Temporal and Spatial Fault Detection for Connected Cyber-Physical Systems

Hugo Araujo[1], Mohammad Reza Mousavi[1], and Shiva Nejati[2]

[1] King's College London, United Kingdom
first.last@kcl.ac.uk
[2] University of Ottawa, Canada
snejati@uottawa.ca

Abstract. Testing connected cyber-physical systems (CPS) is a complex task. Connected CPS feature complex stochastic dynamic behaviour in interaction with the physical and human environment as well as communication over networks. Devising an oracle for testing connected CPS is a challenge; the oracle should be able to quantitatively reason about the stochastic nature of the interactions between the CPS and its environment. The quantitative reasoning should be sensitive to significant deviations in the dynamics and neglect minor deviations, e.g., due to measurement errors. To address this challenge, we provide the mathematical framework for conformance testing of connected CPS. We define a quantitative measure of closeness for two distributions of trajectories (i.e., output distributions from two distinct stochastic systems that are provided with the same input stimuli) that allows for capturing significant temporal and spatial deviations and neglecting subtle ones. This measure forms the basis for our notion of stochastic conformance, which determines when two stochastic systems conform to each other. We implement our proposed notion of stochastic conformance and compare our notion against a state-of-the-art baseline by applying both approaches to a case study involving a platoon of connected vehicles. Our notion detects a variety of different types of faults whilst allowing subtle deviations resulting from naturally occurring perturbations inherent to CPS.

Keywords: Cyber-physical systems · Conformance testing · Stochastic

1 Introduction

Connected Cyber-Physical Systems (CPS) represent an integration of computation, networking, and physical processes, where embedded computers and networks monitor and control physical processes [9]. As these systems are highly prevalent in critical domains such as healthcare, automotive, and aerospace, ensuring their correctness and reliability is paramount [24]. However, verifying CPS is challenging and many techniques have been developed to ensure their correctness [38,6]. One such approach is *conformance testing*, which verifies whether a system complies with its specification by comparing their outputs through a well-defined mathematical relation (i.e., a conformance notion) [27]. Conformance

testing is highly relevant to CPS, as comparing the system against its specification supports automated test oracles. Nonetheless, applying conformance testing to CPS presents several challenges.

One of the key challenges in conformance testing for CPS is to define a quantitative measure that detects significant deviations, yet disregards subtle perturbations in the system behaviour caused by naturally occurring physical phenomena and measurement errors. For instance, sensor noise, mechanical backlash, and communication delays can affect the behaviour of such systems in a negligible way. If not accounted for, this can lead to false negatives, i.e., failing a test case, even when the system is within the expected boundaries [2]. Thus, conformance testing techniques should provide adjustable temporal and spatial bounds (to be defined by domain experts) to allow for detecting significant deviations between a system's output and its specification while neglecting the minor ones. The second key challenge arises from the stochastic nature of connected CPS and their environments, necessitating conformance testing that accounts for the probabilistic distributions of outcomes. The need for such a notion has been identified in the literature: it has been demonstrated that test results for CPS are often stochastic, leading to variability in outputs when the same test is re-executed multiple times [18].

In the literature, existing conformance notions can accommodate for (i) temporal error margins and (ii) spatial error margins [1], or (iii) stochasticity [32], exclusively. However, to our knowledge, there is no conformance testing approach that covers these three aspects simultaneously. The aim of this work is to propose a conformance notion that responds to the identified need and addresses all three aspects. To this end, we define a quantitative measure that compares output distributions and checks whether their distance is within user-defined margins. This lets us uncover deviating stochastic behaviour that indicates a failure whilst allowing for subtle, naturally occurring temporal and spatial deviations. We implement our conformance testing notion and compare it against the state-of-the-art it using a case study of a connected platoon. Our results show that our conformance notion can detect a higher number of inserted faults and common faulty signal patterns (identified by a taxonomy on signal-based properties of CPS [14]), given the same test suite, compared to the alternative.

In summary, the main contributions of this work are as follows. We first introduce a novel conformance notion that allows for (i) reasoning about the stochastic nature of connected CPS (by considering distributions of outputs) and for (ii) quantitative temporal and spatial error margins in the outputs (which are needed to fail major deviations while reducing the number of false negatives). Then, we implement our conformance notion into a publicly-available tool. Lastly, we present the results of an empirical evaluation that compares our notion against a baseline [32]; we make the assets and data resulting from the study publicly available at `https://zenodo.org/records/14906880`.

We assess the performance of our conformance testing approach using the true positive (a *correct* fail verdict) and false positive (an *incorrect* fail verdict) metrics. We devise the following research questions based on these two metrics.

- **RQ1.** Is our conformance testing approach effective in detecting substantial discrepancies between the outputs of two CPS and, hence, yielding true positive verdicts?
- **RQ2.** Is our conformance testing approach adaptable to allow negligible discrepancies between the outputs of two CPS and, hence, avoiding false positive verdicts?

The notion of false positives and negatives is related to the parameters of the conformance notion. That is, any deviating behaviour within the allowed margins should not result in a fail verdict. The opposite must also be true: any deviating behaviour beyond the allowed margins should result in a fail verdict.

2 Related Work

Classical methods for automatic verification of CPS, such as reachability analysis [8] and, more generally, model checking [16] have been extensively studied and applied [22,15,33,3,10]; such methods typically rely on exhaustively exploring the state-space of the system (or its model). However, these methods are prone to the state-space explosion problem and cannot be used for large-scale systems. Classical conformance testing approaches [36,1,19] address this problem by non-exhaustively falsifying a specification through comparing the specification against the system under test. This is typically done using distance metrics such as Euclidean [1] and Skorokhod [19] to compute the degree of dissimilarity between the observed and expected output. However, they fail to account for stochastic behaviour, limiting their applicability to real-world CPS. To address the limitations of traditional verification approaches, stochastic approaches have been proposed. They range from exhaustive model-checking approaches [28] to non-exhaustive formal verification [17,25] and conformance testing [29,32] approaches. In the remainder of this section, we review the most significant contributions in this area and compare them to ours.

