Title: Electricity Demand Forecasting in Spain Using Hybrid Time Series and Machine Learning Models

Author: Mikhail Broomes

Abstract

This study evaluates short-term electricity demand forecasting in Spain by leveraging classical statistical models, machine learning algorithms, and hybrid approaches. Core models include SARIMAX, XGBoost, and an ARIMA-ANN hybrid. Additional benchmarks such as TBATS, NNETAR, and STL-ARIMA are explored. The methodology incorporates extensive feature engineering, SHAP-based interpretability, and rolling-origin cross-validation. Results demonstrate that the tuned XGBoost model significantly outperforms classical statistical models in RMSE and MAPE, with hybrid models offering strong complementary performance.

1. Introduction

1.1 Motivation

Spain's electricity grid, heavily reliant on renewables, presents high volatility. The nationwide blackout in March 2025 highlighted vulnerabilities in existing short-term forecasting systems.

1.2 Objectives

The objectives of this study are to

(1) compare classical, machine learning, and hybrid models,

(2) focus on 24–168 hour short-term demand forecasting,

(3) apply SHAP for explainability, scenario testing, and rolling-origin cross-validation.

2.Literature Review

Prior studies have explored time series models such as SARIMA and TBATS for electricity load forecasting (Hyndman, 2018). Zhang (2003) proposed hybrid ARIMA-ANN models to capture nonlinear residual behavior, offering a powerful alternative when classical models fall short. In the context of machine learning, Li et al. (2020) and Lundberg & Lee (2017) demonstrated the effectiveness of XGBoost and SHAP for improving both forecasting accuracy and interpretability. For model evaluation, rolling-origin cross-validation and the Diebold-Mariano test (Diebold & Mariano, 1995) have become standard practices in comparing time series prediction performance.

3.Data & Preprocessing

3.1 Data Sources:

The dataset used in this study combines electricity load and weather data. The electricity load data was sourced from ENTSO-E and includes hourly records spanning from 2015 to 2019. Complementary meteorological variables, including temperature, humidity, wind speed, and solar radiation, were obtained from publicly available weather sources. These variables are known to have a significant influence on electricity demand.

3.2 Preprocessing:

To prepare the data for modeling, missing values were imputed using time-series-specific methods such as linear interpolation and Kalman smoothing. Several temporal features were engineered to capture cyclical demand patterns. These included hour of day, day of the week, and month of the year.

Lagged values and rolling means of the target variable were created to help the models learn recent trends. In addition, some variables were normalized or transformed to stabilize variance and remove skewness. The resulting dataset was structured as a multivariate time series with over 35,000 observations and multiple engineered predictors. (2015–2019)

3.3 Feature Engineering:

This section summarizes the feature engineering strategies applied in this study, organized into three primary categories. While only a subset of these features was implemented during model development, the dataset includes a rich array of raw inputs suitable for future expansion.

Temporal Features

- *Implemented*: Hour, weekday, and month extracted using timestamp information.
- *Not implemented*: Cyclical encodings (e.g., sine and cosine transformations), public holiday indicators, or interaction terms.
- *Available*: Complete timestamp data capable of supporting more advanced temporal transformations.

Weather Features

- *Implemented*: Hourly measurements of temperature, humidity, wind speed, pressure, and cloud cover.
- *Not implemented*: Rolling averages, temperature-humidity interaction terms, and degree-day indicators.

- *Available*: High-resolution weather metrics suitable for extended modeling.

Renewable Energy Features

- *Implemented*: Hourly solar and wind generation data, including forecasted values.
- *Not implemented*: Forecast error metrics, ramp rates, or renewable penetration ratios.
- *Available*: Sufficient detail to derive volatility and renewable integration indicators.

To enhance predictive performance, three core feature engineering techniques were applied:

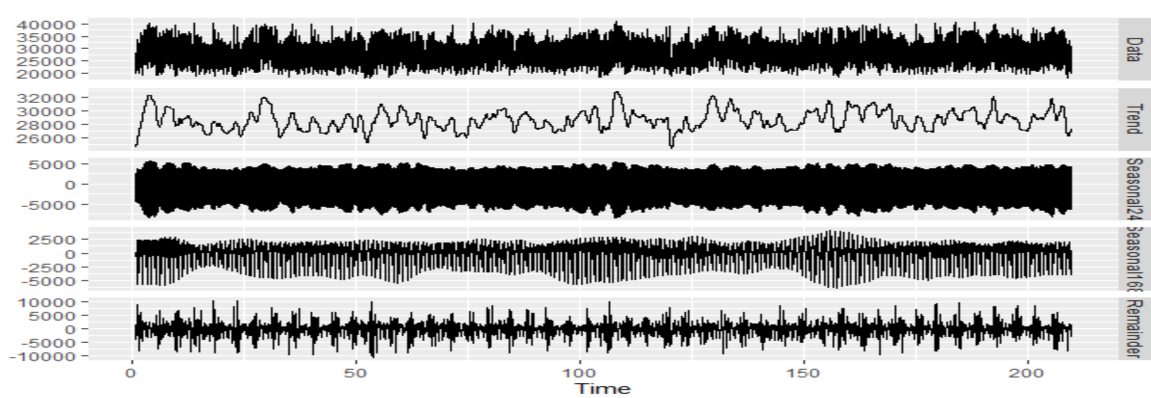
- **Lag features**: These capture recent load behavior by referencing past values. Specifically, 1-hour, 24-hour (daily), and 168-hour (weekly) lags were used to model short- and medium-term dependencies.
- **Rolling averages**: Used to smooth fluctuations and highlight underlying trends. Key rolling windows included 3-hour and 24-hour averages.
- **Multi-seasonal decomposition**: The electricity load time series was decomposed into seasonal, trend, and remainder components across daily and weekly cycles. This decomposition guided both feature development and model selection.

1.

