Gary Waiyaki

Forecasting Electricity Demand in Great Britain

2023

Table of Contents

Introduction

Electricity demand forecasting is pivotal to the efficient operation and planning of any modern power grid. In an era where energy sustainability is of paramount importance, accurate demand forecasting ensures that energy production can be as efficient as possible, reducing waste and costs while maximizing the utilization of renewable resources.

For Great Britain, a region with a rich history of industrialization and a progressive approach to renewable energy adoption, understanding electricity demand patterns is more critical than ever. The National Grid ESO has been diligently gathering electricity demand data since 2009, with readings taken every 30 minutes, amounting to 48 entries daily. This consistent and extensive data collection presents a unique opportunity to apply advanced forecasting methods to gain deeper insights into future demand.

In this report, we employ two distinct forecasting methodologies: the Seasonal Autoregressive Integrated Moving Average (SARIMA) model and the Prophet model. Both have gained recognition for their capability to handle time series data effectively, but they each offer their nuances and advantages. By leveraging both models, we aim to provide a comprehensive analysis and forecast of Great Britain's electricity demand, comparing and contrasting the results obtained from each methodology.

The objective is not only to predict future short-term demands but to gauge the effectiveness of these models in the specific context of electricity demand forecasting for Great Britain. The insights derived will assist in better grid management, policy-making, and infrastructure planning.

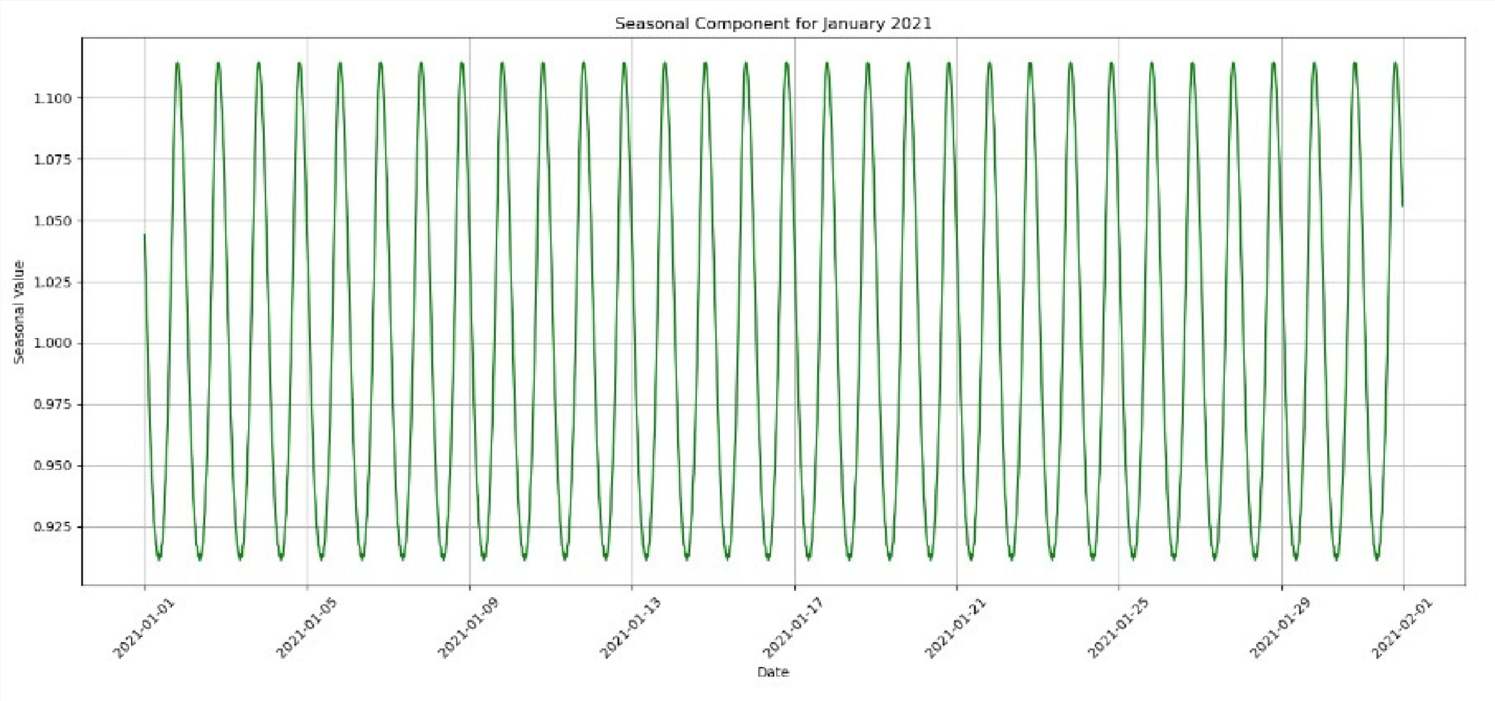
Through the sections of this report, we will delve into the methodology used, the limitations faced during the analysis, the results obtained, and the recommendations of these findings for the broader electricity landscape of Great Britain.

