2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Scenario Approach and Conformal Prediction for Verification of Unknown Systems via Data-Driven Abstractions

Rudi Coppola[1], Andrea Peruffo[1], Lars Lindemann[2], and Manuel Mazo Jr.[1]

Abstract— **Verification of uncertain, complex dynamical systems is crucial in the modern day world. An increasingly common method to verify complex logic specifications for dynamical systems involves symbolic abstractions: simpler, finite-state models whose behaviour mimics the one of the systems of interest. By sampling trajectories of the concrete, unknown system and via robust analysis, we build a data-driven abstraction, related to the underlying model through a probabilistic behavioural inclusion relation. As the distribution from which the trajectories are drawn is unknown, we adopt two distinct distribution-free theories, namely scenario optimization and conformal prediction. We compare and discuss the differences between the two approaches in terms of the type of guarantees that they are able to provide. Furthermore, via experimental benchmarks we outline the efficiency of the two methods with respect to the number of samples available and the tightness of the guarantees.**

I. Introduction

The increasing digitalisation and interconnection of systems is forging a new large class of complex models, often equipped with decision making capabilities and data-driven perception (e.g. self-driving cars). These systems introduce two challenges: their inherent complex nature prevents the use of exact models, along with the verification of desired behaviours with either formal or probabilistic approaches. Thus, one can embrace a black-box model approach, and rely solely on observations of the unknown system. As the resulting models are uncertain, their *verification* becomes ever more essential, yet more complex. Verification aims at checking the correctness of a system against specifications expressed in formal languages; typically, these languages require the knowledge of the underlying distribution to accurately model the transition probabilities. In a data-driven setting however, an accurate distribution is often unknown; we thus turn to *confidence* intervals to provide a similar intuition. Two popular techniques are scenario approach [6], [12] and conformal prediction [29], [33]. The former is an optimisation-based technique that provides probably approximately correct (PAC) guarantees on a user-defined performance metric. The approach relies on independently drawn samples and it is distribution-free, namely it requires no previous knowledge about the underlying distribution driving a system's uncertainty. The latter is a statistical technique providing confidence intervals for general prediction algorithms; similarly to the scenario theory, it foregoes assumptions on the underlying distribution and on the actual prediction mechanism.

The scenario approach (SA) is extensively employed in data-driven verification and synthesis. Among others, in [9], [2], [17] where the authors employ the scenario approach to construct interval MDPs, equipped with probability intervals derived from the scenario approach. One-step transitions are used in [10] to define a PAC alternating simulation relation between abstractions and the concrete system. Modeling of unknown systems is also a prominent area of research: in [21], the authors propose PAC over-approximations of monotone systems, finally employed to build abstractions of the concrete systems; [15] builds abstractions and synthesises controllers based on data-driven growth rates. Data-driven ℓ-complete models are presented in [24] for linear PETC models and in [7] for general systems.

Conformal prediction (CP) is a lightweight statistical technique for uncertainty quantification of complex models [1], [33], [29]. CP has been applied to a wide range of applications, e.g. in drug discovery [8], robotic motion planning [19], and within a variety of machine-learning frameworks [3]. CP has further been used for system verification under temporal logic specifications [20], [5], [26]. In particular, [25], uses CP to estimate the conformance between two stochastic systems, while we are here instead interested in using CP for checking a behavioral inclusion property between two systems. Closest to our work are [13], [22], [31], [4] in which CP is applied to reachability problems where no system knowledge and only a finite number of system observations is available.

Contributions. In this work, we tackle a verification problem for unknown dynamical systems. Our approach constructs a data-driven finite abstraction, belonging to the so-called Strongest Asynchronous ℓ-complete Abstractions (SAℓCA), from a collection of independent samples of the underlying system's trajectories. We define the notion of probabilistic behavioural inclusion to relate the abstraction to the system behaviours stemming from random samples. We leverage two approaches to provide confidence results on the probabilistic inclusion: the scenario theory for non-convex problems, and the conformal prediction approach. Further, we compare the two methods, highlighting their similarities and differences, in terms of the confidence guarantees they offer and of computational requirements. Once the probabilistic behavioural inclusion is established, the desired property can then be verified directly on the finite state abstraction with standard techniques [30].

[1] Rudi Coppola, Andrea Peruffo, Manuel Mazo Jr. are with the Delft University of Technology, Delft, the Netherlands. r.coppola@tudelft.nl

[2] Lars Lindemann is with the University of Southern California, Los Angeles, California, USA.

II. Preliminaries

A. Notation

Let Δ be the event space of a random vector and denote by $(\Delta, \mathcal{F}, \mathbb{P})$ the associated probability space, where $\mathcal{F}$ is a σ-algebra and $\mathbb{P} : \mathcal{F} \to [0,1]$ is a probability measure on Δ. For any fixed $N \in \mathbb{N}$, we consider a sequence $\omega_0, \omega_1, ..., \omega_N$ of independent identically distributed random variables (RVs) with $\omega_i : \Delta \to \Omega$, where $(\Omega, \mathcal{G})$ is a measurable space and, for $x \in \Delta$, $\hat{\omega}_i := \omega(x)$ denotes a realization of it. Our work takes a distribution-free outlook, i.e. $\mathbb{P}$ is unknown.

