Assignment 1

Advanced Analytical Techniques

**Name:** Haiyue Wang

**Email ID:** wanhy149

Content

[Half Hour Solar Radiation Dataset 3](#_Toc144735469)

[Taks1: Getting Frequencies 3](#_Toc144735470)

[Task2: Getting the fourier model 3](#_Toc144735471)

[Task 3: Getting the coefficients for ARMA model 4](#_Toc144735472)

[AR Model 4](#_Toc144735473)

[ARMA Model 5](#_Toc144735474)

[Task 4: Using the ARMA model to forecast 5](#_Toc144735475)

[AR(4) + Seasonality 5](#_Toc144735476)

[ARMA(3,2) + Seasonality 6](#_Toc144735477)

[Task 5: To evaluating the model 7](#_Toc144735478)

[AR(4) + Seasonality 7](#_Toc144735479)

[ARMA(3,2) + Seasonality 8](#_Toc144735480)

[Task 6: Testing the ARMA model using 2018 data 8](#_Toc144735481)

[AR(4) + Seasonality 8](#_Toc144735482)

[ARMA(3,2) + Seasonality 9](#_Toc144735483)

[Solar Farm Dataset 10](#_Toc144735484)

[Task 1: Find frequencies 10](#_Toc144735485)

[Task 2: Make seasonality 10](#_Toc144735486)

[Task 3: ARIMA coefficients 11](#_Toc144735487)

[Task 4: Final model 12](#_Toc144735488)

[Task 5: Error Metric 13](#_Toc144735489)

[Snowtown Wind Farm Dataset 13](#_Toc144735490)

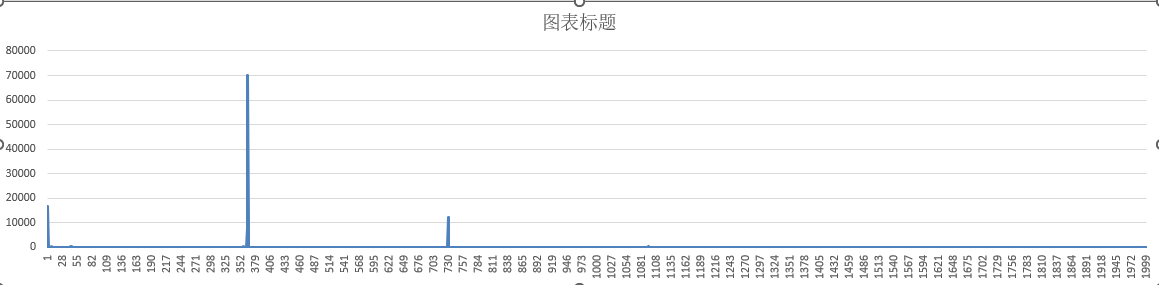
[Task 1: Find frequencies 13](#_Toc144735491)

[Task 2: Compare AR(p) and ARMA(p,q) 13](#_Toc144735492)

# Half Hour Solar Radiation Dataset

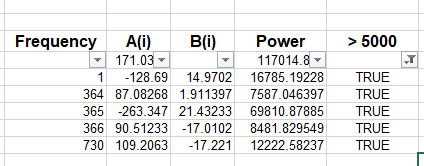
## Taks1: Getting Frequencies

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



***Frequencies power plot***

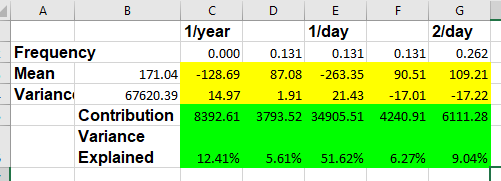
According to this graph, it could easily find that the minimum values of the most important frequencies are around 10000, we use filter power > 5000 to filter all the possible frequencies. The value could be got like the graph below.



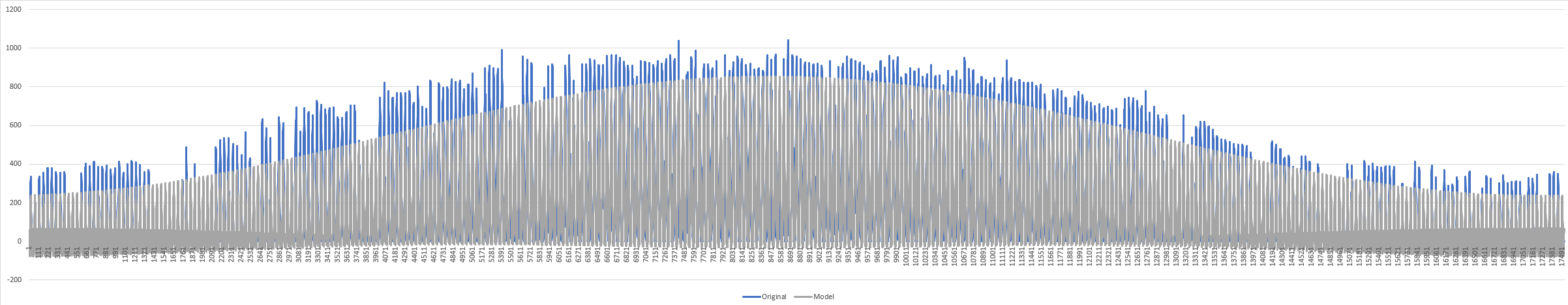
It could easily get that the most important frequencies are 1(1 cycle per year), 364, 365(1 cycle per day), 366 and 730 (2 cycles per day).

## Task2: Getting the fourier model

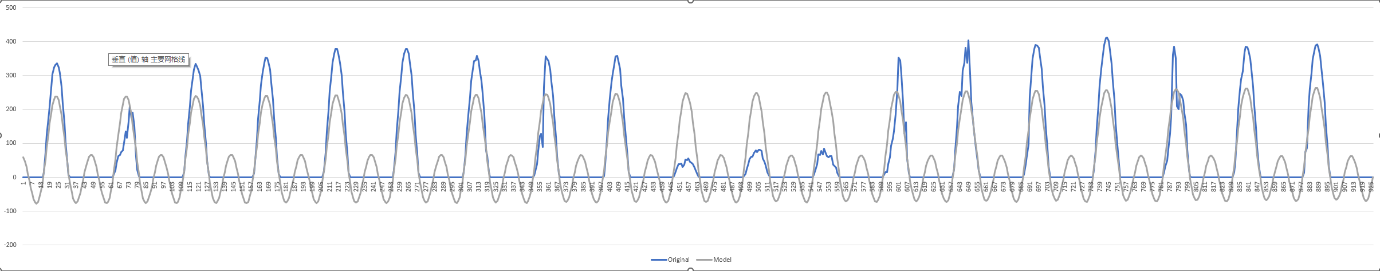
After getting the frequencies, fill them to template file, and then using the solver to minimize the SSE, then the coefficients for the Fourier series will get, just like the picture below (Yellow background).



Sum all the waves and mean, then the seasonality model will get. The fitting results will like the pictures below. I will visualize the fitting result below.



*The whole dataset fitting result*



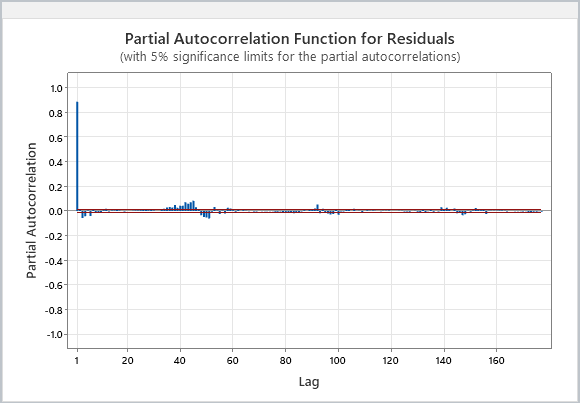
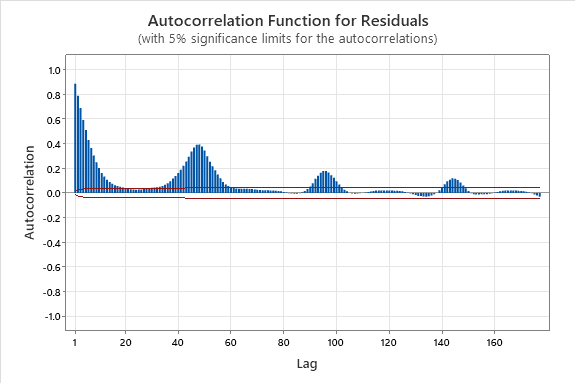
*The first 1000 objects fitting result*

According to the fitting result, it could be easily found that the seasonality model could capture the seasonality pattern not bad.

