Renewable Energy 109 (2017) 529–541

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Hour-ahead wind power forecast based on random forests

A. Lahouar [a, b, *], J. Ben Hadj Slama [b]

[a] National Engineering School of Tunis, University of Tunis El Manar, BP 37, Le Belvedere, 1002, Tunis, Tunisia
[b] Laboratory of Advanced Technology and Intelligent Systems, National Engineering School of Sousse, University of Sousse, BP 264, Sousse Erriadh, 4023, Tunisia

ARTICLE INFO

Article history:
Received 27 May 2016
Received in revised form
22 February 2017
Accepted 19 March 2017
Available online 24 March 2017

Keywords:
Random forest
Wind power forecast
Hour-ahead
Spatially averaged wind speed
Wind direction
Importance of inputs

ABSTRACT

Due to its chaotic nature, the wind behavior is difficult to forecast. Predicting wind power is a real challenge for dispatchers who need to estimate renewable generation in advance to establish their strategies. To achieve an accurate wind power prediction, it is important to determine first which meteorological data need to be included in the predictor. For that purpose, this paper focuses on choosing the appropriate weather factors, namely spatially averaged wind speed and wind direction. These factors are selected according to correlation and importance measures. Then, the random forest method is proposed to build an hour-ahead wind power predictor. The random forest does not need to be tuned or optimized, contrary to most other learning machines. Both point and probabilistic forecasts are performed using the same inputs. The emphasis is put on the effect of wind speed and direction on the model performance, and the immunity of random forest to irrelevant inputs. The wind data used to test the proposed model are taken from Sidi Daoud wind farm in Tunisia. Results show an interesting improvement of forecast accuracy using the proposed model, as well as an important reduction of the different error criteria compared to classical neural network prediction.

1. Introduction

Wind energy has gained much attention by the end of the 20th century. It is one of the most used renewable energy resources worldwide, and it is getting more and more interesting for countries aiming to reduce their dependence on fossil fuels. For researchers, the challenge is to achieve maximum benefits from this resource without generating any disturbance or danger for actual power installations. According to the global wind statistics [1], the global wind power capacity at the end of 2015 is 432419 MW, and it is in continuous expansion for more than 20 years. In Tunisia, the installed capacity reached 245 MW provided by two wind farms with a total contribution of about 3.6% in the national production. China, the world leader, has 145104 MW of wind power installed capacity, including 30500 MW installed only in 2015 [1]. These recent and huge installations prove the world's trend towards renewable energy exploitation.

Since wind power is one of the most promising energy alternative resources, it is important to deal with challenges related to its integration in power grids, namely variability, additional costs, planning, transmission, etc [2]. The main problem often encountered is intermittency of wind, which induces grid instability and may lead either to lack or waste in energy. In general, a complete study about variability of wind should be done before installing any wind farm [3]. It is not possible to manage the wind blow, however, it is possible to predict its behavior, and this is the essence of wind power forecast. As a simple definition, predicting wind power means to estimate the expected power output of wind turbines for a period of time in the future, through one or many time steps. This period of time is called forecast horizon, and according to its length, short and long term are defined. There is no universal definition of what a short or a long term forecast is. But in general, short term ranges from one to few hours of prediction in advance, useful for planing dispatch and keeping the network stability. The long term is extended to few days in advance or even more, and it is useful for wind turbines' maintenance [4]. Obviously, the prediction performance varies from very short to long term [5], as well as the forecasting interests and applications [4]. The forecast uncertainty is handled by insurance strategies in order to mitigate operational risks [6,7].

* Corresponding author. Laboratory of Advanced Technology and Intelligent Systems, National Engineering School of Sousse, University of Sousse, BP 264, Sousse Erriadh, 4023, Tunisia.
E-mail addresses: ali.lahouar@gmail.com (A. Lahouar), bhslama@yahoo.fr (J. Ben Hadj Slama).

http://dx.doi.org/10.1016/j.renene.2017.03.064
0960-1481/

The state of the art shows a certain maturity already achieved in the field of wind power forecasting. Many reviews were written to summarize existing methods [8–12]. Almost all of them lead to two obvious conclusions; 1. Wind is one of the most difficult meteorological data to forecast, 2. There is no ideal strategy; each method has its own advantages and drawbacks and may be suitable in some context and not suitable in other cases. The advantages and drawbacks of each approach, the suitable context for application, the interest of combining methods and the comparison between common used models are depicted and detailed in Ref. [9]. Generally, forecasting methods are divided in two main groups; physical approaches and statistical approaches. The physical approach is related to physical considerations like temperature, pressure, wind farm layout and relief. In general, it utilizes numerical weather prediction (NWP) models which are descriptions of the atmosphere behavior using mathematical equations. When solved by a computer, they provide an accelerated simulation of the future weather states. NWP models may be used either directly or with post-processing to refine results [13–18]. However, this approach is accused to be not suitable for short term and small areas, apart from requiring extensive time simulation and resources [9,15,19]. The statistical approach includes time series methods used to predict the future evolution of wind according to historical measured values. Among pure statistical methods, the well known autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) methods are often used [20,21]. Another statistical approach called the persistence model (PER) assumes that the future wind value is equal to the current one. The persistence is considered a reference for the very short term, and it is used to evaluate the performance of other models [22]. Unlike NWP, pure statistical methods are suitable for short term prediction. Nonetheless, they are unable to learn and to change their prediction strategy, since they establish a strict mathematical relation between inputs and outputs [19]. Artificial intelligence models, and specifically machine learning techniques, are created to surpass this handicap. They are very powerful tools, with nonlinear modeling abilities, and they outperform generally the other models. The artificial neural network (ANN) is the most commonly used method in literature with different architectures: simple [23], high order network trained by Kalman filter [24], combined with wavelet transform [25,26], used with Gaussian distribution mixture [27], tuned by genetic algorithms [28], used with spatial correlation [29,30], and so on. The support vector machine (SVM) is another machine learning technique used originally for classification and then extended to regression problems with SVR (support vector regression). The SVR is widely exploited for wind speed forecast [31,32]. It may also be combined with other methods, such as empirical mode decomposition (EMD) which is a technique to decompose complex time series [33,34] or phase space reconstruction and genetic algorithms [35]. The major problem encountered with ANN and SVR is tuning, since they have many parameters to adjust, and these parameters affect the model accuracy. The hybridization is among improvements proposed to enhance these models. The random forest (RF) is a non-parametric learning machine, that does not require optimization or hybridization. It is an ensemble method that combines the prediction of several weak predictors. However, it is rarely used for wind forecast [36].

The classification of physical and statistical is not unique. Methods may be also classified into four families; namely physical, statistical, artificial intelligence and combined approaches [19]. Other classifications may be also found, considering spatial correlation models as a group [8]. According to forecasting data, methods may be indirect or direct. Indirect methods predict the wind speed (and eventually direction) and use the power curve to determine the power output. Direct methods predict power directly [4,19]. All mentioned papers in the previous paragraph use indirect methods, while several direct methods were indeed elaborated, based either on physical models or statistical models [37–42]. Among the most recent contributions in direct methods: the SVR enhanced by Markov models [43], the SVR combined with morphology operators [44], the NWP enhanced by ANN [45], the ANN tuned by particle swarm optimization (PSO) [46], the adaptive neuro-fuzzy inference system (ANFIS) with wavelet and PSO [47], the phase space reconstruction with resource allocating network [48], the decision trees [49] and the probabilistic wind power forecasts [50]. These recent contributions reflect the researchers' trend towards use of machine learning, numerical and heuristic optimization and probabilistic forecast.

Although power output is related to wind speed, many other atmospheric factors may have also their impact, such as wind direction, pressure, temperature and humidity [51]. The wind direction for example is predicted by the least square method [52] as well as by ARIMA [53] along with wind speed. The direction effect is also highlighted in the case of wind/power conversion through ANN [54]. The wind direction has not gained much attention despite its potential importance. However, predicting wind direction is meaningful only for turbines with horizontal axis.

This paper proposes a random forest (RF) approach achieving hour-ahead wind power forecast. The adopted method is direct. The RF is chosen for all the advantages mentioned of machine learning techniques, and was preferred over SVR and ANN since it doesn't need any optimization. The adoption of a non-parametric method such as RF resolves the major problem often encountered with artificial intelligence methods, which is parameter tuning. An extension to RF, called quantile regression forest, is developed in order to construct confidence intervals for the prediction. The probabilistic forecast, characterized by these prediction intervals, can handle uncertainties induced by intermittent wind power. The additive quantitative information provided by confidence intervals is useful for decision-makings especially when accurate spot predictions are difficult [55]. This paper focuses also on showing the importance of exogenous inputs of the predictor, such as wind speed and direction, and their influence on the model accuracy. A special attention is paid to the importance measures of inputs, and the ability of random forest to resist to non-significant data. The wind direction, the spatially averaged speed and the random forest are the main contributions of this paper over previous wind forecast researches.

The remainder of the paper is organized as follows: section 2 depicts the necessary mathematical tools, section 3 selects the suitable meteorological variables, section 4 describes the model construction, section 5 discusses the forecast results and section 6 concludes the paper.

2. Mathematical tools

The random forest is a method derived from artificial intelligence, like artificial neural network. It is an ensemble method that combines many decision trees using an algorithm called bagging.

2.1. Decision tree

A decision tree, also called Classification And Regression Tree (CART), is a statistical model introduced by Breiman in 1984 [56]. It depicts the different classes or values that an output may take in terms of a set of input features. Generally speaking, a tree is a set of nodes and branches organized in a hierarchy with no loops. A decision tree is a tree whose nodes store a test function to be applied to incoming data. Terminal nodes are called the tree leaves, and

each leaf stores the final test result. The tree is binary if each node has exactly two outgoing branches; called right child and left child. The decision tree is robust, immune to irrelevant inputs and provides good interpretability.

