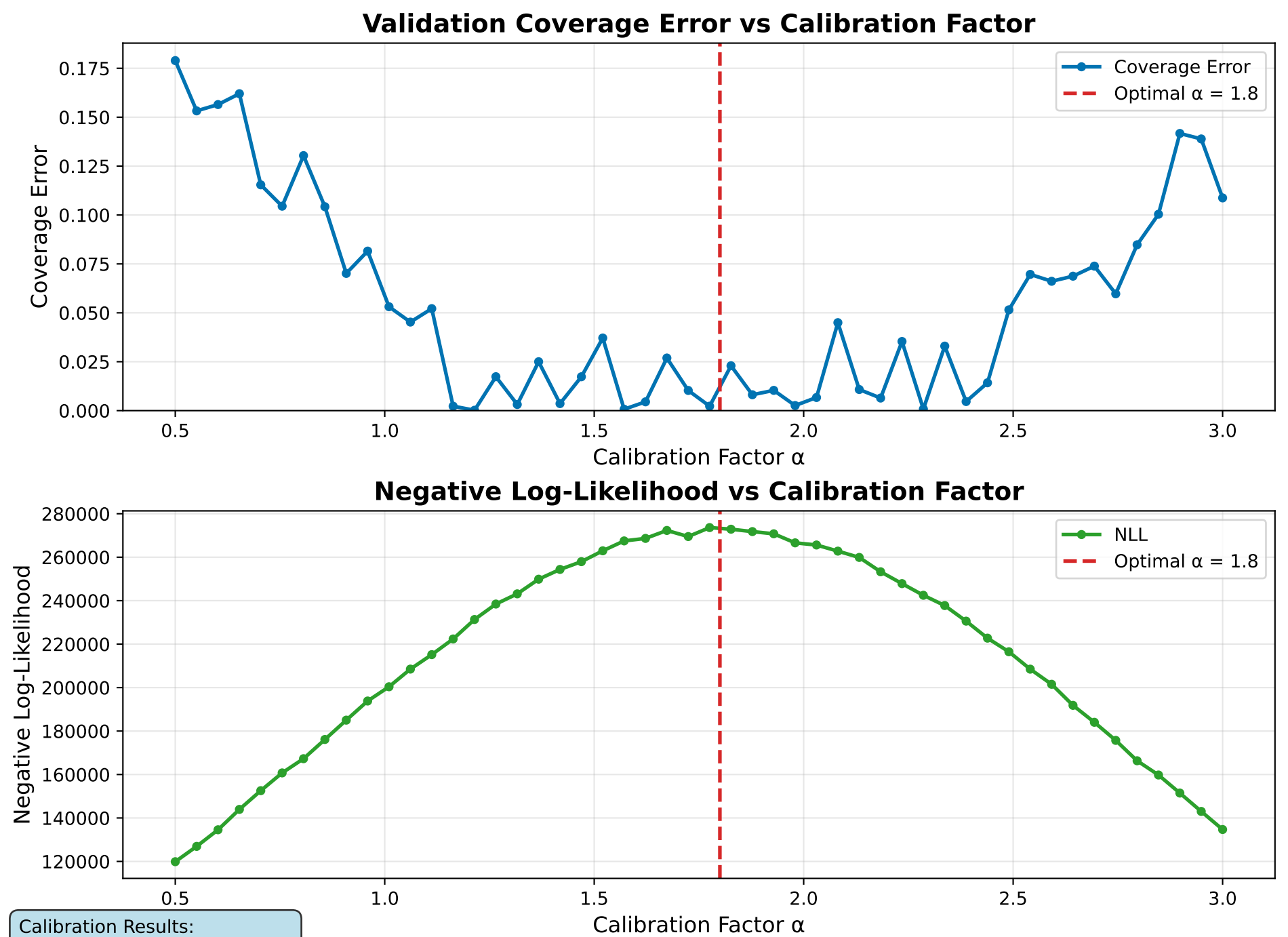
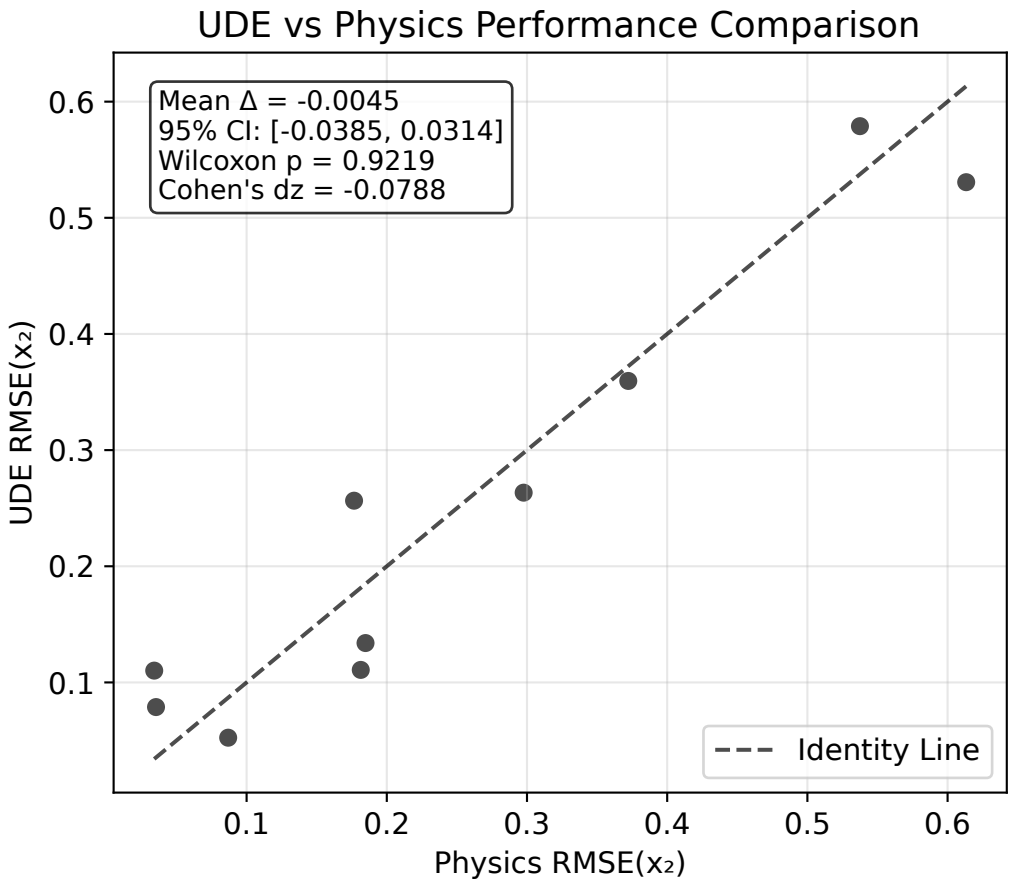
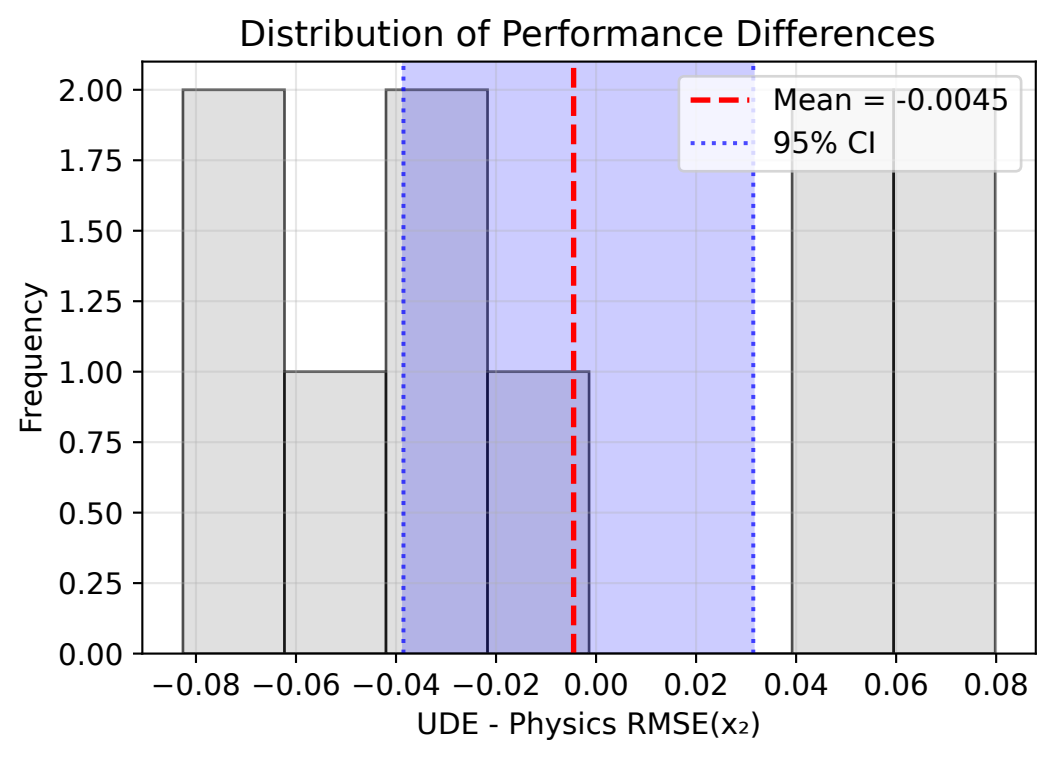
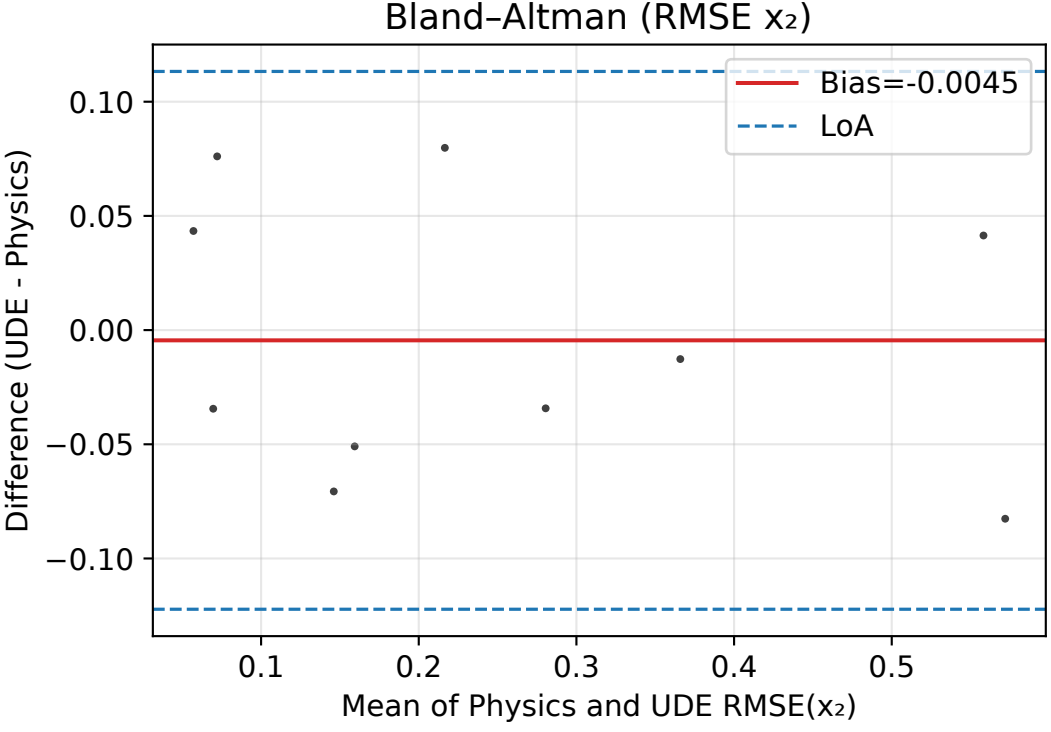