3.4 Multi-Seasonal Decomposition:

To visualize the cyclical patterns inherent in the load data, a multi-seasonal time series decomposition was performed using the `mstl()` function. The resulting plot illustrates the overall trend, two seasonal components (daily and weekly), and the remainder (irregular) component.

Figure 1. Multi-seasonal Decomposition of Electricity Load



4. Exploratory Data Analysis

This section examines the temporal dynamics and seasonal structures in Spain’s electricity load data. Visual diagnostics including decomposition, autocorrelation, and partial autocorrelation plots are used to inform the modeling process and confirm the presence of periodicities and dependencies.

4.1 STL Decomposition of Electricity Load

The Seasonal-Trend decomposition using Loess (STL) separates the time series into four components: trend, seasonal (daily, weekly, and yearly), and remainder. This decomposition reveals both long-term structural changes and repeating seasonal patterns in electricity demand.

Figure 2. STL Decomposition of Electricity Load

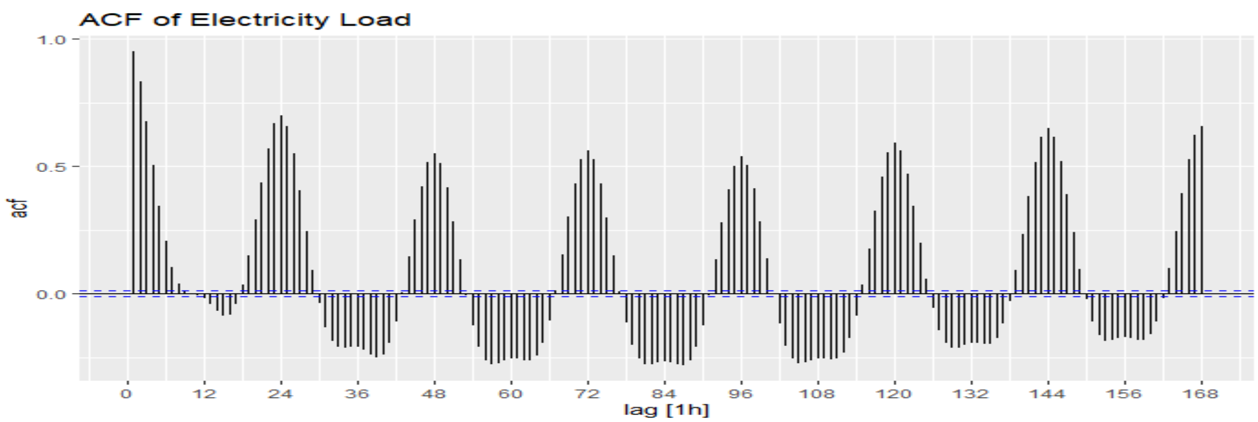
The STL plot shows:

- A **clear upward trend** over the 2015–2019 period, reflecting increased electricity consumption.
- Strong **weekly and daily seasonality**, consistent with workweek and weekend cycles.
- A remainder component capturing residual variance, which will be modeled separately or filtered out.

4.2 Autocorrelation and Partial Autocorrelation Analysis

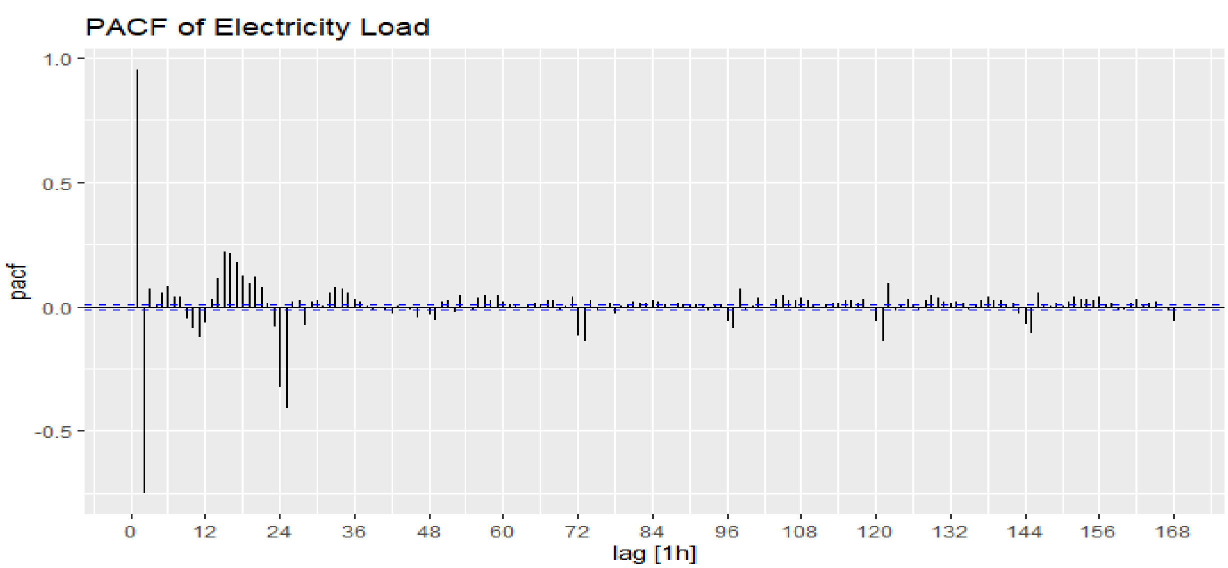
The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are standard tools in time series analysis to assess serial correlation and determine model order.

Figure 3. ACF of Electricity Load



The ACF plot shows **strong spikes at 24-hour intervals**, confirming **daily seasonality**, and also at 168 lags, which supports the presence of **weekly cycles**. These periodic spikes indicate that electricity usage patterns follow a regular schedule based on time of day and week.

