TUM School of Computation, Information and Technology
Professorship of Cyber Physical Systems

Data-Driven Reachability Analysis:
A Contrastive Evaluation of Direct and Indirect Methods for Formal Safety Verification

Ayse Aybüke Ulusarslan

Guided Research - SS 2025

Advisor: M. Sc. Yongkuan Zhang

Supervisor: Prof. Dr.-Ing. Matthias Althoff

Submission: 1. Oktober 2025

Data-Driven Reachability Analysis: A Contrastive Evaluation of Direct and Indirect Methods for Formal Safety Verification

Aybüke Ulusarslan
Technische Universität München
Email: aybuke.ulusarslan@tum.de

***Abstract*—The formal verification of cyber-physical systems demands rigorous state inclusion guarantees using bounded uncertainty. We compare two set-based data-driven approaches for reachability analysis of Linear Time-Invariant systems: the *direct* Data-Driven Reachability Analysis, which provides guaranteed containment bounds by modeling the set of all consistent system dynamics [1], and the *indirect* Reachset-Conformant System Identification, which minimizes conservatism by optimizing uncertainty sets subject to reachset conformance [2]. We evaluate these algorithms under identical, reproducible conditions, with an emphasis on *sample efficiency*. To isolate this axis cleanly, we adopt a unified evaluation protocol in which (i) the process/measurement uncertainty bounds and (ii) the state dimension are kept fixed across methods and experiments. Within this controlled setting, we systematically vary the data budget (trajectory length and number of nominal seeds) and report fidelity (containment) and conservatism (size proxies), along with compute costs. All computations can be reproduced using the code base available at github.com/aiulus/ddra-cedim.**

I. Introduction

The development of highly autonomous cyber-physical systems, such as autonomous vehicles and human-in-the-loop robotics, has created an imperative for formal safety verification. Since no mathematical model can perfectly describe the complexities of real-world dynamics, system identification techniques must incorporate uncertainties to account for discrepancies between the model and the target system [3], [4].

Classical system identification often relies on statistical methodologies, making probabilistic assumptions about the noise and model error. However, for safety-critical applications involving public safety, these probabilistic statements are insufficient, as they cannot provide the absolute guarantees of state inclusion necessary to formally prove that the system provably avoids unsafe sets. This critical requirement motivates a shift toward modeling with bounded uncertainties, a paradigm commonly explored in set-membership estimation (SME) or bounded-error approaches. SME aims to produce a set estimator that guarantees coverage of the ground truth, meaning the true state vector is rigorously contained within the computed set.

Reachability analysis. The central tool common to these bounded-error methodologies is Reachability Analysis: the rigorous computation of the union of all possible trajectories—the reachable set ($\mathcal{R}_k$)—that a system can attain over time, starting from a bounded set of initial conditions. Reachability analysis is foundational to formal safety verification, where safety is certified if the computed reachable set is mathematically disjoint from the predefined unsafe set.

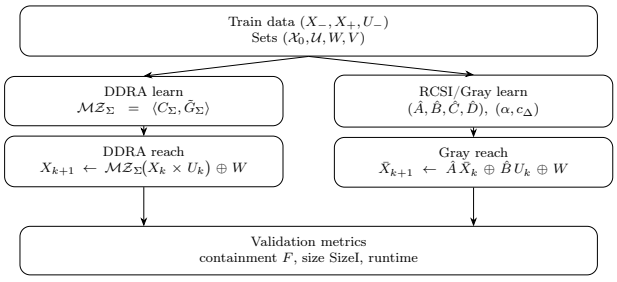

Fig. 1. DDRA / Gray-box RCSI computation pipelines.

All modern set-based approaches seek to reliably compute a guaranteed over-approximation of a system's future states, given bounded uncertainty (e.g., process and measurement noise) and high-level knowledge of the system's class. How the uncertainty of the model itself is handled gives rise to two distinct methodological families:

- **(a) Indirect Approaches** (Model Identification First). These methods first identify an optimal model structure, parameters, and/or uncertainty bounds, and then use the resulting abstract model to perform reachability analysis. This group includes Reachset-Conformant System Identification (RCSI).
- **(b) Direct Approaches** (Set Propagation from Data). These methods bypass the identification of a single model. Instead, they leverage algebraic techniques to characterize the entire set of all possible system models consistent with the available data and a fixed uncertainty bound, and then propagate the reachable set forward using this whole set of systems consistent with the information at hand. This group includes Data-Driven Reachability Analysis (DDRA).

Data-Driven Reachability Analysis. Under bounded uncertainty and persistency of excitation, Data-Driven Reachability Analysis (DDRA) [5] offers containment guarantees: the computed set of feasible system models is guaranteed to contain the target system. This ensures robustness, as the

verification will never lead to a catastrophic false negative in safety certification (proving an unsafe system to be safe). DDRA achieves this by characterizing the full set of consistent models using matrix zonotopes. Conservatism is further reduced by introducing constrained matrix zonotopes and an auxiliary framework for incorporating prior knowledge (side information) about the unknown system. However, this inherent robustness often comes at the cost of conservatism (failing to certify a safe system as safe). If the model uncertainty is unbounded or overly pessimistic, the computed reachable set may become too large, unnecessarily intersecting the unsafe set.

Reachset-Conformant System Identification (RCSI). Developed by Lützow and Althoff, RCSI seeks to identify the minimal model uncertainty bounds (α, c_Δ) and/or model parameters (p) necessary to satisfy reachset conformance. Reachset conformance is the necessary and sufficient property to transfer safety properties from the identified model S_M to the target system S_T, requiring that the reachable set of the model encloses all system measurements $(\mathcal{Y}_k^{(m)}(S_T) \subseteq \hat{\mathcal{Y}}_k^{(m)}(S_M))$. By minimizing the volume of the reachable set subject to this conformance constraint, RCSI is engineered to produce the least conservative model possible for a given architecture. The RCSI framework demonstrates flexibility by addressing identification across various levels of prior knowledge :

- **(a) White-Box Identification.** Model functions are known; only minimal uncertainty bounds (α, c_Δ) are determined, often solved via a scalable Linear Program (LP).
- **(b) Gray-Box Identification.** Model functions contain unknown parameters p; both p and the uncertainty bounds are optimized by solving a linear program.
- **(c) Black-Box Identification.** The entire model function is determined from data using techniques like (reachset-conformant) genetic programming.

