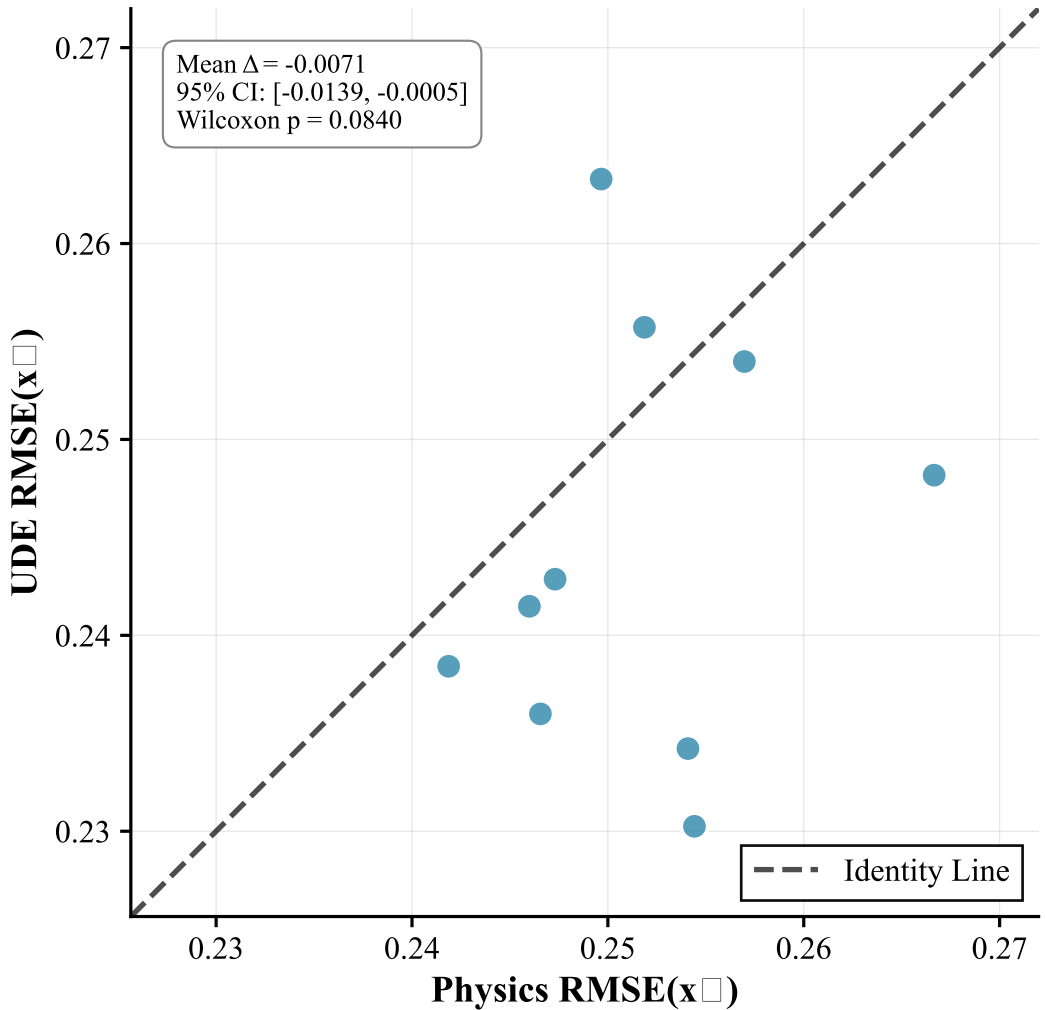
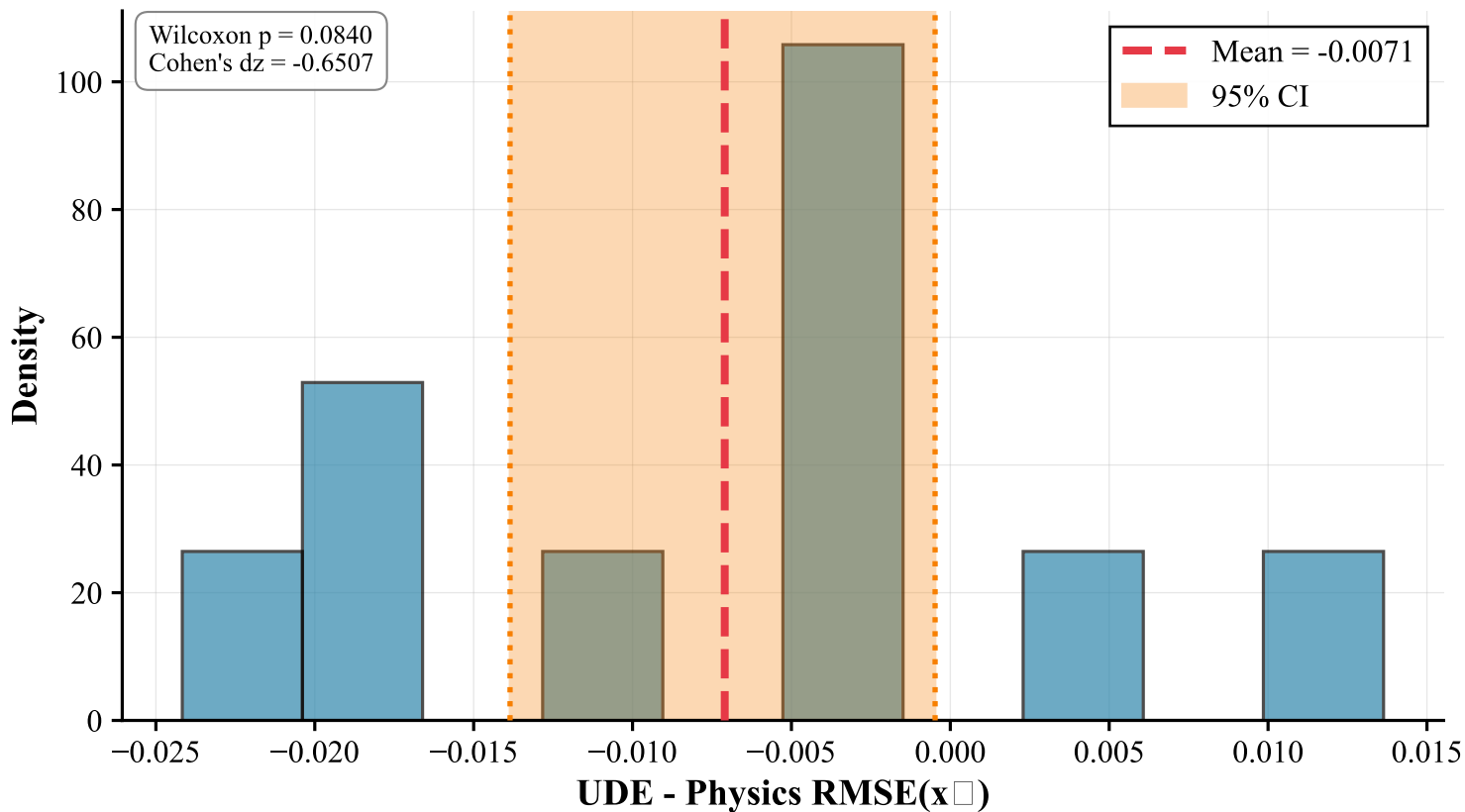


