RESEARCH

Open Access



Bienvenu Macaire Agbomahena¹, Kossoko Babatoundé Audace Didavi^{1*} and Richard Gilles Agbokpanzo^{2†}

*Correspondence: didavia@gmail.com

 Department of Electrical Engineering Polytechnic School of Abomey-Calavi (EPAC), Abomey-Calavi, Benin
 Department of ENSET-Lokossa National University of Science, Technology, Engineering and Mathematics of Abomey (UNSTIM), Abomey, Benin

Abstract

The need to integrate renewable energy sources into the energy mix is felt because of the many advantages they offer over fossil fuels, notably in terms of environmental protection and more uniformly distributed availability. The intermittent and stochastic ones, such as wind power, present many problems to network operators due to the volatile nature of their output power. This work presents a new technique for optimally forecasting the power output of a wind turbine installed at any geographic point located within a very large area. Once the study area is defined, it is gridded and optimally sampled in order to have a truly representative number of geographical points. The study area is then divided into sub-areas by grouping the samples by similarity of variation of meteorological parameters (wind speed and direction). For each sub-area, the optimal production periods are then identified and used for forecasting the power output. The forecasting technique used combines the LSTM model for forecasting meteorological parameters and the linear model for approximating the power curves of wind turbines. The technique was applied to the Beninese territory on which 90 subzones were formed. A 12 h forecasting of wind speed, wind direction and wind power were presented for one of the sub-areas. The clustering results gave a Silhouette score of at least 0.99. The wind speed and direction forecasting gave (0.34 m/s, 7.8 rad) and (93%, 70%) for RMSE and R^2 , respectively.

Keywords: Optimal production, *K*-means model, Wind power, Wind speed, Forecasting, LSTM

Introduction

The announced depletion of conventional energy sources is the main reason for the interest in renewable energies in recent years. Indeed, for the resources identified in 54 years, 63 years, 112 years and 100 years, respectively, according to a report of 2021 of the International Energy Agency (IEA), oil, gas, coal and uranium are expected to run out contrary to the renewable ones which are supposed to be inexhaustible on a human scale [1, 2]. At the same time, the IEA estimates that by 2030, global energy demand is expected to grow by 45% due to the industrialization and development policies of



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/.

various countries. In this context, countries with high production and high self-consumption will find themselves unable to meet all their energy needs with their energy mixes mainly made up of generators running on conventional resources. Countries with low production will consequently see their energy import costs increase and the quantity of imported energy decrease because of the difficulties that exporting countries would encounter. It should also be noted that these conventional resources are less evenly distributed than renewable energy sources and that climate change policies now require us to move toward green energy.

It seems quite obvious that new policies must be put in place, first of all for the exploitation of existing resources, but also for their production. Thus, two major solutions are emerging: energy efficiency and renewable energy sources as a means of energy production. Turning to this last solution, in terms of renewable energy production, Africa has an enormous potential with wind, solar photovoltaic (PV), hydroelectric, geothermal and biomass production of 978,066 TWh/year, 1,449,742 TWh/year, 1478 TWh/year, 105 TWh/year and 2374 TWh/year, respectively [3]. A potential that should allow the entire African population to have access to electricity. But unfortunately, according to a 2018 report by IRENA et al., 47% of the African population, i.e., more than 600 million people, 75% of whom live in rural areas, had no access to electricity [3]. Why this paradox? Looking closely, apart from the financial problems of exploitation of these renewable sources, they generally present problems of a natural and material nature. Indeed, the solar source is intermittent due to the natural movement of the earth around the sun and varies very randomly when it is available. This leads to a random variation of the output power of the PV generators. In the same way, the wind source (the wind) varies very randomly which leads to a random variation of the output power of the wind turbines but also an intermittence of this one due to a minimum of starting speed. In summary, these wind and PV generators have performances and powers that vary not only with time but also with the environment (space) in which they are installed.

In the literature, several solutions are proposed to the intermittence and random variability of these sources. For intermittency, it is possible to associate several generators with complementary productions: this would allow to have a production almost all the time without big interruption [4-10]. Another solution to intermittency is energy storage, but on a large scale, this is expensive from production to operation and recycling [11-14]. We can also think of international exchanges when the different countries are interconnected [15-17] or even the management of consumption by encouraging users to consume during periods of high production. For the random variability of the output power, the adequate solution encountered in the literature remains the forecasting thanks to the artificial intelligence techniques which allow to have an overview of the future evolution of these powers [18-22].

In a few years, it will be of great interest for everyone to be able to identify in his space any exploitable geographical point that can allow the development and exploitation of solar and wind resources mainly. It will also be necessary to be able to make an optimal and adapted forecast according to the type of system and the environment in order to be able to integrate well these fluctuating resources. In this paper, we propose a technique for forecasting and identifying the optimal production periods of a wind generator installed at any geographical point located within a very large area. The second part presents the material used and the methodology adopted, the third part the results obtained and the last part the conclusion.

Methods

Weather parameters influencing the variability of wind power

The random variation of the output power of the wind turbines is due to the randomness of the meteorological parameters. To control this variability, it is necessary to identify and control the evolution of these influencing parameters, as shown in the research flowchart in Fig. 1.

As with photovoltaic modules, there are different models describing the behavior of a wind turbine [23–26]. The output power of the wind generator is most often obtained from its power curve, which represents the output power as a function of the wind speed. It is possible that the curve used corresponds to a real generator or is obtained from a model. The most commonly used model is the Pallabazzer model. It is also possible to use the mechanical power on the shaft of the wind turbine determined from the wind speed, the area swept by the blades and by the power coefficient.

From these different models, we can easily deduce the wind speed and direction as input weather parameters. Thus, these models generally take the wind speed as input to give the associated electrical power for a given wind turbine. The wind direction is used to orient the blades of the wind turbine in order to extract a maximum of energy. We therefore retain these two parameters for the following.



Fig. 1 Research flowchart

Study area definition, gridding and sampling

Once the meteorological influencing parameters have been identified, it is necessary to delimit the study area and to record these parameters on representative samples of the area, which constitute geographical points. For our study, we considered the entire Beninese territory, which covers 114,763 km², as our study area.

The gridding and sampling technique is shown in Fig. 2. We define two types of samples: the green samples are the ones used for the work, and the red ones are the ones used for the results validation. We also define a grid spacing Δt which is the difference in latitude or longitude between two test or validation samples. We will see later on how the gridding step has been optimized.

Data and database used

As mentioned in the previous paragraph, the data used in this work are satellite data obtained from the NASA database. The main characteristics are presented below:

- Database: POWER LARC NASA [27];
- Data frequency: Hourly;
- Data types:
 - V50M: Wind speed at 50 m from the ground;
 - D50M: Wind direction at 50 m from the ground;
- Period: From January 1, 2012, to December 31, 2021.



