Short-Term Forecasting and Uncertainty Analysis of Wind Turbine Power Based on Long Short-Term Memory Network and Gaussian Mixture Model

Zhang Jinhua* 1, 2, Yan Jie2, David Infield3, Liu Yongqian2, Fue-sang Lien4

1. North China University Water Resources and Electric Power, Zhengzhou, 450046, China

2. State Key Laboratory of New Energy Power Systems (North China Electric Power University), Beijing, 102206, China

3. Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G11XW, UK)

4. Department of Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

*Corresponding author. Tel.: +86-13598068812; E-mail address: zhangjhwind@163.com

Abstract

Wind power plays a leading role in the development of renewable energy. However, the random nature of wind turbine power and its associated uncertainty create challenges in dispatching this power effectively in the power system, which can result in unnecessary curtailment of the wind turbine power. Improving the accuracy of wind turbine power forecasting is an effective measure for resolving such problems. This study uses a deep learning network to forecast the wind turbine power based on a long short-term memory network (LSTM) algorithm and uses the Gaussian mixture model (GMM) to analyze the error distribution characteristics of short-term wind turbine power forecasting. The LSTM algorithm is used to forecast the power and uncertainties for three wind turbines within a wind farm. According to numerical weather prediction (NWP) data and historical power data for three turbines, the forecasting accuracy of the turbine with the largest number of training samples is the best of the three. For one of the turbines, the LSTM, radial basis function (RBF), wavelet, deep belief network (DBN), back propagation neural networks (BPNN), and Elman neural network (ELMAN) have been used to forecast the wind turbine power. This study compares the results and demonstrates that LSTM can greatly improve the forecasting accuracy. Moreover, this study obtains different confidence intervals for the three units according to the GMM, mixture density neural network (MDN), and relevance vector machine (RVM) model results. The LSTM method is shown to have higher accuracy and faster convergence than the other methods. However, the GMM method has better performance and evaluation than other methods and thus has practical application value for wind turbine power dispatching.

Keywords: Long short-term memory network, Gaussian mixture model, Wind turbine power, Short-term prediction, Uncertainty analysis

1. Introduction

Owing to the abundance of wind energy and the need for a cleaner environment, wind power presently plays a leading role in the generation of electricity from renewable energy. As of the end of 2017, the newly installed capacity of wind power in China is 19.5 GW, and the cumulative total installed capacity has reached 170.9 GW [1]. For both capacities over seven consecutive years China has ranked first in the world [2]. It is estimated that by 2050, the installed capacity of wind power in China will reach 1.5 to 2 billion kW. China's power system presently exhibits a high proportion of grid-connected renewable generation, which is an important feature of future power grids. However, renewable energy such as wind is random and, to an certain extent, uncertain. This creates challenges for the integration of large-scale wind power such as the need for increased standby generation, potential curtailment of wind power, and difficulties in plant dispatching.

Wind power forecasting technology is an effective method for reducing the adverse impact of wind farms on the power grid and for increasing the penetration of wind power in the power system. Moreover, it provides technical support for the safety, stability, and economic operation of power systems, and wind farms. However, wind power forecasting is uncertain owing to a combination of wind speed uncertainty and variations in wind farm performance. In addition, several phenomena occur in the forecasting error distribution of wind power such as deviation, multiple peaks, and heavy-tailed distribution, which is a probability distribution model in which the tail is thicker than the exponential distribution. These phenomena will affect the normality of the forecasting error [3]. Moreover, the wind power confidence interval is not symmetrical around the central forecast point. Thus far, most research has focused on deterministic numerical forecasting methods that provide central point estimates. For power generation dispatching, however, the evaluation and analysis of the forecasting error can provide more complete bases for decision making [4]. Because large-scale wind farms are integrated into the grid, errors associated with wind power forecasting will reflect significant and important uncertainty. This will inevitably result in problems with the operation and scheduling of the power system; as the amount of wind power increases, so will the magnitude of the forecast errors. It is difficult to improve the forecasting accuracy from a technical perspective. However,

Peer-reviewed, accepted author manuscript of the following research article: Zhang, J., Yan, J., Infield, D., Liu, Y., & Lien, F. (2019). Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Applied Energy, 241, 229-244. https://doi.org/10.1016/j.apenergy.2019.03.044

existing technology can be used to provide a better understanding of the uncertainties which can then lead to improved plant dispatching and power system operation.

Wind power deterministic forecasting can be divided into physical methods, statistical methods, and learning methods [5-7]. Physical methods use weather modelling to forecast wind speed and then apply the wind turbine power curve to convert wind speed forecasts into wind power forecasts. Although this approach is suitable for medium- to long-term forecasting of wind power, the physical models are computationally complex and thus require considerable computing resources. Statistical methods are generally based on historical wind power data over an appropriate time period; statistical models are fitted using this data and are then employed to generate forecasts. Although such models have been found to perform well for ultra-short-term and short-term wind power forecasting, their forecasting accuracy is limited. Learning methods apply artificial intelligence to extract the relationship between input and output. Such models have a wide range of applications. The recent the rapid development of artificial intelligence has resulted in the availability of many deep learning algorithms which have been widely applied.

The concept of deep learning originates from the study of artificial neural networks. Multilayer perceptron with multiple hidden layers is a type of deep learning structure. Deep learning can combine low-level features to form more abstract high-level ones, which can represent an attribute category or feature to discover the distributed representations of data features. Deep learning overcomes the problems of over-fitting and slow training speed in traditional neural networks. The long short-term memory network (LSTM) algorithm represents the development of the recurrent neural network (RNN) circulation neural network, which is a special form of RNN. However, its universality is much better than that of traditional RNN. Many scholars use the LSTM algorithm for forecasting research [8] in many areas. LSTM has been used for short-term traffic forecasting [9] and for forecasting hourly day-ahead solar irradiance based on weather forecasting data [10]. In the latter study, LSTM is used to improve the prediction accuracy of back propagation neural networks (BPNN); however, the predicted time interval is hourly. Wang et al. [11] put forward a deep belief network (DBN) based on wind speed prediction models; their case studies verify that the proposed model is accurate and stable.

Jingjing Song et al. [12] used a novel combined model based on an advanced optimisation algorithm for short-term wind speed forecasting. Their approach combined mode-included improved complete ensemble empirical mode decomposition adaptive noise (ICEEMDAN) [13] and grey wolves optimisation (GWO) [14] algorithms, and their forecasting results were compared with the separate results of ICEEMDAN-BP, ICEEMDAN-Elman neural network (ENN) [15], ICEEMDAN-wavelet neural network (WNN) [16], and ICEEMDAN-general regression neural network (GRNN) [17].

Hui Liu et al. [18] used wavelet packet decomposition (WPD) [19], convolutional neural network (CNN) data [20], and convolutional long short-term memory network (CNNLSTM) data to forecast wind speed, where the historical data of 600-min and the next 100-min wind speed time series were taken as the training set and the test set, respectively. Yanfei Li et al. [21] used empirical wavelet transform (EWT) [22] decomposition, LSTM principal computing, regularised extreme learning machine (RELM) [23] subordinate computing, and inverse empirical wavelet transform (IEWT) reconstruction to forecast wind speed, where the historical data of 900-min wind speed time series at 10 min/data point was taken as the training set, and that of the next 100-min was used as the test set.

Several of the aforementioned studies including [11], [12], [18], and [21] include short-time or super short-time forecasting, and used the data of wind speed time series only; NWP data are not adopted in their methods. Therefore, wind power forecasting for power system dispatch is urgently needed.

Lopez Erick et al. [24] used echo state networks and LSTM to forecast wind power and the following 6 h of wind turbine power. In the present study, we research the short-term forecasting of wind turbine power based on the LSTM algorithm and NWP data.

The number of intervals of the deterministic forecasting error and the probability distribution of this error are collectively referred to as a probabilistic forecasting method. Two distinct types of probabilistic forecasting methods are used. The first obtains the quantiles by analysing the results from a deterministic forecasting model to estimate the probability distribution [25]. Common methods include quartile regression (QR) and similar approaches. The second obtains the probability density function of the forecasting error distribution directly [26]. Common methods applied to achieve this are kernel density estimation [27], Gaussian distribution fitting [28], beta distribution [29] and similar approaches. Al-Awami et al. [30] researched the accommodating wind uncertainty by using the conditional probability distribution function of the wind power output as the forecast level. Mixed Gaussian (MG) distribution was utilised for wind uncertainty.

Noting the impact of applications of artificial intelligence to the wind power industry, the present study uses LSTM algorithms based on a large amount of data to study wind turbine power forecasting and error distribution models based on numerical weather prediction (NWP). This method is also used for power forecasting and is compared with other classical forecasting methods. In addition, the MG model is used for error distribution analysis of the forecasting value obtained by the LSTM method, and the results are compared with those of other forecasting value error distribution methods.

2. Long Short-Term Memory Neural Network

2.1 Principle of LSTM neural network

The RNN constitutes a very powerful computational model that can represent almost all dynamics. However, the degree to which such potential can be exploited in practice is determined by the effectiveness of the adopted training procedures. A gradient-based approach using back propagation has significant limitations because the time evolution of the path integral depends on the weights assigned. This means that the back propagation error either disappears or explodes because the standard RNN cannot learn effectively when the time lag between the relevant input signal and the target signal is greater than 5–10 discrete time steps. The problem of vanishing error raises the issue of whether the standard RNNs can demonstrate practical advantages for feedforward networks based on time windows. A recent model, the LSTM algorithm (Hochreiter & Schmidhuber, 1997), is not affected by this problem [31].

LSTM can learn to bridge a minimum time lag of more than 1000 discrete time steps by implementing a constant error carousel (CEC) in special units called cells. A multiplication gate unit learns to open and close access to the cells. The LSTM learning algorithm is local in space and time; its computational complexity for each time step and weight is O(1). Moreover, it solves complex long-term tasks that have not been solved by previous RNN algorithms.

The basic unit in the hidden layer of the LSTM network is the storage block, which contains one or more memory units and a pair of adaptive multiplier gate control units which are input and output to all units in the block. The core of each storage cell is a loop self-connected linear unit known as the CEC, and its activation is known as the unit state. The CEC solves the problem of error disappearance. In the absence of new input or error signal input units, the local error reflux of CEC remains unchanged; neither growth nor attenuation occurs. CEC protects the forward flow and reverse flow errors by using the input and output gates, respectively. When the door is closed, i.e. when the activation time is about zero, the uncorrelated input and noise cannot enter the cell, and the state of the cell will not disturb the rest of the network. Figure 1 shows a memory block with a single unit. The unit state Sc is updated on the basis of its current state and three input sources, in which net_c input is to the cell itself, and net_{in} and net_{out} are inputs to the input and output gates, respectively.

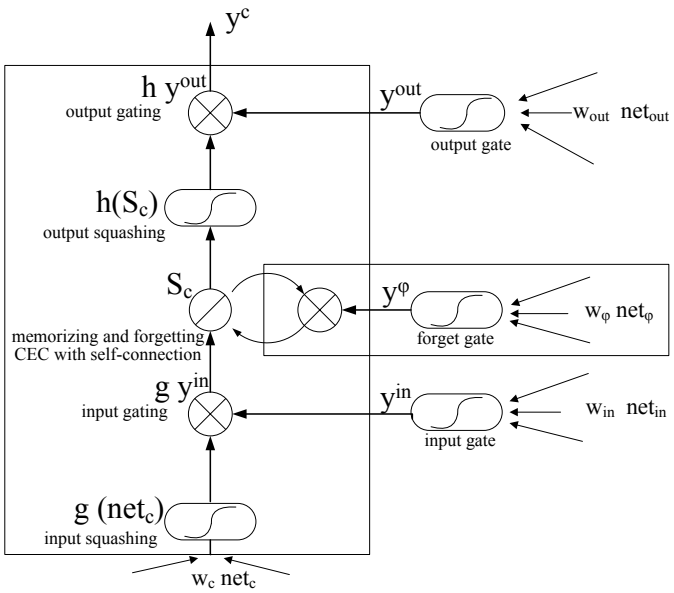


Fig. 1 Standard LSTM unit.

