

Wind Speed Prediction Based on Seasonal ARIMA model

Ilham Tyass^{1*}, Abdelouahad Bellat¹, Abdelhadi Raihani¹, Khalifa Mansouri¹, and Tajeddine Khalili^{1,2}

¹ENSET—SSDIA Laboratory, Hassan II University of Casablanca, Casablanca, Morocco

²ENSA Béni Mellal, TNAI Research Team, SULTAN MOULAY SLIMAN University, Béni Mellal, Morocco.

Abstract. Major dependency on fossil energy resources and emission of greenhouse gases are common problems that have a very harmful impact on human communities. Thus, the use of renewable energy resources, such as wind power, has become a strong alternative to solve this problem. Nevertheless, because of the intermittence and unpredictability of the wind energy, an accurate wind speed forecasting is a very challenging research subject. This paper addresses a short-term wind speed forecasting based on Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The forecasting performances of the model were conducted using the same dataset under different evaluation metrics in terms of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) performance evaluation metrics. The obtained results denote that the used model achieves excellent forecasting accuracy.

1 Introduction

Renewable energies (RE) provide sustainable solutions, to the energy challenges of the 21st century, climate change, air pollution, depletion of resources, and fast demographic evolution. Actually, more than 90% of the world's new renewable energy production capacity is from solar and wind [1]. Wind energy is among the most promising sources [2]. Indeed, this source coupled with interconnected grid networks could provide on the one hand, the energy needs, and on the other hand, prevent the excessive depletion of fossil sources and reduce the harmful gas emission. The development of a wind farm projects makes use of climate data. So, very short-term forecasts ranging from a few minutes to several days will be used for the optimization of electricity market transactions, sizing and management of electricity networks, including reserves, at different spatial and temporal scales. Wind speed is the main factor in the wind power generation system.

Subsequently, accurate wind speed forecasting is critical for system operators to ensure reliable electricity supply, thus operating the system securely, in terms of system stability, auxiliary service and power quality [3]. Because of its nonlinear and non-stationary characteristics, as well as the complex interactions with several meteorological factors, forecasting the wind speed precisely is a difficult task.

The first works on wind speed forecasting problems for the next hours and days appeared in the early 1980s. These methods were presented by the authors in [4], who

proposed to give information on the uncertainty of forecasts. Nowadays, the literature related to wind forecasting and wind energy production is expanding very rapidly. A major part of the developments in forecasting methods is devoted to time series modeling methods (AR, ARMA, ARIMA...etc.). Authors in [5] used an Autoregressive Moving Average (ARMA) model to predict the hourly average wind speed and compared the results with the persistence model, and they concluded that the errors of ARMA model were smaller than the ones of the persistence model. In [6], authors compared the Autoregressive Integrated Moving Average (ARIMA) and the Artificial Neural Network (ANN) methods applied to a time series of wind speed measurements. They found that the ARIMA models presented a better sensitivity to the wind speed adjustment and prediction. Based on the ARIMA and Improved Kalman Filter Algorithm, authors proposed a new wind speed prediction model [7].

The primary purpose of this paper is to analyze the time series of wind speed data in order to extract significant statistics and other data features (Trend, seasonality, noise...). Additionally, we propose a predicting wind speed data model based on seasonal ARIMA model (SARIMA). Finally, we will use three different yearly periods to investigate the forecasting. Different evaluation metrics, such as root mean square error (RMSE), mean absolute error (MAE) and Absolute Percentage Error (MAPE), are used to evaluate the prediction performance.

*E-mail: ilham.tyass-etu@etu.univh2c.ma

2 Background Information of study area and data.

The meteorological data used in this study are collected on the site "Abdelkhalak Torres" in Koudia Al Baida Tetouen located at (Latitude: 35° 45' 35.1, Longitude: -5° 41' 19.9"). This wind farm, ordered by the national office of electricity and drinking water (ONEE) in 2000, has a capacity of 50.4 MW. It is the first wind farm realized in Morocco and on the African continent, a repowering and extension program is developing. Currently, more than 206 GWh/year of wind power are generated by 91 wind turbines of 500 kW and 600 kW. This contributes to a reduction of greenhouse gas emissions equivalent to 140,000 tons of CO₂ per year. The wind speed data used in the present study covers three different months in 2018, March, October and July which correspond respectively to the windiest, moderately windy and less windy months. These data are recorded daily at 10-minute intervals at a height of 100 m [8]. The average wind speed in this site is about 10.3 m/s. Figure 1, shows the distribution of wind speed and direction at the studied location. We observe that there is a predominance of wind in the East-South-East (ESE) direction.