Fig. 2 Gridded and sampled map of Benin

Sample data pre-processing

Once the grid has been completed and the samples defined, meteorological data (wind speed and direction) must be collected on all samples. For this study, we use hourly satellite data taken from the NASA database at each geographical point (test and trial samples) from January 1, 2012, to December 31, 2021.

These data must be pre-processed so that we can derive the correct information that we desire. There are several processing steps for this purpose.

Identification and suppression of outliers

An anomaly or outlier is best described as an observation that differs so much from the rest of the record that it is suspected to have been generated by a different process (Hawkins, 1980). These outliers can be due to a variety of processes, such as measurement error or specific phenomena, such as the occurrence of a fire or weather event. In view of the large number of applied areas that make use of outlier detection, the existing literature offers a considerable number of approaches devoted to it. The easiest way to identify outliers in a dataset is to represent and observe them. This allows us to notice the values that are abnormally far from each other. Box plots, for example, allow us to visualize the distribution of a single variable. These graphs are based on the median, as well as the lower and upper quartiles. An outlier is any extreme value, greater or less than *I* times the interquartile range. Usually *I* is 1.5. We can also perform statistical calculations such as the calculation of the mean, the standard deviation, the maximum or the minimum. These statistics make it possible to quickly identify possible anomalies in the data set.

Check for data stationarity

Since our data are time series, it is important to ensure the conformity of some of their properties. An important property of the time series is its stationarity. If a process is stationary, it means that its statistical properties do not vary over time, namely its mean, its variance (homoscedasticity) or its covariance. This notion of stationarity represents a crucial point in the analysis of time series, where the estimation of non-stationary series leads to spurious or illusory regressions. A stationarity study of the data is therefore important to ensure that the structure of the process that generated these series does not change over time: this is a very important condition for the time series forecast. The augmented Dickey–Fuller test (ADF) is an appropriate statistical tool. The time series considered is stationary if the p value is low (according to the null hypothesis).

Adding new variables to the dataset

It is sometimes important to add new variables to the data set in order to facilitate the learning of the models. For time series such as meteorological data (wind speed, irradiation, temperature...), the notion of periodicity is important for their analysis and interpretation or forecast. In this work, the series undergo a fast Fourier transform in order to identify the highest frequencies. These frequencies are then used to add new variables by the formulas (1) and (2).

$$x_{\rm sin} = \sin\left(t * \left(\frac{2 * \pi}{T}\right)\right) \tag{1}$$

$$x_{\cos} = \cos\left(t * \left(\frac{2 * \pi}{T}\right)\right) \tag{2}$$

 $X_{\rm sin}$ and $X_{\rm cos}$ are the new variables to be added for an identified important frequency f = 1/T and t is the time of data acquisition. It is also important in our case to make the models understand the concept of wind direction. Indeed, wind directions are angles and the models should understand for example that 0° and 360° are identical. Also, the wind direction is not very useful if the wind is weak. So, we associate these two variables (wind speed and direction) to create two new variables as shown in Eqs. (3) and (4).

$$W_{y} = W_{s} \sin(W_{d}) \tag{3}$$

$$W_{\rm x} = W_{\rm s} \cos(W_{\rm d}) \tag{4}$$

 W_x and W_y are the two new variables, W_s and W_d , respectively, the wind speed and direction at time *t*.

Data normalization or standardization

It is then essential, when one has in one's dataset variables of different orders of magnitude, to bring them to the same scale for training the models. For this purpose, two techniques are used: standardization and MinMax normalization. In this work we use the MinMax normalization presented by Eq. (5) which does not require the knowledge of data distribution.

$$x_{\text{norm}} = a + \frac{(x - \min(x))(b - a)}{\max(x) - \min(x)}$$
(5)

 x_{norm} represents the normalized x variable, a and b are the bounds of the new scale.

Data set subdivision

The dataset is then subdivided into three parts: the trainset (70%) to train the models, the validation set (20%) to set the hyper parameters of the models and the test set (10%) to test them.

Study area subdivision

It should be remembered that the objective is to be able to say at any point of the study area how the power output of a wind turbine will evolve in the near future and in which periods of the year a good production could be obtained. For this purpose, we divide the study area into sub-areas (Fig. 3) in which the wind profiles over the study period are identical or nearly so.

This is equivalent to grouping the samples (geographical points) whose meteorological parameters vary identically. For this study, a similarity threshold of at least 90% has been



Fig. 3 Sub-area cutting technique

set. This technique will allow to reduce the number of study samples because we could take without risk a sample in each sub-area to represent the whole sub-area. In the literature, several time series clustering models have been identified [28–35].

Clustering is a statistical analysis method used to organize raw data into homogeneous silos. Within each cluster, data is grouped according to a common characteristic. The ordering tool is an algorithm that measures the proximity between each element based on defined criteria. In this study, we used the k-means algorithm for clustering wind speeds as time series. *K*-means is an unsupervised non-hierarchical clustering algorithm. It allows to group in *K* distinct clusters the observations of the data set. Thus, similar data will be found in the same cluster. Moreover, an observation can only be found in one cluster at a time (exclusivity of membership). The same observation cannot belong to two different clusters.

Choosing a number of clusters K is not necessarily intuitive especially when the dataset is large. A large number K can lead to a too fragmented partitioning of the data. This will prevent the discovery of interesting patterns in the data. On the other hand, a number of clusters that is too small will lead to having, potentially, too generalized clusters containing a lot of data. In this case, there will be no "fine" patterns to discover.

For the same dataset, there is not a single possible clustering. The difficulty will be to choose a number of clusters K that will allow to highlight interesting patterns between the data. The most common method to choose the number of clusters is to run K-means with different values of K and to calculate the variance of the different clusters. The variance is the sum of the distances between each centroid of a cluster and the different observations included in the same cluster. Thus, we try to find a

number of clusters K such that the selected clusters minimize the distance between their centers (centroids) and the observations in the same cluster.

In our case, the number of clusters represents the different number of wind profiles that would have been identified in the whole study area. The search for a maximum number of clusters would allow us to optimize the grid spacing and thus the number of samples representative of the study area. Figure 4 shows the optimization principle that we have developed: we increase the number of samples until the number of clusters (wind profiles) does not increase anymore.



Fig. 4 Grid pitch and cluster number optimization flow chart

Sub-area characterization

Once the sub-areas have been formed, it is now necessary to identify in which period(s) of the year one could optimally produce with a wind turbine. To do this, for each subzone, we must:

- identify the periods where wind speeds are acceptable (medium high with low variance);
- · identify in which direction the wind turbine blades should be oriented.

We use the Weibull distribution to identify the most frequent wind speeds and the wind roses to identify the orientations.

Wind power forecasting

Forecasting technique options

Once the optimal production periods have been identified, the data on these periods are used to train wind power forecasting models, as it will be useless to pretend to make a forecast on periods when the wind turbine cannot be used. In the literature, several wind power forecasting techniques have been identified [20, 36–46]. The first technique is shown in Fig. 5.

