Energy Conversion and Management 118 (2016) 193–203

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Using data-driven approach for wind power prediction: A comparative study

Ehsan Taslimi Renani [a,b], Mohamad Fathi Mohamad Elias [a,*], Nasrudin Abd. Rahim [a]

[a] UM Power Energy Dedicated Advanced Centre (UMPEDAC), Level 4, Wisma R&D, University of Malaya, Jalan Pantai Baharu, 59990 Kuala Lumpur, Malaysia
[b] Institute of Graduate Studies, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form 23 March 2016
Accepted 25 March 2016
Available online 2 April 2016

Keywords:
Adaptive neuro fuzzy inference systems
Back tracking search algorithm
Data mining
Wind power prediction
Wind speed prediction

ABSTRACT

Although wind energy is intermittent and stochastic in nature, it is increasingly important in the power generation due to its sustainability and pollution-free. Increased utilization of wind energy sources calls for more robust and efficient prediction models to mitigate uncertainties associated with wind power. This research compares two different approaches in wind power forecasting which are indirect and direct prediction methods. In indirect method, several times series are applied to forecast the wind speed, whereas the logistic function with five parameters is then used to forecast the wind power. In this study, backtracking search algorithm with novel crossover and mutation operators is employed to find the best parameters of five-parameter logistic function. A new feature selection technique, combining the mutual information and neural network is proposed in this paper to extract the most informative features with a maximum relevancy and minimum redundancy. From the comparative study, the results demonstrate that, in the direct prediction approach where the historical weather data are used to predict the wind power generation directly, adaptive neuro fuzzy inference system outperforms five data mining algorithms namely, random forest, M5Rules, *k*-nearest neighbor, support vector machine and multilayer perceptron. Moreover, it is also found that the mean absolute percentage error of the direct prediction method using adaptive neuro fuzzy inference system is 1.47% which is approximately less than half of the error obtained with the indirect prediction methods.

1. Introduction

Due to the depletion of fossil fuels continuing past few decades, renewable energy sources such as solar and wind are becoming increasingly important since they are clean and sustainable energy sources [1]. However, it is well known that wind power generation is stochastic in nature which leads to great uncertainties, difficulties and challenges in electrical power systems. The variation in wind power can jeopardize power system quality and stability as well as affect the market participants who bear the economic losses due to failure of producing the contracted amount of energy. Hence, there is a great need of tools that accurately forecast wind power generation and mitigate the undesirable influences of integrating wind energy into electric power grids.

In the past, researches on wind power forecasting are generally divided into two groups: physical and statistical approaches. Physical methods employ physical laws governing atmospheric behavior and utilize a great number of meteorology information such as surface roughness, orography, obstacles, pressure and temperature to estimate the local wind speed and direction. Statistical methods focus on vast historical data and attempt to tune model parameters to minimize error between the predicted and the observed power. The statistical models usually exhibit more accurate results than physical approaches in short term prediction [2]. In [3], an adaptive statistical technique such as Kalman filter was employed to improve the wind speed prediction and in turn the wind power forecasting. A novel ARIMA model was proposed in [4] by introducing a limiter into the model to present the upper and lower bound of wind power generation. Prediction of the wind speed and direction tuple using four approaches based on autoregressive moving average (ARMA) model was conducted in [5]. The best results in wind direction forecasting obtained when wind speed decomposed into lateral and longitudinal components, while, in wind speed prediction, the traditional-linked ARMA showed better accuracy. In [6], fractional ARIMA (FARIMA) was applied to forecast wind speeds over 24- and 48-h horizons and significant improvement compared to persistence method was observed.

Wind generation forecasting could progress further with the advent of machine learning. Artificial neural networks (ANNs) [7] and support vector machines (SVMs) [8] are among intelligence

* Corresponding author.
E-mail address: fathi@um.edu.my (M.F.M. Elias).

http://dx.doi.org/10.1016/j.enconman.2016.03.078
0196-8904/

Nomenclature

n	number of samples
$\hat{y}$	predicted value
y	observed value
N	dimension of the search space
Pop	population matrix
PS	population size
U	uniform distribution
Pop_{old}	historical population matrix
F	amplitude control function of search-direction matrix
TR	final population matrix
$Mutant$	trial population matrix
map	binary matrix
$mixrate$	BSA's control parameter
i,j	common indices
p	order of the AR
q	order of the MA
φ	AR coefficient
L	lag operator
θ	MA coefficient
ε_t	white noise
S_t'	smoothed value at time t
b_t	best estimation of the trend at time t
η, λ	smoothing parameters
x	input value of the 5-PL
δ	vector parameter of 5-PL
$H(X)$	entropy of random variable X
$P(X)$	probability function of X
$H(X,Y)$	joint entropy of X and Y
$P(X,Y)$	joint probability distribution of variables X and Y
$H(X\|Y)$	conditional entropy
$I(X;Y)$	mutual information of variables X and Y
WP	predicted wind power
S	wind speed
T	temperature
H	humidity
D	wind direction
NL_{WP}	number of lag order for the output power
NL_S	number of lag order for the wind speed
NL_T	number of lag order for the temperature
NL_D	number of lag order for the wind direction
NL_H	number of lag order for the humidity
t	neural network output
h	number of hidden units
$\alpha_1, \alpha_2, \beta$	coefficient of penalty term
V	number of attributes selected in the first stage of feature selection

algorithms which attempt to discover non-linear relation between input and output data set, widely used in the last decade. Although, they outperform time-series model, the former suffer the overfitting problem and falling into local extremum. Nowadays many researchers tend to combine different models to enhance forecasting accuracy [9]. The integration of physical strategy and ANN was suggested in [10] and it outperformed individual ANN in terms of accuracy, however, it sacrificed operation cost and speed. In [11], hybrid method based on wavelet analysis and SVM was provided which was superior to radial basis function (RBF)-SVM in terms of accuracy and computational time. In [12], authors used imperialistic competitive algorithm to update weights of NN in 3–6 h ahead prediction. In [13], combination of Gaussian process and numerical weather prediction was presented and comparative results demonstrated the accuracy of forecasting improved by at least 9% as compared to ANN. In [14], wavelet NN was employed and in order to boost learning process a new optimization algorithm, named improved Clonal selection algorithm, was proposed. The model achieves improvement compared to other search algorithms such as cuckoo search algorithm, particle swarm optimization and simulated annealing. One of the popular hybrid models which is being focused in this paper is adaptive neuro fuzzy inference systems (ANFIS). ANFIS takes the advantage of both fuzzy control systems and NN and it has been successfully applied in wind power and speed prediction due to its capability for automatic learning [15].

In this paper two different approaches, direct and indirect prediction methods are investigated for predicting wind power. In [16], the author also used indirect power forecasting model, however the model was based on the availability of power in the wind which is cumbersome and give inaccurate result [17]. Hence a different indirect prediction method based on wind turbine power curve modeling is applied. Here, an effective feature selection technique is proposed to identify the relevant features and remove the less significant data. The proposed feature selection method is capable to evaluate the nonlinear dependencies between the wind power and its input features, as well as increase learning accuracy and comprehensibility. Experiments were conducted with one year 5-min interval data sets from a wind farm in Iran.

This paper is organized as follows. Section 2 presents backtracking search algorithm and several wind speed prediction models. It also provides the numerical results based on wind turbine power curve from a real-world case study. In Section 3, at first the proposed feature selection technique is introduced. Then, several direct prediction models are implemented and the results are discussed. Finally concluding remark is given in Section 4.

