Time-Series Analysis of Energy Consumption (2002–2018): Trends, Seasonality, and Forecasting Performance

Group 7: Prateek Singh and Sumanth Reddy Thandra
(MS in Data Science)

Time-series Analysis (MA 5781)
Spring 2025
under
Prof. Yeonwoo Roo

Michigan Technological University

Table of Contents

Abstract

This paper presents a detailed time-series analysis of PJM Interconnection LLC (PJM) hourly energy consumption for the period 2002-2018. PJM manages one of the nation's largest regional transmission organizations that serves all or parts of 13 states and the District of Columbia. The study examines long-run historical evolution of energy consumption over time, unveiling permanent tendencies and showing significant seasonal and cyclical effects—like summer and winter peaking demand caused by cooling and heating loads, reduced energy consumption during holidays, and discernible weekday-weekend patterns.

SARIMA models can be applied for short-term forecasting with normal seasonality, but they do not perform well in longer horizons because they depend on past patterns and stationarity. Even though more advanced models such as machine learning or combination models may outperform SARIMA in long-horizon settings, SARIMA is a viable and explainable option for energy forecasting with minimal preprocessing needs.

Introduction

PJM Interconnection LLC (PJM) is a major regional transmission organization (RTO) in the United States that operates the electric grid and coordinates the wholesale electricity market across a large territory that includes parts of the Mid-Atlantic and Midwest regions, covering all or parts of 13 states and the District of Columbia. The PJM East zone, in particular, is densely populated and economically active, leading to complex and high-volume electricity demand patterns.

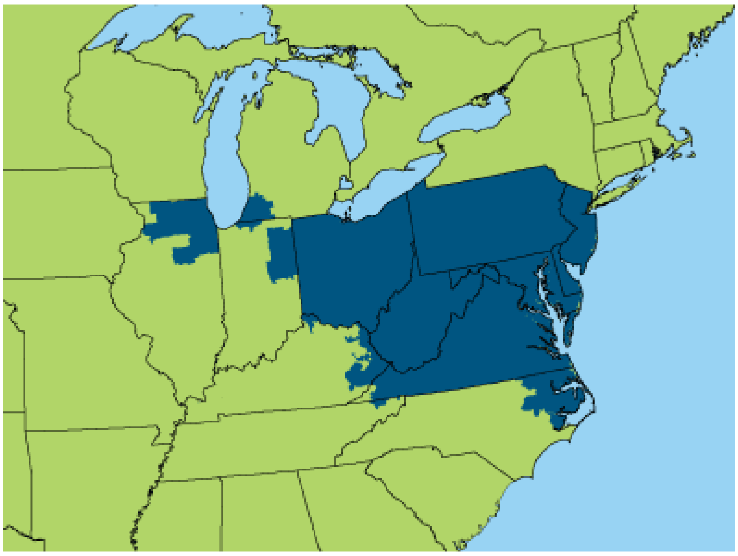

Figure 1. *PJM East regional coverage map across eastern United States.*

As energy consumption continues to evolve due to technological advancements, urban growth, climate variability, and shifts in usage behavior, it becomes increasingly important to understand historical consumption patterns and to accurately predict future demand. These forecasts are vital for effective energy planning, resource allocation, infrastructure investment, and policy formulation.

This study explores 17 years of hourly electricity consumption data with the goal of identifying significant demand patterns and evaluating suitable time-series forecasting techniques. By analyzing both the full data (2002–2018) and a post-2011 subset (2012–2018), the analysis aims to capture long-term structures as well as more recent changes in electricity usage. The modeling process integrates transformation methods, stationarity assessments, and diagnostic tools to inform the selection of predictive models. The resulting forecasts provide insights into expected demand patterns under varying temporal conditions and support decision-making in energy management through 2025.

Research Questions?

1. How has energy consumption in the PJM region evolved from 2002 to 2018?
2. What are the key seasonal and cyclical patterns observed in the data?
3. How accurately can SARIMA models forecast short-term and long-term energy usage?
4. What are the limitations of the SARIMA model?
5. Which forecasting approach or aggregation would be optimal?

Time Series Data

The dataset used in this study comprises hourly electricity consumption data in megawatts (MW) from PJM Interconnection LLC, sourced from its official website. The data spans from 2002 to 2018 and includes consumption figures across various subregions within PJM's operational footprint. Due to changes in PJM's coverage and structure over time, data availability varies by region and date. The data underwent preprocessing to handle missing values, ensure consistency in time intervals, and aggregate or filter by relevant geographic areas where necessary.

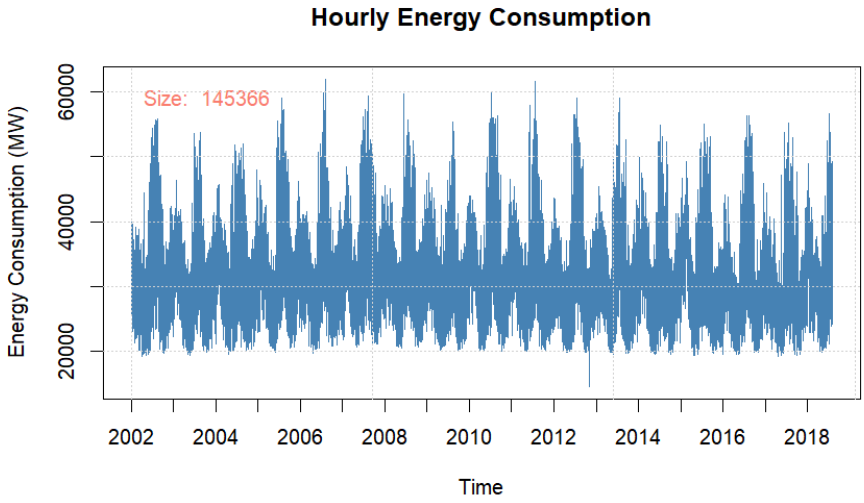

Figure 2. *Hourly electricity consumption in the PJM East region from 2002 to 2018*

The **Hourly Energy Consumption** shows electricity usage in megawatts (MW) from 2002 to 2018 in the PJM East region. The time series consists of 145,366 hourly observations. It reveals a clear seasonal pattern with periodic spikes indicating peak usage. These peaks often align with summer and winter months, when heating and cooling demands are high. There is noticeable cyclicality, with higher frequency oscillations within each year due to daily and weekly usage habits.

After around 2011, the pattern appears to shift slightly, showing reduced peaks and a marginally lower average consumption level. This suggests a structural change in consumption behavior or grid operations. The graph also highlights heteroskedasticity—fluctuations in variance over time—with more pronounced volatility in the earlier years.

The **Daily Energy Consumption** data, covering the same period with over 6,000 data points, presents similar seasonal trends, including weekly and annual cycles. Compared to the hourly series, daily aggregation reduces noise but still exhibits irregular dips and sharp fluctuations that can obscure long-term patterns and introduce variability in model estimation. The **Weekly Energy Consumption** series, consisting of 342 observations from 2002 through early 2008, offers a balance between temporal resolution and noise reduction. However, the plot shows noticeable variability and irregular seasonal structure, especially during the early years, making it harder to detect consistent peaks and troughs for modeling purposes. **Monthly electricity consumption** series exhibits strong seasonal variation with clear annual peaks, corresponding to higher energy demand during summer and winter months, likely due to heating and cooling usage. The overall trend appears relatively stable, with some fluctuations around 2010–2014, but without a strong upward or downward shift over the entire period.