This first technique exploits a history of power readings on an existing wind turbine and forecasts a succession of future values of this power from a succession of past values using a time series forecasting model. It presents a simple architecture but the developed forecasting model is only usable for the installation on which the records have been made. Moreover, this configuration requires a long period of power measurement for the forecast to be effective.

A second technique (Fig. 6) we encountered combines a time series forecasting model and a regression model. The time series forecasting model predicts the meteorological influence parameters (wind speed in our case) and then the regression model associates the wind turbine power values to the predicted influence parameter values. For this purpose, historical data of the influence parameters and the wind power are required. The architecture is more complex and as for the previous one, the model is only usable for the system for which it is developed and, in the environment, where the system is located. Indeed, as the meteorological parameters are a function of the geographical



Fig. 5 Forecasting architecture 1



Fig. 7 Forecasting architecture 3

locations, for the same system in two different environments the forecasting models will be different.

The last technique (Fig. 7) always combines a time series forecasting model to estimate future values of the influencing parameters and a wind power estimation model (Pallabazzer model, linear model, Chang model) that typically takes wind turbine characteristics such as $V_{\text{cut-in}}$ the interlocking speed, $V_{\text{cut-off}}$ the trip speed, V_{rated} the speed for which rated power is obtained, P_{rated} the rated power. Thus, this technique can be easily retrofitted to another wind turbine by changing the characteristics of the wind turbine in the power estimation model. Only the influence parameter prediction model has to be repeated when the installation site changes. This technique has the advantage of being more flexible, easy to adapt and does not require the system to be existing. It is thus the technique retained for the continuation.

Forecasting model's selection

The forecasting technique chosen involves using a wind speed and direction forecasting model as a time series and then estimating future power from the predicted speed and

direction from a wind turbine power curve approach model. In [36], Wang et al. give a fairly comprehensive summary of wind speed and wind power forecasting models. From this summary, we can retain as the most efficient basic wind speed forecasting models, the LSTM (Long short-term memory) and CNN (Convolutional neural network) models.

The convolution neural network is an improvement of the traditional neural network. It was originally designed for image processing and allows to encode the specific features of the image while reducing the number of parameters needed for the model configuration. Globally, CNNs apart from the input and output layers consist of three types of layers: convolutional layers, pooling layers and fully connected layers (Fig. 8).

The LSTM neural network is an improvement of the traditional recurrent neural network to overcome the vanish gradient problem. An LSTM unit is then composed of a cell, a dynamic memory C and three gates: Forget Gate, Input Gate, Output Gate. Indeed, the Forget Gate is an operation which leads the unit to forget or to decrease the weight of an information which was useful at the time t - 1 but which is not any more at the time t whereas the Input Gate allows the storage of new non-existent information or of very weak weight at the time t - 1. Finally, the output Gate controls the information which will be transmitted at time t + 1 according to the dynamic memory C and the activation function. The LSTM cell, thanks to this memory vector C, memorizes the values on arbitrary time intervals and the three gates regulate the flow of information entering and leaving the cell (Fig. 9).

These two models will be compared to select the most efficient one for each wind profile (wind speed variations in each sub-area).

To estimate the power produced by wind generators, we use the characteristic power curve. This curve allows to know the power produced from the wind speed. It is specific for each wind generator. There are two different approaches for modeling wind turbines, namely the application of a power curve model available in the literature on the one hand or the use of real curves to which an interpolation method is applied on the other hand. In the following, we present three power curve modeling methods, as well as different power curves of small and medium power machines available on the market. The characteristic parameters of each power curve are:



Fig. 8 CNN neural network architecture



Fig. 9 LSTM unit architecture

- $V_{\text{cut-in}}$: speed at which the power is switched on;
- *V*_{rated}: speed at which the rated power is obtained;
- *V*_{cut-off}: the speed at which the power is switched off;
- *P*_{rated}: rated power.

The linear model is the simplest and assumes that the power variation between $V_{\text{cut-in}}$ and V_{rated} is linear. Thus, the reduced power curve is expressed by three Eq. (6):

$$P = \begin{cases} 0, \quad V \le V_{\text{cut-in}} \\ a + b * V, \quad V_{\text{cut-in}} < V < V_{\text{rated}} \\ 1, \quad V_{\text{rated}} < V < V_{\text{cut-off}} \end{cases}$$
(6)

where the coefficients a and b are obtained by (7):

$$\begin{cases} a = \frac{V_{\text{cut-in}}}{V_{\text{rated}} - V_{\text{cut-in}}} \\ b = \frac{1}{V_{\text{rated}} - V_{\text{cut-in}}} \end{cases}$$
(7)

Studies show an overestimation of the productivity of wind generators, but despite this, this model is often used in studies of hybrid systems.

We also have the Pallabazzer model [47, 48], which differs from the linear model by the nonlinear shape of the curve between the speed of engagement and that for which we obtain the nominal power. In this part, the reduced power is expressed by:

$$P = \left(\frac{V^2 - V_{\text{cut-in}}^2}{V_{\text{rated}}^2 - V_{\text{cut-in}}^2}\right) \tag{8}$$

Some authors introduce a third degree polynomial in the central part of the curve [49]:

$$P = a_1 V^3 + a_2 V^2 + a_3 V + a_4 \tag{9}$$

where a_1, a_2, a_3 et a_4 are calculated on the basis of the power curve of the wind generator.

The most commonly used model is the Pallabazzer model (Notton et al. 2001; Prasad and Nataeajan2006). It is also possible to use the mechanical power on the shaft of the wind turbine determined from the wind speed (*V*), the area swept by the blades (A_r) and by the power coefficient (C_p):

$$P_{\rm mec} = \frac{1}{2} C_{\rm p} A_r \rho_a V^3 \tag{10}$$

Subsequently, we selected the Pallabazzer model for wind power estimation.

Forecasting model hyperparameters optimization

It is important to find out for which parameters of a model one could obtain better performance. As mentioned above, we have selected the CNN and LSTM models as the meteorological parameter forecasting models. The hyperparameters of these two models are presented in Table 1 along with the ranges of values that these hyperparameters can take for optimization.

We vary the input data sequence between 3 and 72 h, the sequence to forecast between 1 and 24 h. For the CNN network, the number of filters is varied between 50 and 300, the number of units between 100 and 1000. For the LSTM network the number of units is also varied between 100 and 1000. We use three different learning rates: 10^{-2} , 10^{-3} , 10^{-3} . This optimization is done with the Keras Tuner library under Python [50].

Models training

To train the models, we use the meteorological data (wind speed and direction) of the identified optimal production periods as well as the optimal hyperparameters obtained. The trainings are performed under Python3 where the CNN and LSTM models are developed with the TensorFlow library [51].

Clustering and forecasting model's performance evaluation

It is important to choose the right metric to evaluate a model, otherwise the assessment of its performance would be wrong. The following metrics are used to evaluate the models developed in this work.

