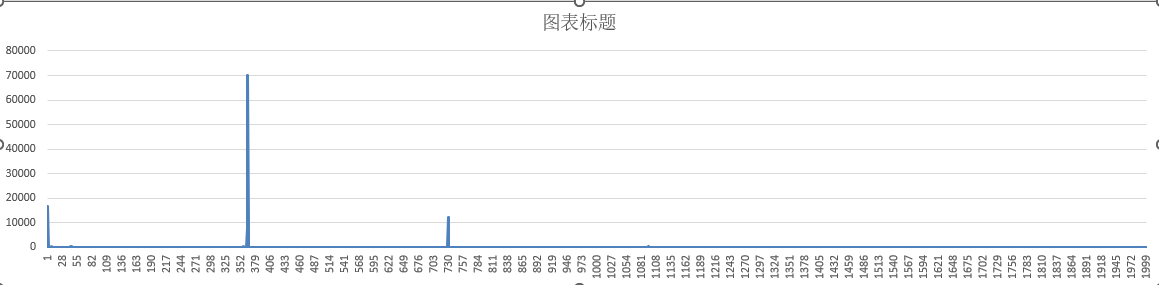
Assignment 1

Content

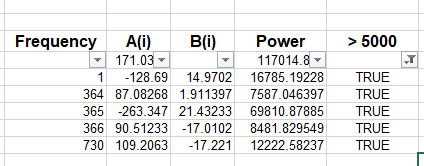
# Half Hour Solar Radiation Dataset

## Taks1: Getting Frequencies

Using powerspectrum excel to get the best frequencies. The parameters used in this case are, number of objects equals 17520 and number of frequencies is 2000. We got a graph like



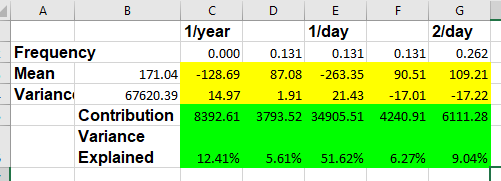
According to this graph, it could easily find that the minimum values of the most important frequencies are around 10000, we use filter power > 5000 to filter all the possible frequencies. The value could be got like the graph below.



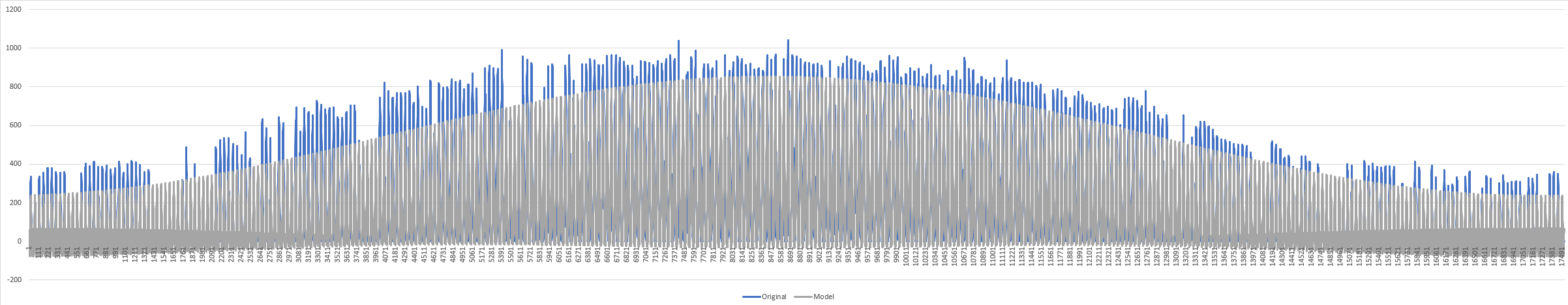
It could easily get that the most important frequencies are 1(1 cycle per year), 364, 365(1 cycle per day), 366 and 730 (2 cycles per day).

## Task2: Getting the fourier model

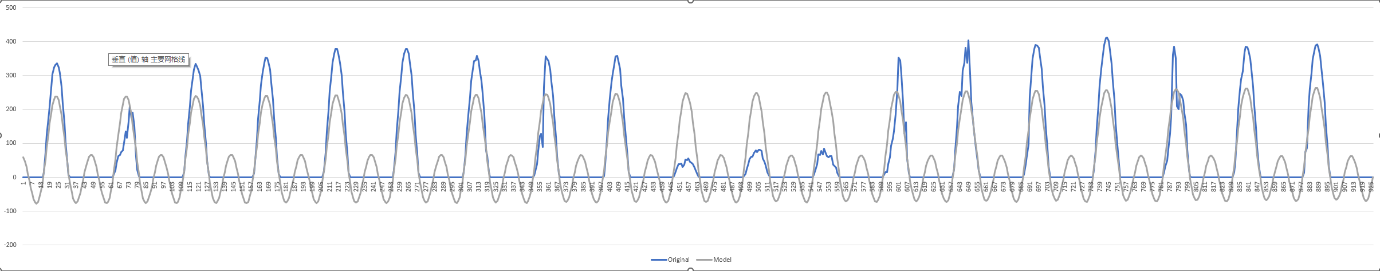
After getting the frequencies, fill them to template file, and then using the solver to minimize the SSE, then the coefficients for the Fourier series will get, just like the picture below (Yellow background).



