Applied Energy 87 (2010) 2313–2320

Contents lists available at ScienceDirect

Applied Energy

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

On comparing three artificial neural networks for wind speed forecasting

Gong Li, Jing Shi *

Department of Industrial and Manufacturing Engineering, North Dakota State University, Department 2485, P.O. Box 6050, Fargo, ND 58108, USA

ARTICLE INFO

Article history:
Received 10 November 2009
Received in revised form 18 December 2009
Accepted 21 December 2009
Available online 21 January 2010

Keywords:
Wind speed
Forecasting
Artificial neural network
Back propagation
Radial basis function
Adaptive linear element

ABSTRACT

Wind speed forecasting is critical for wind energy conversion systems since it greatly influences the issues such as the scheduling of a power system, and the dynamic control of the wind turbine. In this paper, we present a comprehensive comparison study on the application of different artificial neural networks in 1-h-ahead wind speed forecasting. Three types of typical neural networks, namely, adaptive linear element, back propagation, and radial basis function, are investigated. The wind data used are the hourly mean wind speed collected at two observation sites in North Dakota. The performance is evaluated based on three metrics, namely, mean absolute error, root mean square error, and mean absolute percentage error. The results show that even for the same wind dataset, no single neural network model outperforms others universally in terms of all evaluation metrics. Moreover, the selection of the type of neural networks for best performance is also dependent upon the data sources. Among the optimal models obtained, the relative difference in terms of one particular evaluation metric can be as much as 20%. This indicates the need of generating a single robust and reliable forecast by applying a post-processing method.

1. Introduction

Wind energy, being socially beneficial, economically competitive, and environmentally friendly, has become the world's fastest growing renewable energy source of electricity generation. However, due to the stochastic nature of wind speed from time to time and from site to site, accurate information of dynamic wind at the wind farm site is crucial for the operations and management of wind energy conversion systems. For instance, long-term wind speed prediction is vital for the sitting and sizing of wind power applications [1,2], whereas short-term forecasting of wind speed is important for improving the efficiency of a wind power generation systems [3,4] as well as for the integration of wind energy into the power system [5–7].

The forecasting approach can be determined based on the available information and the time scale in question, and thus its application. Short-term wind speed forecast is a subclass of the wind speed forecast. The time scales concerning short-term forecasting range from some seconds to minutes, hours or several days. Short-term forecast mainly helps with the daily and intraday spot market, system management and maintenance scheduling, which are usually of great interest and importance to system operators, electricity companies and wind farm promoters. For periods in the range of a few seconds up to several minutes, the forecasting objective is the control of wind energy conversion systems (WECS); wind speed forecasts in the range of hours target the problem of scheduling in a power system, whereas forecasts in the range of days are related with maintenance and resource planning [8,9].

Much research effort has been made to develop good wind forecasting methods. The approaches found in the literature include many physical methods, statistical methods, hybrid physical-statistical models, artificial intelligence and some other new methods [5,9–11]. Physical methods, such as Numerical Weather Forecast (NWF) and Mesoscale models [12], usually provide the satisfactory forecast precision by combining multiple physical considerations. Statistical methods, such as autoregressive integrated moving average (ARIMA) models, make forecasts by finding the relationship of the observed wind speed time series [13,14]. Often, both physical and statistical models are utilized together, where NWF results are usually regarded as input variables, together with historical data, to train the system on the local conditions according to statistical theories [5]. Recently, some new methods based on artificial intelligence techniques have been developed, including the Artificial Neural Network (ANN) of Multi-Layer Perceptrons (MLP) [15,16], Radial Basis Function [17] and Recurrent Neural Networks [18,19], and Fuzzy Logic [20,21].

ANN is a technique basically used to map random input vector(s) to the corresponding random output vector without pre-assuming any fixed relationship between them. Neural networks can learn from past data, recognize hidden patterns or relationships in historical observations and use them to forecast future values. Additional advantages of the neural network approach over

* Corresponding author. Tel.: +1 701 231 7119; fax: +1 701 231 7195.
E-mail address: jing.shi@ndsu.edu (J. Shi).

0306-2619/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2009.12.013

the conventional forecasting schemes include data error tolerance, ease in adaptability to online measurements, and lack of any excess information (other than time series history of wind speeds) [19].

Different network structures, learning rates, and inputs are believed to result in different forecast accuracies. However, in literature usually only BP models are employed for wind speed forecasting, and thus the obtained conclusions regarding the performance of ANN model in wind speed forecasting may be misleading. A comprehensive investigation is needed on the selection of type of neural network and model parameters with regard to multiple evaluation criteria and multiple datasets from different sites. To bridge the research gap, this paper intends to analyze and compare the performances of three typical ANN techniques, namely, Feed Forward Back-Propagation (FFBP), Radial Basis Function (RBF), and Adaptive Linear Element (ADALINE) Neural Networks, in 1-h-ahead wind speed forecasting.

The remainder of the paper is organized as follows. In Section 2, the neural network based time series methods are explained. In Section 3, the data used in this paper are summarized and explained, followed by the introduction to the statistics used for the measurement of forecast accuracy. Thereafter, the structures of different models as well as the simulation parameters are explained. In Section 5, the simulation results are presented and discussed. Finally, the conclusive remarks are drawn.