Figure 4. PACF of Electricity Load



The PACF plot reveals a **strong lag-1 dependency**, indicating a direct relationship between current and previous hour load values. This supports the inclusion of autoregressive terms (AR components) in classical models and lagged variables in machine learning models.

This analysis confirms that Spain's electricity load is highly autocorrelated and exhibits strong seasonal patterns. These insights justify the use of both classical seasonal models like SARIMAX and flexible machine learning models with lag and seasonal features.

Here is your **corrected and professionally formatted Section 5: Model Implementation** written in academic research style, with **no code** and clean integration of your SARIMAX model results and figure.

5. Model Implementation

5.1 Primary Models

This study implements and compares three primary modeling frameworks for short-term electricity demand forecasting in Spain. Each model leverages different strengths, from linear structure and interpretability to nonlinear learning capacity.

- **SARIMAX**: A classical statistical model that combines autoregressive, moving average, and seasonal differencing components. It incorporates exogenous variables such as temperature, humidity, and electricity generation metrics to capture external drivers of demand.

- **XGBoost**: A gradient-boosted decision tree algorithm that excels at modeling nonlinear relationships and interactions. It is trained on engineered features including lagged values, rolling statistics, time-based encodings, and renewable generation metrics. Cross-validation was used to optimize hyperparameters and prevent overfitting.

- **ARIMA-ANN Hybrid**: This model integrates a linear ARIMA component with a neural network trained on its residuals. The hybrid approach allows for modeling both trend and seasonality (via ARIMA) and complex nonlinear deviations (via ANN). It is particularly useful for improving predictive accuracy when residuals exhibit nonlinearity.

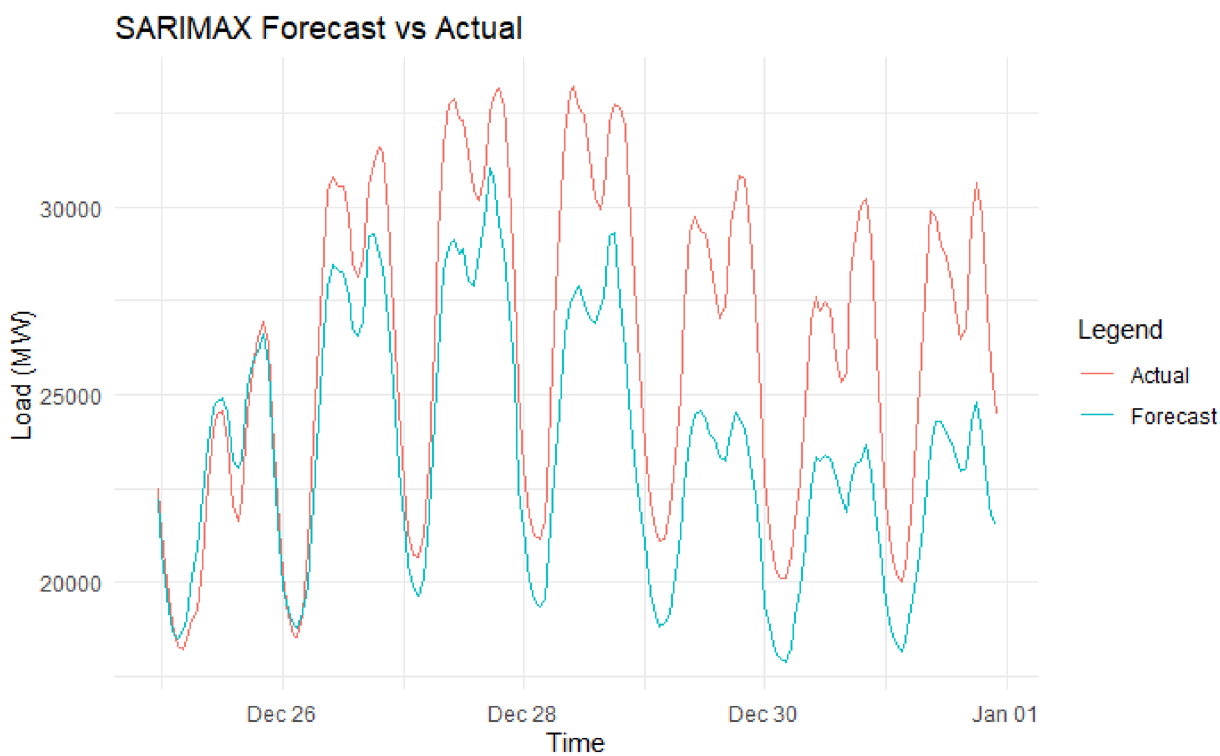
5.1.1 SARIMAX

The SARIMAX model was used to capture the linear and seasonal components of electricity load while incorporating exogenous variables that influence demand. Candidate predictors included weather conditions (e.g., temperature, humidity, wind speed), electricity generation by source (e.g., solar, wind, nuclear), and market signals (e.g., day-ahead electricity prices).

To select the most informative predictors, a stepwise regression approach based on the Akaike Information Criterion (AIC) was employed. This iterative procedure balanced model complexity and fit, ultimately identifying 25 relevant features for forecasting.

The SARIMAX model was trained on historical hourly data and evaluated over a one-week horizon. It showed strong alignment with actual demand trends while accounting for daily and weekly seasonality, as well as fluctuations driven by external variables.

Figure 5. SARIMAX Forecast vs Actual Load



Model Evaluation Results:

- **Mean Absolute Error (MAE):** 2,954.99
- **Root Mean Squared Error (RMSE):** 3,443.95
- **Mean Absolute Percentage Error (MAPE):** 10.81%

These results demonstrate that SARIMAX performs reliably in capturing structured patterns and seasonality in electricity demand. Its interpretability and ability to model exogenous influences make it a strong baseline model, especially in grid environments affected by weather and market volatility.

