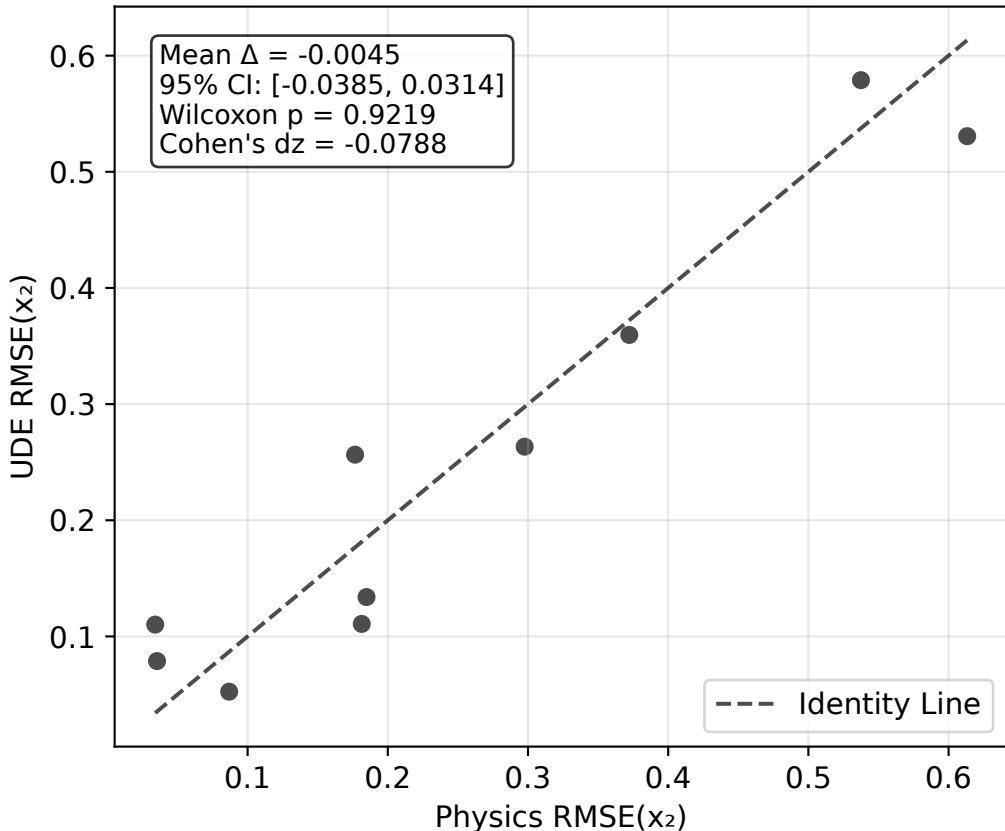
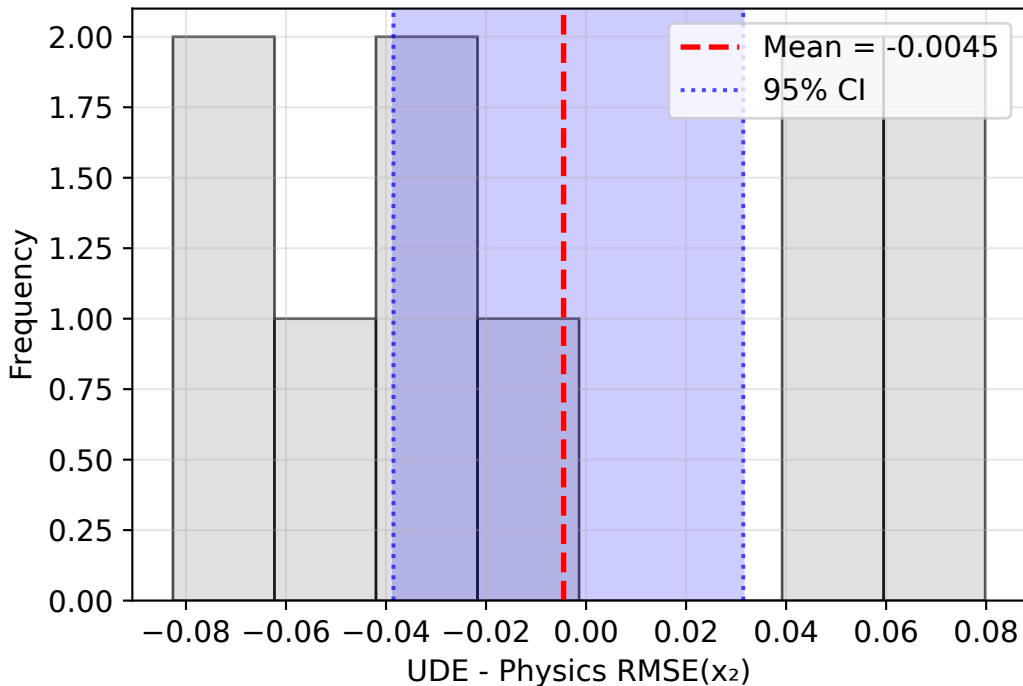


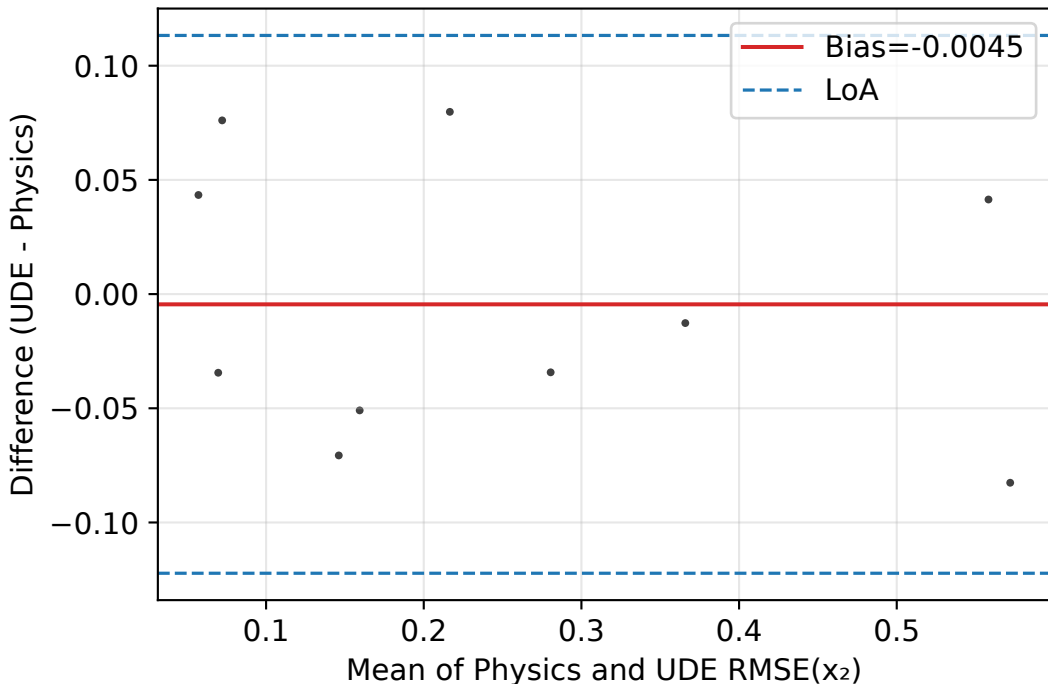
