Journal of Power, Energy, and Control

Research article

Wind Power Prediction Model Using Machine Learning

Paul Waweru, Charles Kagiri, Titus Mulembo*

Mechatronic Department, Dedan Kimathi University of Technology, Private Bag -10143, Dedan Kimathi, Kenya
*Correspondence: titus.mulembo@dkut.ac.ke

https://doi.org/10.62777/pec.v1i1.6

Received: 2 April 2024
Revised: 22 April 2024
Accepted: 25 April 2024
Published: 29 April 2024

Abstract: Before installing a wind turbine, it's essential to conduct wind power forecasting to gauge the effectiveness of the wind power initiative. Conventionally, wind speed measurements have been conducted instantaneously between various points. These measurement points solely indicate the locations where wind turbines will be positioned. However, these locations might exhibit reduced wind speeds, potentially making them less suitable for the optimal placement of the wind turbine. To address location challenges, we suggest conducting wind power predictions in areas where wind measuring instruments are yet to be installed. The study relies on the instantaneous measurements already performed at the site set up at the Dedan Kimathi University of Technology. To this end, a wind power forecasting model has been created. Real-time data from the site was gathered via a wireless sensor node utilising the Internet of Things (IoT). Additionally, a machine learning prediction model based on time series analysis was developed. Our forecasts were moderately aligned with the testing values, showing seasonality throughout the year. Therefore, the developed machine learning model captured the underlying patterns, trends, and seasonality in the wind data, making its forecasts reliable.

Keywords: IoT, forecast, wind power, machine learning, SARIMAX, ARIMA

1. Introduction

Investments in wind energy are increasing [1]. Conventionally, wind turbines are placed where wind speed measurements are conducted [2]. However, this location might not have the highest wind speed across the surveyed area. Hence, it is necessary to forecast wind speeds in the surrounding areas, as they might have the potential to generate higher wind speeds. Moreover, wind power experiences fluctuations that are difficult to control [3]. Decreasing the variance involves consolidating power generation from numerous wind farms. An appropriate model first predicts wind speed. The forecasted wind speed is employed to estimate the anticipated wind power production for a particular wind farm, and the prediction outcome for a wind farm can also be utilised to predict regional output [4].

Several researchers have employed various wind prediction techniques [5], such as the Wind Atlas Analysis and Application Program [6]. It uses the linear atmospheric model to extend wind climate information within a specific area, considering factors such as terrain features and surface roughness [7]. This model applies the linear aspects of the Navier-Stokes equations to compute wind velocities at various points. Wind atlas models rely on simplified assumptions and extrapolation techniques, leading to potential inaccuracies in predicting complex local wind patterns. Another prediction technique is computational fluid dynamics (CFD), which uses fluid dynamic equations to quantify wind climate [8]. CFD simulations require significant computational resources and time, making it challenging to efficiently perform large-scale and long-term wind energy assessments. LiDAR technology (light detection and ranging) is also used for prediction [9]. It uses the pulse from a laser to collect measurements, which can then be used to create 3D models and maps of objects and environments [10]. However, LiDAR devices have a small measurement range, resulting in limited spatial coverage.

To overcome the existing gaps, this work proposes to employ machine learning techniques to improve forecasts. To this end, a model is developed that analyses wind data to make accurate predictions of future wind power generation using the ARIMA and SARIMAX models.

The rest of this paper is structured as follows: Section 2 presents the methodology, highlights the wind data collection setup, and outlines the wind energy prediction methods employed in this study, ARIMA and SARIMAX. Section 3 provides the results and discussion of the model, including the selection of parameters, data cleaning procedures, and analytical techniques. Finally, Section 4 is the conclusion of the paper.

2. Methods

Data was collected using custom-built wireless sensor nodes in two stations located one hundred meters apart. The sensor nodes were fabricated using an Atmega 328P microcontroller and Xbee module, which communicated with each other via the I2C protocol. The sensor nodes were equipped with a ready-made anemometer and wind vane to measure speed and direction. The collected data was then transmitted wirelessly to a Raspberry Pi. The Raspberry Pi received the data from the two stations, formatted it, and sent it to an email. A wind power prediction model developed in this study is also presented.