## Task 3: Getting the coefficients for ARMA model

After getting the seasonality, we should remove it from the original dataset, and then try to find another model to fit the residuals. This part I will do two models for the residuals, the first one is AR model another one is ARMA model.

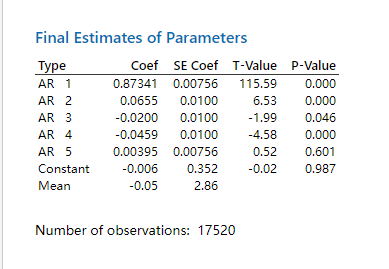
Before using AR model, the auto correlation analysis and partial auto correlation analysis will be used to analysis the residuals.



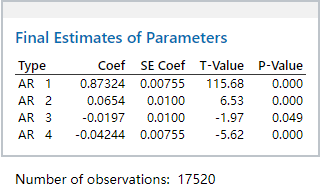
According to the results from the two graphs, it could easily find the values are correlated with the past values. It means that the ARMA model could be used for this dataset.

### AR Model

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.



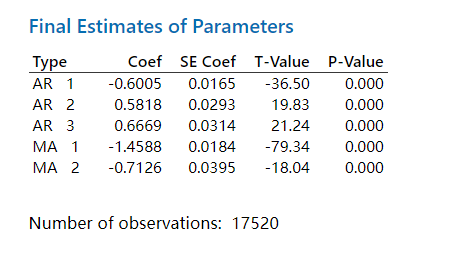
According to the graph above, it could easily be found that the pvalue for AR5 and constant is greater than 0.05, it means not significant, so the constant should be ignored, then try to search other possible coefficients.



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

### ARMA Model

This part I will continue to search the best ARMA model, finally, I got the ARMA(3,2) is the best model for the residuals, coefficients like the picture below.

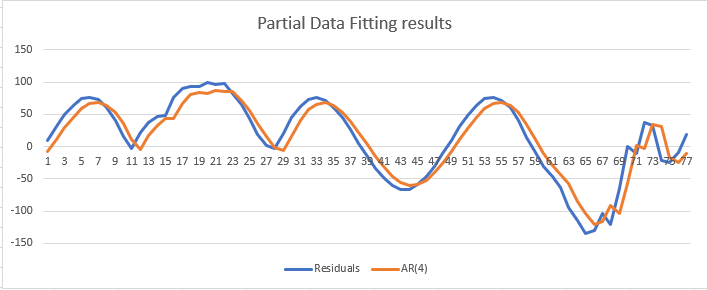


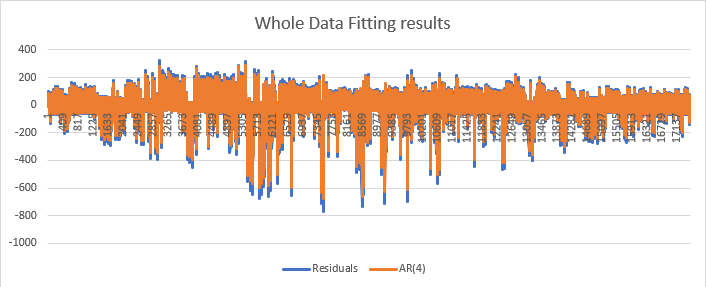
Then we could see the fitting result for the residuals.

## Task 4: Using the ARMA model to forecast

### AR(4) + Seasonality

After getting the coefficients, then we try to forecast values. Then, using plot to view the fitting results.

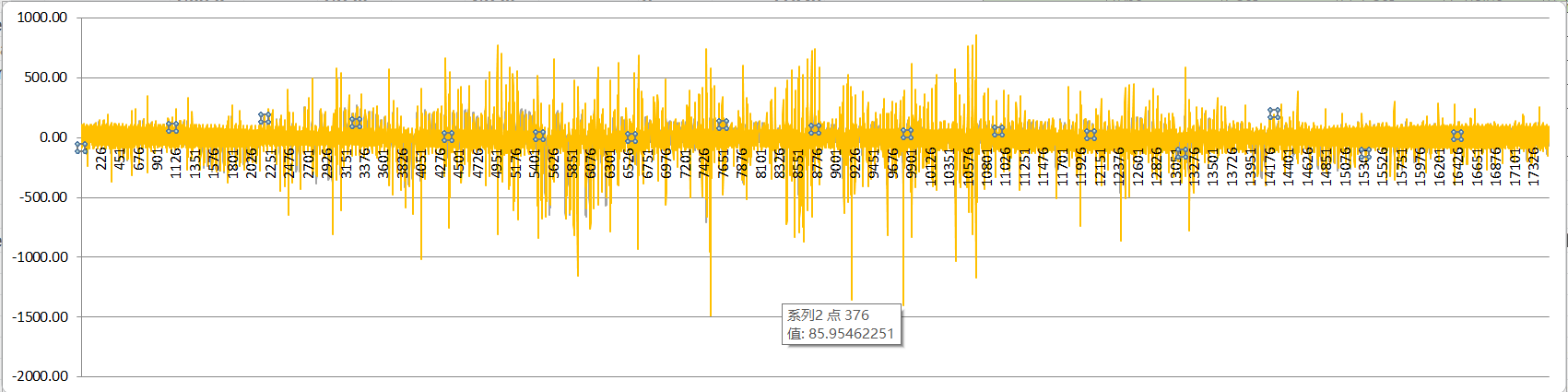




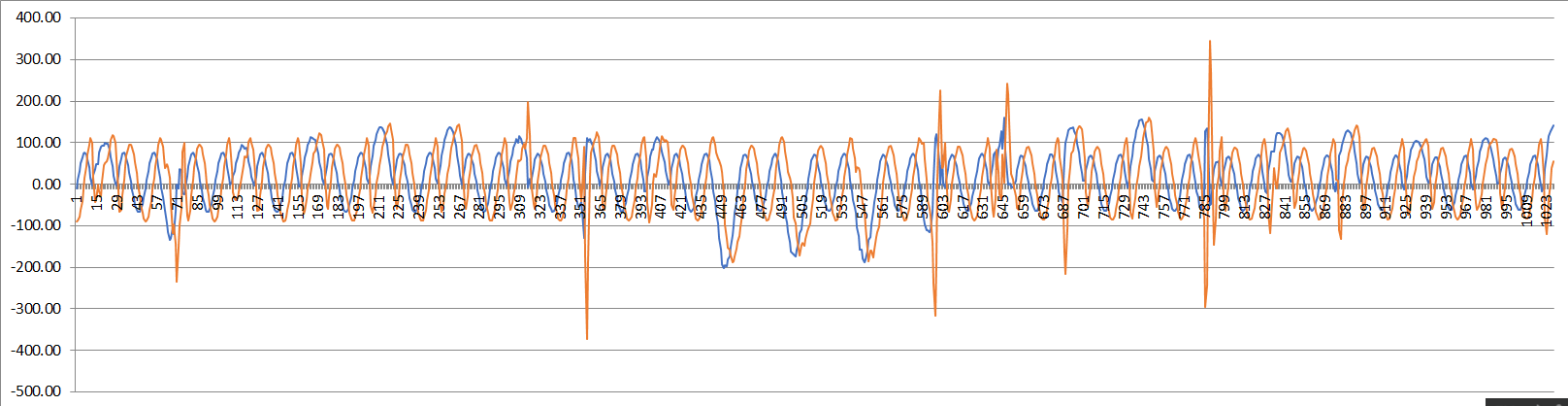
According to the graphs above, the AR(4) model could fit the residuals very well.

### ARMA(3,2) + Seasonality

Similar with the AR(4), we visualize the results firstly, like the pictures below.