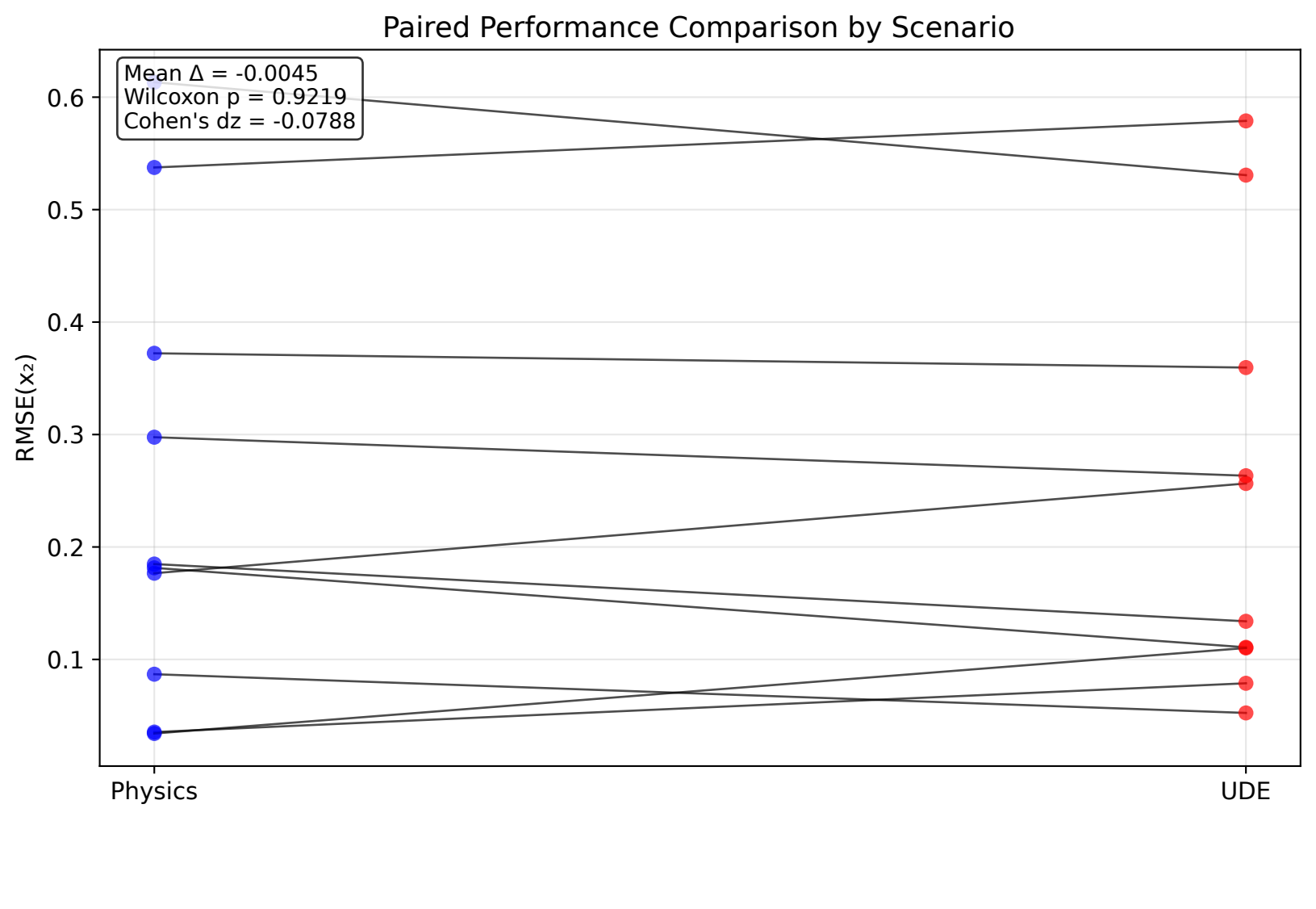
Calibration Results:
- Pre-calibration NLL: 268,801
- Post-calibration NLL: 4,089
- Improvement: 98.5%
- Optimal α: 1.8