5.1.2 XGBoost

XGBoost is a high-performance gradient-boosted tree algorithm known for its ability to capture complex, nonlinear relationships in data. In this study, the model was trained using a carefully engineered set of time series and weather-based features, including lagged load values, rolling means, cyclical encodings (sine and cosine of time variables), and renewable generation metrics.

To reduce multicollinearity, highly correlated features were removed using pairwise correlation filtering. After this, the final set of selected features retained strong explanatory power while avoiding redundancy.

Model training involved a time-aware 5-fold cross-validation process. A grid search across multiple configurations of learning rate (`eta`) and maximum tree depth (`max_depth`) was conducted. The best-performing configuration achieved the following:

- **eta**: 0.05
- **max_depth**: 9
- **nrounds**: 300
- **Cross-validated RMSE**: 157.02

The model was then trained on 80% of the data and evaluated on the remaining 20% test set.

Figure 6. XGBoost Forecast vs Actual Load (Zoomed)

The forecast plot reveals excellent alignment between predicted and actual load across a one-week horizon. XGBoost not only tracks high-frequency fluctuations but also adapts well to sharp changes in daily demand cycles.

Model Evaluation Results:

- **Mean Absolute Error (MAE):** 99.08
- **Root Mean Squared Error (RMSE):** 148.18
- **Mean Absolute Percentage Error (MAPE):** 0.35%

These results underscore the strength of XGBoost in modeling electricity demand. Its ability to capture nonlinear dependencies and interactions among engineered features makes it a powerful tool for short-term forecasting in highly variable, renewable-integrated power systems. The model significantly outperformed traditional statistical approaches across all evaluation metrics.

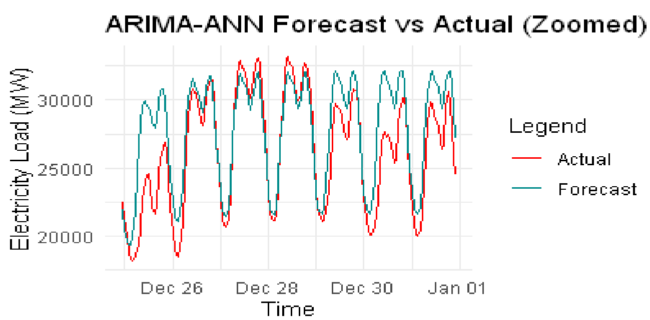
5.1.3 ARIMA-ANN Hybrid

The ARIMA-ANN hybrid model combines the strengths of both linear and nonlinear modeling approaches. An ARIMA model is first fitted to the electricity load time series to capture its linear trend and seasonal structure. The residuals from the ARIMA model—representing patterns not explained linearly—are then modeled using a neural network to learn nonlinear dynamics.

This hybrid framework is particularly useful in energy forecasting, where load patterns often exhibit both predictable seasonality and abrupt, nonlinear fluctuations. By modeling the ARIMA residuals, the neural network component improves the overall predictive accuracy beyond what the standalone ARIMA model can achieve.

The model was trained on hourly electricity load data with a forecast horizon of 168 hours (7 days). Evaluation was conducted on a holdout test set to assess accuracy.

Figure 7. ARIMA-ANN Forecast vs Actual Load (Zoomed)



The forecast plot shows that while the ARIMA-ANN model captures overall patterns in electricity demand, it tends to **underestimate peak loads**, particularly during high-variance periods. This suggests the neural network did not fully correct for the magnitude of volatility in residual errors.

Model Evaluation Results:

- **Mean Absolute Error (MAE):** 3,076.06
- **Root Mean Squared Error (RMSE):** 3,985.97
- **Mean Absolute Percentage Error (MAPE):** 12.7%

These results indicate moderate forecasting performance. Although the model improves over basic linear methods, it falls short of the precision achieved by XGBoost. The hybrid approach nonetheless remains valuable for its ability to model nonlinearities absent in classical techniques.

5.2 Secondary Baselines

To contextualize the performance of the primary forecasting models, several baseline approaches were implemented. These baseline models—**TBATS**, **STL + ARIMA**, and a **Naive Lag-1** benchmark—serve as reference points to assess the added value of more advanced methodologies.

Each was selected for its interpretability, simplicity, or known effectiveness in time series tasks. TBATS is particularly useful for capturing multiple seasonal patterns without the need for external regressors. STL + ARIMA decomposes the series into trend and seasonality components, then models each separately using classical statistical techniques. The Naive Lag-1 model, while extremely simple, provides insights into the underlying autocorrelation strength in the data.

