The Encoding-Learning Tradeoff in Scientific Machine Learning: A Risk-Asymmetric Study of Brusselator Reaction-Diffusion Forecasting

Vejandla Sampath Kumar [1] **Prathamesh Dinesh Joshi** [2] **Raj Dandekar** [2] **Rajat Dandekar** [2] **Sreedath Panat** [2]

Abstract

Hybrid modeling blends physics and data, but the risks of this balance are rarely quantified. Data-driven models struggle to extrapolate, while physics-informed models can fail catastrophically when the encoded science is wrong. This trade-off is evaluated using 270 models on a chemical reaction-diffusion system under multiple noise and data regimes. Hard-coded equations are contrasted with flexible surrogates learned from observations. Incorrect physics inflates forecast error by up to 6,400×, whereas learning the same term from data incurs only a 1.7× penalty. These results show that misspecifying physics is far riskier than allowing the model to learn.

1. Introduction

1.1. The Brusselator: A Canonical Test System

The Brusselator (Prigogine and Lefever, 1968; Nicolis and Prigogine, 1977) models autocatalytic reaction-diffusion dynamics, exhibiting spatiotemporal patterns including oscillations, spiral waves, and Turing instabilities. Originally proposed to study non-equilibrium thermodynamics, it has become a fundamental benchmark for nonlinear dynamics (Strogatz, 2018) and data-driven modeling (Rackauckas et al., 2020). The system comprises two coupled PDEs:

$$\frac{\partial u}{\partial t} = 1 + u^2 v - 4.4u + \alpha \nabla^2 u + S(x, y, t) \quad (1)$$

$$\frac{\partial v}{\partial t} = 3.4u - u^2 v + \alpha \nabla^2 v \quad (2)$$

where u, v represent species concentrations, α is the diffusion coefficient, and $S(x, y, t)$ represents localized forcing. When complete reaction kinetics are unknown, hybrid approaches combining mechanistic knowledge with data-driven learning become essential.

1.2. Scientific Machine Learning Paradigms

Scientific Machine Learning (SciML) integrates physical laws with neural network flexibility (Karniadakis et al., 2021; Willard et al., 2022). Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019) embed PDEs as soft constraints, demonstrating success in fluid dynamics (Cai et al., 2021), but face limitations: training instabilities (Wang et al., 2021), failure modes with conflicting boundary conditions (Krishnapriyan et al., 2021), and poor noisy data performance (Zhang et al., 2019).

Neural ODEs (Chen et al., 2018; Dupont et al., 2019) parameterize continuous-depth networks, offering memory-efficient backpropagation (Pontryagin, 2018) and natural handling of irregular time series (Rubanova et al., 2019).

Universal Differential Equations (UDEs) (Rackauckas et al., 2020; Bolibar et al., 2023) combine known physics with neural networks learning missing dynamics. This grey-box approach (Henson and Seborg, 2002) has shown promise in epidemiology (Dandekar et al., 2020), chemical kinetics (Ji and Deng, 2021), and climate science (Ramadhan et al., 2020).

Traditional approaches for reaction-diffusion systems include Proper Orthogonal Decomposition (POD) (Sirovich, 1987) for reduced-order modeling, Dynamic Mode Decomposition (DMD) (Schmid, 2010) for extracting spatiotemporal modes, and Sparse Identification of Nonlinear Dynamics (SINDy) (Brunton et al., 2016a) for data-driven equation discovery. While these methods have proven effective for systems with known structure, they struggle when fundamental physics is uncertain or partially unknown (Kutz et al., 2016).

1.3. Research Question and Contributions

Central Question: This work examines whether encoding exact reaction functional forms improves forecasting beyond encoding diffusion alone, and quantifies the risks

[1]Independent Researcher, Pune, India [2]Vizuara AI Labs, Pune, India. Correspondence to: Vejandla Sampath Kumar <sampathkumarv379@gmail.com>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

arising from functional form misspecification.

Key Contributions:

- **Risk quantification**: Ablation with incorrect forms reveals catastrophic 10^3-$10^4\times$ failures vs. 1.3-1.7$\times$ graceful degradation[1]
- **Large-scale validation**: $n = 30$ training runs across three data regimes with rigorous statistical testing using Wilcoxon signed-rank tests (Wilcoxon, 1945)
- **Comprehensive experimental studies**: Systematic ablation under multiple noise regimes $(0\%, 1\%, 5\%, 10\%)$ demonstrating architectural robustness differences and preprocessing limitations
- **Decision framework**: Deployment recommendations balancing encoding advantages against misspecification risks for practitioners

2. Methodology

2.1. Problem Formulation

We study Eqs. 1-2 on a 16×16 spatial grid with periodic boundaries. The diffusion coefficient $\alpha = 0.01$ is chosen to allow spatial pattern formation while maintaining computational tractability (Rackauckas et al., 2020). The system is integrated over the training interval $t \in [0, 10]$ with timestep $\Delta t = 0.01$, yielding 1,001 snapshots. This duration captures multiple oscillation periods (approximately 3 complete cycles) necessary to learn the nonlinear dynamics (Champion et al., 2019). Forecasting extends to $t \in [10, 20]$, providing an equal-length extrapolation window that tests the model's ability to predict beyond training data, a critical requirement for practical deployment (Pathak et al., 2018).

