I J C T A, 9(15), 2016, pp. 7681-7690

Wind Speed Forecasting based on Statistical Auto Regressive Integrated Moving Average (ARIMA) method

A. Vishnupriyadharshini*, V. Vanitha and T. Palanisamy*****

ABSTRACT

Wind power forecasting is of greater importance to increase the wind power penetration to the grid as well as to maintain the grid stability. Wind power varies with cubic times the wind speed. Thus, accurate forecasting of wind speed is a preliminary process. This paper projects the analysis of wind speed forecasting using various statistical approaches and describes Auto Regressive Integrated Moving Average (ARIMA) method based wind speed forecasting in detail. Historical time series wind speed data, with a time interval of 3 hours average, collected from Amrita Wind Energy Centre is considered for the analysis. Results show that ARIMA model forecast the wind speed with better accuracy. Consideration of multivariate data and seasonal factors are also suggested to improve the wind speed forecasting accuracy.

Keywords: Wind speed forecasting, Statistical methods, ARIMA model.

1. INTRODUCTION

In India, renewable energy has a share of approximately 13% of the total installed capacity. Among various renewable energy sources, wind energy plays a prominent role, contributing approximately 65% to the total renewable energy installed capacity. National Institute of Wind Energy (NIWE) recently estimated the wind power potential to be 302 GW at a hub height of 100m, considering only the sites suitable for practical wind turbine installations [1], [2]. But this potential remains underutilized due to various factors such as intermittency, inadequate transmission facility, improper planning and scheduling of renewable energy generation. One of the major problems associated with wind energy is its intermittent nature, which reduces the effectiveness of wind energy integration to the grid. To increase the wind energy penetration to the grid, wind electric generation has to be scheduled effectively for proper planning and scheduling of conventional power plants. For this, accurate wind speed forecasting is of greater importance. India, though achieved significant miles in load forecasting, still lags in wind power forecasting, struggling to meet the prescribed standard of Indian Electricity Grid Code (IEGC) [3], which states that wind power forecasting error should be less than 30%.

Wind power forecasting also has cost impact associated with large wind power penetration to the grid [4]. Atmospheric parameters like pressure, temperature, wind direction, air density, etc, have a direct impact on the accuracy of wind power forecasting [5]. Literature studies state that time series wind power forecasting models perform better than persistent models [6]. But there is also a lot of issues associated with statistical forecasting. Some of these issues can be mitigated through the forecasting capabilities of Auto-Regressive Moving Average and Artificial Neural Network model [7].

* Department of Electrical and Electronics Engineering Amrita School of Engineering, Coimbatore Amrita Vishwa Vidyapeetham Amrita University, India, *Email: vishnupriyadharshini@gmail.com*

** Department of Electrical and Electronics Engineering Amrita School of Engineering, Coimbatore Amrita Vishwa Vidyapeetham Amrita University, India, *Email: v_vanitha@cb.amrita.edu*

*** Department of Mathematics Amrita School of Engineering, Coimbatore Amrita Vishwa Vidyapeetham Amrita University, India

1.1. Objective

The main objective of this paper is to analyze the performance of short-term wind speed forecasting based on statistical methods. The impact of input order and seasonal factors on the accuracy of wind speed forecasting is also analyzed. Different statistical methods like Moving Average, Exponential Smoothing, Auto Regressive Integrated Moving Average (ARIMA) are considered for the analysis.

1.2. Overview

Historical time series wind speed data with a time interval of 3 hours average collected from Amrita Wind Energy Centre, (Amrita Vishwa Vidyapeetham University, Coimbatore, India), is considered for the statistical analysis of wind speed forecasting. Simple Moving Average (MA) method, Exponential Smoothing Method are considered for initial analysis without accounting seasonality. Seasonal effect is considered in Auto-Regressive Integrated Moving Average (ARIMA) method. The data are made stationary by deseasonalizing using differencing method. Auto Correlation and Partial Auto Correlation Functions are plotted to identify the order of the model. ARIMA model equations are formulated for different orders with trial and error method. General procedure for the above is described and ARIMA model, which gives a better result, is explained in detail.