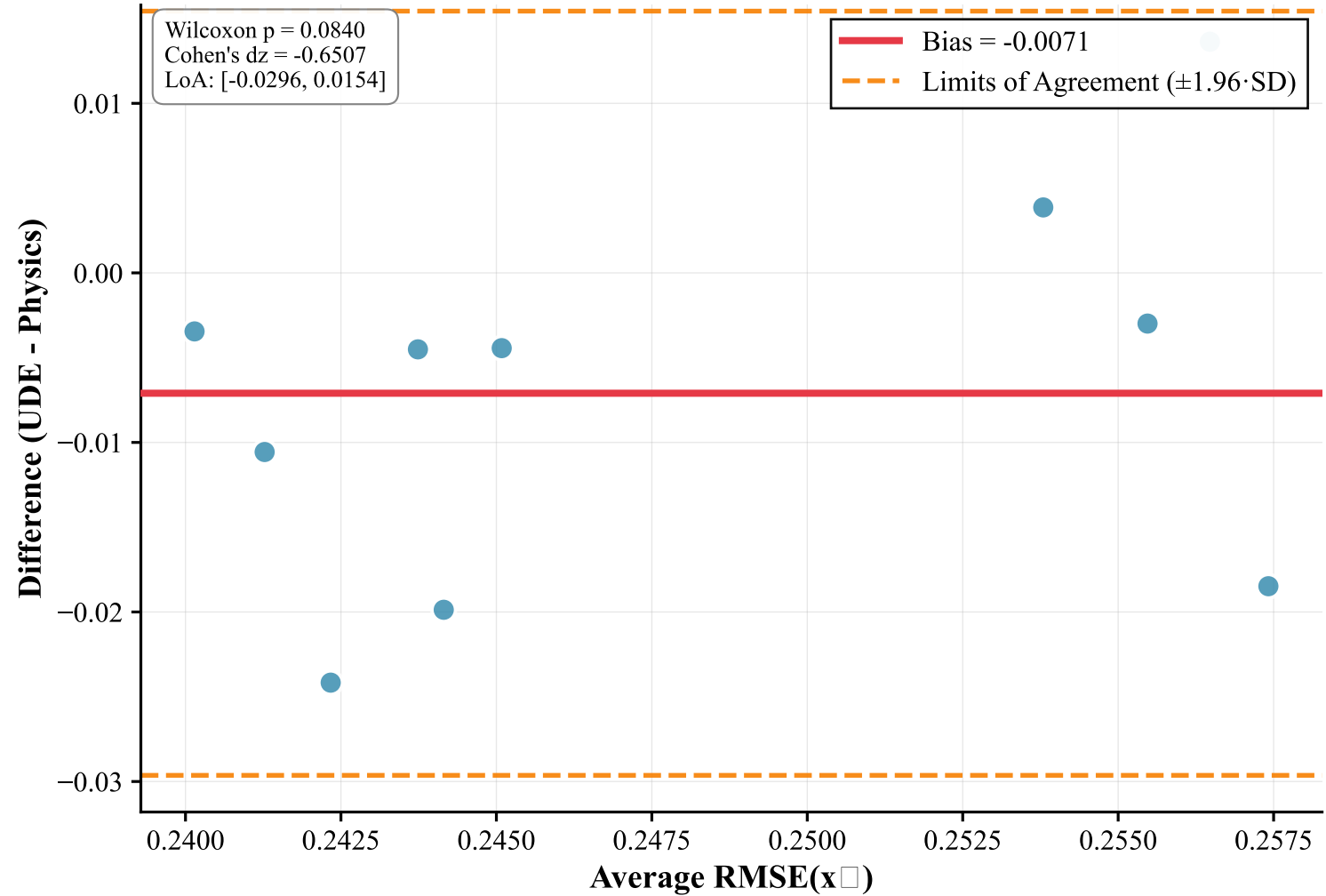
UDE vs Physics RMSE Comparison



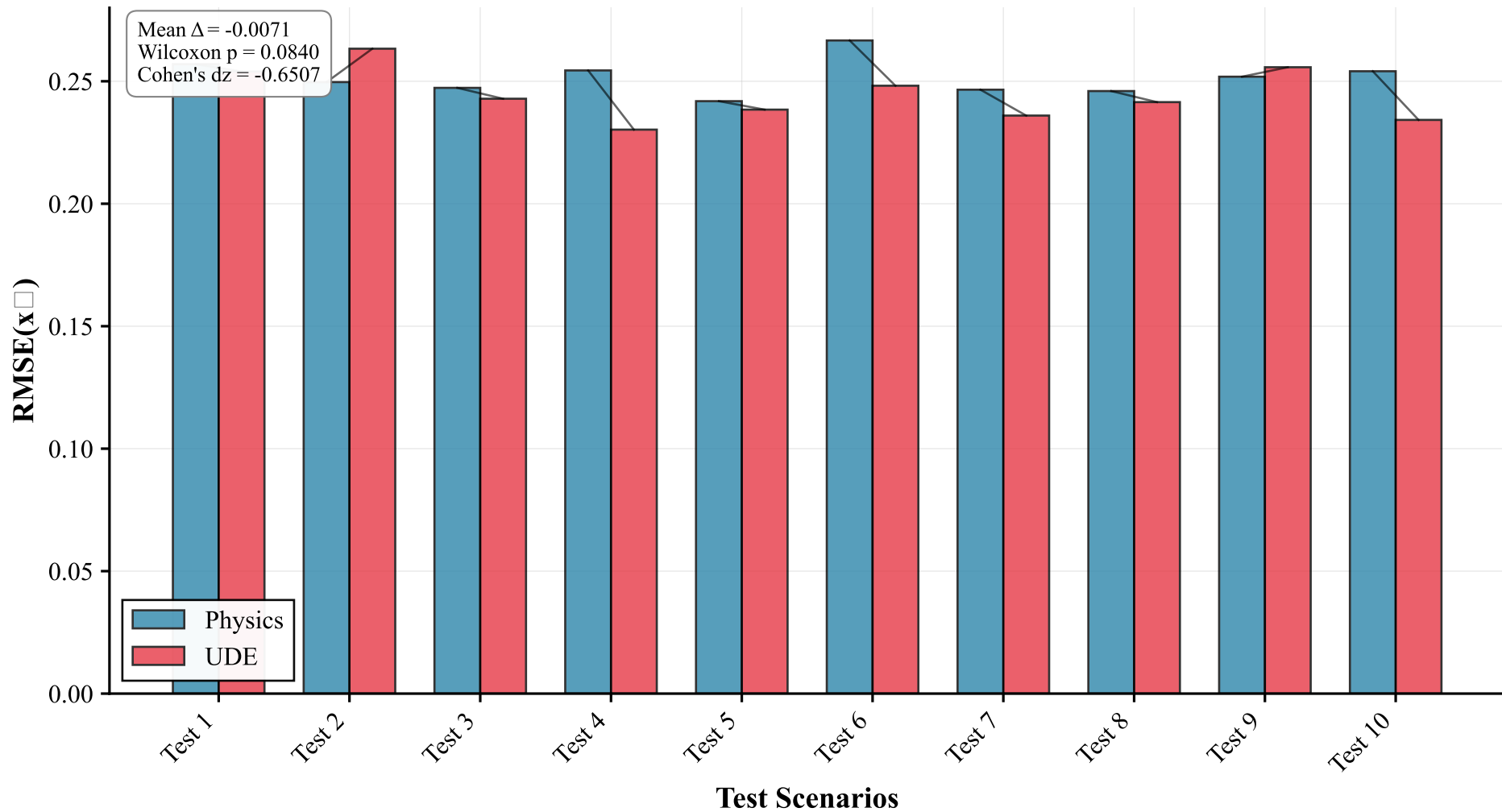
Distribution of