Consider discrete time steps t = 1, 2, …, where each step involves the updating of all units (forward pass) and the calculation of the error signals of all weights. The input gate activation y^{in} and the output gate activation y^{out} are calculated as follows:

$$\mathrm{net}_{out_j}(t) = \sum_m w_{out_j m} y^m(t-1), \quad \mathrm{y}^{out_j}(t) = \sum_m f_{out_j}(\mathrm{net}_{out_j}(t)) \tag{1}$$

$$\mathrm{net}_{in_j}(t) = \sum_m w_{in_j m} y^m(t-1), \quad \mathrm{y}^{in_j}(t) = \sum_m f_{in_j}(\mathrm{net}_{in_j}(t)) \tag{2}$$

where *j* is the pointer memory block and *v* is the storage unit in the index block *j*, so that c_j^v represents the v unit of the *j* storage block. w_{jm} is the weight of the connection from the unit *m* to the unit l. The index *m* scope is specified on all source units, such as the network topology. For the gate part, *f* is a logical function with a range of [0, 1]. The network input is compressed by *g*, and the logical *S* shape function is defined over [-2, 2].

The internal state of the memory cell $s_c(t)$ is calculated by adding the compressed gate control input to the state of the previous time step $s_c(t-1)(t > 0)$:

$$\mathrm{net}_{c_j^v}(t) = \sum_m w_{c_j^v m} y^m(t-1), \quad \mathrm{S}_{c_j^v}(t) = \mathrm{S}_{c_j^v}(t-1) + \mathrm{y}^{in_j}(t) g(net_{c_j^v}(t)) \tag{3}$$

When $\mathrm{S}_{c_j^v}(0)=0$, the internal state S_c is compressed by the output compression function h and is then multiplied by the (gated) output gate to activate y^{out} to calculate the cell output y^c:

$$\mathrm{y}^{c_j^v}(t) = \mathrm{y}^{out_j}(t) h(S_{c_j^v}(t)) \tag{4}$$

Where h is a centralised function with a range of [-1, 1].

Finally, assuming a hierarchical network topology with a standard input layer, hidden layer, and standard output layer composed of memory blocks, the equation of output unit k is

$$\mathrm{net}_k(\mathrm{t}) = \sum_m w_{km} y^m(\mathrm{t}-1), \quad \mathrm{y}^k(\mathrm{t}) = f_k(\mathrm{net}_k(\mathrm{t})) . \tag{5}$$

The M range covers all units of the feed output unit, which is usually the unit of the hidden layer and the input unit but not the memory block gate. As a squeeze function fk, we again use a logical S function with a range of [0, 1]. All equations except for equation (3) will still be valid for extended LSTMs with forget gates [32].

2.2 Training long short-term memory network

The training algorithm of the LSTM network is the reverse error propagation algorithm (BP). Because the LSTM processes sequence data training, the network requires back propagation of errors over the entire time series. Therefore, back propagation through time (BPTT) refers to the use of error back propagation on the graph structure of the LSTM with loops.

According to the LSTM structure in Figure 2, the current state of a cell will be affected by the previous cell state, which reflects the recurrent characteristics of LSTM. Moreover, when the error back propagation is calculated, the error of $y^c(t)$ includes the error at all times after time T in addition to the error of the current time T so that the errors at each time point can be calculated by $y^c(t)$ and Sc(t+1) iterations.

The forward propagation and back propagation processes of LSTM are shown in the figure, where $S_C(t)$ indicates the forgetting threshold for the input. The error of the output $y^c(t-1)$ of the unit at the previous moment is determined by the current unit output $y^c(t)$, and all of the gradients propagated by the gate layer are summed. The cell state Sc(t-1) is determined by the cell state Sc(t), and the error of Sc(t) includes two parts: $y^c(t)$ and Sc(t+1). Therefore, when calculating the Sc(t) back propagation error, yc(t) and Sc(t+1) are imported, whereas $y^c(t)$ needs to add $y^c(t+1)$ when updating. In this way, the gradient at any time from T can be calculated, and random gradient descent is used to complete the update of the weight coefficient [33].

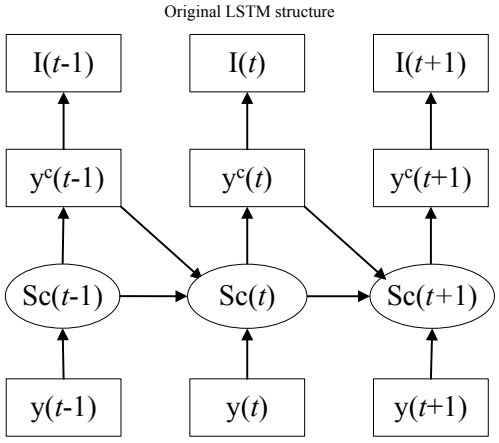


Fig. 2 LSTM network diagram.

3. Short-term wind turbine power forecast error distribution characteristics

Uncertainty analysis of wind power forecasting, also known as wind power probability forecasting or wind power interval forecasting, is used to calculate the range or probability that the result of single point forecasting may deviate from the actual power. Moreover, it is an extension of single point forecasting, which is used to assess the risk of a single point

of forecasting. The new energy power system demands higher single-point forecasting accuracy and the reliability and sharpness of uncertainty information of wind power.

The uncertainty analysis is still in the initial stage. Most of the existing methods assume that the forecasting error or output power is subject to a single probability distribution. Although this reduces accuracy, the known probability distribution is limited by its own mathematical characteristics. This creates difficulties in the solution of subsequent optimisation operation or increases the computational amount, which is not conducive to real-time online management of power systems. In this study, we consider using multiple distribution functions to describe the probability density distribution characteristics of wind power fluctuations, such as the Gaussian mixture model (GMM).

3.1 Gaussian mixture model

When using a GMM [34], the Gaussian probability density functions provide a flexible and accurate model for a wide range of probability distributions. Each GMM consists of *k* Gaussian distributions, and each Gaussian distribution is known as a component. These components are linearly added together to form the GMM probability density function [35]:

$$p(x_i) = \sum_{k=1}^{K} \pi_k p(x_i|k) = \sum_{k=1}^{K} \pi_k N(x_i|u_k, \Sigma_k) \tag{6}$$

where π_k are the weights assigned to the different components, and $p(x|k)$ represents the associated probability density functions.

GMMs normally employ expectation maximum (EM) for parameter evaluation. The maximum likelihood method is used to maximise the probability of the sample point giving the estimated probability density function. Because the probabilities are generally very small when N is large, the result of this multiplication is small and can cause floating point underflow. To avoid this, it is conventional to take logs for solving and derivation, and the objective is then rewritten as to maximise the log-likelihood function:

$$\max \sum_{i=1}^{N} \log p(x_i) = \max \sum_{i=1}^{N} \log(\sum_{k=1}^{K} \pi_k N(x_i|u_k, \Sigma_k)). \tag{7}$$

3.2 Expectation maximisation algorithm

The EM algorithm involves two stages. The first is to assume that the parameters of each Gaussian model are known and are based on these to estimate the weights; the second is to refine the parameters of the Gaussian models based on the estimated weights. These two operations are repeated until convergence occurs. The specific steps used to implement this are as given below.

Step1: Algorithm initialisation

The k-means clustering algorithm [36] is used to cluster the samples by using the mean value of each class as u_k and calculating Σ_k, π_k taking the proportion of each sample to the total number of samples.

Step 2: Estimation step (E-step)

The posterior probability is calculated:

$$\varpi_i(k) = \frac{\pi_k N(x_i|u_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x_i|u_j, \Sigma_j)}. \tag{8}$$

Step 3: M-step

The sample size is N, and the number of samples belonging to K classification is N_1, N_2, ..., N_k, respectively. Each Gaussian parameter is estimated from

$$\begin{aligned} \pi_k &= \frac{N_k}{N} \\ u_k &= \frac{1}{N_k} \sum_{i=1}^{N_k} \varpi_i(k) x_i \\ \Sigma_k &= \frac{1}{N_k} \sum_{i=1}^{N_k} \varpi_i(k)(x_i - u_k)(x_i - u_k)^T \end{aligned}. \tag{9}$$

Step 4: Iterate the steps, updating the above three values repeatedly until $p(x_i)$ converges.

4. Short-term forecasting model wind turbine power

4.1 Long short-term memory network forecasting model parameters

During forecasting, the selected test data are to input into the forecasting model, and the obtained forecasting value is compared with the real value to evaluate the advantages and disadvantages of the model. A model showing the relationship between training and testing is shown in Figure 3.

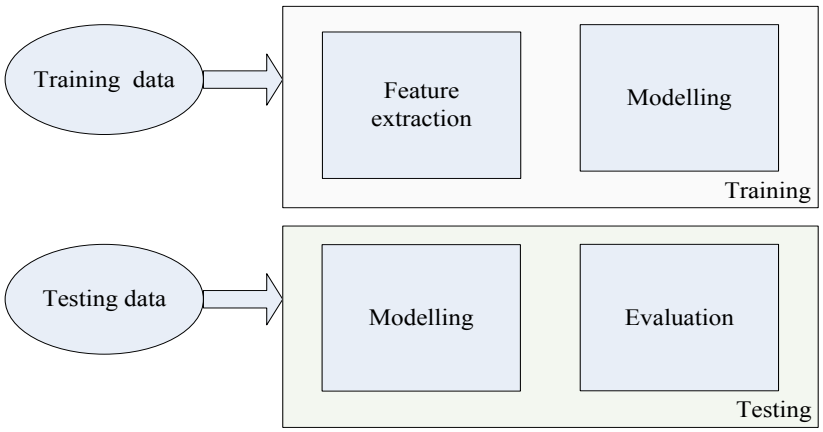


Fig. 3 Steps of modelling and evaluation used in this study.

Establishment of the LSTM network for short-term wind turbine power forecasting proceeds as described below.

The training model includes the training data and the model. The wind speed of the NWP data and the measured wind power are used for input and output training data, respectively. For the testing model, the wind speed and wind direction of the NWP data and the forecasting wind power data are used as the input and the output testing data, respectively.

The dimensions required for the LSTM input layer are three-dimensional and contain three parameters: batch_size, input_dim, and time_step. The hidden layer has one parameter: n_hidden. The output layer takes two parameters: n_hidden and output_dim. The batch_size of the input layer refers to the number of training inputs into the neural network at one time. The time_steps actually refer to the number of times that the constructed LSTM will input data; this is related to the number of previous successive data inputs. The meaning of input_dim is the number of data dimensions for input, which in this case is the number of data in one row. The hidden layer has a parameter, n_hidden, which represents the number of nodes used to remember and store past states for wind turbine power forecasting. Because this study considers only the LSTM layer forecasting model, this parameter is set to 1. The output layer parameter, output_dim, represents means the dimensions of the output data and is the only new parameter.