Clarke et al. [17] developed a strategy for statistical model checking of CPS by combining the Monte Carlo method with temporal logic model checking. They sample simulations of the system model and check their conformance with respect to a temporal formula by applying a statistical estimation technique to compute the probability that the formula is satisfied. Unlike our work, they focus on verifying compliance with respect to properties whereas our methodology works by verifying that the distance between (expected and observed) output distributions is smaller than a pre-defined value. More closely related to our work, Qin et al. [32] propose a notion of conformance for stochastic system that checks whether the probability of the distance between outputs is less than the failure probability. Unlike their work, we consider not only spatial but also temporal distance between outputs. This allows us to cater for delays that naturally occur in CPS. Furthermore, our notion of distance considers point-by-point and not the overall distance between the entire output signal; hence, we propose a more thorough formalism that can catch short bursts of discrepancy that violate

conformance. Similarly, Leemans et al. [29] use the Wasserstein distance to quantify the distance between two stochastic Petri systems. Their notion of distance is based on the "earth movers' distance" and measures the effort to transform the distributions of traces in one model into the distribution of traces in another. Their approach substantially differs from ours, as it has only been considered for discrete systems (using traces based on event logs). The effect of using different distance measures (such as total variation distance [35] or Wasserstein [23]) on the effectiveness of conformance testing can be further investigated.

3 Preliminaries

In this section, we provide the preliminary concepts used to define conformance. We start with a running example of a CPS (Section 3.1), and, in Section 3.2, we provide the mathematical background to describe stochastic systems. Lastly, in Section 3.3, we recall a basic notion of conformance [2] (for non-stochastic systems) that we extend to deal with the stochastic nature of CPS.

3.1 Running example

Consider a system where a convoy of cars autonomously follow a leading human-driven car. The leading car sends its acceleration, velocity, and location via wireless communication channels to the followers. It is critical that the follower cars are up-to-date with the information received from the connected cars; therefore, following the literature, we employ the concept of data age as a safety metric for the platoon [13]. Due to network congestion, the transmission of a packet has a probability distribution. In the remainder, we use this running example to explain the basic concepts and further elaborate on it as our case study.

3.2 Probability theory

To formally model stochastic systems, we start by defining probability spaces.

Definition 1 (Probability Space). *A probability space is a triple, denoted by $(\Omega, \mathcal{F}, P)$, comprising the sample space Ω, a set $\mathcal{F}$ of events that is a σ-algebra of Ω, and a function $P : \mathcal{F} \to [0, 1]$ that provides the probability measure for the set of events. A σ-algebra of a set X is defined as a non-empty collection of subsets of X closed under complement, countable unions, and countable intersections.*

A probability space is a triple that comprises the sample space Ω (i.e., the set of all possible outcomes), a collection of events within the sample space $\mathcal{F}$ that may or may not comprise every outcome in Ω (i.e., the σ-algebra of Ω), and a probability measure P that assigns a probability to each event in $\mathcal{F}$.

Example 1 (Probability Space). A probability space for our running example comprises the set of all possible outcomes for data-age ($\Omega = \mathbb{R}_{>0}$), a possible collection of events is the sets of intervals of size 1 between natural numbers less than 5 ($\mathcal{F} = \{[1, 2], [2, 3], [3, 4], [4, 5]\}$), and the probability of the event where the data-age X falls within $[1, 2]$ is $P(1 \leq X \leq 2) = \frac{2}{3}$.

The concepts of probability space and random variables are closely related; in this work, a random variable is a function that assigns a numerical value to each elementary outcome. We formally define it below.

Definition 2 (Random Variable). *Consider the probability space $(\Omega, \mathcal{F}, P)$ where Ω is the sample space, $\mathcal{F}$ is a σ-algebra of Ω, and $P : \mathcal{F} \to [0, 1]$ is a probability measure. A random variable X is a function $X : \Omega \to \mathbb{R}$ from the sample space Ω into the set of numerical values $\mathbb{R}$.*

In probability theory, the probability of a continuous random variable X taking a specific value is always equal to zero; instead, its probability is measured over a range of values (i.e., $P[a \leq X \leq b]$). For such variables, in order to calculate probabilities, we use a probability density function (pdf), denoted by $f(t)$. Essentially, the area beneath the two points (a, b) in the plot of such a function constitutes the probability density within the range $P[a \leq X \leq b]$. We recall the formal definition of probability density functions below.

Definition 3 (Probability Density Function). *Consider the probability space $(\Omega, \mathcal{F}, P)$, and a random variable $X : \Omega \to \mathbb{R}$. A probability density function of X, denoted by $f(t)$, is a function that obeys the following properties:*

- $P(X \in [a, b]) = \int_b^a f(t)dt$;
- $f(t) \geq 0$ *for all possible values of* t;
- $\int_{-\infty}^{+\infty} f(t)dt = 1$;

We denote by $Dens(V)$ the set of the density functions over the set of random variables V. The first property specifies that the probability of an event X to be within the range $[a, b]$ is equal to the integral of $f(t)$ (the pdf of X) from a to b; this corresponds to the area beneath the plotted line formed by $f(t)$ within $[a, b]$. The second and third properties specify that a pdf is never negative and that the total area beneath the plot is always equals to 1, respectively. The last property essentially states that the probability of an event to be between $[-\infty, +\infty]$ is 1.

Example 2 (Probability density function). Consider our running example; the pdf for the data-age (D) is the normal distributions function, defined by:

$$f_D(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-u}{\sigma})^2},$$

where σ and u are the location and scale parameters and, for this example, are set to 0.25 and 1, respectively. The plot of this function is shown in Figure 1, where we highlight in grey the area that corresponds to $P(0.8 \leq D \leq 1.2) = 0.64$.

In order to determine conformance between two systems, one first needs a way to quantify the degree of similarity between them. For stochastic systems, we can make use of statistical distance metrics such as the total variation distance [35], the Wasserstein metric [23], and the Hellinger distance [12]. Throughout the remainder of this paper, we employ the last definition (defined below) as our

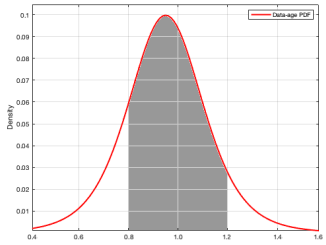


Fig. 1: Running example PDF.

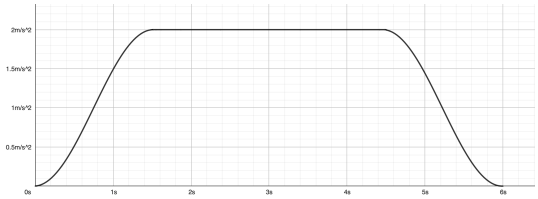


Fig. 2: Example of an acceleration trajectory.

distance metric as it is akin to the Euclidean distance for stochastic system and, hence, more intuitive to understand. In the experiment section (Section 6), however, we also compare the results of using the other two metrics and we show that, in fact, Hellinger distance does perform better than the alternatives.