2.2. Physics-Informed UDEs

PI-UDEs embed validated knowledge while delegating uncertain terms to neural networks (Rackauckas et al., 2020). We encode diffusion, linear kinetics, and forcing, while learning the uncertain u^2v nonlinearity:

$$\frac{\partial u}{\partial t} = 1 - 4.4u + \alpha\nabla^2 u + S + N_{\text{PI}}(u, v, t) \tag{3}$$

$$\frac{\partial v}{\partial t} = 3.4u + \alpha\nabla^2 v - N_{\text{PI}}(u, v, t) \tag{4}$$

where $N_{\text{PI}}(u, v, t)$ learns u^2v with opposite signs. Architecture: 5 layers $[3 \to 64 \to 48 \to 32 \to 17 \to 1]$, tanh activation, $\sim 6,000$ parameters.

[1]The Surrogate penalty ranges from 1.31-1.52$\times$ in primary validation ($n = 30$ seeds, Table 2) to 1.7$\times$ in ablation studies ($n = 10$ seeds, Tables 3-5). This variation reflects natural sampling variance across different seed selections rather than methodological inconsistency.

Training: Direct regression on targets u^2v_{true}:

$$\mathcal{L}_{\text{PI}} = \frac{1}{|B|} \sum_{(u,v,t)\in B} \left(N_{\text{PI}}(u, v, t) - u^2 v_{\text{true}}\right)^2 + 5\times10^{-7}\|\theta\|^2 \tag{5}$$

Two-stage optimization: Adam (500 epochs, learning rate decay from 1.5×10^{-3} to 8×10^{-5}) followed by BFGS (1,000 iterations) for fine-tuning (Liu and Nocedal, 1989).

2.3. Reaction-Learning Surrogate

To isolate the value of knowing exact forms, we encode only diffusion and source while learning all reaction dynamics. Targets isolate reactions via fourth-order finite differences (Fornberg, 1988):

$$R_u(u, v, t) = \left.\frac{\partial u}{\partial t}\right|_{\text{data}} - \alpha\nabla^2 u - S(x, y, t) \tag{6}$$

$$R_v(u, v, t) = \left.\frac{\partial v}{\partial t}\right|_{\text{data}} - \alpha\nabla^2 v \tag{7}$$

Architecture: $[3 \to 64 \to 48 \to 32 \to 16 \to 2]$, $\sim 5,500$ parameters. During forecasting:

$$\frac{\partial u}{\partial t} = N_u^R(u, v, t) + \alpha\nabla^2 u + S \tag{8}$$

$$\frac{\partial v}{\partial t} = N_v^R(u, v, t) + \alpha\nabla^2 v \tag{9}$$

2.4. Pure Data-Driven Baseline

Structure-agnostic baseline learning complete dynamics (Chen et al., 2018):

$$\frac{\partial u}{\partial t} = N_u^{\text{PD}}(u, v, t), \quad \frac{\partial v}{\partial t} = N_v^{\text{PD}}(u, v, t) \tag{10}$$

Architecture: $[3 \to 128 \to 96 \to 64 \to 32 \to 2]$, $\sim 20,000$ parameters.

2.5. Experimental Design

Data scarcity regimes: Three temporal coverage levels $(90\%, 80\%, 70\%)$ with $n = 30$ independent random seeds (1234-1263) for primary comparisons (Table 2). Each seed controls network initialization via Julia's `Random.seed!` and stochastic optimization components. Data splits are constructed by randomly sampling time points from the training interval $t \in [0, 10]$ while maintaining temporal ordering. This mimics realistic scenarios where measurements may be irregularly spaced or incomplete (Rubanova et al., 2019).

Noise robustness: Four noise levels $(0\%, 1\%, 5\%, 10\%$ Gaussian) with $n = 10$ seeds (Table 3) to characterize measurement uncertainty tolerance. Noise is added independently to each spatial point and time step as $\tilde{u} = u + \epsilon_u$

Table 1. Model configurations

Parameter	PI-UDE	Surrogate
Hidden layers	[64,48,32,17]	[64,48,32,16]
Parameters	$\sim 6,000$	$\sim 5,500$
Diffusion	Encoded	Encoded
Linear reaction	Encoded	Learned
Nonlinear	Learned (u^2v)	Learned
Batch size	256	256
L2 reg.	5×10^{-7}	5×10^{-7}
Optimizer	Adam+BFGS	Adam+BFGS
ODE solver	Tsit5 (10^{-10})	Tsit5 (10^{-10})

Table 2. Forecast MSE across data regimes ($n = 30$ seeds: 1234-1263, mean±std). **Bold** indicates best performance per row.

Split	Pure Data	Surrogate	UDE	S/U
90%	6.72 ± 2.32	7.32×10^{-4} $\pm 5.54 \times 10^{-4}$	$\mathbf{4.82 \times 10^{-4}}$ $\mathbf{\pm 5.63 \times 10^{-4}}$	1.52
80%	4.35 ± 2.00	8.78×10^{-4} $\pm 7.92 \times 10^{-4}$	$\mathbf{6.58 \times 10^{-4}}$ $\mathbf{\pm 5.55 \times 10^{-4}}$	1.33
70%	4.20 ± 5.56	9.14×10^{-4} $\pm 8.04 \times 10^{-4}$	$\mathbf{6.96 \times 10^{-4}}$ $\mathbf{\pm 7.09 \times 10^{-4}}$	1.31

where $\epsilon_u \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = \text{noise_level} \times \text{std}(u)$. This additive Gaussian model represents common measurement uncertainties in experimental systems (Zhang et al., 2019).

