Statistical Analysis and Forecasting of Wind Speed

Conference Paper · April 2022

DOI: 10.1109/ICEPE55035.2022.9798358

4 authors, including:

Sakshi Shukla
Birla Institute of Technology and Science Pilani

Sarita Sheoran
Birla Institute of Technology and Science Pilani

Statistical Analysis and Forecasting of Wind Speed

Sakshi Shukla
Department of Mathematics
Birla Institute of Technology and Science, Pilani
Rajasthan, India
p20200443@pilani.bits-pilani.ac.in

Rohit Ramaprasad
Department of Mathematics
Birla Institute of Technology and Science, Pilani
Rajasthan, India
f20180224@pilani.bits-pilani.ac.in

Sumanta Pasari
Department of Mathematics
Birla Institute of Technology and Science, Pilani
Rajasthan, India
sumanta.pasari@pilani.bits-pilani.ac.in

Sarita Sheoran
Department of Mathematics
Birla Institute of Technology and Science, Pilani
Rajasthan, India
p20180439@pilani.bits-pilani.ac.in

***Abstract*—Energy plays a vital role in urbanization and industrialization. Wind energy is highly valuable and accurate forecasts can help determine the best locations to set up windmills. Using a dataset comprising wind speeds from 15 years (2000–2014) within two locations of Rajasthan, namely Jaipur and Jaisalmer, we present a detailed statistical analysis including distribution analysis and forecasting using Moving Average (MA), Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). We show empirically why SARIMA is the best model and why the former four models are inadequate when it comes to forecasting wind speeds.**

***Index Terms*—Wind Speed Forecasting, AR, MA, ARMA, ARIMA, SARIMA**

I. Introduction

With growing industries and populations, the demand for power has considerably increased. Over 79% of India's electricity demand is met from fossil fuel resources. Climate change and rising global average temperatures have urged nations to adopt renewable resources. According to RENs21 report, India has a total of 38.124GW installed wind power capacity, the fourth largest in the world. The government of India has set an ambitious target of installing 175GW of renewable energy capacity by 2022 of which 60GW is from wind power. The Ministry of New and Renewable Energy, India has partnered with Adani Power for setting up new wind farms in India.

In recent years, several algorithms have been proposed for reliable wind energy prediction. These algorithms are broadly divided into three groups: physical methods, statistical methods and artificial intelligence techniques. To date several statistical models have been proposed for wind energy forecasting. Among these, time series models are particular technique used to forecast wind speed. The authors in [1] discussed the time series-based approach for renewable energy modeling on different region of Turkey. The authors in [2] considered univariate and multivariate models for short-term wind speed forecasting using exponential smoothing and ARIMA model.

In [3], the author has proposed the probabilistic and deterministic wind speed forecasting based on non-parametric approaches and wind characteristics information. The authors in [4] studied the forecasting of wind speed and direction tuple. Four approaches based on ARMA model is proposed and wind speed and direction both consider for forecasting. Hocaoglu and Karanfil in [5] discuss a hybrid approach to multi-step, short-term wind speed forecasting using correlated features and Zheyong et al. [6] developed decompostion based time series model to improve the forecasting accuracy.

Using a dataset comprising wind speeds from 15 years (2000–2014) within two locations of Rajasthan namely, Jaipur and Jaisalmer. we present a detailed statistical analysis including distribution analysis and forecasting using Moving Average (MA), Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA).

The structure of the remaining portion is outlined below. The description of time series methods is mentioned in Section 2. The dataset and methodology of the paper are summarized in Section 3. Section 4 presents the results of the implemented statistical models and finally, concluding remarks are presented in Section 5.

II. Time Series Methods

This part presents some basics of time series methods that are used to forecast wind energy. For each method, a brief overview is provided along with a mathematical representation.

978-1-6654-7124-4/22/31.00 ©2022 IEEE

A. Auto-Regressive (AR)

The autoregression in a time series considers that the output variable linearly depends on its own prior values as well as a stochastic component (an unpredictable term). An AR model of order p is given by [7]:

$$X_t = \sum_{i=1}^{p} \psi_i x_{t-i} + \omega_t = \psi_1 x_{t-1} + \psi_2 x_{t-2} + ... + \psi_p x_{t-p} + \omega_t \tag{1}$$

Here x_t denotes the values of time series; ω_t denotes noise; $\psi = (\psi_1, \psi_2, .., \psi_p)$ is the model coefficient vector and p is a positive integer.

