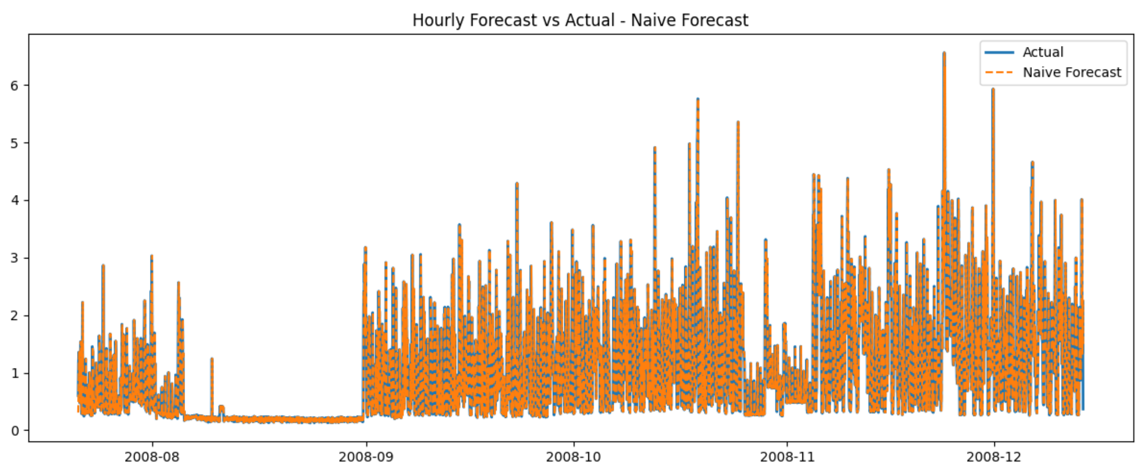
Results & Discussion — Model Comparison

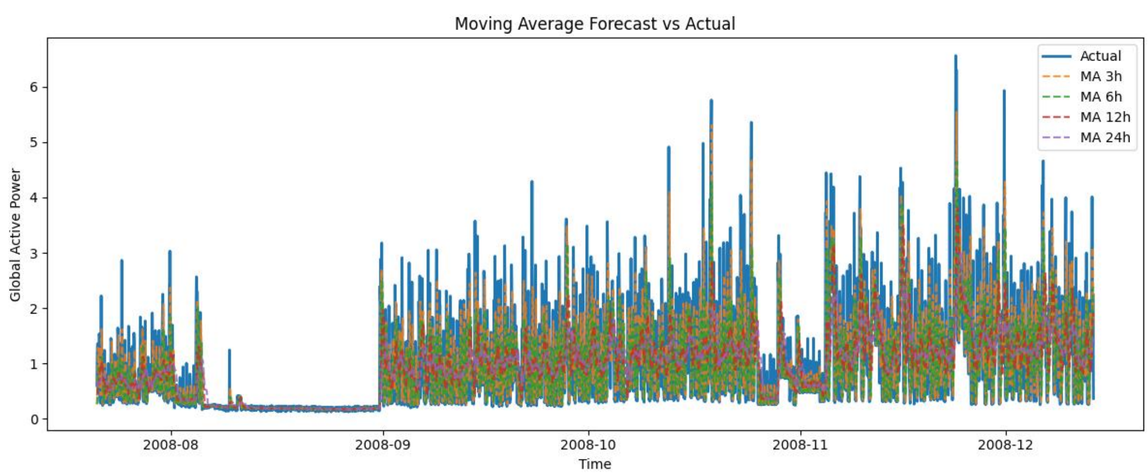
Overview of Experiments

All models were evaluated on an hourly-aggregated Global Active Power series using a chronological 80/20 split. Metrics included MAE, RMSE, MAPE, and sMAPE. Each model predicted the same unseen test window to ensure a fair and controlled comparison across methods.

