Electricity Net Generation in the US and Texas

STAT 6363 - Time Series Analysis Final Project Report

Léon Lee Yuan

2024-05-03

1. Introduction

1.1 Background

The data used in this project is the monthly net generation of electricity in the United States and Texas from January 2001 to January 2024. The monthly net generation is the total amount of electricity generated by all sectors in the United States and Texas. The net generation includes electricity generated by coal, natural gas, nuclear, conventional hydroelectric, wind, all solar. The data is measured in thousand megawatts hours. The data is obtained from the U.S. Energy Information Administration (EIA) website. We will focus on two time series data. The first one is the net generation of electricity in Texas. The each state data link is here. The second one is the total net generation of electricity in the United States. The whole US data link is https://www.eia.gov/electricity/data/browser/.

1.2 Visualization

The first step is to read in the data and convert the Month column into a date format. The second step is to pivot the data into a long format for ggplot. The third step is to plot the data. The first plot Figure 1 is the bar plot of the cumulative every year net generation by each state. For the page limit, I only displayed the first five states. As we can observe, Texas has the highest net generation and it has as twice as the second state, Florida. This is the reason that I choose Texas as the state to analyze. The second plot Figure 2 is the time series plot of the net generation monthly in Texas from January 2001 to February 2024. The third plot Figure 3 is the time series plot of the net generation monthly in the United States from January 2001 to January 2024. The US net generation is the sum of each state net generation.