The input data can be constructed as tensor shape [n_samples, time_steps, input_dim], and the output data must be constructed as tensor shape [n_samples, time_steps, output_dim]. The input and output data could also be a matrix depending on the algorithm design.

The activation function of the input layer is sigmoid or tanh; that of the output layer is sigmoid.

In addition, batch_input_shape indicates the data format to be read for each batch; batch_size is the number of data rows read per batch, step size, and number of attributes; and epochs indicate the temporal structure of the network learning data and the number of cycles. Neurons represent the number of neurons, which is the number of memory cells and is generally in the range 1–5; the output layer is a single neuron and uses the sigmoid activation function to predict the data for the next time step. The loss function uses mean squared error (MSE), and the optimisation algorithm uses adaptive moment estimation (ADAM). ADAM uses the first moment estimation and the secondary moment estimation of the gradient to dynamically adjust the learning rate of each parameter. The advantage of ADAM is that the learning rate of each iteration has a certain range after bias correction so that the parameters are relatively stable. The parameters of ADAM optimisers in the LSTM model are lr, beta_1, beta_2, and epsilon. lr is known as the learning rate or step factor and controls the update rate of weight. A higher value, e.g. 0.3, will result in faster initial learning before the learning rate is updated; a smaller value, e.g. 1.0e-5, will enable the training convergence for better performance. In our program, lr is 0.001. Beta1 is the exponential decay rate of first order moment estimation, and beta_1 is 0.9. Beta_2 is exponential decay rate estimated by second order moments. This superparameter should be set to a number close to 1 in a sparse gradient; thus, beta_2 is 0.999 in our program. The parameter epsilon represents numerical stability. A very small number is needed to prevent dividing by zero in the implementation; thus, epsilon is 1e-08 here.

4.2 Forecasting evaluation index

The most commonly used wind turbine power forecasting performance indicators are root mean square error (RMSE) and mean absolute error (MAE), where the RMSE or is given by

$$e_{RMSE} = \frac{\sqrt{\sum_{t=1}^{N}\left(P_t^{act} - P_t^{pred}\right)^2}}{p_{cap}\sqrt{N}} \tag{10}$$

and the MAE is

$$e_{MAE} = \frac{\sum_{t=1}^{N}\left|P_t^{act} - P_t^{pred}\right|}{p_{cap} \times N} \tag{11}$$

where P_t^{act} is the actual measured value (kW or MW) of the output power of a wind farm or a single unit at time t; P_t^{pred} is the predicted value of output power (kW or MW) of a wind farm or single unit at time t; p_{cap} is the wind farm installed capacity (kW or MW); and N is the total number of samples.

According to the performance requirement in the functional specification of wind power forecasting system reported by State Grid Corporation of China, article IV, the RMSE of short-term forecasting of a single wind farm should be less than 20%. The RMSE of the predicted value of 4 h should be less than 15%.

4.3 Uncertainty evaluation index

The evaluation system reported in the literature [37] was used to verify the forecasting uncertainty analysis model, including three evaluation indicators: reliability, sharpness and skill score.

(1) Reliability tests correctness from the perspective of probability and is the difference between the proportion of effective data points and the preset confidence level. If the actual wind turbine output curve falls within the forecasting interval, the result is valid. If the ratio of the effective result is greater than or equal to the set confidence level, the result is reliable. The reliability index is the difference between the preset probability level and the probability that the actual power of the wind turbine falls within the forecasting interval. Therefore, a reliability index close to zero indicates that the model reliability is closer to the theoretical value of the algorithm. If $r^{(\tau)}$ is greater than zero, the model operation produces positive deviation, and the reliable value is higher than the theoretical level, which is a favorable deviation. On the contrary, the negative deviation of the model is harmful if the reliability value is lower than the theoretical level.

(2) Sharpness refers to the degree to which the forecasting interval is close to the actual output curve. Under the premise of ensuring reliability, a range narrower and closer to the actual curve relates to better performance of the uncertainty analysis model.

(3) Skill score is an index used for comprehensive evaluation of reliability and sharpness. A higher value relates to stronger comprehensive performance of the probabilistic forecasting model; zero represents the best model. If the model meets the requirement of reliability evaluation, the skill score can also be used as an index to evaluate the sharpness. Therefore, reliability index and skill score are selected as the criteria for evaluating the uncertainty analysis model.

The confidence level is τ, reliability index $r^{(\tau)}$, and skill score $S_S^{(\tau)}$ are defined as

$$\vartheta^{(\tau)} = \begin{cases} 1, & if \ \ w_t^{act} \le \hat{q}_t^{(\tau)} \\ 0, & otherwise \end{cases} \tag{12}$$

$$z^{(\tau)} = \sum_{t=1}^{N} \vartheta^{(\tau)} \tag{13}$$

$$r^{(\tau)} = \hat{a}^{(\tau)} - \tau = \frac{z^{(\tau)}}{N} - \tau \tag{14}$$

$$S_s^{(\tau)} = \frac{1}{N}\sum_{t=1}^{N}\left(\vartheta^{(\tau)} - \tau\right)\left(w_t^{act} - \hat{q}_t^{(\tau)}\right) \tag{15}$$

where τ is a predetermined level of confidence or probability; $\hat{q}_t^{(\tau)}$ is the limit of the confidence level interval (kW or MW) at confidence level τ at time t; $\vartheta^{(\tau)}$ is a binary indicator variable which represents whether the measured power

falls within the forecasting interval; $z^{(\tau)}$ is the number of data in which the actual wind power falls within the forecasting interval at the confidence level; and $S_s^{(\tau)}$ is the skill score at confidence level τ .

4.5 Data pre-processing

(1) Data filtering is applied to remove wind speeds where wind turbines are not normally operational, i.e. values below cut-in and above cut-out wind speeds. It is also necessary to remove data when wind turbines are not operational, i.e. those showing shutdown status. Below the cut-in, the power generated by the wind turbine can be negative; such values should be replaced by zero. Above the rated wind speed, the power is assumed to take the rated power value.

(2) Statistical analysis requires a large number of wind speed data points. Because the wind speed is collected in a 15 min sampling period, data analysis can be undertaken without concern for wind turbine inertial effects. The main object of this study is to assess wind speeds from NWP data. By analysing the autocorrelation of wind speed, the characteristics of wind speed are extracted. Forecasting models and modelling methods are selected and designed on the basis of the data analysis.

(3) To facilitate model development, the data normalisation process is commonly applied as

$$p_{0-1}(i) = \frac{p(i) - p(\min)}{p(\max) - p(\min)} * (max - min) + min \quad (16)$$

$$v_{0-1}(i) = \frac{v(i) - v(\min)}{v(\max) - v(\min)} * (max - min) + min \quad (17)$$

where the resulting range is then [0, 1].

5. Example analysis

Wind power forecasting technology is an effective method for reducing the adverse impacts of the wind power grid on the power system and for improving the proportion of wind power. Moreover, this approach provides technical support for the safe, stable, and economical operation of power systems and wind power plants. However, forecasting is uncertain, and single forecast results do not always meet the requirements of power grid and wind farm operation scheduling.

The uncertainty analysis of wind power forecasting, which is also referred to as uncertainty prediction, probability prediction, or interval prediction, reflects supplementation and expansion of certainty forecasting, which can provide a possible range of actual power under any confidence level. This provides an important reference for economical dispatching of a power system, unit commitment, power flow calculation, and other measures. The main points of this approach are detailed below.

(1) Reasonable arrangement of the spare capacity avoids the necessity of reserving a large amount of this capacity to cope with the forecast error which may be large but has a low probability of occurrence. This improves the acceptance ability of the grid to wind power and reduces the system operation cost.

(2) This analysis can provide comprehensive information of wind power availability to obtain the optimal conventional unit commitment, and to avoid the rapid and deep adjustment of conventional units through the estimation of error risk. This improves the economy of the power system and reduces the operation risk and wind curtailment.

5.1 Sources of data

Data from a wind farm of 123 units of the same type in North China for the first quarter of 2010 were used to train and test the forecasting model. The data include actual measured wind speed, measured wind turbine output power, and NWP data. The rated capacity of one wind turbine is 1500 kW, and the sampling period was 15 min. Owing to machine failures and other events, periods of missing data are inevitable. During the three-month period, 7393 wind speed data points were collected from one unit. As a result of the data filtering outlined above, the number of useful data points was reduced to 3072. Figure 4 shows the power output of one turbine over this period.

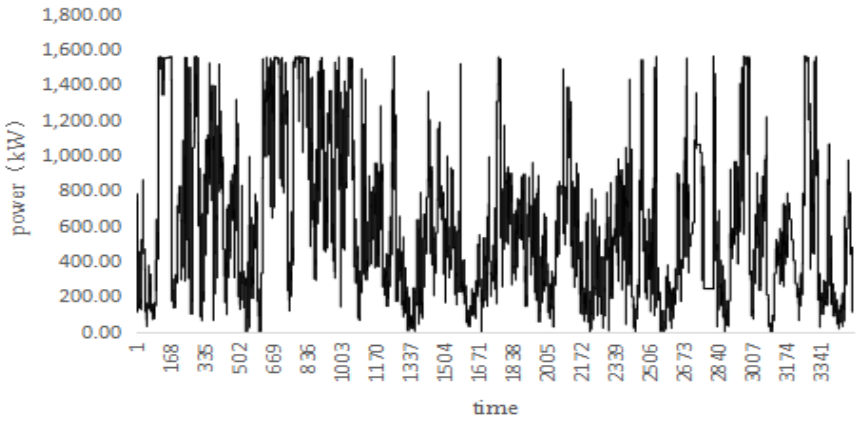


Fig. 4 Wind power curve of the wind farm during the first quarter of 2010.

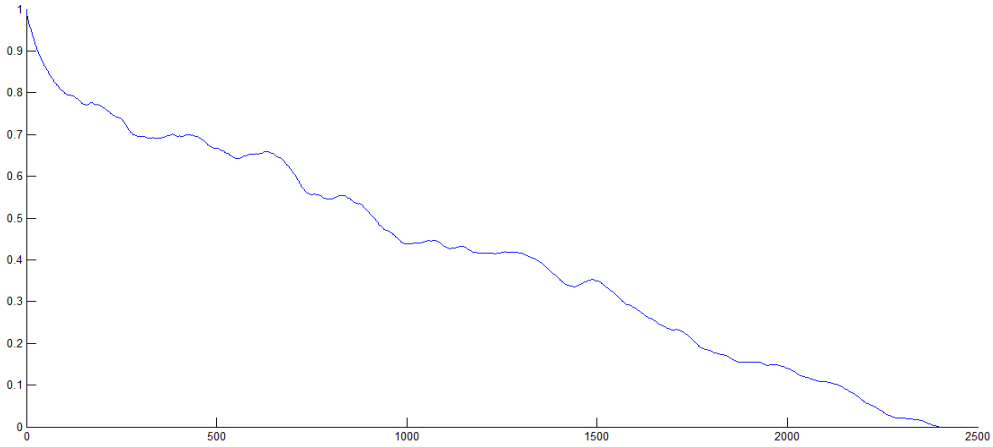


Fig. 5 Autocorrelation curve of wind speed data.

