PROJECT REPORT (Q2)

Group 5(Pearson Group)

RAVI TEJA CHILAKALA	**2016B4A10490P**
PADIPELLY HEMANTH RAO	2017A7PS0159P
PRIYANKA	2016B4TS0957P
SRINIVAS NAVEEN	2017A1PS0801P
SHUBHAM KUMAR	2017B4A40882P
VISHAL KUMAR SINGH	2017B4AB0804P
SHUBHAM SHUKLA	2017A2PS0821P
KETAN CHANDAK	2017B4A40879P

NOVEMBER 22, 2019

1. INTRODUCTION TO RENEWABLE ENERGY

Renewable energy is produced from the sources that are constantly being replenished, such as sunlight, water, and wind. This means that we can use them without having to worry about them running out. Moreover, renewable energy sources are much more environmentally friendly than fossil fuels because they release very few chemicals, like carbon dioxide, that can potentially harm the environment.

In India, MNRE (Ministry of New and Renewable Energy) is the nodal Ministry of government that is concerned with all the matters related to renewable energy. The main aim of the Ministry is to develop and deploy new and renewable energy to meet the energy requirements of the country. Among others, one of the mission of the Ministry is to ensure the increase in the share of clean power i.e. renewable (bio, wind, hydro, solar, geothermal & tidal) electricity to lessen the fossil fuel based electricity generation.

As of 2019, approximately 35% of India's total electricity is produced from renewable resources and it aims to increase it to 57% by 2027. The renewable energy technologies related to reduce the use of nonrenewable electricity include: Solar street lighting systems, Solar lanterns and solar home lighting systems, Solar water heating systems, Akshay Urja / Aditya Solar Shops, Wind pumps, Micro-Hydal plants, and Standalone solar/ biomass based power generators.

According to the research by University of Technology in Finland, there is a great potential for India to move into a fully renewable electricity system by 2050 because of the abundance of renewable resources. This is possible if we can deploy new technologies which is being done by MNRE. Renewable energy's future in India looks bright as around 293 global and domestic companies have signed to generate 266 GW of solar, wind, Mini-Hydal and biomass-based power in India over the next decade. This would require an investment of $310 billion - $350 billion (Euros 27 billion to Euros 30 billion). The International Finance Corporation, the investment arm of the World Bank Group, is planning to invest about $6 billion by 2022 in several sustainable and renewable energy programs in India. With the investment potential of INR 15 trillion over the next four to five years in Indian power sector indicates immense opportunities in power generation, distribution, transmission and equipment.

2. TERMS ASSOCIATED WITH SOLAR ENERGY AND WIND ENERGY

1. DHI- Diffuse horizontal irradiance or diffuse sky radiation is radiation at the earth's surface from light scattered by the atmosphere. It is basically the amount of radiation received per unit area by a surface that does not arrive on a direct path from the sun, but has been scattered by molecules and particles in the atmosphere. Basically, it is the illumination that comes from clouds and the blue sky. It is a term obviously related to the solar energy.

2. DNI- It stands for Direct Normal Irradiance. It is the amount of solar radiation received per unit area by a surface that is always held perpendicular to the rays that come in a straight line from the direction of the sun at its current position. It is also a term related to solar energy. Typically, the amount of irradiance annually received can be maximized by a surface by keeping it normal to incoming radiation. This quantity is of particular interest to concentrating solar thermal installations and installations that track the position of the sun.

3. GHI- It stands for Global horizontal irradiance. It is also a term related to solar energy. The radiation reaching the earth's surface can be represented in a number of different ways. GHI is the total amount of shortwave radiation received from above by a surface horizontal to the ground. This value is of particular interest to photovoltaic installations (related to solar power panels) and includes both DNI and DIF (diffuse horizontal irradiance).

4. Dew point- The atmospheric temperature below which, water droplets begin to condense and dew can form. Both temperature and dew point correlate with solar intensity at higher values: if the temperature or dew point is high, then the solar intensity is likely to be high. However, if the temperature or dew point is low, the solar intensity exhibits more significant variation between high and low values. Hence it is clearly related to solar energy.

5. Temperature- It is just the temperature that is persisting throughout the day. It is strong measure of the solar radiation throughout the day. The intensity of radiations directly influences the amount of solar energy generated. Hence, it is related to solar energy.

6. Pressure- It is a term which is related to both wind and solar energy. Because, wind flows from high pressure to low pressure, if pressure gradient is high, then wind will flow rapidly (e.g. coastal areas). Also, development of pressure gradient depends on the temperature throughout the day. Higher temperature in an area, means pressure there will be low. Higher temperature also means solar radiation intensity was high. Pressure in layman's terms is force exerted by the air above an area, divided by that area.

