Wind Energy Generation Trend Analysis and Forecast

https://github.com/xueying-F/CaiFengYang_ENV790_TSA_FinalProject

Chuqi Cai, Xueying Feng, Longyi Yang

Contents

1 Rationale and Research Questions

There are increasing number of countries seeking ways to sustiana clean environment, and they developed national policy to reduce fossil fuel based energies and increase the integration of renewable based energy sources, for example, wind and solar power. In 2018, there are 95 percent of the new power installation are for renewable in Europe. The wind power energy generation in European counties meets 14 percent of the total electricity demand. The overall onshore capacity reached 160 GW which meets 12% of the total electricity demand, and the offshore wind capacity reached 18.5 GW.

The increasing demand for renewable wind energy raises interest in the economic issues related to the integration of wind energy into the power grid. The wind forecast could provide information including the onshore and offshore wind energy capacity, and the share of wind energy generation within the whole energy market. An accurate wind power forecasting could help decision makers to conduct better grid planning that integrates this renewable energy into the power systems, and to balance the power supply and demand.

Our research goal is to generate a 2021 forecast for onshore and offshore wind generation in Germany.

2 Database Information

We access the wind generation data from the website https://data.open-power-system-data.org/time_series/. This package includes wind and solar load and prices in hourly/half-hourly/quarter-hourly resolution, the geographical coverage of this package includes the EU and some neighbouring countries. This package is updated Oct.2020, and includes the 2015-mid 2020 data from TSOs and ENTSO-E Transparency.

2.1 Data Content Information

We download the dataset and choose the hourly resolution to conduct the forecasting. Among countries in EU, we choose Germany as a sample and focus on the three data field: actual wind generation, actual onshore wind generation and actual offshore wind generation. The summary of the data content we used in this project is listed in the table below.

Table 1: Wind Generation in Germany

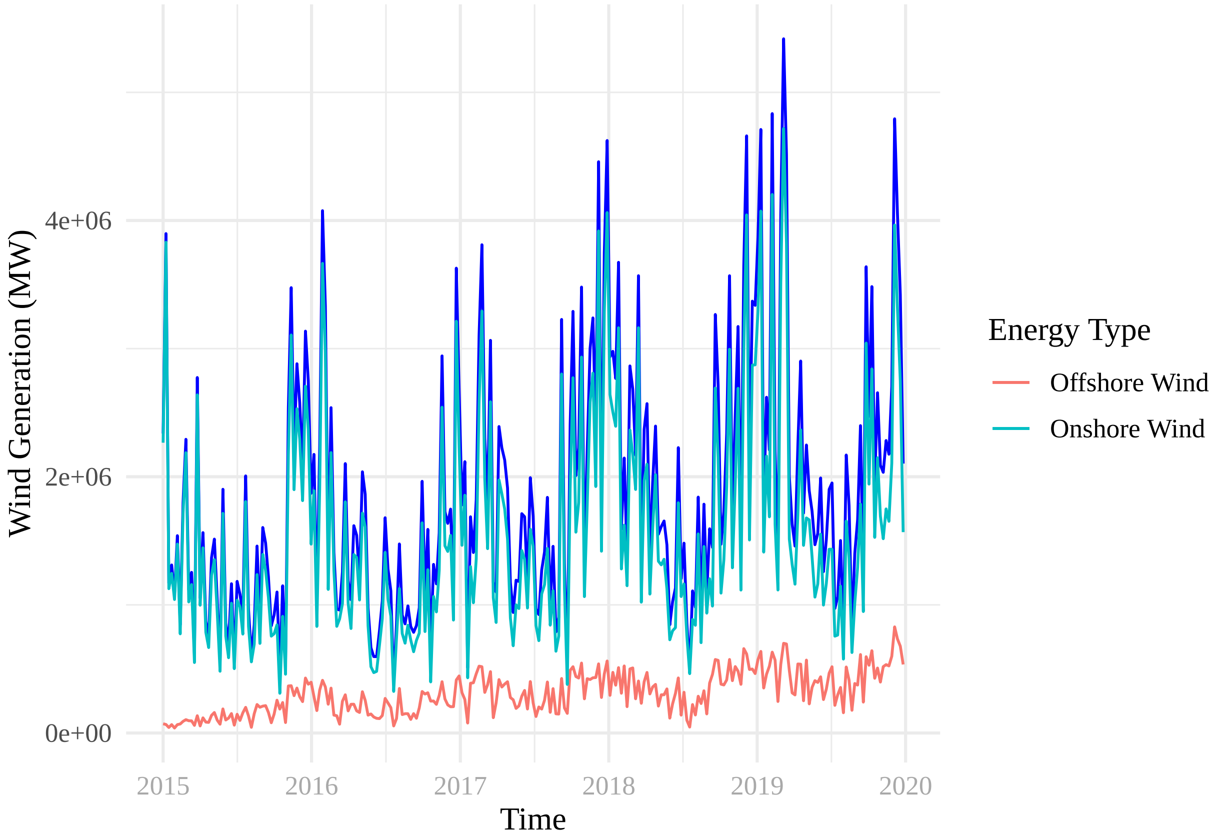
Data	Description
Time	2015 - 2020
Country	Germany
Resolution	Hourly (60 minutes)
DE_wind_offshore_generation_actual	Actual wind_onshore generation in Germany in MW
DE_wind_onshore_generation_actual	Actual wind_offshore generation in Germany in MW
DE_wind_generation_actual	Actual wind generation in Germany in MW

3 Analysis (Methods and Models)

3.1 Data Set Overview

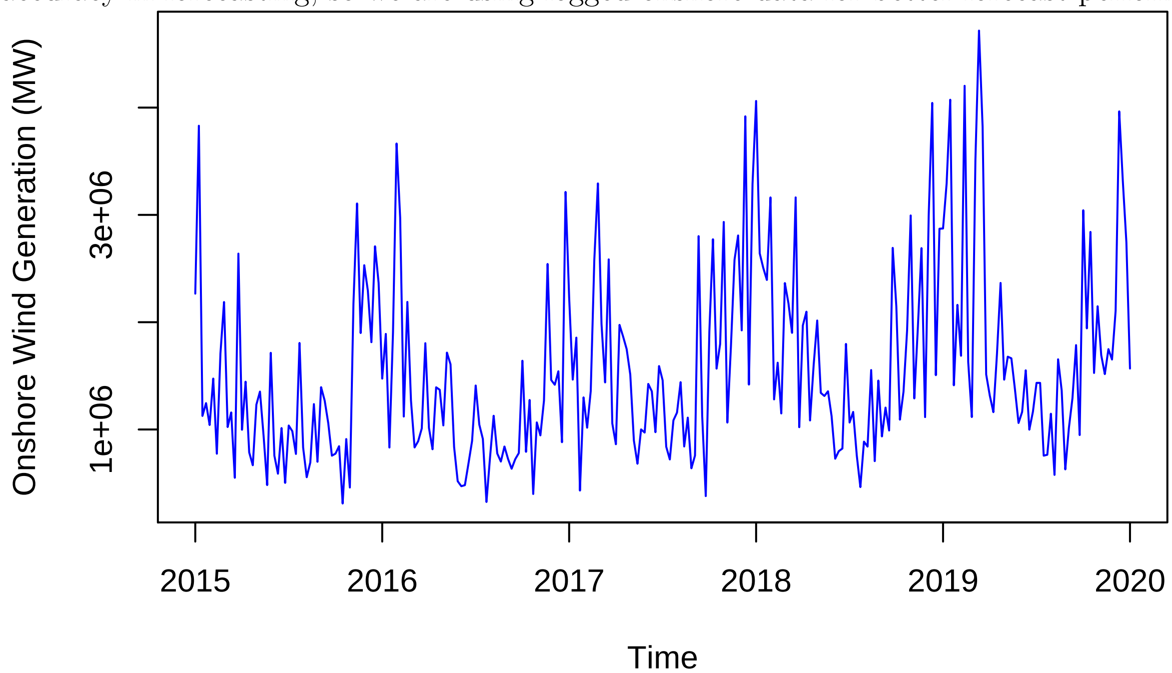
The first step of the project is to wrangling the raw data for these datasets that are useful for this research.The hourly resolution is too high for year-long scale forecasting, so we sum up the actual onshore and offshore wind generation by week, and conduct the time series analysis based on a frequency equals 52.

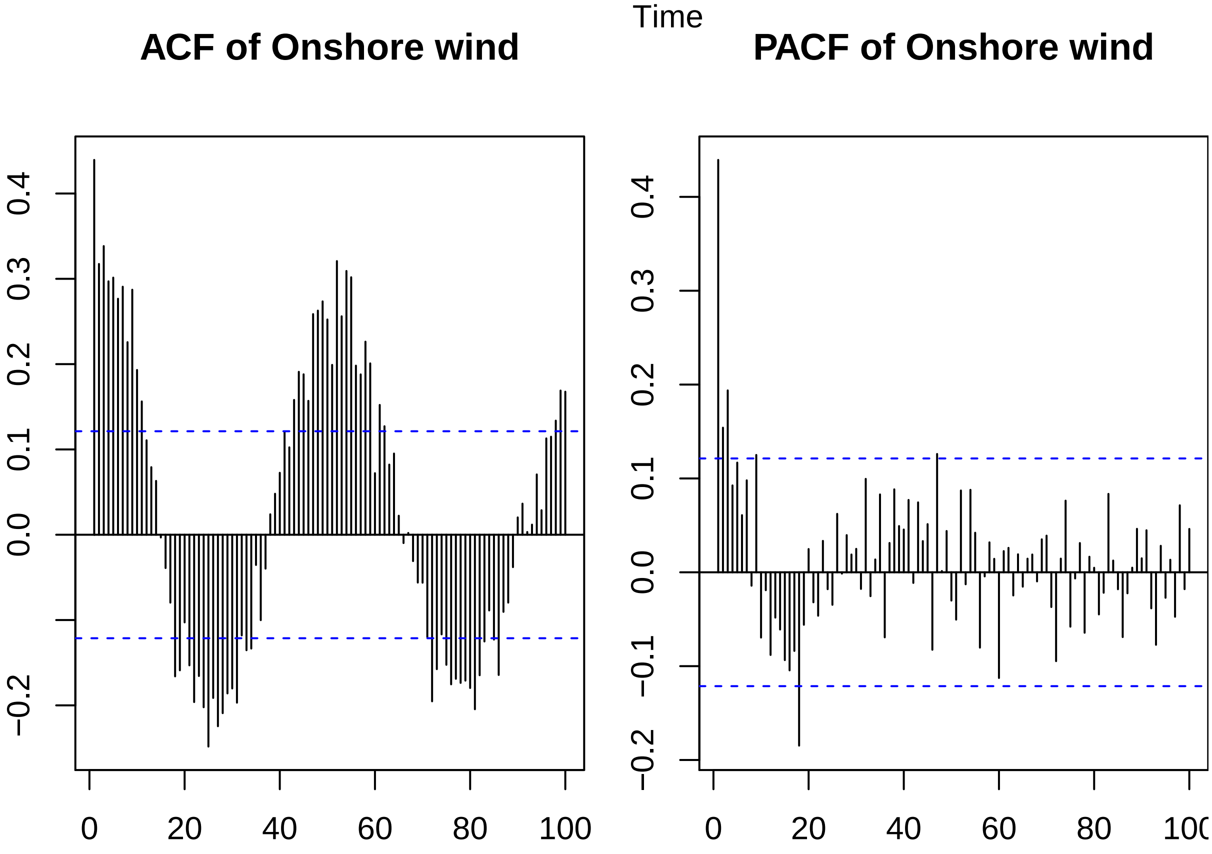
The Figure below shows the total wind generation in Germany from 2015 to 2019. The overall wind generation is the sum of the onshore and offshore wind generation, which is shown in dark blue. Since the offshore wind generation only takes a small portion in overall wind generation, the pattern of the overall wind generation is largely based on the onshore wind generation series. In this project, we conducted forecasting of onshore and offshore wind generation separately.