UDE vs Physics Performance Comparison



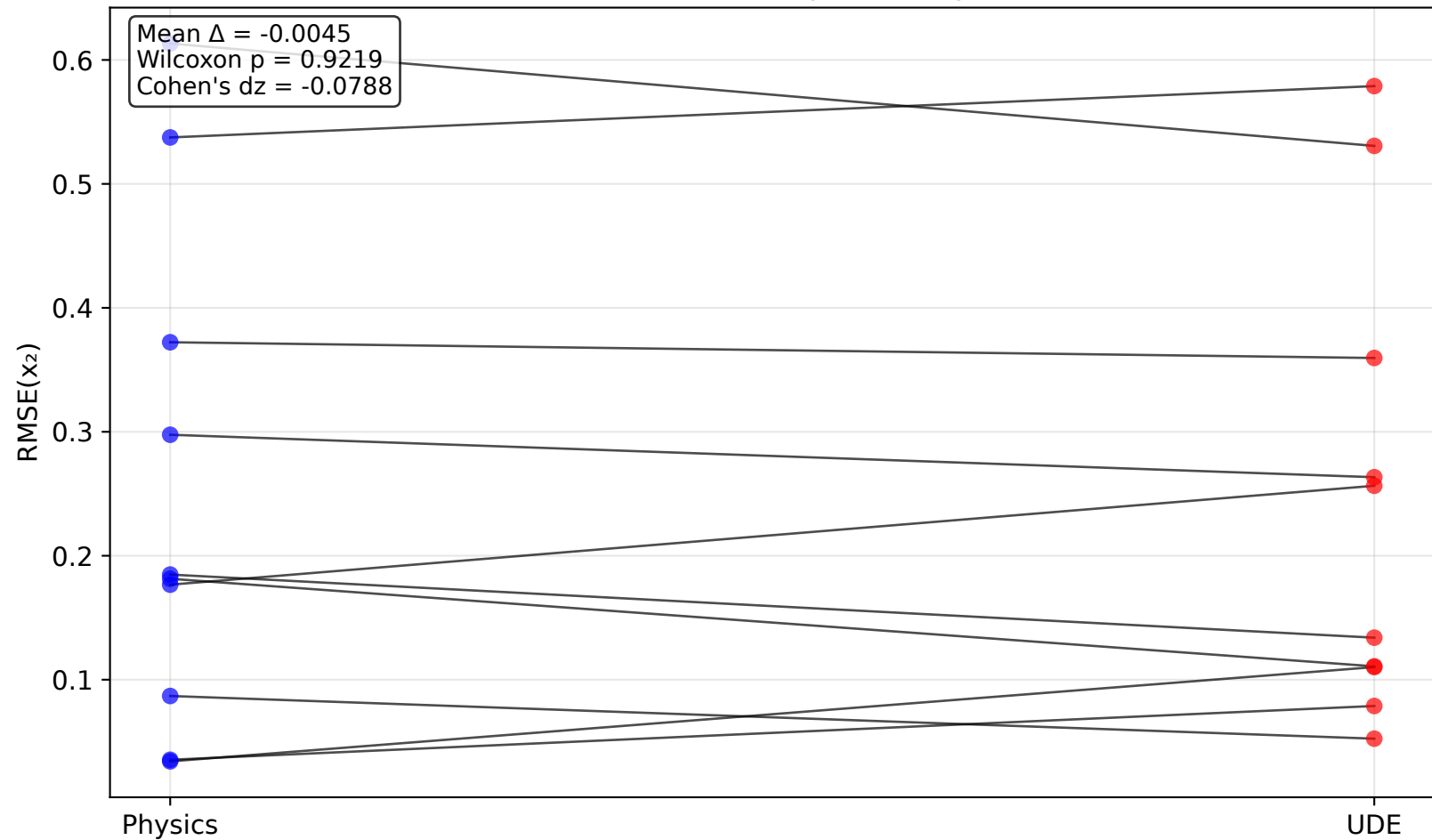
Distribution of Performance Differences