Definition 4 (Hellinger distance). *Given two probability density functions $f(x)$ and $g(x)$, the Hellinger distance $d(f, g)$ is given by the formula:*

$$d(f,g) = \frac{1}{2}\int (\sqrt{f(t)} - \sqrt{g(t)})^2 dt$$

Essentially, the Hellinger distance can be seen as the ℓ_2 norm between two distributions (more specifically, between the square root of the distributions).

3.3 Conformance

In what follows, we recall a basic theory to define stochastic systems and use the notions of pdf (Definition 3) and a measure of distance (Definition 4) to develop our stochastic conformance notion. We start with the notion of valuation, which provides the values for set of variables and serves as the basis for our definitions.

Definition 5 (Valuation). *Given a set of variables $V = \{X_1, \ldots, X_n\}$, we denote by $Val(V) = V \to \mathbb{R}$ the set of all total functions from V to the real domain $\mathbb{R}$.*

In cyber-physical systems, variables have continuous valuation over time. This can be represented using trajectories, which are collections of variable valuations within a time interval.

Definition 6 (Trajectory). *Given a set of variables V, the set of trajectories over V, denoted by $Trajs(V) = \{x_1, \ldots, x_m\}$, is the set of all mappings $T \to Val(V)$, where T is the time domain, assumed to be a convex subset of $\mathbb{R}_{\geq 0}$.*

Example 3 (Acceleration trajectory). Consider our running example and that, in this example, the acceleration can be depicted by a trajectory that increases for the first 1.5 seconds, stays constant for 3 seconds and then decreases for another 1.5 seconds. This trajectory is depicted in Figure 2.

The trajectory of a variable maps the moments in time to values. Given two trajectories, their similarity can be evaluated using a parametric notion that caters for spatial and temporal discrepancies. This allows for conformance verdicts in scenarios where slight deviations are acceptable.

Definition 7 ((τ,ϵ)-closeness [1]). *Consider the maximum temporal and spatial distances $\tau, \epsilon \in \mathbb{R} \mid \tau, \epsilon > 0$, and the time domain T; then, two trajectories y_1 and y_2 are said to be (τ,ϵ)-close, denoted by $y_1 \approx_{(\tau,\epsilon)} y_2$, if*

1. *for all $t \in \mathrm{dom}(y_1)$ with $t \leq T$, there exists $s \in \mathrm{dom}(y_2)$ such that $|t-s| \leq \tau$ and $\|y_1(t) - y_2(s)\| \leq \epsilon$, and*
2. *for all $t \in \mathrm{dom}(y_2)$ with $t \leq T$, there exists $s \in \mathrm{dom}(y_1)$ such that $|t-s| \leq \tau$ and $\|y_2(t) - y_1(s)\| \leq \epsilon$.*

The notion of (τ,ϵ)-closeness [2] is defined based on the continuous behaviour associated with a continuous physical system, and, hence, this notion does not require the output signals to behave in exact synchronisation. In practice, due to physical phenomena such as measurement errors, transport delays, or mechanical backlash, two implementations of the same system will often slightly deviate from each other [2]. The closeness notion that we adopt in this work to determine how close two trajectories are, in terms of valuation, is based on maximum error margins τ (to account for temporal deviations) and ϵ (to account for spatial deviations). Given the notion of trajectory, we define a deterministic cyber-physical system as an input-output relation based on trajectories; the system is deterministic if each input trajectory yields only one possible output trajectory.

Definition 8 (Deterministic Cyber-Physical System). *A continuous cyber-physical system S is described by the input output relation $S : Trajs(I) \rightarrow Trajs(O)$ where I is the set of input variables and O is the set of output variables. The system S is deterministic if, and only if, $\forall x \in Trajs(I), \forall y_1, y_2 \in Trajs(O)$, we have that $S(x) = y_1 \wedge S(x) = y_2 \implies y_1 = y_2$.*

Lastly, we define a parametric notion of conformance between CPS. Essentially, for every input stimuli that is fed to both systems, the corresponding output trajectories must not deviate beyond the τ and ϵ bounds.

Definition 9 ((τ,ϵ)-conformance [1]). *Given two deterministic cyber-physical systems S_1 and S_2, a maximum temporal distance $\tau \in \mathbb{R}_{>0}$, and a maximum spatial distance $\epsilon \in \mathbb{R}_{>0}$, we say that S_1 (τ,ϵ)-conforms to S_2, denoted by $S_1 \approx_{\tau,\epsilon} S_2$, if, and only if, for any input trajectory $x \in dom(S_1) \cup dom(S_2)$, we have that $S_1(x) \approx_{\tau,\epsilon} S_2(x)$.*

4 Stochastic conformance

In this section, we first present the mathematical formalism to compare two stochastic systems given parametric error margins. Then, we provide some intuition about the differences between our notion and the state-of-the-art [32].

4.1 Stochastic conformance

In stochastic systems, instead of specific values, trajectories map a moment in time to a possible distribution of outcomes. In this work, given the continuous nature of our systems, the distribution is represented by density functions.

Definition 10 (Stochastic trajectory). *Given the probability space $(\Omega, \mathcal{F}, P)$ and a set of random variables V, an stochastic trajectory $x : T \to Dens(V)$ is the set of all mappings of the time domain into a set of density functions over V.*

We denote by $STrajs(V)$ the set of all possible stochastic trajectories over the set of random variables V. Next, we define an stochastic cyber-physical system as a system that, given an input trajectory, outputs a stochastic trajectory.

Definition 11 (Stochastic CPS).

Example 4 (Leader-Follower example). Consider that we use the acceleration trajectory a depicted in Example 3 as the input trajectory for our running example. The output for such a system is a stochastic trajectory that represents the data-age (D) and comprises a density function $f_X^t(x)$ (i.e., a distribution of values for D) for each $t \in \text{dom}(a)$.

We note that, even in a stochastic system, some variables may have a deterministic value (a value with probability 1). In this work, we use a Dirac distribution in order to model such variables. Essentially, a Dirac distribution is a function that is mapped to positive infinity for a specific value and is zero at any other point. The integral of any interval containing that one point is equals to one as its density function.

