Assignment 1

Haiyue August 25, 2023 About 8 min Assignment

Requirements

Instructions

Requirements Analysis

1. For this question you will need the files from the website named
   a . HalfHourSolarRadiation2017.xlsx
   b . HalfHourSolarRadiation2018.xlsx
   c . PowerSpectrumGeneric.xlsm
   d . SolarTemplate.xlsm

   The tasks for this question are listed below.

   1. Take the solar radiation data from File a , and copy it into File c , and run the power spectrum tool to find out which frequencies are important.
   2. Use File c to find the Fourier series model for the seasonality.

   > **Info**
   >
   > Note that **the Template is designed for hourly data**. You will have to make some adjustments to use it for half hourly data plus change the relevant frequencies if necessary.

   3. Take the difference between the data and the Fourier model - the residuals - and take them to Minitab and find the best ARMA(p,q) model.
   4. Use the ARMA model to forecast one step ahead for the residuals and add that to the Fourier series model to get the full one step ahead forecast.
   5. Use the error metrics defined below to evaluate the model.
   6. Use the models you have developed for 2017 to see how they perform for the 2018 data, the out of sample data. Comment on the differences in the error metrics.

The Normalised Mean Bias Error (NMBE) is defined by taking the difference between the data y_i and thee model $\hat{y}_i$ for all i and dividing by the number of data values. To normalise it, we divide by the mean of the data.

> **Tips**
>
> Note that for solar radiation, we only do the calculation for solar elevation greater than or equal to 10 degrees. That is why I included the elevation data.

$$NMBE = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{n\bar{y}}$$

We also define the Normalised Mean Absolute Error (NMAE)

$$NMAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n\bar{y}}$$

All works below

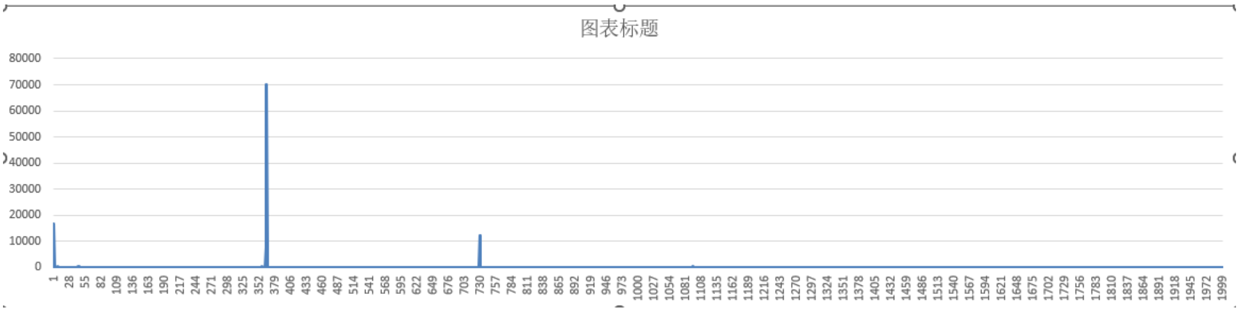
Question 1

The GHI column is selected as the target variable. Because of the data is half hourly dataset, it's a lots of data and not easy to use for analysis. So I convert it into daily data.

Taks1: Getting Frequencies

- DFT
  Using `powerspectrum` excel to get the best frequencies.

> **Info**
>
> Step1: using half-hourly data
> Step2: Number of data: 17520, Number of frequencies: 2000



It could easily be found that the power of frequencies is greater than 500. Then we could select the value like

Frequency	A(i)	B(i)	Power	> 5000
	171.03		117014.8	
1	-128.69	14.9702	16785.19228	TRUE
364	87.08268	1.911397	7587.046397	TRUE
365	-263.347	21.43233	69810.87885	TRUE
366	90.51233	-17.0102	8481.829549	TRUE
730	109.2063	-17.221	12222.58237	TRUE

1. Results

> **Results**
>
> The best frequencies are [1, 364,365, 366, 730].

Task2: Getting the fourier model

- Seasonality

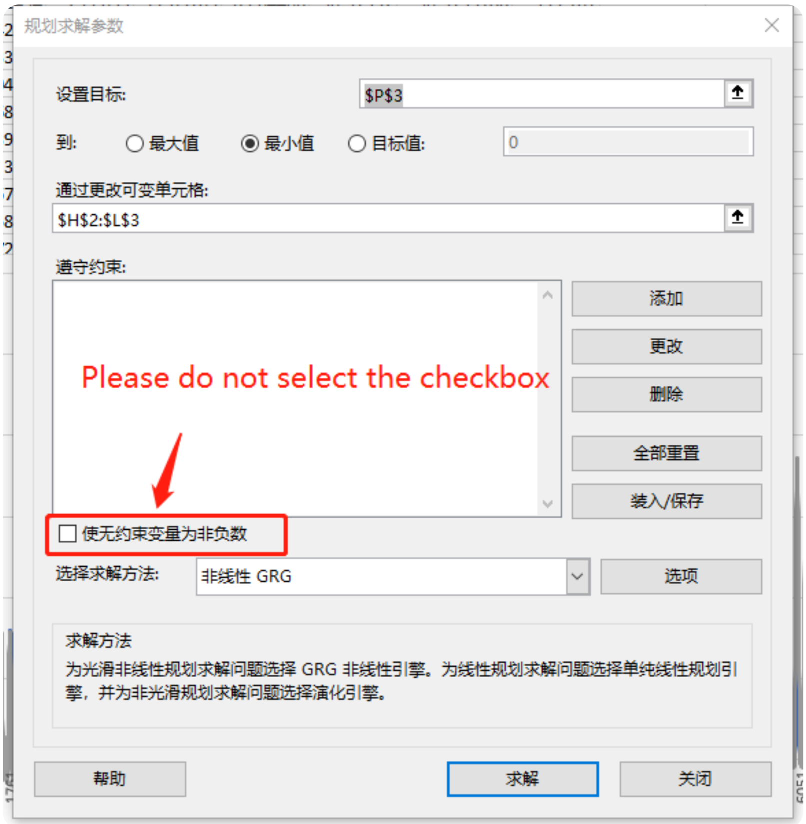
  We could calculate the coefficients for the seasonality using fourier model.

F	G	H	I	J	K	L	M	N	O	P
		0.000359	0.130541	0.1309	0.131258	0.261799				
		-128.69	87.08317	-263.347	90.51197	109.2061				
mean	171.0372	14.96779	1.897821	21.43217	-17.0138	-17.2199			SSE	178226938
T	GHI	1	364	365	366	730	model	final model		
1	0	-128.685	86.58927	-258.297	87.5066	101.0282	59.17967	59.17967041	3502.233	
2	0	-128.679	84.6219	-248.827	82.99575	85.96531	47.11396	47.11396147	2219.725	
3	0	-128.674	81.21454	-235.099	77.05705	65.04406	30.57952	30.5795186	935.107	
4	0	-128.668	76.42517	-217.349	69.79266	39.69017	10.92736	10.92736161	119.4072	
5	0	-128.663	70.33529	-195.88	61.32755	11.63145	-10.2119	0	104.2832	
6	0	-128.658	63.04853	-171.06	51.80736	-17.2199	-31.0442	0	963.7445	
7	0	-128.652	54.68888	-143.312	41.39588	-44.8978	-49.7403	0	2474.1	
8	0	-128.647	45.39861	-113.113	30.27222	-69.516	-64.5674	0	4168.946	
9	0	-128.641	35.3358	-80.9779	18.62775	-89.3967	-74.015	0	5478.216	

  The parameters like below:

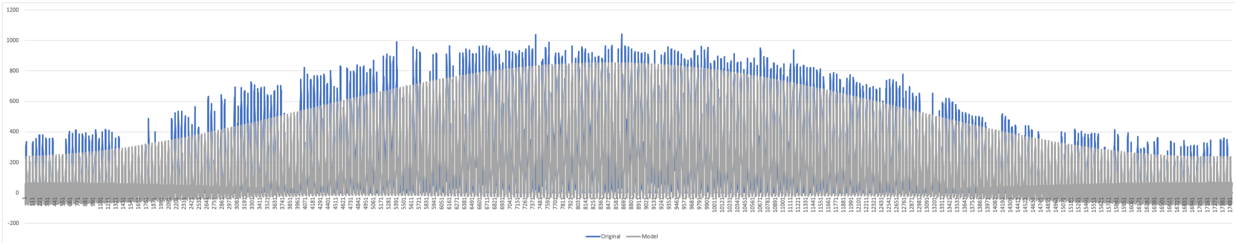
```VBA
P3 = SUM(O5:O17524)
G3 = AVERAGE(G5:G17524)

H1 = 2*PI()/17520*H$4
I1 = 2*PI()/17520*I$4
J1 = 2*PI()/17520*J$4
K1 = 2*PI()/17520*K$4
L1 = 2*PI()/17520*L$4

'The formula should be drag down for filling the necessary cells
H5 = H$2*COS(H$1*$F5)+H$3*SIN(H$1*$F5)
I5 = I$2*COS(I$1*$F5)+I$3*SIN(I$1*$F5)
J5 = J$2*COS(J$1*$F5)+J$3*SIN(J$1*$F5)
K5 = K$2*COS(K$1*$F5)+K$3*SIN(K$1*$F5)
L5 = L$2*COS(L$1*$F5)+L$3*SIN(L$1*$F5)

M5 = SUM(H5:L5)+$G$3
O5 = (G5-M5)^2
```

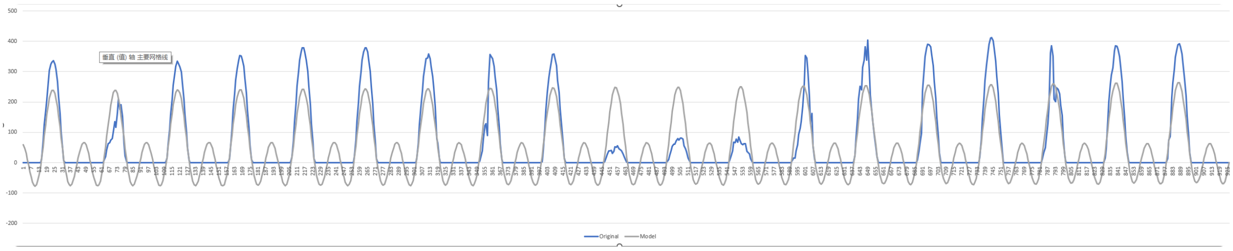
We should use the solver to minimize the target function in `P3`



Alt text

We could see the seasonality results





Part fitting results

Task 3: Getting the coefficients for AR model

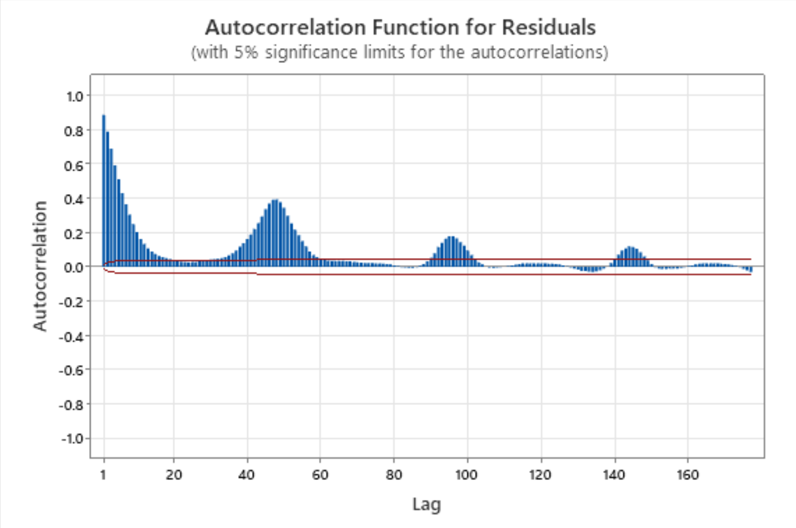
Now we got residuals

		0.000359	0.130541	0.1309	0.131258	0.261799						
		-128.69	87.08317	-263.347	90.51197	109.2061						
mean	171.0372	14.96779	1.897821	21.43217	-17.0138	-17.2199			SSE	178226938		
T	GHI	1	364	365	366	730	model	final model			Residuals	
1	0	-128.685	86.58927	-258.297	87.5066	101.0282	59.17967	59.17967041	3502.233		-59.1797	
2	0	-128.679	84.6219	-248.827	82.99575	85.96531	47.11396	47.11396147	2219.725		-47.114	
3	0	-128.674	81.21454	-235.099	77.05705	65.04406	30.57952	30.5795186	935.107		-30.5795	
4	0	-128.668	76.42517	-217.349	69.79266	39.69017	10.92736	10.92736161	119.4072		-10.9274	
5	0	-128.663	70.33529	-195.88	61.32755	11.63145	-10.2119	0	104.2832		10.21192	
6	0	-128.658	63.04853	-171.06	51.80736	-17.2199	-31.0442	0	963.7445		31.04423	

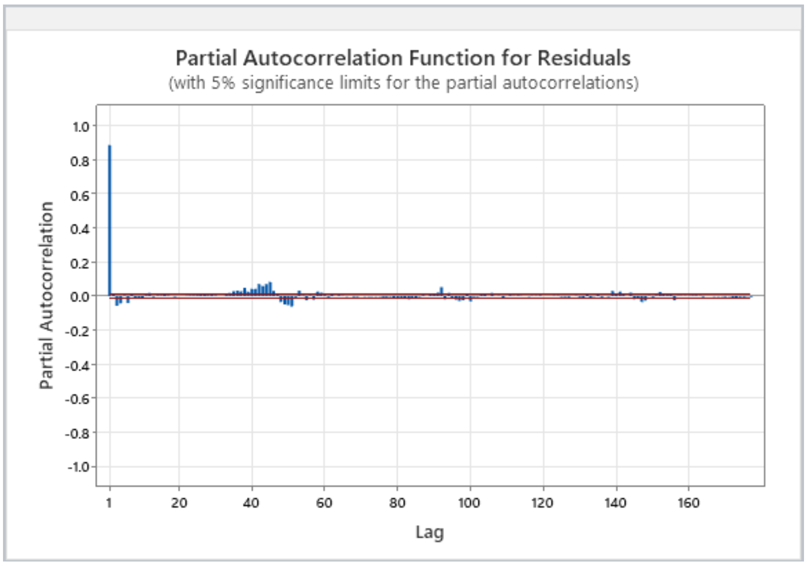
The formula of residuals

```VBA
O5 = G5-M5
```

Then copy all the residuals to minitab for autocorrelation analysis.
Auto Corrrelation Result



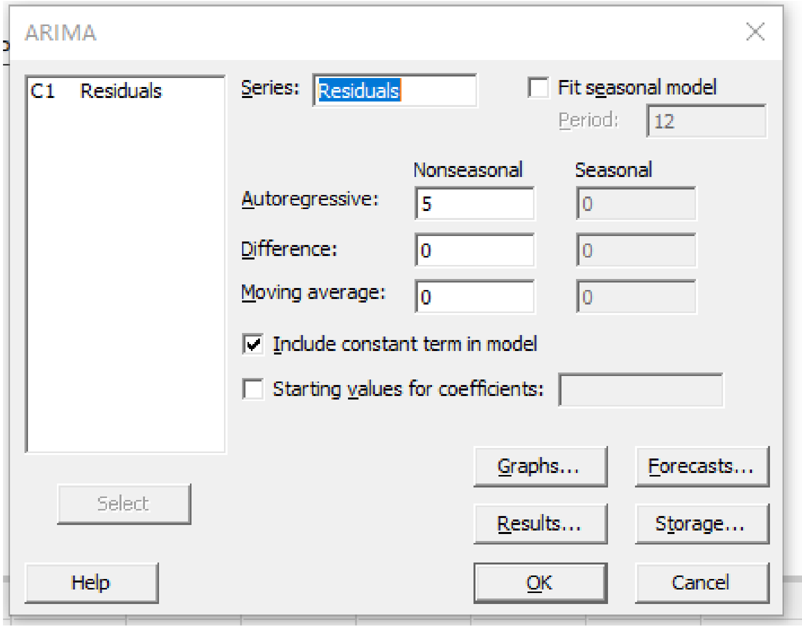
Partial Auto Corrrelation Result



According to the results of Auto correlation and Partial Auto Correlation, it can see the values are correlated with the past values.

