A

Project Report

on

WIND POWER SCENARIO GENERATION USING ARIMA MODEL

Submitted in partial fulfilment of the requirements
of the degree of
Bachelor of Technology
In
Electrical and Electronics Engineering
by

Aman Jain (171230010)

Ambuj Kumar (171230011)

Aniket Pathak (171230012)

Mallika Goel (171230028)

Supervisor:
Dr. Manoj Kumawat
Assistant Professor
Dept. of Electrical & Electronics Engg.

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY DELHI

(Year: 2019-2020)

National Institute of Technology
Delhi

CERTIFICATE

This is to certify that Minor Project Report entitled **WIND POWER SCENARIO GENERATION USING ARIMA MODEL** which is submitted by **Aman Jain** (Roll No. 171230010)**, Ambuj Kumar** (Roll No. 171230011)**, Aniket Pathak** (Roll No. 171230012) **and Mallika Goel** (Roll No. 171230028) in partial fulfilment of the requirements for the award of degree B.Tech. in Department of Electrical and Electronics Engineering at **National Institute of Technology Delhi**, is a record of the candidate's own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree. I hereby declare that this submission was their own work and that, to the best of their knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma from a university or other institute of higher learning, except where due acknowledgment has been made in the text.

Date:

Dr. Manoj Kumawat
Dept. of Electrical & Electronics Engg.
National Institute of Technology
Delhi - 110040

ACKNOWLEDGEMENT

We are very thankful to our project guide Dr. Manoj Kumawat, Assistant Professor, Department of Electrical and Electronics Engineering, National Institute of Technology, Delhi, to introduce us to this topic and for their inspiring guidance, constructive criticism and valuable suggestions throughout the project work.

We would also like to thank Dr. Anmol Ratna Saxena, Assistant Professor and Head of Department, Electrical and Electronics Engineering, National Institute of Technology, Delhi, for providing necessary information and guidance for the completion of this project work.

Last but not the least, we would like to thank all our friends and families for supporting and motivating us to pursuing our interest throughout this project.

Aman Jain (171230010)

Ambuj Kumar (171230011)

Aniket Pathak (171230012)

Mallika Goel (171230028)

Dept. of Electrical and Electronics Engineering
National Institute of Technology
Delhi - 110040

ABSTRACT

The role of wind power is increasing to meet the electricity demand due its environmentally friendly nature. The possibilities with wind as a source of energy are beyond our imagination. Adapting wind as a prime source of energy can help us in solving major global issues like Greenhouse gas emissions, Global warming and ultimately, Climate change. However, wind power exhibits large uncertainty and to a large extent influenced by the geographical and meteorological conditions. Therefore, Scenario generation is a very crucial process in planning as well as operation of power systems with high renewable penetrations. In this process, an approach which is data driven is proposed for scenario generation using Autoregressive Integrated Moving Average model, which is a class of statistical models for analysing and forecasting time series data. For validation, the dataset of "**Global Energy Forecasting Competition 2012 - Wind Forecasting"** time-series data from **Kaggle** is used. It is shown that the proposed method did generate realistic wind power profiles with various behaviours. The process of generating scenarios based on different time ahead forecasting is also illustrated. As this model is statistical in nature, we can generate scenarios efficiently without involving a lot of complex calculations.

TABLE OF CONTENTS

List of Tables

List of Figures

1. INTRODUCTION

High levels of electricity generated from renewable resources have challenges in the scheduling, operation, and planning of power systems. Since renewables are not continuous, steady and have random probability in distribution, accurate modelling the uncertainties in them, is the solution. Time-series scenarios can be an approach to know about the uncertainties in renewable resources. By the use of a set of possible power generation scenarios, renewables producers and system operators are able to make decisions taking uncertainties into consideration, such as stochastic economic dispatch/unit commitment, optimum operation of wind and storage systems, and trading strategies.

Due to the environment friendliness and renewable nature of wind power, it has got the attention of the world. Hence the wind industry is growing into a large-scale business. Therefore, short term reliable wind speed forecasts play an important role in wind energy conversion systems, such as the dynamic control of wind turbines and scheduling of power systems. A precise forecast needs to overcome problems of variable energy production caused by fluctuating weather conditions. Power generated by wind is highly dependent on the wind speed. Though it is highly non-linear, wind speed follows a certain pattern over a certain period of time. This time series pattern is exploited to gain useful information and use it for power prediction [1].

Now, it is known that wind energy prediction is a very interesting problem, let's look into why it is a difficult one. Accurate and reliable wind speed forecasts are a significant challenge due to its random nature with high rates of change, highly nonlinear behaviour with no typical patterns, and dependency on elevation, terrain, atmospheric pressure, and temperature, resulting in large uncertainties of wind speeds. This makes it difficult for any machine learning model to figure out a pattern and give an accurate prediction. In this project efforts are made to interpret this problem as time series forecasting problem as wind follows a particular pattern for a certain period like a day, month or year. Autoregressive Integrated Moving Average (ARIMA) model, which is best known for time series data prediction is used to learn these patterns in wind and make a prediction about power [2]. This prediction problem is divided into two categories:

1. **Estimation**: Given weather conditions like temperature, wind speed, pressure etc. determining the energy power prediction.
2. **Prediction**: Without knowing any details about the weather conditions predicting the power generation using the pattern which it has followed in a certain period of time.

2. LITERATURE SURVEY

Research objectives are prepared after correlating the various works done by previous contemporary researchers. Most of the researchers have developed methods for wind speed-based forecasting. Apart from wind speed forecasting, many other parameters required to assess the wind energy potential are studied. Meteorological, climatological information coupled with topographical data have to be utilized for wind power estimation at a particular place. After this, wind turbine power curves have to be mapped against the wind parameters. So that the number of energy units that can be generated at wind plant level, for a month. In most of the wind plants in India, wind speed measurements taken from a single location for the whole plant is an average indicator of wind power potential at plant level. Deviations from mean position of the wind measurement to the wind turbine are neglected. Hence, Average wind speed indicator continues to contribute for wind 38 power estimation at plant level. Wind speed and wind power generation along with changing values across time series in the current focus of work [14].

Scenario-generation methods can be classified into three categories: path-based methods, matching the movement, and internal sampling. In the first one, scenarios are generated by time-series models. Second one generates a discrete distribution of statistically dependent random variables. While the third one is a process of continuous sampling of original distribution of random variables.

Monte Carlo simulation with the path-based scenario-generation method was used for generation of electric load scenarios in power management problems [17]. This uses two heuristic scenario-reduction algorithms: fast-forward selection and simultaneous backward reduction. For price and load scenarios forward and backward scenario-reduction algorithms have been suggested [18]. The wind speed scenario-generation method may use statistical time-series ARMA models for scenario generation of different sites. This particular approach shows the relation of generated scenarios by autocorrelation plots between every two sites. Thermal ratings of transmission lines have been forecasted by the time-series ARIMA model, considering the effect of weather unpredictability [19]. Price scenarios of the Nordic market are generated by an ARMA model to restore the bidding strategy problem using a stochastic programming approach [20]. Further, the recursive backward scenario-reduction technique has also been used for wind scenario reduction [21].