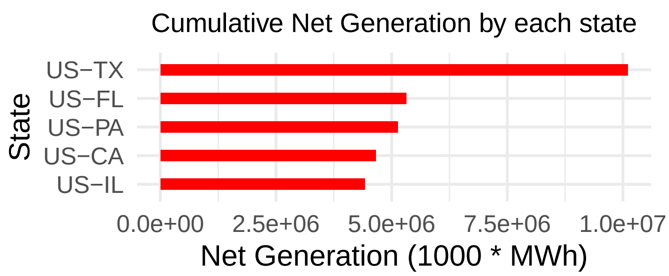

Figure 1: Cumulative Net Generation by each state

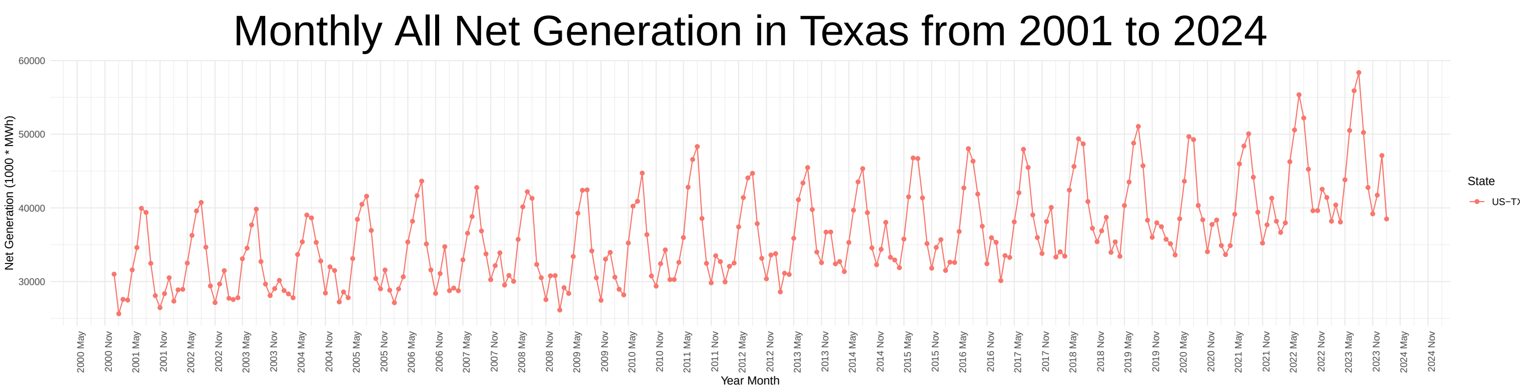

Figure 2: Texas Net Generation from 2001 to 2024

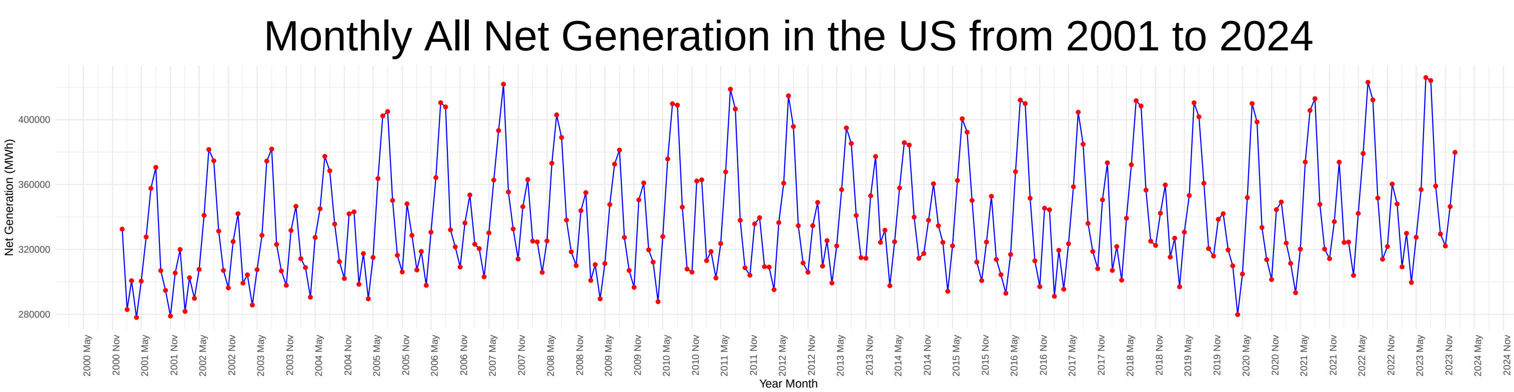

Figure 3: The US Net Generation from 2001 to 2024

Questions

I want to answer the following questions: 1. What is the trend of the net generation in Texas from 2001 to 2024? 2. What is the trend of the net generation in the United States from 2001 to 2024? 3. What is the seasonality of the net generation in Texas and the United States? 4. What is the difference between the net generation in Texas and the United States? 5. What is the forecast of the net generation in Texas and the United States in the next 12 months?

2. Statistical Analysis

2.1 The US net generation 🇺🇸

When we look at the trend, we may first ask if the time series is stationary. The first step is simply to visually check the time series plot. For the US net generation, we can see that the time series plot seem stationary. However, we also need some quantitative analysis to confirm this. The first thing I do is to split the time series into two parts , the first half and the second half. Then I calculate the mean and standard deviation of each part. The mean of the first part is 342216.4 and the mean of the second part is 333535. In the measurement of unit, these two numbers are very close. The standard deviation of the first part is 35246.63 and the second part is 34130.53. The standard deviations are close too. This is a good sign that the time series is stationary. The next step is to use the Augmented Dickey-Fuller test to test the stationarity. As we can see, the p-value is smaller than 0.01 which means we can reject the null hypothesis. The time series is stationary.

Since the time series is stationary, it is not necessary to transform the data. We can directly fit the SARIMA model to it. The next step is to decompose it and detect any seasonality and trend. I use the additive model to decompose it because the variance of it is constant. The decomposition plot Figure 4 shows there is a clear seasonality of 12 months. I also plot the periodogram to further decompose the frequencies with the general guideline, $\sqrt{n}/2$ to smooth it. As we can observe, the smoothed periodgram shows signals at period 12, 24, 36, 48, 60. Among these signal periods, 24 is the strongest. However considering the decomposition result and the practical meaning, I still choose the seasonal period as 12 because the net generation is monthly data and the signal at 12 is only slightly weaker than 24. In light of parsimony, 12 should be the seasonal period. The overall trend is the stationary except for that there has a slight downward from 2017. So I still decide to take the first order difference of seasons and non-seasons to remove that slight trend.