2.1. Wireless Sensor Node Fabrication

The essential embedded systems and the assembly of the complete system device, forming a wireless sensor node aimed at collecting and transmitting data, were set up. Towers standing at a height of sixty meters at Dedan Kimathi University of Technology were used to install the necessary instruments for collecting wind data. The wireless data system was constructed based on the Atmega328P microcontroller due to its ease of prototyping, wireless communications through Bluetooth, and its wide range of libraries. This setup also had an Xbee radio capable of transmitting data via the ZigBee IEEE 802.15.4 protocol to a central station a few meters away [11], as shown in Figure 1. The two stations were set up with a distance of about one hundred meters between them. Each station had:

- Wind sensors for data collection;
- Arduino ATmega328P microcontroller board. It acts as the interface between wind sensors, reading the digital signals produced by the sensors, scheduling sensor readings, and coordinating wireless communication;
- Xbee is used to transmit data to the Raspberry Pi, which reads and formats the data and sends it to the receiving station.

A ready-made cup anemometer and a wind vane were used to increase the accuracy of the data collected. The system in this study was based on an Atmega328P-AU microcontroller. The system also had an XBee radio capable of transmitting data via the ZigBee IEEE 802.15.4 protocol to a central station one hundred meters away [12]. The system was then programmed to understand the signals from sensors, save, display on the screen, and transmit wind data to the email for presentation. This program was written in a C/C++ development environment. The sensors used needed to communicate via the I^2C protocol [13]. Sensor pins and addresses were first defined. Their addresses were assigned, and the necessary sensor libraries were integrated into the program. The initial configurations were set up within a function called only once. Sensor interrupts were specified, internal resistors were activated, sensors were initialised, and serial communication was activated. The iteration process was managed within the primary loop function. Throughout the iteration, the program monitored the passage of time. Sensors were programmed to transmit data via serial communication immediately upon receiving it. Calibration details for the sensors were also accounted for during this process. The wind vane sensor was adjusted to record the wind vane tail direction and transmit the data serially. The declination angle of the wind vane was accounted for by calibration using the tunnel [14]. The speed sensors provided instantaneous revolutions per second. The system was programmed to gather data samples every five seconds and transmit the averaged samples after sixty seconds.

2.2. Wind Data Collection

The collected data from the two stations was transmitted, with each station having the speed and direction of the wind. The test IEEE 802.15.4 sink node, positioned a hundred meters away, excellently received the data every sixty seconds. The collected data represents the actual values of wind speed and direction. In time series, they are referred to as observed values. It includes the combined effects of the trend, seasonality, and any random fluctuations or noise in the data. Figure 1 shows the flowchart of the data collection setup.

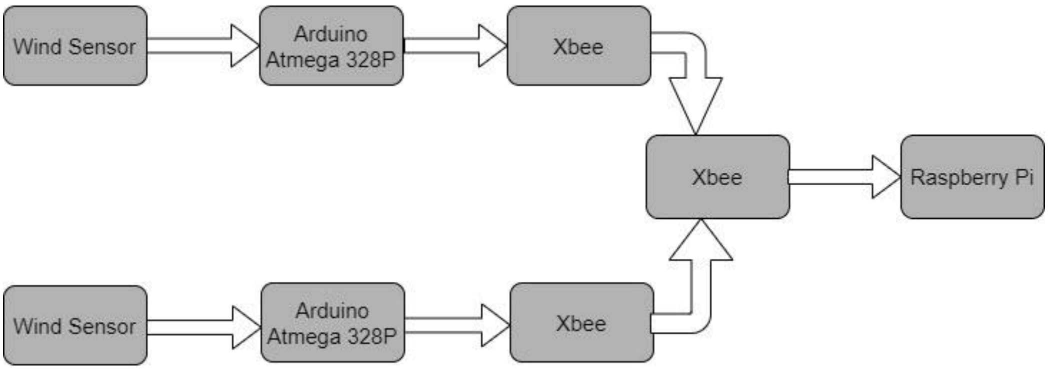


Figure 1. Flowchart of data collection setup.

The data from stations 1 and 2 was recorded, including date, time, speed, and direction. Table 1 shows a sample of data taken over five minutes. Our data was collected for one year, from January 2018 to 2018.