Performance Differences



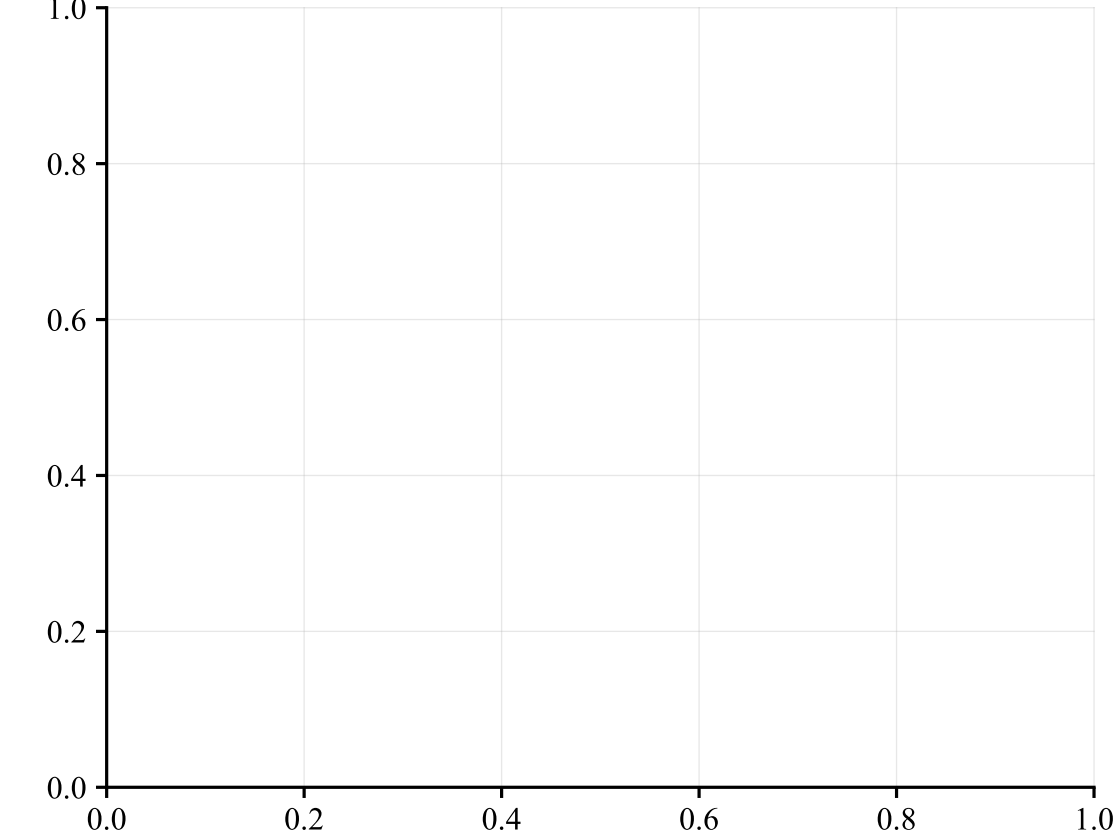
Bland-Altman Analysis of Agreement



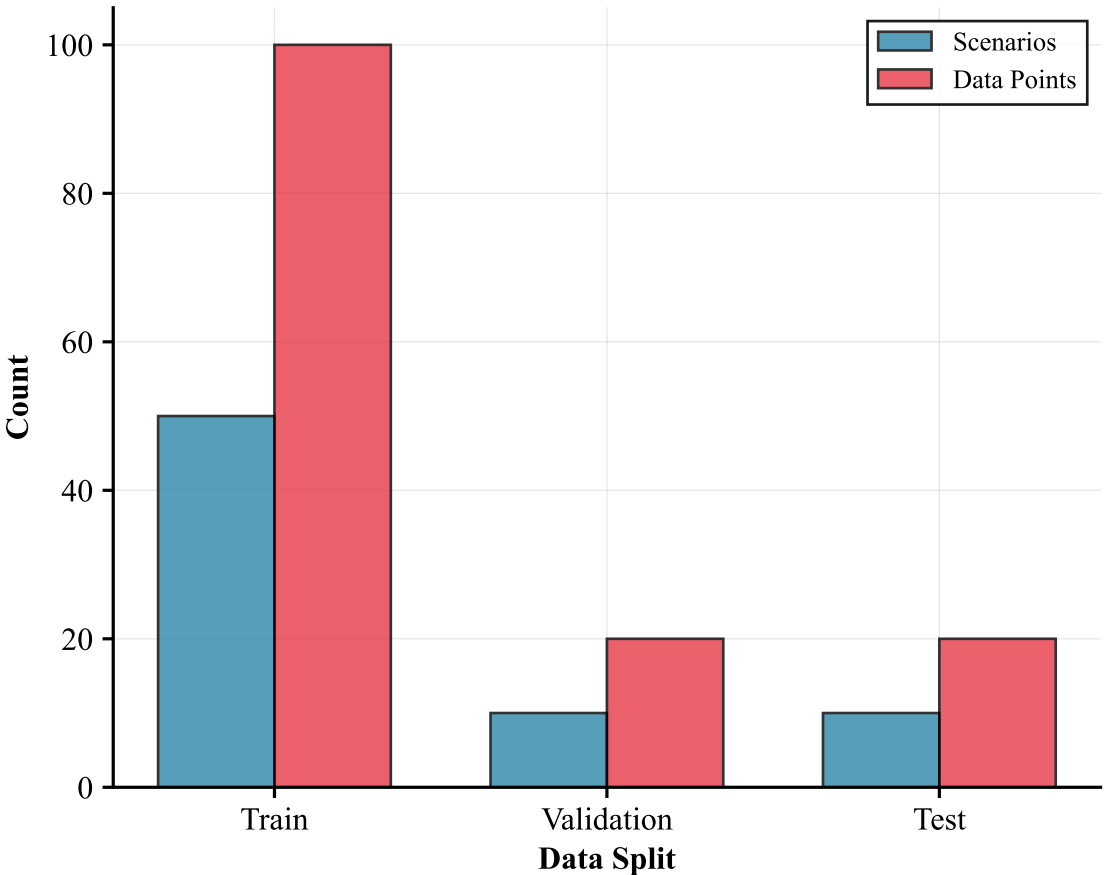