The determination coefficient \mathbf{R}^2 is an indicator that allows to judge the quality of a linear regression, simple or multiple. With a value between 0 and 1, it measures the adequacy between the model and the observed data. In the case of a simple linear

CNN parameters	LSTM parameters
INPUT_WIDTH: 3 to 72 h	
LABEL_WIDTH: 1 to 24 h	
Filters: 10 to 300	Units : 10 to 1000
Units : 10 to 1000	
Learning_rates: 10 ⁻² , 10 ⁻³ , 10 ⁻⁴	Learning_rates: 10^{-2} , 10^{-3} , 10^{-4}

 Table 1
 CNN and LSTM parameters

regression, it is the square of the correlation coefficient. The R^2 is defined as the proportion of variance explained in relation to the total variance, i.e., [1 - (sum of squared residuals/total variance)]. This coefficient applies to both simple and multiple regression.

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} \left(Y_{i} - \widehat{Y}_{i}\right)^{2}}{\sum_{i=1}^{n} \left(Y_{i} - \overline{Y}\right)^{2}}$$
(11)

The mean square error (MSE) is defined as the average of the squares of the errors and allows to evaluate the quality of a prediction model. It penalizes the largest errors and outliers (outliers). Its expression is given by the following formula:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i \right)^2$$
(12)

with Y_i the actual value and \hat{Y}_i the predicted value and *n* the prediction size.

The root mean squared error (RMSE) is an extension of the MSE and allows to reduce the MSE to the same unit as the quantity evaluated. It is defined as the square root of the MSE:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2}$$
(13)

The normalized root mean squared error (NRMSE) is an extension of the MSE and is the most commonly encountered metric in the literature for evaluating wind speed prediction models. It is defined as the quotient of the root mean squared error and the mean or range of the values in the series.

$$NRMSE = \frac{RMSE}{Y_{max} - Y_{min}} ouNRMSE = \frac{RMSE}{\overline{Y}}$$
(14)

The Silhouette score is the metric that indicates the quality of a clustering. It varies between -1 and 1:

- The classification is bad when we obtain a negative co-efficient;
- The classified element is close to the decision boundary or is alone in the cluster when we obtain a coefficient equal to zero;
- The classification is good when we obtain a positive coefficient.

Its expression is given by the formula (15):

$$S_{\rm sil} = \frac{b-a}{\max(a,b)} \tag{15}$$

where *a* and *b* represent, respectively, the average distance that separates the data point from its cluster and the neighboring cluster.

Results and discussion

Grid and sample results

Figure 10 recalls the principle of gridding and shows the range of variation of latitude (l) and longitude (L). It can be noted that the latitude of the samples was varied from 5.625° to 12.6° and the longitude from 0.7° to 4.075° . The optimal grid step Δt obtained is 0.15° . We recall that the optimal grid spacing is the spacing for which the number of clusters does not vary anymore, i.e., the number of different profiles of the meteorological parameters (here the wind speed) does not vary anymore.

Thus, 0.15° separates in latitude or longitude two test samples or two validation samples and 0.075° separates in latitude or longitude a test sample and a validation sample. The test samples are those used for cluster formation and the validation samples are used to evaluate the quality of the formed clusters. We thus obtained 1080 test samples and 1080 validation samples.

Data presentation

Once the 1081 test samples were obtained, the wind speed and direction were downloaded for all of them for the study period (January 1, 2012 to December 31, 2021). Figure 11 shows the evolution of the mean, standard deviation and maximum for the



Fig. 10 Benin map (study area) gridded



Fig. 11 Statistics evolution for all wind speed data



Fig. 12 Silhouette score of 90 clusters

wind speeds. We can see that there is no apparent anomaly. The averages vary between 3.35 and 5.26 m/s with maxima ranging from 7.97 to 15 m/s. These statistics allow us to continue the study over the entire study area because wind turbines generally need a minimum of 3 m/s as a starting speed. We also note a variation of standard deviations between 1.17 and 2.2 m/s. These variations in statistics also show a difference in the different samples.

Clustering results

Once we are sure that the data are usable and that the study is useful for the entire study area, we proceed to clustering. In this step we group the samples whose wind speeds vary identically over the entire study area. In other words, each cluster represents a sub-area within which the wind speed varies identically at all points. We obtained a total of 90 clusters. Figure 12 shows the different Silhouette scores obtained for the 90 clusters.

We can notice that for 89 clusters, the Silhouette score is close to 1 which means that the similarity within these clusters is perfect. For cluster 90, the Silhouette score is 0. This value comes from the fact that within this cluster there is only one sample.

Figures 13 and 14 show the evolution of wind speeds within clusters 56 and 90 for the month of August 2020. We can notice that the curves are perfectly stacked for the cluster 56. For cluster 90 we notice the presence of only one curve which confirms the value 0 of the Silhouette score obtained. For ease of interpretation, the wind speed profiles of the samples have been named as follows: **Wind_Speed_Latitude_Longitude**. Results for the other clusters are available at [52].

Once the clusters are set up, the validation samples are then used to verify the representation of the entire study area by the 90 clusters. Figure 15 shows the Silhouette score (quality of membership in one of the 90 trained clusters) for the 1080 validation samples.

We can notice that the silhouette scores are all above the threshold of 0.9 that we have set (vary between 0.96 and 1). We can therefore conclude that the wind speed variations in the whole study area can be represented by the 90 different profiles identified.

Characterization of each sub-area

After clustering we have 90 sub-zones. Now we have to characterize each sub-area by determining the predominant wind directions and by determining the periods of the year when a wind turbine installed there could optimally produce. Given the number of sub-zones obtained, the following results will only be presented for cluster 56, an arbitrary choice. Indeed, any sub-zone could be chosen.

Figure 16 shows the wind rose for sub-area 56. We can notice that in this zone, the wind blows mainly to the southwest. This allows us to identify a good orientation of the blades in this area. Figure 17 shows the Weibull distribution as well as the frequency histogram of the wind speeds. We can see that the most probable wind speed is 5.4 m/s. We can also notice that the most convincing wind speeds are between 3.2 and 6.4 m/s. This speed range allows the installation of a wind turbine in this area. But in which period(s) of the year would we obtain an optimal production?

Figure 18 shows the monthly wind speeds and standard deviations for sub-area 56, which range from (3.37 m/s and 6.3 m/s) to (1.56 m/s and 2.3 m/s), respectively. In order to determine the optimal production period, we calculate the differences between the monthly wind speeds and monthly standard deviations (Fig. 19).

Knowing that the wind turbines start with minimum speeds between 7.2 and 10 km/h [53], we obtain for this sub-area an optimal production period going from **November to July**.

