Energy Conversion and Management 165 (2018) 681–695

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization

Jie Chen[a], Guo-Qiang Zeng[b], Wuneng Zhou[a,*], Wei Du[a], Kang-Di Lu[a]

[a] School of Information Sciences and Technology, Donghua University, Shanghai 200051, China
[b] National-Local Joint Engineering Laboratory of Digitalize Electrical Design Technology, Wenzhou University, Wenzhou 325035, China

ARTICLE INFO

Keywords:
Wind speed forecasting
Deep learning
Time series prediction
LSTMs (Long Short Term Memory neural networks)
Ensemble learning
Extremal optimization

ABSTRACT

As an essential issue in wind energy industry, wind speed forecasting plays a vital role in optimal scheduling and control of wind energy generation and conversion. In this paper, a novel method called EnsemLSTM is proposed by using nonlinear-learning ensemble of deep learning time series prediction based on LSTMs (Long Short Term Memory neural networks), SVRM (support vector regression machine) and EO (extremal optimization algorithm). First, in order to avert the drawback of weak generalization capability and robustness of a single deep learning approach when facing diversiform data, a cluster of LSTMs with diverse hidden layers and neurons are employed to explore and exploit the implicit information of wind speed time series. Then predictions of LSTMs are aggregated into a nonlinear-learning regression top-layer composed of SVRM and the EO is introduced to optimize the parameters of the top-layer. Lastly, the final ensemble prediction for wind speed is given by the fine-turning top-layer. The proposed EnsemLSTM is applied on two case studies data collected from a wind farm in Inner Mongolia, China, to perform ten-minute ahead utmost short term wind speed forecasting and one-hour ahead short term wind speed forecasting. Statistical tests of experimental results compared with other popular prediction models demonstrated the proposed EnsemLSTM can achieve a better forecasting performance.

1. Introduction

As a promising and practical solution to cut greenhouse gas emissions and build a renewable society, wind energy is becoming more and more popular in various countries. The global wind report, released by the Global Wind Energy Council (GWEC) in 2017, has stated that the 2016 world wind power market was more than 54.6 GW, causing the total global installed capacity to nearly 487 GW, which was still led by China, US, Germany and India [1]. And the capacity of wind energy will continue to grow vastly in next years. However, it can be a difficult task to perform a reliable and seasonable wind power management in electrical power systems due to the natural irregular characteristic of wind speed. The unstable and uncontrollable wind speed influences heavily the generation of wind power and subsequently this will impact wind turbines control, power systems and micro-grid scheduling, power quality and the balance of supply and load demand [2,3]. So, dependable and accurate wind speed forecasting can not only provide a security basis for wind energy generation and conversion, but also reduce the costs of power system operation.

The existing wind speed forecasting approaches can be classified into three groups as physical models, statistical models and artificial intelligence models. Physical models are plain methods, which take advantage of physical information like atmospheric pressure, temperature, obstacles and roughness [4]. Thereinto, NWP (numerical weather prediction) models employ a set of mathematics equations based on physical information to forecast. Moreover, a range of statistical models have been researched to perform wind speed forecasting in the recent decades. The widely used statistical models include autoregressive models (AR), moving average models (MA), autoregressive moving average models (ARMA), autoregressive integrated moving average models (ARIMA) and seasonal autoregressive integrated moving average (SARIMA). Liu [5] proposed a novel method based on recursive ARIMA and EMD (empirical mode decomposition) to perform short term wind speed forecasting for railway strong wind warning system. Kavasseri et al. [6] developed a fraction-ARIMA to predict one-day and two-day ahead wind speed in North Dakota. On the other hand, with the rapid development of soft-computing technologies, artificial intelligence models have been proposed successfully for time series prediction. Among them, ANN (artificial neural networks) such as back propagation neural networks [7], multi-layer perceptron neural networks [8], radial basis function neural networks [9], Bayesian neural networks [10] and extreme learning machine [11] have been applied to

* Corresponding author.
E-mail address: zhouwuneng@163.com (W. Zhou).