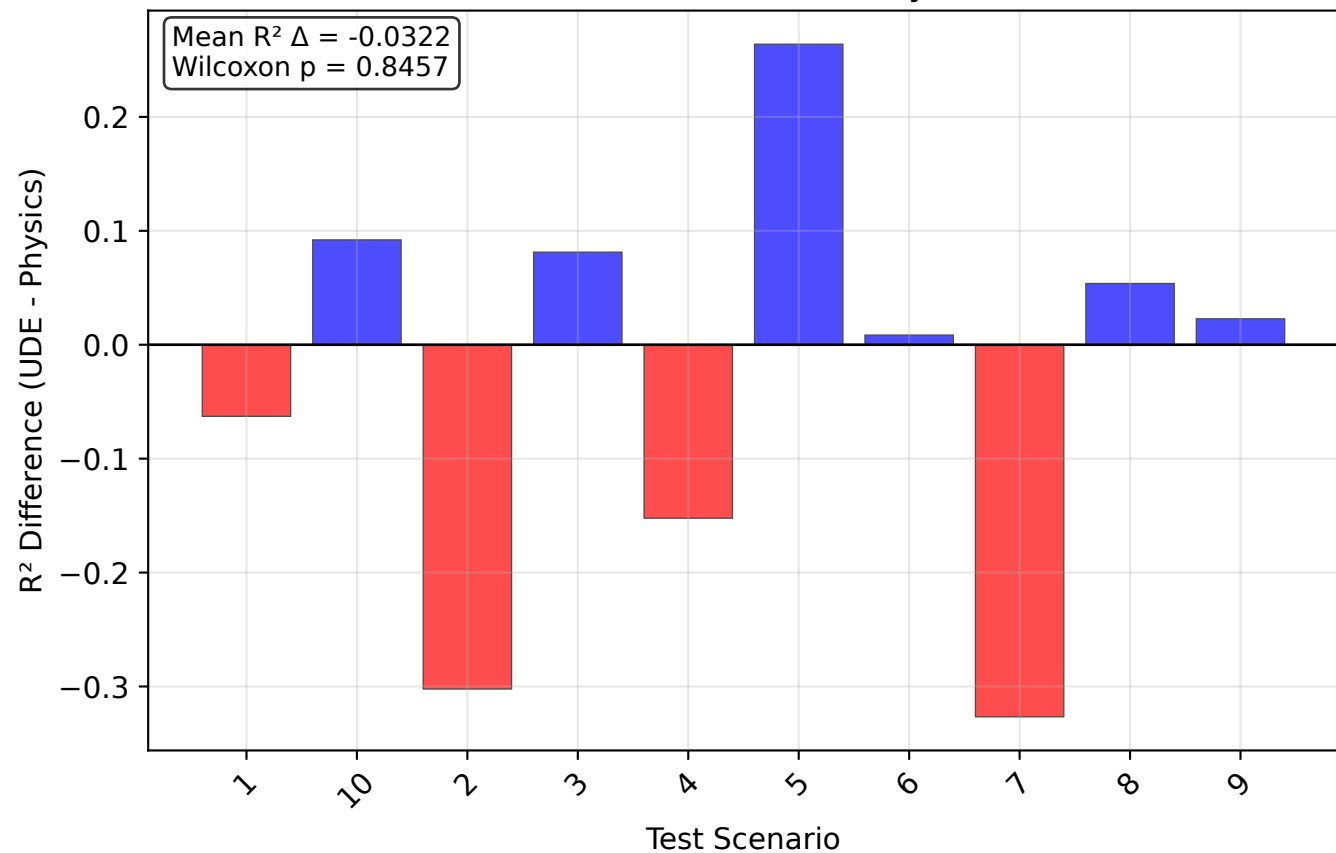
Bland-Altman (RMSE x_2)



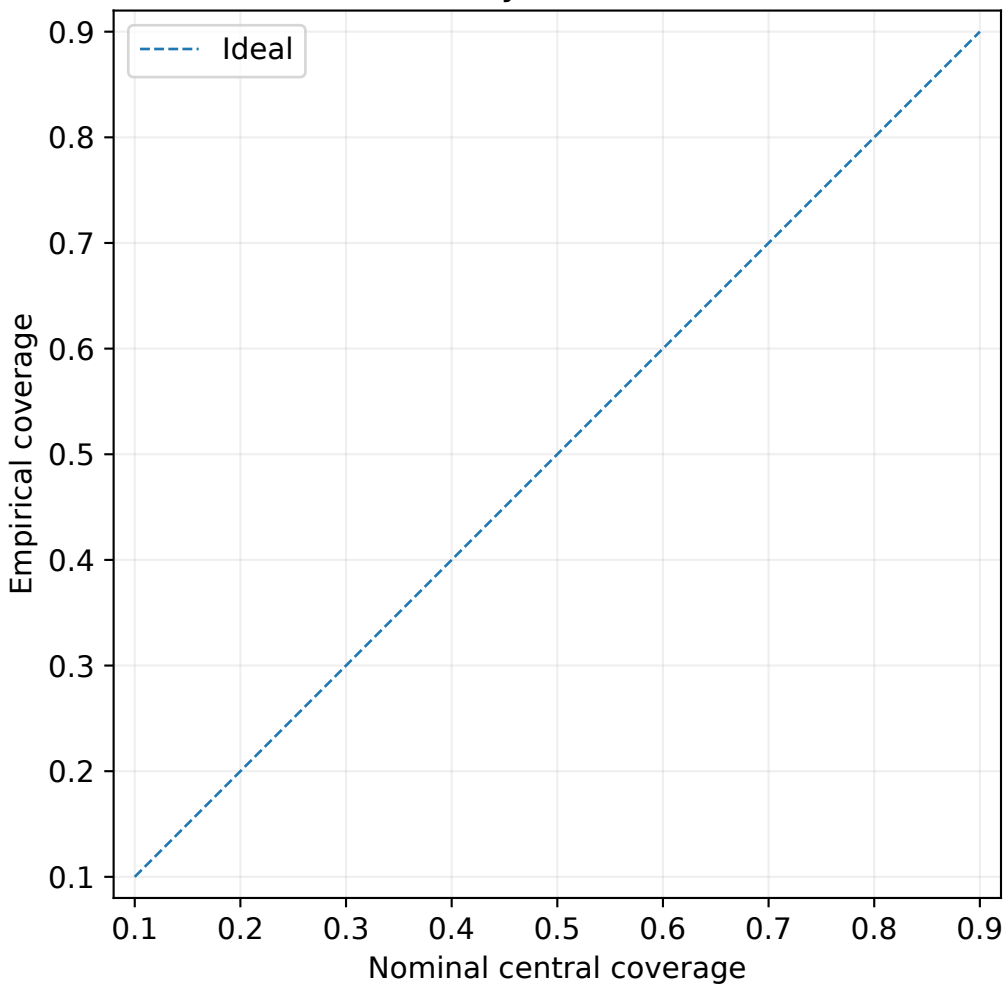
Paired Performance Comparison by Scenario