7. Relative Humidity- The amount of water vapors present in air expressed as a percentage of the amount needed for saturation at the same temperature. This term is related to both wind and solar energy. Because if the solar radiation intensity is more, water from earth's surface will be evaporated more and hence, water in the atmosphere will be more. Thus, more relative humidity means more solar radiation intensity. Also, if relative humidity will be more, the air will become heavier and hence wind speed will decrease. Thus, wind energy output will decrease.

8. Wind Speed- It is just the speed of wind at any point at any time. It is clearly related to wind energy. Higher wind speed, higher wind energy output. Wind speed also depends on pressure gradient, which in turn is related to solar radiation intensity at two different places. Hence in that sense, it can also be related to solar energy.

3. K-S TEST FOR NORMALITY

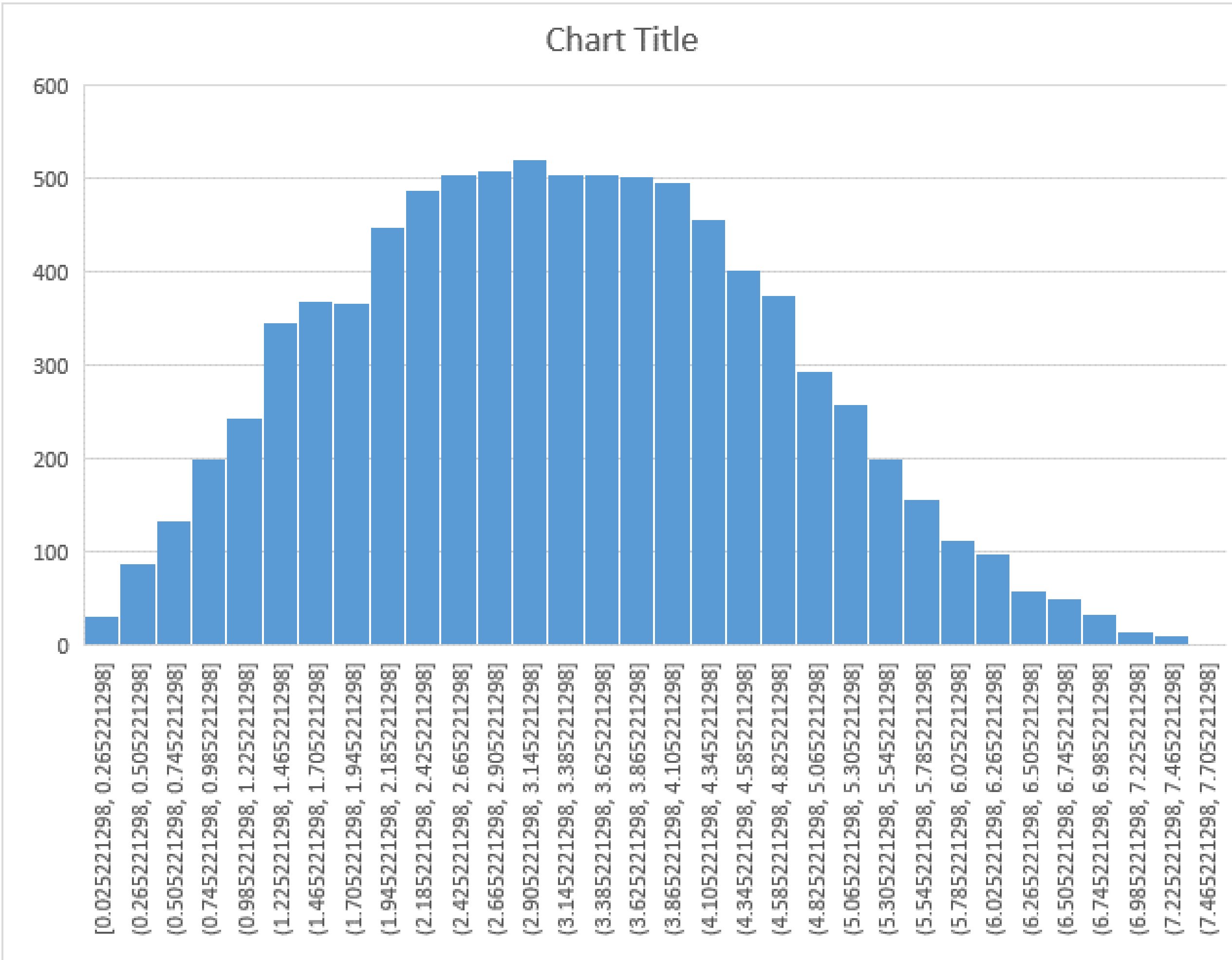


Fig1. Histogram of wind speed data (looks a bit like normal distribution)