After taking the first 1 order difference of seasons and non-seasons at 12, The ACF and PACF Figure 6 show the ACF and PACF estimation. As we can observe, at all non-seasonal lags, both ACF and PACF show tailing off to zero. This suggests that the first order of AR and MA, p = q = 1. At the seasonal lag, the ACF shows a cut-off at lag 2 and the PACF shows a tailing off to zero. This suggests the first order seasonal MA, Q = 1. In summary, the SARIMA model is SARIMA(1,1,1)(0,1,1)[12].

After fitting, the diagnostic plot Figure 7 shows the residuals of the model. The Q-Q plot shows the residuals are normal distributed. The ACF of residuals shows that they are not correlated. The Ljung-Box test shows that almost all lag p-values are much greater than 0.05. This suggests that the residuals are not autocorrelated and they are random like a white noise. This whole diagnostic suggests that model is fitted very well.

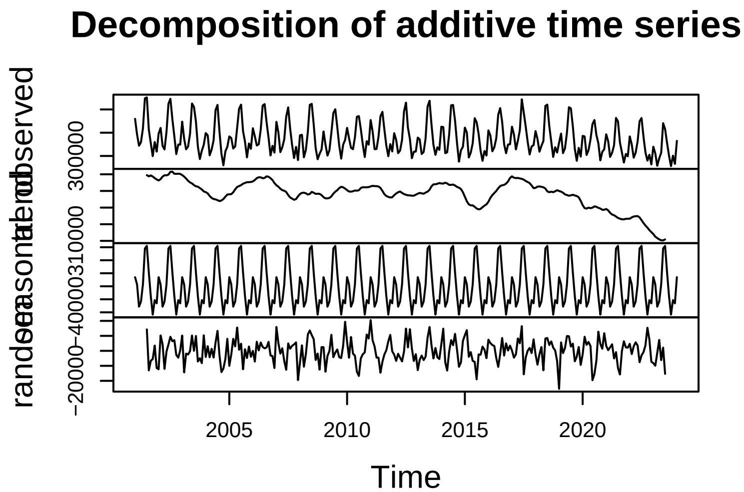

Figure 4: The decomposition of The US Net Generation

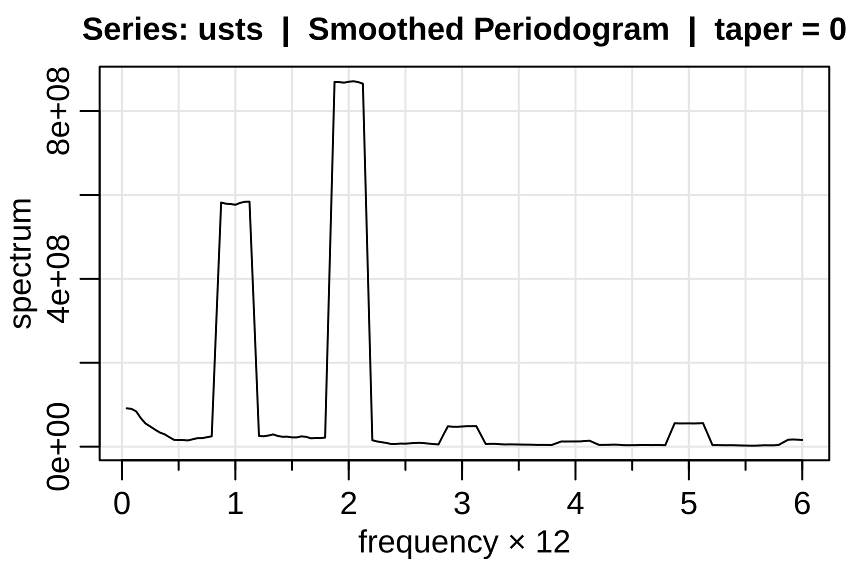

Figure 5: The Periodogram of The US Net Generation

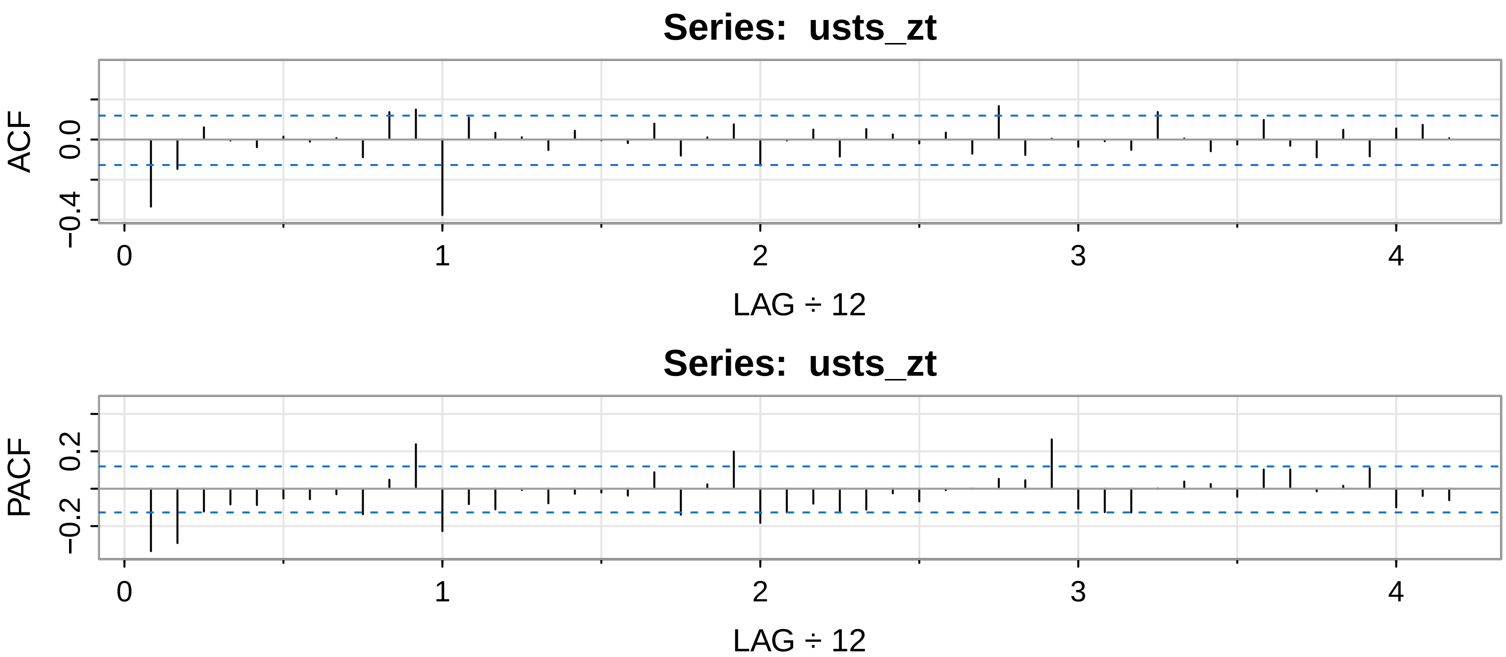

