WHEN HAVE WE LEARNED ENOUGH PHYSICS? TARGET-RELATIVE SUFFICIENCY FOR PDE OPERATOR DISCOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

Equation discovery is often declared complete when the selected support, fitted coefficients, or validation error appears stable. Those criteria describe the behavior of an estimator. They do not answer the scientific question: could another operator that is still compatible with the evidence behave differently in the regime that matters?

We define completion through the version set of all data-compatible operators. A target state envelope, declared before the unknown future response is observed, induces a target-action diameter on this set. This diameter measures the scientific ambiguity that remains after the completed experiments.

We prove an exact conditional ambiguity law and show that the realized diameter cannot increase as valid experiments are added. Under a common one-sided stability bound, we further transfer operator ambiguity to unseen-trajectory ambiguity and obtain a monotone stopping rule. Validated derivative error and bounded model discrepancy enter through explicit uncertainty inflation.

On public $Re_\tau \approx 1000$ turbulent-channel DNS from JHTDB, two acquisition blocks reduce normalized target-action ambiguity from 0.418 to 0.282 for an outer-layer target and from 0.434 to 0.313 for a near-wall target. Pressure and viscous multiplier coordinates nevertheless continue to span their declared prior intervals. We make no trajectory-stopping claim for this turbulent case because no useful validated global one-sided stability constant is available for the declared regime.

For an analytically controlled reaction–diffusion system, both the invariant target envelope and the stability constant are certified. The locked conservative trajectory certificate first crosses its predeclared tolerance at experiment 6. Across 100 independent noise realizations, 99 runs certify by experiment 8, the median stopping step is 6, every certificate path is monotone, and no operational false stop occurs among the 99 stopped runs.

The central distinction is therefore between finding a plausible equation and having enough evidence to declare the scientific uncertainty resolved.

1 INTRODUCTION

Learning governing equations from data is now an important problem in scientific machine learning. Sparse identification of nonlinear dynamics and PDE-FIND established that governing equations can be recovered directly from observed trajectories (Brunton et al., 2016; Rudy et al., 2017). Later work extended this idea to noisy, sparse, and limited observations through weak-form, neural, Bayesian, and online methods (Messenger et al., 2022; Stephany & Earls, 2024; Long et al., 2024). Mechanistic neural architectures also incorporate PDE structure directly into learned systems (Pervez et al., 2025). More broadly, equation discovery, operator learning, and experimental design are increasingly studied together (Brunton & Kutz, 2024).

Two limits are now well understood. Some governing equations remain structurally non-identifiable even when the optimization problem is solved accurately (Casolo et al., 2026). At the same time, active discovery and optimal experimental design can improve the information obtained from new experiments (Jiang et al., 2025; Xu et al., 2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We ask a different question:

> *When have the experiments already supplied enough information for the scientific use at hand?*

Full coefficient identifiability is not always necessary. A poorly resolved parameter direction may have almost no effect in the regime that matters. In that case, resolving it more precisely adds little scientific value. The reverse can also happen. A design matrix may be full rank, yet a weakly constrained direction may still produce large changes on the target regime. Rank, coefficient convergence, and point-estimation accuracy therefore do not determine whether discovery is complete.

A useful completion criterion should reflect the target rather than the parameterization alone. It should vary continuously with the remaining uncertainty, account for measurement error, and be computable before the unseen target response is revealed.

Consider

$$\partial_t u = \mathcal{F}_\theta(u) + g_e(t), \qquad \theta \in \Theta, \tag{1}$$

where e denotes an experiment or regime. For a finite library,

$$\mathcal{F}_\theta(v) = \sum_{j=1}^{p} \theta_j \mathcal{L}_j[v].$$

After the completed experiments, we collect all operators that remain compatible with the observations into a version set. We then declare a target state envelope independently of the unknown future response. Two operators are considered close if they act similarly on every state in that envelope, whether or not their coefficients are close.

This gives a direct notion of completion: the evidence is sufficient once all operators that remain compatible with it agree on the declared target to the required tolerance.