The sum all the waves and mean, then the seasonality model will get. The fitting results will like the pictures below.



The whole dataset fitting result



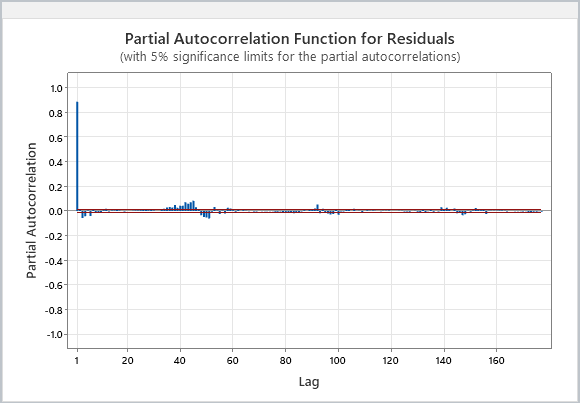
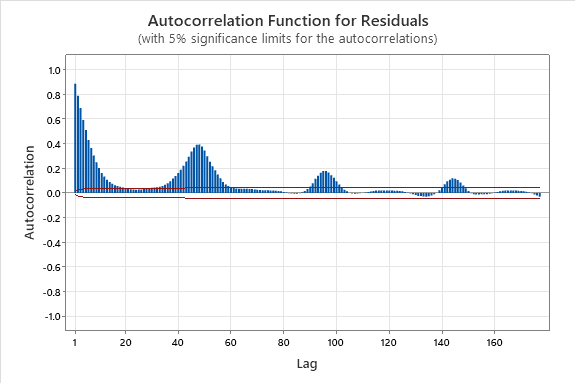
The first 1000 objects fitting result

According to the fitting result, it could be easily found that the seasonality model could capture the seasonality pattern very well.

## Task 3: Getting the coefficients for AR model

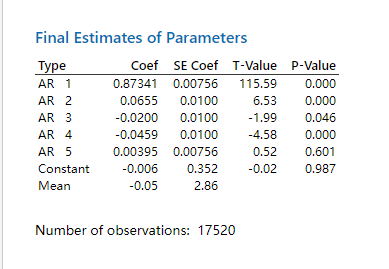
After getting the seasonality, we should remove it from the original dataset, and then try to find another model to fit the residuals. We try to use AR model using this case.

Before using AR model, the auto correlation analysis and partial auto correlation analysis will be used to analysis the residuals.

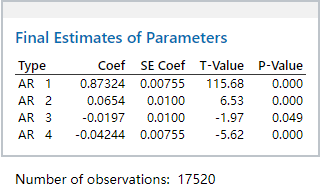


According to the results from the two analysis, it could easily find the values are correlated with the past values. It means that the AR model could be used for this dataset.

Then we could try to search the best coefficients for AR model, for using AR(5), we could get the coefficients like the picture below.



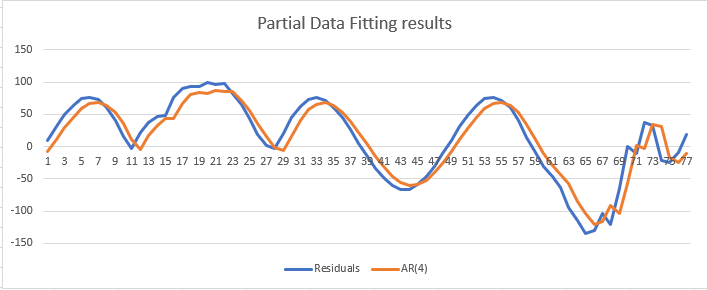
According to the graph above, it could easily be found that the pvalue for AR5 and constant is greater than 0.05, it means not significant, so the constant should be ignored, then try to search other possible coefficients.

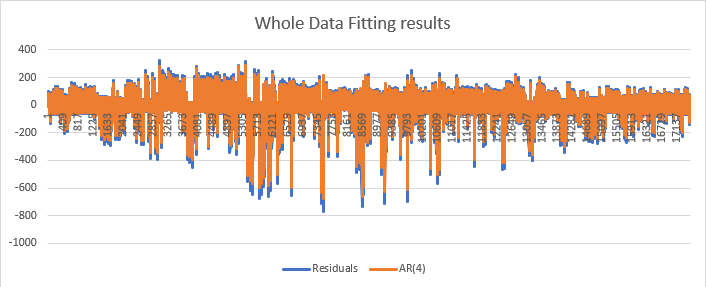


Finally, the coefficients could be got for AR(4).