Both DDRA and RCSI provide essential, complementary tools for the data-driven verification of cyber-physical systems. Given the ubiquity of Linear Time-Invariant (LTI) systems as exact representations or local approximations in engineering applications, we focus our analysis on linear dynamics. This involves studying the foundational DDRA algorithm (LTI-Reachability) against the Gray-Box methods of RCSI.

Ultimately, claiming one approach is universally superior to the other is impractical. Instead, this work aims to provide a structured comparison of these two rigorous set-based methodologies—one prioritizing maximal certainty (DDRA containment) and the other prioritizing minimal conservatism (RCSI optimization)—to determine their respective suitability for safety verification goals.

II. Preliminaries

We exclusively look at discrete-time LTI systems

$$\Sigma_{\text{i/s/o}} : \begin{cases} \mathrm{x}_{k+1} = A\mathrm{x}_k + B\mathrm{u}_k + w_k \\ \mathrm{y}_k = C\mathrm{x}_k + D\mathrm{u}_k + v_k, \end{cases} \tag{1}$$

where $\mathrm{x} \in \mathbb{R}^n, \mathrm{u} \in \mathbb{R}^m, \mathrm{y} \in \mathbb{R}^p$ denote the state, input, and output vectors respectively. We will annotate specific system instances as tuples of systems matrices: $\Sigma_{\text{i/s/o}} \ni (A_0, B_0, C_0, D_0) \coloneqq \mathcal{S}_0$. We assume bounded process and measurement disturbances $w_k \in W \subset \mathbb{R}^n$ and $v_k \in V \subset \mathbb{R}^p$.

a) Persistency of excitation (PE).: Informally, an input is *persistently exciting* if, over any window of L consecutive time steps, it "shakes" the system in enough independent directions so that no nontrivial linear combination of those L input vectors looks the same. Algebraically, this richness is captured by the full-row-rank condition on the block Hankel matrix built from the input (Def. 1). PE matters in identification because it prevents ambiguity: with PE of order L, the corresponding regression/behavior matrices (e.g., $Z = \begin{bmatrix} X_-^\top & U_-^\top \end{bmatrix}^\top$ in our setting) become well-conditioned or full rank in the noiseless limit, so different system responses to different inputs are distinguishable rather than confounded. As a result, estimation problems are well-posed, uncertainty sets contract with more data (instead of plateauing), and the learned models support reliable reachability computations. In practice, broadband signals (e.g., multi-sine, PRBS, filtered noise) are PE to high order, whereas constant or single-frequency inputs fail PE beyond trivial orders. Throughout, we synthesize input suites with a target PE order L so that the stacked data used by DDRA (to build $\mathcal{M}_\Sigma$) and by RCSI (to enforce conformance) are informative rather than degenerate.

Definition 1 (Block Hankel and PE)**.** *Let $U \in \mathbb{R}^{m \times T}$ be the (concatenated) input sequence. The block Hankel of order L is*

$$\mathcal{H}_L(U) = \begin{bmatrix} U_{:,1} & \cdots & U_{:,T-L+1} \\ \vdots & \ddots & \vdots \\ U_{:,L} & \cdots & U_{:,T} \end{bmatrix} \in \mathbb{R}^{(mL)\times(T-L+1)}.$$

We say U is PE of order L if $\operatorname{rank}\big(\mathcal{H}_L(U)\big) = mL$.

b) Set-based Uncertainty Models: As mentioned earlier, the key modeling assumption in this line of research (safety verification, reachability analysis & related) is one of bounded uncertainty, which makes (convex) sets first class citizens. Within various representation modalities (halfspace- and vertex zonotopes, ellipsoids, intervals ...), zonotopes are preferred due to their desirable dimensional scaling properties.

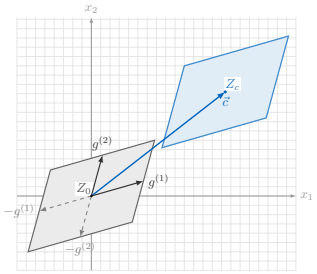

Fig. 2. A two-dimensional zonotope with two generators.

A zonotope is the Minkowski sum of line segments $g^{(i)}$, called the generator vectors, from the origin to $\pm g^{(i)}$, translated by the center vector c. There can be arbitrarily many

generator vectors (heuristically: if we're trying to approximate some nominal set with a zonotope, the more generator vectors we add, the more closely we can approximate the true set, which comes with a runtime trade-off). Formally, we define zonotopes as follows.

Definition 2 (Zonotope)**.** *Given a center $c \in \mathbb{R}^n$ and a generator matrix $G = [g^{(1)}, \dots, g^{(\gamma)}] \in \mathbb{R}^{n\times\gamma}$, the zonotope spanned by (c, G) is*

$$\mathcal{Z} \overset{\text{def.}}{=} \left\{ c + \sum_{i=1}^{\gamma} \beta_i g^{(i)} \;\middle|\; \beta \in [-1,1]^\gamma \right\} = \tag{2}$$

$$c \oplus \bigoplus_{i=1}^{\gamma} \left([-1,1] \cdot g^{(i)} \right) \subset \mathbb{R}^n. \tag{3}$$

We write $\mathcal{Z} = \langle c, G \rangle\rangle$ for short.