Table 1. Data sample from the two stations.

Date and time	Station number	Speed (m/s)	Direction (degrees)
1/10/2018 0:00	1	7.24	190.3
1/10/2018 0:00	2	1.49	273.2
1/10/2018 0:01	1	6.12	190.5
1/10/2018 0:01	2	1.09	238.6
1/10/2018 0:02	1	8.61	187.8
1/10/2018 0:02	2	0.64	248.7
1/10/2018 0:03	1	7.16	191.3
1/10/2018 0:03	2	1.13	247.3
1/10/2018 0:04	1	6.76	182.3
1/10/2018 0:04	2	1.93	246.6
1/10/2018 0:05	1	6.92	191.7
1/10/2018 0:05	2	0.36	283.0

2.3. Wind Power Prediction Model Development

The data was first processed by aggregating the wind speed and direction values. The data was divided into two sets: one for training and the second for testing the model. The training set was utilised to develop the predictive model, while the testing set was employed to assess the model's performance. A prediction model was developed using the Python programming language. The autoregressive integrated moving average (ARIMA) model was selected as a machine learning algorithm. The algorithm was chosen since we had sufficient historical data to accurately capture the underlying patterns and estimate model parameters. ARIMA models are represented using the notation (p, d, q). p is the autoregressive order, representing the number of past observations utilised as predictors. d is the differencing order, representing the frequency of differencing applied to attain stationarity. q is the moving average order, representing the count of past forecast errors employed in the prediction equation. By decomposing the time series into the three components, the individual contributions of trend, seasonality, and noise to the overall behaviour of the wind data were analysed. This decomposition was used to forecast future wind power based on historical patterns and identify any changes or anomalies in the data.

The trend component captures the long-term behaviour of the time series [15]. The component indicates any long-term changes in wind power, such as shifts in prevailing winds over time. The seasonality component accounts for the repetitive patterns or cycles in the time series that occur at fixed intervals within a year [16]. The repetitive patterns were captured and reflected in the model's predictions by incorporating a seasonality component in the ARIMA model. Resid or residual represents the difference between the predicted and observed values [17]. It accounts for the unexplained variation or the remaining noise the model could not capture.

AIC (Akaike Information Criterion) statistical measure was used for model selection and comparison since different models were being compared. It provided a way to evaluate the adequacy of the model to the data of different models while considering their complexity. The model with the lowest AIC is the best fit, considering its ability to explain the data and its complexity. From Table 2, the output suggests that ARIMA (0,1,0) x (0,1,0,12) was chosen since it yields the lowest value of 2.0.

Table 2. Arima forecasting.

ARIMA model	AIC value
ARIMA (0, 0, 0) x (0, 0, 1, 12) 12	AIC:4.0
ARIMA (0, 0, 0) x (0, 1, 0, 12) 12	AIC:2.0
ARIMA (0, 0, 0) x (0, 1, 1, 12) 12	AIC:4.0
ARIMA (0, 0, 0) x (1, 0, 0, 12) 12	AIC:4.0
ARIMA (0, 0, 0) x (1, 0, 1, 12) 12	AIC:6.0
ARIMA (0, 0, 0) x (1, 1, 0, 12) 12	AIC:4.0
ARIMA (0, 0, 0) x (1, 1, 1, 12) 12	AIC:6.0
ARIMA (0, 0, 1) x (0, 0, 1, 12) 12	AIC:6.0
ARIMA (0, 0, 1) x (0, 1, 0, 12) 12	AIC:4.0
ARIMA (0, 0, 1) x (0, 1, 1, 12) 12	AIC:6.0
ARIMA (0, 0, 1) x (1, 0, 0, 12) 12	AIC:6.0
ARIMA (0, 0, 1) x (1, 0, 1, 12) 12	AIC:8.0
ARIMA (0, 0, 1) x (1, 1, 0, 12) 12	AIC:6.0
ARIMA (0, 0, 1) x (1, 1, 1, 12) 12	AIC:8.0
ARIMA (0, 1, 0) x (0, 0, 1, 12) 12	AIC:4.0
ARIMA (0, 1, 0) x (0, 1, 0, 12) 12	AIC:2.0
ARIMA (0, 1, 0) x (0, 1, 1, 12) 12	AIC:4.0
ARIMA (0, 1, 0) x (1, 0, 0, 12) 12	AIC:4.0