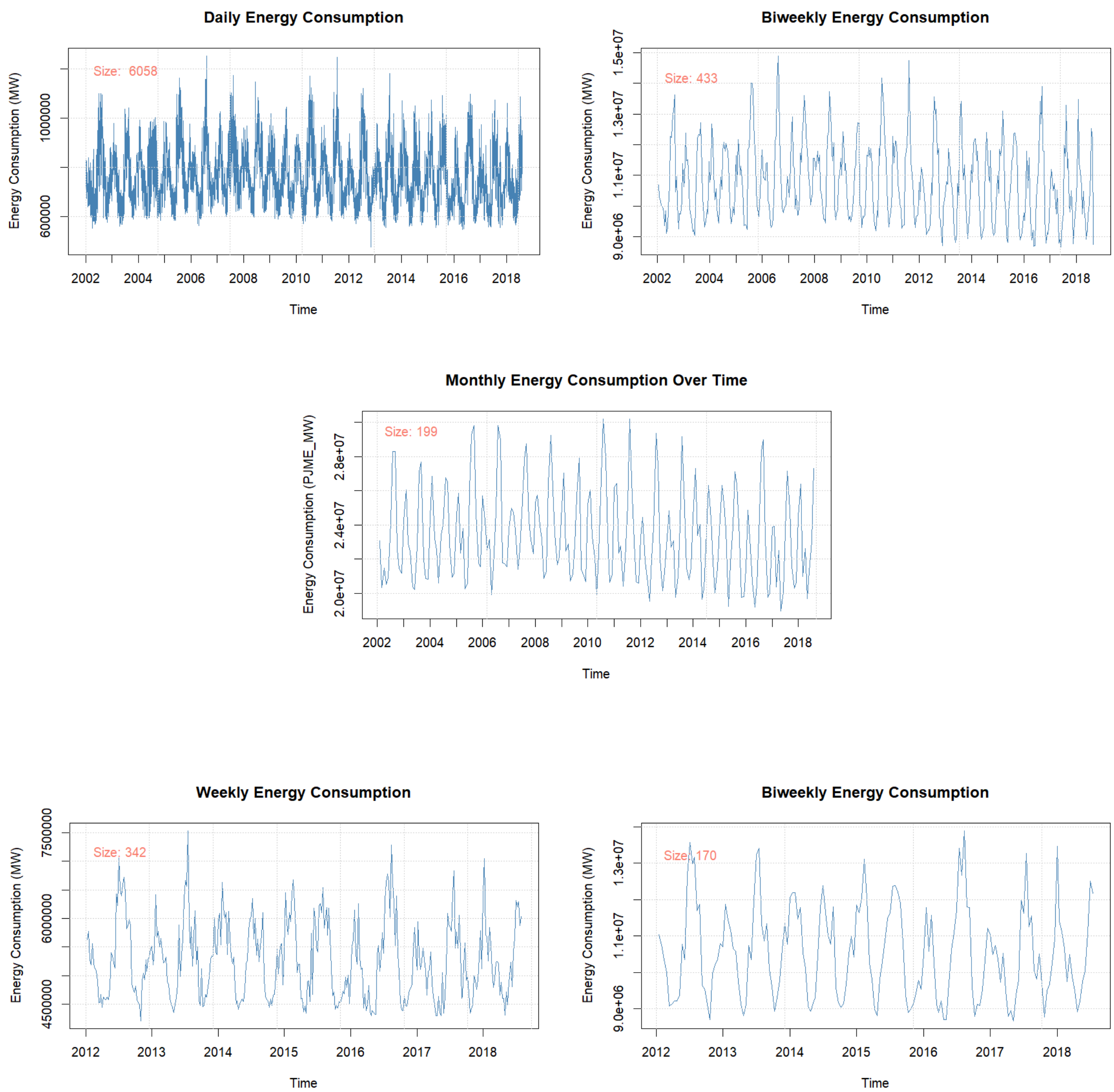

Figure 3. Aggregated electricity consumption in the PJM East region across different time scales: (top-left) daily, (top-right) biweekly (2002–2018), (middle) Monthly(2002-2018)(bottom-left) weekly (2002–2008), and (bottom-right) biweekly (2012–2018)

In contrast, the **Biweekly Energy Consumption** data, which includes 433 points from 2002 to 2018, demonstrates a smoother and more interpretable structure. Aggregating over two-week periods helps expose clearer seasonal trends, with distinct recurring annual peaks and troughs evident in the plot. This made it particularly suitable for capturing medium- to long-term consumption patterns. Finally, the **Post-2011 Biweekly Energy Consumption** subset focuses on the 2012–2018 period, totaling 170 observations. This plot reveals even more consistent seasonal patterns and smoother behavior, likely reflecting changes in usage behavior or improved grid operations in recent years. The series is more homogeneous and structurally stable.

Data Preprocessing

The raw dataset consisted of **145,366 hourly energy consumption observations** recorded between January 1, 2002, and June 1, 2018, measured in megawatts (MW). Daylight Saving Time adjustments were managed by converting timestamps to Coordinated Universal Time (UTC) to ensure temporal consistency. To facilitate modeling and highlight broader patterns, the data was aggregated into multiple time resolutions using a bottom-up approach: **Daily (6,058 entries), Weekly post-2011 (342 entries), Biweekly for the full period (433 entries), Biweekly post-2011 (170 entries)**—excluding incomplete periods—and **Monthly (199 entries)**.

Methodology

Stationarity Testing

To determine the suitability of the energy consumption data for time series modeling, a series of statistical tests were conducted to evaluate stationarity—the property indicating consistent mean and variance over time. The tests applied include the **Augmented Dickey-Fuller (ADF)** test, the **Phillips–Perron (PP)** test, and the **Kwiatkowski–Phillips–Schmidt–Shin (KPSS)** test. These tests assess whether the data follow a stochastic or deterministic trend, which is a critical consideration for selecting and fitting appropriate forecasting models.The ADF and PP tests operate under the null hypothesis that the series possesses a unit root, indicating non-stationarity, while the alternative hypothesis suggests the presence of stationarity. Conversely, the KPSS test assumes stationarity as the null hypothesis and non-stationarity as the alternative. Interpretation of results is based on p-values and test statistics: rejection of the null hypothesis in ADF or PP implies stationarity, whereas rejection of the null in KPSS indicates non-stationarity.

Frequency	**Test**	**Lag order**	**Test Statistic**	**p-value**	**Stationarity**
Hourly	ADF	52	-24.065	0.01	Stationary
	KPSS	24	6.4861	0.01	Non-stationary
	PP	24	-1433.6	0.01	Stationary
Daily	ADF	18	-7.7491	0.01	Stationary
	KPSS	11	0.86469	0.01	Non-stationary
	PP	11	-788.04	0.01	Stationary
Biweekly	ADF	7	-9.51	0.01	Stationary
	KPSS	5	0.29161	0.10	Stationary
	PP	5	-135.66	0.01	Stationary
Monthly	ADF	5	-5.9851	0.01	Stationary
	KPSS	4	0.4485	0.05625	Stationary
	PP	4	-64.248	0.01	Stationary
Weekly (Post-2011)	ADF	6	-7.8954	0.01	Stationary
	KPSS	5	0.049962	0.10	Stationary
	PP	5	-85.679	0.01	Stationary
Biweekly (Post-2012)	ADF	5	-8.571	0.01	Stationary

	KPSS	4	0.04428	0.10	Stationary
	PP	4	-58.024	0.01	Stationary

Table 1. Summary of Stationarity Test Results Across Multiple Temporal Aggregations

The test results identify that **biweekly data beyond 2011** exhibits the **strongest evidence of stationarity** of the aggregated forms and is therefore a better choice for modeling. **Daily data** showed **autocorrelation trends** at a lag of between **2.5 and 1.5 weeks**, while **monthly data** expressed **deterministic seasonality**, which can be modeled by **seasonal decomposition**. Based on this insight, **biweekly series** were selected for **stochastic trend analysis**.

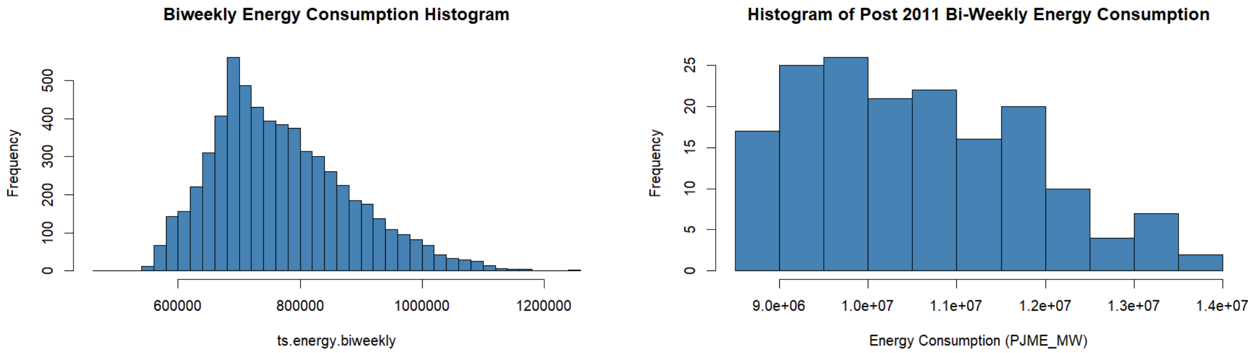

Figure 4. Distribution of Bi-Weekly Energy Consumption (PJME_MW): Full Data vs. Post-2011 Subset

The histogram of bi-weekly energy consumption from 2002 to 2018 shows a distribution with heavier right skew and a higher frequency of extreme consumption values. In contrast, the post-2011 subset (2012–2018) exhibits a more balanced distribution with reduced skewness and fewer extreme values. This suggests greater variance stability and we can see the hypothesis of structural changes in energy consumption behavior after 2011.