2. Time series models based on neural networks

Time series methods make forecasts based solely on historical patterns in the measurement data by using time as independent variable. In a time series, measurements are taken at successive points or over successive periods. The ANN trained with time series have the ability to model arbitrarily linear and nonlinear functions. Being widely utilized in various different fields including transient detection, pattern recognition, approximation, and time series forecast, Artificial Neural Network (ANN) is a promising technology in wind speed forecast. Lapedes and Farber [22] proposed an ANN model along with feed forward and error back-propagation algorithm in wind speed forecast. Song [23] developed an ANN based methodology to perform one-step ahead forecast, and it showed good performance when the wind data do not oscillate violently. Alexiadis et al. claimed that their ANN predictor was about 10% better than persistence model for one-step-ahead forecast [24].

2.1. BP neural networks

Feed forward back-propagation (BP) network is one of the most popular techniques in the field of ANN. The common topology of a BP neural network model is illustrated in Fig. 1. The source nodes in the input layer of the network supply respective elements of the activation pattern or input vector, which constitute the input signals applied to the neurons in the hidden layer. The output signals of the hidden layer are used as inputs to the output layer. The output signals of the neurons in the output layer of the network constitute the overall response of the network to the activation patterns applied by the input layer neurons [23].

With n input neurons, m hidden neurons, and one output neuron, the training process of BP network can be described as follows.

(I) Calculate the outputs of all hidden layer nodes

$$net_j = \sum_{i=0}^{n} w_{i,j} y_i \ (i = 0, 1, \ldots, n; \quad j = 1, \ldots, m), \tag{1}$$

$$z_j = f_H(net_j) \quad (j = 1, 2, \ldots, m), \tag{2}$$

where net_j is the activation value of the jth node, $w_{i,j}$ is the connection weight from input node i to hidden node j, y_i is the ith input with y_0 being the bias b_{IH} (with weight $w_{0,j} = 1$), z_j is the corresponding output of the jth node in the hidden layer, and f_H is called the activation function of a node, which is usually a sigmoid function,

$$f_H(x) = \frac{1}{1+\exp(-x)} \tag{3}$$

(II) Calculate the outputs of all output layer neurons

$$O = f_o\left(\sum_{j=0}^{m} w_{j,k} z_j\right) \quad (j = 0, 1, 2, \ldots, m), \tag{4}$$

where f_O is the activation function, usually a line function, $w_{j,k}$ is the connection weight from hidden node j to output node k (here $k = 1$), z_j is the corresponding output of the jth node in the hidden layer with z_0 being the bias b_{HO} (with weight $w_{0,k} = 1$). All the connection weights and bias values are assigned with random values initially, and then modified according to the results of BP training process.

(III) Minimize the global error E via the training algorithm

$$E = \frac{1}{2}\sum (O - y_t)^2 \tag{5}$$

2.2. RBF neural networks

The RBF neural network is a multi-input, single-output forward network system consisting of an input layer, a hidden layer, and an output layer, as illustrated in Fig. 2.

The RBF network consists of two layers whose output nodes form a linear combination of the basis functions. The basis function in the hidden layer produces a significant nonzero response to the input only when it falls within a small localized region of the input space [24]. It is assumed that, given N n-dimension different points $\{x_i \in R^n, \quad i = 1, 2, \ldots, N\}$ and N real numbers, $\{y_i \in R, \quad i = 1, 2, \ldots, N\}$, a nonlinear function $f(x)$ satisfying $f(x_i) = y_i, \quad i = 1, 2, \ldots, N$ is called an RBF when it depends only on the radial distance $r = ||x - t||$, where t refers to the centre of point x.

The RBF approach chooses f from a linear space of dimension N, depending on the data points $\{x_i \in R^n, \quad i = 1, 2, \ldots, N\}$. The basis of this space is chosen to be the set of functions

$$\{h(\|x - x_i\|), \quad i = 1, 2, \ldots, N\}, \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm on R^n. Therefore, the solution of the above-mentioned interpolation problem has the following form:

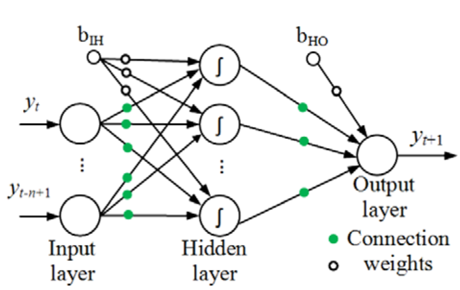

Fig. 1. Topology of backpropagation network.

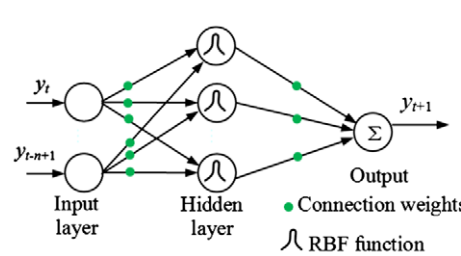

Fig. 2. Topology of RBF neural network.

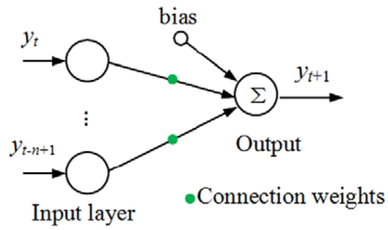

Fig. 3. Topology of a simple ADALINE network.

$$f(x) = \sum_{i=1}^{N} c_i h(\|x - x_i\|), \tag{7}$$

where coefficients c_i can be obtained by imposing the interpolation conditions $f(x_i) = y_i, \quad i = 1, 2, \ldots, N$ on the above equation. Thus, the following solution can be derived:

$$f(x) = \sum_{i=1}^{N} c_i h(\|x_j - x_i\|), \quad j = 1, 2, \ldots, N \tag{8}$$

By defining the vectors $\boldsymbol{y}$, $\boldsymbol{c}$ and the symmetric matrix $\boldsymbol{H}$ as $(\mathbf{y})_j = y_j$, $(\mathbf{c})_j = c_j$, $(\mathbf{H})_{ij} = h(\|x_j - x_i\|)$, the coefficients c_i can be obtained from $\mathbf{c} = \mathbf{H}^{-1}\mathbf{y}$.