Wind power forecasts associated with wind resource assessment methodologies, data and experience through modelling will help the wind power plants to meet the changing grid requirements. Forecasting serves two purposes – (1) allows wind generators to take operational decisions which are aligned to market prices of energy, (2) allows the system operator to better plan ancillary services required to balance generation from variable renewable energy sources.

Literature review of this thesis work aims at finding out publications related to ANNs proposed as wind resource assessment and forecasting techniques.

ANNs have successfully been applied in various engineering areas such as – function approximation, pattern association and associative memory. These are suitable for cases such as – incomplete data sets, data with unknown relationship, imprecise problems. Additionally, they show fault tolerance and robustness. Function approximation means mapping of multiple inputs to a single input. Statistical techniques solve these problems by estimating parameter values. When ANNs are applied as function approximation solvers, output function parameters are estimated based on a model. Such models relate inputs vs. outputs through the activation functions. Various activation functions will be selected and tried by the programmer or designer of ANN. These activation functions do change with respect to the relationship between inputs and outputs and are selected on trial and error basis [15].

Wind power estimation & forecasting models are majorly classified as models based on physical or numerical weather prediction models, statistical models and hybrid models. Physical models are based on spatio-temporal meteorological data and open boundary fluid flow equations. Navier-Stokes and other energy equations are solved as nonlinear solvers in predicting wind energy potential in selected space. Wake flow analysis can be carried out in this for different terrains and topographies. A wind turbine can also be placed on rough terrain inside the simulation environment and wind flow studies can be performed. Major drawback of these models is that they don't assume external governing factors such as electrical parameters, mechanical conversion efficiency and of course the historical 39 time series wind pattern data. Statistical models are based on probabilistic approaches such as Weibull distribution etc. in order to estimate the values of wind speed, potential based on approximations and time series analysis. These are obtained by extrapolations of the previous data and do not consider the flow dynamics, external influencing parameters. Vertical wind profile modelling is based on probabilistic empirical models derived by various researchers.

Cadenas and Rivera [3] proposed a hybrid neural network-based model that uses ARIMA and ANN for simulation. A month data of wind speed time series is utilized to run the neural networks based on ARIMA and ANN. Statistical errors were calculated which showed that model is able to predict the wind speed of new site accurately [16].

ARIMA model is chosen in this project, because of its simplicity and wide acceptability of the model. The effect on prediction accuracy is also studied based on various possible previous period data taken.

3. ARIMA MODEL

An ARIMA model is a part of statistical models used for analysing and forecasting time series data. It explicitly uses standard structures in time series data, and as such provides a basic yet powerful method for making adept time series forecasts [4]. Basically, ARIMA stands for Auto Regressive Integrated Moving Average. It is a generalization of the simpler Auto Regressive Moving Average or ARMA, which has an integration component. Briefly, they are:

- **AR**: Autoregression. A model that uses the dependent relationship between an observation and number of lagged observations.

  A pure **Auto Regressive (AR only) model** is one where Y_t depends only on its own lags. That is, Y_t is a function of the 'lags of Y_t'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_1 \quad (1)$$

  where, $\mathbf{Y}_{t-1}$ is the lag1 of the series, β_1 is the coefficient of lag1 that the model estimates and α is the intercept term, also estimated by the model.

- **I:** Integrated. The use of differencing of raw observations in order to make the time series stationary. It involves subtracting an observation from an observation at the previous time step.
- **MA**: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

  A pure **Moving Average (MA only) model** is one where Y_t depends only on the lagged forecast errors.

$$Y_t = \alpha + \epsilon_t + \varphi_1 \epsilon_{t-1} + \varphi_2 \epsilon_{t-2} + \cdots + \varphi_q \epsilon_{t-q} \quad (2)$$

  where the error terms are the errors of the autoregressive models of the respective lags. The errors Et and E(t-1) are the errors from the following equations:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_0 Y_0 + \epsilon_t \quad (3)$$

$$Y_{t-1} = \alpha + \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \cdots + \beta_0 + \epsilon_{t-1} \quad (4)$$

An **ARIMA** model is one where the time series is differenced at least once to make it stationary and you combine the AR and the MA terms. So, the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \epsilon_t + \varphi_1 \epsilon_{t-1} + \varphi_2 \epsilon_{t-2} + \cdots + \varphi_q \epsilon_{t-q} \quad (5)$$

ARIMA model in words can be expressed as:

$$\begin{aligned} &Predicted\ Yt = Constant + Linear\ combination\ Lags\ of\ Y\ (upto\ p\ lags) \\ &\qquad + Linear\ Combination\ of\ Lagged\ forecast\ errors\ (upto\ q\ lags) \end{aligned} \quad (6)$$

Each of these components are explicitly specified in the model. A slandered notation is used of ARIMA (p, d, q) where the parameters are substituted with integer values. But it does not support seasonal data. That is a time series with a repeating cycle. ARIMA expects data that is either nonseasonal or has the seasonal component removed, e.g. seasonally adjusted by methods such as seasonal differencing. A linear regression model is made including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary. A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model [5]. Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. It helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model [6].

4. EXPERIMENT

A. Data:

The historical wind speed data used for scenario-generation is obtained from publicly available Kaggle platform. The used dataset of “Global Energy Forecasting Competition 2012 - Wind Forecasting” [7]. Hourly data of seven wind farms are collected from July 01, 2009 to June 26, 2012. The wind speed data is converted into wind power data based on anemometer height, shear coefficient, and air density of each farm. This data is also normalized before making it publicly available.

Dataset has a total of 8 columns and glance of values are kept in Table 1. The entries of first column of dataset (i.e. date) indicated the date and time of data collection. For example, 2009070100 represents year 2009, month 09, date 01, and time 12:00am and 2009070101 represents year 2009, month 09, date 01, and time 01:00 am and so on. Remaining seven columns (i.e., "wp1" to "wp7") gather the normalized wind power measurements for the seven wind farms.

Table 1. Description of dataset

S.No.	date	wp1	wp2	wp3	wp4	wp5	wp6	wp7
0	2009070100	0.045	0.233	0.494	0.105	0.056	0.118	0.051
1	2009070101	0.085	0.249	0.257	0.105	0.066	0.066	0.051
2	2009070102	0.020	0.175	0.178	0.033	0.015	0.026	0.000
3	2009070103	0.060	0.085	0.109	0.022	0.010	0.013	0.000
4	2009070104	0.045	0.032	0.079	0.039	0.010	0.000	0.000

As the ARIMA model is mainly for the forecasting of univariate data, so out of seven, data of only one wind farm (named ‘wp1’) is considered to train and test the forecasting model. Data of rest wind farms are eliminated for further study. Now the dataset looks like as shown in Table 2.