B. Moving Average (MA)

In contrast to an AR model, which applies a weighted total of previous values to determine a statistical illustration, the Moving Average (MA) process considers that the output variable is linearly dependent on the present and numerous previous values of a random term. The MA process of order q is as follows [7]:

$$X_t = \sum_{j=1}^{q} \theta_j \omega_{t-j} + \omega_t = \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + ... + \theta_q \omega_{t-q} + \omega_t \tag{2}$$

Here $\theta = (\theta_1, \theta_2, .., \theta_q)$ is the the model coefficient vector and q is a positive integer.

C. Auto-Regressive Moving Average (ARMA)

An ARMA process is integrated by Autoregressive (AR) and Moving Average (MA) method to output a process with a minimal parametrization. An ARMA process of order p and q is expressed as follows [8].

$$X_t = \sum_{i=1}^{p} \psi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \omega_{t-j} \tag{3}$$

Here ψ_i and θ_j are the coefficients of AR and MA part of an ARMA model.

D. Auto-Regressive Integrated Moving Average (ARIMA)

An ARIMA process is preferred when the data exhibits some evidences of non-stationarity. Predictions are based on past values of time-series data in AR models, whereas prior residuals are used for forecasting future values in moving average models. An ARIMA (p, d, q) model is created by combining a stationary ARMA (p, q) process with the d-th difference of a time series. The underlying procedure can be stated as follows [9]:

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \psi_i x_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j} \tag{4}$$

Here d is the degree of differencing to get stationary; p and q respectively denote the order of AR and MA models.

III. Methodology

A. Visualization of Time Series

Before we proceed with any analysis of our data, it is imperative that we obtain some preliminary visualizations and plots to better understand our data. Descriptive statistics such as mean values and other statistical data such as presence or absence of outlier and other major variations (if any) can be identified through these plots.

From the yearly boxplot for Jaipur depicted in Figure 1, it is observed that the 75th quartile value (denoted by the upper edge of each box) lies somewhere around 2.50 m/s and 3.70 m/s, whereas the 25th quartile value (denoted by the lower edge of each box) lies somewhere around 1.7 m/s and 2.2 m/s. The median value (denoted by the middle edge of each box) lies around 2.50 m/s and 3.56 m/s. The upper edge of circles at the top denote the presence of some outliers in the data.

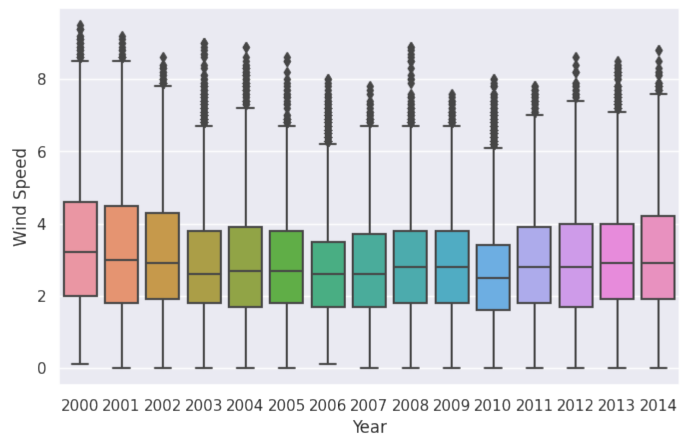

Fig. 1. Yearly boxplot of wind speed in Jaipur

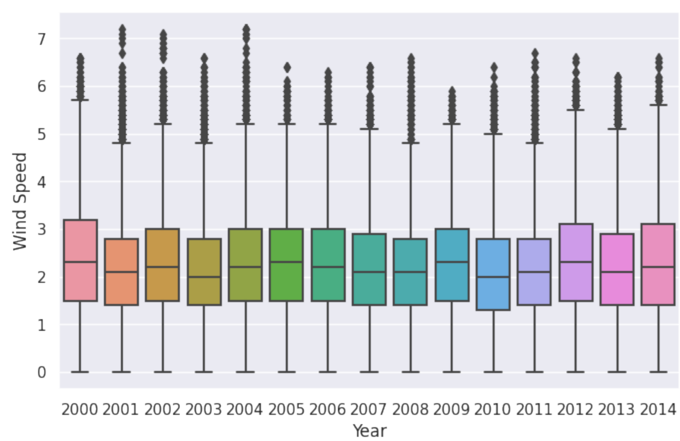

Fig. 2. Yearly boxplot of wind speed in Jaisalmer

From the violin plots in Figure 5,6 it is evident that the median value and the distribution (indicated by the width of the plots) for the wind speed do not seem to vary much