Zonotopes are essentially sets of vectors. In reachability analysis, it's often not just states or inputs that carry uncertainty, but system matrices themselves. We can represent (bounded-)uncertain matrices thorugh matrix zonotopes, which can be interpreted as Minkowski sums of line segments in the matrix space $\mathbb{R}^{n\times p}$. We formally define them as follows.

Definition 3. *[1] Given a center matrix $C_{\mathcal{M}} \in \mathbb{R}^{n\times p}$ and $\gamma_{\mathcal{M}}$ generator matrices $G_{\mathcal{M}}^{(j)} \in \mathbb{R}^{n\times p}$, the matrix zonotope is given by:*

$$\mathcal{MZ} \coloneqq \left\{ C_{\mathcal{M}} + \sum_{j=1}^{\gamma_M} \beta^{(j)} G_{\mathcal{M}}^{(j)} \middle| \beta^{(j)} \in [-1,1] \right\}. \tag{4}$$

We use the shorthand $\mathcal{MZ} = \left\langle C_{\mathcal{M}}, \tilde{G}_{\mathcal{M}} \right\rangle$, where $\tilde{G}_{\mathcal{M}} = \left[G_{\mathcal{M}}^{(1)}, \dots, G_{\mathcal{M}}^{(\gamma_M)} \right]$.

c) Conformant Models: Crucially, reachset conformance is sufficient to transfer universally quantified safety properties: if the model's reachable outputs avoid an unsafe set at all times, then the target system's measurements (being contained in those sets) cannot violate the safety property. There are other stronger but less practical notions of conformance, such as trace conformance, which requires that each measured trajectory from the target system coincides with some model trace under the same experiment.

Definition 4. ***(Reachset Conformance).*** *Let $\mathcal{M}$ be a finite test suite comprising n_m experiments with nominal initial–input pairs $\{(x_0^{(m)}, u_{0:k-1}^{(m)})\}_{m=1}^{n_m}$. For a target system $\mathcal{S}_T$ and a model $\mathcal{S}_M$, let $\hat{\mathcal{Y}}_k^{(m)}(\mathcal{S}_M)$ denote the model-predicted reachable output set at time k, and let $\mathcal{Y}_k^{(m)}(\mathcal{S}_T)$ be the set of measured outputs from $\mathcal{S}_T$ under the same pair. We write*

$$\mathcal{S}_T \;\; \text{conf}_R \;\; \mathcal{S}_M \quad :\Longleftrightarrow$$
$$\forall m \in [n_m],\ \forall k \in [n_k - 1] :\ \mathcal{Y}_k^{(m)}(\mathcal{S}_T) \subseteq \hat{\mathcal{Y}}_k^{(m)}(\mathcal{S}_M).$$

The identification goal is to minimize conservatism (size of predicted sets) subject to conformance.

Definition 5 (Reachset-conformant identification)**.** *Let q collect the unknowns (see cases below), and let $r(\cdot)$ be a monotone size functional (e.g., volume or sum of support widths). Given weights $w_k \geq 0$, solve*

$$q^\star \in \arg\min_q \sum_{m=1}^{n_m} \sum_{k=0}^{n_k-1} w_k \, r\big(\hat{\mathcal{Y}}_k^{(m)}(\mathcal{S}_M(q))\big) \quad \textit{s.t.} \tag{5}$$

$$\forall m, k, s \in [n_m] \times [n_k - 1] \times [n_S] :\ y_k^{(m,s)} \in \hat{\mathcal{Y}}_k^{(m)}(\mathcal{S}_M(q)),$$

where $y_k^{(m,s)}$ are samples from $\mathcal{S}_T$ gathered for test m at time k (sample index s). Depending on available a priori information:

- ***(a) White-box:*** *Dynamics are known; uncertainty lies in $\mathcal{X}_0^{(m)} = \langle c_x + c_{\Delta,x}, G_x \operatorname{diag}(\alpha_x) \rangle$ and $\mathcal{U}_i^{(m)} = \langle c_u + c_{\Delta,u}, G_u \operatorname{diag}(\alpha_u) \rangle$. Let $q = (\alpha_x, \alpha_u, c_{\Delta,x}, c_{\Delta,u})$, with $G_x \in \mathbb{R}^{n\times d_x}$, $G_u \in \mathbb{R}^{m\times d_u}$.*
- ***(b) Gray-box:*** *A parametric family $\mathcal{F}_p$ is known (e.g., LTI with unknown (A, B)). Then $q = (\alpha_x, \alpha_u, c_{\Delta,x}, c_{\Delta,u}, p)$ (e.g., $p = (A, B)$).*
- ***(c) Black-box:*** *$\mathcal{F}$ is unknown. One may (i) search over $\mathcal{F}$ (e.g., via genetic programming) and (ii) perform gray-box identification of p within the selected family.*

d) Data-Driven Reachability Analysis: Given input-state trajectories $D = (U_-, X)$ of a system $\mathcal{S} = (A, B, I_{n\times n}, \mathbf{0})$ such that $\begin{bmatrix} X_-^\top & U_-^\top \end{bmatrix}^\top$ has a full rank, they show that the matrix zonotope

$$\mathcal{M}_\Sigma = (X_+ \;\; -\mathcal{M}_w) \begin{bmatrix} X_- \\ U_- \end{bmatrix}^\dagger$$

contains all matrices $\begin{bmatrix} A & B \end{bmatrix}$ that are consistent with the data $D = (U_-, X)$,

$$X_- = \left[\mathrm{x}^{(1)}(0) \dots \mathrm{x}^{(1)}(T_1 - 1) \dots \mathrm{x}^{(K)}(0) \;\; \dots \mathrm{x}^{(K)}(T_K - 1) \right]$$
$$X = \left[\mathrm{x}^{(1)}(0) \dots \mathrm{x}^{(1)}(T_1) \dots \mathrm{x}^{(K)}(0) \dots \mathrm{x}^{(K)}(T_K) \right] \tag{6}$$
$$U_- = \left[\mathrm{u}^{(1)}(0) \dots \mathrm{u}^{(1)}(T_1 - 1) \dots \mathrm{u}^{(K)}(0) \dots \mathrm{u}^{(K)}(T_K - 1) \right],$$

and the noise bound, where $\mathcal{M}_w = \left\langle C_{\mathcal{M}_w}, \tilde{G}_{\mathcal{M}_w} \right\rangle$ is the matrix zonotope resulting from the concatenation of multiple noise zonotopes $\mathcal{Z}_w$.