To define stochastic closeness, we lift the definition of (τ, ϵ)-closeness (see Section 3.3) to work with stochastic trajectories. Essentially, two stochastic trajectories y_1 and y_2 are close if, for every distribution $y_1(t)$ there exists a point in time $s \in [t - \tau, t + \tau]$ that results in a distribution $y_2(s)$ and the Hellinger distance between $y_1(t)$ and $y_2(s)$ is lower than a predetermined ϵ.

Definition 12 (Stochastic Closeness). *Given two stochastic trajectories y_1 and y_2, a maximum temporal distance τ, the Hellinger distance function $d()$, and a maximum distribution distance ϵ, we say that y_1 is stochastically close to y_2, denoted by $y_1 \approx^s_{\tau,\epsilon} y_2$, iff:*

- *for all $t \in dom(y_1)$, there exists $s \in dom(y_2)$ such that $|t - s| \leq \tau$ and $d(y_1(t), y_2(s)) \leq \epsilon$*
- *for all $t \in dom(y_2)$, there exists $s \in dom(y_1)$ such that $|t - s| \leq \tau$ and $d(y_2(t), y_1(s)) \leq \epsilon$*

Analogously, we lift the definition of (τ, ϵ)-conformance [1] to work with stochastic systems. Two stochastic systems conform to each other when, for all possible inputs, the resulting stochastic trajectories are stochastically close.

Definition 13 (Stochastic Conformance). *Given two stochastic systems S_1 and S_2, a maximum temporal distance τ, and a maximum distribution distance ϵ, we say that S_1 (τ,ϵ)-stochastically conforms to S_2, denoted by $S_1 \approx^s_{\tau,\epsilon} S_2$, if, and only if, for any input trajectory $x \in dom(S_1) \cup dom(S_2)$, we have that $S_1(x) \approx^s_{\tau,\epsilon} S_2(x)$.*

4.2 Comparison with the state of the art

Qin et al. [32] defined a conformance notion for stochastic continuous systems that computes the distribution between trajectory distances. Their models of cyber-physical systems comprise stochastic input and output trajectories. Given a system of probability space $(\Omega, \mathcal{F}, P)$, inputs and outputs are defined as a function $Y : T \times \Omega \to \mathbb{R}^m$, where the sample space is part of the domain but the outcome are specific values. Hence, given an outcome $\omega \in \Omega$, one can produce a specific trajectory $y = Y(\bullet, \omega)$ (called a realisation of Y). The possible values for the outcome ω leads to a distribution of realisations of Y.

Now, consider two stochastic outputs Y_1 and Y_2 and two realisations y_1 and y_2. The authors define a distance metric between two trajectories as $d_p(y_1, y_2) := (\int_T ||y_1(t) - y_2(t)||^p \mathrm{d}t)^{\frac{1}{p}}$. A distribution on the possible trajectories for Y_1 and Y_2 leads to a distribution on their distance, denoted by $d(Y_1, Y_2)$. Thus, given a maximum distance ϵ and a failure rate δ, they define a conformance notions as $P(d(Y_1, Y_2) \leq \epsilon) \geq 1 - \delta$.

There are a few key differences between our and their strategies. Overall, our work aims to compute the distance between probabilities distribution; Qin's work, instead, computes the probability of the distance between distributions of trajectories. Additionally, in Qin's work, the degree of closeness between two trajectories is given by the integral of their distance and produces an overall value. The (τ,ϵ)-closeness relation, instead, checks if conformance holds for every point in the trajectories. We motivate our work by identifying two main distinctions.

Firstly, our conformance notion caters for temporal deviations. As an example, consider two identical trajectories in the shape of a signal that has high frequency and high amplitude. If we apply a tiny delay to one of them, this may result in a significant difference between their overall distance, whereas pointwise the distance may be negligible. Hence, allowing for temporal deviations may result in conformance to hold, and, from the literature [19,1], it seems necessary to accommodate such delays.

Secondly, our conformance notion captures short bursts of high deviation. Consider two trajectories that are identical except during a short time interval where they abruptly deviate from each other significantly. In our approach, this abrupt deviation will be considered and this may result in a non-conforming verdict. In Qin's work however, due to the fact that the two trajectories are mostly identical, this may not result in non conformance.

The literature [14] consider such bursts and spikes as significant fault categories that need to be detected by a notion of conformance. To confirm our intuition, we have devised a controlled experiment (see Section 6) where we evaluate our implementation of both conformance notions.

5 Mechanisation

Our approach is mechanised as illustrated in Algorithm 1. Given an input trajectory x that is fed into two systems S_S and S_I (e.g., the system specification and its implementation), the algorithms estimates the distributions for the outputs of $S_S(x)$ and $S_I(x)$.

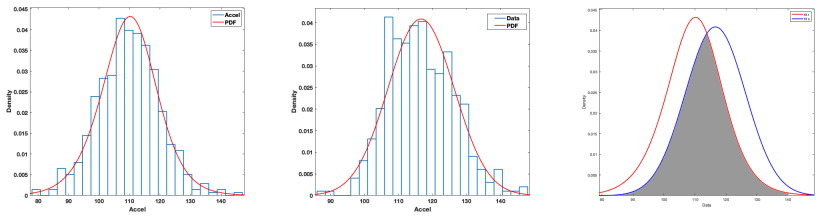
Algorithm 1: Pseudo-code for conformance check algorithm.

```
   input : Trajectory x, Time error margin τ, Value error margin ϵ, ,
           Specification S_S, Implementation S_I;
   output: Boolean conforms;
 1 Function Main() :
 2 |  if StochasticCloseness(x,τ,ϵ,S_S,S_I) then
 3 |  |  return StochasticCloseness(x,τ,ϵ,S_I,S_S);
 4 |  end
 5 |  return False;
 6 end
 7 Function StochasticCloseness(x,τ,ϵ,S_1,S_2) :
 8 |  for t ← 0 to T do
 9 |  |  Boolean conforms = False;
10 |  |  Distribution D_1 = Sample(x,t,S_1);
11 |  |  for s ← Min(0, t − τ) to Max(T, t + τ) do
12 |  |  |  Distribution D_2 = Sample(x,s,S_2);
13 |  |  |  Real distance = HellingerDistance(D_1,D_2);
14 |  |  |  if distance ≤ ϵ then
15 |  |  |  |  return conforms = True;
16 |  |  |  end
17 |  |  end
18 |  |  if !conforms then
19 |  |  |  return False;
20 |  |  end
21 |  end
22 |  return True;
23 end
24 Function Sample(x,t,S) :
25 |  Set [Real] outputs = {};
26 |  for i ← 0 to 50 do
27 |  |  outputs.Add(Execute(S,x,t));
28 |  end
29 |  Distribution D = EstimateDistribution(outputs);
30 |  return D;
31 end
```