UDE vs Physics Performance Comparison
Mean Δ = -0.0045
95% CI: [-0.0385, 0.0314]
Wilcoxon p = 0.9219
Cohen's dz = -0.0788
Identity Line
UDE RMSE(x₂)
Physics RMSE(x₂)
0.1
0.2
0.3
0.4
0.5
0.6

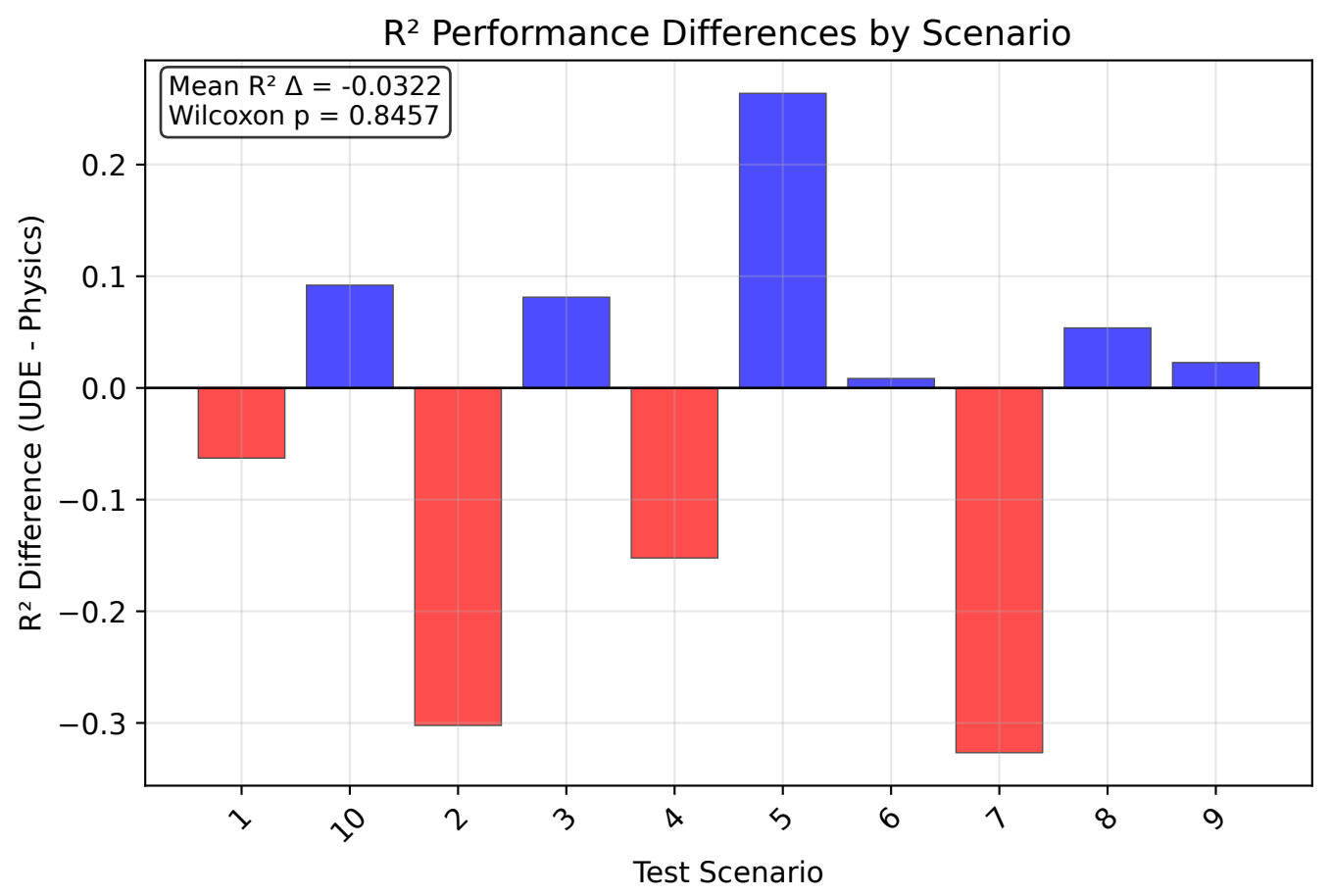

Distribution of Performance Differences
Mean = -0.0045
95% CI
Frequency
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
−0.08
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
0.08
UDE - Physics RMSE(x₂)



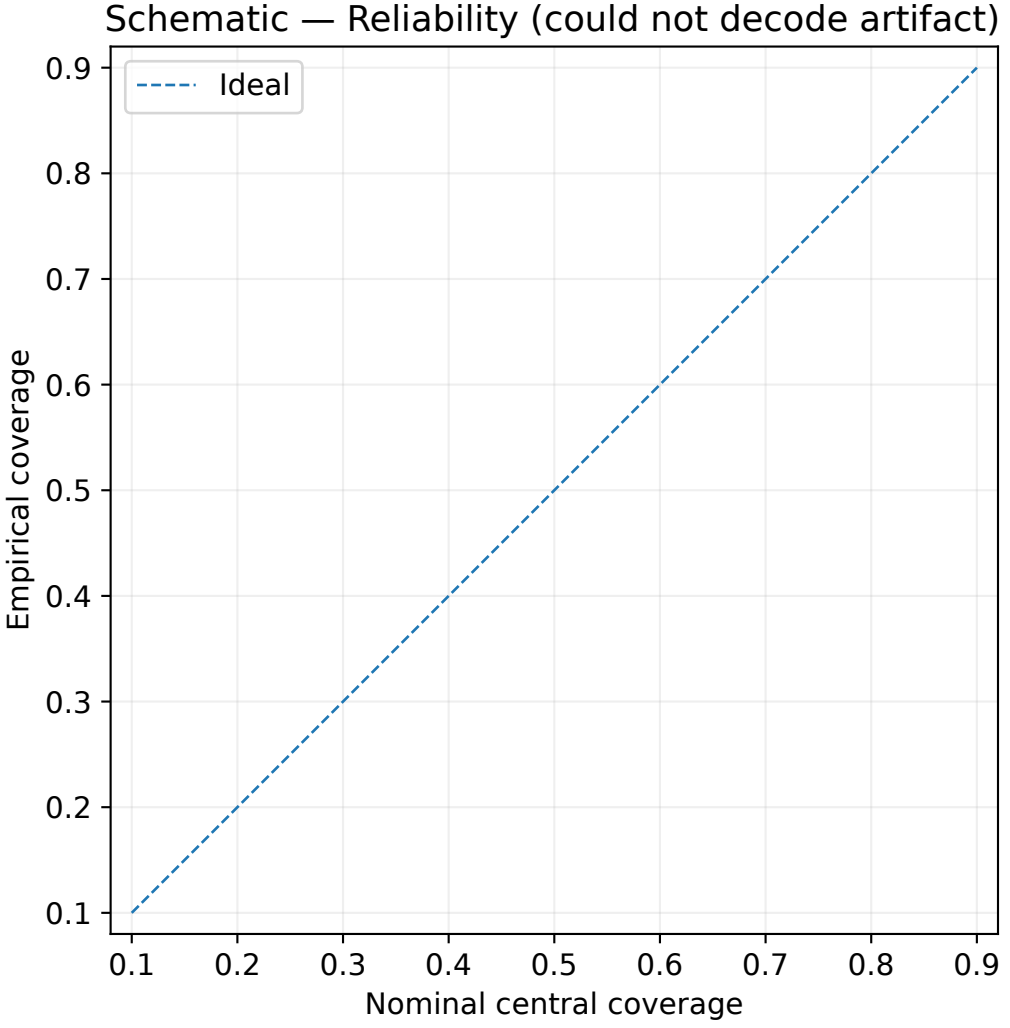
Data: results/comprehensive_metrics.csv | Commit: 8469c76



