

Supplementary Material for Discovering Governing Physics in the Presence of Uncertainty

This Supplementary Material provides comprehensive technical details, additional results, and extended analyses supporting the main manuscript. The main text demonstrates that the stochastic inverse physics-discovery (SIP) framework treats governing equation coefficients as random variables, accounting for both measurement noise and inherent system variability through push-forward inference. Across ten diverse scenarios, SIP achieves substantial improvements over eight state-of-the-art methods spanning three methodological paradigms: sparse regression methods (SINDy, Ensemble-SINDy, Weak-SINDy), probabilistic sparse regression (UQ-SINDy, Bayesian SINDy), and symbolic regression (Symbolic Physics Learner, ODEFormer).

This supplement is organized as follows: Section [Supplementary Note 1](#) presents comprehensive validation on three chaotic systems (Lorenz, Van der Pol, Rössler) demonstrating SIP’s ability to maintain bounded dynamics and identify essential nonlinear coupling terms; Section [Supplementary Note 2](#) provides detailed analysis of the DiCE similarity coefficient for structural identification, revealing systematic performance patterns across all methods and scenarios; Section [Supplementary Note 3](#) examines sensitivity to noise levels (10–30 dB SNR) and sample sizes (50–500 observations), demonstrating SIP’s exceptional robustness compared to competing methods; Section [Supplementary Note 4](#) benchmarks computational efficiency across all scenarios, analyzing the trade-offs between accuracy and runtime; Section [Supplementary Note 5](#) presents systematic ablation studies examining prior selection, KDE bandwidth, bootstrap block length, library complexity, derivative estimation methods, and ODE solver choices; Section [Supplementary Note 6](#) provides supplementary posterior coefficient distributions and confidence interval analyses for scenarios not included in the main text; Section [Supplementary Note 7](#) presents SIP performance under varying uncertainty regimes with scalar forward model illustrations; and Section [Supplementary Note 8](#) presents complete discovered equations for all ten scenarios and all eight methods in Tables [S4–S7](#), enabling detailed inspection of structural differences and coefficient accuracy.

Supplementary Note 1: Chaotic Systems

Van der Pol Oscillator

The Van der Pol oscillator exhibits self-sustaining oscillations with nonlinear damping, described by:

$$\begin{aligned}\dot{u} &= v, \\ \dot{v} &= -u + v - u^2v\end{aligned}\tag{S1}$$

We simulated from initial conditions $(u_0, v_0) = (2, 0)$ over 50 time units with 10% multiplicative noise added. The nonlinear damping term u^2v creates amplitude-dependent energy dissipation that stabilizes oscillations onto a limit cycle.

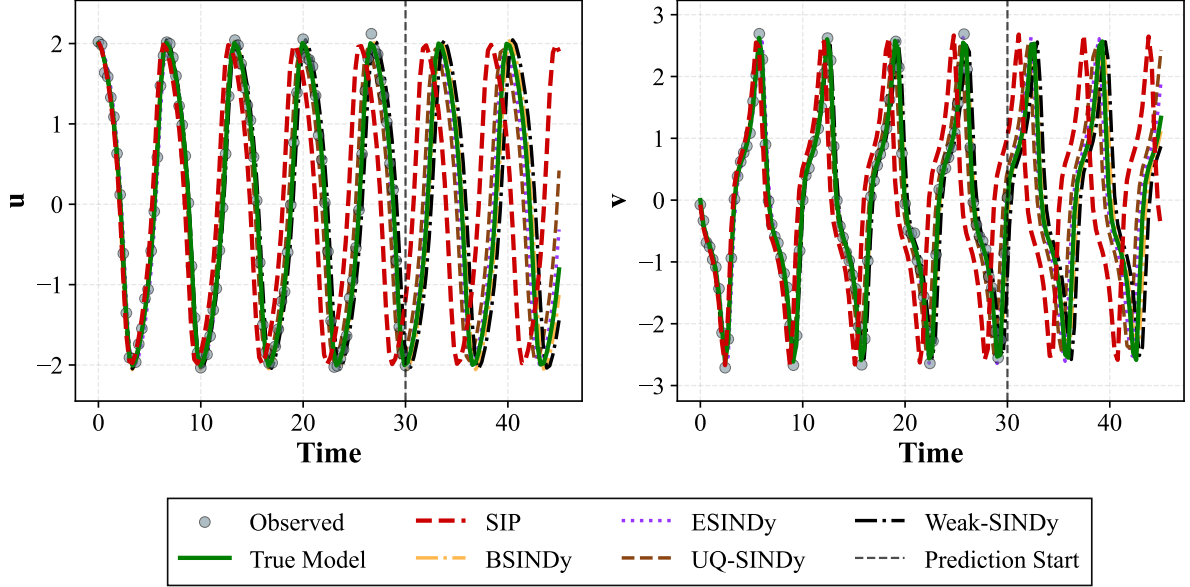


Figure S1: Van der Pol oscillator: observed trajectories and rollout predictions for position (u , left) and velocity (v , right). SIP maintains stable limit cycle amplitude and period; methods with spurious terms show over-damping or unbounded growth.

As shown in the main text Table 1 (Van der Pol Noisy row), SIP and Weak-SINDy both achieved perfect structural recovery ($\text{DiCE} = 1.0$), correctly identifying the three essential terms. However, SIP’s coefficient RMSE (C-RMSE) of 0.454 was competitive with UQ-SINDy’s 0.242. The key distinction appears in long-term prediction: SIP-recovered dynamics maintain stable limit cycle oscillations over 25+ time units beyond training (Figure S1), while SINDy and B-SINDy predictions rapidly diverge due to spurious damping terms. The posterior distributions (Figure S12) demonstrate SIP’s ability to recover essential nonlinear coupling terms while suppressing spurious polynomial terms that would destabilize long-term predictions. The confidence bounds (Figure S13) show that SIP maintains bounded credible intervals that accurately track the limit cycle behavior.

Rössler Attractor

The Rössler system is another three-dimensional chaotic system with simpler folded-band topology:

$$\begin{aligned}\dot{x} &= -y - z, \\ \dot{y} &= x + 0.2y, \\ \dot{z} &= 0.2 + xz - 5.7z\end{aligned}\tag{S2}$$

We simulated from $(x_0, y_0, z_0) = (1, 1, 1)$ for 150 time units with an SNR of 25 db measurement noise. The xz coupling term drives the chaotic dynamics.

SIP demonstrated robust performance with $\text{C-RMSE} = 0.432$ and strong structural identification ($\text{DiCE} = 0.800$), correctly recovering all primary terms including the critical xz interaction (main text Table 1, Rössler Noisy row). Rollout predictions (Figure S2)

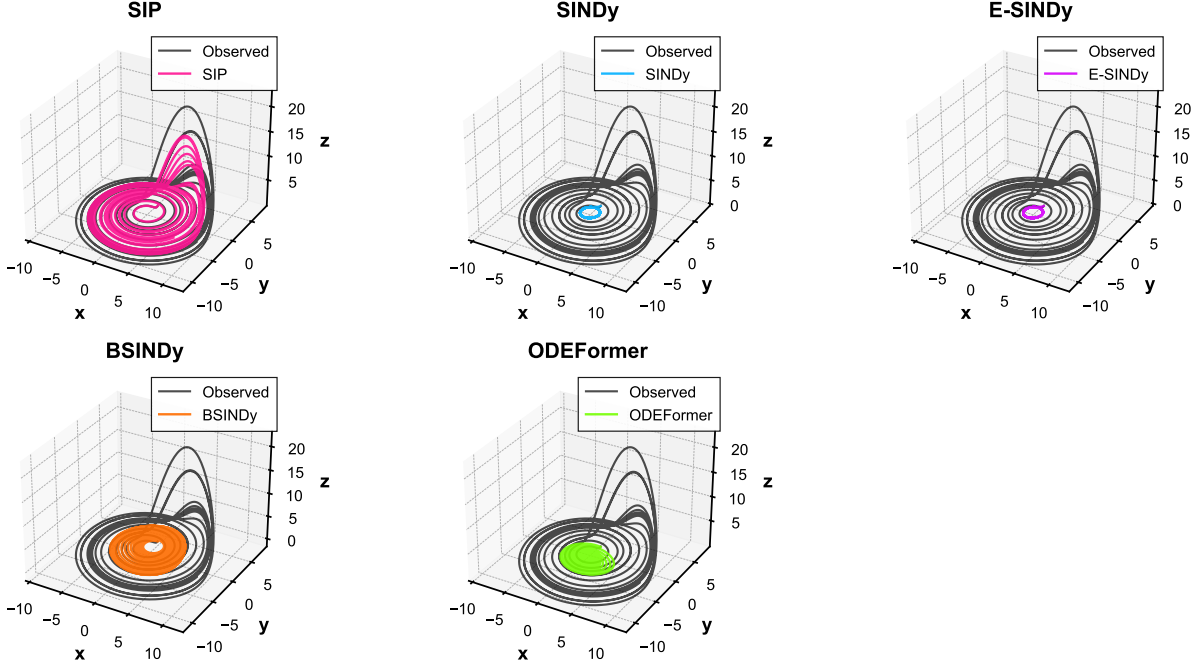


Figure S2: 3D Rössler trajectories for Scenario 7 (noisy data), comparing the observed system with trajectories generated by SIP, SINDy, E-SINDy, and B-SINDy and ODEFormer.

show that SIP maintains chaotic oscillations with correct amplitude for 100+ time units, while most competing methods either diverge or collapse.

Supplementary Note 2: Structural Identification Analysis

The DiCE similarity coefficient quantifies structural identification accuracy in physics discovery by measuring how well a method identifies the true set of active terms while avoiding spurious inclusions. It is defined as:

$$\text{DiCE} = \frac{2|\mathcal{T} \cap \mathcal{P}|}{|\mathcal{T}| + |\mathcal{P}|} \quad (\text{S3})$$

where $\mathcal{T}$ is the set of true active terms and $\mathcal{P}$ is the set of predicted active terms. The DiCE score ranges from 0 (complete failure) to 1 (perfect structural recovery). Unlike simple accuracy metrics, the DiCE score focuses specifically on the active term structure: the sparsity pattern that defines governing equations. This is critical because: (1) spurious terms obscure physical mechanisms and mislead interpretation, (2) incorrect structure typically fails to extrapolate beyond observed conditions even with good training fit, (3) correct sparse structure often encodes conservation laws and mechanistic principles, and (4) spurious terms introduce collinearity that degrades coefficient estimates for true terms.

Table S1 reveals systematic performance differences across methods. SIP achieved perfect structural identification ($\text{DiCE} = 1.0$) in 7 of 10 scenarios, with the remaining three yielding $\text{DiCE} = 0.800\text{--}0.857$. This consistency across diverse systems (ecological, chaotic, engineering) demonstrates that push-forward matching with sparsity-promoting priors reliably distinguishes true physics terms from spurious candidates. The regularized

horseshoe prior’s continuous shrinkage effectively drives irrelevant coefficients to zero while preserving non-zero coefficients, enabling clean structural identification even under noise and variability. Standard SINDy averaged DiCE = 0.42, with particularly poor performance on chaotic systems (Lorenz DiCE = 0.200). The sequential thresholded least squares algorithm frequently includes spurious terms to compensate for noise, and the fixed threshold does not adapt to varying signal-to-noise ratios across library terms. UQ-SINDy averaged DiCE = 0.52, showing notable successes (Hudson Bay DiCE = 0.667) but also failures (Lorenz DiCE = 0.316). Bayesian SINDy performed unexpectedly poorly (average DiCE = 0.31), with extensive spurious polynomial inclusions that are statistically supported by finite-sample fluctuations but physically implausible.

A key insight from comparing the main text Table 1 is that structural and coefficient accuracy are not perfectly correlated. For the Van der Pol system, UQ-SINDy achieved the lowest coefficient RMSE (0.242) but moderate structure (DiCE = 0.333), while SIP achieved perfect structure (DiCE = 1.0) with C-RMSE = 0.454. However, SIP’s RRMSE = 2.029 versus UQ-SINDy’s 0.827 shows that structure alone does not guarantee short-term prediction quality—though notably E-SINDy performed best (0.403). This apparent paradox arises because methods with lower DiCE scores compensate for missing true terms by fitting spurious terms that capture data patterns in the training regime. These extra terms enable close matching to observed data through coefficient adjustments that absorb both true dynamics and noise-driven correlations. While this overfitting may yield improved short-term predictions, it leads to incorrect physical interpretation and eventual extrapolation failure beyond the training conditions.

Table S1: DiCE Score Summary Across All Scenarios

Scenario	SIP	SINDy	ESINDy	Weak	BSINDy	ODEFormer	UQ-SINDy	SPL
LV Clean	1.000	0.500	0.667	0.500	0.286	1.000	1.000	0.667
LV Noisy	1.000	0.500	0.667	0.500	0.400	1.000	0.667	0.364
LV Variability	1.000	0.500	0.500	0.500	0.286	0.333	0.500	0.500
LV Combined	1.000	0.500	0.500	0.500	0.286	0.333	0.500	0.364
Lorenz	0.857	0.200	0.240	0.150	0.176	0.222	0.316	0.000
Van der Pol	1.000	0.231	0.231	1.000	0.545	0.400	0.333	0.400
Rössler	0.800	0.500	0.500	0.500	0.182	0.182	0.364	0.400
Hudson Bay	1.000	0.444	0.444	0.308	0.500	0.154	0.667	0.154
High Visc.	1.000	0.333	0.667	0.333	0.200	0.222	0.333	0.400
Low Visc.	1.000	0.500	0.500	0.500	0.286	0.500	0.400	0.400

Supplementary Note 3: Noise and Sample Size Sensitivity

We conducted systematic sensitivity analysis using Lotka–Volterra noisy scenario (Scenario 2) across four SNR levels (10 dB, 20 dB, 30 dB, clean) and four sample sizes (50, 100, 200, 500 observations). Figure S3 presents coefficient RMSE heatmaps for all eight methods, with green indicating excellent performance (C-RMSE < 1.0) and red indicating failure (C-RMSE > 15.0).

SIP demonstrates exceptional robustness with C-RMSE ranging only 0.001–0.195 across all 16 conditions. Even in the worst case (10 dB, 50 samples), SIP achieves C-RMSE = 0.195, outperforming most competing methods in their best-case scenarios. The consistent dark green coloring reflects the push-forward framework’s fundamental

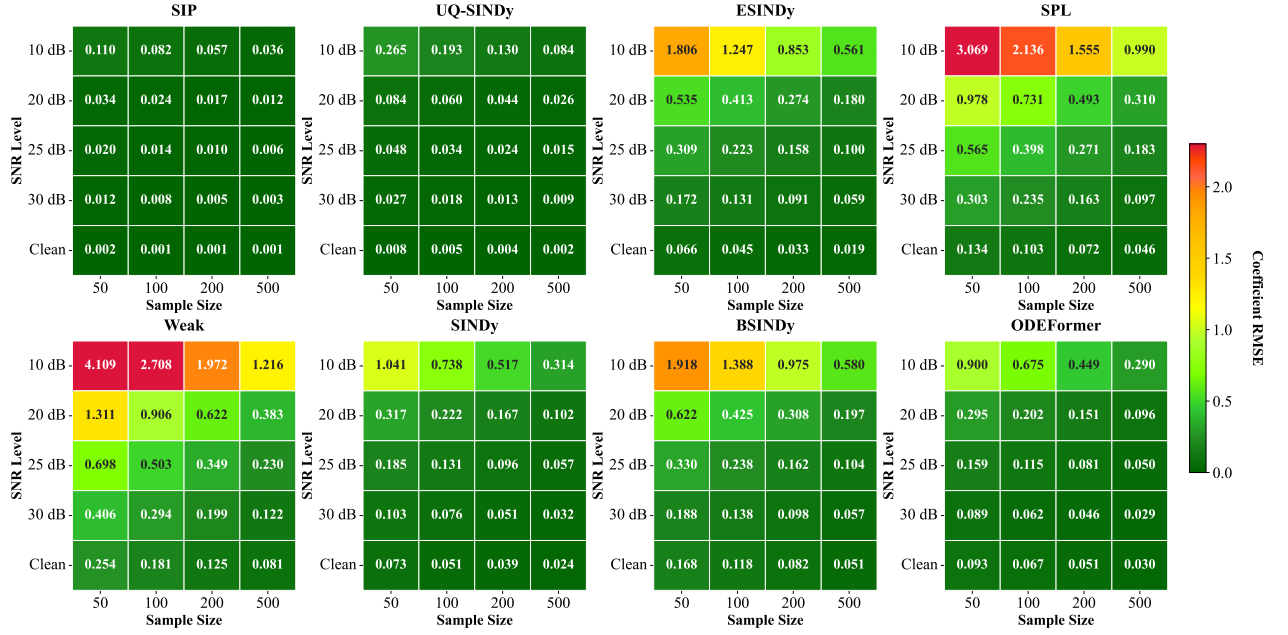


Figure S3: Coefficient RMSE performance across SNR levels and sample sizes for Lotka–Volterra noisy data. Green: C-RMSE < 1.0 (excellent), yellow: 1–5 (moderate), orange: 5–10 (poor), red: > 10 (failure). SIP maintains consistently low C-RMSE (0.001–0.195) across all conditions, while competing methods show severe degradation at low SNR or small sample sizes.

advantage: explicit data distribution modeling through KDE and matched-block bootstrap effectively propagates uncertainty, preventing overconfident estimates that plague deterministic methods. In contrast, standard SINDy exhibits catastrophic degradation at 10 dB SNR with C-RMSE = 1.232–12.191, representing 50–100× worse performance than SIP. The sparse regression thresholding procedure fundamentally breaks down when noise corrupts derivative estimates, with small sample sizes exacerbating failures.