The remainder of this section is restricted to regression problems, since the prediction is a kind of regression. Let X be an input vector containing m features, Y an output scalar and S_n a training set containing n observations (X_i, Y_i).

$$S_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\},\ X \in \mathbb{R}^m,\ Y \in \mathbb{R} \tag{1}$$

During training, an algorithm drives the inputs split at each node, so that the parameters of split functions become optimized to fit with the S_n set. The principle consists of splitting recursively the input space X by searching optimal sub-partitions. More precisely, the first step of CART algorithm has to split at best the root into two different children according to:

$$\left\{X^j < d\right\} \cup \left\{X^j > d\right\} \tag{2}$$

where $j \in \{1, \ldots, m\}$ and $d \in \mathbb{R}$. In order to select the best split, the couple (j, d) should minimize a cost function, which is generally the child node variance. The variance of a node p is defined by:

$$Var(p) = \sum_{i:X_i \in p} \left(Y_i - \overline{Y}_p\right)^2 \tag{3}$$

where $\overline{Y}_p$ is the mean of the scalars Y_i present in the node p. Then, children nodes are also divided in the same way. The development of the tree is stopped by a termination criterion. It is common to stop the tree when a maximum number of levels is reached, or when a node contains less than a predefined number of observations. At the end of this training process, a prediction function $\widehat{h}(X, S_n)$ is constructed over S_n.

The testing process determines an estimation $\widehat{Y}$ of the output Y corresponding to any new input vector X.

$$\widehat{Y} = \widehat{h}(X, S_n) \tag{4}$$

Starting from the root, each node applies its own split function to the new input X. According to the result of the binary test, data are send to right or left child. This process continues until the data reach a leaf (terminal node) [56].

2.2. Bagging and random forest

The random forest is an ensemble method that combines the prediction of several decision trees [57]. The basic principle is called bagging (bootstrap aggregation); where a sample of size n taken from the training set S_n is selected randomly and fitted to a regression tree. This sample is called bootstrap, and it is chosen by replacement, which means that the same observation (X_i, Y_i) may appear several times.

A bootstrap sample is obtained by selecting randomly n observations with replacement from S_n, where each observation has the probability of $1/n$ to be selected. The independent identically distributed random variables Θ_l represent this random selection. The bagging algorithm selects several bootstrap samples $(S_n^{\Theta_1}, \ldots, S_n^{\Theta_q})$, applies the previous CART algorithm to these samples in order to construct a collection of q prediction trees $(\widehat{h}(X, S_n^{\Theta_1}), \ldots, \widehat{h}(X, S_n^{\Theta_q}))$, and then aggregates the output of all these trees.

In addition to bagging, the random forest selects also a predefined number *mtry* among the m features for the split in each node. The RF algorithm tries to find the best split among only the *mtry* selected features. The selection at each node is uniform, thereby each feature has the probability of $1/m$ to be selected. The number *mtry* is the same for all prediction trees, and it is recommended to be the square root or the one third of the features' number m:

$$mtry = \lceil \sqrt{m} \rceil \tag{5}$$

Or:

$$mtry = \left\lceil \frac{m}{3} \right\rceil \tag{6}$$

where $\lceil x \rceil$ denotes the ceiling function of x. The remainder of the algorithm is similar to CART, the best split couple (j, d) is obtained by minimizing a cost function, and the procedure continues until the full development of all trees.

The aggregation is performed by averaging the outputs of all trees. Consequently, the estimation $\widehat{Y}$ of the output matching a new input vector X is as follows:

$$\widehat{Y} = \frac{1}{q} \sum_{l=1}^{q} \widehat{h}\left(X, S_n^{\Theta_l}\right) \tag{7}$$

The main advantage of bootstrap aggregation is immunity to noise, since it generates non-correlated trees through different training samples. A weak predictor (a standalone regression tree) may be sensitive to noise, while the average of several decorrelated decision trees is not. The selection of a random subset *mtry* of features has the same aim of decorrelating trees.

Two main characteristics distinguish the random forest: the out-of-bag error *OOBE* and the measure of variable importance *VI*. The *OOBE*, also called generalization error, is a kind of built-in cross validation. It is the average prediction error of first-seen observations; i.e. using only the trees that did not see these observations while training. More explicitly, for each observation (X_i, Y_i) of S_n, an estimation $\widehat{Y}_i$ of Y_i is achieved by aggregating only the trees constructed over bootstrap samples not containing (X_i, Y_i). The *OOBE* is useful to estimate the generalization capacity of the model.

$$OOBE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2 \tag{8}$$

The variable importance measure is obtained by permuting a feature and averaging the difference in *OOBE* before and after permutation over all trees. Let's define for each bootstrap sample $S_n^{\Theta_l}$ its associated OBB_l, i.e. the set of observations not included in $S_n^{\Theta_l}$. For a fixed j among the m features, the values of the j^{th} variable are permuted randomly over OBB_l to get a disturbed sample called $\widetilde{OOB}_l$. The new $\widetilde{OOBE}_l$ of the disturbed sample is then calculated. These operations are repeated for every bootstrap sample. The importance of the j^{th} variable, called $VI(X^j)$, is defined by the difference between average errors of original OBB_l and disturbed $\widetilde{OOB}_l$ [58].

$$VI\left(X^j\right) = \frac{1}{q} \sum_{l=1}^{q} \left(\widetilde{OOBE}_l - OBBE_l\right) \tag{9}$$

If permutations over the j^{th} variable lead to increasing error, this variable is relevant. The more the score $VI(X^j)$ increases, the more the j^{th} variable becomes important [57]. The number of trees q will be denoted *ntree* for the remainder of the paper. The quantile regression forests will not be detailed here for concision purpose. However, a brief description of their operating principle will be

given in section 5.3.

3. Selection of input meteorological factors

The aim of this section is to determine the appropriate wind factors affecting the power output, by respect to the farm layout. The chosen variables will be the exogenous inputs of the RF predictor.

3.1. Wind farm characteristics

Over 4792 MW of installed power in Tunisia by 2014, only 307 MW are provided by renewable resources; 245 MW of wind power and 62 MW of hydraulic energy. The main resources of production are still gas and steam turbines, making the electricity very dependent on the import of fossil fuels. Over the 245 MW of wind power, 53.5 MW are provided by the farm of Sidi Daoud in the north-east, which is the subject of this study. The wind power provided 507 GWh in 2014, i.e. 3.6% of the total production of 14117 GWh. This contribution is still very low and unable to face the increasing demand every year, but it is expected to rise in the near future.

The offshore wind farm of Sidi Daoud contains 70 turbines aligned from north-east to south-west, with a total capacity of 53.5 MW. It covers an area of 6×5 km^2. The wind speed histogram is given in Fig. 1, showing a distribution centered at 5 m/s. The wind rose of Fig. 2 gives the wind dominant direction, which ranges from west to west north-west. It is noticed that turbines are aligned almost perpendicularly to this direction, as shown in Fig. 3. The nominal power of each individual turbine is between 330 kW and 1320 kW.

3.2. Effects of wind on a single turbine

The power curve is a wind turbine characteristic given by the manufacturer. Defined as a non-linear transfer function binding wind speed and power output, it depends on the squared rotor radius and cubed wind velocity. Three curve points are particular: cut-in speed (beginning of production), rated speed (nominal power) and cut-off speed (stop production). However, this theoretical curve may differ from real behavior, due to the other atmospheric factors like wind direction or temperature. In fact, modeling the wind power curve needs separate study [59,60]. Fig. 4(a) illustrates this difference on a 330 kW wind turbine of Sidi Daoud wind farm.

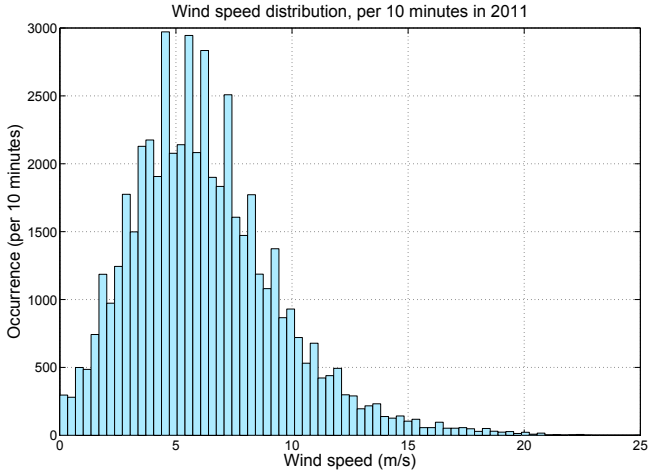

Fig. 1. Distribution of wind speed, sampled at 10 min in 2011.

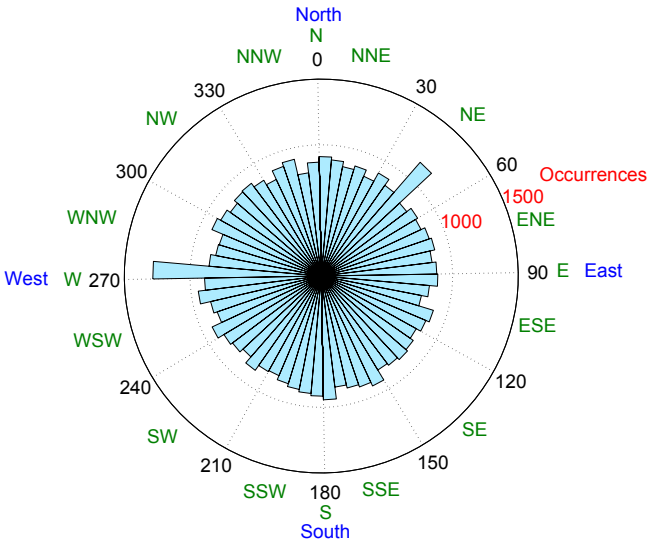

Fig. 2. Wind rose, directions sampled at 10 min in 2011.

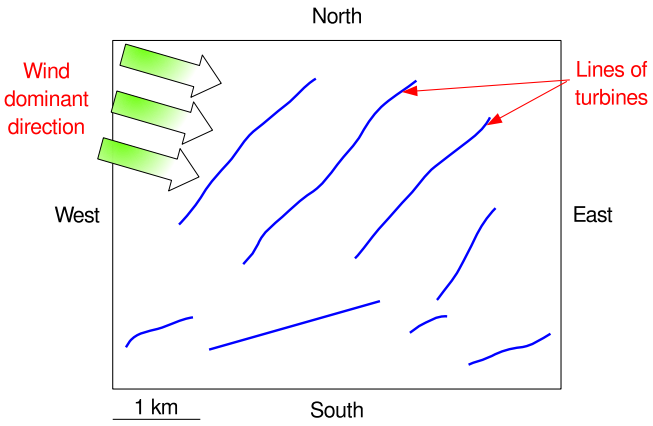

Fig. 3. Layout of the wind farm.