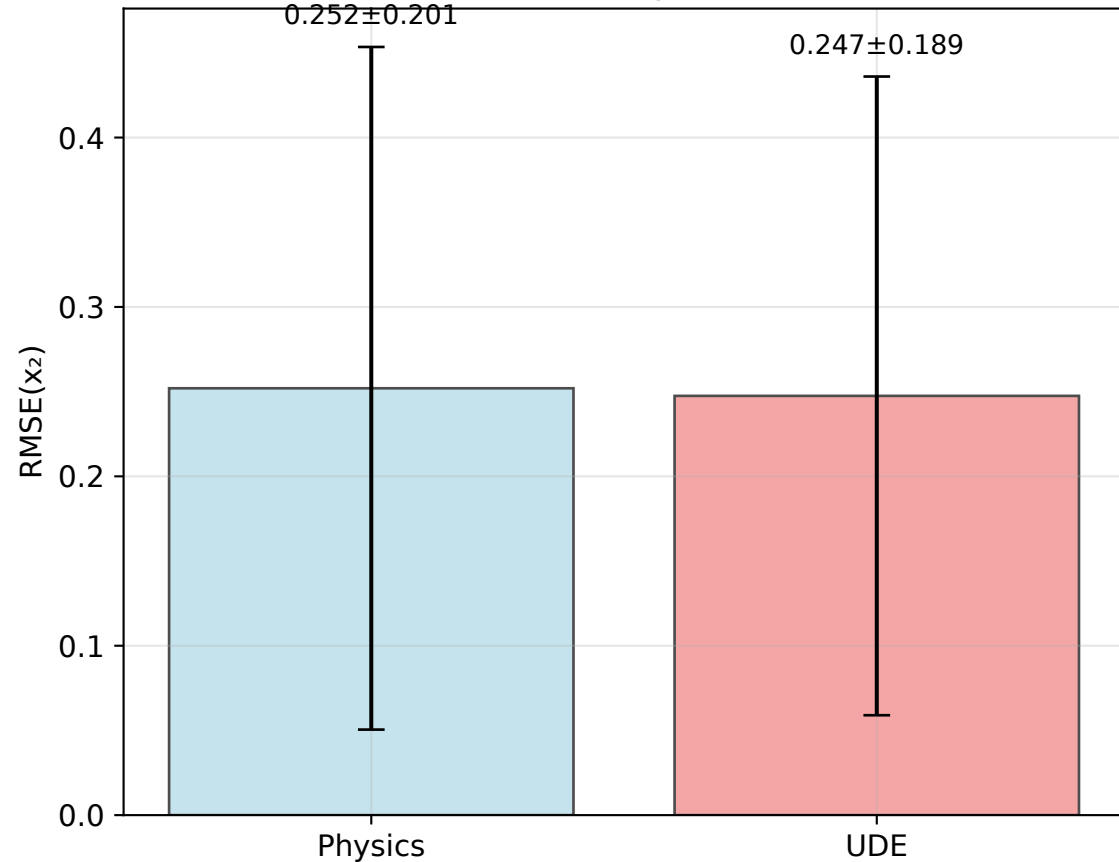
R² Performance Differences by Scenario



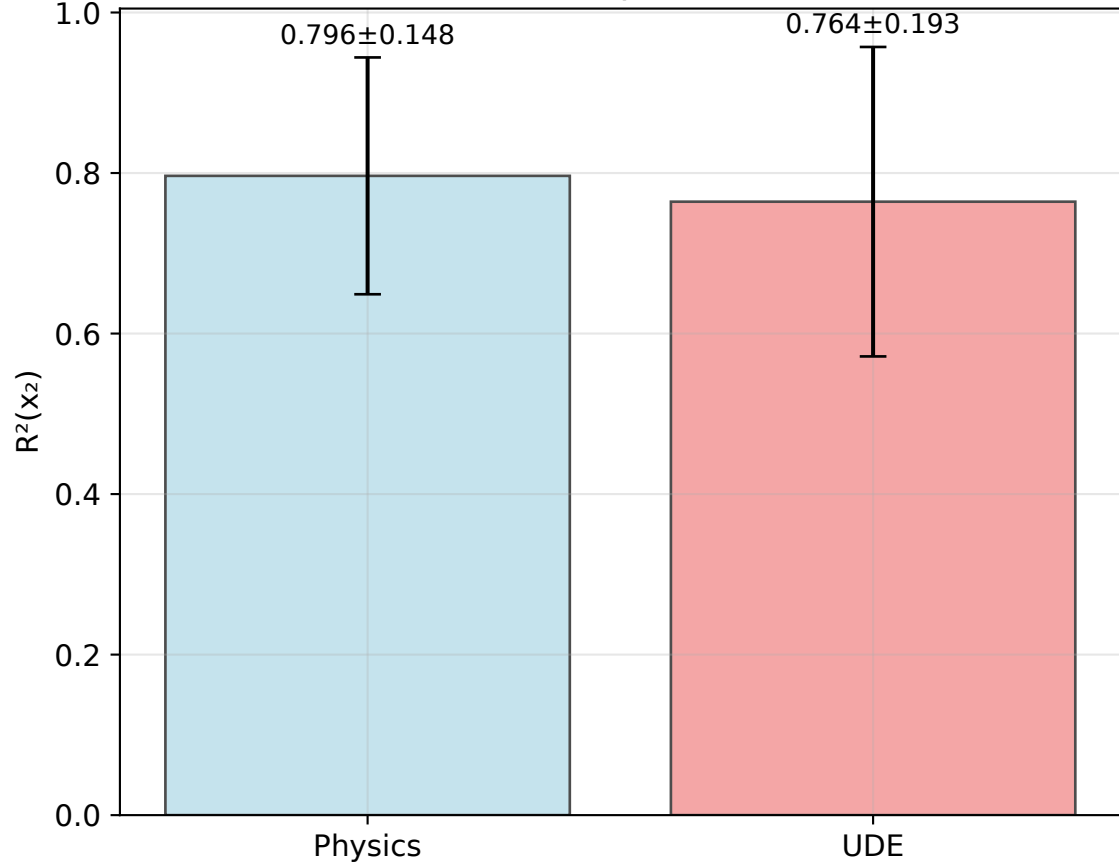