2. STATISTICAL ANALYSIS OF WIND SPEED FORECASTING

Forecasting is the process of estimating the expected outcome of a parameter in the future. The estimation can be based on various factors such as historical data record, physical process involved, application, etc. Wind speed can be estimated through various forecasting techniques. There are ‘n’ number of wind speed and wind power forecasting techniques that can be broadly categorized into persistence method, physical method, statistical method and hybrid models. Time series models and Artificial Neural Network model comes under statistical method [8],[9],[10]. Hybrid models combining any of the methods together, for instance, Fuzzy and ANN models combined using Adaptive Neuro-Fuzzy Inference Systems (ANFIS), proves to give better accuracy [11].

Literature reviews in this area show that statistical method gives better accuracy for short term wind speed forecasting. Thus, different statistical methods are applied for forecasting wind speed.

Evaluation of the model is most commonly done using Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE) or Mean Absolute Percentage Error (MAPE) criteria[12],[13]. The evaluation criteria used in this paper is MAPE, which is given by,

$$MAPE = \frac{y(t) - y_f(t)}{y(t)} \tag{1}$$

Different statistical methods such as MA, ES, ARIMA are used for the purpose of analysis. These techniques can be understood from the materials given in reference [14] through [19].

Moving Average (MA) methods is used for the order analysis and time horizon analysis. This analysis helps in selection of order, (i.e) number of previous observations to be used. It also helps in proving that forecasting error increases with increase in time horizon (i.e) the time period ahead which the forecasting is done [20], [21].

Exponential smoothing (ES) method is similar to the Moving Average method except that the weights are considered for the previous observations in a decreasing order. This analysis helps in proving the fact that recent past observation has more effect on the future value. Optimization of weights also helps in reducing the error. Procedure and results of this analysis are not highlighted in this paper; instead more focus is given on ARIMA model, which proves to give better accuracy.

3. WIND SPEED FORECASTING BASED ON ARIMA MODEL

In statistical analysis, stationary series are considered for developing a sensible model. This seasonality elimination differentiates ARIMA model from ARMA. Auto Regressive Integrated Moving Average, generally termed as ARIMA model, is an extensive forecasting technique used for any data series containing seasonality. General representation of ARIMA model is ARIMA (p, d, q). p, d, q are the orders of Auto-Regressive, Differencing, Moving Average parts respectively. In ARIMA model seasonality is considered but the data is made stationary using differencing operation which leads to the integral term I in ARIMA model. General expression for ARIMA (p, d, q) model is given by,

$$\alpha_p\left(B^k\right)y_t = \beta_q\left(B^k\right)e_t \tag{2}$$

$$e_t = y_t - y_{t-1} \tag{3}$$

$$\left(B^k\right)y_t = y_{t-k} \tag{4}$$

3.1. Seasonal ARIMA model

Model involving seasonal differencing is termed as Seasonal ARIMA model or S-ARIMA model.

There are two types of S-ARIMA model, one is pure S-ARIMA model and another is multiplicative S-ARIMA model.

1) *Pure S-ARIMA model*: If normal differencing is not involved, pure S-ARIMA model is formed only by considering the observation and parameters of seasonal lags. General expression for pure S-ARIMA model is given by,

$$\phi_p\left(B^{sP}\right)w_t = \theta_Q\left(B^{sQ}\right)Z_t \tag{5}$$

$$w_t = \delta_S^D y_t = y_t - y_{t-s} \tag{6}$$

$$z_t = y_t - w_t \tag{7}$$

2) Multiplicative S-ARIMA model: In multiplicative S-ARIMA model, observations and parameters of both normal lags and seasonal lags are involved. General expression for multiplicative S-ARIMA model is given by,

$$\alpha_p\phi_P\left(B\right)w_t = \beta_q\theta_Q\left(B\right)z_t \tag{8}$$

3.2. Procedure and modelling of wind speed forecasting using ARIMA method

Wind speed forecasting using statistical ARIMA model involves various steps as follows [18], [19], [20], [21].

1) *Analysis of Time Series Data using Time Plot:* Time plot is the graph in which observations are plotted against time. Properties of time series data such as trend, seasonality, randomness, discontinuity, etc., which are important factors in model formulation, can be identified using this time plot. Overall pattern of the monthly data and the presence of trend can be analyzed using the time plot for wind speed time series data with 1-day average as shown in the Figure 1.

