Forecasting Electricity Consumption & Losses For Residential, Commercial, and Industrial Sectors

04/26/2023

Lynn Wu, Haru Koga, Miaojun Pang

Introduction and Objectives

Energy consumption is a critical indicator for policymakers, industry stakeholders, and researchers to develop effective energy policies and solutions for an effective use of electricity and power loss reduction. However, accurately forecasting and analyzing energy consumption data can be challenging to obtain, as most data are preliminary and subject to revision.

In this project, we will use time series analysis in R to forecast and analyze energy consumption data for three sectors: residential, commercial, and industrial, in the United States, from February 1973 to November 2022. Motivated by the importance of accurate and reliable energy consumption data, this project aims to contribute to a better understanding of energy trends and help inform policy and decision-making in the energy sector. By forecasting and analyzing the U.S. Energy Information Administration's data on residential, commercial, and industrial uses, we hope to shed light on the drivers of energy consumption in these sectors and identify potential opportunities for reducing energy consumption and improving energy efficiency.

Moreover, in the time range of February 1973 to November 2022, the United States has experienced a number of political and natural disturbances, such as recessions, pandemics, and natural disasters. Those external factors frequently have an impact on electricity consumption. Then this project aims to explore the impacts of external disturbances on energy consumption in each sector. Forecasting the energy consumption could help ensure the reliability of the energy needs in each sector amid the turmoil and further seek the potential use of emerging technologies such as energy storage in the respective sector.

Objectives The objective of this study is to analyze:

(1) Effective use of electricity in the residential, commercial, and industrial sectors while considering the energy loss data

(2) How energy consumption gets affected by natural and political disturbances in each sector

(3) Potential in the use of energy technologies, such as electricity storage, by the sectors

Dataset Information

The Energy Information Administration (EIA) is the best source of energy data in the United States. It puts out detailed monthly and annual reports on energy production, consumption, stocks, trade, and the effects of different energy sources on the environment. The dataset on energy consumption by the three sectors of residential, commercial, and industrial, covering the period from February 1973 to November 2022, has been selected for analysis for the purposes of this study. This data set has five metrics for each sector, including total energy consumption, end-use energy consumption, energy losses in the electrical system, and primary energy consumption. In order to better understand the patterns of electricity use and energy efficiency in the three sectors, our study focuses on two variables: primary energy consumption and energy losses.

Most energy sources, including coal, natural gas, petroleum, geothermal, solar, and biomass, are included in primary energy consumption. Overall, the following steps were applied to our dataset:

(1) Download the dataset from EIA.

(2) Import the dataset to R and delete the unit column.

(3) Keep two variables for three sectors.

(4) Convert the dataset from characters to numerical.

(5) Round all numbers to three decimal places.

(6) Create a date's data frame.

(7) Convert the original dataset into time series data for forecasting and summarize it.

(8) Calculate the mean, standard deviation, maximum values, minimum values, and ranges for primary energy consumption and electrical losses in three sectors.

(9) Make a summary statistics table.

(10) Utilize ggplot() to compare three sectors using two initial plots that are variable-classified.

```
##  Residential.Primary.Energy Residential.Energy.Losses Commercial.Primary.Energy
##  Min.   : 192.3             Min.   : 325.3            Min.   :165.3
##  1st Qu.: 301.0             1st Qu.: 512.7            1st Qu.:212.9
##  Median : 484.6             Median : 651.1            Median :296.7
##  Mean   : 584.3             Mean   : 662.3            Mean   :345.1
##  3rd Qu.: 874.4             3rd Qu.: 782.7            3rd Qu.:476.7
##  Max.   :1488.0             Max.   :1144.8            Max.   :703.6
##  Commercial.Energy.Losses Industrial.Primary.Energy Industrial.Energy.Losses
##  Min.   :258.2            Min.   :1466              Min.   :405.0
##  1st Qu.:445.3            1st Qu.:1712              1st Qu.:526.3
##  Median :639.7            Median :1808              Median :575.4
##  Mean   :605.4            Mean   :1806              Mean   :576.4
##  3rd Qu.:743.9            3rd Qu.:1895              3rd Qu.:621.4
##  Max.   :982.2            Max.   :2267              Max.   :776.8
```

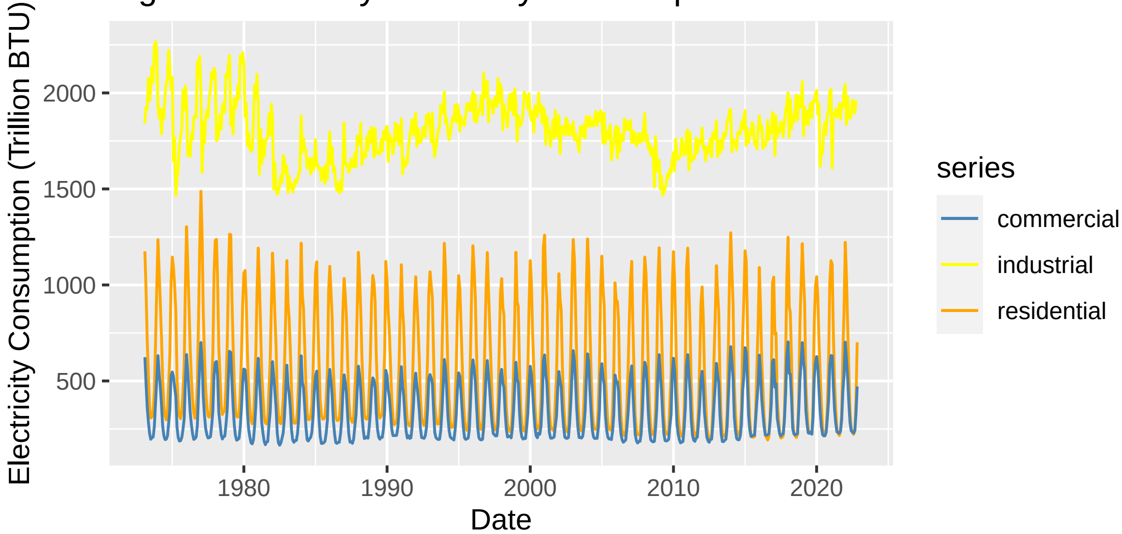
The following table and figures are visualized here:

Table 1: Table 1. Summary Statistics Table used with ts() function

Variable	Mean	StandardDeviation	Max	Min	Range
Residential_Sector_Consumption	584.60	321.54	1488.0	192.3	1295.71
Residential_Sector_Losses	662.45	185.03	1144.8	325.3	819.46
Commercial_Sector_Consumption	345.14	146.90	703.6	165.3	538.31
Commercial_Sector_Losses	605.47	178.14	982.2	258.2	723.97
Industrial_Sector_Consumption	1806.54	146.71	2267.0	1466.0	800.93
Industrial_Sector_Losses	576.51	68.86	776.8	405.0	371.76

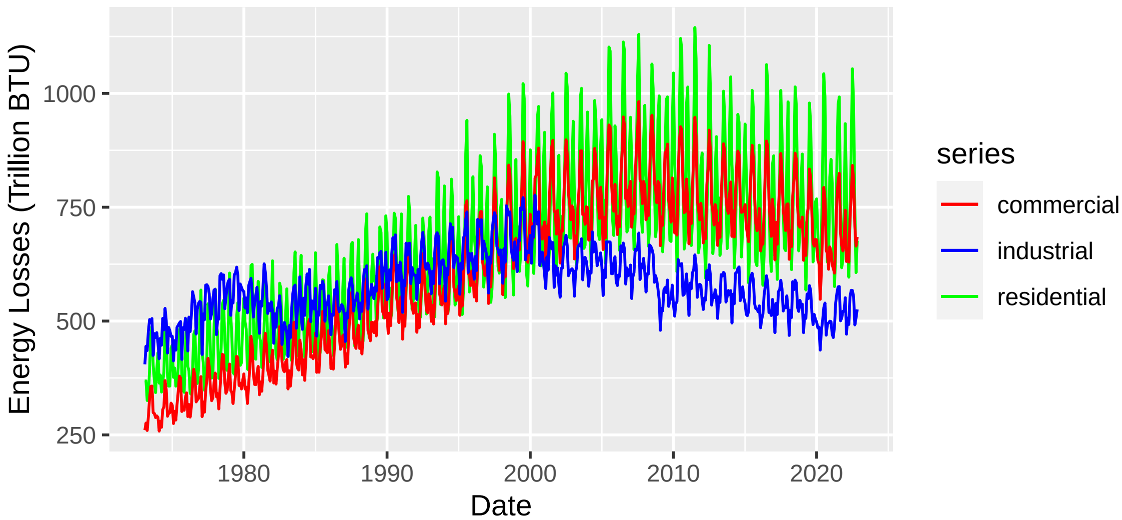
The industrial sector has the highest average consumption (1806.54 trillion BTU) when comparing primary electrical consumption across the three sectors (Table 1). Despite the similar average energy losses across all three sectors, the residential sector nevertheless suffers from the highest energy losses as a result of overcoming transmission line resistance. The residential sector has the highest primary electricity consumption, whereas the industrial sector has the lowest energy losses, according to a comparison of standard deviations that we also conducted. The ranges show the fundamental seasonal shifts over the last few decades. For primary electricity consumption, the residential sector has the strongest seasonality, while the industrial sector exhibits the least seasonality.

Figure 1. Primary Electricity Consumption of Three Sectors



The highest primary energy consumption in Figure 1 is in the industrial sector. Although there is the most seasonality in the residential sector, the commercial sector uses the least primary energy. Strong seasonality is typically more pronounced in the residential and industrial sectors in the summer and winter but less pronounced in the spring and fall. In general, residential and commercial sector patterns are nearly stable; however, exogenous influences throughout time cause the industrial sector to experience increasing and declining trends.

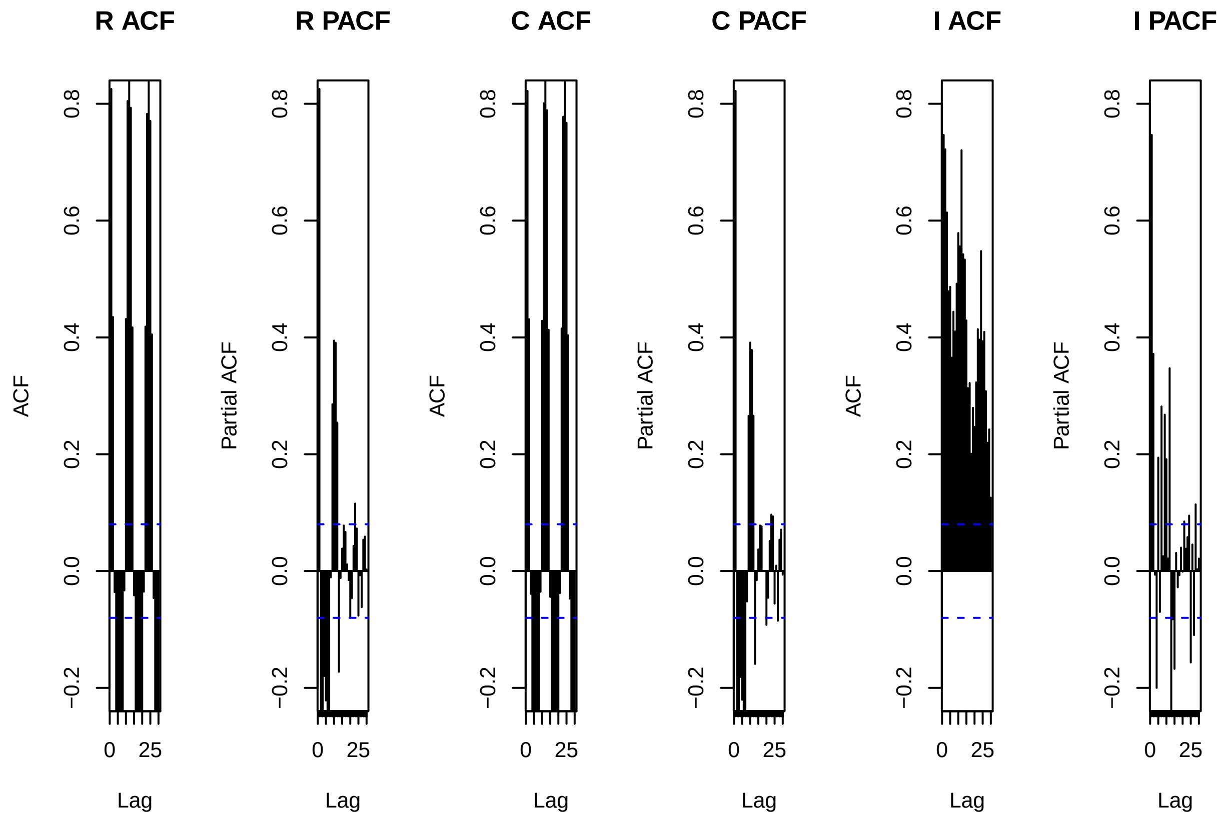
Figure 2. Energy Losses of Three Sectors