Paired Performance Comparison by Scenario



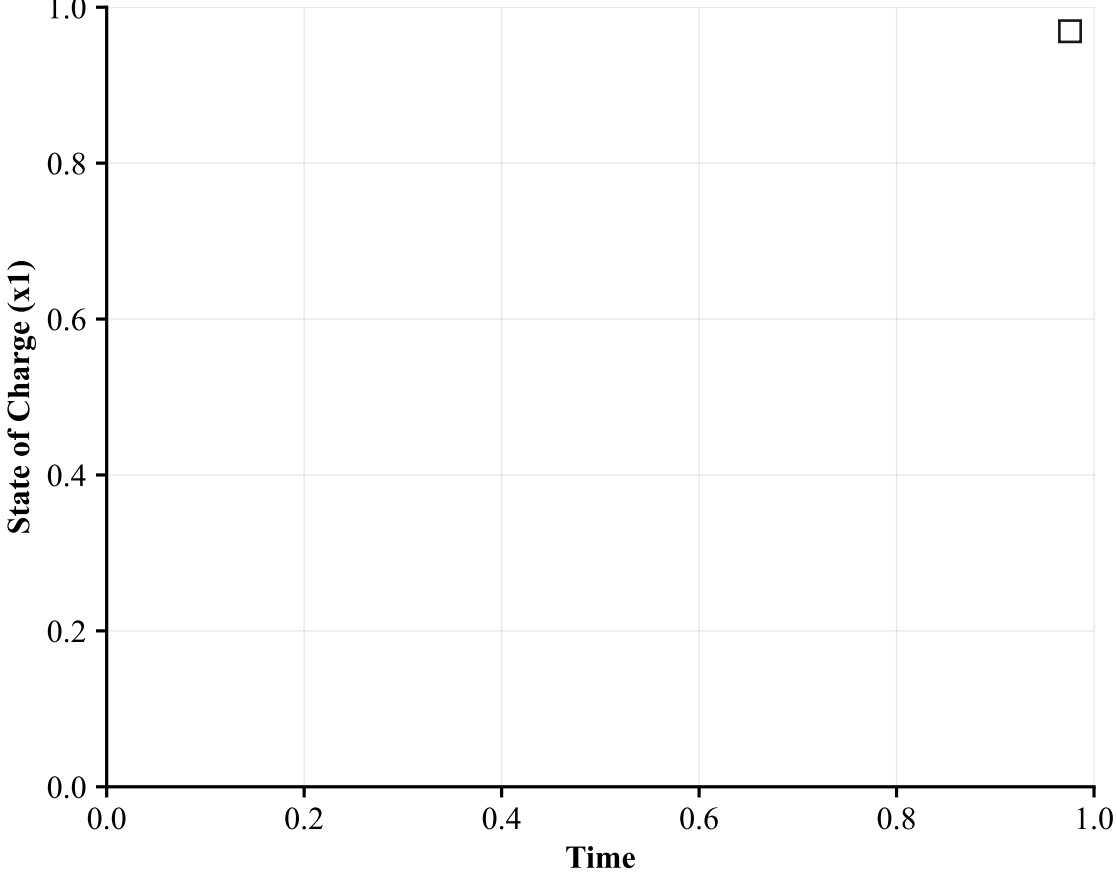
(a) Training Scenarios Parameter Distribution