3.2 Onshore

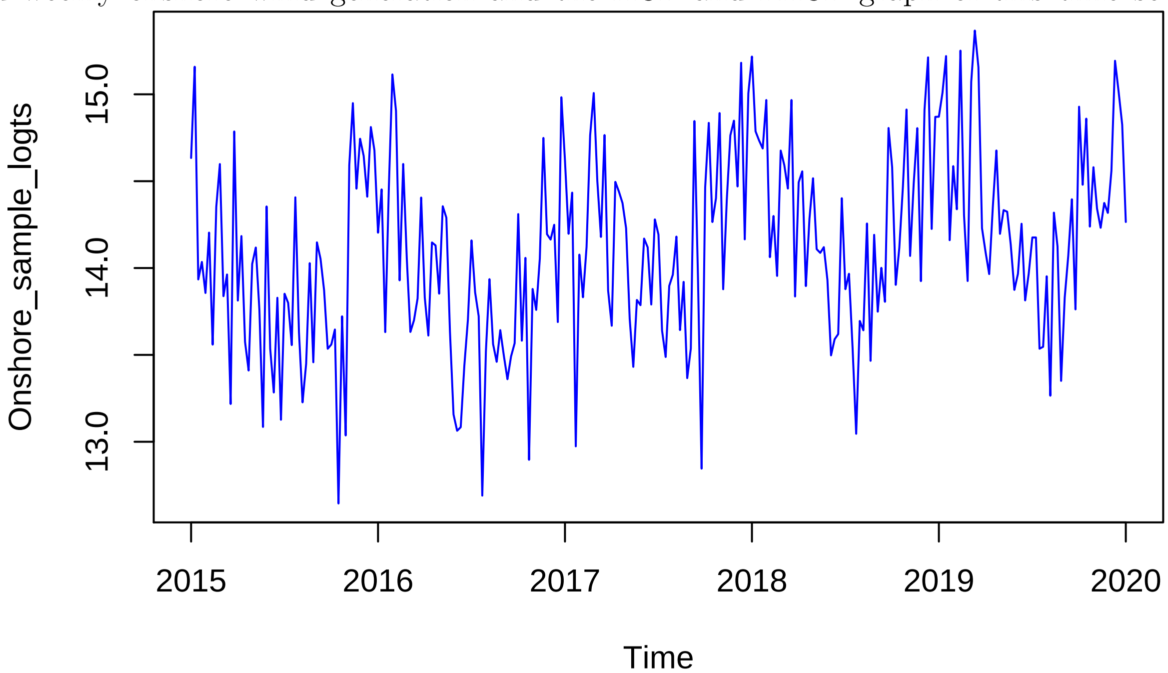
The original onshore wind data series is shown in the following figure, there is a clear overall trends over 5 year spend, and strong seasonality is shown in the figure. There is a slow decay in both ACF and PACF. However, the original data does not form high accuracy in forecasting, so we are using logged onshore data for better forecast performance.

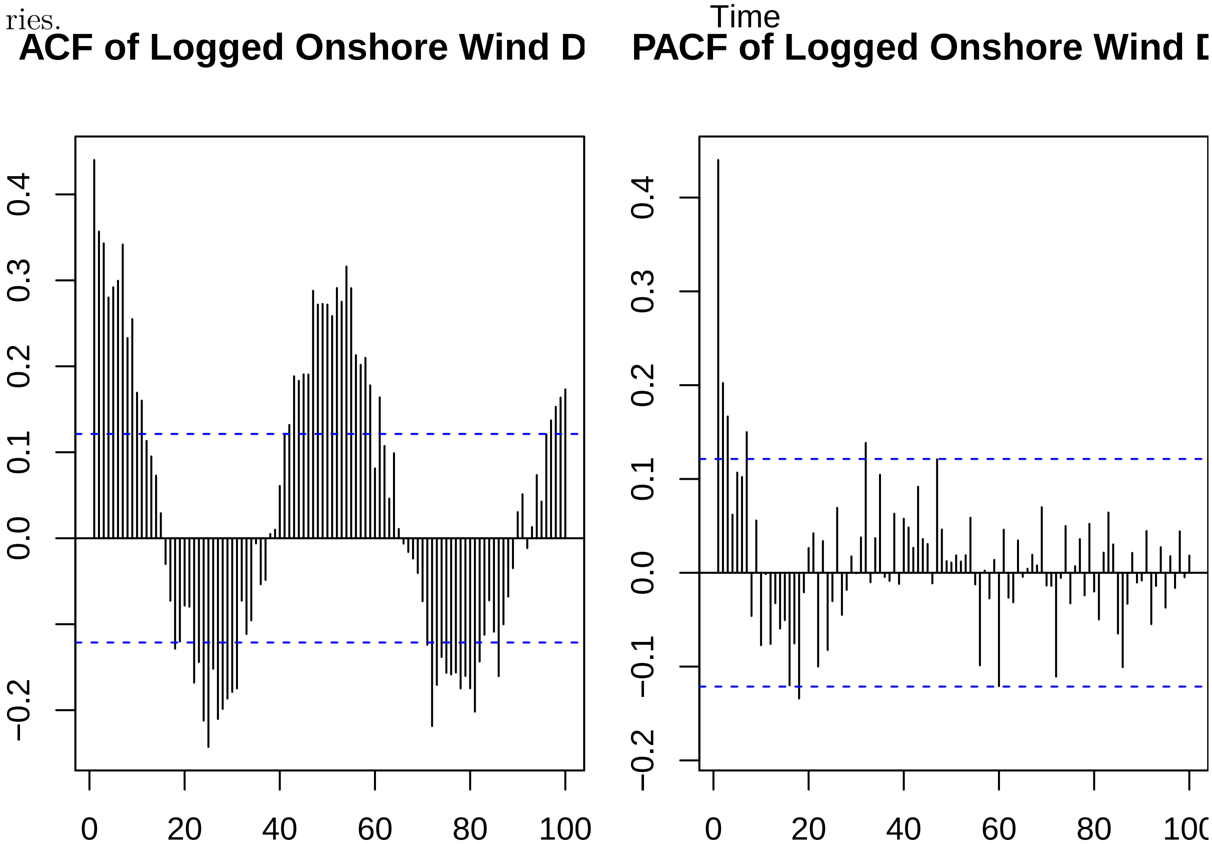




3.3 Onshore Wind data (log)

A multiplicative time series can be converted to additive by taking a log of the time series. We calculated the logged onshore wind generation, and the following figure is the logged weekly onshore wind generation and the ACF and PACF graph of this time series.





The decompose() splits the time series into seasonality, trend and error components.This step is to identify if a time series is additive or multiplicative: IF the variance in the graph is constant through out from central line then its additive else multiplicative. Additive model is used when the variance of the time series doesn't change over different values of the time series. On the other hand, if the variance is higher when the time series is higher then it often means we should use a multiplicative models. In this case, the size of the seasonal and random fluctuations change over time and the level of the time series, it is a multiplicative time series.

The following graph is the decomposition result of the onshore wind generation. The trend component is an increasing pattern. The random component is kind of randomness, so there still are some seasonality on that.