	DHI	DNI	Temperature	Pressure	Wind Speed
DHI	1.000000	0.831906	0.618573	-0.148639	-0.082015
DNI	0.831906	1.000000	0.471489	0.064150	-0.182814
Temperature	0.618573	0.471489	1.000000	-0.609877	0.130209
Pressure	-0.148639	0.064150	-0.609877	1.000000	-0.397750
Wind Speed	-0.082015	-0.182814	0.130209	-0.397750	1.000000

Fig. 3. Correlation heatmap of Jaipur data

	DHI	DNI	Temperature	Pressure	Wind Speed
DHI	1.000000	0.758557	0.585716	-0.143924	-0.064402
DNI	0.758557	1.000000	0.470963	0.076129	-0.149942
Temperature	0.585716	0.470963	1.000000	-0.578090	0.181856
Pressure	-0.143924	0.076129	-0.578090	1.000000	-0.323331
Wind Speed	-0.064402	-0.149942	0.181856	-0.323331	1.000000

Fig. 4. Correlation heatmap of Jaisalmer data

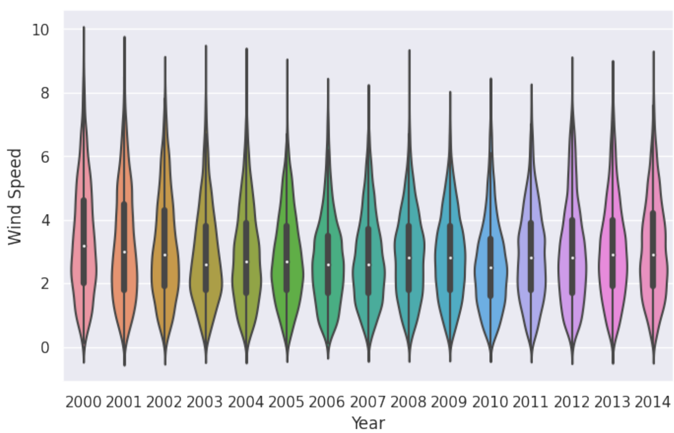

Fig. 5. Yearly violin plot of wind speed in Jaipur

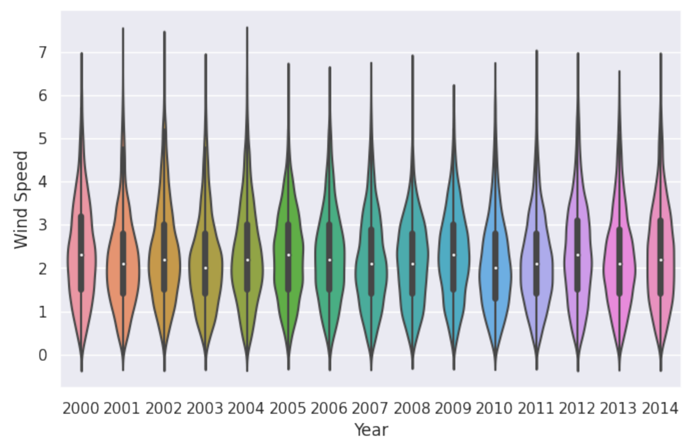

Fig. 6. Yearly violin plot of wind speed in Jaisalmer

over the years for all these two locations, though there are minor variations in the maximum/minimum values for the wind speeds achieved in the particular year. Thus, it may be concluded that for the data analysis of a given location, any year's data can be used with the results in a particular year not differing much from any other year.

Now, we proceed to analyse the various factors provided to us in the dataset. Through a survey of previous works [10], we find that factors affecting wind speed are: pressure gradient, temperature and, topography

While parameters like GHI, DNI, DHI indirectly affect the wind speed by influencing the local climatic conditions, dew point temperature and relative humidity more directly affect precipitation.