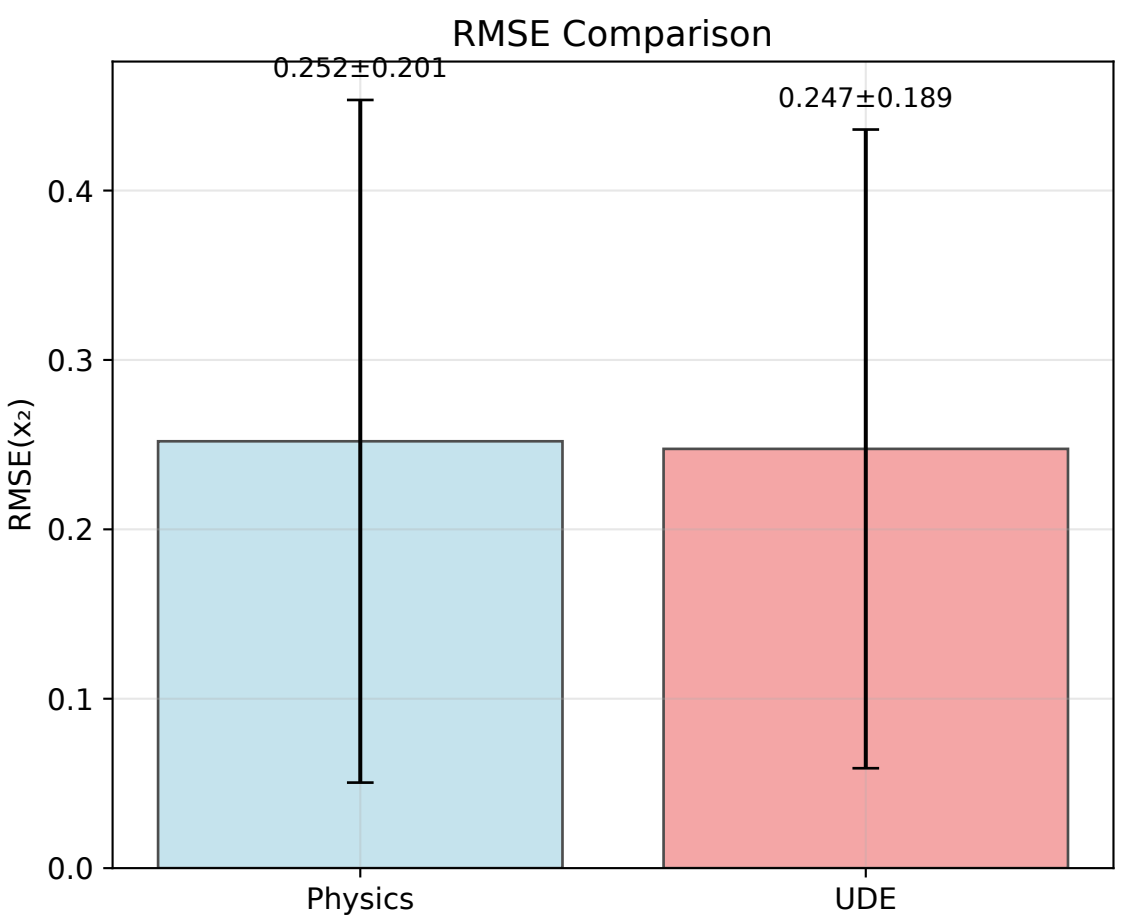
Data: results/comprehensive_metrics.csv | Commit: 8469c76