The algorithm works as follows. As per Definition 12, the closeness notion between two trajectories checks that every point in the first trajectory is close enough to a neighbouring point in the second trajectory; additionally, the reverse

must also hold. Hence, our algorithm performs the closeness check both ways (lines 02 and 03). The closeness check iterates through every $t \in T$ (line 08). For every t, we sample the possible outcomes for S_1 (line 10). Essentially, the sampling process executes S_1 using x as input several times (lines 25 - 27) and estimates its distribution (line 28). Then, we compute the distributions for the other system $S_2(x)$ within the $[t-\tau, t+\tau]$ time interval (lines 11 and 12) and check if there is at least one point within the interval where the Hellinger distance between $S_1(x)$ and $S_2(x)$ is smaller than ϵ (lines 13 and 14). If no such point exists, then the conformance does not hold (line 19).

As an example, Figure 3 shows the comparison between two output distributions. Consider two implementations of the running example. Given the same input to both systems, Figure 3a shows the output distribution for a time t in the first implementation and, analogously, Figure 3b shows the distribution for the same time t in the second implementation. Our algorithm checks whether the distance (Figure 3c) between both distributions (and also, distributions in the neighbourhood) is greater than ϵ, and, if so, the systems are deemed non-conformant with respect to each other.



(a) Implementation 1 (b) Implementation. 2 (c) Hellinger distance.

Fig. 3: Output distribution.

6 Empirical evaluation

In this section, we present the application of the strategy proposed in Section 5 to a case study in which we detect inserted faults in a model of a convoy of autonomous vehicles.

6.1 Research objectives

Our evaluation assesses the effectiveness of our proposed stochastic conformance notion for testing CPS by comparing it to the state-of-the-art [32]. Using a Simulink [20] model of a connected vehicle platoon, we manually introduce faults and analyse test outputs from both correct and faulty models. Our baseline is the approach by Qin et al. [32] (see Appendix 4.2 for its description). To the

best of our knowledge, this approach is the only stochastic conformance testing approach relevant to trajectories. We aim to answer two key questions:

- **RQ1.** Is our conformance testing approach effective in detecting substantial discrepancies between the outputs of two CPS and, hence, yielding true positive verdicts?
- **RQ2.** Is our conformance testing approach adaptable to allow for negligible discrepancies between the outputs of two CPS and, hence, avoiding false positive verdicts?

The above research questions aim to assess if our conformance notion can efficiently identify common types of programming faults (via insertion of mutants) and common types of failures observed in cyber-physical systems (via insertion of anti-patterns), benchmarking against existing alternatives.

6.2 Case study: connected platoon

Vehicular platooning is an autonomous driving technology that uses wireless communication to maintain a close but safe distance between vehicles in a convoy. We use an open-source model from a previous study [5] where a human-driven lead vehicle sets the pace and autonomous followers adjust their speed accordingly. Communication follows the ETSI EN 302 637-2 standard [21], which, among others, describes the rules for the frequency of packet transmission. These packets comprise Cooperative Awareness Messages (CAM), which contain information about the vehicle, such as acceleration and position. We employ the Intelligent Driver Model [34] as the controller for the vehicles.

The platoon model consists of five vehicles moving along a straight road, with followers adapting to the lead vehicle's acceleration via communication. The main input of our system is the behaviour of the driver in the lead vehicle (i.e., its acceleration), and the output is the acceleration of the follower vehicles.

6.3 Experiment Design

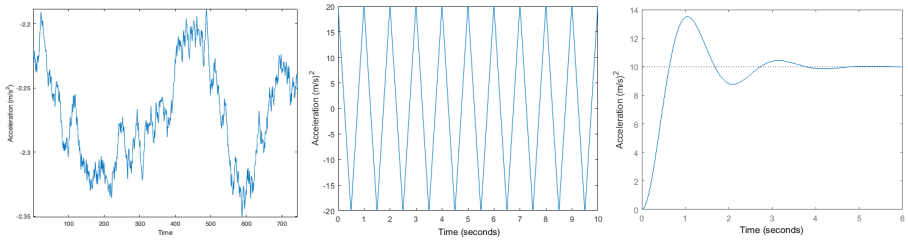
In this section, we explain the experiment design. Particularly, we describe the methodology, and our metrics and hypotheses.

Methodology. An overview of our methodology is as follows. We first generate faulty variants for a correct model of the platooning system. Then, we automatically generate and execute test cases in our models. Lastly, we employ the conformance notions (ours and a state-of-the-art from the literature [31]) on the outputs obtained in the previous step to reach the test verdicts.

We work with two types of fault insertion: (i) code mutation and (ii) signal-based patterns. In the first approach, we manually insert mutations to the correct model in order to create faulty variants. The mutation operators used in this experiment were inserted via a mutation tool for Simulink (FIM [11]). We used the *Delay Operator* (simulates the addition of delays), *Noise Operator* (adds

noise to the signals), *Package Drop* (modifies the value of a variable), and also the *Logical* and *Arithmetic Operator Replacements* (swaps an operator with another of the same type). In total, we inserted 100 faults to the model. As the second method for fault insertion, we insert three types of anti-patterns (i.e., common and significant fault types identified by a taxonomy on signal-based properties of CPSs [14]) to a correct system output, depicted in Figure 4. We incorporate the *spike*, *oscillatory behaviour*, and *overshoot* types of anti-pattern. We have inserted 100 faults for each type by manually modifying the output of the system to match a type of anti-pattern; we consider that the faulty output needs to be identified by the conformance notions.

Fig. 4: Examples of signal-based patterns.



(a) Spike. (b) Oscillation. (c) Overshoot.

To generate inputs for the simulations, we use one random-based generation approach and two multi-objective search-based algorithms (standard simulated annealing [26] and genetic algorithm [30]). The former generates valid but completely random test cases and this can be used as a baseline measurement. The two other options are search-based heuristics which have been shown to generate tests that are more likely to exhibit failures [7,37,4]. Both search heuristics adopted in this case study employ 'fitness functions' to optimise the search, and we consider a notion of closeness, coverage, and diversity as the objectives; they have been demonstrated to lead the SUT towards conformance violations [4].