From the above time plot, no systematic change in mean or variance is found, i.e. mean varies over a constant value and the variance also remains more or less constant. Thus, this wind speed time series data is considered to be a stationary with no trend pattern.

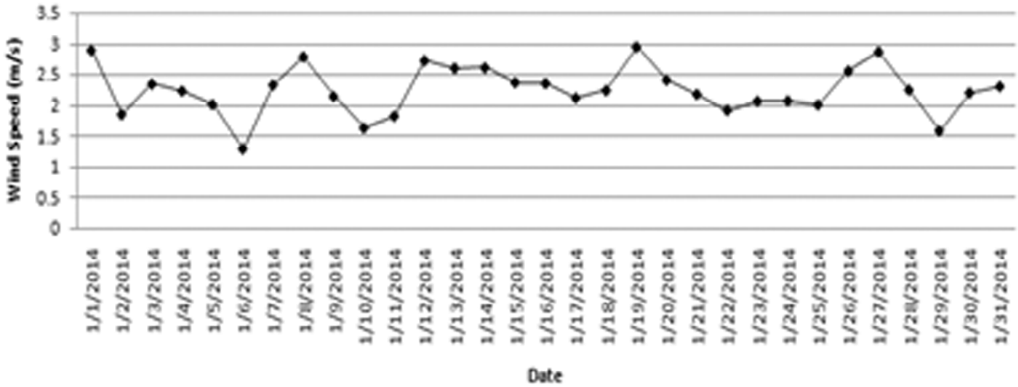

Figure 1: Time plot for 1-month daily average wind speed time series data.

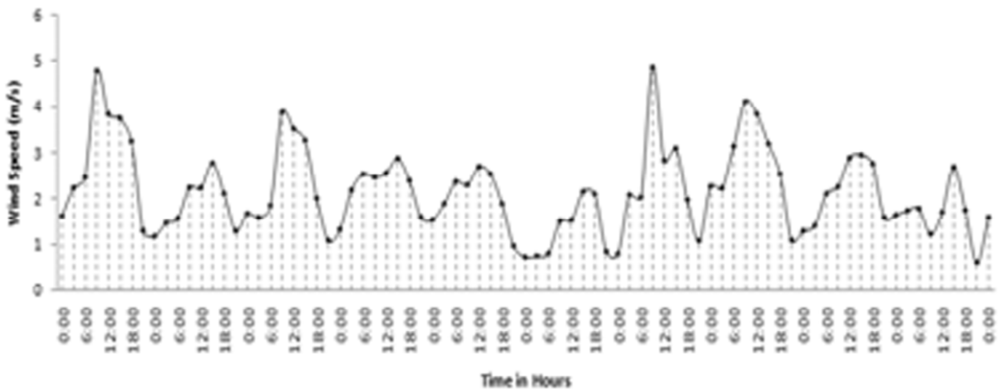

Figure 2: Time plot for 3 hours average wind speed time series data.

Considering 3 hours average for 1-month data leads to 8 observations per day at a regular interval of 3 hours. If the same pattern repeats over other days, it may be termed as seasonal variation. This can be identified from the Figure 2.

In Figure 2, wind speed pattern repeats, thus it is an evidence for the presence of seasonality. Though considering daily seasonality may not prove to be satisfactory for wind speed forecasting, as wind contains annual seasonal patterns, it is sufficient to analyze the statistical wind speed forecasting model.

2) *Check for Stationary Series*: Stationary series is the one in which there is no systematic change in mean and variance [19]. It is purely based on intuitive analysis. Time series analysis is concerned with stationary series. Hence, non-stationary series has to be transformed into stationary series by differencing method as described in Step 3.

i) Intuitive Analysis: In Figure 2, let 2 m/s to 2.5 m/s be considered as approximate mean. Other values more or less revolve around this mean value and there is no systematic change in mean or variance, so it is stationary.

ii) Theoretical analysis using Correlation Function: Auto correlation function and Partial Auto Correlation function plays a major role in statistical modelling. It is used in checking the stationarity of the series, model identification, and parameter estimation.