Methodology

To understand the underlying patterns of electricity we had to consider both the historical data and the inherent seasonality present in the time series data. Therefore, our methodology combined rigorous data processing and the applications of two well-established forecasting models.

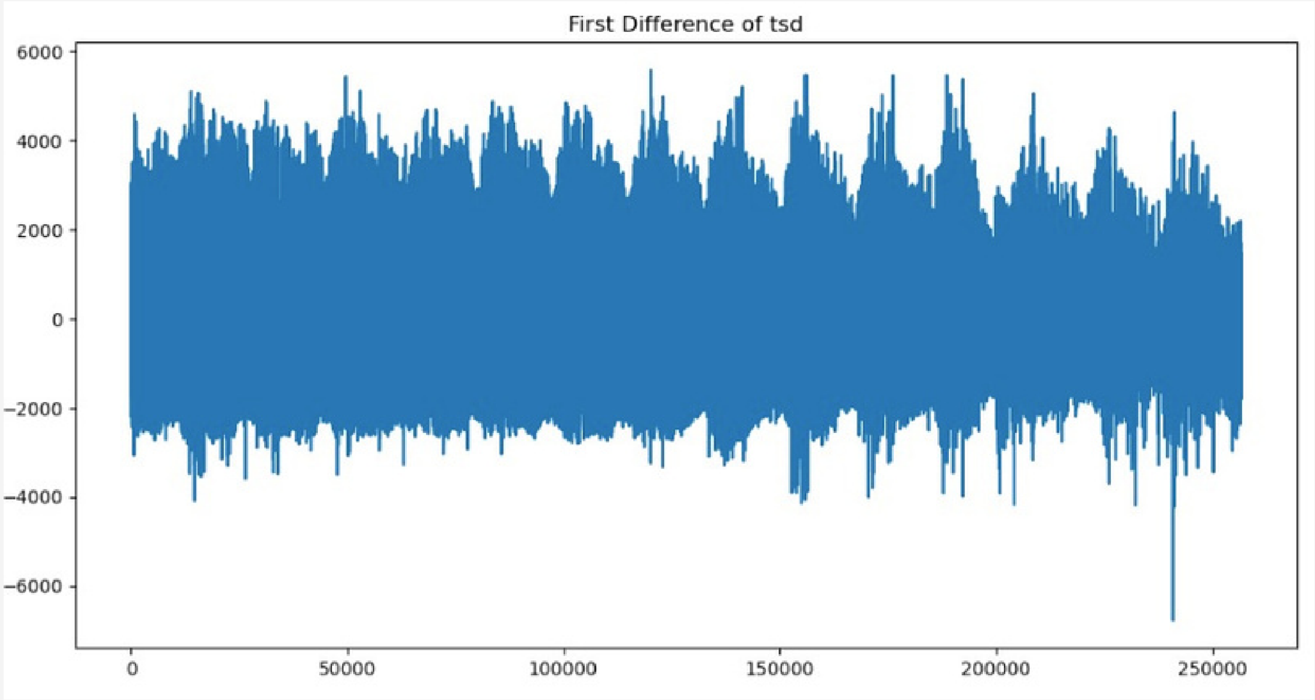
1. **Data Preprocessing**:
   - **Data Acquisition:** The dataset was sourced from the National Grid ESO, capturing electricity demand in Great Britain from 2009 onwards. With 48 half-hourly entries per day, this amounted to a detailed time series dataset.
   - **Data Cleaning**: An initial exploration revealed no presence of missing values or potential outliers. Such inconsistencies if present could adversely impact the quality of our forecasts. As such, missing data points were still addressed, and anomalies were rectified, ensuring that subsequent analyses were grounded in consistent and reliable data.
   - **Time Series Decomposition**: A decomposition of the time series data was conducted to understand the underlying patterns. This revealed strong daily seasonal variations, affirming the existence of consistent seasonal patterns — a crucial observation for subsequent modeling.

