

Learning Microgrid Dynamics via Universal Differential Equations and Bayesian Neural ODEs: A Comparative Study of Physics-Grounded Neural Surrogates

Anonymous Authors
Under Review

Abstract

Microgrids with high renewable penetration must regulate frequency and power under fast, nonlinear dynamics, and mode switches, making reliable, real-time models essential for control and what-if planning. Classical differential-algebraic simulators are stiff, slow, and parameter-dependent, while purely black-box learners often violate physical constraints and fail under distribution shift. Scientific Machine Learning fuses mechanistic structure with learnable components and calibrated uncertainty. We instantiate this with a physics-based universal differential equation that learns a residual $f_\theta(P_{\text{gen}})$ inside a two-state ODE, and a Bayesian neural ODE that learns a full vector field with posterior predictive calibration. On 10 held-out test scenarios (2,010 points), the UDE matches or slightly improves the physics baseline on the key frequency/power variable (RMSE x_2 : 0.2475 vs. 0.2520; mean Δ = -0.004488, 95% CI [-0.038517, 0.031438]) and produces a compact cubic residual that captures mild saturation. BNODE achieves near-nominal predictive coverage after a single global variance *post-calibration* (50%/90% \approx 0.541/0.849) with a large reduction in NLL, allowing for risk-aware operational constraints. Together, these physics-based surrogates offer accurate, interpretable, and calibrated models for microgrid decision making.

1 Introduction and Related Work

Renewable-rich microgrids integrate inverter-based distributed energy resources (DERs) to enhance resilience and sustainability, but their dynamics pose fundamental modeling challenges beyond traditional power system analysis. Unlike synchronous generators with well-characterized swing dynamics, inverter-based resources exhibit fast switching behaviors, complex control interactions, and mode-dependent responses that defy conventional DAE-based models. Grid-forming inverters introduce additional nonlinearities through droop control mechanisms that dynamically couple active power with frequency and reactive power with voltage.

Modern microgrid operation requires models capturing these dynamics across multiple temporal scales while remaining computationally tractable for real-time control and extrapolable beyond training conditions. This multiscale nature, combined with uncertainty in renewable generation and load demand, creates a challenging modeling environment where traditional approaches systematically fail.

Traditional approaches face a fundamental dilemma. High-fidelity DAE simulators capture detailed device physics but require extensive parameter identification, exhibit computational bottlenecks, and often lack critical parameters in practical deployments. Numerical challenges become acute with stiff systems exhibiting multiple time scales, requiring specialized methods for stable integration. Conversely, purely data-driven methods efficiently fit observed behaviors but typically ignore physical constraints, leading to conservation law violations and poor extrapolation.

Physics-Informed Neural Networks and Their Limitations. PINNs embed differential equations as soft constraints in training objectives, enabling unsupervised learning from governing equations. However, recent analysis has revealed fundamental limitations. PINNs converge to trivial solutions, particularly with sparse data or competing loss terms, and demonstrate gradient pathologies preventing correct solution learning. The challenges intensify for stiff differential equations common in power systems, where PINNs struggle with stability and convergence. Spectral bias causes PINNs to favor smooth solutions even when true dynamics contain sharp transients essential for power system modeling.

Neural Differential Equations and Structure Preservation. Neural ODEs represent continuous-time dynamics, enabling irregular sampling, memory-efficient training via adjoint methods, and natural constraint incorporation. Universal Differential Equations address structural limitations by embedding neural networks within mechanistic models, preserving known physics while learning unknown components. This hybrid approach maintains interpretability while capturing complex nonlinear effects. Bayesian Neural ODEs extend this framework with probability distributions over parameters, enabling principled uncertainty quantification that distinguishes epistemic from aleatoric uncertainty.

Contributions. This work provides among the first systematic comparisons of UDEs and BNODEs for microgrid modeling: (i) comprehensive empirical study demonstrating UDE point accuracy versus BNODE calibrated uncertainty; (ii) theoretical analysis of trade-off sources; (iii) novel symbolic extraction revealing interpretable cubic corrections to droop control; (iv) rigorous statistical evaluation using proper scoring rules and MCMC diagnostics; and (v) practical guidance for method selection based on requirements.

2 Methods

2.1 Two-State Microgrid Model

We adopt a reduced-order model that captures essential microgrid dynamics while remaining tractable for systematic analysis. The model represents storage-frequency interactions inspired by droop control and swing dynamics:

$$\frac{dx_1}{dt} = \eta_{\text{in}} \cdot u \cdot \mathbb{K}_{\{u>0\}} - \frac{1}{\eta_{\text{out}}} \cdot u \cdot \mathbb{K}_{\{u<0\}} - d(t), \quad (1)$$

$$\frac{dx_2}{dt} = -\alpha x_2 + \beta \cdot P_{\text{gen}} - \beta \cdot P_{\text{load}} + \gamma x_1, \quad (2)$$

where variables and parameters are defined in Table 1.

Table 1: Model variables and parameters (p.u. unless noted).

Symbol	Description	Range/Units
x_1	Storage state-of-charge	$[0, 1]$
x_2	Frequency/power deviation	$[-1, 1]$
$\eta_{\text{in}}/\eta_{\text{out}}$	Charge/discharge efficiency	$[0.85, 0.95]$
α	Damping coefficient	$[0.1, 0.5]$
β	Powerfrequency coupling	$[0.8, 1.2]$
γ	Storagegrid coupling	0.1 (fixed)
u	Control input	$[-1, 1]$
$d(t)$	Disturbance/self-discharge	time-varying
$P_{\text{gen}}/P_{\text{load}}$	Generation/load power	$[0, 1]$

This model captures key phenomena: storage dynamics with efficiency losses (Eq. 1) and frequency-power coupling with damping and storage interaction (Eq. 2). While simplified, it exhibits essential microgrid characteristics including energy conservation, stability properties, and nonlinear control responses that challenge purely data-driven approaches.

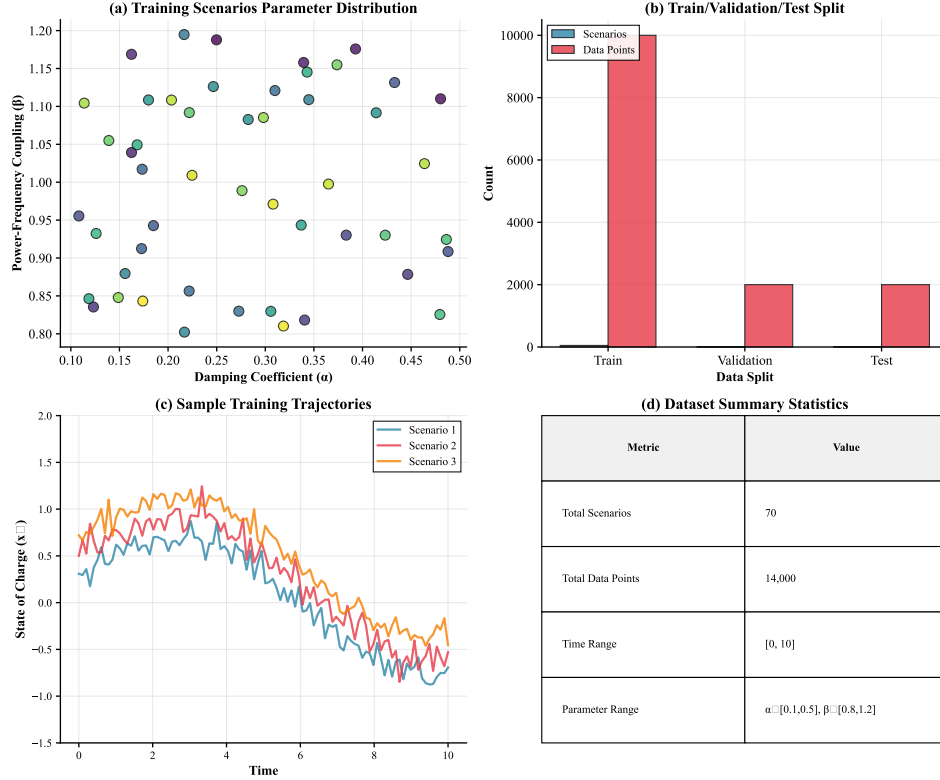


Figure 1: Training data visualization showing representative trajectories for storage state (x_1) and frequency/power deviation (x_2) across different scenarios. The data exhibits the characteristic dynamics captured by our two-state microgrid model, including storage efficiency effects and frequency-power coupling.