We plot a heatmap as shown in Figure 3,4 for correlation between various factors for Jaipur and Jaisalmer data. We observe that the provided factors (DHI, DNI, GHI and so on) do not exhibit a good correlation with wind speed data. Only pressure has a weak correlation with wind speed. This is supported by previous studies [10].

IV. Time Series Decomposition

Before proceeding to decompose the time series, we must ensure that our data is not stationary. We conduct the Augmented Dickey–Fuller test to check for stationarity of our time series. When applied on the entire data for year 2014 of Jaipur (similar results are obtained for Jaisalmer as well), we obtain the below results.

$$H0 : \text{Time series is non-stationary}$$

$$Ha : \text{Time series is stationary}$$

$$ADF Statistic : -6.454084$$

$$p - value : 1.495e - 08$$

As the P-value is less than 0.05, the results seem to indicate that our time series is stationary. We claim that this is because the ADF test fails to detect seasonal non-stationarity due to the presence of almost 1,00,000 data points (all 14 years data) making the data very dense. For the months of August and September 2014 in Jaipur, the ADF test gives us p-values of 4.199e-09 indicating the presence of stationarity in our time series. Also, we use a additive model as no seasonal component in wind speeds are available.

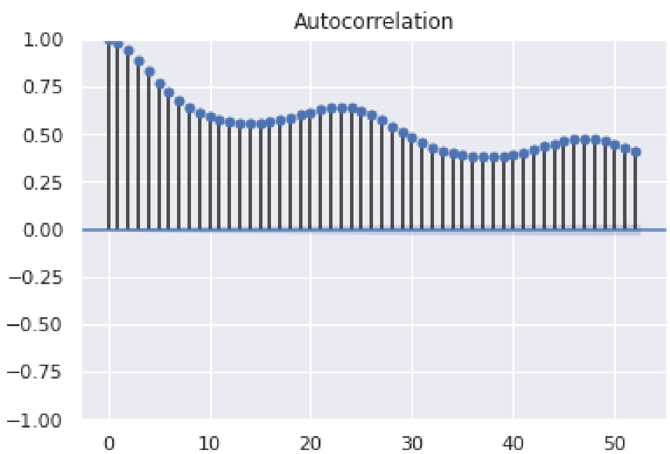

Fig. 7. ACF plot for Jaipur for non-differenced data

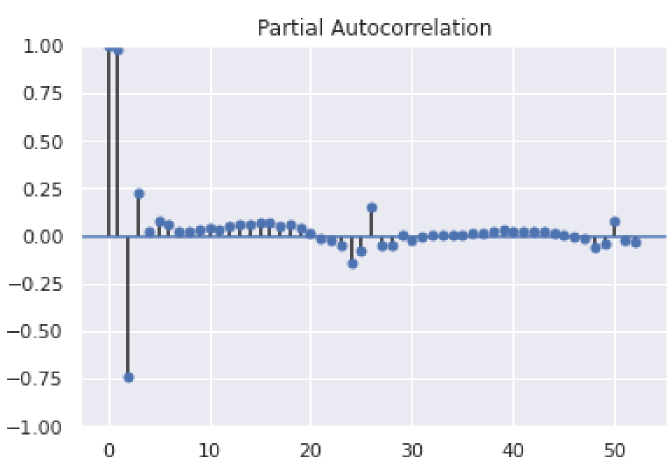

Fig. 8. PACF plot for Jaipur for non-differenced data

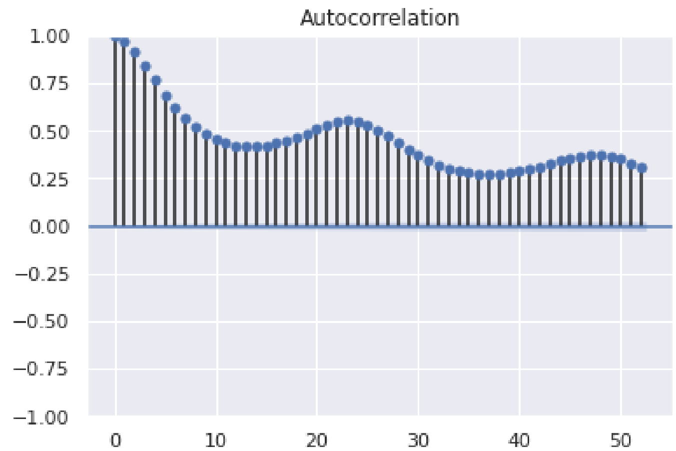

Fig. 9. ACF plot for Jaisalmer for non-differenced data

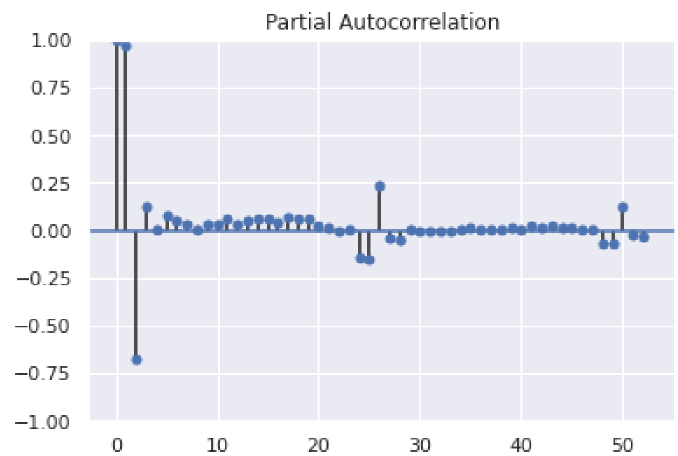

Fig. 10. PACF plot for Jaisalmer for non-differenced data

V. Wind Speed Forecasting