Table 2. Description of dataset taking one windfarm

Dataframe	
S.No.	wp1(Wind farm 1)
1	0.045
2	0.085
3	0.020
4	0.060

As the goal of study to generate a realistic scenario of a wind farm for n-day ahead (where n>=1), hourly data are converted into daily data by taking the mean of 24 hours as one day. Thus, we have a total of 782 values which represent the daily wind power generation of a single wind farm. After taking a sum of 24 hours data, the values of daily data are no longer in the range of zero to one. So normalization is done to better fit the model and make it easy to learn the various learning parameters during training.

Normalisation:

Using **Min-Max Scaler** the data is normalised to get rid of the dependencies of scales.

$$X_{norm}(i) = \frac{\left(x_i - min(x)\right)}{max(x) - min(x)} \qquad (7)$$

After normalization of daily data, it will look as shown in table 3 and all the entries are in range of zero to one.

Table 3. Data after Normalisation

Dataset	
S.No	**wp1**
0	0.002042
1	0.016339
2	0.023192
3	0.208914
4	0.358825

Train-Test Split:

To train and evaluate the purposed model, whole dataset is divided into train and test data. First 700 values are used to train the ARIMA model and rest 82 values are used to test the performance.

$$Train\ Dataset\ Shape : (700,)$$

$$Test\ Dataset\ Shape : (82,)$$

Various statistical parameters of train and test data are mentioned in table 4. By observing Table 4, it is found that all parameters are almost equal for both data (train and test). Mean value of train data is 0.2638 and 0.3666 for test data. Which is not too much close, but models should learn accurate future scenarios which can handle such scenarios. Other parameters like 25% which shows top 25% values are less the 0.100 in training data and 0.1653 for test data, are also neither too close nor too far. Same is happening for other parameters too.

Table 4. Statistical description of Train and Test Data

Parameter	Train Data	Test Data
count	700.000000	82.000000
mean	0.263811	0.366567
std	0.206511	0.231576
min	0.000000	0.002269
25%	0.100157	0.165286
50%	0.208551	0.313598
75%	0.393364	0.535912
max	1.000000	0.907820

Visualisation of Data:

Visualisation is done mainly for looking the pattern and conclude some inference about data. Plot of training and test data are shown in below figure 3 and in figure 4.

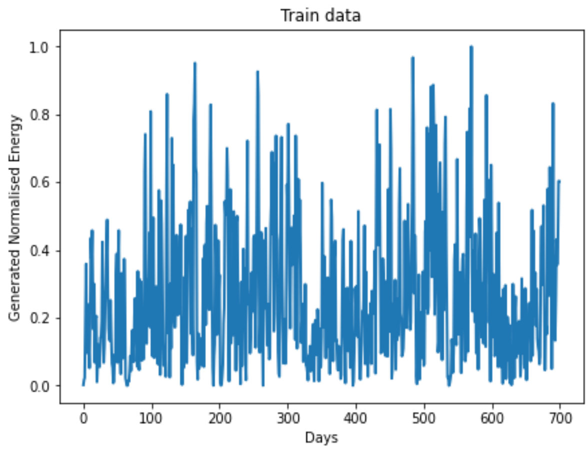

Figure 3. Plot of Train data (Generated Normalised Energy vs No. of Days)

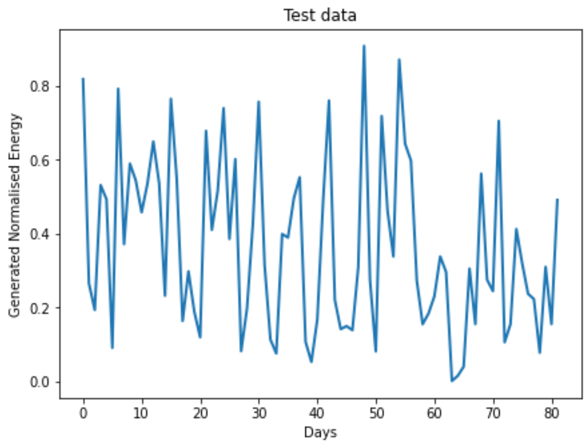

Figure 4. Plot of Test data (Generated Normalised Energy vs No. of Days)

After observing Figure 3 and Figure 4, it can say that there are negligible trends and seasonality in the data.

To compute trends and seasonality, data is decomposed using sklearn library of python.

Checking for Seasonality in Data:

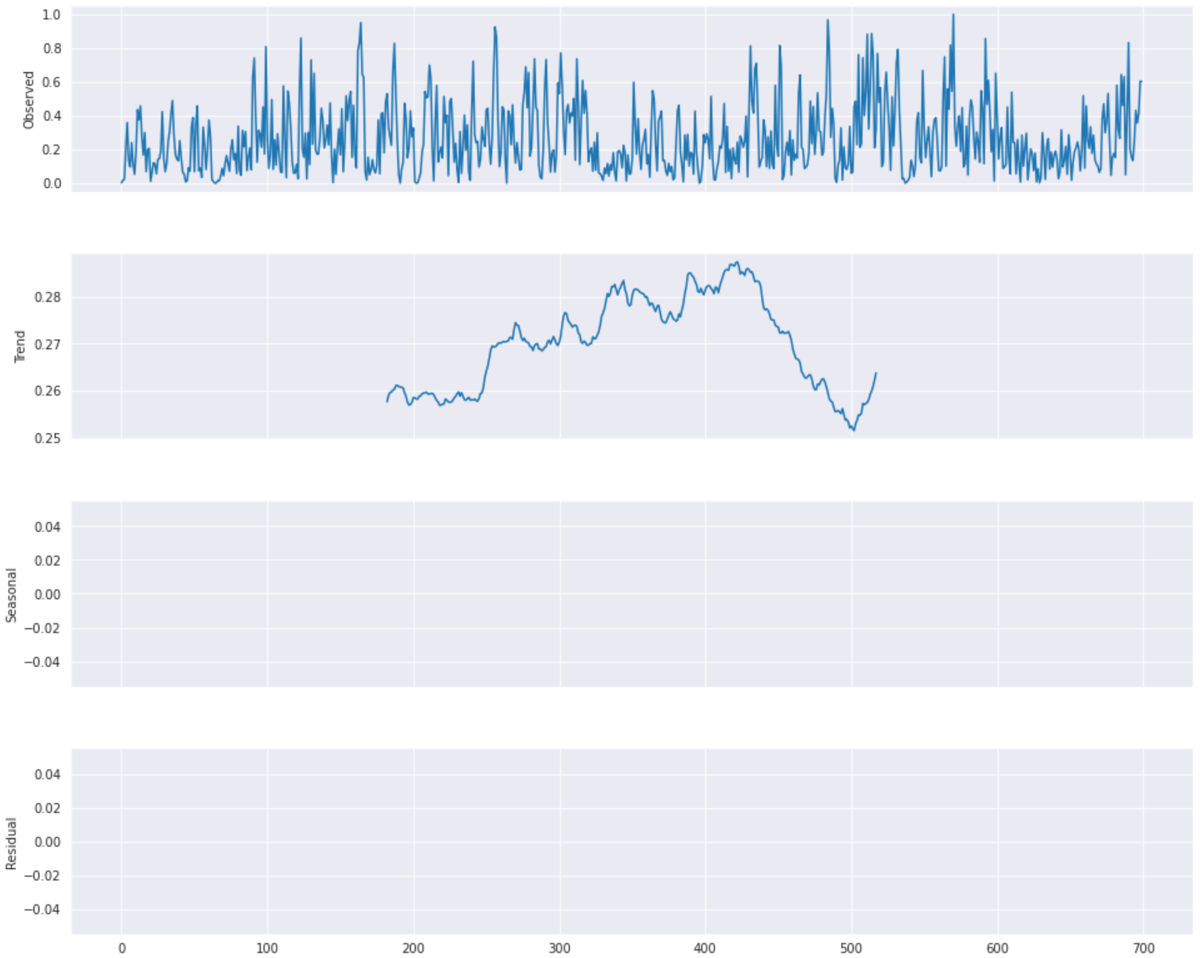

Figure 5. Seasonal Decomposition of Training Dataset

From the above Seasonal Decomposition of Training Data, an inference can be made that the data is having trend but is **not seasonal**.