Now we try to use the ARIMA to find the coefficients of the forcasting model.

The first step we try to set the autoregressive is 5.



The results are here.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.87341	0.00756	115.59	0.000
AR 2	0.0655	0.0100	6.53	0.000
AR 3	-0.0200	0.0100	-1.99	0.046
AR 4	-0.0459	0.0100	-4.58	0.000
AR 5	0.00395	0.00756	0.52	0.601
Constant	-0.006	0.352	-0.02	0.987
Mean	-0.05	2.86		

Number of observations: 17520

We could see the pvalue of AR5 and Constant are more then 0.05, so we need the exclude the constant and decrease the number of lags, now we try to set autoregressive is 4.

Final Estimates of Parameters

Type		Coef	SE Coef	T-Value	P-Value
AR	1	0.87324	0.00755	115.68	0.000
AR	2	0.0654	0.0100	6.53	0.000
AR	3	-0.0197	0.0100	-1.97	0.049
AR	4	-0.04244	0.00755	-5.62	0.000

Number of observations: 17520

According to the results we could see all the pvalues are less than 0.05, so we could use AR(4) model to model the residuals.

Task 4: Using the ARMA model to forecast

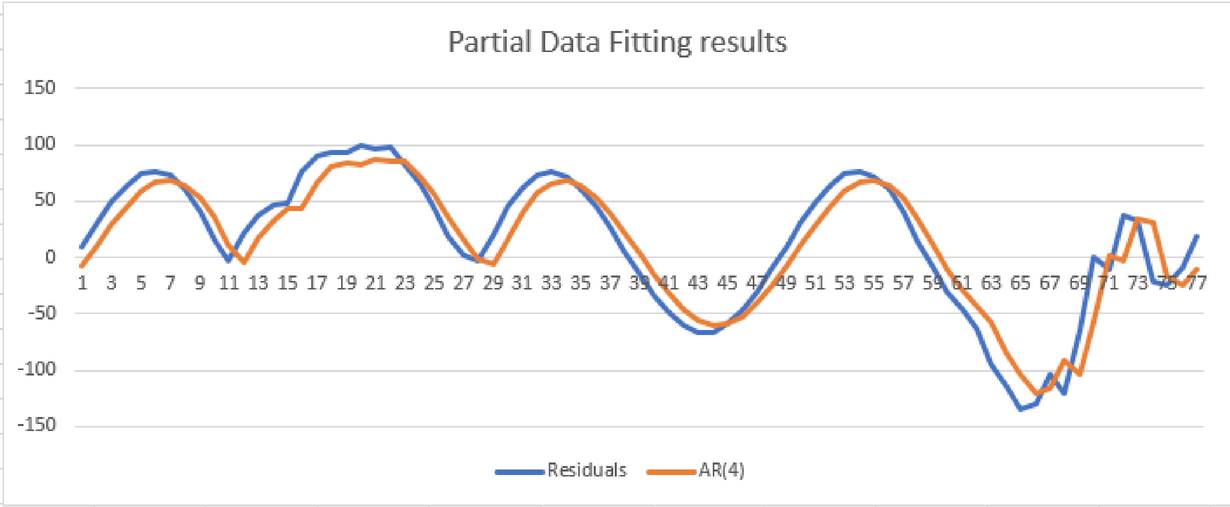
Now we copy the coefficients to excel, and try to model the AR(4) like the picture below.

178226938						Final Estimates of Parameters				
	Residuals		AR(4)			Type	Coef	SE Coef	T-Value	P-Value
	-59.1797					AR 1	0.87324	0.00755	115.68	0
	-47.114					AR 2	0.0654	0.01	6.53	0
	-30.5795					AR 3	-0.0197	0.01	-1.97	0.049
	-10.9274					AR 4	-0.04244	0.00755	-5.62	0
	10.21192		-8.10238							
	31.04423		10.80474							
	49.74033		29.28999							
	64.56738		45.72812							
	74.01497		58.59087							
	76.90652		66.55814							
	[illegible]		[illegible]							

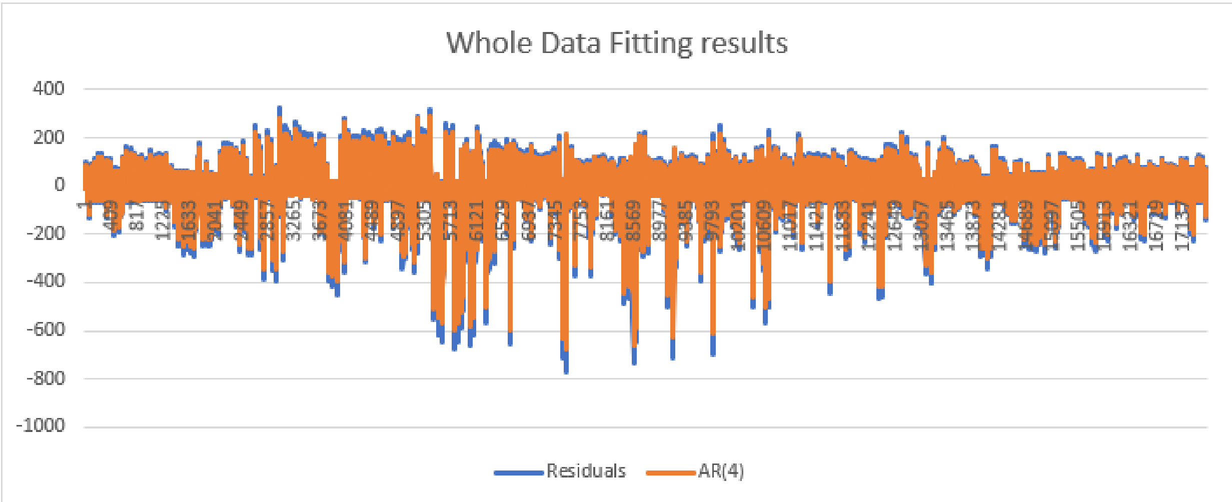
```VBA
S9 = Q8*$V$5 + Q7*$V$6 + Q6*$V$7 + Q5*$V$8
```

We could to visualize the fitting result of AR(4)

Partial Fitting Result



Whole Data Fitting Result



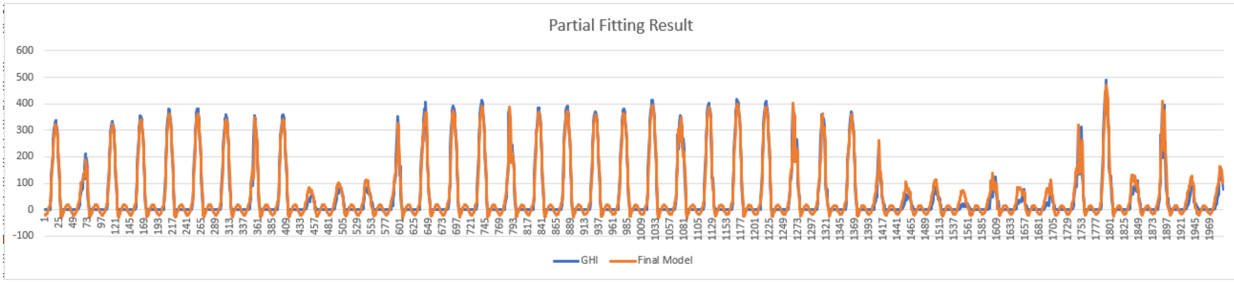
According to the results above, we could know, the seasonality and AR(4) could fitting very well on our data. Now we need to combine the two components to see the final result.