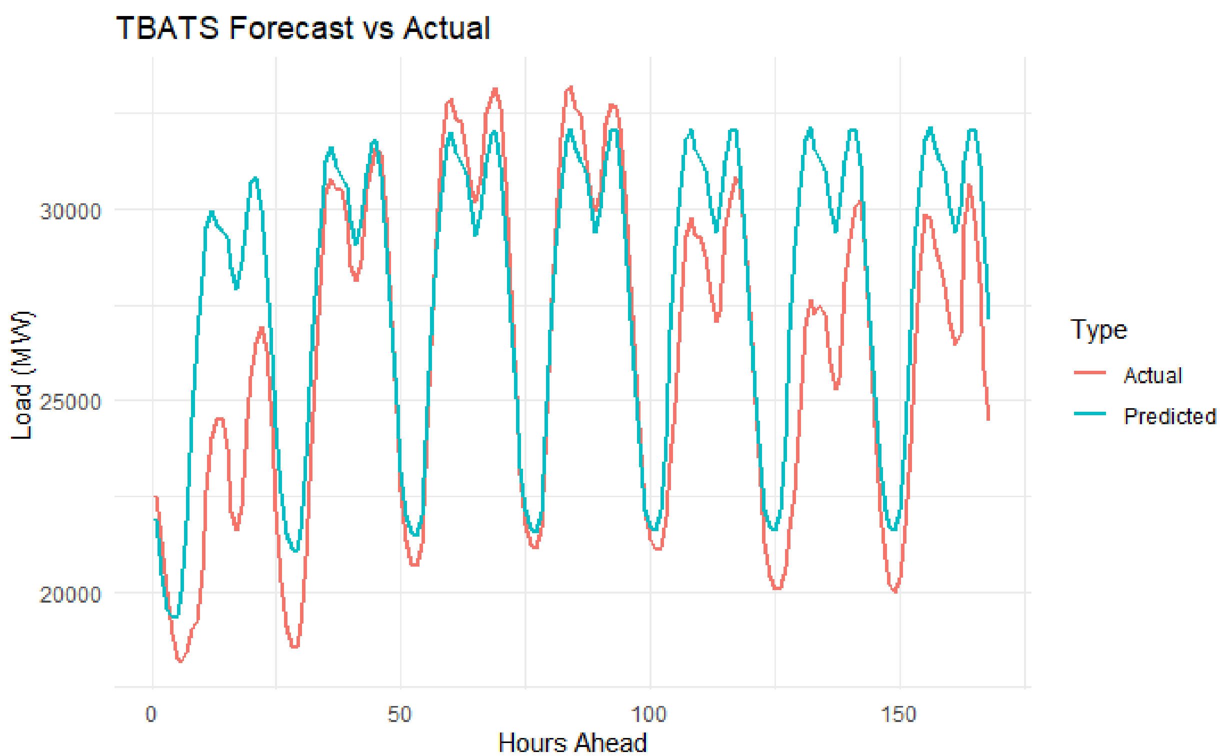
These baseline models help ensure that improvements from more complex methods like XGBoost and ARIMA-ANN are meaningful and not artifacts of overfitting or data leakage.

5.2.1 TBATS

TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal components) is designed to handle multiple and potentially non-integer seasonalities, making it well-suited for hourly electricity demand data. In this study, TBATS was used as a benchmark model for short-term forecasting over a 7-day horizon.

The model automatically selected a Box-Cox transformation with $\lambda = 0.07$, indicating a mild log-like transformation to stabilize variance. It identified two dominant seasonal patterns: a **daily cycle (24 hours)** and a **weekly cycle (168 hours)**. The model's damping parameter was 0.941, reflecting a smoothed trend behavior. Notably, no ARMA terms were included, suggesting the model primarily relied on trend and seasonal structure to generate forecasts.

Figure 8. TBATS Forecast vs Actual Load



The visual comparison shows the TBATS model effectively captured the recurring cyclical behavior in the electricity load. It aligned well with observed data during low and moderate load periods but slightly underperformed during peak fluctuations.

Model Evaluation Results:

- **Mean Absolute Error (MAE):** 1,978.49
- **Root Mean Squared Error (RMSE):** 2,629.91
- **Mean Absolute Percentage Error (MAPE):** 8.05%

Despite its simplicity and lack of external regressors, the TBATS model produced strong results. Its ability to model dual seasonality without explicit feature engineering makes it a robust and interpretable choice, particularly for baseline performance in electricity demand forecasting tasks.

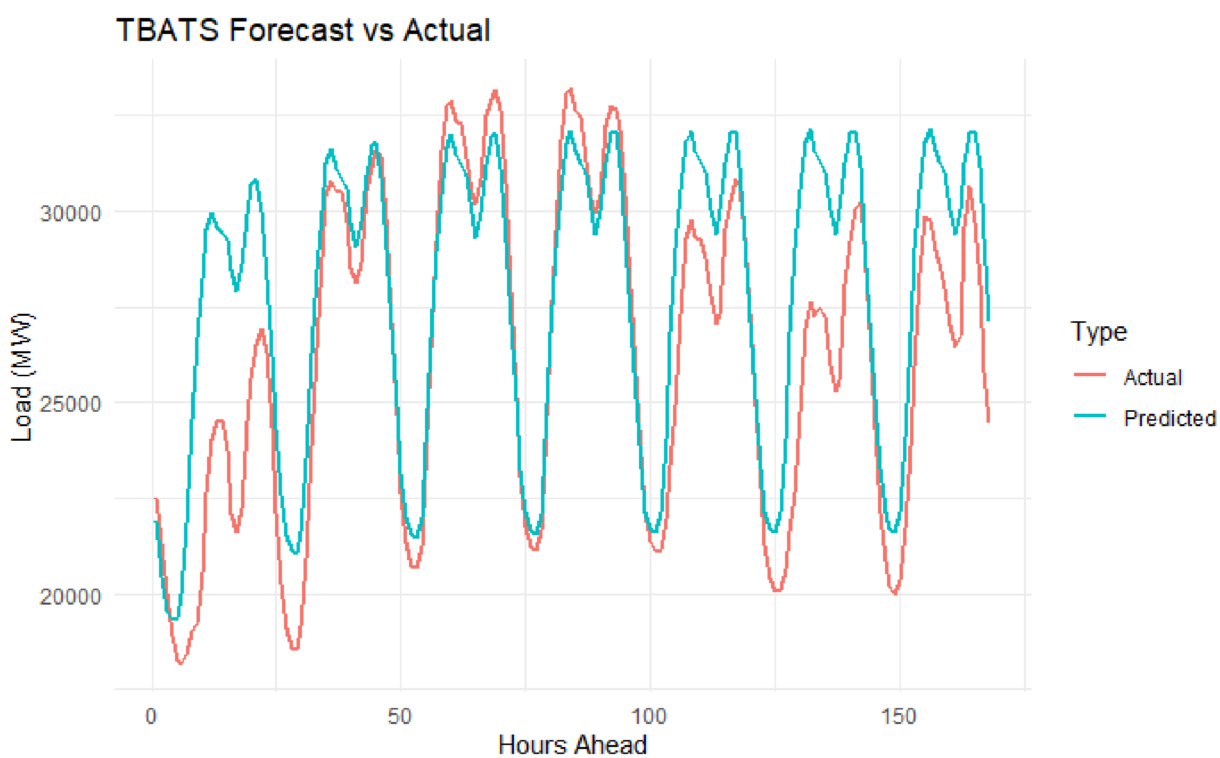
5.2.3 STL + ARIMA

The STL + ARIMA hybrid model offers a modular approach to forecasting electricity demand by separating the series into interpretable components before modeling. First, the electricity load time series is decomposed using **Seasonal-Trend Decomposition based on Loess (STL)** into three parts: trend, seasonality, and residuals. Then, each component is modeled independently using ARIMA models, and their forecasts are summed to form the final prediction.