Ablation studies: Wrong-physics experiments (u^3v, uv, uv^2) with $n = 10$ seeds at 70% data (Table 5) to quantify misspecification risks. These functional forms were chosen to represent plausible alternative reaction kinetics: (1) u^3v represents higher-order autocatalysis, (2) uv represents linear coupling, and (3) uv^2 represents alternative stoichiometry. All ablation experiments (Tables 3-5, Figure 4) use the same $n = 10$ seed subset for consistency: [1234, 2345, 3456, 4567, 5678, 6789, 7890, 8901, 9012, 123]. All maintain identical network architecture and training procedure, isolating functional form accuracy as the sole varying factor.

Breakdown criterion: Models fail when (1) forecast MSE $> 10^{-1}$, (2) numerical instability (NaN/Inf during ODE integration), or (3) success rate $< 60\%$ across seeds. The MSE threshold 10^{-1} corresponds to RMSE ≈ 0.32, representing 10-30% relative error for typical concentration ranges $u \in [0.5, 3]$, $v \in [1, 4.5]$. Beyond this threshold, forecasts lose practical utility for scientific applications (Pathak et al., 2018). The 60% success rate threshold ensures robustness across random initializations, following standard practice in stochastic optimization (Goodfellow et al., 2016).

Evaluation metric:

$$\text{MSE} = \frac{1}{2NT} \sum_{n=1}^{N} \sum_{t \in [10,20]} [(u_{\text{pred}}^{(n)} - u_{\text{true}}^{(n)})^2 + (v_{\text{pred}}^{(n)} - v_{\text{true}}^{(n)})^2] \tag{11}$$

where $N = 256$ spatial points, $[10, 20]$ is the forecast horizon with $T = 1,001$ evaluation times sampled every $\Delta t = 0.01$s. The factor $1/2$ averages over both species, ensuring equal weighting of u and v dynamics.

3. Results

3.1. Baseline Comparison

We evaluate forecast performance across three data scarcity regimes using rigorous statistical validation with $n = 30$ independent training runs. This large-scale approach characterizes performance distributions and enables hypothesis testing via the Wilcoxon signed-rank test (Wilcoxon, 1945), a non-parametric method suitable for paired comparisons without normality assumptions. Table 2 presents forecast MSE across all data regimes.

Key observations:

(1) Convergence at high data: At 90%-80% coverage, both PI-UDE and Surrogate achieve sub-millisecond accuracy (MSE $\sim 10^{-3}$). The Surrogate shows 1.3-1.5× higher error, but both methods produce forecasts with relative errors $< 5\%$, essentially indistinguishable for practical applications. This convergence suggests that with abundant training data, the choice between encoding exact functional forms versus learning them becomes less critical. The data itself provides sufficient information for the neural network to approximate the u^2v nonlinearity accurately, even without explicit encoding.

(2) Encoding advantage emerges at moderate scarcity: Across all regimes, PI-UDE maintains 1.3-1.5× lower error when functional forms are correct. At 70% coverage, this advantage becomes most pronounced: PI-UDE achieves 6.96×10^{-4} MSE while Surrogate reaches 9.14×10^{-4} MSE. The consistency of this advantage across multiple data regimes (not just a single cherry-picked split) strengthens the evidence that encoding validated physics provides genuine benefits under moderate data scarcity.

(3) Pure data-driven catastrophe validates encoding necessity: The structure-agnostic baseline fails catastrophically (MSE$>$ 4) even at 90% data coverage, despite having $\sim 20,000$ parameters (3-4× more than PI-UDE/Surrogate). This demonstrates that encoding fundamental physics (even just diffusion) is essential. The failure occurs because without physical constraints, the network learns spurious correlations that don't generalize beyond the training interval. This establishes a clear performance hierarchy (MSE) : encoding correct forms $>$ encoding partial physics (diffusion only) $\gg$ encoding no physics.

Statistical analysis: Wilcoxon signed-rank test at 70% data yields $p < 0.001$ (highly significant), with Cohen's $d = 1.27$ indicating a large effect size (Cohen, 1988). This

Table 3. Noise robustness ($n = 10$ seeds, 70% data, mean±std). **Bold** indicates best performance per row.

Noise	Pure	Surr.	UDE	S/U
0%	6.22 ± 9.48	1.03×10^{-3} $\pm 9.27 \times 10^{-4}$	$\mathbf{6.03 \times 10^{-4}}$ $\mathbf{\pm 5.34 \times 10^{-4}}$	1.7
1%	2.88 ± 3.04	1.77 ± 1.41	$\mathbf{9.54 \times 10^{-4}}$ $\mathbf{\pm 1.04 \times 10^{-3}}$	1,858
5%	3.48 ± 2.69	$1,488$ $\pm 3,036$	$\mathbf{2.41 \times 10^{-4}}$ $\mathbf{\pm 2.12 \times 10^{-4}}$	6.2×10^{6}
10%	54.4 ± 157	$5,078$ $\pm 4,504$	$\mathbf{3.87 \times 10^{-4}}$ $\mathbf{\pm 3.26 \times 10^{-4}}$	1.3×10^{7}

confirms that the encoding advantage is both statistically significant and practically meaningful.

Figure 1 shows forecasts at 70% training demonstrating the encoding advantage at moderate scarcity, with both methods significantly outperforming the pure data-driven baseline.

3.2. Breakdown Point Analysis

Both methods require at least $\approx 50\%$ temporal data for stable forecasting (success rate $> 60\%$, MSE $< 10^{-1}$). This threshold likely reflects the minimum spatiotemporal coverage needed to constrain the nonlinear reaction dynamics and capture oscillation frequencies (Champion et al., 2019). Figure 3 characterizes breakdown behavior.