Auto Correlation Coefficient measures the correlation between observations at different distance apart. Considering N observation on time series with each observation separated by 1-time interval, correlation coefficient or the relationship between the observations can be measured using the expression,

$$r_k = \frac{\sum_{t=1}^{N-k}\left(y_t - Y\right)\left(y_{t+k} - Y\right)}{\sum_{t=1}^{N}\left(y_t - Y\right)^2} \tag{9}$$

Partial Auto Correlation coefficient measures the excess correlation between the observation k steps apart. In autocorrelation, if y_t and y_{t-1} are related, relationship between y_t and y_{t-2} will have the impact on y_t and y_{t-1}. But in partial auto correlation, impact of other variable is removed. General expression for PAC is

$$c_k = \frac{\left(r_k - r_{k-t}^2\right)}{\left(1 - r_{k-t}^2\right)} \tag{10}$$

Using ACF, stationarity of a series can be analyzed. Upper and lower critical values are given by 2/√N. Only if the Auto correlation coefficient liesoutside this these limits, it is considered to be significant. For stationary series only very few values of AC will be significant.

From the Figure 3, all the values of ACF lie within the limit, thus, it can be considered as insignificant proving that the data series is stationary. Hence, normal differencing is not required.

From the Figure 4, ACF at seasonal lags of 8,16,24,.. is significant. Thus, the wind speed data series contain seasonal which has to be removed through seasonal differencing.

3) *Transformation of data using iterative differencing method*: Differencing is a type of filtering that helps in transforming non-stationary data into stationary data. It is also employed to eliminate the seasonality from a seasonal time series data. Ordinary differencing is used for non-seasonal data and seasonal differencing is used for seasonal data. In both the cases first order differencing is sufficient, sometimes second order differencing may be required.

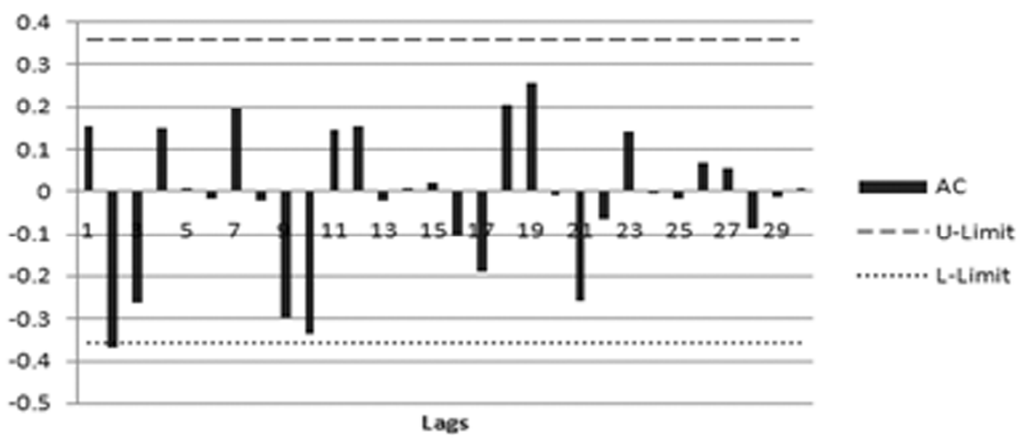

Figure 3: Auto correlation plot to check the stationarity of observed wind speed.