Baseline Performance

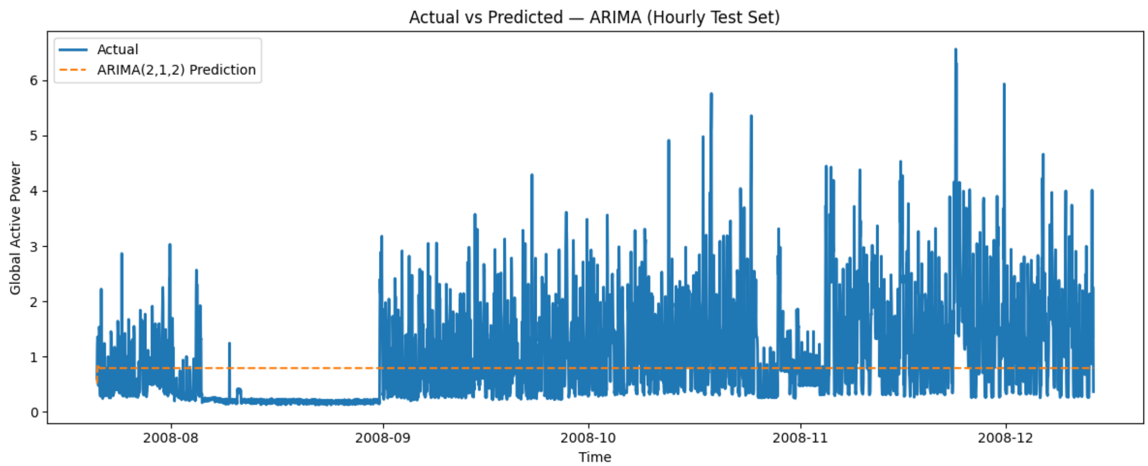
The naive persistence model achieved strong performance, confirming high hour-to-hour autocorrelation in residential demand. Moving-average baselines produced higher error due to oversmoothing, which removed informative short-term variation such as appliance activation events.