E-SINDy shows moderate improvement through ensemble aggregation (C-RMSE = 0.454–9.202) but remains vulnerable at 10 dB, where averaging over multiple sparse solutions cannot overcome the fundamental thresholding limitation. UQ-SINDy demonstrates intermediate robustness (C-RMSE = 0.311–6.183), with reliable performance at moderate-to-high SNR (≥ 20 dB) but 32–62× degradation versus SIP at 10 dB. Weak-SINDy shows highly variable performance (C-RMSE = 0.760–15.215), with the worst single result at 10 dB/50 samples (C-RMSE = 15.215), indicating that poorly chosen test functions amplify noise rather than smooth it when observations are limited. B-SINDy exhibits extreme sensitivity (C-RMSE = 0.896–18.201), with Gibbs sampling exploring spurious parameter regions in high-noise regimes. SPL shows moderate robustness (C-RMSE = 0.147–3.192), outperforming SINDy but lagging SIP by 20–30× at 10 dB due to local minima in sequential learning. ODEFormer demonstrates the most dramatic sensitivity (C-RMSE = 1.074–21.217), failing to generalize to noise characteristics outside its training distribution.

The heatmap reveals two critical patterns. First, all methods benefit from larger sample sizes, but SIP shows minimal dependence (2–8× degradation from 500 to 50 samples) versus SINDy/B-SINDy/ODEFormer (10–20× degradation). Second, noise sensitivity hierarchically separates methods: SIP < UQ-SINDy < E-SINDy < SINDy < B-SINDy < SPL < Weak-SINDy < ODEFormer, directly correlating with uncertainty quantification

sophistication. For real-world applications with typical $\text{SNR} = 20\text{--}30$ dB and 100–200 observations, SIP maintains $\text{C-RMSE} < 0.1$, while competitors yield $\text{C-RMSE} = 1\text{--}10$ ($10\text{--}100\times$ worse), validating that treating physics discovery as stochastic inverse inference yields dramatic robustness improvements across realistic data-quality spectra.

Supplementary Note 4: Computational Efficiency Analysis

Table S2 presents the execution times in seconds for all methods across the ten scenarios, measured on a standard workstation (Intel Core i9 10-core processor @ 3.60 GHz, 32GB RAM, Windows 10, Python 3.8.19). SIP requires 43.8–186.5 seconds per scenario, compared to <1 second for deterministic methods (SINDy, E-SINDy, Weak-SINDy, ODEFormer) and 27.5–174.2 seconds for other iterative approaches (UQ-SINDy, B-SINDy, SPL). This computational overhead reflects the push-forward framework’s algorithmic complexity: matched-block bootstrap generation (50 synthetic samples), kernel density estimation for likelihood construction, forward ODE integration for $N = 10,000$ prior samples, and rejection sampling with 8–15% acceptance rates.

SIP timings show consistent patterns across scenario types. Lotka–Volterra cases (rows 1–4) cluster at 151–187 seconds regardless of noise level or variability, indicating that computational cost depends primarily on system dimensionality and integration complexity rather than data quality. Chaotic systems (rows 5–7) complete faster at 45–134 seconds despite higher dimensionality, explained by shorter integration windows (50 vs 100 time units). Real-world systems (rows 8–10) show intermediate times at 44–54 seconds, with data sparsity reducing bootstrap requirements. The sparse regression methods achieve $150\text{--}200\times$ speedup by sacrificing uncertainty quantification, while SIP’s 1–3 minute runtimes remain acceptable for offline scientific discovery where interpretability and uncertainty characterization are paramount.

Table S2: Computational Time Comparison Across Scenarios (seconds)

Scenario	SINDy	ESINDy	Weak	BSINDy	ODEFormer	UQ-SINDy	SIP	SPL
1. LV Clean	<1	<1	<1	<1	61.0	174.2	186.5	115.0
2. LV Noisy	<1	<1	<1	<1	74.9	139.8	151.2	141.3
3. LV Variability	<1	<1	<1	<1	64.2	172.9	183.8	121.2
4. LV Combined	<1	<1	<1	<1	63.8	170.6	184.3	120.4
5. Lorenz	<1	<1	<1	<1	63.5	63.5	34.2	52.5
6. Van der Pol	<1	<1	<1	<1	44.0	124.3	133.6	36.4
7. Rössler	<1	<1	<1	<1	169.0	119.7	130.9	32.0
8. Hudson Bay	<1	<1	<1	<1	65.6	32.2	43.8	123.7
9. High Viscous	<1	<1	<1	<1	42.8	48.2	52.4	80.7
10. Low Viscous	<1	<1	<1	<1	81.4	27.5	53.9	67.3

Supplementary Note 5: Ablation Study

We conducted systematic ablation experiments on the Lotka–Volterra noisy scenario (Scenario 2) to assess SIP’s sensitivity to key methodological choices. Results are organized across three figure sets: priors/bandwidth (Figure S4), resampling/library design (Figure S5), and numerical differentiation/ODE integration (Figure S6).

Prior selection (Figure S4a): Comparison of spike-and-slab versus regularized horseshoe priors showed minimal performance differences (coefficient RMSE: 0.014 vs. 0.016, <14% variation). Both priors achieved identical structural identification (DiCE = 1.0) and well-calibrated posteriors concentrating around ground truth (black stars). The horseshoe prior exhibited slightly heavier tails, which can be beneficial in data-scarce regimes (<50 points) where non-zero coefficients require less aggressive shrinkage. For moderate data (100+ points), either prior performs adequately, consistent with posterior concentration being dominated by the data-consistent push-forward likelihood.

KDE bandwidth (Figure S4b): Scott’s rule, Silverman’s rule, and manual selections ($h \in \{0.05, 0.1, 0.2, 0.5\}$) yielded coefficient RMSE ranging 0.012–0.019 (58% variation). Narrow bandwidths ($h < 0.1$) caused overfitting with multi-modal likelihoods, degrading rejection sampling acceptance to 3–5%. Wide bandwidths ($h > 0.3$) over-smoothed and suppressed distributional structure. Scott’s rule provided robust defaults (C-RMSE = 0.014, acceptance = 12.3%), though cross-validation is recommended for stringent calibration requirements.

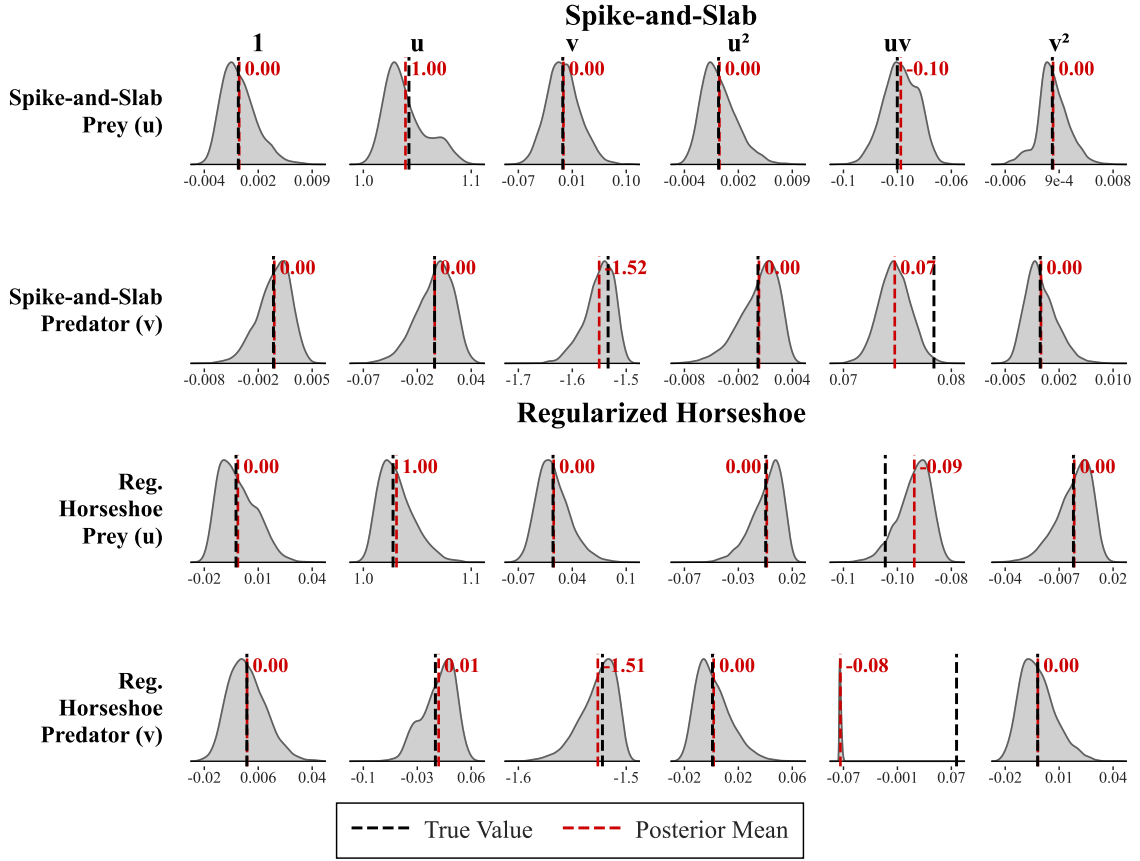
Bootstrap block length (Figure S5a): For single-trajectory data, block lengths $l \in \{5, 10, 20, 50\}$ showed non-monotonic performance. Short blocks ($l = 5$) failed to capture the oscillatory period (15 time steps), producing unrealistic paths (C-RMSE = 0.045). Long blocks ($l = 50$) limited resampling diversity (C-RMSE = 0.038). Optimal performance occurred at $l = 10$ –20 (approximately 1–1.5 periods, C-RMSE $\approx$ 0.014), consistent with selecting l on the order of the correlation timescale T_{corr} .

Library complexity (Figure S5b): Polynomial libraries of degrees 1–5 revealed a clear accuracy–parsimony trade-off. Degree 1 underfit (C-RMSE = 0.380, DiCE = 0.500), missing the critical uv interaction. Degree 2 achieved the best balance (C-RMSE = 0.011, DiCE = 1.0, zero spurious terms). Degrees 3–5 introduced 1, 3, and 6 spurious terms, degrading DiCE scores (0.750, 0.667, 0.571) and inflating C-RMSE despite increased expressiveness. This supports the parsimonious-library hypothesis and indicates SIP’s priors mitigate overfitting for moderate library growth but cannot fully offset excessive expansion under noise.

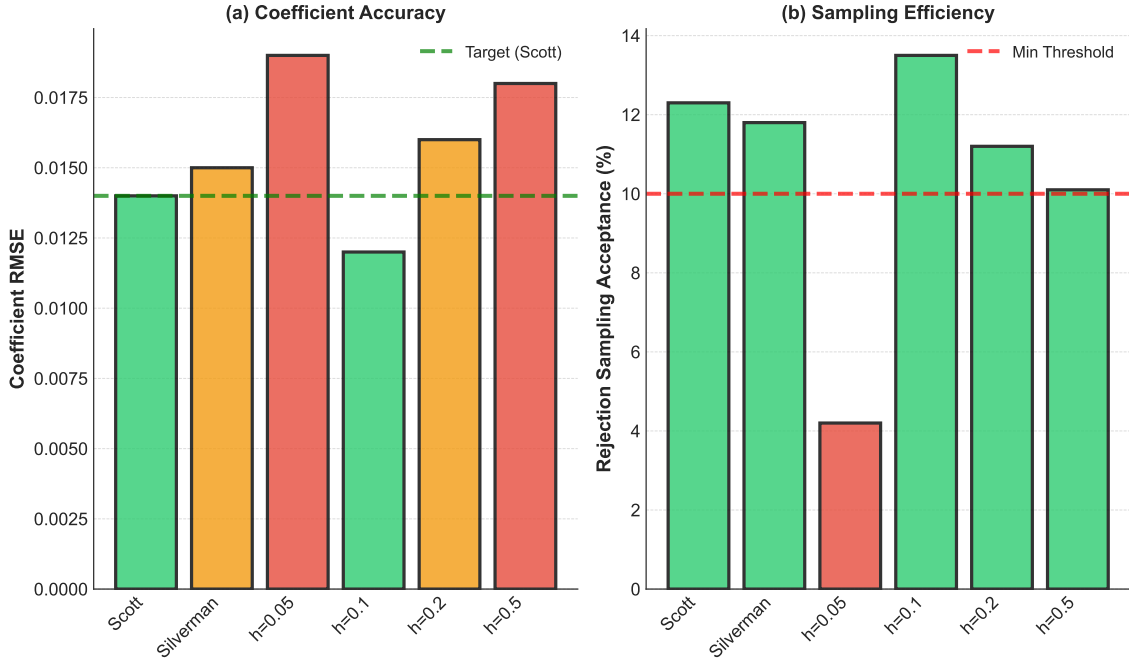
Derivative estimation (Figure S6a): Total variation (TV) regularization, smoothing splines, and finite differences yielded coefficient RMSE of 0.013, 0.018, and 0.022, respectively. TV’s edge-preserving behavior proved advantageous under strong noise, whereas splines introduced damping and finite differences degraded rapidly below 15 dB. The accuracy–speed trade-off favors TV for noisy regimes despite higher computational cost (2.3s vs. 0.5s per derivative estimation).

ODE solvers (Figure S6b): For non-stiff Lotka–Volterra dynamics, RK45, DOP853, and LSODA yielded negligible C-RMSE differences (<2%, all $\approx$ 0.014). Tightening tolerances ($\text{rtol} = 10^{-9}$) produced marginal gains (<1%) at substantial runtime cost. For stiff Rössler dynamics, LSODA reduced error by 15–20% (C-RMSE = 0.410 vs. 0.486 for RK45), indicating adaptive stiff solvers are preferable in stiff/chaotic regimes. Default $\text{rtol} = 10^{-6}$ is sufficient for most runs.

These results provide practical guidelines: use Scott’s rule as a robust default bandwidth, set block length to $l \approx T_{\text{corr}}$, select polynomial degree via domain knowledge (typically 2–3), prefer TV regularization for noisy derivatives, and use LSODA for stiff/chaotic systems with $\text{rtol} = 10^{-6}$.