Wind speed and direction forecasting

Once the optimal production periods have been determined for each sub-area, the meteorological data for these periods are used to train the models according to the structure shown in Fig. 7. To do this, the wind speed and direction data are transformed using Eqs. (3) and (4) and then the prevailing frequencies of the speeds are used to add new variables to the data set according to Eqs. (1) and (2). Figure 20 shows the results of the fast Fourier transform (FFT) of the wind speed of sub-area 56.





Fig. 13 Wind speed evolution in August 2020 for Cluster 56





Fig. 15 Validation Silhouette score



Fig. 16 Wind rose of sub-area 56

We can notice the attendance of the annual, daily, half-day, third day and quarter day periods. Figure 21 shows the appearance of the dataset after adding the new columns.

Table 2 and 3 present the optimization results of the CNN and LSTM hyperparameters for wind speed and direction forecasts for sub-area 56. We can notice that for both models, the best forecast performances were obtained for a 12-h forecast from 18 h of past data. We therefore present here only the results for the 18_12 models.

Thus, it can be noticed that optimal (filters, units, learning rate) parameters of (70, 590, 10^{-3}) and (50, 300, 10^{-3}) were obtained for the W_x and W_y predictions for the CNN,



Fig. 17 Weibull distribution of sub-area 56



Fig. 18 Sub-area 56 monthly wind speeds and standard deviations

respectively. For the LSTM, optimal (units, learning rate) parameters of (460, 10^{-3}) and (160, 10^{-3}) were obtained for the W_x and W_y forecasts, respectively.

For the prediction of W_x and W_y , we can notice that the best performing model is the LSTM with coefficients of determination \mathbb{R}^2 and Root Mean Square Error RMSE of (86%, 0.47 m/s) and (91%, 0.33 m/s), respectively. The LSTM model is therefore retained for the forecast of wind speed and direction for this sub-area. Figures 22 and 23 show the forecast results of the parameters W_x and W_y for 12 h on three randomly selected parts of the test set.

We can notice from these forecast results that the forecast trend is often well predicted. Nevertheless, discrepancies can be noticed between the predicted and true W_x and W_y



Fig. 19 Optimal production period determination



parameters. This is due to the very random variation of the wind speed which makes learning difficult.

From the forecast values of W_x and W_y we can then deduce the values of wind speed and direction from Eqs. (15) and (16). We thus obtain for the 12 h of forecast of W_x and W_y , 12 h of forecast of wind speed and direction.

$$\begin{cases} W_y = W_s \sin(W_d) \\ W_x = W_s \cos(W_d) \end{cases} \Rightarrow \frac{W_y}{W_x} = \tan W_d \\ W_d = \tan^{-1} \left(\frac{W_y}{W_x} \right) \end{cases}$$
(16)

	WX	Wy	Day sin	Day cos	Day/2 sin	Day/2 cos	Day/3 sin	Day/3 cos	Day/4 sin	Day/4 cos	Year sin	Year cos
0	4.973031	4.028519	-9.697354e-12	1.000000	-1.939471e-11	1.000000e+00	2.911560e-11	1.000000e+00	-3.878941e-11	1.0	-0.003183	0.999995
-	4.758725	4.383542	2.588190e-01	0.965926	5.000000e-01	8.660254e-01	7.071068e-01	7.071068e-01	8.660254e-01	0.5	-0.002466	0.999997
N	4.517008	4.743326	5.000000e-01	0.866025	8.660254e-01	5.000000e-01	1.000000e+00	-2.655226e-11	8.660254e-01	-0.5	-0.001749	0.999998
3	4.102172	5.054927	7.071068e-01	0.707107	1.000000e+00	2.195805e-11	7.071068e-01	-7.071068e-01	4.391610e-11	-1.0	-0.001032	0.999999
4	3.540498	5.510660	8.660254e-01	0.500000	8.660254e-01	-5.000000e-01	-2.398891e-11	-1.000000e+00	-8.660254e-01	-0.5	-0.000315	1.000000

Fig. 21 Header of the data set after adding the new variables

CNN parameters		LSTM parameters	
INPUT_WIDTH: 18 h			
LABEL_WIDTH: 12 h			
Filters	70	Units	460
Units	590		
Learning_rates	10-3	Learning_rates:	10 ⁻³
RMSE	0.62 m/s	RMSE	0.43 m/s
R ²	83%	R ²	85%

Table 2 CNN and LSTM parameters and performance for W_x forecasting

Table 3 CNN and LSTM parameters and performance for Wy forecasting

CNN parameters		LSTM parameters	
INPUT_WIDTH: 18 h			
LABEL_WIDTH: 12 h			
Filters	50	Units	160
Units	300		
Learning_rates	10 ⁻³	Learning_rates:	10 ⁻³
RMSE	0.54 m/s	RMSE	0.39 m/s
R ²	85%	R ²	90%



Fig. 22 W_x forecast for 12 h on three portions of the test set

$$W_{\rm s} = \frac{W_y}{\sin\left(\tan^{-1}\frac{W_y}{W_x}\right)} \tag{17}$$

A validation of the forecasting performance of the models was performed on the data of 31 December 2020 and 1 January 2021. Figure 24 shows the 12 rows of input



Fig. 23 W_y forecast for 12 h on three portions of the test set

	Wx	Wy	Day sin	Day cos	Day/2 sin	Day/2 cos	Day/3 sin	Day/3 cos	Day/4 sin	Day/4 cos	Year sin	Year cos
78922	1.53933	3.48519	0.50000	-0.86603	-0.86603	0.50000	1.00000	0.00000	-0.86603	-0.5	0.01805	0.99984
78923	2.26591	2.45382	0.25882	-0.96593	-0.50000	0.86603	0.70711	-0.70711	-0.86603	0.5	0.01876	0.99982
78924	2.61445	1.89672	0.00000	-1.00000	-0.00000	1.00000	0.00000	-1.00000	-0.00000	1.0	0.01948	0.99981
78925	2.50808	1.80489	-0.25882	-0.96593	0.50000	0.86603	-0.70711	-0.70711	0.86603	0.5	0.02020	0.99980
78926	2.16080	1.91912	-0.50000	-0.86603	0.86603	0.50000	-1.00000	-0.00000	0.86603	-0.5	0.02091	0.99978
78927	1.99021	1.99787	-0.70711	-0.70711	1.00000	0.00000	-0.70711	0.70711	0.00000	-1.0	0.02163	0.99977
78928	2.16458	1.92992	-0.86603	-0.50000	0.86603	-0.50000	-0.00000	1.00000	-0.86603	-0.5	0.02235	0.99975
78929	2.82103	1.75048	-0.96593	-0.25882	0.50000	-0.86603	0.70711	0.70711	-0.86603	0.5	0.02306	0.99973
78930	4.10588	1.86597	-1.00000	0.00000	-0.00000	-1.00000	1.00000	-0.00000	0.00000	1.0	0.02378	0.99972
78931	5.45020	2.31123	-0.96593	0.25882	-0.50000	-0.86603	0.70711	-0.70711	0.86603	0.5	0.02450	0.99970
78932	6.18785	2.93821	-0.86603	0.50000	-0.86603	-0.50000	-0.00000	-1.00000	0.86603	-0.5	0.02521	0.99968
78933	6.06794	3.68647	-0.70711	0.70711	-1.00000	-0.00000	-0.70711	-0.70711	0.00000	-1.0	0.02593	0.99966
78934	5.22511	4.49113	-0.50000	0.86603	-0.86603	0.50000	-1.00000	0.00000	-0.86603	-0.5	0.02665	0.99964
78935	3.82205	5.24519	-0.25882	0.96593	-0.50000	0.86603	-0.70711	0.70711	-0.86603	0.5	0.02736	0.99963
78936	2.24899	5.72403	-0.00000	1.00000	-0.00000	1.00000	-0.00000	1.00000	-0.00000	1.0	0.51771	0.85556
78937	0.58613	5.79041	0.25882	0.96593	0.50000	0.86603	0.70711	0.70711	0.86603	0.5	0.51832	0.85519
78938	-0.84348	5.61702	0.50000	0.86603	0.86603	0.50000	1.00000	-0.00000	0.86603	-0.5	0.51893	0.85482
78939	-2.08524	5.32637	0.70711	0.70711	1.00000	0.00000	0.70711	-0.70711	0.00000	-1.0	0.51954	0.85444