M	O	P	Q	R	S	AA
	SSE	178226938				
model			Residuals		AR(4)	Final Model
59.17967	3502.233		-59.1797			
47.11396	2219.725		-47.114			
30.57952	935.107		-30.5795			
10.92736	119.4072		-10.9274			
-10.2119	104.2832		10.21192		-8.10238	-18.3142953
-31.0442	963.7445		31.04423		10.80474	-20.2394978
-49.7403	2474.1		49.74033		29.28999	-20.4503346
-64.5674	4168.946		64.56738		45.72812	-18.8392618
-74.015	5478.216		74.01497		58.59087	-15.424099
-76.9065	5914.613		76.90652		66.55814	-10.3483879
-72.4886	5254.599		72.48861		68.61548	-3.87313495
-60.4918	3659.256		60.49178		64.13131	3.63952298

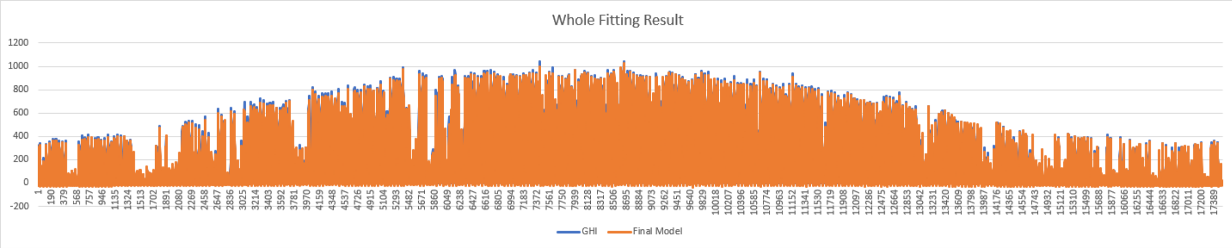
Alt text

```VBA
AA9 = M9+S9
```

Partial Final Model Fitting Result



Whole Final Model Fitting Result



According to the graph, we could know the final model could fit the data very well.

Task 5: To evaluating the model

I switch to template excel to implement all the steps. So, the postion for cells will totally change. You could change all values above to the correct position.

We could know the error metric include several indicators.

- indicators

> **Formulas**
>
> $$MeAPE = MEDIAN(|\frac{\hat{y}_i - y_i}{y_i}| * 100)$$
> $$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$$
> $$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}}{\bar{y}}$$
> $$NMBE = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{n\bar{y}}$$
> $$NMAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n\bar{y}}$$

K	L	M	N	O	P	Q	R	S	T	U	V
	AR coefficients	Type	Coef	SE Coef	T-Value	P-Value			Error Metric	MeAPE	8.378448548
		AR 1	0.87324	0.00755	115.68	0				MBE	0.228434102
		AR 2	0.0654	0.01	6.53	0				NRMSE	0.174231879
		AR 3	-0.0197	0.01	-1.97	0.049				NMAE	0.106781018
		AR 4	-0.04244	0.00755	-5.62	0				NMBE	0.000565729
Valid_Count	7266							Final Mean	403.79		
				$y_i - \hat{y}_i$	$\lvert y_i - \hat{y}_i \rvert$		$\left\lvert \frac{\hat{y}_i - y_i}{y_i} \right\rvert$				
Valid_Elevatior	*Residuals*	AR(4)	Final Mod	Diff	ABS(Diff	Diff^2	Diff/yi				
1	47.40	33.0788145	157.93	14.33	14.32569124	205.2254294	8.316319				
1	47.55	43.48033625	202.81	4.07	4.07301817	16.58947701	1.968751				
1	75.98	42.98578957	232.50	32.99	32.99301503	1088.539041	12.42706				
1	90.71	66.95514036	280.59	23.76	23.75538825	564.3184707	7.805372				
1	93.68	81.23242752	311.50	12.45	12.44693046	154.9260779	3.842278				
1	92.69	84.22208432	322.64	8.47	8.468289633	71.71192931	2.557546				

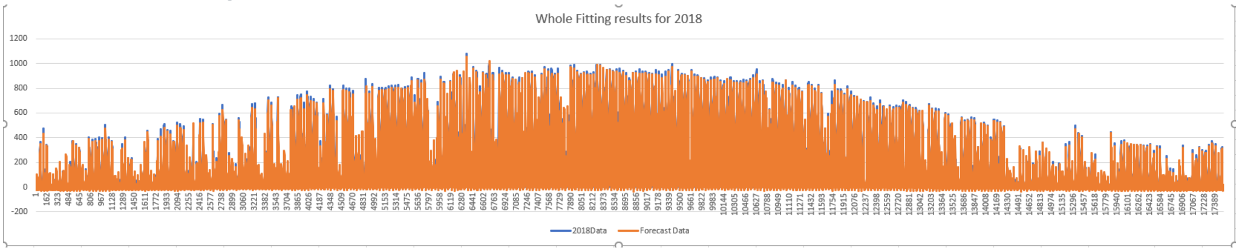
Solar2017

```vba
L7 = SUBTOTAL(2,K15:K17530)
T7 = SUBTOTAL(1,N15:N17530)
V1 = AGGREGATE(12,1,R15:R17530)
V2 = SUBTOTAL(109, O15:O17530)/L7
V3 = SQRT(SUBTOTAL(109,Q15:Q17530)/L7)/T7
V4 = SUBTOTAL(109, P15:P17530)/(L7*T7)
V5 = SUBTOTAL(109, O15:O17530)/(L7*T7)

# H column are the data of elevations
K1  = IF(H11>=10,1,0)

O15 = (A15-N15)
P15 = ABS(O15)
Q15 = O15^2
R15 = IF(A15=0,0,ABS(K15*P15/A15)*100)
```

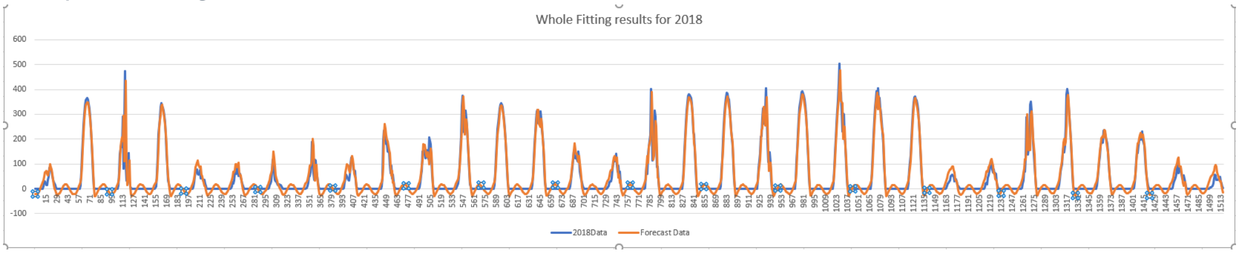
If we want to select the valid metric according to our data, should filter all the data where `K` column equals 1.
Then we get.