Whole dataset fitting results



Partial dataset fitting results

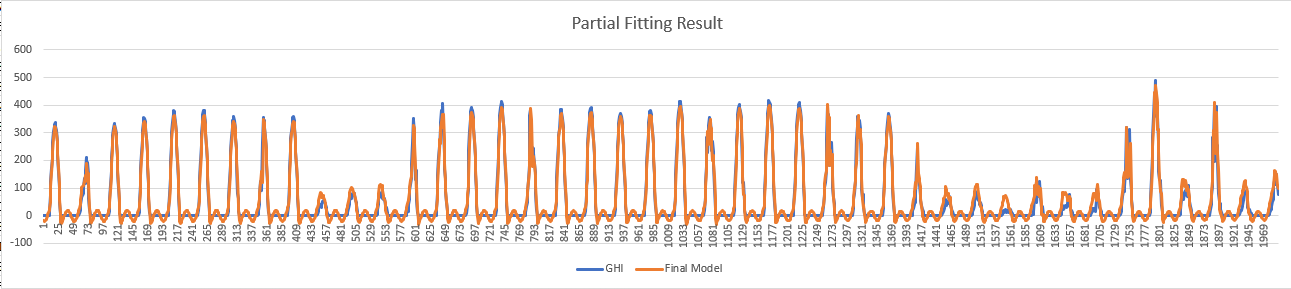
According to the graphs above, the ARMA(3,2) could fit the residuals very well.

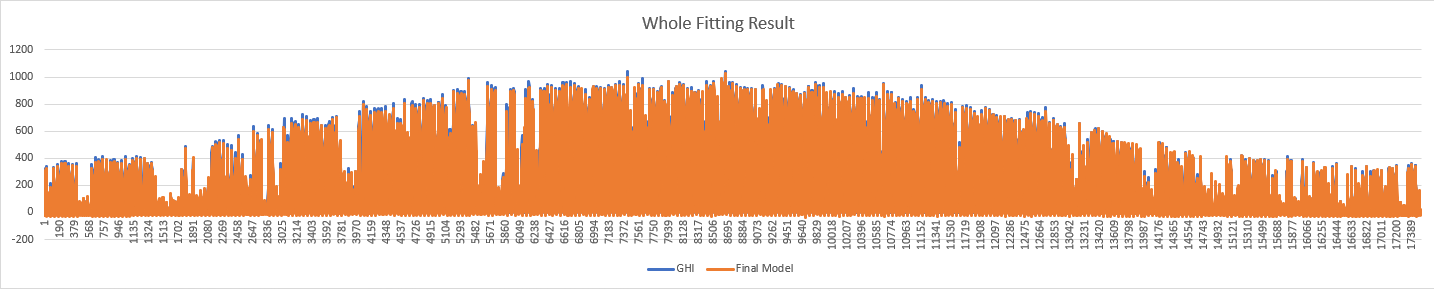
## Task 5: To evaluating the model

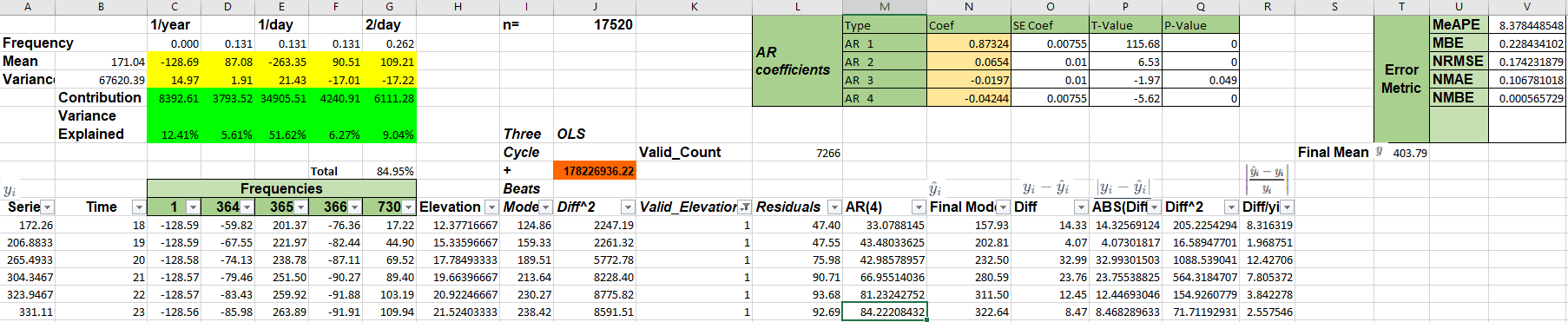
According the steps above, the original dataset has been split into two components, the seasonality and AR(4) /ARMA(3,2) model, now combine the two components to form the final model. Then try to use error metric to evaluate the model.

### AR(4) + Seasonality

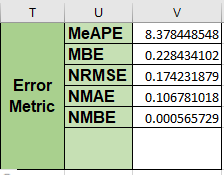
Before calculating the error metric, visualize the fitting result of the final model.



  
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.



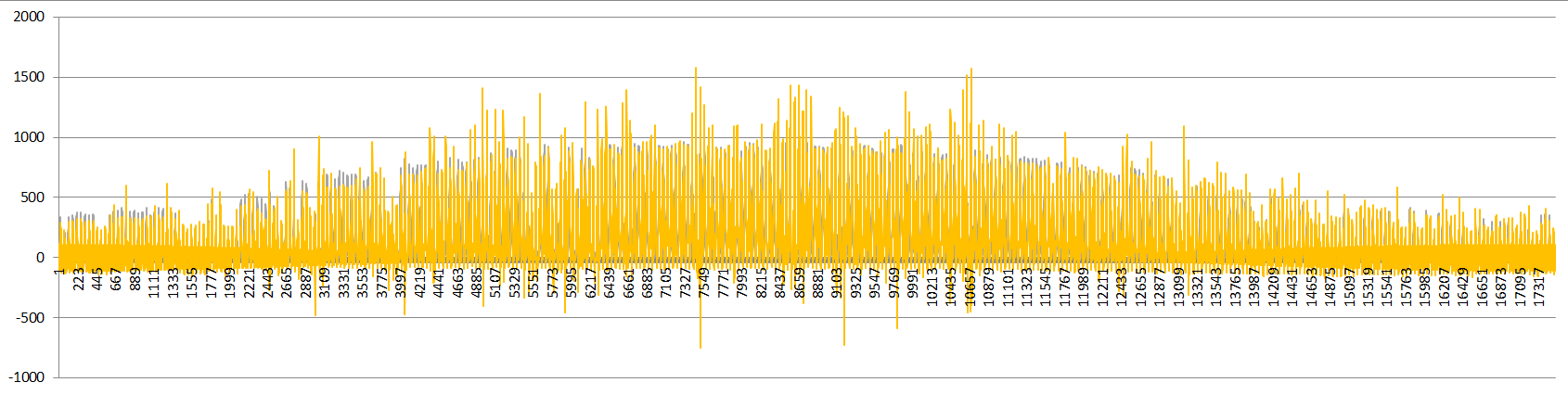
Finally, getting the error metric like



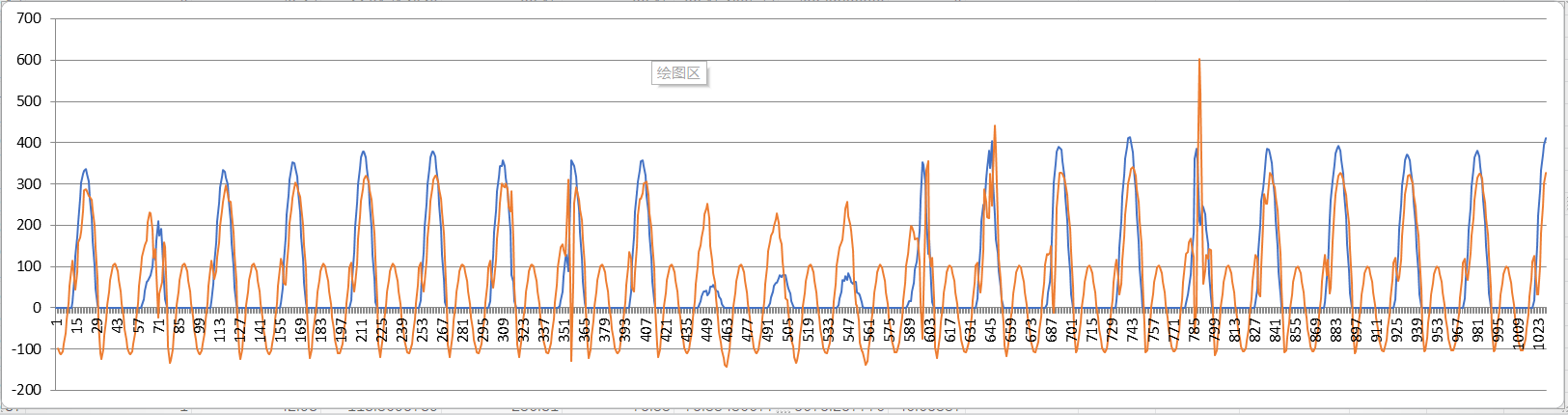
MBE measures the average bias between the model's predictions and the actual observations. NRMSE quantifies the model's prediction errors relative to the range of the actual observations. A smaller NRMSE indicates better model performance. NMAE measures the average absolute error of the model relative to the range of the actual observations. NMBE quantifies the average bias of the model relative to the range of the actual observations. All the indicators of the error metric, the lower value means the better performance. All other evaluations below will use the four indicators for evaluation. According to the value of the four indicators, all of them are normally low, especially the NMBE is very close to 0.