The autocorrelation analysis of the wind speed at the wind farm, shown in Figure 5, indicated an expected unity autocorrelation for zero time lag. In addition, a steady decrease in correlation with the time lag occurred, although some features reflected daily periodicity at 24 h with harmonics at 48 h and 72 h. Moreover, an exponential decrease in autocorrelation occurred during the first 6 h, which suggests that statistically based short-term forecasting is likely effective for look-ahead periods of less than 6 h.

5.2 Analysis of wind turbine power forecast results based on long short-term memory network

The input value is the wind speed data of NWP, and the output value is the wind power forecasting value of one unit. The range of wind speed is from 0 to 25 m/s because at wind speeds greater than 25 m/s, the wind turbine will be stopped. The range of wind power is from 0 to 1650 kW because the greatest power of the wind turbine is 1650 kW, which is 1.1 times the rated power (Pe = 1500 kW). First, we normalised the wind speed data of NWP and the real wind power value into [0, 1], and we then determined the training and testing data sizes. The training data include the input data (wind speed of NWP) and output data (wind power of each unit) to train the LSTM model, and the testing data use the input data (wind speed of NWP) and the trained LSTM model to forecast the wind power of each unit. The actual wind power value is used to check the accuracy of the LSTM model.

Because the input data of the LSTM algorithm are three-dimensional, the one-dimensional time-series data, including the training and testing data, are reshaped into three-dimension matrix data. The output data of the LSTM algorithm are one-dimensional. The look_back is the number of previous time steps to use as input variables for predicting the next time period. In the processor, the look_back value is 3. The input and output data include three dimensions of wind speed and one dimension of wind power, respectively, in the LSTM algorithm. In the time series of wind speed and wind power, the wind power corresponds to the third wind speed of three-dimensions of data. We forecasted that 192 points of wind turbine power in the future need 198 NWP data points (= 192 + 2*look_back). Therefore, the time horizon of the following figures includes 192 points at 15-min intervals, which reflects two-day short-term wind power forecasting. This forecasting value could be used for day-ahead optimal dispatching in a wind farm or power system.

5.2.1 Experiment I: comparison with different units

The training sample for unit #10 includes 1141 data points from February 1 to February 25; that for unit #16 includes 1149 data points from January 1 to January 23; and that for unit #58 training includes 9021 data points from January 1 to April 28. The NWP data of three units are classified as good data, general data, and bad data.

The forecasting of the three units was repeated five times. We obtained five results for each unit in one computation; each result included a range 1000 epochs and 5 updates in 1 repeat. The RMSE and running time results of the three units are shown in Table 1. The model of the LSTM algorithm is stable; the data are larger, and the result is closer to the actual value, showing high accuracy. Table 1 shows the results of five repeats, where the mean is the average value; std is the standard deviation; min and max are the minimum and maximum; 25%, 50%, 75% are the confidence levels of the results, time is the running time of the LSTM, data is the size of the unit's data. The specific forecasting results are indicated in Table 3. According to the results given in Tables 1–2 and Figures 6–10, the following points were considered.

Table 1. Comparison results of five forecasting operations of three units by LSTM.

Parameter	#10 RMSE	#16 RMSE	#58 RMSE
mean	0.123126	0.064306	0.315713
std	0.001410	0.000422	0.000147
min	0.121063	0.063704	0.315544
25%	0.122730	0.064001	0.315602
50%	0.122956	0.064369	0.315728
75%	0.124289	0.064540	0.315771
max	0.124593	0.064839	0.315919
time	550S	544S	3140S
data	1152	1149	9021

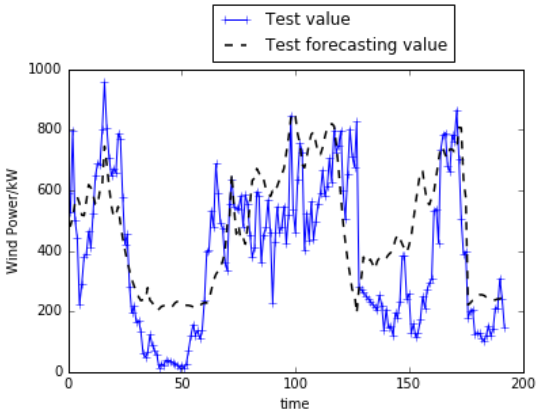
(a) Table 1 indicates that more sample data relates to longer program runs such as 9021 s, 540 s, and 50 s for units 58 16, and 10, respectively. In addition, the RMSEs of the 192 test values for units 10, 16, and 58 were 12.10%, 6.37%, and 31.55%, respectively. The data of unit #58 was larger than that of #10 and #16, and the operation time of unit #58 was longer than that of #10 and #16. Because the program runs five times, the operation time is longer than that in other models.

(b) The results shown as time series in Figures 6–8 reflect a 15-min interval between each point. In the figures, (a) shows the forecasting value and the test value; (b) shows the loss of the training forecasting value and the training value, represented by val_loss; and (c) shows the box plots of five-times repeated forecasting results of each unit. The LSTM algorithm converges steadily by 1000 epochs in (b) in all three figures, and the convergence of the test forecasting value is better than that of the training forecasting value in Figures 6 and 7(b). However, the convergence of the test forecasting value is not better than that of train forecasting value in Figure 8(b).

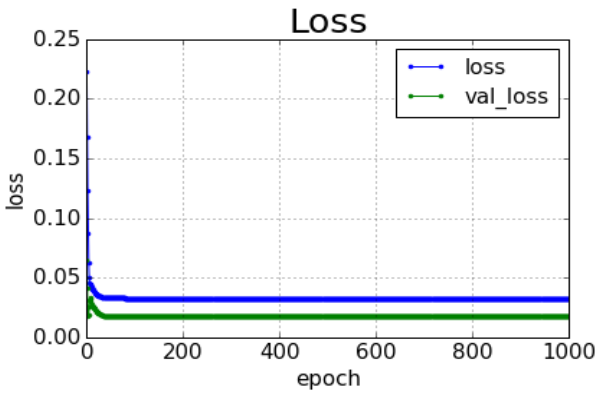
(c) A comparison of Figure 9(c) and Figure 8(a) revealed that the wind speed curve of NWP data is similar to the curve of wind turbine power forecasting and is different from the actual (test) value curve. A comparison of Figure 9(b) and Figure 7(a) revealed that the wind speed curve of NWP data is almost the same as the curve of the wind turbine power forecasting and the actual (test) value curve. A comparison of Figure 9(a) and Figure 6(a) revealed that the wind speed curve of the NWP data is similar to the curve of the wind turbine power forecasting and is somewhat similar to the actual (test) value curve. The main reason for the difference is the NWP data, the accuracy of which is important in wind turbine power forecasting. We can conclude that the NWP data of unit #16 was better than those of units 10 and 58 and that the NWP data of unit #58 was the worst of the three units.

(d) The three RMSE values of the wind turbine power forecasting are shown in (c) of Figures 6–8, which are drawn in accordance with Table 1. The max and min RMSE values of units 16 and 58 showed little difference; therefore, the std values of these units are smaller than that of unit #10. The performances of the forecasting models for the three units will be discussed subsequently in section 5.3.1.

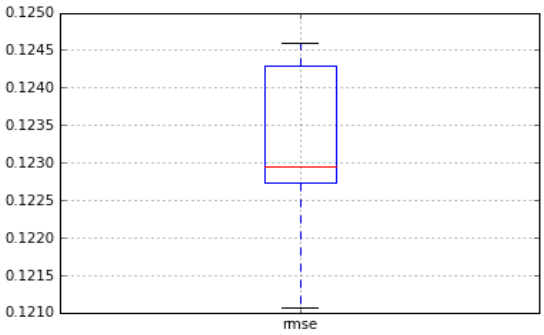
(e) The forecasting results in Table 2 show that the tests of three units have similar laws. In addition, the five repeated time steps in the forecasting results of LSTM model are the iteration of the LSTM learning process. The results were stable in each step.



(a) Wind power forecasting

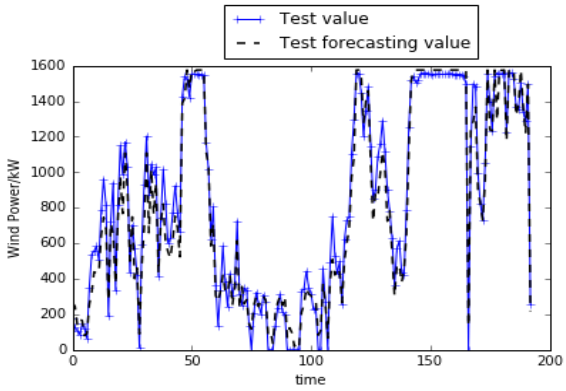


(b) Loss and val_loss of wind power forecasting

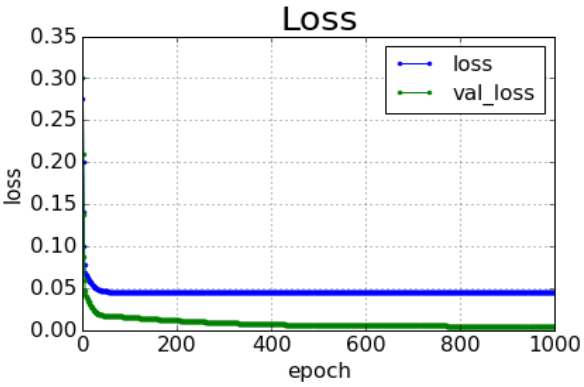


(c) RMSE of wind power forecasting

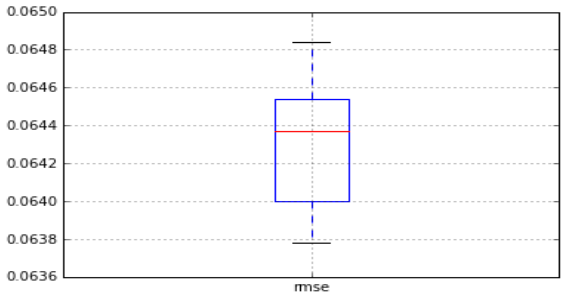
Fig. 6 Results of time series of wind power forecasting for unit #10.



(a) Wind power forecasting

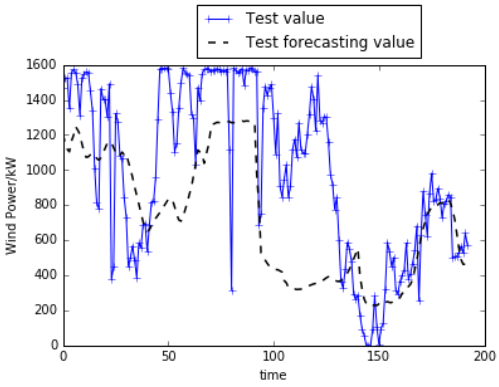


(b) Loss and val_loss of wind power forecasting

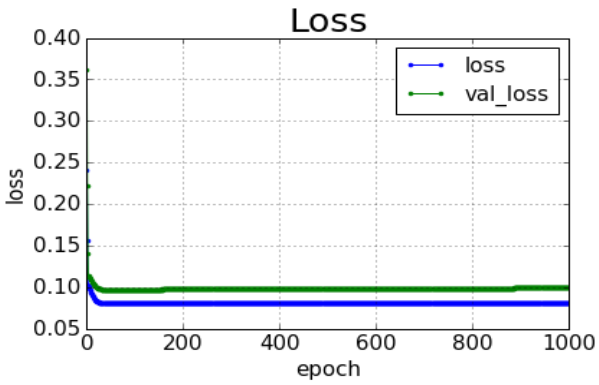


(c) RMSE of wind power forecasting

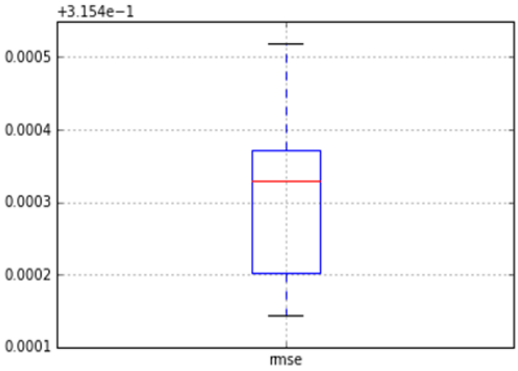
Fig. 7 Results of time series of wind power forecasting for unit #16.