When only *noisy* states $y_k = x_k + v_k$ are available, standard DDRA—which regresses a model set from (X_-, X_+, U_-)—is no longer directly applicable because the transported noise term $V_+ - AV_-$ leaks into X_+ and X_-. The measurement-aware variant replaces the noiseless stacks by (Y_-, Y_+) and *explicitly compensates* the transport of V in one of two ways: (i) if a bound on $\hat{O} \overset{\text{def.}}{=} V_+ - AV_-$ is known, subtract it during regression to obtain a model set $\mathcal{M}_{\tilde{\Sigma}}$; (ii) if such a bound is unavailable, first fit a nominal $\tilde{M}$ on (Y_-, Y_+, U_-), then learn a one-step residual zonotope $\mathcal{Z}_{AV}$ that absorbs both AV_- and model mismatch. In both cases, reachability proceeds as in DDRA but with these modified objects. *Key difference to standard DDRA:* the regression uses (Y_-, Y_+) instead of

(X_-, X_+) and the propagation is inflated by measurement-induced residuals (either via $\mathcal{M}_{\tilde{\Sigma}}$ built with $\mathcal{M}_o$ or via $\mathcal{Z}_{AV}$ and $\mathcal{Z}_v$), while the same PE/rank conditions are checked on $\begin{bmatrix} Y_-^\top & U_-^\top \end{bmatrix}^\top$.

Algorithm 1 Measured-state DDRA

Inputs : (U_-, Y); $\mathcal{X}_0$, $\mathcal{Z}_w$; optional $\mathcal{Z}_v$; either a matrix-zonotope bound $\mathcal{M}_o \ni \hat{O}$ or none

Outputs: Reachable sets $\hat{\mathcal{R}}^{\mathrm{m}}_{1:k}$

If $\mathcal{M}_o$ is available (subtractive regression):

`// model set from noisy stacks`

$\mathcal{M}_{\tilde{\Sigma}} \leftarrow (Y_+ - \mathcal{M}_o - \mathcal{M}_w) \begin{bmatrix} Y_- \\ U_- \end{bmatrix}^\dagger$

$\hat{\mathcal{R}}^{\mathrm{m}}_0 \leftarrow \mathcal{X}_0$

for for $t = 0{:}k-1$: **do**

  $\hat{\mathcal{R}}^{\mathrm{m}}_{t+1} \leftarrow \mathcal{M}_{\tilde{\Sigma}}\big(\hat{\mathcal{R}}^{\mathrm{m}}_t \times \mathcal{U}_t\big) \oplus \mathcal{Z}_w.$

Else (data-driven residual):

`// nominal LS on noisy stacks`

$\tilde{M} \leftarrow (Y_+ - C_{\mathcal{M}_v} - C_{\mathcal{M}_w}) \begin{bmatrix} Y_- \\ U_- \end{bmatrix}^\dagger$

`// Compute residual columns`

$e_j \leftarrow Y_{+,j} - \tilde{M} \begin{bmatrix} Y_{-,j} \\ U_{-,j} \end{bmatrix}.$

$\mathcal{Z}_{AV} \leftarrow \mathrm{interval\_hull}(\{e_j\}) \ominus \mathcal{Z}_w \ominus \mathcal{Z}_v$ $\hat{\mathcal{R}}^{\mathrm{m}}_0 \leftarrow \mathcal{X}_0$

for $t = 0{:}k-1$ **do**

  $\hat{\mathcal{R}}^{\mathrm{m}}_{t+1} \leftarrow \tilde{M}\big((\hat{\mathcal{R}}^{\mathrm{m}}_t \oplus \mathcal{Z}_v) \times \mathcal{U}_t\big) \oplus \mathcal{Z}_{AV} \oplus \mathcal{Z}_w.$

e) Reachability Analysis (discrete-time).: Given bounded sets for the initial condition $\mathcal{X}_0 \subset \mathbb{R}^n$, inputs $\mathcal{U}_k \subset \mathbb{R}^m$, process noise $W \subset \mathbb{R}^n$ and measurement noise $V \subset \mathbb{R}^p$, the (state) reachable set at step k is the set of all states that can arise under *any* admissible choices of $(\mathrm{x}_0, \mathrm{u}_{0:k-1}, u_{k-1}, w_{0:k-1}, v_{0:k-1})$. For the LTI model from (1), this set is computed by unrolling the dynamics or, more conveniently, by the recursion

$$\mathcal{R}_0 = \mathcal{X}_0,$$
$$\mathcal{R}_{k+1} = A\mathcal{R}_k \oplus B\mathcal{U}_k \oplus W, \qquad k \geq 0,$$

with the output reachable set

$$\hat{\mathcal{Y}}_k = C\mathcal{R}_k \oplus D\mathcal{U}_k \oplus V. \tag{7}$$

Here $\oplus$ denotes the Minkowski sum and linear images like $A\mathcal{R}_k = \{Ax \mid x \in \mathcal{R}_k\}$ implement the dynamics. This set-based propagation is the core of reachability analysis: compute $\mathcal{R}_k$ (and $\hat{\mathcal{Y}}_k$) so that they *contain* all behaviors consistent with the modeling assumptions. Exact reachable sets are available only for special classes; in general, one uses *sound* over-approximations and certifies safety if the over-approximate reachable sets are disjoint from the unsafe set.