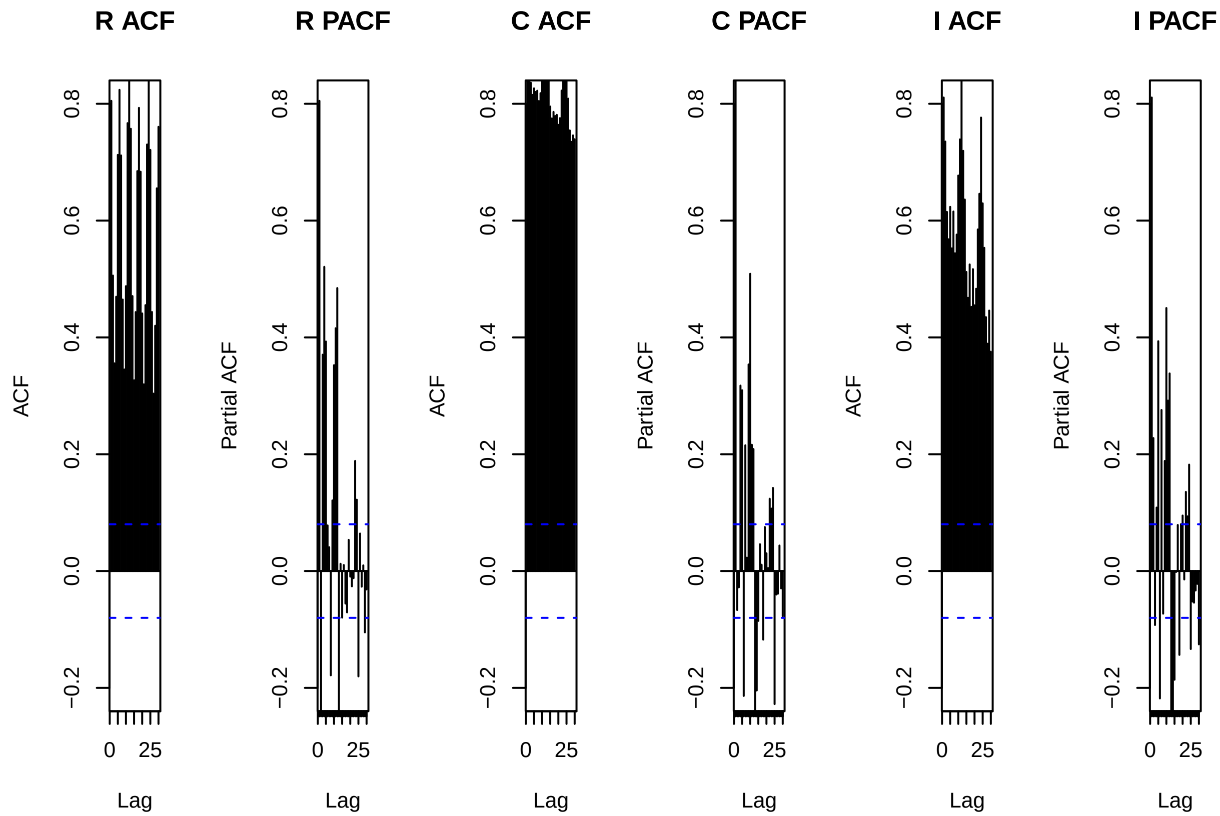
In contrast to the commercial and residential sectors in Figure 2, the electrical losses in the industrial sectors are decreasing. In three sectors, electricity losses are highest in the residential sector while being lowest in the industrial sector. Three industries saw an increasing tendency in energy losses prior to 2000, but after that year, they saw decreasing trends.

We initially build three sub-data frames containing the two variables primary electricity consumption and energy losses for each of the three sectors. All data were converted to time series data. To compute five one seasonality models, forecast consumption and losses from 2017 to 2022, and determine the most effective forecasting models for the next 10 years, all three sectors also contain one testing dataset and one training dataset between 1973 and 2017.

Before running a time series forecast model and selecting models for time series forecasting, we examine ACF and PACF plots of the original, un-transformed time series (shown below). For the following plots, we use R, C, and I to stand for residential, commericial, and industrial sectors, respectively.



For our first variable–primary energy consumption, both ACF and PACF not only show strong seasonality but also non-stationary due to slow decays.



For our second variable–energy loss, both ACF and PACF still show significant values beyond thresholds, especially at lags=12 and 24.

Analysis (Methods and Models)

For our forecast training for all the three sectors, we run five models: Arithmetic Mean (MEAN), Seasonal Naive (SNAIVE), Seasonal ARIMA (SARIMA), State Space Exponential Smoothing (SSES), and Basic Structure Model (BSM). We created a training dataset for each sector which has observations from February 1973 to November 2017, and the forecast testing window is 5 years from November 2017 to November 2022. We fit the single-seasonality time series data from each sector to the above five models for the primary energy consumption and energy losses 5-year forecasting. Having the five models run, we identify the best model from the result values of RMSE from an accuracy test. Finally, the identified best models are used to forecast the next ten years (2022-2032) of primary energy consumption and energy losses in the three sectors. The results and discussion will be given below by sectors.

(1) The arithmetic mean model is a simple forecasting method that assumes that the future values of a time series will be equal to the average of past values. The code implements this model to forecast primary energy consumption and energy losses.

(2) The seasonal naive model was implemented to forecast primary energy consumption and energy losses, and the plot shows that there is a significant seasonal trend in the data. The seasonal naive model takes into account the seasonality in the data and forecasts future values based on the same seasonal pattern observed in the historical data.

(3) The plots of the SARIMA forecasts for both primary energy consumption and energy losses show a continuation of the recurring pattern observed in the historical data, indicating that the SARIMA

model is capturing the seasonality in the data. However, it is important to assess the accuracy of the forecasts by checking the residuals and other model diagnostics before making any conclusions about the effectiveness of the model.

(4) We were fitting state space models using exponential smoothing to the original (seasonal) time series data for primary energy consumption and energy losses. The SSES model uses exponential smoothing to capture the seasonality and trend in the data.

(5) Model 5 fits state space models using StructTS() to the original (seasonal) time series data for primary energy consumption and energy losses. The StructTS() function can estimate various state space models with different types of time-varying components, including trend, seasonality, and cycle. The forecasts generated by the models are plotted for both primary energy consumption and energy losses.

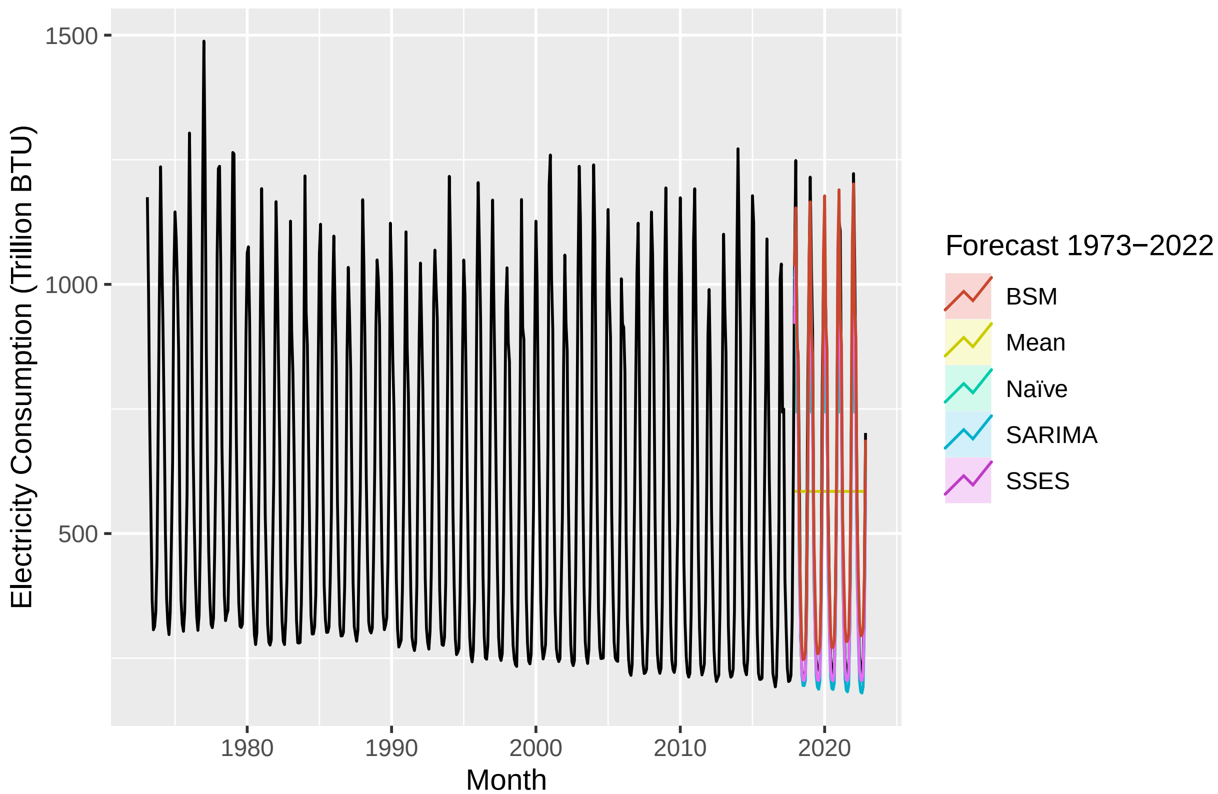
Residential sector

We begin with the residential sector.

These two following plots of primary energy consumption for the residential sector show a strong relationship and seasonality, with the BSM model performing the best among the five models. The first plot shows the original data from 1973 to 2022, along with the forecasts generated by the five models. The second plot only shows the last three years of the data (2019-2022), along with the same forecasts generated by the five models.

In both plots, the BSM model forecast shows a close fit to the observed data, capturing the strong seasonality and trend in the data. The other models also capture the seasonality to varying degrees, but they may not perform as well in capturing the trend. The MEAN model and the seasonal naive model capture the seasonality but assume a constant mean, which may not be appropriate for data with a trend. The SARIMA and SSES models can capture both seasonality and trend, but they may require more parameter tuning and may be more complex than the BSM model.

Primary Energy Consumption Plots (1973-2022 vs. 2017-2022)



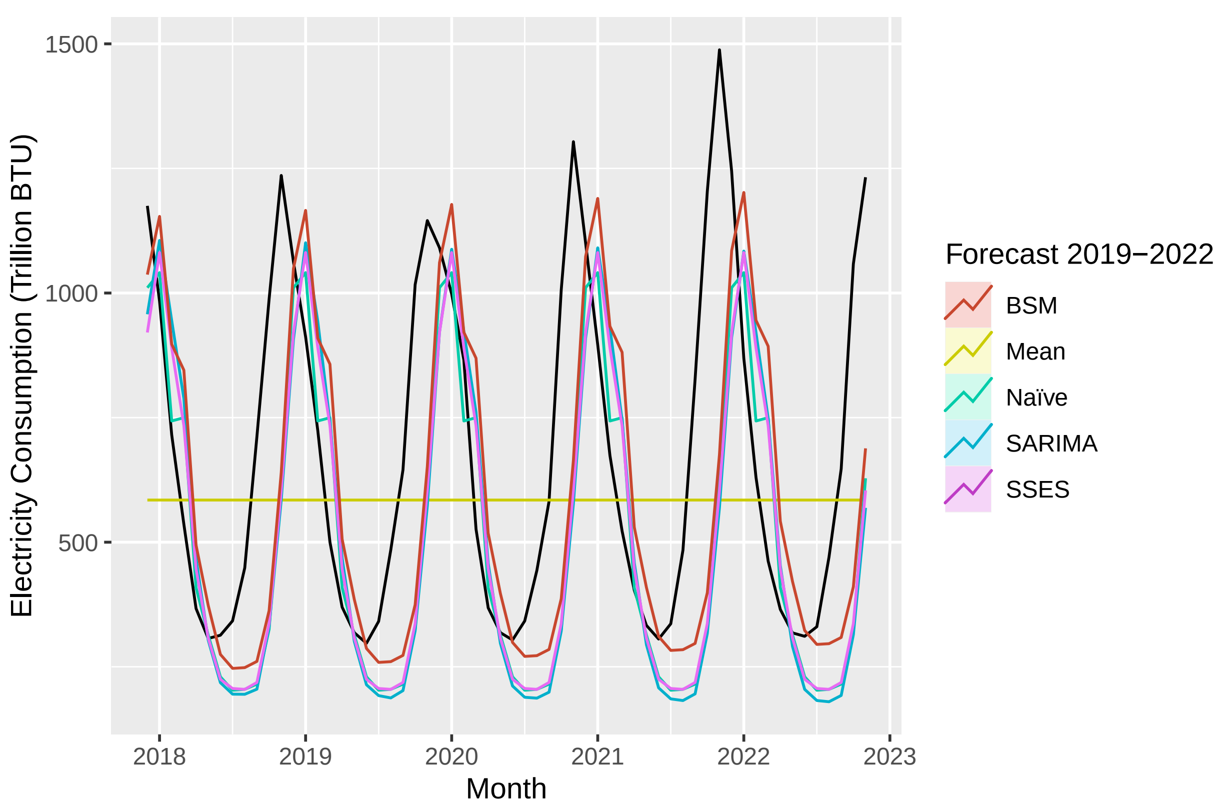
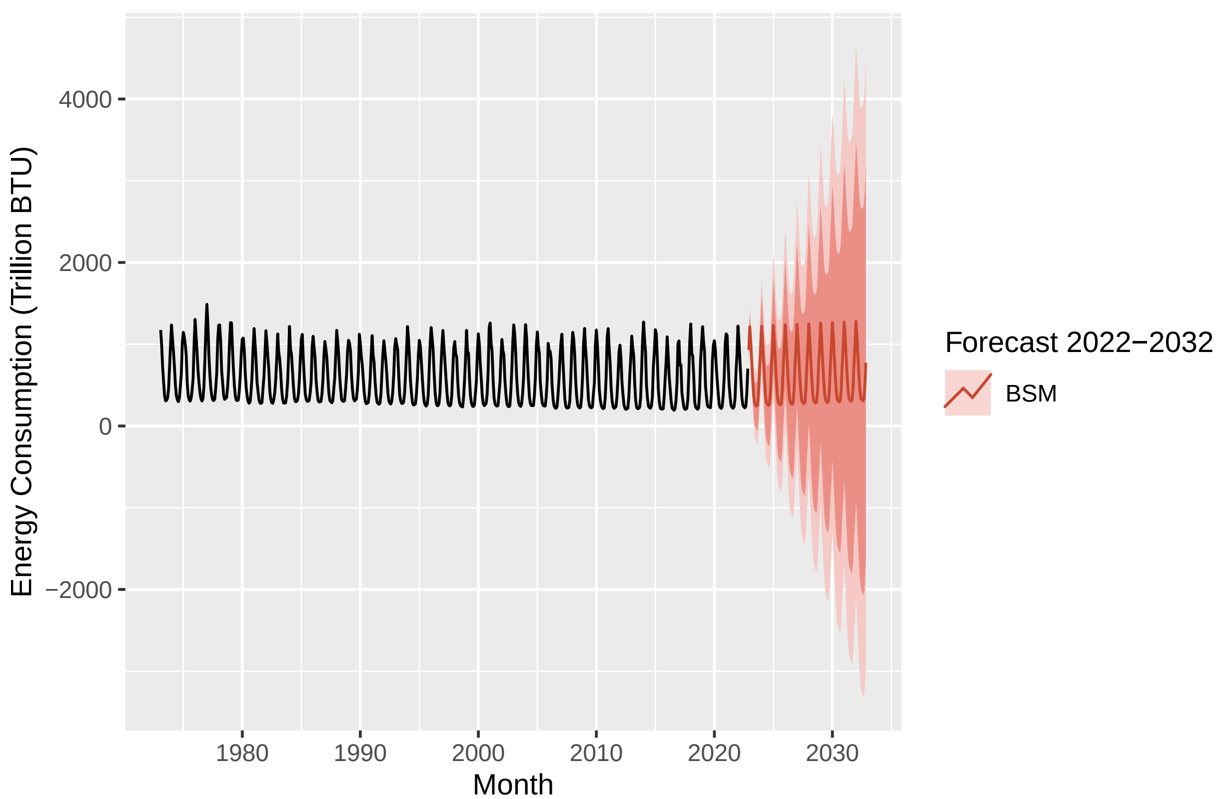