The model was then trained and evaluated. The seasonality technique was refined to improve the model's accuracy. The technique incorporated seasonal differencing and seasonal orders into the ARIMA model to form SARIMAX. The SARIMAX forecasting model modifies the ARIMA model to include exogenous variables, which is appropriate for forecasting time series that are influenced by external factors. It accounts for both the temporal dependencies and the seasonal patterns associated with wind data, enabling accurate predictions. Various parameters (seasonality, trend, and noise) were integrated, and exogenous variable parameters were incorporated to form the SARIMAX function.

3. Results and Discussion

This study aimed to predict wind speed and direction using time series machine learning models, specifically ARIMA and SARIMAX. The ARIMA model was first applied to the dataset, which consisted of historical wind speed and direction data. The model successfully captured the data patterns and exhibited moderate predictive performance for both wind speed and direction. However, ARIMA did not account for the potential effects of external factors on wind patterns. The SARIMAX model, which incorporates exogenous variables, was employed to address this constraint. By including additional weather variables, such as topology, as exogenous input, the SARIMAX model improved the prediction accuracy.

3.1. Data Analysis and Visualisation

Data analysis was carried out to determine the best model for the data. Time series indexing was used for efficient retrieval, manipulation, and analysis of data based on specific time intervals of one month. The downsampling and aggregation method for indexing was chosen since we dealt with large datasets. The technique uses averaging or summing values within fixed time intervals without sacrificing critical information. Due to the complexity of the data and for better visualisation, the average wind speeds and directions for that month were used. The beginning of each month was used as the timestamp.

Statistical measures provided essential insights into the characteristics of the data. It helped to understand the characteristics and patterns within the data. The mean, range, and dispersion of the data were identified. This helped visualise any outliers, trends, or seasonal patterns that existed. These measures enhanced the accuracy and reliability of predictions and contributed to a deeper understanding of the underlying patterns and dynamics in the data. Our data did not have significant outliers, as shown in Table 3, making the data reliable for training the model.

Table 3. Statistical analysis of data.

	Station 1		Station 2	
	Speed (m/s)	Direction (degrees)	Speed (m/s)	Direction (degrees)
Count	115663	163845	73313	155659
Mean	6.50	176.46	4.12	167.64
Standard Deviation	3.92	53.22	3.75	62.79

To investigate the wind data further, the average wind direction for each month from stations 1 and 2 was plotted against the start of every month. The plotted wind direction is seen in Figure 2 (a). The average wind speed from stations 1 and 2 for each month was plotted against the start of every month, as shown in Figure 2 (b). Different wind speeds and directions were recorded in different months of the year due to different seasons.

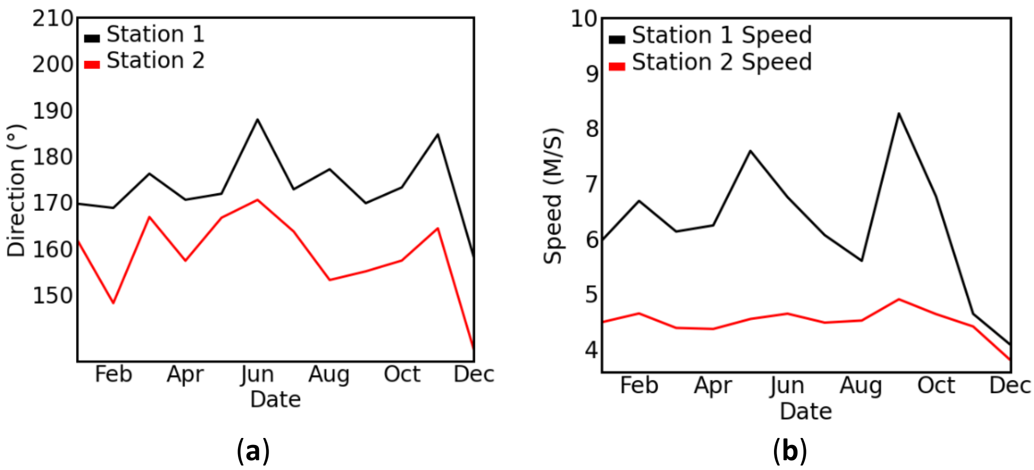


Figure 2. Plotted wind data: (**a**) wind direction; (**b**) wind speed.