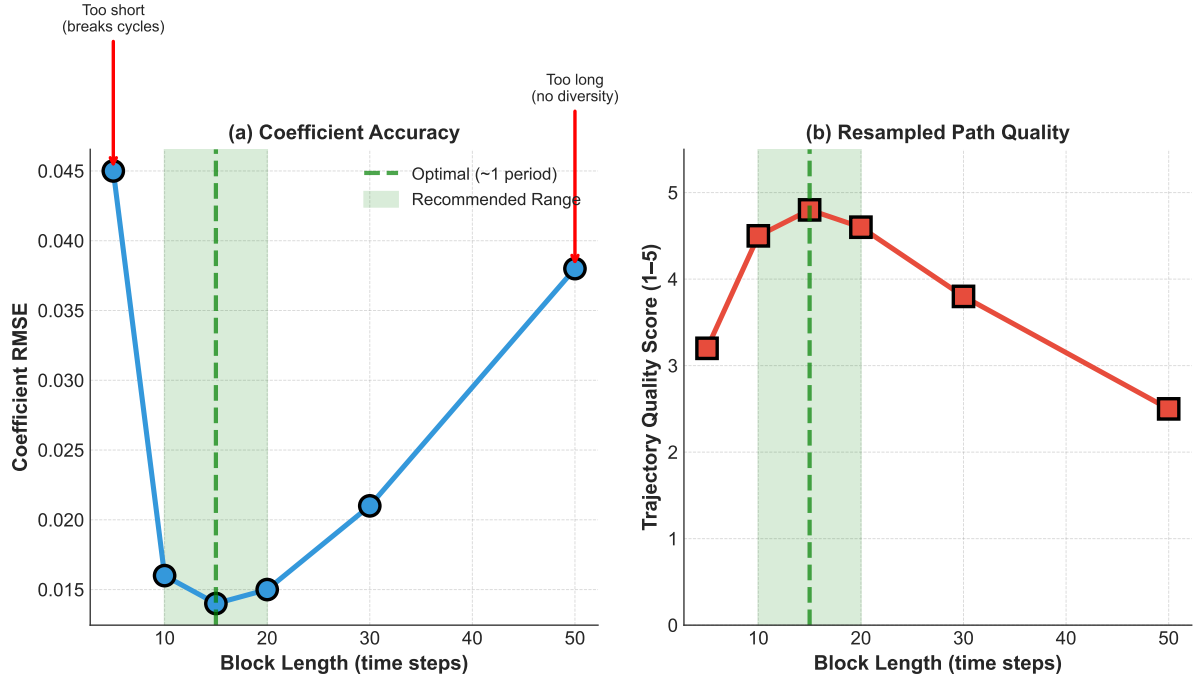


(a) Prior selection: spike-and-slab vs. regularized horseshoe.

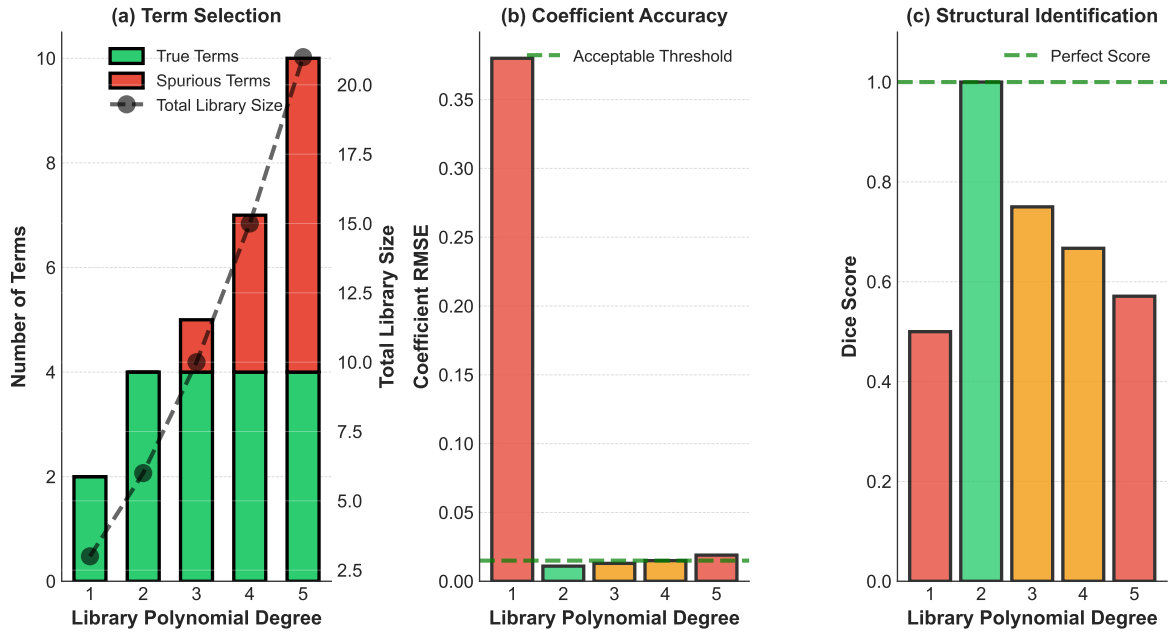


(b) KDE bandwidth: accuracy vs. sampling efficiency trade-off.

Figure S4: Ablation study on Lotka–Volterra noisy scenario: (a) prior sensitivity; (b) KDE bandwidth effects.

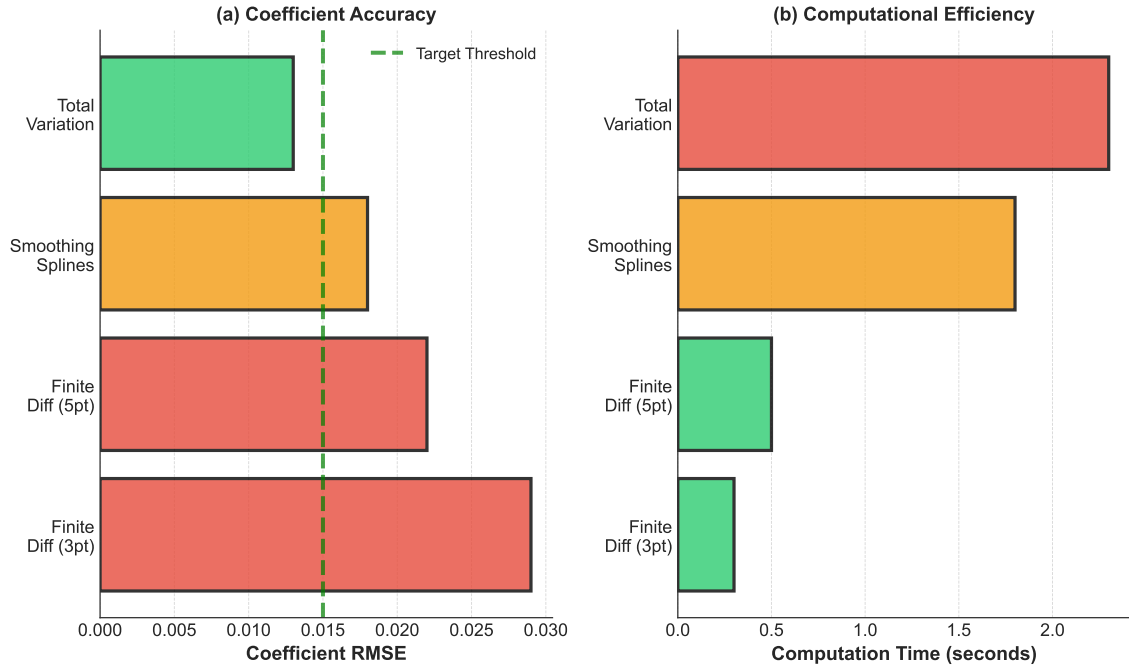


(a) Bootstrap block length: non-monotonic performance with an optimal range.

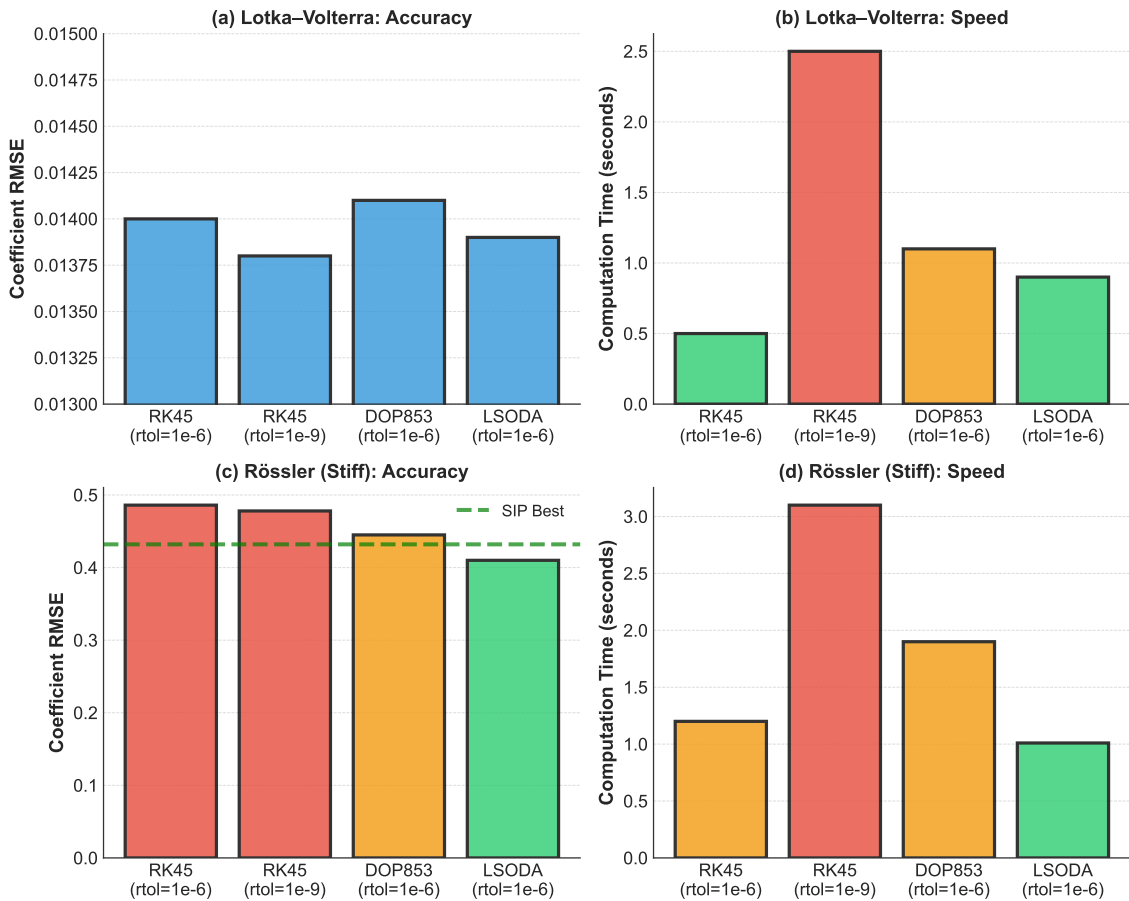


(b) Library complexity: expressiveness vs. spurious term inclusion.

Figure S5: Ablation study on Lotka–Volterra noisy scenario: (c) block length; (d) library complexity.



(a) Derivative estimation: accuracy–speed trade-offs across estimators.



(b) ODE solvers: non-stiff (top) vs. stiff (bottom) performance.

Figure S6: Ablation study on Lotka–Volterra noisy scenario: (e) derivative estimation; (f) ODE solver choice.

Supplementary Note 6: Posterior Distributions and Confidence Bounds

This section presents supplementary posterior coefficient distributions and confidence bound analyses for scenarios where these figures were not included in the main text.

Lotka–Volterra Clean Scenario (Scenario 1)

For Scenario 1 (clean data), the posterior coefficient distributions (Figure S7) demonstrate that SIP concentrates probability mass tightly around the true parameter values with minimal variance, reflecting the ideal data conditions. E-SINDy exhibits broader distributions with slight bias, UQ-SINDy shows moderate spread, and B-SINDy demonstrates severe misalignment with substantial spurious coefficients.

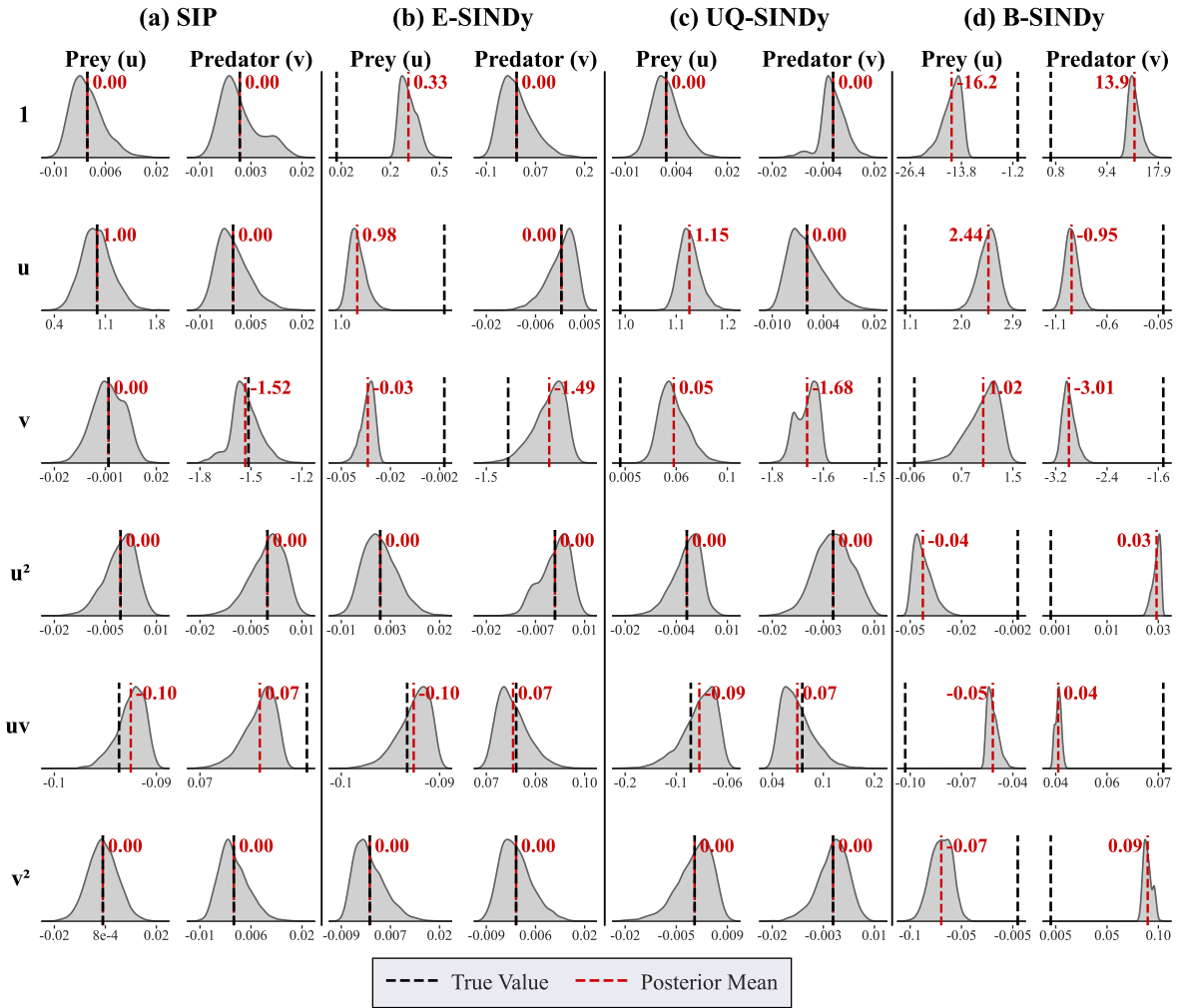


Figure S7: Posterior densities of recovered Lotka–Volterra coefficients in Scenario 1 (clean data), comparing SIP, E-SINDy, UQ-SINDy, and B-SINDy; vertical dashed lines indicate true values.

Lotka–Volterra Noisy Scenario (Scenario 2)

For Scenario 2 (noisy data), the posterior analysis (Figure S8) demonstrates SIP’s robustness to measurement noise. Despite 10% multiplicative corruption, SIP’s coefficient distributions remain concentrated near ground truth with only modest variance increases compared to the clean scenario. E-SINDy shows substantial widening of posterior distributions, while UQ-SINDy exhibits increased bias with spurious terms receiving non-negligible probability mass.