Contributions. We do not propose another sparse-regression solver. Version sets, moduli of continuity, optimal recovery, and the distinction between parameter-oriented and prediction-oriented design are classical (Micchelli & Rivlin, 1977; Donoho, 1994; Pukelsheim, 2006). We use these ideas to address a missing question in PDE discovery:

1. We define a target state envelope without using the unknown future response.
2. We give an exact conditional ambiguity law for the largest target-relevant difference that can remain unresolved by the completed experiments.
3. We prove a sharp transfer from operator ambiguity to unseen-trajectory ambiguity whenever a validated common stability bound is available.
4. We show how validated derivative error and bounded model discrepancy enter the certificate.
5. We test the framework on real turbulent DNS and on a controlled reaction–diffusion system, followed by a numerical-resolution audit and 100 independent noise repetitions.

The practical point is simple. A fitted equation may look convincing while the data still permit other operators with materially different scientific consequences.

2 TARGET-RELATIVE INFORMATION

Let Θ be a nonempty compact admissible class. For each completed experiment $e \in \mathcal{E}$, suppose the observed state has already been used to construct the chosen strong, weak, or projected library map. We write

$$y_e = A_e\theta^\star + \eta_e, \qquad \|\eta_e\|_2 \le \sigma_e. \tag{2}$$

We keep the experiment-wise constraints separate:

$$\mathcal{V}_{\mathcal{E}}(y) = \{\theta \in \Theta : \|A_e\theta - y_e\|_2 \le \sigma_e \quad \forall e \in \mathcal{E}\}. \tag{3}$$

A new experiment therefore adds a new constraint rather than changing the uncertainty budget assigned to earlier experiments.

For Gaussian response noise,

$$\eta_e \sim N(0, s_e^2 I_{m_e}),$$

the choice

$$\sigma_e = s_e \sqrt{\chi^2_{m_e,1-\delta_e}}$$

gives block coverage $1-\delta_e$. A union bound gives simultaneous coverage of at least $1-\sum_e \delta_e$.

Declaring the target. The target is a state envelope $\mathcal{U}_\mathcal{T} \subset \mathcal{X}$ chosen independently of $\theta^\star$ and of the unseen target response. It may come from a certified invariant or reachable set, from state-only target observations with a validated containment radius, or from fixed probe states when only operator-action claims are required.

Assume

$$M_\mathcal{T} := \sup_{\theta,\theta' \in \Theta} \sup_{v \in \mathcal{U}_\mathcal{T}} \|\mathcal{F}_\theta(v) - \mathcal{F}_{\theta'}(v)\|_\mathcal{X} < \infty. \tag{4}$$

Define

$$d_\mathcal{T}(\theta,\theta') = \sup_{v \in \mathcal{U}_\mathcal{T}} \|\mathcal{F}_\theta(v) - \mathcal{F}_{\theta'}(v)\|_\mathcal{X}, \qquad D_\mathcal{T}(\mathcal{V}) = \sup_{\theta,\theta' \in \mathcal{V}} d_\mathcal{T}(\theta,\theta'). \tag{5}$$

The diameter $D_\mathcal{T}$ measures the largest target-action difference that is still compatible with the evidence. Unlike a rank test, it remains informative when the design is technically full rank but poorly conditioned in scientifically important directions.

For the completed experiments define

$$d_\mathcal{E}(\theta,\theta') = \max_{e \in \mathcal{E}} \frac{\|A_e(\theta-\theta')\|_2}{\sigma_e}, \qquad \omega_{\mathcal{T}|\mathcal{E}}(\rho;\Theta) = \sup_{d_\mathcal{E}(\theta,\theta') \le \rho} d_\mathcal{T}(\theta,\theta'). \tag{6}$$

Theorem 1 (Exact conditional ambiguity law)**.** *For fixed realized experiment maps A_e,*

$$\sup_{y:\mathcal{V}_\mathcal{E}(y) \neq \varnothing} D_\mathcal{T}(\mathcal{V}_\mathcal{E}(y)) = \omega_{\mathcal{T}|\mathcal{E}}(2;\Theta). \tag{7}$$

For every realized dataset with $\theta^\star \in \mathcal{V}_\mathcal{E}(y)$ and every feasible $\widehat{\theta} \in \mathcal{V}_\mathcal{E}(y)$,

$$d_\mathcal{T}(\widehat{\theta},\theta^\star) \le D_\mathcal{T}(\mathcal{V}_\mathcal{E}(y)).$$

Adding a valid experiment cannot increase the realized diameter.

The constant 2 is exact. Two operators in the same version set differ by at most twice each experiment's uncertainty radius. Conversely, any pair satisfying that bound can be placed symmetrically inside a common response ball. The reverse construction ranges over the declared response sets rather than over dynamically realizable responses, so the identity is conditional on the realized maps A_e. Restricting to a smaller physically realizable response class can only reduce the worst-case diameter. The complete proof is in the supplement.