3.2. Wind Power Prediction using ARIMA

Using the time series decomposition method, the model was decomposed into three distinct components: trend, seasonality, and resid.

3.2.1. Trend Analysis

Long-term changes in wind power, such as shifts in prevailing winds over time, were analysed and plotted to investigate trends in the wind data. The trend component captured the long-term behaviour of wind power. Some unique patterns appear when the wind speed and wind direction data are plotted against different times of the year due to different seasons, as shown in Figures 3 (a) and 3 (b).

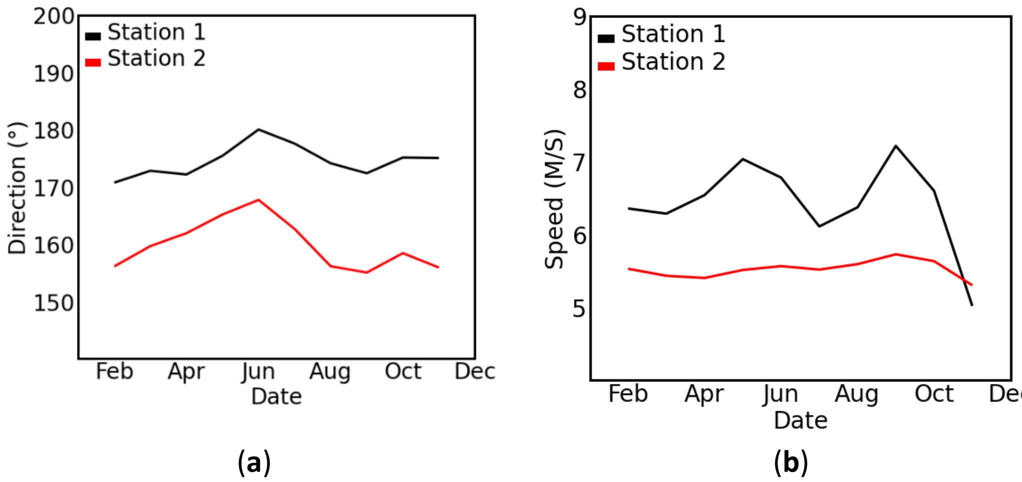


Figure 3. Trend: (**a**) wind direction; (**b**) wind speed.

3.2.2. Seasonality Analysis

The repetitive patterns or cycles in the time series that occur at fixed intervals within a year were analysed and plotted for better visualisation to investigate seasonality in the wind data. Figure 4 (a) and Figure 4 (b) show repetitive patterns in the wind direction and wind speed data, respectively, at different months of the year. Seasonality corresponds to variations in wind power due to different seasons, such as prevailing winds shifting between different seasons at different months. This shows that the model accurately represents the underlying dynamics of the data, increasing confidence in our predictions.