2. Indirect wind power prediction

2.1. Real data

The data used in this study was collected from a wind farm in north of Iran. The observation was reported at 5-min interval including wind speed, wind direction, temperature and generated power over the period of 12 months starting from 09.04.2013 15:05:00 P.M to 04.03.2014 23:55:00 P.M, providing 94,270 samples in total. The observations were divided into two data set which are data set 1 and data set 2 comprising 86,206 and 8064 instances respectively. Data set 1 is used to develop wind speed and wind power forecasting models. Data set 2 contains four weeks test data which are randomly selected, corresponding to the four seasons in a year: first week of November (fall), third week of May (spring), third week of August (summer), fourth week of February (winter). Note that the collected data includes some noises due to several reasons such as maintenance issues, sensor malfunction, and environment issues (dirt, ice). Therefore, in order to obtain an accurate prediction, noisy data should be filtered out. To do so, the wind speeds are divided into small and equal intervals (0.2 m/s) as shown in Fig. 1(a). Then the mean μ and standard deviation σ of the corresponding wind power for each wind speed interval are calculated and those wind powers located outside the range of $[\mu-\sigma, \mu+\sigma]$ are discarded (blue circles). In the next step, for each wind speed interval, wind power is equally divided into ten intervals. The probability of each wind power interval is calculated and sorted in descending order. Given a threshold TH, the first z intervals ($z < 10$) representing greater occurence of power are selected as shown with pink star in Fig. 1(b).

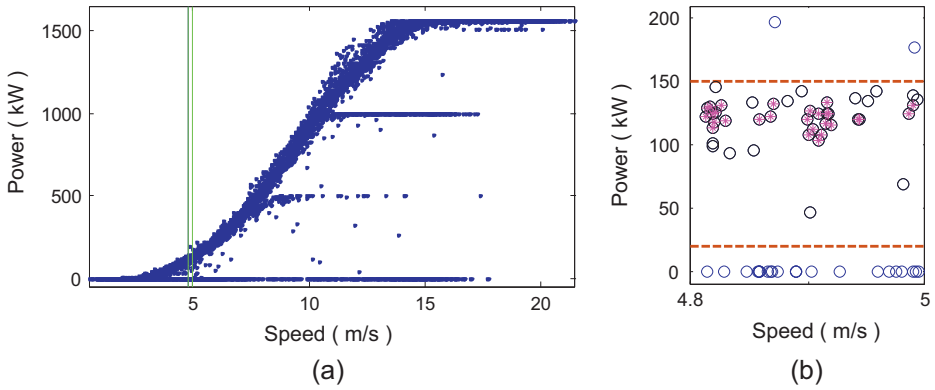

Fig. 1. Process of filtering, (a) 5-min average value of observed wind speed and power; (b) outside the range of $[\mu-\sigma, \mu+\sigma]$ (blue circle), unselected (black circle), selected (pink asterisk). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To quantify the prediction performance several evaluation criteria are employed; mean absolute error (MAE), mean absolute percentage error (MAPE), and standard deviation error (SDE) as given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{1}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{\hat{y}_i - y_i}{y_i}\right|\right) * 100 \tag{2}$$

$$\mathrm{SDE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i - \mathrm{MAE})^2 \tag{3}$$

where $\hat{y}_i$ is the predicted power, y_i is the observed power and n is the number of samples forecasted.

2.2. Backtracking search optimization algorithm (BSA)

BSA is an effective stochastic search algorithm for solving numerical optimization problems which is capable of mitigating issues associated with other evolutionary algorithms such as premature convergence and sensitivity to parameters. BSA has a simple structure which consists of a single control parameter that is insensitive in terms of initial value [18]. The procedure of BSA is described as follows:

1. *Representation of the population:* In an N-dimensional search space, the individual is represented as a vector length of N. The complete population with PS individuals randomly initialized within the maximum and minimum limits of the variables is shown in matrix as below:

$$Pop = [X_1, X_2, \ldots, X_i, \ldots, X_{PS}] \tag{4}$$

where $X_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]^T$ illustrates i-th individual as one of the solutions to parameter estimation.

2. *Selection-I:* In this stage BSA determines historical population or old population which is later used to calculate the search direction. The initial old population is defined as:

$$Pop_{old} \sim U(l_k, u_k) \tag{5}$$

where U is the uniform distribution and $k = \{1, 2, \ldots, N\}$. There is an option in BSA algorithm to update the historical population matrix at the beginning of each iteration through if-then rule. In other words, BSA has a memory which can store a population from randomly chosen previous generation as historical population. After an update on historical population representing experiences obtained in the previous generations, the order of its individuals randomly changed to determine the Pop_{old}.

3. *Mutation:* BSA's mutation strategy employs only one individual from a previous population. To generate the initial form of offspring, (6) is applied in the mutation process.

$$Mutant = Pop + F \cdot (Pop_{old} - Pop) \tag{6}$$

where F controls the amplitude of the search direction. The scale factor F substantially enhances the ability of BSA in solving optimization problems because it can produce greater amplitude essential in searching global optimum and low amplitude essential in searching local optimum.

4. *Crossover:* In this stage, initial form of offspring, *Mutant*, changes to the final form of offspring, *TR*, through crossover operator. Crossover mechanism determines those individuals that would be manipulated by a binary integer matrix (*map*) of size *P.N*. For instance, where $j = \{1, 2, \ldots, i, \ldots, PS\}$ and $k = \{1, 2, \ldots, N\}$, if $map = 1$, the individual TR_{ij} is updated to Pop_{ij} which means $TR_{ij} = Pop_{ij}$. There is a parameter called '*mix-rate*' which controls the number of elements of individuals to be engaged in the crossover operation. At the end of the crossover process those individuals that exceed the search-space limit would be generated through boundary control strategy.
5. *Section-II:* All individuals of population *TR* and *Pop* are evaluated by the objective function. Individuals of population *TR* having better fitness value than the corresponding individuals of population *Pop* are used to update *Pop*.
6. *Termination:* In BSA the maximum number of iteration can be chosen as stopping criteria or when there is a failure in finding better solution than the existing one during the last particular number of function evaluation.

2.3. Wind speed prediction models

Generally speaking, there are two approaches for wind speed forecasting, namely, time series-based and weather based. The latter employs physical data such as temperature, pressure and topography information to forecast future wind speed, however, it does not yield reliable results in a short forecasting horizon. Hence, since the focus of this paper is on the short term prediction, time series is used to forecast the wind speed. Time series-based methods attempt to predict future value through recursive techniques.

2.3.1. Auto-regressive moving average (ARMA)

This model incorporates into prediction not only past values of the data but also past value of the prediction residual. This model,

Table 1
Parameters of DES model for data set 2.

Parameter	Test week			
	Winter	Spring	Summer	Fall
η	0.998	0.996	0.999	0.998
λ	0.014	0.011	0.002	0.003

Table 2
Statistical error measures of the four models in wind speed prediction for data set 2.

Model	Error	Test week				Average
		Winter	Spring	Summer	Fall	
DES	MAPE	0.34	0.27	0.68	0.45	0.43
	MAE	0.03	0.03	0.09	0.02	0.04
	SDE	0.06	0.06	0.10	0.02	0.06
ARMA	MAPE	4.7	6.12	7.02	8.25	6.52
	MAE	0.50	0.78	0.93	0.33	0.63
	SDE	0.88	1.32	1.56	0.57	1.08
Persistence	MAPE	6.93	6.85	7.17	8.34	7.32
	MAE	0.46	0.85	1.00	0.48	0.69
	SDE	0.87	1.46	1.68	0.92	1.23

known as ARMA (p, q), is traditionally well suited to capture short range correlation which is expressed as follows:

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{7}$$

where p is order of the auto-regression (AR) part, q is order of the moving average (MA) part, φ_i is the AR coefficient, θ_i is the MA coefficient, L is the lag operator and ε_t is the white noise process with zero mean and variance σ^2.