III. Experiments

The purpose of our study is to compare *Data-Driven Reachability Analysis* and Gray-box *Reachset-Conformance System Identification*– respectively representing direct and indirect reachability analysis methods– under controlled, reproducible conditions. We focus on the question of *sample efficiency*, i.e. how performance scales with data budget. Throughout, we evaluate on identical train/validation test suites and enforce a unified noise policy.

A. Computation Pipeline

Each experiment proceeds as:

1) **Data suite construction.** Fix the discrete-time LTI ground truth $\mathcal{S}_0 = (A, B, C, D)$, sampling time Δt, initial set $\mathcal{X}_0 = \langle c_{\mathrm{x}_0}, G_{c_{\mathrm{x}_0}} \rangle$, and an input set $\mathcal{U} = \langle c_u, G_u \rangle$. Generate n_m *nominal* input sequences $\left\{ \left(u_1^{(*,i)}, \ldots, u_{n_k}^{(*,i)} \right) : i \in [n_m] \right\}$ with PE order L and amplitude s. For each nominal input sequence $u_{1:n_k}^{(*,i)}$, create n_s stochastic samples by drawing $\mathrm{x}_0^{(m,s)} \in \mathcal{X}_0$ and per-step input perturbations $\delta u_k^{(m,s)} \in \mathcal{U}$. Validation suites have separate budgets $\left(n_m^{\mathrm{val}}, n_k^{\mathrm{val}}, n_s^{\mathrm{val}}\right)$.
2) **Unified noise policy.** With process uncertainty enabled, define the state-space noise as the pushforward of $\mathcal{U}$ through B:

$$\mathcal{Z}_w \overset{\text{def.}}{=} \langle 0, BG_u \rangle \subset \mathbb{R}^n.$$

For models with disturbance channels and map E, choose a conservative disturbance set E_d with $E_d \supseteq \mathcal{Z}_w$.
3) **DDRA learning.** Build the stacked TRAIN triples (X_-, X, U_-) with $N = n_m n_s n_k$, and regress a *matrix zonotope* over $[A\ B]$:

$$\mathcal{M}_\Sigma \overset{\text{def.}}{=} (X^+ - \mathcal{M}_w)\, Z^\dagger, \quad Z = \begin{bmatrix} X^- \\ U^- \end{bmatrix},$$

where $\mathcal{M}_w$ is the matrix zonotope induced by $\mathcal{Z}_w$ across the N columns. If Z is rank-deficient and ridge is enabled, use $Z^\dagger = Z^\top (ZZ^\top + \lambda I)^{-1}$ and widen either $\mathcal{M}_\Sigma$ or $\mathcal{Z}_w$ accordingly. *Measured-state* DDRA instead learns a center AB_c and one-term residual zonotope AV_{one} that bounds linearization/mismatch plus W and (optional) V.
4) **RCSI/Gray identification.** Solve a conformance program that finds model parameters p (gray) and minimal uncertainty scales (α, c_Δ) s.t. the model's reachable output sets contain all TRAIN measurements:

$$\forall (m, k, s): \quad y_k^{(m,s)} \in \widehat{\mathcal{Y}}_k^{(m)}(\mathcal{S}_M(p), \alpha, c_\Delta),$$

while minimizing a size functional of $\widehat{\mathcal{Y}}$ (e.g. interval width sum). For linear models we identify a new LTI $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ and uncertainty.
In our application of gray-box LTI, q collects $\{A, B\}$ and uncertainty scales/centers; reachability is carried out with the generalized linear output (GLO) approximation along nominal traces, with linearization errors bounded by sets $\widehat{\mathcal{E}}_k^{(m)}$:

$$\hat{\mathcal{Y}}_k^{(m)}(S_{\mathrm{M}}) \;=\; \bar{\mathcal{Y}}_k^{(m)}(S_{\mathrm{M}}) \oplus \widehat{\mathcal{E}}_k^{(m)},$$

where $\bar{\mathcal{Y}}_k^{(m)}$ is the GLO reach. Optimization

$$\min_q \quad \sum_{m,k} r\big(\hat{\mathcal{Y}}_k^{(m)}(S_{\mathrm{M}}(q))\big) \quad \text{s.t.}$$
$$y_k^{(m,s)} \in \hat{\mathcal{Y}}_k^{(m)}(S_{\mathrm{M}}(q)),\ \forall m,k,s.$$

is solved by alternating nonlinear updates of $p=(A,B)$ with LP updates of the uncertainty (α, c_Δ)

5) **Validation & metrics.** On VAL, both pipelines produce per-time pre-update output sets

$$\mathcal{Y}_k = C\mathcal{R}_{k-1} \oplus D\mathcal{U}_k,$$

where $\mathcal{R}_k$ is propagated by DDRA

$$\mathcal{R}_k \supseteq \mathcal{M}_\Sigma\left(\mathcal{R}_{k-1} \times \mathcal{U}_k\right) \oplus \mathcal{Z}_W$$

or, in the measured-state variant,

$$\mathcal{R}_k \supseteq AB_c\left(\mathcal{R}^*_{k-1} \times \mathcal{U}_k\right) \oplus AV_{\text{one}},$$

with $\mathcal{R}^*_{k-1} = \begin{cases} \mathcal{R}_{k-1} & \text{if } k=1 \\ \mathcal{R}_{k-1} \oplus V & \text{if } k>1, \end{cases}$ and the gray-box variant of RCSI. All metrics are computed on the validation suite $\{(x_0^{(j)}, u_{0:k-1}^{(j)}, y_{0:k}^{(j)})\}_{j=1}^N$ with $N = n_m^{\text{val}} n_s^{\text{val}}$.

a) **Containment (fidelity).** A scalar coverage score in $[0,100]$:

$$F = \frac{100}{N\, n_k^{\text{val}}} \sum_{j=1}^{N} \sum_{k=1}^{n_k^{\text{val}}} \mathbf{1}\left\{ y_k^{(j)} \in \text{interval\_hull}\big(\hat{\mathcal{Y}}_k^{(\cdot)}\big) \right\},$$

b) **Conservatism (size proxy).** Let $I_k = \text{interval\_hull}(\hat{\mathcal{Y}}_k)$ with componentwise widths $w_{k,j} = \overline{I}_{k,j} - \underline{I}_{k,j}$. Define the mean $s_k = \frac{1}{p}\sum_{j=1}^p w_{k,j}$ and report the global average $\text{SizeI} = \frac{1}{N\, n_k^{\text{val}}} \sum_{j,k} s_k^{(j)}$ and the per-step average $\bar{s}_k = \frac{1}{N}\sum_j s_k^{(j)}$.