Checking whether Data is Stationary or not:

The mean of the series should be time independent. The graph on the right below is time stationary because the mean increases over time [8].

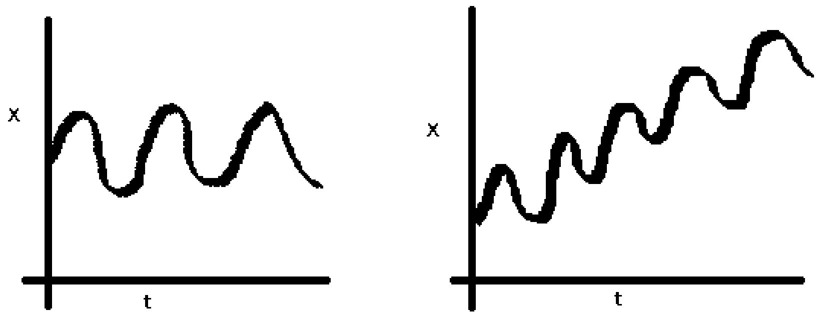

Figure 6. Mean of Stationary Series vs Non-Stationary Series

The variance of the series should not be a function of time. This property is known as homoscedasticity. In the graph on the right the varying spread of data over time.

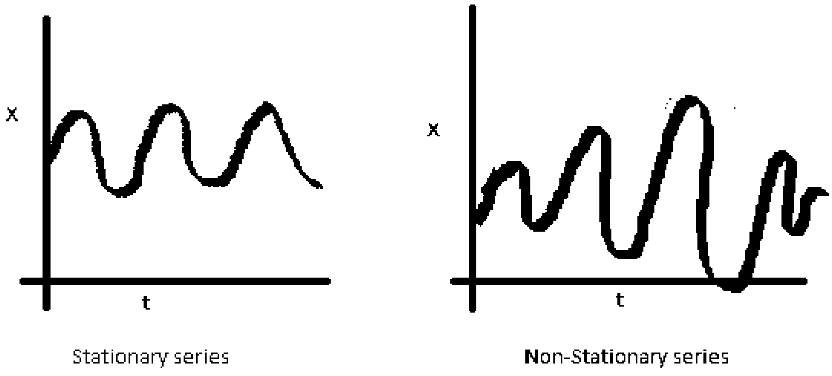

Figure 7. Variance of Stationary Series vs Non-Stationary Series

Finally, the covariance of the i^{th} term and the $(i + k)^{th}$ term should not be a function of time. In the following graph, spread becomes closer as the time increases. Hence, the covariance is not constant with time for the series on the right [9].

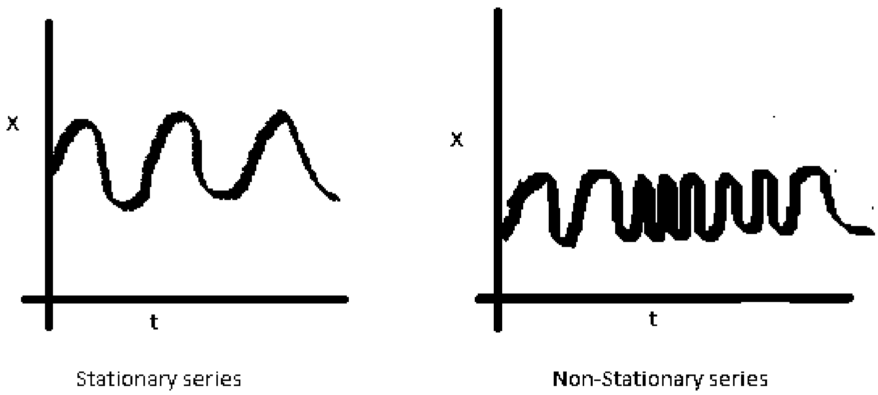

Figure 8. Covariance of Stationary Series vs Non-Stationary Series

<u>Augmented Dickey-Fuller Test for checking Stationarity of series:</u>

Statistical tests take strong assumptions about the data. They can only be used to tell the degree to which a null hypothesis can be rejected or fail to be reject. The result must be interpreted for a given problem to be meaningful.

It uses an autoregressive model and optimizes an information criterion across multiple different lag values.

The null hypothesis of the test is that the time series can be represented by a unit root, that it is not stationary i.e. it has some time-dependent structure. The alternate hypothesis rejecting the null hypothesis, is that the time series is stationary [10].

- **Null Hypothesis (H0)**: If failed to be rejected, it tells the time series has a unit root, meaning it is non-stationary. It definitely has some time dependent structure.

- **Alternate Hypothesis (H1)**: The null hypothesis is rejected; it tells the time series does not have a unit root, meaning it is stationary. It does not have time-dependent structure.