2.3. Adaptive linear element networks

The structure of a simple ADALINE used in this paper is illustrated in Fig. 3. The network output is

$$f = \mathbf{Wp} + b = \sum_{i=1}^{n} w_i y_i + b = \sum_{i=0}^{n} w_i y_i, \tag{9}$$

where y_0 represents threshold bias b with weight $w_0 = 1$, $\mathbf{W}$ represents the weight matrix corresponding to the one-column input vector $\mathbf{p}$.

First developed by Widrow and Lehr [25], the ADALINE networks can only solve linearly separable problems. However, the least mean squared (LMS) or Widrow–Hoff learning rule can minimize the mean squared error (MSE) and search for the global minimum point in space, thus moving the decision boundaries as far as it can from the training patterns. During the learning process, the LMS rule diminishes MSE, a mathematical function defined in the multi-dimension space of weights for a set of given training patterns as follows,

$$MSE = \frac{1}{2L} \sum_{l=1}^{L} \varepsilon_l^2 = \frac{1}{2L} \sum_{l=1}^{L} (d_l - f_l)^2, \tag{10}$$

where L is the number of patterns in the training dataset, d_l and f_l represents the desired value and the forecast one of the network, respectively.

The changes in weights at time $(t + 1)$ and time t are proportional to the descendent gradient of the error function, which is commonly defined as the learning rate α

$$w_i(t+1) = w_i(t) + \alpha(d_l - f_l) y_{l_i}. \tag{11}$$

3. Wind speed data and performance metrics for forecast accuracy

3.1. Wind speed data

The state of North Dakota has the greatest wind resource among the lower 48 states of the United States. Since mid 1990s, the state government and regional utility companies have expanded the observation and assessment of wind energy potential to a number of potential sites throughout North Dakota. In this paper, in order to investigate the performance consistency of ANN models with different wind datasets, two representative sites, namely Hannaford and Kulm, in North Dakota were selected. At the two sites, the wind speeds were measured using anemometers at fixed positions having different heights, and the continuously recorded wind data were averaged over 1 h and stored as hourly values. For each selected site, a complete one-year data records were extracted. Table 1 summarizes the geographical information of the two sites and the wind speed characteristics. Fig. 4 illustrates the time series plots of the hourly wind speeds at the two sites from January 1, 2002 to December 31, 2002. Note that wind speed at the height of 10 m was used as recommended by the World Meteorological Organization (WMO) [26].

3.2. Performance metrics of forecast accuracy

To date, a number of performance measures have been proposed and employed to evaluate the forecast accuracy, but no single performance measure has been recognized as the universal standard. This actually complicates the performance comparison

Table 1
Information of two ND sites and their wind speed characteristics.

Site	Latitude (North)	Longitude (West)	Elevation (m)	Wind speed at 10 m above ground level (m/s)		
				Mean	Min	Max
Hannaford (Hann)	47°19′39″	98°12′34″	448	7.40	0.36	25.56
Kulm (Kulm)	46°17′56″	98°51′58″	600	8.79	0.36	23.86

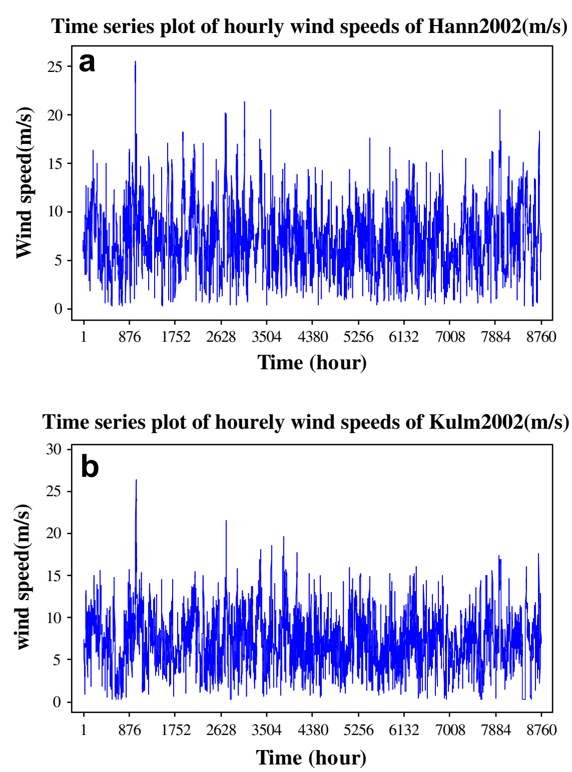

Fig. 4. Time series plots of hourly wind speeds at: (a) site Hann, (b) site Kulm.

of different forecasting models. As a result, we need to assess the performance based on multiple metrics, and it is interesting to see if different metrics will give the same performance ranking for the models to be tested. The metrics included in this study are: mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE).

$$\mathrm{MAE}_j = \frac{1}{T}\sum_{t=1}^{T}|y_t - f_{j,t}|, \tag{12}$$

$$\mathrm{RMSE}_j = \sqrt{\frac{1}{T}\sum\nolimits_{t=1}^{T}(y_t - f_{j,t})^2}, \tag{13}$$

$$\mathrm{MAPE}_j = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - f_{j,t}}{y_t}\right|, \tag{14}$$

where y_t and $f_{j,t}$ denote the observations and the forecast value from model j, respectively, T is the number of data used for performance evaluation and comparison.