We carry out hourly forecasting for two locations in Rajasthan. For hourly forecasting we carry out a comparative analysis of five time series models - AR, MA, ARMA, ARIMA and SARIMA. Our models are fitted on the last year's (2014) data excluding the last two days, that is 48 hours, which we use for testing our models.

AR, MA and ARMA are not expected to do well on our non-stationary data as they don't involve any differencing of data and hence do not take care of any seasonality in the data. However, we still apply these three models in hourly forecasting to back-up this claim.

To find out the parameters of the AR and MA models we make use of the ACF and PACF plots for all the four states. We use the fact that if the PACF displays a sharp cutoff or the first lag autocorrelation is positive, then the point at which the PACF cuts or the number of significant lags in the PACF plot is our AR order. Similarly, if the autocorrelation function (ACF) of the series displays a sharp cutoff or the first lag autocorrelation is negative then the point at which the ACF cuts or the number of significant lags is the order of our MA model.

We show the ACF and PACF plots for Jaipur and Jaisalmer on the non differenced data in figure 9,8,10,7. The ACF plots show a very gradual decline and an overall scalloped shape, suggesting that any feasible MA model won't perform well on it. The PACF plots on the other hand have 3 or 4 significant lags for all the regions suggesting that an AR(3) or AR(4), i.e., AR models with lag orders 3 or 4 could do well on this data.

The results of AR, MA, ARMA and ARIMA for Jaipur for weekly data are shown in Figure 11, 12, 13, 15, 14. The results for other location is similar. With the exception of ARIMA and SARIMA, the forecasts obtained from the remaining models were not satisfactory. We obtain best hourly forecasts using the SARIMA model, it outperforms the ARIMA model, but not by much, an important consideration when choosing models for further analysis. Results from our SARIMA model are shown in Figure 14.