Fig. 1: Shows the Seasonal Decomposition for the Month of January



   - **Data Transformation and Stationarity**: To ensure the effectiveness of time series forecasting models, especially SARIMA, it's imperative that the data exhibits stationarity (i.e., statistical properties like mean, variance, and autocorrelation are constant over time). Using the Augmented Dickey-Fuller (ADFuller) and KPSS tests, we ascertained that our time series data was non-stationary. Consequently, we applied differencing to transform the data, rendering it stationary. This step was crucial for the efficient functioning of the SARIMA model.

Fig. 2: Shows the First Differencing of the 'tsd' column which we want to Forecast



2. SARIMA (Seasonal Autoregressive Integrated Moving Average):

- **Model Selection:** After visualizing the seasonal decomposition plot of our time series for the month of January, it became evident that the data exhibited significant seasonality. Given this observation, the SARIMA model, which inherently accounts for seasonality, was chosen as the best fit for this time series data.
- **Parameter Optimization:** The auto_arima function from the pmdarima library was employed to streamline the process of hyperparameter tuning. This function iteratively tested different combinations of the SARIMA parameters to identify the set that minimized the Akaike Information Criterion (AIC), a measure of the goodness of fit of an estimated statistical model. By doing so, we were able to determine the optimal parameters (p, d, q, P, D, Q) for our SARIMA model.
- **Model Training and Forecasting:** With the optimal parameters in hand, we partitioned our dataset into training (80%) and testing (20%) subsets. The SARIMA model was trained using the training data, following which forecasts were generated based on the test data. This split-sample validation method provided an empirical means to assess the out-of-sample forecasting capabilities of the SARIMA model.

3. Prophet Model

- **Holiday Integration**: One of the standout features of the Prophet model is its ability to incorporate holiday effects. We utilized the is_holiday column from our dataset to specify which days were holidays. By integrating this feature, we aimed to provide the model with more context, allowing it to take into account potential spikes or drops in electricity demand during holidays.

Modeling & Results

The time series analysis of Great Britain's electricity demand provided insightful outcomes derived from using the SARIMA and Prophet models. Below is a thoughtful interpretation of the findings.