Table 2: Forecast Accuracy for Residential Primary Energy Consumption's Seasonal Data

	ME	RMSE	MAE	MPE	MAPE
MEAN	-12.65689	338.25702	305.34068	-47.89962	77.25445
SNAIVE	65.83145	110.51893	74.43798	10.53433	11.76199
SARIMA	60.00183	82.83334	66.79698	12.00626	12.88058
SSES	57.32929	83.87768	61.99349	9.46394	10.14529
BSM	-21.59000	75.18453	60.06963	-8.98415	13.54302

After checking five models accuracy, we created a data frame to compare the performance metrics of the five models for primary energy consumption. In this case, the best model by RMSE is the state space model using the basic structural model (BSM). Next, we created a formatted table of the forecast accuracy for the models applied to the primary energy consumption's seasonal data.



This plot shows the forecasted primary energy consumption for the entire 50-year period (1973-2032), with the BSM model generating a 10-year forecast (120 periods) for the residential sector. The forecast indicates a continuation of the strong seasonality and trend observed in the historical data, with a slight increase in consumption towards the end of the forecast horizon.

Energy Loss Plots (1973-2022 vs. 2017-2022)

These following two plots of energy losses for the residential sector show a strong relationship and seasonality, and the SSES model performs the best among the five models. The first plot shows the original data from 1973 to 2022, along with the forecasts generated by the five models. The second plot only shows the last year of the data (2017-2022), along with the same forecasts generated by the five models.

In both plots, the SSES model forecast shows a close fit to the observed data, capturing the strong seasonality and trend in the data. The other models also capture the seasonality to varying degrees, but they may not perform as well in capturing the trend. The MEAN model and the seasonal naive model capture the seasonality but assume a constant mean, which may not be appropriate for data with a trend. The SARIMA and BSM models can capture both seasonality and trend, but they may require more parameter tuning and may be more complex than the SSES model.

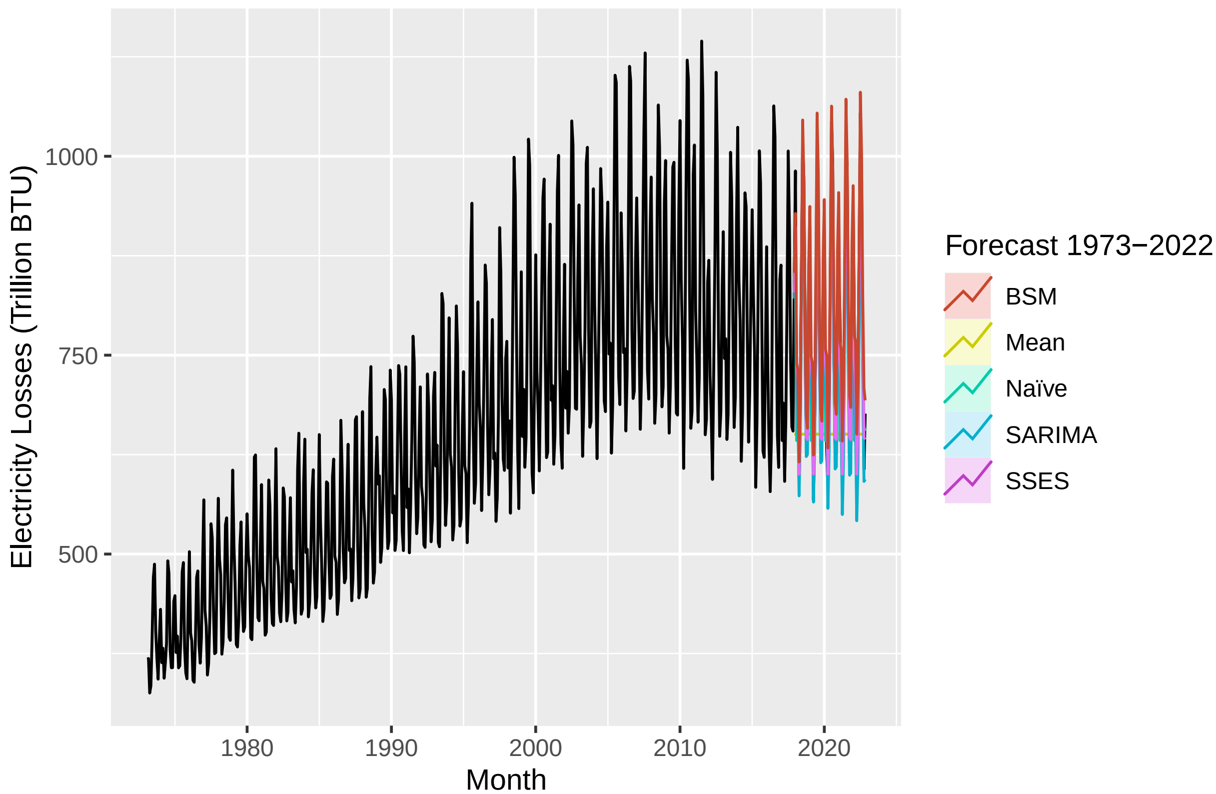
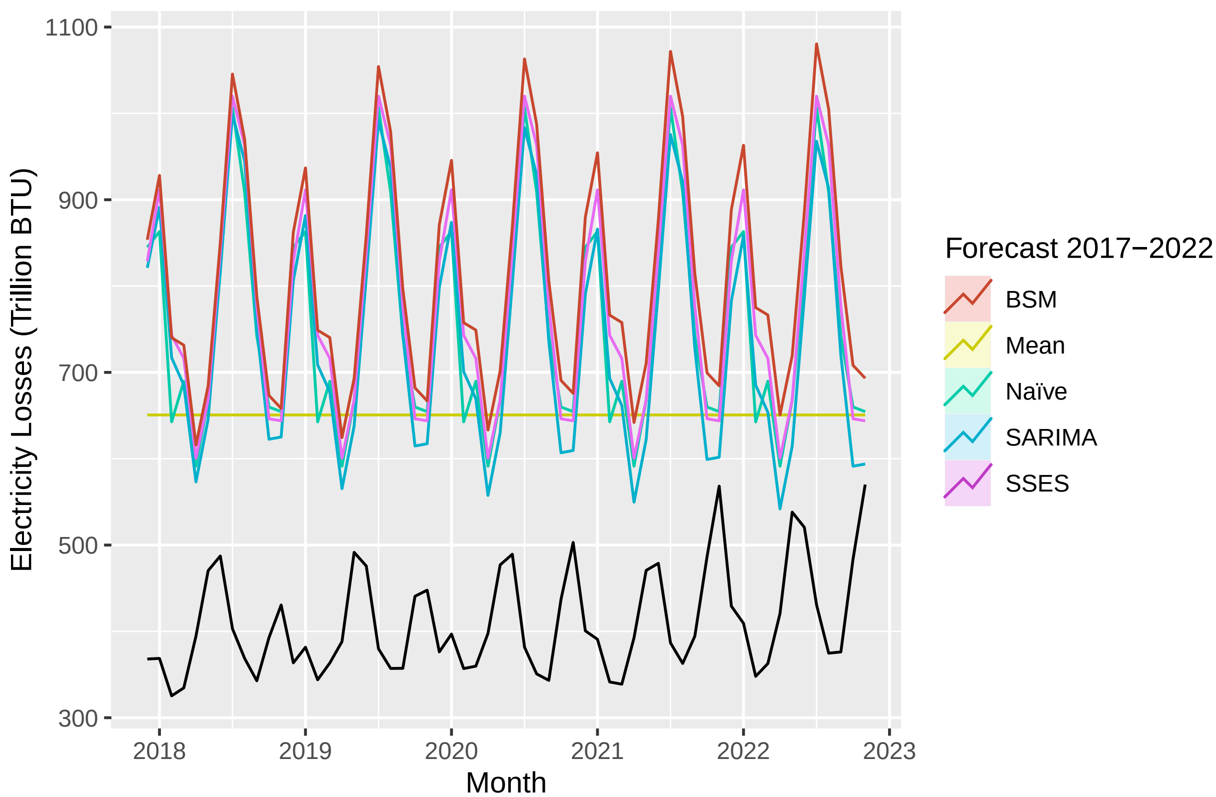
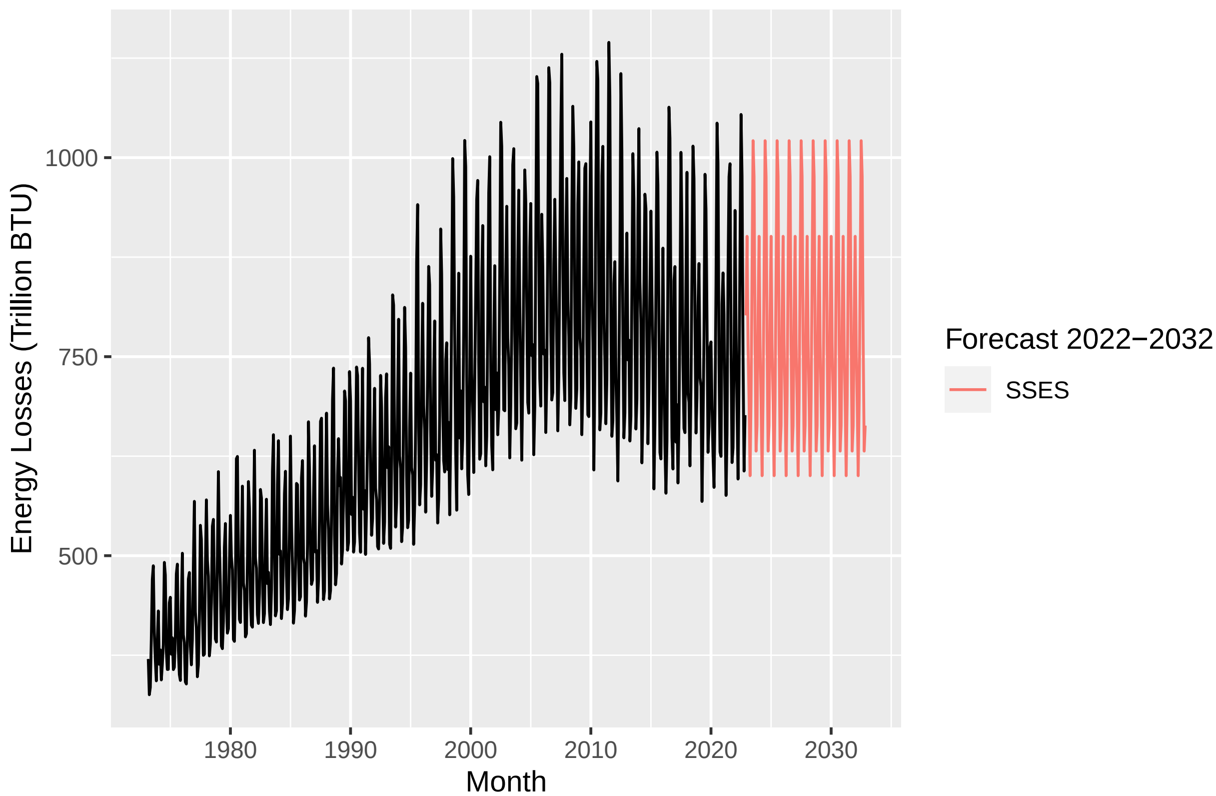


Table 3: Forecast Accuracy for Residential Enegy Losses' Seasonal Data

	ME	RMSE	MAE	MPE	MAPE
MEAN	120.92807	180.58483	137.55002	13.19234	15.97748
SNAIVE	14.37412	53.62351	40.88775	1.44559	5.20327
SARIMA	29.98819	50.03421	40.76336	3.83462	5.27084
SSES	-8.18260	40.19296	31.73947	-1.31055	4.18805
BSM	-40.78226	59.95601	47.41864	-5.67828	6.49470



From the Table 3, the SSES model has the smallest RMSE, so we continually fit this model to forecast future 10 years consumption and losses.