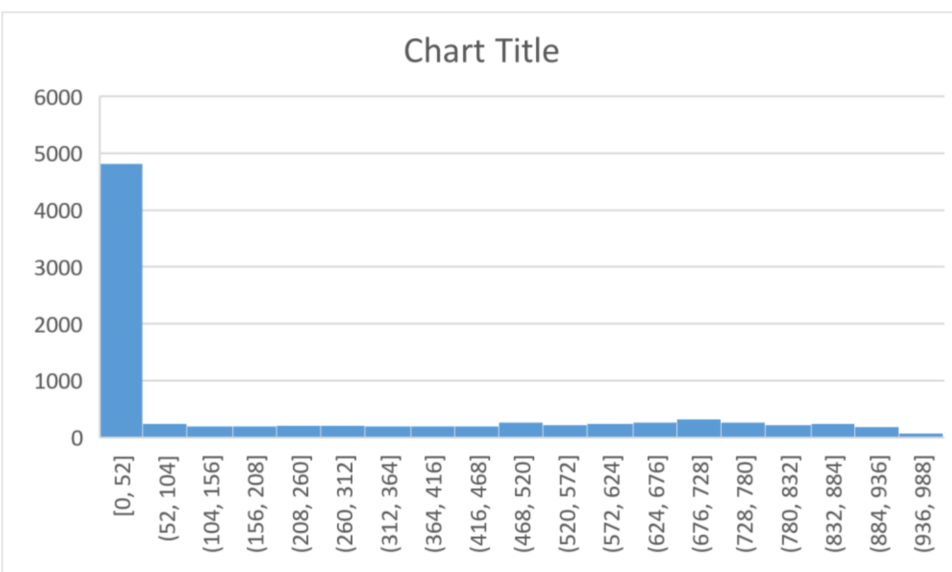


Fig2.

Histogram plot of GHI for the year 2000(does not look like normal distribution)

As can be seen from the plots, the wind speed data look a lot like a normal distribution but the GHI data does not look like a normal distribution. Let us prove it mathematically.

For, finding out whether the dataset follows normal distribution or not, we are going to use K-S test for normality. This test was performed using excel.

Procedure followed:

1) The cumulative probabilities of values in the data are compared with cumulative probabilities in a theoretical normal distribution.

2) The critical value is found from the K-S table values. If test statistic value is less than critical value, then we accept the null hypothesis.

3) $D = \max_{1 \leq Y \leq N} \left[F(Yi) - \frac{i-1}{N} , \frac{i}{N} - F(Yi) \right]$ (Formula for the K-S test statistic)

$F(Y_i)$ = input numbers in the ascending order

i= count from 1 to N

Various columns are used to calculate this final value.

4) The seventh value in the graphs below gives the value in the square braces above. Its maximum is found and that is the test statistic.

5) This test statistic is then compared with the critical value.

H_0: It follows a normal distribution. (If test statistic is smaller than critical value)

H_1: It does not follow a normal distribution. (If test statistic is less than critical value).

The critical value for sample sizes more than n=50 is given by $1.36/n^{0.5}$. (For a level of significance 0.05)

Critical value = 0.014531

The test statistic value comes out to be 0.027018 for the wind speed data

The test statistic value comes out to be 0.285966 for the GHI data

As, the test statistic is more than the critical value for both the cases, the null hypothesis can be rejected and hence, it is concluded that wind speed data and the GHI data for the year 2000 do not follow normal distribution.