(a) Wind power forecasting

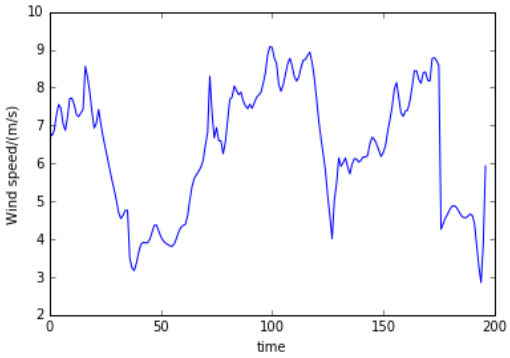


(b) Loss and val_loss of wind power forecasting

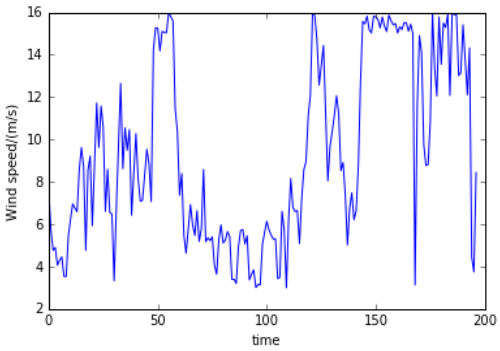


(c) RMSE of wind power forecasting

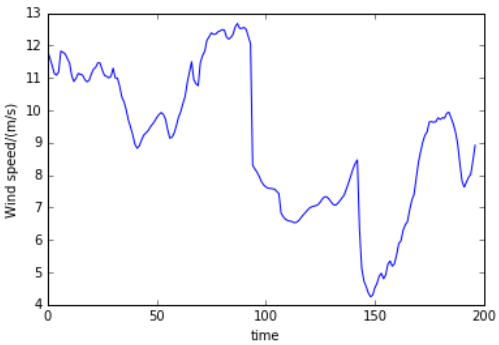
Fig. 8 Results of time series of wind power forecasting for unit #58.



(a)



(b)



(c)

Fig. 9 Wind speed of NWP for units (a) 10, (b) 16, and (c) 58.

Table 2 Error evaluation results of diff erent units for wind power.

Unit #10	Time step 1	Time step 2	Time step 3	Time step 4	Time step 5
count	5.000000	5.000000	5.000000	5.000000	5.000000
mean	0.123547	0.123869	0.123619	0.122706	0.123353
std	0.001657	0.001362	0.001841	0.001252	0.001234
min	0.121261	0.122293	0.121645	0.121217	0.121774
25%	0.122439	0.122577	0.121772	0.121747	0.122460
50%	0.124156	0.124297	0.124034	0.122740	0.123496
75%	0.124565	0.124870	0.124945	0.123583	0.124327
max	0.125315	0.125310	0.125698	0.124243	0.124705
Unit #16	Time step 1	Time step 2	Time step 3	Time step 4	Time step 5
count	5.000000	5.000000	5.000000	5.000000	5.000000
mean	0.064322	0.064306	0.064386	0.064338	0.064318
std	0.000465	0.000423	0.000247	0.000464	0.000437
min	0.063761	0.063783	0.064030	0.063818	0.063805
25%	0.064001	0.064004	0.064251	0.064001	0.064001
50%	0.064370	0.064366	0.064481	0.064370	0.064369
75%	0.064515	0.064519	0.064515	0.064492	0.064493
max	0.064961	0.064857	0.064655	0.065009	0.064923
Unit #58	Time step 1	Time step 2	Time step 3	Time step 4	Time step 5
count	5.000000	5.000000	5.000000	5.000000	5.000000
mean	0.314020	0.313983	0.313957	0.314050	0.313989
std	0.001084	0.000988	0.000994	0.001069	0.000951
min	0.312905	0.312930	0.312924	0.312908	0.313021
25%	0.313427	0.313412	0.313454	0.313547	0.313415
50%	0.313741	0.313777	0.313610	0.313756	0.313744
75%	0.314288	0.314288	0.314290	0.314296	0.314292
max	0.315740	0.315510	0.315507	0.315743	0.315471

5.2.2 Experiment II: comparison of classic models

By using the data of the three units, the RBF [38] method, the wavelet [39] method, the DBN method, the BP method, and the ELMAN method are compared with the predicted power value of the LSTM. According to Figures 9 and 10 and Table 3, the following points are considered.

(a) The different units adopt different methods and will thus show different prediction errors. The meteorological data of the three units were taken from different time periods. The most important factor influencing the accuracy of forecasting is the accuracy of wind speed in the NWP data. The NWP data of units 10 and 16 were better than that of unit #58; therefore, the accuracy of wind power forecasting by units 10 and 16 was better than that of unit #58. However, the three units are in different in geographic locations; therefore, the power generated by them will be different during the same period according to the wind speed, wake effect, topography, and shadow effect of each location.

(b) The RMSE of the wind power forecast value obtained by the RBF method was 7.8%; that of the wavelet, DBN, BP, ELMAN, and LSTM was 7.56%, 9.21%, 7.52%, 6.53%, and 6.37%, respectively. Thus the forecasting accuracy has been greatly improved. The parameter of the RBF model for unit #58 was adjusted to obtain the result of 32.25%, although the accuracy decreased and the operation time was reduced. If the same parameter as that for units 10 and 16 is, different results will be obtained.

(c) We chose the best results of each method shown in Table 3. The results of the LSTM model was the best of the six methods, and the results of the ELMAN model were also good. The running time of the BP model was the shortest among the methods.

(d) As shown in Figure 10(a) and (b) regarding the NWP data, the accuracy of the forecasting curves of the three units was similar. The forecasting curve of the Wavelet method was higher than the others in Figure 10(c); therefore it can be concluded that the Wavelet has better forecasting performance in high-speed areas of NWP data. Others methods showed better forecasting performance in middle–low-speed areas of the NWP data.

The results of this experiment indicate that LSTM has the best forecasting performance of the three units; the operation time of LSTM method was longer than that in other methods. Moreover, the differences in performance of NWP data strongly influenced the accuracy of the wind power forecasting. Therefore, the uncertainty of the wind power forecasting should be discussed.

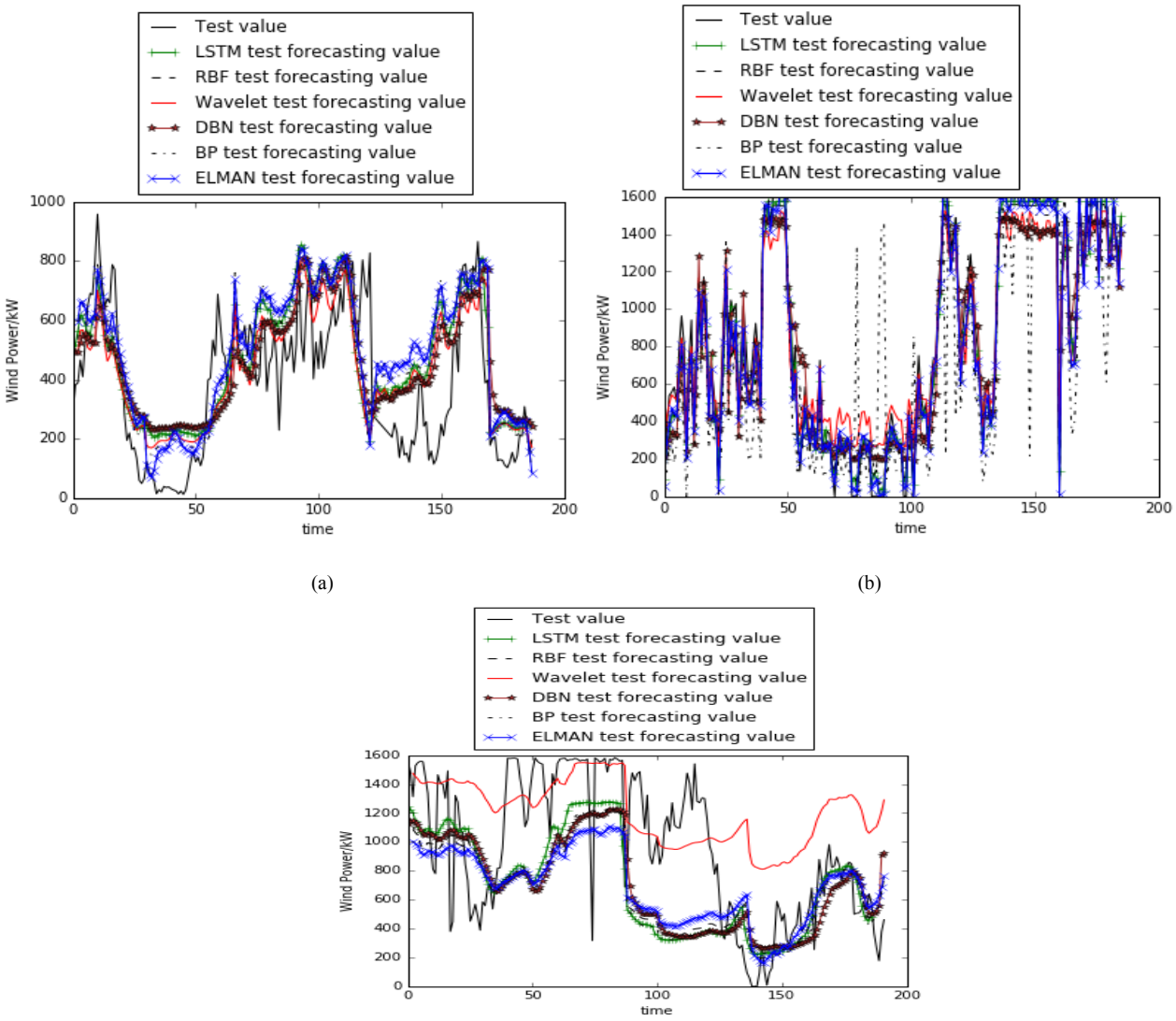


Fig. 10 Comparison of wind power forecasting for units (a) 10, (b) 16, and (c) 58.

Table 3. Comparison results of units 10, 16, and 58 among the different methods.