Given a set $\mathcal{Q}$ we denote its H-th cartesian product by $\mathcal{Q}^H$, unless otherwise specified. For $q_H \in \mathcal{Q}^H$ we denote its $(k+1)$-th element by $q_H(k)$ for $k \in [0, H-1]$. We denote the restriction of q_H to the interval $I = [k_1, k_2]$, with $k_1 < k_2 < H$ by $q_H|_{[k_1,k_2]} := q_H(k_1)...q_H(k_2)$. Given two sequences $q_{H,1} \in \mathcal{Q}^H$ and $q_{H',2} \in \mathcal{Q}^{H'}$ we denote their concatenation by $q_{H,1} \cdot q_{H',2}$. For $\infty > H \geq H'$, we say that $q_{H,1}$ exhibits $q_{H',2}$ if there exists $k \geq 0$ such that $q_{H,1}(k+i) = q_{H',2}(i)$ for $i = \{0, ..., H'-1\}$, denoted $q_{H,1} \models \Diamond q_{H',2}$. Given a set of sequences $Q \subseteq \mathcal{Q}^H$ we say that Q exhibits $q_{H',2}$ if there exists $q_H \in Q$ such that $q_H \models \Diamond q_{H',2}$, denoted $Q \models \Diamond q_{H',2}$.

B. Scenario Optimization

The scenario approach (see [6] for more details) constructs an optimisation program

$$\begin{aligned} \min_{\theta} \quad & c^T \theta \\ \text{s.t.} \quad & g(\theta, \hat{\omega}_i) \leq 0, \quad \text{for all } i = 1, ..., N, \end{aligned} \tag{1}$$

where $\theta \in \mathbb{R}^d$ represents the optimisation variable, $c \in \mathbb{R}^d$ represent the cost, $g : \mathbb{R}^d \times \Omega \to \mathbb{R}$ is a constraint function. Once the optimal solution θ^*_N is computed, the scenario theory allows to obtain *high-confidence* bounds on the probability of constraint violation. The value θ^*_N depends on the N collected samples $\hat{\omega}_1, ..., \hat{\omega}_N$; the following theorem allows to determine the probability that θ^*_N would violate the new constraint given by a realization of ω_0, $g(\theta^*_N, \hat{\omega}_0)$.

Theorem 1 (PAC bounds [12, Theorem 1])**.** *Given a confidence parameter $\beta \in (0,1)$ and the solution θ^*_N, it holds*

$$\mathbb{P}^N(\mathbb{P}\{g(\theta^*_N, \omega_0) > 0\} \leq \epsilon(s^*_N, \beta, N)) \geq 1 - \beta, \tag{2}$$

where $\epsilon(\cdot)$ can be computed via a polynomial equation (omitted here for brevity) and s^*_N is the so-called complexity of the solution, representing the minimum number of constraints ($m \leq N$) that yield the same solution θ^*_N. □

C. Conformal Prediction

Conformal prediction (CP) is a statistical technique providing confidence intervals for general prediction algorithms. Let us consider the aforementioned random[1] variables $\omega_0, ..., \omega_N$. In this work we employ a formulation of CP where we have sampled $\hat{\omega}_1, ..., \hat{\omega}_N$; based on that alone, we want to predict ω_0. With this, we emphasize that no additional information or *feature* of ω_0 is available at the time of prediction. Additionally, we adopt the formulation of CP known as *split conformal prediction* [23], [11], where the set of observations is divided into two subsets as described below. The set of outcomes is usually formalized through the notion of a bag. A bag $\hat{\mathcal{I}} := \wr \hat{\omega}_1, ..., \hat{\omega}_N \int = \wr \hat{\omega}_i \int_{i=1}^N$ is a collection of elements, or examples, in which repetition is allowed and any information about the ordering of the list $\hat{\omega}_1, ..., \hat{\omega}_N$ is removed. We partition $\hat{\mathcal{I}}$ in a training set $\hat{\mathcal{I}}_{\text{train}}$ and a calibration set $\hat{\mathcal{I}}_{\text{cal}}$, for simplicity assume $\hat{\mathcal{I}}_{\text{train}} = \{\hat{\omega}_i : i = 1, ..., M\}$. The first step is the definition of a *nonconformity measure* $A(\hat{\mathcal{I}}_{\text{train}}, \omega)$, a real valued function that assigns a measure to how different an element $\omega \in \Omega$ is from the training examples, see [33] for details. We adopt this measure to assign to every observation in the calibration set a score indicating how different the observation is when compared to the training examples, known as the *nonconformity score* $R_j := A(\hat{\mathcal{I}}_{\text{train}}, \hat{\omega}_j)$ for $\hat{\omega}_j \in \hat{\mathcal{I}}_{\text{cal}}$. We are interested in predicting the nonconformity score R_0 from the bag $\wr \hat{\omega}_i \int_{i=1}^N$. Formally, given a failure probability δ, we want to estimate a prediction interval $\gamma^\delta(\omega_1, ..., \omega_N)$ such that

$$\mathbb{P}^{N+1}[R_0 \leq \gamma^\delta(\omega_1, ..., \omega_N)] \geq 1 - \delta, \tag{3}$$

where $\mathbb{P}^{N+1}$ is the product measure on the $N+1$ i.i.d. RVs. It can be shown [29] that $\gamma^\delta(\omega_1, ..., \omega_N)$ is the $(1-\delta)$-th quantile of the empirical distribution of the values $R_{M+1}, \dots, R_N$, and $R_{N+1} := \infty$. By defining $p := M + \lceil (N - M + 1)(1-\delta) \rceil$ we set $\gamma^\delta(\omega_1, ..., \omega_N) = R_p$, that is the p-th smallest nonconformity score is an upper bound for R_0; note that if $p = N+1$, inequality (3) is trivially valid but uninformative.