Task 6: Testing the ARMA model using 2018 data

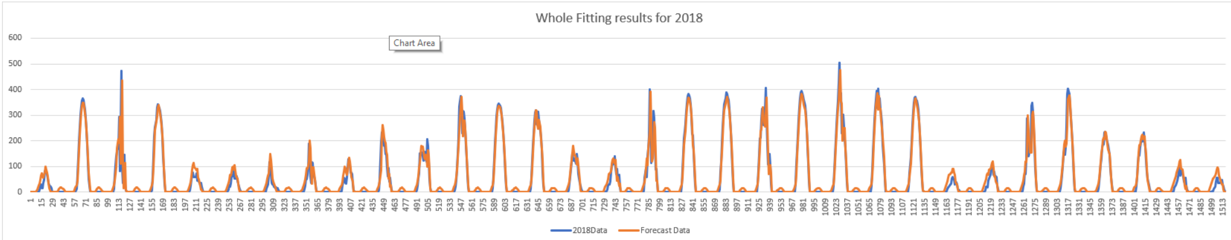
Using the model get from the 2017 dataset, to predict the data in 2018.
The whole fitting result



The part fitting result



Actually, when the data if not possible be zero, so we could reset all negative number to 0. so we could get.



K	L	M	N	O	P	Q	R	S	T	U	V
	AR coefficients	Type	Coef	SE Coef	T-Value	P-Value			**Error Metric**	**MeAPE**	9.955924989
		AR 1	0.87324	0.00755	115.68	0				**MBE**	-1.786323383
		AR 2	0.0654	0.01	6.53	0				**NRMSE**	0.193146513
		AR 3	-0.0197	0.01	-1.97	0.049				**NMAE**	0.119774747
		AR 4	-0.04244	0.00755	-5.62	0				**NMBE**	-0.004674086
Valid_Count	7266							**Final Mean**	382.18		
				$y_i - \hat{y}_i$	$\lvert y_i - \hat{y}_i \rvert$		$\left\lvert \frac{\hat{y}_i - y_i}{y_i} \right\rvert$				
Valid_Elevatior	***Residuals***	**AR(4)**	**Final Mod**	**Diff**	**ABS(Dif**	**Diff^2**	**Diff/yi**				
1	-84.72	-49.73565221	65.00	-34.98	34.98499122	1223.949611	116.5389				
1	-134.30	-76.59601906	72.62	-57.70	57.70244235	3329.571853	386.832				
1	-164.50	-120.3802007	59.02	-44.12	44.11947715	1946.528264	296.1039				
1	-168.25	-148.4951531	55.03	-19.76	19.75613407	390.3048336	56.01399				
1	-172.83	-151.4408092	68.71	-21.39	21.39164866	457.6026325	45.20636				
1	-145.20	-152.9875793	75.32	7.78	7.782804112	60.57203984	9.365589				
1	-150.97	-127.805944	99.69	-23.16	23.16450475	536.5942802	30.27117				
1	-140.31	-130.7844429	86.99	-9.53	9.52966177	90.81445345	12.30322				

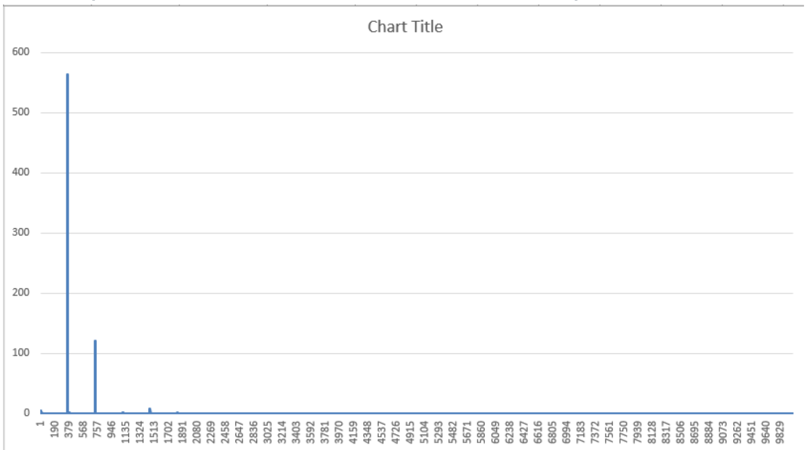
- Important Node

> **Important**
>
> Some steps above does not use the template file, it's a little ugly and unclear. I will redo it in the solar template, the files could download via the like below.

Question 2

Task 1: Find frequencies

Copy data from `solarfarm` to `powerspectrum`, using Number of objects: 105120, number of frequencies: 10000. We could see the power barchart like below.

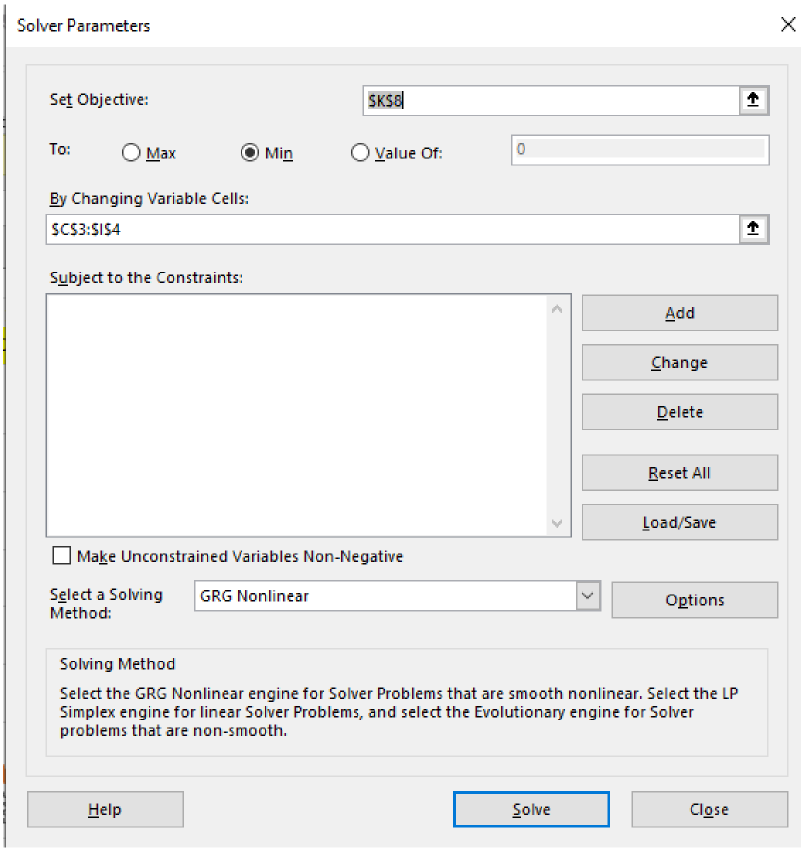


According to the graph above, we could know the best frequencies is around 10, here we regards the frequencies that the value of power are greater than 2, the best frequencies could get like the picture below.