Autocorrelation Function analysis

Model Identification for Biweekly Energy Consumption

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x o x  x  x  x
1 x x x x x x x x x o x  x  x  x
2 x x o o o o o o o o o  o  o  o
3 x x o o o o o o o o x  o  o  o
4 x x x o o o o o o o o  o  o  o
5 x x x x o o o o o o x  o  o  o
6 x x o x x x o o o o x  o  o  o
7 x x x x o x o o o o x  o  o  o
```

Figure 5: *Extended Autocorrelation Function (EACF) Table for Biweekly Energy Consumption*

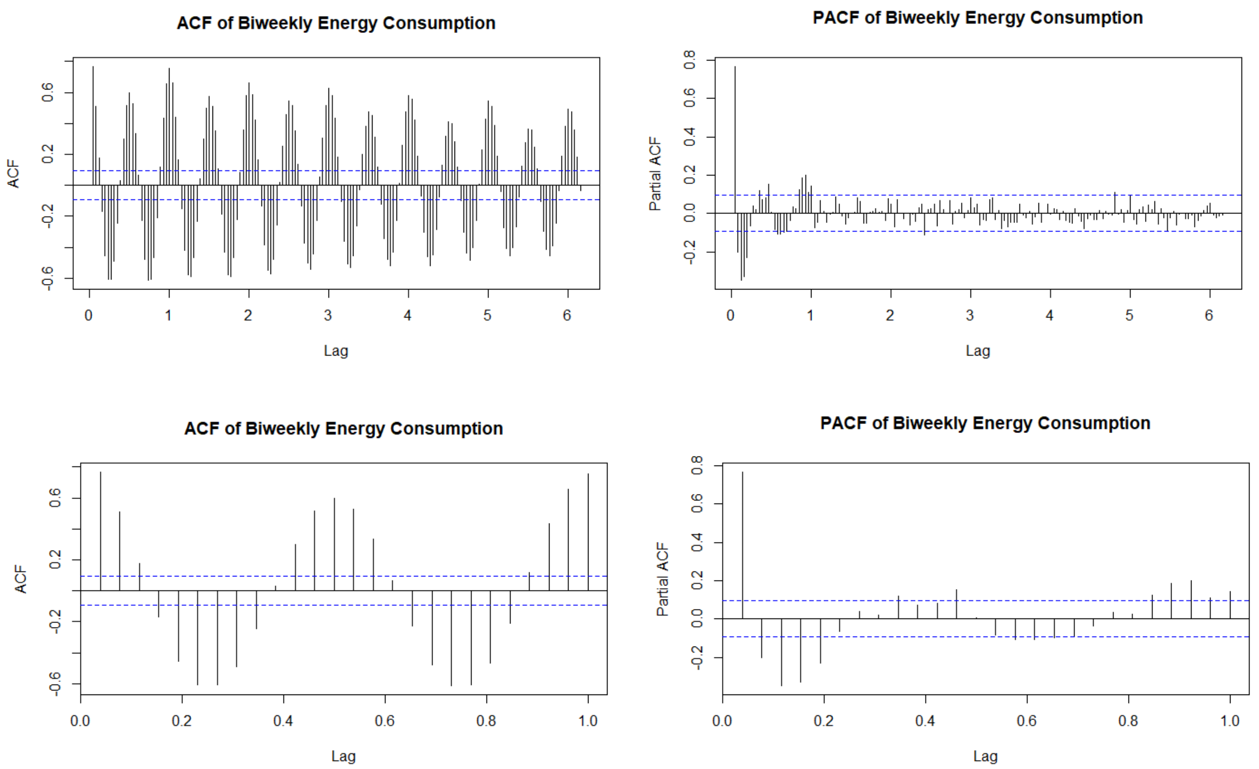

***Figure 5**: Autocorrelation and Partial Autocorrelation Analysis of Biweekly Energy Consumption*
Top: Full-lag ACF and PACF plots (2002–2018), showing strong seasonal patterns.

For the **non-seasonal component**, the ACF displays a slow decay pattern with no clear cutoff, which is indicative of a moving average (MA) process—specifically suggesting an **MA(1)** or **MA(2)** model. The PACF shows a sharp cutoff at lag 1, pointing toward an autoregressive (AR) structure of order 1. The EACF (Extended Autocorrelation Function) supports this, as insignificant values dominate beyond the (2,2) cell, reinforcing the suitability of an **ARMA(2,2)** structure.The significance of AR and MA drops off beyond lag 2, suggesting **AR(2)** and **MA(2)**. For the **seasonal component**, the ACF reveals a slow decay with persistent peaks at seasonal lags, implying the need for seasonal differencing. Meanwhile, the PACF exhibits a sharp cutoff after the first seasonal lag, indicating that a seasonal AR(1) term may adequately capture the seasonal structure.

Model Identification Post-2011 Biweekly Data

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	o	x	x	x	x	x	x	o	x	x	x	x
1	x	x	o	x	x	x	x	x	x	o	x	x	x	x
2	x	x	o	o	o	o	o	x	o	o	o	o	o	o
3	x	o	o	o	o	o	o	o	o	o	o	o	o	o
4	x	x	o	o	o	o	o	o	o	o	o	o	o	o
5	x	x	o	o	o	o	x	o	o	o	o	o	o	o
6	x	o	x	o	o	o	x	o	o	o	o	o	o	o
7	x	x	x	o	x	x	x	o	o	o	o	o	o	o

***Figure 7:** EACF Table for Post-2011 Biweekly Stationary Data*

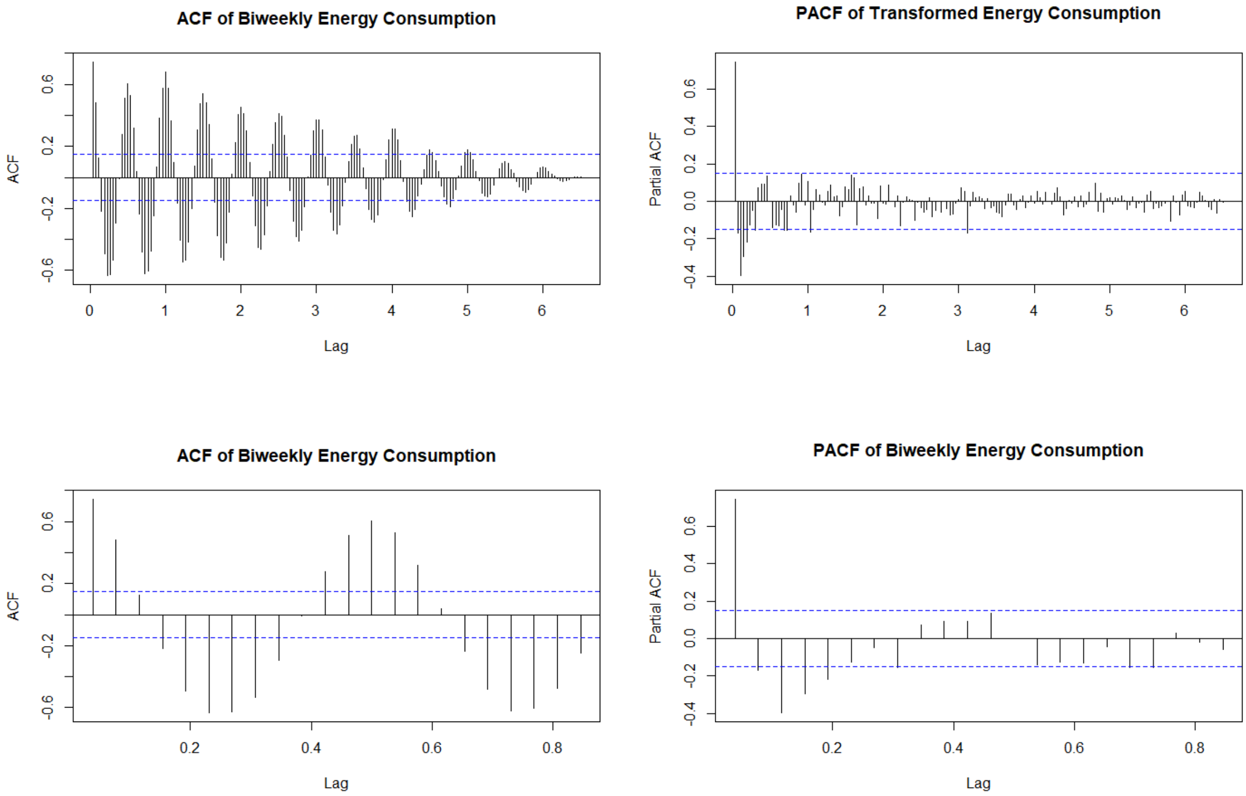

Figure 8: ACF and PACF of Post-2011 Biweekly Energy Consumption Data

For the non-seasonal part of the series, the ACF shows a slow decay without a sharp cutoff, suggesting the presence of a moving average process, with **MA(1)** or **MA(2)** being appropriate candidates. The PACF exhibits a sharp cutoff at lag 1, indicating an autoregressive process of order **AR(1)**. The EACF matrix supports this interpretation, with insignificant entries dominating beyond the (2,2) point, implying that higher-order AR and MA components are unnecessary. Both AR and MA terms beyond lag 2 become insignificant, suggesting that **AR(2)** and **MA(2)** are sufficient for modeling the short-term dependencies.. For the seasonal part, the ACF decays slowly without a sharp cutoff up to lag 14, indicating the presence of seasonality and the potential need for seasonal differencing. The PACF cuts off after lag 1, which supports the use of a seasonal **AR(1)** component.

A key difference in Biweekly Data and Post-2011 Biweekly Data was the **faster decay in ACF for the AR terms**, reflecting stronger stationarity in post-2011 data. Based on this diagnostic analysis, **SARIMA(2, 0, 1)(0, 1, 1)** and **SARIMA(1, 0, 1)(0, 1, 1)** models emerged as strong candidates. To determine the most suitable model, all configurations were evaluated using **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)**, prioritizing both model fit and parsimony for reliable forecasting.

Candidate Modelling

	Model <chr>	AIC <dbl>	BIC <dbl>
3	ARMA(2,0,2)	12831.61	12856.02
7	Auto.ARIMA	12831.61	12856.02