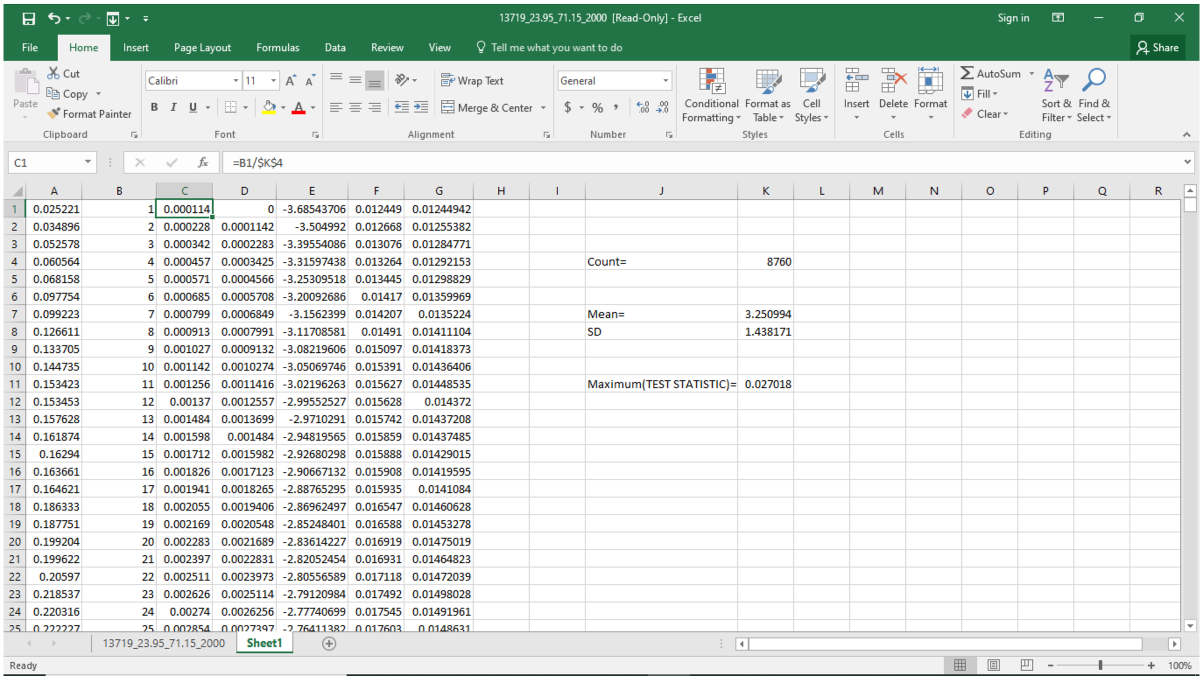


Fig. K-S test being done in excel for the wind speed data set

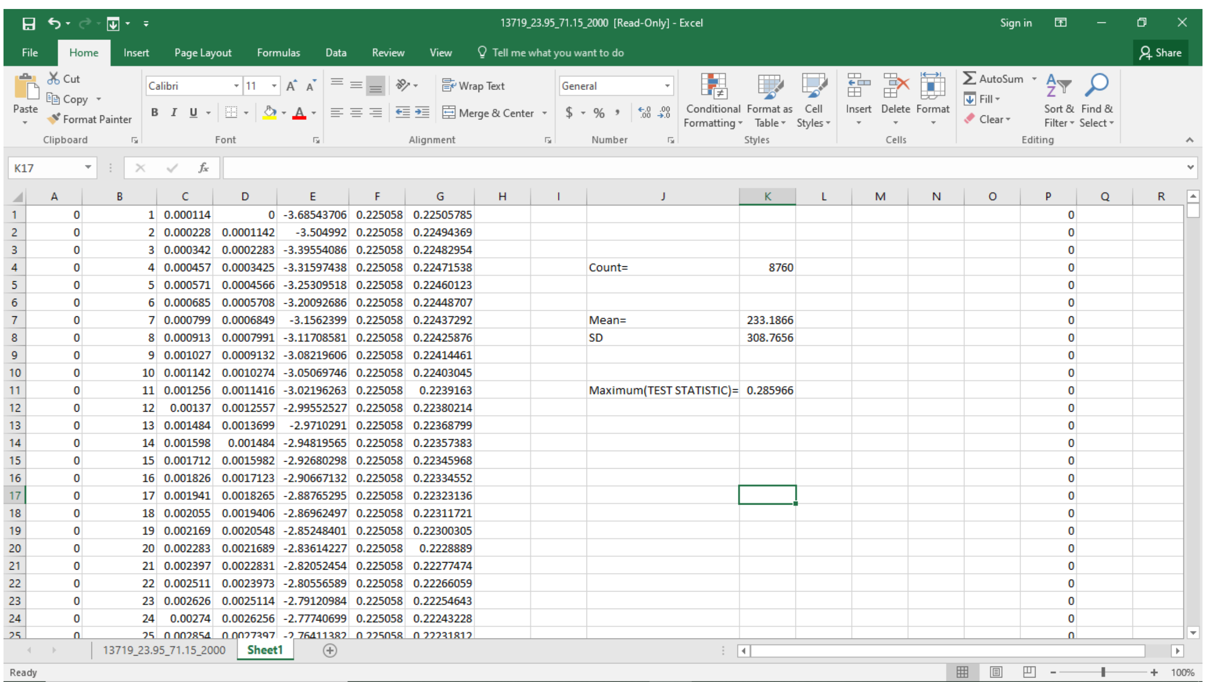


Fig. K-S test being done on the GHI data set using excel

<u>For the GHI Data: -</u>

This data set cannot be lognormal, or gamma or box-cox transformation or Weibull because this data set contains non-negative values (0).