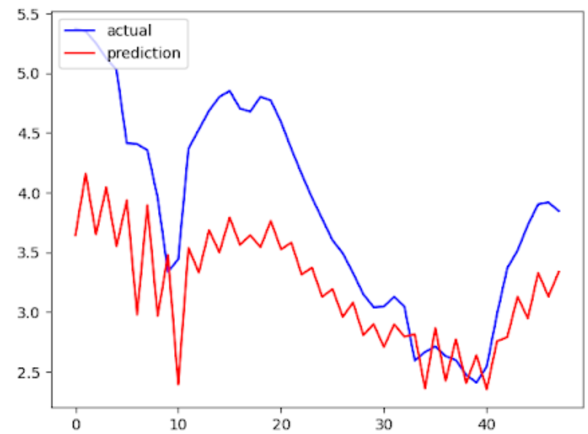

Fig. 12. MA based forecast of Jaipur

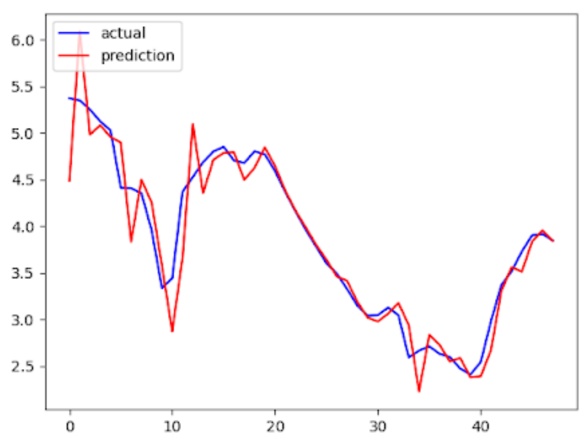

Fig. 11. AR based forecast of Jaipur

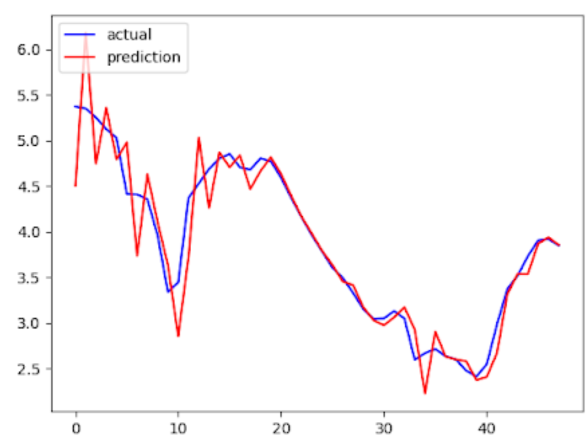

Fig. 13. ARMA based forecast of Jaipur

Judging from the results of the hourly forecasts, it is evident that ARIMA and SARIMA models outperform AR, MA and ARMA models because the differencing involved in those two models makes the time series data trend stationary, an important prerequisite of any time series forecasting. As is evident, the MA model is a very poor fit (RMSE = 0.5994). AR and ARMA models perform much better, with RMSE values of 0.204 and 0.199 respectively.

VI. Summary

In this paper. we present a detailed analysis of forecasting wind speeds for two locations in Rajasthan. We empirically evaluate five statistical models, Moving Average (MA), Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). We demonstrate that SARIMA is the best

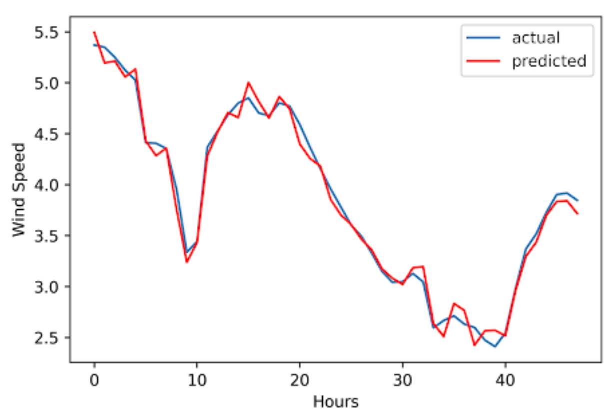

Fig. 14. SARIMA based forecast of Jaipur

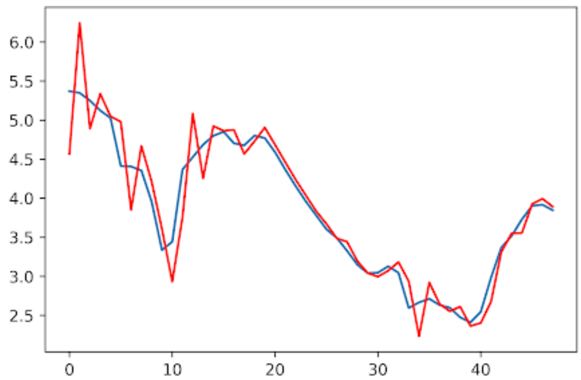

Fig. 15. ARIMA based forecast of Jaipur

model to be used for forecasting as it takes into account both the non-stationarity and seasonality of the wind speeds.