Proposition 2 (What the diameter certifies)**.** *Let $\mathfrak{T}(\theta)$ denote the target operator action on $\mathcal{U}_\mathcal{T}$, and define*

$$r_\mathcal{T}(\mathcal{V}) = \inf_z \sup_{\theta \in \mathcal{V}} \|z - \mathfrak{T}(\theta)\|_{\infty,\mathcal{U}_\mathcal{T}}. \tag{8}$$

Then

$$\frac{1}{2} D_\mathcal{T}(\mathcal{V}) \le r_\mathcal{T}(\mathcal{V}) \le D_\mathcal{T}(\mathcal{V}). \tag{9}$$

Hence $D_\mathcal{T}(\mathcal{V}) \le \tau$ makes every feasible operator τ-accurate relative to every possible truth in $\mathcal{V}$. If $D_\mathcal{T}(\mathcal{V}) > 2\tau$, no single target estimator can guarantee error at most τ uniformly over that same version set.

The upper bound follows by choosing any feasible target action as the center. The lower bound follows from the triangle inequality applied to two maximally separated feasible actions. Thus a large target diameter is not simply imprecise coefficient estimation. It is an obstruction to a uniform target guarantee.

3 FROM OPERATOR AMBIGUITY TO FUTURE BEHAVIOR

Let $\mathcal{R}_{\mathcal{T}}$ be a declared family of future inputs. For $r \in \mathcal{R}_{\mathcal{T}}$, let u_θ^r solve

$$\dot{u} = \mathcal{F}_\theta(u) + g_r(t), \qquad u(0) = u_{0,r}, \qquad 0 \le t \le T. \tag{10}$$

Assume every admissible target trajectory remains inside $\mathcal{U}_{\mathcal{T}}$. Also assume that a common one-sided stability constant exists:

$$\langle \mathcal{F}_\theta(v) - \mathcal{F}_\theta(w), v - w \rangle_{\mathcal{X}} \le \mu \, \|v - w\|_{\mathcal{X}}^2 \quad \forall \theta \in \Theta, \; v, w \in \mathcal{U}_{\mathcal{T}}. \tag{11}$$

Let

$$D_{\mathrm{traj}}(S) = \sup_{\theta,\theta' \in S} \sup_{r,t} \|u_\theta^r(t) - u_{\theta'}^r(t)\|_{\mathcal{X}} .$$

Theorem 3 (Sharp trajectory transfer)**.** *For every nonempty $S \subseteq \Theta$,*

$$D_{\mathrm{traj}}(S) \le C_\mu(T) D_{\mathcal{T}}(S), \qquad C_\mu(T) = \begin{cases} (e^{\mu T} - 1)/\mu, & \mu \ne 0, \\ T, & \mu = 0. \end{cases} \tag{12}$$

The factor $C_\mu(T)$ is unimprovable under only (11)*.*

The proof applies a comparison argument to the upper-right Dini derivative of

$$\|u_\theta - u_{\theta'}\| .$$

Using the Dini derivative avoids dividing by this norm at the common initial state, where it is zero. A scalar affine system attains equality, so the factor cannot be improved without additional dynamical assumptions.

On the coverage event,

$$\boxed{C_\mu(T) D_{\mathcal{T}}(\mathcal{V}_{\mathcal{E}}(y)) \le \tau_{\mathrm{traj}} \quad \Longrightarrow \quad \text{stop.}} \tag{13}$$

Every feasible operator then predicts every declared target trajectory within τ_{traj} of the unknown truth. Because new valid experiments only shrink the version set, a certified stopping decision cannot later be reversed.

The trajectory rule requires the stability assumption. If no useful common bound is available, (13) cannot be used. The target-action diameter remains valid, but it should then be interpreted only as an operator-level statement.