## Task 4: Using the ARMA model to forecast

After getting the coefficients, then we try to forecast the current values. Then using the fitting plot to evaluate the result.



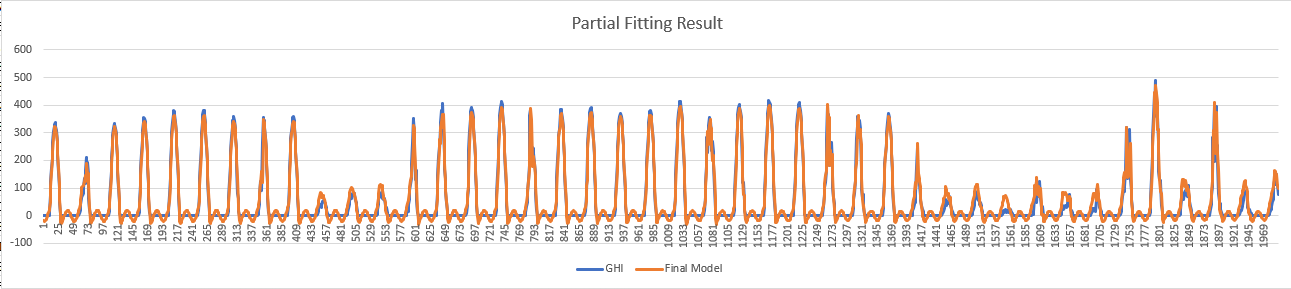


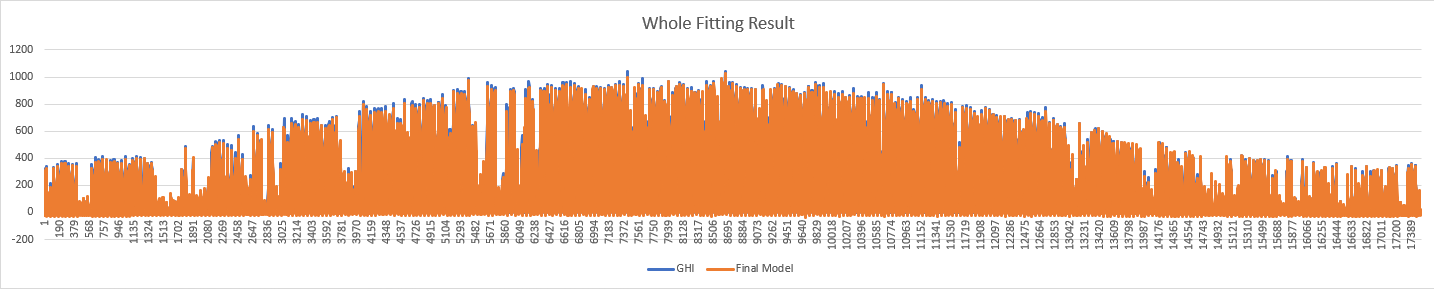
According to the graphs above, the AR(4) result fit the residuals very well.

## Task 5: To evaluating the model

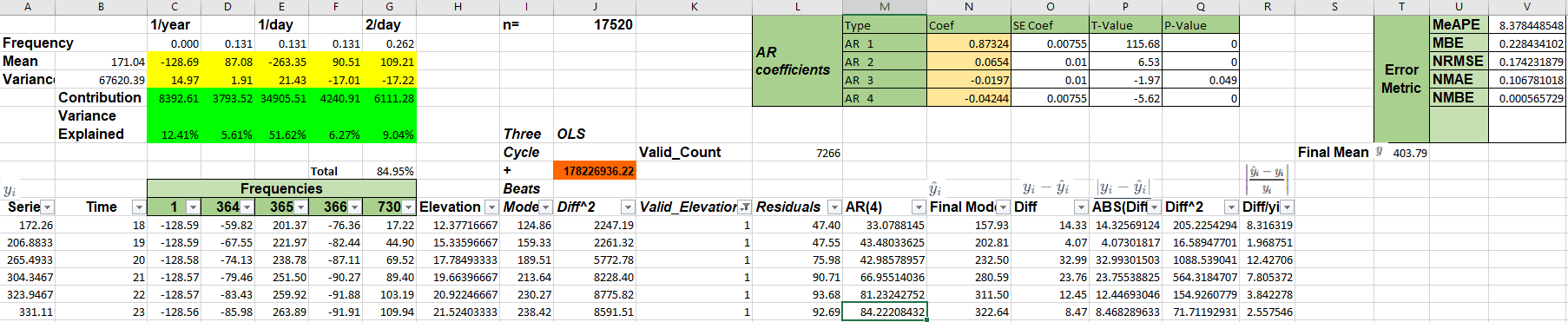
According the steps above, the original dataset has been split into two components, the seasonality and AR(4) model, now combine the two components to form the final model. Then try to use error metric to evaluate the model.