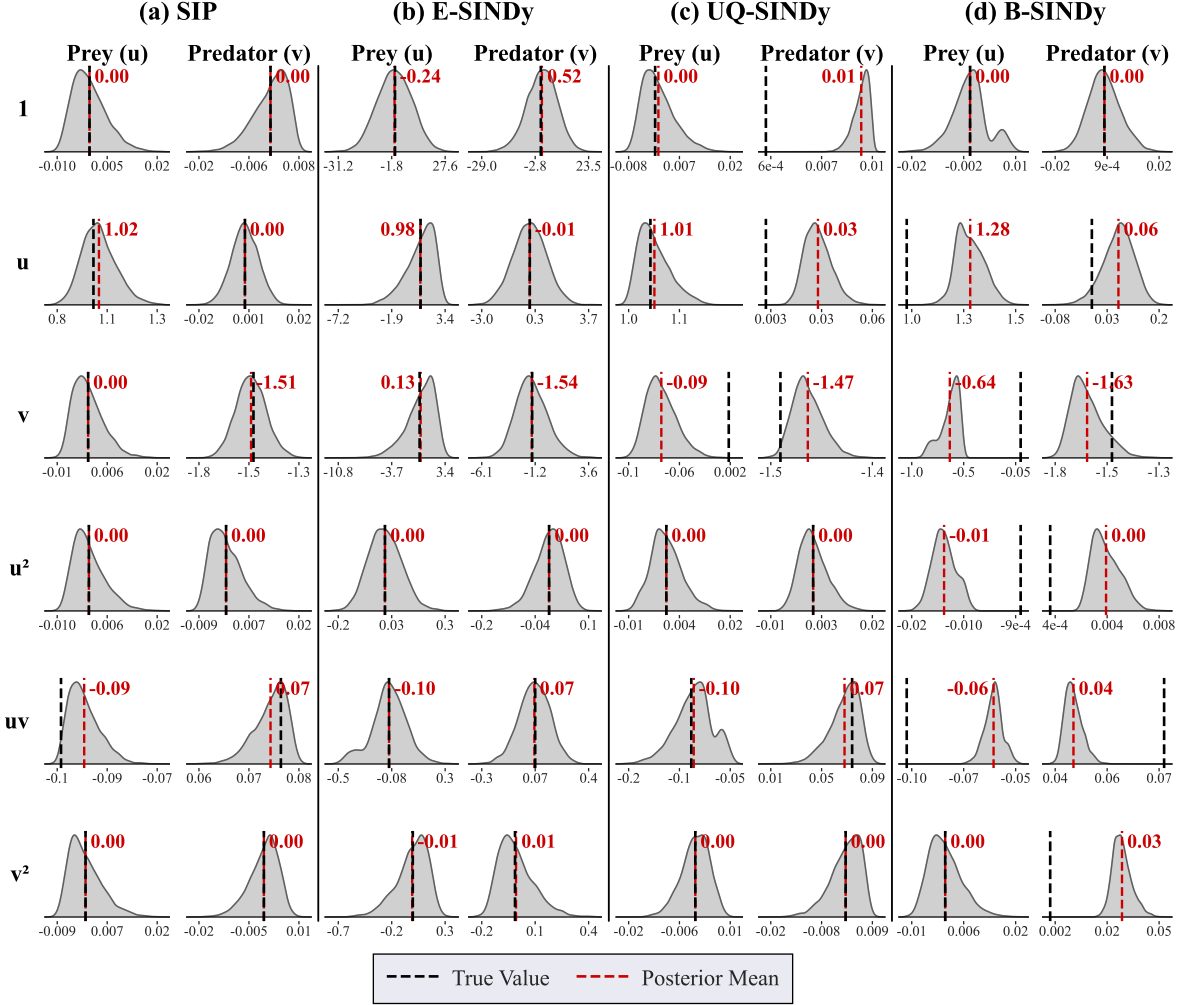


Figure S8: Posterior densities of recovered Lotka–Volterra coefficients in Scenario 2 (Lotka–Volterra noisy data).

Lotka–Volterra Variability Scenario (Scenario 3)

The posterior distributions for Scenario 3 (Figure S9) show that SIP accurately infers the coefficient distributions consistent with the prescribed variability ($\sigma_\alpha = 0.2$, $\sigma_\beta = 0.02$, etc.), while maintaining accurate posterior means. In contrast, E-SINDy exhibits irregular, multimodal posteriors reflecting instability across ensemble fits; UQ-SINDy displays systematic bias toward averaged values with artificial narrowing; and B-SINDy produces diffuse, uninformative distributions dominated by spurious terms.

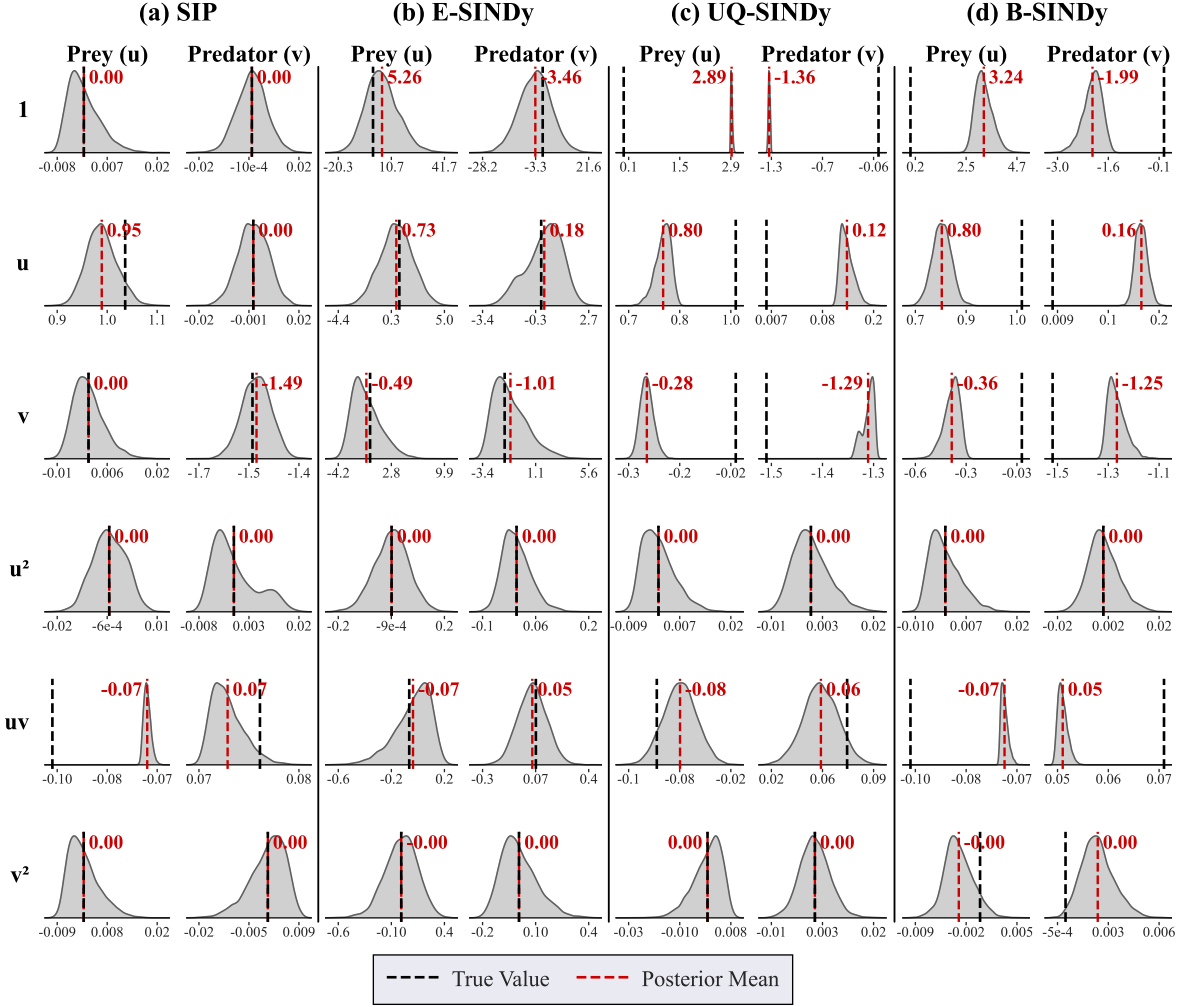


Figure S9: Posterior densities of recovered Lotka–Volterra coefficients in Scenario 3 (Lotka–Volterra with input variability only).

Chaotic Systems

For the chaotic systems (Lorenz, Van der Pol, and Rössler), the posterior distributions (Figures S11, S12, S14) demonstrate SIP’s ability to recover essential nonlinear coupling terms while suppressing spurious polynomial terms that would destabilize long-term predictions.

Hudson Bay Dataset

The posterior distributions for Hudson Bay (Figure S15) compare uncertainty quantification across methods. SIP correctly identifies the four essential Lotka–Volterra terms (u , v , uv for both equations) with coefficients concentrated around ecologically meaningful values, while E-SINDy, UQ-SINDy, and B-SINDy include spurious constant and additional polynomial terms that compromise model interpretability.

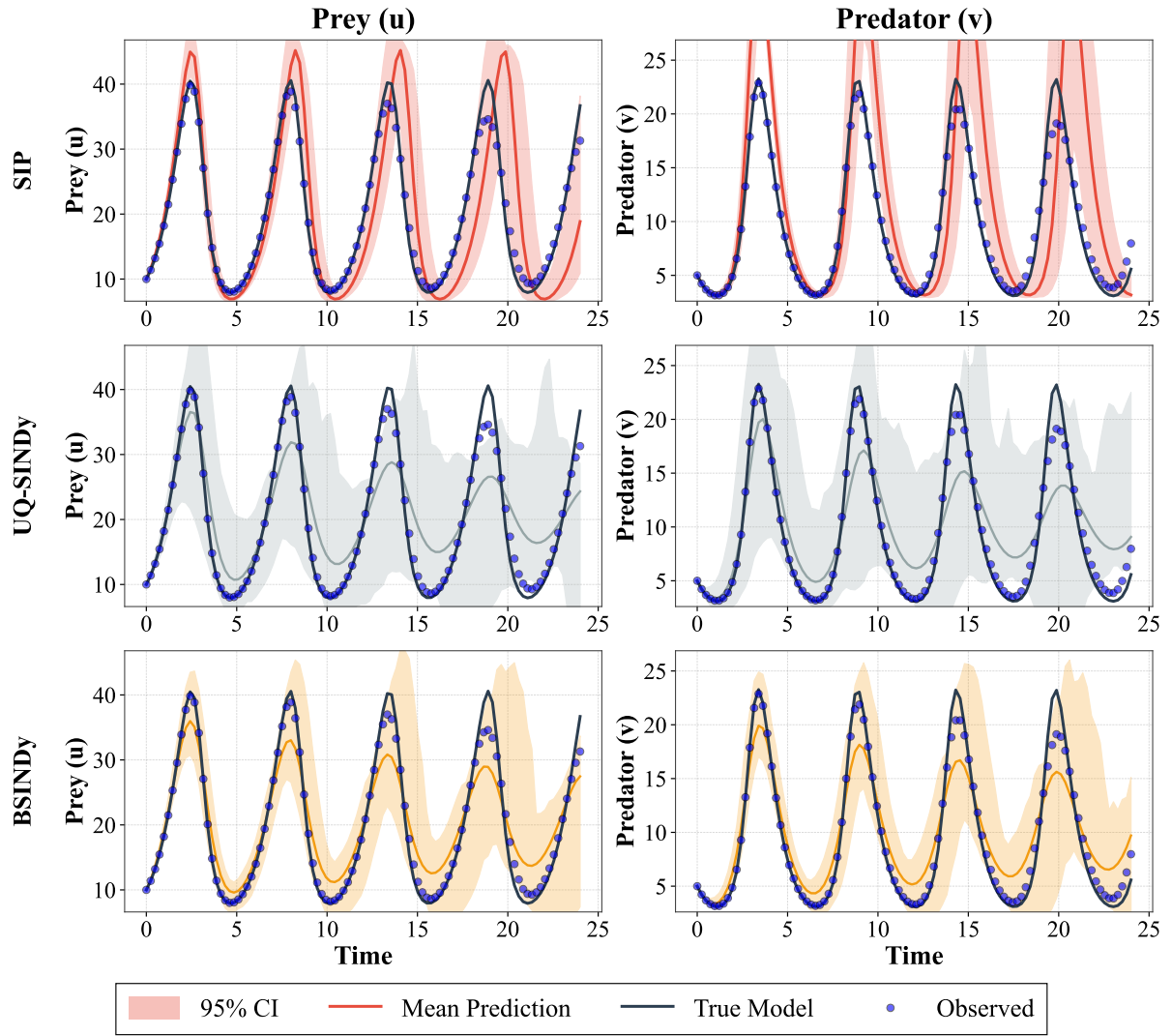


Figure S10: 95% confidence intervals for the Lotka–Volterra system under input variability across replicate experiments. Rows correspond to SIP, UQ-SINDy, and BSINDy; columns correspond to Prey u and Predator v .

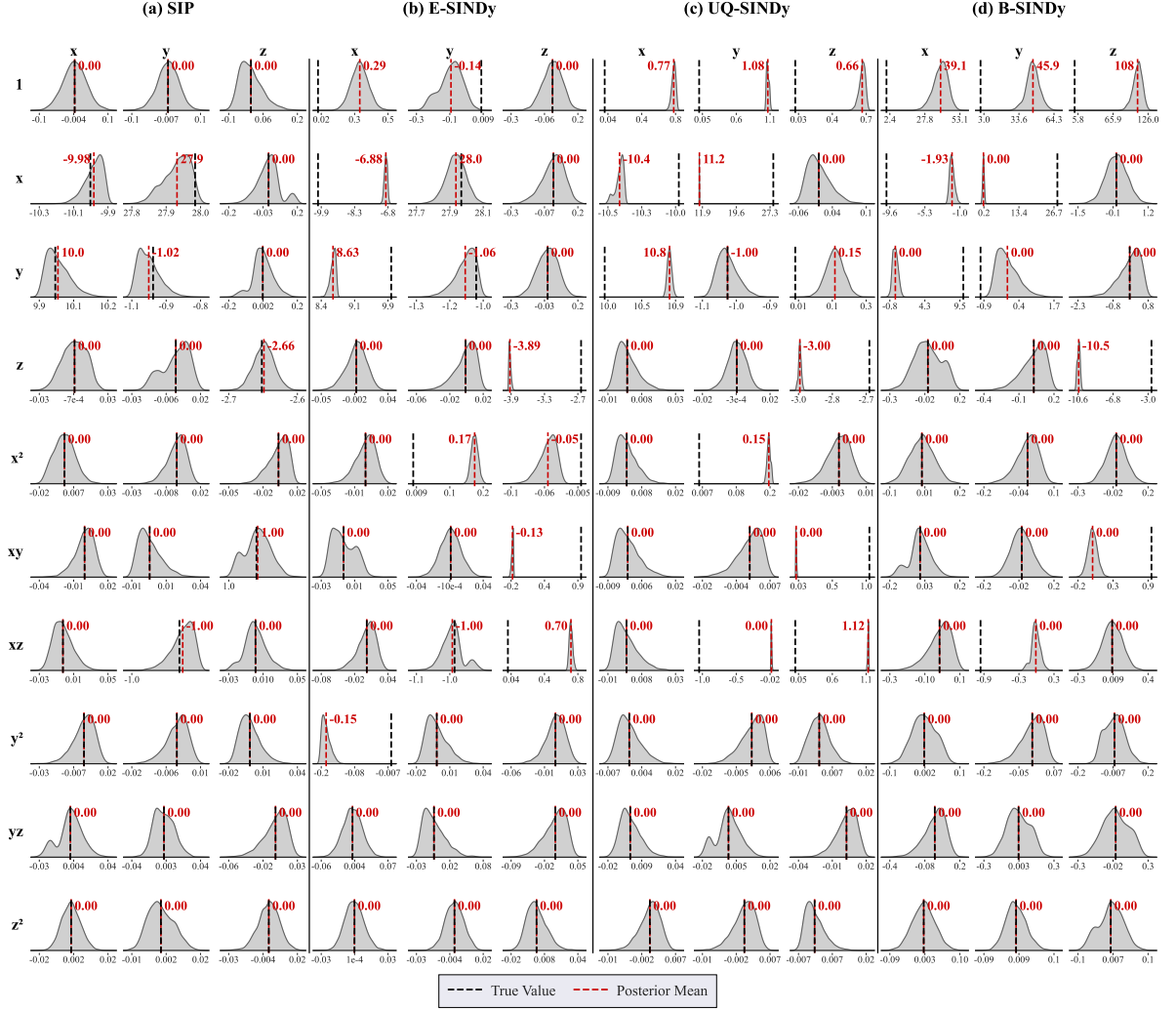


Figure S11: Posterior densities of recovered Lorenz system coefficients under different methods.