III. Problem Formulation

A. System Description and Verification Problem

Consider a time-invariant dynamical system with symbolic outputs described by

$$\Sigma(x) := \begin{cases} x_{k+1} = f(x_k), \\ y_k = h(x_k), \\ x_0 = x, \end{cases} \tag{4}$$

where $x_k \in \mathcal{X} \subset \mathbb{R}^{n_x}$ is the plant's state at time $k \in \mathbb{N}_+$, n_x is the state-space dimension, x_0 is the initial state, $y_k \in \mathcal{Y}$ is the system output with $|\mathcal{Y}| < \infty$. The expression of the flow $f(\cdot)$ and of the output map $h(\cdot)$ are unknown, but we assume that given an initial condition x_0 we can observe the output sequence or *behaviour* $y_0, y_1, ...$ generated by $\Sigma(x_0)$. We denote with $\mathcal{B}_H(\Sigma(x_0)) \in \mathcal{Y}^H$ the behaviour for the time interval $k = [0, H-1]$ generated by $\Sigma(x_0)$, and by $\mathcal{B}_H(\Sigma)$ the set of all behaviours for the same time interval generated by all possible initial conditions. The map $h(\cdot)$ can be regarded as a *partitioning* map, that returns the partition label (or index) corresponding to any state x_k. This observation relates to the notion of *equivalence class* [30]:

$$[y] = \{x \in \mathcal{X} \mid y = h(x)\},$$

[1] The framework of CP requires a slightly weaker assumption than i.i.d., namely it is sufficient that they are *exchangeable*, see [29] for details.

and similarly, we define the equivalence class for an output sequence $y_{\ell,i} = y_i(0)y_i(1)...y_i(\ell-1) \in \mathcal{Y}^\ell$ as

$$[y_{\ell,i}] = \{x |\ y_i(j) = h(f^j(x)) \text{ for } j = 0, ..., \ell-1\}, \quad (5)$$

with $f^0(x) = x$. Equation (5) states that for $i = 1, ..., |\mathcal{Y}|^\ell$, i.e. for every ℓ-sequence $y_{\ell_i} \in \mathcal{Y}^\ell$ the output equivalence class $[y_{\ell,i}]$ is the set of points x such that if the dynamical system is initialized at x, then the output sequence over the time interval $[0, \ell-1]$ corresponds to $y_{\ell,i}$. Further, for all $\ell \geq 1$, the set of all $[y_{\ell,i}]$ forms a partition of the domain $\mathcal{X}$.

Our goal consists of verifying whether an unknown system (4) satisfies a given *specification* in a probabilistic sense. Examples of specifications can be found in [30].

B. Transition Systems, ℓ-complete Models, System Relations

We solve the verification problem with an abstraction-based technique: we represent the concrete system Σ by a (finite-)state transition system, such that it is amenable to algorithmic iterative verification frameworks.

Definition 1 ((Autonomous) Transition System (TS) [30])**.** *A transition system (TS) S is a tuple $(\mathcal{X}, \mathcal{X}_0, \sigma, \mathcal{Y}, \mathcal{H})$, where $\mathcal{X}$ is the set of states, $\mathcal{X}_0 \subseteq \mathcal{X}$ is the set of initial states, $\mathcal{Y}$ is the set of outputs, $\sigma \subseteq \mathcal{X} \times \mathcal{X}$ is a transition relation, and $\mathcal{H} : \mathcal{X} \to \mathcal{Y}$ is an output map.*

One can represent system (4) equivalently in the form of a TS, where the state space is given by $\mathbb{R}^{n_x}$, the transition relation is dictated by f, and the output map is dictated by $h(\cdot)$. Denote by S the transition system equivalent to (4). We extend our notation $S(x_0)$, $\mathcal{B}^H(S(x_0))$, and $\mathcal{B}^H(S)$ for TSs using this equivalence. S and Σ have identical behaviours, that is $b_H \in \mathcal{B}^H(\Sigma) \iff b_H \in \mathcal{B}^H(S)$. From now on, we assume that for the concrete system it holds $\mathcal{X}_0 = \mathcal{X}$.

As we assume to have access solely to S's behaviours, i.e. sequences of elements of $\mathcal{Y}$, we first discuss which abstraction class suits this scope. In [27], [28], the authors present a particular class of abstractions known as *Strongest Asynchronous ℓ-complete Approximations* (SAℓCA). We motivate our interest in the SAℓCA of S because it can be constructed directly by knowing the set of all behaviours[2] of a system, $\mathcal{B}_H(S)$, bypassing the need for the knowledge of the internal mechanisms of the underlying model.

We now illustrate how to construct such abstractions. Let us define the set of all ℓ-long subsequences of all H-long behaviours in the TS S as

$$\Pi_{\ell,H} := \bigcup_{b \in \mathcal{B}_H(S)} \ \bigcup_{0 \leq k \leq H-\ell} b|_{[k,k+\ell-1]}. \quad (6)$$

Note that $\Pi_{\ell,H} \subseteq \mathcal{Y}^\ell$, and from the knowledge of $\Pi_{\ell+1,H}$, all the $(\ell+1)$ subsequences, we can easily obtain $\Pi_{\ell,H}$, all the ℓ subsequences.

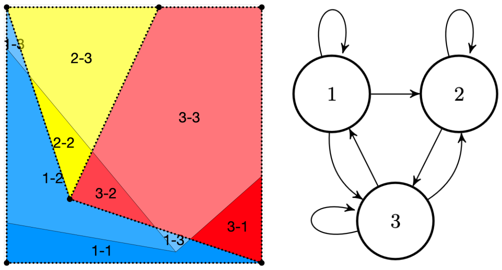

Fig. 1: Partition of the domain based on the set $\Pi_{2,H}$ sequences (left), and the resulting SAℓCA for $\ell = 1$ (right).