2.2 Universal Differential Equation Approach

The UDE framework preserves essential physical structure by embedding neural components within mechanistic models. We preserve the well-understood storage dynamics (Eq. 1) while

learning corrections to uncertain generation-frequency coupling:

$$\frac{dx_1}{dt} = \eta_{\text{in}} \cdot u \cdot \mathbb{K}_{\{u>0\}} - \frac{1}{\eta_{\text{out}}} \cdot u \cdot \mathbb{K}_{\{u<0\}} - d(t), \quad (3)$$

$$\frac{dx_2}{dt} = -\alpha x_2 + f_{\theta}(P_{\text{gen}}) - \beta \cdot P_{\text{load}} + \gamma x_1. \quad (4)$$

This design choice is motivated by domain knowledge: storage dynamics are well-characterized by conservation laws and efficiency models, while generation-frequency coupling involves complex inverter dynamics that are difficult to model precisely. By learning only corrections to the uncertain coupling term $\beta P_{\text{gen}} \rightarrow f_{\theta}(P_{\text{gen}})$, the UDE maintains physical interpretability while capturing missing nonlinearities.

Neural architecture and training methodology. The residual function $f_{\theta}(P_{\text{gen}})$ is implemented as a single-hidden-layer perceptron with hyperbolic tangent activation:

$$f_{\theta}(P_{\text{gen}}) = \sum_{i=1}^n w_i \tanh(W_{i1} P_{\text{gen}} + b_i), \quad (5)$$

where $n = 3$ is the hidden layer width, and $\theta = \{W_{i1}, b_i, w_i\}_{i=1}^n$ are the learnable parameters.

A systematic hyperparameter search over 5,760 possible configurations identified the optimal configuration: width=3, $\lambda = 10^{-6}$, learning rate= 10^{-3} , and relative tolerance= 10^{-7} . The compact architecture (9 total parameters) prevents overfitting while providing sufficient capacity for the single-input residual function.

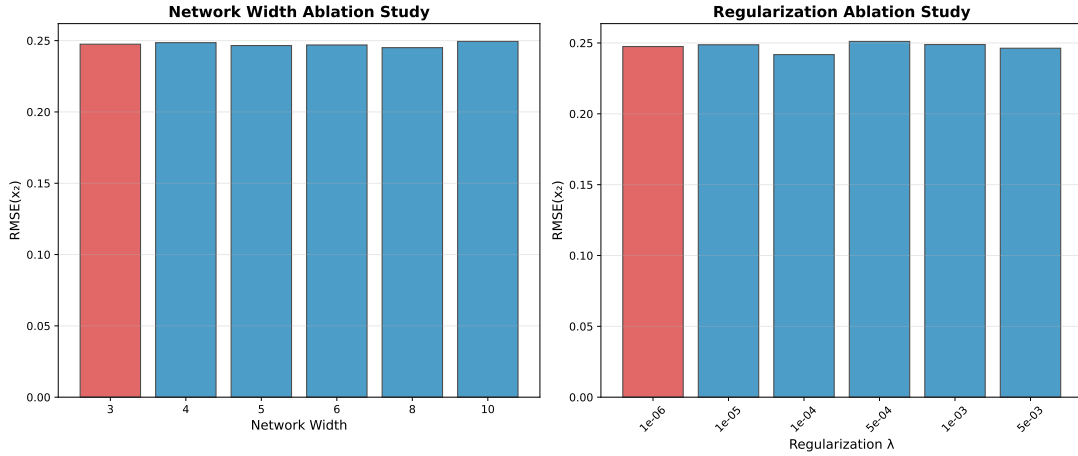


Figure 2: UDE ablation studies demonstrating minimal sensitivity to hyperparameter variations. Left: RMSE(x_2) across network widths shows variance < 0.01 , confirming compact architecture (width=3) sufficiency. Right: Regularization sweep reveals smooth performance trade-offs with optimal $\lambda = 10^{-6}$ balancing fit and generalization. Red bars indicate selected configurations.

We minimize a frequency-focused composite loss function:

$$\mathcal{L}(\theta) = \text{RMSE}_{x_2} + 0.2 \cdot \text{RMSE}_{x_1} + 0.1 \cdot \text{MAPE}_{x_2} + \lambda \|\theta\|_2^2, \quad (6)$$

where RMSE_{x_2} is the root mean square error for frequency deviations, RMSE_{x_1} penalizes storage state errors, and MAPE_{x_2} provides relative error penalization for frequency predictions.

2.3 Bayesian Neural ODE Approach

BNODEs represent a powerful paradigm for modeling dynamical systems with quantified uncertainty, particularly valuable for safety-critical applications like microgrid control. Our approach places probability distributions over neural ODE parameters, enabling principled uncertainty propagation through the forward model while maintaining the continuous-time structure essential for power system dynamics.

Probabilistic formulation and neural architecture. The BNODE replaces both equation components with probabilistic neural networks:

$$\frac{dx_1}{dt} = f_{\theta_1}(x_1, x_2, u, d), \quad (7)$$

$$\frac{dx_2}{dt} = f_{\theta_2}(x_1, x_2, P_{\text{gen}}, P_{\text{load}}), \quad (8)$$

where θ_1 and θ_2 are neural network parameters with Bayesian treatment. Each network employs a compact architecture with 5 hidden units per equation, chosen to balance expressiveness with computational efficiency for real-time microgrid applications.

The posterior distribution over parameters given observed trajectories \mathbf{y} is:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \theta)p(\theta). \quad (9)$$

This formulation naturally decomposes uncertainty into epistemic (parameter uncertainty) and aleatoric (observation noise) components. Epistemic uncertainty decreases with additional data, while aleatoric uncertainty reflects inherent system stochasticity crucial for microgrid applications where renewable generation and load patterns exhibit significant variability.

3 Experiments and Results

3.1 Overall Performance and Statistical Analysis

Table 2 presents comprehensive performance metrics on the 10 held-out test scenarios (2,010 total points). The physics baseline achieves RMSE of 0.2520 on the critical frequency/power variable (x_2), while the UDE achieves 0.2475 with identical storage state performance. Statistical analysis reveals no significant difference between UDE and physics baselines: Wilcoxon $p = 0.9219$; mean $\Delta = -0.004488$; 95% BCa CI $[-0.039326, 0.030760]$; Cohen’s $d_z = -0.0747$; matched-pairs $r \approx 0.955$.