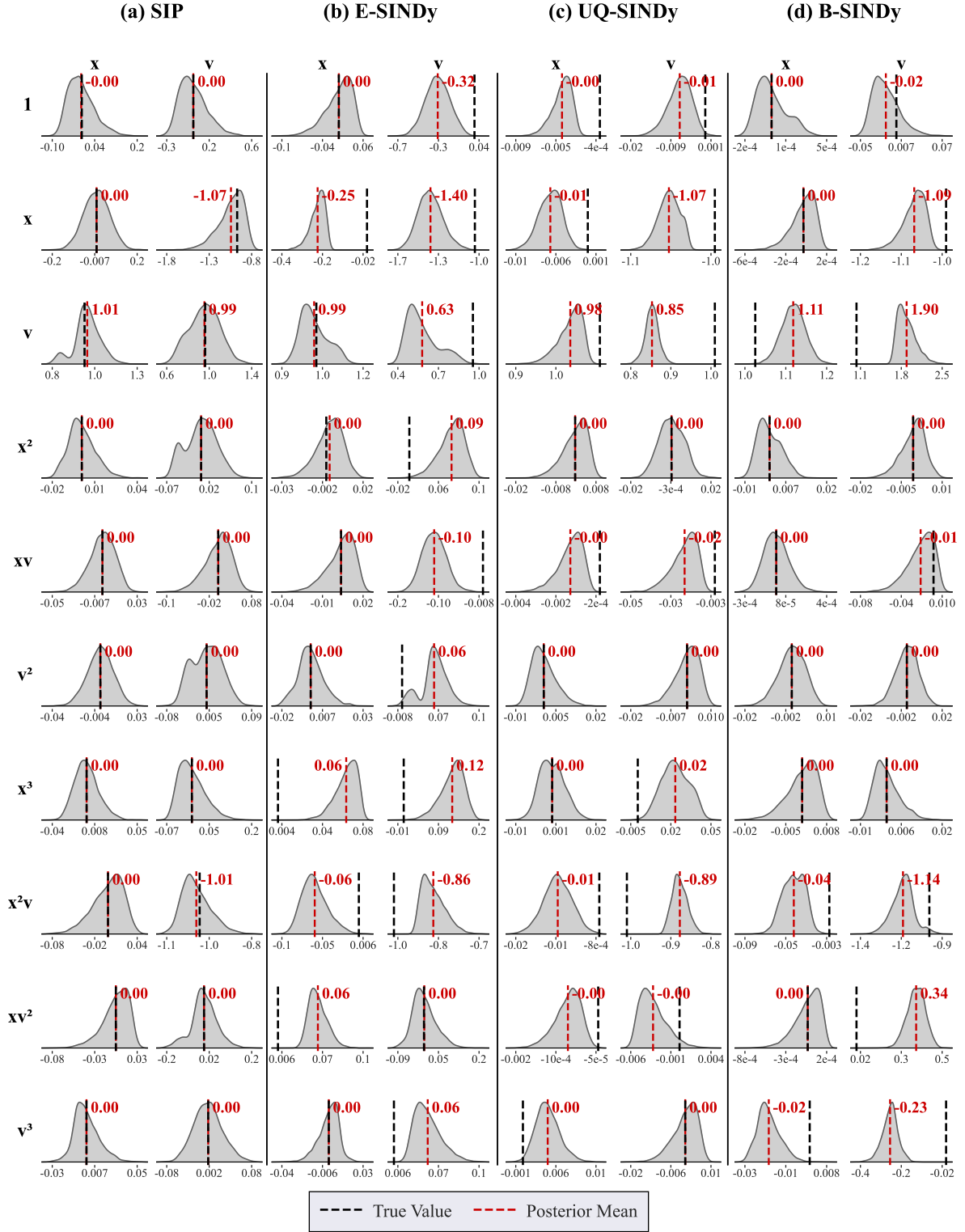


Figure S12: Posterior densities of recovered Van der Pol oscillator under different methods.

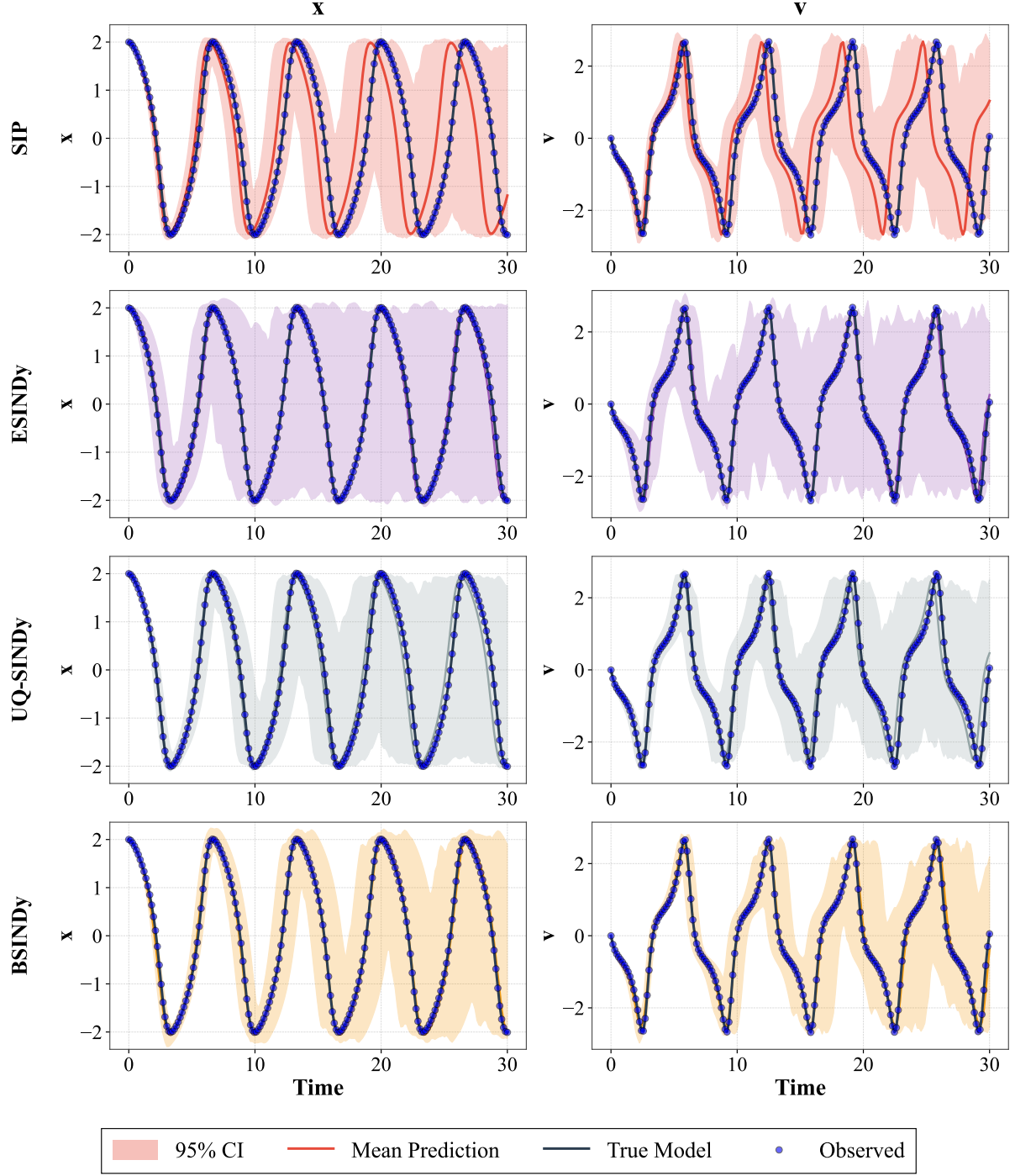


Figure S13: 95% confidence intervals for the Van der Pol oscillator under measurement noise. Rows correspond to SIP, ESINDy, UQ-SINDy, and BSINDy.

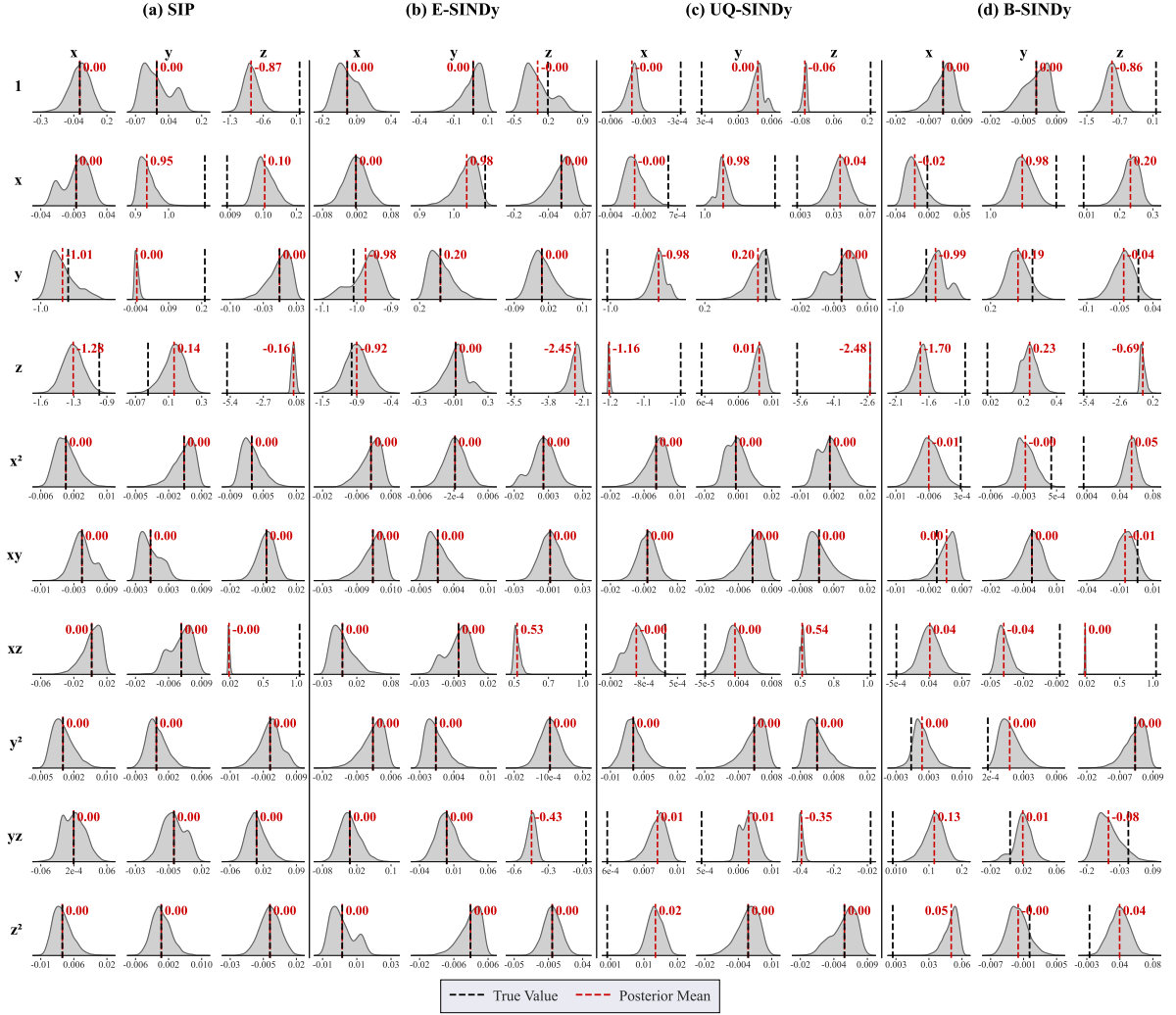


Figure S14: Recovered Posterior densities of coefficients for Rössler attractor.

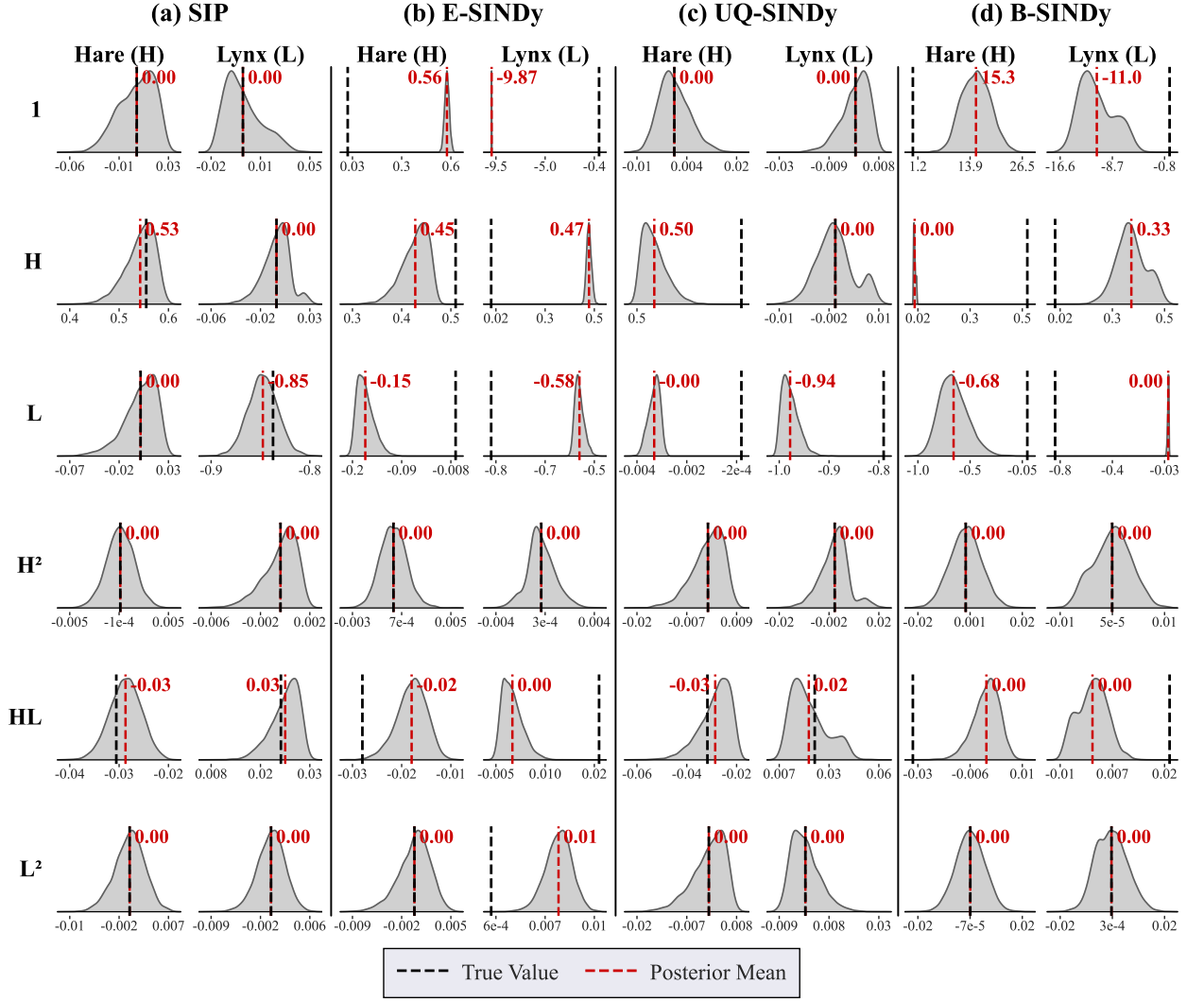


Figure S15: Recovered Posterior distributions of coefficients for Hudson Bay data.