Validated numerical and model error. Suppose the computed experiment map $\widetilde{A}_e$ satisfies

$$\sup_{\theta \in \Theta} \left\| (A_e - \widetilde{A}_e)\theta \right\| \le \varepsilon_{A,e}.$$

Then the experiment radius can be inflated from σ_e to

$$\sigma_e + \varepsilon_{A,e}.$$

If the computed target action has error at most $\varepsilon_{\mathcal{T}}$ and the true target dynamics contain an additional bounded term $r^\star$ satisfying

$$\sup \|r^\star\| \le \varepsilon_{\mathrm{model}},$$

while remaining inside the same target envelope, then

$$\sup_{r,t} \left\| u_{\widehat{\theta}}^r - u_{\mathrm{true}}^r \right\| \le C_\mu(T) \left[\widetilde{D}_{\mathcal{T}}(\widetilde{\mathcal{V}}) + 2\varepsilon_{\mathcal{T}} + \varepsilon_{\mathrm{model}} \right]. \tag{14}$$

These approximation bounds must be justified independently. The result only shows how they propagate into the final certificate.

4 COMPUTATION AND EMPIRICAL PROTOCOL

For a finite library, the target action is linear in coefficient differences. If the target is represented by scalar probes $b_r^\top \theta$ and $\mathcal{V}$ is convex, then

$$D_{\mathcal{T}}(\mathcal{V}) = \max_r \left(\max_{\theta \in \mathcal{V}} b_r^\top \theta - \min_{\theta \in \mathcal{V}} b_r^\top \theta \right). \tag{15}$$

The certificate can therefore be computed by support optimization over the version set. No numerical-rank threshold is required. Small nonconvex sparse classes can be treated by support enumeration. Large nonlinear operator classes may require certified upper approximations.

For a finite Euclidean target map $B\theta$, let $U\Sigma V^\top$ be an SVD of B. After absorbing the singular values into the coordinates, let w_i denote the support width of the ith target coordinate over $\mathcal{V}$. Then

$$\max_i w_i \le \sup_{\theta, \theta' \in \mathcal{V}} \|B(\theta - \theta')\|_2 \le \left(\sum_i w_i^2 \right)^{1/2}. \tag{16}$$

The JHTDB analysis reports both bounds and uses the upper bound for certification. In the final outer-layer version set they are 0.2775 and 0.2817, so the reported plateau is not explained by a loose coordinate bound.

The two experiments test different parts of the framework. JHTDB tests whether target-action ambiguity contracts on real turbulent DNS under a validated numerical error model. The reaction–diffusion experiment tests the full trajectory certificate because its target envelope and common stability constant can be proved analytically.

Exploratory JHTDB preprocessing was completed before the confirmatory split. The final acquisition order, target sets, uncertainty rules, and hidden-response barriers were then fixed before hidden evaluation. The conventional rank, residual, and coefficient-stabilization comparisons were added later to the same frozen reaction–diffusion paths. They are descriptive diagnostics rather than preregistered baselines. The full chronology is included with the reproducibility material.

5 REAL TURBULENT DNS: TARGET AMBIGUITY CONTRACTS WITHOUT A TRAJECTORY CERTIFICATE

We use the public JHTDB turbulent-channel DNS at $Re_\tau \approx 1000$ (Li et al., 2008; Graham et al., 2016; Johns Hopkins Turbulence Database, 2026). The domain is

$$8\pi \times 2 \times 3\pi,$$

with

$$2048 \times 512 \times 1536$$

grid nodes and viscosity

$$\nu = 5 \times 10^{-5}.$$

The imposed mean pressure-gradient drive is 0.0025.

JHTDB provides both velocity and pressure. According to the channel documentation, the stored pressure output is obtained through the database's pressure-Poisson computation (Johns Hopkins Turbulence Database, 2026). We query pressure directly and treat the imposed mean drive separately as the known forcing

$$f = (0.0025, 0, 0).$$

The learned library is

$$\partial_t u - f = q_a[-(u\cdot\nabla)u] + q_p[-\nabla p] + q_\nu[\nu\Delta u], \qquad q \in [0.25, 1.75]^3. \tag{17}$$

The coefficient box is a broad, sign-preserving admissible class around the canonical multiplier 1. Its $\pm 75\%$ width was fixed before the confirmatory evaluation. It is not a confidence interval and is not estimated from JHTDB.

Because the reported diameters are conditional on this class, we also performed a post hoc prior-box sensitivity study. The audit reruns the box-dependent numerical radius and every validation gate for 11 symmetric boxes with half-widths $h = 0.10$–0.95. All 11 remain admissible. For $h = 0.30$–0.95, the relative contraction is stable: 32.4–32.8% for the outer target and 26.9–30.7% near the wall. Absolute diameters still change with the admissible box, as expected.