### ARMA(3,2) + Seasonality

Firstly, we need to visualize the fitting results of this final model.

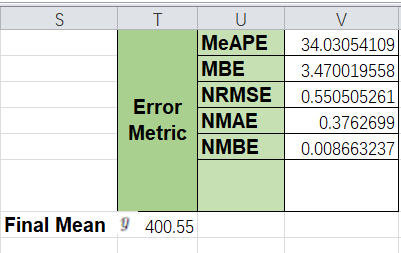


***Whole dataset fitting results***



***Partial dataset fitting results***

According to the graphs above, indicate the ARMA(3,2) + Seasonality could fit the dataset very well also. The error metric likes the picture below.

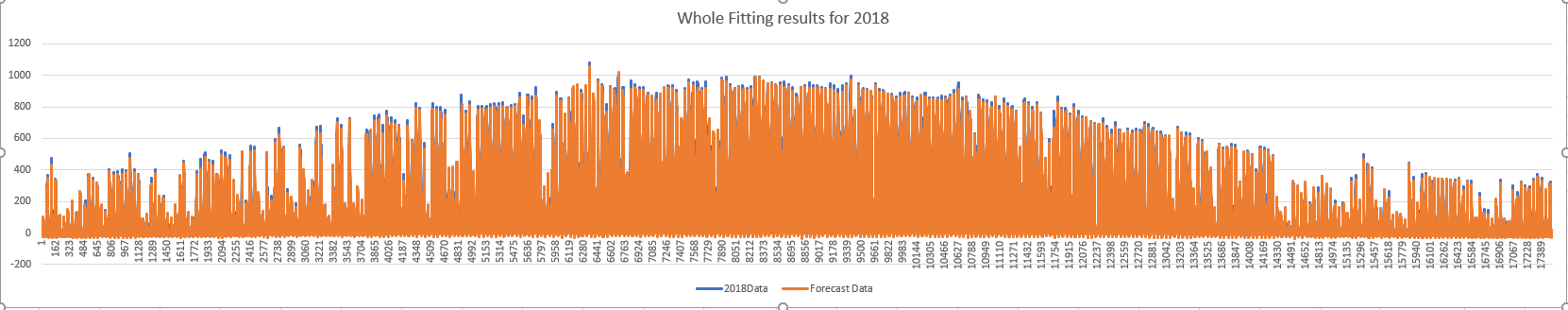


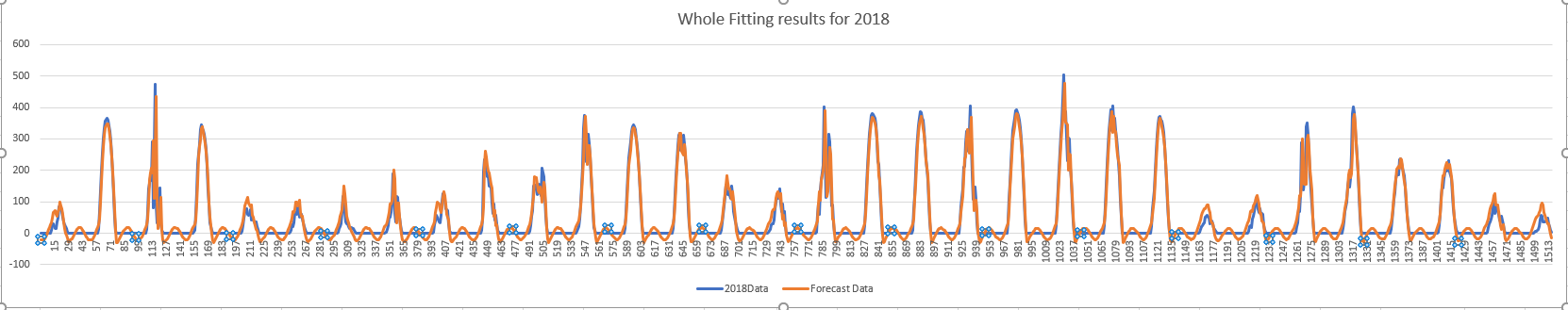
According to the error metric, the value of each indicators looks low also, but compare to the value of AR(4) + Seasonality, the performance is a little lower .

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

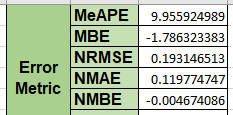
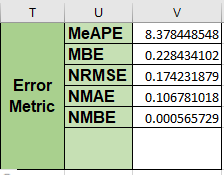
### AR(4) + Seasonality

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.





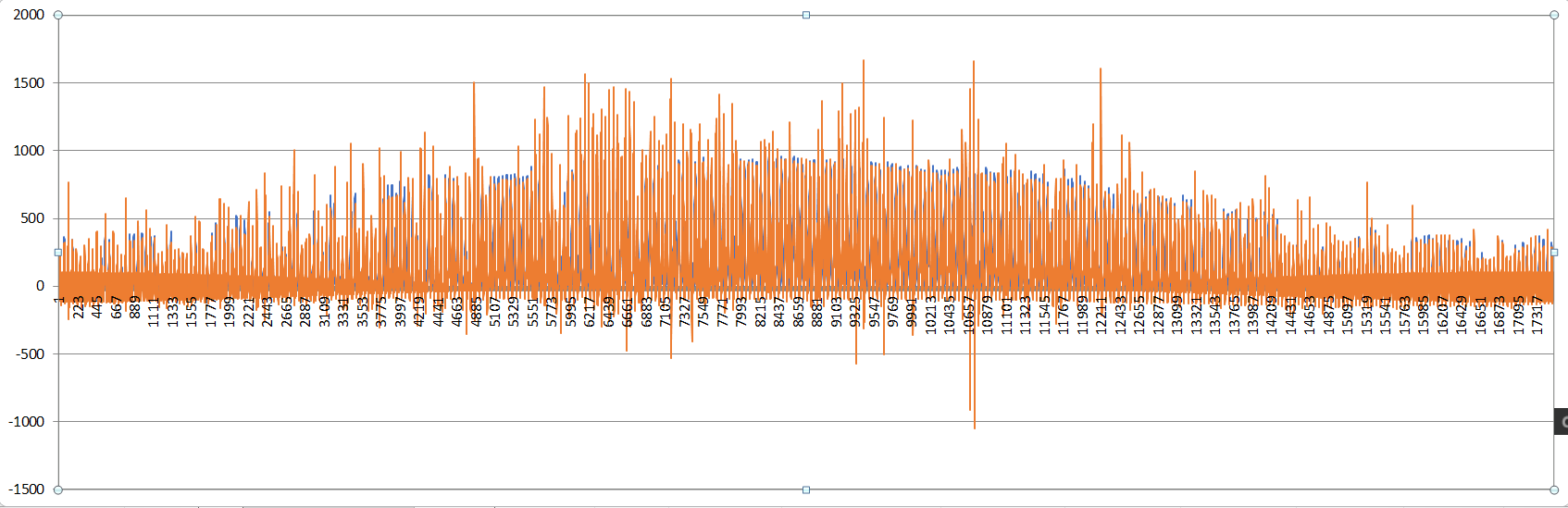
According to the graphs above, the model could work well, then we calculate the error metric,