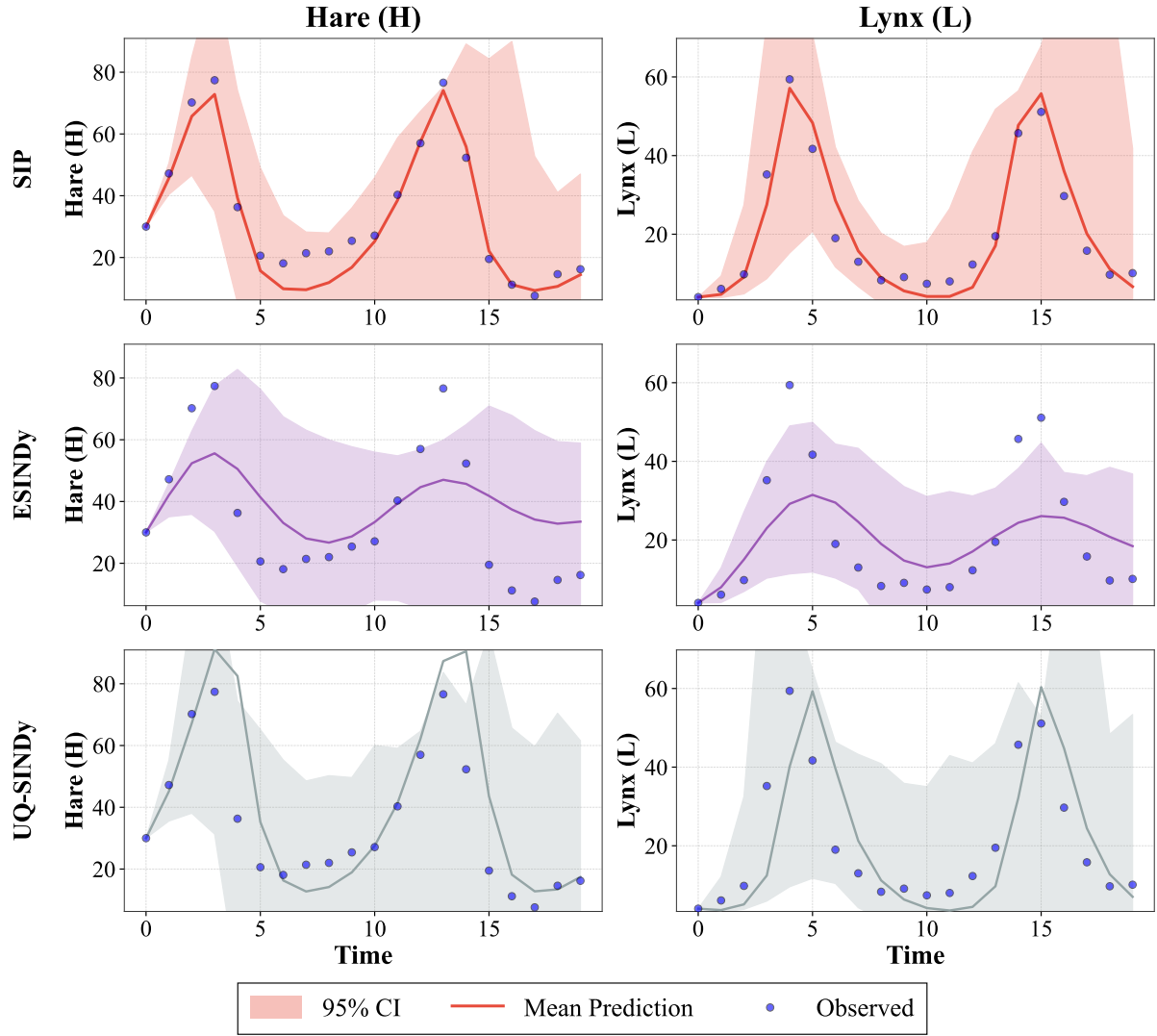


Figure S16: 95% confidence intervals for the Hudson Bay hare-lynx system using real ecological time-series data. Rows correspond to SIP, ESINDy, and UQ-SINDy; columns correspond to Hare H and Lynx L .

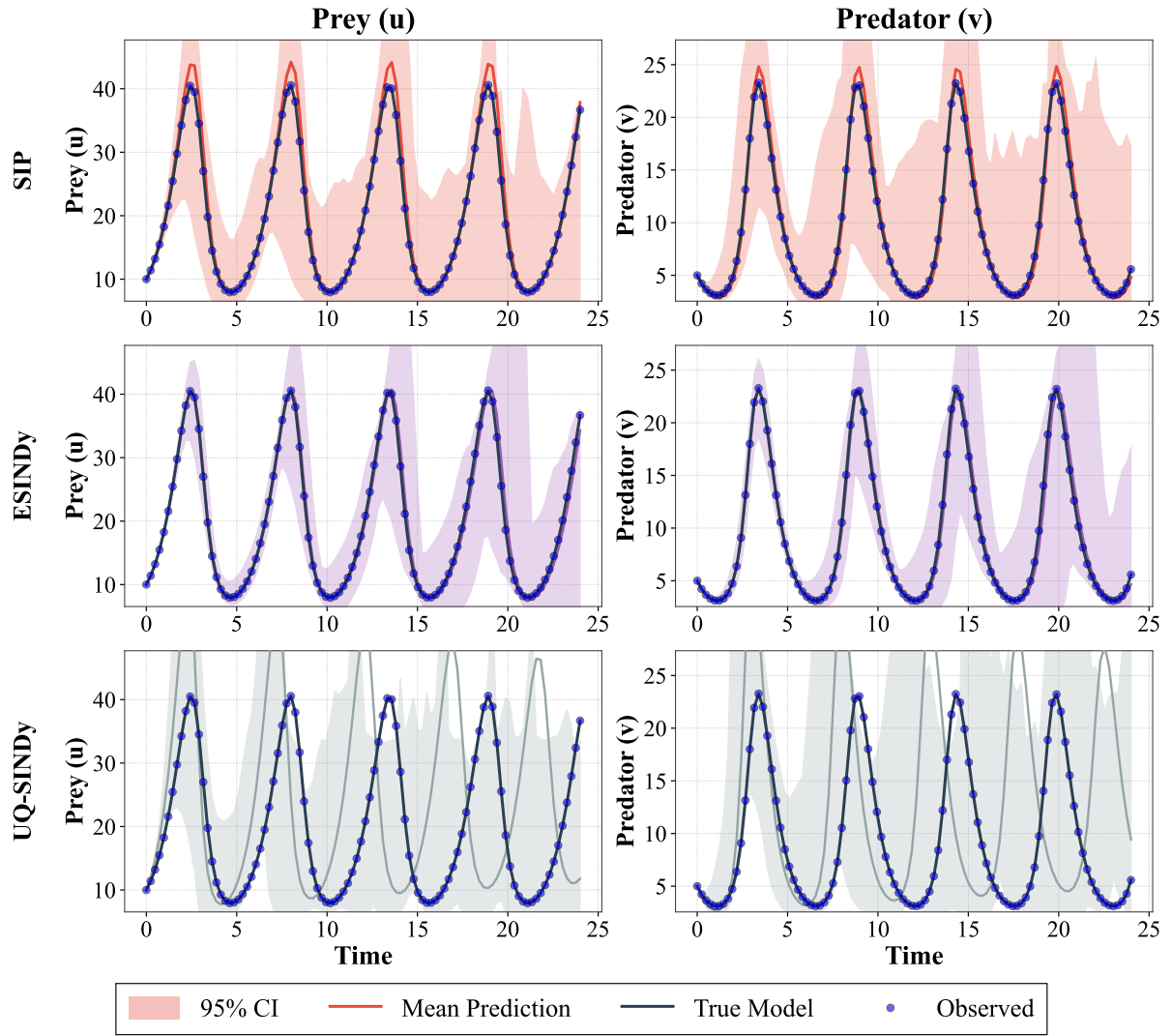


Figure S17: 95% confidence intervals for the Lotka–Volterra system under clean (noise-free) conditions. Rows correspond to SIP, ESINDy, and UQ-SINDy; columns correspond to Prey u and Predator v .

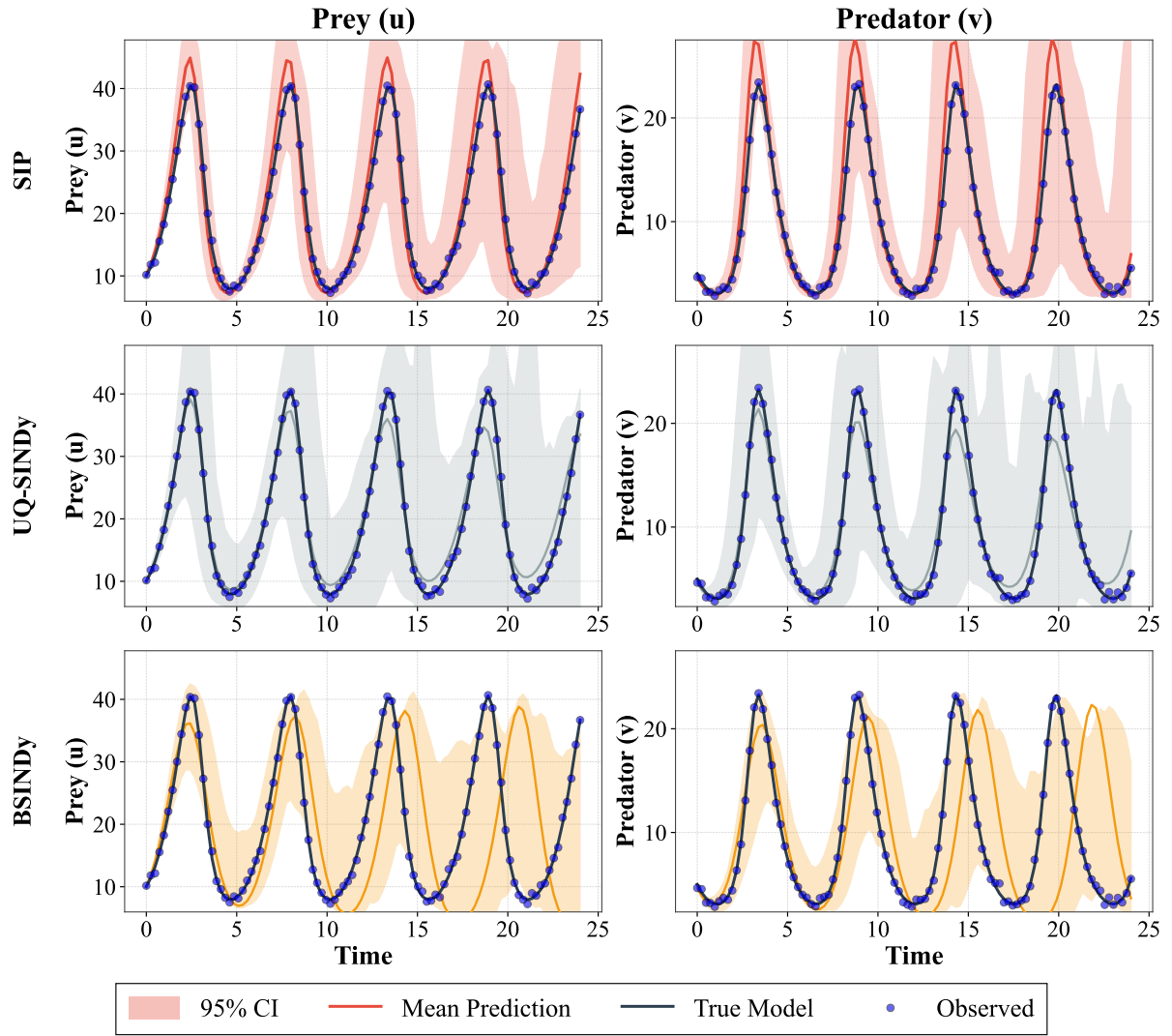


Figure S18: 95% confidence intervals for the Lotka–Volterra system under measurement noise. Rows correspond to SIP, UQ-SINDy, and BSINDy; columns correspond to Prey u and Predator v .

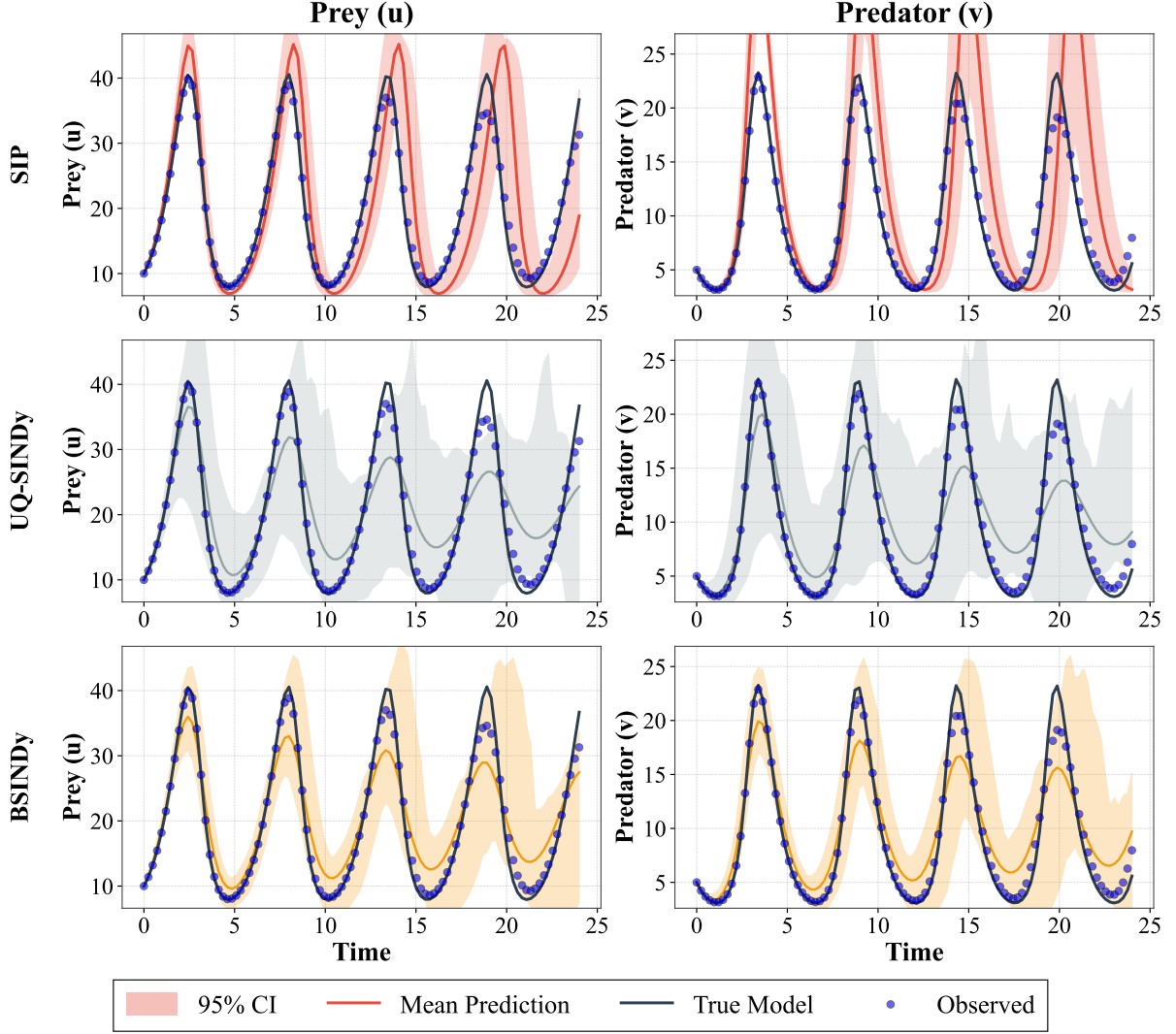


Figure S19: 95% confidence intervals for the Lotka–Volterra system under input variability across replicate experiments. Rows correspond to SIP, UQ-SINDy, and BSINDy; columns correspond to Prey u and Predator v .

Supplementary Note 7: SIP Performance Under Different Uncertainty Configurations

To provide intuition for SIP’s performance advantage, we construct three illustrative scenarios using the scalar forward model $Q(\lambda) = \lambda^p$ with $\lambda \in [-1, 1]$ for $p \in \{1, 5\}$. Each scenario demonstrates a distinct configuration of uncertainty sources, (1) input variability only, (2) measurement noise only, and (3) both simultaneously, and compares the push-forward posteriors produced by Standard Bayesian (SB) inference and SIP against the observed data distribution. This controlled setting reveals how each framework attributes observed spread to measurement noise versus input variability, and why SIP’s joint inference yields better distributional matching even when it does not perfectly recover the true noise level.

We denote input variability by σ_λ (the standard deviation of the input parameter distribution) and measurement noise by σ_ε . In Scenario 1, λ is drawn from a scaled Beta(2, 5) distribution on $[-1, 1]$, yielding input variability $\sigma_\lambda \approx 0.319$ with no measurement noise ($\sigma_\varepsilon = 0$). In Scenario 2, λ is sampled on a uniform grid over $[-1, 1]$ and Gaussian noise with $\sigma_\varepsilon = 0.1$ is added, with no input variability beyond the fixed grid. In Scenario 3, both sources are present simultaneously ($\sigma_\lambda \approx 0.319$, $\sigma_\varepsilon = 0.1$).

Supplementary Note 7.1: Scenario 1: Input Variability Only

In this scenario, λ is drawn from a scaled Beta(2, 5) distribution on $[-1, 1]$ ($\sigma_\lambda \approx 0.319$) with no measurement noise ($\sigma_\varepsilon = 0$). All spread in the observed quantity Q therefore arises from input variability alone.