MAE measures the average magnitude of the errors in a set of forecasts. To be more specific, it is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. MAE is a linear score which means that all the individual differences are weighted equally in the average. RMSE is a quadratic scoring rule which measures the average magnitude of the error. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable. MAPE is measure of accuracy in a fitted time series value in statistics, specifically trending. The difference between actual value and the forecast value is divided by the actual value. The absolute value of this calculation is summed for every forecast point in time and divided again by the total number of forecast points.

4. Model inputs and parameters

The wind forecast problem aims to find an estimate $f(t+k)$ of the wind vector $y(t+k)$ based on the previous n measurements $y(t), y(t+1), \ldots, y(t-m+1)$. In order to have accurate wind speed forecast, k is chosen to be small and this is called short-term wind speed forecast [3].

Training an ANN model by using time series implies necessarily to know the relation that exists between the series and their lags. There are statistics tools that orient on the matter. That is the case of the autocorrelation function (ACF) and the partial autocorrelation function (PACF), both of which help to determine the input variables of the neural network [27]. For example, Fig. 5 illustrates the ACF plot and PACF plot for the hourly wind speed data observed at site Kulm in the entire year of 2002. Obviously, the previous six observations are correlated with and significantly affect the next possible value of wind speed. For site Hann, the previous seven observations seem to significantly affect the next observations.

For purpose of comparison, however, the inputs of the ANN models varied from previous 1–8 observations for all models. Correspondingly, each previous n ($n = 1, 2, \ldots, 8$) observations were preprocessed and formed into one input vector. The preprocessed dataset was further divided into three subsets: training, evaluation, and testing datasets, respectively. With the last 120 (five day) input vectors being separated as the testing dataset, 5000 other input vectors were 'randomly' selected as the training dataset. Actually, in order to consider the variability among different months, the training patterns or vectors were randomly selected from each

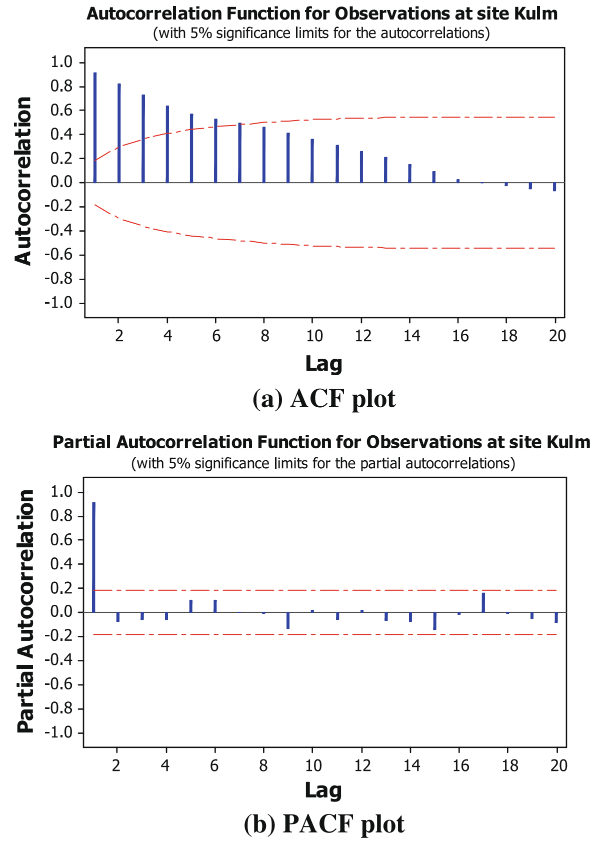

Fig. 5. ACF and PACF plots for wind speeds at site Kulm.

month, respectively [3]. The remaining data were used to construct the validation dataset.

The models adopted have the structures as illustrated in Figs. 1–3, respectively. The number of neurons of the first (input) layer is the same as that of the input vector. According to [28], the optimal number of hidden layer neurons for BP and ADALINE models can be selected as the integer number close to log(T), where T is the number of training vectors. In this study, their numbers of hidden layer neurons was thus selected as four. As for RBF model, the number will be optimized iteratively during training process. The output layer of all three types of ANN models only contains one neuron representing the forecast value of next hourly average wind speed.

The FFBP are trained using the Levenberg–Marquardt (LM) optimization technique since it is more powerful than the conventional gradient descent techniques [29]. BP technique has some drawbacks such as long training duration with a high number of iterations. The LM weight optimization approach is believed to be a solution in tackling this problem. For RBF models, different spread constants are examined in the study, namely, from 0.5 to 1.5 with 0.1 increments. For BP and ADALINE models, different learning rate constants are examined in the study. For BP, it is from 0.025 to 0.5 with 0.025 increments, while for ADALINE, it is from maximum learning rate (MLR) to one tenth of MLR with one tenth of MLR decrement.

5. Results and discussion

5.1. Site Hann

Through extensive calculation, the results confirmed that different learning rates or spread constants, as well as different number of inputs, can result into different values in terms of MAE, RMSE, and MAPE. Fig. 6 only shows the influences of learning rates and observation number on the forecasting performances of BP models. Since usually MAE follows the same trend as MAPE, only RMSE and MAPE are presented in the figure. It can be observed that with six observation inputs, the effect of learning rate on RMSE is not as significant as that on MAPE. The relative variation of MAPE is close to 10% in this case. Also, with a fixed learning rate of 0.1, the effects of observation number on both RMSE and MAPE are relatively less significant.