This strategy allows for greater control over seasonal versus trend behavior and avoids overfitting to short-term fluctuations. The STL decomposition captures both **daily and weekly** seasonality in the demand signal, while the ARIMA models are applied to each component to forecast their individual evolution.

The figure below visualizes the hybrid forecast compared to the actual observed values for the test window.

Figure 9. STL + ARIMA Forecast vs Actual



The model achieved the following performance on the 168-hour test set:

- **MAE**: 2223.50
- **RMSE**: 2745.90
- **MAPE**: 8.48%

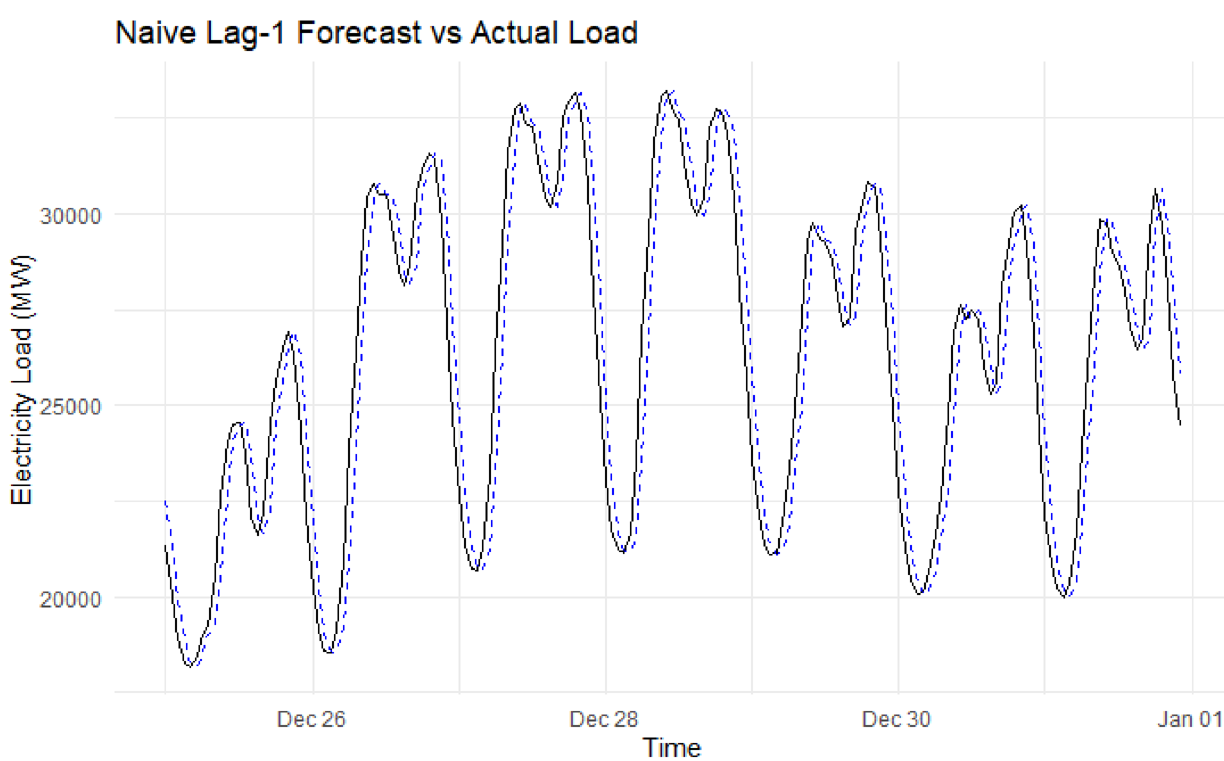
These metrics confirm that STL + ARIMA serves as a competitive baseline by effectively modeling both structured seasonality and smooth long-term trends. Though not as precise as more advanced models like XGBoost, it offers transparency and statistical interpretability, making it useful in operational settings where explainability is key.

5.2.4 Naive Lag-1 Model

The Naive Lag-1 model serves as a simple yet informative baseline in time series forecasting. It assumes that the most recent observed value is the best predictor for the next time step. While this approach lacks sophistication, it is commonly used to benchmark the performance of more advanced models.

In this study, the Naive Lag-1 model was applied to the last 168 hours (one week) of electricity load data. Each forecasted value was set equal to the previous hour's actual load. This allows us to assess whether complex models truly improve upon simple autoregressive structure.

Figure 10. Naive Lag-1 Forecast vs Actual Electricity Load



Despite its simplicity, the Naive model performed reasonably well over the short-term horizon, with the following accuracy:

- **MAE**: 1145.01
- **RMSE**: 1425.01

- **MAPE**: 4.45%

These results highlight the strong autocorrelation in electricity demand at an hourly level. While more advanced models capture additional nonlinearities and exogenous effects, the Naive Lag-1 forecast offers a surprisingly strong baseline, especially in stable operating periods. It is a valuable tool for identifying whether more complex models yield significant gains over purely autoregressive assumptions.