REFERENCES

[1] Agbulut U, "A novel stochastic model for very short-term wind speed forecasting in the determination of wind energy potential of a region: A case study from Turkey", *Sustainable Energy Technologies and Assessments*, 2021, pp 1-15.

[2] Priya C B, Arulanand N, "Univariate and multivariate models for Short-term wind speed forecasting", *Materials Today: Proceedings*, 2020, pp 1-9.

[3] Heng J, Hong Y, Hu J, Wang S, "Probabilistic and deterministic wind speed forecasting based on non-parametric approaches and wind characteristics information", *Applied Energy*, 2021, pp 1-13.

[4] Erdem E, Shi J, "ARMA based approaches for forecasting the tuple of wind speed and direction", Applied Energy, 2011, pp 1405-1414.

[5] Hocaoglu F, Karanfil F, "A time series-based approach for renewable energy modeling", *Renewable and Sustainable Energy Reviews*, 2013, pp 204-214.

[6] Jebli I, Belouadha FZ, Kabbaj MI, Tilioua A, "Deep Learning based Models for Solar Energy Prediction", *Advances in Science, Technology and Engineering Systems Journal*, 2021, pp 349-355.

[7] Moon J, Hossain Md B, Chon KH, "AR and ARMA model order selection for time-series modeling with ImageNet classification", *Signal Processing*, 2021 pp 1-11.

[8] Chen T, Chen G, Chen W, Houd S, Zhen Y "Application of decoupled ARMA model to modal identification of linear time-varying system based on the ICA and assumption of short-time linearly varying", *Journal of Sound and Vibration*, 2021, pp 1-21.

[9] Yongganga D, Huan W, Mingqiang W,Linjianga T, Tao Y, "Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction", *Petroleum*, 2021, pp 1-8.

[10] Wooten RD, "Statistical Analysis of the Relationship Between Wind Speed, Pressure and Temperature." *Journal of Applied Sciences*, 2011, pp 2712–2722.

[11] Dorvlo, Atsu S S, "Estimating Wind Speed Distribution", *Energy Conversion and Management*, 2002, pp 2311–2318.

[12] Cheng, Yung H, "Hybrid solar irradiance now-casting by fusing Kalman filter and regressor", *Renewable Energy*, 2016, pp 434-441.

[13] Zhao Y, Ye L, Li Z, Song X, Lang Y, Su J, "A novel bidirectional mechanism based on time series model for wind power forecasting", *Applied Energy*, 2016, pp 793-803.

[14] Lopes, Raul H C, "Kolmogorov-Smirnov Test", *International Encyclopedia of Statistical Science*, 2011, pp. 718–720.

[15] Kaoga D K, Danwe R, Yamigno S D, Djongyang N, "Performance Analysis of Methods for Estimating Weibull Parameters for Wind Speed Distribution in the District of Maroua." *Journal of Fundamental and Applied Sciences*, 2015, pp 153–174.

[16] Löning M, Bagnall A, Ganesh S, Kazakov V, Lines J, Király F J, "Sktime: A Unified Interface for Machine Learning with Time Series", ArXiv Preprint ArXiv:1909.07872, 2019, pp 1-9.

[17] Kreinovich V, Nguyen H T, Ouncharoen R, "How to Estimate Forecasting Quality: A System- Motivated Derivation of Symmetric Mean Absolute Percentage Error (SMAPE) and Other Similar Characteristics", 2014.

[18] Orosa J A, Oliveira A C, Costa A M, "New Procedure for Wind Farm Maintenance", *Industrial Management and Data Systems*, 2010, pp 861–882.

[19] Rajeevan A K, Shouri P V, Nair U, "ARIMA Modeling of Wind Speed for Wind Farm Reliability Analysis", *Emerging Research Areas: Magnetics, Machines and Drives (AICERA/ICMMD), Annual International Conference*, 2014, pp 1–5.

[20] Cadenas E, Rivera W, Amezcua RC, Heard C, "Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model", *Energies*, 2016, pp 1-18.

[21] Wang X, Guo P, Huang X, "A Review of Wind Power Forecasting Models" Energy Procedia, 2011, pp 770–778.