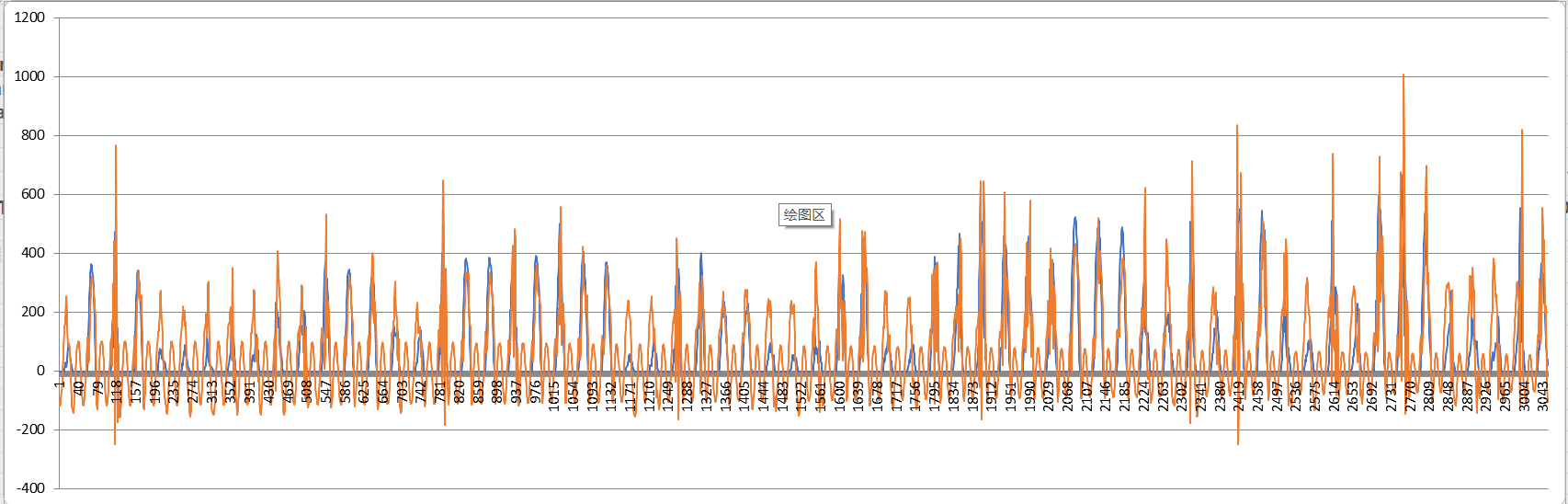
2018 Error metric 2017 Error metric

According to the error metric, the value of each indicator is very similar. So the performance on the both datasets is well.

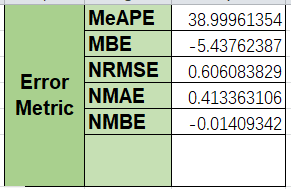
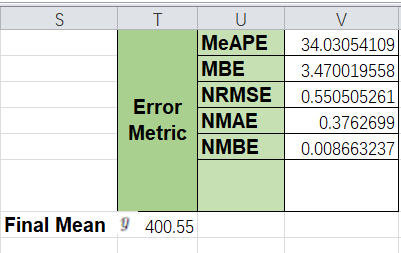
### ARMA(3,2) + Seasonality



***Whole dataset fitting result***



***Partial dataset fitting results***

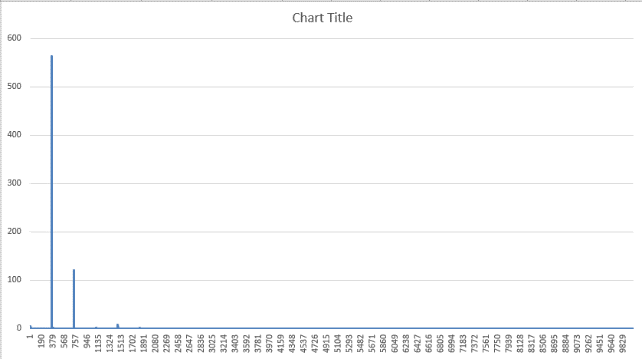
2018 Error metric 2017 Error metric

According to the error metrics, the values of the indicator are very similar. So the performance on the both datasets is well.

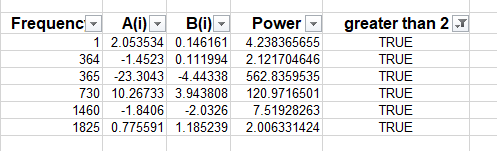
# Solar Farm Dataset

## 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.

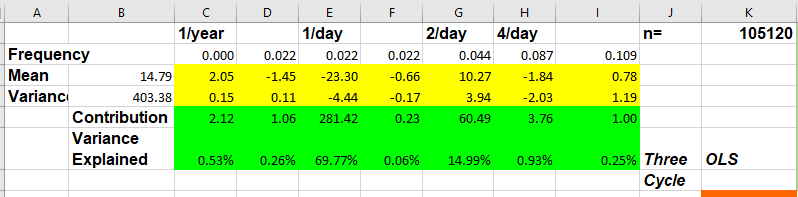


It could be easily found that the most important frequencies are around 100, it seems there are some frequencies around 0 also hold higher power value. We use the value 2 to filter the most important frequencies, the results could be get like below.

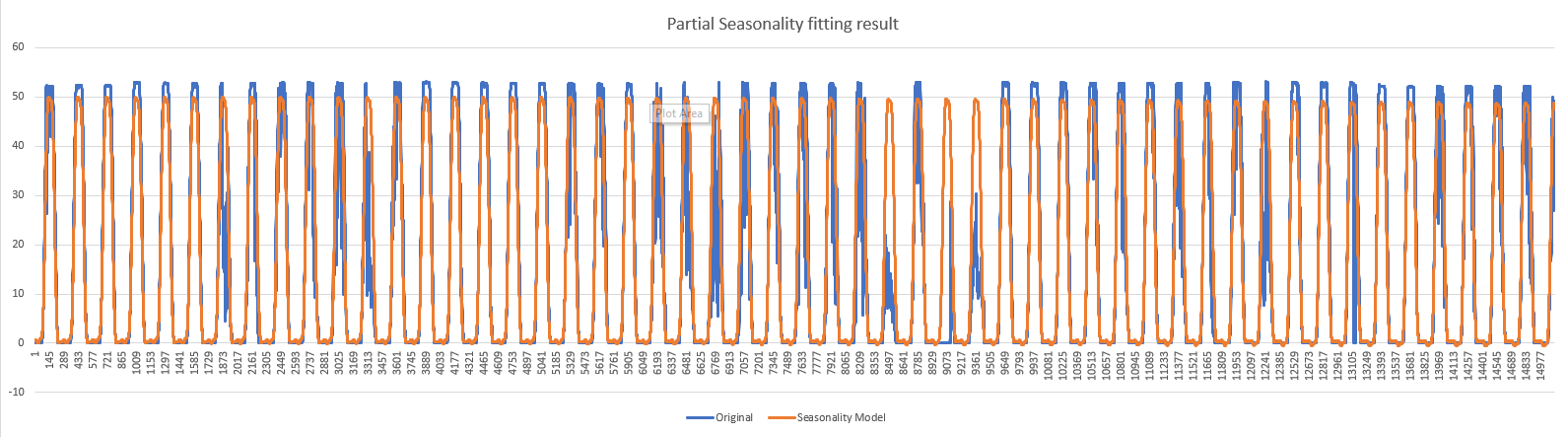


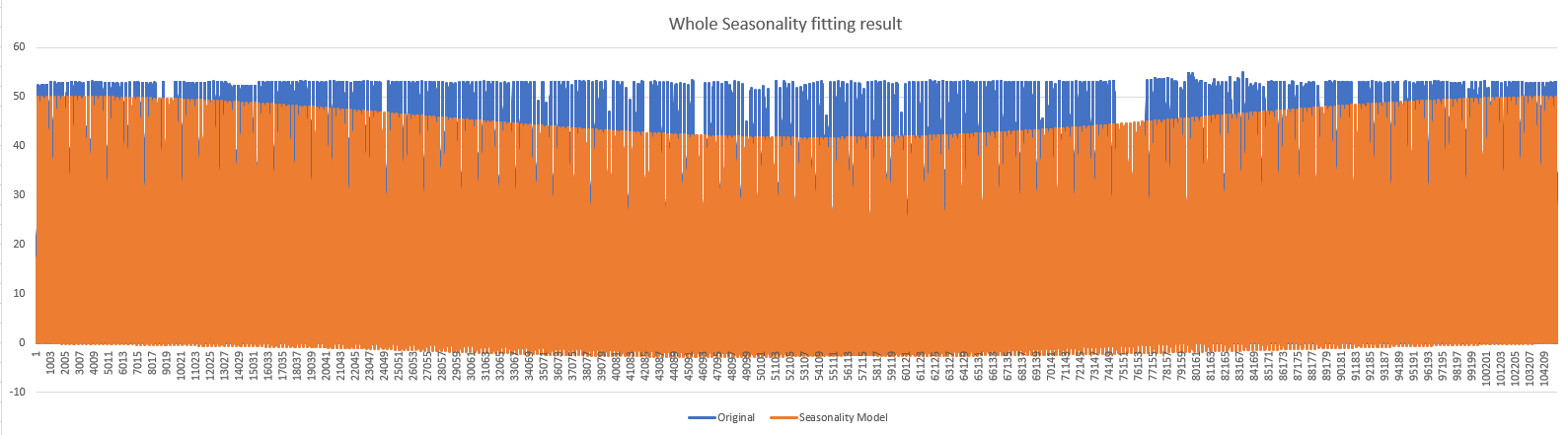
## Task 2: Make seasonality

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.