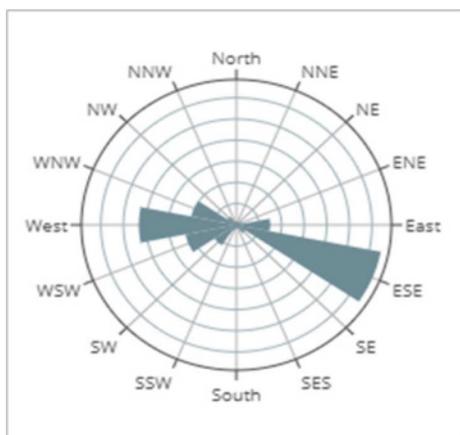


Fig. 1. Wind rose diagram of al koudia al baida [8]

3 Time Series Wind Data diagnostics

3.1. Data mining

In order to prepare the data for processing, the original data were carefully reviewed for errors, so columns that were not used were removed, missing and null values were replaced by the average of the five previous values, and afterwards data were synthesized by calculating the average wind speed for each hour. Thus, each month totals 744 data samples whose evolution is depicted in Figure 2.

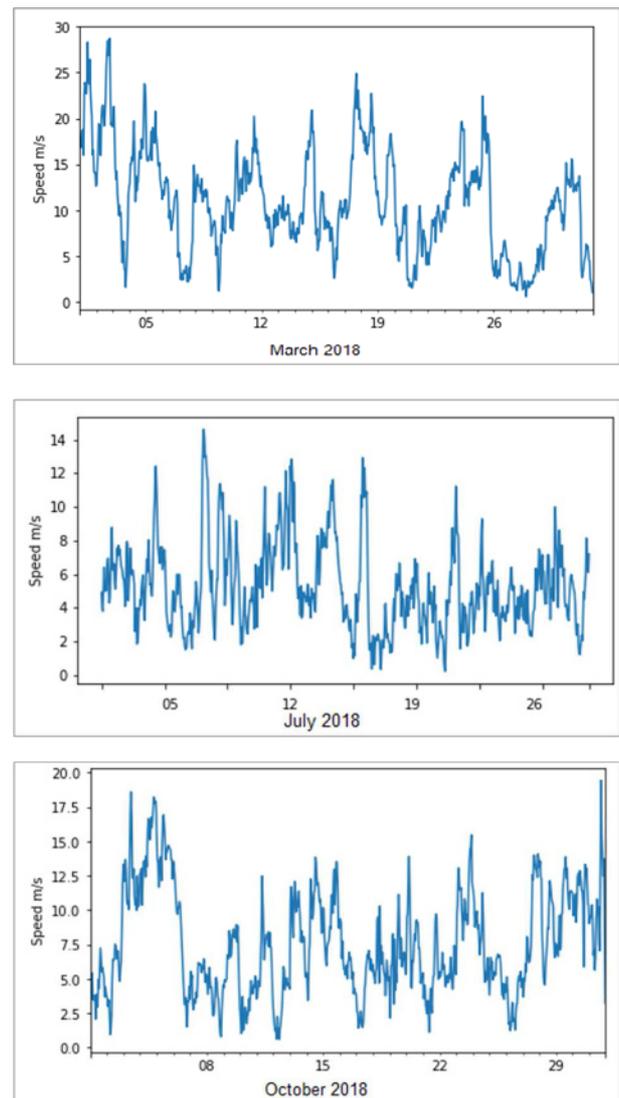


Fig. 2. Time series of wind data visualization for three months

3.2. Time Series Wind Data analysis

Time series are sequences of data points measured over successive time intervals. Their main specificities compared to the most common domains of machine learning are their time dependency and the seasonal behaviors that appear in their evolution [9]. In order to analyze a time series, a first step of decomposition of the series is necessary. Decomposition provides a useful abstract model for better understanding problems during time series analysis and forecasting. A time series (Y_t) is commonly decomposed into:

Trend (T_t): A trend corresponds to a long-term evolution of the series.