2.3.2. Double exponential smoothing (DES)

In this technique, the old data are assigned with smaller weight than recent observations. The idea behind is that the future might be more dependent on recent past observation than on distant past [19]. DES attempts to compute local estimates of level and trend using two constant parameters, defined as follows:

$$\begin{aligned} S'_t &= \eta y_t + (1-\eta)(S'_{t-1}+b_{t-1}) \\ b_t &= \lambda(S'_t - S'_{t-1}) + (1-\lambda)b_{t-1} \\ \hat{y}_{t+1} &= S'_t + b_t \\ S'_1 &= y_1 \\ b_1 &= [(y_2-y_1)+(y_3-y_2)+(y_4-y_3)]/3 \end{aligned} \tag{8}$$

where S'_t is the smoothed value at time t, b_t is the best estimation of the trend at time t, $\hat{y}_{t+1}$ is the predicted value, η and λ are two smoothing parameters with values between 0 and 1.

2.3.3. Persistence method

This model is a classical benchmark model, known as naïve predictor, which simply assume that forecasted value at time t is similar to the last measurement, ($\hat{y}_t = y_{t+1}$). It does not require any parameter setting or exogenous variables and usually outperforms NWP in short horizon prediction [20].

2.4. Performance of time-series models

To confirm that the data used in ARMA model is stationary or not, visual inspection of the auto-correlation function (ACF) is performed in this study. If there is a fast uptrend in ACF plot this means that the data is stationary, but if the ACF presents gradual downward trend the data is non-stationary which differencing is required to make it stationary in the mean. After structure of ARMA (p, q) is defined, the estimation of model parameters is performed by maximum likelihood estimation using the Kalman filter in conjunction with Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. Based on close inspection performed, ARMA (2,1) is chosen for wind speed prediction. In order to obtain the unknown

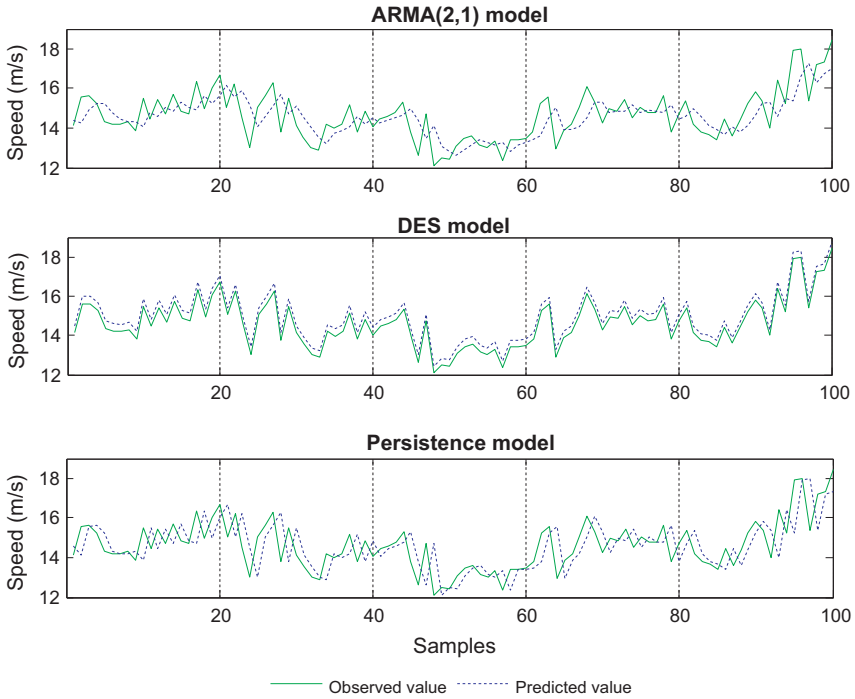

Fig. 2. The actual and predicted wind speed for the first 100 samples of the third week of August.

Table 3
Parameters of 5-PL model based on BSA for data set 2.

Test week	Vector parameter (δ)				
	a_1	a_2	a_3	a_4	a_5
Winter	1559.8	7.22	52.04	4.15	1559.8
Spring	1559.7	2.44	87.98	3.46	1559.7
Summer	1558.4	17.39	52.95	4.26	1559.7
Fall	1559.5	7.89	51.78	4.20	1559.5

Table 4
Statistical error measures of 5-PL algorithm in wind power prediction over four time intervals.

Error	Time (min)	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	5	1.83	1.60	1.11	9.09	3.41
	15	2.14	1.40	1.05	8.11	3.17
	30	2.31	1.32	1.00	8.25	3.22
	60	2.64	1.39	1.68	8.78	3.62
MAE	5	18.23	18.83	14.33	20.44	17.96
	15	21.33	16.42	13.53	18.40	17.42
	30	21.95	15.49	12.87	18.49	17.20
	60	26.32	16.30	21.67	19.66	20.98
SDE	5	37.40	35.36	24.81	33.69	32.81
	15	38.34	30.53	32.86	29.38	32.78
	30	42.31	29.83	30.75	26.89	32.44
	60	53.05	30.37	49.17	28.25	40.21

parameters of DES model, (9) is minimized by using BSA. In this paper for BSA, the maximum number of iteration is 2000, population size is 30, *mixrate* = 0.95, $F = 3.rndn$, whereby stopping criteria is based on the maximum number of iteration. Noted that *rndn* is the normal distribution with mean equal to 0 and standard deviation equal to 1.

$$\begin{aligned} &\text{Objective function}: \ \min \sum_{t=1}^{n} (y_{t+1} - \hat{y}_{t+1})^2 \\ &\hat{y}_{t+1} = (1-\eta-\lambda)(S'_{t-1} + b_{t-1}) + \eta y_t + \lambda S'_t + b_{t-1} \\ &0 < \eta, \lambda < 1 \end{aligned} \tag{9}$$

The best parameters of DES obtained are shown in Table 1. Table 2 presents the performance of DES and ARMA models based on the data set 2 and also the performance of persistence model based on the original data set in Section 2.1 that does not require any training samples. According to this table, the most accurate model in wind speed prediction is DES whereas persistence performs the worst for the 5-min interval data. Therefore, DES model is selected for wind speed forecasting. The same procedure is carried out for 15-, 30-, and 60-min ahead wind speed predictions. According to the data set in Section 2.1, three, six, and twelve consecutive data are averaged to make 15-, 30-, and 60-min ahead prediction respectively. Fig. 2 shows observed and predicted wind speed by time series model for 5-min interval average data.