(b) Train/Validation/Test Split



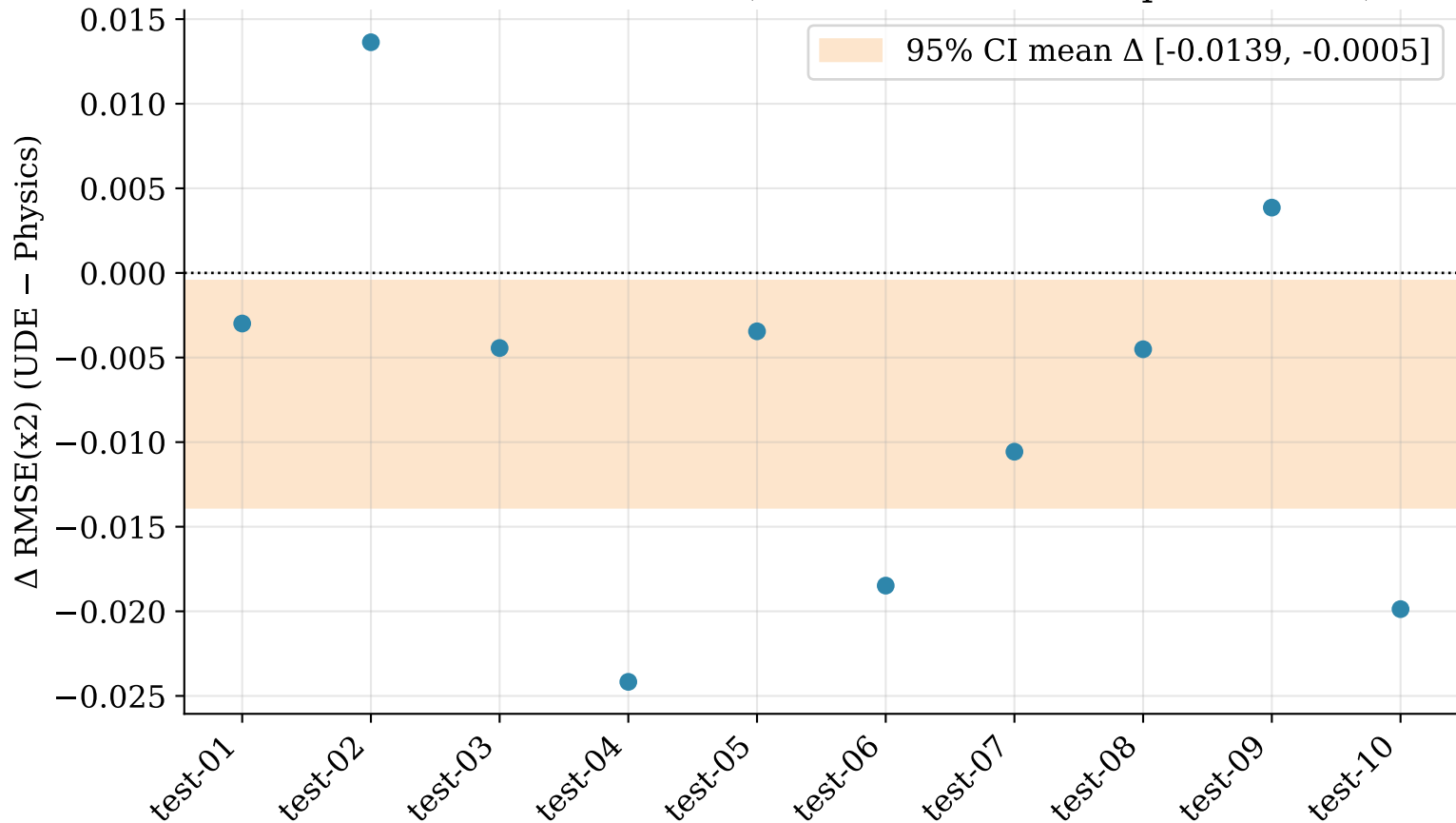
(c) Sample Training Trajectories



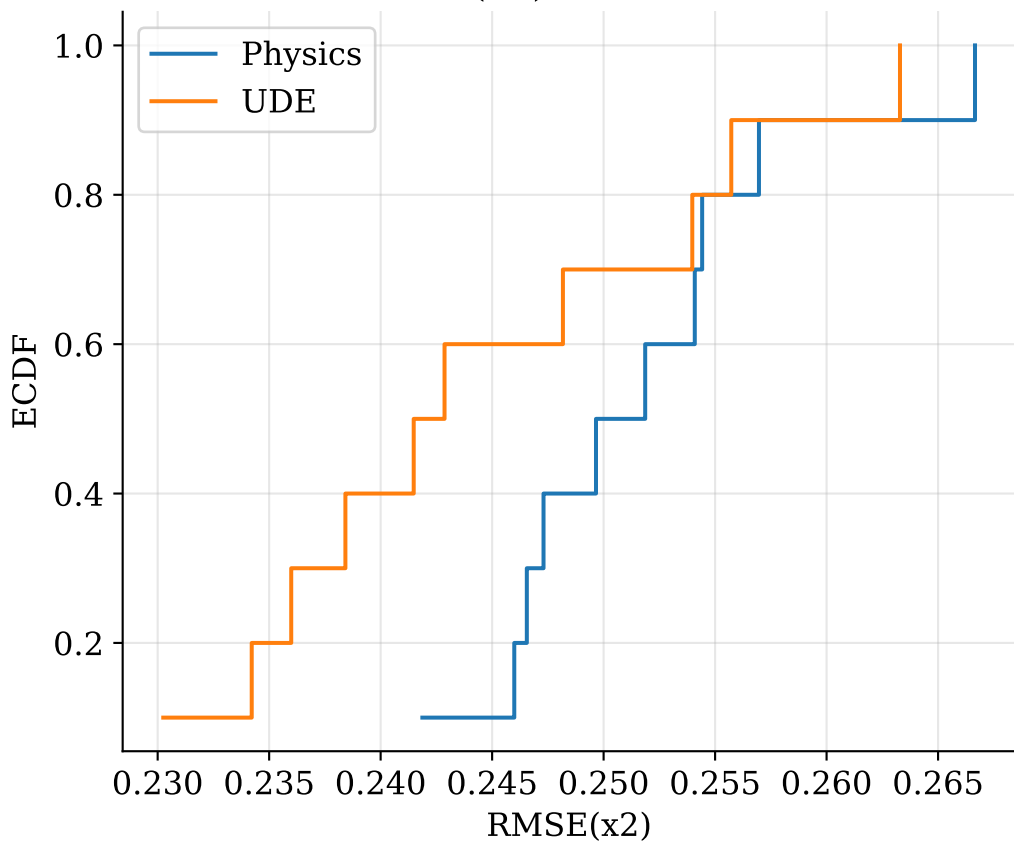
(d) Dataset Summary Statistics