Fig. 24 Validation input data

data and Fig. 25 the 12 rows of true values of the parameters W_x and W_y that the forecasts should give.

Figures 26 and 27 show the representative curves of target and predicted W_x and target and predicted W_y , respectively. We can notice that the forecasting is quite good.

The forecasting performances are presented in Table 4.

From the forecasting of W_x and W_y we deduce the values of wind speed and direction. Figure 28 shows the true (target) values of wind speed and direction that the forecasting should give.

	Wx	Wy
78940	-3.02534	4.97189
78941	-3.52837	4.52750
78942	-3.64065	4.10923
78943	-2.76312	3.17636
78944	-1.68198	2.72230
78945	-1.49250	3.96863
78946	-0.59339	3.76351
78947	-0.14825	3.08644
78948	-0.06255	2.55924
78949	-0.27307	2.18299
78950	-0.71162	1.98638
78951	-1.26318	1.93403

Fig. 25 Validation output data



Fig. 26 Validation target and forecasted W_x

Figures 29 and 30 show the representative curves of target and predicted wind speed and direction, respectively.

Table 5 presents the forecasting performances. We can note coefficients of determination of 93% and 70%, respectively, for the wind speed and direction with root mean square error of 0.34 m/s and 7.8 rad, respectively.

With these performances, we can qualify the LSTM model for medium term forecasting of wind speed and direction. In the literature, forecasting is mostly very short or short term with Root Mean Square Error above 0.7 m/s with deep learning networks [36, 54–57]. As an example, we list in Table 6 the results found in some recent articles. It must be said that the forecast horizons, the datasets and the computers used are rarely the same. This makes it difficult to compare performances in terms of accuracy and computation time. We can notice that for most of the proposed models for wind speed prediction, despite a horizon up to 12 h, our proposed LSTM model remains more efficient.



Fig. 27 Validation target and forecasted W_y

Table 4 W_x and W_y forecasting performance

W_x forecasting performance	
RMSE	0.47 m/s
R^2	86%
W_y forecasting performance	
RMSE	0.33 m/s
R^2	91%

	WS	WD
78940	5.82	121.32
78941	5.74	127.93
78942	5.49	131.54
78943	4.21	131.02
78944	3.20	121.71
78945	4.24	110.61
78946	3.81	98.96
78947	3.09	92.75
78948	2.56	91.40
78949	2.20	97.13
78950	2.11	109.71
78951	2.31	123.15

Fig. 28 Target values of wind speed and direction



Fig. 29 Validation target and forecasted wind speed



Fig. 30 Validation target and forecasted wind direction

 Table 5
 Wind speed and wind direction forecasting performance

Wind speed forecasting performance	
RMSE	0.34 m/s
R^2	93%
Wind direction forecasting performance	
RMSE	7.8 rad
	70%

Power output estimation of a wind turbine installed in sub-area 56 from predicted speeds

From the predicted wind speed values and characteristics, we can now estimate future values of the power output using Eq. (6). As an example, we make this

Refs.	Forecast type Method		Forecast horizon	Obtained accuracy		
				RMSE	MAPE (%)	R ²
Yatiyana et al. [58]	Wind speed	ARIMA	6 h	-	4.90	_
	Wind direction			-	15.60	-
Çevik et al. [59]	Wind speed	SWD-ANFIS, ANN and Ensemble methods	3 h	0.3842 m/s	-	-
Liu et al. [60]	Wind speed	ICEEMDAN-MMODA	3 h	0.35 m/s	5.02	
Liu et al. [56]	Wind speed	Jaya-SVM	3 h	1.01 m/s	15.34	0.85
Duan et al. [57]	Wind speed	ICEEMDAN-GRU- ARIMA	15 min	0.09 m/s	2.34	
Neshat et al. [61]	Wind speed	ED-HGNDO-BiLSTM	1 h	1.57 m/s		0.91
Yun Wang et al. [36]	Wind speed	CNN	1 h	0.73	9.84	0.96
		LSTM		0.77	10.51	0.95
Sheng-Xiang and Lin [<mark>62</mark>]	Wind speed	HTD- MOBBSA- Seq2Seq	1 h	1.58 m/s		
Our proposed model	Wind speed	LSTM	12 h	0.34 m/s	8.22	0.93
	Wind direction			7.8 rad	6.09	0.70

Table 6 Performance comparison

estimation on the 2 MW model of the manufacturer WindEnergy Lebanon whose characteristics are presented below:

- Power (kW) 2000
- Rated Power (kW) 2050
- Diameter (m) 88
- Cut-in speed (m/s) 3
- Cut-out speed (m/s) 25
- Rated speed (m/s) 12
- Maximum Cp 0.41
- Hub height (m) 80

Because the height of the wind turbine is 80 m and the data used for our prediction were taken at a height of 50 m, we extrapolated the velocities to a height of 80 with the logarithmic wind profile as defined below [63]:

$$\nu_2 = \nu_1 \frac{\ln\left(\frac{h_2}{z_0}\right)}{\ln\left(\frac{h_1}{z_0}\right)} \tag{18}$$

The reference velocity v_1 is measured at the reference height h_1 . v_2 is the wind speed at height h_2 . z_0 is the roughness length. We used a z_0 value of 0.0024 m.

Figure 31 shows the power obtained for the 12 h forecasting with the three estimation models.