2.5. Wind power prediction based on power curve

Wind turbine power curve presents the output power of the turbine for different wind speeds. It includes three regions: cut-in speed (V_c), rated speed (V_r) and cut-off speed (V_f). In V_c turbine starts producing power; in V_r turbine is generating rated power and the power remains constant until V_f is attained [21]. Wind turbine power curve is usually necessary to monitor anomalies and performance of wind turbines but if the wind speed predicted is available it also can facilitate the estimation of wind energy. Although wind turbine manufacturers supply power curves under ideal condition, there is a significant discrepancy between theoretical and empirical power curve. This is due to the fact that manufacturers do not take into account the wear and tear of wind turbine components and also local turbulence. Therefore, it is essential to model wind power curve. In this study five-parameter logistic function (5-PL) [22] was employed for wind turbine power curve modeling and expressed as follows:

$$\hat{y} = f(x,\delta) = a_1 + (a_2 - a_1) \Big/ \left(1 + \left(\frac{x}{a_3}\right)^{a_4}\right)^{a_5} \tag{10}$$

where x is the wind speed predicted in Section 2.3.2 by DES over 5-, 15-, 30- and 60-min interval and $\delta = (a_1, a_2, a_3, a_4, a_5)$ is the vector parameter of 5-PL. BSA is used to optimize (11) and to search for the best estimation of δ.

$$\text{Objective function}: \ \min \sum_{t=1}^{n} (\hat{y}(x_i) - y(x_i))^2 \tag{11}$$

where $\hat{y}$ is the predicted power by (10), is the observed power and n is the sample size. The parameters of 5-PL obtained by BSA for different seasons are tabulated in Table 3. Table 4 summarizes the power predicted by 5-PL for four different time intervals in four seasons. Fig. 3 illustrates the performance of 5-PL for the last 100 samples in the third week of May.

3. Direct wind power prediction

3.1. Proposed feature selection technique

Feature selection is an important preprocessing step commonly used in image processing, anomaly detection, and also machine learning that involve an enormous amount of features. Most of these attributes contain noisy, irrelevant and redundant

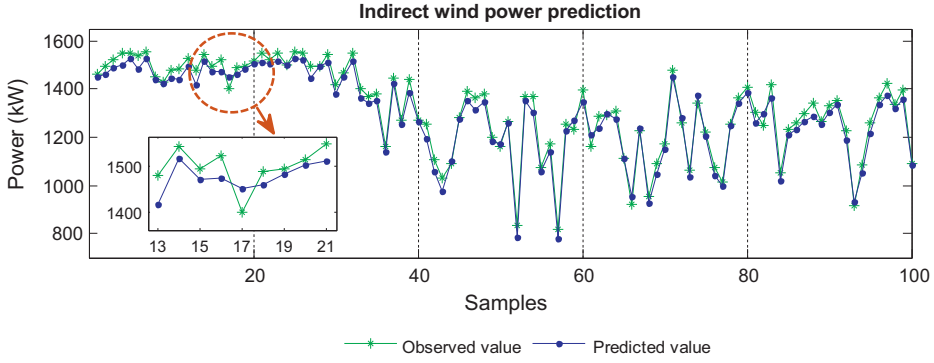

Fig. 3. The real and predicted wind power using 5-PL for the last 100 samples in the third week of May.

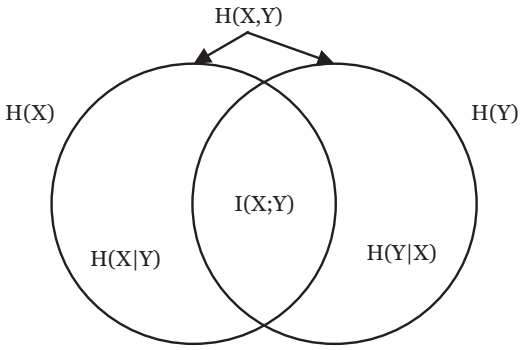

Fig. 4. Graphical representation of the conditional entropy and the mutual information.

information. The main objective of a feature selection technique is to evaluate relevancy and redundancy of the input features for selecting the best subset of features representing the most important information of the original feature set. Generally feature selection methods can be categorized into three types: filter methods, embedded methods and wrapper methods. Filter methods evaluate the relevance of features by using only intrinsic characteristic of the data. In embedded approaches selection of the features is performed simultaneously with model construction. In other words, training process and the feature selection parts cannot be separated. Wrapper methods apply learning algorithm to assess the quality of features [23]. One of the effective attribute selection strategies is mutual information operating based on the concept of entropy. Recently, mutual information is widely used in pattern recognition such as face and fingerprint identification, statistics, data mining, and time series modeling. Here, firstly the principal of information theory based on mutual information and entropy is introduced. Indeed, in the first stage mutual information aims to filter out irrelevant inputs. In the second stage, the neural network is used to remove the redundant features.

The entropy of random variable regardless of discrete or continuous presents the average amount of information that can be learnt from the random variable X and this is a measure of its uncertainty [24,25]. The entropy of discrete random variable $X = (X_1, X_2, \ldots, X_n)$ denoted by (X), is defined as:

$$H(X) = -\sum_{i=1}^{n} P(X_i)\log(P(X_i)) = -\mathbb{E}[\log(P(X))] \tag{12}$$

where $P(X)$ is the probability function which is obtained by division of the number of samples with value of X_i to the total number of samples (n). Although the base of logarithm function is two which results in $H(X)$ varying between 0 and 1, choosing the base is arbitrary since it only changes the unit of entropy. If X is continuous random variable with probability function of $P(X)$ then $H(X)$ is expressed as follows:

$$H(X) = -\int P(X)\log(P(X))dX \tag{13}$$

If X and Y are two continuous random variables, then the joint entropy of X and Y is defined as:

$$H(X,Y) = -\iint P(X,Y)\log(P(X,Y))dXdY \tag{14}$$

where $P(X,Y)$ is the joint probability distribution of variables and $H(X,Y)$ presents the total amount of uncertainty of two random variables X and Y. Conditional entropy measures the remaining uncertainty of variable X when the value of Y is known. Conditional entropy is typically equal to or greater than zero, but it is equal to entropy of variable X and Y when both variables are absolutely independent. Conditional entropy denoted by (X,Y), is expressed as:

$$H(X|Y) = -\iint P(X,Y)\log(P(X|Y))dXdY \tag{15}$$

The relationship between conditional entropy and joint entropy is known as chain-rule and defined as:

$$H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X) \tag{16}$$

This states that the total uncertainty of variable X and Y is equal to the uncertainty of X plus the remaining entropy of Y when X is known.

Mutual information is a measure of the amount of information that one variable contains about another variable and it is expressed as follows:

$$\begin{aligned} I(X;Y) &= \iint P(X,Y)\log\left(\frac{P(X,Y)}{P(X)P(Y)}\right)dXdY = H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y) = I(Y;X) \end{aligned} \tag{17}$$

Mutual information has two important properties. First, it is capable of measuring any kind of relationship between variables. Second it is invariant to space transformation due to the fact that logarithm function used in (17) is non dimensional. Venn diagram in Fig. 4 illustrates mutual information of two variable X and Y.

In a wind power forecasting process, wind power is a function of diverse metrological variables such as temperature, wind speed, wind direction and humidity. It is well known that the power produced in time t depends not only on the metrological variables at time t but also on their past values and even the past values of the power generated which is expressed as follows:

$$\begin{aligned} WP(t) = f(&S(t), S(t-1), \ldots, S(t-NL_S), T(t), T(t-1), \ldots, \\ &\times T(t-NL_T), D(t), D(t-1), \ldots, D(t-NL_D), H(t), \\ &\times H(t-1), \ldots, H(t-NL_H), WP(t-1), \ldots, WP(t-NL_{WP})) \end{aligned} \tag{18}$$

where $S(t), T(t), D(t)$, and $H(t)$ present the current wind speed, temperature, wind direction and humidity, $WP(t)$ is the predicted power at time t, NL_S presents the number of lag order for the wind speed and similarly NL_T, NL_D, NL_H, and NL_{WP} denote other variables lag order. Though aforementioned exogenous variables have relationship with wind power, it is neither efficient and nor feasible to employ all of them as inputs to forecast engine. Assuming 25 lagged values of metrological variables ($NL_S = NL_T = NL_D = NL_H = NL_{WP} = 25$) are used to forecast wind power, these constitute 128 attributes including 100 past values of power, temperature, wind speed, wind direction and humidity plus the current values of temperature, wind speed, and humidity as inputs to forecast engine. If such a huge amount of features is applied on any machine learning algorithms, this will not only slowing down the learning process but also giving rise to a poor performance and overfitting the training data. In other words, although these factors are of vital importance to wind power forecasting, only those features exhibiting significant influence on the output should be picked.