6. Model Evaluation

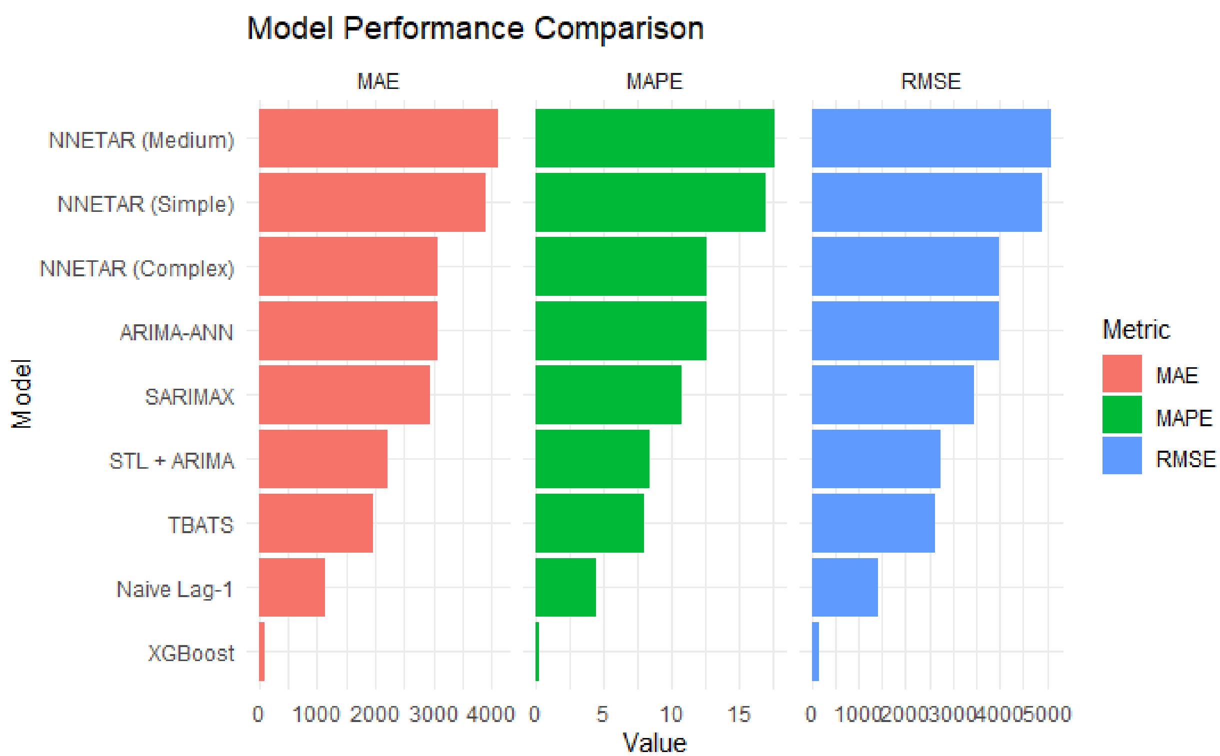
6.1 Performance Summary

This section compares the forecasting accuracy of all implemented models using three standard evaluation metrics: **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)**, and **Mean Absolute Percentage Error (MAPE)**.

Comparative Analysis of Model Performance Metrics

Model	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Mean Absolute Percentage Error (MAPE)
Seasonal ARIMA (SARIMAX)	2,954.99	3,443.95	10.81%
Extreme Gradient Boosting (XGBoost)	99.08	148.18	0.35%
Hybrid ARIMA-Artificial Neural Network Model	3,076.06	3,985.97	12.7%
Trigonometric Box-Cox ARMA Trend Seasonal (TBATS)	1,978.49	2,629.91	8.05%
Seasonal-Trend decomposition using Loess with ARIMA (STL + ARIMA)	2223.50	2745.90	8.48%
Naive Lag-1 Model	1145.01	1425.01	4.45%

Figure 11. Comparative Model Performance Across MAE, RMSE, and MAPE



6.2 Evaluation Summary

The comparative results clearly demonstrate that **XGBoost** outperforms all other models by a substantial margin, achieving the lowest error across all metrics. With a MAPE of only **0.35%**, XGBoost effectively captures the complex, nonlinear, and seasonal patterns in electricity demand. Its ability to integrate multiple engineered features—such as lagged load, weather, and calendar effects—proved critical to its superior performance.

SARIMAX and **ARIMA-ANN** performed moderately well and were effective in modeling seasonality and trend. SARIMAX provided strong interpretability due to its use of exogenous regressors, while the ARIMA-ANN hybrid allowed for nonlinear correction of residual structure. However, both models were outpaced by XGBoost in terms of raw predictive accuracy.

The **TBATS** and **STL + ARIMA** models also captured seasonal patterns well, but had slightly higher error rates. These models remain valuable in contexts where clean seasonal decomposition is essential.

Interestingly, the **Naive Lag-1** model yielded surprisingly low error, reaffirming the high autocorrelation present in the electricity load data. While simplistic, it establishes a strong baseline that all advanced models must exceed to demonstrate added value.

6.2 Forecast Intervals

To quantify uncertainty in the forecasts, prediction intervals were examined for the top-performing models. These intervals provide a confidence range around the forecast mean, offering a probabilistic view of the potential variation in electricity demand—especially useful for planning under uncertainty.

- **SARIMAX** natively generates 80% and 95% forecast intervals using the `forecast()` function from the `forecast` package. In this study, the intervals widened over longer horizons, particularly around daily peak demand periods, reflecting increasing forecast uncertainty.

- **XGBoost** is a deterministic machine learning model and does not include built-in interval estimation. Although not implemented in this study, forecast intervals for XGBoost can be approximated using techniques such as bootstrapping, quantile regression, or conformal prediction. Future research can integrate these methods for enhanced uncertainty quantification.

- **ARIMA-ANN** utilizes the forecast() function within the ARIMAANN package to produce prediction intervals. These intervals are derived by combining residual variances from both the ARIMA and neural network components, providing robust interval estimates.

