco et al., 2011). To sense the surrounding environment, an ideal frontal vision system (e.g., a radar) with a horizontal field of view of $\pm$ 50 [deg] and a detection range of 150 m (Bosh Mobility Solutions) is employed. Furthermore, to realistically mimic the behavior of a real vehicle, we consider the following constraints acting on the vehicle dynamics in Eq. (8) and limiting its capabilities (European Parliament Regulation 2019/2144, 2019; Lengyel et al., 2020): *i*) maximum acceleration $a_{max} = 2.5$ [m/s$^2$]; *ii*) maximum deceleration $b_{max} = 9$ [m/s$^2$]; *iii*) maximum jerk $j_{max} = 20$ [m/s$^3$].

The Ego Vehicle model is implemented in Simulink.

### 3.3. SUT: Autonomous Emergency braking

The Autonomous Emergency Braking (AEB) (Yang et al., 2022) is an advanced assistant system that automatically starts braking in case of an impending collision when the driver fails to respond. Its primary objective is to prevent collisions with other road users or, if inevitable, minimize the impact by reducing the collision speed. Various AEB solutions are currently available in the market, and although there may be variations in the hardware and algorithm implementation, the system's fundamental operation is well-established. The functional specification of the SUT used in this study is as follows:

i. a vision system (e.g., a radar) calculates the relative distance and relative speed between the Ego Vehicle and the detected objects;
ii. a control logic activates the braking system when the Time-To-Collision (TTC) reaches a threshold of 1.5 [s]; according to Wang et al. (2021), the TTC is the inverse of the ratio between the spacing and the relative speed and assumes relevant values when the speed of the vehicle that follows is greater than the one of the leading vehicle;
iii. when activated, the braking system applies a deceleration of 6 [m/s$^2$], a value that ensures consistency with an Anti-Block System (ABS).

The SUT model and the radar system model are implemented in Simulink, exploiting the Automated Driving Toolbox.

### 3.4. Source of uncertainty

One of the objective of this study is to assess the relative influence of the input factors on the performance of the SUT. Hence, the identification and the characterization of the source of uncertainty is a crucial step.

Given the set of conditions and the elements defined by the CBNA-50 scenario (Section 3.1), we define the uncertain factors to investigate and the parameters to keep constant (are expected to not have an impact on the test.). For example, light conditions remain unchanged during this test as they have no effect on the radar. Each simulation trial also maintains certain distinct critical aspects of the experiment, such as the type of vehicles involved (a car and a bicycle), the path of the vehicles, and the expected point of impact. Conversely, all remaining aspects are varied and assume the role of input parameters for the considered use case. It is important to respect and leverage the relationships among the parameters in the real world. For example, in this use case, the conflict point, where the collision is anticipated to occur, remains the same for any initial conditions (position and speed) of the two road users considered and any combination thereof. As such, only two of these parameters can be modified arbitrarily and are hence independent. Independent parameters are the factors, while fixed parameters such as light conditions are not identified as factors.

For the considered use case, we have identified seven input factors, listed in Table 1. The longitudinal slope is the slope of the road in the direction of the vehicle's trajectory. The Ego Vehicle speed is the velocity the car maintains from the start of the simulation to the beginning of the braking maneuver. The bicycle speed is constant throughout the test. The length and width of the bicycle are the longitudinal and lateral dimensions of the bicycle. The obstacle's position (X and Y in Fig. 2) is the distance between the bicycle and the obstacle. For each considered input factor we have considered its physical-meaning range of variation, whose upper and lower bound

**TABLE 1**
Input Factors and their range of variability.

| Factors | Lower bound | Upper bound | Unit of measure |
|---|---|---|---|
| Longitudinal Slope | −3.45 | +3.45 | degree [deg] |
| Ego Vehicle Speed | 20.00 | 60.00 | kilometers per hour [km/h] |
| Bicycle Speed | 10.00 | 40.00 | kilometers per hour [km/h] |
| Bicycle Length | 1.40 | 2.00 | meters [m] |
| Bicycle Width | 0.50 | 0.65 | meters [m] |
| Obstacle position X (see Fig. 1) | 0.00 | 10.00 | meters [m] |
| Obstacle position Y (see Fig. 1) | 2.00 | 20.00 | meters [m] |

are reported in Table 1. Accordingly, these factors are modelled as random variables assumed to be independently and uniformly distributed within their feasible ranges of variation (unknown PDFs). As a consequence, factors can vary with continuous values, producing a continuous and homogeneous seven-dimensional space.

### 3.5. The quantity of interest

To evaluate the influence of each input factor to the performance of the SUT, i.e., evaluate the rate of success/failure, we compute a scalar output serving as *quantity of interest*. Specifically, we calculate the distance between the front bumper of the Ego Vehicle and the expected point of impact serving as the output of interest. This distance is calculated after the Ego Vehicle speed reaches zero. Positive distances indicate that the Ego Vehicle successfully avoided the collision by detecting the obstacle and activating the braking system in time (success). Conversely, negative distances indicate that the Ego Vehicle failed to avoid the collision, with the stop occurring after the expected point of impact (failure).

### 3.6. Model evaluation

The model evaluation step allows papping the PDFs of factors into the output distribution. To this end, the simulation model composed of the vehicle dynamics, the SUT, the cyclist and the obstacle is run several times. As preliminary step, a quasi-random sample in the multi-dimensional space of input factors, following a uniform distribution law, is generated via the LHS approach. Each set of the sample consists of a tuple (in this case a vector of $M$ elements) containing a random value of each uncertain input factor. This approach ensures an exhaustive input mapping by evenly exploring the whole input space. For each run, the quantity of interest (output) is calculated so to estimate its distribution. For the implementation, we used the *lhsdesign* function in MATLAB. Then, the simulation model is run in Simulink environment for each set of the sample, and the quntity of interest (output) is calculated. In our case, the total number of simulation runs is $N = 4000$.

### 3.7. Pawn

As described in Section 2.1, the PAWN method estimates the sensitivity indexes by distance between the conditional and unconditional CDFs of the output, as described by Pianosi and Wagener (2018). In our case, the $N = 4000$ simulation runs are used to estimate the unconditional distribution of the output (stop distance). Starting from this generic input–output dataset $<$ x,y$>$, the PAWN indexes are computed following the approximation procedure described in Section 2.1. Specifically, we split the range of variation of each input factor $x_p$ into $n = 20$ equally spaced intervals and define the conditional sample $YU_{pk}$ ($\forall p \in M$) accordingly. Then, applying Eq. (5) – Eq. (7) and the intermediate tasks, the approximated PAWN indexes are computed.