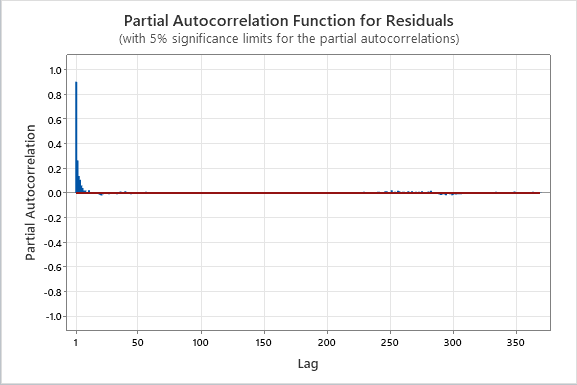
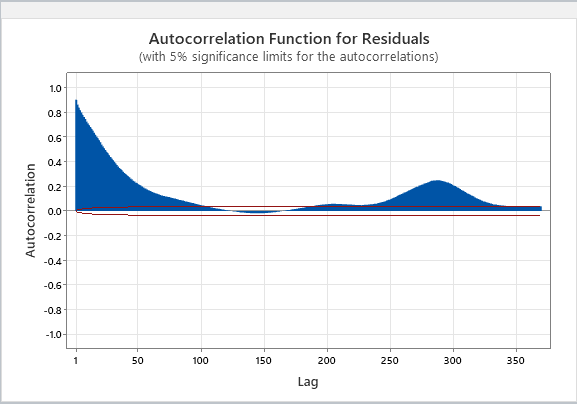




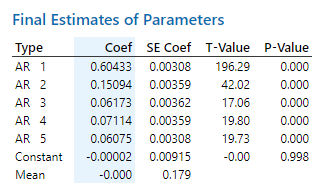
According to the graphs above, the seasonality could capture the pattern of the original dataset, but it seems has a big gap in the middle of the datasets.

## Task 3: ARIMA coefficients

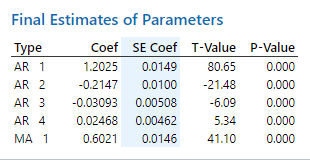
After getting the seasonality, remove it from the original dataset will get the residuals. We try to model the residuals for better forecasting. The autocorrelation and partial autocorrelation will be used for analysing the residuals firstly.



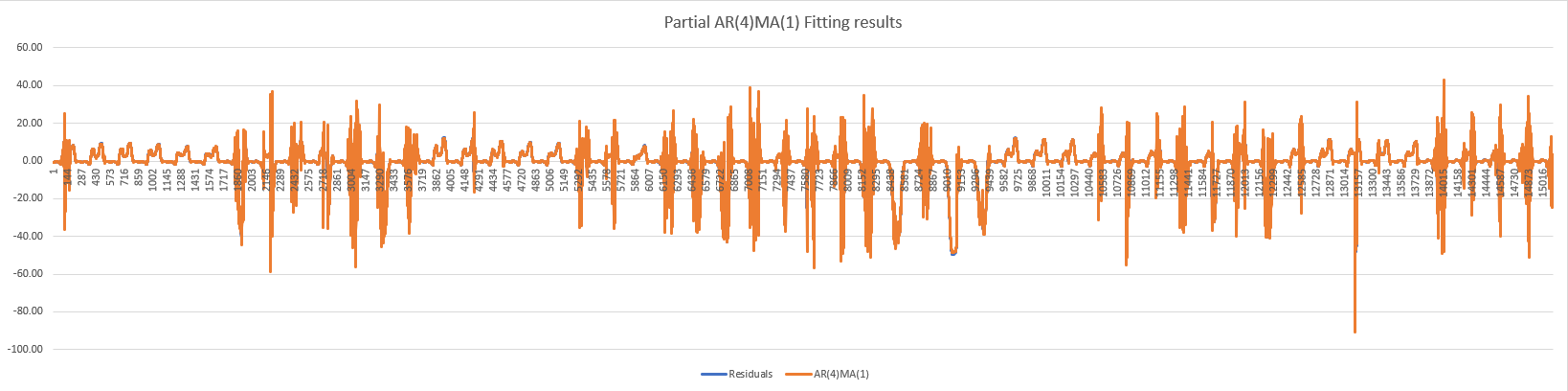
According the graphs the dataset is related to the past values, so the ARMA model could be used for the dataset. Then try to search the proper coefficients of ARMA model.

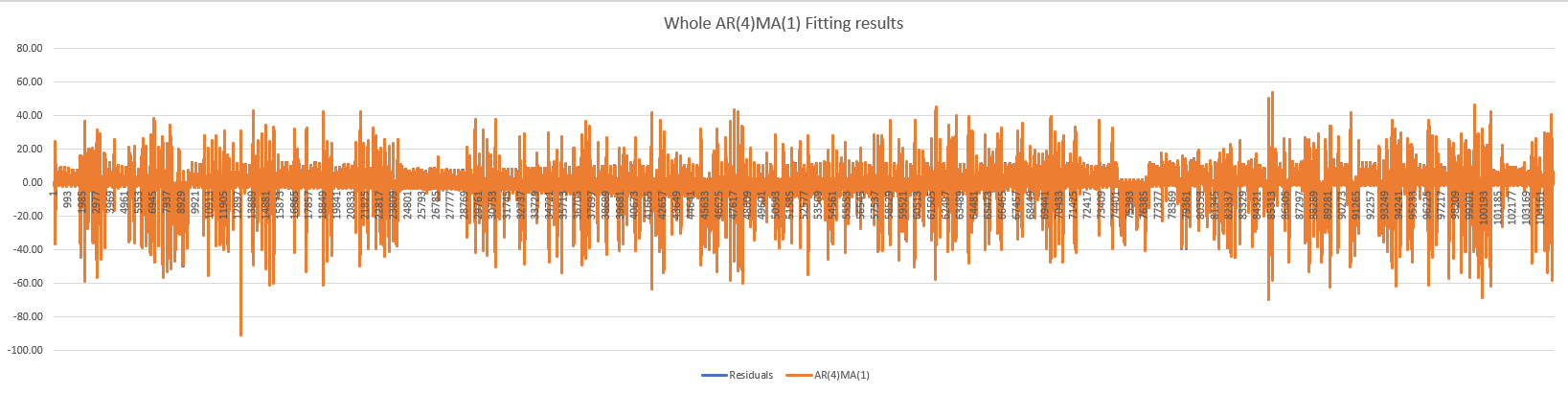


According to the graph above, we could know the pvalue of constant is greater than 0.05, which means no significant in this case, so the constant should be ignored. Continue to search the coefficients could get the result like the picture below