This plot shows the forecasted energy losses for the same 50-year period, with the SSES model generating a 10-year forecast (120 periods) for the residential sector. The forecast indicates a continuation of the strong seasonality and trend observed in the historical data, with a slight increase in losses towards the end of the forecast horizon.

Commercial Sector

Primary Energy Consumption Plots (1973-2022 vs. 2017-2022)

These following two plots of primary energy consumption for the residential sector show a strong relationship and seasonality, with the SARIMA model performing the best among the five models. The first plot shows the original data from 1973 to 2022, along with the forecasts generated by the five models. The second plot only shows the last three years of the data (2019-2022), along with the same forecasts generated by the five models. In both plots, the SARIMA model forecast shows a close fit to the observed data, capturing the strong seasonality and trend in the data. The other models also capture the seasonality to varying degrees, but they may not perform as well in capturing the trend.

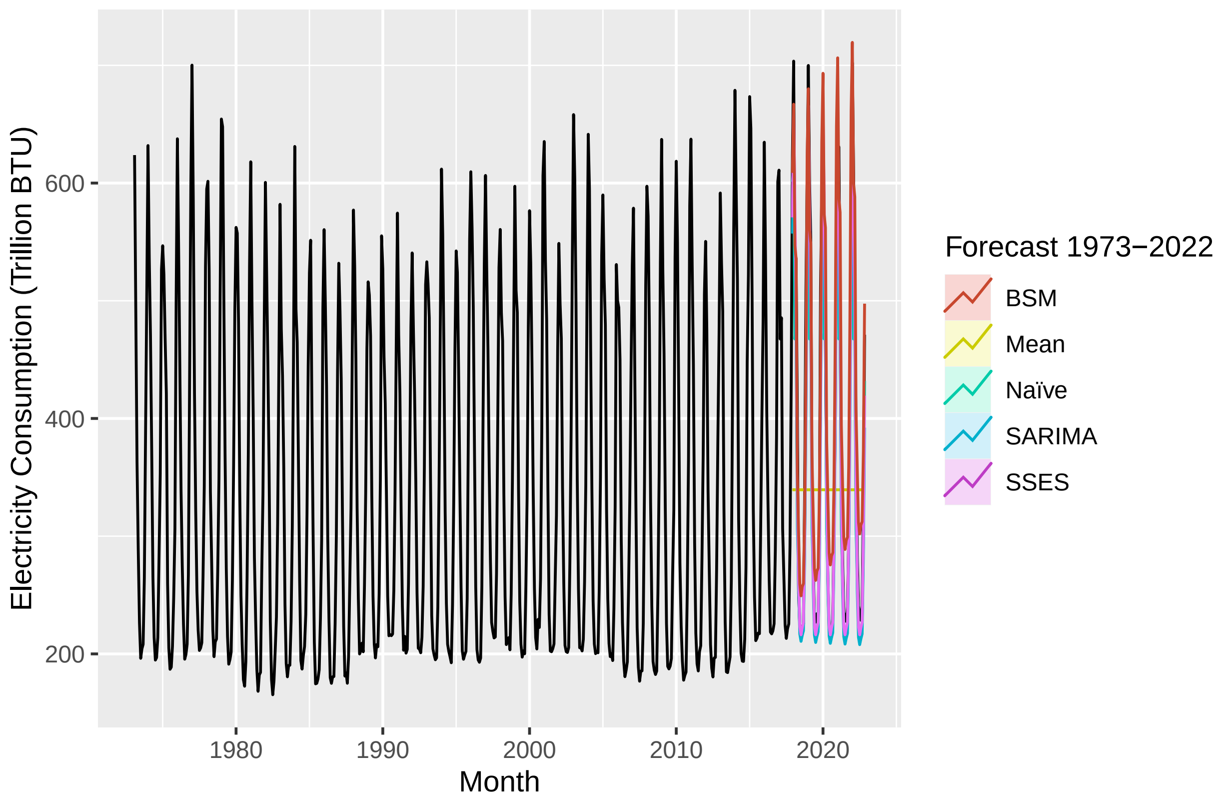
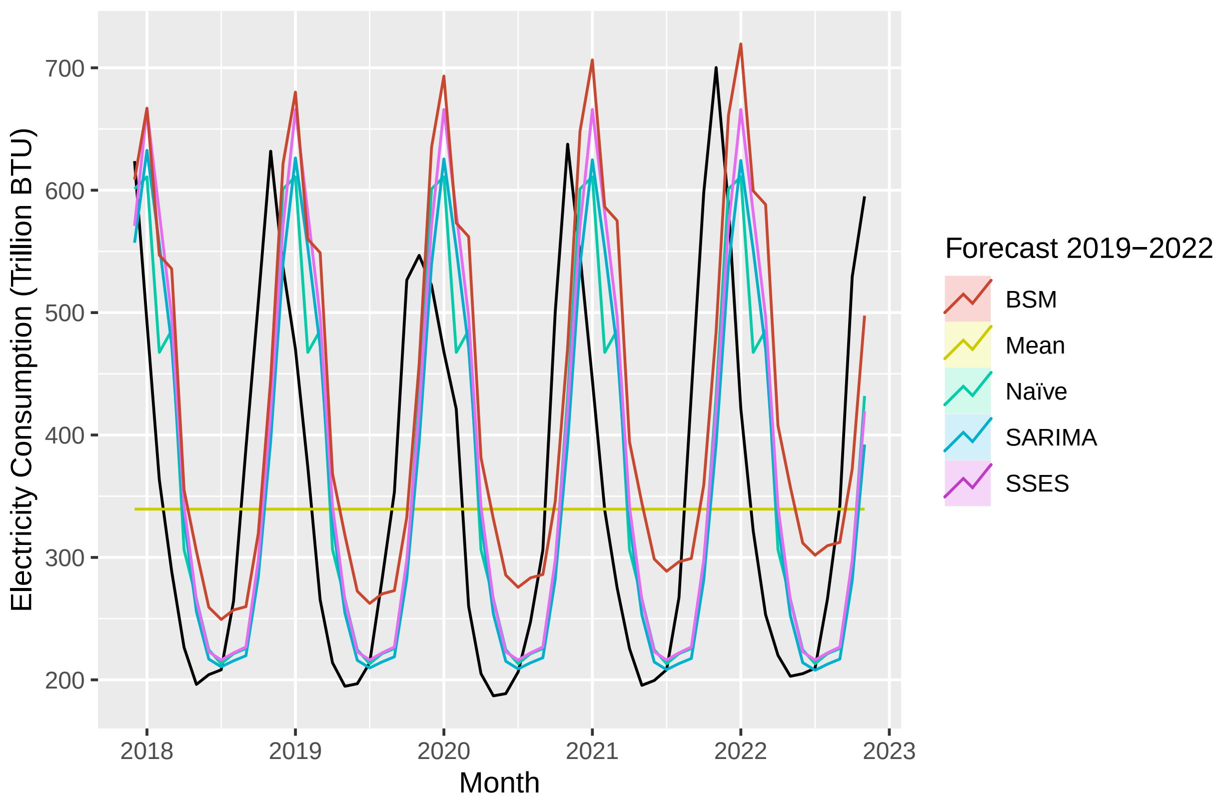
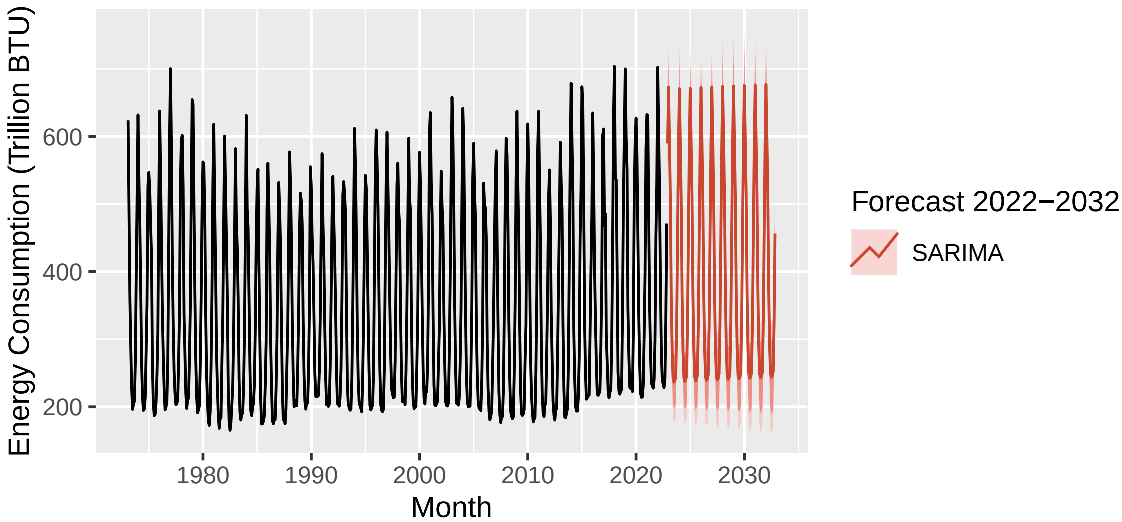

Electricity Consumption (Trillion BTU)
600
400
200
1980
1990
2000
2010
2020
Month
Forecast 1973−2022
BSM
Mean
Naïve
SARIMA
SSES

Table 4: Forecast Accuracy for Commercial Primary Energy Consumption's Seasonal Data

	ME	RMSE	MAE	MPE	MAPE
MEAN	51.56288	169.02419	140.99802	-2.36476	35.75435
SNAIVE	28.83988	52.43772	35.56298	6.22350	7.83247
SARIMA	31.86433	43.89807	33.38297	7.69468	8.02837
SSES	57.32929	83.87768	61.99349	9.46394	10.14529
BSM	-35.97851	54.28067	47.88303	-12.84780	15.11202



The SARIMA model has the lowest RMSE, and SNAIVE model has the lowest MAPE. But we use the value of RMSE as the threshold.

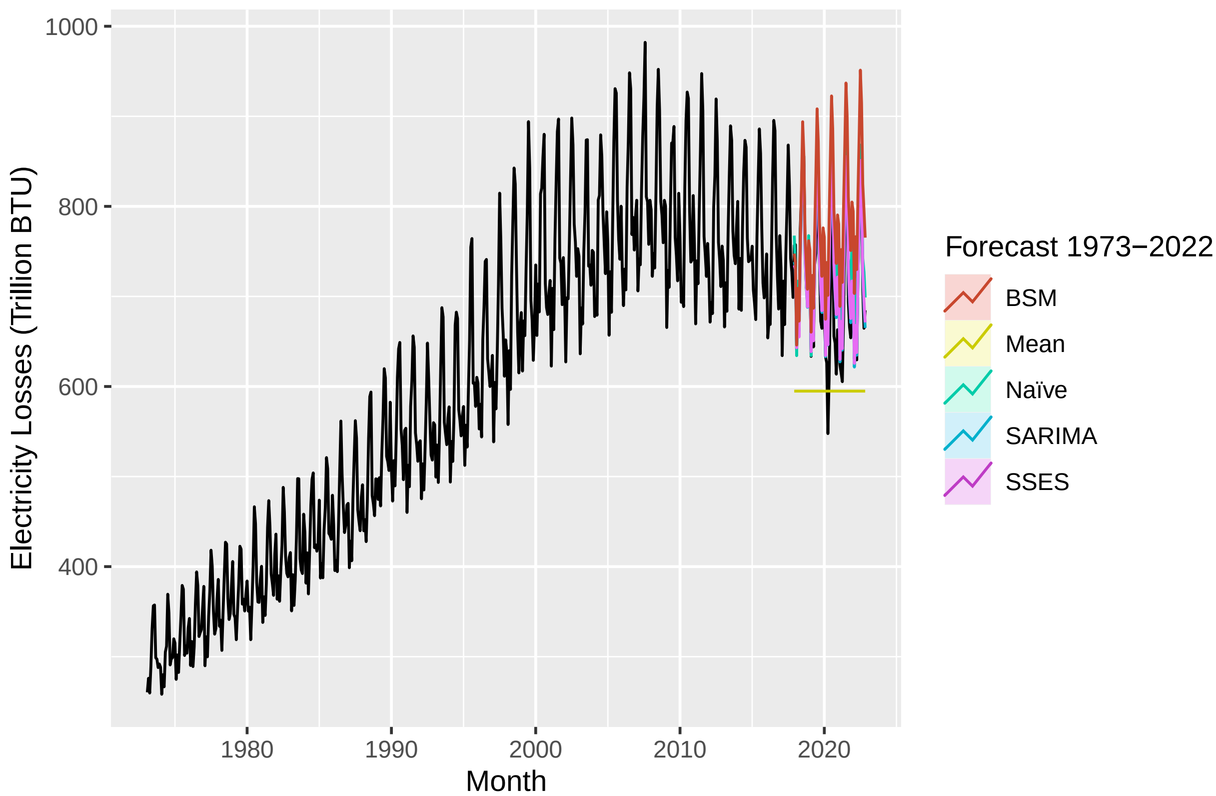


This plot shows the forecasted primary energy consumption for the entire 50-year period (1973-2032), with the SARIMA model generating a 10-year forecast (120 periods) for the commercial sector. The forecast indicates a continuation of the strong seasonality and trend observed in the historical data, with a significant increase in consumption towards the end of the forecast horizon.

Energy Loss Plots (1973-2022 vs. 2017-2022)

These two following plots of commercial energy losses show a strong seasonality in the data, with the SARIMA model performing the best among the five models. The first plot shows the original data from 1973 to 2022, along with the forecasts generated by the five models. The second plot only shows the last five years of the data (2017-2022), along with the same forecasts generated by the five models.