The PAWN analysis is implemented using the MATLAB adaptation of the latest version of the SAFE toolbox developed by (Noacco

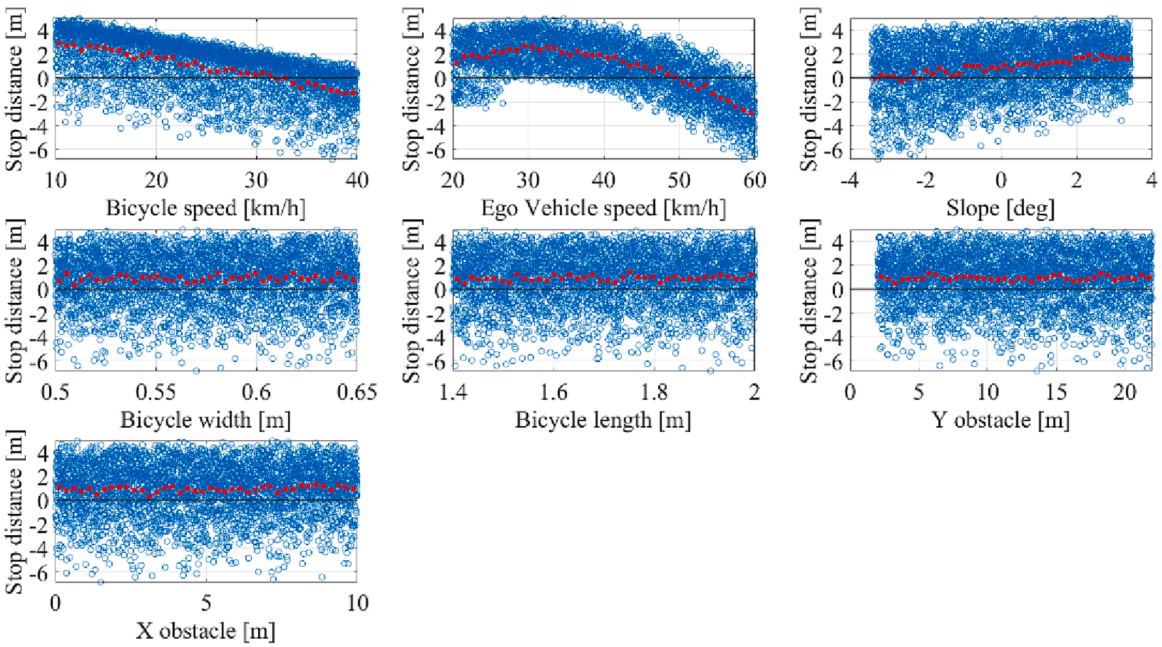


**Fig. 3.** Scatter plots of each input factor $x_p$ ($p = 1, \ldots, 7$) against simulations output. The black line is the zero value. The red points represent the average output calculated at each bin into which each input's range was divided. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

et al., 2019).

## 4. Results and discussion

In this section, we will examine the results of the $N = 4000$ simulations. We will analyze the relationship between the uncertain input factors (Table 1) and the output both qualitatively and quantitatively. It is worth recalling that the output is the stop distance, which is the distance between the front bumper of the Ego Vehicle and the expected point of impact, expressed in [m]. A negative value indicates a system failure resulting in a crash.

### 4.1. Visualization of results

The qualitative analysis is assessed through specific visual analyses, involving scatterplots and histograms. These visualizations provide straightforward and intuitive evidence of the relationship between the input factors and the output.

The procedure tests the seven factors $x_p$ ($p = 1, \ldots,7$) listed in Table 1. Fig. 3 illustrates the related scatter plots, which refer to the analysis of all the simulation runs and the same number of points (4000, one for each run) compose each plot. The y-axe of each sub-plot corresponds to the output variable $y$ (stop distance [m]), whereas the x-axe refers to a different factor. The plots show the average values of the output distributions for each factor (red points), whereas the black line represents the zero value. The system's success/failure is represented by points above/below the black line. The results demonstrate that the speed of the bicycle and of the Ego Vehicle significantly affect the success/failure. Indeed, the higher the bicycle speed (the same is for the Ego Vehicle speed), the more the SUT fails, as expected. Moreover, the average value distribution of the output for the bicycle speed highlights a threshold of approximately 32 [km/h], which identifies the separation between predominantly negative and predominantly positive values. With respect to the Ego Vehicle speed, the threshold for the separation between predominantly negative and predominantly positive values is around 50 [km/h]. Another significant factor is the slope of the road, but it appears to be less relevant, as the average line is slightly angled and there is significant dispersion across the average. As expected, the road slope has a negative impact on the SUT when the Ego Vehicle is going downhill (i.e., negative slope values) since it hinders the braking action of the AEB. The remaining factors do not seem to have a significant effect on the output.

The results of the regional analysis are in agreement with the previous findings (see Fig. 4). This analysis divides the output into two categories: successful (positive values, shown in blue) and failed (negative values, shown in orange). As previously discussed, the bicycle speed and the Ego Vehicle speed have the most significant impact on the distribution of the output histogram. In particular, the rate of failed tests surpasses that of successful tests at approximately 30 [km/h] and 48 [km/h] for the two respective input factors. Therefore, for this use case, the most unfavorable scenarios are those characterized by high values of these two input factors. The road slope also has some impact on the results, as expected. The lowest slope values, corresponding to a downward slope, result in the worst cases.

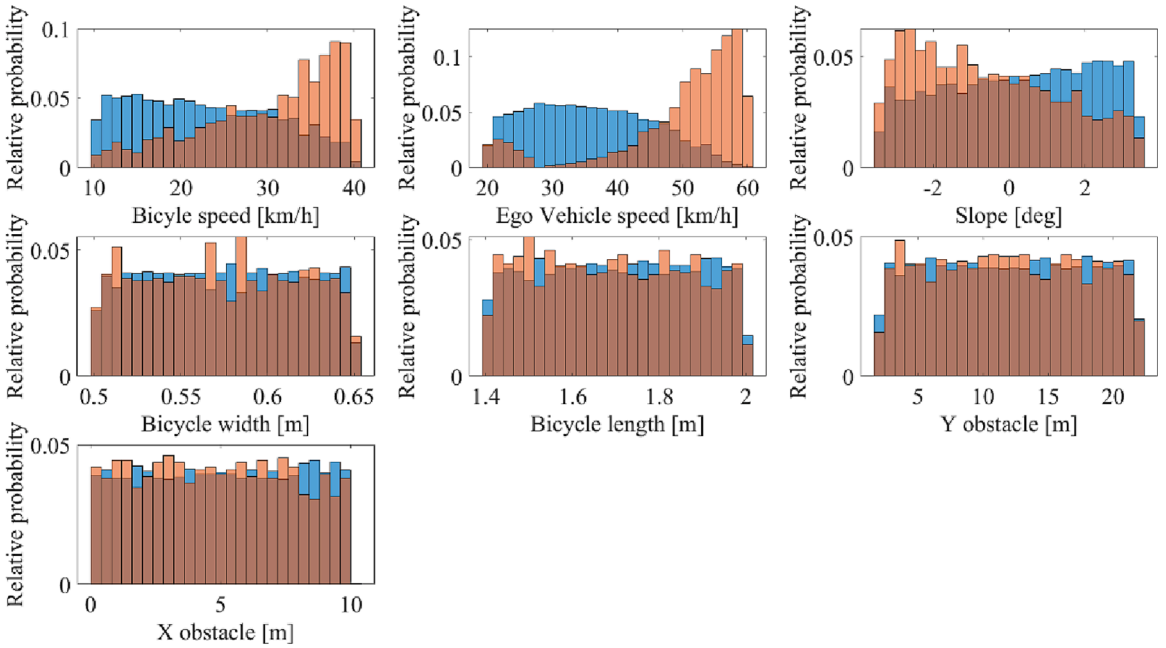


**Fig. 4.** Regional analysis. Histograms of each factor $x_p$ ($p = 1, \ldots,7$) in behavioral tests (succeeded) against CDF of factors in non-behavioral tests (failed). Behavioral tests are shown in blue, while non-behavioral tests in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 4.2. PAWN sensitivity analysis results

Here we discuss the impact of each input factor $x_p$ on the output $y$ (stop distance [m]) from the perspective of GSA. The PAWN method, introduced in Section 2.1, is employed to quantify the relative influence of input factors by computing their CDFs.