Schematic — Reliability (could not decode artifact)



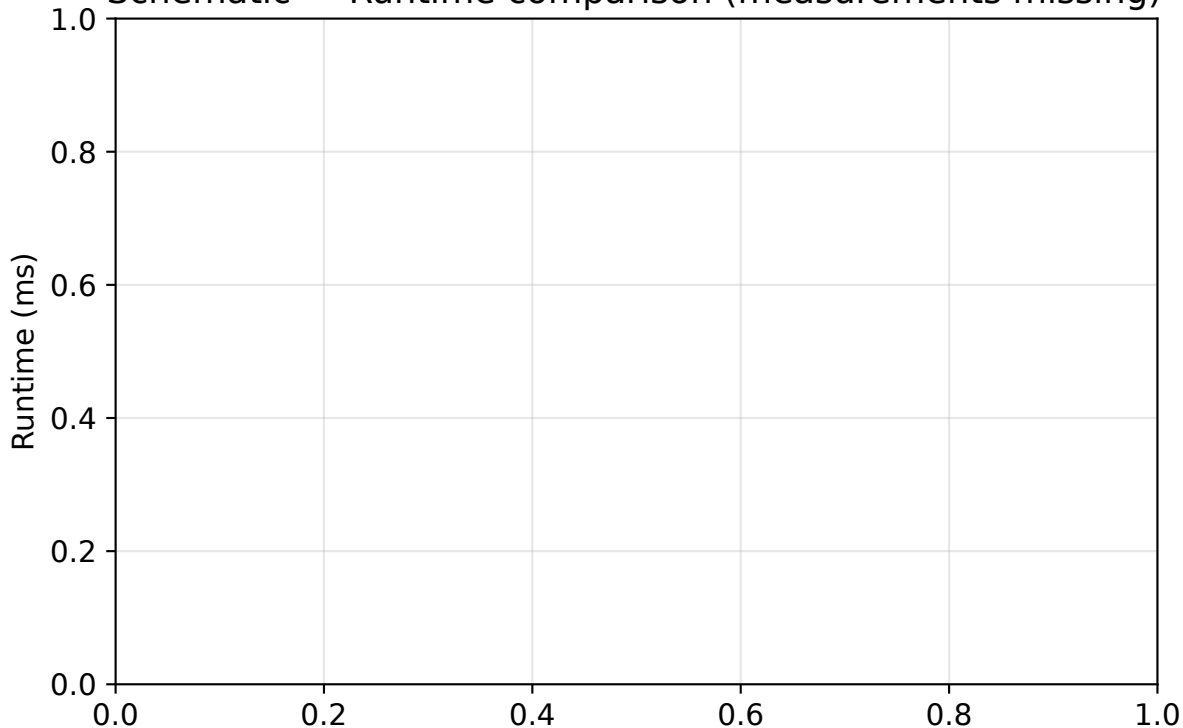
RMSE Comparison