Let $X = \{S(t), \ldots, S(t-NL_S), T(t), \ldots, T(t-NL_T), D(t), \ldots, D(t-NL_D), H(t), \ldots, H(t-NL_H), WP(t-1), \ldots, WP(t-NL_{WP})\}$ denote a vector that is composed of all the inputs features and let $Y = WP(t)$ denote a target or output feature. In the first stage of the proposed feature selection algorithm according to (17), the mutual information between each individual input $\forall X_i (1 < i < NL_S + NL_T + NL_D + NL_H + NL_{WP})$ and output feature is computed, e.g., $MI(X_2;Y)$ presents the mutual information between the first lagged value of wind speed and the current output power. Based on the value of $MI(X_i,Y)$, input features are sorted in descending order where the higher value of mutual information shows the stronger dependency between each variable of X and Y. Input features having lower value of mutual information than the chosen threshold TH

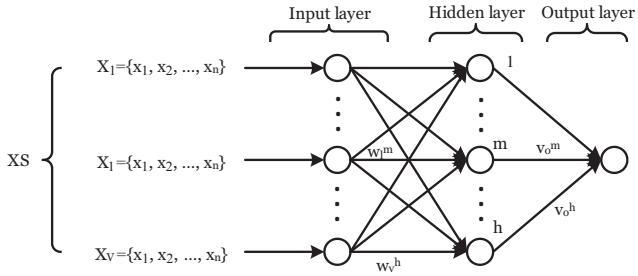

Fig. 5. NN model for the second stage of proposed feature selection technique.

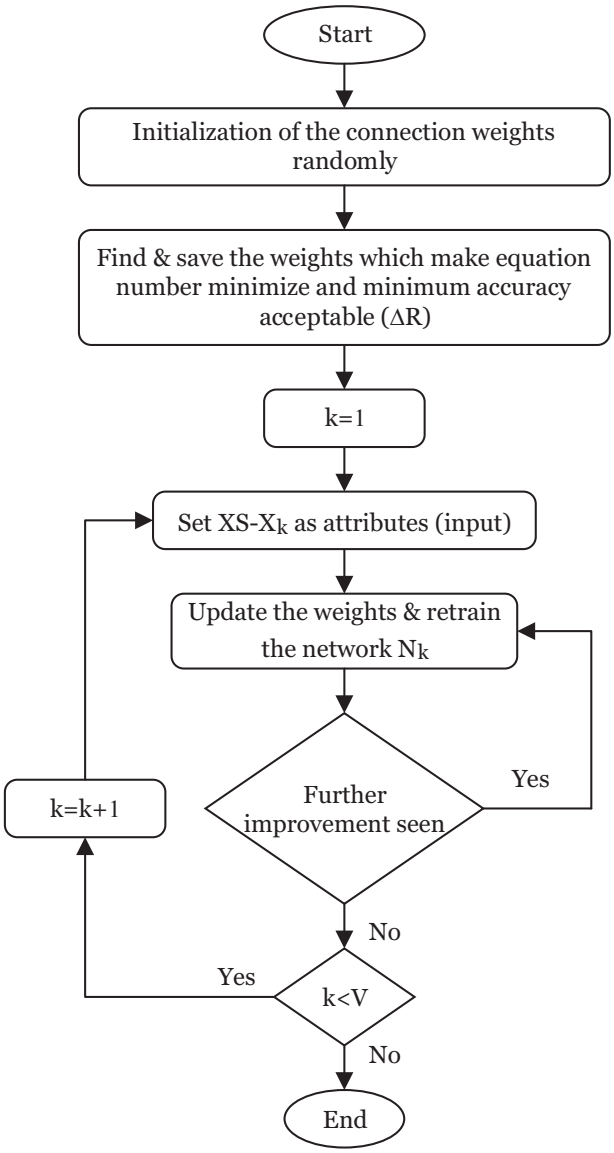

Fig. 6. Structure of the second stage of proposed feature selection technique.

presenting less significant influence on the output are eliminated. The variables with mutual information above the given *TH* are considered as relevant features and form a subset $XS \subset X$ which remains for the next stage. Note that, although the threshold *TH* is provided by a user, a high value of *TH* results in missing lots of information. Whereas, a low value of *TH* may include too many features, either relevant or irrelevant leading to a large computational burden. Here, several thresholds are applied and the best one is selected. However, different feature sets may require different value of threshold. In other words, data collected from different wind farms may need different thresholds because these data considerably vary in the number of irrelevant and redundant features.

In general, the most relevant features to target feature may not lead to optimal results as it may include redundant features. In fact, the inclusion of redundant data in the model building process phase can result in a poor predictive performance and increased computation cost. Indeed, the selected *m* best features may not result in highest accuracy which can be obtained with the best *m* features. Thus, the second stage of two-stage algorithm focuses on finding and removing those features which are highly correlated to other features.

In order to do so, three-layer feedforward neural networks is applied as shown in Fig. 5. The network is trained so that the connections from redundant inputs have smaller magnitudes. On the other hand, salient attributes are differentiated by the strength of their connections from the input layer to hidden layer and the hidden layer to output layer. The connections with smaller magnitude can be eliminated since they have less significant effect on the network accuracy. The accuracy remains remarkably preserved even after removing the small weight and if the accuracy of the network drops, it can be recovered by retraining the network.

Generally, the error function measured during the training process is defined as follows:

$$S = \frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2 \tag{19}$$

where *n* is the number of samples, *t* is the network output and *y* is the actual value. In order to detect unnecessary attributes, a penalty function is added to (19) which is expressed as:

$$P(w) = \alpha_1\left(\sum_{m=1}^{h}\sum_{l=1}^{V}\frac{\beta(w_l^m)^2}{1+\beta(w_l^m)^2} + \sum_{m=1}^{h}\frac{\beta(v_o^m)^2}{1+\beta(v_o^m)^2}\right) + \alpha_2\left(\sum_{m=1}^{h}\sum_{l=1}^{V}(w_l^m)^2 + \sum_{m=1}^{h}(v_o^m)^2\right) \tag{20}$$

where α_1, α_2 and β are coefficients that control the influence of the penalty term, *h* is the number of hidden units, *V* is the number of attributes selected in the first stage, w_l^m is the weight connecting from *l*-th attribute to *m*-th hidden unit, and v_o^m is the weight connecting from *m*-th hidden unit to network output. The idea behind this algorithm is that, the accuracy of trained network $\mathbb{N}$ with the set of attributes $XS = \{X_1, \ldots, X_V\}, XS \subset X, V < (NL_S + NL_T + NL_D + NL_H + NL_{WP})$ as input is first measured. Then the number of attributes sequentially reduced to form a new model. Assuming $k = \{1,2,\ldots,V\}$, the accuracy of the network $\mathbb{N}_k$ having *k* less attributes is computed and at the end algorithm decides whether more attributes can be deleted. The main steps of the algorithm are illustrated in Fig. 6 and outlined below in details.

Table 5
Selected attributes by proposed feature selection technique for the third week of August.