Frequenc	A(i)	B(i)	Power	greater than 2
1	2.053534	0.146161	4.238365655	TRUE
364	-1.4523	0.111994	2.121704646	TRUE
365	-23.3043	-4.44338	562.8359535	TRUE
730	10.26733	3.943808	120.9716501	TRUE
1460	-1.8406	-2.0326	7.51928263	TRUE
1825	0.775591	1.185239	2.006331424	TRUE

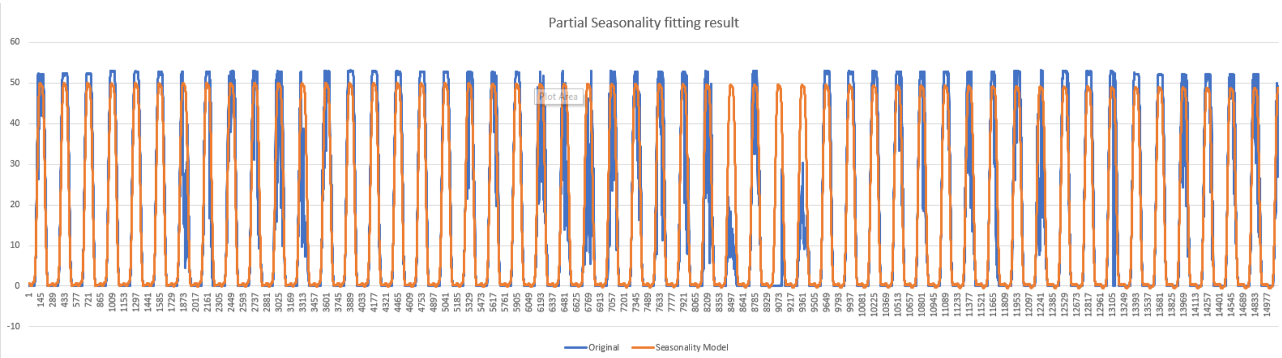
Task 2: Make seasonality

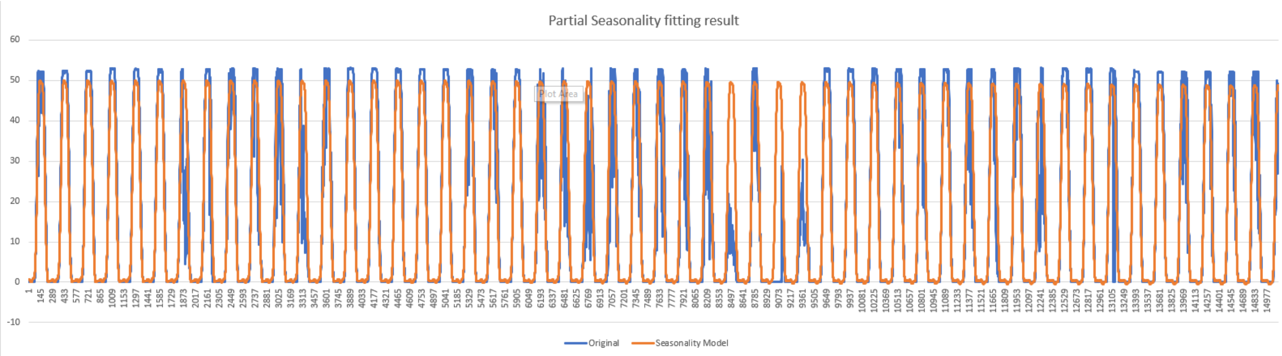
After getting the best frequencies, we could using solver to minimize the SSE(Sum square of error), to find the best coefficients to our model.

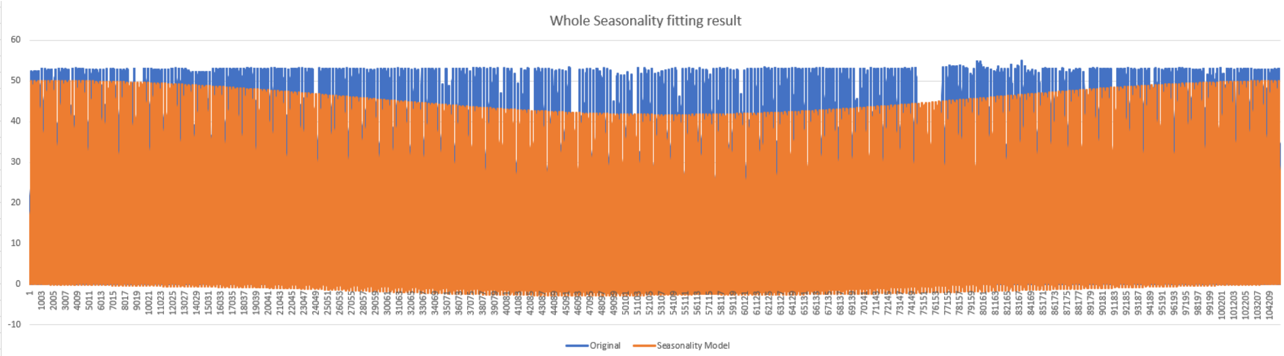


A	B	C	D	E	F	G	H	I	J	K
		1/year		1/day		2/day	4/day		n=	105120
Frequency		0.000	0.022	0.022	0.022	0.044	0.087	0.109		
Mean	14.79	2.05	-1.45	-23.30	-0.66	10.27	-1.84	0.78		
Varianc	403.38	0.15	0.11	-4.44	-0.17	3.94	-2.03	1.19		
	Contribution	2.12	1.06	281.42	0.23	60.49	3.76	1.00		
	Variance Explained	0.53%	0.26%	69.77%	0.06%	14.99%	0.93%	0.25%	*Three*	*OLS*
									Cycle	

We could visualize the seasonality model fitting results.





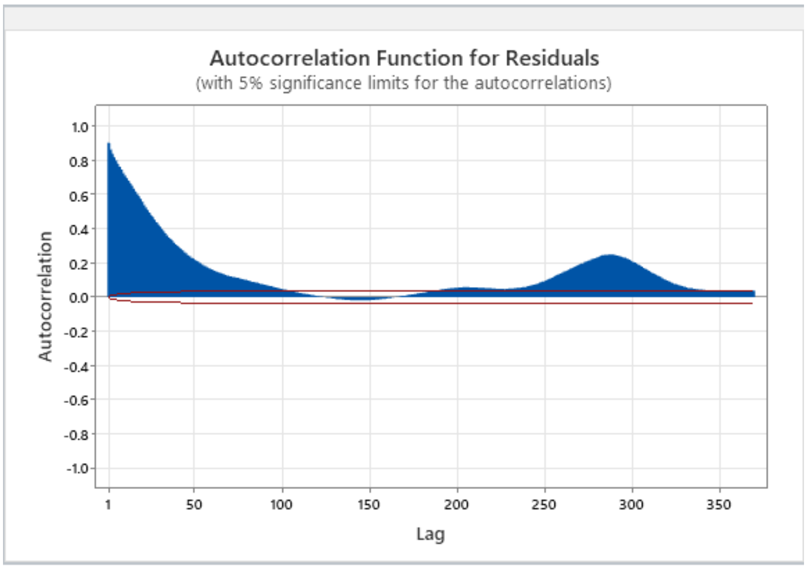


According to the graph, we could see the seasonality could fit the data well, but still have a large gap in the middle of the data.

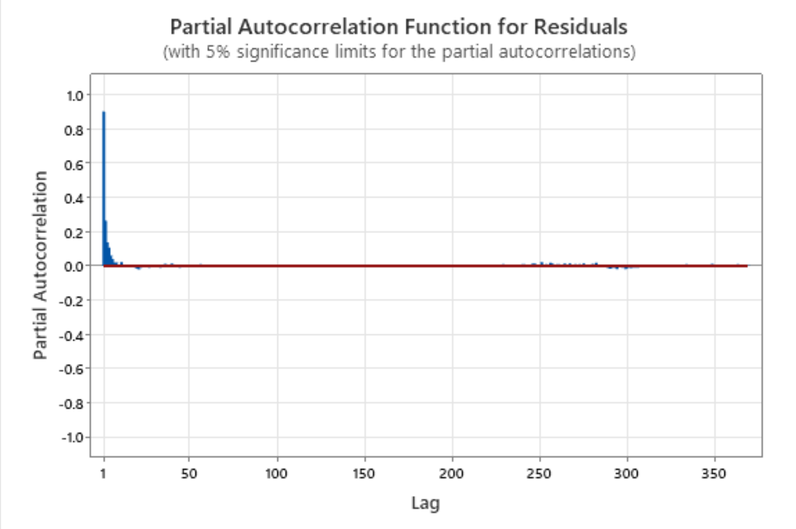
Task 3: ARIMA coefficients

After getting the seasonality model, then could get the residuals to find the cofficients for ARIMA model.