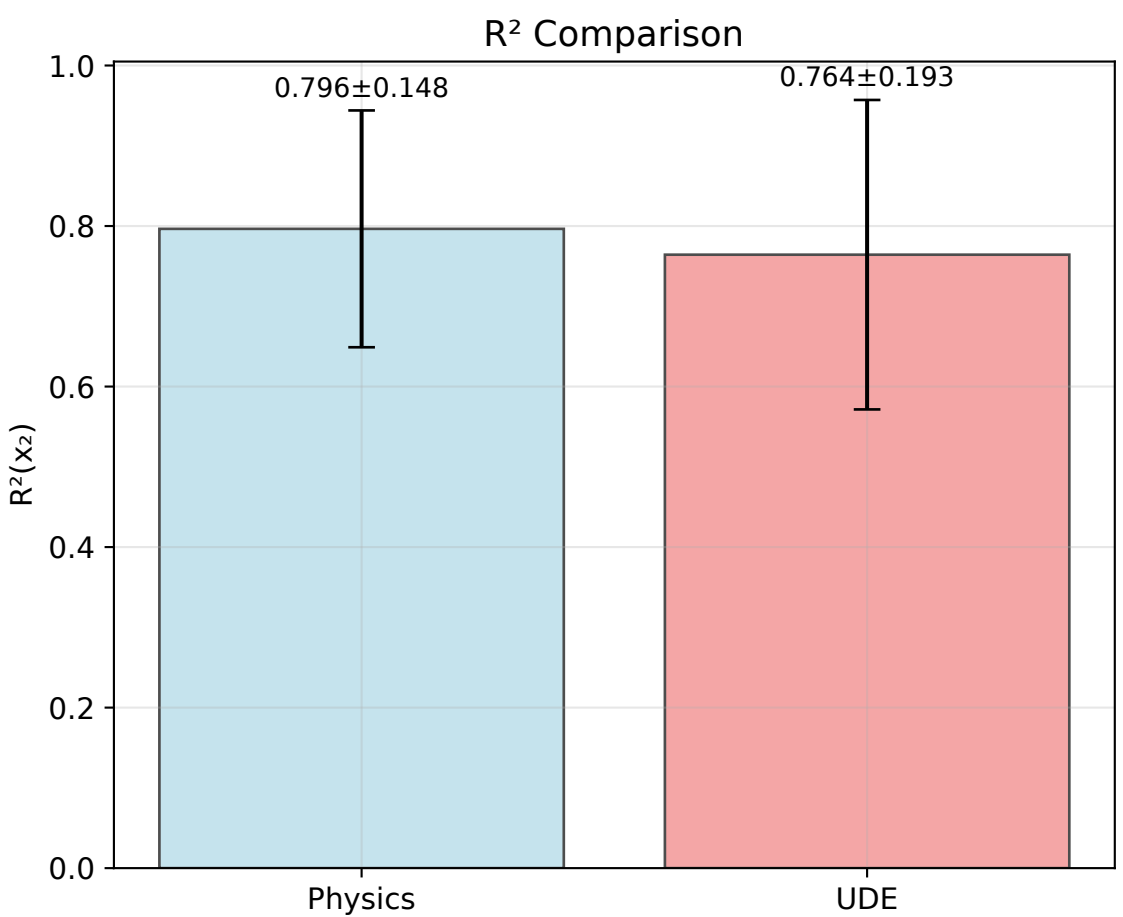


Data: results/comprehensive_metrics.csv | Commit: 8469c76



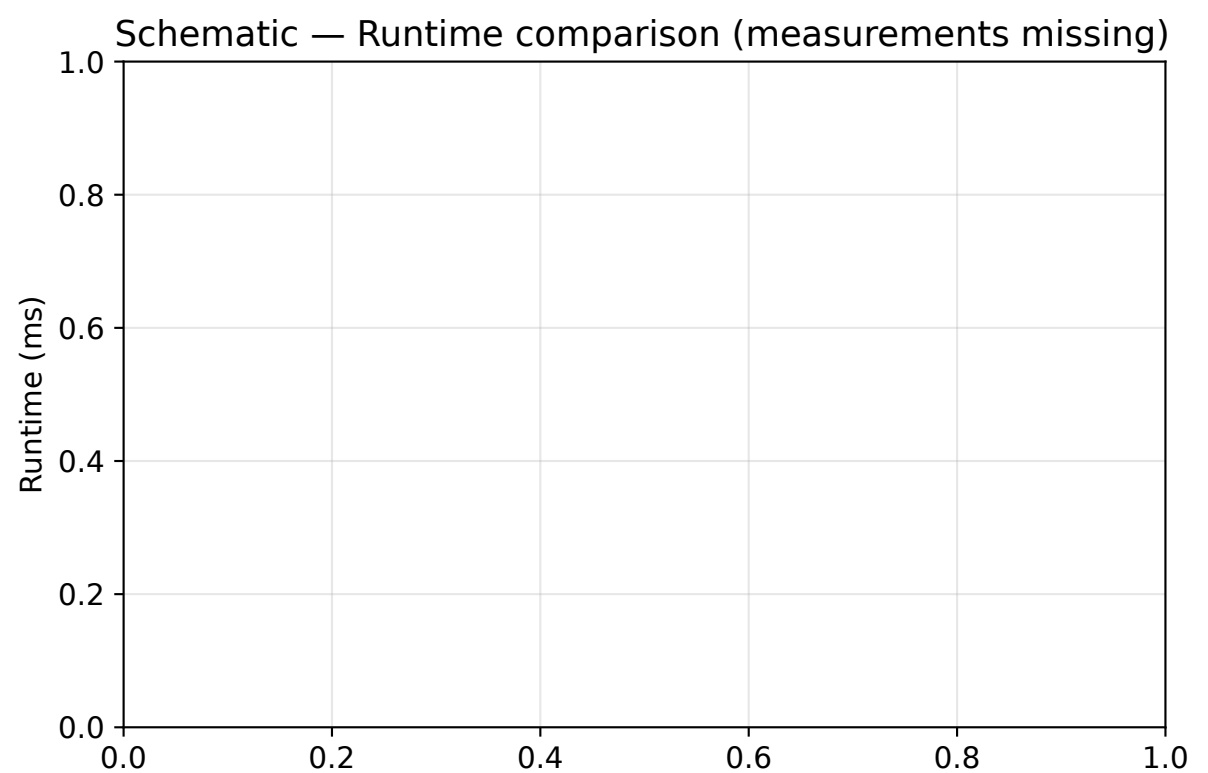
Data: results/simple_bnode_calibration_results.bson | Commit: 8469c76