Algorithm 2 Unified Evaluation Pipeline

Inputs : $(x_0^{(b)}, u_{0:k-1}^{(b)}, y_{1:k}^{(b)})$; sets $(\mathcal{X}_0, \mathcal{U}, W)$; PE order L_{eff}
Outputs: Containment and size metrics for DDRA and Gray

DDRA: $X \leftarrow \mathcal{X}_0$
for $k = 1$ **to** n_k^{val} **do**
  $\hat{\mathcal{Y}}_k^{\text{DDRA}} \leftarrow C\hat{\mathcal{X}}_k \oplus D\mathcal{U}_k^{(b)}$
  $\hat{\mathcal{X}}_{k+1} \leftarrow \mathcal{M}_\Sigma\big(\hat{\mathcal{X}}_k \times \mathcal{U}_k^{(b)}\big) \oplus W$
Gray: Compute system model S_{M}, and uncertainty parameters (α, c_Δ), return reachable sets used for conformance checking
Evaluate: Compute global averages and per-step means;
return

B. Sample Efficiency

1) Trajectory Length: We evaluate the dependence of both algorithms on data quantity by varying the **trajectory length** (n_k) across the sweep grid $\mathbf{n_k} = 4:4:100$, while fixing the state dimension to $D = 2$, the number of training trajectories $n_m = 10$, and the samples per trajectory $n_s = 5$ (full configuration details can be found in the Appendix). This experiment assesses the required temporal extent of data needed to reliably constrain the model sets and propagate uncertainty.

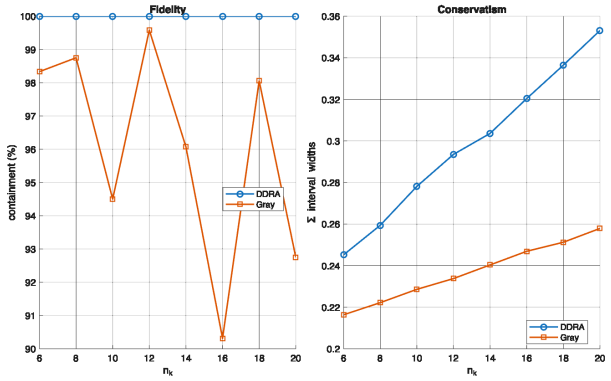

Fig. 3. Left: evolution of containment scores with varying trajectory length $n_k \in \{4, 6, \ldots, 20\}$, Right: Estimated reachset sizes over varying n_k. Mean over 5 runs; training noise off, validation noise $\alpha_W = 1.5$.

DDRA attains $\approx 100\%$ containment at all horizons under sufficiently exciting control inputs and correct bounded-noise assumptions. This is consistent with its construction: as long as the data matrix $Z = \begin{bmatrix} X_- & U \end{bmatrix}^\top$ is invertible, and the disturbances bounded, the predicted sets cover the true reachable sets. Gray-box RCSI ranges between 90% and 100% with small non-monotone wiggles across n_k. Two factors explain this: (i) the model and uncertainty sets are fitted on unperturbed data, while the validation set is injected with noise; (ii) our containment test uses the interval hull of $\hat{\mathcal{Y}}_k$, which can deviate from pointwise membership.

The average interval width grows roughly linearly with trajectory length n_k for both methods, which is expected due to uncertainty accumulation forwards in time. RCSI's sets are consistently smaller since the uncertainty scales (α, c_Δ) are tuned to the data. DDRA propagates the entire model set and thus remains more conservative.

RCSI-learn time increases steeply with n_k (the number of conformance constraints grows with n_m, n_s, n_k), so repeated LP/NLP solves appear "exponential" over this range. Empirically, this is a higher-order polynomial. DDRA-learn (one-off regression / build of $\mathcal{M}_\Sigma$) time is nearly flat in n_k. At inference, the trends swap: DDRA grows superlinearly with horizon (generator growth and repeated Minkowski sums), whereas RCSI inference remains small and near-linear (linear maps with negligible remainder on LTI tasks).

2) Number of Unique Seeds: We evaluate the dependence of both algorithms on data quantity by varying the amount of unique input trajectories (n_m) across the sweep grid $\mathbf{n_m} = 4:4:100$, while fixing the state dimension to $D = 2$, the number of training trajectories $n_m = 10$, and the samples per trajectory $n_s = 5$ (full configuration details can be found in the Appendix). This experiment assesses the required volume of