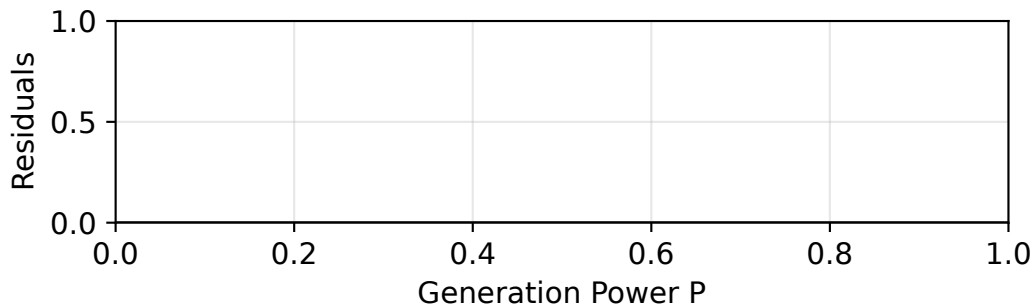
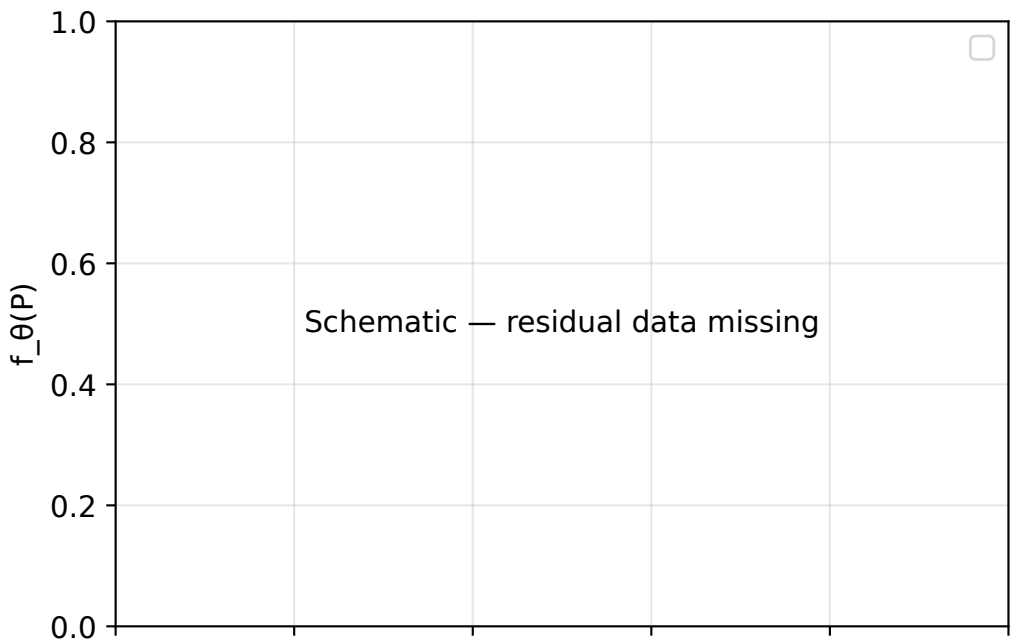


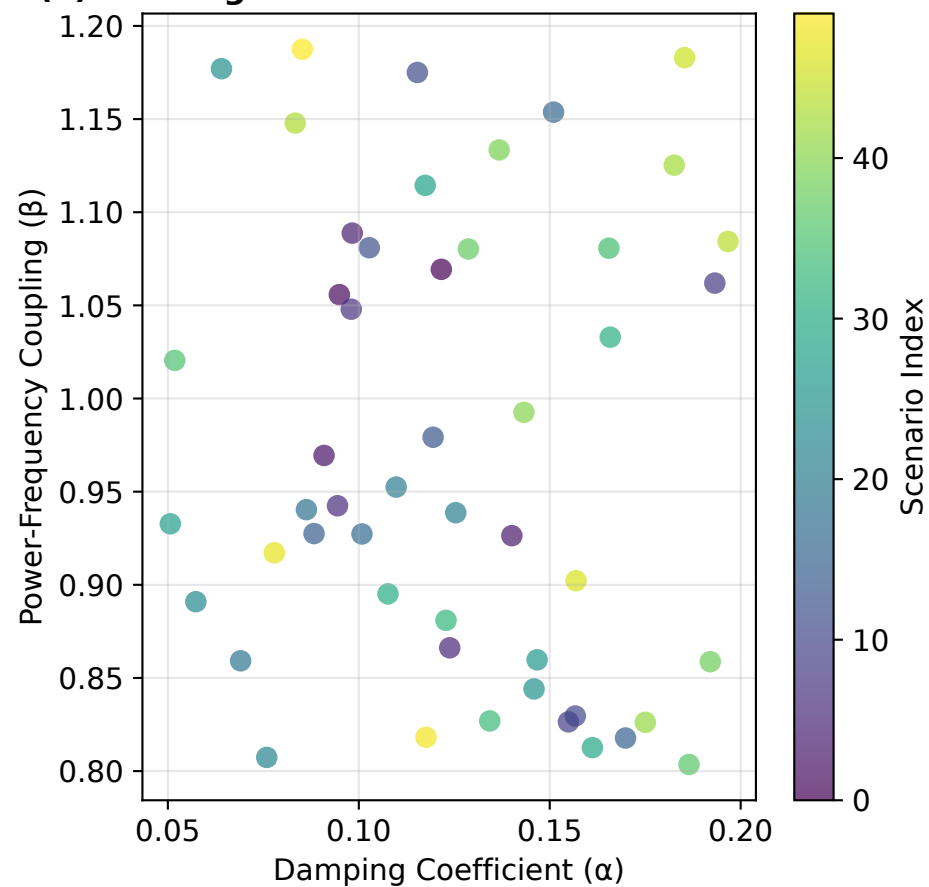
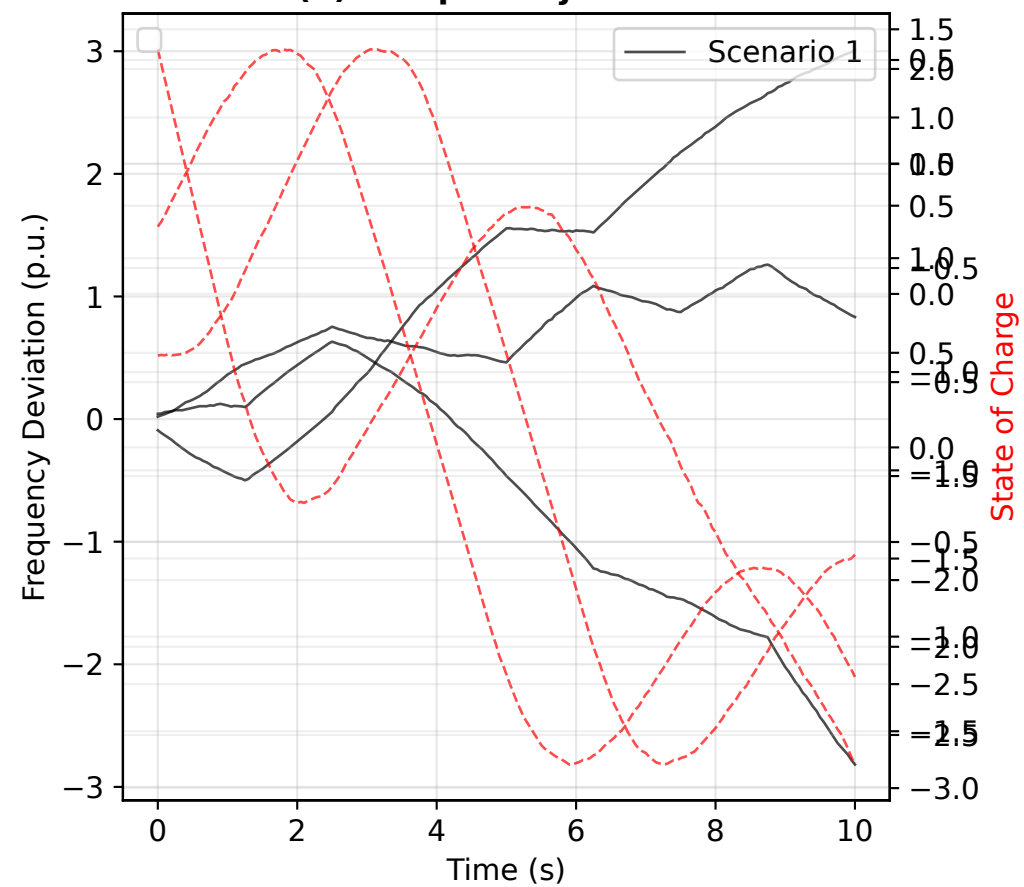