Selected Attributes	Rank	Selected attributes	Rank	Selected attributes	Rank
$WP(t-1)$	1	$T(t-1)$	7	$S(t-12)$	13
$S(t)$	2	$S(t-4)$	8	$D(t-8)$	14
$S(t-1)$	3	$WP(t-5)$	9	$WP(t-17)$	15
$WP(t-3)$	4	$D(t-2)$	10	$S(t-18)$	16
$WP(t-8)$	5	$S(t-9)$	11	$T(t-10)$	17
$S(t-2)$	6	$T(t-5)$	12		

1. Given input vector $XS = \{X_1, \ldots, X_V\}$, $XS \subset X$ with the size of $V * n$ is separated into two data set: training set, S_T and testing set S_C. The network $\mathbb{N}$ is trained to minimize (19) and (20). It also computes the accuracy of the training set R_T, the testing set R_C and also the maximum acceptable decrease (ΔR) on set S_C. It should be noted that in the first procedure of training equal value of α_1, α_2 are set for the weights from input layer to hidden layer. Based on the research conducted in [26], $\alpha_1 = 10^{-1}$, $\alpha_2 = 10^{-4}$, $\beta = 10$ and the maximum acceptable decrease (ΔR) is set to 3%.
2. Suppose features $XS - XS\{1, \ldots, k\} = XS\{k+1, \ldots, V\}$ are input features of the network $\mathbb{N}_k$, i.e., $\mathbb{N}_k$ does not include first k attributes. For instance, $XS\{4, 5, \ldots, V\}$ are input features of network $\mathbb{N}_3$. The network $\mathbb{N}_k$ is trained while the connection from attribute $XS(k)$ to hidden layer is set to zero and all weights from other attributes are set equal to weights of network$\mathbb{N}$. The accuracy of training and testing set for all k is measured and called R_T^k and R_C^k respectively.
3. Rank networks $\mathbb{N}_k$ based on testing set as $R_C^1 \geqslant R_C^2 \geqslant \ldots \geqslant R_C^V$. Then, compute the average of this rate R_C^{avg}.
4. Algorithm updates the penalty parameter of attributes. If the accuracy of the network $\mathbb{N}_k$ denoted by R_C^k is smaller than R_C^{avg}, only the weights from attribute $XS(k)$ are multiplied by 1.1. In fact, the expectation is that with the larger penalty parameters, network $\mathbb{N}_k$ will produce a smaller magnitude for the weights connected to $XS(k)$. On the contrary, if R_C^k is higher than R_C^{avg} all network connections from input $XS(k)$ are divided by 1.1. This allows salient inputs have connections with higher value in magnitude after network is retrained. On the other hand, algorithm removes network connections having a small magnitude representing unimportant attributes.

It is worth noting that selecting a general initial value for parameters α_1 and α_2 may not work best for all problems. Whereby, based on this study and the work conducted in [30], the recommended settings for the initial value of α_1 and α_2 are 10^{-1} and 10^{-4} respectively.

3.2. Wind power prediction models

In this section, ANFIS [27] and five data mining algorithms namely, k-nearest neighbor (k-NN) [28], M5Rules [22], random forest [29], SVM [30] and multilayer perceptron (MLP) [31] are employed to predict the wind power generated. ANFIS models are capable to map nonlinear relationship between input and output by setting up a set of fuzzy rules and tuning the membership function parameters in the training phase. k-NN is an instance-based learning in which each new instance is compared with existing ones and k closest existing instances are used to assign the class to the new ones. Euclidean distance function is often applied to compute difference between instances. M5Rules is a non-parametric algorithm which generates propositional regression rules in IF-THEN format rule from model tree. Random forest can be applied for classification, regression purposes as well as ranking candidate attributes. It is an ensemble method that combines the prediction of many decision trees. SVMs map given data set from input space into high dimensional feature space through the use of kernel function. MLP is an algorithm with multi-layer perceptron structure.

The most important features selected in Section 3.1 are used to evaluate the accuracy of these algorithms. In this study, 82 features, containing 20 lagged values of wind speed, wind direction, temperature, power generated and the current values of wind speed and temperature are considered where almost all the informative features are captured. Humidity data is not considered

Table 7
Error measures of the direct wind power prediction's algorithms for data set 2.

Error	Algorithm	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	k-NN (k = 50)	0.81	1.43	0.86	3.12	1.55
	M5Rules	1.11	1.55	0.84	3.32	1.70
	Random forest	1.00	1.69	1.06	4.88	2.15
	SVM	0.87	1.41	1.13	3.19	1.65
	MLP	0.82	1.43	0.87	3.13	1.56
	ANFIS	0.76	1.34	0.83	2.96	1.47
MAE	k-NN (k = 50)	8.07	16.87	11.12	7.03	10.77
	M5Rules	11.15	18.28	10.91	7.48	11.95
	Random forest	10.02	19.93	13.68	10.97	13.65
	SVM	8.70	16.58	14.61	7.18	11.76
	MLP	8.19	16.78	11.27	7.03	10.81
	ANFIS	7.59	15.76	10.79	6.66	10.20
SDE	k-NN (k = 50)	17.99	32.77	20.98	13.00	21.18
	M5Rules	21.89	34.40	21.12	13.66	22.76
	Random forest	20.93	37.47	26.26	18.49	25.78
	SVM	24.86	31.81	23.05	13.16	23.22
	MLP	17.97	32.61	21.79	13.00	21.34
	ANFIS	16.09	29.04	22.09	12.76	19.99

Table 8
Prediction accuracy of k-NN with different number of k.

Error	k-NN algorithm	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	k = 50	0.81	1.43	0.86	3.12	1.55
	k = 100	0.84	1.42	0.95	3.14	1.58
	k = 150	0.88	1.41	1.10	3.15	1.63
	k = 200	0.94	1.41	1.29	3.15	1.69
	k = 250	1.05	1.41	1.47	3.22	1.78
MAE	k = 50	8.07	16.87	11.12	7.03	10.77
	k = 100	8.44	16.73	12.31	7.07	11.13
	k = 150	8.83	16.63	14.29	7.90	11.91
	k = 200	9.39	16.60	16.71	7.10	12.45
	k = 250	10.46	16.59	19.05	7.26	13.34
SDE	k = 50	17.99	32.77	20.98	13.00	21.18
	k = 100	18.76	32.46	21.85	13.16	21.55
	k = 150	19.60	32.28	25.31	13.16	22.58
	k = 200	20.79	32.18	31.14	13.23	24.33
	k = 250	23.21	23.05	37.65	13.64	24.38

Table 6
MAPE (%), MAE and SDE of proposed feature selection, CA, PCA and Relief algorithms for data set 2.

Season	PCA			CA			Relief			MI + NN		
	MAPE	MAE	SDE	MAPE	MAE	SDE	MAPE	MAE	SDE	MAPE	MAE	SDE
Winter	0.88	8.83	17.28	0.96	8.81	18.05	0.86	8.61	17.04	0.76	7.59	16.09
Spring	1.52	17.80	31.39	1.38	16.14	29.02	1.36	15.96	29.77	1.34	15.76	29.04
Summer	1.06	13.67	24.82	0.91	11.74	23.50	0.94	12.18	24.07	0.83	10.79	22.09
Fall	3.47	7.86	13.91	3.49	7.90	14.46	3.04	6.88	13.28	2.96	6.66	12.76
Average	1.73	12.04	21.85	1.69	11.15	21.26	1.55	10.90	21.04	1.47	10.20	19.99

Table 9
Error comparison of ANFIS model in four time intervals.