Unit #	Method	RMSE (%)	Time (s)
10	LSTM	12.10	550
	RBF	13.28	90
	Wavelet	12.29	25
	DBN	12.11	555
	BP	13.01	11
	ELMAN	13.69	12
16	LSTM	6.37	540
	RBF	7.70	82
	Wavelet	7.56	24
	DBN	9.21	510
	BP	7.52	10
	ELMAN	6.53	15
58	LSTM	31.29	3140
	RBF	36.25	10
	Wavelet	32.16	305
	DBN	32.51	1498
	BP	32.53	24
	ELMAN	31.69	45

5.3 Analysis of wind turbine power forecast result uncertainty based on the Gaussian mixture model

5.3.1 Experiment I: comparison with different units

Confidence interval comparison

Information on the surrounding environments of the wind farm, unit operation status, historical data, and NWP data affects the forecasting accuracy. The performance of a given predictive model will vary among wind farms to reflect the different turbines and array characteristics. Therefore, it is necessary to understand the factors that influence forecasting error such as the accuracy of the input data, the array design and the wind turbine rotor characteristics, and the likelihood for unit faulting. The error distribution function of the actual and predicted values of the wind turbine power using the GMM algorithm was established according to the training data of the wind farm unit, as shown in Figures 11–13.

Table 4. Comparison of different confidence intervals for forecast results of the three units.

Confidence interval	Unit #10	Unit #16	Unit #58
95%	±366.188 kW	±262.148 kW	±344.273 kW
90%	±364.469 kW	±259.145 kW	±341.118 kW
85%	±361.990 kW	±257.188 kW	±339.063 kW

As shown in Table 4, the 95% confidence interval for the future predicted value of each unit was ±366.188 kW, ±262.148 kW, and ±344.273 kW for units 10, 16, and 58, respectively. The 90% confidence interval for each unit's future predicted value was: ±364.469 kW, ±259.145 kW, and ±341.118 kW and the confidence interval for 85% of the predicted value of each unit was: ±361.990 kW, ± 257.188 kW, and ±339.063 kW for units 10, 16, and 5, respectively. The confidence interval of unit #10 was the greatest of the three units; this value is related to the std value in Table 2.

Uncertainty evaluation index comparison

Owing to the influence of uncontrollable factors which leads to differences in accuracy in numerical weather forecasting, operating conditions of wind turbines, wind gusts and other parameters, the output power of the wind farm fluctuates greatly, as does the fluctuation range of the predicted power. However, the deviation within a given confidence level originates from system error of the model itself.

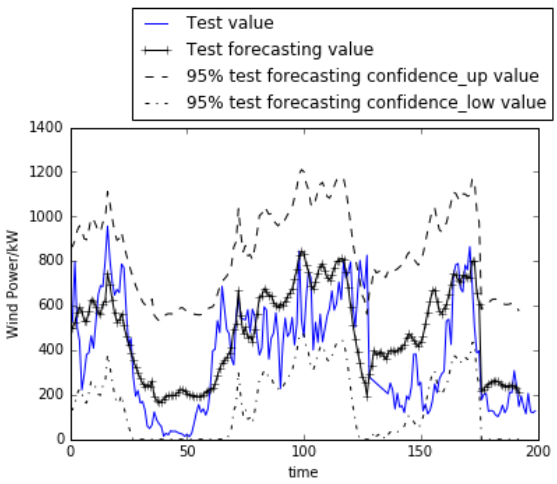
As shown in Table 5, we analysed three confidence interval values of 95%, 90% and 85%. Confidence represents the confidence interval, z_up is the number of data in which the actual wind turbine power fall within the forecasting up interval at the particular confidence level, r_up is the reliability index of z_up, z_low is the number of data in which the actual wind turbine power fall within the forecasting down interval at the particular confidence level, r_low is the reliability index of z_low, z is the number of data in which the actual wind turbine power fall within the forecasting interval at both confidence levels, and r is the reliability index of z. According to the table, the following points can be concluded.

(a) The probability indices of units 10 and 16 were almost greater than zero, including those for parameters r, r_up, and r_low. This shows that the model operation produces positive deviation and that the reliable values are higher than those at the theoretical level. The r_up values of units 10 and 16 were close to the confidence value. For unit #16, the r_low values were 0.050, 0.100, and 0.150 and the r_up values were 0.040, 0.090, and 0.140 at confidence levels of 95%, 90%, and 85%, respectively. In addition, the skill score of unit #16 was closer to zero than that in other units. Because the RMSE value of wind turbine power forecasting for unit #58 was greater than 30%, the probability index is not good. Moreover, the up and down intervals were not symmetric.

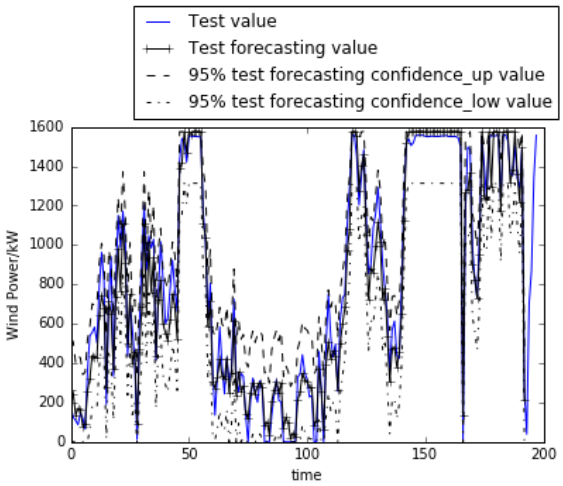
(b) The value range of each unit is related to the RMSE of the unit and the sample data amount. However, if the GMM analysis of the group data of the unit is considered, different confidence intervals are given for different error distributions, and the uncertainty analysis is more targeted and practical. This topic will be the focus of future work.

Table 5 Values of uncertainty analysis and evaluation under different confidence levels.

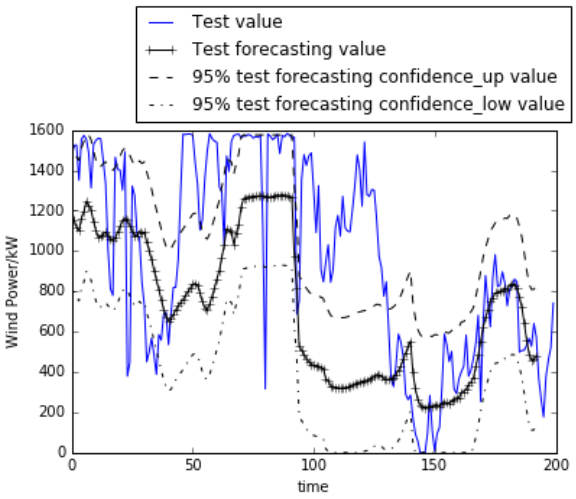
Parameter	**Unit #10**			**Unit #16**			**Unit #58**		
confidence	95%	90%	85%	95%	90%	85%	95%	90%	85%
skill score	-0.017	-0.032	-0.046	-0.006	-0.011	-0.017	-0.083	-0.086	-0.089
z_up	189	189	189	191	191	191	120	120	120
r_up	0.029	0.079	0.129	0.040	0.090	0.140	-0.328	-0.278	-0.228
z_low	186	186	186	192	192	192	184	184	184
r_low	0.014	0.064	0.114	0.050	0.100	0.150	0.003	0.0053	0.103
z	182	182	182	191	191	191	111	111	111
r	-0.007	0.043	0.093	0.040	0.090	0.140	-0.375	-0.325	-0.275



(a)

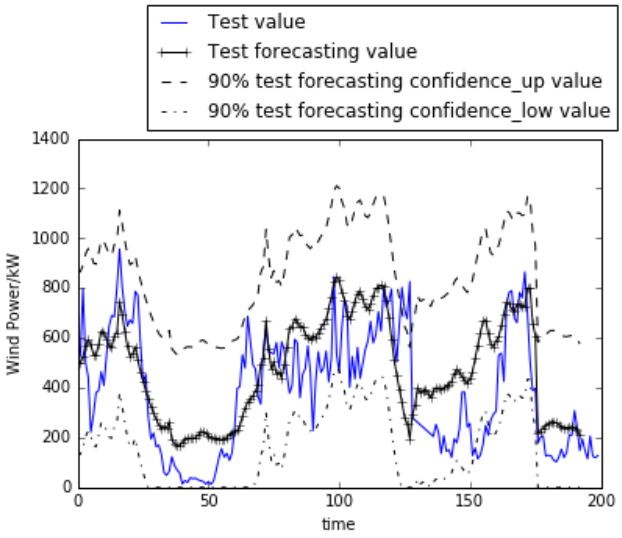


(b)

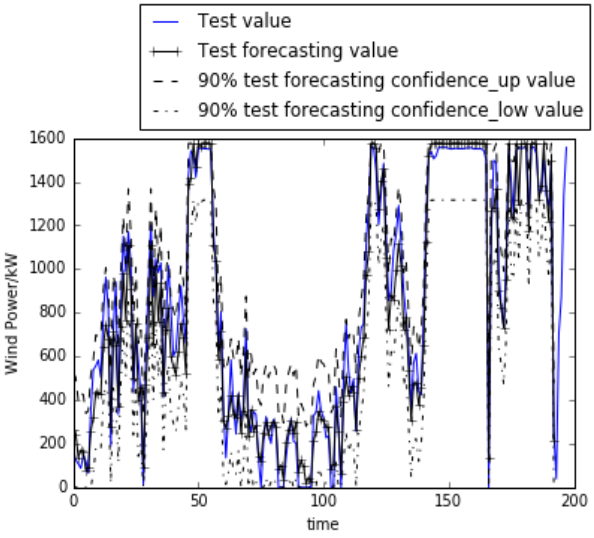


(c)

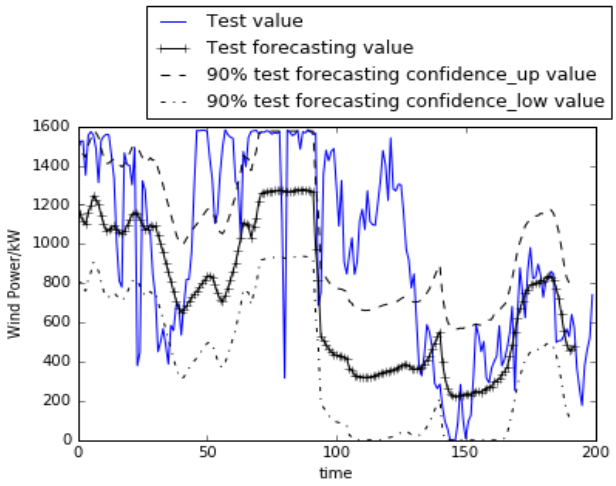
Fig. 11 Wind power forecasting at the 95% confidence interval for units (a) 10, (b), 16, and (c) 58.