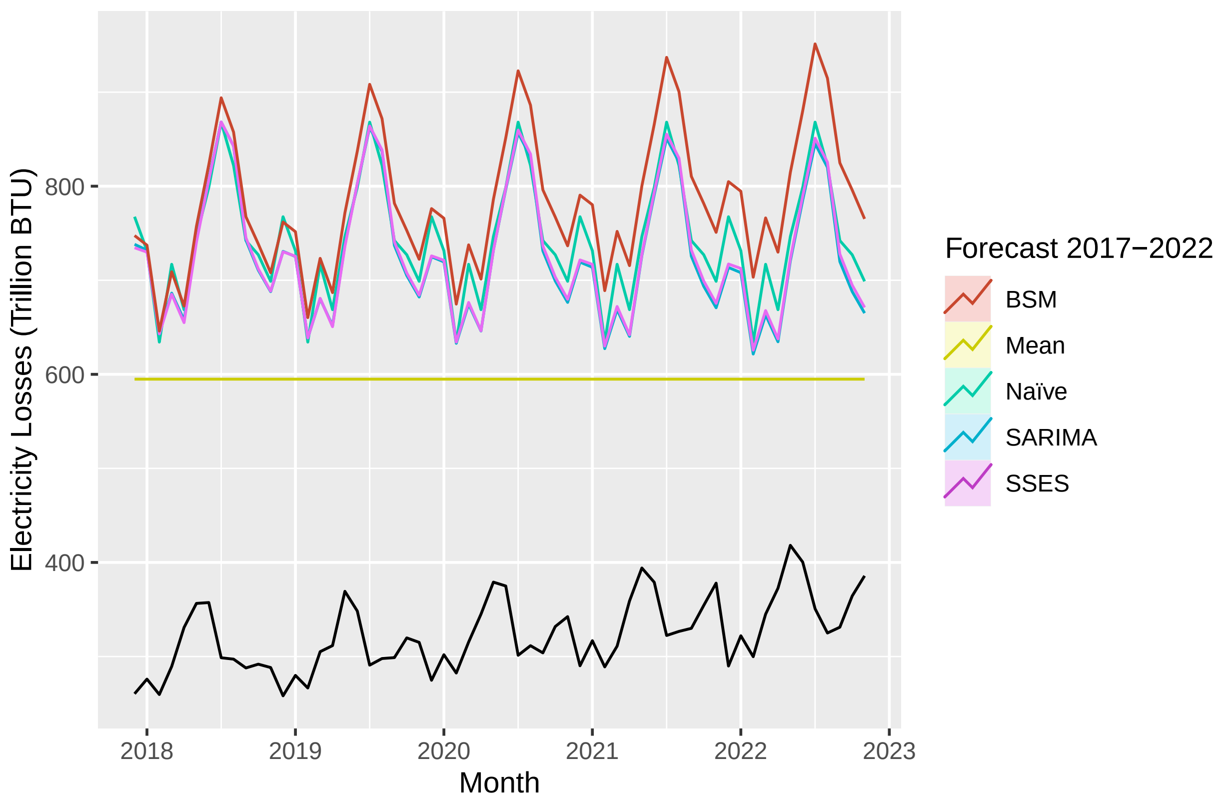
In both plots, the SARIMA model forecast shows a close fit to the observed data, capturing the strong seasonality and trend in the data. The other models also capture the seasonality to varying degrees, but they may not perform as well in capturing the trend.



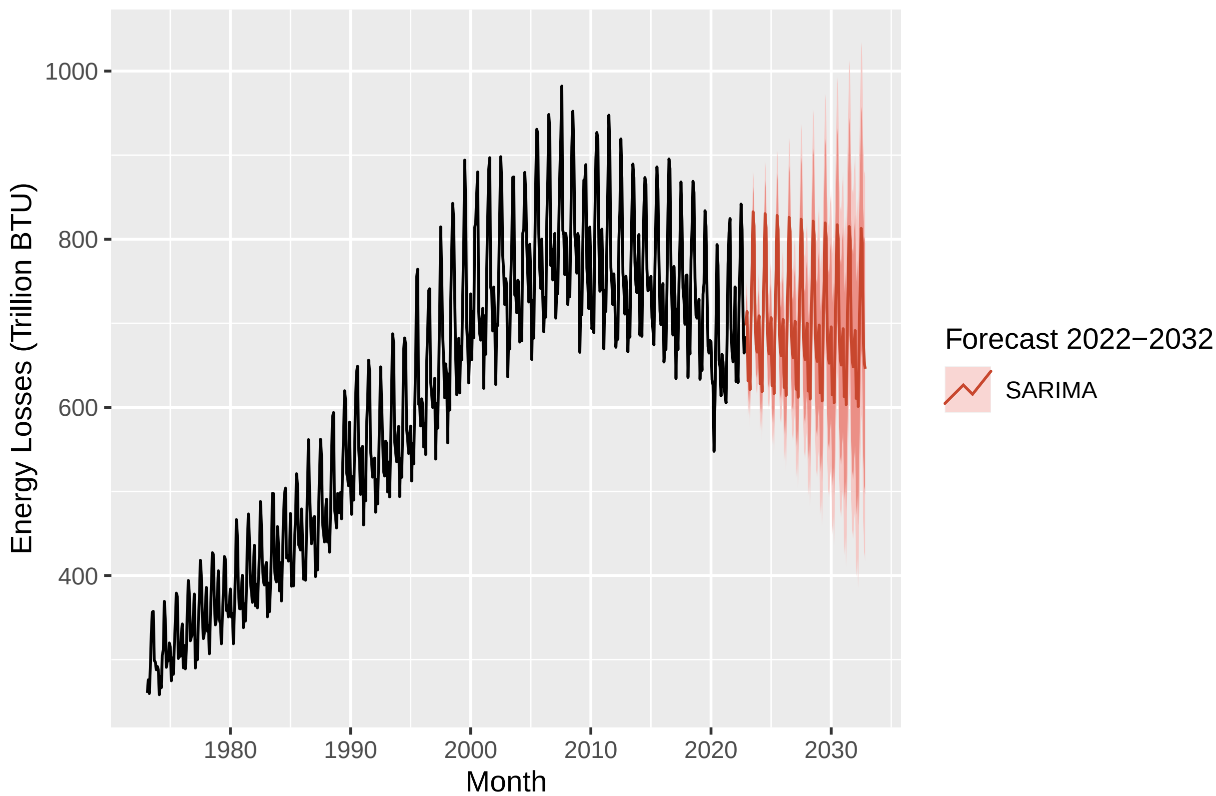
Electricity Losses (Trillion BTU)
1000
800
600
400
1980
1990
2000
2010
2020
Month
Forecast 1973−2022
BSM
Mean
Naïve
SARIMA
SSES

Table 5: Forecast Accuracy for Commercial Energy Losses' Seasonal Data

	ME	RMSE	MAE	MPE	MAPE
MEAN	110.86060	132.39779	112.43234	14.83786	15.12483
SNAIVE	-37.82263	54.65647	42.38680	-5.71676	6.29919
SARIMA	-21.19513	39.16264	28.22249	-3.22158	4.17409
SSES	-23.25252	40.72791	30.06601	-3.51722	4.43541
BSM	-77.77498	91.94242	79.35527	-11.38418	11.59101



From the Table 5, SARIMA model has both the lowest RMSE and MAPE. Therefore, it is the perfect fitting model to forecast future ten years' variables in the commercial sector.



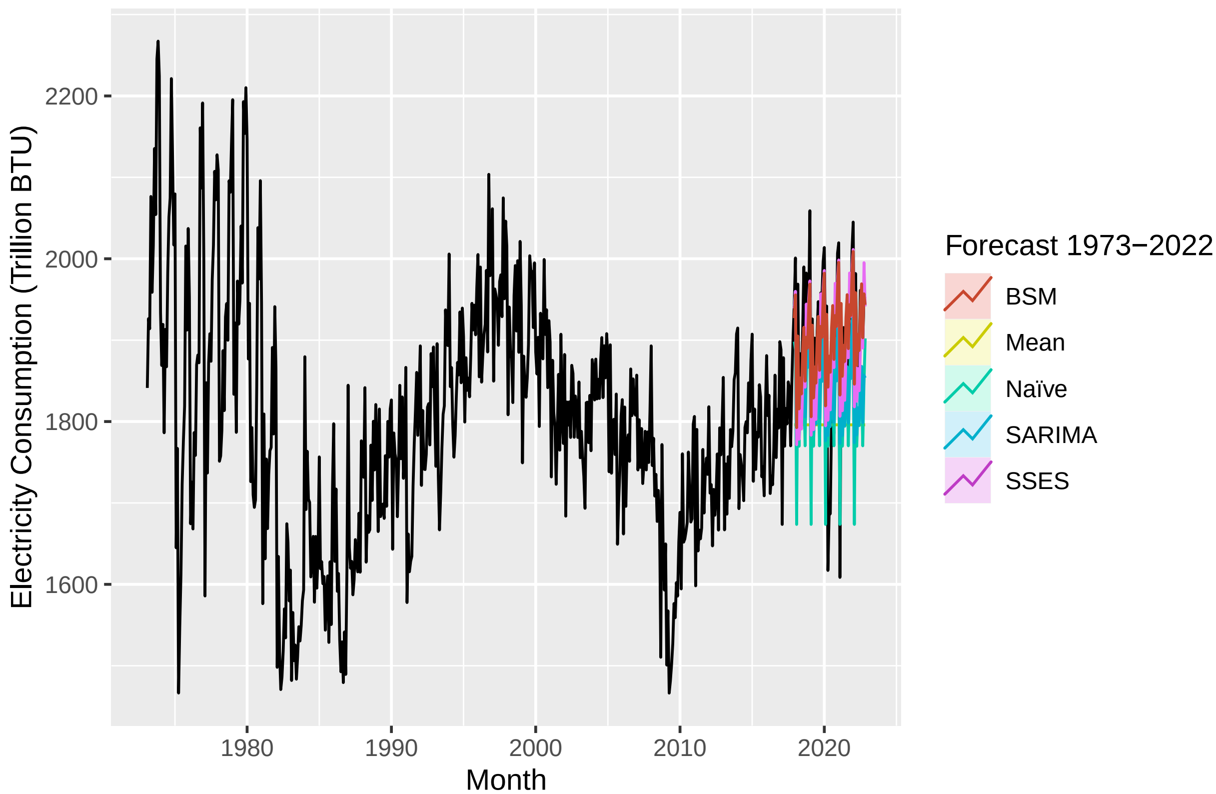
We create a time series object for the entire energy loss data, fit a SARIMA model to it using the auto.arima() function, check the residuals, and generate a forecast for the next 10 years using the forecast() function. This plot shows the forecasted energy losses for the same 50-year period, with the SARIMA model generating a 10-year forecast (120 periods) for the commercial sector. The forecast indicates a continuation of the strong seasonality and trend observed in the historical data, with a significant increase in losses towards the end of the forecast horizon.

Industrial Sector

Industrial Primary Energy Consumption

First, the primary energy consumption time series data has been fit into the five models for 5 year forecasting from 2017 to 2022 for the testing purpose. While the MEAN model did not quite work well and showed no forecasting pattern, the other models turned out a messy wavy pattern with great variability. Yet, compared to the original data from 2017 to 2022, the forecasts are not as accurate as we expected. While the forecasts display the same repeating pattern every year in the five years, the actual data has a huge variability. With such big variability, it is hard to demonstrate very accurate forecasting. We determined that the best model for this data is SSES, given the small values of RMSE and MAPE from the accuracy test below.

Primary Energy Consumption Plots (1973-2022 vs. 2017-2022)