The autocorrelation result



The partial autocorrelation result



According to the data graph, we could find the data the correlated with the past data. Next we try to find the best coefficients for the forecasting model.

Final Estimates of Parameters

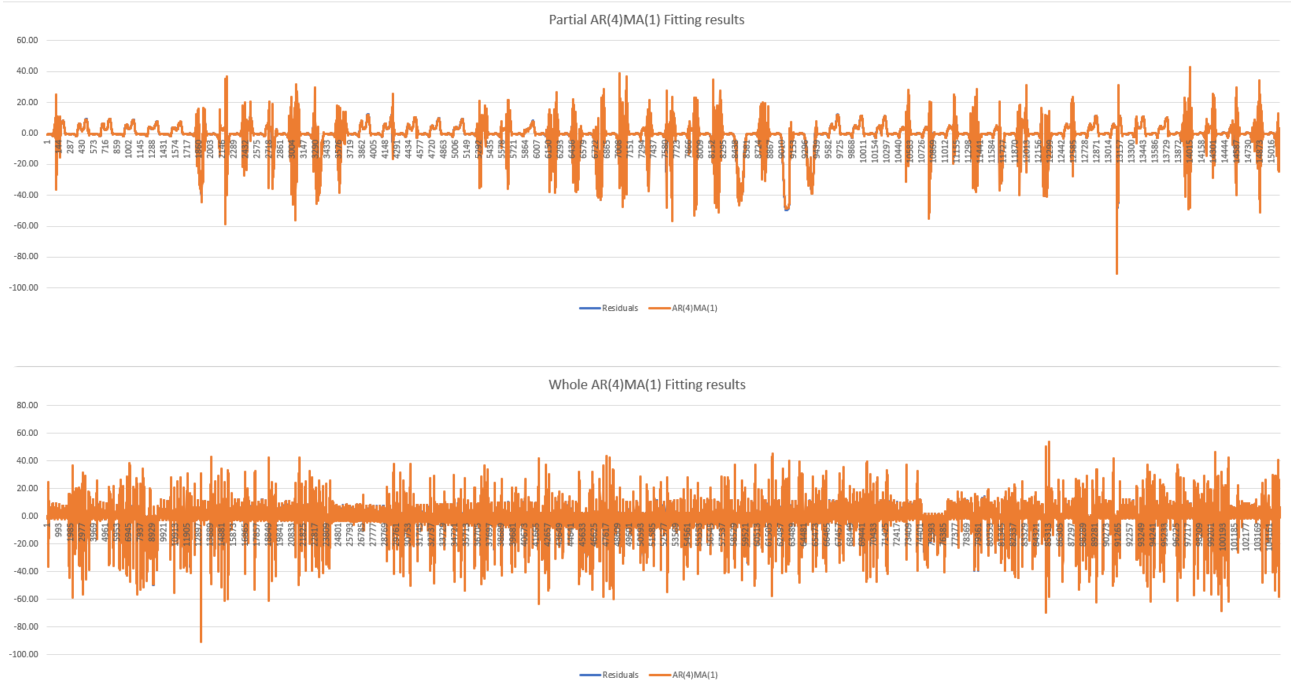
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.60433	0.00308	196.29	0.000
AR 2	0.15094	0.00359	42.02	0.000
AR 3	0.06173	0.00362	17.06	0.000
AR 4	0.07114	0.00359	19.80	0.000
AR 5	0.06075	0.00308	19.73	0.000
Constant	-0.00002	0.00915	-0.00	0.998
Mean	-0.000	0.179		

According to the graph above, we could know the constant is no significant with the model, so we should remove it during searching the coefficients.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.2025	0.0149	80.65	0.000
AR 2	-0.2147	0.0100	-21.48	0.000
AR 3	-0.03093	0.00508	-6.09	0.000
AR 4	0.02468	0.00462	5.34	0.000
MA 1	0.6021	0.0146	41.10	0.000

After getting the coefficients, we could using the model to forecast the data, the pictures below show the forecasting results.

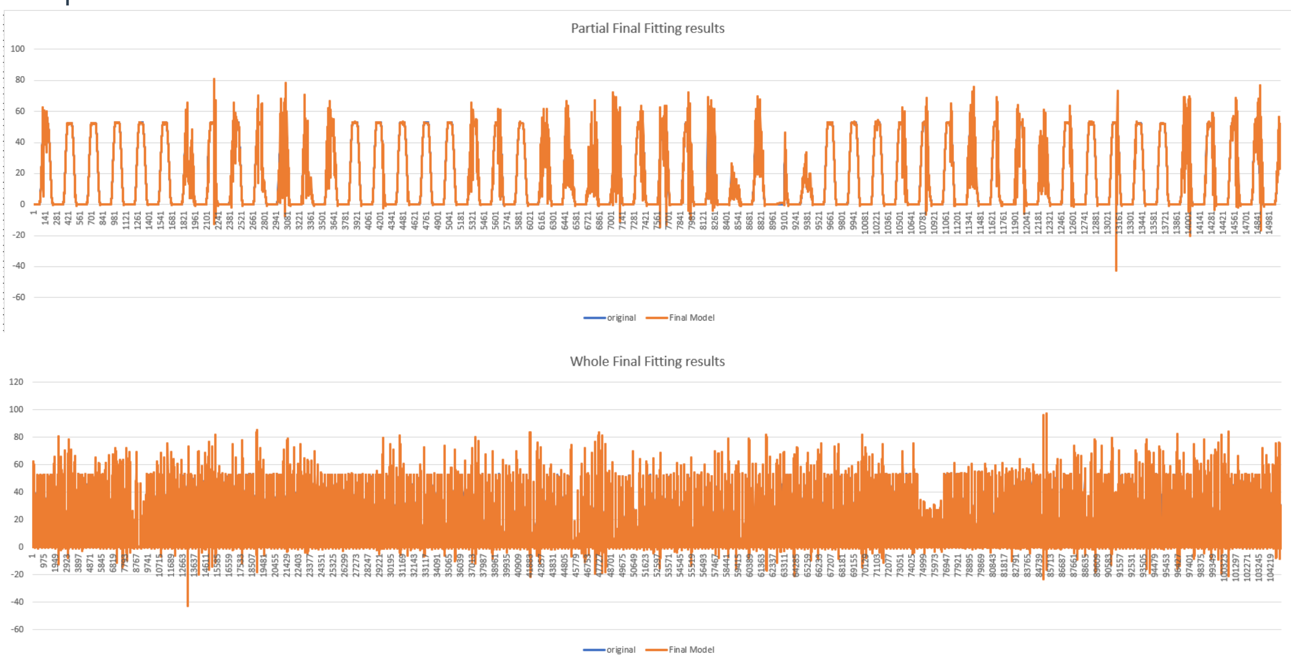


Alt text

Task 4: Final model

The previous has been split the data analysis into two components, in this step, combine the two components to form the final model.

The picture below will show the final results.