Error	Time (min)	Test week				Average
		Winter	Spring	Summer	Fall	
MAPE (%)	5	0.76	1.34	0.83	2.96	1.47
	15	0.71	1.27	0.92	2.59	1.37
	30	0.77	1.23	0.89	2.33	1.30
	60	1.03	1.26	0.91	2.72	1.48
MAE	5	7.59	15.76	10.79	6.66	10.20
	15	7.13	14.96	11.94	5.88	9.97
	30	7.72	14.51	11.59	5.23	9.76
	60	10.29	14.78	11.81	6.10	10.74
SDE	5	16.09	29.04	22.09	12.76	19.99
	15	14.78	28.82	29.63	11.01	21.06
	30	16.32	27.77	27.34	9.79	20.30
	60	25.89	26.78	25.93	12.38	22.74

because it was not available in the collected data. Considering $TH = 0.64$, 33 features are selected as the most relevant attributes in the first stage. In the second stage, 16 features of the feature set being reduced in the first stage are removed since they are not comprehensive representation of the characteristics of the target feature. Finally, 17 features with most relevancy and maximally dissimilar are selected as inputs for the mentioned six algorithms. Selected features are illustrated in Table 5.

In order to evaluate the effectiveness of the proposed feature selection method, it has been compared with the correlation analysis (CA) [32], principal component analysis (PCA) [33] and relief feature selection techniques [34]. CA measures only the degree of linear association between two variables. PCA linearly transforms the original inputs into new uncorrelated features and relief is a feature weight-based algorithm inspired by instance-based learning algorithm. Noted that neither of these feature selection methods helps with redundant attributes, however, this imperfection is overcome through second stage of the proposed feature selection technique. The accuracy of each model for its corresponding test week is presented in Table 6. To have a better comparison, the average value of statical error of each model also is shown. The test results indicate that, the proposed feature selection technique outperforms others for the same testing data. By taking into consideration that the features with higher rank have greater influence on power, the wind speed is the most important variable among other metrological data in wind power forecasting.

Table 7 provides a comparison between ANFIS and data mining algorithms in terms of three statistical measures. For the sake of a fair comparison, the same data as in Section 2 are considered in this section. The table reveals that k-NN and MLP models perform quiet well and they produce almost the same results. Moreover, the table discloses ANFIS yields the best accuracy among other models while the accuracy of random forest is the worst. Hence, the ANFIS model is selected to predict the generated power over 5–60 min ahead in the future.

It is worth noting that in ANN perspective, parameter selection has been always a challenging issue because there is indefinite method to obtain the optimum parameters. The worst is that, an effective NN model performing well in a typical problem does not guarantee the well performance in other problems [35]. Thus, in this study fifteen different structures are repeated to obtain the minimum value of error. For instance, the performance of MLP is observed by changing the number of hidden layers, the number of nodes, transfer functions, etc. Eventually, MLP was developed with the following structure: backpropagation learning algorithm, two hidden layers, six nodes in the first hidden layer and three nodes in the second hidden layer, hyperbolic tangent sigmoid transfer function, and the number of iteration of 1000. The SVM was trained with RBF kernel function and scale factor of 1 while ANFIS, Sugeno type, was set up with 8 cluster centers. The k-NN also indicated different behavior for different numbers of k. Thus several experiments were performed to obtain the best number of k. To be concise, the obtained results for only changes in the free parameters of k-NN model are shown in Table 8.

Table 9 indicates the performance of ANFIS model over 5-, 15-, 30-, and 60-min ahead interval. The table shows ANFIS algorithm produces less accuracy in a longer horizon of prediction. Moreover, it can be seen that there is a difference between errors measured in different test weeks corresponding to the four seasons of a year. This might be due to impact of weather conditions on the mechanical efficiency of the turbine and wind speed distribution. The latter was verified by observing the value of shape parameter (k) and scale parameter (c) of the collected data. When $1 < k < 2$, this represents wind speed follows the Weibull distribution in the winter and fall, whereas it follows the normal distribution in the spring and summer, $k > 3$. Low temperatures impact adversely on different materials used in the fabrication of wind turbines such as composite and steel materials. Steels become more brittle and

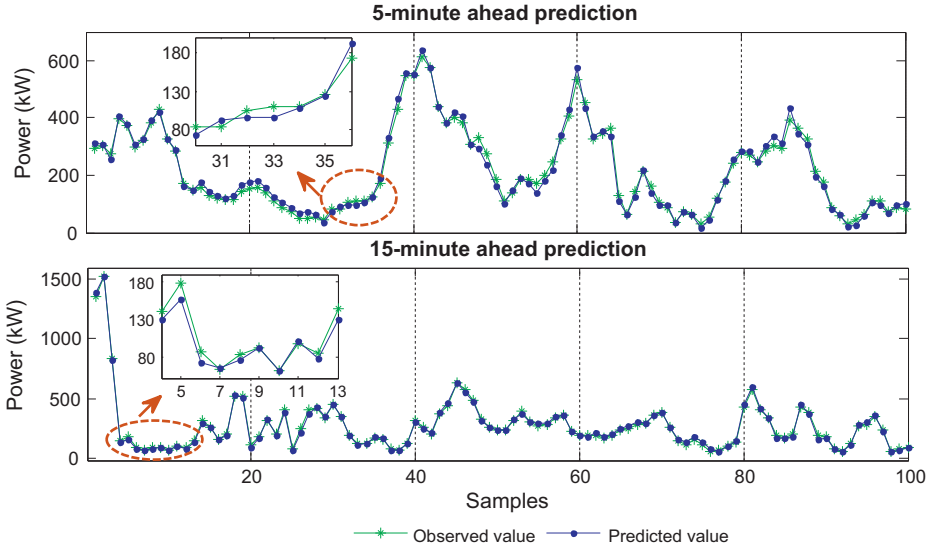

Fig. 7. The real value and predicted wind power using ANFIS for the last 100 samples of first week of November.

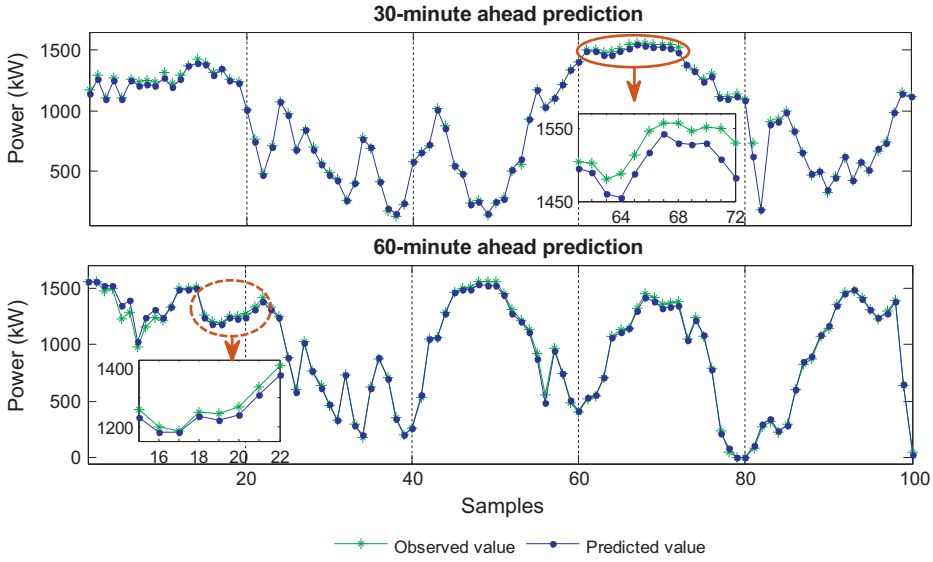

Fig. 8. The real value and predicted wind power using ANFIS for the last 100 samples of third week of May.

composite materials will be subjected to a residual stress. Sufficiently high stresses can cause microcracks in the material which result in stiffness and impermeability reduction of the materials. Low temperature can also damage electrical equipment such as yaw drive motors and transformers and the winding can suffer from a thermal shock. Moreover, long exposure to cold can damage gearboxes, hydraulic couplers. The colder temperature become, the more viscosity of the lubricant increase and will damage gears because oil cannot freely circulate. Furthermore, seals and rubber parts loose at low temperature. All these may not necessarily cause part's failure but can result in a general decline in performance. Figs. 7 and 8 provide a 2-D scatter plot of the performance of the ANFIS algorithm over 5-, 15-, 30- and 60-min interval average data (data set 2).