Fig. 31 Estimated wind power

Conclusion

It is very important to be able to identify a favorable geographical site because of its characteristics when it comes to installing a wind generator. Also, when the generator is installed, it is necessary to have tools that can give an idea of how its output power will evolve given the stochastic character of the wind speed. The results of this work have shown that it is possible, based on artificial intelligence clustering techniques, to highlight the potential of wind power production at any point within a large geographical area and to delimit the periods of the year during which the production could be optimal. Moreover, these results led to the forecasting for 12 h of the wind speed and direction as well as of the output power of a wind generator by associating a time series forecasting model (LSTM) and an approximation model of the power curves of wind turbines. These results will be very useful for the identification of any potential site of installation of wind turbines for a country and the control in advance of the possible behavior of the wind turbine during all periods of the year.

Abbreviations	
ANN	Artificial neural network
RMSE	Root mean square error
MAPE	Mean absolute percentage error
MSE	Mean square error
LSTM	Long short-term memory
ED-HGNDO-BiLSTMHybrid	Evolutionary decomposition generalized normal distribution of bidirectional long- term memory model
CNN	Convolutional neural network
IHTSDS	Improved hybrid time series decomposition strategy
ANFIS	Artificial neuro-fuzzy inference system
SWD	Stationary wavelet decomposition
MMODA	Modified multi-objective dragonfly algorithm
ARIMA	Autoregressive integrated moving average
GRU	Gated recurrent unit
ICEEMDAN	Improved complete ensemble empirical mode decomposition with adaptive noise
RES	Renewable energy sources
SVM	Support vector machine
R ²	Regression coefficient

Acknowledgements

Not applicable.

Author contributions

The first draft of the manuscript was written by AM. DA and AR performed the conceptualization of the research idea, participated in the interpretation of the results, and reviewed the edited manuscript. All authors have made a substantial contribution to the manuscript. All authors read and approved the final manuscript.

Funding

No funding was received for conducting this study.

Availability of data and materials

The references related to the downloaded data and materials used have been specified. The other data used in the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 29 December 2022 Accepted: 23 March 2023 Published online: 12 April 2023

References

- 1. IEA (2022) Africa Energy Outlook 2019—analysis, IEA, Feb. 2022
- 2. IEA—International Energy Agency (2022) Energy transitions: Tracking progress in clean energy transitions through key indicators across fuels and technologies, IEA, Aug. 2022
- 3. IRENA, GIZ, KFW (2021) The Renewable Energy Transition in Africa, 2021
- 4. Weschenfelder F, Leite GNP, da Costa ACA, de Castro Vilela O, Ribeiro CM, Ochoa AAV, Araujo AM (2020) A review on the complementarity between grid-connected solar and wind power systems. J Clean Prod 257:120617
- Jurasz J, Piasecki A, Wdowikowski M (2016) Assessing temporal complementarity of solar, wind and hydrokinetic energy. In: E3S web of conferences, EDP Sciences: 00032, 2016
- Guezgouz M, Jurasz J, Chouai M, Bloomfield H, Bekkouche B (2021) Assessment of solar and wind energy complementarity in Algeria. Energy Convers Manag 238:114170
- Naeem A, Hassan NU, Arshad N (2020) Design of solar-wind hybrid power system by using solar-wind complementarity. In: 2020 4th international conference on green energy and applications (ICGEA). IEEE, pp 100–105
- Gonzalez-Salazar M, Poganietz WR (2021) Evaluating the complementarity of solar, wind and hydropower to mitigate the impact of El Niño Southern Oscillation in Latin America. Renew Energy 174:453–467
- Naeem A, Ul Hassan N, Yuen C, Muyeen SM (2019) Maximizing the economic benefits of a grid-tied microgrid using solar-wind complementarity. Energies 12(3):395
- Widén J, Carpman N, Castellucci V, Lingfors D, Olauson J, Remouit F, Bergkvist M, Grabbe M, Waters R (2015) Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. Renew Sustain Energy Rev 44:356–375
- 11. Toledo OM, Oliveira Filho D, Diniz ASAC (2010) Distributed photovoltaic generation and energy storage systems: a review. Renew Sustain Energy Rev 14(1):506–511
- 12. Vieira FM, Moura PS, de Almeida AT (2017) Energy storage system for self-consumption of photovoltaic energy in residential zero energy buildings. Renew Energy 103:308–320
- Zahedi A (2011) Maximizing solar PV energy penetration using energy storage technology. Renew Sustain Energy Rev 15(1):866–870
- Amrouche SO, Rekioua D, Rekioua T, Bacha S (2016) Overview of energy storage in renewable energy systems. Int J Hydrogen Energy 41(45):20914–20927
- 15. Shah D, Chatterjee S (2020) A comprehensive review on day-ahead electricity market and important features of world's major electric power exchanges. Int Trans Electr Energy Syst 30(7):e12360
- 16. Siano P, De Marco G, Rolán A, Loia V (2019) A survey and evaluation of the potentials of distributed ledger technology for peer-to-peer transactive energy exchanges in local energy markets. IEEE Syst J 13(3):3454–3466
- Khan KR, Rahman M, Masrur H, Alam MS (2019) Electric energy exchanges in interconnected regional utilities: A case study for a growing power system. Int J Electr Power Energy Syst 107:715–725
- Mellit A, Pavan AM, Lughi V (2021) Deep learning neural networks for short-term photovoltaic power forecasting. Renew Energy 172:276–288. https://doi.org/10.1016/j.renene.2021.02.166
- Mellit A, Massi Pavan A, Ogliari E, Leva S, Lughi V (2020) Advanced methods for photovoltaic output power forecasting: a review. Appl Sci 10(2):487. https://doi.org/10.3390/app10020487
- Hanifi S, Liu X, Lin Z, Lotfian S (2020) A critical review of wind power forecasting methods—past, present and future. Energies 13(15):3764
- 21. Sharma R, Diksha S (2018) A review of wind power and wind speed forecasting. Journal of Engineering Research and Application 8(7):1–9
- Ozkan MB, Karagoz P (2019) Data Mining-based upscaling approach for regional wind power forecasting: regional statistical hybrid wind power forecast technique (RegionalSHWIP). IEEE Access 7:171790–171800. https://doi.org/10. 1109/ACCESS.2019.2956203