Before delving into the analysis, a preparatory step involves assessing the resulting empirical distribution of the output $y$ (Fig. 5). The results indicate that the stop distance falls within the range of [-6.843, 4.975] [m] depending on the input values according to the parameter uncertainty. The mean value is $\mu_y = 0.914$ [m], the median is $\mu e_y = 1.191$ [m], and the variance is $\sigma^2{}_y = 4.665$ [$m^2$]. According to the variance and the mean, the standard deviation is $\sigma_y = 2.160$ [m], and the coefficient of variation ($\sigma_y / \mu_y$, also known as relative standard deviation) is $cv_y = 3.363$. The shape of the output distribution in Fig. 5 indicates that it is negatively skewed with a left tail. To examine the asymmetry of the distribution appropriately, we calculate the third-order central moment $\mu_y{}^3 = E[(y - \mu_y)^3]$, and its standardized version $\gamma = E[(y - \mu_y)^3/\sigma_y{}^3]$ referred to as skewness: if the distribution of the considered variable is symmetric concerning the mean, then all odd-order moments are null. In this case, $\mu_{s,y} = -6.123$ [$m^3$] and $\gamma_y = -0.607$ confirm the asymmetry of the distribution, hence supporting our choice of adopting the PAWN method.

Then, we estimate the PAWN sensitivity indexes according to Eq. (5). The results of this analysis are presented in Fig. 6. The *Output CDFs* panels of Fig. 6 display the unconditional and conditional CDFs of the output variable $y$ (stop distance) in response to the uncertainty associated with each input factor $x_p$ ($p = 1, \ldots, 7$). Herein, the red lines represent the empirical unconditional output distribution $F_y(y)$, while the grey lines represent the conditional ones $F_{y|x}(y|x_p \in \mathfrak{I}_k)$. The more the conditional distributions are shifted w.r.t. the unconditional one, the higher the influence of $x_p$ on the output variability. In our case, left-shifted distributions indicate a negative impact, as they increase the probability of a negative value for the stop distance, while right-shifted distributions indicate a positive impact. The *Sensitivity Index* panels of Fig. 6 report the *KS* statistics for all the $n = 20$ conditioning points of each input factor as defined in Section 2.1. The dashed horizontal red lines in the sensitivity index panels indicate the threshold of significant sensitivity, which is the value assumed by the dummy variable according to its definition given in Section 2.1 We use grayscale for the circles representing the $n$ conditioning levels of each input factor $x_p$ corresponding to the grayscale used in the plots of the Output CDFs panels. The sensitivity index plots also report two statistics for *KS*: the median (as a summary statistic) and the maximum (to spot extreme behaviors). A significant dispersion and separation from the threshold line can be appreciated in the *KS* axis of bicycle speed and Ego Vehicle speed; this outcome indicates that these parameters significantly influence the stop distance. This trend is also apparent by examining the distance between conditional and unconditional CDFs. Upon considering the *median KS* PAWN sensitivity indexes, it is notable that they are equal to $S_m^{bs} = 0.365$ and $S_m^{es} = 0.320$ for bicycle speed and Ego Vehicle speed, respectively. This implies that, on average, the speed of the bicycle is the most crucial factor, followed by that of the Ego Vehicle. In the case of road slope, there exists a minor disparity between the *KS* and the threshold line (conditional CDFs are slightly shifted from the unconditional one). In this case, the PAWN index value $S_m^{sl} = 0.113$ suggests that this parameter is of secondary relevance. The average results of the other input parameters imply a low influence on the output variable $y$.

#### 4.2.1. Robustness of the results

With the aim to evaluate the reliability of the aforementioned findings, we conduct a robustness analysis utilizing the bootstrapping technique. Furthermore, a convergence analysis is performed on the selected sample size $N$.

We perform 50 bootstraps resamples without replacement to estimate the distribution of sensitivity indexes of input factors. The results in Fig. 7 confirm that the most influential input factor is the bicycle speed, followed by the Ego Vehicle speed and, finally the road slope. The median sensitivity values of the remaining parameters are below the threshold line, indicating that they may have a low influence on the output or that their estimated sensitivity is due to random effects. Therefore, these parameters are classified as potentially uninfluential. Regarding the influential factors, the distribution boxes indicate that the bicycle speed sensitivity index is less dispersed and skewed than the other two input factors. The Ego Vehicle speed, on the other hand, has the most dispersed distribution. This outcome highlights how the *median KS* sensitivity index of each factor could be affected by the estimation errors and/or