Figure 6: The ACF and PACF of The US Net Generation

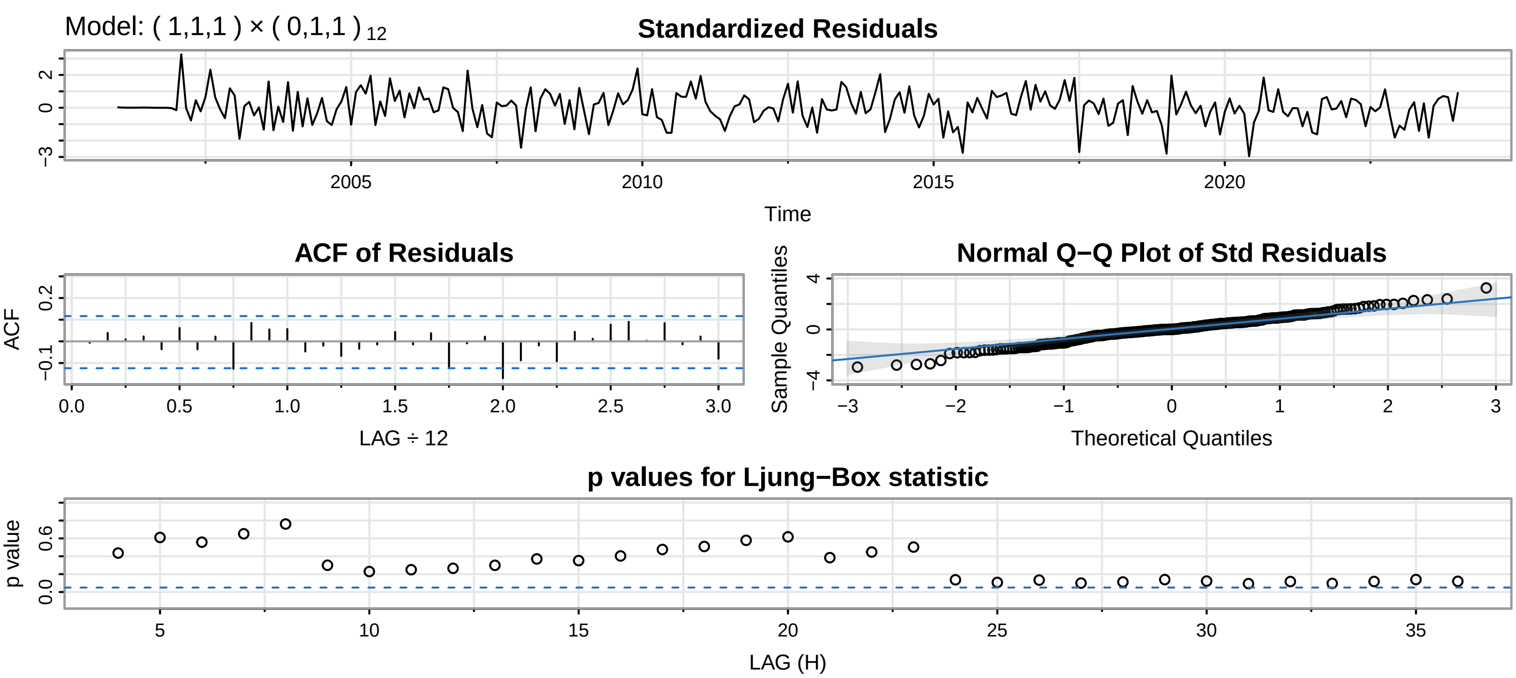

Figure 7: The diagnostic of the US Net Generation

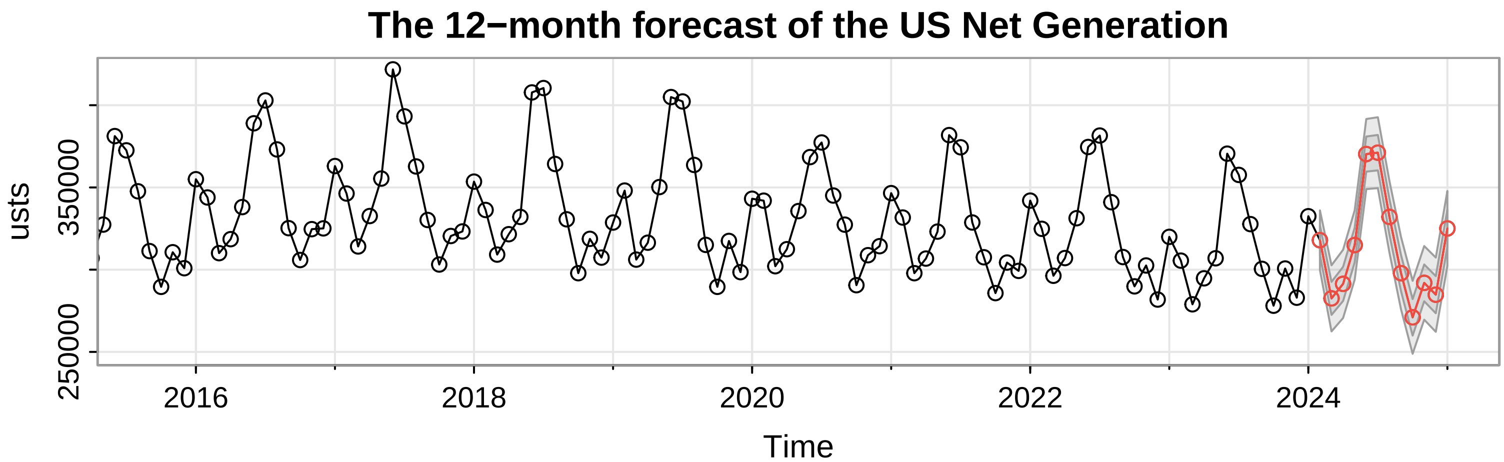

Figure 8: The forecast of the US Net Generation

After fitting the SARIMA model and checking the diagnostic, the next step is to forecast the next 12 months. The forecast plot Figure 8 shows the forecast of the next 12 months and the 95% confidence interval. The forecast shows that the net generation will continue to slightly decrease following the previous downward trend. The 95% confidence interval is quite narrow which suggests that the forecast is quite reliable. As we can observe, the red point forecast generally follows the regular seasonal pattern very well: going up to a peak and going down. This shows in general the US net generation is produced more during the summer and less during the winter. The US net generation is quite stable and predictable.

2.2 Texas Net Generation

The second part is to analyze the net generation in Texas. Due to the page limit, I only describe the main steps and analysis results in this report. The plots are suppressed. The codes and plots are available in the R quarto source file.