The wind speed in the x-axis is locally measured by an anemometer above the nacelle, at a height of 30 m. It can be seen that a wide range of power values (270–385 kW) correspond to a single wind speed of 15 m/s. The wind angular direction effect appears in Fig. 4(b) showing maximum production points for NW since it is the dominant direction. Despite equipped with yaw system orienting the rotor towards the wind, the turbine does not produce much power for NE and SW directions. This minimum production is a consequence of the farm layout (depicted in the previous section 3.1). Fig. 4(c) shows an extrapolation surface that fits power to wind speed and direction. For the rated speed of 15 m/s, several peaks and valleys appear according to the angular value. Because of this effect, the wind direction will be considered in the forecasting model.

3.3. Effects of wind on a set of turbines

The wind farm contains 70 turbines, aligned from NE to SW, perpendicularly to the wind dominant direction. Despite the small farm area of 6×5 km^2, the 70 measures of wind speed are not equal, as illustrated in Fig. 5(a). The difference between them, appearing only for some wind directions, leads to the question of which speed should be utilized as input. To deal with this problem,

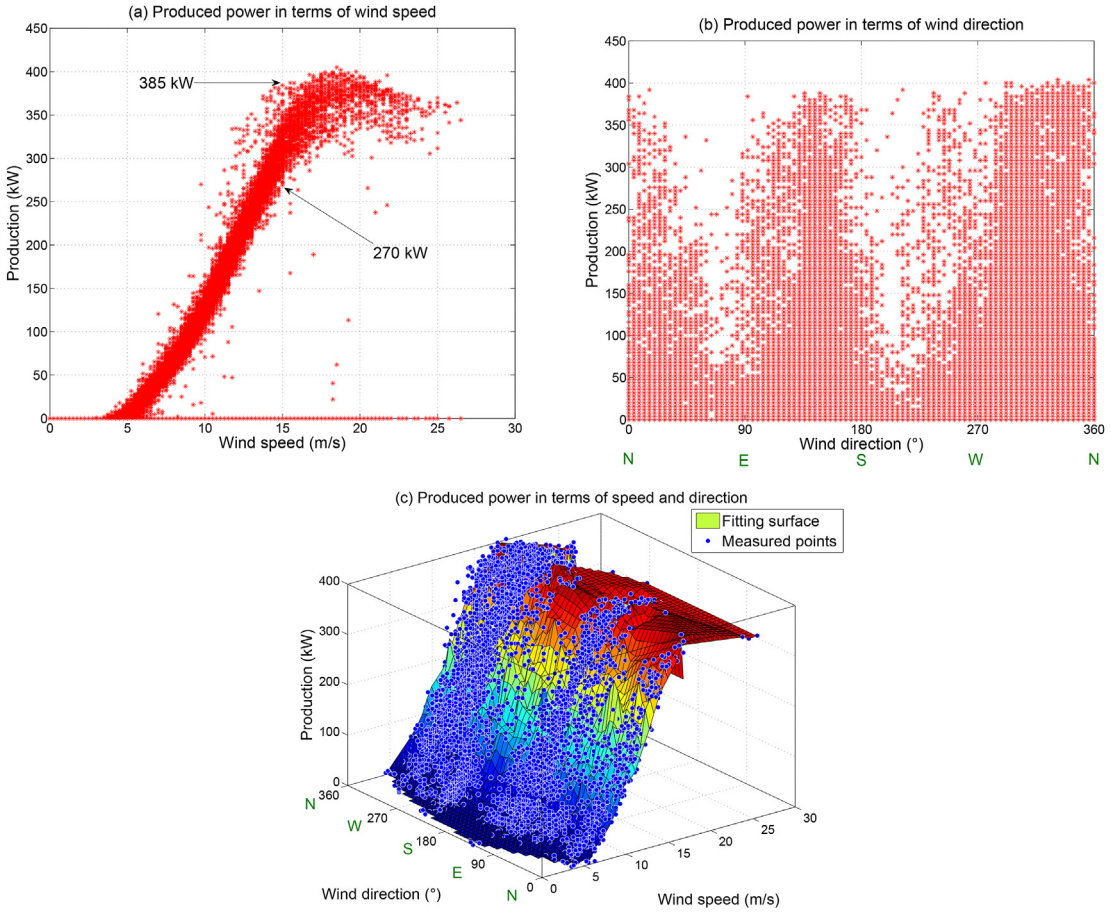

Fig. 4. Produced power of a 330 kW turbine in terms of: (a) wind speed, (b) wind direction, (c) speed and direction (2011).

it is relevant to pick the spatial average, defined by the average of all the 70 measures. This choice is inspired by the spatial smoothing effect, known to reduce forecast errors [51,61]. The spatial average of wind speed has similar power curve shape to that of single turbines. Fig. 5(b) shows a main curve reaching the nominal production of 53.5 MW, along with small other curves with the same shape. The small curves appear obviously when some turbines are shut down, i.e. when the farm is not working at full capacity. If the wind speed of the x-axis was recorded from one single location (from the meteorological tower for example), the curve would be much thicker; thus it would be harder to establish a production-speed relationship. Fig. 6 shows the correlated curves of overall power and wind spatial average. The linear correlation coefficient is computed in order to assess the similarity between the two curves. But since the relation between speed and power is not linear, only the samples between 5 and 18 m/s will be considered. The [5–18] m/s interval specifies almost the linear part of the power curve. The linear correlation coefficient between power and spatial average in 2011 is equal to 0.89 within the specified interval. The same coefficient calculated between power and speed metered by the meteorological tower is 0.69 within the same interval. Thus, the speed spatial average is much more correlated to the global production than the tower measures. As a consequence, the wind spatial average will also be considered in the predictor.

The wind direction has also an impact on the overall power output. As it can be seen through Fig. 5(c), the entire farm production is maximum for some directions and minimum elsewhere. This is caused by the wind farm layout characterized by large spacing between turbines in the direction of the dominant wind. This spacing weakens the turbulence behind the pales known as the wake effect, causing maximum power for dominant direction NW and opposite SE. In the other hand, the spacing in the perpendicular direction is much smaller, which explains the weak generation for SW and NE directions. Fig. 5(d) contains the surface fitting power to speed spatial average and direction. Very similar to Fig. 4(c), the main difference is that peaks and valleys are more visible, which asserts the importance of wind direction in forecasting power generation. The correlation between power and direction is not linear, thus the linear correlation coefficient is not needed in this case. However, similarities will be detected in the next section through correlation plots.

4. Forecast technique

In general, forecasting wind power means predicting the power that the farm is expected to produce while operating under a constant wind for a given time step [62]. During this time step, which is equal to one hour in this study, the power output $P(h+1)$ is assumed to be constant and equal to its average value $\tilde{P}(h+1)$. The estimated hour-ahead power value will then be the output of the random forest predictor, and denoted $\widehat{\tilde{P}}(h+1)$, where $h+1$ stands for one hour in advance. The inputs of the predictor may be endogenous or exogenous. The endogenous inputs are past power values P, and they are mandatory for building time series predicting models. The exogenous inputs are external factors that have an influence on the future power, for instance wind spatially averaged

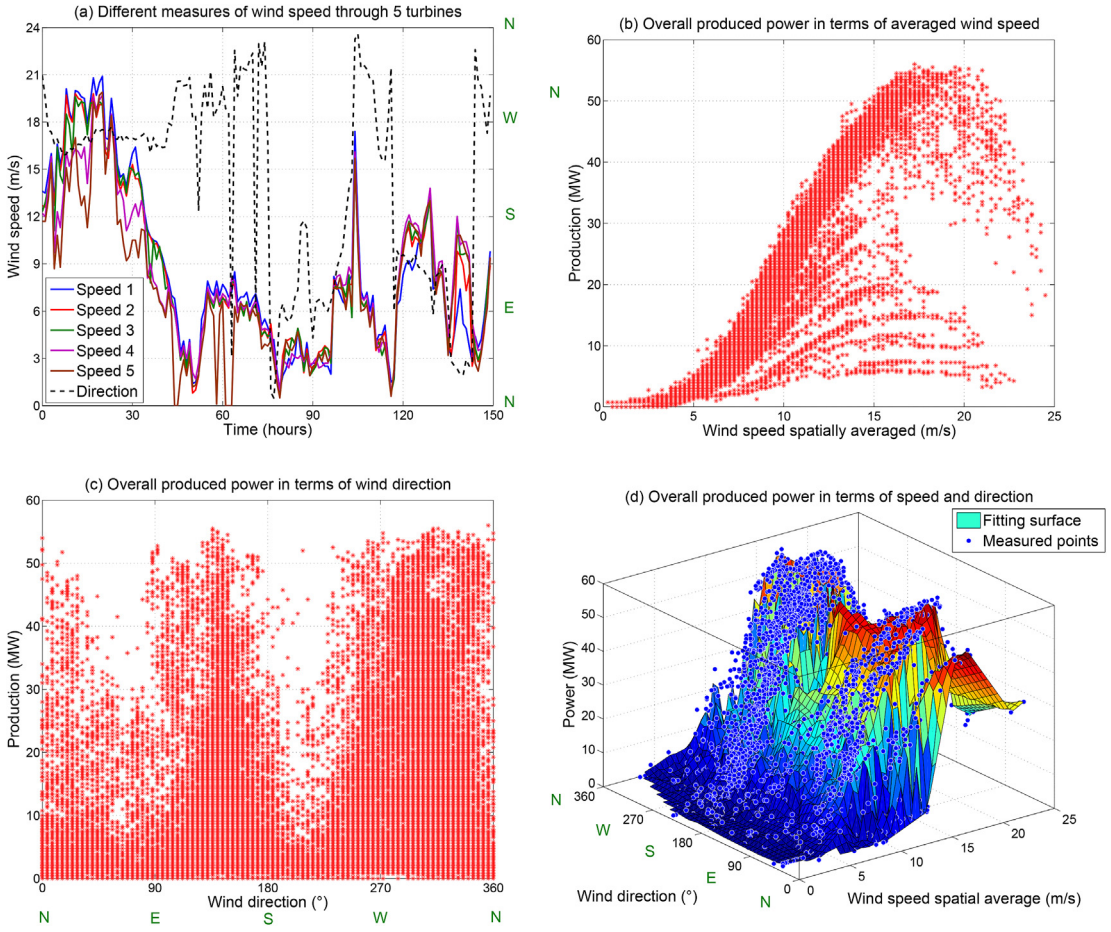

Fig. 5. Different measures of wind speed through 5 turbines (a), overall produced power of the wind farm in terms of: (b) wind speed, (c) wind direction, (d) speed and direction (2011).