Definition 2 ((Strongest asynchronous) ℓ-complete abstraction (adapted from [28]))**.** *Let $S := (\mathcal{X}, \mathcal{X}_0, \sigma, \mathcal{Y}, \mathcal{H})$ be a TS, and let $\Pi_{\ell+1,H}$ be defined as in* (6)*. Then, the TS $S_\ell := (\mathcal{X}_\ell, \mathcal{X}_{\ell 0}, \sigma_\ell, \mathcal{Y}, \mathcal{H}_\ell)$ is called the strongest asynchronous ℓ-complete abstraction (SAℓCA) of S, where $\mathcal{X}_\ell := \Pi_{\ell,H}$, $\mathcal{X}_{\ell 0} := \mathcal{X}_\ell$, $H_\ell(x_\ell) := x_\ell(0)$ and*

$$\sigma_\ell := \{(x_\ell, x'_\ell) : x_\ell \cdot x'_\ell(\ell-1) \in \Pi_{\ell+1,H}\} \quad (7)$$

Note that, the set of states $\mathcal{X}_\ell$ consists of output sequences of the original system, that is, if $[y_{\ell,i}] \neq \emptyset$ then $y_{\ell,i} \in \Pi_{\ell,H}$. By construction, and intuitively, it holds that $\mathcal{B}_H(S) \subseteq \mathcal{B}_H(S_\ell)$ [28]. The advantage of knowing the SAℓCA of a system is that we have obtained a finite-state machine containing all the behaviors of the original (possibly infinite-state) system at the expense of accepting the existence of a set of spurious behaviours, that is, behaviours that the abstraction contains but the concrete system doesn't. We provide an example of the construction of the SAℓCA for the bi-dimensional linear system

$$\begin{cases} 3x_{k+1}^{(1)} = x_k^{(1)} + 2x_k^{(2)}, \\ 3x_{k+1}^{(2)} = x_k^{(2)} - 1.8x_k^{(1)}, \end{cases}$$

with $\mathcal{Y} = \{1, 2, 3\}$ in Figure 1. The set $\Pi_{2,H}$ is independent of H for $H \geq 2$; for more details about this system, the interested reader may refer to [7].

Definition 3 (Behavioural inclusion [30])**.** *Consider two systems S_a and S_b with $\mathcal{Y}_a = \mathcal{Y}_b$. S_a is behaviourally included in S_b until horizon H if this holds until horizon H, i.e. $\mathcal{B}_H(S_a) \subseteq \mathcal{B}_H(S_b)$, denoted $S_a \preceq_{\mathcal{B}_H} S_b$.* □

From the discussion above and Definition 3 we state that any dynamical system S defined as in (4) is behaviorally included by its SAℓCA S_ℓ until horizon H. Therefore, any property satisfied by all the behaviours of the abstraction is necessarily satisfied by all the behaviours of the underlying system; the converse is however not true.

IV. Data-Driven Abstractions

From here on for the probability space $(\Delta, \mathcal{F}, \mathbb{P})$ we fix the sample space to be $\Delta = \mathcal{X}$ where $\mathcal{X}$ is a compact subset of $\mathbb{R}^{n_x}$ and $\mathbb{P} = \mathbb{P}_x$, where $\mathbb{P}_x$ represents a probability distribution over the domain $\mathcal{X}$, which might be known or unknown. As previously mentioned, we assume to collect the

[2] Originally in [27] the authors construct the SAℓCA using *infinite* length behaviours, but under the assumption that $\mathcal{X}_0 = \mathcal{X}$ it is possible to use finite length behaviours equivalently.

behaviours of (4) by sampling initial conditions. In practice, we interact with the following random object

$$\Sigma_r(x) := \begin{cases} x_{k+1} = f(x_k), \\ y_k = h(x_k), \\ x_0 \sim \mathbb{P}_x. \end{cases} \tag{8}$$

Let us denote by S_r the equivalent representation of the system above as TS. Hence, the behaviour $\mathcal{B}_H(S_r(x_0))$ represents a RV originating from $\mathbb{P}_x$; we denote by $b_H \in \mathcal{B}_H(S_r)$ the set of all behaviours of S_r with strictly positive probability measure. Consider a sequence of N i.i.d. RVs $x_{0,1}, ..., x_{0,N} \sim \mathbb{P}_x$ and let $\omega_i : \mathcal{X} \to \mathcal{Y}^H$ be defined as $\omega_i := \mathcal{B}_H(S_r(x_{0,i}))$. We sample N initial conditions in the dynamical system, and consider the resulting H-long behaviours displayed by S_r, denoted by $\hat{\mathcal{I}} := \Lbag \mathcal{B}_H(S_r(\hat{x}_{0,i})) \Rbag_{i=1}^N = \Lbag \hat{\omega}_i \Rbag_{i=1}^N$. Now, let us define a data-driven version of (6) for finite behaviours using this set

$$\hat{\Pi}_{\ell,H} := \bigcup_{\hat{\omega} \in \hat{\mathcal{I}}} \bigcup_{k \in [0,H-1]} \hat{\omega}|_{[k-\ell+1,k]}, \tag{9}$$

where the hat remarks that this set is derived from samples.