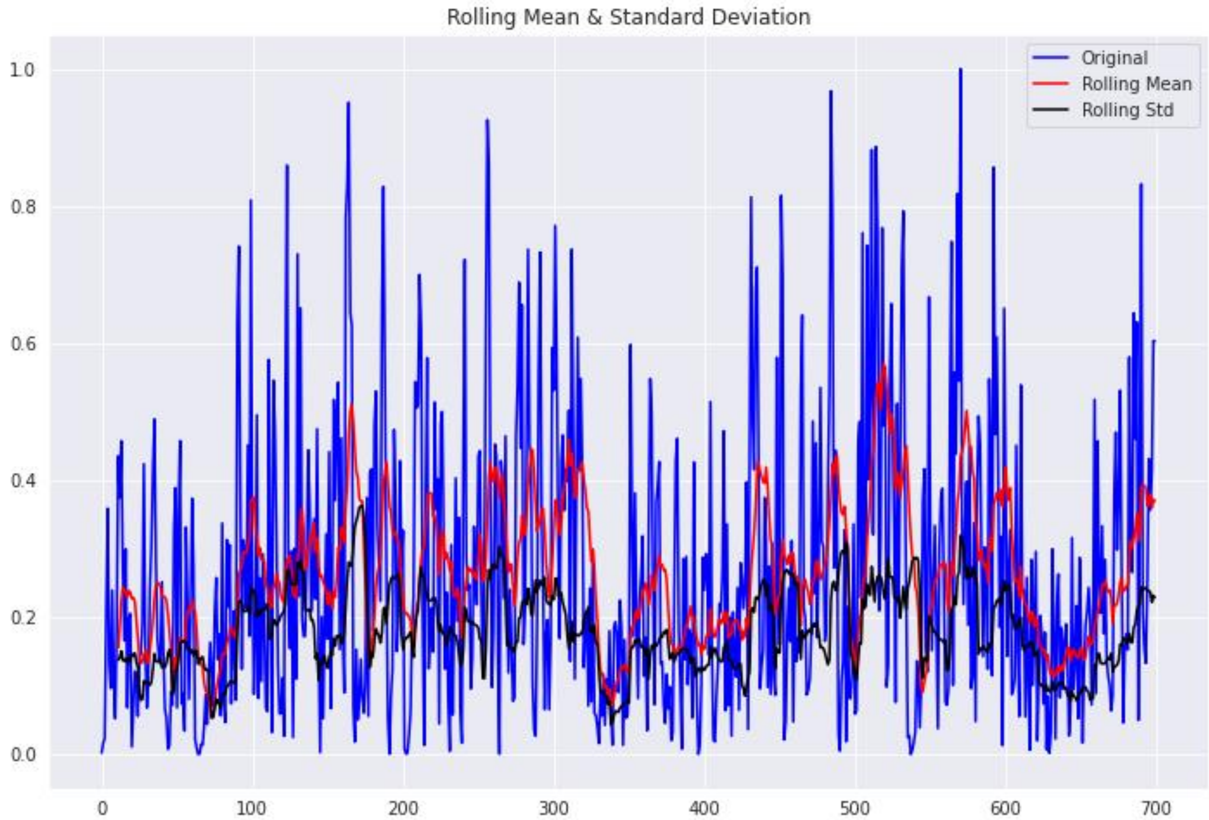

Figure 9. Results of Augmented Dickey Fuller Test

The result is interpreted using the p-value from the test. A p-value below a threshold (such as 5% or 1%) suggests to reject the null hypothesis (stationary), while a p-value above the threshold suggests that rejection of null hypothesis failed (non-stationary).

- **p-value > 0.05**: Fail to reject the null hypothesis (H0), the data has a unit root and is non-stationary.
- **p-value <= 0.05**: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

Table 5. Results of Augmented Dickey Fuller Test

Results of Dickey-Fuller Test	
p-value = 0.0000 (The series is likely stationary)	
Test Statistic	-1.134089e+01
p-value	1.055299e-20
Number of Lags Used	2.000000e+00
Number of Observations Used	6.970000e+02
Critical Value (1%)	-3.439767e+00
Critical Value (5%)	-2.865696e+00
Critical Value (10%)	-2.568983e+00

From the ADF Test we got p-value = 1.055299e-20 ~ 0.000 (< **0.05**).

So, it is concluded that the series is Stationary and no further test is required [11].

B. Methodology:

First step is to apply any statistical model on the dataset, one common practice is to determine the **Lags** value. Lag is a common term used in time-series analysis which signifies the number of steps (apart from the pattern learned) a model will use to predict the next result.

<u>ACF and PCF:</u>

Autocorrelation function of this data, which helps in identify the no of lags for a particular data, is plotted below

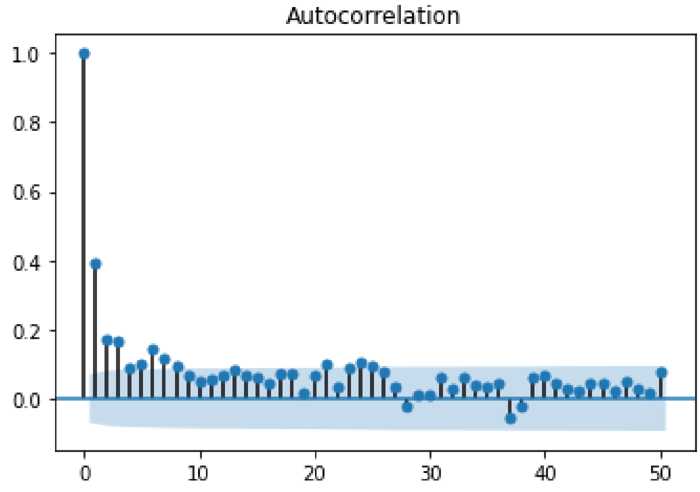

Figure 10. Autocorrelation function (ACF) of data.

It is found that lag values till 27 may have significant change in output results.

To get more familiar with appropriate lags **partial autocorrelation function** is plotted which is shown below

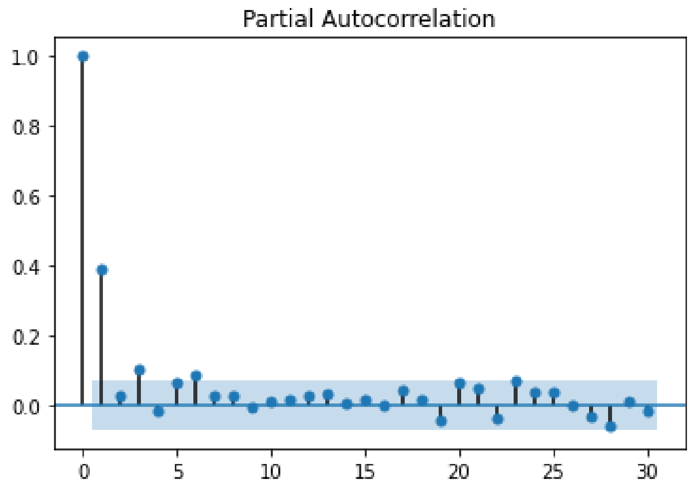

Figure 11. Partial Autocorrelation function (PACF) of data

It was found that the lags values till 6-8 have much influence on output.

Then mean and deviation of two half centred parts is calculated. It is found that mean and deviation of 2nd parts is more. Hence, differencing 1 is also considered.

Table 6. Summary Statistics of Data

Mean1 = 0.253646	**Mean2** = 0.295525
Variance1 = 0.040260	**Variance2** = 0.048196

Building ARIMA Model:

In this case, it is observed that no differentiation is required because the series is already stationary (So, **I = 0**).

Due to a little bit of trends in the data, the results are also investigated with one step differencing (i.e. **I = 1**).

AR model might be investigated first with lag length selected from the PACF or via empirical investigation.

In this case, it is clear that within 6 lags the AR is significant. Which means, $AR = 6$**.**

To avoid the potential for incorrectly specifying the MA order (in the case where the MA is first tried then the MA order is being set to 0), it may often make sense to extend the lag observed from the last significant term in the PACF.

What is interesting is that when the AR model is appropriately specified, the the residuals from this model can be used to directly observe the uncorrelated error. This residual can be used to further investigate alternative MA and ARMA model specifications directly by regression.