speed $\overline{S}$ and wind direction D. The term $\overline{S}$ is chosen to make the difference with speed metered by the meteorological tower, denoted S. The number of past samples is set to 6 for each feature; P, $\overline{S}$ and D. This number is chosen assuming that only recent values of inputs have an impact on future power in the short term. This choice will be justified by correlation and importance measures. All samples are measured at hour h (current hour) and 10 min spaced. The term $h:i$ stands for hour h and i minutes, where $i \in [0, 10, 20, 30, 40, 50]$. By adopting the notation of section 2, the model is depicted by these equations:

Power input:

$$P(h) = [P(h:00), P(h:10), \ldots, P(h:50)] \tag{10}$$

Speed input:

$$\overline{S}(h) = \left[\overline{S}(h:00), \overline{S}(h:10), \ldots, \overline{S}(h:50)\right] \tag{11}$$

Direction input:

$$D(h) = [D(h:00), D(h:10), \ldots, D(h:50)] \tag{12}$$

Input vector:

$$X = \left[P(h), \overline{S}(h), D(h)\right]^T \tag{13}$$

Measured output:

$$Y = \tilde{P}(h+1) \tag{14}$$

Predicted output:

$$\widehat{Y} = \widehat{\tilde{P}}(h+1) \tag{15}$$

The $P(h)$ component is mandatory, while $\overline{S}(h)$ and $D(h)$ are

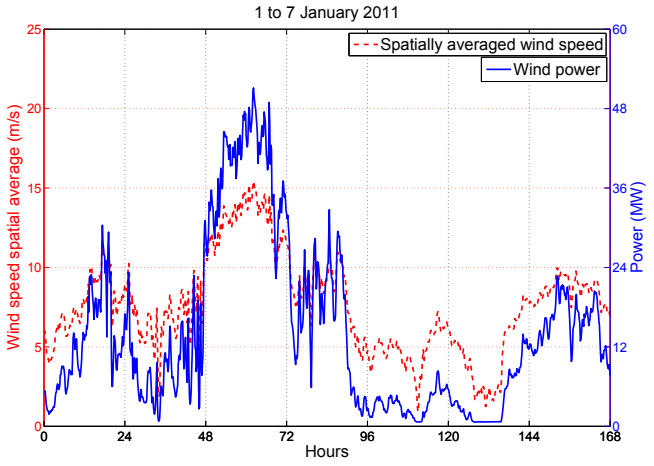

Fig. 6. Overall produced power of the farm, and spatial average of wind speed, from 1 to 7 January 2011.

optional, and added just to assess their influence on the model accuracy. The total number of input m is then 6, 12 or 18, according to the availability of $\bar{S}(h)$ and $D(h)$ in the input vector X. There is no rule to choose the number of past samples, but the correlation functions can give an idea. Fig. 7 shows the autocorrelation plot of the produced power during 7 days in January 2011, sampled at 10 min. The idea is to determine the similar past samples, i.e. the lags with maximum correlation. As the curve is dropping fast, there is no need to get more than 6 samples, that ensure a minimum correlation equal to 0.93. In other words, there is a strong correlation between the signal and its one-hour-lagged version. The blue dashed line represents the maximum correlation minus 0.1, and it intersects the curve in the tenth lag. This lag is considered the last input acceptable for hour-ahead predicting. An extended two-years autocorrelation function shows that there is no obvious seasonality or cycles in the power signal. The same procedure is thereafter followed to detect the similarities between power and speed spatial average, through the cross correlation plot of Fig. 8 (only the positive part of the x-axis is given). In fact, this plot can be informative if the power curve is linear. Thus, the nonlinear parts of the power curve are simply neglected, since power is proportional to speed most of the time (most speed measures are within [5–18] m/s). The cross correlation plot ensures, by 6 past samples, a minimum correlation of 0.83 between power and spatial average. The minimum acceptable correlation is detected at lag 15, so there is no need to take more than 15 past samples. Indeed, lags are considered acceptable if they are between the green dashed line (maximum correlation) and the blue one (maximum correlation minus 0.1).

The same procedure cannot be simply repeated with wind direction, since there is no linear correlation between power and direction, as it can be seen through Fig. 5(c). However, there is a noticeable increase in power production between 230° (SW) and 360° (N). This part of the scatter plot can be considered linear just to establish the cross correlation plot of Fig. 9. Although the approximation is a bit coarse, it is generally enough to identify the most influential past samples of direction. Fig. 9 shows the cross correlation plot from March 28th (at 5 a.m.) to April 2nd, a period within which the direction was between 230° and 360°. The sixth lag ensures a correlation of 0.33, while the minimum acceptable correlation is detected at lag 50. The 6 chosen past samples are

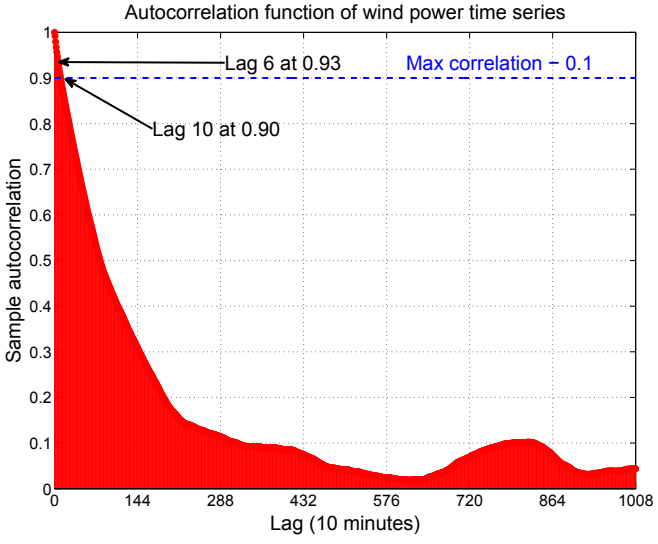

Fig. 7. Autocorrelation function of overall produced power of the farm through 7 days in 2011.

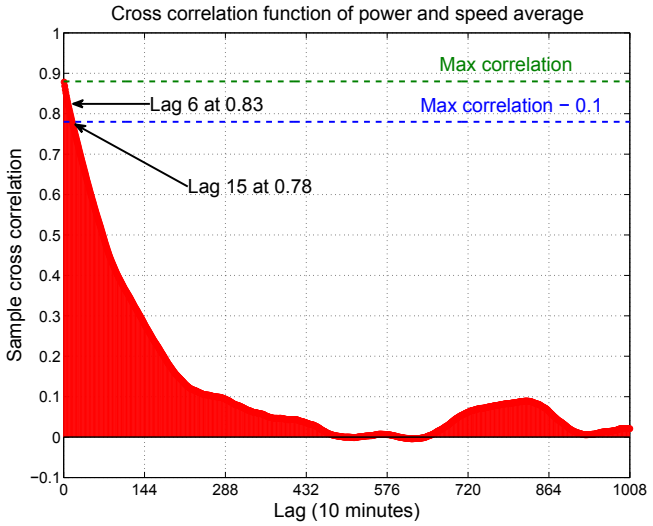

Fig. 8. Cross correlation function between overall produced power and spatially averaged wind speed through 7 days in 2011 (only the positive part of the x-axis is shown).

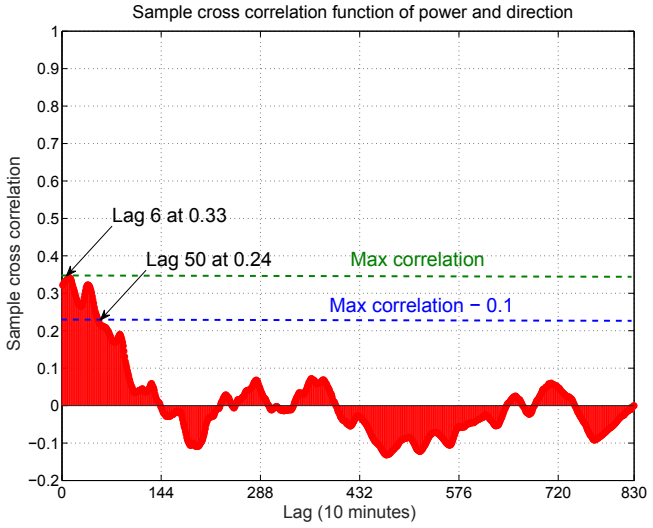

Fig. 9. Cross correlation function between overall produced power and wind direction, from 28 March (5 a.m.) to 2 April 2011 (only the positive part of the x-axis is shown).

within the limits, whether for P, $\bar{S}$ or D. The measure of importance, depicted in section 2 and given in Table 1, helps also to determine a suitable number of past values. The values of this table will be explained in details in section 5.

The main drawback of the proposed approach is incapability of forecasting wind gust, since wind is considered constant during the whole time step, though sudden wind variations may be harmful for turbines. In addition, accuracy decreases as the time step increases, making the model valid only for the short term.

5. Case study

5.1. Data preprocessing

The data available in this study are gathered from Sidi Daoud wind farm. They are in raw format, as they are metered by the sensors. For each wind turbine, the produced power is measured in

Table 1
Importance measure of the RF model inputs.

Importance ($\times 10^{10}$)	1 January 2011	1 April 2011	1 July 2011	1 October 2011
$P(h:00)$	1.7067	0.8217	1.1021	0.9179
$P(h:10)$	3.5701	2.3557	2.4630	2.1143
$P(h:20)$	7.6746	6.6662	5.7677	6.4738
$P(h:30)$	19.5569	14.8396	14.2989	10.9115
$P(h:40)$	37.4365	42.2595	38.2738	31.4237
$P(h:50)$	70.1238	71.4938	58.4888	61.1984
$\overline{S}(h:00)$	0.7091	0.6118	0.5980	0.5429
$\overline{S}(h:10)$	0.7878	0.6332	0.6077	1.0069
$\overline{S}(h:20)$	2.0886	2.6560	1.2208	1.4390
$\overline{S}(h:30)$	5.0169	6.0880	2.6230	4.8386
$\overline{S}(h:40)$	13.9133	15.4094	17.6812	17.6111
$\overline{S}(h:50)$	43.9048	36.6618	36.9445	33.8879
$D(h:00)$	0.3174	0.3561	0.3237	0.3245
$D(h:10)$	0.3121	0.3357	0.3143	0.3196
$D(h:20)$	0.3115	0.3161	0.2925	0.3007
$D(h:30)$	0.3014	0.3215	0.3072	0.2957
$D(h:40)$	0.3136	0.3245	0.3080	0.3154
$D(h:50)$	0.3324	0.3474	0.3206	0.3228

kW, as well as the wind speed in m/s by an anemometer above the nacelle. In addition, the meteorological tower provides wind speed (in m/s), wind direction (between 0° and 359.9°) and air temperature (in °C). All data are given by one decimal place and by step of 10 min, for the years 2010 and 2011.