Metric	Value
Total Scenarios	50
Total Data Points	100

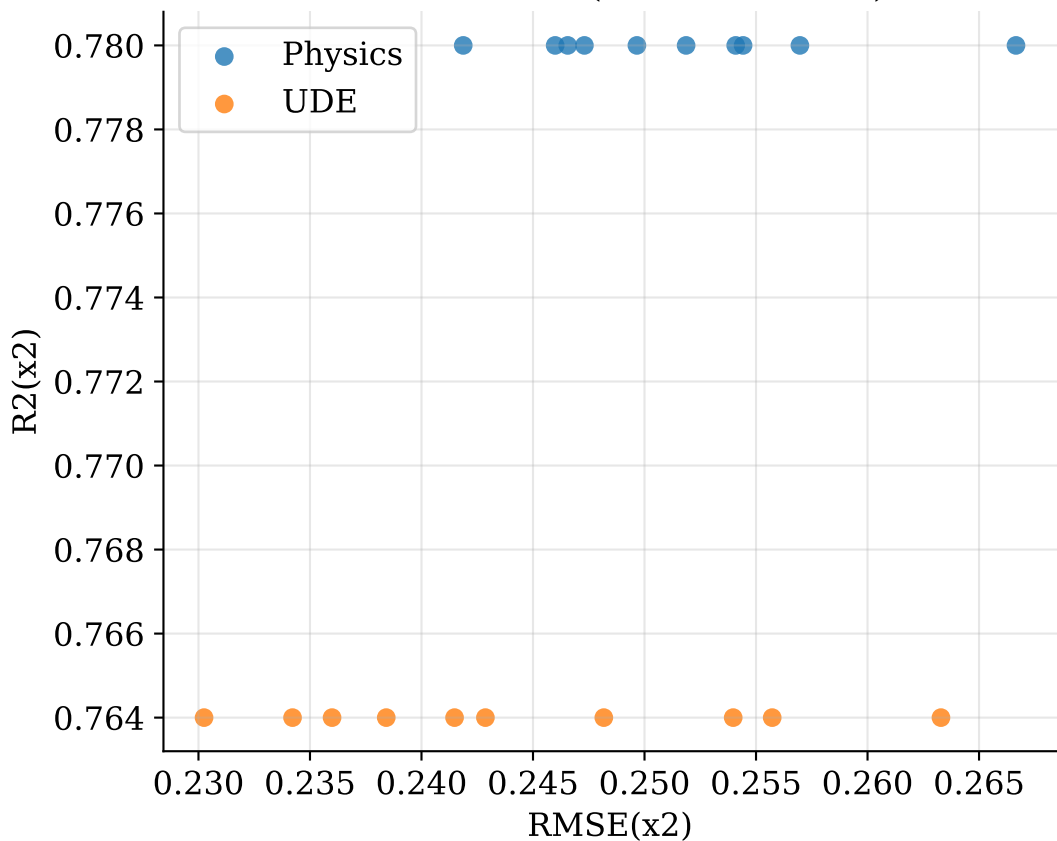
Per-Scenario Differences (with overall bootstrap CI shaded)