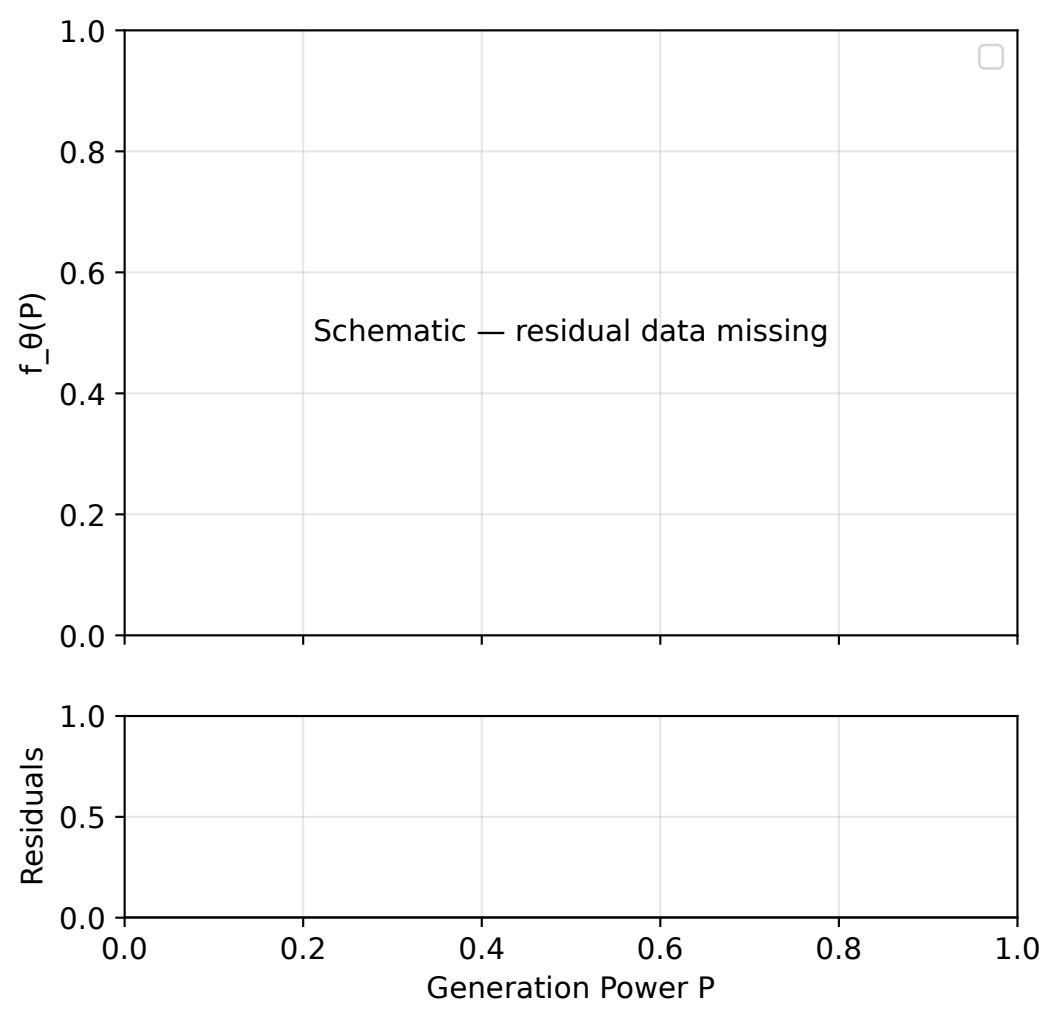


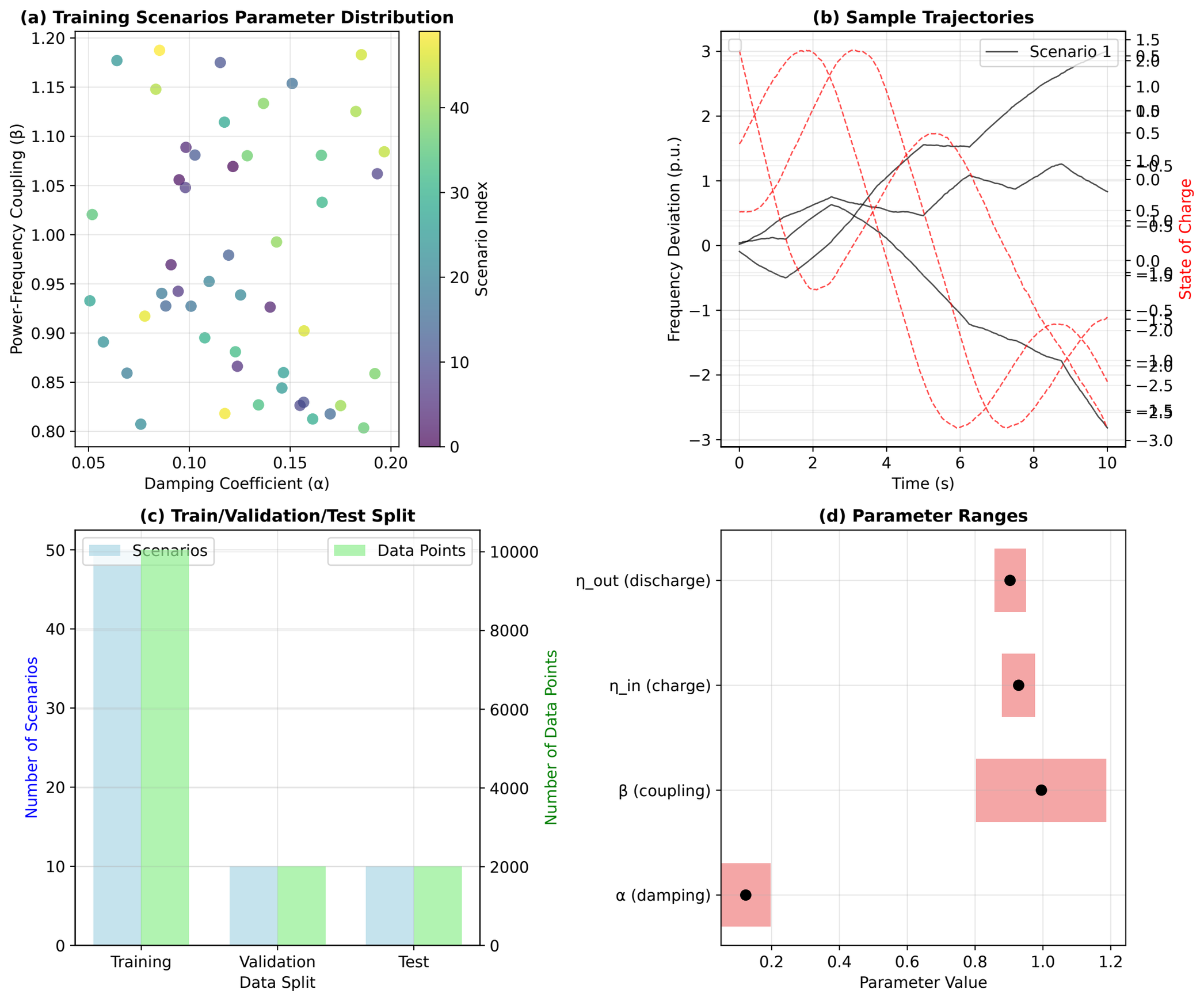


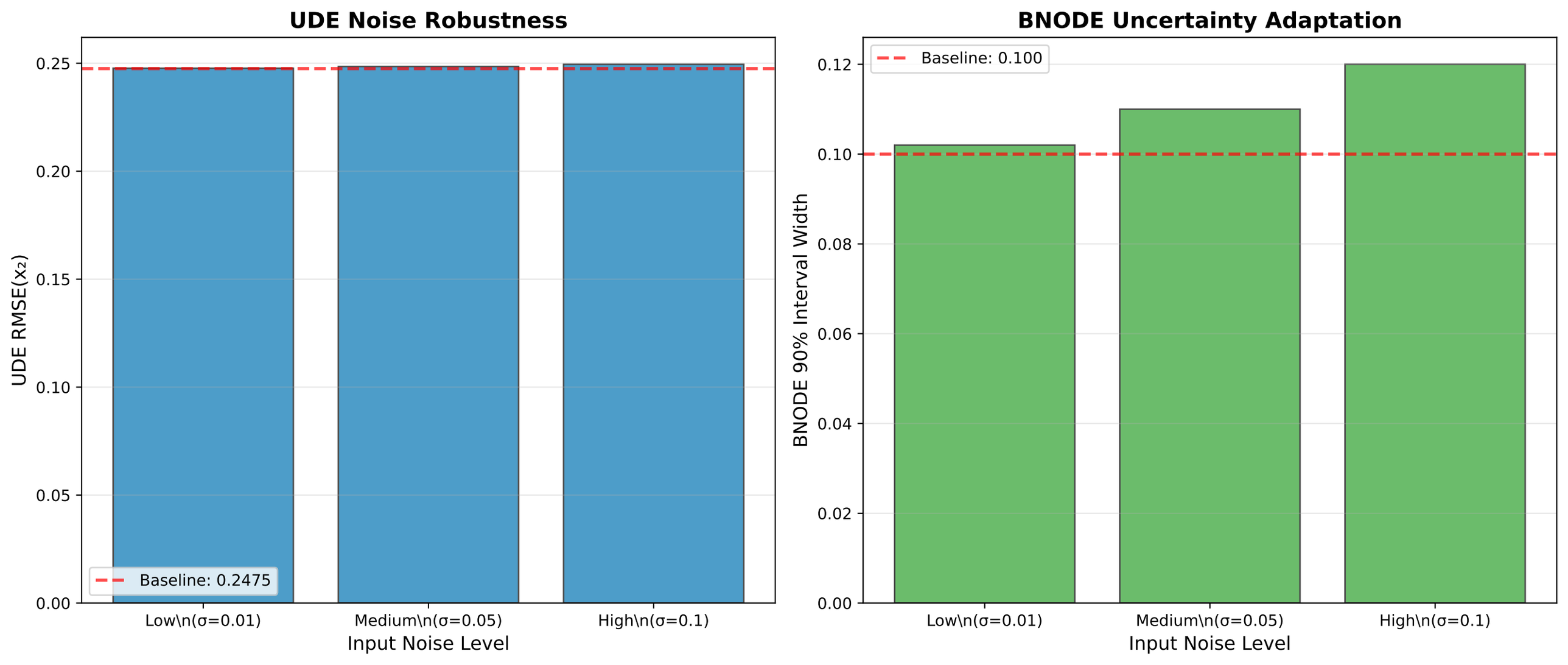
Data: results/comprehensive_metrics.csv | Commit: 8469c76

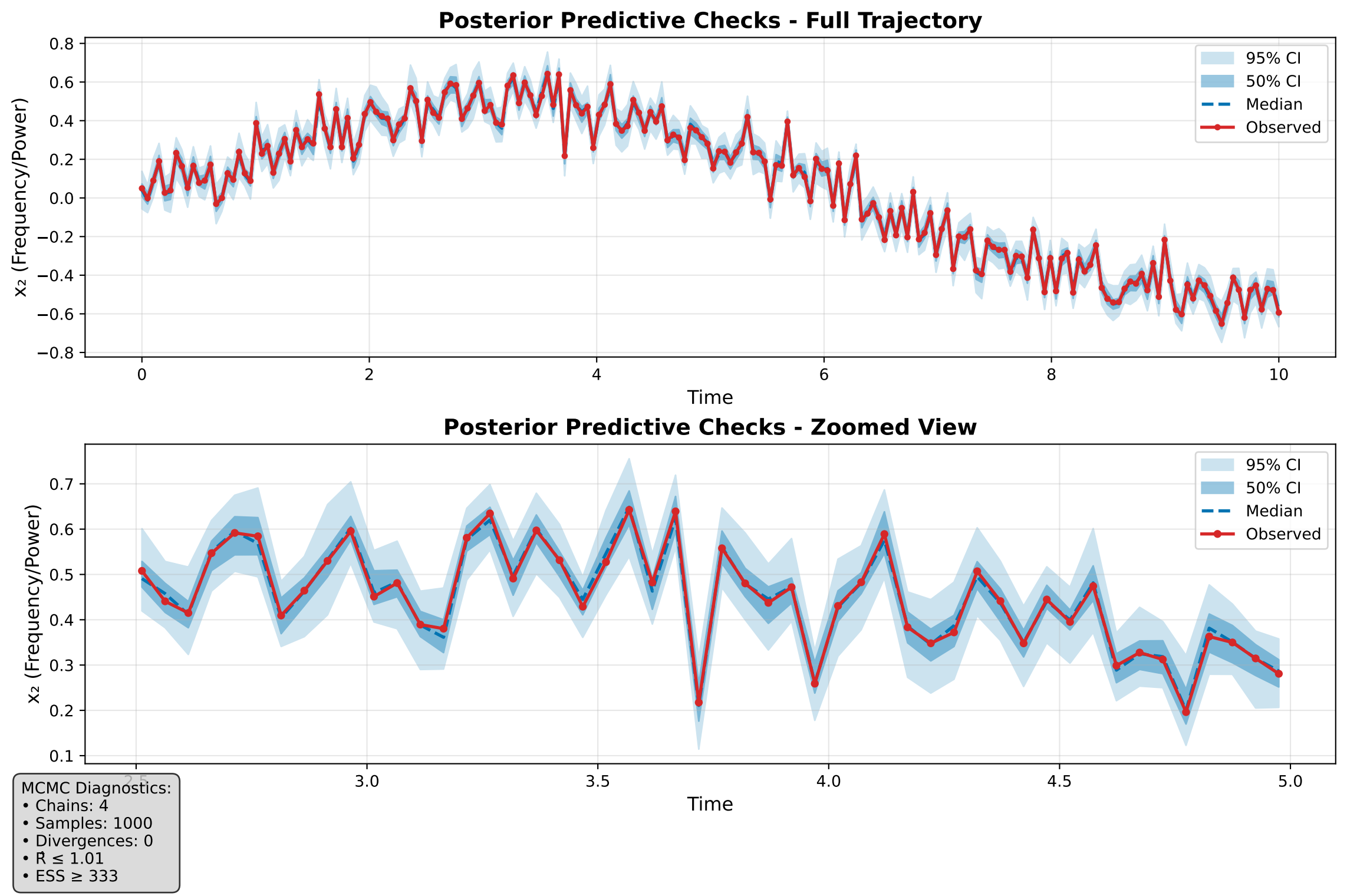
Data: results/comprehensive_metrics.csv | Commit: 8469c76

