

# Prediction of Monthly Wind Velocity Using Machine Learning

Ahmed T. Al-Hasani<sup>1</sup>, Ban Jaber Ednan Al-Juburi<sup>2</sup>, Fouad Hussein Hasan<sup>1</sup>, Ali J. Ramadhan<sup>3\*</sup> and Adelaja Oluwaseun Adebayo<sup>4</sup>

<sup>1</sup>AL-Furat Al-Awsat Technical University, Najaf, Iraq

<sup>2</sup>General Directorate for Education in Najaf, Najaf, Iraq

<sup>3</sup>University of Alkafeel, Najaf, Iraq

<sup>4</sup>Lagos State University, Lagos, Nigeria

**Abstract.** The utilization of non-renewable energy resources necessitates the power sector's adoption of alternative energy sources, including photovoltaic and wind power generation systems. This academic investigation utilizes two machine learning methodologies, in particular, the study utilizes the random forest and support vector machine algorithms. to conduct its analyses. predict the velocity of the wind in the Diyala governorate of Iraq for the subsequent time interval. This is achieved solely by utilizing historical monthly time series data as input predictors. The three performance metrics employed encompass the coefficient of assurance (R<sup>2</sup>), root cruel square mistake (RMSE), and cruel outright blunder (MAE). The findings demonstrate that utilizing a lag of 12 months in the time series data (the maximum lag duration tested) as input predictors leads to the most accurate predictions in terms of performance. However, the prediction performance of the two algorithms used was almost similar (RF's RMSE, MAE, and R<sup>2</sup> were 0.237, 0.180, and 0.836, while for SVM were 0.223, 0.171, and 0.836). The capacity to anticipate wind speed constitutes a paramount advantage to Iraq, given its current predicament in the electric power industry, and this has the potential to enable stakeholders to forecast oversupply or undersupply and implement pre-emptive measures.

## 1. Introduction

Wind, a cost-effective and renewable energy resource, holds significant importance in reducing reliance on non-renewable (fossil) fuels and mitigating carbon dioxide emissions. Consequently, it plays a vital part in alleviating the adverse impacts of anthropogenic climate change [1]. The power of wind is defined by its speed, and wind speed is influenced by various factors, such as the pressure gradient, jet streams, and regional weather patterns. So, one of the challenges encountered in implementing wind energy is the critical variances within the accessibility of wind control [2]. Consequently, the precise prediction of changes in the availability of wind resources is indispensable to the strategic planning and functioning of energy systems [3].

\* Corresponding author: [ali.j.r@alkafeel.edu.iq](mailto:ali.j.r@alkafeel.edu.iq)

The field of machine learning presents a promising multidisciplinary domain in which novel techniques for wind speed can be effectively utilized [4]. Consequently, in the literature, a variety of machine learning-based tools have been explored and analyzed by various scholars aimed at predicting wind speed. For example, the utilization of artificial neural networks (ANNs) for the development of time series models meant for predicting wind speed/power are presented by [5]–[8]. The present study employs historical data on wind speed and power to predict their prospective values. In the study [9], the authors developed three distinct models employing machine learning algorithms to estimate wind speed and direction. These models were constructed utilizing temperature, pressure, relative humidity, and local time as input variables.

Specifically, the considered machine learning algorithms were multilayer feed-forward ANN, SVM, and adaptive neuro-fuzzy inference system (ANFIS). Reference [10] presents a comparative analysis between ANN and SVM in predicting wind power generation. The two machine learning algorithms were trained using datasets consisting of wind power generated from 10 wind turbines, as well as wind speed and direction data spanning a period of 2.5 years. The aim of this training was to predict wind power generation for the day ahead accurately. The study of [11] employs a range of meteorological parameters, including but not limited to the time of day, year, geographical coordinates (latitude and longitude), air temperature, wind direction, humidity, and air pressure, as input variables to the ANN algorithm for the purpose of predicting the speed of the wind. In accordance with the findings reported in [12], the utilization of input data pertaining to elevated wind speed at the height of 50 meters, wind direction at 63 meters, and wind direction at 48 meters can facilitate the prediction of wind speed at the height of 65 meters through the application of three distinct ANN models.[29] These models include a multi-layer perceptron neural arrange, a spiral premise work (RBF) neural arrange, and a categorical relapse. In a related academic request[13], Lavenberg-Marquardt (LM) learning algorithm was employed to construct an ANN for predicting wind speed at a specific site using data from a different location. The work published in [14] entails the construction of an ANN model that leverages the supervisory control and data acquisition (SCADA) method to incorporate wind speed, nacelle orientation, yaw error blade pitch angle, and ambient temperature as inputs to predict wind power generation.[28]