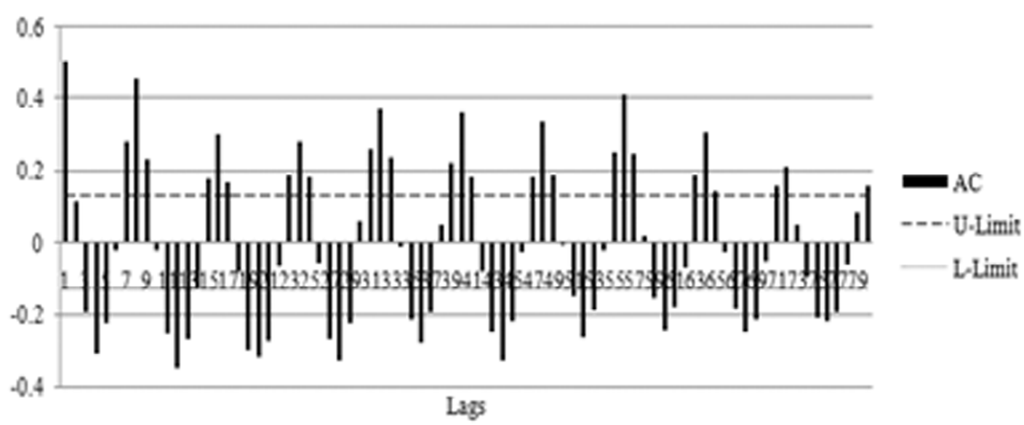

Figure 4: Auto correlation plot to check the seasonality in observed wind speed.

i) *Seasonal differencing:* Seasonal differencing is used for the time series data with seasonal pattern. Generally, seasonal differencing will be of the form

$$d_t = \delta^d \delta_S^D y_t \tag{11}$$

On considering 1-month daily wind speed data with 3 hours average, there are 249 data (N = 249) and seasonal factor s will be 8 as there are eight (24/3 = 8) 3 hours average observations in a day (s = 8).

First order seasonal differencing represents the difference between wind speed at time t and wind speed at time t-s. It is because wind speed at time t and t-s are considered to be similar in seasonal time series data. The general expression for first order seasonal differencing is as follows,

$$d_t = \delta_s y_t = y_t - y_{t-s} \tag{12}$$

ACF for the wind speed time series data obtained after seasonal first order differencing is shown in Figure 5. Here only a few seasonal lags are significant and it has transformed the actual wind speed into a stationary wind speed by eliminating the seasonal pattern.

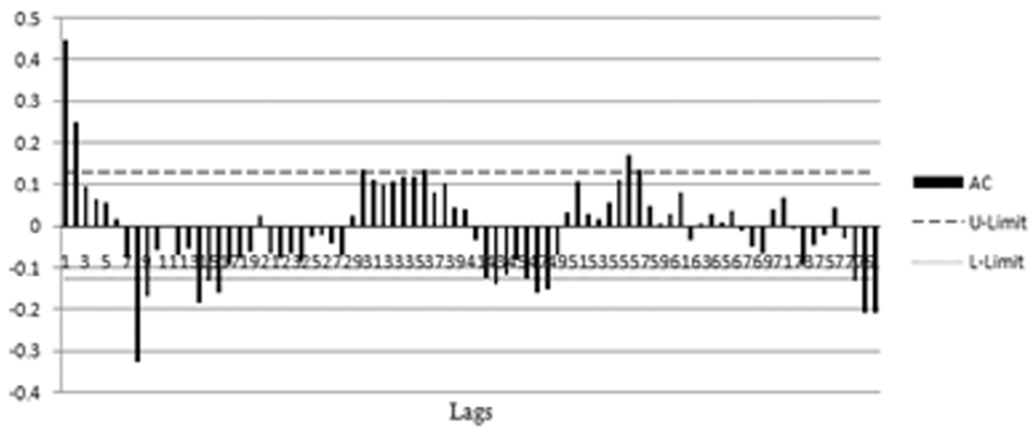

Figure 5: Auto correlation plot for 1st order seasonal differenced wind speed.

4) *Differencing order identification.:* Analysis of time series wind speed data confirmed the presence of seasonality. Statistical modelling of time series data deals with stationary series, thus trend or non stationarity can be eliminated using normal differencing of order d and seasonality can be eliminated using seasonal differencing of order D.

i) *Order of d:* Order of normal differencing 'd' can be identified from the Figure 3, absence of trend is clear as the series is almost stationary, as there is no significant change in mean of variance. Thus differencing is not required for the transformation of non-stationary series or for the elimination of trend. Therefore, order of normal differencing is considered to be zero (d = 0).

ii) *Order of D:* Order of seasonal differencing 'D' can be identified from the Figure 5, in which mean and variance of the data series vary only at seasonal lags of 8, 16, 24 and so on. Thus, first order seasonal differencing is sufficient (D = 1). Further differencing i.e second order seasonal differencing can also be considered. But sometimes this may result in over differencing error, which can be identified by iterative procedure.