Seasonality (S_t): This is the property of a time series displaying periodic behaviors repeating at a constant frequency.

Noise (ϵ_t): Statistical noise is irregular, random and unexplained fluctuations

This decomposition can be additive $Y_t = T_t + S_t + \mathcal{E}_t$ or multiplicative $Y_t = T_t * S_t * \mathcal{E}_t$. It is also possible to combine these two decompositions $Y_t = T_t * S_t + \mathcal{E}_t$

Sometimes we add another component, the cycle C_t which corresponds to a regular repetitive phenomenon (thus predictable) of unknown or changing period. The graph illustrated in figure 3, shows separately the three components obtained following the decomposition of the series, the variance of the series has been stabilized in order to have a stable periodic component over time.

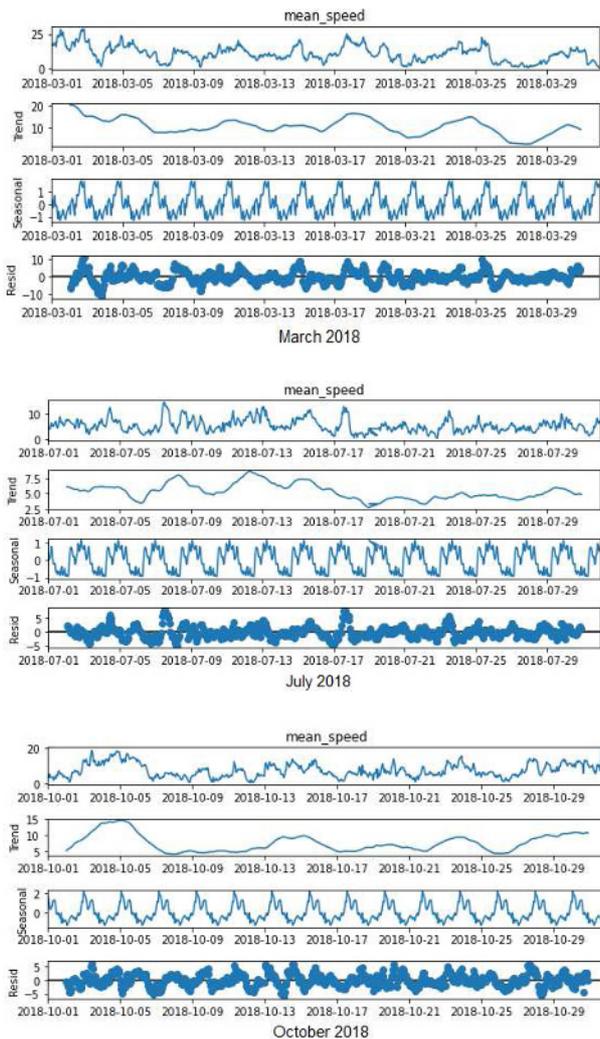


Fig. 3. Decomposition of time series of wind data

The graph presented on figure 3 shows clearly that the wind speed for the three months study is unstable, with some seasonality, also the residuals are interesting, which means that there are important random fluctuations in the data. According to the trend line, the series has no long-term evolution.

4 Wind speed forecasting based on SARIMA model

4.1. The forecasting autoregressive models of time series

Time series prediction models are different from classical models, because they permit to use the series historical values to estimate the future ones, using lagged variables. In other words, to make predictions at time t , we suppose to know the past values at time $(t-1)$. The first models developed for the prediction of time series are univariate models based on the principle of auto-regression. The most popular models of this type are; AR (Auto-Regressive), MA (Moving Average), ARMA (Auto-Regressive Moving Average), and ARIMA (Auto-Regressive Integrated Moving Average) [10].

AR model: Most current models are inspired from the autoregressive principle. In an auto regression model, we forecast the variable of interest using a linear combination of *past values of the variable*. The term auto regression indicates that it is a regression of the variable against itself. Thus, an autoregressive model of order p can be written as:

$$Y(t) = \alpha_0 + \alpha_1 Y(t-1) + \dots + \alpha_p Y(t-p) + \mathcal{E}(t) \quad (1)$$

Changing the parameters $\alpha_1 \dots \alpha_p$, we obtain different time series models. The variance of the error term $\mathcal{E}(t)$ will only change the scale of the series, not the patterns.