The two narrowest boxes behave differently. At $h = 0.10$, the normalized target diameter remains essentially equal to its prior width, so the experiment has little room to produce additional normalized contraction. At $h = 0.20$, contraction has begun but has not yet reached the stable broad-box regime. The complete symmetric and one-at-a-time sweeps are reported in the supplement.

The canonical coefficients $(1, 1, 1)$ are used only to audit the numerical measurement model and later for hidden evaluation.

Numerical consistency before certification. The first preprocessing design used an interpolated derivative pipeline with a crude error radius. It covered the outer-layer blocks but failed the known-PDE closure test near the wall. We rejected it.

A second design used nearest-grid high-order derivatives. Its apparent discretization ratios looked better, but the known-PDE residual exceeded the declared radius by a factor of 3.58. The cause was the moving Galilean grid used by the channel database: evaluating the temporal stencil at the nearest stored grid point changed the physical location across time.

We therefore rebuilt the confirmatory pipeline at fixed wall-frame coordinates. Primitive fields are evaluated with Lag8 interpolation. Spatial derivatives are computed with nested fourth-order differences at three resolutions, and time derivatives use a sixth-order centered stencil.

An independent calibration set contains 32 outer and 32 near-wall points. It fixes a factor-2 multiresolution uncertainty envelope before confirmatory acquisition begins. The confirmatory sequence then uses eight new blocks of 96 physical points each. Blocks 1–4 sample the outer layer; blocks 5–8 sample the near-wall region. The two target maps each use 192 held-out physical points.

The version-set residual and its radius use the same per-component RMS normalization. With $m_e = 3n_e$ scalar response entries,

$$\bar{A}_e = A_e^{\mathrm{raw}}/\sqrt{m_e}, \qquad \bar{y}_e = y_e^{\mathrm{raw}}/\sqrt{m_e}, \qquad \left\|\bar{A}_e q - \bar{y}_e\right\|_2 \le \bar{\sigma}_e. \tag{18}$$

The radius is

$$\bar{\sigma}_e = 2\left[\frac{\|D_6 - D_4\|_2}{\sqrt{m_e}} + \sup_{q\in\Theta} \frac{\|(A_{\mathrm{fine}} - A_{\mathrm{mid}})q\|_2}{\sqrt{m_e}}\right]. \tag{19}$$

This is exactly equivalent to the corresponding raw ℓ_2 constraint after multiplication by $\sqrt{m_e}$.

The coarser temporal and spatial differences are used only as convergence diagnostics. They never enlarge the radius after the result is observed. Calibration known-PDE residual/radius is 0.095 in the outer region and 0.144 near the wall. The largest confirmatory ratio is 0.129. Temporal fine/coarse ratios lie in $[0.169, 0.201]$, while spatial ratios lie in $[0.226, 0.478]$.

The target map has rank at most three because the coefficient vector has three components. The SVD simply provides an orthogonal basis in target-response space.

After block 2, the target-support extrema no longer move at reported precision. From blocks 2–8, their changes are below 8×10^{-16}, while the coefficient-width changes are below 4×10^{-14}. The constrained least-squares center continues to move over the same blocks.

This is not a caching artifact. The version set and all support problems are recomputed cumulatively after every acquisition. The later blocks shrink parts of the feasible set, but they do not change the extreme points that determine the reported worst-case target widths. The final outer lower and upper bounds, 0.2775 and 0.2817, are also very close, so the plateau is not caused by the coordinate upper bound.

The normalized coefficient diagnostic reports the widest individual coordinate interval relative to its prior width. At block 8 the three widths are

$$0.284, \qquad 1.000, \qquad 1.000$$

for advection, pressure, and viscosity. A width of 1.000 means that the corresponding coefficient can still reach both faces of its prior interval in some feasible solutions. It does not mean that the entire three-dimensional box remains feasible.

The pressure and velocity terms are physically coupled, so compensating coefficient changes are plausible. We do not infer a unique structural null direction from coordinate widths alone. The supplement reports the one-at-a-time prior sensitivity analysis.

Meanwhile, the point estimate changes from

$$(1.0013, 1.0325, 1.7500)$$

at block 1 to

$$\widehat{q}_8 = (0.9982, 0.9849, 0.9898).$$

After the certificate is frozen, hidden evaluation gives target-action errors of 0.00123 in the outer region and 0.00181 near the wall. The corresponding held-out temporal-response relative errors are 3.38% and 6.46%.