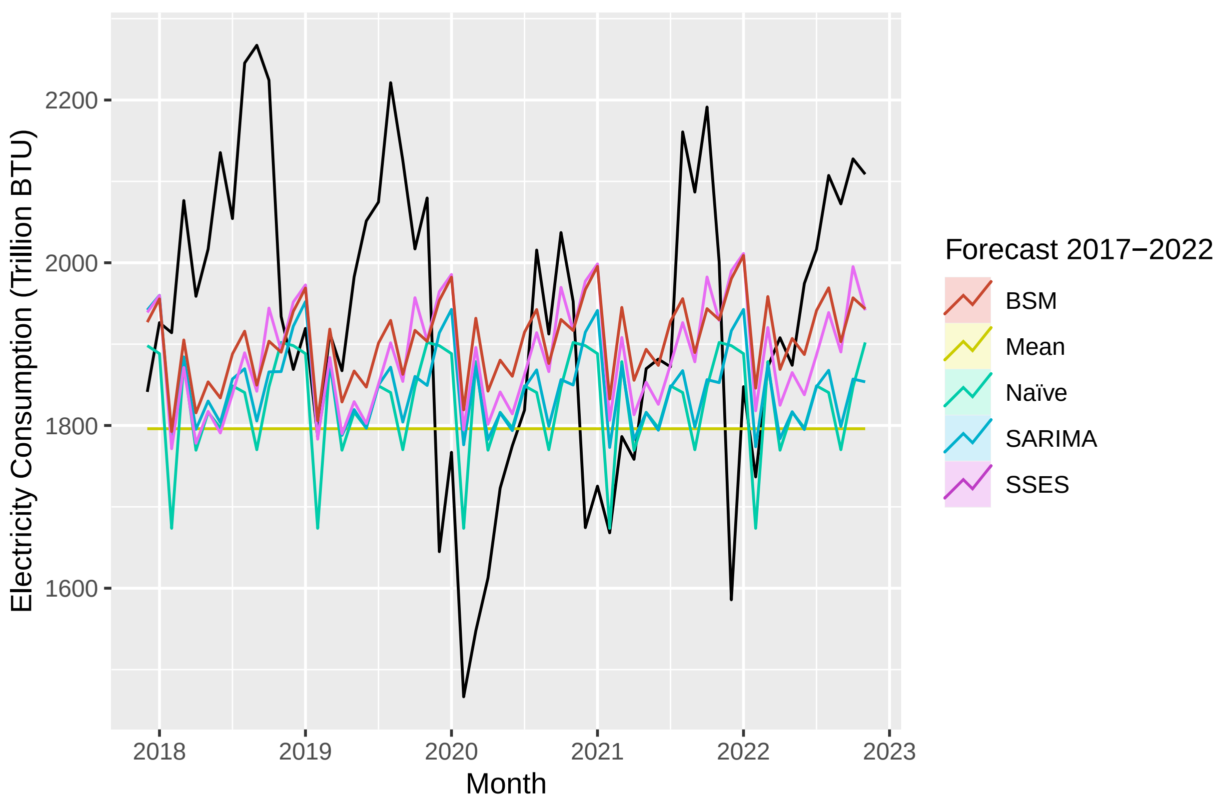
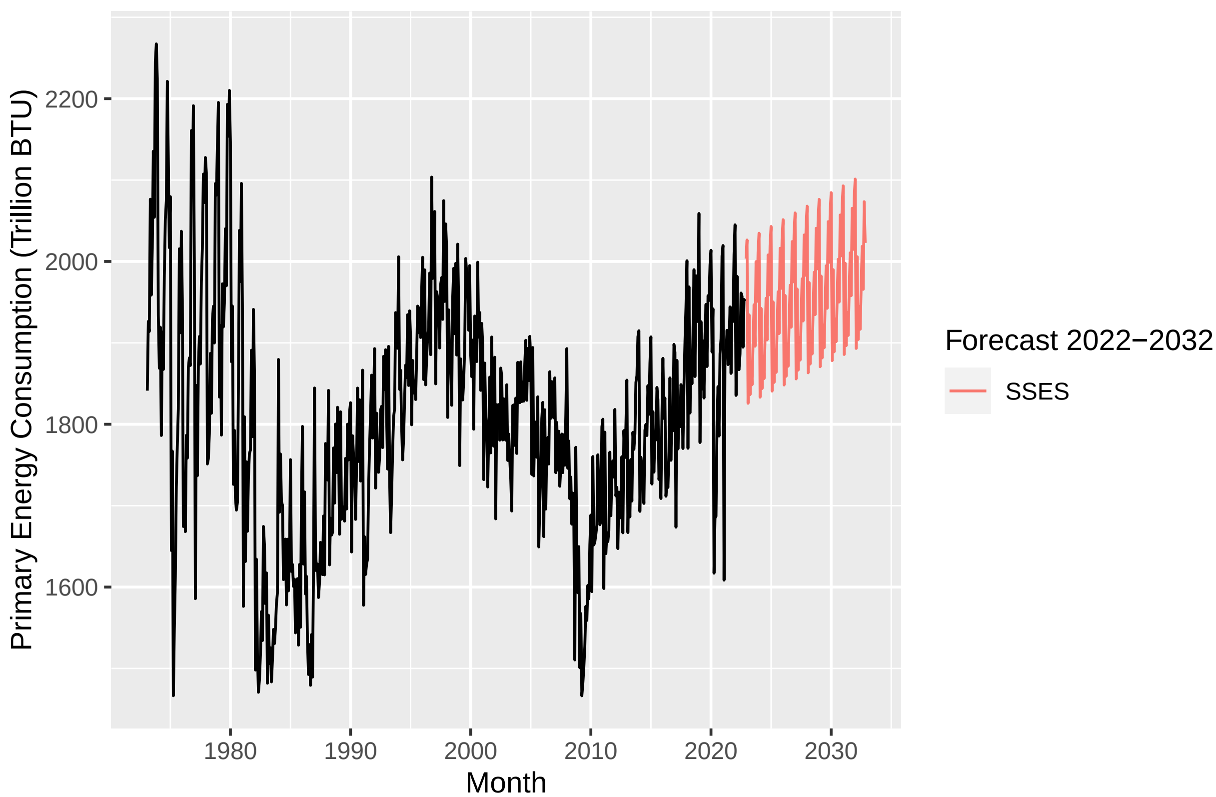


Table 6: Forecast Accuracy for Industrial Primary Energy Consumption's Seasonal Data

	ME	RMSE	MAE	MPE	MAPE
MEAN	101.47143	137.57833	122.82318	5.10577	6.39965
SNAIVE	69.94750	97.26449	86.84143	3.55857	4.57356
SARIMA	50.80107	81.45935	73.15008	2.53073	3.87272
SSES	11.85278	64.13764	48.94701	0.49903	2.65083
BSM	-6.72821	65.76714	41.68324	-0.51284	2.29347

Then, here's the ten year forecasting for the industrial sector primary energy consumption. SSES model clearly shows the increasing trend of energy consumption in the industrial sector from 2022 to 2032. This can be understood as a sign of expanding electrification in the sector. Compared to the other two sectors, the industrial sector has experienced a rapid installation of electrified machines, energy-intensive manufacturing systems, and increasing production due to growing demand in general. Thus, the primary energy consumption in the industrial sector will likely increase in the next ten years. A possible challenge could be how the sector will accommodate the increasing electricity needs while achieving their carbon emission reduction objectives.



Industrial Energy Loss

Second, the energy losses forecasts showed similar results as the energy consumption. BSM, Naive, SARIMA, and SSES models displayed the similar patterns for the 2017-2022, and the MEAN model did not do a good job. The industrial sector energy losses data contain an odd variability which is not quite the same as a simple seasonality. Although the forecast models captured the odd pattern to some extent, the forecast data are not completely matching with the original data. From the accuracy test, we identified the best model for this forecasting is SARIMA as the training SARIMA model had the smallest RMSE and MAPE values.

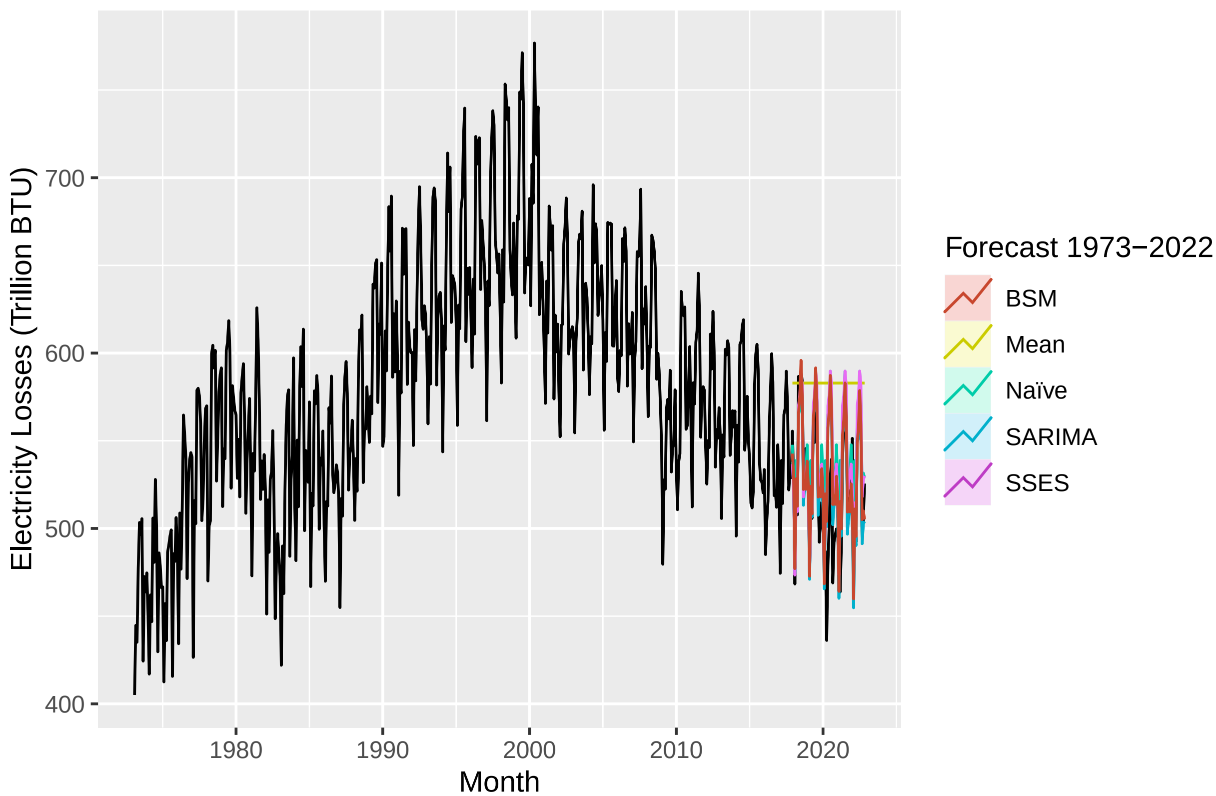
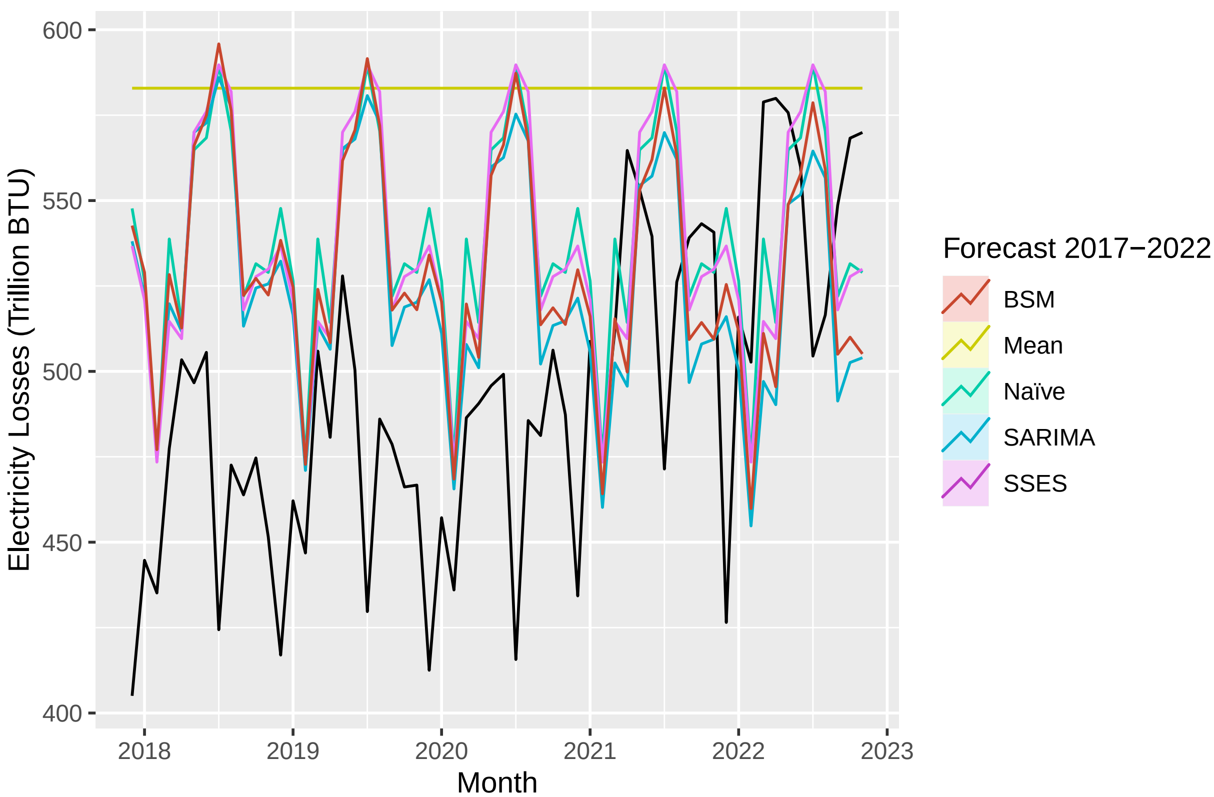
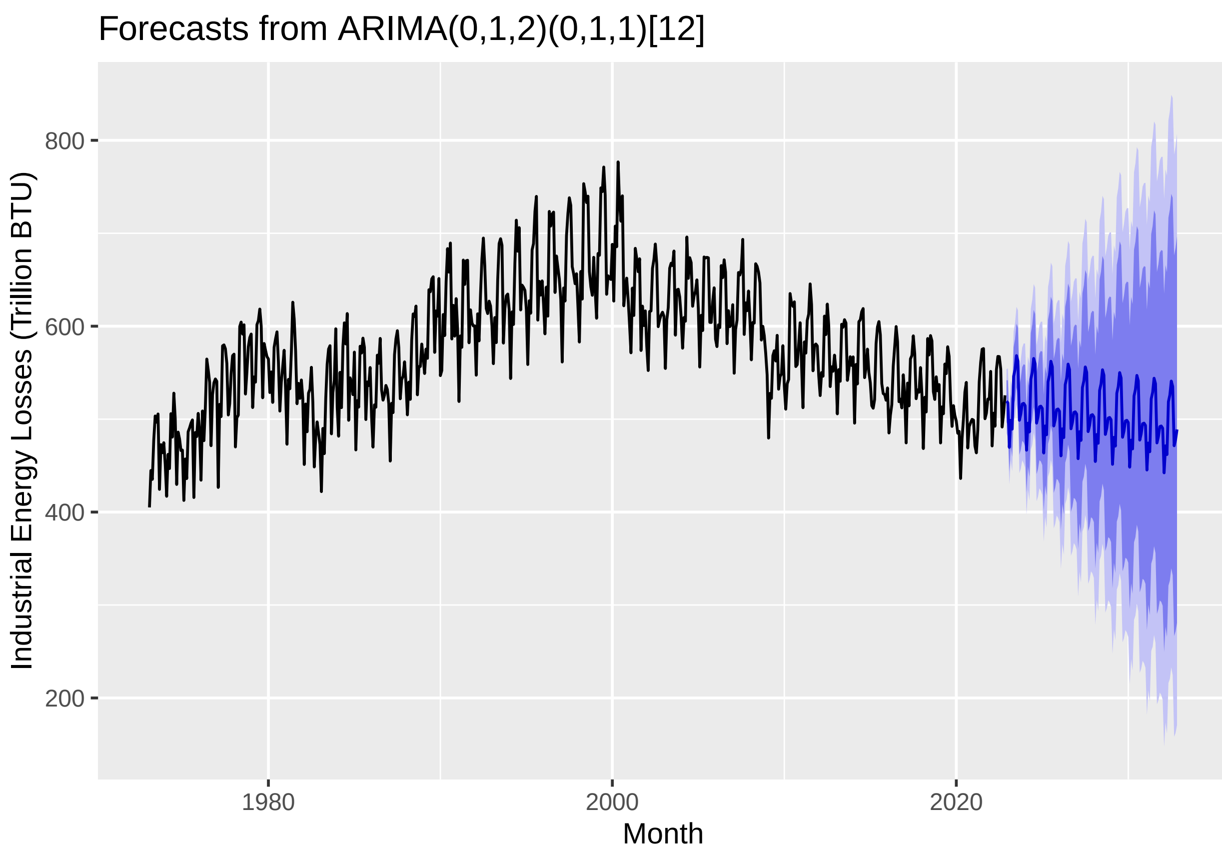