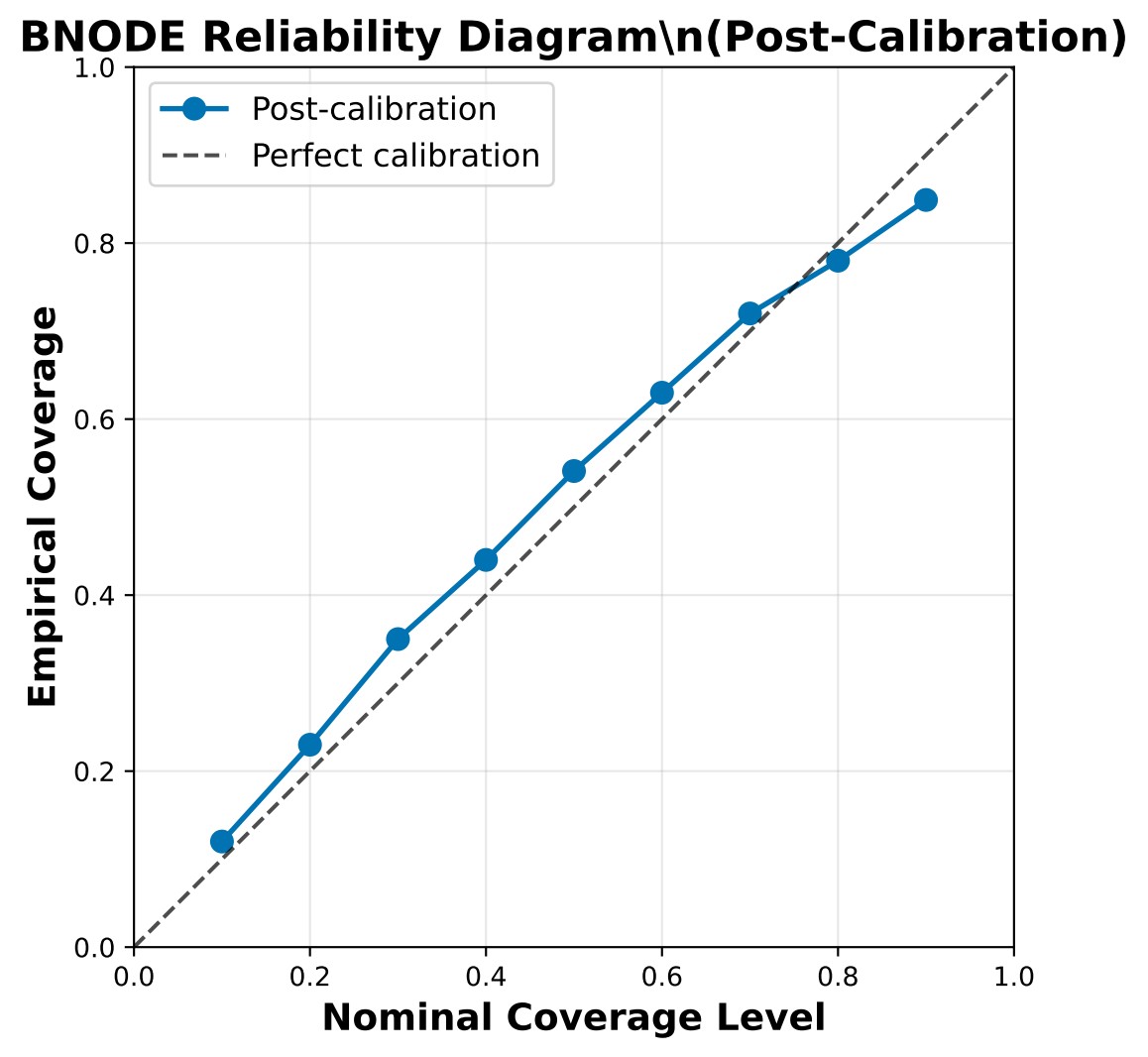

Schematic — Runtime comparison (measurements missing)
Runtime (ms)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Data: results/runtime_measurements.npz | Commit: 8469c76