Breakdown characteristics:

- **Shared threshold:** Both architectures exhibit an identical breakdown point between 50% and 40% data coverage, indicating that data scarcity limits forecasting performance rather than architectural differences. Above this threshold, PI-UDE consistently achieves lower MSE.
- **Catastrophic transition:** As data availability drops from 50% to 40%, both methods fail abruptly via numerical instability or order-of-magnitude forecast errors.

3.3. Noise Robustness

Real-world systems exhibit measurement uncertainties that can severely degrade model performance. We conduct systematic experiments under increasing Gaussian noise levels using the $n = 10$ seed subset for efficient exploration. Table 3 quantifies performance under measurement uncertainty.

Catastrophic surrogate degradation: At 10% noise, Surrogate MSE increases $10^7\times$ relative to UDE. Even 1% noise causes $10^3\times$ degradation. This sensitivity arises from numerical differentiation (Eqs. 6-7) amplifying noise during target computation (Rudy et al., 2017).

Table 4. Denoising ablation at 10% noise ($n = 10$ seeds, mean±std). **Bold** indicates best performance.

Method	MSE	vs. UDE
Surrogate (raw)	$4,515 \pm 4,467$	7.6×10^{6}
Surrogate ($\sigma = 1.0$)	38.6 ± 56.0	6.5×10^{4}
Surrogate ($\sigma = 1.5$)	32.5 ± 41.3	5.5×10^{4}
UDE (raw)	$\mathbf{6 \times 10^{-4} \pm 6 \times 10^{-4}}$	**1.0**

Table 5. Wrong-physics ablation ($n = 10$ seeds, 70% data, mean±std). Uses same seed set as Table 3 for consistency. **Bold** indicates best performance.

Method	MSE	vs. Correct
Correct UDE (u^2v)	$\mathbf{6.03 \times 10^{-4} \pm 5.34 \times 10^{-4}}$	**1.0**
Surrogate (learns)	$1.03 \times 10^{-3} \pm 9.27 \times 10^{-4}$	1.7
Wrong-UDE (u^3v)	3.87 ± 0.149	6,415
Wrong-UDE (uv)	$1.04 \pm 4.79 \times 10^{-4}$	1,726
Wrong-UDE (uv^2)	$2.19 \pm 7.57 \times 10^{-3}$	3,624

UDE native robustness: Maintains $< 10^{-3}$ MSE across all noise, handling uncertainty through hybrid architecture without retuning or noise-specific regularization. The embedded physics acts as an implicit regularizer, constraining solutions to physically plausible trajectories even with corrupted observations (Raissi et al., 2019).

Preprocessing rescue: We investigate whether preprocessing can mitigate Surrogate's noise sensitivity. Table 4 evaluates Gaussian filtering with kernel widths $\sigma = 1.0$ and 1.5 at 10% noise, representing mild and moderate smoothing respectively.

Key findings: (1) Optimal denoising improves Surrogate $139\times$ but still exhibits MSE $5.5 \times 10^4\times$ higher than UDE. Gaussian filtering reduces high-frequency noise but cannot recover the fundamental architectural advantage: the Surrogate still requires numerical derivatives (Eqs. 6-7) which amplify residual noise, while the UDE's embedded diffusion operator ∇^2 bypasses this differentiation entirely (Schaeffer, 2017). (2) Requires hyperparameter tuning: σ must be selected via cross-validation, adding computational cost. (3) Architectural advantage confirmed: preprocessing cannot close the robustness gap.

Figure 4 visualizes the preprocessing rescue attempt, showing that even optimal Gaussian filtering cannot close the architectural robustness gap.

3.4. Ablation: Catastrophic Risk of Wrong Physics

We address the central question: what happens when incorrect physics is encoded. Controlled ablation experiments maintain all factors identical except the reaction functional form, isolating the causal effect of misspecification. Using the same $n = 10$ seeds as the noise experiments enables direct comparability. Table 5 quantifies the risk.