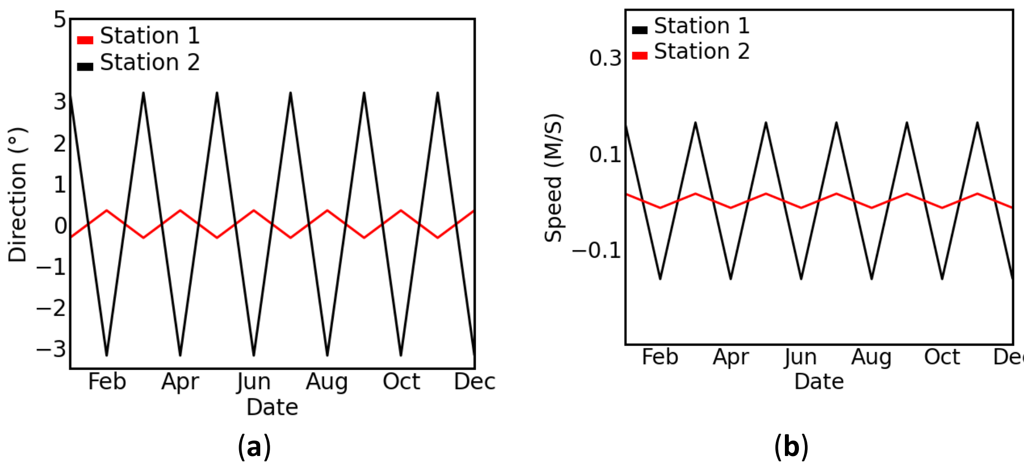
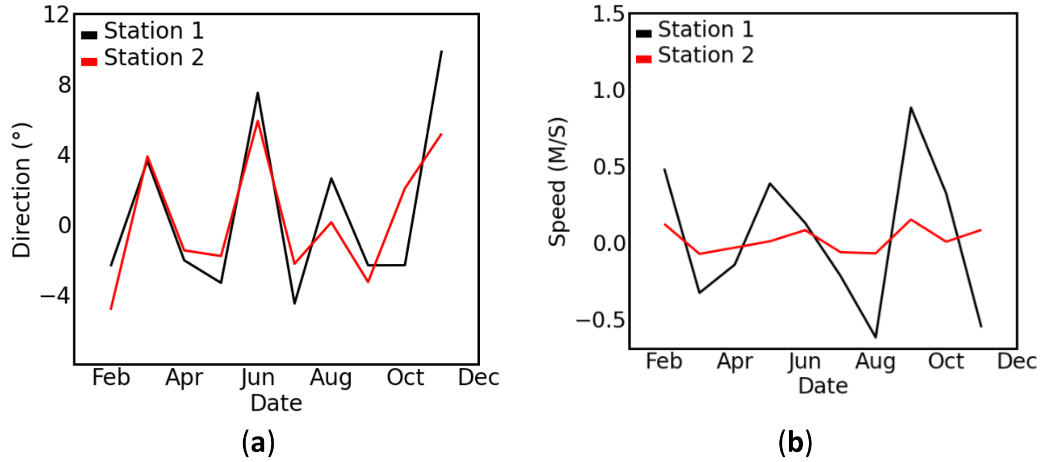


Figure 4. Seasonality: (**a**) wind direction; (**b**) wind speed.

3.2.3. Resid Analysis

Resid represents the difference between the predicted and observed values. It accounts for the unexplained variation or the remaining noise the model could not capture. Wind data were analysed and plotted to investigate the difference between the predicted and observed values. It was analysed to determine noise and variability in the wind data. From Figures 5 (a) and 5 (b), the residuals are closely distributed; hence, the model adequately accounts for the mean and variability of the data points. Thus, our model is forecasting correctly.

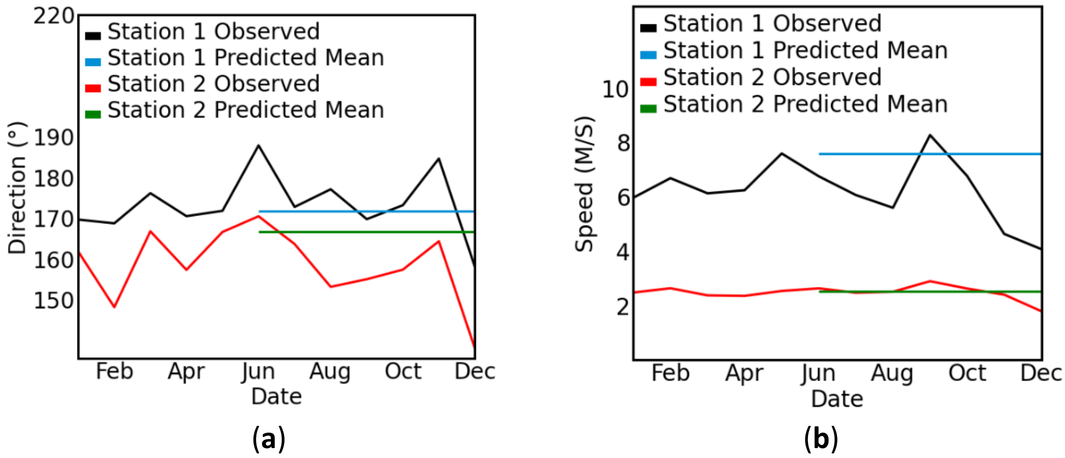
Figure 5. Resid: (**a**) wind direction; (**b**) wind speed.