The verdict of a test is given by a conformance notion. For this study, to be considered an actual failure, the deviations need to be above error margins. The values chosen here are based on a maximum spatial deviation of $0.5m/s^2$ for the acceleration trajectories and a maximum temporal deviation of $1s$, which are used in a study conducted by domain experts [5]. With respect to our conformance notion, we chose to replicate the specific values of maximum allowed deviation and, hence, we have chosen $\tau = 1s$ and $\epsilon = 0.5$. Qin's conformance notion, however, requires two parameters: the overall difference between error margins for the entire trajectory (λ) and a failure rate (δ), which are set to 2.5 and 0.25, respectively. If distance between two trajectories is set to constant 0.5 for 10 seconds, this results in lambda value (i.e., the integral of the distance between the two trajectories) of 5. Hence, we chose 2.5 as it is half of this value.

Metrics and hypotheses. In this experiment, we make use of the number of true positives (TP) and false positives (FP) verdicts. These metrics essentially quantify the fault detection rate of each conformance notion. Detecting more faults is generally the goal of any testing approach. We have defined one hypothesis for each of our research questions. With respect to RQ1, which focusses on fault detection capabilities of the conformance notion, we have devised the hypotheses $H_{A0} : TP_{our} \leq TP_{alt}$ and $H_{A1} : TP_{our} > TP_{alt}$. Essentially, a test suite will be generated and the same test suite will be fed to both the correct and faulty models. The conformance notion that detects a higher number of mutants correctly is deemed more effective in the True Positive rate. The null hypothesis (H_{A0}) states that the number of True Positives detected by our conformance notion (TP_{our}) is lower or equal to the one obtained by the alternative notion (TP_{alt}). This experiment aims to refute such a hypothesis. Thus, an alternative hypothesis (H_{A1}) is also defined, which has a complementary role to the null one, and can be accepted in case their counterpart is rejected. Analogously, we have defined the hypotheses $H_{B0} : FP_{our} \geq FP_{alt}$ and $H_{B1} : FP_{our} < FP_{alt}$ to compare the number of false positives resulting from ours (FP_{our}) and the alternative (FP_{alt}) conformance notion.

6.4 Results

We split the main results into two tables. Table 1 shows the false and true positive rates (as well as other metrics) for the mutation operators and Table 3 focuses on the detection of anti-patterns. We show the results of applying our three variants of our conformance notion: Hellinger distance, Wasserstein metric, and Total Variation Distance (TVD) and the baseline to the three types of input generation: Random, Simulated Annealing (SA), and Genetic Algorithms (GA).

The numbers shown in Table 1 represent the number of false positives (FP), true positives (TP), false negatives (FN), and true negatives (TN). A false positive occurs when a test fails when it should not have. On the other hand, true positives represent tests that have failed correctly. Analogously, true and false negatives are tests that have correctly and incorrectly passed, respectively. The numbers represent how many verdicts are in each category. Moreover, we also display the values for Accuracy, Precision, Recall and F1 metrics.

The results indicate a small but significant difference for the true positive rates for the detection of mutations in favour of our approach. The difference increases with the complexity of the input generation approach. In terms of distance metric, Hellinger and Total Variation distances seemed to yield similar and better results compared to Wasserstein metric. By analysing the results more closely (see Table 2 for results per operator), we note how our conformance notion tends to detect the *Noise* more predominantly than the alternative. This is because the *Noise* operator is more likely to lead to a short burst of deviation between the results from the correct and faulty implementations. The operators *Package Drop* and *Logical Operator Replacement* showed little difference between the notions, however.

Table 1: Detection of mutation operators.

(a) Random Input

	FP	TP	FN	TN	Acc.	Prec.	Rec.	F1
Ours - Hellinger	0	71	19	10	0.81	1.00	0.79	0.88
Ours - Wasserstein	0	70	20	10	0.80	1.00	0.78	0.88
Ours - TVD	0	71	19	10	0.81	1.00	0.79	0.88
State-of-the-art	5	70	20	5	0.75	0.93	0.78	0.85

(b) Simulated Annealing

	FP	TP	FN	TN	Acc.	Prec.	Rec.	F1
Ours - Hellinger	0	81	9	10	0.91	1.00	0.90	0.95
Ours - Wasserstein	0	79	11	10	0.88	1.00	0.87	0.93
Ours - TVD	0	82	8	10	0.92	1.00	0.91	0.95
State-of-the-art	6	77	23	4	0.74	0.93	0.77	0.84

(c) Genetic Algorithms

	FP	TP	FN	TN	Acc.	Prec.	Rec.	F1
Ours - Hellinger	0	85	5	10	0.95	1.00	0.94	0.97
Ours - Wasserstein	0	82	8	10	0.91	1.00	0.90	0.95
Ours - TVD	0	84	6	10	0.94	1.00	0.93	0.97
State-of-the-art	4	79	11	6	0.84	0.94	0.88	0.91

Table 2: TP per mutant operator.

	Noise	PD	LOR	AOR	Delay
Ours - RA	16	08	17	16	14
Alt - RA	15	08	17	16	14
Ours - SA	16	13	20	17	15
Alt - SA	14	12	20	16	15
Ours - GA	19	15	20	17	14
Alt - GA	15	14	20	16	14

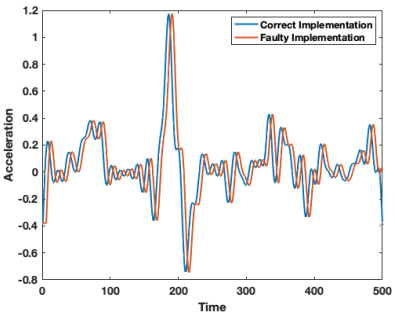


Fig. 5: Trajectory with a delay.

As for the false positives (shown in Table 1), most of them occurred with the *Delay* operator. The reasoning is that by allowing for temporal error margins, our conformance notions tends to disregard negligible deviations resulting from small *Delay* mutations (below the error margin). Such mutants are not killed and, hence, we avoid false positives. As an example, Figure 5 shows the output of the correct and of an implementation with a small *Delay* operator. Even though both outputs are very similar, this slight delay in time has led to a drastic

distance using Qin's notion due to a high degree of accumulated variation. Our conformance notion, however, permits such divergences depending on the value of τ, which gives flexibility to the verification process.