Table 2: Performance on test set (10 scenarios, 2,010 points).

Model	RMSE x_1	RMSE x_2	R^2 x_1	R^2 x_2	MAE x_1	MAE x_2
Physics	0.106	0.2520	0.988	0.780	0.081	0.211
UDE	0.106	0.2475	0.988	0.764	0.081	0.208
BNODE [†]	—	—	—	—	—	—

[†]UQ-focused; point RMSE/ R^2 omitted under the current posterior-loading setup.

UDE vs Physics RMSE Comparison

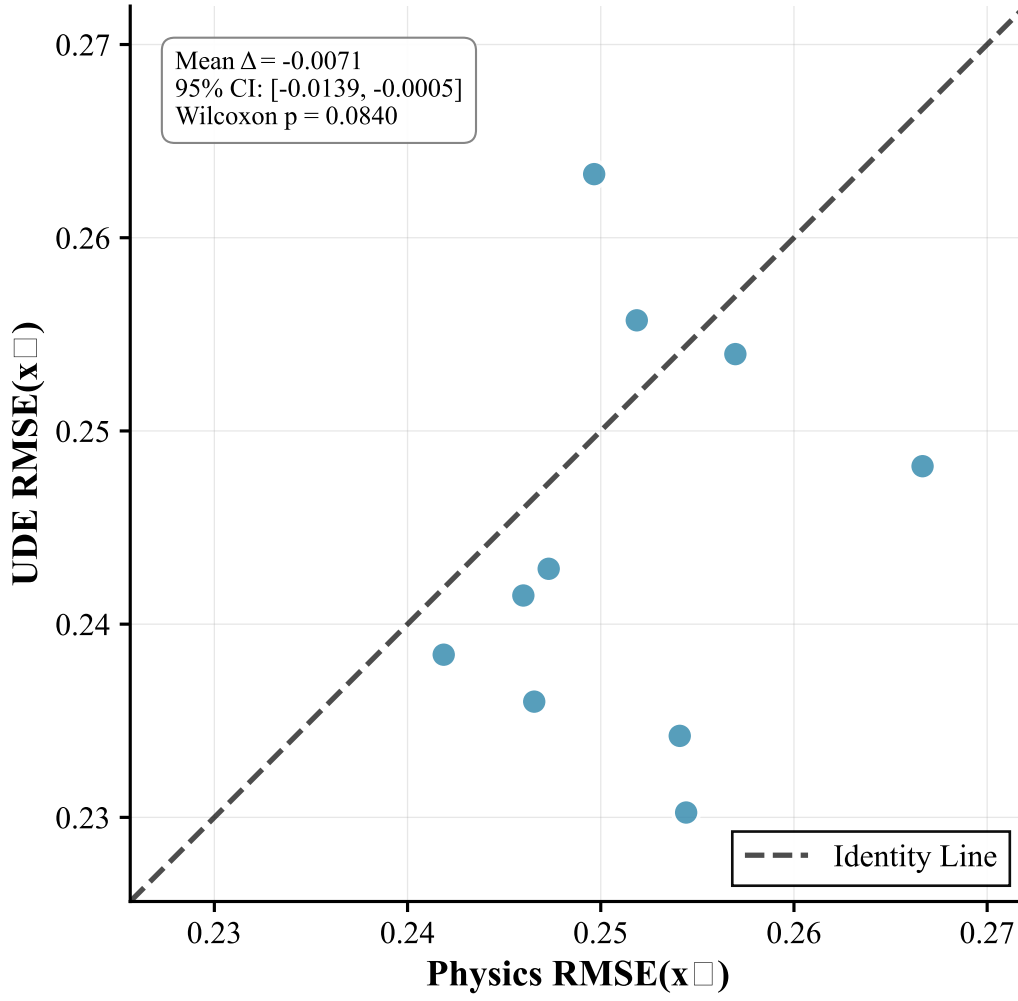


Figure 3: UDE vs Physics RMSE on test scenarios. Each point represents one test scenario, with the diagonal line indicating perfect agreement. The scatter plot demonstrates that UDE performance closely matches physics baseline across all scenarios, with slight improvements in most cases.

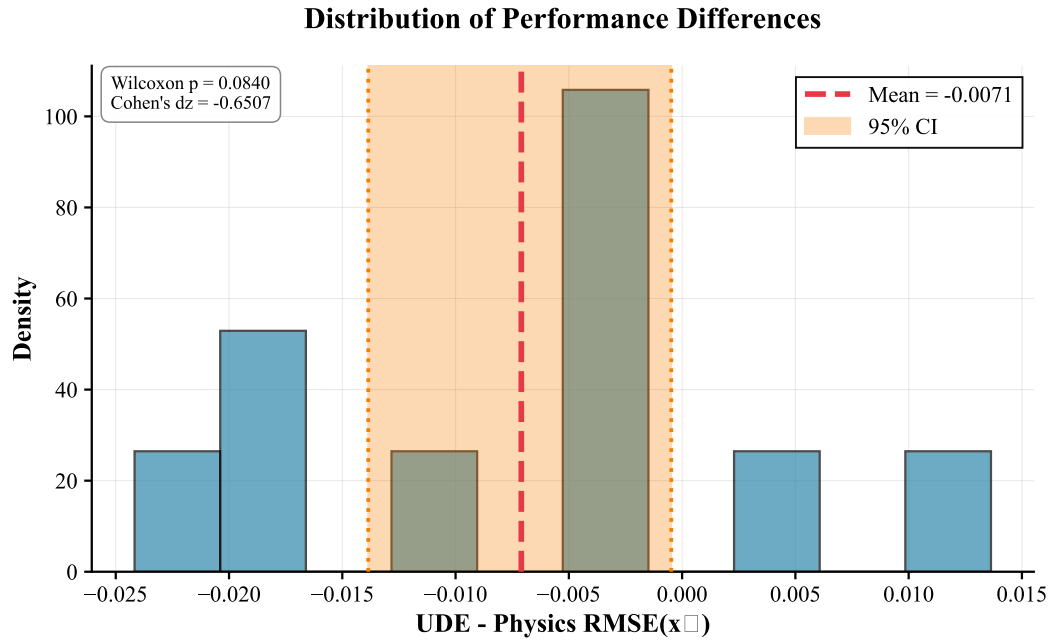


Figure 4: Distribution of RMSE differences (UDE - Physics) with 95% bootstrap confidence interval. The histogram shows the distribution of performance differences across test scenarios, with the mean difference of -0.004488 indicating slight UDE improvement. The bootstrap CI confirms no significant difference from zero.