The scholarly objective of this research is to establish a robust predictive modelling framework for monthly wind speed in the Diyala governorate of Iraq. This endeavour employs two sophisticated machine-learning methodologies, namely Random Forest (RF) and Support Vector Machine (SVM). The precise forecasting of monthly wind speeds in this locale holds the potential to significantly optimise the administration and functioning of the electrical system. Consequently, it can facilitate the seamless integration of additional wind vitality sources into the control framework.

## **2. Wind Speed Data Used**

The wind speed records for the Diyala governorate in Iraq were obtained from the ERA5 atmospheric reanalysis dataset, which is a comprehensive collection of meteorological data compiled by the European Centre for Medium-Range Weather Forecasts (ECMWF). These records cover the period from 1981 to 2022 and were measured at a height of 10 meters above ground level. The ERA5 dataset provides global coverage with a spatial resolution of  $0.25^{\circ} \times 0.25^{\circ}$  (approximately 30 km) and captures atmospheric conditions across 137 levels, ranging from the Earth's surface up to an altitude of 80 km. Access to the ERA5 data is available from 1979 up to three months prior to the present real-time period.

### 3. Methods

#### Random Forest (RF):

A Random Forest (RF) constitutes an ensemble machine-learning technique grounded in tree structures. This algorithm is extensively applied for modeling data with high dimensionality, addressing both regression and classification challenges. In essence, the RF algorithm constructs and amalgamates the outcomes of multiple decision trees, which are generated randomly. This amalgamation forms an ensemble that enhances the stability and precision of predictions [16].

Uniformity exists in the bias across all trees, while variances can be mitigated by diminishing the coefficients of relationships [17]. In the context of regression tasks, the random forest relies on a randomly generated vector, and through the growth of trees, it employs the predictor  $d(x, \phi)$  to derive numerical values. The RF algorithm assumes statistical independence within the training data samples. The generalized crucial square mistake of the numerical indicator  $d(x)$  can be communicated as follows:

$$E_{x,y}(Y - d(x))^2 = \arg \text{minimum}_{\delta} \sum_{i=1}^k \Psi(y_i, F_{m-1}(x_i) + \delta d_m(x_i)) \quad (1)$$

The Random Forest (RF) prediction is subsequently constructed by aggregating the outputs of  $k$  trees, denoted as  $d(x, \phi)$ . For a more in-depth exploration of the theoretical underpinnings of the RF model, comprehensive information can be referenced in [16].

#### Support Vector Regression (SVR)

SVR is an expansion calculation to the directed machine-learning strategy known as Back Vector Machine (SVM) for relapse issues (foreseeing persistent yield values based on the input) [18]. In relapse issues, the SVR models attempt to diminish the mistake rate between the target and anticipated test by fitting around the hyperplane (relapse line) inside a certain limit esteem ( $\epsilon$ ) such that all information focuses inside  $\epsilon$  are not penalized for their blunder. The issue can be defined as takes after:

where  $n$  is the number of preparing tests, the slack variable  $\xi$  is the deviation for any esteem that falls exterior  $\epsilon$ , and  $C$  is the punishment calculate that decides the tradeoff between minimizing the preparing mistake and minimizing show complexity. As  $C$  increments, the resistance for focuses exterior  $\epsilon$  to increments. The execution of SVR at that point depends on the choice of parameters  $\epsilon$  and  $C$ .