Definition 4 (Data-driven SAℓCA). *Given $\hat{\Pi}_{\ell+1,H}$ the TS $\hat{S}_\ell := (\hat{\mathcal{X}}_\ell, \hat{\mathcal{X}}_{\ell 0}, \hat{\sigma}_\ell, \mathcal{Y}, \mathcal{H}_\ell)$ is called the* data-driven *(strongest asynchronous) ℓ-complete abstraction (SAℓCA) of S_r, where $\hat{\mathcal{X}}_\ell := \hat{\Pi}_{\ell,H}$, $\hat{\mathcal{X}}_{\ell 0} := \hat{\Pi}_{\ell,H}$, $H_\ell(x_\ell) := x_\ell(0)$*

$$\hat{\sigma}_\ell := \{(x_\ell, x'_\ell) : x_\ell \cdot x'_\ell(\ell - 1) \in \hat{\Pi}_{\ell+1,H}\} \tag{10}$$

Remark 1. *To distinguish between the data-driven SAℓCA viewed as a function of the bag of RVs $\mathcal{I} := \Lbag \mathcal{B}_H(S_r(x_{0,i})) \Rbag_{i=1}^N = \Lbag \omega_i \Rbag_{i=1}^N$ or viewed as a function of the bag of realizations $\hat{\mathcal{I}} = \Lbag \hat{\omega}_i \Rbag_{i=1}^N$ we use the notation $\hat{S}_\ell(\mathcal{I})$ and $\hat{S}_\ell(\hat{\mathcal{I}})$ respectively.*

By comparing (6) and (9) it is easy to see that $\hat{\Pi}_{\ell+1,H} \subseteq \Pi_{\ell+1,H}$. If $\hat{\Pi}_{\ell+1,H} = \Pi_{\ell+1,H}$ the data-driven SAℓCA would be identical to the "true" SAℓCA (modulo zero-measure behaviours), and as such we could conclude that we have obtained a data-driven abstraction which behaviorally includes the original system. However, in general, it holds that $\hat{\Pi}_{\ell+1,H} \subset \Pi_{\ell+1,H}$. Recall that the set of all $[y_{\ell+1,i}]$ forms a partition of the domain, and thus so does $\Pi_{\ell+1,H}$. Therefore, unless every $[y_{\ell+1,i}]$ has been visited at least once by some of the state trajectories initialized at one of the N initial conditions $\Lbag \hat{x}_{0,i} \Rbag_{i=1}^N$, the set $\hat{\Pi}_{\ell+1,H}$ will not be equal to $\Pi_{\ell+1,H}$. Moreover, we have no way of knowing the missing subsequences $\Pi_{\ell+1,H} \setminus \hat{\Pi}_{\ell+1,H}$. However, we show that we can use either the scenario approach or conformal prediction to upper bound the probability measure of the equivalence classes of the sequences belonging to $\Pi_{\ell+1,H} \setminus \hat{\Pi}_{\ell+1,H}$.

For this reason, we provide a generalization of Definition 5 which bridges the random nature of our abstractions with the formal verification of specifications on the original system.

Definition 5 (Probabilistic Behavioural Inclusion). *Consider a TS S_a, a sequence of $N+1$ i.i.d. RVs $x_{0,0}, ..., x_{0,N} \sim \mu$, let $\omega_i : \mathcal{X} \to \mathcal{Y}^H$ be defined as $\omega_i := \mathcal{B}_H(S_a(x_{0,i}))$ and define $\mathcal{I} = \Lbag \omega_i \Rbag_{i=1}^N$. Let $S_b(\mathcal{I})$ be a TS as per Definition 4, with $\mathcal{Y}_a = \mathcal{Y}_b$. We say that S_a is behaviourally included in S_b with probability greater or equal than $1-\epsilon$ until horizon H with respect to μ if it holds:*

$$\begin{gathered} \mu\left[\mathfrak{B}(S_a, S_b(\mathcal{I}))\right] \geq 1 - \epsilon, \\ \mathfrak{B}(S_a, S_b(\mathcal{I})) := \mathcal{B}_H(S_a(x_{0,0})) \subseteq \mathcal{B}_H(S_b(\mathcal{I})), \end{gathered} \tag{11}$$

where $S_a(x_{0,0})$ denotes the internal behaviour of system S_a starting from x_0.

The characterisation provided by (11) describes the behaviours emerging from system (8) which are in fact RVs. Thus, the probabilistic behavioural inclusion does not merely 'count' the behaviours, but rather weights them by their associated probability. In other words, relation (11) defines, or rather provides an upper bound, the maximum probability mass of unseen (i.e. unpredictable) behaviours by ϵ. Considering the deterministic behavioural inclusion, all behaviours in S_a ought to lie within $\mathcal{B}(S_b)$. If, instead, two systems satisfy the probabilistic behavioural inclusion, the total sum of the probability assigned by μ to the behaviours that do not lie within $\mathcal{B}(S_b)$ should be smaller than ϵ.

Ultimately, we aim at relating the concrete system S_r with its data-driven SAℓCA through a probabilistic behavioural inclusion. By constructing an abstraction based on the collected behaviours of a system, we are implicitly asking how predictable a system is based on its past realizations. We employ two different techniques to provide guarantees about relation (11), outlined in the following.