MA model: The Moving Average has the same structure as the AR model, but considering the error terms instead of the previous values of the series. The MA model (q) can be defined as follows:

$$Y(t) = \theta_0 + \theta_1 \mathcal{E}(t-1) + \dots + \theta_p \mathcal{E}(t-p) + \mathcal{E}(t) \quad (2)$$

ARMA model (p, q): Combines both processes AR (p) and MA (q) by considering both the error terms and the previous values of the series:

$$Y(t) = Y_0 + \sum_{i=1}^p \alpha_i Y(t-i) + \sum_{i=1}^q \theta_i \mathcal{E}(t-i) + \mathcal{E}(t) \quad (3)$$

ARIMA model: For non-stationary time series, the ARIMA (p, d, q) model is more appropriate. It consists in applying the ARMA model on the series transformed by differentiation of order d , calculating d times the differences between the consecutive observations.

But, ARIMA does not support seasonality. The Seasonal Autoregressive Integrated Moving Average, SARIMA, is an extension of ARIMA that explicitly supports the seasonal component. Seasonality is treated as an ARIMA itself, with terms similar to the non-seasonal components of the model by adding the parameter S which is the period and the parameters P, D and Q which correspond to the factors p, d and q of the differentiated series times [11]. The parameters p, d, q, S, P, D and Q are determined by optimizing the Akaike information criteria (AIC). This criteria measures the compromise between the complexity of a model and its fitting quality [12]. There are also neural network models and deep learning model that use delayed predictors, especially in the Recurrent

Neural Network (RNN) family, such as the Long and Short and Term Memory (LSTM) [13] and Gate Recurrent Unit (GRU) networks [14].

4.2. Application of SARIMA model to forecast the hourly average speed

4.2.1 The parameters selection

The aim of this phase is to find the optimal set of parameters that achieves the best performance for the regression model. Primarily, the data is prepared by a degree of differencing in order to make it stationary.

The optimal combinations of SARIMA parameters that give the lowest Akaike Information Criteria (AIC) are:

For March:

SARIMA (1, 0, 1) x (1, 0, 0, 12)₁₂ –AIC= 2369.72

For July:

SARIMA (1, 0, 1) x (0, 1, 1, 12)₁₂ –AIC= 2504.64

For October:

SARIMA (1, 0, 1) x (0, 1, 1, 12)₁₂ –AIC= 3011.65.

Figure 4 illustrates the model diagnostics obtained for each month.