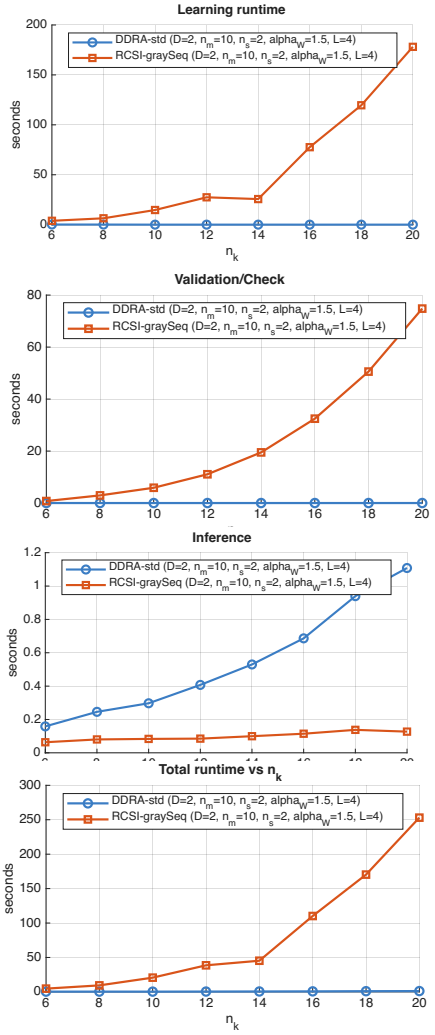

Fig. 4. Runtime scores separated into learning, validation and inference time, measured against varying n_k. Mean over 5 runs; training noise off, validation noise $\alpha_W = 1.5$.

data needed to reliably constrain the model sets and propagate uncertainty.

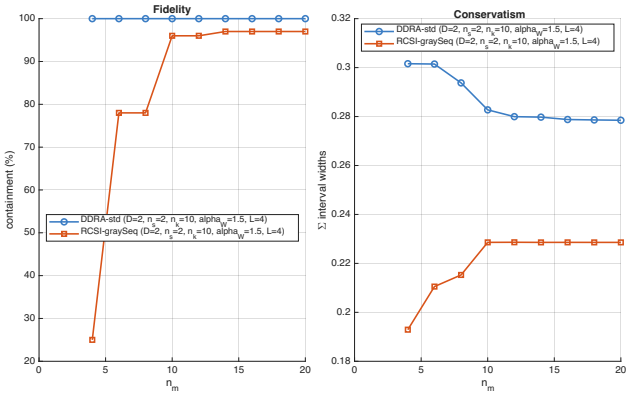

Fig. 5. Left: evolution of containment scores with varying number of nominal seeds $n_m \in \{4, 6, \ldots, 20\}$, Right: Estimated reachset sizes over varying n_m.Mean over 5 runs; training noise off, validation noise $\alpha_W = 1.5$

DDRA again stays at $\approx$ 100% across the sweep. RCSI starts below 100% for small n_m and rises toward 100% as n_m increases. This matches the intuition that more nominal cases expose more of the underlying system dynamics, which leads to more variability and forces larger uncertainty sets, ultimately improving coverage on validation.

RCSI's size increases with n_m initially and plateaus near $\sim$ 0.24: adding cases reveals rarer residuals, which the conformance program absorbs by enlarging α until no new violations appear. DDRA's size decreases with n_m and then stabilizes near $\sim$ 0.24: with more columns, the data matrix $Z = \begin{bmatrix} X_- & U \end{bmatrix}^\top$ is better conditioned and the feasible model set $\mathcal{M}_\Sigma$ shrinks, yielding tighter reachsets.

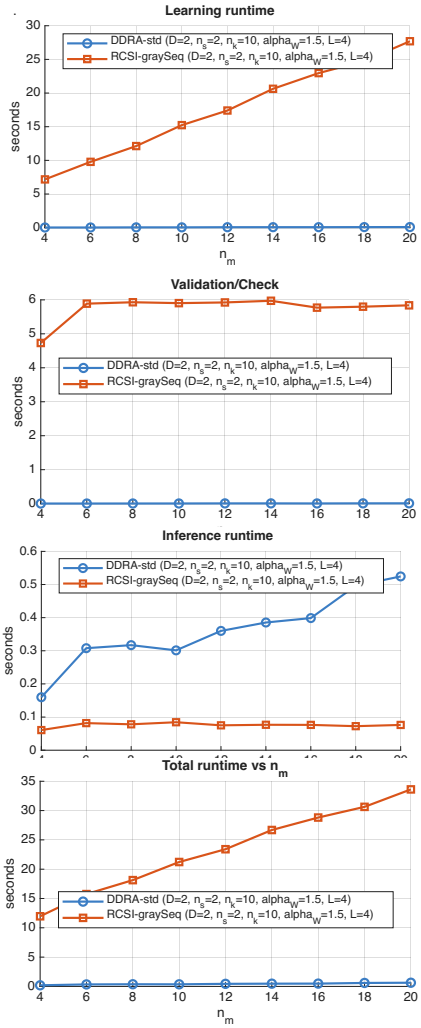

Fig. 6. Runtime scores separated into learning, validation and inference time, measured against varying n_m. Mean over 5 runs; training noise off, validation noise $\alpha_W = 1.5$

RCSI-learn grows roughly linearly in n_m; DDRA-learn remains near constant. Inference panels follow the same pattern as in the n_k sweep.

IV. Conclusion

Across all sweeps, DDRA maintains $\approx$100% containment

(by construction, given PE and bounded sets), with conservatism that decreases then plateaus as data improves the conditioning of the regression matrices. Gray-box RCSI reaches near-100% containment as n_m or n_k grows, while consistently producing smaller reachable sets because uncertainty scales (α, c_Δ) are optimized to the data. Computation-wise, DDRA training is essentially one-off and insensitive to horizon, but inference cost grows with n_k due to generator accumulation; RCSI exhibits the opposite trend (heavier training due to LP/NLP iterations that scale with the number of conformance constraints, lighter inference due to fixed linear maps). Within this class we study sample efficiency under a fixed uncertainty envelope and fixed dimension, so trends can be attributed unambiguously to data budget.