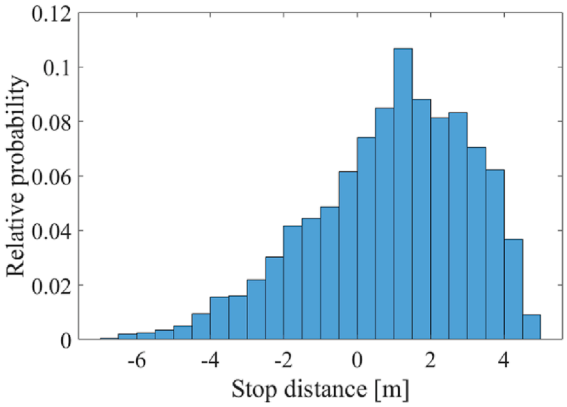


**Fig. 5.** Empirical unconditional PDF $f_y(y)$ of the output y.

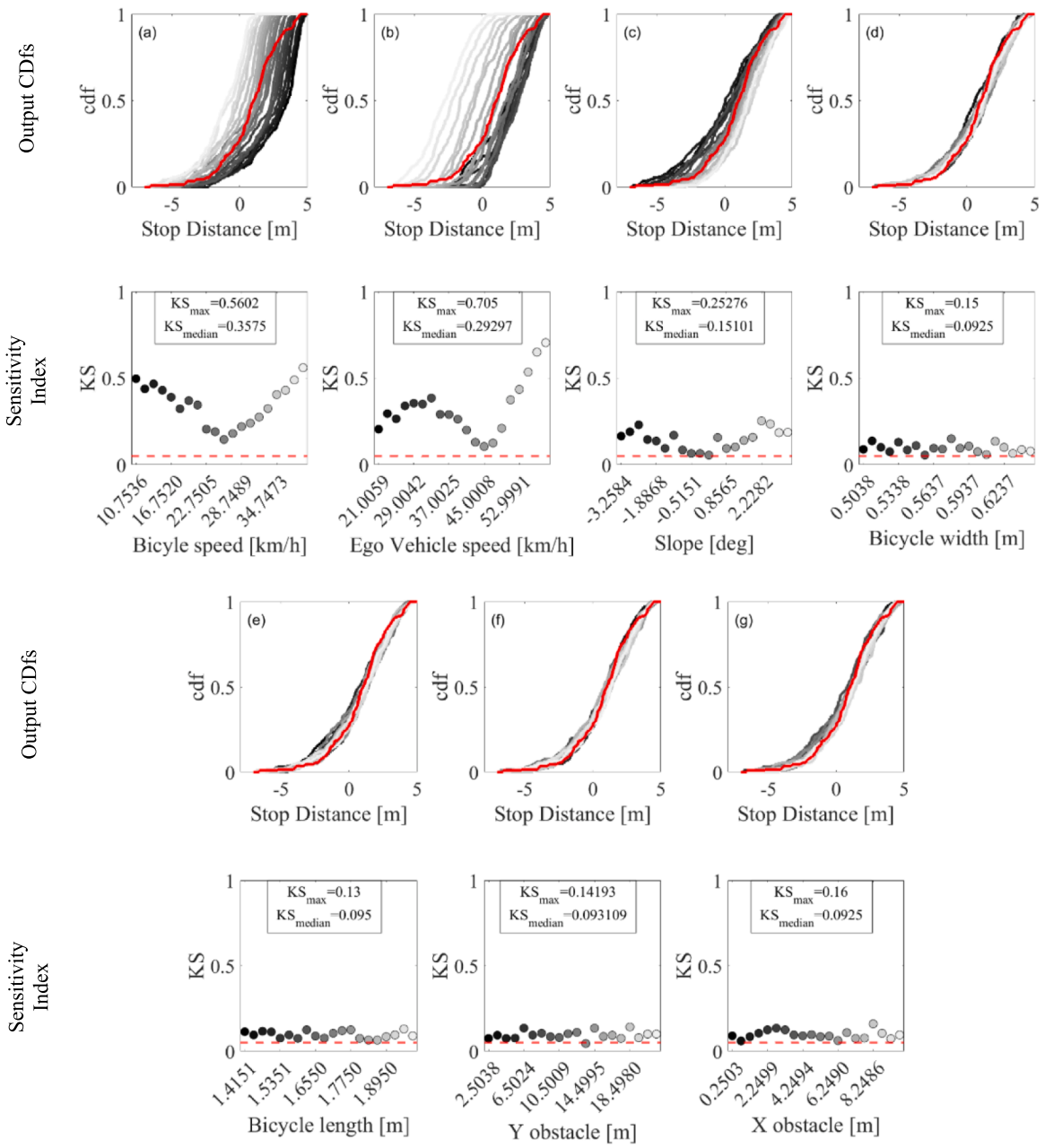


**Fig. 6.** Sensitivity behavior of the output y in response to each input factor $x_p$ (p = 1, …,7): (a) bicycle speed, (b) Ego Vehicle speed, (c) slope, (d) bicycle length, (e) later position of the obstacle, (f) bicycle width, (g) longitudinal distance of the obstacle. Output CDFs panels (first and third rows): empirical unconditional output distribution $F_y(y)$ (red lines) and conditional ones $F_{y|x}$ ($y|x_p \in \Im_k$) for each input factor $x_p$ (p = 1, …,7). Sensitivity Index panels (second and fourth): Kolmogorov-Smirnov (KS) statistics at different conditioning levels of each input factor $x_p$ (p = 1, …,7); the dashed red lines represent the threshold of significant sensitivity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the variability of the index over the conditioning point range (i.e., over the range of variability of the input factor).

Furthermore, we conduct a convergence analysis of the sensitivity indexes to verify that the sample size is adequate to capture the effects of all the input factors (Markkula et al., 2018). The purpose of this analysis is to determine the minimum number of runs required to obtain a stable ranking of the input parameters. We compute the median values of the sensitivity indexes for sample sizes ranging from 500 to 5000, incrementing the size by 500 at each iteration. As shown in Fig. 8, the sensitivity index values decrease as the sample size increases until they stabilize at around 3500 model runs. Furthermore, the separation between influential and non-influential input factors becomes more pronounced with increasing sample size. It is noteworthy that, for sample sets with<2000

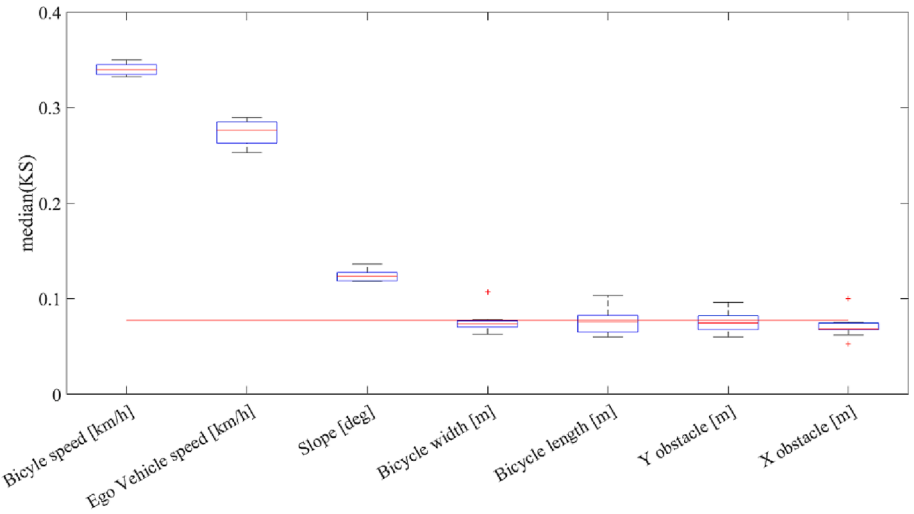


**Fig. 7.** PAWN sensitivity indexes for each input factor $x_p$ (p = 1, …,7). Bootstrapping estimates the distribution of each PAWN index. Boxplots are employed to display variation in the sample: the bottom and top edges of a box indicate the 25th and 75th percentiles, while the red line inside the box the median value. The horizontal red line represents the KS of the dummy parameter. Input parameters are sorted according to their PAWN index values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model runs, the sensitivity indexes are overestimated, a phenomenon that has been observed in other studies in different application domains (Chaturvedi and Rajasekar, 2022; Sarrazin et al., 2016; Yang et al., 2019). Nevertheless, results in Fig. 8 clearly demonstrate that the input factors with the greatest influence (bicycle speed and Ego Vehicle speed) consistently exhibit higher and stable sensitivity index values, even for smaller sample sets of model evaluations. On the other hand, input factors with negligible effects do not maintain a stable ranking order, as their sensitivity indexes decrease with an increasing dimension of the sample set. Their sensitivity indexes decrease with an increasing dimension of the sample set. The road slope exhibits a distinct behavior, as for a minimal sample set dimension (below 500), its sensitivity index is comparable to those of other non-influential input factors. However, it becomes an influential factor when the sample size increases to over 500 runs and up to 1500 runs. As the number of runs increases, the divergence between slope and non-influential input factors becomes more evident, as the slope increasingly influences the output, while the other factors become increasingly non-influential. Overall, the ranking of input factors at $N = 1500$ is consistent with those at $N = 4000$ and $N = 5000$ (regarding influential factors). We argue that, at least in our case, $N = 4000$ is a reliable sample size. The ranking of

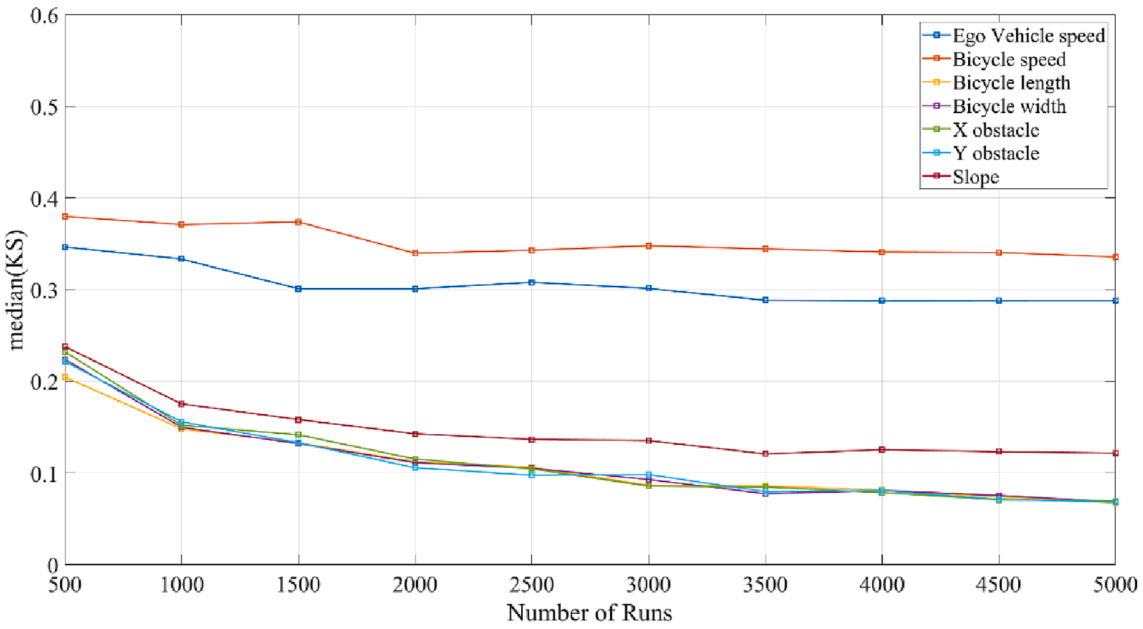


**Fig. 8.** Convergence assessment of PAWN sensitivity indexes for stop distance estimating by increasing model evaluations.

influential parameters obtained after bootstrapping and convergence analysis is consistent with the one in Fig. 6, thus confirming the robustness of the PAWN results.