Table 7: Forecast Accuracy for Industrial Enegy Losses' Seasonal Data

	ME	RMSE	MAE	MPE	MAPE
MEAN	-61.62735	71.29813	62.03478	-12.35443	12.42369
SNAIVE	-235.98558	258.03045	235.98558	-44.96051	44.96051
SARIMA	-4.38849	22.60478	15.32718	-0.97952	3.03061
SSES	-16.11516	28.23179	20.81630	-3.23995	4.10684
BSM	-9.68020	24.43230	17.07850	-2.00786	3.38398

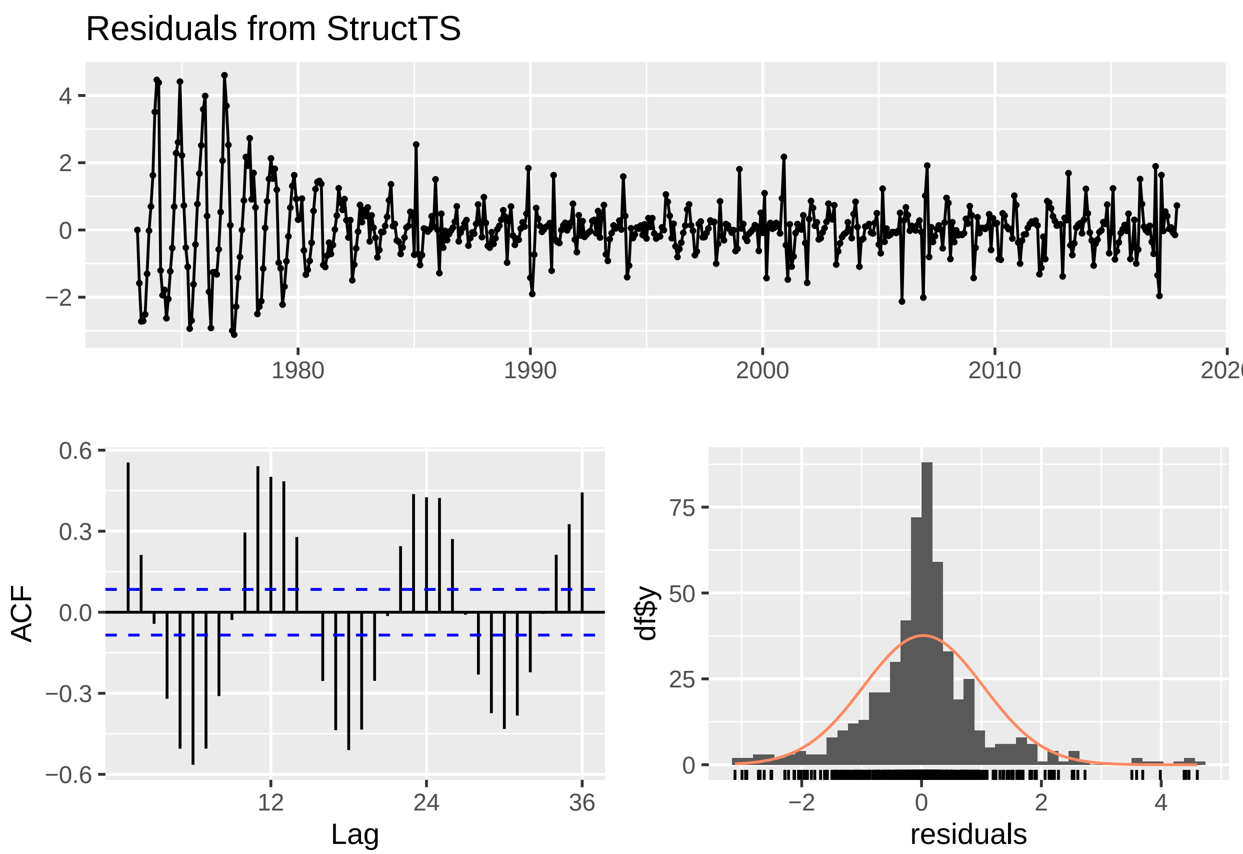


The 2022-2032 ten year forecasting for the industrial sector energy losses is shown below. In contrast to the energy consumption forecasting, the SARIMA model turns out having a decreasing trend forecasting for the energy losses. Given that the energy losses have been decreasing especially since after 2000, we can assume that increased energy efficiency in the industrial systems have contributed to the decline. Technological advancement in manufacturing energy efficiency has so much potential to reduce the energy losses in the industrial sector. It should be a good incentive for many industrial sector performers to enhance their energy efficiency for more effective use of electricity with lower energy losses.

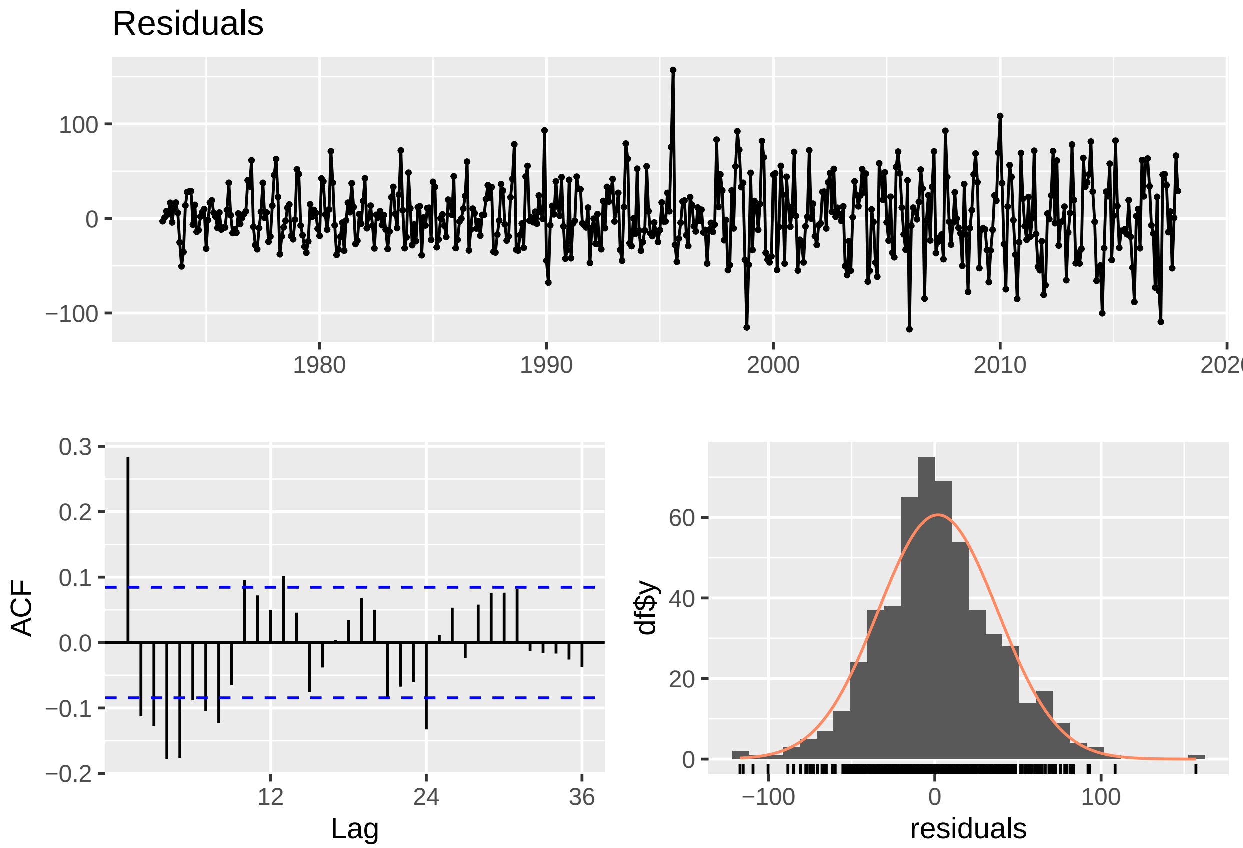


Best Models' Residuals 1973-2022

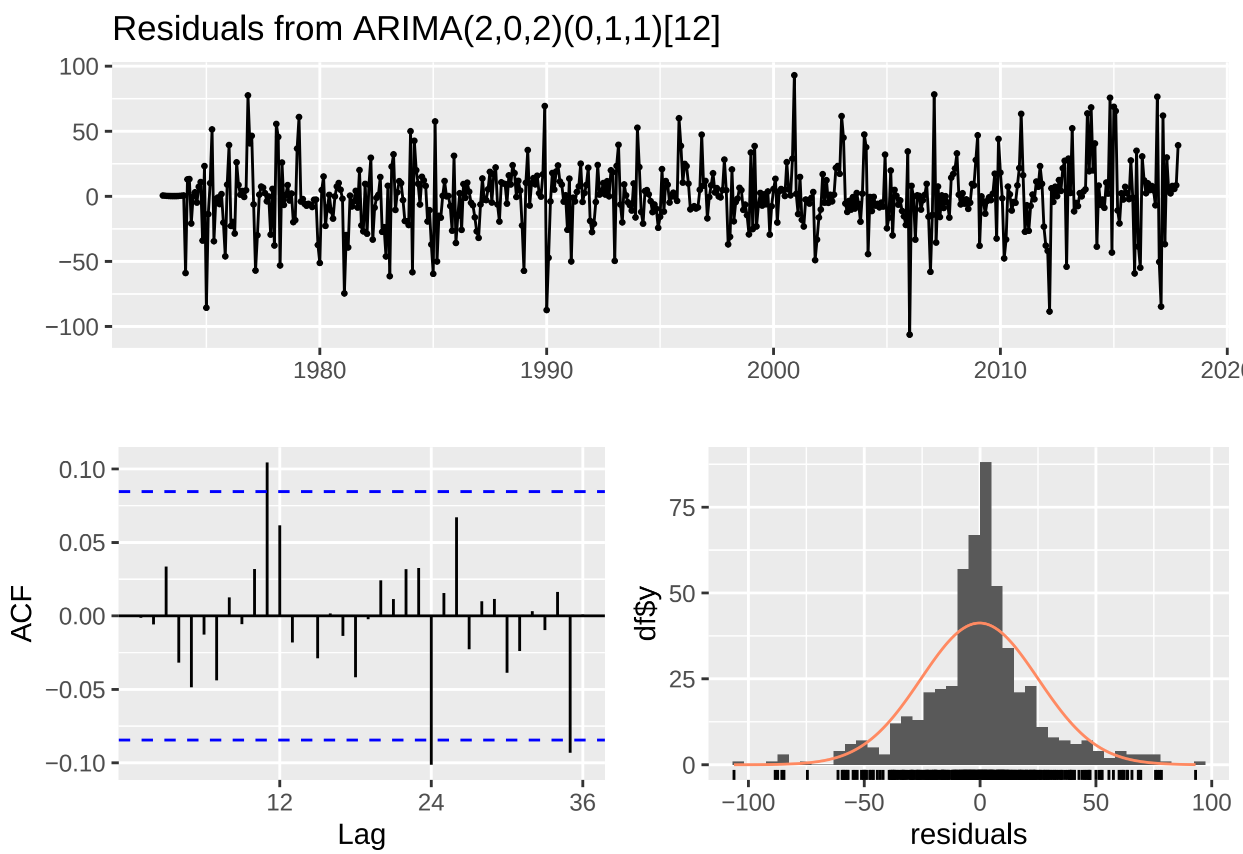
The following six figures show residuals of all six best models in three sectors, and they are BSM, SSES, SARIMA, SARIMA, SSES, and SARIMA. It is not hard to tell all residuals indicate nearly normal distributions that they could fit the best forecasting models.