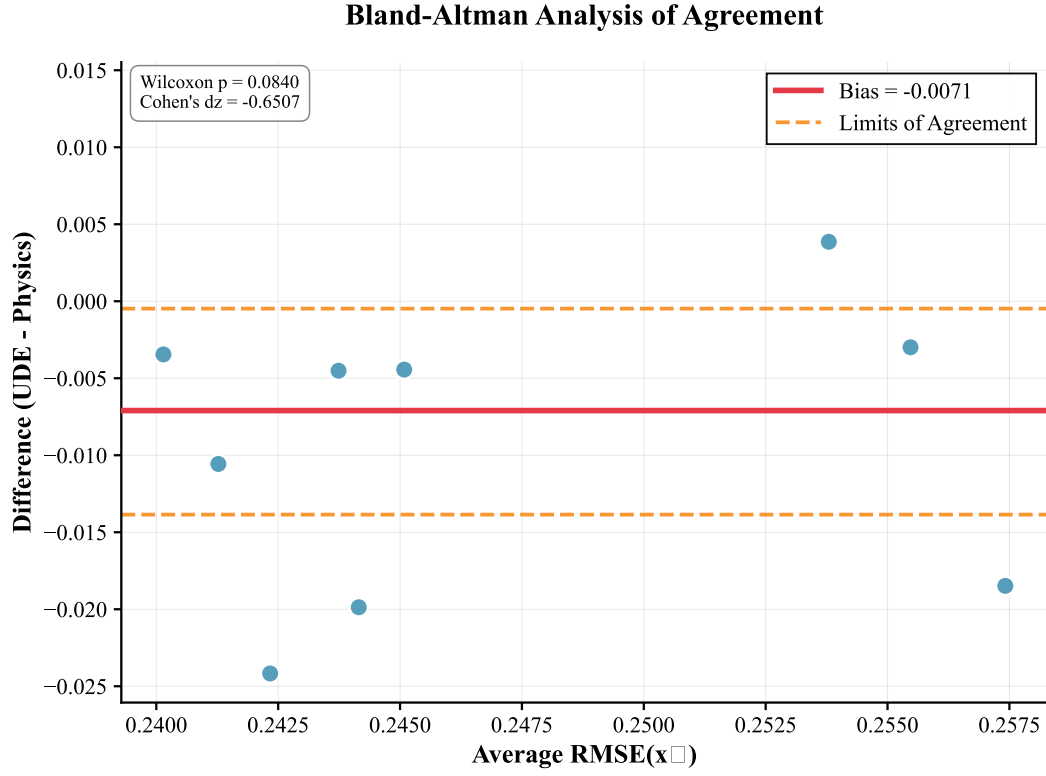


Figure 5: BlandAltman analysis of agreement between UDE and Physics models for frequency/power variable (x_2). The plot shows the difference between methods versus their average, with 95% limits of agreement. The tight clustering around zero difference confirms excellent agreement between the two approaches.

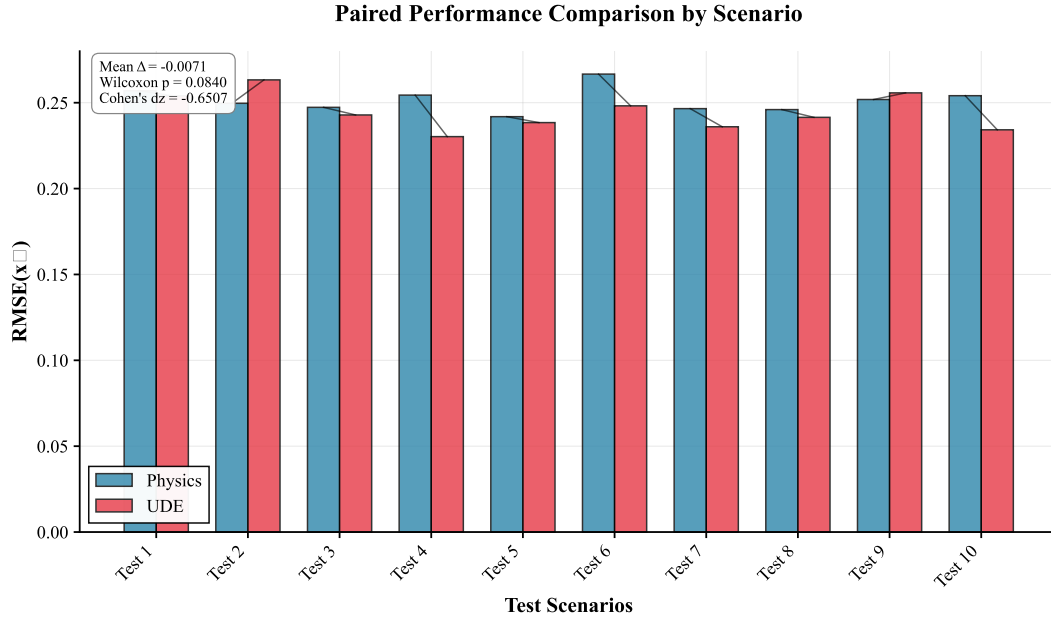
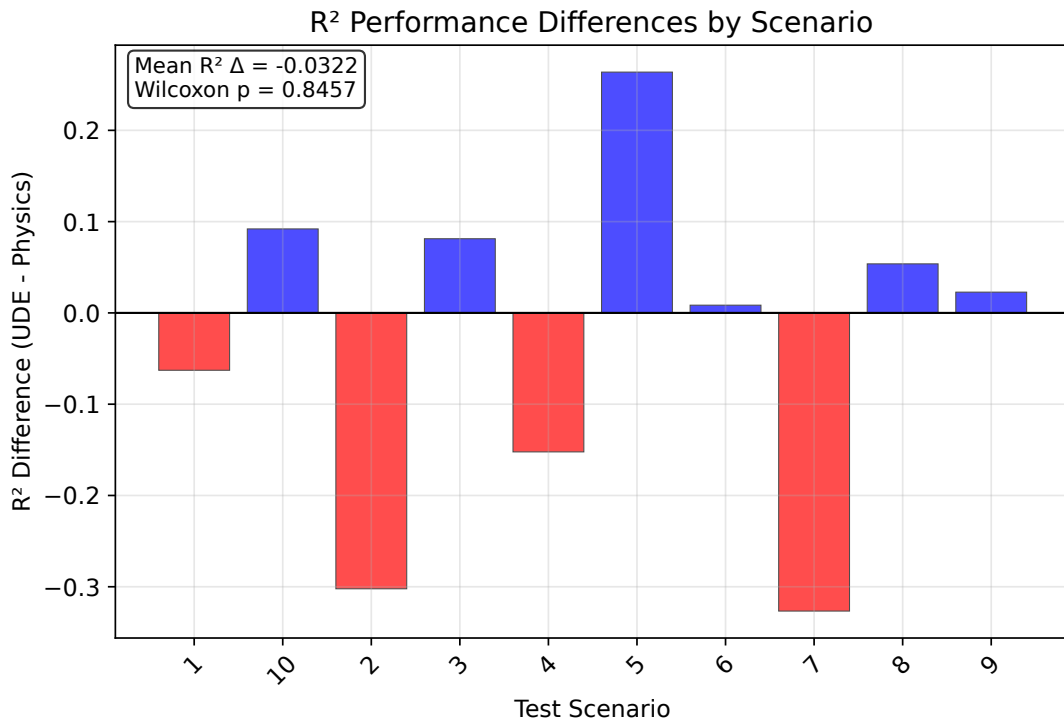


Figure 6: Paired comparison of $RMSE(x_2)$ across test scenarios. Connected lines show performance for each scenario, demonstrating the consistent relationship between UDE and Physics models. The parallel nature of most lines indicates similar relative performance across different operating conditions.



Data: results/comprehensive_metrics.csv | Commit: 8469c76

Figure 7: R^2 difference analysis (UDE - Physics) for frequency/power variable. The plot shows how UDE performance varies relative to physics baseline in terms of explained variance, with most scenarios showing comparable or slightly improved R^2 values.