The robustness of indices could be theoretically checked by varying the number of conditioning intervals $n$. However, Pianosi and Wagener (2018) and Wang et al. (2020) have shown that the choice of $n$ has a limited impact on the estimation of sensitivity indexes if it assumes a sufficiently high value (equal to or higher than 10). In particular, as $n$ increases, only a slight increase in the indices of less-influential factors can be observed. This trend mainly concerns input parameters with low sensitivity indices and does not affect the overall input ranking. Thus, the choice of $n$ does not impact the key conclusion that these parameters are non-influential, as their indexes remain below the one of the dummy parameter. Given the above discussion, the analysis of the effect of the tuning parameter $n$ is not considered, as a value of $n$ higher than the suggested one has already been chosen, and further insights would not be gained.

#### 4.2.2. Regional response probabilistic sensitivity analysis

As previously discussed in Section 2.1.1, the PAWN sensitivity indexes can be applied to a particular sub-range of the output distribution rather than the entire range. This allows for the identification of regions within the input space that lead to specific output values. In this study, we focus on regions within the input space that result in a negative output value ($y < 0$), which corresponds to a system failure.

Some of the findings presented in Section 4.2 suggest that a localized analysis would be beneficial. Specifically, the results in Fig. 6 indicate that using the max *KS* leads to an increased importance of certain input factors. Notably, the position of the two most influential factors is swapped, with the Ego Vehicle speed becoming the most significant input factor, followed by the bicycle speed. More in detail, Fig. 6b demonstrates that for Ego Vehicle speeds exceeding 50 [km/h], the corresponding sensitivity indexes are very high, and the associated conditional CDFs are significantly shifted to the left of the unconditional CDF. This suggests that a sub-range of these input parameters has a significant negative impact on the output and, consequently, on the system's performance. In contrast, while the bicycle speed (Fig. 6a) has a greater influence on the output variation overall, there is no sub-range of bicycle speed that has as significant an impact on the output as the Ego Vehicle speed does. Lastly, the ranking of the road slope remains unchanged. Based on the previous discussion, we conduct a detailed analysis of the localized impact of input factors on the output to achieve the following objectives: (*i*) gain insights into the input–output mapping; (*ii*) reinforce the visual inspection of scatter plots with quantitative evidence of local effects. To accomplish this, we compute the PAWN sensitivity indexes for a specific subset of the output. Specifically, we examine failed cases by applying Eq. (4) with $y < 0$.

The results in Fig. 9 validate the findings in Fig. 6a-b. confirming that bicycle speed, Ego Vehicle speed, and road slope are influential factors, and revealing a ranking swap between bicycle speed and Ego Vehicle speed when analyzing only failed cases. The boxes in the distributions show that the *median KS* of Ego Vehicle speed is the most dispersed, indicating greater sensitivity to estimation errors and index variability in each conditioning interval. Similar trends are observed for bicycle speed, while the distribution

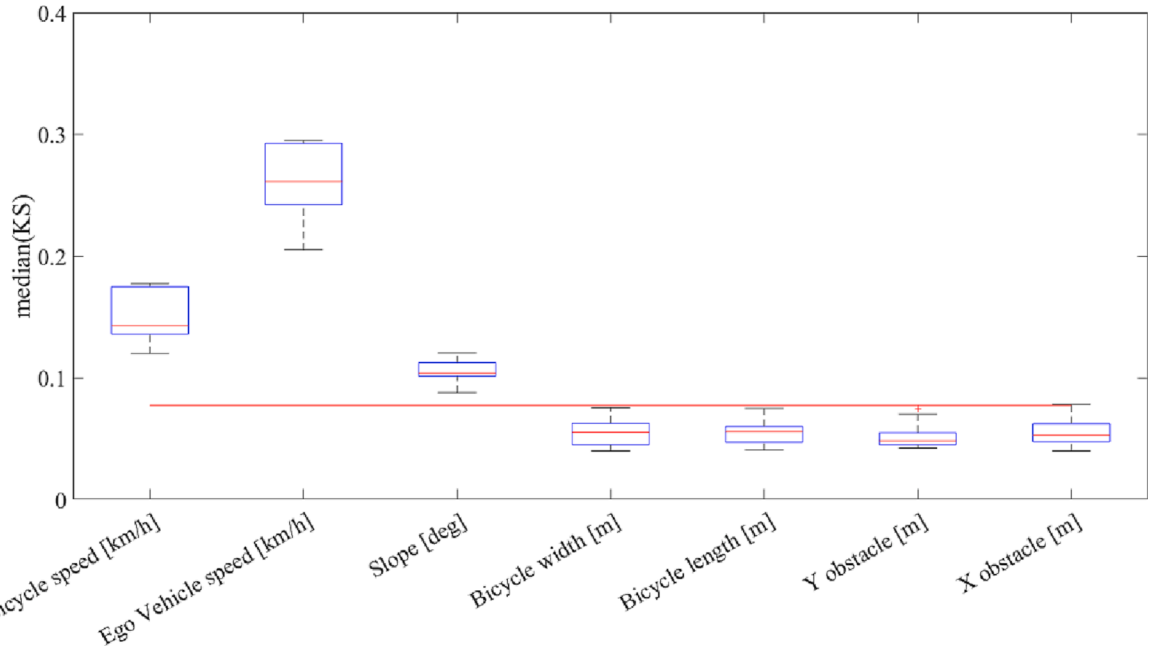


**Fig. 9.** PAWN sensitivity indexes for each input factor $x_p$ ($p = 1, \ldots, 7$) over the sub-range $y < 0$. Bootstrapping estimates the distribution of each PAWN index. Boxplots are employed to display variation in the sample: the bottom and top edges of a box indicate the 25th and 75th percentiles, while the red line inside the box the median value. The horizontal red line represents the KS of the dummy parameter. Input parameters are sorted according to their PAWN index value as in Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for road slope is less dispersed. The distributions of the remaining factors are below the red threshold line, indicating that they are uninfluential, and their ranking remains unchanged. These findings provide further quantitative evidence of the localized effects of input factors on the output, and reinforce the insights gained from visual inspection in Fig. 6.

The presented results underscore two crucial aspects of the proposed assessment methodology. Firstly, the median values of sensitivity indices computed over the entire range of input factors should be considered as a summary statistic, as they can be influenced by specific sub-ranges of input factors. Hence, localized evaluations are necessary to identify extreme behaviors. Secondly, the range uncertainty of input factors has a significant impact on the related sensitivity indices, as highlighted by Wang et al. (2020). The broader the range of variation, the more prone the input factor is to uncertainty. Therefore, identifying the feasible (or plausible) range of uncertainty for each input factor in practical applications is a crucial step.

## 5. Comparison with EuroNCAP results

In this section, we compare the results obtained from the proposed testing approach with those achieved using the EuroNCAP approach. It is important to note that the testing conditions employed for the EuroNCAP approach are the same as those described in Section 3.1.