6	ARMA(2,1,2)	12852.81	12873.14
2	ARMA(2,0,1)	12890.38	12910.72
1	ARMA(2,0,0)	12968.22	12984.49
4	ARMA(2,1,0)	12993.40	13005.60
5	ARMA(2,1,1)	12995.34	13011.60

	Model <chr>	AIC <dbl>	BIC <dbl>
2	SARIMA(2,0,2)(1,1,1)	11982.01	12010.05
3	SARIMA(2,0,2)(1,1,2)	11982.73	12014.79
4	SARIMA(2,0,1)(0,1,1)	11983.02	12003.05
5	SARIMA(2,0,1)(1,1,1)	11983.71	12007.75
6	SARIMA(2,0,1)(1,1,2)	11984.29	12012.33
1	SARIMA(2,0,2)(0,1,1)	11984.88	12008.92
7	Auto.ARIMA	12022.14	12042.17

Figure 9*: Seasonal and Non-Seasonal model's AIC and BIC scores*

Based on the AIC and BIC scores, these SARIMA models performed best for the full biweekly dataset:

1. **SARIMA(2,0,1)(0,1,1)[26]** – Best lowest AIC and BIC values.
2. **SARIMA(2,0,2)(1,1,1)[26]** – Competitive performance, offering improved seasonal fit at the cost of added complexity.
3. **SARIMA(2,0,1)(1,1,1)[26]** – Good seasonal component representation with efficient parameterization.

	Model <chr>	AIC <dbl>	BIC <dbl>
8	ARMA(2,0,2)	5064.824	5083.638
9	Auto.ARIMA	5064.824	5083.638
7	ARMA(2,0,1)	5085.821	5101.500
5	ARMA(1,0,2)	5114.198	5129.877
3	ARMA(1,0,0)	5127.421	5136.828
6	ARMA(2,0,0)	5124.838	5137.381
4	ARMA(1,0,1)	5127.187	5139.731
2	ARMA(0,0,2)	5135.915	5148.458
1	ARMA(0,0,1)	5179.251	5188.659

	Model <chr>	AIC <dbl>	BIC <dbl>
7	Auto.ARIMA	4294.834	4306.713
5	SARIMA(2,0,1)(1,1,2)	4290.744	4308.562
6	SARIMA(0,1,2)(0,0,1)	4288.874	4309.662
4	SARIMA(2,0,1)(1,1,1)	4296.010	4310.859
2	SARIMA(2,0,2)(0,1,1)	4291.909	4312.698
3	SARIMA(2,0,2)(1,1,1)	4290.867	4314.625
1	SARIMA(2,0,2)(1,0,2)	4296.876	4314.695

Figure 10*: Post 2011 - Seasonal and Non-Seasonal model's AIC and BIC*

For the biweekly dataset limited to the post-2011 period, SARIMA models were evaluated using AIC and BIC scores to identify optimal configurations. The following models demonstrated the best performance:

- **SARIMA(0,1,2)(0,0,1)** – Top-performing model with strong seasonal adjustment and minimal overfitting.
- **SARIMA(1,0,1)(0,1,1)** – Balanced structure and supported both by performance metrics and auto.arima() recommendation.
- **SARIMA(2,0,1)(1,1,2)** – Captured seasonal nuances effectively, though slightly more complex.

The SARIMA analysis for the **full dataset** indicated that **no differencing was necessary**, suggesting that the biweekly data is **already stationary**. As a result, the top-performing model with differencing (**SARIMA(0,1,2)(0,0,1)**) was **eliminated from consideration**.

For the **post-2011 subset**, the auto.arima() function also selected **SARIMA(1,0,1)(0,1,1)**, reinforcing its robustness and consistency. These findings support the conclusion that **SARIMA(2,0,1)(0,1,1)** and **SARIMA(1,0,1)(0,1,1)** emerged as **strong candidate models**, especially given the **sharper AR component decay** observed post-2011.

Prototype models selected for further analysis:

- **Full data**: SARIMA(2,0,1)[0,1,1][26]
- **Post-2011 data**: SARIMA(1,0,1)(0,1,1)[26]

Residual Analysis for Prototype Models(Full Data)

The energy consumption data, recorded in tens of millions, exhibited high variance. To stabilize variance and improve model performance, a **heavy double-log transformation** was applied prior to modeling.

To validate model adequacy, residual diagnostics were conducted to ensure that residuals behaved like white noise—uncorrelated, normally distributed, and with constant variance.

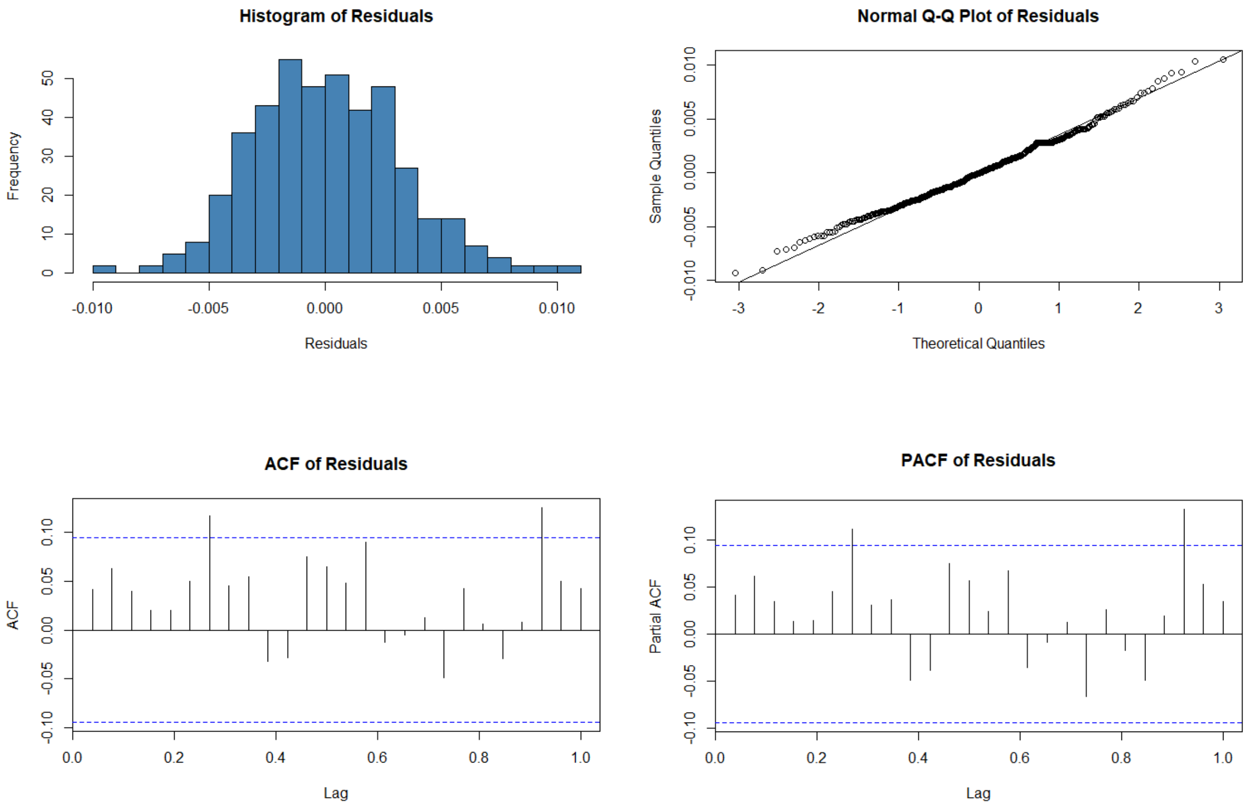

Figure 11: *Residual Diagnostics of the Final SARIMA Model(2,0,1)[0,1,1][26].The top-left (histogram).The (top-right) Q-Q plot.The bottom-left and bottom-right (display the ACF and PACF of residuals).*