The huge quantity of data contains necessarily some mistakes caused generally by defective sensors. Any missing information are directly detected by numerical calculus software, and they are replaced by previous values. Redundant data are detected by the number of rows of the data table, which should contain $365 \times 24 \times 6 = 52560$ rows for each year. These faults are often scarce and do not affect the prediction quality. Some defective sensors provide a constant value for a long period. When detected, this constant value is considered false and its corresponding period is avoided. Additional information may also be calculated, such as the overall produced power by summing the output of all turbines, the spatially averaged wind speed by averaging all the speed measures, and the hourly averaged power.

5.2. Evaluation criteria

The evaluation criteria of forecast methods are procedures to quantify the prediction error. They are used to estimate through different approaches the distance between original output P and estimated output $\widehat{P}$ (The notations P and $\widehat{P}$ are used here instead of $\tilde{P}(h+1)$ and $\hat{P}(h+1)$ to lighten the equations). The chosen evaluation criteria in this paper are the following: the mean absolute error (*MAE*), the root mean squared error (*RMSE*), the maximum absolute error (*MXE*), the normalized mean absolute error (*NMAE*), the mean absolute percentage error (*MAPE*), the mean absolute scaled error (*MASE*) and the correlation coefficient (*CC*).

$$MAE = \frac{1}{s}\sum_{i=1}^{s}\left|\widehat{P}_i - P_i\right| \tag{16}$$

$$RMSE = \sqrt{\frac{1}{s}\sum_{i=1}^{s}\left(\widehat{P}_i - P_i\right)^2} \tag{17}$$

$$MXE = \max_{1\le i\le s}\left|\widehat{P}_i - P_i\right| \tag{18}$$

$$NMAE = \frac{1}{s}\sum_{i=1}^{s}\frac{\left|\widehat{P}_i - P_i\right|}{P_{installed}} \times 100 \tag{19}$$

$$MAPE = \frac{1}{s}\sum_{i=1}^{s}\frac{\left|\widehat{P}_i - P_i\right|}{P_i} \times 100 \tag{20}$$

$$MASE = \frac{1}{s}\sum_{i=1}^{s}\left(\frac{\left|\widehat{P}_i - P_i\right|}{\frac{1}{s-1}\sum_{j=2}^{s}|P_j - P_{j-1}|}\right) \tag{21}$$

$$CC = corr\left(P, \widehat{P}\right) = \frac{cov\left(P, \widehat{P}\right)}{\sigma_P \sigma_{\widehat{P}}} \tag{22}$$

where s stands for the total number of errors over the testing period, $P_{installed}$ means the total installed power of the wind farm (53.5 MW in the case of this paper), *cov* denotes the covariance and σ_P stands for the standard deviation of P.

These criteria cover the most common error measurements used by researchers. The *MAE* is the most natural criterion, since it is simply the average of all found errors, through their absolute values to avoid error offset. The *RMSE* does the same thing, except avoiding the error sign by squaring instead of absolute value. The *RMSE* is similar to the *RMS* of alternating electric current, and it is generally greater than *MAE*. These two criteria assess the prediction accuracy by raw and not normalized quantification, and this may be irrelevant. For instance, an error of 2 MW has not the same impact on a farm of 1000 MW of installed capacity as it may has on a 50 MW farm. To overcome this problem, it is possible to use the *MAPE*. The *MAPE* divides each error by the real power value before averaging, in order to get a percentage. However, this criterion has also its drawbacks, especially when the real value tends towards zero. The *MAPE* will then tend towards infinity, which is not relevant, and this is particularly the case of intermittent time series such as wind. The *MASE* is proposed in this context of intermittence. The *MASE*, with no unit, reflects better accuracy whenever it gets closer to zero. It does never give infinite values except in the case where all real values are equal, which is not significant. Finally,

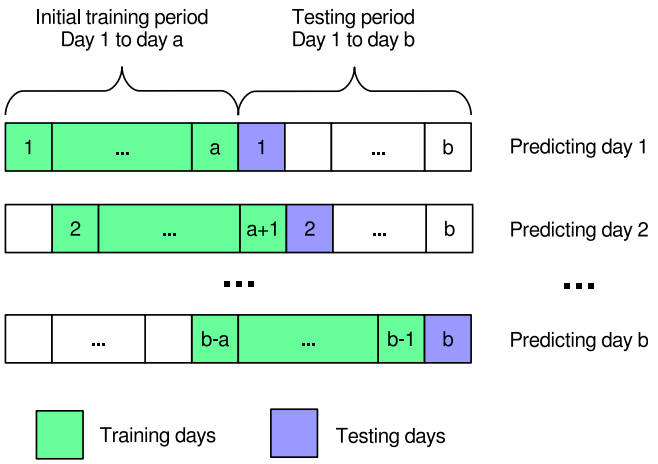

Fig. 10. The online learning principle.

the *NMAE* is another proposed criterion, which divides each error by the overall installed capacity instead of the real power value. It avoids then the division by values close to zero, and differs from the *MASE* since it represents a percentage. But since it is related to the installed power, the *NMAE* remains dependent on the wind farm capacity. Thus, it has the drawback of getting smaller in larger farms.

The remaining criteria are a bit special. The *MXE* determines the highest error through its absolute value over all the test period. It is especially useful to detect error peaks. The *CC* is simply the linear correlation coefficient between measured and predicted time series, and it is the only criterion that should be close to one to reflect better accuracy.

5.3. Results and interpretations

To test the proposed model on the available data, hour-ahead wind power forecast is performed throughout the year 2011. The RF predictor uses the set of equations (10)–(15) for modeling, and equation (7) for predicting. The number of trees *ntree* is set to default 500, and the *mtry* parameter follows equation (6). The learning process is online, with 1 year of training and 1 year for testing. For example, to predict the 24 h of 1st January 2011, 1st January to 31st December 2010 are used as training period. The training set S_n contains then 8759 observations:

$$n = 365 \times 24 - 1 = 8759 \tag{23}$$

The minus 1 means that the first hour of the training period can not be used as observation, since its corresponding inputs are outside S_n. The testing set contains 24 input vectors which will be applied to the RF predictor. The online learning means repeating the training process for each new day of the testing period. It is introduced to capture the seasonal components. As shown in Fig. 10, the training period is sliding in order to learn the new weather patterns. This implies 24 new observations appended everyday to the training set S_n, while the oldest 24 observations are removed. Hence, the training period contains always one year just

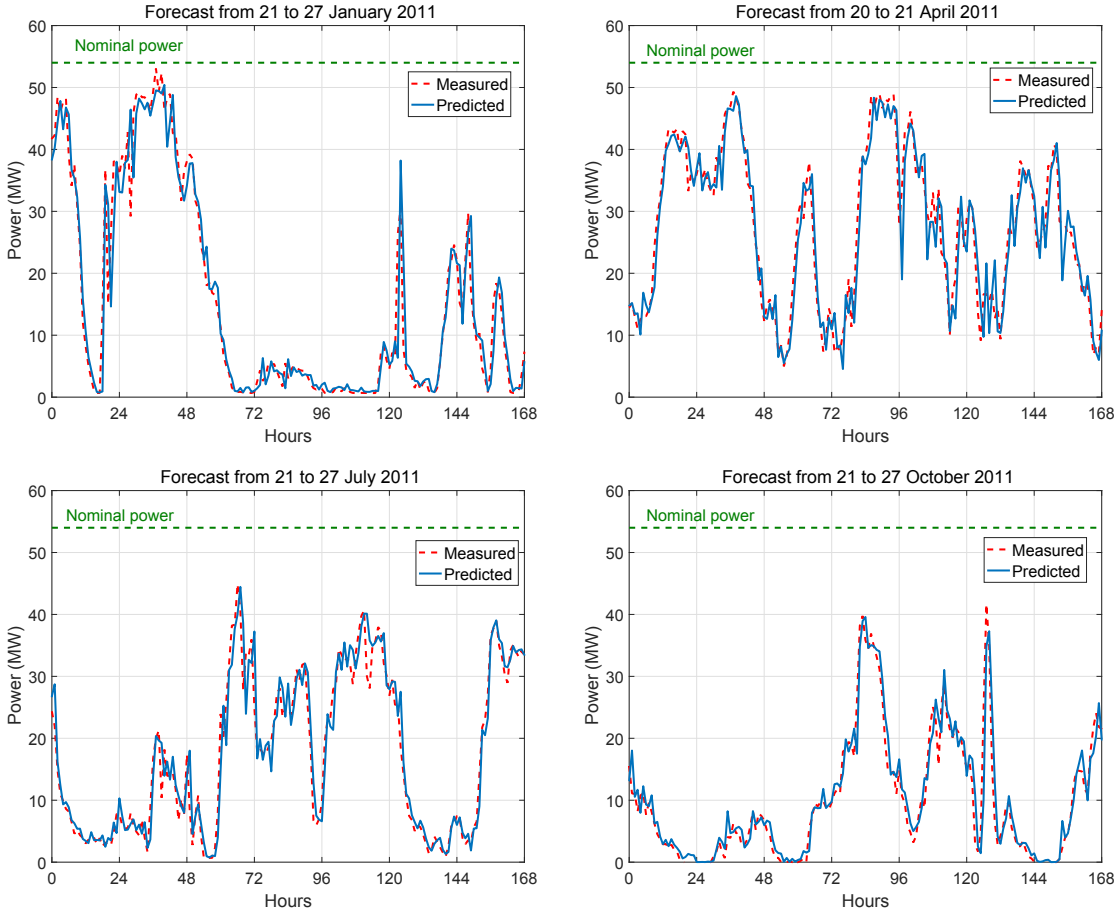

Fig. 11. Hour-ahead forecast results over four testing periods.

Table 2
Forecast evaluation criteria of all methods throughout the year 2011.