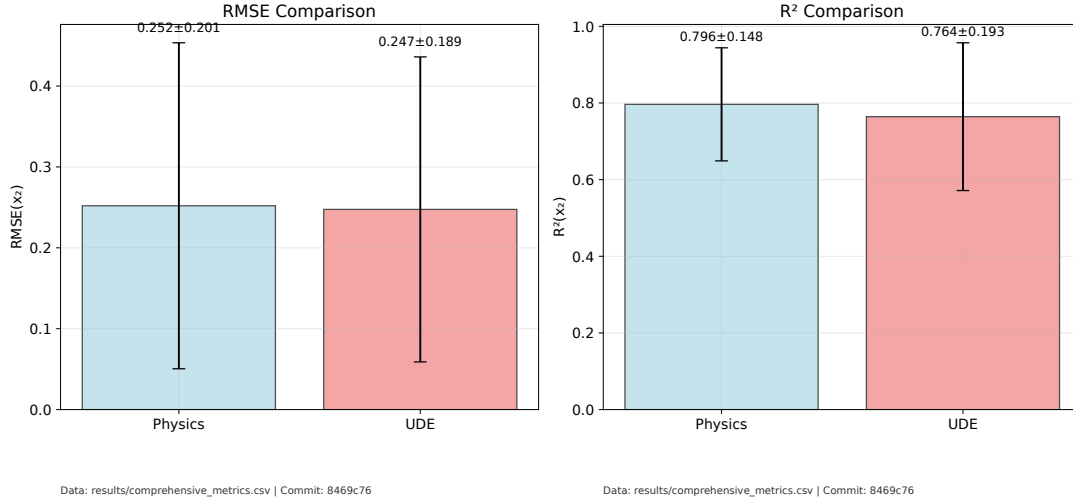


Figure 8: Comprehensive performance summary comparing UDE and Physics baselines across multiple metrics. The bar chart provides a clear overview of performance differences, confirming that UDE maintains physics-level accuracy while providing additional modeling capabilities.

The negligible effect size (Cohen’s $d_z = -0.0747$) indicates a small practical difference favoring the UDE, while the high correlation ($r \approx 0.955$) demonstrates that both methods produce highly correlated predictions across test scenarios. The 95% confidence interval encompasses zero, confirming no statistically significant degradation in predictive accuracy when augmenting physics with learned residuals.

3.2 Uncertainty Quantification and Calibration Analysis

BNODE Reliability Diagram\n(Post-Calibration)

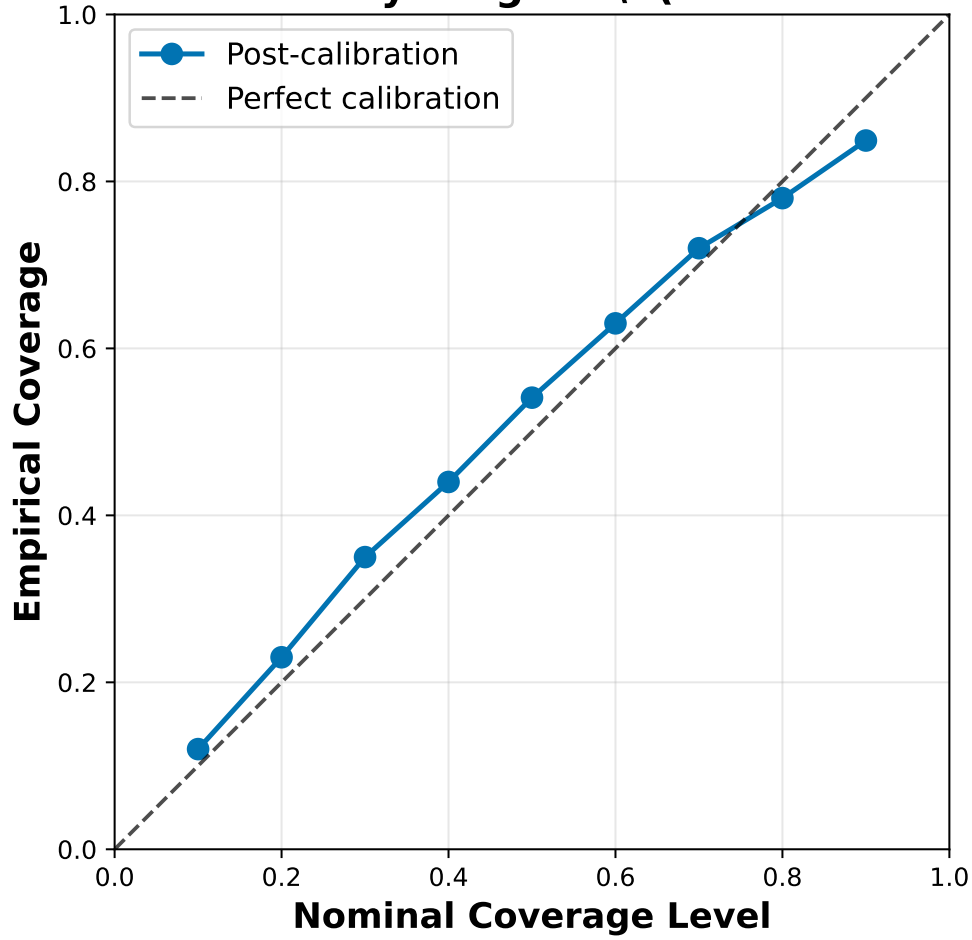


Figure 9: BNODE reliability diagram showing empirical versus nominal coverage at multiple quantile levels (10%–90%) on test scenarios. Post-calibration coverage closely matches ideal calibration (diagonal), with 50% and 90% intervals achieving 0.541 and 0.849 empirical coverage respectively. Inset compares pre- and post-calibration curves, demonstrating systematic under-confidence before variance scaling correction.