For the rest of the distributions, Minitab shows the following p-values for each of the distributions: -

Goodness of Fit Test

Distribution	AD	P
Normal	962.853	<0.005
3-Parameter Lognormal	1470.691	*
2-Parameter Exponential	16982.567	<0.010
3-Parameter Weibull	1446.580	<0.005
Smallest Extreme Value	906.184	<0.010
Largest Extreme Value	1084.608	<0.010
3-Parameter Gamma	1915.758	*
Logistic	905.314	<0.005
3-Parameter Loglogistic	1305.133	*

↓	C1	C2	C3	C4
	0			
1	0			
2	0			
3	0			
4	0			
5	0			
6	0			
7	0			
8	142			
9	334			

Sheet1

None of the p-values are greater than 0.05. Hence, the GHI data does not follow any of the distributions given above.

For the Wind speed data: -

The following result was obtained when goodness of fit test was done on Minitab using the 14 distributions available in Minitab:

As can be seen from below, none of the distribution gives a p-value more than 0.05. Hence, this wind speed data does not follow any of the following distributions as well.

Distribution Identification for C1

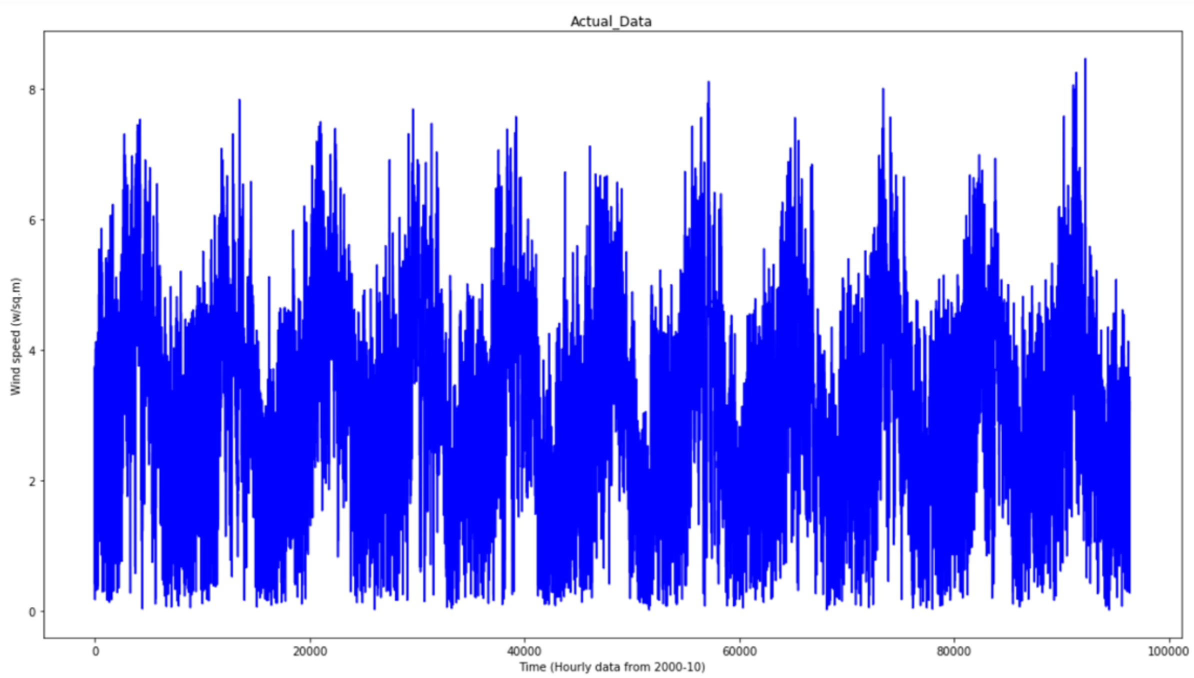
Goodness of Fit Test

Distribution	AD	P	LRT P
Normal	13.625	<0.005	
Box-Cox Transformation	9.756	<0.005	
Lognormal	171.753	<0.005	
3-Parameter Lognormal	11.892	*	0.000
Exponential	1148.221	<0.003	
2-Parameter Exponential	1130.500	<0.010	0.000
Weibull	10.366	<0.010	
3-Parameter Weibull	5.445	<0.005	0.000
Smallest Extreme Value	110.046	<0.010	
Largest Extreme Value	53.899	<0.010	
Gamma	67.434	<0.005	
3-Parameter Gamma	12.009	*	0.000
Logistic	26.372	<0.005	
Loglogistic	101.168	<0.005	
3-Parameter Loglogistic	25.798	*	0.000

	C1	C2	C3	C4	C5	C6
1	3.73680					
2	3.70279					