(a)

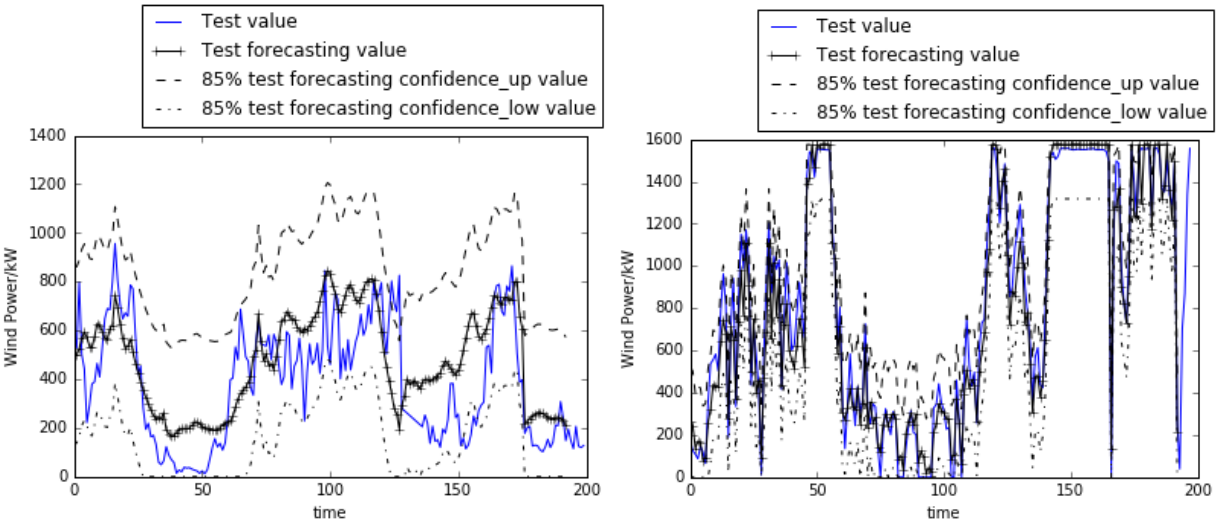


(b)

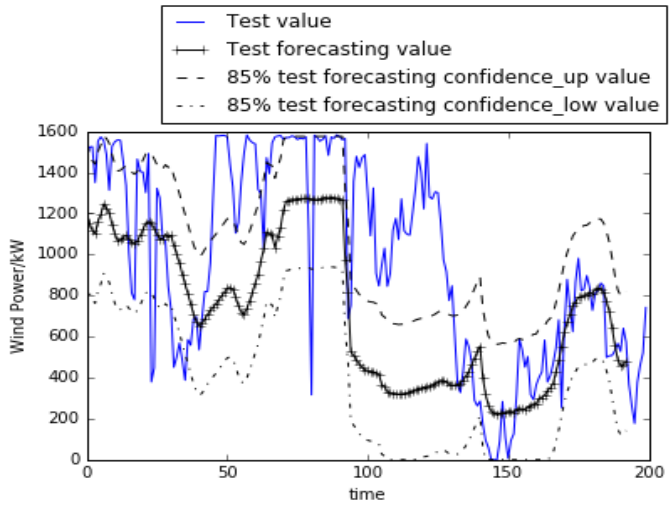


(c)

Fig. 12 Wind power forecasting at the 90% confidence interval for units (a) 10, (b), 16, and (c) 58.



(a) (b)



(c)

Fig. 13 Wind power forecasting at the 85% confidence interval for units (a) 10, (b), 16, and (c) 58.

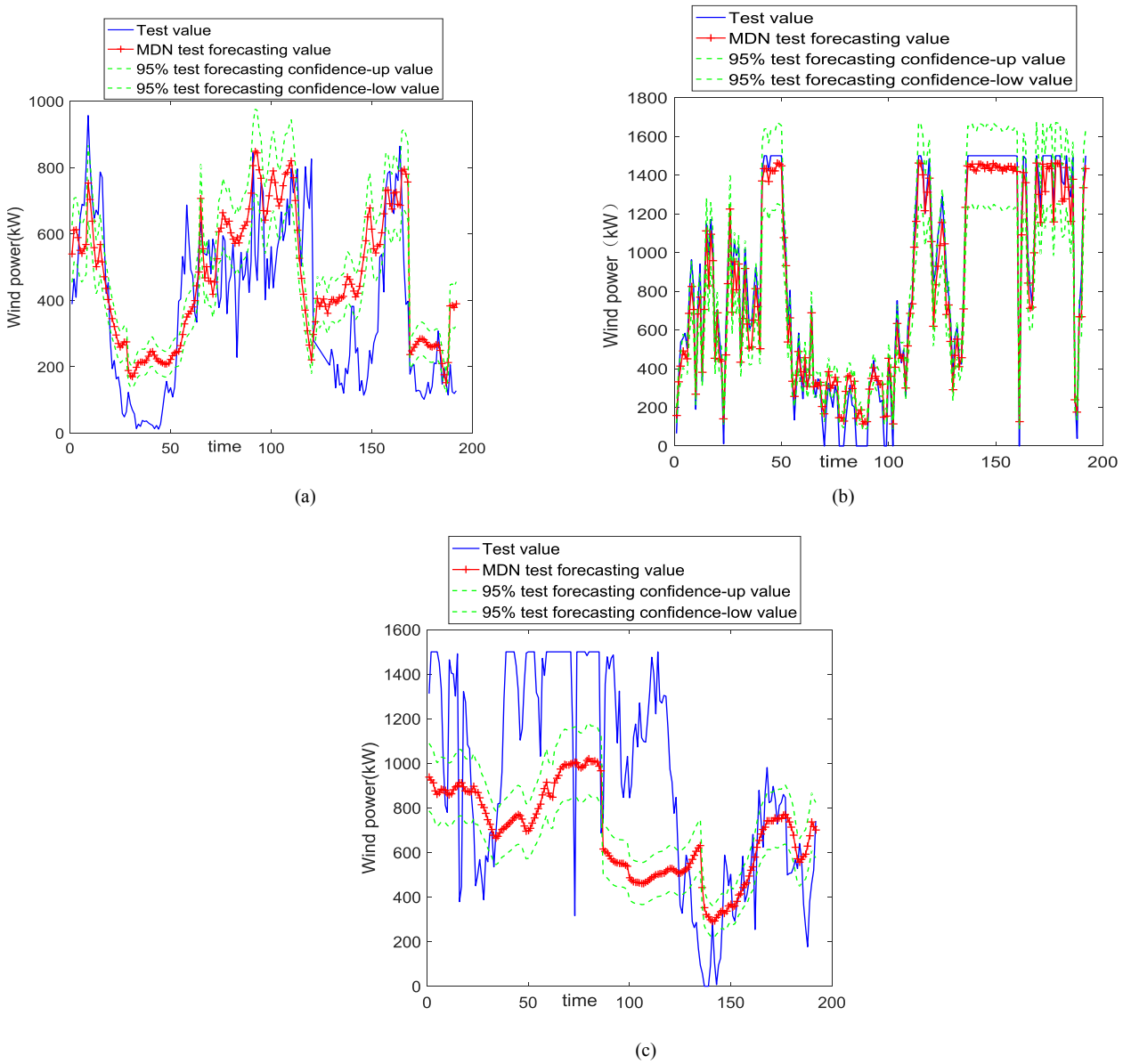


Fig. 14: Wind power forecasting at the 95% confidence interval by MDN model for units (a) 10, (b) 16, and (c) 58.

5.3.2 Experiment II: comparison with mixture density neural networks and relevance vector machine models

Mixture density neural networks (MDN) [40] can be used to estimate or learn the probabilistic relationship between the input and output values of unknown mapping by using a training data set consisting of a set of examples of the input and output values of the mapping [41]. Relevance vector machine (RVM) is a new supervised learning method similar to SVM. Although it has the same functional form as SVM, RVM provides uncertainty analysis capabilities [42]. MDN and RVM were used for comparison with the GMM method for uncertainty analysis and evaluation. The research focuses mainly on that 95% confidence interval of wind turbine power forecasting by GMM, MDN, and RVM methods as shown in Figures 14 and 15 and Table 6.

(a) In Table 6, the results of MDN method are divided into three types according to the numbers of epochs. The running times of the three types are different; the epoch is greater, and the running time is longer. The 100 epoch had the longest running time and the most accurate result of three types. RVM had a better forecasting effect on small sample data, and its forecasting and generalisation ability were better. Moreover, the running time of RVM method was very short, and the result accuracy was the best..

(b) The green curves show the 95% confidence interval of the wind turbine power in Figure 15; these curves are narrower than those of the wind turbine power in shown in Figure 14. The forecasting value was close to the actual value; moreover, the confidence interval of the forecasting value was narrower, and the performance of the confidence interval was better. As shown in Table 7, the GMM method was the best of the three methods. The indices of reliability of the three units were close to zero, and the number of data in which the actual wind turbine power fall within the forecasting interval

was the largest. The skill score also showed a value close to zero, and the performance of the confidence interval was better.

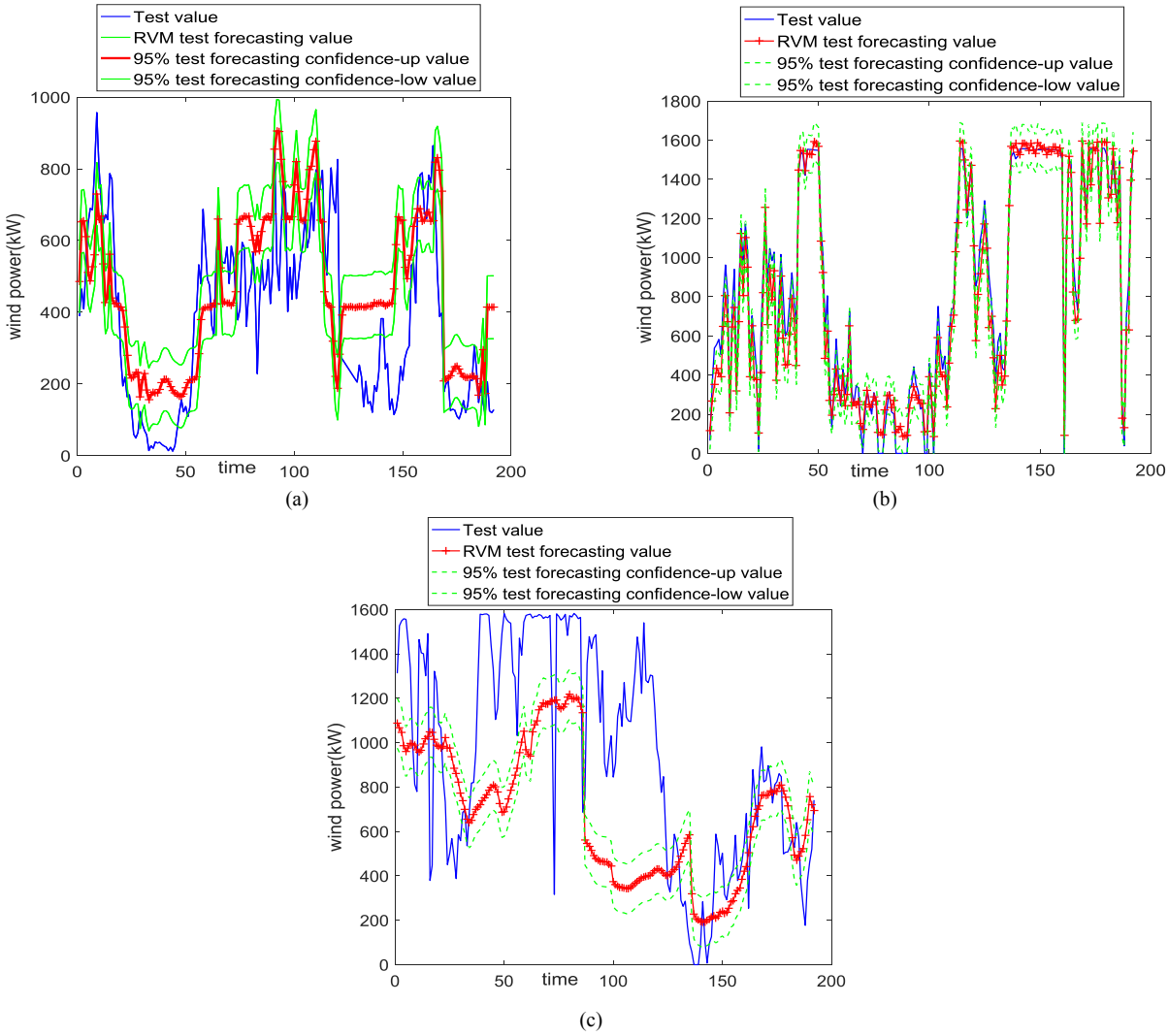


Fig. 15: Wind power forecasting at the 95% confidence interval by RVM Model for units (a) 10, (b) 16, and (c) 58.