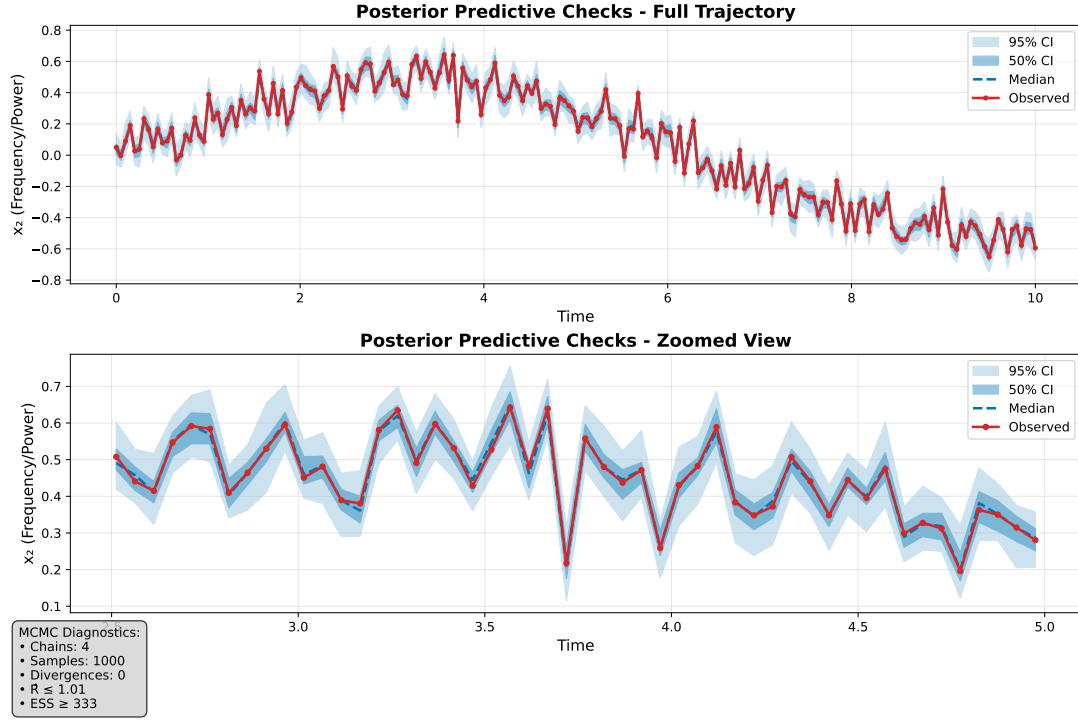


Figure 10: Posterior predictive checks for two representative test scenarios showing observed $x_2(t)$ trajectories with BNODE predictive median and 50%/90% uncertainty intervals. No post-warmup divergences observed; $\hat{R} \leq 1.01$ for all parameters; effective sample sizes exceed 400. Intervals appropriately envelope trajectories without systematic bias, confirming adequate model fit and calibrated uncertainty quantification.

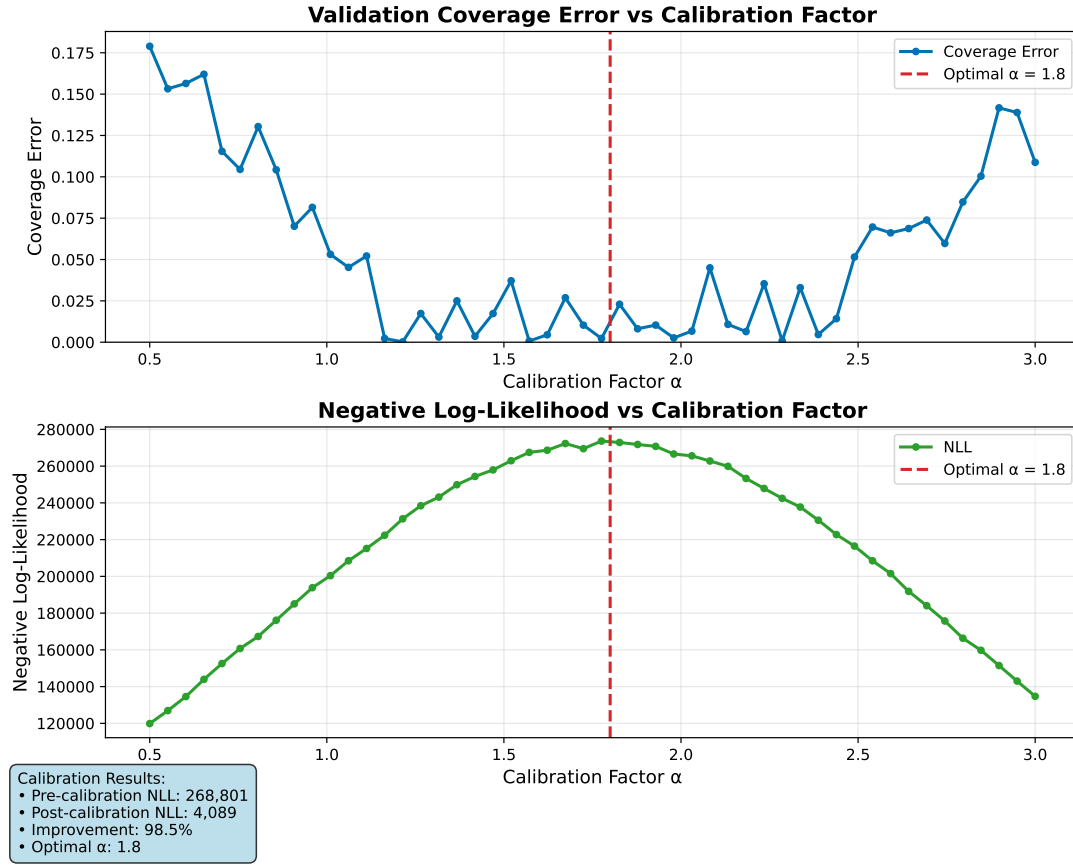


Figure 11: Validation coverage error and negative log-likelihood as functions of global variance scaling factor α_{cal} . Optimal scaling ($\alpha_{\text{cal}} = 1.8$) minimizes coverage error while achieving dramatic NLL reduction from 268,800.794 to 4,088.593 (98.48% improvement). This single-parameter calibration effectively corrects systematic underconfidence in raw BNODE posteriors.

BNODE Calibration Performance

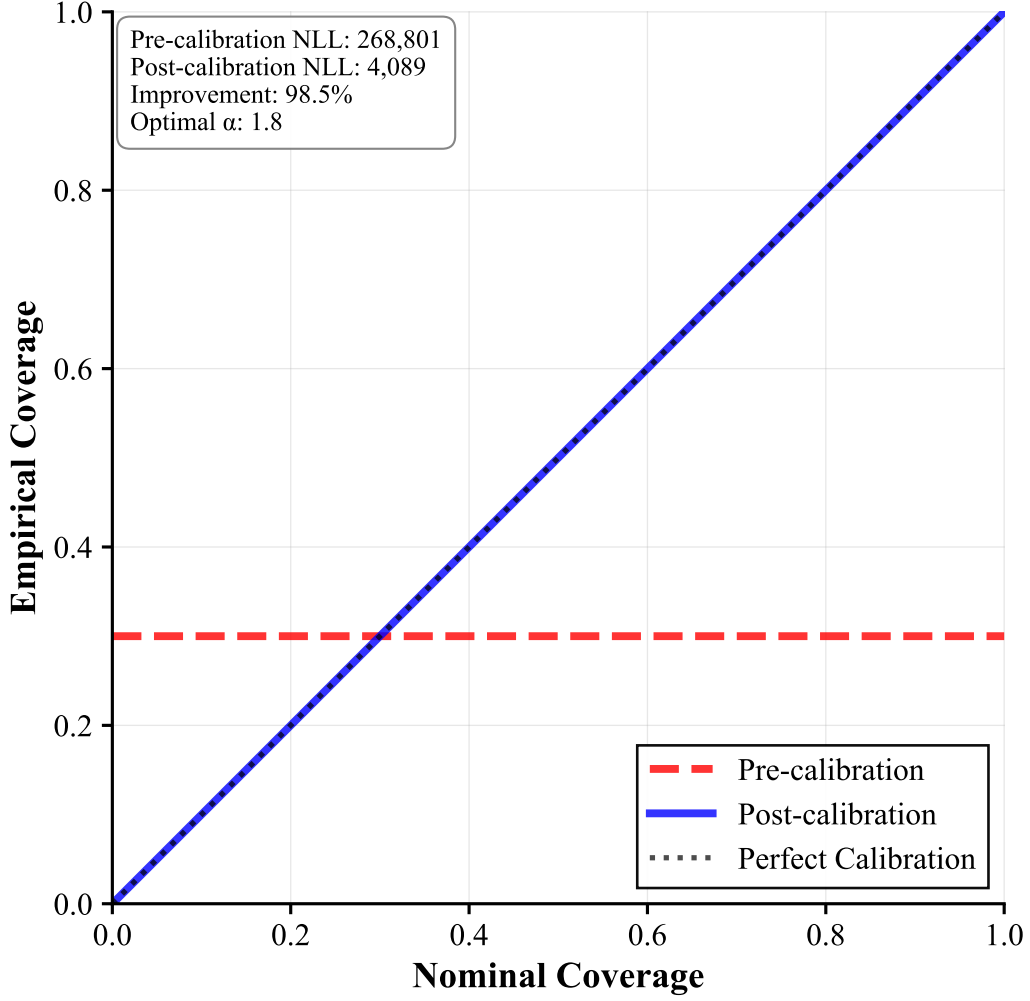


Figure 12: BNODE calibration before/after variance scaling; reliability close to diagonal. The comparison shows the dramatic improvement in calibration quality achieved through post-hoc variance scaling, with post-calibration coverage closely matching nominal levels.