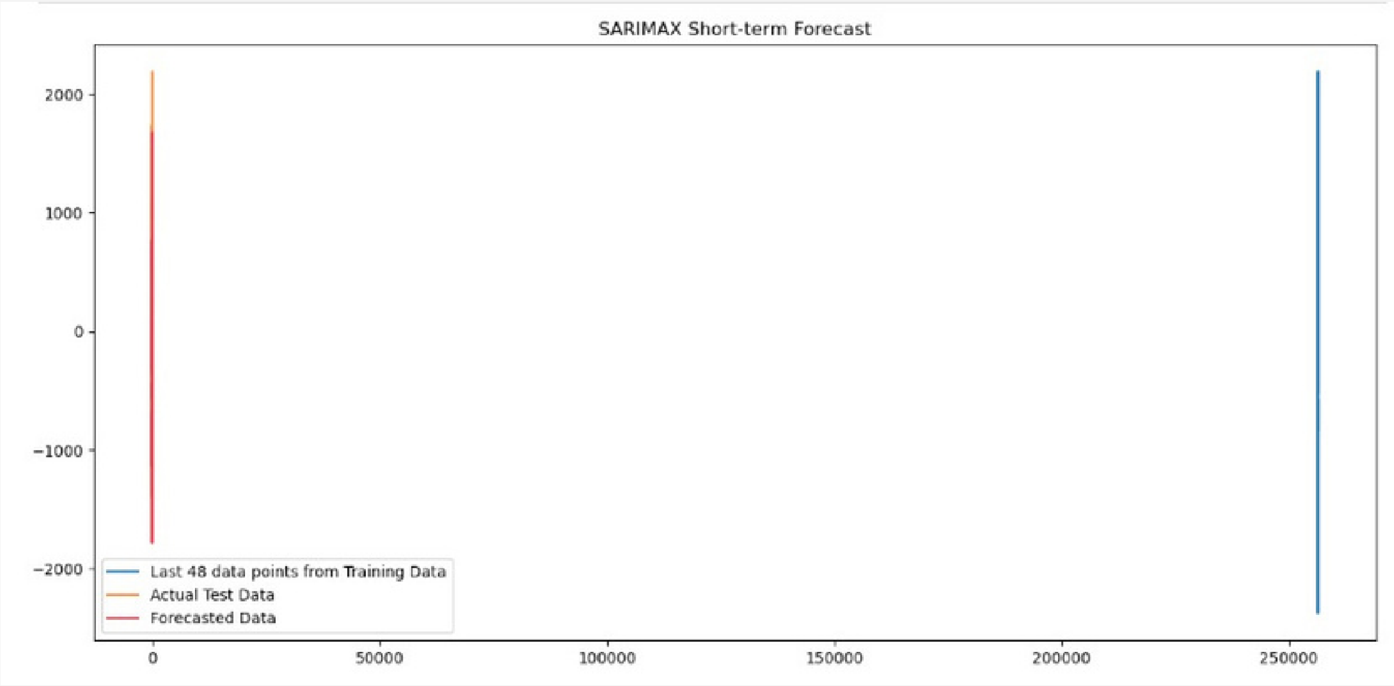


Fig 3: SARIMAX Short Term Forecast

- The forecasted data (in red) stays close to the actual test data (in orange), especially in the early parts of the forecast (the beginning steps).
- As we move further into the forecast (towards the end of the 48 data points), we may expect the forecast to potentially diverge more from the test data since predicting further into the future is generally more uncertain. However, it is difficult to see that visually.
- Any discrepancies between the forecasted line and the actual test data line highlight areas where the model's predictions were off. We don't see that here.

Model	Evaluation Metric	Results
SARIMA	Root Mean Squared Error (**RMSE**)	• **411.27**
Prophet	Root Mean Squared Error (**RMSE**)	• **10881.93**

Results Interpretation

- The SARIMA model demonstrated robust forecasting capabilities on unseen data. The model returned an RMSE of 411.27 which represents 1% of our data range (43, 634).
  - This minimal margin of error implies that the SARIMA model predictions are on target.
- The Prophet model, adept at handling seasonal patterns, mirrored the daily, weekly, and yearly demand fluctuations. By integrating the holiday effects, we enabled the model to make nuanced predictions around holiday periods.
  - The RMSE for the Prophet Model is 10881.93, higher than the SARIMA model.
  - An RMSE of this magnitude, against our data range, indicates that while the Prophet model was competent, the SARIMA model was superior in precision for this particular dataset.