4. TIME SERIES PLOT FOR WIND SPEED

Now, we will be working on the wind speed data from 2000-2010. Firstly we will consider it as time-series data. The plot of the time series data against time is as follows:



From the above plot, we can observe that there is seasonality in this data. This can be understood as there are certain points of time every year where we get a similar range of values for the wind speed.

Every year, a pattern is being repeated. If we observe keenly, we notice that a slightly decreasing trend is present during the middle of the period. This means that the pattern which is more or less repeating every year has gone a bit down compared to the previous year.
Now, as we have observed seasonality component in the data, we will try to decompose the data into trend, seasonality and residual components. The main purpose of this decomposition is to get a stringent analysis of each component separately so that we can get the forecast as accurate as possible.

After decomposition, we have plotted the trend, seasonal and residual components with time (in hours). Which looks as shown below:

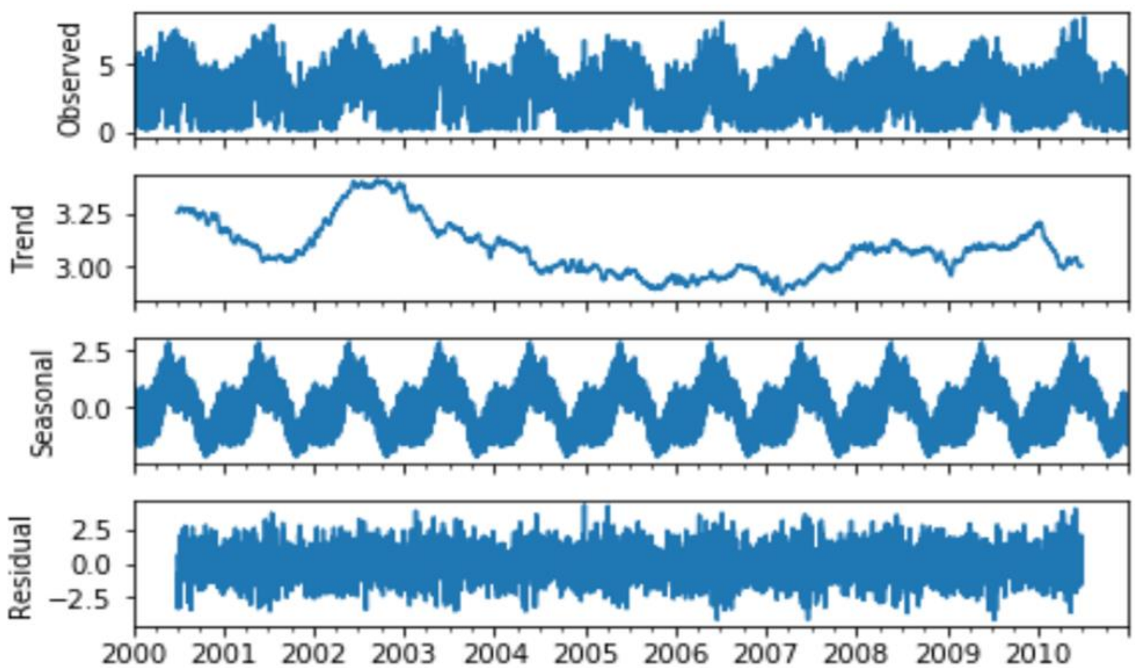


Fig. showing the decomposition of time series into trend, seasonal and residual Components when the Additive model is used

As we know, the decomposition of a time series can be done by using a multiplicative model and an additive model. An additive model is a good tool to do the decomposition if the magnitude of the seasonal fluctuation does not vary much with the level of the series. This is the case with our wind speed data. So, we went ahead to decompose the series using the additive model.