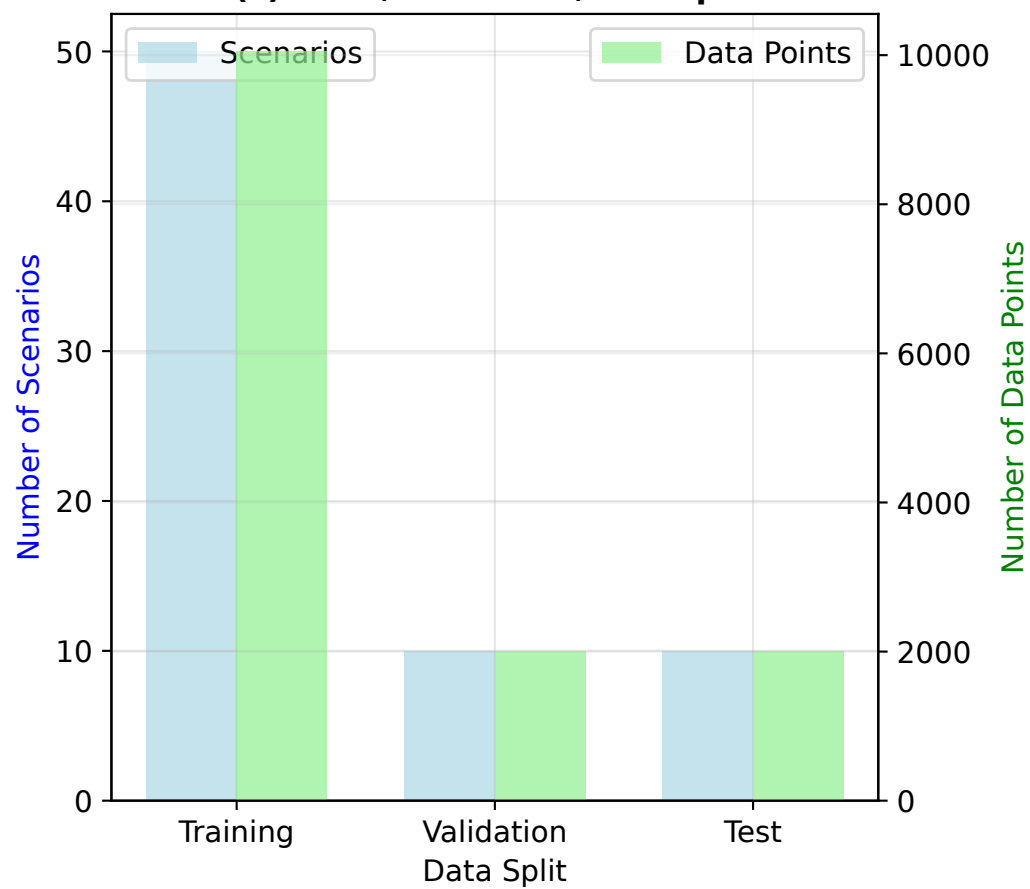
R² Comparison



Schematic — Runtime comparison (measurements missing)





(a) Training Scenarios Parameter Distribution**(b) Sample Trajectories****(c) Train/Validation/Test Split****(d) Parameter Ranges**