Literature Review — Electricity Consumption Forecasting

Introduction

Short-term electricity load forecasting (STLF) has become a cornerstone of modern energy management, particularly within the frameworks of smart-grid optimization and demand-side response. At the household level, however, the "signal-to-noise" ratio is notoriously low. Unlike aggregate grid data, residential consumption is highly stochastic, driven by individual behavioral routines and the intermittent activation of high-power appliances. Literature suggests that these series are characterized by strong short-term autocorrelation but lack the stable, monotonic trends found in industrial sectors. Consequently, the challenge lies in selecting a model that balances sensitivity to local spikes with the stability required for broader temporal patterns. This review evaluates four dominant methodologies: persistence baselines, classical statistical frameworks, trend-decomposition models, and deep learning architectures.

Baseline & Persistence Forecasting

In the context of STLF, the "Persistence" or Naive model—which assumes the next value will equal the current one ($y_{t+1} = y_t$)—is frequently cited as a surprisingly formidable opponent. Because electricity usage often transitions gradually between adjacent hours (e.g., a sustained heating cycle), persistence captures the immediate state of the system effectively. While moving-average filters can smooth out the "jitter" of small appliance usage, they often introduce lag and inadvertently erase the sharp "ramps" of major load events. Most researchers agree that any "advanced" model failing to outperform persistence is functionally obsolete for real-time applications.

Classical Statistical Models (ARIMA & SARIMA)

The Autoregressive Integrated Moving Average (ARIMA) framework remains the benchmark for linear stochastic modeling. By regressing a variable on its own lagged values and errors, ARIMA excels at capturing the momentum of a series. However, its reliance on stationarity often limits its utility in residential settings where demand is non-linear and volatile.

To address periodicity, Seasonal ARIMA (SARIMA) introduces seasonal terms to account for daily (24-hour) or weekly cycles. While SARIMA is more robust in identifying recurring "morning peaks" or "evening surges," its performance degrades when human behavior deviates from the mean—such as on holidays or during irregular work-from-home schedules. Essentially, these models are "rigid"; they struggle to adapt when the underlying "seasonal" pattern shifts abruptly.

Trend-Decomposition: The Prophet Approach

Prophet, developed by Meta, treats forecasting as a curve-fitting exercise rather than a traditional time-series analysis. It decomposes the signal into trend, seasonality, and holidays. While highly effective for business metrics with clear growth trajectories, the literature highlights a specific weakness in residential energy: **oversmoothing**. Because Prophet assumes a smooth, piecewise linear trend, it often fails to track the erratic, non-monotonic "burstiness" of household data. It is generally viewed as a "macro" tool that struggles with the "micro" volatility of individual meters.

Deep Learning: LSTM Architectures

Long Short-Term Memory (LSTM) networks have redefined the state-of-the-art by addressing the "memory" limitations of traditional RNNs. LSTMs utilize a gating mechanism to decide which past information (e.g., a peak from two hours ago) is relevant for the current prediction. Studies consistently show that LSTMs outperform classical models in high-variance scenarios because they can learn complex, non-linear mappings. However, this accuracy comes at the cost of interpretability and high computational overhead. Recent findings suggest that while LSTMs offer lower error metrics, the margin of improvement over a well-tuned SARIMA is often narrower than expected, raising questions about the trade-off between complexity and performance.

Research Gap

A persistent issue in existing literature is the lack of "apples-to-apples" comparisons. Many studies evaluate a single "hero" model against weak baselines or use disparate datasets with varying degrees of preprocessing. There is a clear need for a unified evaluation framework that subjects persistence, classical statistics, Prophet, and LSTMs to the same rigorous preprocessing and "unseen" testing conditions. This project fills that gap by providing a controlled comparative analysis to determine which architecture truly captures the nuances of residential demand.