Method	PER	RF			ANN		
Criterion	1 input	6 inputs	12 inputs	18 inputs	6 inputs	12 inputs	18 inputs
MAE	2.2928	1.8823	1.8877	1.8900	1.9795	2.0926	2.2607
RMSE	3.7171	3.0354	3.0131	2.9989	3.2345	3.3828	3.4997
MXE	32.1926	30.1021	28.0905	28.2117	32.1493	45.1402	41.4134
NMAE	4.2459	3.4858	3.4957	3.5000	3.6658	3.8751	4.1864
MAPE	26.2563	14.3833	15.4588	16.5648	8.1082	14.4940	51.7620
MASE	1.0000	0.8210	0.8233	0.8243	0.8634	0.9127	0.9860
CC	0.9583	0.9718	0.9723	0.9726	0.9680	0.9653	0.9628

before the testing day. Consequently, in order to predict 2nd January 2011, 2nd January 2010 to 1st January 2011 are used as training period. The procedure continues until 31st December 2011. Table 1 shows the importance of each input in four different days of the year. The importance is measured following equation (9). The most important input is P, especially its last components $P(h:40)$ and $P(h:50)$, then the last components of $\overline{S}$. The direction inputs are not very important regardless of their position.

The hour-ahead forecast results appear in Fig. 11, which shows the hourly predicted signal through four testing weeks in different seasons. The predicted signal is the overall produced power of the farm, while taking all the 18 possible inputs, P, $\overline{S}$ and D. The chaotic nature of wind is obvious through the curves. The forecast is accurate in most cases, except sharp and sudden variations of the measured signal.

The results of Fig. 11 are not necessarily optimal, since some inputs may be disturbing. For instance, the direction inputs have weak correlation with the output and low importance measure, so they may be irrelevant. However, they are kept to observe their effect. Table 2 tries to assess the contribution of each input, by computing the seven proposed evaluation criteria while operating with 6, 12 and 18 inputs, respectively. The 6 inputs are $P(h)$ (all endogenous), the 12 inputs are $[P(h),\overline{S}(h)]$ (6 endogenous and 6 exogenous) and the 18 inputs are $[P(h),\overline{S}(h),D(h)]$ (6 endogenous and 12 exogenous). As for evaluation criteria, *MAE*, *RMSE* and *MXE* are in MW, *NMAE* and *MAPE* are in percent, and *MASE* and *CC* are with no unit. Table 2 shows the results for the whole year 2011. It compares the proposed RF model to classical neural network (ANN) and persistence (PER). The ANN is a generic feedforward network with one hidden layer of 30 neurons. It has the same inputs of RF. The persistence has one input which is the current average power $\tilde{P}(h)$. ANN and RF are better than PER in terms of accuracy in most cases. With only 6 inputs, ANN and RF have similar predictions. With 12 inputs, the RF remains almost the same; most criteria have changed a little bit. However, the ANN prediction quality decreases considerably. With 18 inputs, the quality of prediction dropped more in the ANN case, while the RF prediction is almost not affected. This reflects the immunity of RF to irrelevant inputs, which makes all its strength. Hence, the user of the RF predictor does neither need to remove non-significant inputs, nor need to search optimal configurations for its machine. These are two major advantages of RF over ANN. By respect to most evaluation criteria, the RF predictor achieved a decrease of 17% compared to persistence. To conclude, the RF profits from spatially averaged speed $\overline{S}$ and direction D once appended to the input vector. And even when they have no contribution, they do not degrade significantly the prediction accuracy.

It is quite difficult to compare results with data from other farms, since forecast models are generally site-specific and related to the prediction horizon [9]. However, it is known that the *NMAE* is generally high, ranging from 15% to 20% for day-ahead predictions, contrary to load forecast whose errors are typically in the range of 1%–3% [43]. Naturally, hour-ahead prediction errors will be lower, in the range of 2%–4%. The contribution of this paper is not to get only the minimum error, but also to assess the effect of new exogenous features like wind speed or direction, and to avoid harmful effects of any irrelevant input.

In addition to point forecast, it is possible to predict with confidence intervals. The quantile regression forest is used for this purpose. The concept is quite simple; instead of recording the average response of each leaf in the tree, it stores all responses in the leaf. The prediction does not then return only the mean of responses, but the full distribution of response values for each input. Through this distribution, it is possible to create prediction intervals by using the appropriate percentiles [63]. Quantile regression forests are tested on the whole year 2011, using the same online learning procedure. Fig. 12 gives the obtained prediction intervals from 4th to 7th February 2011. Results are pretty satisfying; the measured curve does not get over boundaries. The largest confidence interval contains almost all samples. Furthermore, in order to quantitatively assess the forecast quality, Table 3 gives the following criteria, for the whole year 2011: the number of measured samples inside each prediction interval, their percentages (by dividing by the overall number of samples; $365\times 24 = 8760$ samples), the average thickness of each interval, and the relative average thickness (by dividing by the wind farm capacity; 53.5 MW). As it can be seen, only 7.64% of measured samples are outside prediction intervals. Also, the relative average thickness of the largest interval does not exceed 15.90%. Naturally, prediction intervals should be as narrow as possible.

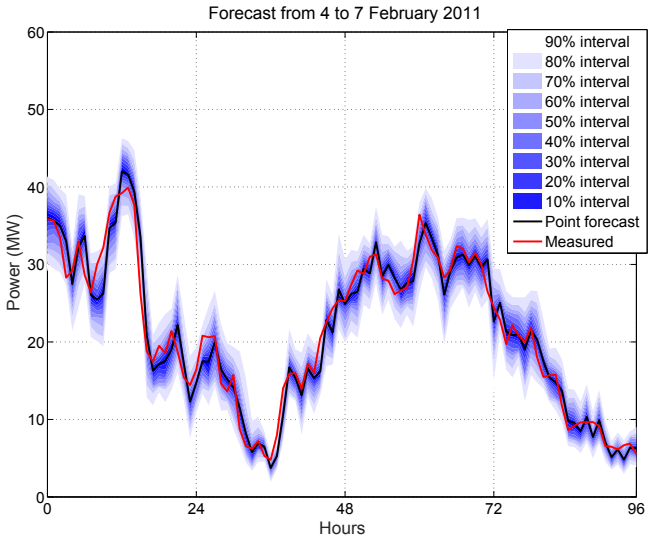

Fig. 12. Prediction intervals for hour-ahead forecast.

Table 3
Quantitative assessment of probabilistic forecast.

Interval	Number of samples	Percentage of samples	Average thickness (kW)	Relative thickness
10% interval	956	10,91%	545	1,01%
20% interval	1966	22,44%	1100	2,05%
30% interval	2915	33,27%	1680	3,14%
40% interval	3880	44,29%	2316	4,32%
50% interval	4760	54,33%	3012	5,62%
60% interval	5643	64,41%	3819	7,13%
70% interval	6475	73,91%	4810	8,99%
80% interval	7296	83,28%	6170	11,53%
90% interval	8090	92,35%	8509	15,90%
Out of bounds	670	7,64%		

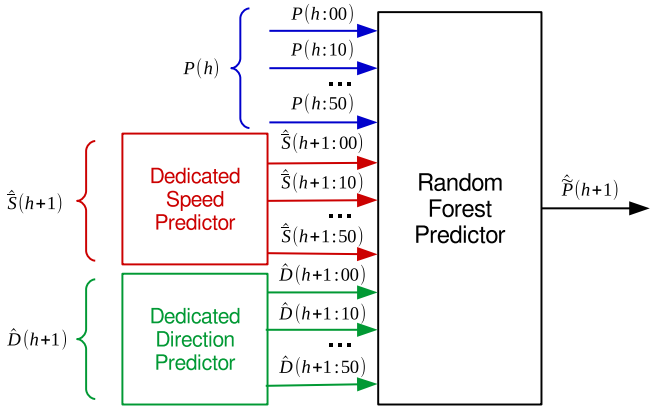

Fig. 13. Hybrid hour-ahead wind power forecasting model.

Table 2 does not assert that speed and direction have an effective contribution in the forecast accuracy, since the decrease in most criteria is not very interesting. The *RMSE* for example dropped from 3.03 to 2.99 after appending speed and direction to the inputs. Hence, a second test is performed by adding the future values of spatial average speed $\widehat{\bar{S}}(h+1)$ and direction $\widehat{D}(h+1)$ instead of current values. The future values are predicted by dedicated forecasters that may be physical or statistical, as shown in Fig. 13. Here, these values are exact assuming perfect forecasters, just to assess the real contribution of exogenous inputs. The model is then modified as follows:

$$X = \left[P(h), \widehat{\bar{S}}(h+1), \widehat{D}(h+1)\right]^T \quad (24)$$

where:

$$\widehat{\bar{S}}(h+1) = \left[\widehat{\bar{S}}(h+1:00), \widehat{\bar{S}}(h+1:10), \ldots, \widehat{\bar{S}}(h+1:50)\right] \quad (25)$$

$$\widehat{D}(h+1) = \left[\widehat{D}(h+1:00), \widehat{D}(h+1:10), \ldots, \widehat{D}(h+1:50)\right] \quad (26)$$

The modified model is considered hybrid, since it includes direct and indirect methods. The results of hybrid forecast are given in Table 4, for the whole year 2011. The effectiveness of exogenous inputs in this case is obvious, the *NMAE* for example dropped from 3.48 to 1.76 after adding $\widehat{\bar{S}}(h+1)$ and to 1.72 after adding $\widehat{D}(h+1)$. However, the ANN did not profit properly from these additional inputs. The advantage of RF here is the ability to gather information from any extra inputs without getting disturbed if these information are irrelevant. Furthermore, the hybrid model is a way to exploit the other sophisticated forecasting methods, such as optimized ANN/SVM, in order to benefit from their advantages in addition to RF.

6. Conclusion

Hour-ahead wind power forecast is possible using machine learning techniques such as ANN and RF, that outperform classic statistical methods and avoid the use of NWP complex models. The RF is an ensemble method for classification and regression problems, and it is used in this paper to predict wind power. It is relevant to select the suitable meteorological factors to include in the predictor, by analyzing their effects and by measures of correlation and importance. The spatial average of the wind speed and the wind direction in addition to past power values are chosen as inputs. After preprocessing data and selecting evaluation criteria, four periods of test across four seasons are chosen to assess the proposed model. Results show accurate forecast in most cases, when spatially averaged speed and direction are added to the inputs. However, adding these features to classical methods such as neural networks degrades enormously the results. The main contribution of this study is to demonstrate the ability of random forests to benefit from exogenous inputs that may carry additional information, without getting disturbed when these inputs are irrelevant. The user of the model can get the best from the random forest with no optimization process, which is a major advantage over other learning machines.