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

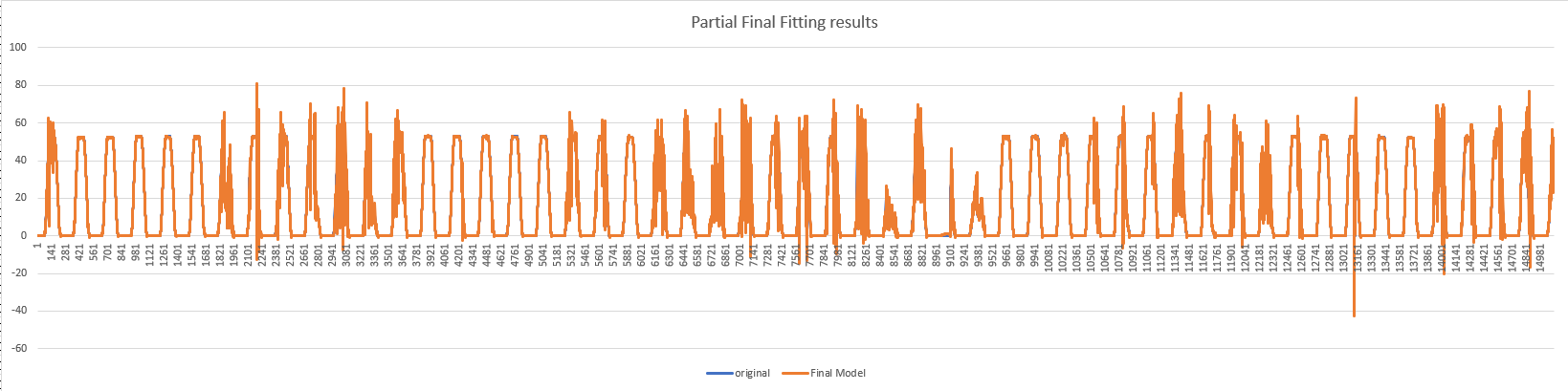


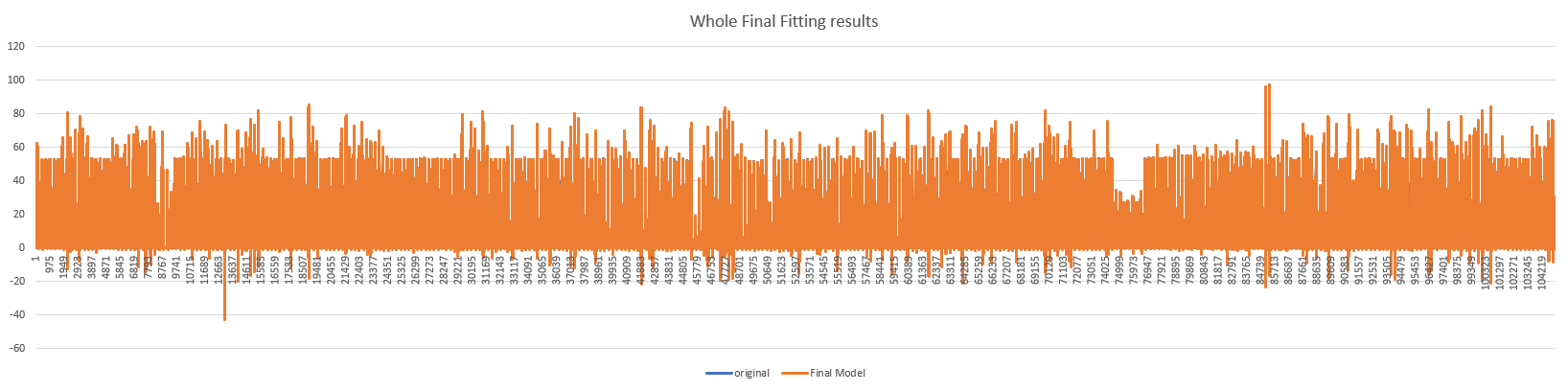


According to the graphs above, the ARMA(4,1) could fit the residuals very well.

## 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

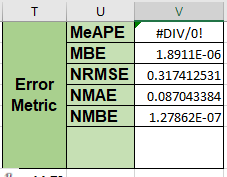




According to the graphs above, the final model could fit the original dataset very well.

## Task 5: Error Metric

This part is mainly to calculate the error metric, the values of the error metric like the picture below.

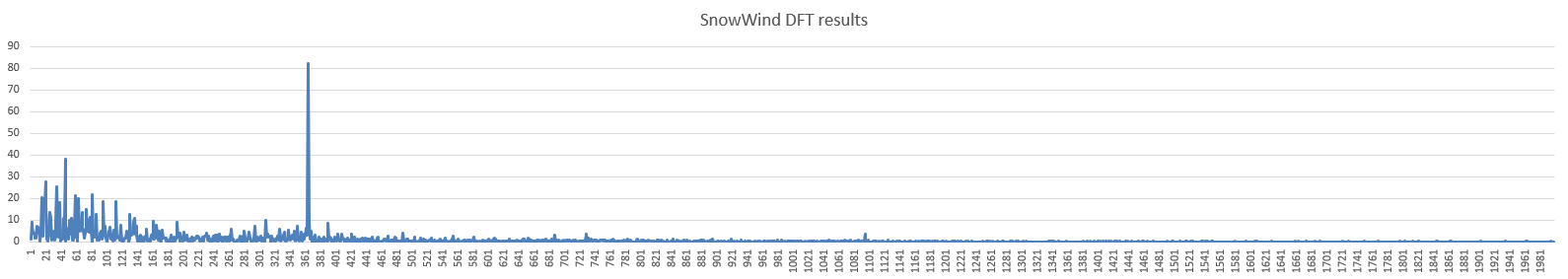


According to the error metric, each value of the indicators is quite low, the MBE, NMAE and NMBE are both close to 0. It means the performance of the model on this dataset is quite high.

# Snowtown Wind Farm Dataset

## Task 1: Find frequencies

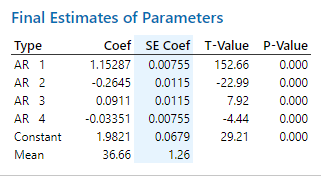
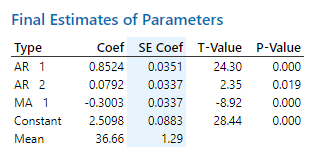
The part will to find the frequencies for Snowtown Wind Farm dataset. Copy the dataset to the power spectrum excel file, the parameters used are, number of objects is 17520, and the number of frequencies is 2000. The result looks like picture below.



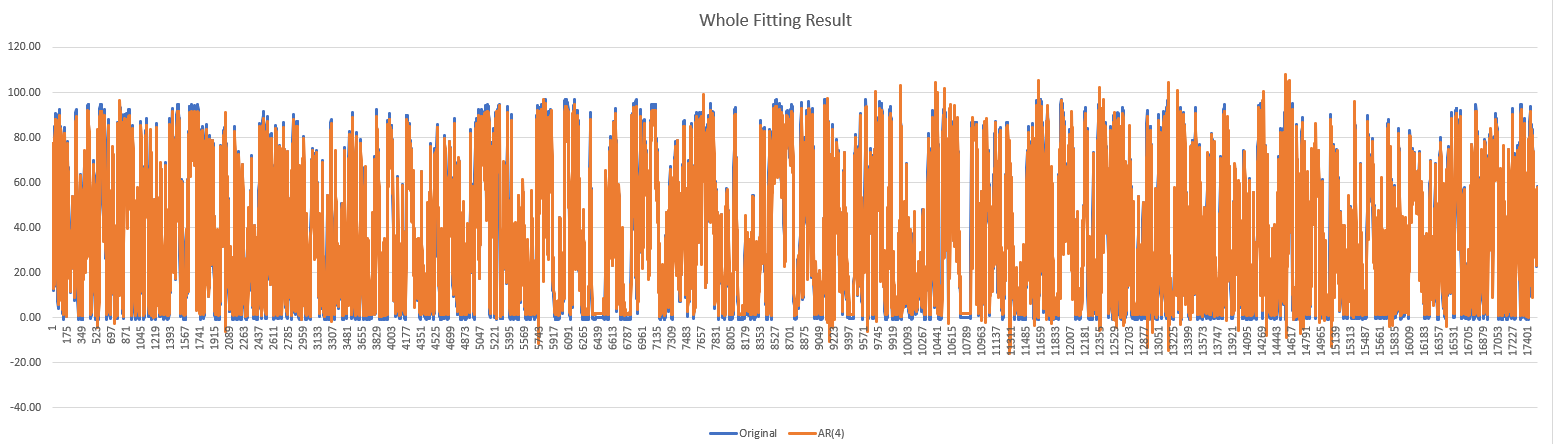
According to the graph above, it not easy to find the few frequencies that holds the important position. It seems that there are lots of frequencies that are important. So, we could say there is no significant seasonality in the data. So for this case, I will not use the seasonality to model.