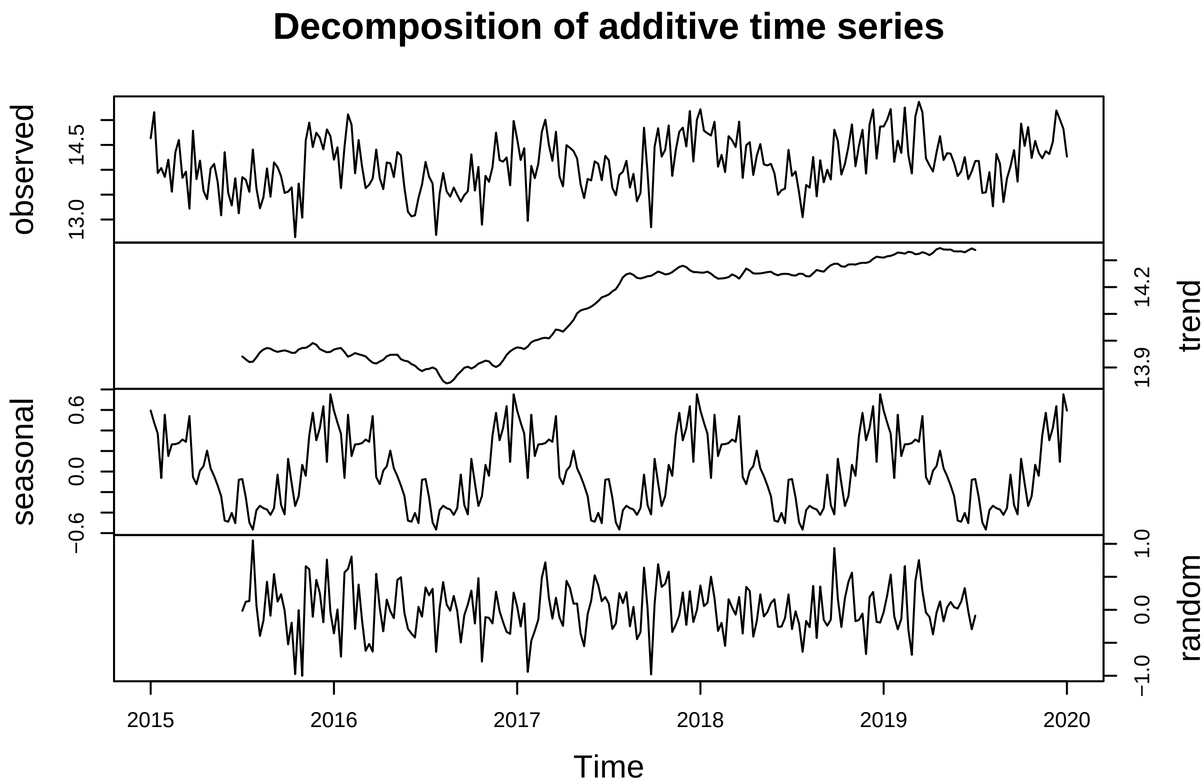
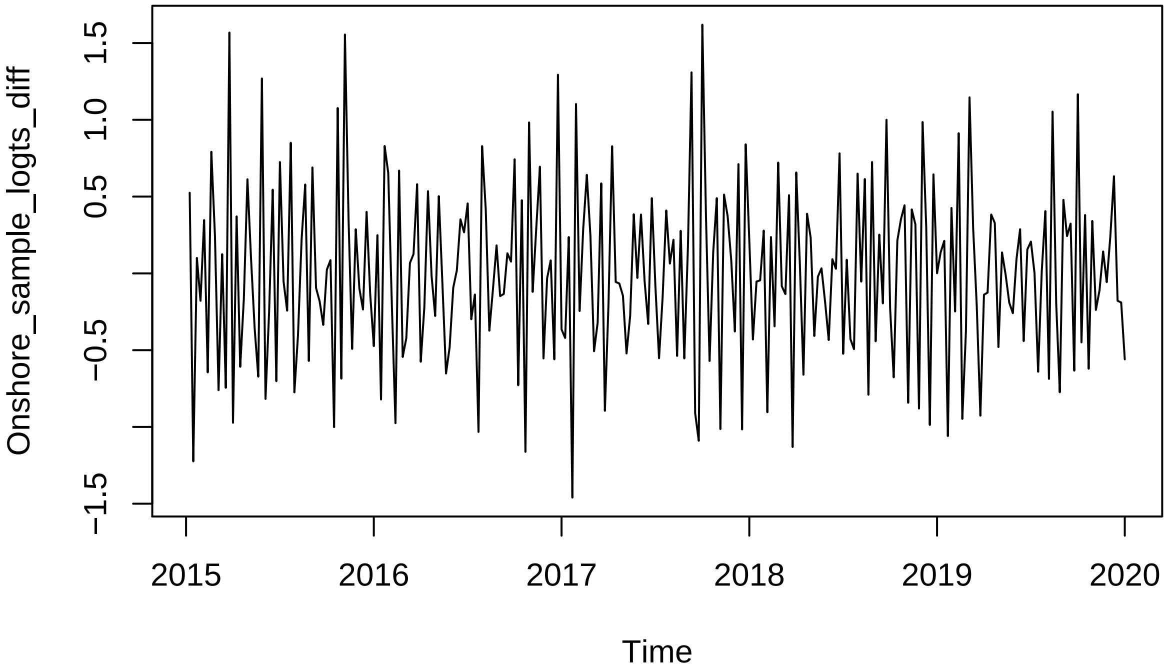


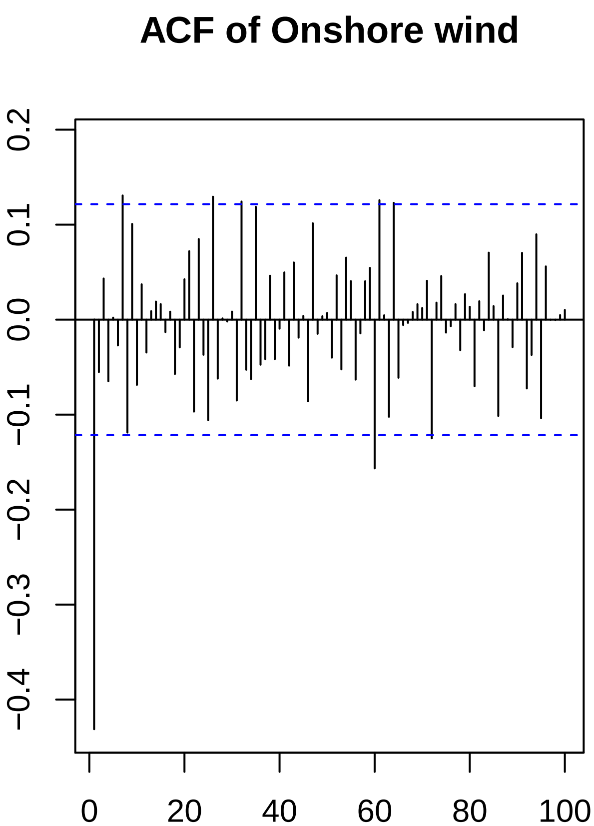
Figure 1: Decomposition of Logged Onshore Wind Generation

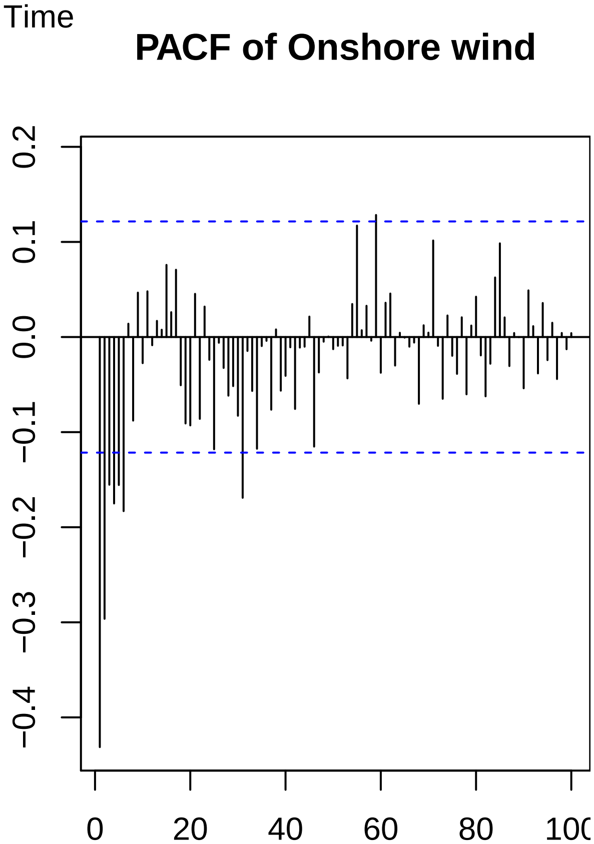
To test whether there is a drift or a trend for the return, we use the Augmented Dickey–Fuller (ADF) test.Start by running ADF to check for unit root. Unit root is related to stochastic trend. This test hypothesis setting is that : H0: data has unit root; H1: data is stationary. The result shows that p-value greater then 0.05, so we accept H0, meaning the series does have a stochastic trend (H0). Next step is to difference the series to remove the trend and we set hte difference of the data at lag 1

For the Mann Kendall test, since there is a seasonality in data, "seasonal Mann Kendall" is suitable for this case. The results shows that p-value less < 0.05, we reject null hypothesis. time and onshore wind generation have significant positive correlation, and data has a derterministic trend.

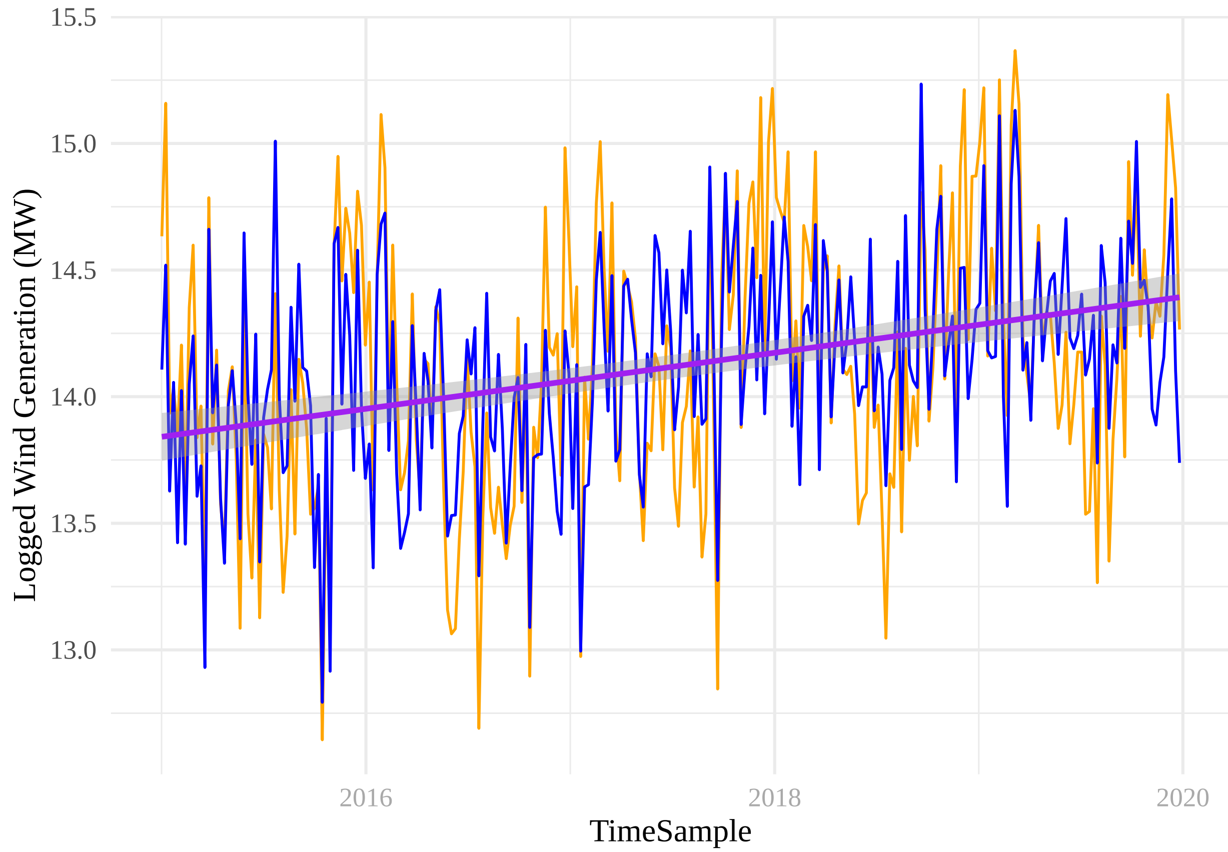
We run the ndiffs tool to determine the number of differencing needed, in this case, first-differencing is need to remove a liner trend. Following is the logged onshore wind generation after first differencing and corresponding ACF and PACF.