Table 4
Forecast results of the hybrid model.

Method	PER	RF			ANN		
Criterion	1 input	6 inputs	12 inputs	18 inputs	6 inputs	12 inputs	18 inputs
MAE	2.2928	1.8823	0.9554	0.9325	1.9795	1.1983	1.4587
RMSE	3.7171	3.0354	1.7637	1.7615	3.2345	2.0686	2.5792
MXE	32.1926	30.1021	19.7152	19.5989	32.1493	22.8434	38.9105
NMAE	4.2459	3.4858	1.7693	1.7269	3.6658	2.2190	2.7012
MAPE	26.2563	14.3833	10.0375	10.2980	8.1082	23.4266	38.0565
MASE	1.0000	0.8210	0.4167	0.4067	0.8634	0.5226	0.6362
CC	0.9583	0.9718	0.9911	0.9911	0.9680	0.9877	0.9807

Acknowledgment

The authors want to thank the Tunisian Company of Electricity and Gas (STEG) for providing power and meteorological data of Sidi Daoud wind farm, which are used in this work.

References

[1] GWEC, Global Wind Statistics 2015, Tech. rep., Global Wind Energy Council, February 2016 http://www.gwec.net/wp-content/uploads/vip/GWEC-PRstats-2015_LR.pdf.

[2] P.S. Georgilakis, Technical challenges associated with the integration of wind power into power systems, Renew. Sustain. Energy Rev. 12 (3) (2008) 852–863, http://dx.doi.org/10.1016/j.rser.2006.10.007. http://www.sciencedirect.com/science/article/pii/S1364032106001250.

[3] M.M. Alam, S. Rehman, L. Al-Hadhrami, J. Meyer, Extraction of the inherent nature of wind speed using wavelets and {FFT}, Energy Sustain. Dev. 22 (0) (2014) 34–47, http://dx.doi.org/10.1016/j.esd.2014.02.004 wind Power Special Issue, http://www.sciencedirect.com/science/article/pii/S0973082614000167.

[4] X. Zhao, S. Wang, T. Li, Review of evaluation criteria and main methods of wind power forecasting, Energy Procedia 12 (0) (2011) 761–769, http://dx.doi.org/10.1016/j.egypro.2011.10.102. http://www.sciencedirect.com/science/article/pii/S187661021101928X.

[5] I. Colak, S. Sagiroglu, M. Yesilbudak, Data mining and wind power prediction: a literature review, Renew. Energy 46 (0) (2012) 241–247, http://dx.doi.org/10.1016/j.renene.2012.02.015. http://www.sciencedirect.com/science/article/pii/S0960148112001541.

[6] M. Qadrdan, J. Wu, N. Jenkins, J. Ekanayake, Operating strategies for a gb integrated gas and electricity network considering the uncertainty in wind power forecasts, Sustain. Energy, IEEE Trans. 5 (1) (2014) 128–138, http://dx.doi.org/10.1109/TSTE.2013.2274818. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6589136.

[7] H. Yang, J. Qiu, K. Meng, J.H. Zhao, Z.Y. Dong, M. Lai, Insurance strategy for mitigating power system operational risk introduced by wind power forecasting uncertainty, Renew. Energy 89 (2016) 606–615, http://dx.doi.org/10.1016/j.renene.2015.12.007. http://www.sciencedirect.com/science/article/pii/S0960148115305176.

[8] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, Z. Yan, A review on the forecasting of wind speed and generated power, Renew. Sustain. Energy Rev. 13 (4) (2009) 915–920, http://dx.doi.org/10.1016/j.rser.2008.02.002. http://www.sciencedirect.com/science/article/pii/S1364032108000282.

[9] A. Tascikaraoglu, M. Uzunoglu, A review of combined approaches for prediction of short-term wind speed and power, Renew. Sustain. Energy Rev. 34 (0) (2014) 243–254, http://dx.doi.org/10.1016/j.rser.2014.03.033. http://www.sciencedirect.com/science/article/pii/S1364032114001944.

[10] A.M. Foley, P.G. Leahy, A. Marvuglia, E.J. McKeogh, Current methods and advances in forecasting of wind power generation, Renew. Energy 37 (1) (2012) 1–8, http://dx.doi.org/10.1016/j.renene.2011.05.033. http://www.sciencedirect.com/science/article/pii/S0960148111002850.

[11] J. Jung, R.P. Broadwater, Current status and future advances for wind speed and power forecasting, Renew. Sustain. Energy Rev. 31 (2014) 762–777, http://dx.doi.org/10.1016/j.rser.2013.12.054. http://www.sciencedirect.com/science/article/pii/S1364032114000094.

[12] Y. Zhang, J. Wang, X. Wang, Review on probabilistic forecasting of wind power generation, Renew. Sustain. Energy Rev. 32 (2014) 255–270, http://dx.doi.org/10.1016/j.rser.2014.01.033. http://www.sciencedirect.com/science/article/pii/S1364032114000446.

[13] L. Lazic, G. Pejanovic, M. Zivkovic, Wind forecasts for wind power generation using the eta model, Renew. Energy 35 (6) (2010) 1236–1243, http://dx.doi.org/10.1016/j.renene.2009.10.028. http://www.sciencedirect.com/science/article/pii/S0960148109004546.

[14] W. Mahoney, K. Parks, G. Wiener, Y. Liu, W. Myers, J. Sun, L. Delle Monache, T. Hopson, D. Johnson, S. Haupt, A wind power forecasting system to optimize grid integration, Sustain. Energy, IEEE Trans. 3 (4) (2012) 670–682, http://dx.doi.org/10.1109/TSTE.2012.2201758. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6237561.

[15] S. Al-Yahyai, A. Gastli, Y. Charabi, Probabilistic wind speed forecast for wind power prediction using pseudo ensemble approach, in: power and Energy (PECon), in: IEEE International Conference on, 2012, 2012, pp. 127–132, http://dx.doi.org/10.1109/PECon.2012.6450191. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6450191.

[16] F. Cassola, M. Burlando, Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output, Appl. Energy 99 (0) (2012) 154–166, http://dx.doi.org/10.1016/j.apenergy.2012.03.054. http://www.sciencedirect.com/science/article/pii/S0306261912002747.

[17] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, G. Kallos, Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering, J. Wind Eng. Ind. Aerodyn. 96 (12) (2008) 2348–2362, http://dx.doi.org/10.1016/j.jweia.2008.03.013. http://www.sciencedirect.com/science/article/pii/S0167610508001074.

[18] C. Stathopoulos, A. Kaperoni, G. Galanis, G. Kallos, Wind power prediction based on numerical and statistical models, J. Wind Eng. Ind. Aerodyn. 112 (0) (2013) 25–38, http://dx.doi.org/10.1016/j.jweia.2012.09.004. http://www.sciencedirect.com/science/article/pii/S0167610512002486.

[19] Z. Zheng, Y. Chen, M. Huo, B. Zhao, An overview: the development of prediction technology of wind and photovoltaic power generation, Energy Procedia 12 (0) (2011) 601–608, http://dx.doi.org/10.1016/j.egypro.2011.10.081 the Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011), http://www.sciencedirect.com/science/article/pii/S1876610211019072.

[20] J. Torres, A. Garcia, M.D. Blas, A.D. Francisco, Forecast of hourly average wind speed with {ARMA} models in navarre (Spain), Sol. Energy 79 (1) (2005) 65–77, http://dx.doi.org/10.1016/j.solener.2004.09.013. http://www.sciencedirect.com/science/article/pii/S0038092X04002877.

[21] A. Sfetsos, A novel approach for the forecasting of mean hourly wind speed time series, Renew. Energy 27 (2) (2002) 163–174, http://dx.doi.org/10.1016/S0960-1481(01)00193-8. http://www.sciencedirect.com/science/article/pii/S0960148101001938.

[22] L. Xie, Y. Gu, X. Zhu, M. Genton, Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch, Smart Grid, IEEE Trans. 5 (1) (2014) 511–520, http://dx.doi.org/10.1109/TSG.2013.2282300. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6616027.

[23] M.A. Mohandes, S. Rehman, T.O. Halawani, A neural networks approach for wind speed prediction, Renew. Energy 13 (3) (1998) 345–354, http://dx.doi.org/10.1016/S0960-1481(98)00001-9. http://www.sciencedirect.com/science/article/pii/S0960148198000019.

[24] L.J. Ricalde, G. Catzin, A.Y. Alanis, E.N. Sanchez, Higher order wavelet neural networks with kalman learning for wind speed forecasting, in: computational Intelligence Applications in Smart Grid (CIASG), in: IEEE Symposium on, 2011, 2011, pp. 1–6, http://dx.doi.org/10.1109/CIASG.2011.5953332. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5953332.

[25] W. Wei, G. Wu, M. Yang, Y. Zhang, S. Qiu, A. Sun, Short-term forecasting for wind speed based on wavelet decomposition and lmbp neural network, in: electric Utility Deregulation and Restructuring and Power Technologies (DRPT), in: 4th International Conference on, 2011, 2011, pp. 1126–1131, http://dx.doi.org/10.1109/DRPT.2011.5994064. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5994064.

[26] K. Bhaskar, S. Singh, Awnn-assisted wind power forecasting using feed-forward neural network, Sustain. Energy, IEEE Trans. 3 (2) (2012) 306–315, http://dx.doi.org/10.1109/TSTE.2011.2182215. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6170987.

[27] Z. Men, E. Yee, F.-S. Lien, D. Wen, Y. Chen, Short-term wind speed and power forecasting using an ensemble of mixture density neural networks, Renew. Energy 87 (Part 1) (2016) 203–211, http://dx.doi.org/10.1016/j.renene.2015.10.014. http://www.sciencedirect.com/science/article/pii/S0960148115303645.

[28] H. Liu, H. Tian, X. Liang, Y. Li, New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks, Renew. Energy 83 (2015) 1066–1075, http://dx.doi.org/10.1016/j.renene.2015.06.004. http://www.sciencedirect.com/science/article/pii/S0960148115300276.

[29] T. Barbounis, J. Theocharis, Locally recurrent neural networks for wind speed prediction using spatial correlation, Inf. Sci. 177 (24) (2007) 5775–5797, http://dx.doi.org/10.1016/j.ins.2007.05.024. http://www.sciencedirect.com/science/article/pii/S0020025507002514.