A. Scenario Theory Guarantees

Let us sample N i.i.d. initial conditions in the dynamical system, and consider the resulting H-long behaviours displayed by S_r, denoted by $\hat{\mathcal{I}} := \Lbag \mathcal{B}_H(S_r(\hat{x}_{0,i})) \Rbag_{i=1}^N = \Lbag \hat{\omega}_i \Rbag_{i=1}^N$. We construct the data-driven SAℓCA as described in the previous section. We are now interested in using the scenario approach to show that the concrete system is behaviorally included in such an abstraction with a probability greater or equal than some constant. To define the scenario program, we encode every collected H-long behaviour as a binary vector where every entry encodes whether one (out of the possible $|\mathcal{Y}|^{\ell+1}$) $\ell+1$ sequences is exhibited by the H-long behaviour, as per [7]. Let $y_{\ell+1,j} \in \mathcal{Y}^{\ell+1}$ for $j \in \{1, ..., |\mathcal{Y}|^{\ell+1}\}$ be a sequence of length $\ell+1$, and define $v : \mathcal{Y}^H \to \{0,1\}^{|\mathcal{Y}|^{\ell+1}}$ to be the map encoding an H-sequence as a binary vector whose j-th entry is equal to 1 if $y_{\ell+1,j}$ appears in $\mathcal{B}_H(S_r(\hat{x}_{0,i}))$. Formally

$$v^{(j)}(\hat{w}_i) := \begin{cases} 1 \text{ if } \hat{\omega}_i \models \Diamond \mathrm{y}_{\ell+1,j}, \\ 0 \text{ else }, \end{cases}$$

for $j \in \{1, ..., |\mathcal{Y}|^{\ell+1}\}$. We define the scenario program as

$$\begin{aligned} \min_{\theta \in \Theta} \quad & \mathbf{1}^{\mathrm{T}}_{|\mathcal{Y}|^{\ell+1}} \cdot \theta \\ s.t. \quad & (\theta - v(\hat{w}_i)) \geq 0, \quad i = 1, \ldots, N, \end{aligned} \tag{12}$$

where $\mathbf{1}_{|\mathcal{Y}|^{\ell+1}}$ is a column vector of ones, and $\theta \in \mathbb{R}^{|\mathcal{Y}|^{\ell+1}}$. It can be shown [7] that the solution θ^*_N represents which

$(\ell+1)$-sequences were witnessed within the sample set $\hat{\mathcal{I}}$, and that the complexity s_N^* is equal to the cardinality of the smallest subset of the N collected trajectories that includes all the $\ell+1$-sequences. After setting a desired confidence parameter β, we employ Theorem 1, in order to provide an upper probability bound $\epsilon(s_N^*, \beta, N)$. This value sets an upper limit to the violation of the scenario constraints, i.e. the probability of witnessing an unseen $\ell+1$-sequence. We can now state the following proposition [7].

Proposition 1. *Consider a confidence β, a sequence of $N+1$ i.i.d. RVs $x_{0,0}, ..., x_{0,N} \sim \mu$, let $\omega_i : \mathcal{X} \to \mathcal{Y}^H$ be defined as in Definition 5 and $\hat{\mathcal{I}} = \wr \hat{\omega}_i \wr_{i=1}^N$. It holds that*

$$\mathbb{P}_x^N \left[\; \mathbb{P}_x \left[\mathfrak{B}(S_r, \hat{\mathcal{S}}_\ell(\hat{\mathcal{I}})) \right] \geq 1 - \epsilon(s_N^*, N, \beta) \right] \geq 1 - \beta. \tag{13}$$

In practical terms, after the construction of $\hat{\mathcal{S}}_\ell$, the probability of sampling a new (initial condition leading to a) behaviour that is not included in its behaviours is bounded by ϵ, with confidence β. According to Definition 5, with confidence not smaller than $1-\beta$, S_r is behaviorally included in $\hat{S}_\ell$ with probability greater or equal than $1-\epsilon$ until horizon H with respect to the distribution $\mathbb{P}_x$, as the probability is related to the next, single, sample.

B. Conformal Prediction Guarantees

Consider again the set $\hat{\mathcal{I}}$. We split the data set into two parts, denoted $\hat{\mathcal{I}}_{\text{train}}$ of cardinality M and $\hat{\mathcal{I}}_{\text{cal}}$ of cardinality $N-M$; we use the former to construct the set $\hat{\Pi}_{\ell+1,H}$ according to (9) and derive the data-driven SAℓCA $\hat{S}_\ell$ as shown in Definition 4 and the latter to compute the nonconformity scores. According to Section II-C for every $\hat{\omega}_i \in \hat{\mathcal{I}}_{\text{cal}}$ we define the following nonconformity score

$$R_i := A(\hat{\mathcal{I}}_{\text{train}}, \hat{\omega}_i) := \begin{cases} 0 \text{ if } \hat{\omega}_i \in \mathcal{B}_H(\hat{S}_\ell(\hat{\mathcal{I}}_{\text{train}})), \\ 1 \text{ else.} \end{cases} \tag{14}$$

Without loss of generality let the R_i's be ordered in non-decreasing order, we compute the smallest δ that returns a nonconformity score of 0, i.e. $\delta_{\min} := \min\{\delta : p = M + \lceil (N-M+1)(1-\delta) \rceil, R_p = 0\}$. We conclude that

$$\mathbb{P}_x^{N+1}[R_0 \leq R_p] \geq 1 - \delta_{\min}, \tag{15}$$

and, since R_p is 0, we have obtained a bound on the probability that the next initial condition will generate a behaviour that is already captured by the SAℓCA $\hat{S}_\ell$.

Proposition 2. *Consider a sequence of $N+1$ i.i.d. RVs $x_{0,0}, ..., x_{0,N} \sim \mu$, let $\omega_i : \mathcal{X} \to \mathcal{Y}^H$ and $\mathcal{I}$ be defined as in Definition 5. Then it holds that*

$$\mathbb{P}_x^{N+1} \left[\mathfrak{B}(S_r, \hat{\mathcal{S}}_\ell(\mathcal{I})) \right] \geq 1 - \delta_{\min}. \tag{16}$$

Proof. The proof follows from (15). □

In other words, the probability that the behaviour arising from $x_{0,0}$ is not included in the behaviours of $\hat{\mathcal{S}}_\ell$ is bounded by $\delta_{\min}$. With respect to Definition 5, S_r is behaviorally included in $\hat{S}_\ell$ with probability greater or equal than $1-\delta_{\min}$ until horizon H with respect to the distribution $\mathbb{P}_x^{N+1}$, as the probability is defined on the whole sequence of RVs.