For the purpose of brevity, the details of each possible combination of parameters are not provided. Table 2 only gives the results about the optimal ANN models that generated the smallest MAE, RMSE, and/or MAPE values, for the dataset from the site of Hann. Among all the 160 BP models tested, using six previous observations and learning at a rate of 0.1 generated the smallest RMSE value (1.254), while using eight previous observations and a 0.075 learning rate resulted into the smallest MAE (0.945) and MAPE (0.206). For all the ADALINE models (indicated by LNN in the table), using four previous observations and a learning rate of 4.49e − 8 gives the smallest MAPE (0.194), while using two previous observations and a learning rate of 8.86e − 8 gives the smallest MAE (0.965) and RMSE (1.271). The situation is more complicated with the RBF models, in which each smallest MAE, RMSE, or MAPE metric corresponds to a different combination of parameters.

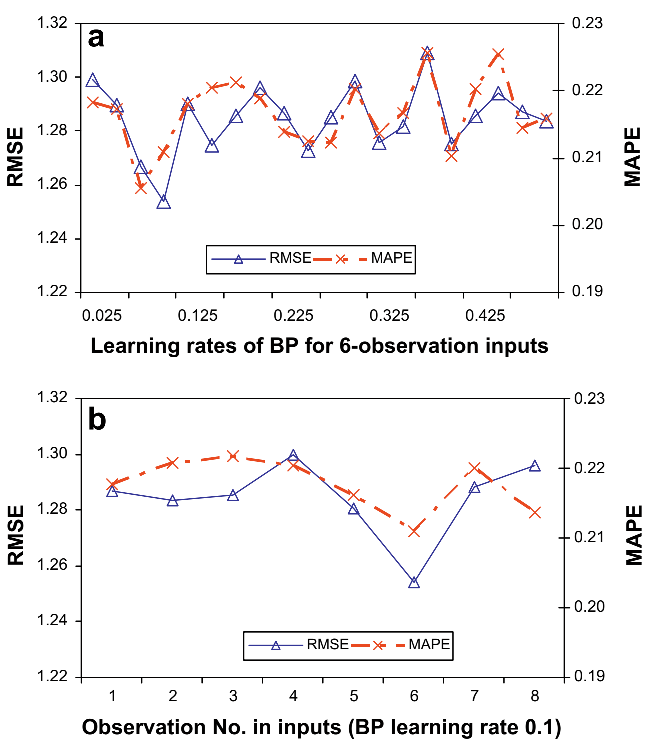

Fig. 6. Effect of (a) learning rate and (b) observation number on forecasting performances for site Hann.

Table 2
Optimal ANN models in terms of MAE, RMSE and/or MAPE, for site Hann.

Models	MAE	RMSE	MAPE	CT(s)
BP_Obs_8_LR_0.075	0.945	1.269	0.206	3.323
***BP_Obs_6_LR*_0.10**	***0.951***	***1.254***	0.211	***2.777***
RBF_Obs_7_spread_0.7	1.058	1.390	0.221	0.047
RBF_Obs_2_spread_0.5	0.989	1.297	0.232	0.047
RBF_Obs_3_spread_0.5	0.997	1.294	0.234	0.047
LNN_Obs_4_lr_4.49e − 8	0.980	1.290	**0.194**	1.529
LNN_Obs_2_lr_8.86e − 8	0.965	1.271	0.196	1.482

Besides, Table 2 also presents the convergence time (CT) for the seven models. It can be seen that all three optimal RBF models converged most swiftly whereas the two optimal BP models demonstrated the slowest convergence rate compared with RBF and ADALINE models.

More importantly, Table 2 indicates that different types of ANN models indeed lead to different performances. BP model using previous six observations and trained at a learning rate of 0.1 seems to be the best in terms of MAE and RMSE, while ADALINE model using four previous observations and learning at 4.49e − 8 learning rate appears to be the best in term of MAPE. It seems that the best ADALINE model has a clear advantage in terms of MAPE – it outperforms the best BP model by 4.8% and the best RBF model by 14.0%. On the other hand, the performance lead of BP model is terms of MAE and RMSE is not as significant. In addition, RBF models do not perform as well as ADALINE or BP models for site Hann.

For purpose of illustration, the forecast time series from the BP and RBF models that generated the smallest and the largest RMSE values, respectively, are plotted as shown in Fig. 7. Obviously, the BP model is more consistent with the observation curve compared to the RBF model. In the meantime, Fig. 8 provides the time series plots of the corresponding forecast errors for the two specific models. It can be observed that although the BP model performs better than the RBF model in terms of RMSE (as well as MAE/MAPE) for the total 120-time-point forecasts, the RBF model is still able to generate smaller forecast errors at some specific time points, e.g. at round time point 85. This verifies that the forecasting performance of ANN models changes with not only time as well as the selected prediction lengths.

5.2. Site Kulm

Similarly, for site Kulm, we also observed that learning rates or spread constants, as well as the number of inputs, affect the forecasting performance in terms of MAE, RMSE and MAPE. In this

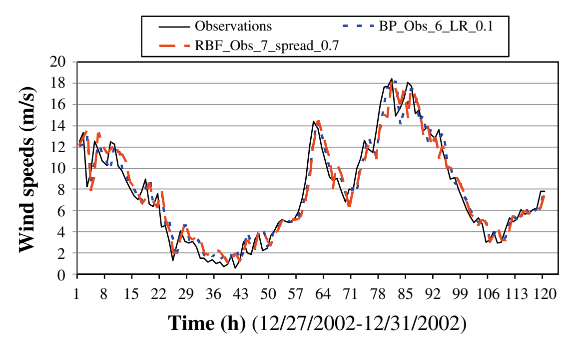

Fig. 7. Forecasts of the two models generating the smallest/largest RMSE values for site Hann.