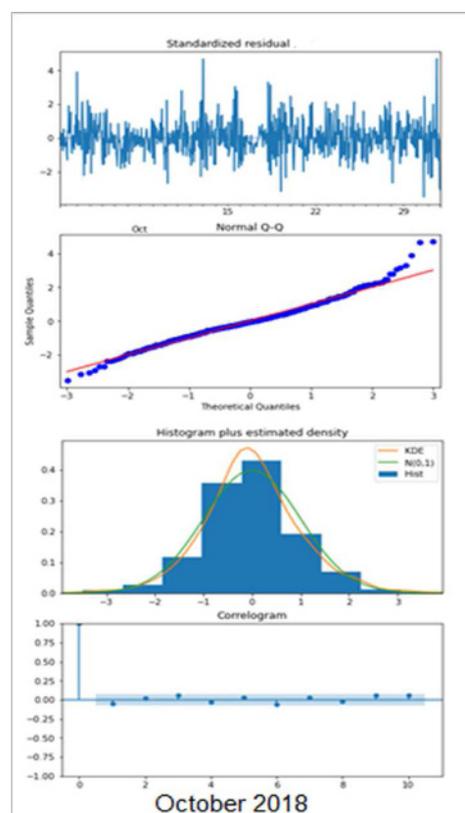
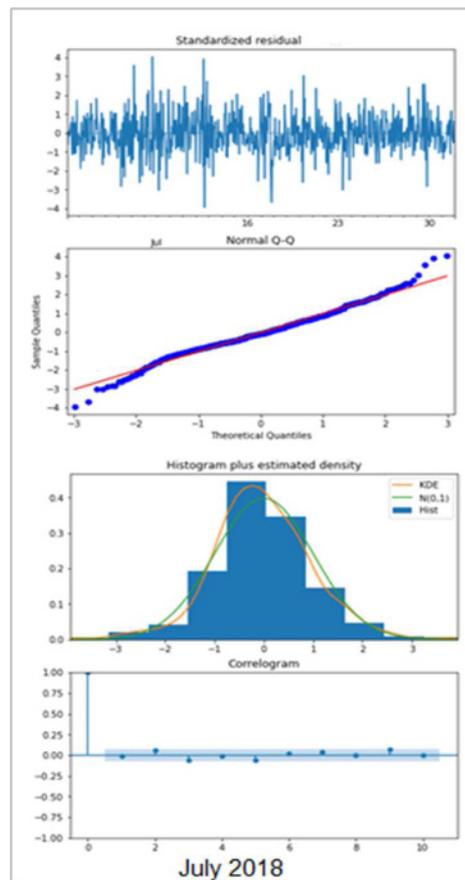
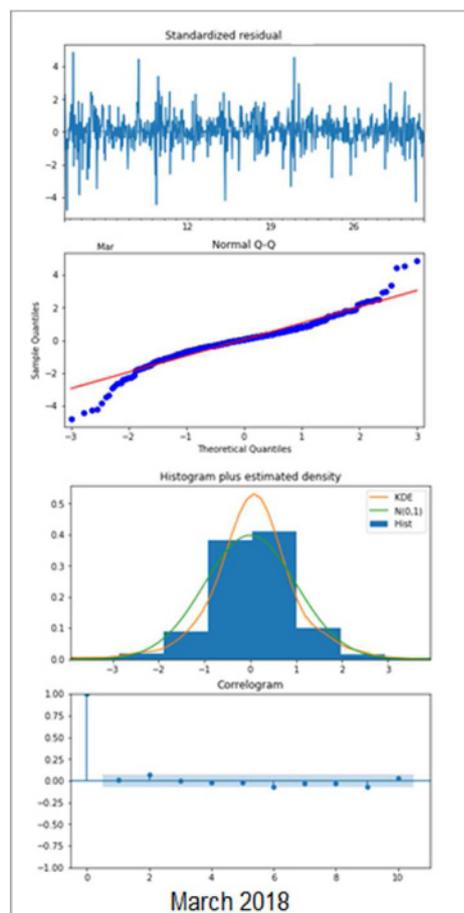


Fig 4. The model diagnostics

The model diagnostics reveals that the model residuals are normally distributed. Moreover, the time series residuals have low correlation with lagged versions of itself. The line of KDE (Kernel Density Estimator) follows the N (0, 1) line (The standard notation for a normal distribution with mean “0” and standard deviation of “1”), this is another indication that the residuals are normally distributed. The quantile-quantile (Q-Q) plot, assesses if the data came from some theoretical distribution, it shows that the ordered distribution of residuals follows the linear trend in the middle of the graph, but they curve at the ends, which means that the data have several extreme values. The Correlogram shows that the time series residuals have low correlation with lagged versions of itself.

Consequently, the model produces a satisfactory fit that helps to understand the time series data and forecast future values.

4.2.2 Forecasting results

We split the dataset into train and test sets, the train set allows to fit the model, and generates a prediction for each element on the test set. We fix forecasts to start respectively at 24-03-2018, 24-07-2018, 24-10-2018 to the end of the data. A rolling forecast is established, given the dependence of the observations of the previous time steps. We produced one-step ahead forecasts, meaning that forecasts at each point are generated using the full history up to that point. The graph in figure 5 shows the observed values compared to the forecast values. On the whole, the forecasts align well with actual values.