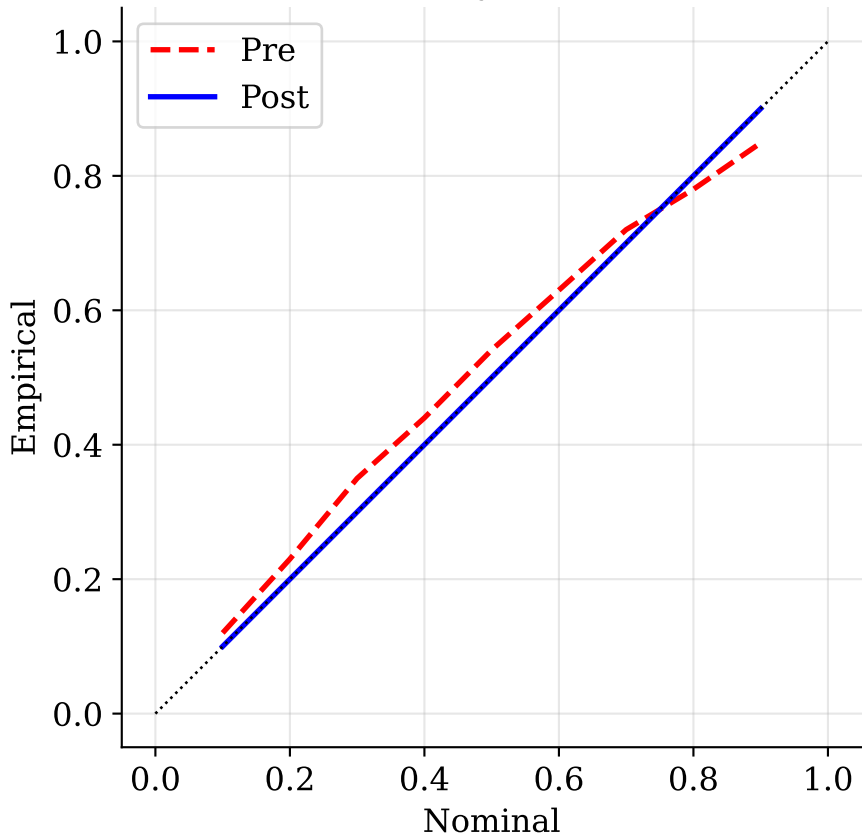
ECDF of RMSE(x2) across test scenarios



RMSE vs R2 (test scenarios)

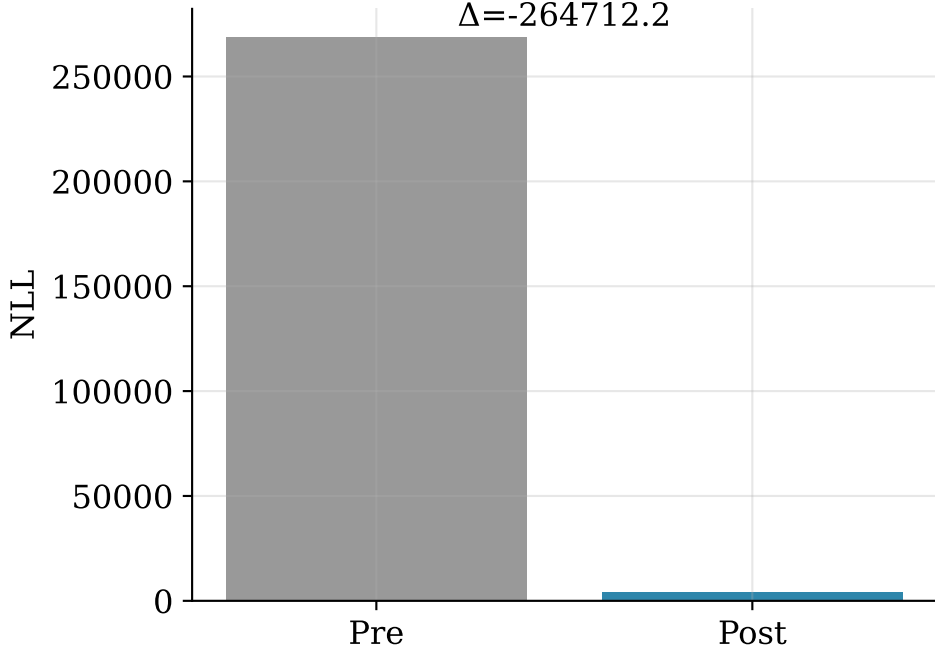


BNODE Reliability (50% illustrative)