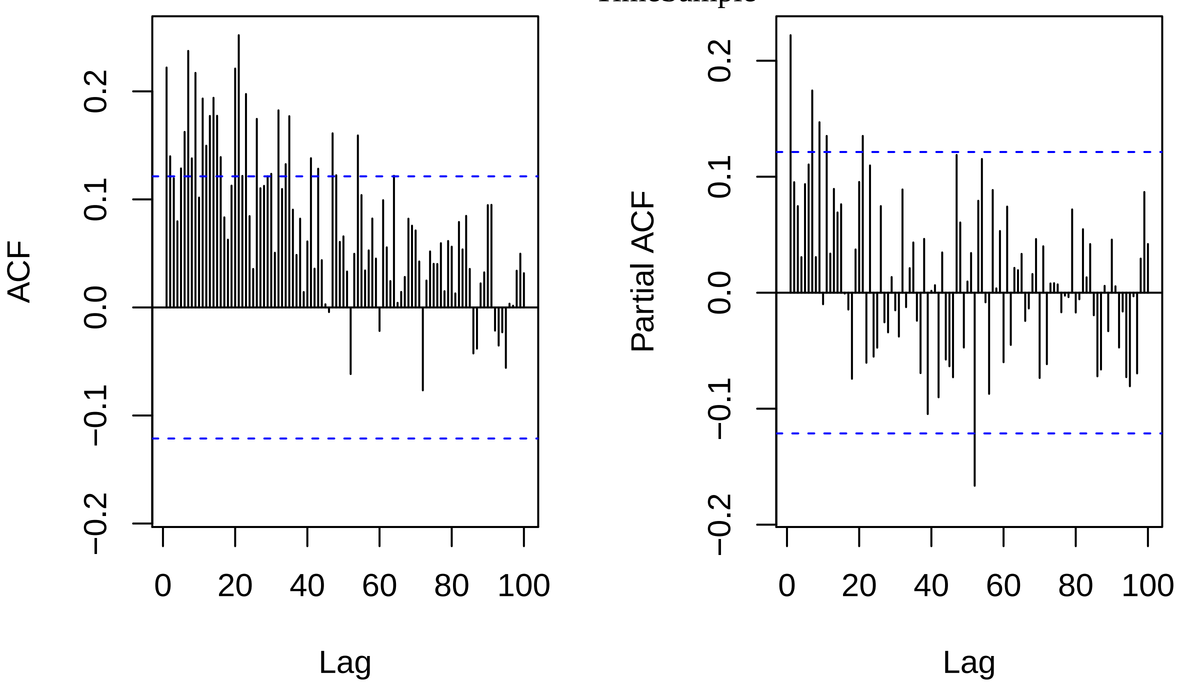






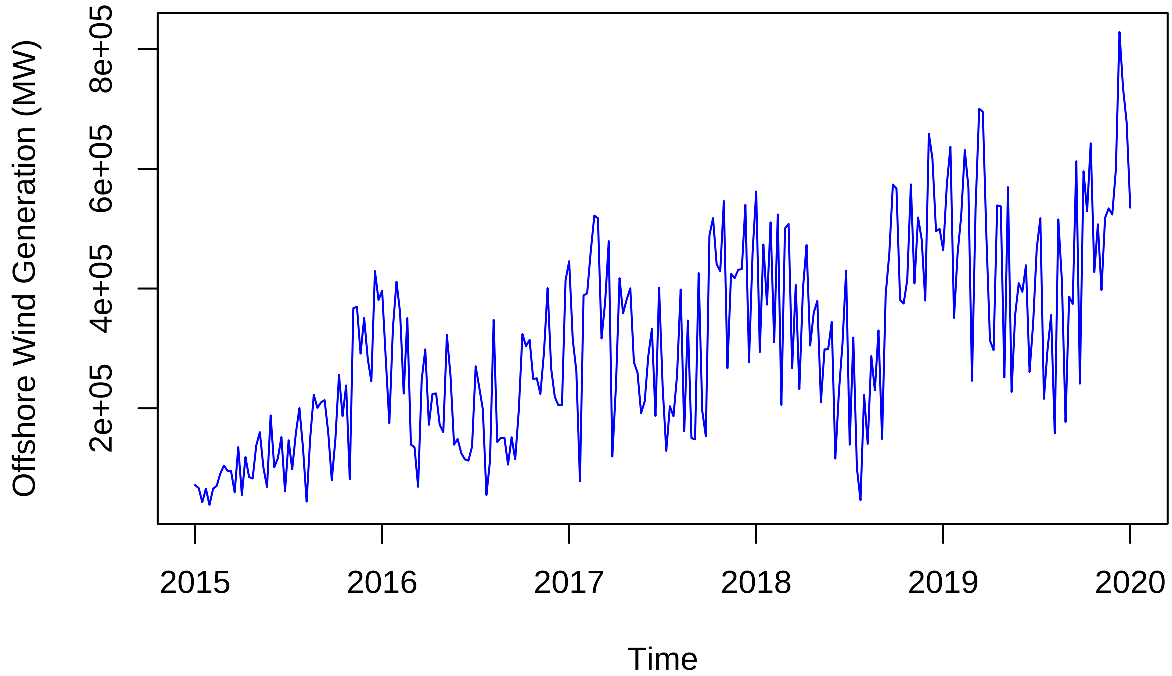
We also deseasonalize the logged data using the seasadj(), following figure shows the difference between the depersonalized logged onshore data and the original data.

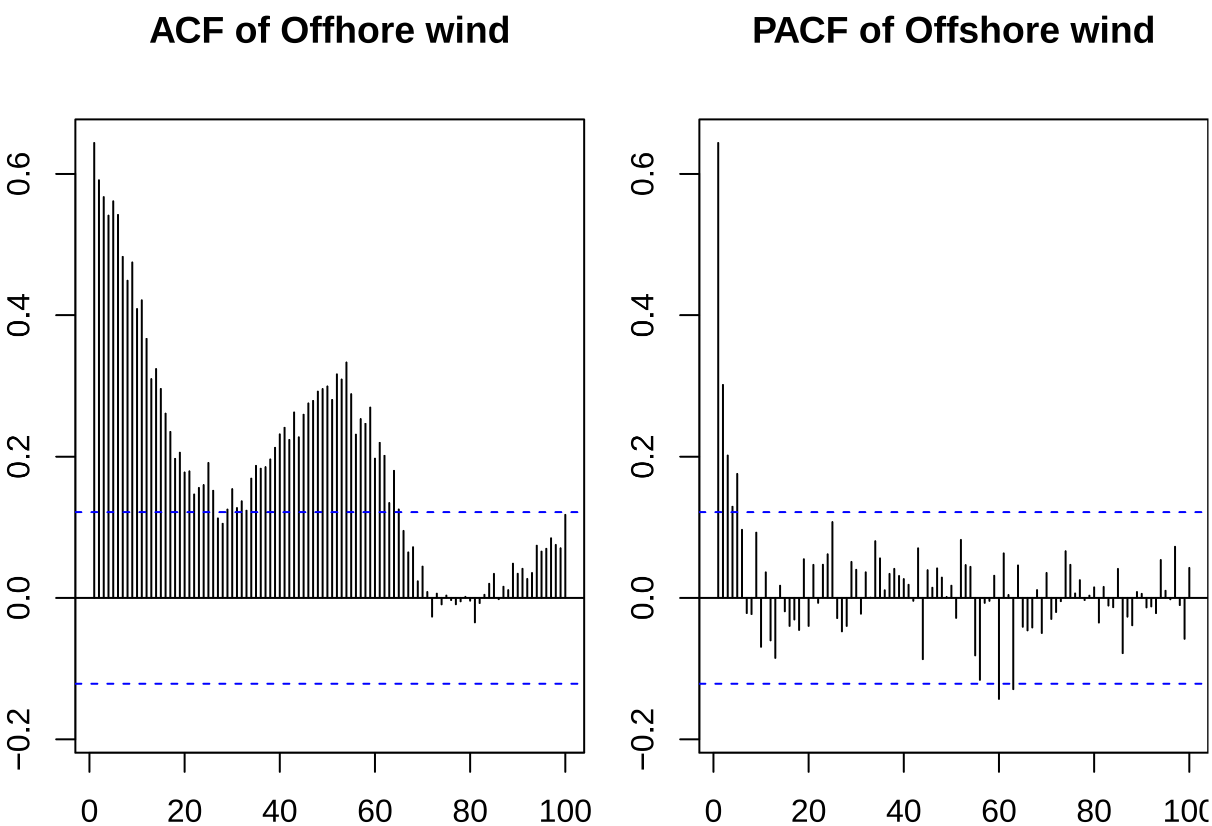

15.5
15.0
14.5
14.0
13.5
13.0
Logged Wind Generation (MW)
2016
2018
2020
TimeSample


ACF
0.2
0.1
0.0
−0.1
−0.2
0
20
40
60
80
100
Lag
Partial ACF
0.2
0.1
0.0
−0.1
−0.2
0
20
40
60
80
100
Lag