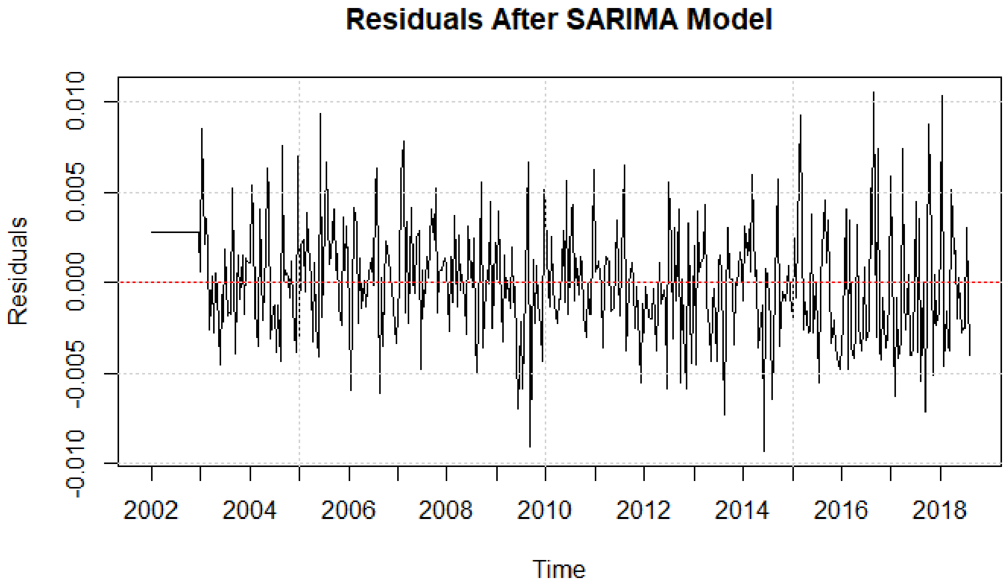

Figure 12: Time Plot of Residuals from the Final SARIMA Model(2,0,1)[0,1,1][26]

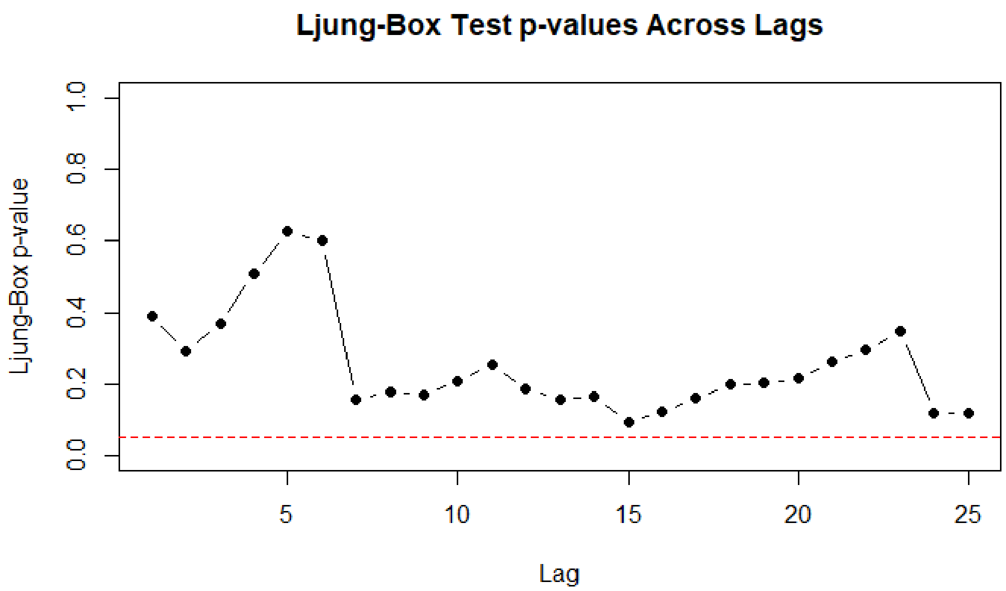

Figure 13: Ljung-Box Test SARIMA(2,0,1)[0,1,1][26]

The residuals from the final SARIMA(2,0,1)[0,1,1][26] model were evaluated for key statistical properties to validate model adequacy. The **residual mean** was found to be close to zero, indicating that the model does not exhibit systematic bias in its predictions. **Homoscedasticity** was generally maintained, though there was evidence of increased variance in the residuals after 2010, suggesting potential heteroskedastic behavior in later years. The **Shapiro-Wilk test** returned a p-value of 0.4306, implying that the residuals are approximately normally distributed, though the result is borderline. Additionally, both **ACF and PACF plots** of residuals showed no strong evidence of autocorrelation, and the **Box-Ljung test** (p-value = 0.2432) supported the conclusion of residual independence.

However, residual behavior post-2010 suggests potential **structural changes** that may not be fully captured by the existing model, warranting further investigation through either decomposition or segmentation-based modeling approaches.

Residual Analysis Analysis Post-2011 Data:

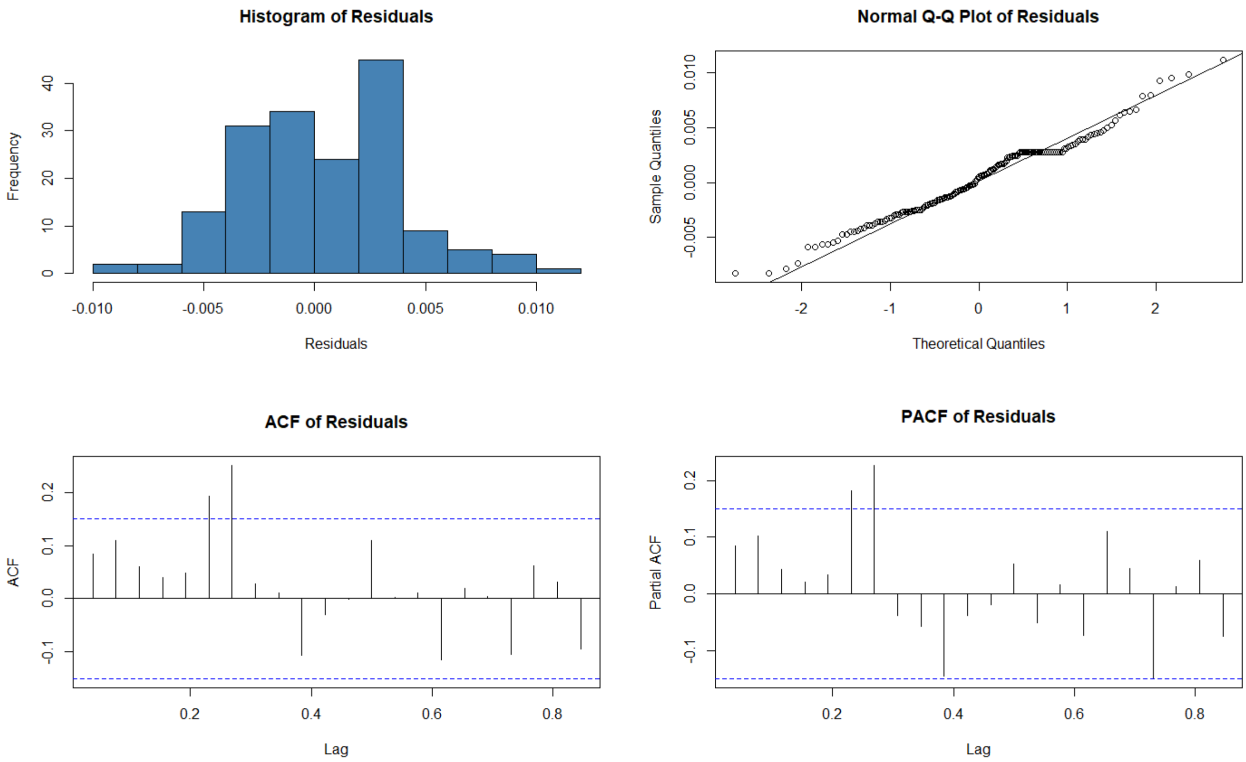

Figure 14: *Residual Diagnostics of the Final SARIMA(1,0,1)(0,1,1)[26].The top-left (histogram).The (top-right) Q-Q plot.The bottom-left and bottom-right (display the ACF and PACF of residuals).*

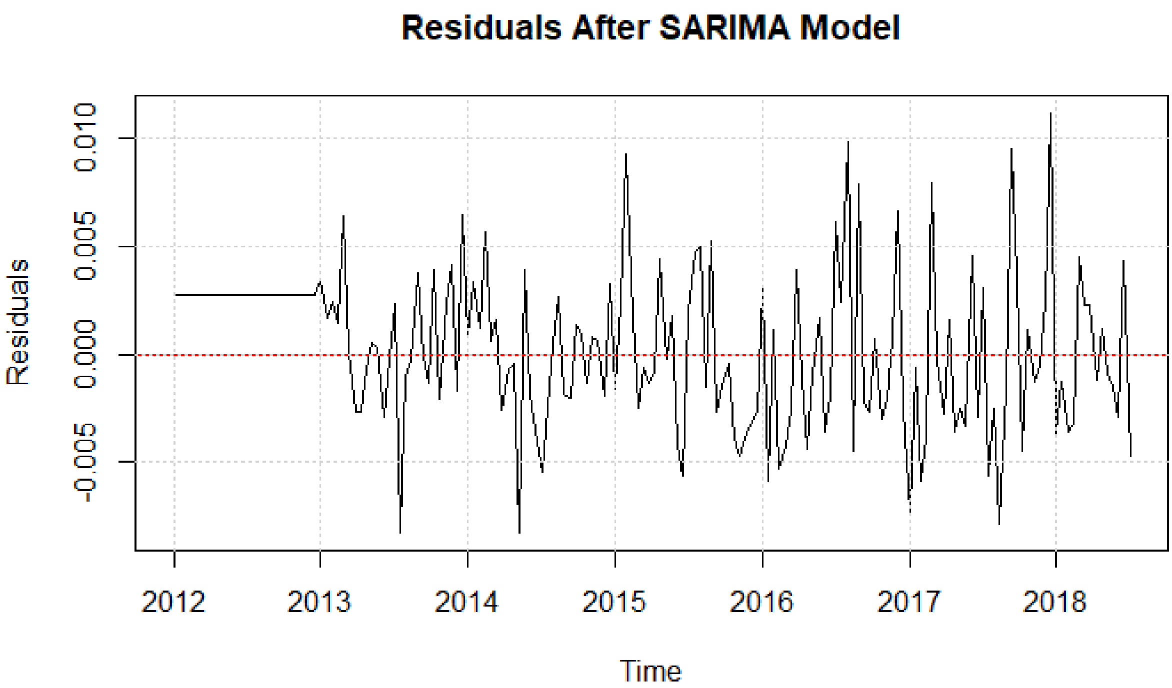

Figure 15: Time Plot of Residuals from the Final SARIMA(1,0,1)(0,1,1)[26]

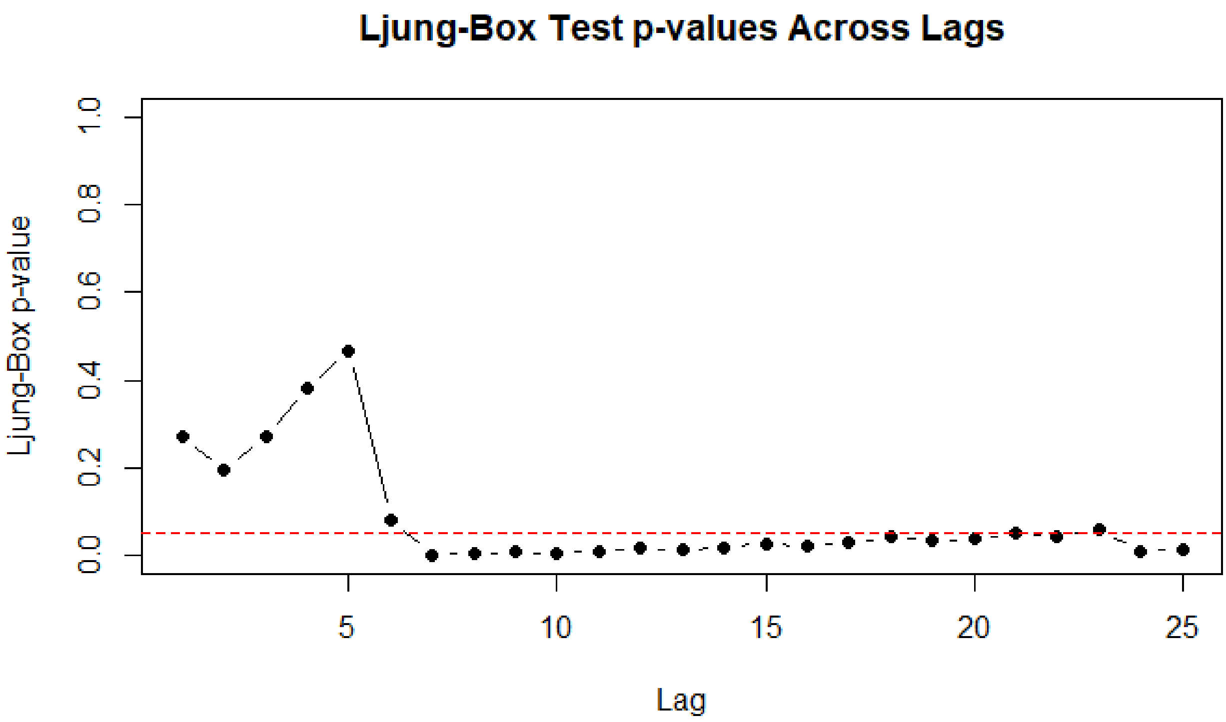

Figure 13: Ljung-Box Test for SARIMA(1,0,1)(0,1,1)[26]

After fitting the SARIMA model(1,0,1)(0,1,1)[26] for post -2011 data, diagnostic checks were performed to evaluate the residual behavior and confirm the model's adequacy. The residuals exhibited a mean close to zero, indicating no systematic prediction bias. Variance appeared generally constant across time, although some noticeable fluctuations were observed after 2010, suggesting mild heteroscedasticity. The normality of residuals was tested using the Shapiro-Wilk test, which produced a borderline p-value of 0.03193, indicating that the residuals are approximately normally distributed but not perfectly so. Autocorrelation analysis via ACF and PACF plots showed some signs of dependence beyond lag 5. The Box-Ljung test returned a p-value of 0.06157, which is marginally above the typical 0.05 threshold—implying residual independence, though results remain borderline.

Double log transformation succeeded in variance stabilization of energy consumption data for facilitating more reliable model fitting and residual analysis. Diagnostic checks validated that the chosen SARIMA models enjoy desirable statistical properties—residuals hover around zero, homoscedastic, more or less normally distributed, and reasonably uncorrelated. The findings suggest that the models are promising candidates for forecasting.

Residual behavior after 2011 indicates the existence of structural breaks in the data, possibly from exogenous policy shocks like the U.S. renewable energy bill enacted in 2011 and PJM's phased-in implementation of renewable energy policies around mid-2013. These events could have introduced new short-term dynamics to the trend of energy consumption while the overall long-term trend is still maintained by the model.

Training Model