https://doi.org/10.1016/j.enconman.2018.03.098
Received 26 November 2017; Received in revised form 21 March 2018; Accepted 31 March 2018
Available online 06 April 2018
0196-8904/

wind speed forecasting. Chang et al. [12] provided an improved neural network based approach with error feedback to predict short term wind speed and power. Noorollahi et al. [13] used ANN models to perform temporal and spatial wind speed forecasting in Iran with success. In [14], Ma et al. proposed a generalized dynamic fuzzy neural network optimized by BSO (brain storm optimization) to forecast short-term wind speed. Another popular group is SVM (support vector machine) with high generalization ability. Jiang et al. [15] presented a hybrid short term wind speed forecasting model using ν-SVM optimized by cuckoo search algorithm. Chen et al. [16] developed a state-space based SVM with unscented Kalman filter for wind speed prediction. Additionally, to improve the forecasting performance of a single model, combination or hybrid models are investigated recently to solve this problem [17,18]. For combined methods, different individual models are used to predict and their predicted results are combined to give the final prediction with the corresponding weight coefficients. Xiao et al. [18] proposed a novel combined model based on no negative constraint theory and artificial intelligence algorithm, in which the chaos particle swarm optimization algorithm was used to find the optimal weight coefficients. To simultaneously obtain high accuracy and strong stability, Wang et al. [19] developed a combined forecasting model using multi-objective bat algorithm for wind speed forecasting. In [20], Wang et al. presented a robust combined model adopting ARIMA, SVM, ELM and LSSVM (least square support vector machine) for short term probabilistic wind speed prediction, in which GPR (Gaussian process regression) is utilized to combine the results of individual predictors. The recent researches have demonstrated that combined forecasting mechanism can achieve better prediction performance than single models. However, it should be noted that the commonly accepted combination strategy of weight coefficients is a linear approach, which could not find the non-linear relationship of individual models. Besides, more advanced prediction approaches need to be introduced to enhance the forecasting performance rather than conventional machine learning algorithms like ANN and SVM.

In recent years, the utilization of deep learning in time series modeling has aroused many people's great research interest [21]. Lv et al. [22] performed traffic flow prediction with big data in a deep learning approach. Qiu et al. [23] proposed an ensemble deep learning method for electrical load forecasting. Moreover, advanced deep learning methods have also been successfully applied into wind speed forecasting field. In [24], Hu et al. provided a deep auto-encoder based model using transfer learning for short term wind speed prediction. Khodayar [25] proposed a rough deep neural network architecture with auto-encoders to perform short-term wind speed forecasting. Wang [26] developed a new deterministic and probabilistic wind speed forecasting method using deep belief network models. Furthermore, ensemble learning has been acknowledged widely that the learning performance could be promoted by combing paralleling learning models intelligently [27]. Although the existing combined models for wind speed forecasting can be regarded as one type of ensemble prediction, almost of their forecasting results are a linear combination of individual predictors. From the perspective of general ensemble learning, ensemble prediction based on non-linear learning should be more explored and researched. Thus, in this study, a novel method using nonlinear-learning ensemble of deep learning time series prediction based on LSTMs (Long Short Term Memory neural networks), SVRM (support vector regression machine) and EO (extremal optimization algorithm) named EnsemLSTM is proposed for wind speed forecasting. LSTMs, as a breakthrough variant of RNNs (recurrent neural networks), can learn the temporal and long term dependencies from time series data deeply and solve the vanishing gradient problem effectively compared with traditional RNNs [31,32]. EO is a novel promising intelligent optimization algorithm from the statistical physics field and has been applied to a lot of combinatorial and continuous optimization problems, which shows its superiority over commonly used ones like GA and PSO [37–41]. Inspired by ensemble learning, a cluster of LSTMs with diverse hidden layers and neurons are firstly introduced to explore and exploit the hidden information of wind speed time series. To overcome the shortcomings of linear representation of traditional combined models, the predictions of LSTMs are aggregated into a nonlinear-learning regression top-layer to give the final ensemble prediction in this paper rather than a linear combination. ANN is a classic artificial intelligence method, but it is unstable and its performance depends on data vastly, which make it difficult to predefine the network construction. Additionally, due to limitations of training algorithms, ANN may easily fail into local minima [28]. On the contrary, SVRM has superiority in solving complex nonlinear regression and prediction problems and has achieved extensive application and remarkable success in forecasting field [29,30]. Accordingly, the nonlinear-learning top-layer used in this paper is composed of SVRM to get rid of the weakness of ANN and the EO will be introduced to search for the optimal parameters of this top-layer. Therefore, the main differences between the proposed EnsemLSTM and traditional combined models are summarized: (a): LSTMs, a kind of deep learning method is introduced as the forecasting engine in EnsemLSTM while predictors of traditional combined models are conventional machine learning algorithms like ANN and SVM; (b): To overcome the defects of liner representation of traditional combined models, a nonlinear-learning regression top-layer is adopted in EnsemLSTM to give the final ensemble prediction; (c): The application of a novel promising intelligent optimization algorithm i.e. EO is performed to find the optimal parameters of the top-layer in EnsemLSTM.