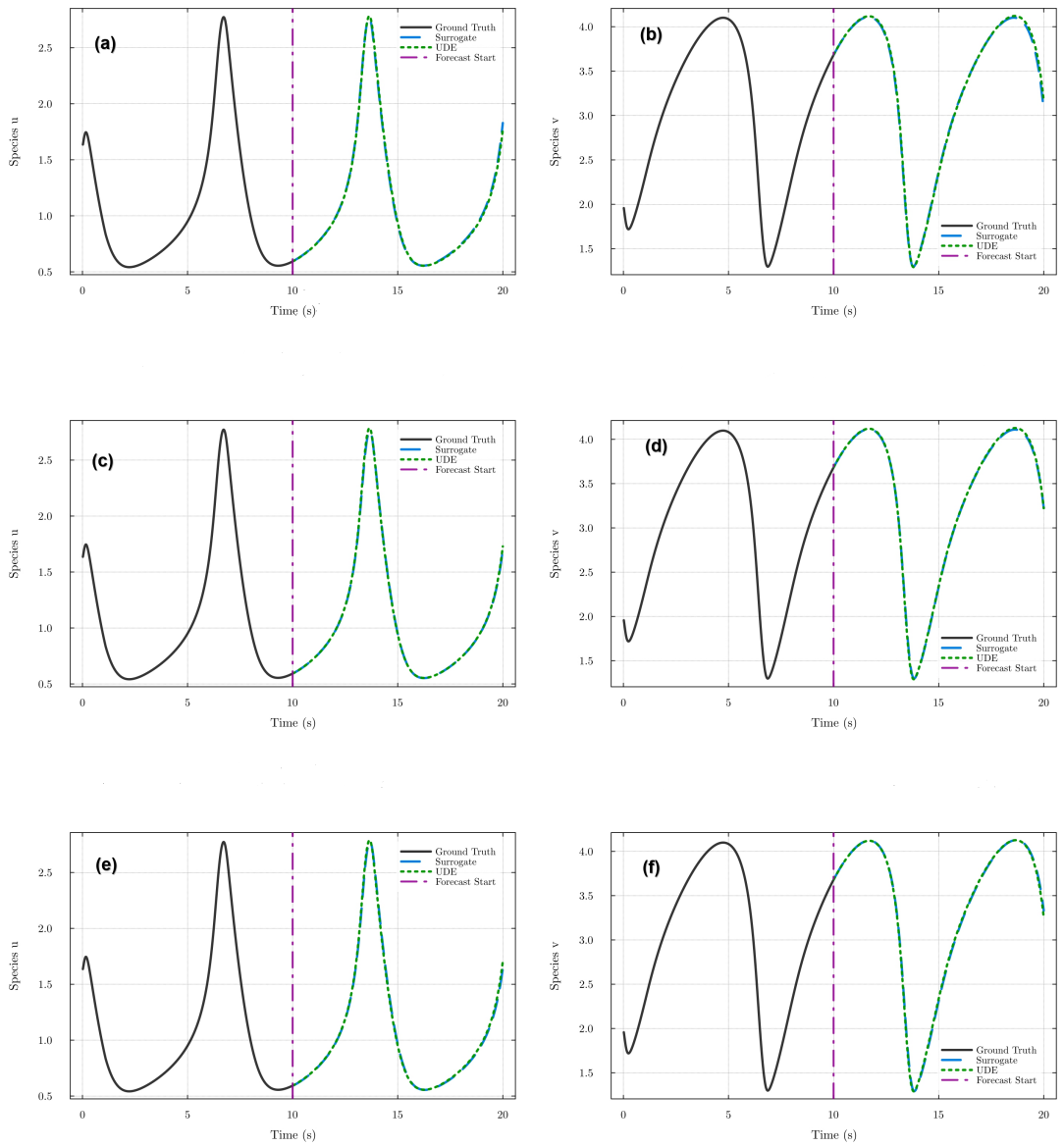

Figure 1. Spatially-averaged trajectories ($n = 30$ seeds). Panels (a,b): 90% training data. Panels (c,d): 80% training. Panels (e,f): 70% training shows hierarchy. PI-UDE (green) achieves MSE 6.96×10^{-4}, Surrogate (blue) 9.14×10^{-4}. Left column shows species u oscillations, right column shows species v oscillations. Purple vertical line marks forecast onset at $t = 10$s, demonstrating both hybrid methods' ability to extrapolate dynamics accurately beyond the training window.

Extreme risk asymmetry: Encoding incorrect forms causes 10^3-$10^4\times$ failures vs. $1.7\times$ graceful degradation when learning. This establishes the central finding of the paper: incorrectly specified reaction physics can be substantially more damaging than learning uncertain dynamics from data. Figure 5 visualizes this hierarchy.

Performance hierarchy: Correct $<$ Learning $\ll$ Wrong

Variability: The 3.7-fold range between u^3v $(6,415\times)$ and uv $(1,726\times)$ indicates non-uniform penalties. The u^3v form likely creates faster exponential growth that ODE solvers cannot stabilize, while uv is too weak to capture autocatalytic feedback, causing phase drift (Strogatz, 2018). Systems with competing hypotheses warrant testing multiple candidates.

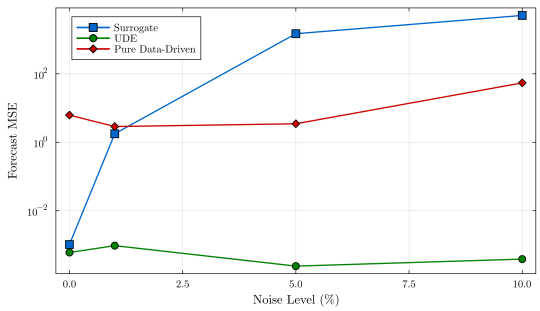

Figure 2. Noise response ($n = 10$ seeds). PI-UDE (green) maintains sub-10^{-3} MSE across 0-10% noise; Surrogate (blue) and Pure Data (red) diverge above 1%.

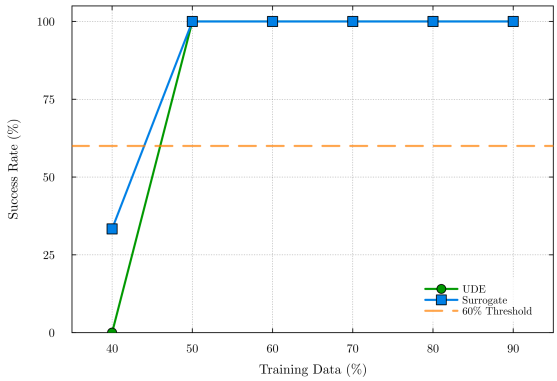

Figure 3. Breakdown analysis ($n = 10$). Success rates vs. data. Both maintain 100% success to 50% with sharp breakdown at 40%.