4. Conclusion

In this paper two different approaches, direct and indirect wind power predictions, were investigated. Since wind speed forecasting is a prerequisite in indirect wind power forecasting, several 5-min time series methods were applied. DES method achieved more accurate results and is found appealing due to its robustness and the simplicity of the model. To set up direct prediction of power, ANFIS along with the five data mining algorithms were employed. To select the best subset from 82 attributes as input for the direct prediction methods, a new feature selection technique with a combination of mutual information and neural network was proposed in this study. The proposed technique is able to remove irrelevant and redundant attributes and this makes the prediction becomes more robust. Case studies on actual wind farm illustrated that ANFIS outperforms others, random forest performs worst and MLP and k-NN mostly have the same accuracy. From the comparison results, direct prediction yields better accuracy than indirect prediction. This is due to the integration of two errors in indirect wind power prediction, DES error and 5-PL error. Using numerical weather prediction to forecast wind speed in indirect wind power prediction model and comparison with direct model in long term will be considered in future work.

Acknowledgements

This work has been supported by High Impact Research Secretariat (HIR) at University of Malaya through the "Campus Network Smart Grid System for Energy Security" Project (Under grant number: H-16001-00-D000032).

References

[1] Hu X, Moura SJ, Murgovski N, Egardt B, Cao D. Integrated optimization of battery sizing, charging, and power management in plug-in hybrid electric vehicles, control systems technology. IEEE Trans 2015. 1–1.
[2] Prasad RD, Bansal RC, Sauturaga M. Some of the design and methodology considerations in wind resource assessment. Renew Power Generation, IET 2009;3:53–64.
[3] Louka P, Galanis G, Siebert N, Kariniotakis G, Katsafados P, Pytharoulis I, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. J Wind Eng Ind Aerodyn 2008;96:2348–62.
[4] Peiyuan C, Pedersen T, Bak-Jensen B, Zhe C. ARIMA-based time series model of stochastic wind power generation. Power Syst, IEEE Trans 2010;25:667–76.
[5] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88:1405–14.
[6] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34:1388–93.
[7] Mabel MC, Fernandez E. Estimation of energy yield from wind farms using artificial neural networks. Energy Convers, IEEE Trans 2009;24:459–64.
[8] Zhou J, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers Manage 2011;52:1990–8.
[9] Osório GJ, Matias JCO, Catalão JPS. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. Renew Energy 2015;75:301–7.
[10] Peng H, Liu F, Yang X. A hybrid strategy of short term wind power prediction. Renew Energy 2013;50:590–5.
[11] Jianwu Z, Wei Q. Short-term wind power prediction using a wavelet support vector machine. Sustain Energy, IEEE Trans 2012;3:255–64.
[12] Ghadi MJ, Gilani SH, Afrakhte H, Baghramian A. A novel heuristic method for wind farm power prediction: a case study. Int J Electr Power Energy Syst 2014;63:962–70.
[13] Niya C, Zheng Q, Nabney IT, Xiaofeng M. Wind power forecasts using gaussian processes and numerical weather prediction. Power Syst, IEEE Trans 2014;29:656–65.
[14] Chitsaz H, Amjady N, Zareipour H. Wind power forecast using wavelet neural network trained by improved clonal selection algorithm. Energy Convers Manage 2015;89:588–98.
[15] Potter CW, Negnevitsky M. Very short-term wind forecasting for tasmanian power generation. Power Syst, IEEE Trans 2006;21:965–72.
[16] Kusiak A, Haiyang Z, Zhe S. Short-term prediction of wind farm power: a data mining approach. Energy Convers, IEEE Trans 2009;24:125–36.
[17] Thapar V, Agnihotri G, Sethi VK. Critical analysis of methods for mathematical modelling of wind turbines. Renew Energy 2011;36:3166–77.
[18] Yuan X, Ji B, Yuan Y, Ikram RM, Zhang X, Huang Y. An efficient chaos embedded hybrid approach for hydro-thermal unit commitment problem. Energy Convers Manage 2015;91:225–37.
[19] Taylor JW, McSharry PE. Short-term load forecasting methods: an evaluation based on European data. Power Syst, IEEE Trans 2007;22:2213–9.
[20] Bludszuweit H, Dominguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast error. Power Syst, IEEE Trans 2008;23:983–91.

[21] Chourpouliadis C, Ioannou E, Koras A, Kalfas AI. Comparative study of the power production and noise emissions impact from two wind farms. Energy Convers Manage 2012;60:233–42.
[22] Lydia M, Selvakumar AI, Kumar SS, Kumar GEP. Advanced algorithms for wind turbine power curve modeling. Sustain Energy, IEEE Trans 2013;4:827–35.
[23] Maldonado S, Carrizosa E, Weber R. Kernel penalized K-means: a feature selection method based on Kernel K-means. Inf Sci 2015;322:150–60.
[24] Hu X, Li SE, Jia Z, Egardt B. Enhanced sample entropy-based health management of Li-ion battery for electrified vehicles. Energy 2014;64:953–60.
[25] Hu X, Jiang J, Cao D, Egardt B. Battery health prognosis for electric vehicles using sample entropy and sparse bayesian predictive modeling, industrial electronics. IEEE Trans 2015. 1–1.
[26] Setiono R, Huan L. Neural-network feature selector. Neural Networks, IEEE Trans 1997;8:654–62.
[27] Schlechtingen M, Santos IF, Achiche S. Using data-mining approaches for wind turbine power curve monitoring: a comparative study. Sustain Energy, IEEE Trans 2013;4:671–9.
[28] Yesilbudak M, Sagiroglu S, Colak I. A new approach to very short term wind speed prediction using k-nearest neighbor classification. Energy Convers Manage 2013;69:77–86.
[29] Lahouar A, Ben Hadj Slama J. Day-ahead load forecast using random forest and expert input selection. Energy Convers Manage 2015;103:1040–51.
[30] Lei Y, Miao H, Junshan Z, Vittal V. Support-vector-machine-enhanced markov model for short-term wind power forecast. Sustain Energy, IEEE Trans 2015;6:791–9.
[31] Velo R, López P, Maseda F. Wind speed estimation using multilayer perceptron. Energy Convers Manage 2014;81:1–9.
[32] Hong Y-Y, Chang H-L, Chiu C-S. Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. Energy 2010;35:3870–6.
[33] Kong X, Liu X, Shi R, Lee KY. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[34] Koutanaei FN, Sajedi H, Khanbabaei M. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. J. Retailing Consumer Services 2015;27:11–23.
[35] Hu X, Li S, Yang Y. Advanced machine learning approach for lithium-ion battery state estimation in electric vehicles, transportation electrification. IEEE Trans 2015. 1–1.