Additive Model can be represented as follows,

$$\mathbf{Y_t = S_t + T_t + E_t}$$

However, in order to compare the results, we also did the decomposition using the multiplicative model.
Multiplicative Model can be represented as follows,

$$\mathbf{Y_t = S_t \times T_t \times E_t}$$

As our data does not follow any specific trend throughout the period, we should use some smoothing method such as moving average or exponential smoothing.

For our time series decomposition, we have taken the help of the statsmodels library in python. From, statsmodels.tsa.seasonal we imported seasonal_decompose() function which uses moving average approach for the smoothing purpose and further decomposes the series into different components. For reference, we have provided the python script used. In the appendix.

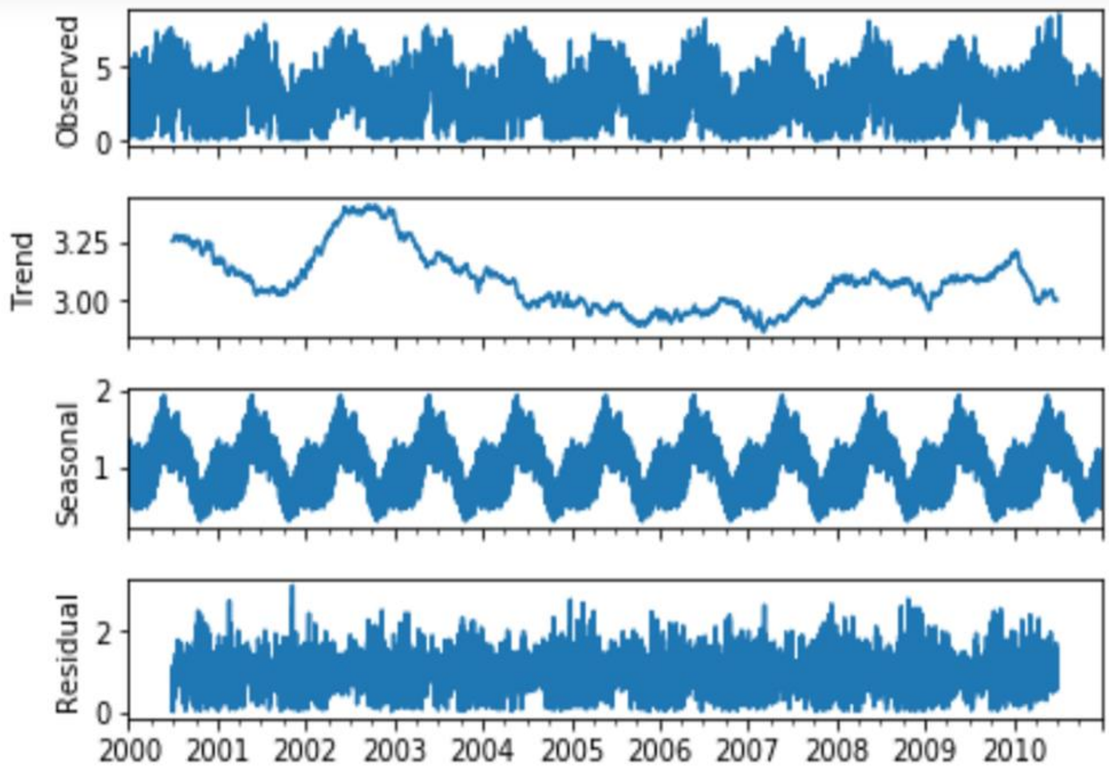


Fig. showing the decomposition of time series into trend, seasonal and residual Components when the Multiplicative model is used.

5. FORECASTING OF WIND SPEED

5.1 ARIMA Forecasting

ARIMA is an acronym that stands for Auto Regressive Integrated Moving Average. It is most extensive forecasting method for univariate time series data forecasting. It does not support time series with a seasonal component.

ARIMA expects data to be non-seasonal or has seasonality component removed.

Some of parameters of ARIMA model are:

- p: trend autoregression order
- d: trend difference order
- q: trend moving average order

To overcome limitations of ARIMA, SARIMA is used.