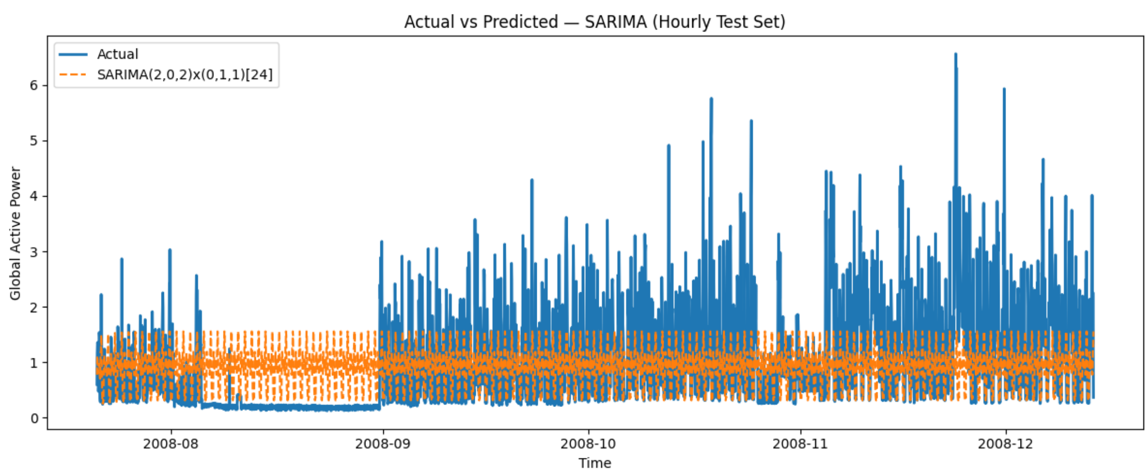




Statistical Models

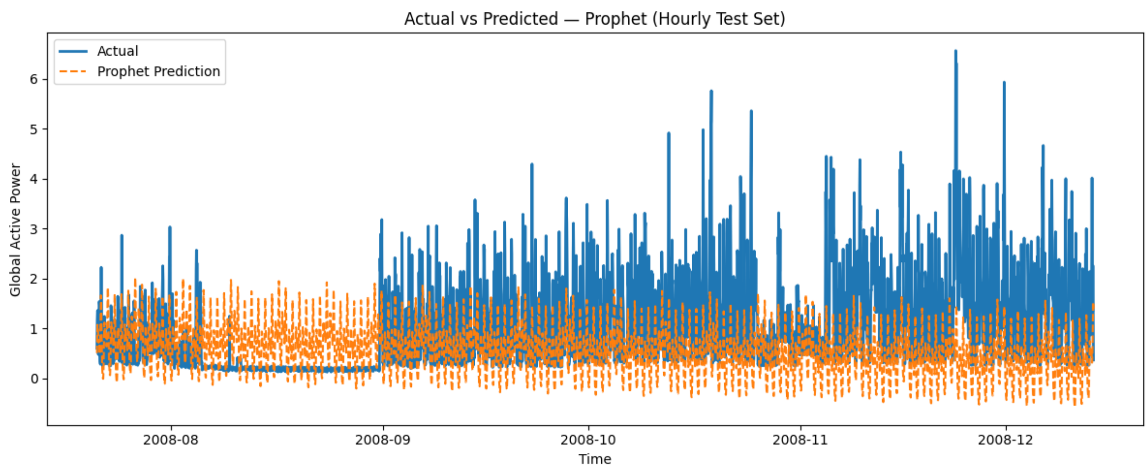
ARIMA underperformed relative to persistence, reflecting its difficulty modeling nonlinear and volatile demand. SARIMA improved upon ARIMA by incorporating daily seasonality, but the gains were insufficient to surpass the baseline, indicating that seasonal structure is present but not dominant in short-term behavior.





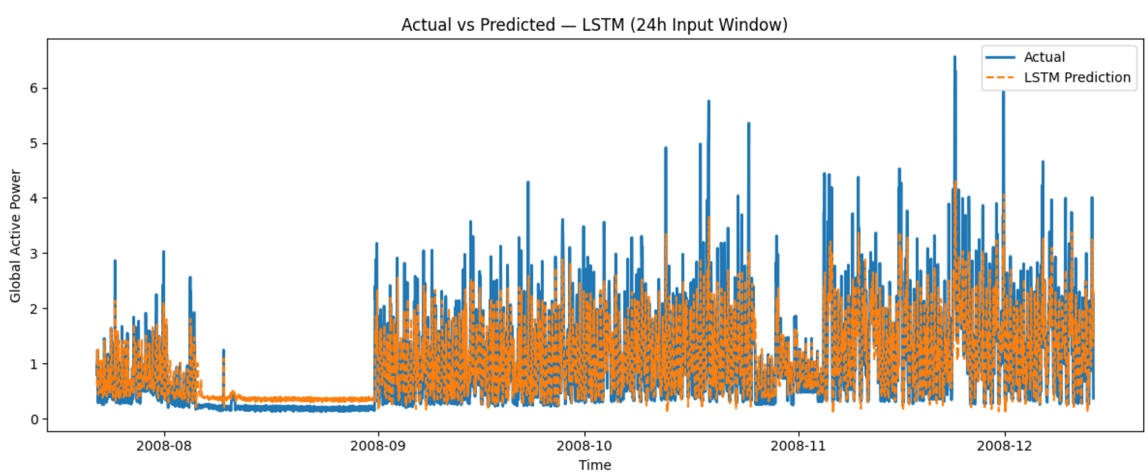
Prophet Model

Prophet exhibited the weakest performance, consistent with prior work showing that trend-decomposition techniques oversmooth noisy residential load and fail to adapt to abrupt fluctuations caused by appliance usage.



LSTM Model

The LSTM model with a 24-hour sliding window achieved the lowest MAE and RMSE and the best sMAPE values, demonstrating the ability to capture nonlinear temporal dependencies and intra-day variations. Improvements were incremental yet meaningful over the naive baseline.



Interpretation

Results suggest that short-horizon electricity consumption is largely persistence-dominated, but deep learning offers additional predictive capability where nonlinear structure is present. Model choice should reflect noise level, time-scale, and behavioral characteristics of the load.