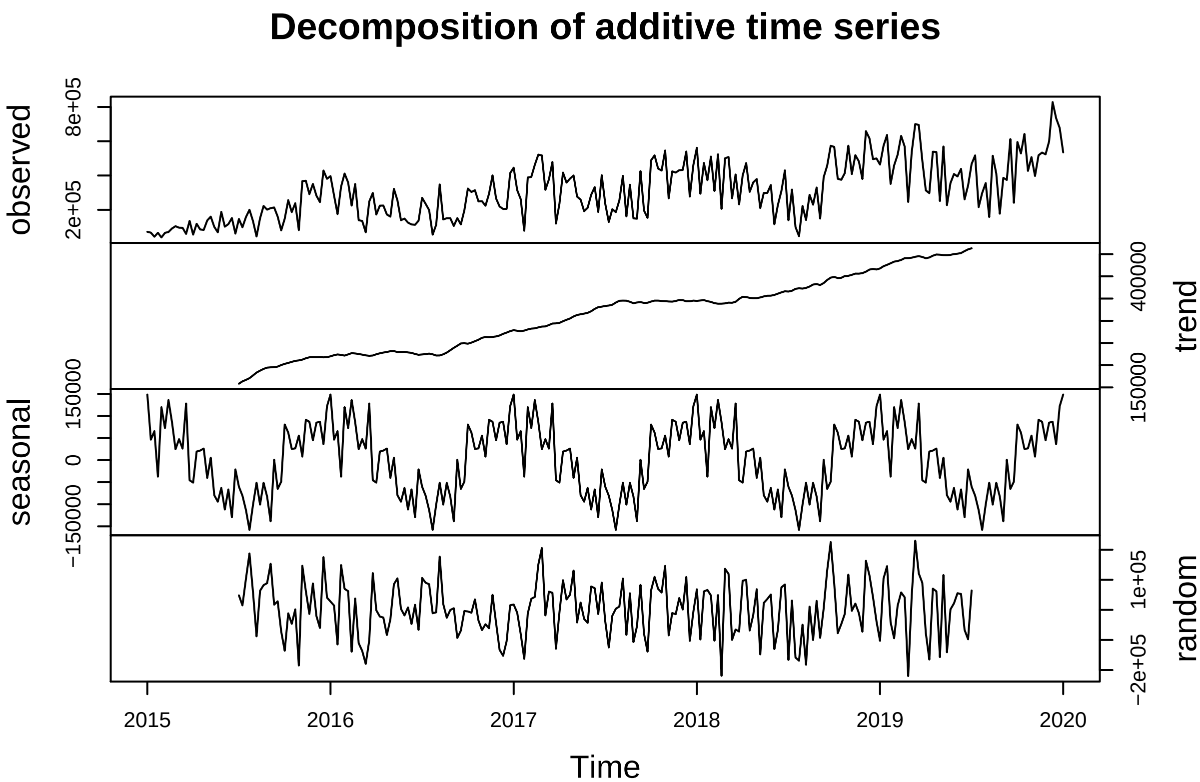
3.4 Offshore

The original offshore wind generation has shown in below figure. Since the offshore wind generation only make up a small portion of the total wind generation, the trends is not obvious when its plotted with the onshore wind. From the plot of the offshore wind generation data, there is a clear increasing trend but the seasonality is not clear. There is a slow decay in both the ACF and PACF graph.



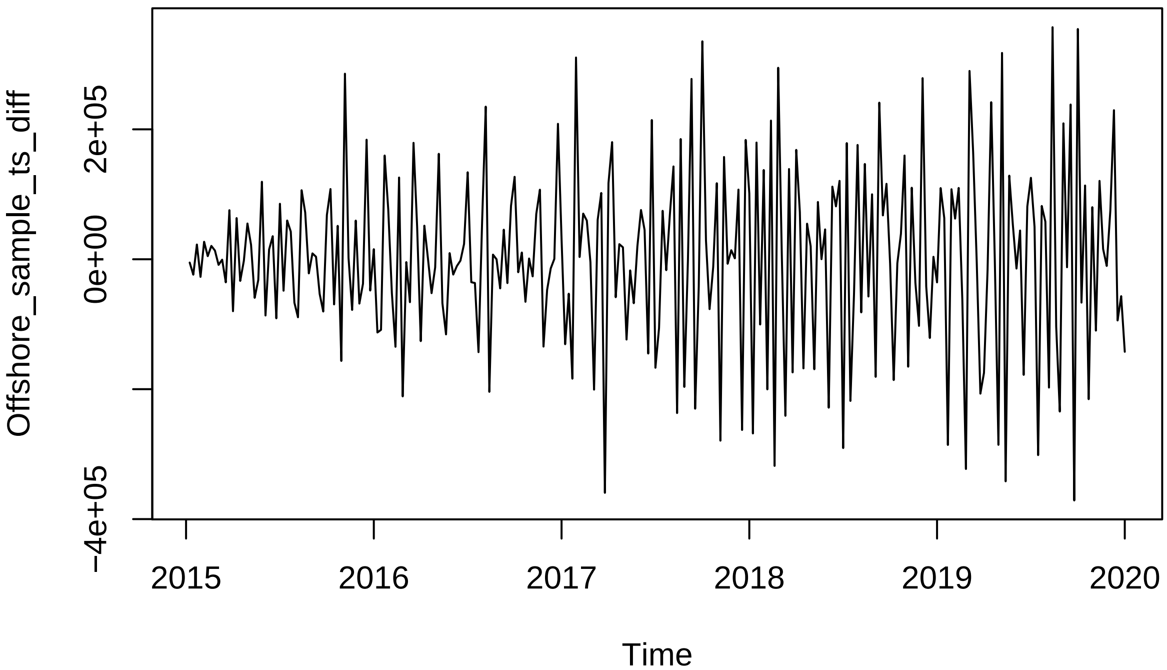


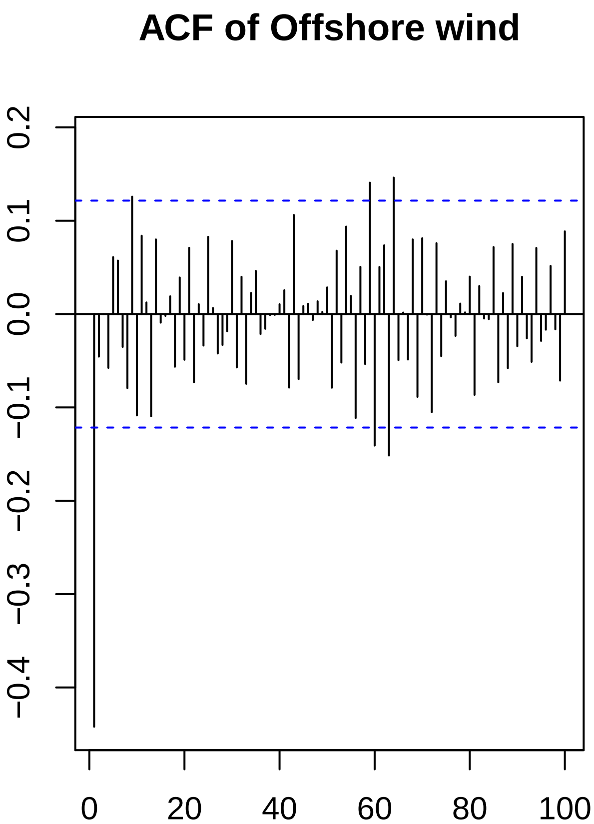
The following graph is the decomposition result of the onshore wind generation. The trend component is an increasing pattern. The random component is kind of randomness that looks like white noise.

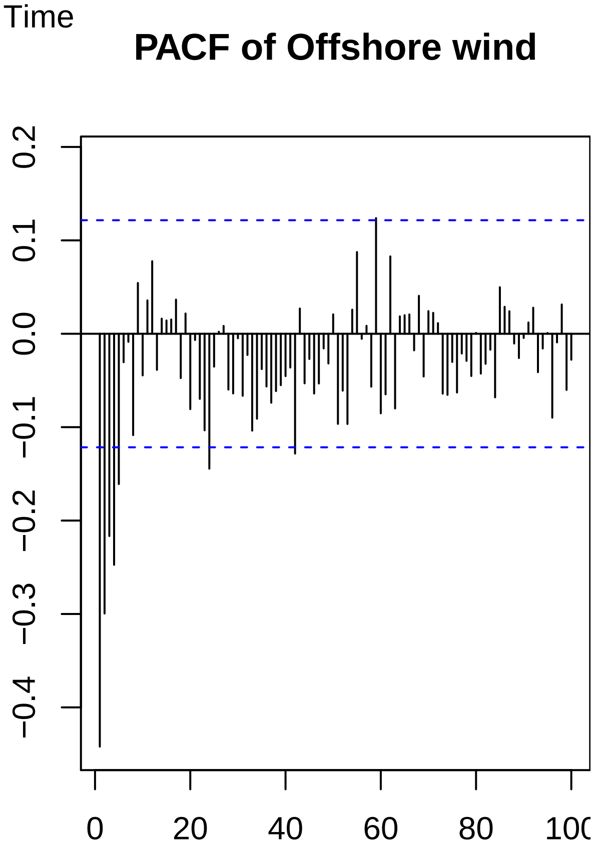


To test whether there is a drift or a trend for the return, we use the Augmented Dickey–Fuller (ADF) test.The result shows that p-value greater then 0.05, so we accept H0, meaning the series does have a stochastic trend (H0). Next step is to difference the series to remove the trend. For the Mann Kendall test, pvalue less < 0.05, so it's safe to reject null hypothesis. Data has a deterministic trend.

We run the ndiffs() tool to determine the number of differencing needed, in this case, first-differencing is need to remove a liner trend. Following is the offshore wind generation after first differencing and corresponding ACF and PACF.