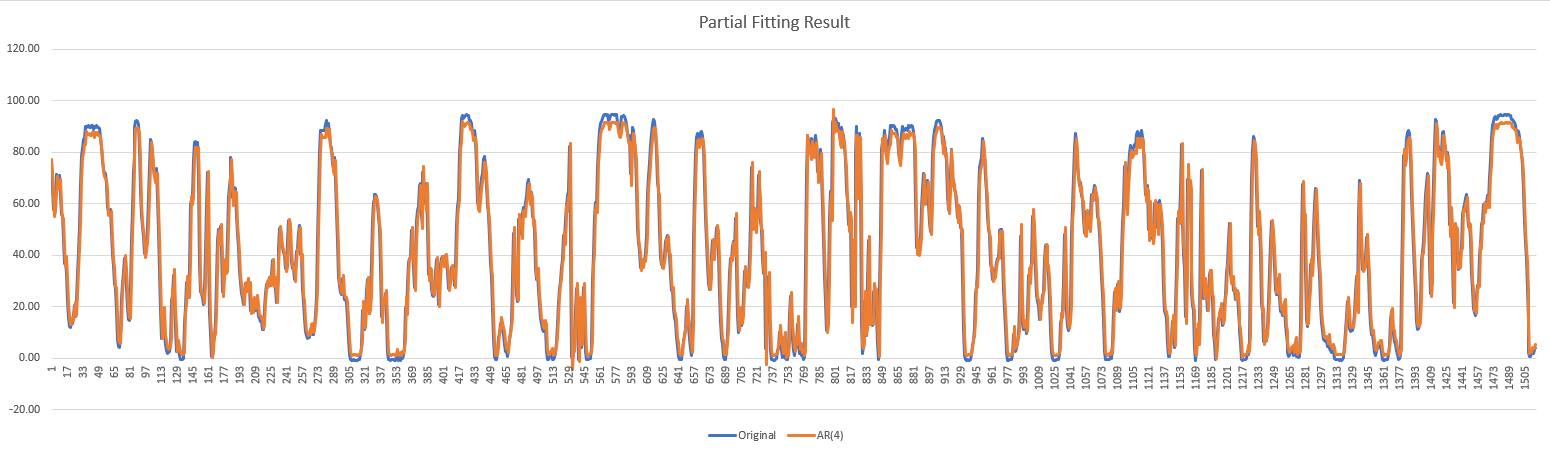
## Task 2: Compare AR(p) and ARMA(p,q)

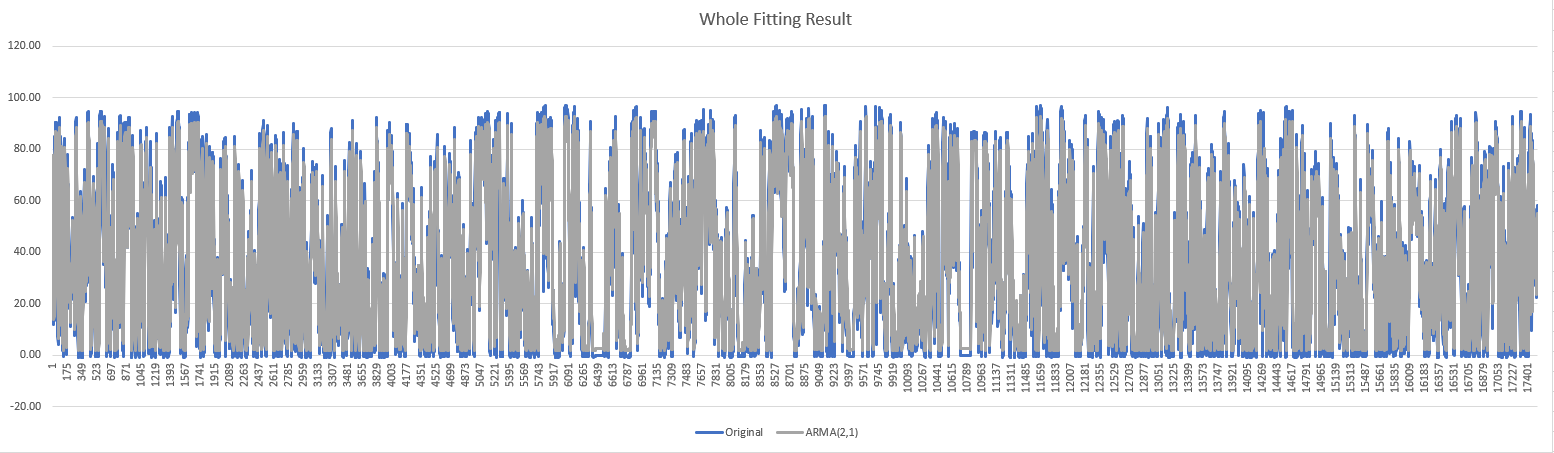
This part will try to find the best AR model and ARMA model for this dataset. After several searching steps, will get two models that are the best AR(4) and ARMA(2,1) suit for this dataset. The coefficients for the two models like the pictures below.

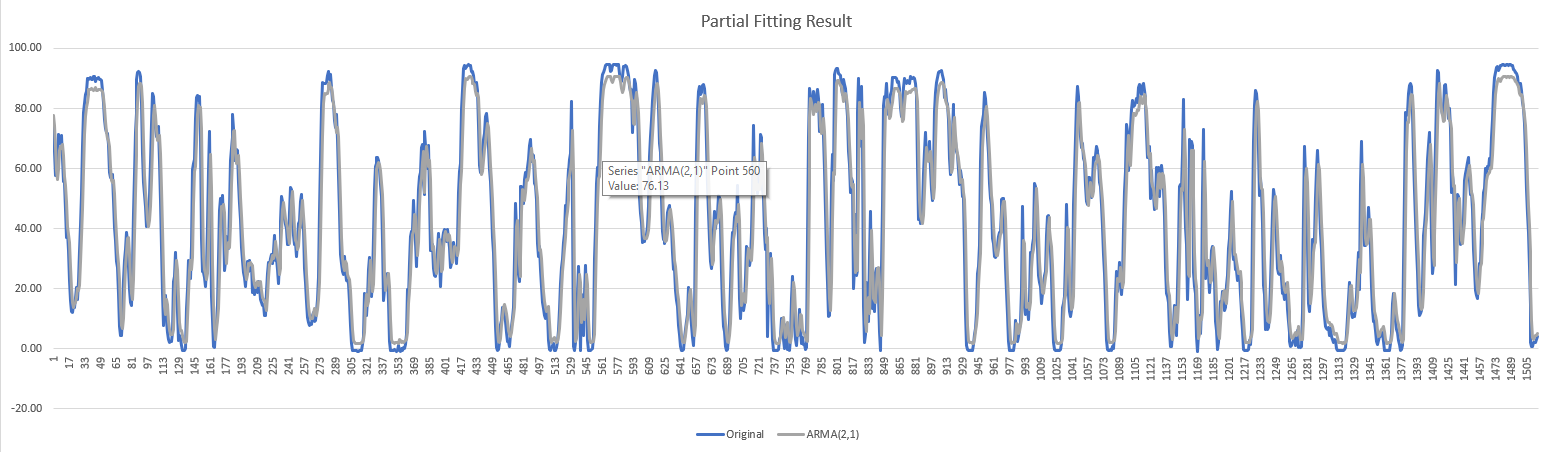


Using the two coefficients to forecast the data, the results will get like the pictures below.

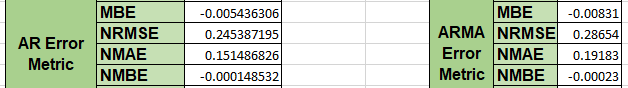








According to the pictures above, we could see the two models both fit the dataset very well. Next steps will compare the error metrics for the two models on the dataset.



According to the picture above, it could be easily found that the all the values of the indicators are similar, which means the performance of the two model is similar. The values of MBE, NRMSE, NMAE and NMBE for AR model are greater than ARMA model, the performance of AR model is a bit better than ARMA, actually the difference is quite small.

## Task 3: Number of the coefficients

According to the coefficients of the two model, the AR(4) contains one more parameter than ARMA(2,1), According to the comparison of the two models, the performance of AR(4) is a little higher than ARMA(2,1), thus, it’s worth to use the one with extra parameters.