Table 6 Wind power forecasting by MDN model for different epochs and by the RVM method.

Unit #	Method	MDN of 20 epochs	MDN of 50 epochs	MDN of 100 epochs	RVM
10	RMSE	0.1365	0.1341	0.1286	0.1306
	Time (s)	198	792	1769	2.4
16	RMSE	0.0797	0.0729	0.0615	0.0639
	Time (s)	142	408	1480	2.1
58	RMSE	0.3899	0.3388	0.3125	0.3223
	Time (s)	2172	9063	20371	9.5

Table 7 Uncertainty analysis and evaluation of the different methods.

Parameter	Unit #10			Unit #16			Unit #58		
method	GMM	MDN	RVM	GMM	MDN	RVM	GMM	MDN	RVM
skill score	-0.017	16.670	12.383	-0.006	2.205	-0.806	-0.083	-18.928	-27.104
z_up	189	164	146	191	166	128	120	80	45
r_up	0.029	-0.096	-0.190	0.040	-0.085	-0.283	-0.328	-0.533	-0.658
z_low	186	76	66	192	161	174	184	152	146
r_low	0.014	-0.554	-0.606	0.050	-0.112	-0.044	0.003	-0.158	-0.190
z	182	48	20	191	135	110	111	40	10

r	-0.007	-0.700	-0.846	0.040	-0.247	-0.377	-0.375	-0.742	-0.898

6. Conclusion

In this paper, the LSTM model was used to realise wind turbine power forecasting, and GMM was used to describe the uncertainty and confidence intervals. To verify and compare the prediction performance of the LSTM model, the RBF, and wavelet model were used. The MDN and RVM models were used to compare the confidence interval performance of the GMM. According to experimental tests, the following results were noted.

(1) LSTM implements short-term wind turbine power forecasting based on wind speed data in NWP, realising multi-dimensional data input and multi-point data output. The data related to wind speed in NWP can be used as input data in subsequent studies.

(2) Among all involved models, the LSTM model is robust and effective in forecasting the wind turbine power time series two days ahead and has the best performance in one to five time-step forecasting.

(3) The GMM provides an uncertainty quantification for wind turbine power forecasting in the form of a confidence interval at a prescribed confidence level. It has the best performance among the GMM and the MDN and RVM models.

(4) The GMM has an uncertainty analysis capability that reflects the confidence intervals. However, if the analyst can classify the weather data and then predict its uncertainty, the forecasting this should be improved. This topic will be the focus of future research.

Acknowledgements

The study is fully supported by the National Natural Science Foundation of China funded project - The multi-scale uncertainty of wind power forecasting and its impact mechanism on power system economic dispatch (51707063), Key scientific research projects of Henan colleges and universities (17A470002, 162102310470); The study is partly supported by Program for Innovative Research Team (in Science and Technology) in University of Henan Province.

References

[1] China Wind Energy Professional Committee of the Renewable Energy Society. Wind Power Capacity Statistics for 2017 [R]. 2018.

[2] Global Wind Energy Council. The Global Wind Energy Statistics 2017[R]. 2018.

[3] Liu Yanhua, Li Weihua, Liu Chong, Zhang Dongying. Mixed skewed distribution model of short-term wind power forecasting error [J]. Proceedings of the CSEE, 2015, 35(10): 2375-2382.

[4] Wang Caixia, Lu Zongxiang. Application of wind power forecast information in daily unit combination [J]. Automation of Electric Power Systems, 2011, 35(7): 13-18.

[5] Peng Xiaosheng, Xiong Lei, Wen Jinyu, Cheng Shijie, Deng Diyuan, Feng Shuanglei, Wang Bo. A summary of improved methods for short-term and ultra-short-term power forecasting accuracy of wind power clusters [J]. Proceeding of the CSEE, 2016, 36(23): 6315- 6326.

[6] Jung J, Broadwater RP. Current Status and Advances for Wind Speed and Power Forecasting[J]Renewable and Sustainable Energy Reviews,2014,31:762-777.

[7] Qian Zheng, Pei Yan, Cao Lijun, WANG Jingyi, JING Bo. Review of Wind Power Forecasting Methods [J]. High Voltage Technology, 2016, 42(4): 1047-1060.

[8] Hauser, Michael; Fu, Yiwei; Phoha, Shashi, Ray, Asok. Neural Probabilistic Forecasting of Symbolic Sequences with Long Short-Term Memory [J]. OURNAL OF DYNAMIC SYSTEMS MEASUREMENT AND CONTROL-TRANSACTIONS OF THE ASME 2018, 140(8): 084502.

[9] Zhao, Zheng; Chen, Weihai; Wu, Xingming, Chen, Peter C. Y., Liu, Jingmeng. LSTM network: a deep learning approach for short-term traffic forecast [J]. IET INTELLIGENT TRANSPORT SYSTEMS, MAR 2017,11(2): 68-75

[10] Qing, Xiangyun; Niu, Yugang. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM [J]. Energy, APR 2018(148):461-468.

[11] Wang HZ, Wang GB, Li GQ, Peng JC, Liu YT. Deep belief network based deterministic and probabilistic wind speed forecasting approach [J]. Applied Energy, 2016; 182:80–93.

[12] Jingjing Song, Jianzhou Wang, Haiyan Lu. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting [J].Applied Energy, FEB 2018(215):643-658.

[13] Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: a suitable tool for biomedical signal processing [J]. Biomed Signal Process Control, 2014; 14:19–29. http://dx.doi.org/10.1016/j.bspc.2014.06.009.

[14] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. [J]. Adv Eng Softw 2014; 69:46–61. http://dx.doi.org/10.1016/j.advengsoft.2013.12.007.

[15] Liu H, Tian H, Liang X, Li Y. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. [J]. Applied Energy, Aug 2015; 157:183–94. http://dx.doi.org/10.1016/j.apenergy.2015.08.014.

[16] Xiao L, Shao W, Yu M, Ma J, Jin C. Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting [J].Applied Energy, Apr 2017; 198:203–22. http://dx.doi.org/10. 1016/j.apenergy.2017.04.039.

[17] Marvuglia A, Messineo A. Monitoring of wind farms' power curves using machine learning techniques [J]. Applied Energy, Apr 2012; 98:574–83. Http: //dx.doi.org/10.1016/j. apenergy.2012.04.037.

[18] Liu Hui, Mi Xiwei Li Yanfei. Smart multi-step deep learning model for wind speed forecasting based on vibrational mode decomposition, singular spectrum analysis, LSTM network and ELM [J]. Energy Conversion and Management, MAR 2018(166):120-131. DOI: 10.1016/j.enconman.2018.01.010

[19] Meng A, Ge J, Yin H, Chen S. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm[J]. Energy Conversion and Management, 2016; 114:75–88. http://dx.doi.org/10.1016/j. enconman.2016.02.013

[20] Sun W, Wang Y. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network [J]. Energy Conversion and Management, 2018; 157:1–12. http://dx.doi.org/10.1016/j.enconman.2017.11.067.

[21] Li Yanfei; Wu Haiping; Liu Hui. Multi-step wind speed forecasting using EWT ecomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction [J]. Energy Conversion and Management, JUL 2018(167): 203-219.

[22] Empirical Gilles J, Transform Wavelet [J]. IEEE Trans Signal Process, 2013; 61:3999–4010.

[23] Sun W, LiuM.Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China [J]. Energy Conversion and Management, Apr 2016; 114:197–208.

[24] Lopez, Erick, Valle, Carlos, Allende Hector. Wind Power Forecasting Based on Echo State Networks and Long Short-Term Memory [J]. ENERGIES, MAR 2018, 11(4):526, doi: 10.3390/en11030526.

[25] Zhang Yao, Wang Jianxue, Wang Xifan. Review on probabilistic forecasting of wind power generation [J]. Renewable & Sustainable Energy Reviews, 2014, 32(4): 255-270.

[26] Yan Jie, Liu Yongqian, Han Shuang, Wang, Yimei, Feng, Shuanglei. Reviews on uncertainty analysis of wind power forecasting [J]. Renewable & Sustainable Energy Reviews, 2015, 52(12):1322-1330

[27] Sun Jianbo, Wu Xiaoshan. Zhang Buhan. Wind Power Interval Forecasting Based on Nonparametric Kernel Density Estimation [J]. Hydroelectric Energy Science, 2013, 31(9):233-235.

[28] Yang Mao, Dong Juncheng. Real-time forecasting error analysis of wind power based on mixed Gaussian distribution [J]. Chinese Journal of Solar Energy, 2016, 37(6): 1594-1602.

[29] Yang Hong, Yuan Jinsha, Zhang Tiefeng. A model and algorithm for the minimum probability interval of wind power forecasting error based on Beta distribution [J]. Proceedings of the CSEE, 2015, 35(9):2135-2142.

[30] Al-Awami, Ali T.; Ansari, M. Abdul Hafeez; Bennett, Brian J . Stochastic Dynamic Economic Dispatch for Grids with Significant Wind Using Mixed Gaussian Distribution [J]. Arabian Journal for Science and Engineer G, FEB 2016, 41(2): 545-553.

[31] Sepp Hochreiter; Jürgen Schmidhuber. Long short-term memory [J]. Neural Computation. 1997, 9 (8): 1735-1780

[32] . Felix A. Gers; Jürgen Schmidhuber; Fred Cummins. Learning to Forget: Continual Prediction with LSTM [J]. Neural Computation. 2000, 12 (10): 2451–2471.

[33] Qing Xiangyun, Niu Yugang, Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM [J]. Energy. 2018, 148(4): 461-468.

[34] Bishop, Christopher. Pattern recognition and machine learning. New York: Springer. 2006, ISBN 978-0-387-31073-2.

[35] Sarabi, Siyamak; Davigny, Arnaud; Courtecuisse, Vincent. Potential of vehicle-to-grid ancillary services considering the uncertainties in plug-in electric vehicle availability and service/localization limitations in distribution grids [J].Applied Energy, 2016, 171(1): 523-540.

[36] Anil K J. Data clustering: 50 years beyond K-Means [J]. Pattern Recognition Letters, 2010, 31(8): 651-666.

[37] P. Pinson, H.A. Nielsen, J.K. Møller, H. Madsen, G. Kariniotakis. Nonparametric probabilistic forecasts of wind power: required properties and evaluation [J]. Wind Energy, 2007, 10: 497-516.

[38] Chang, Wen-Yeau. An RBF Neural Network Combined with OLS Algorithm and Genetic Algorithm for Short-Term Wind Power Forecasting [J]. Journal of applied mathematics, Article Number: 971389, 2013.

[39] Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil [J]. Applied Energy, AUG 15 2018, 224: 550-566

[40] Bishop, C. M. Mixture density networks [R]. Technical Report NCRG/94/004, Published: 1994

[41] Men Zhongxian, Yee Eugene, Lien Fue-Sang. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks [J]. Renewable Energy, March 2016(87):203-211.

[42] M.E. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine [J]. Journal of Machine Learning Research, 2001, 1:211-244.