To identify the optimal forecasting models, we trained our prototype SARIMA models with differing training sizes and periods. Initial experiments involved varying the training window to assess model strength and prediction ability. The following configurations were identified as giving the best performance metrics when tested and are therefore suitable candidates for both long-term and short-term trend analysis.

Long-Term Trend Analysis
To analyze the long-term structure of electricity consumption, the **SARIMA(2,0,1)(0,1,1)[26]** model was employed. This model was trained on biweekly data covering the period from January 2002 to December 2010, which accounts for approximately 50% of the full dataset (about 208 observations). The testing period included data from 2011 onward, enabling the model to forecast future demand based on historical seasonal and non-seasonal patterns. This configuration captures broader structural trends and long-term cyclic behavior observed over the full span of the dataset.

Short-Term Trend Analysis
To better understand recent dynamics and potential changes in consumption behavior, the **SARIMA(1,0,1)(0,1,1)[26]** model was fitted on the post-2011 biweekly subset. The training period extended from January 2012 to December 2017, encompassing approximately 94% of the post-2011 data (around 160 observations). The model was tested on data from 2018, up to June(post-June 2018 data was missing), which was the latest available segment in the dataset. This short-term model focuses on capturing recent consumption trends with improved responsiveness to post-structural changes, offering more adaptive forecasting performance in the current operational context

Results

Forecasting Results

Biweekly Forecast (2002–2018)

SARIMA model's forecast for biweekly electricity consumption in the PJM East region. The model was trained on data up to 2010 and used to generate forecasts through 2018. Actual energy consumption values are depicted in gray, forecasted values in blue, and the 95% confidence intervals as red dashed lines. The model effectively captures the strong seasonal patterns evident in the data, with forecasted peaks and troughs closely aligning with historical trends. However, the widening confidence intervals toward the end of the forecast horizon reflect increasing uncertainty, which is typical in long-range time series predictions. This visualization underscores the model's capability to forecast complex seasonal demand while also highlighting the importance of regular model updates to maintain predictive accuracy.

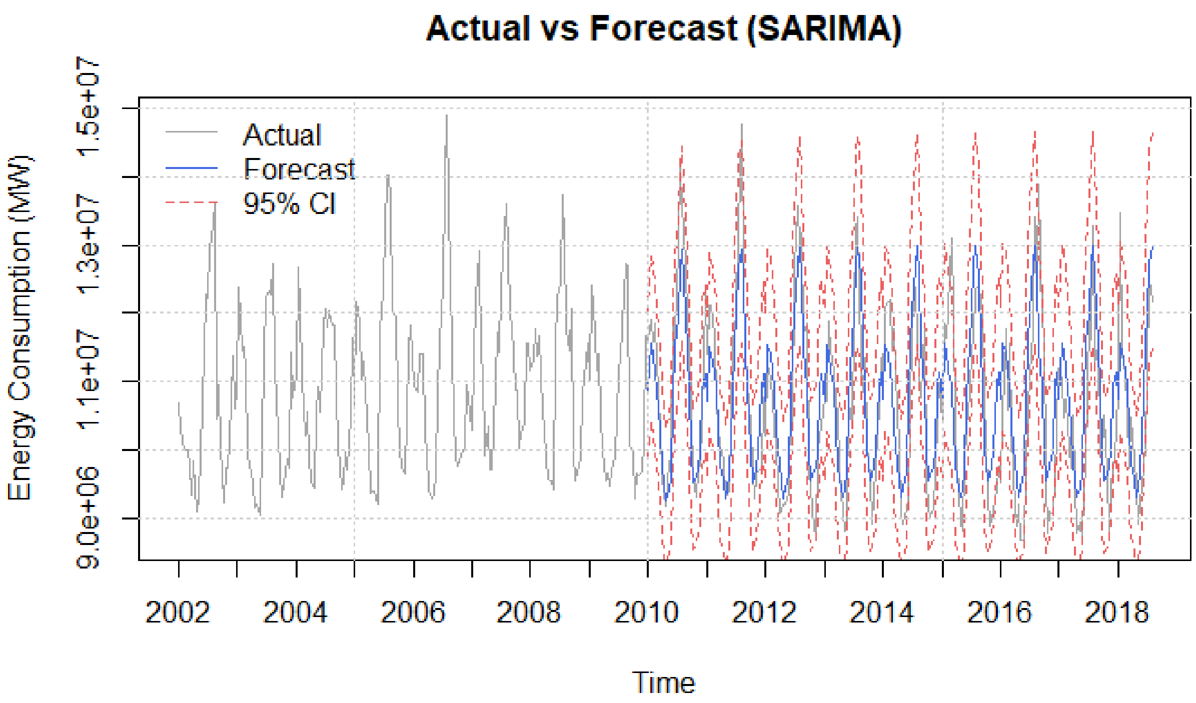

Figure 14 *.SARIMA Forecast on Full Biweekly Data (2002–2018)*

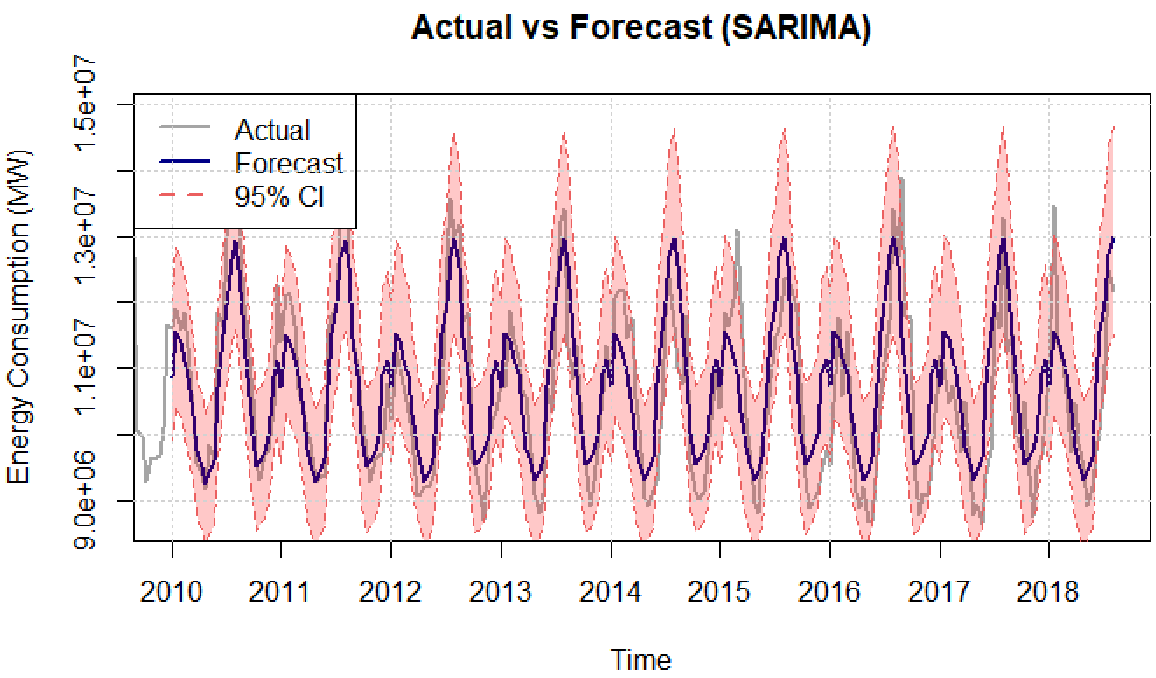

Figure 15*: SARIMA Forecast on Post-2011 Biweekly Subset (2011–2018)*

Post-2011 Biweekly Forecast (2011–2018)

SARIMA model forecast applied to a post-2011 subset of biweekly electricity consumption data for the PJM East region. The actual observed values are plotted in gray, while the model's predictions appear in dark blue, with a shaded pink ribbon representing the 95% confidence interval. This visualization provides a clearer and more focused view of recent consumption trends, capturing the recurring seasonal peaks and troughs effectively. The close alignment between the forecasted and actual values indicates the model's strong performance in short- to mid-term forecasting. The shaded confidence region helps highlight areas of higher uncertainty, offering valuable insights into the model's reliability over time.

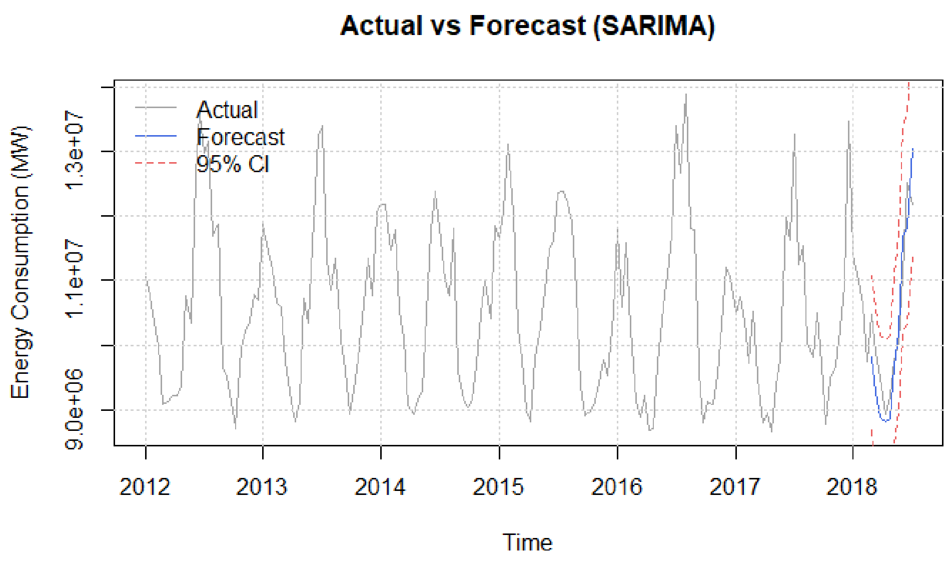

Figure 16. Final SARIMA Forecast on Post-2011 Biweekly Data (Last Observations)