Analogously, our methods (especially Hellinger and TVD) consistently produce fewer false negatives than the state-of-the-art, across all types of input generation. Hence, they are less likely to incorrectly pass faulty systems. The difference is more pronounced as input generation gets more sophisticated (from Random to SA to GA). For instance, with Simulated Annealing, the FN count dropped from 23 (baseline) to as low as 8 (ours), showing a significant improvement. Wasserstein performs slightly worse than Hellinger and TVD in terms of false negatives, though still better than the baseline. Improved FN rates lead to higher recall, which is reflected in the metrics: when using simulated annealing, recall improves from 0.77 (baseline) to > 0.90 (ours - Hellinger and TVD).

Lastly, Table 3 shows the results for the detection rates of anti-patterns. Similarly to noise mutation operator, the spike anti-pattern tends to result in short bursts of discrepancy, which leads to our conformance notion detecting more of them compared to the alternative. The oscillation anti-pattern has been efficiently detected by both approaches and, as for overshoots, there is a small but significant difference in favour of our strategy.

Table 3: Detection of anti-patterns.

(a) Spike

	FP	**TP**	**FN**	**TN**	**Acc.**	**Prec.**	**Rec.**	**F1**
Ours - Hellinger	0	82	8	10	0.89	1.00	0.89	0.94
Ours - Wasserstein	0	71	19	10	0.78	1.00	0.78	0.88
Ours - TVD	0	82	8	10	0.89	1.00	0.89	0.94
State-of-the-art	0	65	25	10	0.72	1.00	0.72	0.84

(b) Oscillation

	FP	**TP**	**FN**	**TN**	**Acc.**	**Prec.**	**Rec.**	**F1**
Ours - Hellinger	0	88	2	10	0.95	1.00	0.95	0.97
Ours - Wasserstein	0	86	4	10	0.93	1.00	0.93	0.96
Ours - TVD	0	87	2	10	0.94	1.00	0.94	0.97
State-of-the-art	0	88	2	10	0.95	1.00	0.95	0.97

(c) Overshoot

	FP	**TP**	**FN**	**TN**	**Acc.**	**Prec.**	**Rec.**	**F1**
Ours - Hellinger	0	60	29	10	0.67	1.00	0.67	0.80
Ours - Wasserstein	0	60	29	10	0.67	1.00	0.67	0.80
Ours - TVD	0	59	31	10	0.66	1.00	0.66	0.80
State-of-the-art	0	56	34	10	0.63	1.00	0.63	0.77

6.5 Threats to validity

As threats to the validity of our experiment, we note that the choice of mutant operators is made by domain experts, but a thorough and formal study of their relation to real faults needs to be conducted. To mitigate this issue, we have also introduced anti-patterns (spike, overshoot, and oscillation) common in CPS to mimic failures. Furthermore, the error margin values in the oracle impact the results: small values would detect all mutants and large ones would detected none. As a mitigation measure, the values we have chosen throughout the work (e.g., τ, ϵ, and δ, as well as the mutation operators) are based on prior experiments and domain knowledge. Moreover, we fit the distributions of the outputs in the experiment to a normal distribution. This is an assumption based on data gathered from the experiments; we have chosen a distribution fit that most closely matches with the observed ones. Lastly, this experiment only considers one (albeit, complex) example of connected CPS. This makes it hard to generalise the outcome of this experiment for a general class of cyber-physical systems. This is mitigated by the large number of mutants that were inserted into this system.

7 Conclusions

We have developed a novel stochastic conformance notion to test connected Cyber-Physical Systems (CPS) that takes error margins into account. Our approach is well-suited to test CPS, where the interaction between computational, physical, and environmental components may lead to a probabilistic distribution of outcomes. Our notion is adaptable so that negligible perturbations (e.g., subtle measurement errors) are not mistakenly flagged as failures. Our approach verifies CPS by checking whether the distance between two output distributions (the observed and the expected one) falls within safety bounds. Our notion is able to detect deviations in stochastic behaviour that manifest under faulty conditions, while accommodating for natural temporal and spatial variations. In the mechanisation of our approach, we implemented our conformance as a tool and evaluated its effectiveness through a case study involving a connected platoon of autonomous vehicles. We show that our approach can detect faulty behaviour, such as oscillation and spikes (common types of anti-patterns in CPS) more reliably than alternatives found in the literature.

Acknowledgements

Hugo Araujo and Mohammad Reza Mousavi have been partially supported by the UKRI Trustworthy Autonomous Systems Node in Verifiability, Grant Award Reference EP/V026801/2, EP- SRC project on Verified Simulation for Large Quantum Systems (VSL-Q), grant reference EP/Y005244/1 and the EP-SRC project on Robust and Reliable Quantum Comput- ing (RoaRQ), Investigation 009 Model-based monitoring and calibration of quantum computations (ModeMCQ), grant reference EP/W032635/1 and ITEA/InnovateUK projects GENIUS and GreenCode.

References

1. Abbas, H., Hoxha, B., Fainekos, G.E., Deshmukh, J.V., Kapinski, J., Ueda, K.: WiP abstract: Conformance testing as falsification for cyber-physical systems. In: Proceedings of the ACM/IEEE 5th International Conference on Cyber-Physical Systems (ICCPS 2014). p. 211. IEEE CS (2014), available online: http://arxiv.org/abs/1401.5200
2. Abbas, H., Mittelmann, H., Fainekos, G.: Formal property verification in a conformance testing framework. In: Formal methods and models for codesign (memocode), 2014 twelfth acm/ieee international conference on. pp. 155–164. IEEE (2014)
3. Althoff, M.: Reachability analysis and its application to the safety assessment of autonomous cars. Ph.D. thesis, Technische Universität München (2010)
4. Araujo, H., Carvalho, G., Mousavi, M., Sampaio, A.: Multi-objective search for effective testing of cyber-physical systems. In: Proceedings of the 17th International Conference on Software Engineering and Formal Methods. Springer (2019)
5. Araujo, H., Hoenselaar, T., Mousavi, M.R., Vinel, A.: Connected automated driving: A model-based approach to the analysis of basic awareness services. In: 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. pp. 1–7. IEEE (2020)
6. Araujo, H., Mousavi, M.R., Varshosaz, M.: Testing, validation, and verification of robotic and autonomous systems: a systematic review. ACM Transactions on Software Engineering and Methodology **32**(2), 1–61 (2023)
7. Arrieta, A., Wang, S., Markiegi, U., Sagardui, G., Etxeberria, L.: Search-based test case generation for cyber-physical systems. In: 2017 IEEE Congress on Evolutionary Computation (CEC). pp. 688–697 (2017). https://doi.org/10.1109/CEC.2017.7969377
8. Asarin, E., Dang, T., Frehse, G., Girard, A., Le Guernic, C., Maler, O.: Recent progress in continuous and hybrid reachability analysis. In: 2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control. pp. 1582–1587 (2006). https://doi.org/10.1109/CACSD-CCA-ISIC.2006.4776877
9. Baheti, R., Gill, H.: Cyber-physical systems. The impact of control technology **12**(1), 161–166 (2011)
10. Bak, S., Chaki, S.: Verifying cyber-physical systems by combining software model checking with hybrid systems reachability. In: Proceedings of the 13th International Conference on Embedded Software. pp. 1–10 (2016)
11. Bartocci, E., Mariani, L., Ničković, D., Yadav, D.: Fim: fault injection and mutation for simulink. In: Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. p. 1716–1720. ESEC/FSE 2022, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3540250.3558932, https://doi.org/10.1145/3540250.3558932
12. Beran, R.: Minimum hellinger distance estimates for parametric models. The annals of Statistics pp. 445–463 (1977)
13. Böhm, A., Kunert, K.: Data age based retransmission scheme for reliable control data exchange in platooning applications. In: 2015 IEEE International Conference on Communication Workshop (ICCW). pp. 2412–2418. IEEE (2015)