$$\text{Min. } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n |\xi_i| \quad (2)$$

$$\text{s.t. } |y_i - w_i x_i| \leq \epsilon + |\xi_i| \quad i = 1, 2, \dots, n$$

### 4. Methodology

The proposed predictive framework employing machine learning encompasses six pivotal stages:

1. **Data Ingestion** The initial stage involves the assimilation of monthly time series data related to wind speed.
2. **Data Preprocessing** Subsequently, the second stage pertains to data preprocessing, a fundamental step where the meticulous selection of relevant and precise information significantly influences the learning capacity of the model. In this study, a comprehensive examination was undertaken on up to 12 consecutive lagged time series of wind speed, serving as input predictions to enhance the effectiveness and efficiency of the forecasting process.
3. **Data Partitioning** The third stage entails the partitioning of data into two subsets: the training set and the test set. This procedural step aims to evaluate the performance of machine learning algorithms when applied to predict outcomes on a data sample not

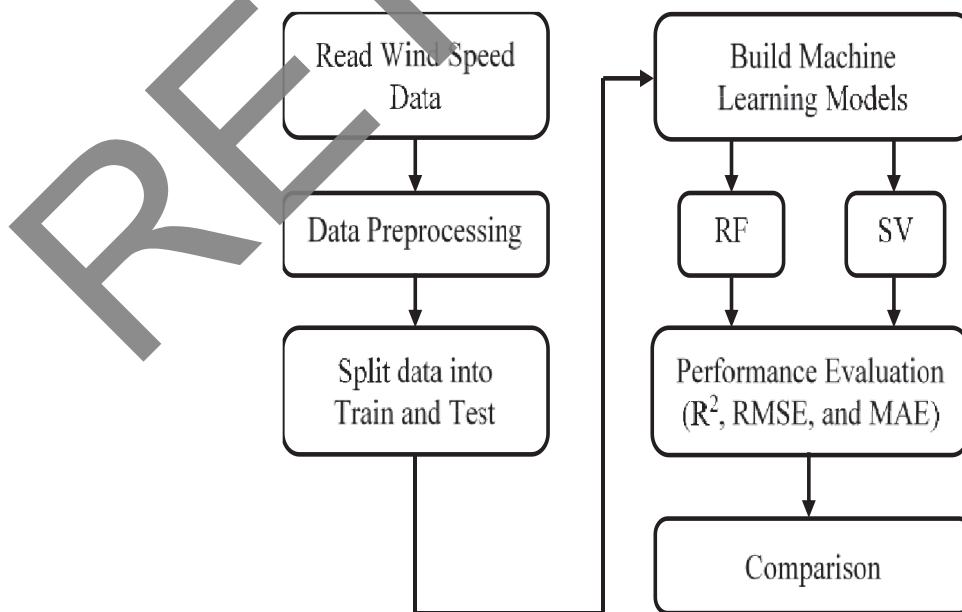
- utilized during the model training phase. The data spanning from 1981 to 2015, constituting 85% of the dataset, were allocated for model training (calibration), while the remaining 15% of the data from 2016 to 2021 were reserved for model verification.
4. **Model Development** Moving on to the fourth stage, machine learning models are meticulously developed, utilizing the Random Forest (RF) and Support Vector Machine (SVM) algorithms, with the primary objective of forecasting wind speed one month in advance.
  5. **Performance Evaluation** The fifth arrangement includes execution assessment, where three quantitative measures—coefficient of assurance (R<sup>2</sup>) (Eq. 3), Root Cruel Square Mistake (RMSE) (Eq. 4), and Cruel Outright Mistake (MAE) (Eq. 5)—are utilized to survey the adequacy of the forecast models [33].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n \left( y_i - \left( \frac{1}{n} \sum_{i=1}^n \hat{y}_i \right) \right)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Ultimately, in the ultimate phase of the analysis, a comparative evaluation of the empirical performance metrics is conducted between the two prediction models. This scrutiny is undertaken to discern the model exhibiting the most optimal performance. The flowchart of the proposed prediction system used in this study is shown in Figure 1.