The first step is simply to visually check the time series plot. For the Texas net generation, we can see that the time series plot seem stationary but with slightly upward trend. However, we also need some quantitative analysis to confirm this. The first thing I do is to split the time series into two parts , the first half and the second half. Then I calculate the mean and standard deviation of each part. The mean of the first part is 33398 and the mean of the second part is 39525. In the measurement of unit, these two numbers are close but the second half is higher than the first half. The standard deviation of the first part is 5063 and the second part is 6094. The standard deviations are relative close. This is a good sign that the time series is weakly stationary. The next step is to use the Augmented Dickey-Fuller test to test the stationarity. As we can see, the p-value is smaller than 0.01 which means we can reject the null hypothesis. The time series is stationary.

Since the time series is stationary, it is not necessary to transform the Texas Net Generation. We can directly fit the SARIMA model to it. The next step is to decompose it and detect any seasonality and trend. I use the additive model to decompose it because the variance of it does not change significantly. The decomposition plot and the smoothed periodogram plot show there is a clear seasonality of 12 months as the US net generation. The periodogram also shows the 12-month period has the strongest signal. The overall trend is the stationary except for that there has a slight upward from 2017. So I still decide to take the first order difference of seasons and non-seasons to remove that slight upward trend.

After taking the first 1 order difference of seasons and non-seasons at 12, The ACF and PACF show the ACF and PACF estimation. At all non-seasonal lags, both ACF and PACF show tailing off to zero. This suggests that the first order of AR and MA, p = q = 1. At the seasonal lag, the ACF shows a cut-off at lag 2 and the PACF shows a tailing off to zero. This suggests the first order seasonal MA, Q = 1. In summary, the SARIMA model is SARIMA(1,1,1)(0,1,1)[12].