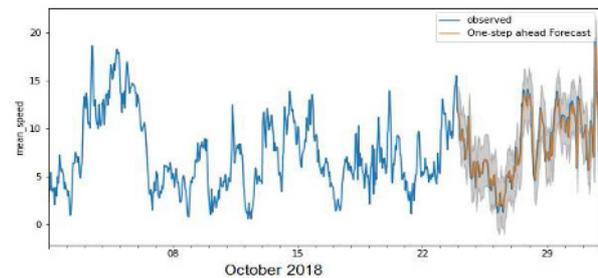


Fig 5. The observed values compared to the forecast values

5 Results analysis

5.1 The statistical errors metrics used to evaluate method performance

The proposed method performance is analyzed based on statistical error metrics: The mean square error (MSE), the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) which are defined as follows:

$$RMSE_{y,\hat{y}} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (4)$$

$$MAE_{y,\hat{y}_i} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (5)$$

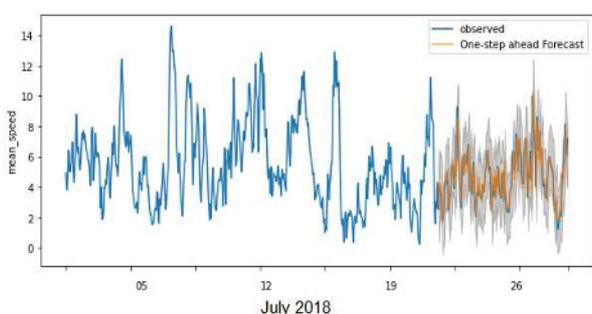
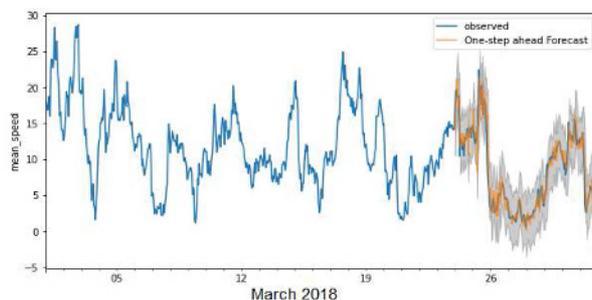
$$MAPE_{y,\hat{y}} = \frac{1}{n} \sum_{j=1}^n \left| \frac{y - \hat{y}}{y} \right| \quad (6)$$

Where “n” is forecasting number, “ŷ” is the forecasting wind speed and “y” is the corresponding actual value.

5.2 Results analysis

The forecasting performance in three different periods is presented in Table 1. Following the interpretation for the typical MAPE value which was explained in [15], the error measures confirm the satisfactory performance of SARIMA models, MAPE values are between 11% and 20% means that the model offers good accuracy forecast. Moreover, compared to several wind speed prediction models cited in the literature based on the classical ARIMA method, the present SARIMA model provides good accuracy values of MAPE comparing to results observed in the bibliography, for one hour ahead forecast [16] [17] [18].

Furthermore, we notice that in march, we recorded the lowest MSE, RMSE and MAE errors, the Seasonal ARIMA model is more appropriate for periods with high velocity. The MAPE values are different from MAE and RMSE because the analyzed wind speed data is unstable. However, it should be noted that MAPE is not the best measurement criterion to be used for the wind speed analysis, because some of wind speed data values are equal or close to 0 m/s. That is why it is better to use



RMSE and MAE instead of MAPE to verify if the model is suitable for wind speed prediction. According to October errors, we conclude that SARIMA forecast quality depends on invariability of the data. When there are sudden changes in the analyzed data, higher errors occur. The reason behind this is the use of previous wind speed values in SARIMA model prediction.

Table 1: SARIMA prediction errors

| | March | July | October |
|-------------|--------|--------|---------|
| MSE | 1.09 | 1.23 | 2.23 |
| RMSE | 1.04 | 1.11 | 1.68 |
| MAE | 0.75 | 0.88 | 1.01 |
| MAPE | 14.99% | 14.12% | 16.02% |

Conclusion

Improving the wind speed prediction accuracy is of great significance for wind energy conversion, SARIMA based models are developed in this paper to predict the wind speed. The model is verified by three wind speed examples data sets. It has been shown that the prediction error for the one hour ahead forecast is less than 16%, which represents a good result regarding the short compilation time which does not exceed 5 minutes. To improve the accuracy of power forecasting, it will be useful to develop a model that integrates the direction of wind flow, as will be investigated in the future work. Also, it is possible to build a hybrid forecasting method based on SARIMA method in order to adopt more seasonal and trending changes and to take care of both the high frequency variation as well as low frequency variation in wind speed.