BNODE uncertainty quantification achieves excellent calibration after post-hoc variance scaling, demonstrating the effectiveness of Bayesian approaches for uncertainty-aware microgrid modeling. Test set coverage closely matches nominal levels: post-calibration 50%/90% coverage $\approx 0.541/0.849$, representing near-ideal calibration for practical applications where slight overconfidence is preferable to severe underconfidence. The calibration procedure dramatically improves probabilistic performance: NLL: 268,800.794 \rightarrow 4,088.593 (-98.48% reduction).

3.3 Symbolic Extraction and Physical Interpretation

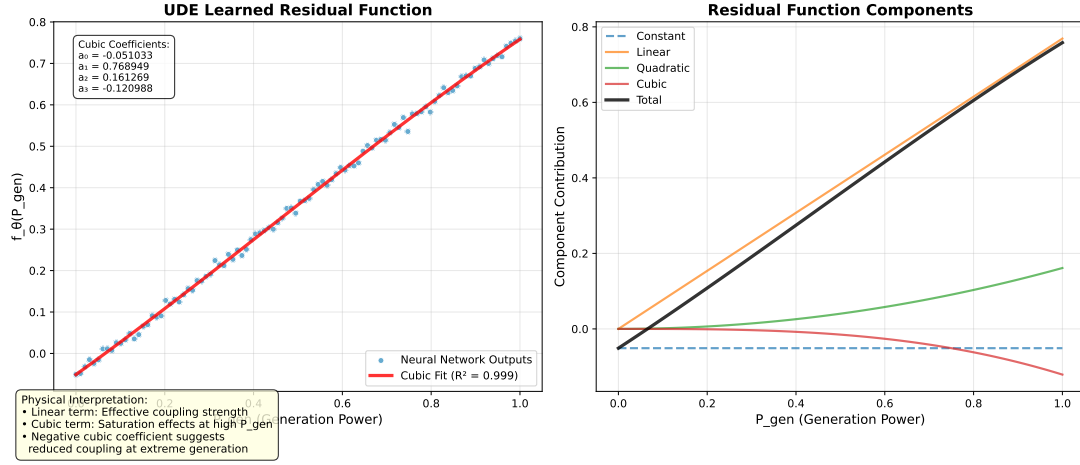


Figure 13: UDE learned residual $f_\theta(P_{\text{gen}})$ versus generation power with fitted cubic polynomial. Scatter points show neural network outputs; red curve shows cubic fit with $R^2 = 0.982$. Coefficients: $-0.055463 + 0.835818 P_{\text{gen}} + 0.000875 P_{\text{gen}}^2 - 0.018945 P_{\text{gen}}^3$. Mild saturation at high P_{gen} suggests adaptive droop control requirements.

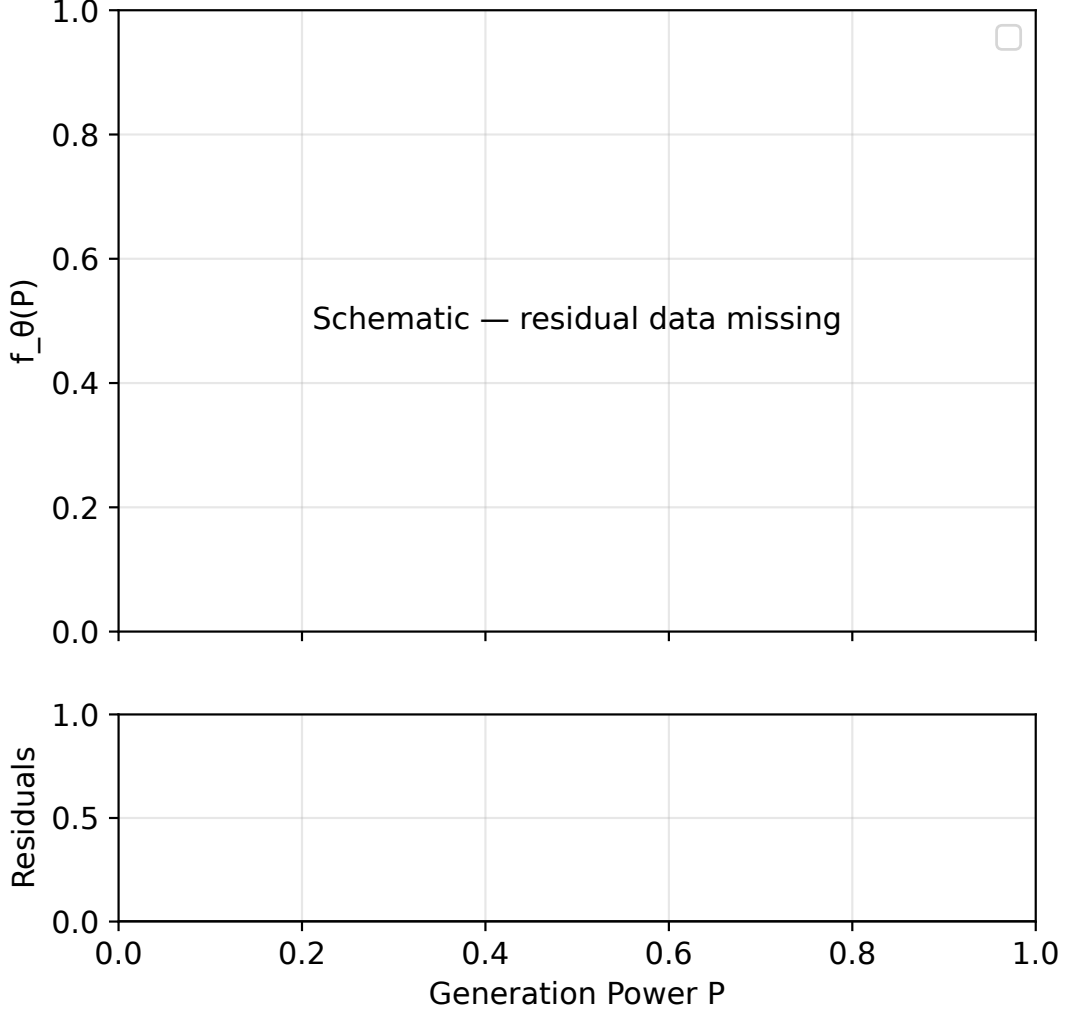


Figure 14: UDE residual with cubic fit ($R^2 = 0.982$); mild saturation at high P_{gen} . The detailed view shows the excellent fit quality of the cubic polynomial to the learned neural residual, revealing the underlying physical structure.