- Carrillo C, Obando Montaño AF, Cidrás J, Díaz-Dorado E (2013) Review of power curve modelling for wind turbines. Renew Sustain Energy Rev 21:572–581. https://doi.org/10.1016/j.rser.2013.01.012
- 24. Seo S, Oh S-D, Kwak H-Y (2019) Wind turbine power curve modeling using maximum likelihood estimation method. Renew Energy 136:1164–1169. https://doi.org/10.1016/j.renene.2018.09.087
- Wang Y, Hu Q, Li L, Foley AM, Srinivasan D (2019) Approaches to wind power curve modeling: a review and discussion. Renew Sustain Energy Rev 116:109422. https://doi.org/10.1016/j.rser.2019.109422
- Wang Y, Hu Q, Srinivasan D, Wang Z (2019) Wind power curve modeling and wind power forecasting with inconsistent data. IEEE Trans Sustain Energy 10(1):16–25. https://doi.org/10.1109/TSTE.2018.2820198
- 27. Power | Data Access Viewer, Aug. 2022
- 28. Aghabozorgi S, Shirkhorshidi AS, Wah TY (2015) Time-series clustering-a decade review. Inf Syst 53:16-38
- 29. Peng K, Leung VC, Huang Q (2018) Clustering approach based on mini batch kmeans for intrusion detection system over big data. IEEE Access 6:11897–11906
- Huang X, Ye Y, Xiong L, Lau RY, Jiang N, Wang S (2016) Time series k-means: a new k-means type smooth subspace clustering for time series data. Inf Sci 367:1–13
- 31. Erdem E, Shi J, Peng Y (2017) Short-term forecasting of wind speed and power-a clustering approach. In: IIE annual conference. proceedings, institute of industrial and systems engineers (IISE), vol. 3501, 2014
- Dong W, Sun H, Li Z, Zhang J, Yang H (2020) Short-term wind-speed forecasting based on multiscale mathematical morphological decomposition, K-means clustering, and stacked denoising autoencoders. IEEE Access 8:146901–146914
- 33. Yang L, Zhang Z (2021) A deep attention convolutional recurrent network assisted by k-shape clustering and enhanced memory for short term wind speed predictions. IEEE Trans Sustain Energy 13(2):856–867
- Jarábek T, Laurinec P, Lucká M (2017) Energy load forecast using S2S deep neural networks with k-Shape clustering. In: 2017 IEEE 14th international scientific conference on informatics, IEEE, pp 140–145
- Liu J, Liu X, Wang S, Zhou S, Yang Y (2021) Hierarchical multiple kernel clustering. In: Proceedings of the AAAI conference on artificial intelligence, pp 8671–8679
- 36. Wang Y, Zou R, Liu F, Zhang L, Liu Q (2021) A review of wind speed and wind power forecasting with deep neural networks. Appl Energy 304:117766
- 37. González-Sopeña JM, Pakrashi V, Ghosh B (2021) An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renew Sustain Energy Rev 138:110515
- Demolli H, Dokuz AS, Ecemis A, Gokcek M (2019) Wind power forecasting based on daily wind speed data using machine learning algorithms. Energy Convers Manag 198:111823. https://doi.org/10.1016/j.enconman.2019.111823
- Kisvari A, Lin Z, Liu X (2021) Wind power forecasting—a data-driven method along with gated recurrent neural network. Renew Energy 163:1895–1909
- 40. Lin Z, Liu X (2020) Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network. Energy 201:117693
- Niu Z, Yu Z, Tang W, Wu Q, Reformat M (2020) Wind power forecasting using attention-based gated recurrent unit network. Energy 196:117081
- 42. Du P, Wang J, Yang W, Niu T (2019) A novel hybrid model for short-term wind power forecasting. Appl Soft Comput 80:93–106
- Ouyang T, Huang H, He Y, Tang Z (2020) Chaotic wind power time series prediction via switching data-driven modes. Renew Energy 145:270–281
- 44. Ding M, Zhou H, Xie H, Wu M, Nakanishi Y, Yokoyama R (2019) A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing 365:54–61
- 45. Qian Z, Pei Y, Zareipour H, Chen N (2019) A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Appl Energy 235:939–953
- Sun G, Jiang C, Cheng P, Liu Y, Wang X, Fu Y, He Y (2018) Short-term wind power forecasts by a synthetical similar time series data mining method. Renew Energy 115:575–584
- 47. Pallabazzer R (1995) Evaluation of wind-generator potentiality. Sol Energy 55(1):49–59
- 48. Pallabazzer R, Gabow AA (1992) Wind generator potentiality in Somalia. Renew Energy 2(4–5):353–361
- Chang T-J, Tu Y-L (2007) Evaluation of monthly capacity factor of WECS using chronological and probabilistic wind speed data: a case study of Taiwan. Renew Energy 32(12):1999–2010
- 50. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, others (2019) KerasTuner
- 51. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Y Jia, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I et al (2015) TensorFlow: large-scale machine learning on heterogeneous systems
- 52. IRJIET_Wind_Speed_Clustering—Google Drive, Dec. 2022
- 53. B. Ltd Wind turbine systems and Renewable energy 2022
- Bali V, Kumar A, Gangwar S (2019) Deep learning based wind speed forecasting—a review. In: 2019 9th international conference on cloud computing, data science & engineering (confluence), pp. 426–431, 2019. https://doi.org/10. 1109/CONFLUENCE.2019.8776923
- 55. Jiang P, Liu Z, Niu X, Zhang L (2021) A combined forecasting system based on statistical method, artificial neural networks, and deep learning methods for short-term wind speed forecasting. Energy 217:119361
- Liu M, Cao Z, Zhang J, Wang L, Huang C, Luo X (2020) Short-term wind speed forecasting based on the Jaya-SVM model. Int J Electr Power Energy Syst 121:106056. https://doi.org/10.1016/j.ijepes.2020.106056
- 57. Duan J, Zuo H, Bai Y, Duan J, Chang M, Chen B (2021) Short-term wind speed forecasting using recurrent neural networks with error correction. Energy 217:119397
- Yatiyana E, Rajakaruna S, Ghosh A (2017) Wind speed and direction forecasting for wind power generation using ARIMA model. In: 2017 Australasian universities power engineering conference (AUPEC), pp 1–6, 2017. https://doi. org/10.1109/AUPEC.2017.8282494

- 59. Çevik HH, Çunkaş M, Polat K (2019) A new multistage short-term wind power forecast model using decomposition and artificial intelligence methods. Physica A 534:122177
- 60. Liu Z, Jiang P, Zhang L, Niu X (2020) A combined forecasting model for time series: Application to short-term wind speed forecasting. Appl Energy 259:114137. https://doi.org/10.1016/j.apenergy.2019.114137
- 61. Neshat M, Nezhad MM, Abbasnejad E, Mirjalili S, Tjernberg LB, Garcia DA, Alexander B, Wagner M (2021) A deep learning-based evolutionary model for short-term wind speed forecasting: a case study of the Lillgrund offshore wind farm. Energy Convers Manag 236:114002
- 62. Lv S-X, Wang L (2022) Deep learning combined wind speed forecasting with hybrid time series decomposition and multi-objective parameter optimization. Appl Energy 311:118674. https://doi.org/10.1016/j.apenergy.2022.118674
- 63. Windenergie-Daten der Schweiz, Sep. 2021
- 64 Notton G, Cristofari C, Poggi P, Musseli M (2001). Wind hybrid electrical supply system: behaviour simulation and sizing optimization. Wind Energy 4:43–59
- 65 Prasad AR, Natarajan E (2006). Optimization of integrated photovoltaic–wind power generation systems with battery storage. Energy 31:1943–1954
- 66 Hawkins DM (1980). Identification of outliers 11. Springer

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[™] journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com