References

- [1] "IRENA_RE_Capacity_Statistics_2021.pdf."
- [2] A. Dabbaghiyan, F. Fazelpour, M. D. Abnavi, and M. A. Rosen, "Evaluation of wind energy potential in province of Bushehr, Iran," *Renew. Sustain. Energy Rev.*, vol. 55, pp. 455–466, Mar. 2016, doi: 10.1016/j.rser.2015.10.148.
- [3] A. Devis, N. P. M. V. Lipzig, and M. Demuzere, "Should future wind speed changes be taken into account in wind farm development," *Environ. Res. Lett.*, vol. 13, no. 6, p. 064012, May 2018, doi: 10.1088/1748-9326/aabff7.
- [4] B. G. Brown, R. W. Katz, and A. H. Murphy, "Time Series Models to Simulate and Forecast Wind Speed and Wind Power," *J. Clim. Appl. Meteorol.*, vol. 23, no. 8, pp. 1184–1195, 1984.
- [5] J. L. Torres, A. García, M. D. Blas, and A. D. Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Sol. Energy*, vol. 1, no. 79, pp. 65–77, 2005, doi: 10.1016/j.solener.2004.09.013.
- [6] E. Cadenas and W. Rivera, "Wind speed forecasting in the South Coast of Oaxaca, México," *Renew. Energy*, vol. 32, no. 12, pp. 2116–2128, 2007.
- [7] S. Yuan and Y. Shen, "A Wind Speed Prediction Model Based on ARIMA and Improved Kalman Filter Algorithm," *J. Phys. Conf. Ser.*, vol. 1650, p. 032095, Oct. 2020, doi: 10.1088/1742-6596/1650/3/032095.
- [8] "Wind Resource Data for Wind Farm Developments | Vortex FDC," *VORTEX*. <https://vortexfdc.com/>
- [9] W. Velicer and J. Fava, "Time Series Analysis," vol. 2, 2003. doi: 10.1002/0471264385.wei0223.
- [10] C. Liu, S. C. H. Hoi, P. Zhao, and J. Sun, "Online ARIMA Algorithms for Time Series Prediction," The Thirtieth AAAI Conference on Artificial Intelligence, Feb. 2016. <https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12135>
- [11] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal ARIMA approach," *J. Earth Syst. Sci.*, vol. 129, no. 1, p. 149, Jun. 2020, doi: 10.1007/s12040-020-01408-x.
- [12] P. Mantalos, K. Mattheou, and A. Karagrigoriou, "Forecasting ARMA models: a comparative study of information criteria focusing on MDIC," *J. Stat. Comput. Simul.*, vol. 80, no. 1, pp. 61–73, Jan. 2010, doi: 10.1080/00949650802464137.
- [13] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting," *Acta Geophys.*, vol. 67, no. 5, pp. 1471–1481, Oct. 2019, doi: 10.1007/s11600-019-00330-1.
- [14] B. Liu, C. Fu, A. Bielefeld, and Y. Q. Liu, "Forecasting of Chinese Primary Energy Consumption in 2021 with GRU Artificial Neural Network," *Energies*, vol. 10, no. 10, Art. no. 10, Oct. 2017, doi: 10.3390/en10101453.
- [15] R. J. C. Chen, P. Bloomfield, and J. S. Fu, "An Evaluation of Alternative Forecasting Methods to Recreation Visitation," *J. Leis. Res.*, vol. 35, no. 4, pp. 441–454, Dec. 2003, doi: 10.1080/00222216.2003.11950005.
- [16] K. R. Nair, V. Vanitha, and M. Jisma, "Forecasting of wind speed using ANN, ARIMA and Hybrid models," in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Jul. 2017, pp. 170–175. doi: 10.1109/ICICICT1.2017.8342555.
- [17] E. Cadenas, W. Rivera, R. Campos-Amezcuca, and C. Heard, "Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model," *Energies*, vol. 9, no. 2, Art. no. 2, Feb. 2016, doi: 10.3390/en9020109.
- [18] N. H. Hussin, F. Yusof, 'Aaishah Radziah Jamaludin, and S. M. Norrulashikin, "Forecasting Wind Speed in Peninsular Malaysia: An Application of ARIMA and ARIMA-GARCH Models," *Pertanika J. Sci. Technol.*, vol. 29, no. 1, Jan. 2021, doi: 10.47836/pjst.29.1.02.