14. Boufaied, C., Jukss, M., Bianculli, D., Briand, L.C., Parache, Y.I.: Signal-based properties of cyber-physical systems: Taxonomy and logic-based characterization. Journal of Systems and Software **174**, 110881 (2021)
15. Chen, X., Sankaranarayanan, S.: Reachability analysis for cyber-physical systems: Are we there yet? In: NASA formal methods symposium. pp. 109–130. Springer (2022)
16. Clarke, E.M.: Model checking. In: Foundations of Software Technology and Theoretical Computer Science: 17th Conference Kharagpur, India, December 18–20, 1997 Proceedings 17. pp. 54–56. Springer (1997)
17. Clarke, E.M., Zuliani, P.: Statistical model checking for cyber-physical systems. In: International symposium on automated technology for verification and analysis. pp. 1–12. Springer (2011)
18. Corso, A., Moss, R., Koren, M., Lee, R., Kochenderfer, M.: A survey of algorithms for black-box safety validation of cyber-physical systems. Journal of Artificial Intelligence Research **72**, 377–428 (2021)
19. Deshmukh, J.V., Majumdar, R., Prabhu, V.S.: Quantifying conformance using the skorokhod metric. In: Computer Aided Verification: 27th International Conference, CAV 2015, San Francisco, CA, USA, July 18-24, 2015, Proceedings, Part II 27. pp. 234–250. Springer (2015)
20. Documentation, S.: Simulation and model-based design (2020), `https://www.mathworks.com/products/simulink.html`
21. ETSI EN 302 637- 2; Intelligent Transport Systems (ITS); vehicular communications; basic set of applications; part 2: Specification of cooperative awareness basic service (2013)
22. Gerking, C., Schäfer, W., Dziwok, S., Heinzemann, C.: Domain-specific model checking for cyber-physical systems. In: MoDeVVa@ models. pp. 18–27 (2015)
23. Givens, C.R., Shortt, R.M.: A class of wasserstein metrics for probability distributions. Michigan Mathematical Journal **31**(2), 231–240 (1984)
24. Hamzah, M., Islam, M.M., Hassan, S., Akhtar, M.N., Ferdous, M.J., Jasser, M.B., Mohamed, A.W.: Distributed control of cyber physical system on various domains: A critical review. Systems **11**(4), 208 (2023)
25. Hashemi, N., Lindemann, L., Deshmukh, J.V.: Statistical reachability analysis of stochastic cyber-physical systems under distribution shift. arXiv preprint arXiv:2407.11609 (2024)
26. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. science **220**(4598), 671–680 (1983)
27. Krichen, M., Tripakis, S.: Conformance testing for real-time systems. Formal Methods in System Design **34**(3), 238–304 (2009)
28. Kwiatkowska, M., Norman, G., Parker, D.: Prism 4.0: Verification of probabilistic real-time systems. In: Computer Aided Verification: 23rd International Conference, CAV 2011, Snowbird, UT, USA, July 14-20, 2011. Proceedings 23. pp. 585–591. Springer (2011)
29. Leemans, S.J., Syring, A.F., van der Aalst, W.M.: Earth movers' stochastic conformance checking. In: Business Process Management Forum: BPM Forum 2019, Vienna, Austria, September 1–6, 2019, Proceedings 17. pp. 127–143. Springer (2019)
30. Mitchell, M.: An introduction to genetic algorithms. MIT press (1998)
31. Qin, X., Aréchiga, N., Deshmukh, J., Best, A.: Robust testing for cyber-physical systems using reinforcement learning. In: Proceedings of the 21st ACM-IEEE International Conference on Formal Methods and Models for System Design. pp. 36–46 (2023)

32. Qin, X., Hashemi, N., Lindemann, L., Deshmukh, J.V.: Conformance testing for stochastic cyber-physical systems. In: Conference on formal methods in computer-aided design, FMCAD. p. 294 (2023)
33. Sirjani, M., Lee, E.A., Khamespanah, E.: Model checking software in cyberphysical systems. In: 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). pp. 1017–1026. IEEE (2020)
34. Treiber, M., Hennecke, A., Helbing, D.: Congested traffic states in empirical observations and microscopic simulations. Physical review E **62**(2), 1805 (2000)
35. Verdú, S.: Total variation distance and the distribution of relative information. In: 2014 Information Theory and Applications Workshop (ITA). pp. 1–3. IEEE (2014)
36. Woehrle, M., Lampka, K., Thiele, L.: Conformance testing for cyber-physical systems. ACM Transactions on Embedded Computing Systems (TECS) **11**(4), 1–23 (2013)
37. Zhang, M., Ali, S., Yue, T.: Uncertainty-wise test case generation and minimization for cyber-physical systems. Journal of Systems and Software **153**, 1–21 (2019)
38. Zheng, X., Julien, C., Kim, M., Khurshid, S.: Perceptions on the state of the art in verification and validation in cyber-physical systems. IEEE Systems Journal **11**(4), 2614–2627 (2015)