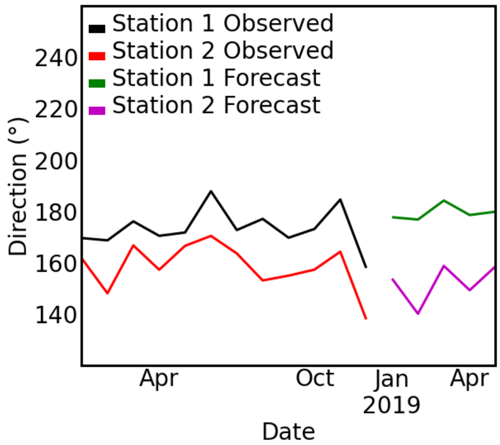
3.3. Wind Power Prediction using SARIMAX

Figure 6 (a) provides an estimate of the expected wind direction, and Figure 6 (b) provides an estimate of the predicted wind speed and wind direction, respectively. Figure 6 (a) and Figure 6 (b) showed the significance of the differences between predicted and observed or actual values from the training set. The predicted mean in SARIMAX represents the central tendency of the predictions generated by the trained model. It indicates the anticipated average wind energy production at a specific future time.

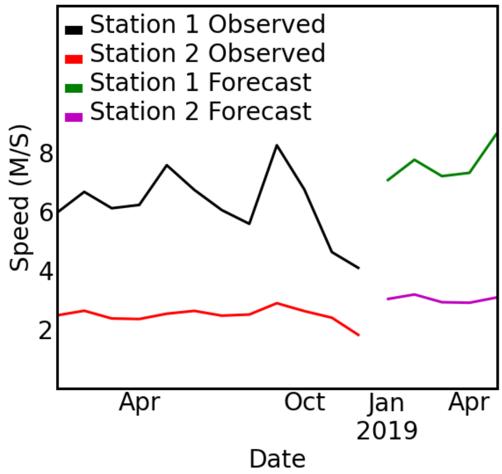
Figure 6. SARIMAX prediction (mean): (**a**) direction; (**b**) speed.



Figures 7 (a) and 7 (b) represent the direction and speed forecasts for wind, respectively. Forecast refers to the predicted value of our wind speed and direction. The line plots show the observed values compared with forecast predictions. Our projections correspond to the actual values, revealing seasonality across the year. This indicates that our model captures the underlying patterns and trends in the wind data. This provides confidence in the reliability of our forecasts.



(a)



(b)

Figure 7. SARIMAX prediction (forecast): (**a**) direction; (**b**) speed.

4. Conclusions

This research Led to the development of a time series-based machine learning model capable of forecasting wind speed and direction. It was demonstrated that wind speed and direction predictions using time series machine learning, specifically ARIMA and SARIMAX, were relatively close to the testing set. The ARIMA model was initially applied to the dataset, which consisted of historical wind speed and direction data. The model successfully captured the temporal patterns and exhibited moderate predictive performance for both wind speed and direction. Then, the SARIMAX model was employed, which incorporated exogenous variables, improving the prediction accuracy. This research holds significant relevance as contemporary forecasting methods are in demand, and computational time is constrained in practical applications.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AIC	*Akaike information criterion*
ANN	*Artificial neural network*
ARIMA	*Autoregressive integrated moving average*
CFD	*Computational fluid dynamics*
IoT	*Internet of things*
RMSE	*Root mean squared error*
SARIMAX	*Seasonal autoregressive integrated moving average with exogenous factor*
SODAR	*Sound detection and ranging*
WASP	*Wind atlas analysis and application programme*

References

[1] Wind Europe, 'Wind Energy in Europe: Scenarios for 2030', Brussels, Belgium, 2017.

[2] P. Gipe, *Wind power : renewable energy for home, farm, and business*, Rev. and Expanded ed. White River Junction, Vermont, USA: Chelsea Green Publishing Company, 2004.