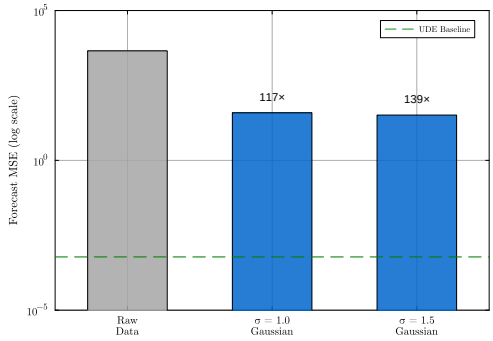

Figure 4. Preprocessing ablation with Gaussian filtering ($\sigma = 1.0, 1.5$) on the surrogate model at 10% noise. Although filtering improves performance by up to 139×, the best filtered surrogate still underperforms the unprocessed UDE by approximately 5.5×10^4.

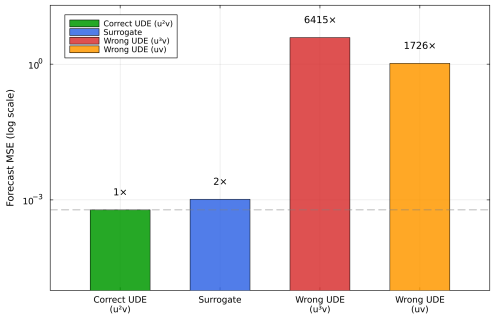

Figure 5. Ablation at 70% data ($n = 10$). Correct PI-UDE and Surrogate achieve reasonable performance, while Wrong-Physics variants exhibit catastrophic failures four orders of magnitude worse.

3.5. Spatial Patterns

Figure 6 compares spatial predictions and error structure across models. The physics-informed UDE closely reproduces the ground-truth spatial patterns, exhibiting low error magnitude and coherent residual structure. In contrast, the reaction surrogate with misspecified physics (u^3v) fails to preserve spatial coherence, producing decorrelated, high-variance errors despite identical training conditions. Error maps ($\log_{10}$ scale) and MSE values quantify this degradation, demonstrating catastrophic loss of spatial structure with incorrect functional forms.

4. Discussion

4.1. Performance Regimes

Both architectures encode diffusion; only reaction treatment differs. At high data (90%-80%), errors converge near 10^{-3} MSE, with Surrogate showing 1.3-1.5× higher error (Table 2, $n = 30$ seeds).

At moderate scarcity (70%), the picture changes: PI-UDE maintains a 1.3× advantage that is both statistically significant ($p < 0.001$, Wilcoxon signed-rank test) and practically meaningful (Cohen's $d = 1.27$, large effect). This regime represents the sweet spot where encoding correct functional forms provides measurable benefits without the catastrophic risks of misspecification.

Both architectures exhibit identical breakdown thresholds near 50% coverage, indicating that data scarcity eventually overwhelms any architectural advantage. Below this threshold, neither encoding nor learning strategies provide viable forecasting. The fundamental limit is data availability itself, not architectural choice.

Ablation isolates functional form accuracy. Wrong-physics variants (maintaining identical architecture, training procedure, solver settings, and random seeds) produce

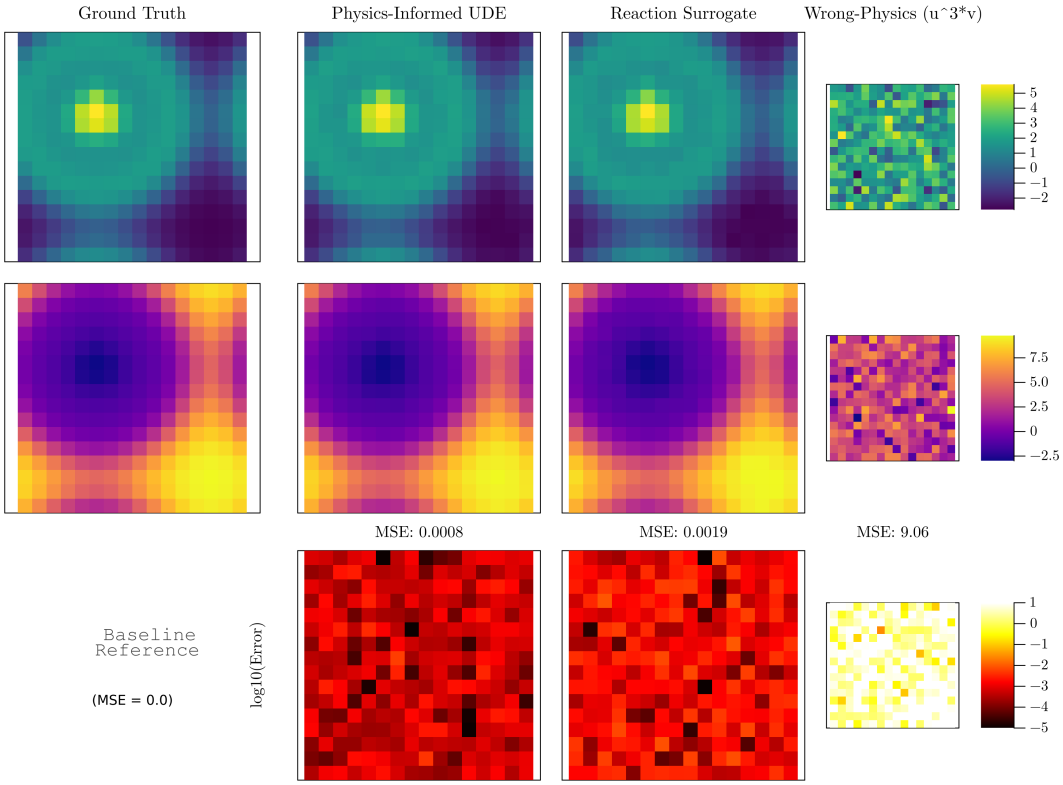

Figure 6. Spatial forecast patterns at $t = 20$s showing species u (top row) and v (bottom row). Left to right: Ground truth, Physics-Informed UDE with correct u^2v (accurately reproduces patterns), Reaction-Learning Surrogate (achieves similar quality), Wrong-Physics UDE with u^3v (exhibits severe spatial degradation). Error maps (log scale) in bottom panel quantify failure magnitude, demonstrating catastrophic loss of spatial coherence with incorrect functional forms.