5) *AR and MA order identification using Correlation function*: ACF and PACF plots play a major role in model identification. ACF and PACF of seasonal first order differencing are considered as it is involved in transformation process. PACF plot is considered for identification of Auto-Regressive order as the impact of other observations between any two observation is nil. ACF plot is considered for identification of

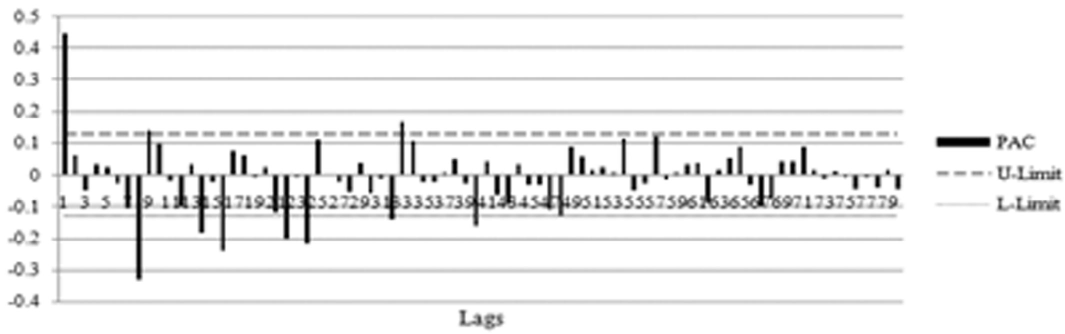

Figure 6: Partial Auto Correlation plot for 1st order seasonal differenced wind speed.

Moving Average order as other observations inbetween any two observation will impact the correlation between those two observations. This is explained in step 2, stationarity check.

i) *AR Order Identification:* Auto Regressive order (p, P) can be identified from PACF plot shown in Figure 6. p represents the number of significant normal lags and P represents the number of significant seasonal lags. PAC coefficient significant only at 1 normal lag. Thus order of p = 1. On considering seasonal lags (8, 16, 24..), PAC coefficients are significant for 1st 6 seasonal lags, (i.e) at lag 8, 16, 24, 32, 40 and 48. Hence order of P = 6.

ii) *MA Order Identification:* Moving Average order (q, Q) can be identified from ACF plot shown in Figure 5. q represents the number of significant normal lags and Q represents the number of significant seasonal lags. Only 2 AC coefficients at lag 1 and 2 are significant after which it decreases. Thus order of q = 2. On considering seasonal lags (8, 16, 24,.), AC coefficients are significant for 1st 2 seasonal lags, (i.e) at lag 8 and 16. Hence order of Q = 2.

6) Parameter Estimation: α, β, θ are the parameters to be estimated for modeling of ARIMA-based wind speed forecasting. α is the coefficient normal AR part, (i.e) significant normal lags of PACF. is the coefficient of seasonal AR part, (i.e) significant seasonal lags of PACF. β is the coefficient of normal MA part, (i.e) significant normal lags of ACF. θ is the coefficient of seasonal MA part, (i.e) significant seasonal lags of ACF. Based on the p, P, q, Q order coefficient parameters are estimated and it is shown in Table I.

7) Model building: Based on the identification of orders and coefficient parameters, ARIMA method based wind speed forecasting model is formulated. It is a purely iterative process. It involves a number of iteration based on the differencing involved. Sometimes, it requires more than 8 iterations for each model considering various differencing options. Models shown here are the one for which wind speed forecasting error is less.

i) *Pure S-ARIMA (6, 1, 2):* In pure S-ARIMA model, only seasonal orders (P, D, Q) and seasonal parameters (Õ and è) are considered. General expression for the model is given by,