**Fig. 1.** Illustrates the composite chart depicting the proposed forecasting framework

## 5. Results and Discussion

The assessment of the predictive capabilities of the models during the verification phase is showcased in Table 1 for the two employed machine learning methodologies. The findings suggest that with an increase in the number of lagged time series predictors for wind speed, the performance of both algorithms exhibits enhancement.

Notably, employing the maximum number of lagged time series (12) yields the most favorable predictive performance for the algorithms. Moreover, it is evident that the support vector machine (SVM) algorithm outperforms the random forest (RF) algorithm, as indicated by its lower root mean square error (RMSE) of 0.352, mean absolute error (MAE) of 0.286, and coefficient of determination ( $R^2$ ) of 0.707, in comparison to the RF algorithm's RMSE of 0.414, MAE of 0.322, and  $R^2$  of 0.599.

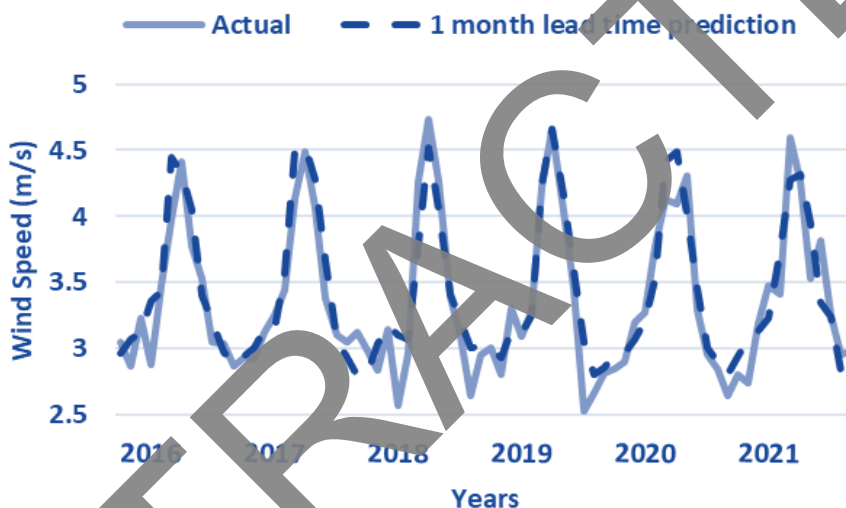
**Table 1.** Presents the performance of the models for different numbers of lagged inputs during the verification phase

No. Lags	RF			SVM		
	<i>RMSE</i>	<i>MAE</i>	$R^2$	<i>RMSE</i>	<i>MAE</i>	$R^2$
1	0.407	0.298	0.241	0.331	0.256	0.307
2	0.453	0.354	0.399	0.417	0.323	0.468
3	0.377	0.271	0.564	0.377	0.272	0.565
4	0.315	0.232	0.697	0.335	0.247	0.655
5	0.306	0.219	0.717	0.314	0.226	0.698
6	0.295	0.217	0.739	0.308	0.221	0.709
7	0.302	0.214	0.725	0.304	0.229	0.719
8	0.303	0.224	0.737	0.283	0.208	0.766
9	0.278	0.211	0.788	0.272	0.201	0.795
10	0.259	0.200	0.814	0.237	0.176	0.841
11	0.258	0.197	0.815	0.242	0.181	0.834
12	0.237	0.180	0.836	0.223	0.171	0.856

Figures 2 and 3 illustrate graphically a comparison between the observed time series of the wind speed using the two prediction algorithms.



**Fig. 2.** Displays the watched and anticipated time arrangement of wind speed throughout the verification period utilizing the RF model.



**Fig. 3.** Illustrates the observed and predicted time series of wind speed during the verification period employing the SVM model.

However, although the results of the prediction performance of the two algorithms were to some extent good, it can be said that these models did not fit the wind speed time series data well. This affirms the previously stated challenge that predicting wind speed is intricate due to its chaotic and stochastic nature. [19-21].