V. APPENDIX

VI. FURTHER EXPERIMENTAL DETAILS

a) Physical Model (Mass-Spring-Damper).: We consider a chain of D point masses connected by springs and linear dampers (see Fig. 7). Let X_i denote the absolute displacement of mass i, m_i its mass, k_i the spring stiffness between $(i, i{+}1)$ with natural length l_i, and b_i the viscous damping of mass i. With fixed boundary at the left and an offset L at the right terminal, the ith mass obeys

$$m_i\ddot{X}_i = k_i\left[X_{i+1}(t) - X_i(t) - l_i\right] - \dots \\ - k_{i-1}\left[X_i(t) - X_{i-1}(t) - l_{i-1}\right] - b_i\dot{X}_i(t).$$

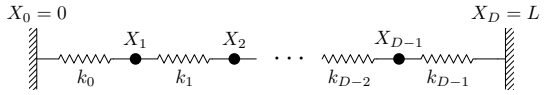

Fig. 7. Mass–spring system

Stacking $X(t) = [X_1, \dots, X_D]^\top$, we obtain

$$M\ddot{X}(t) + B\dot{X}(t) + KX(t) = s,$$

where $M = \mathrm{diag}(m_1, \dots, m_D)$, $B = \mathrm{diag}(b_1, \dots, b_D)$, and $K \in \mathbb{R}^{D\times D}$ is tridiagonal with $K_{ii} = k_{i-1}{+}k_i$, $K_{i,i+1} = K_{i+1,i} = -k_i$. The affine preload $s \in \mathbb{R}^D$ collects the natural-length and boundary offsets:

$$s_1 = k_0 l_0 - k_1 l_1,$$
$$s_i = k_{i-1} l_{i-1} - k_i l_i \ (2{\le}i{\le}D{-}1),$$
$$s_D = k_{D-1} l_{D-1} - k_D l_D - k_D L.$$

Using the augmented state $z(t) = \begin{bmatrix} X(t)^\top & \dot{X}(t)^\top \end{bmatrix}^\top \in \mathbb{R}^{2D}$, the continuous-time dynamics read

$$\dot{z}(t) = A_c z(t) + v,$$
$$A_c \overset{\text{def.}}{=} \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}B \end{bmatrix},$$
$$v \overset{\text{def.}}{=} \begin{bmatrix} 0 \\ M^{-1}s \end{bmatrix}.$$

External generalized forces (inputs) act through a constant matrix B_c; in our tests we use a standard, localized actuation pattern (the exact placement does not affect the methodology).

b) Test-suite generation and persistency of excitation.: For each sweep row we deterministically create a *training* suite and an *independent validation* suite with the same PE configuration: (i) draw a nominal input $\mathrm{u}_{0:k-1}^{(m)}$ per trajectory m using a PE generator (Gaussian or sinusoidal) with order target L, amplitude selected relative to $\mathcal{U}$'s half-width, and a fixed seed; (ii) for each nominal trajectory, create n_s sample paths by perturbing the initial state and input within $\mathcal{X}_0$ and $\mathcal{U}$ (extreme-point sampling with probability p_{extr}); (iii) simulate the true system to collect y_k. For linear systems we gate conformance checks on an *effective* PE order L_{eff} computed from the first nominal input Hankel rank, and we skip grid points that do not satisfy the requested PE order (keeping indices aligned).

c) Reproducibility and aggregation.: Each sweep is repeated N_{rep} times with distinct global seeds. For repetition r, we reuse the same train/val seeds across DDRA and Gray and record a per-row stable seed for determinism within a row. Per repetition we write a CSV of row-level metrics (timings, fidelity, conservatism, evaluation metrics). Aggregation forms group means and standard deviations by the varying key (e.g., n_k), writing a separate CSV for plotting. All reported figures in the main text are computed from these aggregated summaries.

TABLE I
TRAJECTORY LENGTH.

Quantity	Symbol	Default / Sweep
State dimension	n	2 (fixed per sweep)
Nominal trajectories	n_m	e.g. 10 (fixed)
Samples / traj	n_s	e.g. 2 (fixed)
Trajectory length	n_k	4:2:20
Process-noise scale	α_W	1.50 (fixed)
Reduction cap	ord	25 (streaming) / 100 (CORA)
Ridge (DDRA)	$(\lambda, \gamma, \text{policy})$	(10^{-8}, 1.0, MAB)

TABLE II
NOMINAL TRAJECTORY COUNT.

Quantity	Symbol	Default / Sweep
State dimension	n	2 (fixed per sweep)
Nominal trajectories	n_m	4:2:20
Samples / traj	n_s	e.g. 2 (fixed)
Trajectory length	n_k	e.g. 10 (fixed)
Process-noise scale	α_W	1.50 (fixed)
Reduction cap	ord	25 (streaming) / 100 (CORA)
Ridge (DDRA)	$(\lambda, \gamma, \text{policy})$	(10^{-8}, 1.0, MAB)

d) Compute environment.: All experiments were executed on a single Apple M1 system running macOS 15.6.1 (build 24G90; Darwin 24.6.0, `arm64`). The CPU is an Apple M1 with 8 cores (8 logical / 8 physical). The machine has 16 GB of RAM. Unless otherwise stated, all timings and results reported in this paper reflect this configuration.

REFERENCES

[1] A. Alanwar, "Data-driven-reachability-analysis: Source Code [Software]," 2024, commit 4804362, Accessed: 2025-07-05. [Online]. Available: https://github.com/aalanwar/Data-Driven-Reachability-Analysis

[2] L. Lützow and M. Althoff, "Reachset-conformant system identification," arXiv preprint arXiv:2407.11692, 2024.

[3] H. Roehm, A. Rausch, and M. Althoff, "Reachset conformance and automatic model adaptation for hybrid systems," Mathematics, vol. 10, no. 19, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/19/3567

[4] M. Althoff, "Checking and establishing reachset conformance in cora 2023," in Proc. of 10th International Workshop on Applied Verification of Continuous and Hybrid Systems, 2023.

[5] A. Alanwar, A. Koch, F. Allgöwer, and K. H. Johansson, "Data-driven reachability analysis from noisy data," IEEE Transactions on Automatic Control, vol. 68, no. 5, pp. 3054–3069, 2023.