Assuming an AR(s) model were computed, then I would suggest that the next step in identification is to estimate an MA model with s-1 lags in the uncorrelated errors derived from the regression. The parsimonious MA specification might be considered and this might be compared with a more parsimonious AR specification. Then ARMA models might also be analysed.

n-day Ahead Forecast:

The aim is to predict the wind energy generated by the system ahead of time. For this it is required to know that if one start from today and predict energy for n days ahead how much accuracy the model gives [12].

For example: Here a forecast of 82 days is being generated, in n days ahead forecasting one is required to do the prediction for the next n days, evaluate results, roll on to the next n days and then repeat the same until the forecast of whole 82 days is obtained. Thus 82 days ahead data was predicted with different look ahead (like a sliding window). When n equals to 82, model forecasts the data in only one look which can be named as one-look forecast (i.e. 82 days ahead forecast here). In this experiment fallowing values for n were chosen and the results were analysed.

1) One-day ahead forecast ($n = 1$)
2) Three-day ahead forecast ($n = 3$)
3) Five-day ahead forecast ($n = 5$)
4) One-week ahead forecast ($n = 7$)
5) One-month ahead forecast ($n = 30$)
6) One-look Forecast or 82 days ahead forecast ($n = 82$)

These values of n are enough to check the performance of ARIMA model in this experiment.

5. RESULTS

Future scenario of 82 days is forecasted using ARIMA model. This model is fitted on 700 days normalised training data. The PACF plot of sample data determined the order of AR terms. According to Figure 11, the order of AR is set to 6. ACF plot is used to determine the order of moving average term which is set to 6 as the Figure 10 shows. To determines the order of differencing, Augmented Dickey Fuller (ADF) test is performed. Differencing term is selected to 0 according to ADF test.

The generated wind power scenarios are shown in Figure 12-17. To test the performance of ARIMA model, forecasted value and actual values are plotted together in Fig 12-17.