The result is therefore not a stopping claim. The fitted coefficients become very accurate, but the data-compatible set still contains operators whose target actions differ by more than the locked 0.10 operator-action criterion. Because no useful validated global stability constant is available for this turbulent regime, we do not convert that operator ambiguity into a future-trajectory certificate.

6 Reaction–diffusion: a certified stopping transition

We next consider the periodic reaction–diffusion system on $[0, 1]$,

$$u_t = \nu u_{xx} + au - bu^3 + g_r(x), \qquad \nu = 0.02, \tag{20}$$

with

$$(a, b) \in [0.6, 1.4] \times [0.8, 1.2]$$

and truth

$$(a^\star, b^\star) = (1.1, 1.0).$$

The target inputs satisfy

$$\|u_0\|_\infty \le 1.4, \qquad \|g_r\|_\infty \le 0.12.$$

Since

$$0.8(1.4)^3 - 1.4(1.4) - 0.12 = 0.1152 > 0, \tag{21}$$

the maximum principle gives the invariant envelope

$$\|u(t)\|_\infty \le 1.4$$

uniformly over the admissible coefficient class. In L^2, the cubic term is dissipative and the common one-sided stability constant is $\mu = 1.4$. Thus

$$C_\mu(0.4) = 0.536195.$$

The trajectory tolerance is fixed at

$$\tau_{\mathrm{traj}} = 0.12.$$

Each of the eight acquisitions contains 384 noisy response entries after known diffusion and forcing are removed. Gaussian noise has standard deviation 0.01. Each block uses the exact chi-square radius with

$$\delta_e = 0.01/8,$$

which gives simultaneous coverage of at least 0.99 across the eight blocks.

The acquisition sequence gradually increases nonlinear excitation. On the target envelope,

$$D_{\mathcal{T}}(\mathcal{V}) \le D_{\mathrm{env}}(\mathcal{V}) := \max_{\theta,\theta' \in \mathcal{V}} \left[1.4|a - a'| + 1.4^3|b - b'|\right]. \tag{22}$$

The confirmatory stopping rule was fixed using the conservative certificate

$$C_\mu(0.4) D_{\mathrm{env}}.$$

The resulting certificate path is

$$0.6452,\ 0.6452,\ 0.6431,\ 0.3200,\ 0.1519,\ \mathbf{0.1127},\ 0.1126,\ 0.1027.$$

The first certified stop is therefore experiment 6.

The fitted coefficients appear convincing much earlier. At experiment 2,

$$\widehat{\theta} = (1.1017, 0.9751),$$

close to the truth $(1.1, 1.0)$. Yet the relative coefficient diameter is still 1.0, and the trajectory certificate remains more than five times the allowed tolerance.

After the confirmatory analysis was complete, we also evaluated the exact target diameter on the same frozen version sets. The tighter certificate would cross the tolerance at experiment 5, with value 0.0889. We nevertheless retain experiment 6 as the confirmatory stopping point because the conservative D_{env} rule was the rule fixed before hidden evaluation.

At experiment 6, the relative coefficient diameter is 0.145. The hidden extreme-pair trajectory spread is 0.00148, and the truth-to-center trajectory error is 3.71×10^{-5}.

A separate grid-resolution audit compares 192 and 384 spatial points. The target-trajectory discrepancy is 2.16×10^{-5}, or 1.80×10^{-4} of the scientific tolerance. The discrepancy for the most nonlinear acquisition is 8.42×10^{-5}.

The conventional completion rules in **??** are descriptive comparisons on these same frozen paths. Full numerical rank occurs at experiment 1, although the cumulative singular-value ratio is only 0.0092. Across the 100 noise realizations, a 1% residual-plateau rule triggers at median experiment 2, while a 1% least-squares coefficient-stabilization rule triggers at median experiment 4. These occur before the theorem-backed certificate in 98/99 and 96/99 stopped runs, respectively.

Earlier triggering does not imply that these heuristic decisions are wrong on the realized hidden trajectories. It means only that they do not certify the same worst-case target statement. A coefficient-width rule also depends strongly on its chosen threshold: 25% triggers at experiment 5, 15–20% at experiment 6, and 10% does not trigger by experiment 8.

7 One hundred independent noise realizations

We next ask whether the stopping transition depends strongly on one particular noise draw. Every PDE parameter, acquisition state, observation location, target, tolerance, noise scale, radius rule, and solver setting is held fixed. Only the Gaussian observation noise is resampled.