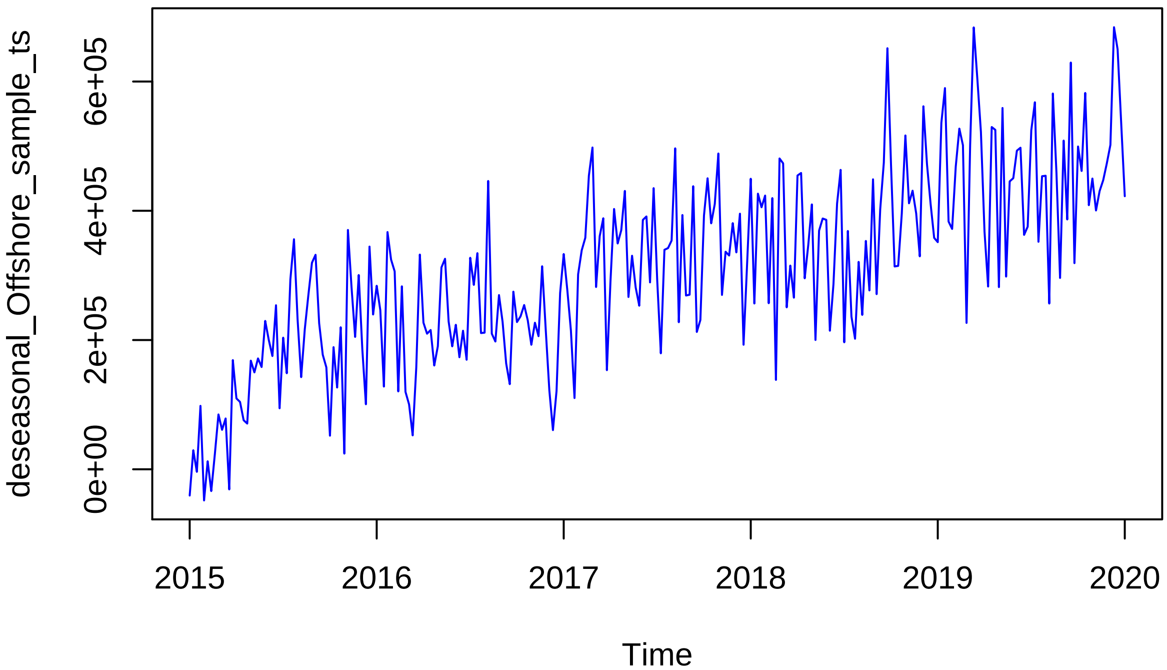


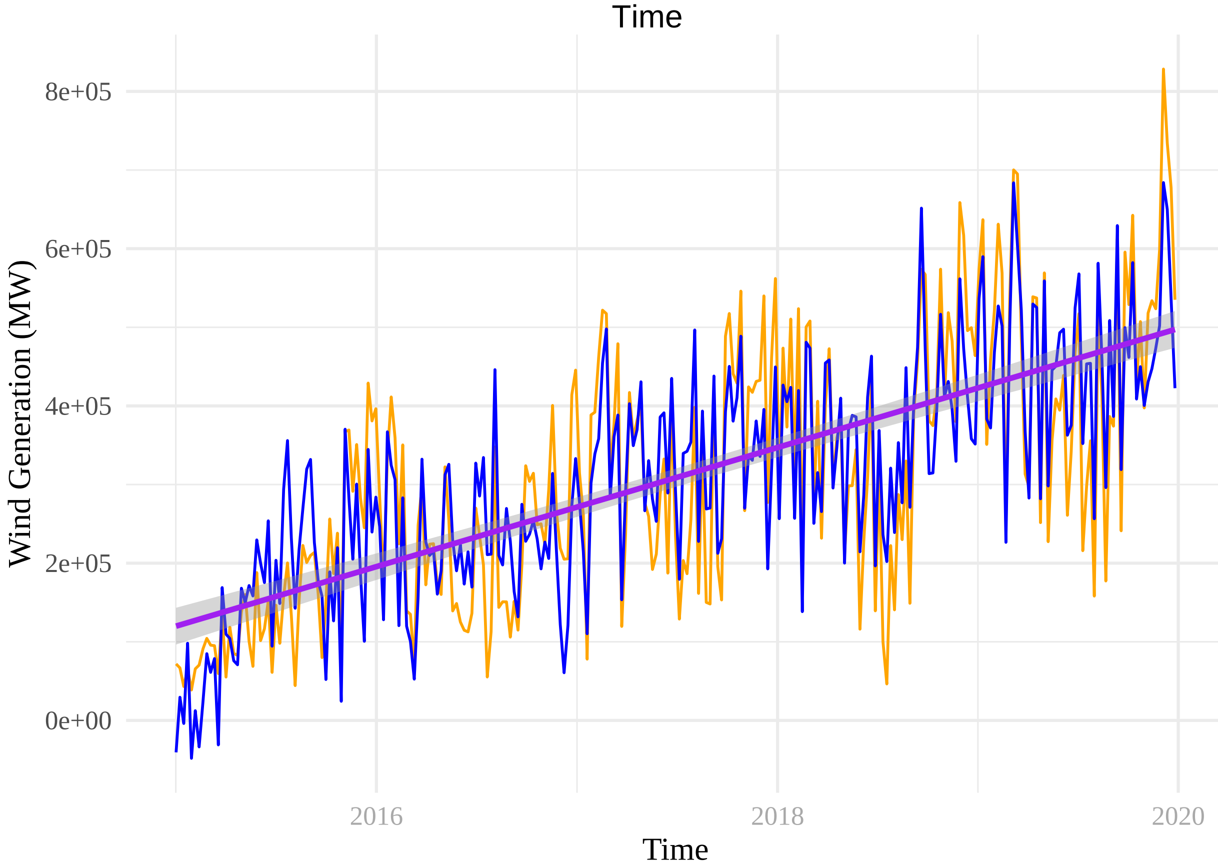


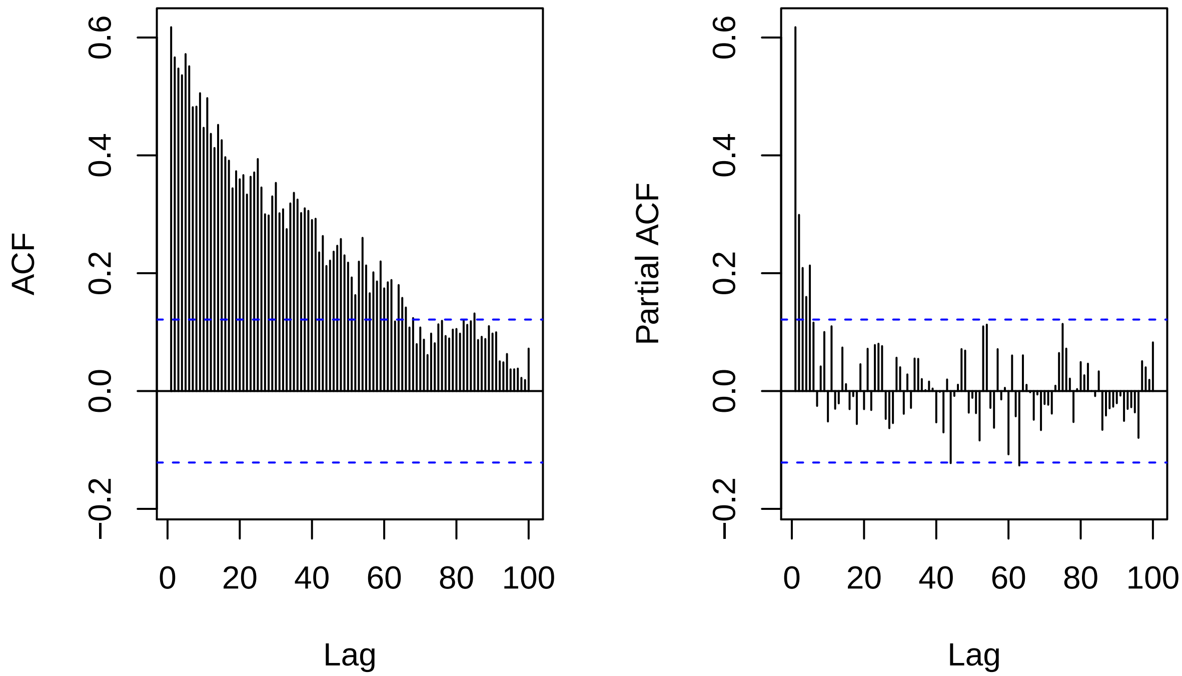


We also deseasonalize the offshore data using the seasadj(), following figure shows the difference between the depersonalized offshore data and corresponding ACF and PACF.

```
## [1] 0
```


deseasonal_Offshore_sample_ts
0e+00
2e+05
4e+05
6e+05
2015
2016
2017
2018
2019
2020
Time


Wind Generation (MW)
8e+05
6e+05
4e+05
2e+05
0e+00
2016
2018
2020
Time


ACF
Partial ACF
Lag
Lag
0.6
0.4
0.2
0.0
−0.2
0
20
40
60
80
100

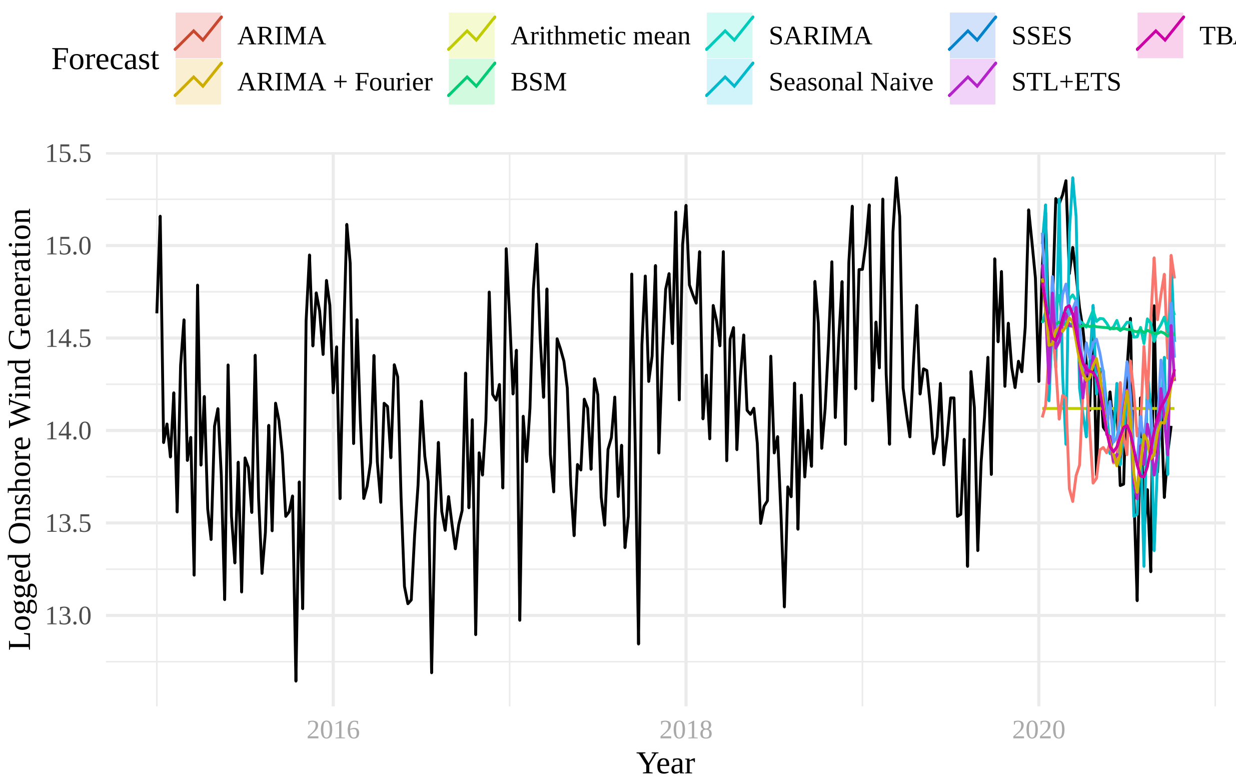
4 Modelling and Forecasting

To better perform the forecasting for 2021 wind generation, we fit the time series from 2015 to 2019 to different models, and conduct a forecast for 2020. We compare our forecasting with the actual wind generation in 2020 to determine the best model that fit the onshore and offshore wind generationn time series.

4.1 Onshore Model and Forecasting

For the onshore forecasting, we are fitting models to the original (sersonal) series, we fit Arithmetric mean model using meanf(), Seasonal naive model using snaive(), ARIMA model and SARIMA model using auto.arima() to perform the forecast. We also fitted State Space Models to the original (seasonal) series, one with StructTS() and another with Exponential smoothing using es(). For better performance, we fit Complex Seasonality models to the original (seasonal) series, we have the STL + ETS, and stlf() function could fit the time series to the model. And since we have multiple seasonalities, the SARIMA model will not work. But we can work with an ARIMA model with Fourier terms for each seasonal period. And we also include TBATS, which is a trigonometric seasonal variation of BATS, and using the tbats() function to achieve it.