Task 5: Error Metric

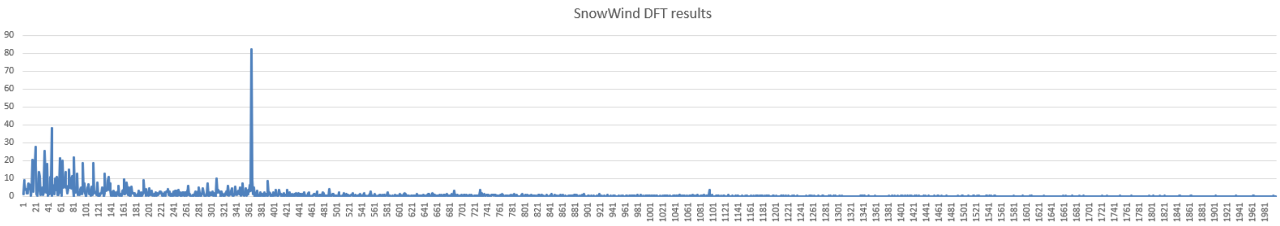
According to the source data, which contains zero value, so the MeAPE is not suit for evaluating the forecasting model.

T	U	V
Error Metric	MeAPE	#DIV/0!
	MBE	1.8911E-06
	NRMSE	0.317412531
	NMAE	0.087043384
	NMBE	1.27862E-07

Question 3

No significant seasonality

When we use the power spectrum, the result should get like



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

Compare AR(p) and ARMA(p,q)

According to the previous results, we just need to find the coefficients for AR(p) and ARMA(p,q) model on the original dataset.

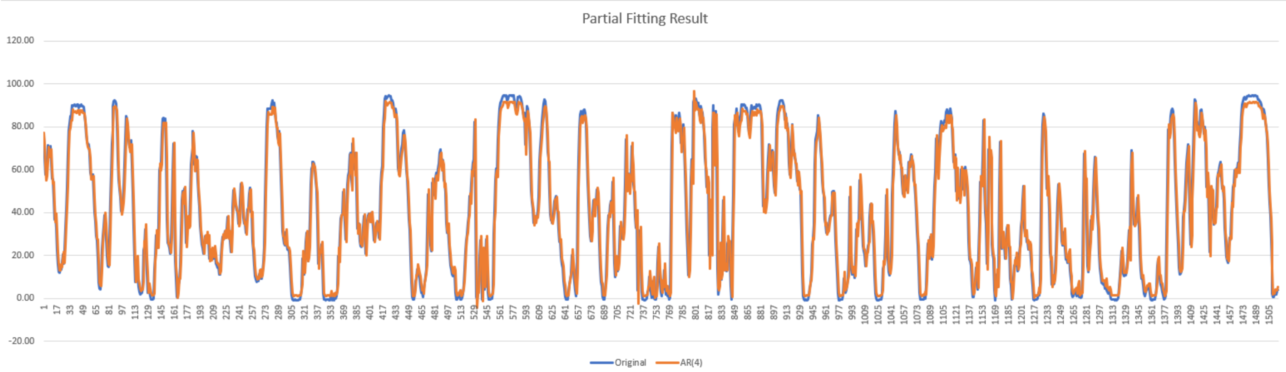
After searching, the best AR model should be AR(4), and ARMA model should be ARMA(2,1).

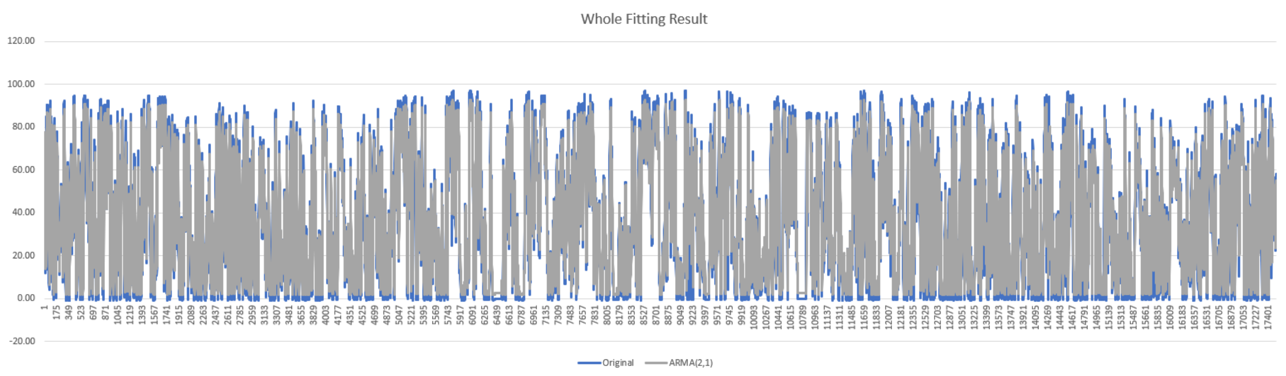
Final Estimates of Parameters

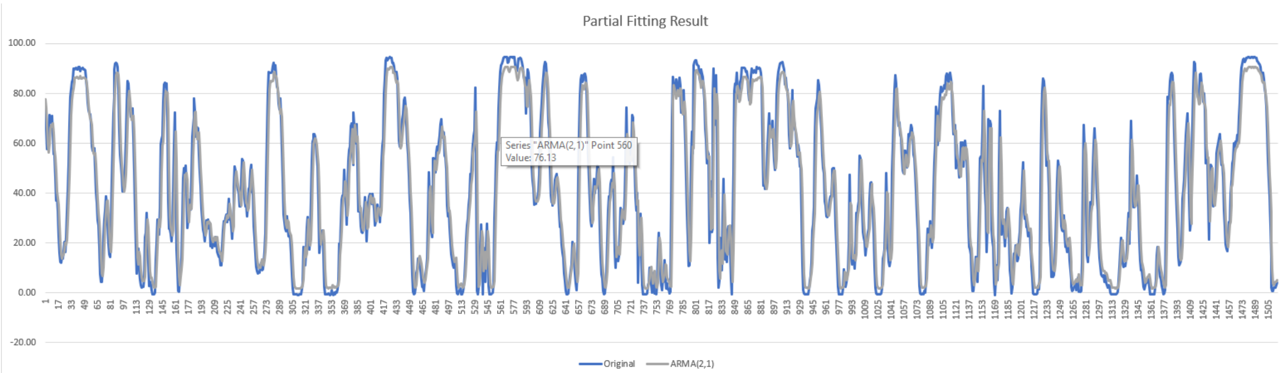
Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.8524	0.0351	24.30	0.000
AR 2	0.0792	0.0337	2.35	0.019
MA 1	-0.3003	0.0337	-8.92	0.000
Constant	2.5098	0.0883	28.44	0.000
Mean	36.66	1.29		

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	1.15287	0.00755	152.66	0.000
AR 2	-0.2645	0.0115	-22.99	0.000
AR 3	0.0911	0.0115	7.92	0.000
AR 4	-0.03351	0.00755	-4.44	0.000
Constant	1.9821	0.0679	29.21	0.000
Mean	36.66	1.26		







AR Error Metric	MBE	-0.005436306
	NRMSE	0.245387195
	NMAE	0.151486826
	NMBE	-0.000148532

ARMA Error Metric	MBE	-0.00831
	NRMSE	0.28654
	NMAE	0.19183
	NMBE	-0.00023

[Download the solarRadiation Process](#)
[The final Excel](#)

References

1. Seasonality Analysis and Forecast in Time Series (https://medium.com/swlh/seasonality-analysis-and-forecast-in-time-series-b8fbba820327)
2. Seasonality: What It Means in Business and Economics, Examples (https://www.investopedia.com/terms/s/seasonality.asp)
3. Using Python and Auto ARIMA to Forecast Seasonal Time Series (https://medium.com/@josemarcialportilla/using-python-and-auto-arima-to-forecast-seasonal-time-series-90877adff03c)
4. A Guide to Time Series Analysis in Python (https://builtin.com/data-science/time-series-python)
5. Error Metrics: How to Evaluate Your Forecasts (https://www.jedox.com/en/blog/error-metrics-how-to-evaluate-forecasts/)

Last update: 9/3/2023, 9:26:19 PM

Contributors: Haiyue,Wang