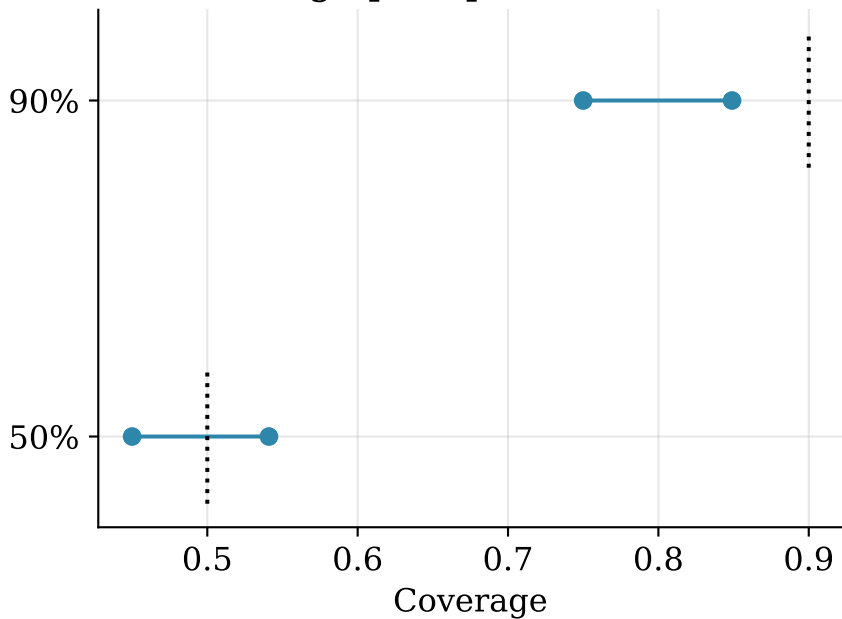


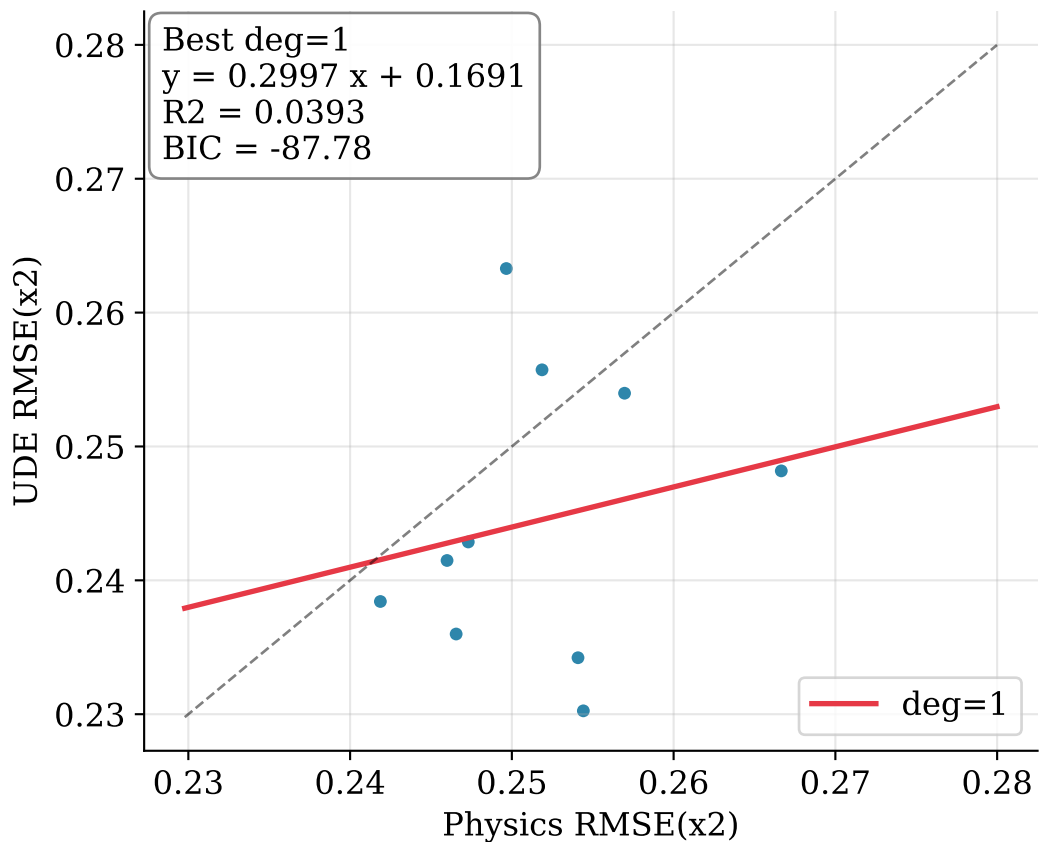
BNODE NLL Pre vs Post Calibration

$\Delta = -264712.2$

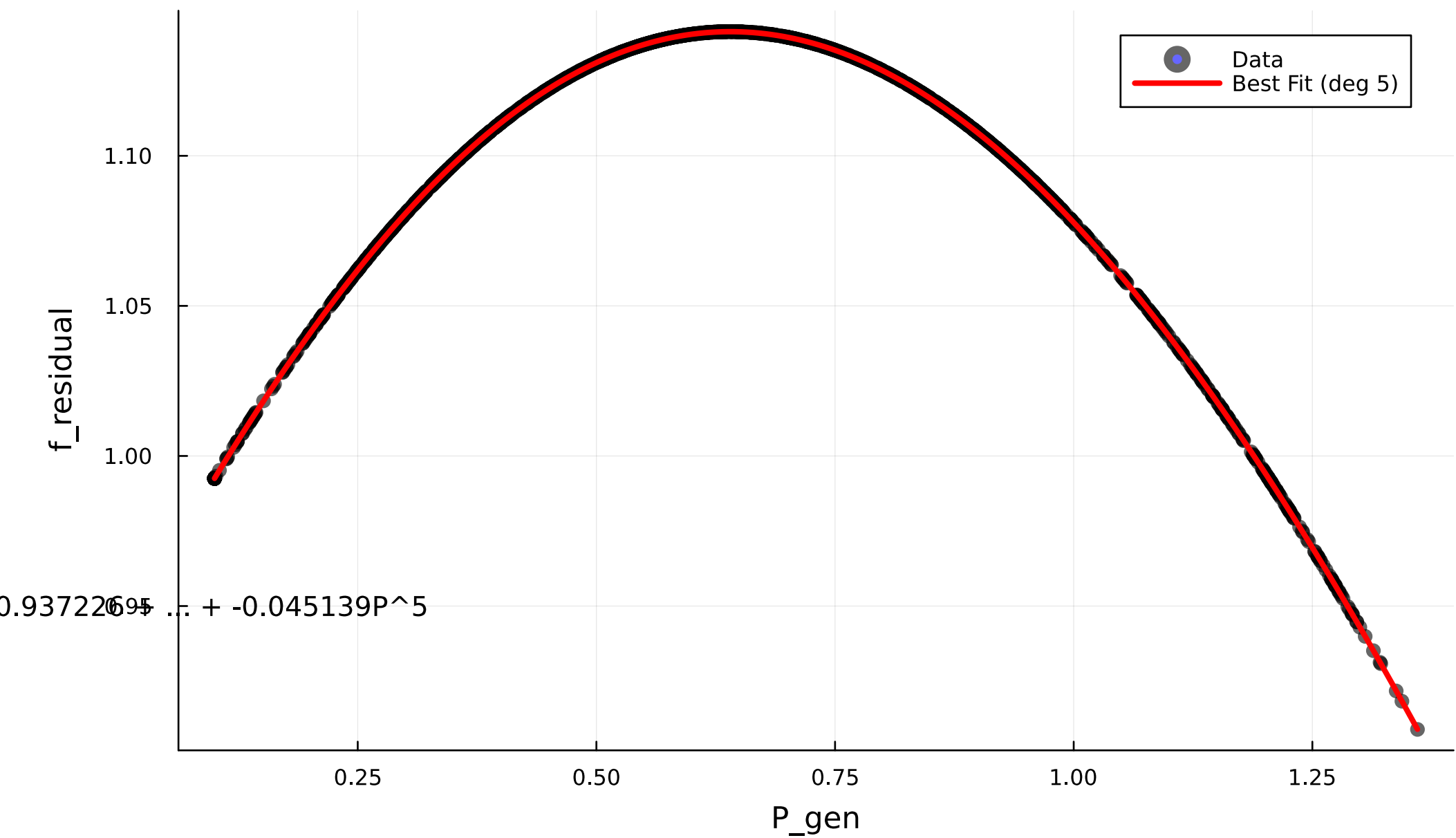


Coverage pre→post vs nominal





True UDE Residual Symbolic Regression



Residuals vs P_{gen}