[30] T. Barbounis, J. Theocharis, A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation, Neurocomputing 70 (7–9) (2007) 1525–1542 advances in Computational Intelligence and Learning, 14th European Symposium on Artificial Neural Networks 2006. doi:http://dx.doi.org/10.1016/j.neucom.2006.01.032. URL, http://www.sciencedirect.com/science/article/pii/S0925231206002505.

[31] S. Salcedo-Sanz, E.G. Ortiz-Garcia, A.M. Perez-Bellido, A. Portilla-Figueras, L. Prieto, Short term wind speed prediction based on evolutionary support vector regression algorithms, Expert Syst. Appl. 38 (4) (2011) 4052–4057, http://dx.doi.org/10.1016/j.eswa.2010.09.067. http://www.sciencedirect.com/science/article/pii/S0957417410010249.

[32] M. Mohandes, T. Halawani, S. Rehman, A.A. Hussain, Support vector machines for wind speed prediction, Renew. Energy 29 (6) (2004) 939–947, http://dx.doi.org/10.1016/j.renene.2003.11.009. http://www.sciencedirect.com/science/article/pii/S0960148103003860.

[33] J. Hu, J. Wang, G. Zeng, A hybrid forecasting approach applied to wind speed time series, Renew. Energy 60 (0) (2013) 185–194, http://dx.doi.org/10.1016/j.renene.2013.05.012. http://www.sciencedirect.com/science/article/pii/S0960148113002577.

[34] Y. Ren, P. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, Sustain. Energy IEEE Trans. 6 (1) (2015) 236–244, http://dx.doi.org/10.1109/TSTE.2014.2365580. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6979269.

[35] G. Santamara-Bonfil, A. Reyes-Ballesteros, C. Gershenson, Wind speed forecasting for wind farms: a method based on support vector regression, Renew. Energy 85 (2016) 790–809, http://dx.doi.org/10.1016/j.renene.2015.07.004. http://www.sciencedirect.com/science/article/pii/S0960148115301014.

[36] Y. Lin, U. Kruger, J. Zhang, Q. Wang, L. Lamont, L.E. Chaar, Seasonal analysis and prediction of wind energy using random forests and arx model structures, IEEE Trans. Control Syst. Technol. 23 (5) (2015) 1994–2002, http://dx.doi.org/

10.1109/TCST.2015.2389031. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7027784.
[37] S. Alessandrini, S. Sperati, P. Pinson, A comparison between the {ECMWF} and {COSMO} ensemble prediction systems applied to short-term wind power forecasting on real data, Appl. Energy 107 (0) (2013) 271–280, http://dx.doi.org/10.1016/j.apenergy.2013.02.041. http://www.sciencedirect.com/science/article/pii/S0306261913001499.
[38] H. Pousinho, V. Mendes, J. Catalao, A hybrid pso-anfis approach for short-term wind power prediction in Portugal, Energy Convers. Manag. 52 (1) (2011) 397–402, http://dx.doi.org/10.1016/j.enconman.2010.07.015. http://www.sciencedirect.com/science/article/pii/S0196890410003080.
[39] L. LI, Y. Qian LIU, Y. Ping YANG, S. HAN, Y. Mei WANG, A physical approach of the short-term wind power prediction based on {CFD} pre-calculated flow fields, J. Hydrodyn. Ser. B 25 (1) (2013) 56–61, http://dx.doi.org/10.1016/S1001-6058(13)60338-8. http://www.sciencedirect.com/science/article/pii/S1001605813603388.
[40] J. Wu, B. Zhang, H. Li, Z. Li, Y. Chen, X. Miao, Statistical distribution for wind power forecast error and its application to determine optimal size of energy storage system, Int. J. Electr. Power Energy Syst. 55 (0) (2014) 100–107, http://dx.doi.org/10.1016/j.ijepes.2013.09.003. http://www.sciencedirect.com/science/article/pii/S0142061513003761.
[41] X. An, D. Jiang, C. Liu, M. Zhao, Wind farm power prediction based on wavelet decomposition and chaotic time series, Expert Syst. Appl. 38 (9) (2011) 11280–11285, http://dx.doi.org/10.1016/j.eswa.2011.02.176. http://www.sciencedirect.com/science/article/pii/S0957417411004003.
[42] M. Ozkan, P. Karagoz, A novel wind power forecast model: statistical hybrid wind power forecast technique (shwip), Ind. Inf. IEEE Trans. PP 99 (2015), http://dx.doi.org/10.1109/TII.2015.2396011, 1–1, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=7018961.
[43] L. Yang, M. He, J. Zhang, V. Vittal, Support-vector-machine-enhanced markov model for short-term wind power forecast, Sustain. Energy IEEE Trans. 6 (3) (2015) 791–799, http://dx.doi.org/10.1109/TSTE.2015.2406814. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7081774.
[44] J. Wu, T. Ji, M. Li, P. Wu, Q. Wu, Multistep wind power forecast using mean trend detector and mathematical morphology-based local predictor, Sustain. Energy IEEE Trans. 6 (4) (2015) 1216–1223, http://dx.doi.org/10.1109/TSTE.2015.2424856. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7106520.
[45] Q. Xu, D. He, N. Zhang, C. Kang, Q. Xia, J. Bai, J. Huang, A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining, Sustain. Energy IEEE Trans. 6 (4) (2015) 1283–1291, http://dx.doi.org/10.1109/TSTE.2015.2429586. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7116614.
[46] N. Amjady, F. Keynia, H. Zareipour, Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization, Sustain. Energy IEEE Trans. 2 (3) (2011) 265–276, http://dx.doi.org/10.1109/TSTE.2011.2114680. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5713273.
[47] G. Osrio, J. Matias, J. Catalo, Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information, Renew. Energy 75 (2015) 301–307, http://dx.doi.org/10.1016/j.renene.2014.09.058. http://www.sciencedirect.com/science/article/pii/S0960148114006284.
[48] L. Han, C.E. Romero, Z. Yao, Wind power forecasting based on principle component phase space reconstruction, Renew. Energy 81 (2015) 737–744, http://dx.doi.org/10.1016/j.renene.2015.03.037. http://www.sciencedirect.com/science/article/pii/S0960148115002190.
[49] J. Heinermann, O. Kramer, Machine learning ensembles for wind power prediction, Renew. Energy 89 (2016) 671–679, http://dx.doi.org/10.1016/j.renene.2015.11.073. http://www.sciencedirect.com/science/article/pii/S0960148115304894.
[50] S. Alessandrini, L.D. Monache, S. Sperati, J. Nissen, A novel application of an analog ensemble for short-term wind power forecasting, Renew. Energy 76 (2015) 768–781, http://dx.doi.org/10.1016/j.renene.2014.11.061. http://www.sciencedirect.com/science/article/pii/S0960148114007915.
[51] J. Yan, Y. Liu, S. Han, M. Qiu, Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine, Renew. Sustain. Energy Rev. 27 (0) (2013) 613–621, http://dx.doi.org/10.1016/j.rser.2013.07.026. http://www.sciencedirect.com/science/article/pii/S1364032113004711.
[52] T.H.M. El-Fouly, E. El-Saadany, M. Salama, One day ahead prediction of wind speed and direction, Energy Convers. IEEE Trans. 23 (1) (2008) 191–201, http://dx.doi.org/10.1109/TEC.2007.905069. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4433186.
[53] E. Erdem, J. Shi, {ARMA} based approaches for forecasting the tuple of wind speed and direction, Appl. Energy 88 (4) (2011) 1405–1414, http://dx.doi.org/10.1016/j.apenergy.2010.10.031. http://www.sciencedirect.com/science/article/pii/S0306261910004332.
[54] S. Velzquez, J.A. Carta, J. Matas, Influence of the input layer signals of {ANNs} on wind power estimation for a target site: a case study, Renew. Sustain. Energy Rev. 15 (3) (2011) 1556–1566, http://dx.doi.org/10.1016/j.rser.2010.11.036. http://www.sciencedirect.com/science/article/pii/S1364032110004053.
[55] Y. Zhang, J. Wang, X. Wang, Review on probabilistic forecasting of wind power generation, Renew. Sustain. Energy Rev. 32 (2014) 255–270, http://dx.doi.org/10.1016/j.rser.2014.01.033. http://www.sciencedirect.com/science/article/pii/S1364032114000446.
[56] L. Breiman, et al., Classification and Regression Trees, Chapman & Hall, New York, 1984.
[57] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324. http://dx.doi.org/10.1023/A%3A1010933404324.
[58] R. Genuer, Forts alatoires : aspects thoriques, slection de variables et applications, Ph.D. thesis, Universit Paris Sud - Paris XI, 2010, https://hal.archives-ouvertes.fr/tel-00550989/document.
[59] M. Lydia, A. Selvakumar, S. Kumar, G. Kumar, Advanced algorithms for wind turbine power curve modeling, Sustain. Energy IEEE Trans. 4 (3) (2013) 827–835, http://dx.doi.org/10.1109/TSTE.2013.2247641. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6491505.
[60] F. Pelletier, C. Masson, A. Tahan, Wind turbine power curve modelling using artificial neural network, Renew. Energy 89 (2016) 207–214, http://dx.doi.org/10.1016/j.renene.2015.11.065. http://www.sciencedirect.com/science/article/pii/S096014811530481X.
[61] U. Focken, M. Lange, K. Mnnich, H.-P. Waldl, H.G. Beyer, A. Luig, Short-term prediction of the aggregated power output of wind farms-a statistical analysis of the reduction of the prediction error by spatial smoothing effects, J. Wind Eng. Ind. Aerodyn. 90 (3) (2002) 231–246, http://dx.doi.org/10.1016/S0167-6105(01)00222-7. http://www.sciencedirect.com/science/article/pii/S0167610501002227.
[62] M. Lydia, S.S. Kumar, Wind farm power prediction: an overview, in: The Fifth International Symposium on Computational Wind Engineering (CWE2010), Chapel Hill, North Carolina, USA, 2010. ftp://ftp.atdd.noaa.gov/pub/cwe2010/Files/Papers/291_edwin.pdf.
[63] N. Meinshausen, Quantile regression forests, J. Mach. Learn. Res. 7 (2006) 983–999. http://dl.acm.org/citation.cfm?id=1248547.1248582.