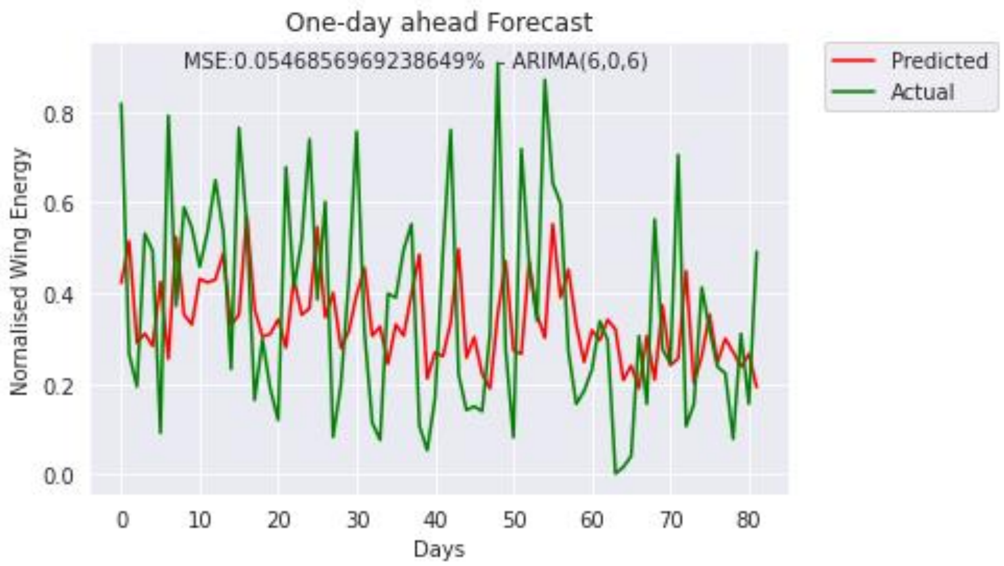

Figure 12. Forecasted vs Actual Plot of One- day ahead forecasting result

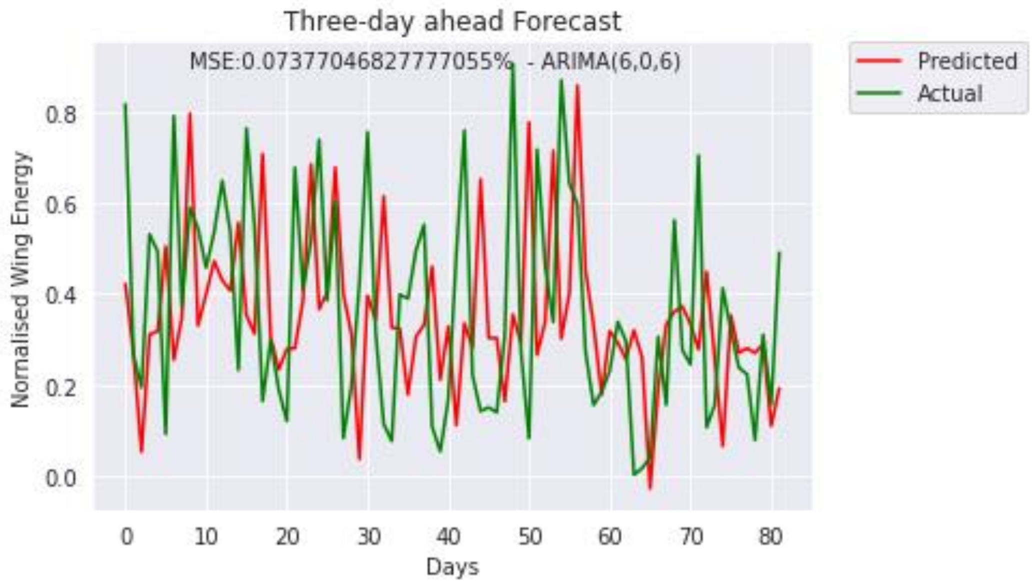

Figure 13. Forecasted vs Actual Plot of Three day ahead forecasting result

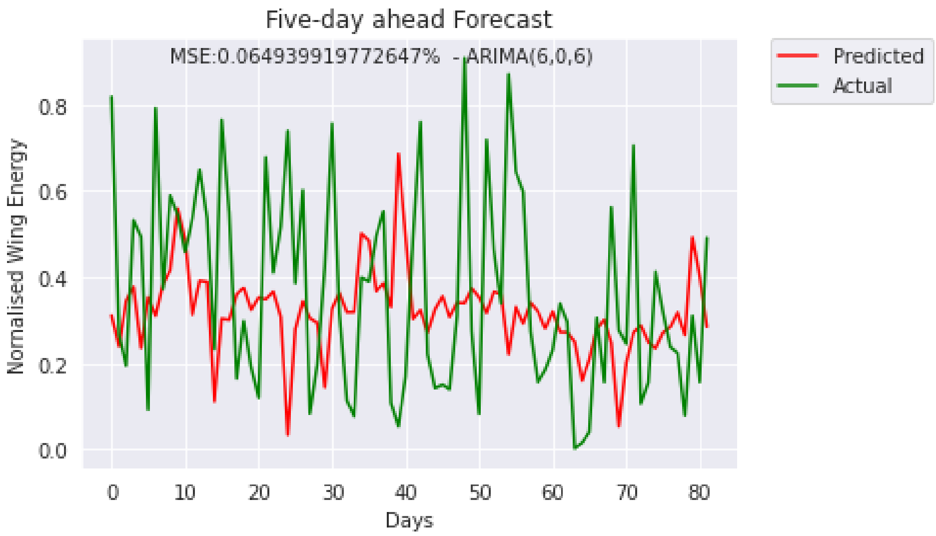

Figure 14. Forecasted vs Actual Plot of Five day ahead forecasting result

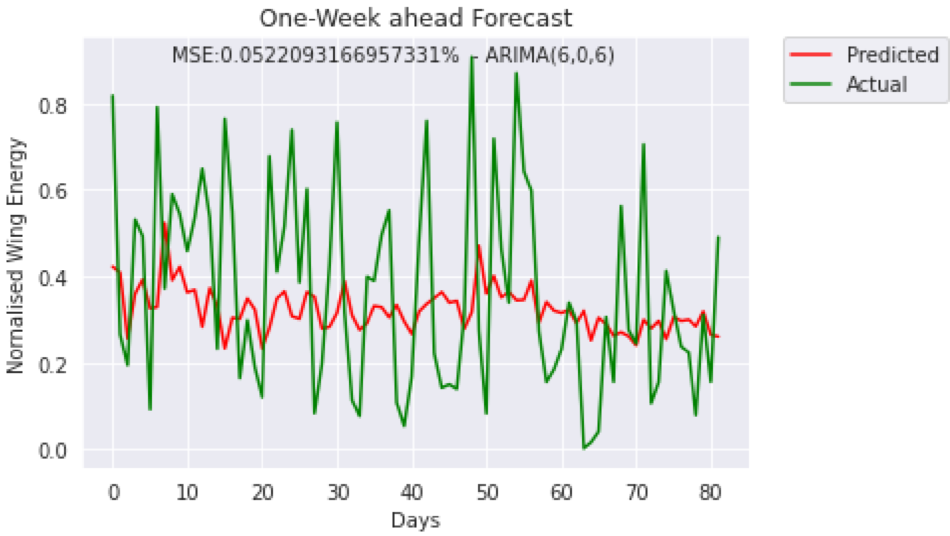

Figure 15. Forecasted vs Actual Plot of One-week ahead forecasting model

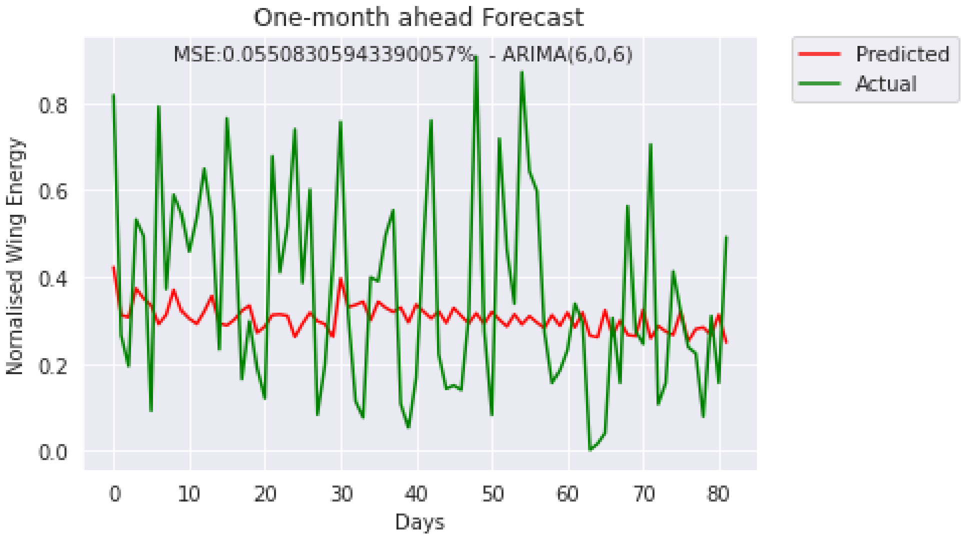

Figure 16. Forecasted vs Actual Plot of One-month ahead forecasting result

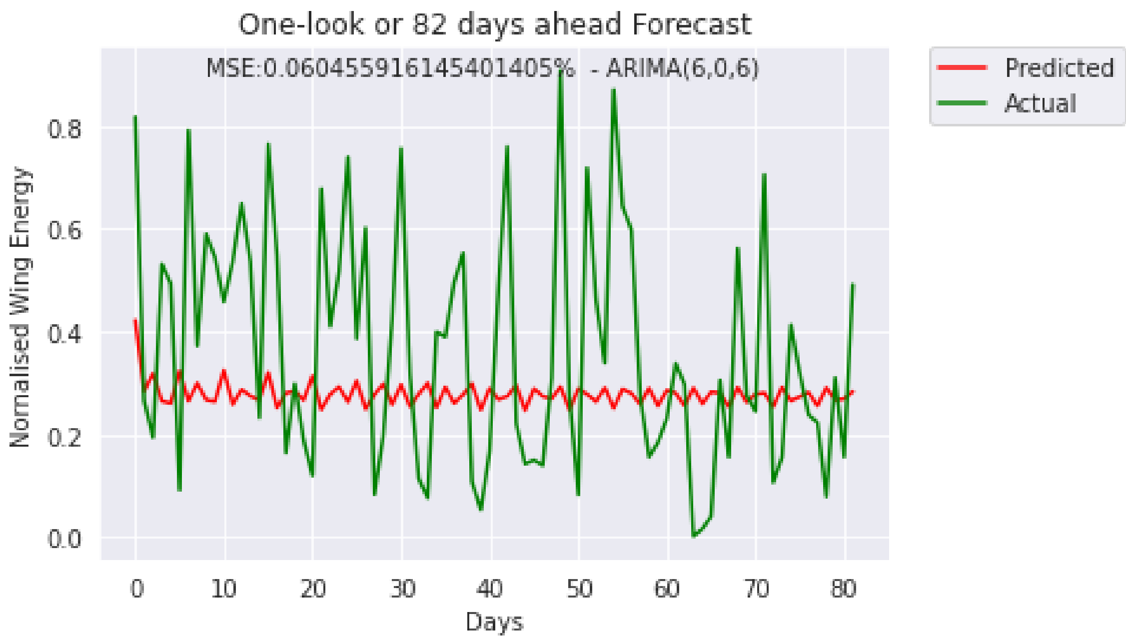

Figure 17. Forecasted vs Actual Plot for One-look or total 82days ahead forecast

After observing the plot of Forecasted and Actual values, it is found that ARIMA model generated future scenarios with acceptable performances. It is also seen, when forecasting is done for short time, it gives more accurate values. for example, plot of one day ahead forecast seems to be more realistic as compared to one-week or one-month.