Under independent block noise, the exact simultaneous coverage probability is

$$(1 - 0.01/8)^8 = 0.99004.$$

The empirical coverage over 100 repetitions is $98/100$, with a 95% Clopper–Pearson interval of

$$[0.9296, 0.9976].$$

With only 100 repetitions, this result should be read as consistent with the nominal 99% coverage, not as a precise empirical estimate of it.

All 100 certificate paths are monotone. By experiment 8, 99 runs have certified. The 95% Clopper–Pearson interval for this proportion is $[0.9455, 0.9997]$. The first stopping step is 5 in 14 runs, 6 in 76 runs, and 7 in 9 runs; one run does not stop by experiment 8. Among the stopped runs, the median and both quartiles equal 6.

The hidden target trajectories are evaluated only after all 100 stopping decisions are stored. Every stopped run satisfies the theorem's coverage event through its stopping time. No operational false stop is observed among the 99 stopped runs, with 95% interval $[0, 0.0366]$.

Every tested extreme-pair trajectory spread lies below its certificate. The largest hidden-spread-to-certificate ratio is 0.0286. At stopping, the median truth-to-feasible-midpoint trajectory error is 1.43×10^{-4} and the maximum is 5.27×10^{-4}.

8 Relation to existing work and scope

Equation-discovery methods primarily address estimation: which governing terms should be selected, and how can they be recovered reliably from imperfect observations (Brunton et al., 2016; Rudy et al., 2017; Stephany & Earls, 2024; Long et al., 2024; Pervez et al., 2025)? Other methods address acquisition. Online WSINDy processes observations sequentially (Messenger et al., 2022), while active discovery and optimal experimental design choose which information to collect next (Jiang et al., 2025; Xu et al., 2026). Identifiability theory characterizes what cannot be recovered from the available experiment at all (Casolo et al., 2026).

Our stopping question is different. It is asked after the data have been acquired: does the evidence already rule out every scientifically relevant alternative?

The mathematical ingredients have classical roots. Version sets and moduli of continuity are central to optimal recovery (Micchelli & Rivlin, 1977; Donoho, 1994). Target dependence is also related to the distinction between parameter-oriented and prediction-oriented experimental design (Pukelsheim, 2006). Our contribution is their use as a completion criterion for PDE discovery, together with the transfer from target-action ambiguity to future trajectories when a suitable dynamical bound is available.

Bayesian equation discovery gives a complementary uncertainty description (Long et al., 2024). A posterior depends on a probabilistic model and prior. Our version set instead gives a coverage-conditional worst-case statement over all operators satisfying the declared experiment constraints.

The method is also distinct from held-out validation. Low validation error for one fitted operator does not exclude another data-compatible operator with a different target action. Nor is target sufficiency an acquisition rule. An active or online method can choose the next experiment; our criterion can then be applied to the updated version set to decide whether another experiment is still needed.

Limitations. The guarantee depends on the declared target envelope and on the validity of the uncertainty bounds. If either is wrong, the certificate need not cover the true system.

The trajectory theorem also requires a useful stability bound. This condition can be verified for the reaction–diffusion example but not for the turbulent JHTDB regime considered here. Extending the trajectory certificate to chaotic systems will require different growth controls, such as finite-time, local, trajectory-dependent, or probabilistic bounds.

The JHTDB absolute diameters depend on the admissible coefficient class and on the chosen acquisition ordering. The prior-box audit shows that the fractional contraction is stable over a broad family of symmetric boxes, but it does not make the absolute diameter prior-independent.

Finally, exact support optimization is straightforward for the small libraries studied here. It may become difficult for large nonlinear operator classes, where certified upper approximations will be needed.

9 Conclusion

A convincing fitted equation is not the same as sufficient evidence.

Our framework asks whether any operator still compatible with the completed experiments can behave differently on the declared scientific target. The target-action diameter measures that remaining ambiguity. When a suitable dynamical stability bound is available, the same quantity yields a certificate for unseen future trajectories.

The experiments show both levels of the framework. Real turbulent DNS exhibits substantial contraction of target-action ambiguity even though some coefficient directions remain poorly resolved. The controlled reaction–diffusion system goes further and produces an actual theorem-backed stopping transition that is stable across repeated noise realizations.

The principle is:

> *Do not stop because one discovered equation looks right. Stop when every equation still compatible with the evidence is harmless for the declared scientific target.*

Reproducibility Statement

The supplement contains the complete proofs, target-envelope arguments, numerical conventions, experimental tables, statistical calculations, the exact reaction–diffusion target-diameter audit, the JHTDB coefficient-box sensitivity study, and the protocol history.