The framework we propose explores different combinations of factors, and the results indicate that test failure occurs in approximately 30% of cases (1213 fails on 4000 simulation runs), as appreciable from the cumulative distribution in Fig. 10. In contrast, simulations performed using the EuroNCAP reference values (represented by red dots in Fig. 10) result in test failure in only about 18% of cases (10 fails on 54 simulation runs). The findings suggest that EuroNCAP does not sufficiently investigate the SUT's functioning since its coverage of the explored cases is inadequate in dealing with input factor variability. Notably, when a failure occurs, the stop distance never assumes values lower than $-2.71$ [m] in the EuroNCAP simulations, while values up to $-6.84$ [m] are obtained using our proposed testing framework. Thus, the EuroNCAP does not allow to correctly estimate the extent of the failed cases.

To conduct a more detailed analysis of the differences between the two testing approaches, below we report the failed cases in the input factor plan. For the sake of clarity, only the parameters that play a significant role in causing a failure are reported, as depicted in Fig. 11. This allows us to emphasize the failure conditions (in terms of input parameter combinations) that are captured by the EuroNCAP approach and those that are not. The results in Fig. 11 clearly highlight that the EuroNCAP testing approach fails to identify input factor combinations that lead to critical scenarios. For instance, in Fig. 11a such test can spot failures due only to high values of the Ego Vehicle speed (the bicycle speed if kept constant at 15 [km/h]), while the proposed testing approach demonstrates that the test failure can be due to different combinations of the two considered input factors. Similar considerations also apply to the results in Fig. 11b and c.

In conclusion, the proper functioning of a given AEB system could result in being non-tested for some conditions in which it fails, giving rise to an incorrect evaluation of the SUT.

## 6. Conclusions

In this paper, we proposed a methodology that utilizes stochastic simulation and uncertainty modeling techniques to identify unknown unsafe operation regions for ADSs in accordance with ISO 21448. We employed a quasi-random sampling approach to exhaustively explore the range of possible situations that a vehicle may encounter, and the PAWN density-based sensitivity method to

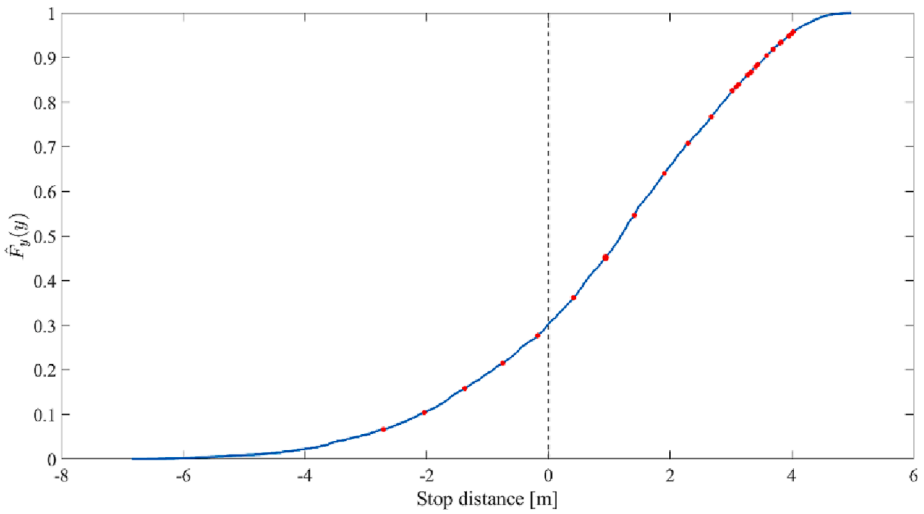


**Fig. 10.** Cumulative Distribution Function of the simulation outputs, marking with circles the outputs of the exploration done with the EuroNCAP standards.

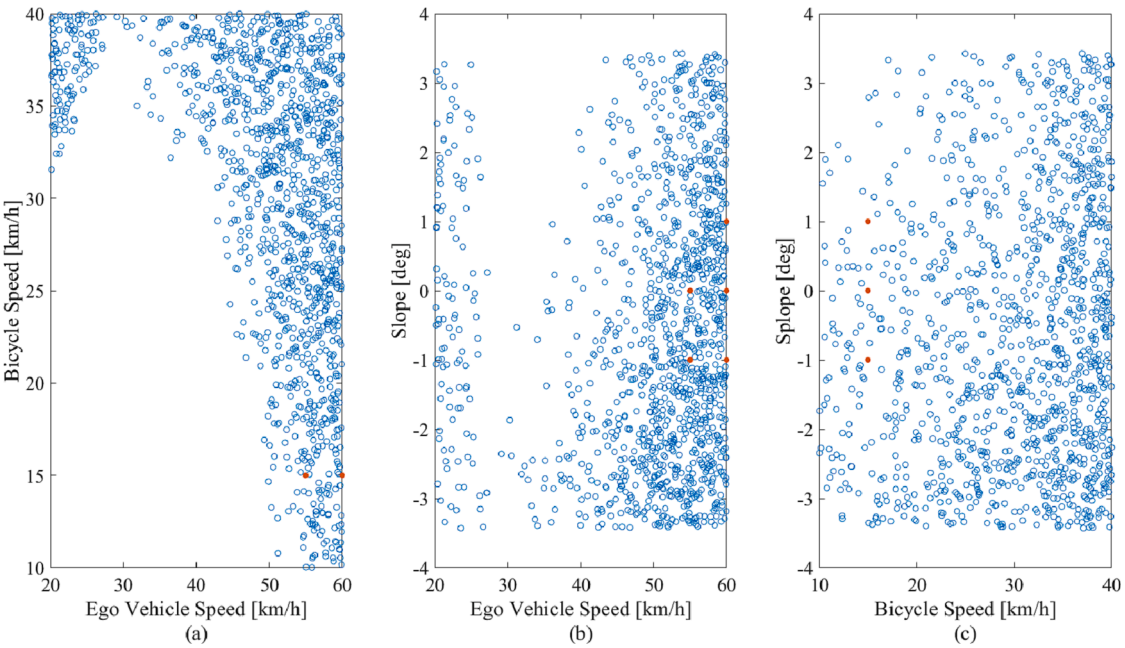


**Fig. 11.** Scatter plots of failed cases in the input factors plane with the proposed approach (blue circle) and with the EuroNCAP one (red dots): (a) Ego Vehicle speed-bicycle speed plane; (b) Ego Vehicle speed – slope plane; (c) Bicycle speed-slope plane. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

analyze the performance of an exemplary automated driving assistance system. In so doing, we identified the situational variables that most strongly influence the pass-fail rate variability and the corresponding triggering conditions. The PAWN sensitivity method was chosen because it is independent from the shape of the output distribution, ensuring unbiased results even for multi-modal or highly skewed distributions, it is easy applicable to nonlinear models and provides robust results also for a relatively low sample size.

The proposed methodology was applied to the well-known use case Euro NCAP AEB Car-to-Bicyclist Nearside Adult 50% in order to prove its robustness. Specifically, the CBNA-50 scenario used to test an AEB system was adapted to the proposed framework by eliminating some of the deterministic assumptions about the testing conditions. The study investigated the influence of vehicle speed, bicycle speed and size, obstacle position, and road slope on the pass-fail rate of the trial. These factors were assumed to be independently and uniformly distributed within feasible physical ranges that reflect a broad range of real-world scenarios. The stop distance was used as an output variable to quantify the pass-fail rate of the trial. The EuroNCAP approach was also used as a baseline for comparison in its original form. The simulation analysis was conducted using the MATLAB/Simulink environment, which allows for the emulation of vehicle motion, virtual road environments, and the AEB functioning. A total of 4000 simulation runs were performed.