Before calculating the model, visualize the fitting result of the final model.

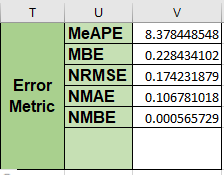


  
According to the graph above, it could easily be found that the final model could fit the dataset better.

Next, the error metric will be calculated. Because of when the value of evaluation is greater than 10, the error metric will be calculated.to calculate the error metric. We could make a valid\_elevation column to filter the rows for calculating the error metric.



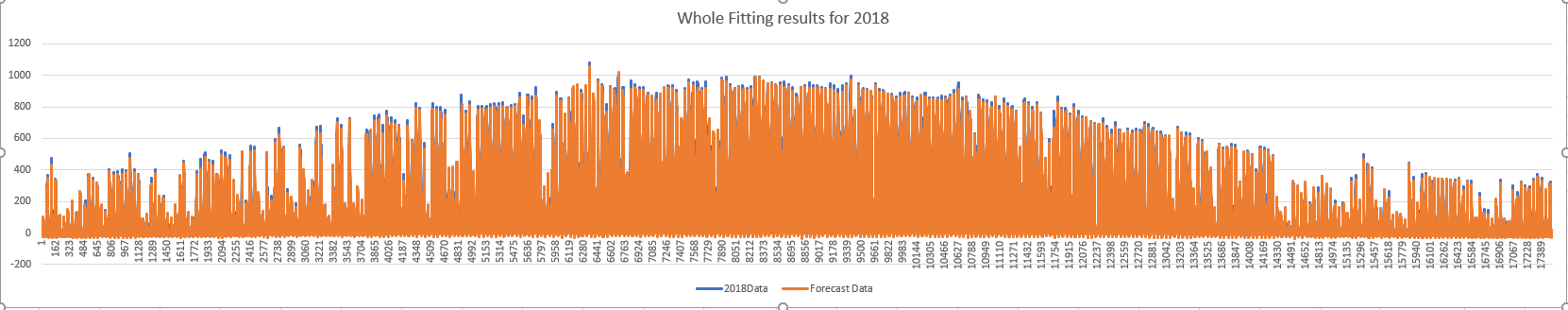
Finally, getting the error metric like

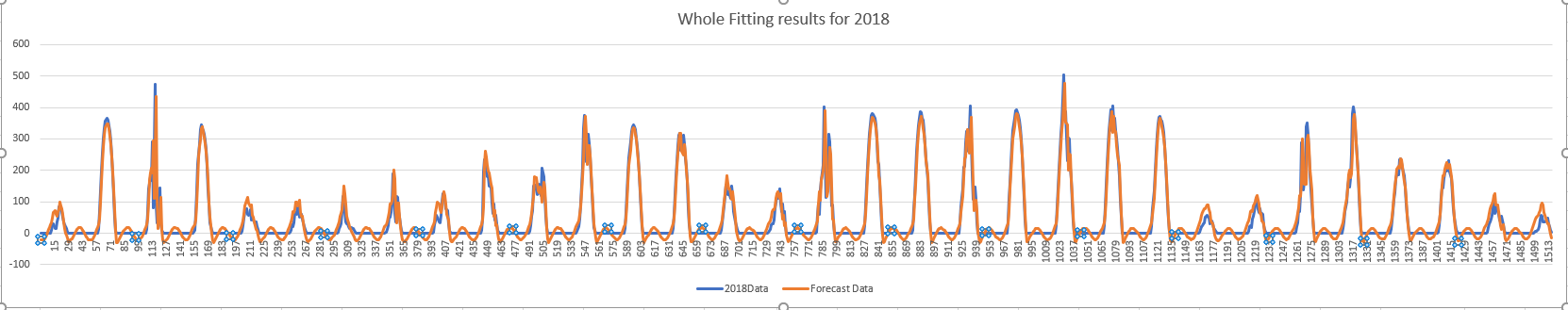


According to the error metric,

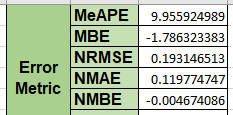
## Task 6: Testing the ARMA model using 2018 data

We have got the final model, then using the model to forecast the 2018 dataset. Then we could visualize the fitting result like the pictures below.