## 6. Conclusions

The precise prediction of wind constitutes a critical aspect in environmental planning, maintenance of energy system balancing, operation and control of wind farms, planning of power systems, and augmentation of system reliability. In this study, the RF and SVM algorithms were used for predicting wind speed one month in advance for Diyala governorates, Iraq. Different lagged time series values (up to 12) of the factor are used to find the best prediction model for the selected governorate. The optimal models were determined by assessing RMSE values. Additionally, other statistical parameters, including R2 and MAE, were computed to enhance the comprehensive evaluation of the models' proficiency. The results showed that both of the predictive machine-learning methods could be used for predicting monthly wind speed. The results showed also that the prediction performance of the SVM algorithm is a little bit better than that of the RF algorithm. However, the prediction accuracy needs to be further optimized. Therefore a future work to this research work is to optimize the current results by using other prediction techniques (such as deep learning modes) and combining lagged values from the other

meteorological factors (such as temperature, relative humidity, pressure, and many others) and even combining global climate indices (climate teleconnections such as the El Nino Southern Wavering (ENSO), Atlantic Multidecadal Wavering (AMO), North Atlantic Swaying (NAO), Dipole Mode List (DMI), and many others) as inputs for developing models.

## References

1. X. He et al., "Solar and wind energy enhances drought resilience and groundwater sustainability," *Nature Communications*, vol. **10**, no. 1, p. 4893, Nov. 2019.
2. B. François, M. Borga, J. D. Creutin, B. Hingray, D. Raynaud, and J. F. Sauterleite, "Complementarity between solar and hydro power: Sensitivity study to climate characteristics in Northern-Italy," *Renewable Energy*, vol. **86**, pp. 543–553, Feb. 2016.
3. P. Zeng, X. Sun, and D. J. Farnham, "Skillful statistical models to predict seasonal wind speed and solar radiation in a Yangtze River estuary case study," *Scientific Reports*, vol. **10**, no. 1, p. 8597, May 2020.
4. M. S. Hanoon et al., "Wind speed prediction over Malaysia using various machine learning models: potential renewable energy source," *Engineering Applications of Computational Fluid Mechanics*, vol. **16**, no. 1, pp. 1673–1689, Dec. 2022.
5. Z. Ma et al., "Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction," *Energy Conversion and Management*, vol. **205**, p. 112345, Feb. 2020.
6. G. Memarzadeh and F. Keynia, "A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets," *Energy Conversion and Management*, vol. **213**, p. 112824, Jun. 2020.
7. Hussein Alkattan, Sanjar Abdullaev, El-Sayed M. El-Kenawy. (2023). The «Climate in Weathers» Approach to Processing of Meteorological Series in Mesopotamia: Assessment of Climate Similarity and Climate Change using Data Mining. *Journal of Intelligent Systems and Internet of Things*, **10** (1), 48-65.
8. C. Wang, H. Zhang, and P. Ma, "Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network," *Applied Energy*, vol. 259, p. 114139, Feb. 2020.
9. A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system," *Sustainable Energy Technologies and Assessments*, vol. **25**, pp. 146–160, Feb. 2018.
10. D. Zafraakis, G. Tzanes, and J. K. Kaldellis, "Forecasting of Wind Power Generation with the Use of Artificial Neural Networks and Support Vector Regression Models," *Energy Procedia*, vol. **159**, pp. 509–514, Feb. 2019.
11. T. Brahim, F. Alhebshi, H. Alnabils, A. Bensenouci, and M. Rahman, "Prediction of Wind Speed Distribution Using Artificial Neural Network: The Case of Saudi Arabia," *Procedia Computer Science*, vol. **163**, pp. 41–48, 2019.
12. R. K. B. Navas, S. Prakash, and T. Sasipraba, "Artificial Neural Network based computing model for wind speed prediction: A case study of Coimbatore, Tamil Nadu, India," *Physica A: Statistical Mechanics and its Applications*, vol. **542**, p. 123383, Mar. 2020.
13. S. M. Lawan, W. A. W. Z. Abidin, and T. Masri, "Implementation of a topographic artificial neural network wind speed prediction model for assessing onshore wind power potential in Sibuluan, Sarawak," *The Egyptian Journal of Remote Sensing and Space Science*, vol. **23**, no. 1, pp. 21–34, Apr. 2020.