After fitting, the diagnostic plot shows the residuals of the model. The Ljung-Box test shows that almost all lag p-values are much greater than 0.05. This suggests that the residuals are not autocorrelated. This whole diagnostic suggests that model is fitted very well.

Finally the forecast plot is generated. The forecast plot shows that the Texas net generation will continue to increase following the previous upward trend and the seasonal pattern repeats.

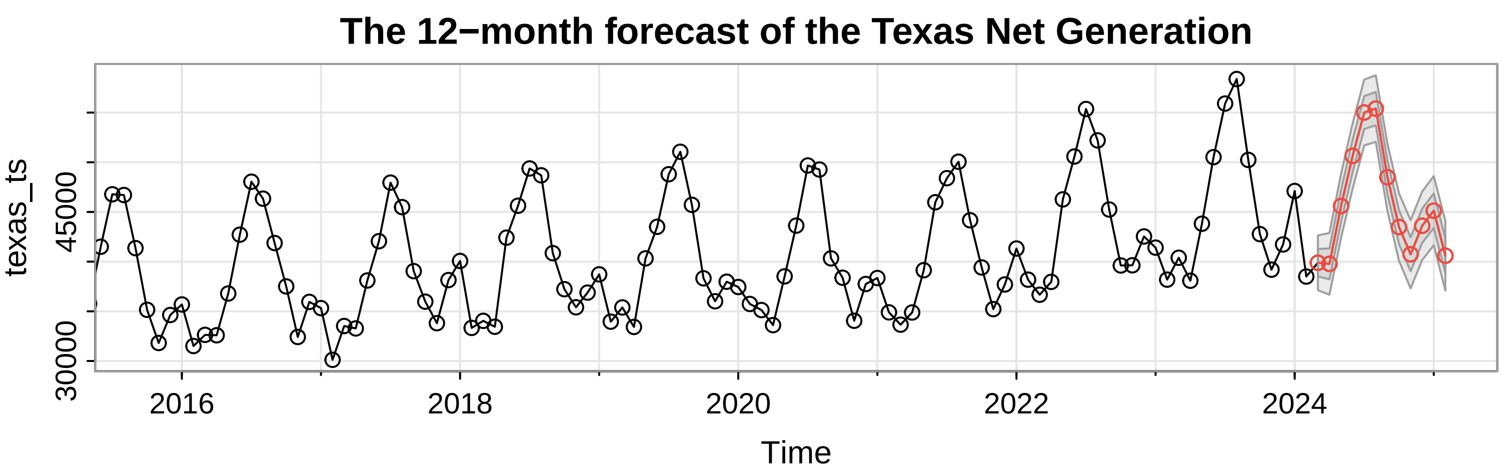

Figure 9: The forecast of the Texas Net Generation

3. Conclusion

The monthly US electricity net generation from January 2001 to January 2024 shows a very clear seasonal pattern of 12 months. The net generation is higher in the summer and lower in the winter. The US net generation is stable in general but with a slight downward trend. The SARIMA(1,1,1)(0,1,1)[12] model is fitted very well and the forecast by this model is reliable.

The monthly Texas electricity net generation from January 2001 to January 2024 also shows a obvious seasonal pattern of 12 months as the US one. The net generation is higher in the summer and lower in the winter as well. The US net generation is overall stable but with a slight upward trend. The SARIMA(1,1,1)(0,1,1)[12] model fits very well to the Texas data and the forecast by this model is reliable to predict the 2024 net generation.

There are a lot in common between the US net generation and the Texas net generation. Texas is the most productive state in the US, so it has strong influence on the US net generation and higher correlation. The only difference is that the US net generation is slightly decreasing while the Texas one is slightly increasing. As a result, the forecast of the US one is keep decreasing while the Texas one is keep increasing. It would be interesting to see if the US net generation will flip to increase in the future because of the Texas influence once more future data is available.