As shown in Figure S20 ($p = 1$) and Figure S21 ($p = 5$), SB infers $\hat{\sigma}_\varepsilon = 0.291$ and $\hat{\sigma}_\varepsilon = 0.179$ respectively, misattributing input-driven spread as measurement noise and producing a push-forward that fails to match the observed distribution. SIP, with $\hat{\sigma}_\varepsilon = 0$ (correctly recovering the absence of measurement noise), produces a push-forward that closely matches the observed distribution in both cases.

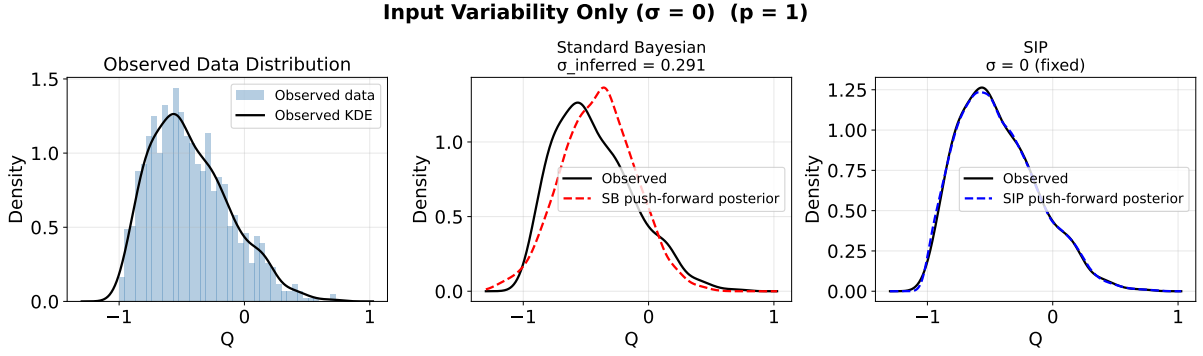


Figure S20: Scenario 1 ($p = 1$, input variability only, $\sigma_\varepsilon = 0$, $\sigma_\lambda \approx 0.319$): observed data distribution (left), SB push-forward posterior with $\hat{\sigma}_\varepsilon = 0.291$ (centre), and SIP push-forward posterior with $\hat{\sigma}_\varepsilon = 0$ (right).

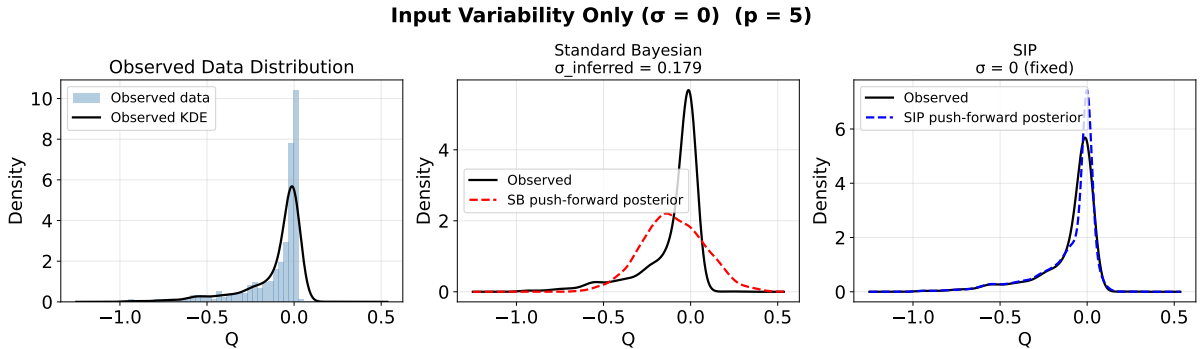


Figure S21: Scenario 1 ($p = 5$, input variability only, $\sigma_\varepsilon = 0$, $\sigma_\lambda \approx 0.319$): observed data distribution (left), SB push-forward posterior with $\hat{\sigma}_\varepsilon = 0.179$ (centre), and SIP push-forward posterior with $\hat{\sigma}_\varepsilon = 0$ (right).

Supplementary Note 7.2: Scenario 2: Measurement Noise Only

Here λ is sampled on a uniform grid over $[-1, 1]$ (no input variability) and Gaussian noise with $\sigma_\varepsilon = 0.1$ is added. All spread beyond the forward model output is attributable to measurement noise.

As shown in Figure S22 ($p = 1$) and Figure S23 ($p = 5$), SB infers $\hat{\sigma}_\varepsilon = 0.299$ and $\hat{\sigma}_\varepsilon = 0.294$ respectively — nearly three times the true noise level — by conflating the spread of the uniform input distribution with noise. SIP recovers $\hat{\sigma}_\varepsilon = 0.158$ ($p = 1$) and $\hat{\sigma}_\varepsilon = 0.160$ ($p = 5$). While these estimates are closer to $\sigma_\varepsilon = 0.1$ than SB, they are not exact, reflecting SIP’s inability to fully isolate the noise contribution from input variability. Nevertheless, SIP’s push-forward aligns substantially better with the observed distribution than SB in both cases.

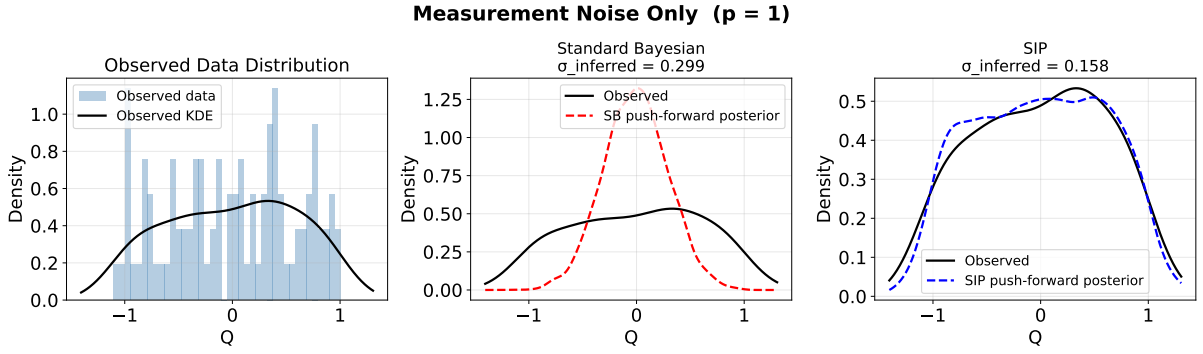


Figure S22: Scenario 2 ($p = 1$, measurement noise only, $\sigma_\varepsilon = 0.1$): observed data distribution (left), SB push-forward posterior with $\hat{\sigma}_\varepsilon = 0.299$ (centre), and SIP push-forward posterior with $\hat{\sigma}_\varepsilon = 0.158$ (right).

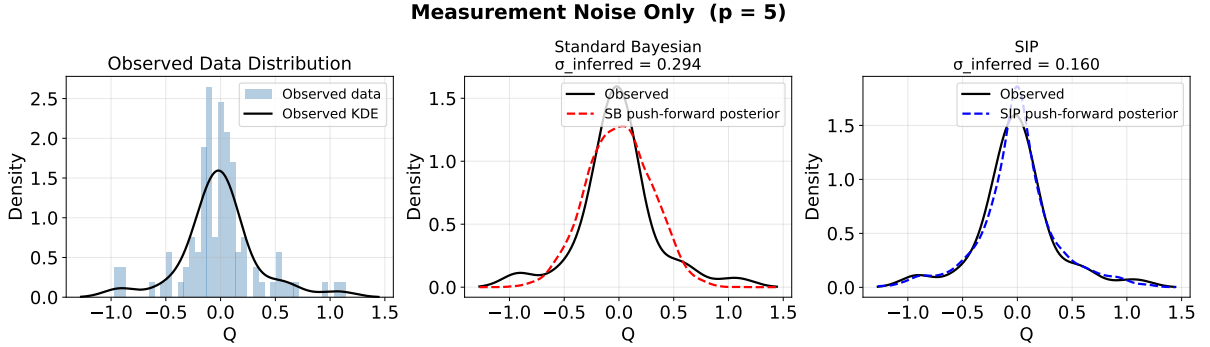


Figure S23: Scenario 2 ($p = 5$, measurement noise only, $\sigma_\varepsilon = 0.1$): observed data distribution (left), SB push-forward posterior with $\hat{\sigma}_\varepsilon = 0.294$ (centre), and SIP push-forward posterior with $\hat{\sigma}_\varepsilon = 0.160$ (right).

Supplementary Note 7.3: Scenario 3: Both Input Variability and Noise

Here λ is drawn from the scaled Beta(2, 5) distribution ($\sigma_\lambda \approx 0.319$) and Gaussian noise with $\sigma_\varepsilon = 0.1$ is added simultaneously. As shown in Figure S24 ($p = 1$) and Figure S25 ($p = 5$), SB infers $\hat{\sigma}_\varepsilon = 0.291$ and $\hat{\sigma}_\varepsilon = 0.207$ respectively, overestimating the true noise

by conflating both uncertainty sources. SIP infers $\hat{\sigma}_\varepsilon = 0.153$ ($p = 1$) and $\hat{\sigma}_\varepsilon = 0.150$ ($p = 5$), closer to $\sigma_\varepsilon = 0.1$, though again not exact. Importantly, SIP’s push-forward more faithfully matches the observed distribution in both cases, demonstrating that its joint $(\lambda, \sigma_\varepsilon)$ framework better accommodates mixed uncertainty sources than SB, even without explicitly separating them.

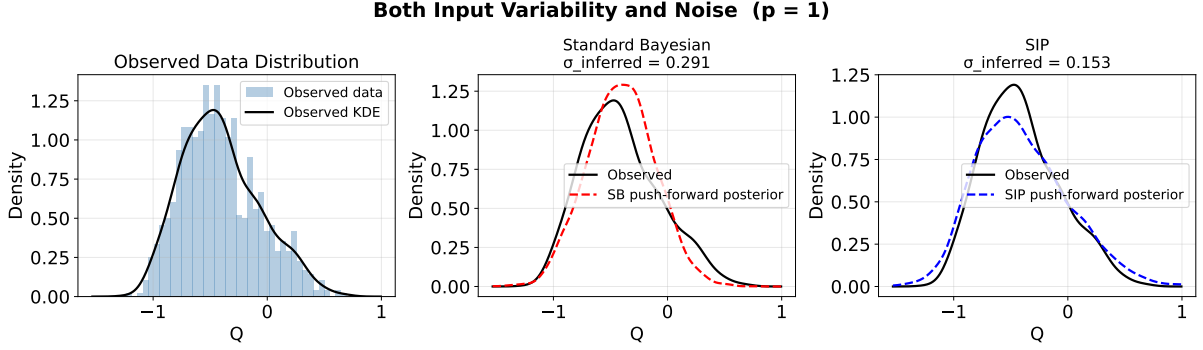


Figure S24: Scenario 3 ($p = 1$, both input variability and measurement noise, $\sigma_\lambda \approx 0.319$, $\sigma_\varepsilon = 0.1$): observed data distribution (left), SB push-forward posterior with $\hat{\sigma}_\varepsilon = 0.291$ (centre), and SIP push-forward posterior with $\hat{\sigma}_\varepsilon = 0.153$ (right).

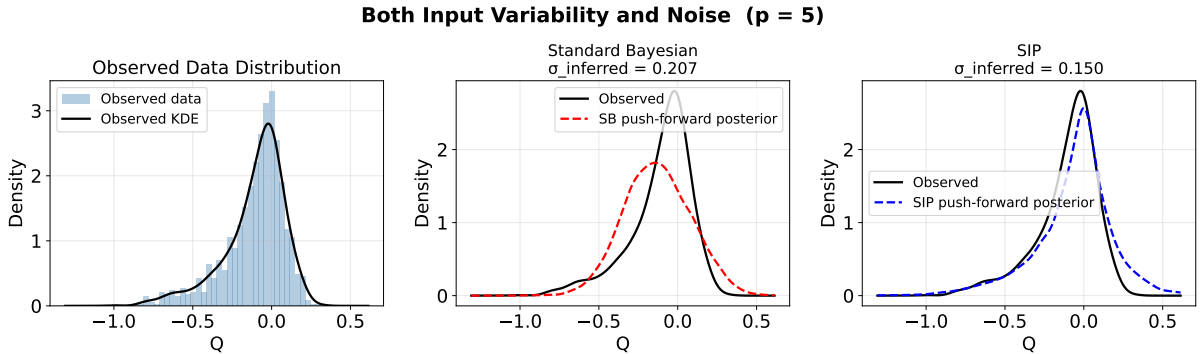


Figure S25: Scenario 3 ($p = 5$, both input variability and measurement noise, $\sigma_\lambda \approx 0.319$, $\sigma_\varepsilon = 0.1$): observed data distribution (left), SB push-forward posterior with $\hat{\sigma}_\varepsilon = 0.207$ (centre), and SIP push-forward posterior with $\hat{\sigma}_\varepsilon = 0.150$ (right).

Table S3 summarises the inferred $\hat{\sigma}_\varepsilon$ values across all scenarios. SB consistently overestimates σ_ε by conflating input variability with noise, whereas SIP recovers substantially closer estimates in all cases. We note, however, that SIP does not explicitly decouple the two uncertainty sources: the joint $(\lambda, \sigma_\varepsilon)$ posterior (Equation 6 in the main text) accommodates both simultaneously, with σ_ε absorbing variability not explained by the inferred input distribution. As a result, SIP’s $\hat{\sigma}_\varepsilon$ values do not precisely match σ_ε , but its push-forward posteriors consistently match the observed data distributions more faithfully than SB across all scenarios, which is the operationally relevant criterion for physics discovery. Developing hierarchical inference to explicitly separate measurement noise from input variability is identified as an important direction for future work.

Table S3: Inferred noise parameter $\hat{\sigma}_\varepsilon$ for SB and SIP across all three scenarios and both values of p . Input variability $\sigma_\lambda \approx 0.319$ (from scaled Beta(2, 5) on $[-1, 1]$) is present in Scenarios 1 and 3. True measurement noise σ_ε is 0 for Scenario 1 and 0.1 for Scenarios 2 and 3.

Scenario	p	σ_λ	σ_ε	SB $\hat{\sigma}_\varepsilon$	SIP $\hat{\sigma}_\varepsilon$
1 (variability only)	1	0.319	0.000	0.291	0.000
1 (variability only)	5	0.319	0.000	0.179	0.000
2 (noise only)	1	—	0.100	0.299	0.158
2 (noise only)	5	—	0.100	0.294	0.160
3 (variability + noise)	1	0.319	0.100	0.291	0.153
3 (variability + noise)	5	0.319	0.100	0.207	0.150

Supplementary Note 8: Discovered Dynamics for All Scenarios

This section presents the complete discovered equations for all ten scenarios and all eight methods, enabling detailed inspection of structural differences and coefficient accuracy.