5.2 SARIMA Forecasting

SARIMA acronym for Seasonal Auto Regressive Integrated Moving Average. This method is used to forecast from data containing trend and seasonality. There are four seasonal element that are not part of ARIMA and is a part of SARIMA:

- P: Seasonal Autoregressive order
- D: Seasonal Difference order
- Q: Seasonal Moving Average order
- m: number of times steps for a single seasonal period

5.3 Prediction

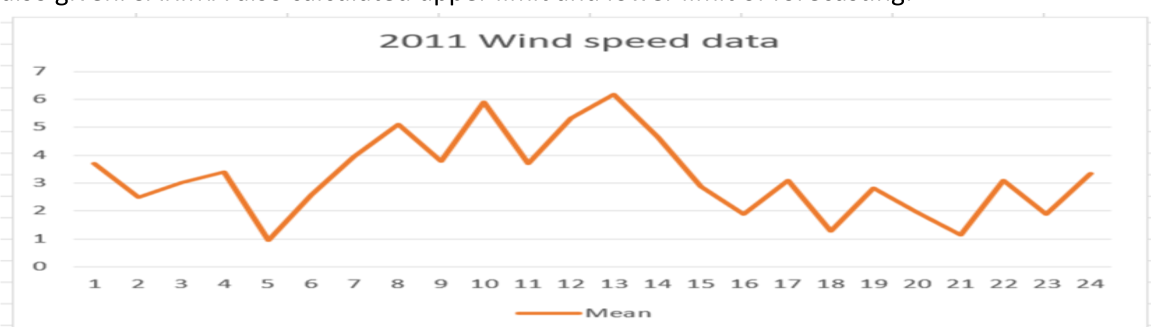
We used NumXL for forecasting. At first, we used all data of all months, weeks, days, hours. We used Weekly dataset for forecasting since it followed yearly seasonal period. The thing we used for forecasting is SARIMA in NumXL. Since our data consisted seasonal pattern so we used SARIMA.
The results of SARIMA are shown below. SARIMA takes care of trend as well as seasonality, that's why this is a good method. As you can see the results calculated are pretty close to target values.

SARIMA(0,1,1)(0,1,1)24	
Param	Value
μ	0.00
θ_1	0.00
Θ_1	0.00
σ	0.25
d	1
s	24
D	1

Goodness-of-fit		
LLF	AIC	CHECK
-223.49	454.97	1

Residuals (standardized) Analysis							
	AVG	STDEV	SKEW	KURTOSIS	Noise?	Normal?	ARCH?
	0.00	1.00	0.07	1.36	FALSE	FALSE	TRUE
Target	0.00	1.00	0.00	0.00			
SIG?	FALSE	FALSE	TRUE	TRUE			

Graph of predicted values are also plotted. This is prediction for a YEAR 2011. 1st prediction stands for half a month, 2nd prediction stands for a month, and so on. Standard deviation in predicted values are also given. SARIMA also calculated upper limit and lower limit of forecasting.



Step	Mean	STD	UL	LL
1	3.69889	2.48339	8.56624	-1.16846
2	2.48087	3.51204	9.36434	-4.40261
3	3.01006	4.30135	11.4406	-5.42044
4	3.39055	4.96677	13.1253	-6.34415
5	0.94075	5.55302	11.8245	-9.94298
6	2.56663	6.08303	14.4892	-9.35589
7	3.94924	6.57043	16.827	-8.92856
8	5.10842	7.02408	18.8754	-8.65853
9	3.79201	7.45016	18.3941	-10.81
10	5.90389	7.85316	21.2958	-9.48802
11	3.70771	8.23646	19.8509	-12.4355
12	5.32079	8.60271	22.1818	-11.5402
13	6.17283	8.95398	23.7223	-11.3766
14	4.63253	9.29198	22.8445	-13.5794
15	2.88667	9.61812	21.7378	-15.9645
16	1.8972	9.93355	21.3666	-17.5722
17	3.08611	10.2393	23.1547	-16.9825
18	1.282	10.5361	21.9324	-19.3684
19	2.82448	10.8248	24.0408	-18.3918
20	1.96873	11.106	23.7362	-19.7987
21	1.15983	11.3803	23.4648	-21.1452
22	3.10706	11.6481	25.9369	-19.7228
23	1.90156	11.9099	25.2445	-21.4414
24	3.35204	12.1661	27.1971	-20.493

6. APPENDIX

```
from pandas import read_csv
import pandas as pd
from matplotlib import pyplot

from statsmodels.tsa.seasonal import seasonal_decompose
import numpy as np

time = pd.date_range('1/1/2000', periods=96360, freq='60min')

series = read_csv('graph.csv',header=None)
series.index=time

series = pd.DataFrame(series)
# series.plot()
# pyplot.show()

#print(series)

result = seasonal_decompose(series, model='additive',freq=365*24)

#result = seasonal_decompose(series, model='multiplicative',freq=365*24)

result.plot()
pyplot.show()
```