According to the graphs above, the model could work well, then we calculate the error metric,

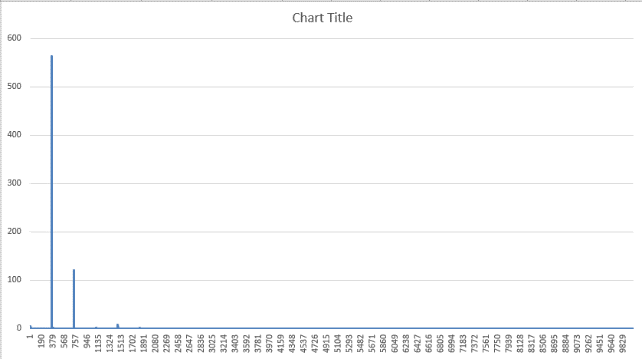


According to the error metric,

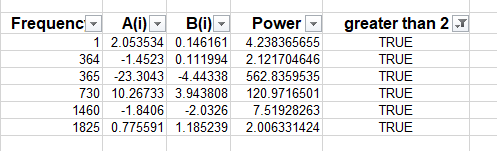
# Solar Farm Dataset

## Task 1: Find frequencies

Copy the Farm Dataset to Power\_SpectrumGeneric file to get the frequencies, the number of objects equals 105120, and the frequencies is 10000, we got the frequencies like the graph below.

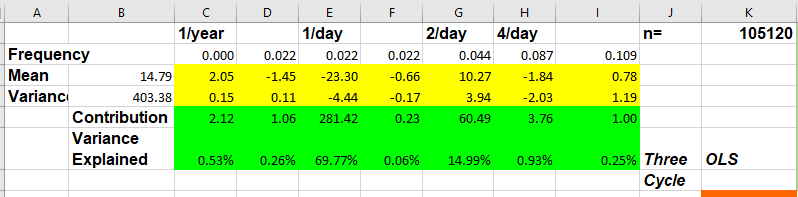


It could be easily found that the most important frequencies are around 100, it seems there are some frequencies around 0 also hold higher power value. We use the value 2 to filter the most important frequencies, the results could be get like below.