The preliminary step involved the examination of the empirical distribution of the output variable, which turned out to be negatively skewed with a left tail. This outcome strengthened the rationale behind adopting the PAWN approach. Then, PAWN sensitivity indexes were computed leveraging two statistics for the Kolmogorov-Smirnov test, namely the median (*median KS*, as a summary statistic for average behaviors) and the maximum (max *KS*, to spot extreme behaviors). The results indicated that the most influential factors for the AEB system were the bicycle's speed and the Ego Vehicle's speed. The road slope was found to have a low but still significant importance, while other factors such as the obstacle's position did not significantly affect the system's performance. The robustness of the obtained indexes was verified through bootstrapping technique, and a convergence analysis demonstrated that the selected sample size of runs enabled the attainment of a stable ranking of input parameters. Results highlight that, when using the max *KS* instead of *the median KS*, a swap in the ranking position between the bicycle speed and the Ego Vehicle one was observed. This suggested that a subset of the input parameters strongly and negatively affected the output and, consequently, the system's performance. To investigate this localized effect, PAWN indexes were recomputed over the sub-range of the output $y < 0$ (failed cases).

Considering this sub-range, the Ego Vehicle speed was found to be the most influential input factor, while the bicycle speed ranked second. The road slope still had a low but significant influence, while the other factors remained insignificant. Two important implications arose from this finding. Firstly, median sensitivity index values computed over the whole range of the output should be considered as a summary statistic, necessitating further evaluations of localized effects to identify possible extreme behaviors. Secondly, the range uncertainty of input factors significantly affects the related sensitivity indexes. The larger the variation range, the higher the input factor's exposure to uncertainty. Therefore, detailed exploration of the plausible uncertainty range of each input factor in practical applications is crucial. Lastly, the traditional testing approach, exemplified by the EuroNCAP test, produced very different results from those obtained by extensively exploring the factor space. The EuroNCAP test only analyzed a small subset of the input space based on predetermined assumptions about the phenomenon. Despite the relatively simple automated system and tested use case, the uncertainty was sufficient to invalidate the traditional methods.

The results of the analysis confirmed that the proposed methodology satisfies the coverage requirements specified in the ISO 21448, enabling the reliable verification of ADSs. It facilitates the identification of safe/unsafe operating regions and trigger conditions, which, in turn, improve the development of automated driving functions and enhance ADS safety. Therefore, the methodology presented in this study is deemed appropriate for identifying the factors that influence ADS performance efficiently. Notably, tests conducted with an appropriate uncertainty analysis tools produce much more robust results than those based solely on statistical databases, particularly when human driver system databases are used to investigate the worst-case scenarios for an ADS. As automation systems' complexity is anticipated to increase, the use of stochastic approaches, as proposed here, is strongly recommended to obtain reliable results.

The methods employed in this experiment are subject to certain limitations stemming from their underlying assumptions. The PAWN method assumes that factors are continuous and uncorrelated, while in real-world problems input factors can be correlated and their joint distribution can take various forms. Ignoring correlation effects and multivariate distributional features can significantly bias or even falsify the sensitivity analysis results (Razavi et al., 2021). Another limitation concerns the choice of size of simulation runs, which requires a trade-off between the need for a small size and reliable results. This trade-off is strongly influenced by the number of considered input factors, their possible correlations, and the model complexity, making the selection of this simulation parameter non-trivial. Moreover, even with a large sample size, some critical scenarios may be missing.

Considering the above discussion, future development of this study could include assessing more sophisticated automated driving systems; exploiting sensitivity methods accounting for correlation effects and multivariate distributions; and more accurate sampling methods, such as the Modified Latin Hypercube Sampling.

## CRediT authorship contribution statement

**Angelo Coppola:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Claudio D'Aniello:** Conceptualization, Investigation. **Luigi Pariota:** Conceptualization, Methodology, Validation, Supervision, Writing – original draft, Writing – review & editing. **Gennaro Nicola Bifulco:** Conceptualization, Methodology, Formal analysis, Validation, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was supported by the Italian Ministry of Research (Ministero dell'Università e della Ricerca - MUR) within the framework of the project DIGIT-CCAM (grant E67G22000010005).

## References

Alghodhaifi, H., Lakshmanan, S., 2021. Autonomous vehicle evaluation: a comprehensive survey on modeling and simulation approaches. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3125620.

Amersbach, C., Winner, H., 2019. Defining required and feasible test coverage for scenario-based validation of highly automated vehicles. In: 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019. https://doi.org/10.1109/ITSC.2019.8917534.

APROSYS Dissemination and Use Plan M1-M60 Deliverable D8.2.10, n.d.

Au, S.K., Beck, J.L., 2003. Important sampling in high dimensions. Struct. Saf. 25 https://doi.org/10.1016/S0167-4730(02)00047-4.

Batsch, F., Kanarachos, S., Cheah, M., Ponticelli, R., Blundell, M., 2021. A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles. J. Intell. Transp. Syst. Technol. Plann. Oper. https://doi.org/10.1080/15472450.2020.1738231.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., Winner, H., 2014. Three decades of driver assistance systems: Review and future perspectives. IEEE Intell. Transp. Syst. Mag. https://doi.org/10.1109/MITS.2014.2336271.

Borgonovo, E., Castaings, W., Tarantola, S., 2011. Moment Independent Importance Measures: New Results and Analytical Test Cases. Risk Anal. 31 https://doi.org/10.1111/j.1539-6924.2010.01519.x.

Chaturvedi, S., Rajasekar, E., 2022. Application of a probabilistic LHS-PAWN approach to assess building cooling energy demand uncertainties Article History. Building Simulation. Springer, in, pp. 373–387.

Coppola, A., Costanzo, L.D., Pariota, L., Santini, S., Bifulco, G.N., 2022. An Integrated Simulation Environment to test the effectiveness of GLOSA services under different working conditions. Transp Res Part C Emerg Technol 134. https://doi.org/10.1016/j.trc.2021.103455.

Coppola, A., Lui, D.G., Petrillo, A., Santini, S., 2022. Eco-Driving Control Architecture for Platoons of Uncertain Heterogeneous Nonlinear Connected Autonomous Electric Vehicles. IEEE Trans. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2022.3200284.

Dantsuji, T., Hoang, N.H., Zheng, N., Vu, H.L., 2022. A novel metamodel-based framework for large-scale dynamic origin–destination demand calibration. Transp Res Part C Emerg Technol 136, 103545. https://doi.org/10.1016/J.TRC.2021.103545.

de Nicolao, G., Ferrara, A., Giacomini, L., 2007. Onboard sensor-based collision risk assessment to improve pedestrians' safety. IEEE Trans Veh Technol 56. https://doi.org/10.1109/TVT.2007.899209.

Efron, B., Tibshirani, R., 1994. An introduction to the bootstrap.

Eidehall, A., Petersson, L., 2008. Statistical threat assessment for general road scenes using Monte Carlo sampling, in: IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2007.909241.

Euro NCAP, 2021. EUROPEAN NEW CAR ASSESSMENT PROGRAMME (Euro NCAP) TEST PROTOCOL-AEB Car-to-Car systems EUROPEAN NEW CAR ASSESSMENT PROGRAMME (Euro NCAP).

Euro NCAP Test Protocol-AEB/LSS VRU systems, 2021.

Favarò, F.M., Nader, N., Eurich, S.O., Tripp, M., Varadaraju, N., 2017. Examining accident reports involving autonomous vehicles in California. PLoS One 12. https://doi.org/10.1371/journal.pone.0184952.