$\epsilon(0.1)$	$\epsilon(10^{-2})$	$\epsilon(10^{-3})$	$\epsilon(10^{-6})$	$\epsilon(10^{-9})$	δ	$\hat{\delta}$
5.8	9.1	11.8	19.8	27.3	5.9	1.6
34.1	39.4	44.0	55.9	66.4	3.9	1.3

TABLE I: Value of ϵ for various values of β (in brackets), δ and the empirical $\hat{\delta}$ – for the Fighter F-16 benchmark (top) and the TCL (bottom). Values are multiplied by 10^{-4}.

C. Discussion of SA and CP guarantees

Let us discuss and highlight the critical differences between the guarantees that we derive using these two distinct approaches. Recalling the scenario approach guarantees in (13), we notice that the inner probability layer can be derived as a conditional statement on the sequence of RVs $x_{0,0}, ..., x_{0,N}$, i.e. we rewrite the inner probability as

$$\mathbb{P}_x \left[\mathfrak{B}(S_r, \hat{\mathcal{S}}_\ell(\hat{\mathcal{I}})) \right] = \mathbb{P}_x^{N+1} \left[\mathfrak{B}(S_r, \hat{\mathcal{S}}_\ell(\mathcal{I})) \mid \mathcal{I} = \hat{\mathcal{I}} \right] \tag{17}$$

whereas the outer probability layer is defined on the product space $\mathcal{X}^N$ and defines a lower bound on the probability of drawing samples $\hat{x}_{0,1}, ..., \hat{x}_{0,N}$ such that (17) holds. In contrast, conformal prediction returns a bound directly on the *joint probability* (16) defined by the RVs $x_{0,0}, ..., x_{0,N}$. In [32] the author establishes how to derive PAC-type guarantees for conformal prediction, obtaining a validity result of split CP conditional to the calibration set. Further, in [18], the authors establish a link between a particular formulation of scenario optimization (not applicable in our case) and conformal prediction. We plan to further explore the details and connections between the two fields.

V. Experimental Evaluation

F-16 Fighter Jet. We employ the F-16 model from [14], providing an accurate 13-dimensional nonlinear representation of a fighter jet. We collect $N = 10^4$ trajectories, with randomness in the initial conditions and collect altitude data. We are interested in verifying that the altitude should always be greater than 3575 feet; the domain $[3575, 3625]$ is divided into 20 partitions. To construct the SAℓCA with $\ell = 4$ we split the trajectories into sequences of length 5, resulting in 52 different sequences. Let us first consider the scenario guarantees: by Theorem 1 we obtain the bound ϵ varying the value of β, as reported in Table I. Next, we tackle the problem from the CP outlook: we split the dataset into a training and calibration sets, and we compute the nonconformity scores R_i as per (14) on a calibration dataset, composed of $M = N/2$ samples. We pick the highest index p (see (3)) such that the non-conformity score is 0, resulting in a value of $\delta_{\min} = 5.9 \cdot 10^{-4}$. The confidence β plays a significant role in the final values of ϵ: setting $\beta = 10^{-1}$ returns $\epsilon \approx \delta$, whilst as β decreases towards more typical values (10^{-6}, 10^{-9}), the ϵ bound deteriorates. To validate the CP results, we sample an additional $5 \cdot 10^4$ initial conditions, and compute the empirical bound $\hat{\delta}$ by counting the number of sequences that violate the nonconformity score condition.

The resulting data-driven abstraction certifies that the unsafe state (representing the plane altitude exiting the domain) is unreachable from the initial states, thus satisfying the safety property, within the probability bounds of Table I.

Thermostatically Controlled Load. We consider a thermostatically controlled load (TCL) benchmark [16]. A (simple) deterministic room temperature evolution model is

$$\theta_{k+1} = a \cdot \theta_k + (1-a) \cdot (\theta_{amb} - m_k \cdot T_r \cdot P), \qquad (18)$$

where a, T_r and P are thermal parameters, θ_k and θ_{amb} are respectively the room temperature at time k and the external temperature, m_k is a binary control input (ON or OFF) designed to maintain the temperature at $20 \pm 0.5°C$. The domain is $\mathcal{D} = [19.25, 20.75]$, partitioned into 20 regions.

We fix $\ell = 4$ and $H = 50$, we sample $N = 10^4$ initial conditions uniformly, obtaining 171 5-sequences. We report in Table I ϵ and β (for the SA) along with the minimum value of δ (for the CP), validated empirically by sampling additionally $5 \cdot 10^4$ initial conditions. We verify on the SAℓCA that the system temperature remains in the prescribed bounds, with the aforementioned probability bounds.

VI. Conclusions

We have presented a method to construct a data-driven abstraction of a deterministic system with unknown dynamics. The abstraction can be used to verify safety specifications, equipped with specific probability guarantees. Under two different perspectives, the scenario approach and the conformal prediction, we characterise a probabilistic behavioural inclusion relation. These two approaches provide probability guarantees with different interpretations, the former including a conditional probability and a confidence parameter, the latter stemming from joint probability distributions. Future work includes further investigations on the existing links between SA and CP. We plan to extend the provision of guarantees for infinite-horizon properties under the conformal prediction framework, and formulate control synthesis problems.