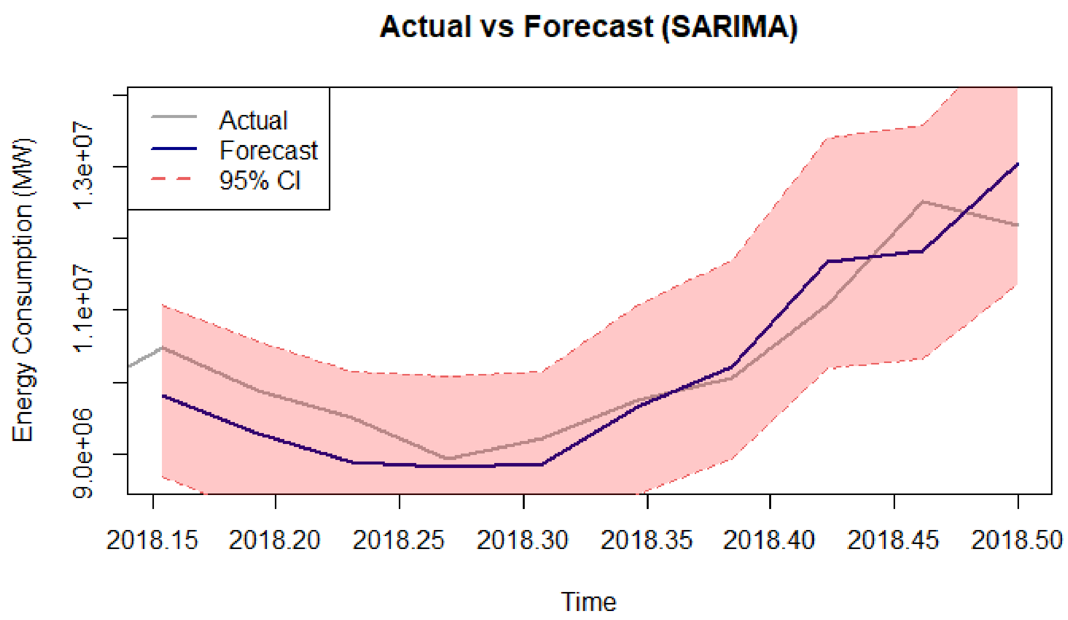

Figure 17. Final SARIMA Forecast with Shaded Confidence Interval (Post-2011 Biweekly Data)

Final SARIMA model applied to the most recent segment of post-2011 biweekly electricity consumption data in the PJM East region. The forecast focuses on predicting the last few observations, making it particularly useful for evaluating short-term accuracy. The gray curve represents the actual observed data, while the blue line shows the SARIMA predictions and the red dashed lines indicate the 95% confidence bounds. This plot demonstrates the model's strong fit in the final window, effectively capturing the recent upswing in consumption levels. The relatively narrow forecast intervals suggest high model confidence, which is critical for short-term operational planning and energy resource management.

Short-Term Forecast with Shaded Confidence Interval

Final SARIMA model forecast applied to the most recent segment of the post-2011 biweekly electricity consumption data in the PJM East region. The plot presents actual values in gray, forecasted values in dark blue, and a 95% confidence interval shaded in pink with dashed red borders. This enhanced visualization provides a more intuitive representation of uncertainty, allowing for a clearer understanding of the forecast's reliability. The model effectively captures the upward trend in recent consumption levels and the widening of the shaded region over time indicates growing uncertainty as the forecast horizon extends. This visualization supports short-term energy planning by providing both accurate predictions and a transparent view of the expected variability in demand.

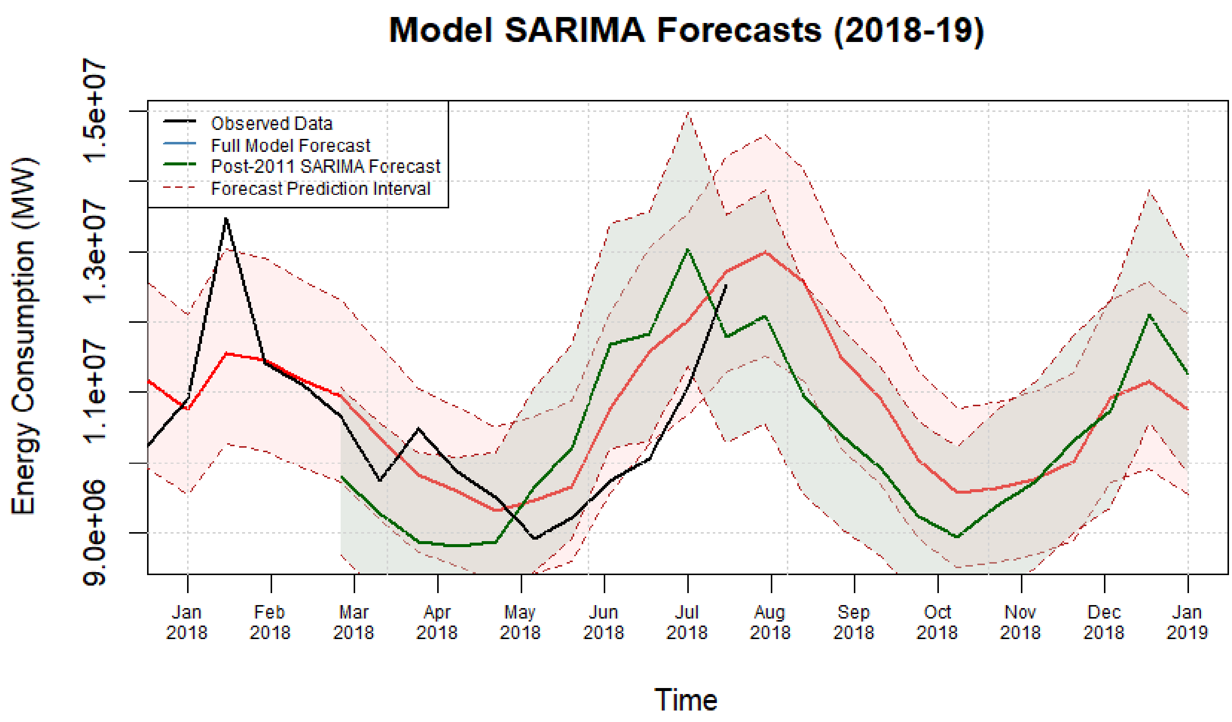

***Figure 18**. Comparison of Full and Post-2011 SARIMA Forecasts (2017–2018)*

The forecasting performance of two SARIMA models on biweekly electricity consumption data from 2017 to 2018 in the PJM East region. The gray line represents the observed data, the blue line shows the forecast from the full-data SARIMA model (2002–2018), and the green line displays the forecast based on the post-2011 subset. The pink shaded region corresponds to the confidence interval from the full model, while the green-shaded region highlights the interval from the post-2011 forecast. Both models capture the seasonal fluctuations effectively, but the post-2011 model demonstrates a tighter alignment with the actual data in the final months, suggesting enhanced responsiveness to recent consumption behavior. This side-by-side comparison helps assess the benefits of updating training windows for short-term forecasting precision.

Evaluation Metrics

Metric	Post-2011 Biweekly Data	Full Dataset
Mean Absolute Error (MAE)	474,926.190	585,420.028
Root Mean Squared Error (RMSE)	544,114.244	739,120.764
Mean Absolute Percentage Error (MAPE)	4.451%	5.492%
Coefficient of Determination (R-squared)	0.778	0.681

The performance measures evidently indicate that the **short-term model**, learned on **post-2011 biweekly data**, is **more predictive and accurate** compared to the l**ong-term model** with considerably **lower MAE, RMSE, and MAPE, and a better R-squared measure**. This indicates that **current consumption behavior** is explained better by the **short-term SARIMA model**.