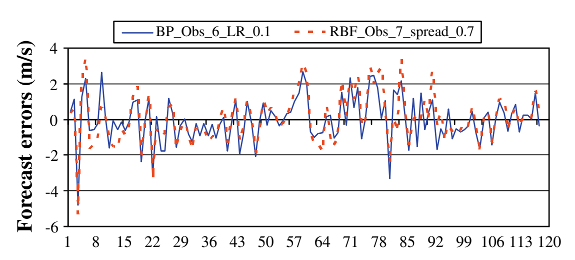

Fig. 8. Forecast errors of the two models generating the smallest/largest RMSE values for site Hann.

case, the influences of different spreads and inputs on the forecasting performances are illustrated only for RBF models. As shown in Fig. 9, the spread clearly affects MAPE values and its effect on RMSE is more significant. The variation of change in RMSE is approximately 7%. With four observation inputs, a 0.7 spread generated the lowest RMSE and MAPE values. On the other hand, the effect of observation number is pronounced in terms of both RMSE and MAPE. In particular, the relative change of MAPE is about 15%. With a fixed spread value of 0.7, four observations inputs can provide the best performance in terms of both RMSE and MAPE.

Similar to site Hann, only the ANN models that generated either the smallest MAE, or RMSE, or MAPE values are provided for site Kulm. Together with the model convergence time, the results are shown in Table 3. It can be seen that the RBF model using four previous observations and trained at a 0.7 learning rate performs the best in terms of all the three statistics. However, the performance lead for the RBF model is less than 5% in terms of any metrics compared with the best BP and ADALINE models. In fact, the BP model using eight previous observations and a learning rate of 0.45 has close performance with the best RBF model in terms of RMSE and MAPE. Again, RBF models converges the fastest whereas BP seems to be the slowest one among the three types of ANN models.

The forecast time series from the RBF and BP models that generated the smallest and the largest RMSE values, respectively, are shown in Fig. 10. In this case, the difference is really minor almost across the entire time span. Similarly, Fig. 11 shows the time series

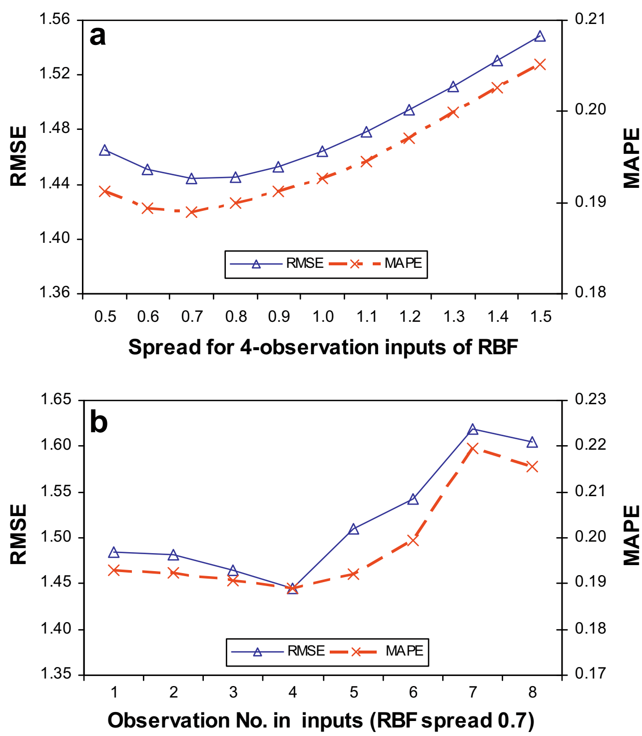

Fig. 9. Effect of (a) learning rate and (b) observation number on forecasting performances for site Kulm.

Table 3
Optimal ANN models in terms of MAE, RMSE and MAPE, respectively, for site Kulm.

Models	MAE	RMSE	MAPE	CT(s)
BP_Obs_8_LR_0.45	1.176	1.469	0.189	3.463
BP_Obs_1_LR_0.45	1.156	1.487	0.194	2.340
RBF_Obs_4_spread_0.7	**1.112**	**1.444**	**0.189**	0.047
LNN_Obs_6_lr_4.59e − 8	1.158	1.485	0.194	***1.404***
LNN_Obs_6_lr_6.88e − 8	1.163	1.485	0.195	1.466

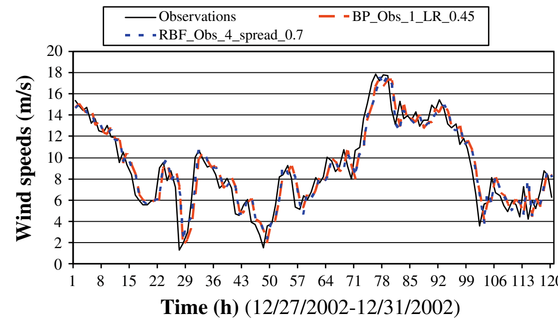

Fig. 10. Forecasts of the two models generating the smallest/largest RMSE values for site Kulm.