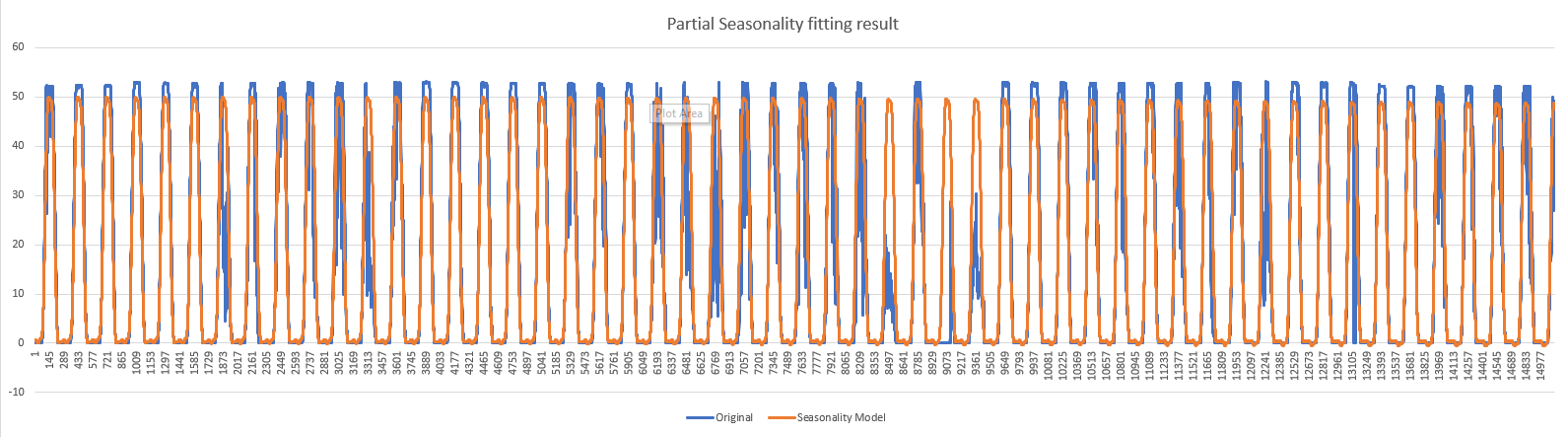


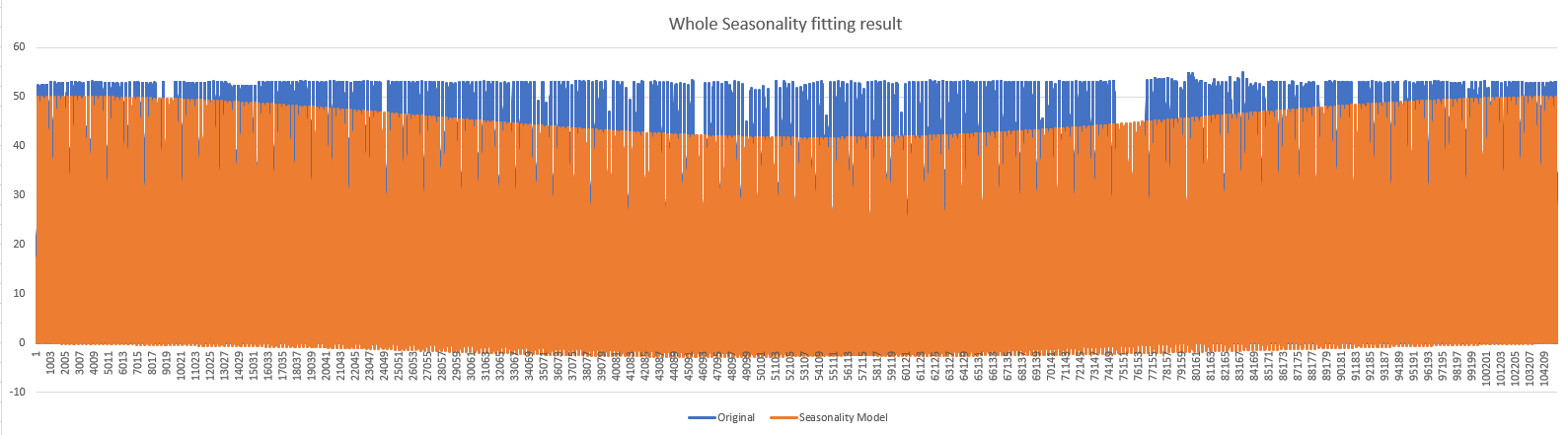
## Task 2: Make seasonality

After getting the most important frequencies, then try to get the coefficients for seasonality, after minimizing the SSE, we got the seasonality coefficients like below.



We could visualize the seasonality result, like the pictures below.

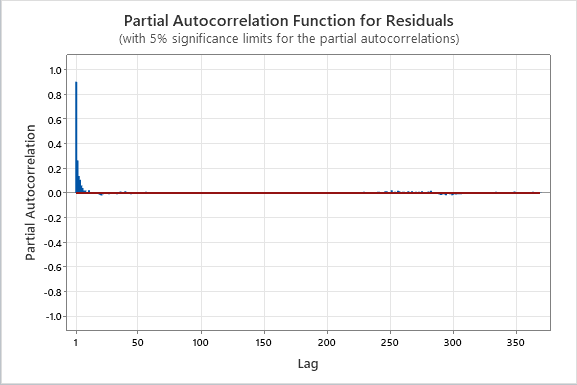
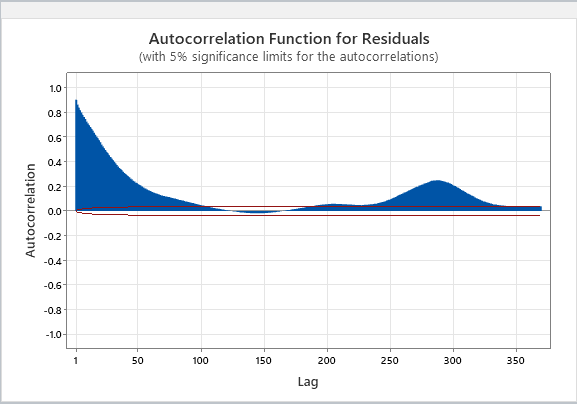




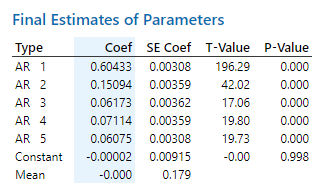
According to the graphs above, the seasonality could capture the pattern of the original dataset, but it seems has a big gap in the middle of the datasets.

## Task 3: ARIMA coefficients

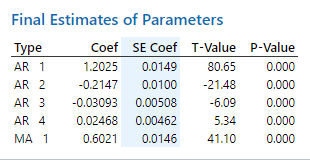
After getting the seasonality, remove it from the original dataset will get the residuals. We try to model the residuals for better forecasting. The autocorrelation and partial autocorrelation will be used for analysing the residuals firstly.