Table 1
Estimated parameters for ARIMA modelling

Coefficient	ARIMA model parameters			
	ℑ	â	Õ	È
1	0.4473	0.4473	-0.3291	-0.3302
2	-	0.2497	-0.2383	-0.1612
3	-	-	-0.2175	-
4	-	-	-0.1385	-
5	-	-	-0.1635	-
6	-	-	-0.1289	-

$$\phi_P\left(B^s\right)w_t = \theta_Q\left(B^s\right)z_t \tag{13}$$

$$\phi_6\left(B^{Ps}\right)w_t = \theta_2\left(B^{QS}\right)z_t \tag{14}$$

$\phi_6\left(B^{Ps}\right)$ and $\theta_2\left(B^{Qs}\right)$ are the polynomials with s = 8, P = 1, 2, .., 6, Q =1, 2 and are expressed as,

$$\phi_6\left(B^{Ps}\right) = \left(1-\phi_1 B^8 - \phi_2 B^{16} - \phi_3 B^{24} - \phi_4 B^{32} - \phi_5 B^{40} - \phi_6 B^{48}\right) \tag{15}$$

$$\theta_2\left(B^{Qs}\right) = \left(1+\theta_1 B^8 + \theta_2 B^{16}\right) \tag{16}$$

On solving the general expression considering Backshift operator and polynomials as in (4), (15), (16) final equation of pure S-ARIMA(6,1,2) is expressed as,

$$\begin{aligned} w_t &= \phi_1 w_{t-8} + \phi_2 w_{t-16} + \phi_3 w_{t-24} \\ &+ \phi_4 w_{t-32} + \phi_5 w_{t-40} \\ &+ \phi_6 w_{t-48} + z_t + \theta_1 z_{t-8} \\ &+ \theta_2 z_{t-16} \end{aligned} \tag{17}$$

This model is built and the result obtained is 49.39 % MAPE which is high when compared to the required 30 %. Here the normal observations (i.e) previous observations of seasonal data are not considered. On considering previous observations along with the seasonal observations, the error may reduce. To evaluate this, multiplicative seasonal ARIMA of the form S-ARIMA (p, d, q) (P, D, Q) is formulated.

ii) Multiplicative S-ARIMA (1, 0, 2) (6, 1, 2): In multiplicative S-ARIMA (p, d, q) (P, D, Q) model observations of both seasonal and normal lags are considered. General expression for the model is given by,

$$\alpha_p(B)\phi_P\left(B^s\right)w_t = \beta_q(B)\theta_Q\left(B^s\right)z_t \tag{18}$$

$$\alpha_1(B)\phi_6\left(B^{Ps}\right)w_t = \beta_2(B)\theta_2\left(B^{Qs}\right)z_t \tag{19}$$

$\alpha_1(B), \beta_2(B), \phi_6\left(B^{Ps}\right), \theta_2\left(B^{Qs}\right)$ are the polynomials with s = 8, P = 1, 2, .., 6, Q = 1, 2 and are expressed as,

$$\alpha_1(B) = \left(1-\alpha_1 B^1\right) \tag{20}$$

$$\beta_2(B) = \left(1+\beta_1 B^1 + \beta_2 B^2\right) \tag{21}$$

$$\phi_6\left(B^{Ps}\right) = \left(1-\phi_1 B^8 - \phi_2 B^{16} - \phi_3 B^{24} - \phi_4 B^{32} - \phi_5 B^{40} - \phi_6 B^{48}\right) \tag{22}$$

$$\theta_2\left(B^{Qs}\right) = \left(1+\theta_1 B^8 + \theta_2 B^{16}\right) \tag{23}$$

On solving the general expression considering the Backshift operator and polynomials as in (4), (20) through (23), final equation of multiplicative S-ARIMA (1, 0, 2) (6, 1, 2) is complex containing 'n' number of terms. To simplify, it is written as follows,

$$f = \left(\phi_1 w_{t-8} + \phi_2 w_{t-16} + \phi_3 w_{t-24} + \phi_4 w_{t-32} + \phi_5 w_{t-40} + \phi_6 w_{t-48}\right) \tag{24}$$

$$g = \left[\alpha_1\left(w_{t-1} - \phi_1 w_{t-9} - \phi_2 w_{t-17} - \phi_3 w_{t-25} - \phi_4 w_{t-33} - \phi_5 w_{t-41} - \phi_6 w_{t-49}\right)\right] \tag{25}$$

$$h = \left(z_t + \theta_1 z_{t-8} + \theta_2 z_{t-16}\right) \tag{26}$$

$$i = \left[\beta_1 \left(e_{t-1} + \theta_1 z_{t-8} + \theta_2 z_{t-16}\right)\right] \tag{27}$$

$$j = \left[\beta_2 \left(e_{t-2} + \theta_1 z_{t-11} + \theta_2 z_{t-18}\right)\right] \tag{28}$$

Actual expression for the multiplicative S-ARIMA(1, 0, 2) (6, 1, 2) can be written by combining equations (24) through (28), as

$$w_t = f + g + h + i + j \tag{29}$$

The Result obtained is 42.8256 % MAPE which is still greater than the required 30 %. But on considering both seasonal and normal terms, error reduces.