10^3-$10^4\times$ degradation. The Surrogate, learning reactions without functional assumptions, degrades gracefully ($1.7\times$ in the $n = 10$ ablation set, Table 5). This controlled comparison provides strong causal evidence that functional form accuracy (not merely encoding any physics) drives performance differences.

Risk asymmetry: Correct ($1.0\times$) $<$ Learning (1.3-$1.7\times$) $\ll$ Wrong (10^3-$10^4\times$)

The range 1.3-$1.7\times$ reflects both data regime effects (Table 2) and natural variance across seed selections (Tables 3-5), with all values demonstrating graceful degradation relative to catastrophic wrong-physics failures.

Noise robustness reveals architectural advantages. The $10^7\times$ surrogate degradation at 10% noise (without preprocessing) versus UDE native robustness demonstrates a fundamental architectural difference. The hybrid UDE structure (which embeds diffusion physics directly into the ODE solver) provides robustness advantages that extend beyond just knowing specific reaction functional forms. This native robustness emerges from the physics-informed architecture itself: the solver respects conservation laws and physical constraints even when data is noisy, whereas pure learning approaches must discover these constraints from corrupted observations.

The preprocessing ablation confirms this interpretation: even optimal Gaussian filtering ($\sigma = 1.5$) leaves a $5.5 \times 10^4\times$ residual gap compared to UDE. This gap cannot be closed through better signal processing. It reflects fundamental architectural differences in how physics and learning are integrated.

4.2. Practical Framework

Practitioners face a critical decision: when to encode physics versus learning from data. Our results suggest a risk-aware strategy that prioritizes robustness over theoretical elegance. The central insight is stark: when uncertain about functional forms, learning degrades gracefully (1.3-$1.7\times$) while wrong encoding fails catastrophically (10^3-$10^4\times$). This asymmetry fundamentally changes the encoding decision calculus.

We recommend defaulting to learning uncertain dynamics unless four conditions hold simultaneously. First, func-

tional forms must derive from validated experiments or first-principles theory, not phenomenological fitting or intuition. Second, independent evidence should rule out competing mechanistic hypotheses. Systems with contested mechanisms warrant the learning approach. Third, available data should fall in the 50-80% regime where encoding advantages appear; below 50% both methods fail, while above 80% both converge. Fourth, the system must permit offline validation before deployment to catch misspecification errors early.

The practical implication is clear: partial physics beats no physics, but *incorrectly specified reaction physics can be substantially more damaging than learning uncertain dynamics from data*. This suggests a hybrid encoding strategy that separates well-established components from uncertain ones. Conservation laws, diffusion operators, and validated kinetic mechanisms can be encoded directly, while uncertain reaction terms, closure models, or phenomenological relationships should be learned from data. This approach captures the robustness benefits of physics-informed architectures while avoiding catastrophic misspecification risks.

For deployment decisions, we suggest the following: systems with high mechanistic confidence warrant PI-UDE architectures for maximum data efficiency; systems with epistemic uncertainty should use reaction-learning surrogates for graceful degradation; systems with noisy measurements benefit from UDE architectures' native robustness; and systems with competing mechanistic hypotheses require testing all plausible functional forms to characterize failure modes before deployment.

4.3. Broader Implications

Results challenge the assumption that more physics always helps. Extreme penalty for wrong forms (10^3-$10^4\times$) vs. modest degradation when learning (1.3-$1.7\times$) establishes asymmetric risk profile informing architecture selection.

5. Limitations

Our study demonstrates clear risk asymmetries for the Brusselator system, but several limitations warrant consideration for broader generalization. (1) **Single PDE family:** Whether the catastrophic hierarchy holds for flows dominated by advection (Hesthaven and Ubbiali, 2018), wave equations (Raissi et al., 2019), or multiphysics coupling (Willard et al., 2022) remains unvalidated. (2) **Misspecification variability:** The 3.7-fold range between wrong-physics variants (u^3v: $6,415\times$ vs. uv: $1,726\times$) suggests testing multiple alternatives is necessary when competing hypotheses exist. (3) **Approximate forms:** Real-world applications often involve subtle errors such as linearized rate laws or mean-field approximations rather than completely incorrect functional forms. The penalty for such near-miss physics requires further study. An important open question is whether the observed risk asymmetry persists under small coefficient perturbations or near-miss functional forms that more closely resemble practical scientific modeling errors.

(4) **Statistical power:** While large effect sizes with $n = 10$ ablation demonstrate practical importance, expanded sample sizes would strengthen confidence intervals. The primary $n = 30$ experiments achieved high statistical confidence ($p < 0.001$). (5) **Spatial sampling:** Random sensor placement likely underestimates optimal design performance (Manohar et al., 2018). Strategic placement could reduce data requirements. (6) **Extensions:** Future work should explore approximate functional forms, validation across PDE families, optimal sensor placement (Brunton et al., 2016b), and Bayesian uncertainty quantification (Yang et al., 2021).