Mean Squared error is another vary important parameter model to evaluate the performance of model. Mean Squared Error of each type of implementation are described in Table 7. One-day and one week ahead forecast scheme gives lower value of mean squared error. The plot of Three-day ahead forecast seems much better forecast but mean squared error for this is more as compared to others. So, it is required to investigate other performance evaluating parameters. For each type of forecasting schemes, these parameters are described in Table 8. According to Table 8, three-day ahead and one-day ahead forecast give **better mean forecasted value**. Three-day ahead forecast and one-day ahead forecast cover **more standard deviation** hence, close to actual value. Three-day ahead forecast is also close the actual minimum and maximum value so in spite of not good RMSE, three-day ahead forecasting model capture more span/range of actual scenario. Parameters of 25/50/75% shows, 25/50/75% value is less than mentioned value like in actual test data 25% values are less than 0.1653 and 50% value is less than 0.3136 and 75% values are less than 0.5359 along with maximum value 0.9078. In the same way 25% values of one-day forecast are less than 0.2671 and 50% values are less than 0.3233 and 75% values are less than 0.3984 with maximum value of 0.5695.

Value of mean squared error and variance between actual and predicted values are mentioned in Table 7.

Table 7. Performances of each type of model

S. No	Type of Model	Mean Squared Error	Variance
1	**One-day ahead Forecast**	0.0546	-0.032
2	**Three-day ahead Forecast**	0.0737	-0.393
3	**Five-day ahead Forecast**	0.0649	-0.226
4	**One- week ahead Forecast**	0.0522	0.014
5	**One- month ahead Forecast**	0.0550	-0.040
6	**One-look or 82 day ahead Forecast**	0.0604	-0.141

After observing the performance of various types of day ahead forecast using ARIMA model we can say that it can give more realistic future scenario of wind power.

Various statistical parameters for each type forecast are described in Table 8.

Table 8. Statistical parameters for different classes of forecast

S. No	Statistical Parameter	Actual Value	Forecasted Value					
			1-day ahead	3-day ahead	5-day ahead	1-week ahead	1-month ahead	1-look ahead
1	**Mean**	0.3666	0.3381	0.3499	0.3206	0.3237	0.3065	0.2789
2	**Standard deviation**	0.2316	0.0928	0.1666	0.0965	0.0518	0.0312	0.0247
3	**Minimum**	0.0022	0.1897	-0.028	0.0345	0.2330	0.2486	0.2467
4	**25%**	0.1653	0.2671	0.2727	0.2794	0.2901	0.2877	0.2634
5	**50%**	0.3136	0.3233	0.3249	0.3189	0.3186	0.3062	0.2782
6	**75%**	0.5359	0.3984	0.4006	0.3599	0.3505	0.3212	0.2908
7	**Maximum**	0.9078	0.5695	0.8591	0.6861	0.5237	0.4225	0.4224

One another important observation, as the value of n increased, span pf forecasting is converged. Like 82-days ahead forecasting has less standard deviation as compared to one-day ahead forecast.

6. CONCLUSION

This study presents comprehensive and well-defined algorithms for generation and reduction of wind power scenarios. Wind power scenarios have been generated using ARIMA model. These proposed scenario-generation algorithms can be used by wind power producers to optimize their strategy under unpredictability in wind availability, market prices, demand, etc. Accurate prediction of wind power reduces imbalance cost, but technical and environmental factors hamper its accurate forecasting. In this study we observed that the ARIMA model was able to extract and learn from the pattern in the wind energy data. We can trust on the future values forecasted by a well optimized ARIMA model. Its performance, computation time and ability to handle trends and non-stationarity makes it a good reliable model for researchers. We also observed that it gives more accurate forecasting for one-day ahead forecast as compared to one-week or one month as for one day forecast root mean square error, standard deviation and pattern of actual and forecasted values are more satisfactory. So, it can be considered as the best model for real-time implementation in short-term forecasting.

This ARIMA model can be successfully used in not only wind energy generation but also for bidding strategy formulation, transmission expansion planning and operation of various types of smart grids. The presented scenario-generation algorithms could be improved further by using advanced forecasting models. If a hybrid new model is created which can work as the combination of statistical models and machine learning models then we can improve upon these results as well as the computation costs.

7. REFERENCES

1. Anupriya, Shruti Garg, "Autoregressive Integrated Moving Average Model based Prediction of Bitcoin Close Price", *Smart Systems and Inventive Technology (ICSSIT) 2018 International Conference on*, pp. 473-478, 2018.
2. M.Y.Cho, J.C.Hwang, C.S.Chen, "Customer short term load forecasting by using ARIMA transfer function model" *Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD '95.2002.*
3. Cadenas, E.; Rivera, W. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificialneural networks. Renew. Energy 2009, 34, 274–278.
4. P. Sen, M. Roy and P. Pal, "Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization", *Energy*, vol. 116, pp. 1031-1038, Dec. 2016.
5. E. Erdogdu, "Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey", *Energy Policy*, vol. 35, no. 2, pp. 1129-1146, Feb. 2007.
6. https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06
7. https://www.kaggle.com/c/GEF2012-wind-forecasting/data
8. https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
9. http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/
10. https://online.stat.psu.edu/stat510/lesson/3/3.3
11. https://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test
12. https://medium.com/wwblog/stationarity-testing-using-the-augmented-dickey-fuller-test-8cf7f34f4844
13. Alfredo Nespoli, Marco Mussetta, Emanuele Ogliari, Sonia Leva, Luis Fernández-Ramírez, Pablo García-Triviño, "Robust 24 Hours ahead Forecast in a Microgrid: A Real Case Study", *Electronics*, vol. 8, pp. 1434, 2019.
14. Carta, J. A., Ramirez , P., and Velázquez, S., "A review of wind speed probability distributions used in wind energy analysis : Case studies in the Canary Islands ," Renew. Sustain. Energy Rev.,Vol.13,pp.933-955,2009.
15. University of Massachusetts, Amherst, "Wind energy center," available at: http://www.umass.edu/windenergy/resoursedata.php/
16. Morales, J. M., Minguez, R., and Conejo, A. J., "A methodology to generate statistically depended wind speed scenarios," Appl. Energy, Vol. 87, pp. 843–855, 2010.
17. Dupacova, J., Consigli, G., and Wallance, S. W., "Scenarios for multistage stochastic programs," Ann. Operat. Res., Vol. 100, pp. 25–53, 2000.

18. Growe-Kuska, N., Heitsch, H., and Romesh, W., “Scenario reduction and scenario tree construction for power management problems,” Proceedings of the 2003 IEEE Bologna Power Technology Conference, Bologna, Italy, 23–26 June 2003.
19. Salehain, A., “ARIMA time series modelling for forecasting thermal rating of transmission lines,” Proceedings of the 2003 IEEE Power Engineering Society Transmission and Distribution Conference and Exposition, Dallas, TX, 7–12 September 2003.
20. Olsson, M., and Soder, L., “Generation of regulating price scenarios,” Proceedings of the 2004 International Conference on Probability Methods Applied to Power Systems, Ames, IA,12–16 September 2004.
21. Muhamad Razali, N. M., and Hashim, A. H., “Backward reduction application for minimizing wind power scenarios in stochastic programming,” Proceedings of the 2003 IEEE International Power Engineering and Optimization Conference, Selangor, Malaysia, 23–24 June 2010.