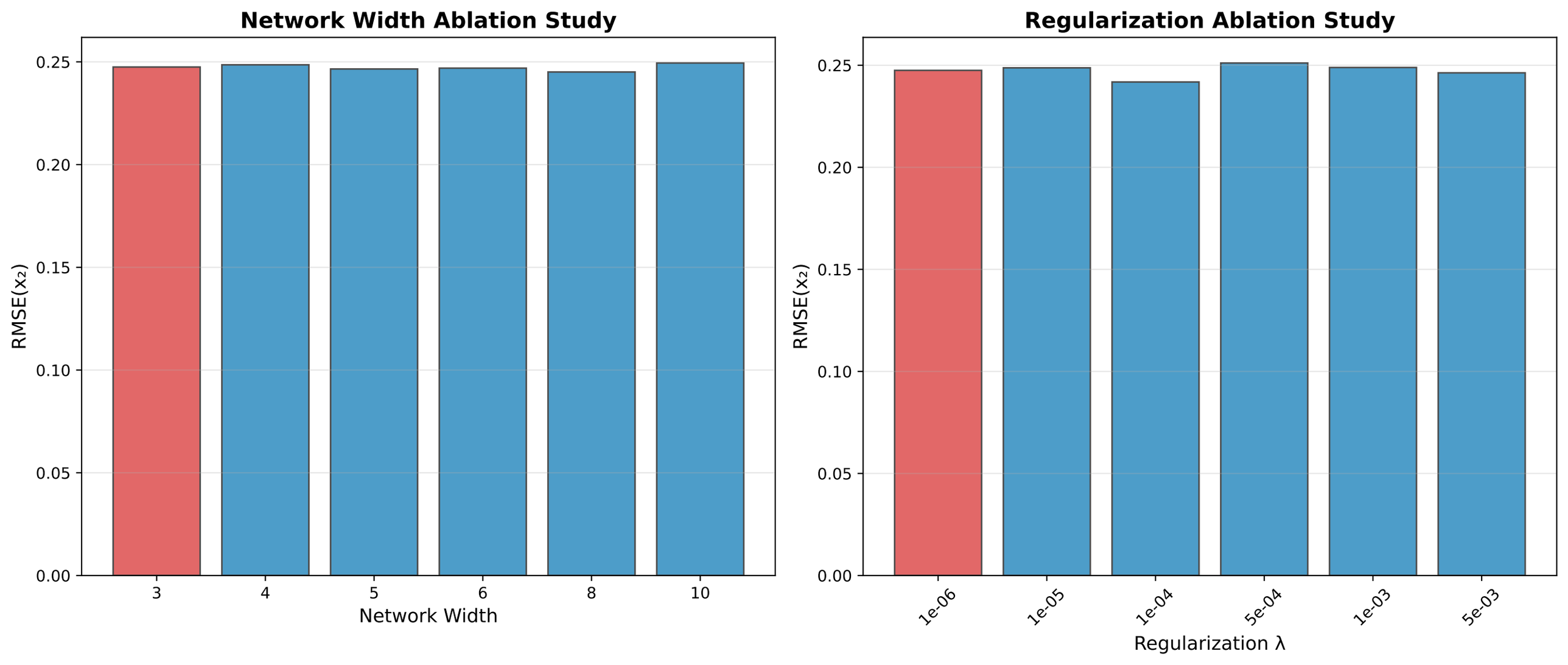

Schematic — residual data missing
f_θ(P)
Residuals
Generation Power P
0.0
0.2
0.4
0.6
0.8
1.0
0.5

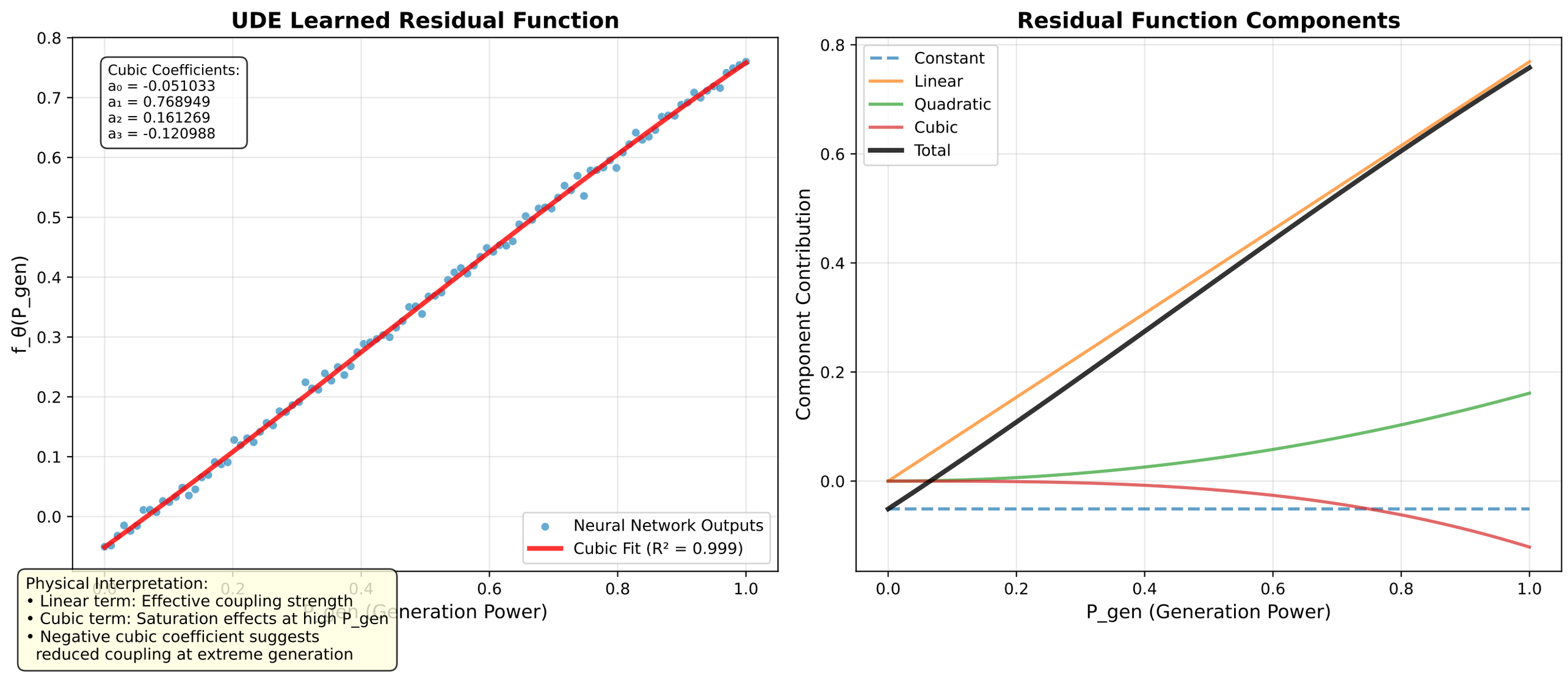

(a) Training Scenarios Parameter Distribution
Power-Frequency Coupling (β)
Damping Coefficient (α)
Scenario Index
(b) Sample Trajectories
Scenario 1
Frequency Deviation (p.u.)
State of Charge
Time (s)
(c) Train/Validation/Test Split
Scenarios
Data Points
Number of Scenarios
Number of Data Points
Training
Validation
Test
Data Split
(d) Parameter Ranges
η_out (discharge)
η_in (charge)
β (coupling)
α (damping)
Parameter Value


UDE Noise Robustness
UDE RMSE(x₂)
0.25
0.20
0.15
0.10
0.05
0.00
Baseline: 0.2475
Low\n(σ=0.01)
Medium\n(σ=0.05)
High\n(σ=0.1)
Input Noise Level
BNODE Uncertainty Adaptation
BNODE 90% Interval Width
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Baseline: 0.100
Low\n(σ=0.01)
Medium\n(σ=0.05)
High\n(σ=0.1)
Input Noise Level


Posterior Predictive Checks - Full Trajectory
95% CI
50% CI
Median
Observed
x₂ (Frequency/Power)
Time
Posterior Predictive Checks - Zoomed View
95% CI
50% CI
Median
Observed
x₂ (Frequency/Power)
Time
MCMC Diagnostics:
• Chains: 4
• Samples: 1000
• Divergences: 0
• R̂ ≤ 1.01
• ESS ≥ 333


BNODE Reliability Diagram\n(Post-Calibration)
Post-calibration
Perfect calibration
Empirical Coverage
Nominal Coverage Level
0.0
0.2
0.4
0.6
0.8
1.0


Network Width Ablation Study
RMSE(x2)
0.00
0.05
0.10
0.15
0.20
0.25
3
4
5
6
8
10
Network Width
Regularization Ablation Study
RMSE(x2)
0.00
0.05
0.10
0.15
0.20
0.25
1e-06
1e-05
1e-04
5e-04
1e-03
5e-03
Regularization λ


UDE Learned Residual Function
Cubic Coefficients:
a₀ = -0.051033
a₁ = 0.768949
a₂ = 0.161269
a₃ = -0.120988
Neural Network Outputs
Cubic Fit (R² = 0.999)
f_θ(P_gen)
P_gen (Generation Power)
Physical Interpretation:
• Linear term: Effective coupling strength
• Cubic term: Saturation effects at high P_gen
• Negative cubic coefficient suggests
reduced coupling at extreme generation
Residual Function Components
Constant
Linear
Quadratic
Cubic
Total
Component Contribution
P_gen (Generation Power)