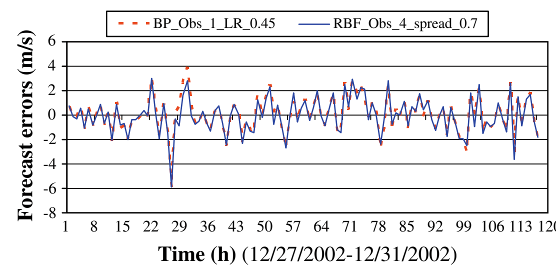

Fig. 11. Forecast errors of the two models generating the smallest/largest RMSE values for site Kulm.

plots of corresponding forecast errors. Again, it can be observed that the performance advantage of the RBF model is not that significant. Although it performs slightly better than the BP model in terms of RMSE as well as MAE/MAPE for the total 120-time-point forecasts, the BP model still can produce smaller forecast errors at some specific time points, e.g. at round time point 113.

5.3. Discussion

From the above results, it can be seen that artificial neural networks trained with different inputs or at different learning rates demonstrate varying levels of accuracy in forecasting the hourly mean wind speed 1-h-ahead. For instance, for site Kulm, the BP models (shown in Table 3) trained with 1 and 8 previous observations produce inconsistent MAE, RMSE and MAPE values, although both models were trained with the same LM algorithm and at the same training rate (0.45). Similarly, the two ADALINE models trained with same inputs (six previous observations) but at different learning rates also demonstrated slightly different performances. Therefore, in building the ANN models for forecasting wind speed, factors such as the model inputs and learning rates, should be properly determined since this decision directly affects the forecast accuracy.

It can also be observed that the types of ANN models may significantly affect the forecasting accuracy. In the literature, the widely-used type of ANN model for wind speed forecasting is BP, and other types have been less investigated. The results confirm the need of this study – multiple types of ANN models should be evaluated before the most suitable type can be determined. Nevertheless, the complicated issue is that different evaluation metrics often give inconsistent ranking among candidate models. This presents a major challenge on which metric to use in practice. Note that there is existing research in literature about developing a single general performance index by combining different evaluation criteria for measuring forecasting accuracy [30,31]. They may be adopted to handle this in the process of model selection.

Meanwhile, Tables 2 and 3 indicate that, among all tested models, the optimal one selected for a specific site might not be suitable for another site. For example, the BP model trained at a learning rate of 0.10 with six previous observations is considered as the best model for site Hann, and it has an RMSE value of 1.254, whereas the RBF model trained by using four previous observations appears to be the best for site Kulm, with an RMSE of 1.444. Therefore, it is not recommended to employ only one type of ANN model in forecasting the wind speed at different sites.

Actually, in many practical situations of the wind energy industry, a final single forecast that could take advantage of a set of plausible forecasts has to be produced [32]. For example, forecasts from alternative forecast agencies should be used since there is not a superior agency. Meanwhile, the forecast agencies themselves might make forecasts for the client by adopting alternative models or procedures such as different ANN models as introduced in this paper. In order to provide a single forecast, the agency has to combine all the available information. Therefore, it is apparent that an efficient forecast combination procedure might be of great importance for wind speed forecast. In this case, the Bayesian model averaging (BMA) method [33], an adaptive model combination method, might be used for post-processing different plausible forecasts to form a single forecast.

In addition, the algorithm efficiency of a model needs to be considered for short-term wind speed forecasting. The promptness of computation is essential when the forecast is in the range of minutes, while it is less critical but still desirable for 1-h ahead forecast. It is dependent upon many computationally important aspects of the network model such as the network topology, the signal propagation method, the activation function of the neurons, the input weights, the node numbers or nodal complexity in the network, etc. Feed forward networks usually terminate very quickly, while recurrent networks may never halt or convergence [24,34]. Conventional BP algorithms may fall into local optima, while RBF has been proven to ensure the global optima. According to the CTs given in Tables 2 and 3, it can be seen that the RBF model demonstrated the fastest convergence speed. This is because that RBF model does not perform parameter learning as in the BP networks, but just performs linear adjustment of the weights for the radial bases. Note that many improved BP algorithms have been developed to speed up the training and avoid the local optimum solution, but in general they are still not as efficient as RBF models.

It should be pointed out that, limited by the available data source, only one-year's data were adopted although the inter-annual variability might influence the forecast performance of ANN models, especially for monthly or longer forecast horizon. It is believed that the influence of the inter-annual variability on 1-h-ahead wind speed forecasting is relatively small. However, it will be ideal to have more training data from different years to address this influence, and this is actually a limitation of the current models obtained. Meanwhile, Figs. 7 and 10 illustrate that the predicted value, to some degree, lags behind the corresponding observation, indicating that the forecast value is closed related to previous, especially most recent, observations. However, this is still different from the persistence model, which assumes current observed value as the forecast value for the next hourly wind speed. In literature, some publications compared ANN models with the persistence model, and it was found that ANN models generally outperform the persistence model, but sometimes the advantage is not very significant [5,7,9].

6. Conclusive remarks

As wind energy is becoming the world's fastest growing source of clean and renewable energy, the predictability of wind power generation is essential for the integration of wind energy into the power system. This is highly related to accurate and reliable short-term wind speed forecasting, which remains a challenge due to the continuous fluctuations of the wind resource. In this paper, for the first time, the performances of different ANN based time series models for 1-h-ahead forecast of wind speed were comprehensively investigated based on the wind data sets from two observation sites in North Dakota. For each type of ANN models (i.e., BP, ADALINE, and RBF), the effects of different inputs and learning rates were examined in terms of multiple performance metrics. It was discovered that, different inputs and learning rates, as well as model structures, directly influence the forecast accuracy. Even with the same wind data set from a specific site, the choice of best-performing model may not be the same with different evaluation criteria. With different wind data sets, the choice becomes more inconsistent, even in term of the same statistics. Therefore, a robust method of combining forecasts from different ANN models is needed to overcome the inconsistency issue in model selection.