[3] B. Fox *et al.*, *Wind Power Integration: Connection and System Operational Aspects*. Institution of Engineering and Technology, 2014. doi: 10.1049/PBRN014E.

[4] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, 'Current methods and advances in

forecasting of wind power generation', *Renew Energy*, vol. 37, no. 1, pp. 1–8, Jan. 2012, doi: 10.1016/j.renene.2011.05.033.

[5] M. Lackner, A. Rogers, and J. Manwell, 'Uncertainty Analysis in Wind Resource Assessment and Wind Energy Production Estimation', in *45th AIAA Aerospace Sciences Meeting and Exhibit*, Reston, Virigina: American Institute of Aeronautics and Astronautics, Jan. 2007. doi: 10.2514/6.2007-1222.

[6] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, 'A Critical Review of Wind Power Forecasting Methods—Past, Present and Future', *Energies (Basel)*, vol. 13, no. 15, p. 3764, Jul. 2020, doi: 10.3390/en13153764.

[7] A.J. Bowen and N.G. Mortensen, 'Exploring the limits of WAsP the wind atlas analysis and application program', in *European Union wind energy conference*, Göteborg, Sweden, May 1996, pp. 584–587.

[8] L. A. M. Tossas and S. Leonardi, 'Wind Turbine Modeling for Computational Fluid Dynamics: December 2010 - December 2012', Golden, CO (United States), Jul. 2013. doi: 10.2172/1089598.

[9] M. Courtney, R. Wagner, and P. Lindelöw, 'Commercial lidar profilers for wind energy. A comparative guide'. 2008. Accessed: Mar. 21, 2022. [Online]. Available: https://www.nrgsystems.com/assets/resources/Commercial-Lidar-Profilers-for-Wind-Energy-Whitepaper..pdf

[10] S. Emeis, M. Harris, and R. M. Banta, 'Boundary-layer anemometry by optical remote sensing for wind energy applications', *Meteorologische Zeitschrift*, vol. 16, no. 4, pp. 337–347, Aug. 2007, doi: 10.1127/0941-2948/2007/0225.

[11] K. F. Haque, A. Abdelgawad, and K. Yelamarthi, 'Comprehensive Performance Analysis of Zigbee Communication: An Experimental Approach with XBee S2C Module', *Sensors*, vol. 22, no. 9, p. 3245, Apr. 2022, doi: 10.3390/s22093245.

[12] B. Champaty, S. K. Nayak, G. Thakur, B. Mohapatra, D. N. Tibarewala, and K. Pal, 'Development of Bluetooth, Xbee, and Wi-Fi-Based Wireless Control Systems for Controlling Electric-Powered Robotic Vehicle Wheelchair Prototype', in *Robotic Systems*, IGI Global, 2020, pp. 1048–1079. doi: 10.4018/978-1-7998-1754-3.ch052.

[13] I. Calvo *et al.*, 'Design and Performance of a XBee 900 MHz Acquisition System Aimed at Industrial Applications', *Applied Sciences*, vol. 11, no. 17, p. 8174, Sep. 2021, doi: 10.3390/app11178174.

[14] S. Milrad, *Synoptic Analysis and Forecasting: An Introductory Toolkit*. Elsevier, 2018. doi: 10.1016/C2015-0-05604-0.

[15] B. Kenmei, G. Antoniol, and M. di Penta, 'Trend Analysis and Issue Prediction in Large-Scale Open Source Systems', in *2008 12th European Conference on Software Maintenance and Reengineering*, IEEE, Apr. 2008, pp. 73–82. doi: 10.1109/CSMR.2008.4493302.

[16] T. N. Stockdale *et al.*, 'Understanding and Predicting Seasonal-to-Interannual Climate Variability - The Producer Perspective', *Procedia Environ Sci*, vol. 1, pp. 55–80, 2010, doi: 10.1016/j.proenv.2010.09.006.

[17] M. Lu and Y. Chen, 'Improved Estimation and Forecasting Through Residual-Based Model Error Quantification', *SPE Journal*, vol. 25, no. 02, pp. 951–968, Apr. 2020, doi: 10.2118/199358-PA.