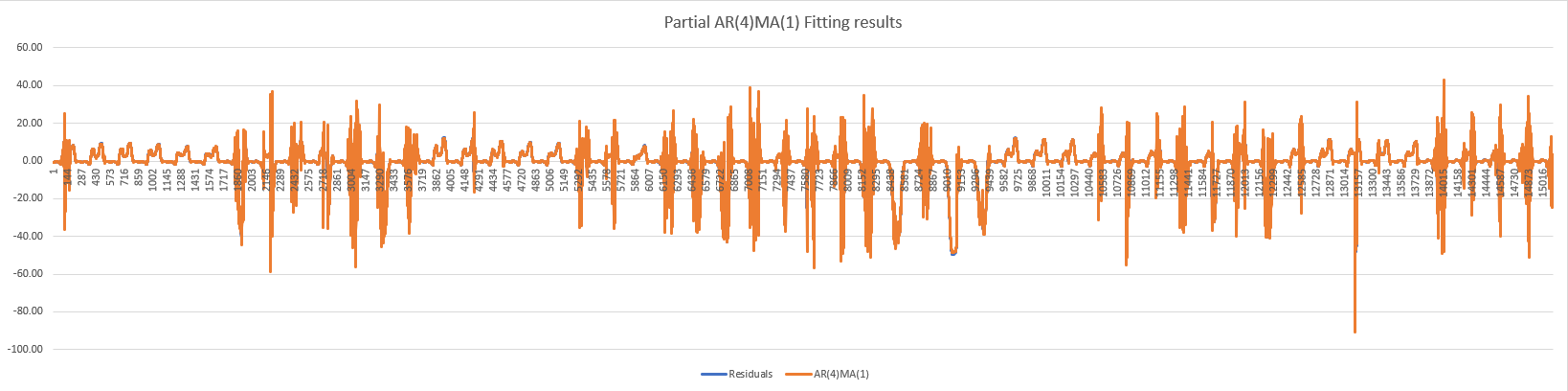
According the graphs the dataset is related to the past values, so the ARMA model could be used for the dataset. Then try to search the proper coefficients of ARMA model.

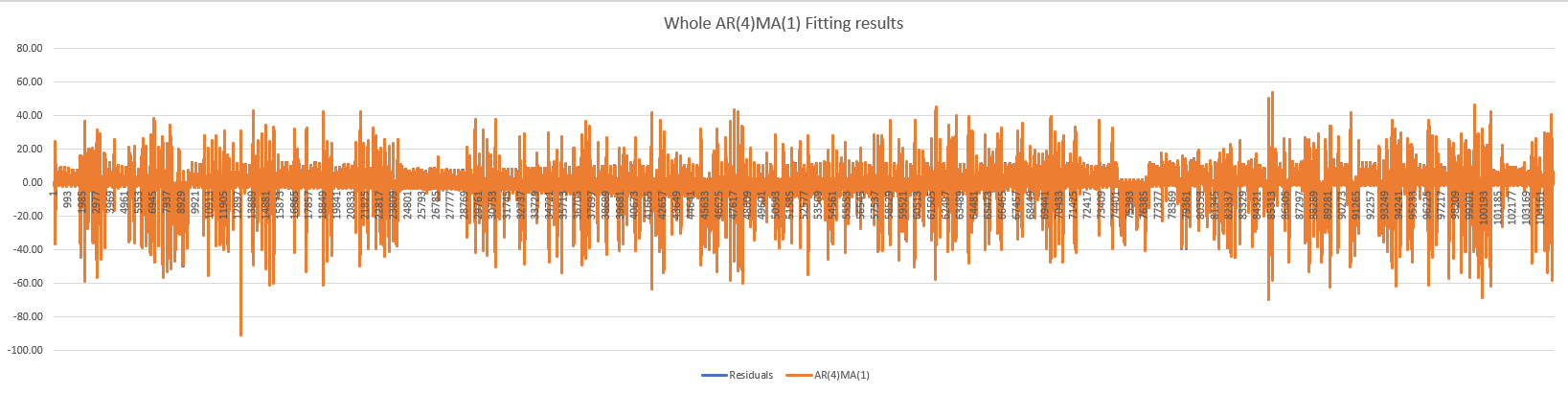


According to the graph above, we could know the pvalue of constant is greater than 0.05, which means no significant in this case, so the constant should be ignored. Continue to search the coefficients could get the result like the picture below



Using the coefficients to model the residuals, we could visualize the results like