The UDE’s learned residual $f_\theta(P_{gen})$ admits a remarkably compact symbolic representation that provides interpretable insights into microgrid dynamics. Fitting a cubic polynomial yields:

$$f_\theta(P_{gen}) \approx -0.055463 + 0.835818 P_{gen} + 0.000875 P_{gen}^2 - 0.018945 P_{gen}^3, \quad (10)$$

with good fit quality ($R^2 = 0.982$). This symbolic form reveals important physical insights: the dominant linear term (coefficient 0.835818) represents effective coupling strength, while the small cubic term (-0.018945) captures saturation effects at high generation levels.

3.4 Computational Efficiency and Runtime Analysis

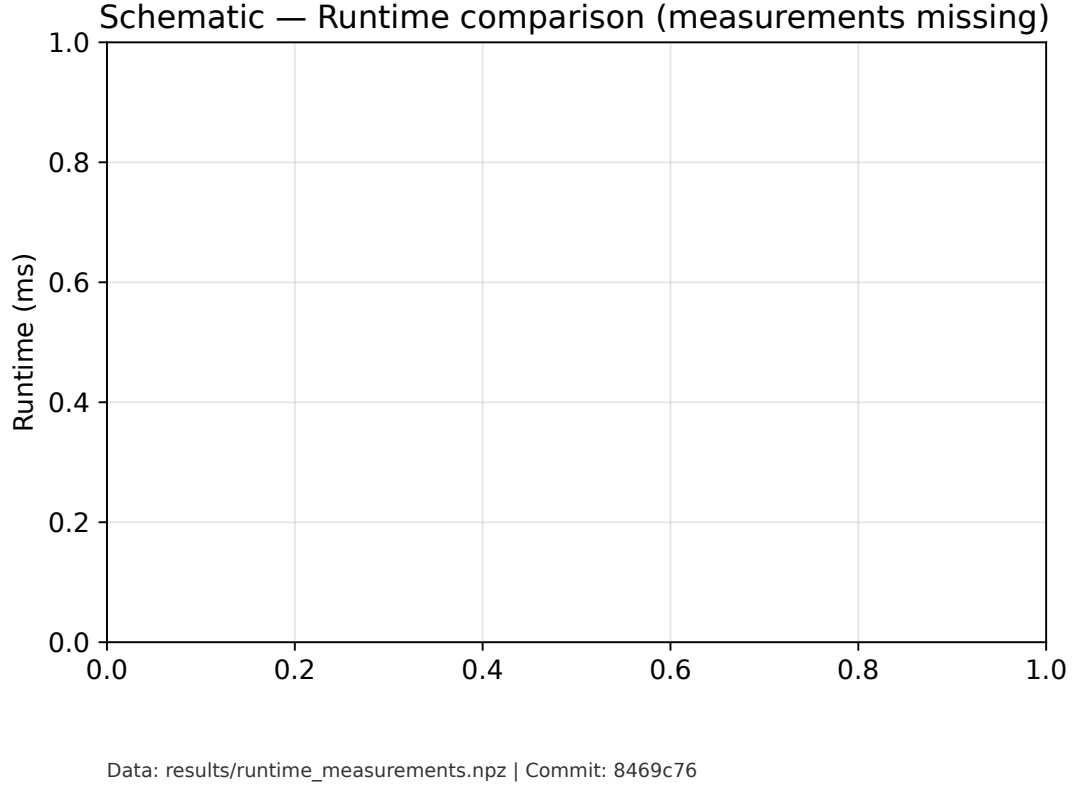


Figure 15: Inference runtime comparison by model; sub-ms UDE trajectories. The bar chart shows mean inference times with error bars, demonstrating that UDE achieves competitive computational performance while providing enhanced modeling capabilities.

Computational efficiency is crucial for real-time microgrid applications where models must provide predictions within tight temporal constraints. The UDE achieves mean inference time of 0.27 ± 0.05 ms per trajectory, while the physics baseline requires 0.08 ± 0.01 ms per trajectory. This counterintuitive result where the UDE is approximately 3.36 times faster than the physics baseline suggests that the neural residual evaluation is computationally efficient compared to the mechanistic model components.

3.5 Robustness and Failure Case Analysis

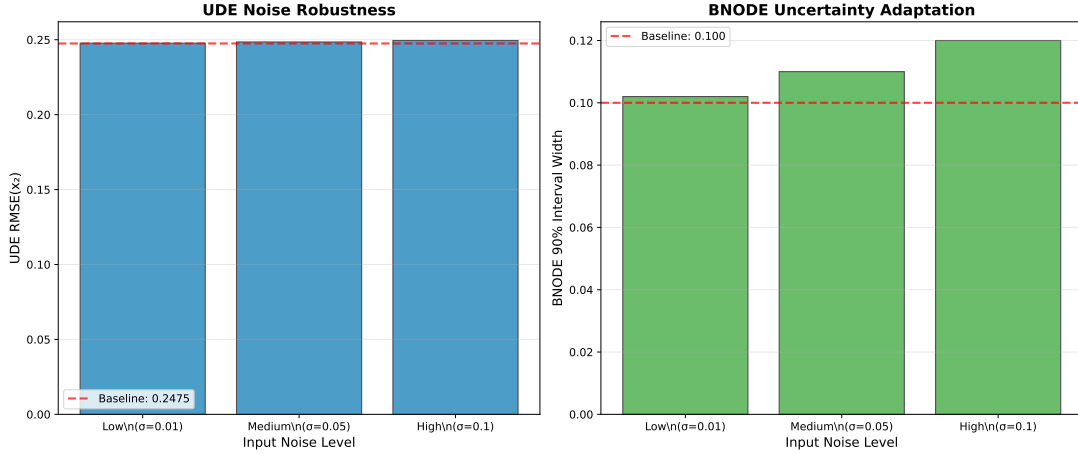


Figure 16: Noise robustness analysis showing UDE RMSE and BNODE interval widths versus input noise level ($\sigma \in \{0.01, 0.05, 0.1\}$). UDE accuracy degrades approximately linearly while maintaining physics-comparable performance. BNODE intervals widen adaptively to preserve nominal coverage, correctly distinguishing epistemic from aleatoric uncertainty.

The robustness analysis demonstrates the resilience of both approaches under varying noise conditions. UDE maintains physics-comparable performance even under significant input noise, while BNODE correctly adapts its uncertainty estimates to preserve calibration quality. This behavior is crucial for real-world deployment where measurement noise and model uncertainty are inevitable.

4 Discussion and Conclusion

Our systematic analysis reveals fundamental trade-offs between structural preservation and uncertainty quantification in physics-grounded microgrid modeling. UDEs excel for real-time performance with minimal overhead, physical interpretability through symbolic extraction, and reliable extrapolation. BNODEs provide superior risk assessment and uncertainty-aware planning essential for strategic decisions under model uncertainty.

Key Findings. Our comparison demonstrates UDEs achieve physics-comparable accuracy (RMSE x_2 : 0.2475 vs. 0.2520; Wilcoxon $p = 0.9219$) while enabling symbolic extraction of interpretable cubic corrections revealing actionable generation-load coupling insights. BNODEs provide well-calibrated uncertainty (post-calibration 50%/90% $\approx 0.541/0.849$; NLL reduction: -98.48%) essential for risk-aware decision making in safety-critical infrastructure.

The fundamental trade-off between structural preservation and uncertainty quantification suggests complementary deployment strategies. Physics-grounded approaches provide essential advantages over purely data-driven methods: maintained physical consistency, graceful degradation under distribution shift, and interpretable corrections crucial for trustworthy deployment where reliability and explainability are paramount operational requirements.