Gao, F., Duan, J., Han, Z., He, Y., 2020. Automatic Virtual Test Technology for Intelligent Driving Systems Considering Both Coverage and Efficiency. IEEE Trans Veh Technol 69. https://doi.org/10.1109/TVT.2020.3033565.
Glynn, P.W., Iglehart, D.L., 1989. Importance Sampling for Stochastic Simulations. Manage Sci 35. https://doi.org/10.1287/mnsc.35.11.1367.
Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 81. https://doi.org/10.1016/S0951-8320(03)00058-9.
Highway Traffic Safety Administration, N., 2008. National Motor Vehicle Crash Causation Survey: Report to Congress.
International Organization for Standardization, 2011. ISO26262 - Road vehicles – Functional safety 2005.
ISO, 2022. ISO 21448:2022(en), Road vehicles — Safety of the intended functionality [WWW Document]. accessed 2.24.23. https://www.iso.org/obp/ui/#iso:std:iso:21448:ed-1:v1:en.
Kalra, N., Paddock, S.M., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? Transp Res Part A Policy Pract 94. https://doi.org/10.1016/j.tra.2016.09.010.
Koo, H., Iwanaga, T., Croke, B.F.W., Jakeman, A.J., Yang, J., Wang, H.H., Sun, X., Lü, G., Li, X., Yue, T., Yuan, W., Liu, X., Chen, M., 2020. Position paper: Sensitivity analysis of spatially distributed environmental models- a pragmatic framework for the exploration of uncertainty sources. Environ. Model. Softw. 134, 104857 https://doi.org/10.1016/J.ENVSOFT.2020.104857.
Kuhn, D.R., Kacker, R.N., Lei, Y., Locke, G., Gallagher, P., 2010. Practical Combinatorial Testing. National Institute of Standards and Technology 800 (142).
Lemmen, P., Fagerlind, H., Unselt, T., Rodarius, C., Infantes, E., van der Zweep, C., 2012. Assessment of Integrated Vehicle Safety Systems for Improved Vehicle Safety. Procedia Soc Behav Sci 48, 1632–1641. https://doi.org/10.1016/J.SBSPRO.2012.06.1138.
Lengyel, H., Tettamanti, T., Szalay, Z., 2020. Conflicts of automated driving with conventional traffic infrastructure. IEEE Access 8. https://doi.org/10.1109/ACCESS.2020.3020653.
Liu, K., Gao, H., Liang, Z., Zhao, M., Li, C., 2021. Optimal charging strategy for large-scale electric buses considering resource constraints. Transp Res D Transp Environ 99, 103009. https://doi.org/10.1016/J.TRD.2021.103009.
Markkula, G., Romano, R., Jamson, A.H., Pariota, L., Bean, A., Boer, E.R., 2018. Using driver control models to understand and evaluate behavioral validity of driving simulators. IEEE Trans Hum Mach Syst 48. https://doi.org/10.1109/THMS.2018.2848998.
Menzel, T., Bagschik, G., Maurer, A.M., 2018. Scenarios for Development. Test and Validation of Automated Vehicles, in: IEEE Intelligent Vehicles Symposium, Proceedings. https://doi.org/10.1109/IVS.2018.8500406.
Najm, W.G., Smith, J.D., (U.S.), J.A.V.N.T.S.C., 2007. Development of crash imminent test scenarios for Integrated Vehicle-Based Safety Systems. https://doi.org/10.21949/1503647.
Noacco, V., Sarrazin, F., Pianosi, F., Wagener, T., 2019. Matlab/R workflows to assess critical choices in Global Sensitivity Analysis using the SAFE toolbox. MethodsX 6, 2258–2280. https://doi.org/10.1016/J.MEX.2019.09.033.
Parliament, E., 2019. European Regulation 2019/2144. Off. J. Eur. Union.
Pianosi, F., Sarrazin, F., Wagener, T., 2020. How successfully is open-source research software adopted? Results and implications of surveying the users of a sensitivity analysis toolbox. Environ. Model. Softw. 124 https://doi.org/10.1016/j.envsoft.2019.104579.
Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model. Softw. 67, 1–11. https://doi.org/10.1016/J.ENVSOFT.2015.01.004.
Pianosi, F., Wagener, T., 2018. Distribution-based sensitivity analysis from a generic input-output sample. Environ. Model. Softw. 108 https://doi.org/10.1016/j.envsoft.2018.07.019.
Punzo, V., Zheng, Z., Montanino, M., 2021. About calibration of car-following dynamics of automated and human-driven vehicles: Methodology, guidelines and codes. Transp Res Part C Emerg Technol 128, 103165. https://doi.org/10.1016/J.TRC.2021.103165.
Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J.H.A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H.R., 2021. The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. Environ. Model. Softw. 137 https://doi.org/10.1016/J.ENVSOFT.2020.104954.
Rocklage, E., Kraft, H., Karatas, A., Seewig, J., 2018. Automated scenario generation for regression testing of autonomous vehicles, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. https://doi.org/10.1109/ITSC.2017.8317919.
Roy, C.J., Oberkampf, W.L., 2011. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Comput Methods Appl Mech Eng 200, 2131–2144. https://doi.org/10.1016/J.CMA.2011.03.016.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., 2008. Global sensitivity analysis: the primer.
Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global Sensitivity Analysis of environmental models: Convergence and validation. Environ. Model. Softw. 79, 135–152. https://doi.org/10.1016/J.ENVSOFT.2016.02.005.
Singh, S., 2015. Critical reasons for crashes investigated in the national motor vehicle crash causation survey.
Sun, J., Zhang, H., Zhou, H., Yu, R., Tian, Y., 2022. Scenario-Based Test Automation for Highly Automated Vehicles: A Review and Paving the Way for Systematic Safety Assurance. IEEE Trans. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2021.3136353.
Varga, B., Doba, D., Tettamanti, T., 2023. Optimizing vehicle dynamics co-simulation performance by introducing mesoscopic traffic simulation. Simul Model Pract Theory 125. https://doi.org/10.1016/J.SIMPAT.2023.102739.
Wadud, Z., MacKenzie, D., Leiby, P., 2016. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. Transp Res Part A Policy Pract 86. https://doi.org/10.1016/j.tra.2015.12.001.
Wang, Y., Huang, J., Tang, H., 2020. Global sensitivity analysis of the hydraulic parameters of the reservoir colluvial landslides in the Three Gorges Reservoir area, China. Landslides 17, 483–494. https://doi.org/10.1007/S10346-019-01290-9/FIGURES/12.
Wang, C., Xie, Y., Huang, H., Liu, P., 2021. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. Accid Anal Prev 157, 106157. https://doi.org/10.1016/J.AAP.2021.106157.
Xu, Y., Zou, Y., Sun, J., 2018. Accelerated testing for automated vehicles safety evaluation in cut-in scenarios based on importance sampling, genetic algorithm and simulation applications. Journal of Intelligent and Connected Vehicles 1. https://doi.org/10.1108/jicv-01-2018-0002.
Yang, H.H., Peng, H., 2010. Development and evaluation of collision warning/collision avoidance algorithms using an errable driver model, in: Vehicle System Dynamics. https://doi.org/10.1080/00423114.2010.515745.
Yang, J., Ma, Y., Fu, J., Shu, J., Liu, J., 2019. Parametric study of gasoline properties on combustion characteristics of gasoline compression engines using reaction kinetics simulation and density-based global sensitivity analysis. Appl Energy 255, 113858. https://doi.org/10.1016/J.APENERGY.2019.113858.
Yang, L., Yang, Y., Wu, G., Zhao, X., Fang, S., Liao, X., Wang, R., Zhang, M., 2022. A systematic review of autonomous emergency braking system: impact factor, technology, and performance evaluation. J. Adv. Transp. 2022 https://doi.org/10.1155/2022/1188089.
Zhang, X., Tao, J., Tan, K., Törngren, M., Sánchez, J.M.G., Ramli, M.R., Tao, X., Gyllenhammar, M., Wotawa, F., Mohan, N., Nica, M., Felbinger, H., 2021. Finding critical scenarios for automated driving systems: a systematic literature review. https://doi.org/10.48550/arxiv.2110.08664.
Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D.J., Nobukawa, K., Pan, C.S., 2017. Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. IEEE Trans. Intell. Transp. Syst. 18 https://doi.org/10.1109/TITS.2016.2582208.