The following graph is all the forecast we made for onshore wind generation in 2020 and compare it to the original actual wind generation we get from the database. For more detailed look, we remove the arithmetic mean, sarima and BSM models which clearly have poor performance. The rest of the plot shows similar pattern with the true data.

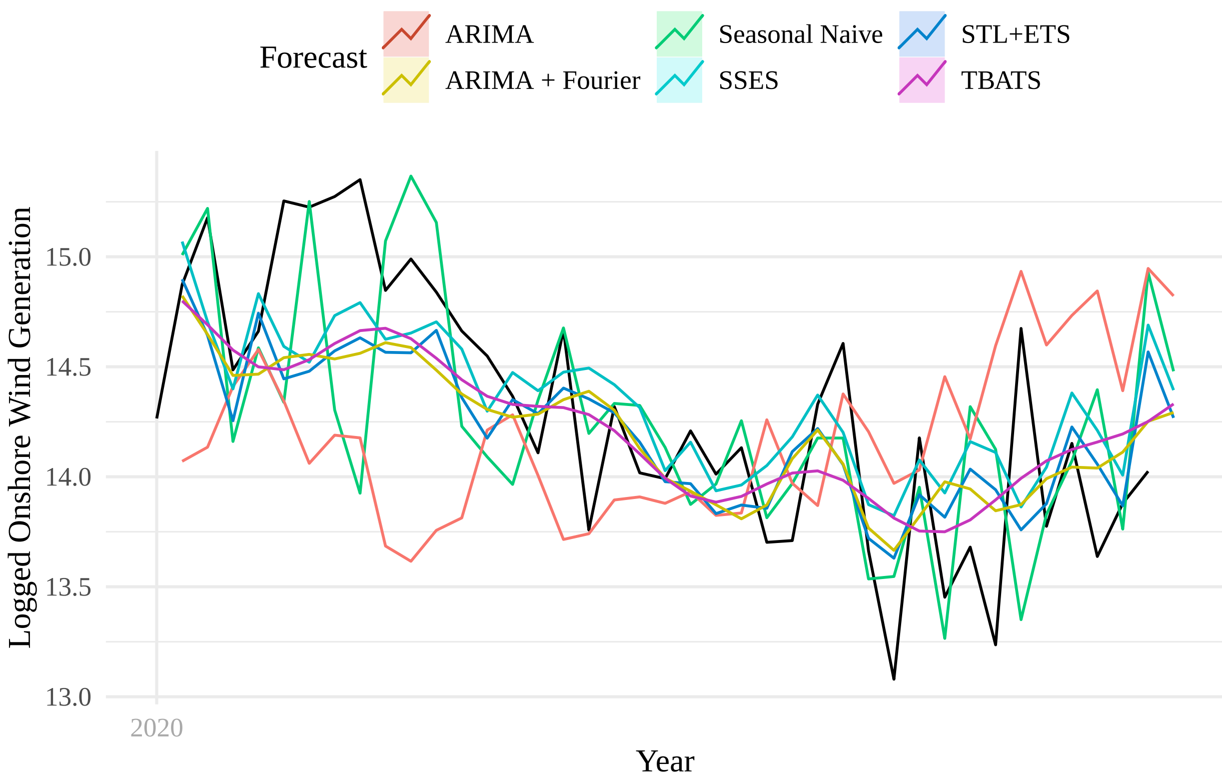


The TBATS, ARIMA + Fourier, SSES and STL+ETS shows similar pattern that mitigate

Table 2: Forecast Accuracy for Data (log)

	ME	RMSE	MAE	MPE	MAPE
SNAIVE	0.0334354	0.5752705	0.4625863	0.1350268	3.225512
ARIMA	0.1143417	0.7581439	0.6128471	0.6145558	4.268905
SSES	-0.0426122	0.4759530	0.3901595	-0.4115544	2.744201
STL+ETS	0.0849052	0.4781349	0.4077674	0.4841429	2.841899
ARIMA+Fourier	0.0919321	0.4409978	0.3743017	0.5379007	2.607984
TBATS	**0.0754294**	**0.4131354**	**0.3541056**	**0.4283995**	**2.471349**

the flutuation of the time series. The ARIMA model shows flatter estimation, which underestimate the generation at the beginning of the year, and overestimate in the later month. The graph shows more accurate in the plot.



To statistically analysis the accuracy of the forecast based on logged data, we generate a table indicating the error of each model. We also conduct the accuracy test for exponential logged data forecasting. Based on the results of these two tables, we choose the TBATS as our forecast model for onshore wind generation.

4.2 Offshore Model and forecasting

The offshore time series is fitted to the Sarima model using the auto.arima() fuction. And the offshore time series can also fitted to the State Space Models, one with StructTS() and another with Exponential smoothing using es(). For better performance, we fit Complex

Table 3: Forecast Accuracy for Data (exp)

	ME	RMSE	MAE	MPE	MAPE
SNAIVE	111767.9	1120180.3	830979.8	-12.6542399	47.18510
ARIMA	358158.8	1442861.0	1084025.6	-18.5732032	66.67605
SSES	133309.6	953331.6	716807.0	-16.7401083	44.03144
STL+ETS	340621.4	993684.4	746566.0	-2.5648842	40.34037
ARIMA+Fourier	364751.0	942192.9	690256.6	-0.1457860	36.54769
TBATS	**335258.7**	**883216.1**	**652066.0**	**-0.8206589**	**34.95617**

Seasonality models to the original series, we have the STL + ETS, and stlf() function could fit the time series to the model. And since we have multiple seasonalities, the SARIMA model will not work. But we can work with an ARIMA model with Fourier terms for each seasonal period. And we also include TBATS, which is a trigonometric seasonal variation of BATS, and using the tbats() function to achieve it.

The following graphs are all the forecast we made for offshore wind generation in 2020 and compare it to the original actual wind generation we get from the database. For more detailed look, only the data for 2020 are shown on the second map, and the BSM, Sarima and SSES model are removed due to poor performance. The Arima_founier model underestimate the first half year and overestimate the last half year. The STL+ETS and TBATS models are having same trends, but STL+ETS model have more flutuation.

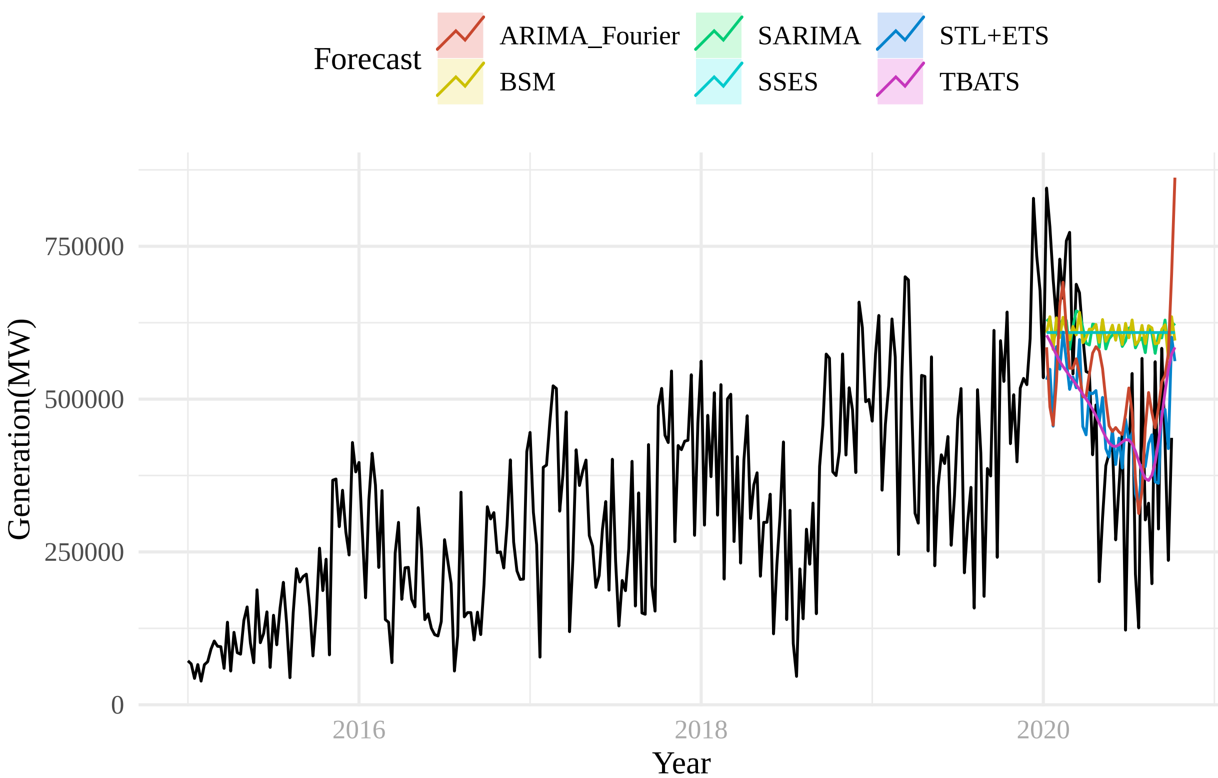
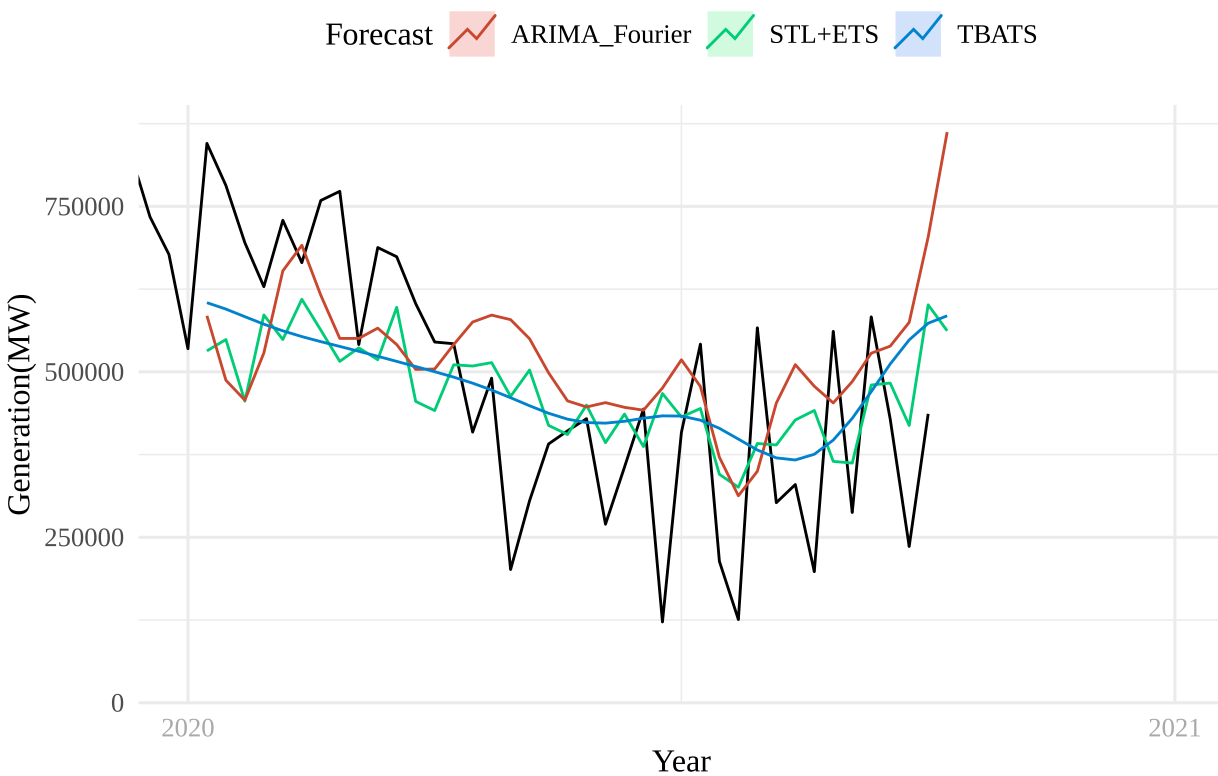