Still, though, **visual examination** of the forecast plots **after mid-2018—the time when data was unavailable**—proves that the **long-run model** is still the one that **best catches broad trends**, presumably because it has the **longer training period (2002–2010, with 208 observations)**. The short-run model, however, trained on 2**012–2017 data (160 observations)** and is **more in line** with **recent structural shifts** in energy demand, such as **policy initiatives like the 2011 U.S. renewable energy bill** and **PJM's adoption of renewables after 2013.**

Taken together, these results imply that a synthesis of both models or an **optimization of the training window** may result in **improved forecast accuracy**, particularly for both **long-term seasonality and recent trend changes in energy use behavior.**

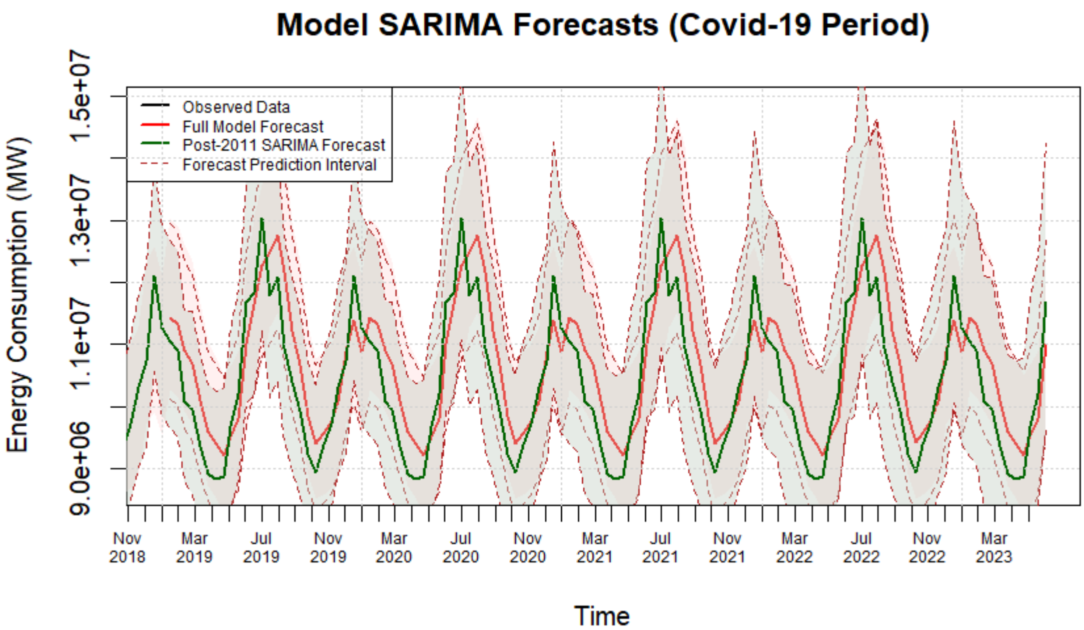

Figure 19. SARIMA Forecast Using Weighted Predictions (2012–2025)

Using a weighted SARIMA model approach. Historical observed data from 2012 to 2018 is shown in gray, and future forecasts from 2019 to 2025 are represented in blue, with the 95% prediction intervals shown as red dashed lines. The weighted forecast was constructed by combining two SARIMA model outputs: the full model (2002–2018) and the post-2011 model, assigning respective weights of 0.45 and 0.55. This hybrid approach aims to balance the long-term trend-capturing capability of the full model with the recency-sensitive behavior of the model.

Further Work

For further work, several improvements can be explored to enhance model performance and accuracy of predictions:

- **Model Optimization:** Use techniques like **GridSearchCV** to optimize SARIMA hyperparameters for better model selection.
- **Training Window Selection:** Experiment with optimal training sizes, balancing between long-term trends and recent shifts.
- **Trend Resolution:** Consider **quarter-yearly (66 observations)** or **four-monthly (~50 observations)** trends as a possible alternative since **yearly (16 obs)** and **half-yearly (33 obs)** resolution appears too **low for capturing deterministic trends**.
- **Stationarity Considerations:** Monthly trends were found to be non-stationary without any perceivable stochastic trend, and therefore not so favorable for time series modeling.
- **Environmental Factors:** Due to increasing global warming, **seasonal drift pattern** changes can emerge, especially in **quarterly trends**, that represent possible changes in **weekly energy cycles** to be incorporated in ensuing models.

Conclusion

Our comparison of long- and short-term SARIMA models to forecast fortnightly U.S. energy consumption showed a high R-squared of 0.681 for the long-term model trained (2002–2010) and an even higher R-squared of 0.778 for the short-term model trained (2012–2017). Both models performed well in capturing underlying trends and were tested on post-2018 data, including the COVID-19 period, which introduced changes similar to those post-2011 with regulation and environmental changes.

To improve forecasting, we could fine-tune hyperparameters using grid search, improve training window selection, and study trend behavior on the quarterly level to better reflect shifting seasonal patterns. And hey—**a close 80% accuracy** for energy demand forecasting during a **pandemic and climate change? Not bad.**

Literature review

1. Deb. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable & Sustainable Energy Reviews.*, *74*. https://doi.org/10.1016/j.rser.2017.02.085
2. Islam. (2013). Financial development and energy consumption nexus in Malaysia: a multivariate time series analysis. *Economic Modelling.*, *30*. https://doi.org/10.1016/j.econmod.2012.09.033
3. Abdel-Aal. (1997). Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis. *Energy*, *22*(11). https://doi.org/10.1016/S0360-5442(97)00032-7

Appendices