References

[1] A. N. Angelopoulos and S. Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

[2] T. S. Badings, A. Abate, N. Jansen, D. Parker, H. A. Poonawala, and M. Stoelinga. Sampling-based robust control of autonomous systems with non-gaussian noise. In *Workshops at the Thirty-Sixth AAAI Conference on Artificial Intelligence*, 2022.

[3] V. Balasubramanian, S.-S. Ho, and V. Vovk. *Conformal prediction for reliable machine learning: theory, adaptations and applications*. Newnes, 2014.

[4] L. Bortolussi, F. Cairoli, N. Paoletti, S. A. Smolka, and S. D. Stoller. Neural predictive monitoring. In *Runtime Verification: 19th International Conference, RV 2019, Porto, Portugal, October 8–11, 2019, Proceedings 19*, pages 129–147. Springer, 2019.

[5] F. Cairoli, N. Paoletti, and L. Bortolussi. Conformal quantitative predictive monitoring of stl requirements for stochastic processes. In *Proceedings of HSCC*, pages 1–11, 2023.

[6] M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[7] R. Coppola, A. Peruffo, and M. Mazo. Data-driven abstractions for verification of linear systems. *IEEE Control Systems Letters*, 7:2737–2742, 2023.

[8] I. Cortés-Ciriano and A. Bender. Concepts and applications of conformal prediction in computational drug discovery. *arXiv preprint arXiv:1908.03569*, 2019.

[9] M. Cubuktepe, N. Jansen, S. Junges, J.-P. Katoen, and U. Topcu. Scenario-based verification of uncertain mdps. In *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, pages 287–305. Springer, 2020.

[10] A. Devonport, A. Saoud, and M. Arcak. Symbolic abstractions from data: A pac learning approach. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 599–604, 2021.

[11] M. Fontana, G. Zeni, and S. Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

[12] S. Garatti and M. C. Campi. The risk of making decisions from data through the lens of the scenario approach. *IFAC-PapersOnLine*, 54(7):607–612, 2021.

[13] N. Hashemi, X. Qin, L. Lindemann, and J. V. Deshmukh. Data-driven reachability analysis of stochastic dynamical systems with conformal inference. *arXiv preprint arXiv:2309.09187*, 2023.

[14] P. Heidlauf, A. Collins, M. Bolender, and S. Bak. Verification challenges in f-16 ground collision avoidance and other automated maneuvers. In *ARCH@ ADHS*, pages 208–217, 2018.

[15] M. Kazemi, R. Majumdar, M. Salamati, S. Soudjani, and B. Wooding. Data-driven abstraction-based control synthesis. *arXiv preprint arXiv:2206.08069*, 2022.

[16] S. Koch, J. L. Mathieu, D. S. Callaway, et al. Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services. In *Proc. PSCC*, pages 1–7. Citeseer, 2011.

[17] A. Lavaei, S. Soudjani, E. Frazzoli, and M. Zamani. Constructing mdp abstractions using data with formal guarantees. *IEEE Control Systems Letters*, 2022.

[18] A. Lin and S. Bansal. Verification of neural reachable tubes via scenario optimization and conformal prediction. *Proceedings of Machine Learning Research vol vvv*, 1:16, 2024.

[19] L. Lindemann, M. Cleaveland, G. Shim, and G. J. Pappas. Safe planning in dynamic environments using conformal prediction. *IEEE Robotics and Automation Letters*, 2023.

[20] L. Lindemann, X. Qin, J. V. Deshmukh, and G. J. Pappas. Conformal prediction for stl runtime verification. In *Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2023)*, pages 142–153, 2023.

[21] A. Makdesi, A. Girard, and L. Fribourg. Data-driven models of monotone systems. 2022.

[22] A. Muthali, H. Shen, S. Deglurkar, M. H. Lim, R. Roelofs, A. Faust, and C. Tomlin. Multi-agent reachability calibration with conformal prediction. *arXiv preprint arXiv:2304.00432*, 2023.

[23] H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman. Inductive confidence machines for regression. In *Machine learning: ECML 2002: 13th European conference on machine learning Helsinki, Finland, 2002 proceedings 13*, pages 345–356. Springer, 2002.

[24] A. Peruffo and M. Mazo. Data-driven abstractions with probabilistic guarantees for linear petc systems. *IEEE Control Systems Letters*, 2022.

[25] X. Qin, N. Hashemi, L. Lindemann, and J. V. Deshmukh. Conformance testing for stochastic cyber-physical systems. *arXiv preprint arXiv:2308.06474*, 2023.

[26] X. Qin, Y. Xia, A. Zutshi, C. Fan, and J. V. Deshmukh. Statistical verification of cyber-physical systems using surrogate models and conformal inference. In *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*, pages 116–126. IEEE, 2022.

[27] A.-K. Schmuck and J. Raisch. Asynchronous l-complete approximations. *Systems & Control Letters*, 73:67–75, 2014.

[28] A.-K. Schmuck, P. Tabuada, and J. Raisch. Comparing asynchronous l-complete approximations and quotient based abstractions. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 6823–6829. IEEE, 2015.

[29] G. Shafer and V. Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[30] P. Tabuada. *Verification and control of hybrid systems: a symbolic approach*. Springer Science & Business Media, 2009.

[31] A. Tebjou, G. Frehse, et al. Data-driven reachability using christoffel functions and conformal prediction. In *Conformal and Probabilistic Prediction with Applications*, pages 194–213. PMLR, 2023.

[32] V. Vovk. Conditional validity of inductive conformal predictors. In *Asian conference on machine learning*, pages 475–490. PMLR, 2012.

[33] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.