14. Z. Lin, X. Liu, and M. Collu, "Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks," *International Journal of Electrical Power & Energy Systems*, vol. **118**, p. 105835, Jun. 2020.
15. H. Hersbach et al., "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. **146**, no. 730, pp. 1999–2049, Jul. 2020.
16. L. Breiman, "Random Forests," *Machine Learning*, vol. **45**, no. 1, pp. 5–32, 2001.
17. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edi. Springer New York, NY, 2009.
18. S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. USA: Cambridge University Press, 2014.
19. V. Nikolić, V. V. Mitić, L. Kocić, and D. Petković, "Wind speed parameters sensitivity analysis based on fractals and neuro-fuzzy selection technique," *Knowledge and Information Systems*, vol. **52**, no. 1, pp. 255–265, Jul. 2017.
20. J. Wang, Q. Li, and B. Zeng, "Multi-layer cooperative combined forecasting system for short-term wind speed forecasting," *Sustainable Energy Technologies and Assessments*, vol. **43**, p. 100946, Feb. 2021.
21. L. Wang, Y. Guo, M. Fan, and X. Li, "Wind speed prediction using measurements from neighboring locations and combining the extreme learning machine and the AdaBoost algorithm," *Energy Reports*, vol. **8**, pp. 1508–1518, Nov. 2022.
22. Al-Mahdawi, H. K., Albadran, Z., Alkattan, H., Abotaleb, M., Alakkari, K., & Ramadhan, A. J. (2023, December). Using the inverse Cauchy problem of the Laplace equation for wave propagation to implement a numerical regularization homotopy method. *AIP Conference Proceedings* (Vol. 2977, No. 1). AIP Publishing.
23. Kenea U, Adeba D, Regasa MS, Nones M. Hydrological responses to land use land cover changes in the Fincha'a watershed. *Ethiopia Land*. 2021;**10**(9):910.3390/land10090916.
24. Aga HT. Effect of land cover change on water balance components in Gilgel Abay catchment using swat model. Netherlands: University of Twente; 2019.
25. Al-Nuaimi, B. T., Al-Mahdawi, H. K., Albadran, Z., Alkattan, H., Abotaleb, M., & El-kenawy, E. S. M. (2023). Solving of the inverse boundary value problem for the heat conduction equation in two intervals of time. *Algorithms*, **16**(1), 33.
26. Gyamfi C, Ndambuki JM, Salim RW. Hydrological responses to land use/cover changes in the Olifants Basin, South Africa. *Water*. 2016;**8**(12):588.10.3390/w8120588.
27. Worku T, Khare D, Tripathi S. Modeling runoff–sediment response to land use/land cover changes using integrated GIS and SWAT model in the Beressa watershed. *Environ Earth Sci*. 2017;**76**:1–14.10.1007/s12665-017-6883-3.
28. Akbar, E., Mollajafari, M., Al-Khafaji, H. M. R., Alkattan, H., Abotaleb, M., Eslami, M., & Palani, S. (2022). Improved salp swarm optimization algorithm for damping controller design for multimachine power system. *IEEE Access*, **10**, 82910-82922.
29. Ehsan khodadadi, S. K. Towfek, Hussein Alkattan. (2023). Brain Tumor Classification Using Convolutional Neural Network and Feature Extraction. *Fusion: Practice and Applications*, **13**(2), 34-41.
30. Da Silva VDP, Silva MT, Souza EPD. Influence of land use change on sediment yield: a case study of the sub-middle of the São Francisco river basin. *Eng Agríc*. 2016;**36**:1005–15.10.1590/1809-4430-eng.agric.v36n6p1005-1015/2016
31. dos Santos JYG, Montenegro SMGL, Silva RM, Santos CAG, Quinn NW, Dantas APX, et al. Modeling the impacts of future LULC and climate change on runoff and sediment yield in a strategic basin in the Caatinga/Atlantic forest ecotone of Brazil. *Catena*. 2021;**203**:105308.10.1016/j.catena.2021.105308.