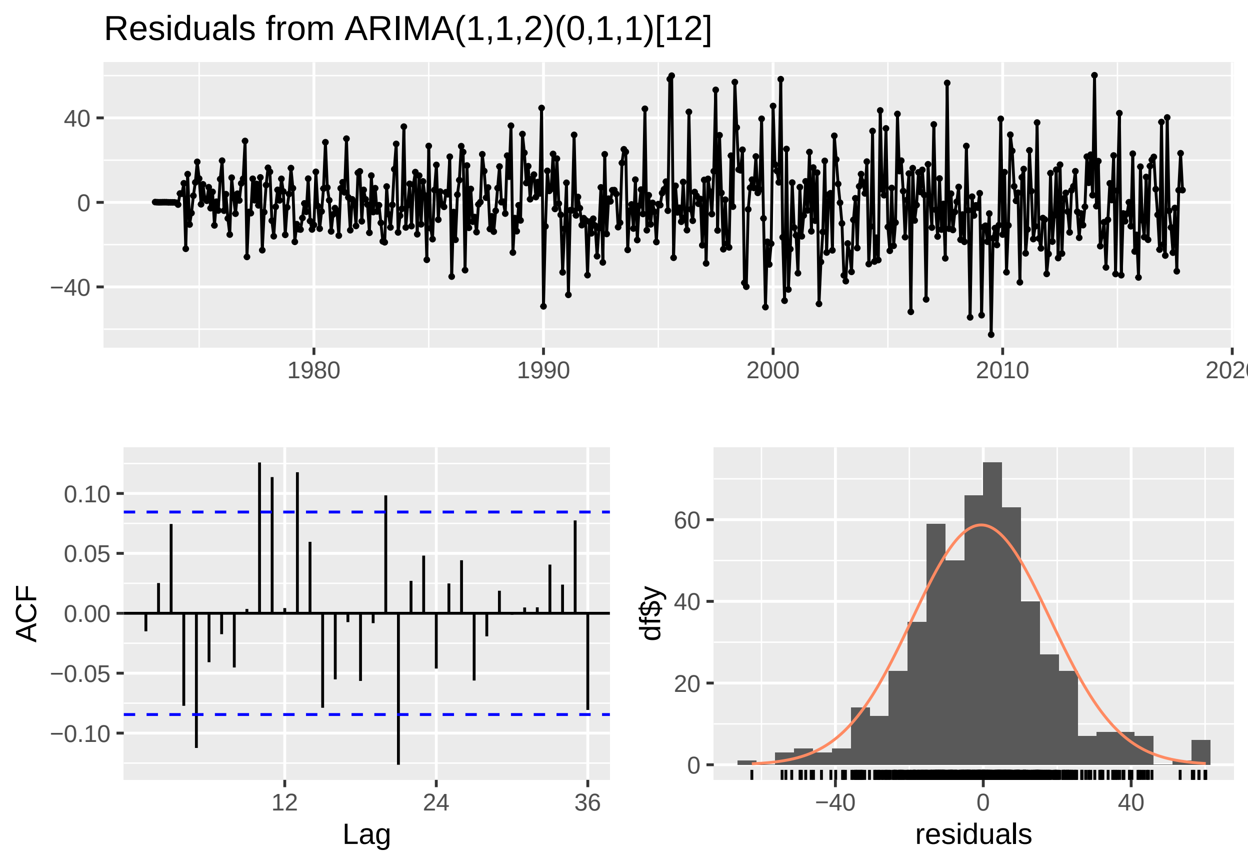

```
##
##  Ljung-Box test
##
## data:  Residuals from StructTS
## Q* = 1947.3, df = 24, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 24
```



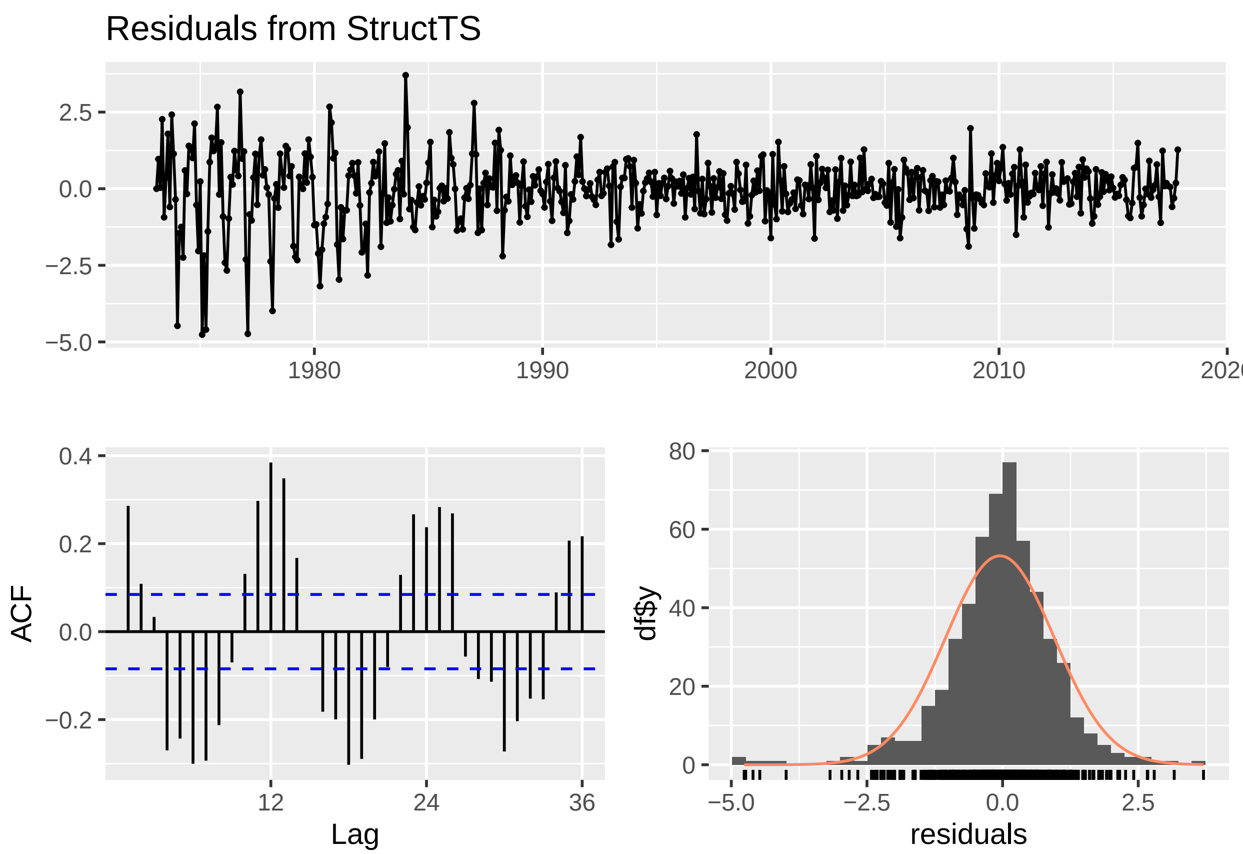

```
##
##  Ljung-Box test
##
## data:  Residuals
## Q* = 157.92, df = 24, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 24
```



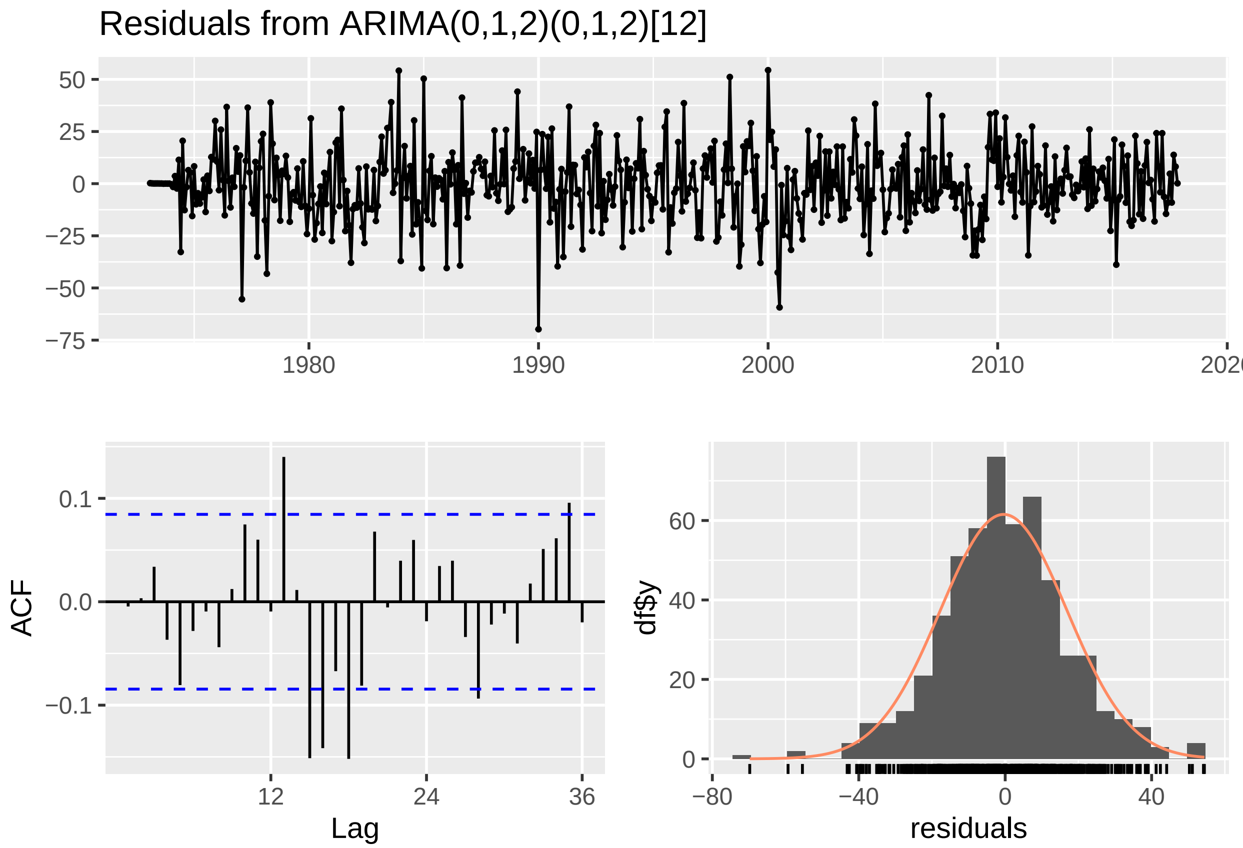

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,2)(0,1,1)[12]
## Q* = 21.49, df = 19, p-value = 0.3104
##
## Model df: 5.   Total lags used: 24
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,2)(0,1,1)[12]
## Q* = 65.617, df = 20, p-value = 9.304e-07
##
## Model df: 4.   Total lags used: 24
```




```
##
##  Ljung-Box test
##
## data:  Residuals from StructTS
## Q* = 715.44, df = 24, p-value < 2.2e-16
##
## Model df: 0.   Total lags used: 24
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,2)(0,1,2)[12]
## Q* = 71.236, df = 20, p-value = 1.143e-07
##
## Model df: 4.   Total lags used: 24
```

Conclusion

Limitation

One of the challenges we faced in this study was energy losses forecasting accuracy. For all three sectors, when we ran the 2017-2022 forecast testing, the results were not quite accurate and even showed some patterns in opposite directions. We have tried different testing time frames, before and after the COVID year of 2020, and we ended up using the 2017-2022 5-year testing window because of the better accuracy test results.

Another limitation is that our models did not consider the impacts from the external factors, such as an economic recession, political disruptions, and pandemics, on the energy consumption and energy losses data. As mentioned in the dataset information description, the industrial sector has been prominently affected by such external factors. In the next ten years, it is hardly possible to predict what will happen in economics and politics. As we have observed, electricity consumption significantly decreased during the early phase of the COVID pandemic, so extraordinary events could disturb the forecast patterns.

Moreover, the ten-year forecasting for energy consumption in the residential and commercial sectors and the energy losses in the residential sector demonstrate the constant trend. These constant trends are likely stemming from the constant trends in previous years, but that could not be the case for future years.

All three sectors show perfect performance on our first variable, primary energy consumption, but terrible performance on our second variable, energy losses, especially in the residential and commercial sectors, when we zoom in and make a detailed comparison between forecasts of historical data and the 2017-2022 prediction. Unexpected losses throughout the transmission procedure are most likely to blame for our inability to predict the precise energy loss statistics. Due to its comparatively high voltage requirements, the industrial sector has the shortest path in terms of the transmission process. However, further step-down processes occur in the commercial and residential sectors, which result in significant energy losses. There are two solutions that we could use to further enhance the precision of energy loss predictions: adding exogenous factors and adopting multiple seasonality models.

Discussion

We can see that the projection for the next ten years differs in three areas depending on several models. This project successfully investigated a variety of models that are appropriate for primary energy use. We came to the conclusion that the SARIMA model outperformed the other models since it accounted for the best model in three forecasts when utilizing performance measures (RMSE, MAPE, MSE, and MPE) as well as a visual comparison. We were able to anticipate the future of two variables, which indicate roughly declining energy losses and rising demands for primary energy consumption. The limits of the two models we choose for future forecasting, as well as those of time series forecasting, must be acknowledged. Nevertheless, many stakeholders and policymakers in the energy and utility sectors can still benefit from anticipating energy usage and losses. We can anticipate that energy use will decline in the near future as a result of increased energy awareness and technical improvements that will reduce energy losses.

Nevertheless, there are a lot of unknown variables causing the abrupt changes in all three areas. For instance, increased electrification, a need for renewable energy, and efficient battery storage may all contribute to an increase in energy consumption in the industrial sector. It is simpler to achieve the requirements of minimal energy losses and energy needs in the era of carbon neutralization. Additionally, technological advancement is a decision factor. Energy loss dramatically decreased after the year 2000, and the emergence of new energy technology is most likely to blame for this. In five years, technology might advance, especially with regard to renewable energy power plants. Our prediction won't be accurate then, but we're happy to predict a different model and witness progress.