Table 4: Forecast Accuracy

	ME	RMSE	MAE	MPE	MAPE
SARIMA	-132875.150	228015.4	185457.3	-60.98677	67.92169
SSES	-134058.632	234383.7	192989.0	-62.40080	70.22752
BSM	-134541.071	232001.3	191281.8	-61.96696	69.47020
ETS	**6282.420**	**158142.4**	**131381.7**	**-19.13028**	**39.74869**
ARIMA_FOURIER	-38995.182	180484.2	150601.3	-31.69249	47.97466
TBATS	-1794.138	153563.4	127930.9	-21.67029	40.57876

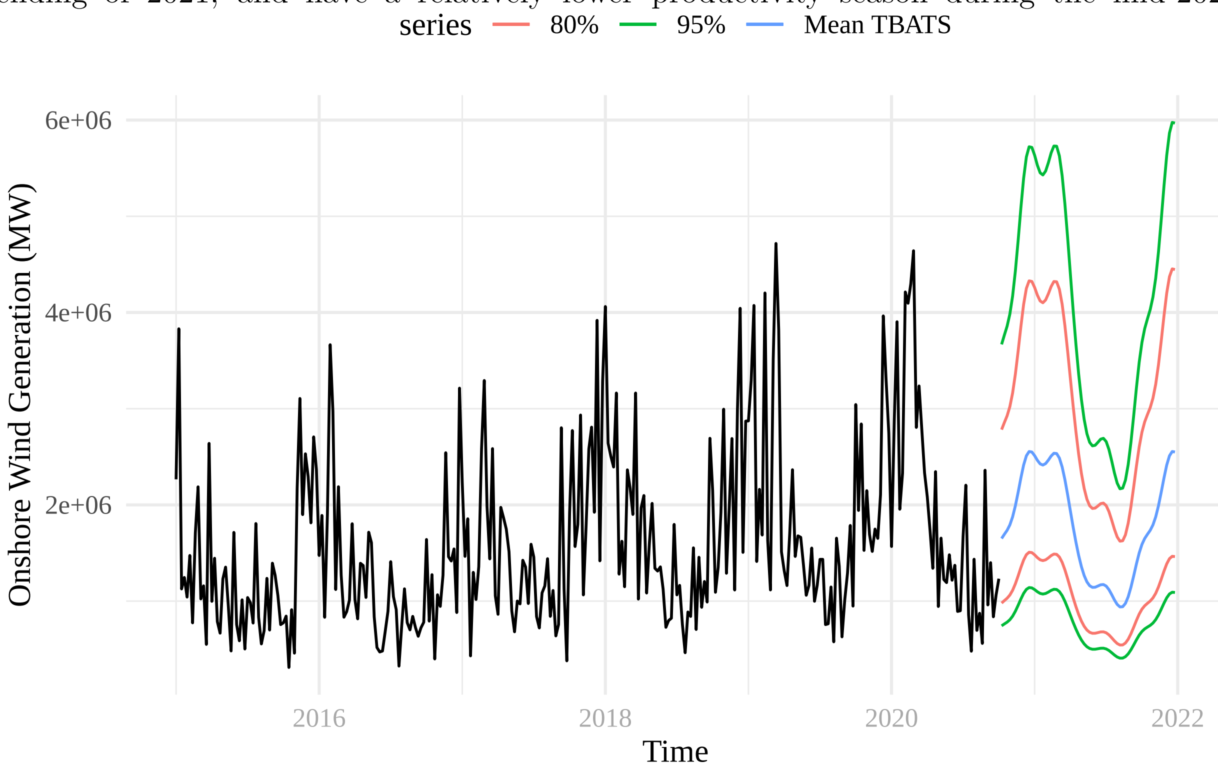


To statistically analysis the accuracy of the forecast based on the offshore wind generation data, we generate a table indicating the error of each model. Based on the results, we choose the TBATS as our forecast model for onshore wind generation due to the smallest Mean Absolute Percentage Error.

5 Forecast Output

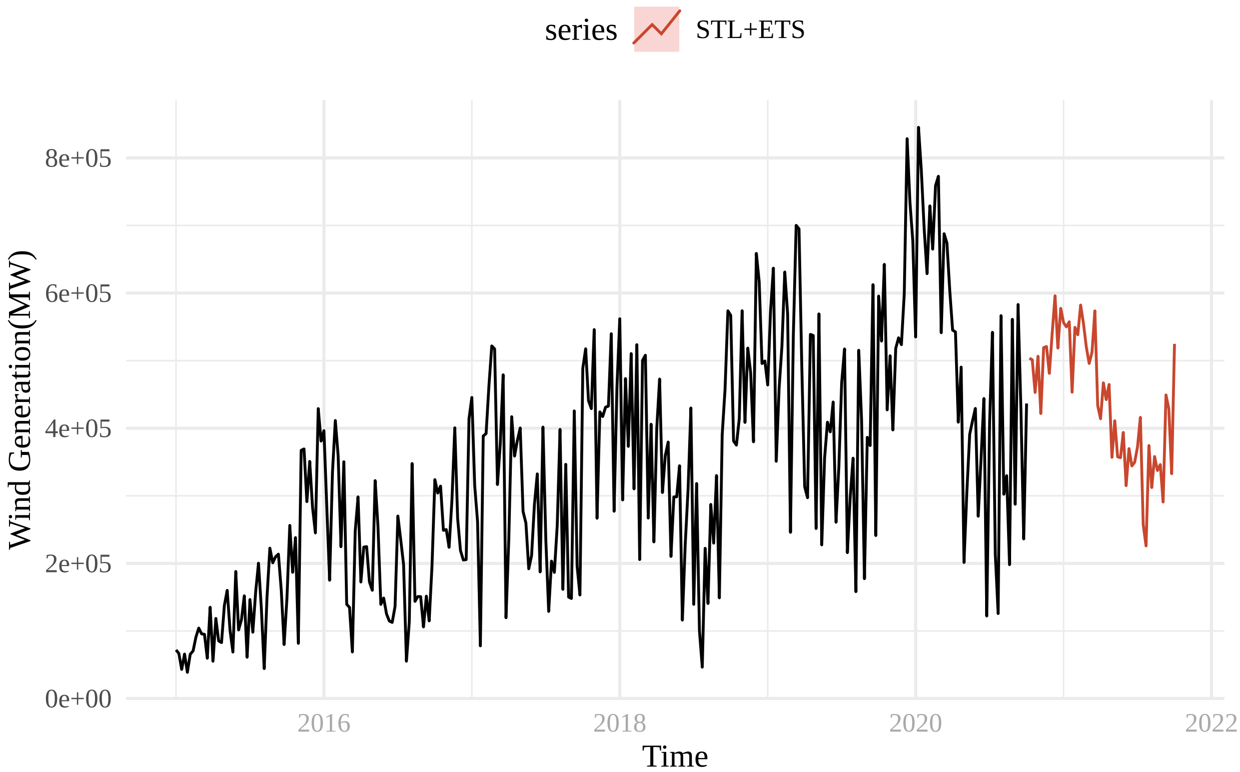
A 52 weeks forecast is conducted based on the actual onshore and offshore wind energy generation in Germany. ## Forecast Onshore Wind Generation in 2021

According to the analysis of the forecasting performance, we choose the TBATS model to fit the 2015 to 2020 time series dataset and conduct a forecast for the onshore wind generation in next 52 weeks after September 2020. The following figure shows our forecast for the onshore wind generation in 2021. It have strong performance with the seasonality but weaker performance in randomness and overall trends. According to our forecasting, the 2021 will have high wind generation at the beginning and the ending of 2021, and have a relatively lower productivity season during the mid-2021.



5.1 Forecast Offshore Wind Generation in 2021

According to the analysis of the forecasting performance, we choose the STL+ETS model to fit the 2015 to 2020 time series dataset and conduct a forecast for the offshore wind generation in next 52 weeks after September 2020. The following figure shows our forecast for the offshore wind generation in 2021. It have strong performance with the ramdomness of the data but have less strength in seasonality and overall trends. According to our forecast, the wind generation in late 2020 and early 2021 will be similar to the mid-2020, and it will decrease during the mid-2021, and acsending back to the early 2021 level at late 2021.


series
STL+ETS
Wind Generation(MW)
8e+05
6e+05
4e+05
2e+05
0e+00
2016
2018
2020
2022
Time

6 Forecast Limitation

There are a few limitation of this project. The first one is that the model and forecast option we conducted in this project only works in Germany due to the climate variation. For other European countries, the time series data should be fitted to other model for more accurate forecasts. The second one is that because of the development of the renewable energy, there are considerable number of the new wind power infrustructure are built in Germany, and this is a influence on the overall trend of this set of data. The future installation is hard to forecast due to possible economic recession or pandemic, for example the COVID crisis. The last limitation that is important is the intermitte nature of the wind, we are not able to include the adjustment about the unexpected extreme weather in our wind generation forecast.