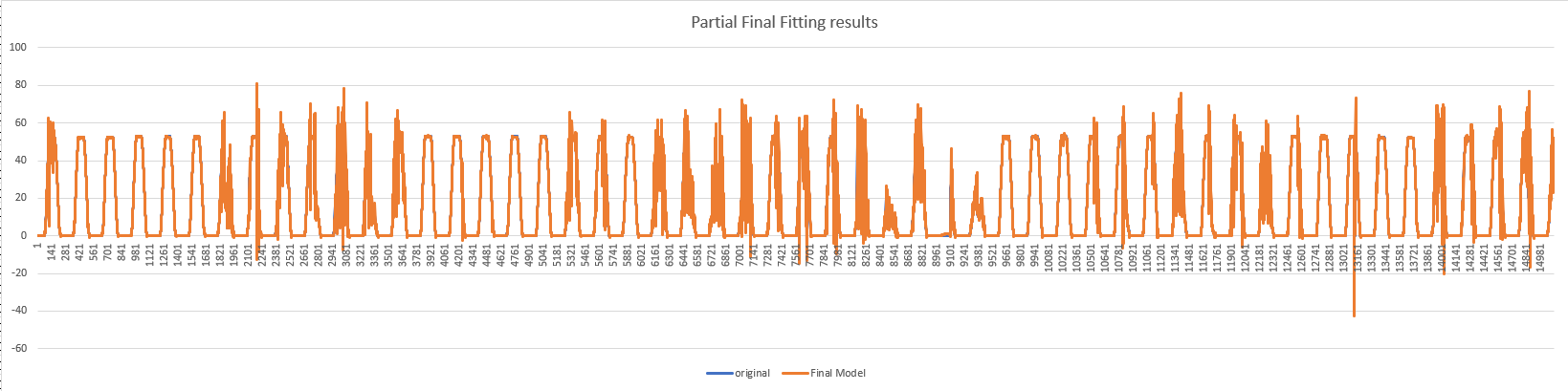


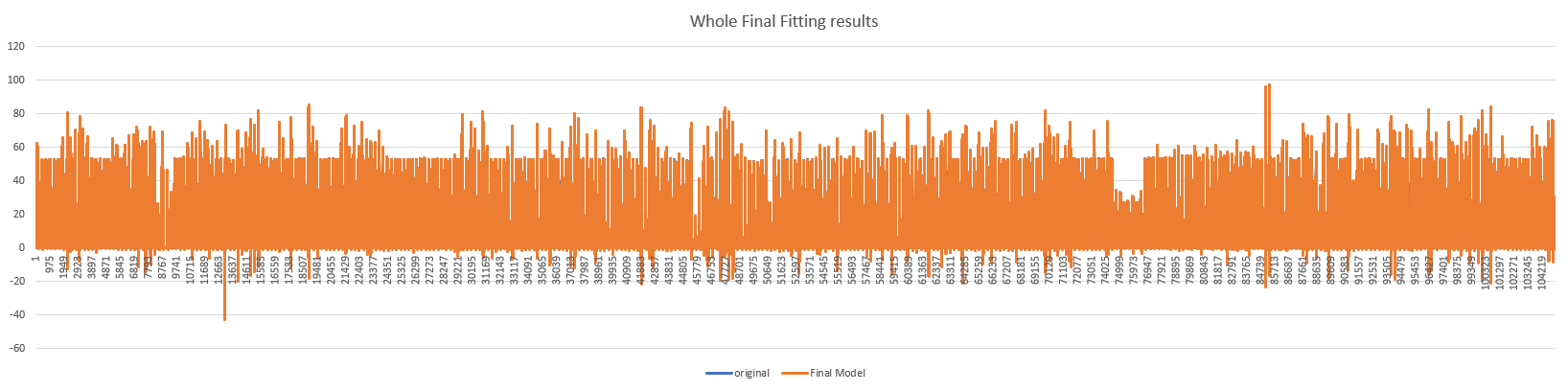


According to the graphs above, the ARMA(4,1) could fit the residuals very well.

## Task 4: Final model

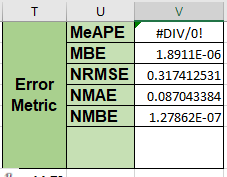
The steps above has split the data into two components, this part will combine them to form the final model. After combining the two components, the final model will fit the original dataset like the picture below





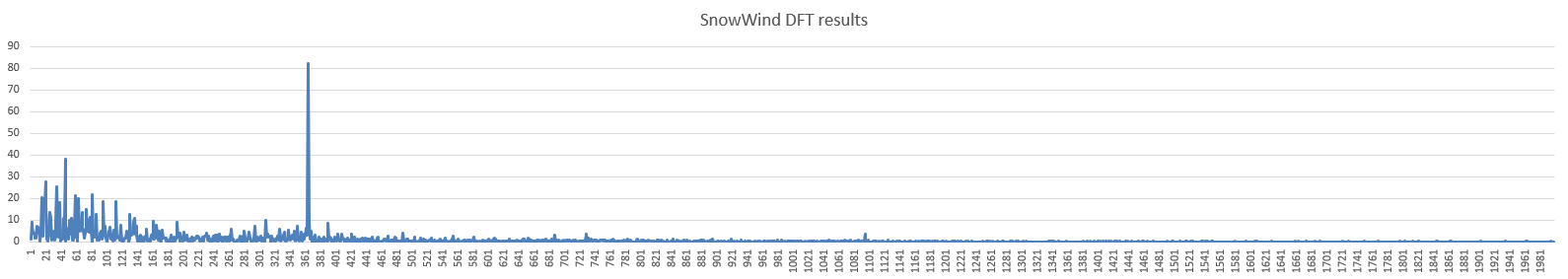
According to the graphs,

## Task 5: Error Metric



# Snowtown Wind Farm Dataset

## Task 1: Find frequencies



## Task 2: Compare AR(p) and ARMA(p,q)