4. CONCLUSION AND FUTURE SCOPE

Based on the analysis, wind speed forecasting using statistical ARIMA method, gives better result. This is because of the consideration of seasonal factors. Analysis also proves the importance of selection of the number of previous data and the identification of model which gives correct representation of the state.

Univariate data having only time series wind speed is considered, which may not be sufficient to represent the actual state of wind speed in a future time. Wind speed depends on different variable such as pressure, temperature, humidity, surface roughness, terrain features, long-term climatological parameters, geographical features etc. Thus, including all these factors in multivariate analysis may improve the accuracy of wind speed forecasting to a greater extent.

REFERENCE

[1] Wind Energy Potential in India, National Institute of Wind energy, India. http://niwe.res.in/department_wra_100m%20agl.php

[2] Birol, Fatih. "India Energy Outlook 2015 – World Energy Outlook Special Report," International Energy Agency, 2015.

[3] Central Electricity Regulatory Commission, "Indian Electricity Grid Code (IEGC) 2010," Apr. 2010.

[4] Ahlstrom, Mark L., and Robert M. Zavadil, "The Role of Wind Forecasting in Grid Operations & Reliability," *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES. IEEE*, 2005.

[5] Dongmei, Zhao, Zhu Yuchen and Zhang Xu, "Research on wind power forecasting in wind farms," *Power Engineering and Automation Conference (PEAM), IEEE*, Vol. 1, 2011.

[6] Milligan,Michael, M. N. Schwartz and Yih-huei Wan, "Statistical wind power forecasting for US wind farms," National Renewable Energy Laboratory 1617, 2004.

[7] Gomes, Pedro, and Rui Castro. "Comparison of statistical wind speed forecasting models," *Sustainable Technologies (WCST), 2011 World Congress on IEEE*, 2011.

[8] Soman, Saurabh S., et al, "A review of wind power and wind speed forecasting methods with different time horizons," *North American Power Symposium (NAPS), IEEE*, 2010.

[9] Chang, Wen-Yeau, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering*, Vol. 161, Apr. 2014.

[10] Wang, Xiaochen, Peng Guo, and Xiaobin Huang, "A review of wind power forecasting models," *Energy procedia*, Vol. 12, pp. 770-778, 2011.

[11] Kishore, SR Nandha, and V. Vanitha, "Wind speed forecasting for grid code compliance," *Journal of Renewable and Sustainable Energy 5*, Vol. 6, p. 063125, 2013.

[12] Zhao, Xin, Shuangxin Wang, and Tao Li, "Review of evaluation criteria and main methods of wind power forecasting," *Energy procedia*, Vol. 12, pp. 761-769, 2011.

[13] Willmott, Cort J., and Kenji Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, Vol.79, Jan. 2005.

[14] Indian Meteorological Department, "Forecasters guide".

[15] Choong, Joe, "Powerful Forecasting With MS Excel," Kindle Edition, 2012.

[16] Arora P.N., et al, *Comprehensive Statistical Methods*, S. Chand, 2007.

[17] T.M.J.A. Cooray, "*Applied Time Series Analysis and Forecasting.*"

[18] Box, George EP, et al., *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.

[19] Chatfield, Chris, *The analysis of time series: an introduction*, CRC press, 2013.

[20] Conejo, Antonio J., et al. "Forecasting electricity prices for a day-ahead pool-based electric energy market," *International Journal of Forecasting*, Vol. 3, pp. 435-462. 2005.

[21] Contreras, Javier, et al. "ARIMA models to predict next-day electricity prices," *Power Systems, IEEE Transactions*, pp. 1014-1020, 2003.