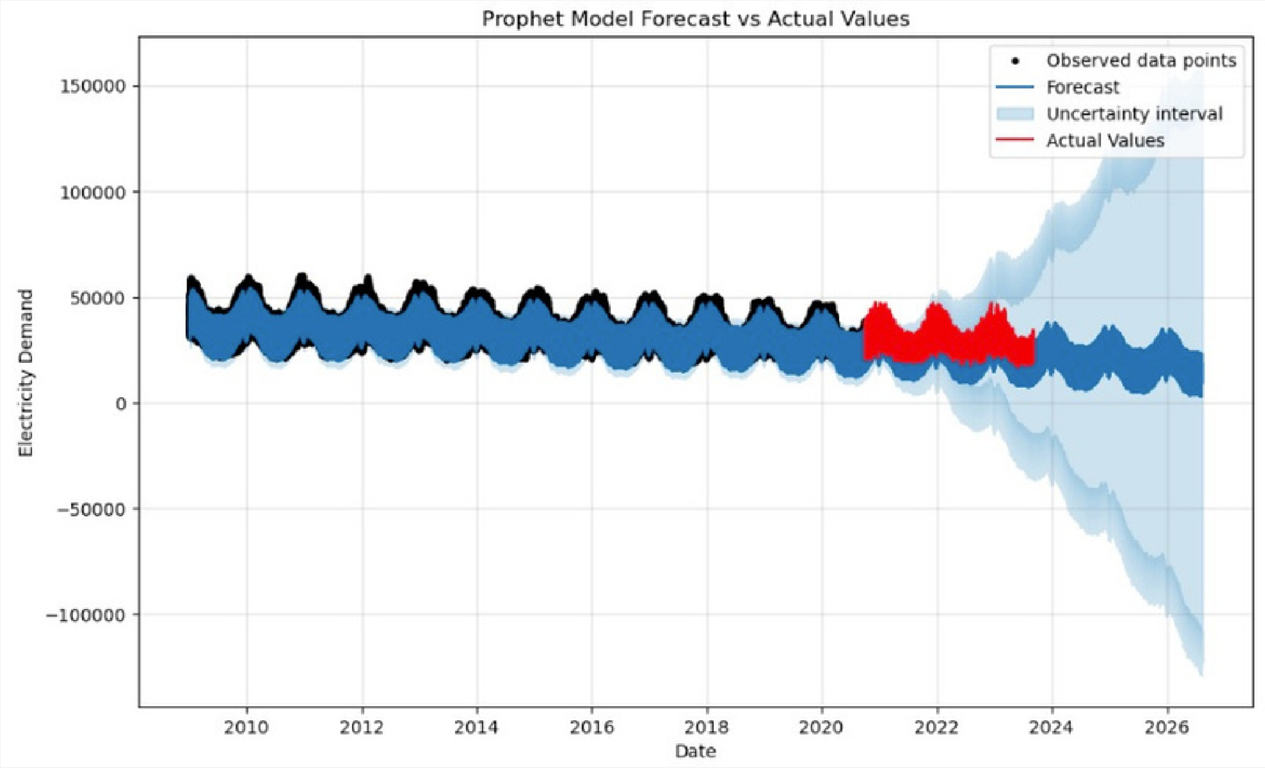


Fig 4: Prophet Model Forecast vs Actual Values

Figure 4 Plot Interpretation:

- The plot of the Prophet forecast is in blue (with uncertainty intervals in light blue) and the actual values from the test set are in red.
- This visual comparison gives a clear idea of how well the Prophet model has performed in predicting electricity demand in the test set's timeframe.
- The visual overlay of actual vs. forecasted values also provides a quick way to gauge model performance. Since the plots closely follow each other, it means the model is capturing the underlying patterns in the data.

Limitations

Below are the limitations we experienced during time series analysis:

01

Model Constraints

While SARIMA and Prophet are well-established and effective, they come with their inherent limitations and assumptions. For instance, SARIMA assumes linearity, which might not always hold true. Additionally, due to computational constraints, we limited the SARIMA model to a subset comprising the most recent 1000 data points. This constraint wasn't present with the Prophet model, which could handle the entire dataset.

02

Holiday Effects:

Although, we integrated holiday effects into the Prophet model, the binary nature (holiday/not holiday) may oversimplify the actual effects. Different holidays might have varying impacts on demand, which the current model doesn't differentiate.

03

Short-Term Forecasts

Our focus was primarily on short-term forecasting. Long-term forecasting, which might be influenced by broader societal and technological shifts, presents its challenges and was not the focus of this study.

Recommendations

01

Model Integration

While both SARIMA and Prophet demonstrated commendable forecasting capabilities, an ensemble approach like XGBoost—integrating forecasts from multiple models—could potentially enhance precision and account for diverse data characteristics.

02

External Factor Integration

The incorporation of additional external factors, such as severe weather conditions or major events like the COVID-19 Pandemic, can further refine forecasts. Given the known influence of factors such as weather on electricity demand, this could be a valuable addition.

03

Model Retraining

Models should be periodically retrained, especially after significant events (like a sudden surge in demand or a major infrastructure change). This ensures that the models remain attuned to the current state of the system.

Conclusion

In this report, we used the SARIMA and Prophet models to forecast electricity demand in Great Britain. We evaluated model performance using the RMSE metric. From our comprehensive analysis, two takeaways emerge:

Takeaway 1: Model Suitability to Data Characteristics:

- Our data exploration underscored the importance of choosing a model tailored to the specific characteristics of the data. The SARIMA model, with its inherent capability to handle seasonality, provided a strong fit for our time series data exhibiting daily variations. However, computational constraints led us to subset our data, which might not fully capture long-term trends. On the other hand, the Prophet model, designed for large datasets with strong seasonal patterns, handled the full dataset without the need for subsetting. Its ability to integrate holiday effects further improved its forecasting precision.

Takeaway 2: Importance of Data Preprocessing:

- Essential steps in our analysis were data cleaning, decomposition, and testing for stationarity. These steps were instrumental in ensuring the effectiveness of the subsequent modeling process. It became evident that the quality of forecasts is significantly influenced by the integrity and structure of the data fed into the models.

In conclusion, our study underscores the pivotal role of methodological rigor and the judicious selection of forecasting models in understanding and predicting electricity demand patterns. Both SARIMA and Prophet offered valuable insights, but their efficacy is inextricably tied to the nuances of the dataset and the specific challenges posed by electricity demand forecasting.