The `reproducibility/` directory contains all scripts used for the JHTDB and reaction–diffusion experiments, together with the CSV and JSON outputs used to produce the reported numbers and figures. Public JHTDB fields are queried through the database interface and are not redistributed. All manuscript graphics are stored in the shared `figures/` directory.

AI Use Statement

Generative AI (OpenAI ChatGPT) was used primarily as a language-editing and writing aid. It assisted with improving grammar, sentence structure, clarity, and concision in some parts of the manuscript.

The authors reviewed all AI-assisted writing and take full responsibility for the content and conclusions of the manuscript.

References

Steven L. Brunton and J. Nathan Kutz. Promising directions of machine learning for partial differential equations. *Nature Computational Science*, 4:483–494, 2024. doi: 10.1038/s43588-024-00643-2.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016. doi: 10.1073/pnas.1517384113.

Cecilia Casolo, Sören Becker, and Niki Kilbertus. Identifiability challenges in sparse linear ordinary differential equations. In *International Conference on Learning Representations*, 2026. URL https://iclr.cc/virtual/2026/poster/10010939.

David L. Donoho. Statistical estimation and optimal recovery. *The Annals of Statistics*, 22(1): 238–270, 1994. doi: 10.1214/aos/1176325367.

J. Graham, K. Kanov, X. I. A. Yang, M. K. Lee, N. Malaya, C. C. Lalescu, R. Burns, G. Eyink, A. Szalay, R. D. Moser, and C. Meneveau. A web services accessible database of turbulent channel flow and its use for testing a new integral wall model for les. *Journal of Turbulence*, 17(2):181–215, 2016. doi: 10.1080/14685248.2015.1088656.

Nan Jiang, Md Nasim, and Yexiang Xue. Active symbolic discovery of ordinary differential equations via phase portrait sketching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17626–17634, 2025. doi: 10.1609/aaai.v39i17.33938.

Johns Hopkins Turbulence Database. Channel flow at $re_\tau = 1000$: Dataset and readme documentation. https://turbulence.idies.jhu.edu/datasets/wallBoundedTurbulence/channelFlow, 2026. Accessed September 2026.

Yi Li, Eric Perlman, Minping Wan, Yunke Yang, Charles Meneveau, Randal Burns, Shiyi Chen, Alexander Szalay, and Gregory Eyink. A public turbulence database cluster and applications to study lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, 9:N31, 2008. doi: 10.1080/14685240802376389.

Da Long, Wei Xing, Aditi Krishnapriyan, Robert Kirby, Shandian Zhe, and Michael W. Mahoney. Equation discovery with bayesian spike-and-slab priors and efficient kernels. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2413–2421. PMLR, 2024. URL https://proceedings.mlr.press/v238/long24a.html.

Daniel A. Messenger, Emiliano Dall'Anese, and David M. Bortz. Online weak-form sparse identification of partial differential equations. In *Proceedings of Mathematical and Scientific Machine Learning*, volume 190 of *Proceedings of Machine Learning Research*, pp. 241–256. PMLR, 2022. URL https://proceedings.mlr.press/v190/a-messenger22a.html.

Charles A. Micchelli and Theodore J. Rivlin. A survey of optimal recovery. In Charles A. Micchelli and Theodore J. Rivlin (eds.), *Optimal Estimation in Approximation Theory*, pp. 1–54. Plenum Press, 1977.

Adeel Pervez, Efstratios Gavves, and Francesco Locatello. Mechanistic pde networks for discovery of governing equations. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 48962–48973. PMLR, 2025. URL https://proceedings.mlr.press/v267/pervez25a.html.

Friedrich Pukelsheim. *Optimal Design of Experiments*. SIAM, Philadelphia, 2006.

Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017. doi: 10.1126/sciadv.1602614.

Robert Stephany and Christopher J. Earls. Weak-pde-learn: A weak form based approach to discovering pdes from noisy, limited data. *Journal of Computational Physics*, 506:112950, 2024. doi: 10.1016/j.jcp.2024.112950.

Xingzi Xu, Johann Guilleminot, and Vahid Tarokh. Data generation with optimal experimental design for operator learning. *Computer Methods in Applied Mechanics and Engineering*, 450:118675, 2026. doi: 10.1016/j.cma.2025.118675.