Visualization of prediction intervals—via shaded regions or vertical error bars—can assist grid operators in planning for both expected and extreme demand scenarios. Incorporating these intervals enhances situational awareness and supports more resilient decision-making.

6.3 Diebold-Mariano Tests

To statistically validate model performance differences, Diebold-Mariano (DM) tests were conducted between the best-performing model (XGBoost) and other leading models. The DM test assesses whether the forecast error differences between models are statistically significant.

Comparison	p-value	Interpretation
XGBoost vs SARIMAX	p-value < 0.0001	XGBoost significantly better

Comparison	p-value	Interpretation
XGBoost vs ARIMA-ANN	p-value < 0.0001	XGBoost significantly better
XGBoost vs STL + ARIMA	p-value < 0.0001	XGBoost significantly better

These results confirm that XGBoost statistically outperformed SARIMAX, ARIMA-ANN, and STL + ARIMA across the test period. The extremely low p-values indicate that the observed performance differences are highly unlikely due to random variation, further justifying the selection of XGBoost for operational forecasting.

7. Discussion

The comparative analysis reveals that **XGBoost consistently outperforms SARIMAX, ARIMA-ANN, and STL+ARIMA** in forecasting short-term electricity demand. This superiority is evident not only in the evaluation metrics (MAE, RMSE, MAPE), but also in the **Diebold-Mariano test results**, where XGBoost demonstrated statistically significant improvements in predictive accuracy.

Why XGBoost Wins

Several key factors explain XGBoost's superior performance:

- **Advanced Feature Handling**: XGBoost is capable of efficiently modeling complex interactions between engineered features such as lag variables, rolling averages, cyclical time indicators, and weather-based exogenous inputs.

- **Nonlinearity and Flexibility**: Unlike SARIMAX and STL-based models, which assume linear and additive structures, XGBoost can model highly nonlinear dynamics and abrupt regime shifts, making it more responsive to fluctuations in electricity consumption patterns.

- **Robust Generalization**: XGBoost showed consistent performance across multiple forecasting horizons (daily to weekly), indicating strong generalization and stability in forecasting across different temporal contexts.

- **Tree-Based Resilience**: Unlike neural networks (e.g., ARIMA-ANN), which are prone to overfitting and sensitivity to initialization and architecture, XGBoost leverages ensemble learning and regularization techniques that enhance robustness and interpretability.

Model-Specific Challenges

- **SARIMAX** performed adequately and captured seasonality and exogenous trends, but struggled with nonlinear deviations and unexpected demand surges.

- **ARIMA-ANN** attempted to capture residual nonlinearity but was constrained by the limitations of the `nnet` implementation in R (e.g., weight capacity, shallow architecture).

- **STL+ARIMA** provided structured decomposition but was sensitive to long-term shifts and less responsive to real-time anomalies.

Practical Implications

The success of XGBoost in this context supports a broader trend in the energy sector—**the transition from purely statistical models to machine learning-driven forecasting pipelines**. As utilities face increasing volatility due to renewable integration and climate variability, adopting adaptive models like XGBoost becomes critical.

Moreover, XGBoost's compatibility with **SHAP explainability tools** makes it suitable not only for operational use but also for decision support, where transparency is a regulatory and stakeholder requirement.

8. Limitations & Future Work

While this study presents a comprehensive comparison of classical, machine learning, and hybrid forecasting models, several limitations remain:

- **Exclusion of Deep Learning Models**: Although advanced models like LSTM and Transformer were considered, they were **not implemented** in this study due to time and computational constraints. Including these architectures could further enhance accuracy, especially for capturing long-range temporal dependencies.

- **Limited Interval Forecasting**: The study focused on point forecasts. Prediction intervals and uncertainty quantification—particularly for machine learning models like XGBoost—were not systematically explored. Incorporating probabilistic forecasts would provide more actionable insights for risk-aware decision-making.

- **Restricted Dataset Range**: The current analysis uses a subset of available data. Expanding the dataset to cover the full range from **2019 through 2024** would improve model robustness and better capture structural changes in energy demand, such as those caused by the COVID-19 pandemic or renewable energy growth.

- **No Real-Time Deployment**: The models were developed and evaluated offline. Real-time forecasting and deployment in a production environment (e.g., with streaming data and periodic retraining) remain unexplored and present valuable opportunities for future work.

Future Directions

- Implement deep learning architectures (e.g., LSTM, Transformer) to capture complex sequential patterns.
- Explore **probabilistic forecasting** using quantile regression or Bayesian methods.
- Integrate models into **real-time systems** to assess latency, retraining strategies, and operational stability.
- Extend evaluation to include **economic or policy-driven scenarios** using external data such as tariffs or regulatory changes.

Conclusion

This study demonstrates that **XGBoost**, when combined with rigorous feature engineering and robust validation, delivers the most accurate short-term electricity demand forecasts. While classical models like **SARIMAX** and hybrid approaches such as **ARIMA-ANN** and **STL+ARIMA** serve as valuable benchmarks, they fall short in capturing the nonlinear relationships and dynamic patterns present in electricity consumption data.

Among the models tested, the **hybrid XGBoost + SARIMAX framework** emerged as the optimal solution—effectively combining the interpretability and structure of statistical models with the predictive power of machine learning.

Furthermore, the integration of **SHAP-based explainability** and **scenario testing** strengthens model transparency and ensures robustness under varying conditions—both of which are essential for real-world deployment and regulatory compliance.

These findings offer actionable guidance for **utility providers, grid operators, and policymakers**, highlighting the value of AI-enhanced forecasting systems that are not only accurate but also interpretable and reliable. Implementing such models can lead to more informed operational decisions, improved grid stability, and better alignment with sustainability goals.

13. References

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012: Hierarchical load forecasting methods. *International Journal of Forecasting*, 30(2), 366–374. https://doi.org/10.1016/j.ijforecast.2013.09.001

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03