References

[1] Xydis G, Koroneos C, Loizidou M. Exergy analysis in a wind speed prognostic model as a wind farm sitting selection tool: a case study in Southern Greece. Appl Energy 2009;86(11):2411–20.
[2] Ucar A, Balo F. Investigation of wind characteristics and assessment of wind-generation potentiality in Uludag-Bursa, Turkey. Appl Energy 2009;86(3):333–9.
[3] Monfared M, Rastegar H, Kojabadi H. A new strategy for wind speed forecasting using artificial intelligent methods. Renew Energy 2009;34(3):845–8.
[4] Pinson P, Nielsen HA, Madsen H, Kariniotakis G. Skill forecasting from ensemble predictions of wind power. Appl Energy 2009;86(7–8):1326–34.
[5] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13(4):915–20.
[6] Luickx PJ, Delarue ED, D'haeseleer WD. Considerations on the backup of wind power: operational backup. Appl Energy 2008;85(9):787–99.
[7] Sfetsos A. A comparison of various forecasting techniques applied to mean hourly wind speed time series. Renew Energy 2000;21(1):23–35.
[8] Kaldellis JK, Kavadias KA, Filios AE. A new computational algorithm for the calculation of maximum wind energy penetration in autonomous electrical generation systems. Appl Energy 2009;86(7–8):1011–23.
[9] Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. Renew Sustain Energy Rev 2008;12(6):1725–44.
[10] Morales JM, Mínguez R, Conejo AJ. A methodology to generate statistically dependent wind speed scenarios. Appl Energy 2010;87(3):843–55.
[11] Fadare DA. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. Appl Energy 2010;87(3):934–42.
[12] Watson SJ, Landberg L, Halliday JA. Application of wind speed forecasting to the integration of wind energy into a large scale power system. IEE Proc Gen Transm Distrib 1994;141(4):357–62.
[13] Torres J, Garcia A, Deblas M, Defrancisco A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Sol Energy 2005;79(1):65–77.
[14] Kavasseri R, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34(5):1388–93.
[15] Lin L, Eriksson JT, Vihriala H, Soderlund L. Predicting wind behavior with neural networks. In: Proceeding the 1996 European wind energy conference. Goteborg, Sweden; 1996. p. 655–8.
[16] Alexiadis M. Short-term forecasting of wind speed and related electrical power. Sol Energy 1998;63(1):61–8.
[17] Beyer HG, Degner T, Haussmann J, Homan M, Rujan P. Short term forecast of wind speed and power output of a wind turbine with neural networks. In: Proceeding the second european congress on intelligent techniques and soft computing. Aachen (Germany); 1994.
[18] Kariniotakis G, Stavrakakis GS, Nogaret EF. Wind power forecasting using advanced neural network models. IEEE Trans Energy Convers 1996;11(4):762–7.
[19] More A, Deo MC. Forecasting wind with neural networks. Mar Struct 2003;16(1):35–49.
[20] Kariniotakis G, Stavrakakis GS, Nogaret EF. A fuzzy logic and neural network based wind power model. In: Proceeding the 1996 European wind energy conference. Goteborg (Sweden); 1996. p. 596–9.
[21] Wang X, Sideratos G, Hatziargyriou N, Tsoukalas LH. Wind speed forecasting for power system operational planning. In: 2004 International conference on probabilistic methods applied to power systems; 2004. p. 470–4.
[22] Lapedes A, Farber R. Nonlinear signal processing using neural networks: forecast and system modeling. Los Alamos National Lab. Los Alamos, NM. Technical report LAUR872662; 1987.

[23] Song YD. A new approach for wind speed prediction. Wind Eng 2000;24(1):35–47.
[24] Kisi O. Comparison of different ANN techniques in river flow prediction. Civ Eng Environ Syst 2007(24):211–31.
[25] Widrow B, Lehr MA. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. Proc IEEE 2002;78(9):1415–42.
[26] Mathew S. Wind energy-fundamentals, resource analysis and economics. Berlin: Springer-Verlag; 2006.
[27] Brockwell PJ, Davis RA. Introduction to time series and forecasting. New York: Springer; 2002.
[28] Wanas N, Auda G, Kamel MS, Karray F. On the optimal number of hidden nodes in a neural network. In: Proceeding of 1998 IEEE Canadian conference on electrical and computer engineering, vol. 2; 2002. p. 918–21.
[29] Cigizoglu HK, Kisi O. Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. Nord Hydrol 2005;36(1):1–16.
[30] Tseng CM, Jan CD, Wang JS, Wang CM. Application of artificial neural networks in typhoon surge forecasting. Ocean Eng 2007;34(11–12):1757–68.
[31] Daga M, Deo MC. Alternative data driven methods to estimate wind from waves by inverse modeling. Nat Hazards 2009;49(2):293–310.
[32] Sanchez I. Adaptive combination of forecasts with application to wind energy. Int J Forecast 2008;24(4):679–93.
[33] Wasserman L. Bayesian model selection and model averaging. J Math Psychol 2000;44(1):92–107.
[34] Orponen P. Computational complexity of neural networks: a survey. Nord J Comput 1994;1(1):94–110.