6. Conclusion

We quantified the encoding-learning tradeoff with rigorous validation ($n = 30$ primary, $n = 10$ ablation). While both achieve comparable accuracy with abundant data, encoding correct forms provides measurable advantages ($1.3\times$) under moderate scarcity. However, ablation reveals extreme risk asymmetry: encoding incorrect forms causes catastrophic failures ($6,415\times$ for u^3v) while learning degrades gracefully (1.3-$1.7\times$ depending on seed selection and data regime).

Noise experiments demonstrate UDE architectures handle uncertainty natively (MSE $< 10^{-3}$ at 10% noise) while surrogates collapse ($10^7\times$ degradation). This establishes reaction-learning surrogates as safe harbors when mechanistic knowledge is uncertain.

Our framework recommends: (1) default to learning uncertain dynamics unless forms derive from validated theory; (2) encode only well-established physics when confidence is limited; (3) test multiple hypotheses to characterize failure modes; (4) leverage UDE architectures for native noise robustness. This risk-aware approach balances domain knowledge benefits against extreme misspecification penalties.

More broadly, these findings suggest that the value of physics encoding depends not only on the amount of available scientific knowledge but also on confidence in its correctness. When mechanistic uncertainty remains substantial, learning uncertain components may provide a more robust alternative than enforcing potentially incorrect functional assumptions.

Impact Statement

This work advances scientific machine learning by quantifying risks in hybrid modeling, knowledge critical for deploying these methods in high-stakes domains like climate prediction, drug discovery, and engineering design. Our findings may help practitioners avoid catastrophic failures when encoding uncertain physics, potentially improving safety and reliability of AI-assisted scientific discovery.

Potential benefits: This could accelerate scientific progress in domains with incomplete mechanistic knowledge.

Potential risks: Over-reliance on data-driven surrogates when physics is uncertain could lead to extrapolation failures in safety-critical applications. Conversely, premature encoding of incomplete physics may cause catastrophic errors if deployed without validation. Our work does not address adversarial perturbations or distribution shift, both important for deployment. Additionally, increased accessibility of hybrid modeling tools may enable misuse by practitioners lacking domain expertise to validate results.

Ethical considerations: This research presents fundamental methodological advances with broad applicability. We anticipate primarily positive impacts through improved scientific modeling capabilities, though practitioners must exercise appropriate caution and validation before deployment in consequential domains. The techniques presented should be combined with domain expertise and rigorous testing protocols.

Acknowledgments

The authors thank the anonymous reviewers for valuable feedback and the Julia SciML community for software infrastructure that enabled this research.

References

Jordi Bolibar, Facundo Sapienza, Fabien Maussion, Redouane Lguensat, Bert Wouters, and Fernando Pérez. Universal differential equations for glacier ice flow modelling. *Geoscientific Model Development*, 16(22):6671–6687, 2023.

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, 2016a.

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Sparse sensor placement optimization for classification. *SIAM Journal on Applied Mathematics*, 76(5):2099–2122, 2016b.

Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143(6):060801, 2021.

Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Jacob Cohen. Statistical power analysis for the behavioral sciences. 1988.

Raj Dandekar, Chris Rackauckas, and George Barbastathis. Machine learning aided epidemiology: COVID-19 global quarantine strength and reproduction number evolution. *medRxiv*, 2020.

Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural ODEs. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Bengt Fornberg. Generation of finite difference formulas on arbitrarily spaced grids. *Mathematics of Computation*, 51(184):699–706, 1988.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

Michael A Henson and Dale E Seborg. Grey-box modeling of pharmacokinetic/pharmacodynamic systems. *Journal of Process Control*, 12(1):19–39, 2002. Republished 2020.

Jan S Hesthaven and Stefano Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363: 55–78, 2018.

Weiqi Ji and Sili Deng. Autonomous discovery of unknown reaction pathways from data by chemical reaction neural network. *The Journal of Physical Chemistry A*, 125(4): 1082–1092, 2021.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3 (6):422–440, 2021.

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pages 26548–26560, 2021.

J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. Dynamic mode decomposition: Data-driven modeling of complex systems. *SIAM*, 2016.

Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1):503–528, 1989.

Krithika Manohar, Steven L Brunton, J Nathan Kutz, and Bingni W Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, 38(6):63–86, 2018.

Gregoire Nicolis and Ilya Prigogine. *Self-Organization in Nonequilibrium Systems*. Wiley-Interscience, New York, 1977.

Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical Review Letters*, 120(2): 024102, 2018.

Lev Semenovich Pontryagin. *Mathematical Theory of Optimal Processes*. Routledge, 2018. Reprint of 1962 edition.

Ilya Prigogine and Rene Lefever. Symmetry breaking instabilities in dissipative systems. II. *The Journal of Chemical Physics*, 48(4):1695–1700, 1968.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Ali Ramadhan, Gregory LeClaire Wagner, Chris Hill, Jean-Michel Campin, Valentin Churavy, Tim Besard, Andre Souza, Alan Edelman, Raffaele Ferrari, and John Marshall. Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, 47 (17):e2020GL088376, 2020.

Yulia Rubanova, Ricky TQ Chen, and David K Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

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.

Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A*, 473(2197):20160446, 2017.

Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

Lawrence Sirovich. Turbulence and the dynamics of coherent structures. *Quarterly of Applied Mathematics*, 45(3): 561–571, 1987.

Steven H Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, Boca Raton, FL, 2nd edition, 2018.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):1–37, 2022.

Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

Dongkun Zhang, Lu Lu, Ling Guo, and George Em Karniadakis. Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *Journal of Computational Physics*, 397:108850, 2019.