Table S4: Discovered equations for Scenarios 1-3 (Lotka-Volterra systems: Clean, Noisy, and Variability)

Scenario	Method	Discovered Equations
Scenario 1: Lotka-Volterra (Clean Data)		
	True	$\dot{u} = 1.0u - 0.1uv$ $\dot{v} = -1.5v + 0.075uv$
	SIP	$\dot{u} = 0.997u - 0.0975uv$ $\dot{v} = -1.521v + 0.072uv$
	SINDy	$\dot{u} = 0.329 + 0.979u - 0.034v - 0.098uv$ $\dot{v} = -0.251 + 0.013u - 1.467v + 0.073uv$
	BSINDy	$\dot{u} = -16.24 + 2.44u + 1.03v - 0.041u^2 - 0.053uv - 0.073v^2$ $\dot{v} = 13.93 - 0.947u - 3.01v + 0.027u^2 + 0.039uv + 0.090v^2$
	ESINDy	$\dot{u} = 0.329 + 0.979u - 0.034v - 0.098uv$ $\dot{v} = -1.487v + 0.074uv$
	UQ-SINDy	$\dot{u} = 1.147u + 0.054v - 0.091uv$ $\dot{v} = -1.684v + 0.069uv$
	Weak-SINDy	$\dot{u} = 0.721 + 0.958u - 0.074v - 0.096uv$ $\dot{v} = -0.551 + 0.028u - 1.432v + 0.072uv$
	ODEFormer	$\dot{u} = 0.945u - 0.084uv$ $\dot{v} = -1.660v + 0.092uv$
	SPL	$\dot{u} = 1.050u - 0.063v - 0.002u^2 - 0.097uv$ $\dot{v} = -0.031u - 1.442v + 0.001u^2 + 0.072uv$
Scenario 2: Lotka-Volterra (Noisy Data)		
	True	$\dot{u} = 1.0u - 0.1uv$ $\dot{v} = -1.5v + 0.075uv$
	SIP	$\dot{u} = 1.023u - 0.0925uv$ $\dot{v} = -1.513v + 0.0731uv$
	SINDy	$\dot{u} = 0.332 + 0.982u - 0.012v - 0.099uv$ $\dot{v} = 0.050 - 1.420v + 0.071uv$
	BSINDy	$\dot{u} = 1.281u - 0.637v - 0.013u^2 - 0.061uv$ $\dot{v} = 0.060u - 1.626v + 0.004u^2 + 0.044uv + 0.030v^2$
	ESINDy	$\dot{u} = -0.244 + 0.983u + 0.134v - 0.100uv$ $\dot{v} = 0.519 - 1.536v + 0.070uv$
	UQ-SINDy	$\dot{u} = 0.001 + 1.007u - 0.085v - 0.097uv$ $\dot{v} = 0.012 + 0.031u - 1.473v + 0.069uv$
	Weak-SINDy	$\dot{u} = -0.540 + 1.018u + 0.034v - 0.101uv$ $\dot{v} = -1.303 + 0.058u - 1.320v + 0.067uv$
	ODEFormer	$\dot{u} = 0.947u - 0.079uv$ $\dot{v} = -1.714v + 0.096uv$
	SPL	$\dot{u} = 1 + 0.669u + 0.041u^2v^{-1} - 0.092uv + 0.004v^2$ $\dot{v} = 0.075uv - 1.510v$
Scenario 3: Lotka-Volterra (Variability)		
	True	$\dot{u} = 1.0u - 0.1uv$ $\dot{v} = -1.5v + 0.075uv$
	SIP	$\dot{u} = 0.952u - 0.073uv$ $\dot{v} = -1.488v + 0.073uv$
	SINDy	$\dot{u} = 2.140 + 0.895u - 0.203v - 0.087uv$ $\dot{v} = -1.023 + 0.047u - 1.379v + 0.069uv$
	BSINDy	$\dot{u} = 3.244 + 0.803u - 0.365v + 0.0004u^2 - 0.073uv - 0.003v^2$ $\dot{v} = -1.991 + 0.161u - 1.250v + 0.054uv + 0.002v^2$
	ESINDy	$\dot{u} = 5.257 + 0.732u - 0.495v - 0.072uv$ $\dot{v} = -3.463 + 0.176u - 1.010v + 0.051uv$
	UQ-SINDy	$\dot{u} = 2.892 + 0.796u - 0.279v - 0.076uv$ $\dot{v} = -1.358 + 0.119u - 1.295v + 0.057uv$
	Weak-SINDy	$\dot{u} = 0.167 + 1.009u - 0.024v - 0.094uv$ $\dot{v} = -0.649 + 0.032u - 1.416v + 0.071uv$
	ODEFormer	$\dot{u} = 0.947u - 0.071uv$ $\dot{v} = 1.031u - 19.411(1 + 0.0007u)^2$
	SPL	$\dot{u} = 1 + 0.942u - 0.219v - 0.084uv$ $\dot{v} = -1.343v + 0.002u^2 + 0.061uv$

Table S5: Discovered equations for Scenarios 4-5 (Lotka-Volterra combined noise and variability, and Lorenz System)

Scenario	Method	Discovered Equations
Scenario 4: Lotka-Volterra (Combined)		
	True	$\dot{u} = 1.0u - 0.1uv$ $\dot{v} = -1.5v + 0.075uv$
	SIP	$\dot{u} = 0.995u - 0.105uv$ $\dot{v} = -1.512v + 0.073uv$
	SINDy	$\dot{u} = 0.342 + 0.991u - 0.001v - 0.088uv$ $\dot{v} = -0.646 + 0.018u - 1.410v + 0.068uv$
	BSINDy	$\dot{u} = 3.372 + 0.768u - 0.330v + 0.002u^2 - 0.074uv - 0.004v^2$ $\dot{v} = -2.027 + 0.169u - 1.261v + 0.053uv + 0.003v^2$
	ESINDy	$\dot{u} = 2.456 + 0.888u - 0.186v - 0.081uv$ $\dot{v} = -5.518 + 0.282u - 0.897v + 0.044uv$
	UQ-SINDy	$\dot{u} = 0.740 + 0.896u - 0.050v - 0.079uv$ $\dot{v} = -0.882 + 0.096u - 1.345v + 0.057uv$
	Weak-SINDy	$\dot{u} = 0.723 + 0.963u - 0.070v - 0.092uv$ $\dot{v} = -0.918 + 0.046u - 1.375v + 0.069uv$
	ODEFormer	$\dot{u} = 0.947u - 0.071uv$ $\dot{v} = 1.031u - 19.411(1 + 0.0007u)^2$
	SPL	$\dot{u} = 1 + 0.949u - 1.730vu^{-1} - 0.090uv$ $\dot{v} = -1.219v - vu^{-1} + 0.002u^2 + 0.058uv$
Scenario 5: Lorenz System (Noisy)		
	True	$\dot{x} = -10x + 10y$ $\dot{y} = 28x - y - xz$ $\dot{z} = xy - 2.667z$
	SIP	$\dot{x} = -10.00x + 10.11y$ $\dot{y} = 27.95x - 1.108y - 0.957xz + 0.05y$ $\dot{z} = 1.02xy - 2.672z$
	SINDy	$\dot{x} = 4.549 - 10.98x + 11.12y - 0.162z + 0.146xy + 0.113xz - 0.122y^2 - 0.119yz$ $\dot{y} = 9.122 + 9.921x + 5.059y - 0.717z + 0.172x^2 - 0.132xy - 0.470xz - 0.130yz$ $\dot{z} = 14.22 - 3.580x + 2.306y - 1.481z + 1.012x^2 - 0.136xy + 0.112xz + 0.162y^2 - 0.080yz - 0.074z^2$
	BSINDy	$\dot{x} = 39.14 - 1.930x + 2.367y - 2.331z + 0.298xy - 0.205y^2 + 0.028z^2$ $\dot{y} = 45.89 - 3.282z + 0.290xy - 0.080xz - 0.194y^2 + 0.079yz + 0.049z^2$ $\dot{z} = 108.3 - 10.45z - 0.368x^2 + 0.400xy + 0.228z^2$
	ESINDy	$\dot{x} = 3.822 - 6.878x + 8.629y - 0.048z + 0.165xy - 0.132y^2 - 0.053yz$ $\dot{y} = -2.216 + 9.482x + 5.540y - 0.129z - 0.007x^2 - 0.452xz - 0.149yz$ $\dot{z} = 32.91 + 0.380x - 0.786y - 3.894z + 0.696x^2 + 0.151xy + 0.099y^2 + 0.011yz$
	UQ-SINDy	$\dot{x} = 0.766 - 10.44x + 10.77y + 0.191xy$ $\dot{y} = 1.082 + 11.21x + 4.151y + 0.164x^2 - 0.502xz$ $\dot{z} = 0.660 + 0.153y + 1.121x^2 - 0.276xy + 0.202y^2$
	Weak-SINDy	$\dot{x} = 57.41 - 7.494x + 6.209y - 5.344z - 0.500x^2 + 0.793xy + 0.127xz - 0.343y^2 - 0.076yz + 0.127z^2$ $\dot{y} = 89.65 - 4.493x + 3.471y - 8.998z - 0.642x^2 + 0.879xy - 0.328y^2 + 0.210z^2$ $\dot{z} = 100.0 - 1.925y - 9.927z + 0.327xy + 0.074yz + 0.191z^2$
	ODEFormer	$\dot{x} = -4.976x + 4.653y$ $\dot{y} = 22.88(-0.052 + 0.063z)^{-1} - 0.958z$ $\dot{z} = -9.559x + 9.372y$
	SPL	$\dot{x} = 141.20xz^{-2}$ $\dot{y} = -x + 224.06xz^{-2}$ $\dot{z} = 1256.39y^2z^{-4}$

Table S6: Discovered equations for Scenarios 6-7 (Van der Pol and Rössler systems)

Scenario	Method	Discovered Equations
Scenario 6: Van der Pol Oscillator (Noisy)		
	True	$\dot{x} = v$ $\dot{v} = -x + v - x^2v$
	SIP	$\dot{x} = -0.002 + 1.013v$ $\dot{v} = -1.067x + 0.993v - 1.012x^2v$
	SINDy	$\dot{x} = -0.251x + 0.992v + 0.061x^3 - 0.059x^2v + 0.059xv^2$ $\dot{v} = -0.315 - 1.398x + 0.626v + 0.089x^2 - 0.105xv + 0.058v^2 + 0.123x^3 - 0.859x^2v + 0.056v^3$
	BSINDy	$\dot{x} = 1.107v - 0.037x^2v - 0.019v^3$ $\dot{v} = -0.015 - 1.091x + 1.898v - 0.014xv - 1.145x^2v + 0.336xv^2 - 0.235v^3$
	ESINDy	$\dot{x} = -0.253x + 0.992v + 0.062x^3 - 0.059x^2v + 0.060xv^2$ $\dot{v} = -0.316 - 1.400x + 0.628v + 0.089x^2 - 0.105xv + 0.058v^2 + 0.123x^3 - 0.860x^2v + 0.055v^3$
	UQ-SINDy	$\dot{x} = -0.004 - 0.007x + 0.981v - 0.001xv - 0.010x^2v - 0.001xv^2 + 0.004v^3$ $\dot{v} = -0.006 - 1.068x + 0.851v - 0.020xv + 0.025x^3 - 0.890x^2v - 0.003xv^2$
	Weak-SINDy	$\dot{x} = 0.986v$ $\dot{v} = -1.034x + 1.492v - 1.049x^2v + 0.184xv^2 - 0.134v^3$
	ODEFormer	$\dot{x} = 0.921v + 0.855 \sin(1.111 + 2.138v)$ $\dot{v} = -0.965x + 1.346 \sin(-1.134 + 2.188x)$
	SPL	$\dot{x} = v$ $\dot{v} = -0.293x^3 - 0.205x^4v$
Scenario 7: Rössler System (Noisy)		
	True	$\dot{x} = -y - z$ $\dot{y} = x + 0.2y$ $\dot{z} = 0.2 + xz - 5.7z$
	SIP	$\dot{x} = -1.007y - 1.281z$ $\dot{y} = 0.947x + 0.139z$ $\dot{z} = -0.867 + 0.101x - 0.157z - 0.001xz$
	SINDy	$\dot{x} = -0.979y - 0.920z$ $\dot{y} = 0.943x$ $\dot{z} = -2.446z + 0.534xz - 0.434yz$
	BSINDy	$\dot{x} = -0.017x - 0.988y - 1.699z - 0.007x^2 + 0.003xy + 0.038xz + 0.002y^2 + 0.129yz + 0.047z^2$ $\dot{y} = 0.979x + 0.191y + 0.227z - 0.003x^2 - 0.038xz + 0.002y^2 + 0.015yz - 0.002z^2$ $\dot{z} = -0.858 + 0.198x - 0.040y - 0.690z + 0.053x^2 - 0.009xy - 0.076yz + 0.037z^2$
	ESINDy	$\dot{x} = -0.979y - 0.920z$ $\dot{y} = 0.981x$ $\dot{z} = -2.446z + 0.534xz - 0.434yz$
	UQ-SINDy	$\dot{x} = -0.004 - 0.002x - 0.979y - 1.165z - 0.001xz + 0.009yz + 0.015z^2$ $\dot{y} = 0.005 + 0.983x + 0.198y + 0.010z + 0.004xz + 0.007yz$ $\dot{z} = -0.058 + 0.040x - 2.481z + 0.537xz - 0.347yz$
	Weak-SINDy	$\dot{x} = -0.974y - 0.709z$ $\dot{y} = 0.937x$ $\dot{z} = 0.274x$
	ODEFormer	$\dot{x} = -1.015y - 6.342z$ $\dot{y} = 0.873x - 1.436(1.090 + 20.45z)^{-1}$ $\dot{z} = 0.141z(4.857x - 4.134z)$
	SPL	$\dot{x} = -0.984y - 1.214z + 0.037z^2$ $\dot{y} = x + 0.180y$ $\dot{z} = 0.036x^2 - 0.200yz$

Table S7: Discovered equations for Scenarios 8–10 (Hudson Bay and Infiltration systems). For infiltration scenarios (Scenarios 9–10), SIP recovers the full second-order PDE governing capillary-driven flow, whereas other methods recover a reduced first-order ODE approximation. Both representations are reported in their native forms. The DiCE score for SIP is computed by comparing the identified PDE terms against the true PDE structure.

Scenario	Method	Discovered Equations
Scenario 8: Hudson Bay System (Real Data) where H and L are the Hare and Lynx population respectively		
	True	$\dot{H} = 0.545H - 0.0282HL, \quad \dot{L} = -0.842L + 0.0261HL$
	SIP	$\dot{H} = 0.530H - 0.0261HL, \quad \dot{L} = -0.853L + 0.0272HL$
	SINDy	$\dot{H} = 0.555 + 0.450H - 0.154L - 0.018HL, \quad \dot{L} = 3.035 + 0.112H - 1.261L + 0.016HL + 0.011L^2$
	BSINDy	$\dot{H} = 15.3 - 0.682L, \quad \dot{L} = -11.0 + 0.334H$
	ESINDy	$\dot{H} = 0.555 + 0.450H - 0.154L - 0.018HL, \quad \dot{L} = -9.868 + 0.469H - 0.583L + 0.009L^2$
	UQ-SINDy	$\dot{H} = 0.502H - 0.003L - 0.025HL, \quad \dot{L} = -0.944L + 0.023HL$
	Weak-SINDy	$\dot{H} = 31.13 + 0.011H - 2.431L + 0.030L^2, \quad \dot{L} = -8.304 + 0.296H - 0.066L$
	ODEFormer	$\dot{H} = 2.065(0.051 + 0.002L)^{-1} - 40.99 \sin(1.081 + 0.033H), \quad \dot{L} = 0.452H - 1.234L$
	SPL	$\dot{H} = -1 - H^{-1} + 0.639H - 0.549L - 0.020HL + 5.469L^2H^{-2} + L^2H^{-3}, \quad \dot{L} = 1 - 19.106LH^{-1} + 0.015HL$
Scenario 9: High-Viscous Infiltration (Real Data), where H is the infiltration height		
	SIP	$979.8 H \ddot{H} + 991.5 \dot{H}^2 + 2.258 \times 10^4 H \dot{H} + 9613 H = 99.97$
	SINDy	$\dot{H} = -0.064 + 91.33H - 10583.87H^2 + 325580.27H^3$
	BSINDy	$\dot{H} = -0.195 + 201H - 3.58 \times 10^4 H^2 + 2.47 \times 10^6 H^3 - 6.08 \times 10^7 H^4$
	ESINDy	$\dot{H} = 0.0473 + 25.87H - 1837.68H^2$
	UQ-SINDy	$\dot{H} = -7.089H - 0.311H^2 - 0.027H^3 - 0.083H^4$
	Weak-SINDy	$\dot{H} = -0.061 + 90.06H - 10450.22H^2 + 321362.98H^3$
	ODEFormer	$\dot{H} = 157.34 \sin(13.29 + 0.139H) - 2.741H$
	SPL	$\dot{H} = 0.103 - H$
Scenario 10: Low-Viscous Infiltration (Real Data)		
	SIP	$673.8 H \ddot{H} + 680.5 \dot{H}^2 + 5057.74 H \dot{H} + 6977.3 H = 49.4$
	SINDy	$\dot{H} = 0.000307 + 0.0351H - 25.90H^2 + 1962.21H^3$
	BSINDy	$\dot{H} = 0.000365 - 11.1H^2 + 8.22 \times 10^4 H^4$
	ESINDy	$\dot{H} = 0.000329 + 0.0352H - 25.92H^2 + 1962.22H^3$
	UQ-SINDy	$\dot{H} = -0.0532H + 1.161H^2 + 0.0004H^3 + 0.0381H^4$
	Weak-SINDy	$\dot{H} = 0.035H - 25.92H^2 + 1962.22H^3$
	ODEFormer	$\dot{H} = 0.000359 + 0.0352H - 25.92H^2 + 1962.22H^3$
	SPL	$\dot{H} = H^4$