The principal contributions of this paper are as follows: (1) A deep learning time series prediction based on LSTMs is introduced to explore and exploit the implicit information of wind speed time series for wind speed forecasting; (2) To improve the generalization capability and robustness of a single deep learning approach, nonlinear-learning ensemble of deep learning time series prediction consisting of a cluster of LSTMs with diverse hidden layers and neurons and one nonlinear-learning regression top-layer composed of SVRM optimized by the EO is developed; (3) The performance of the proposed EnsemLSTM is successfully validated on two case studies data collected from a wind farm in Inner Mongolia, China, to perform ten-minute ahead utmost short term wind speed forecasting and one-hour ahead short term wind speed forecasting. Statistical tests of experimental results have demonstrated the proposed EnsemLSTM can achieve a better forecasting performance when compared with other prediction models.

The remainder of this article is arranged as follows. In Section 2, the optimization problem formulation of nonlinear-learning ensemble of deep learning time series prediction for wind speed forecasting is proposed and the related basic learning and optimization algorithms are introduced. Section 3 presents the proposed EnsemLSTM. Section 4 describes the evaluation indices of model forecasting performance. In Section 5, two case studies are performed and the discussion and comparison of forecasting models are also given in this section. Finally, conclusion and future work of this paper are given in Section 6.

2. Problem formulation

2.1. Deep learning time series prediction

As one distinctive class of RNNs, LSTMs utilize special units named memory blocks to take the place of the traditional neurons in the hidden layers [31,32]. Moreover, there exist three gates units called input gates, output gates and forget gates in memory blocks and hence LSTMs have the ability to update and control the information flow in the block through these gates. The schema of LSTMs is displayed in Fig. 1. And the implementation of updating the state of the cell and calculating the output of LSTMs can be followed below.

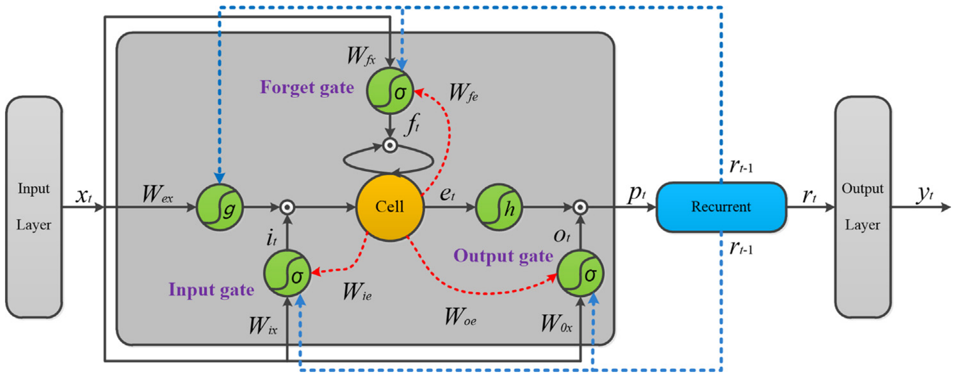

Fig. 1. The schema of LSTMs.

$$
\begin{aligned}
&\text{input gate: } i_t = \sigma(W_{ix}x_t + W_{ip}p_{t-1} + W_{ie}e_{t-1} + b_i)\\
&\text{output gate: } o_t = \sigma(W_{ox}x_t + W_{op}p_{t-1} + W_{oe}e_t + b_o)\\
&\text{forget gate: } f_t = \sigma(W_{fx}x_t + W_{fp}p_{t-1} + W_{fe}e_{t-1} + b_f)\\
&\text{temporary cell state: } \tilde{e}_t = g(W_{ex}x_t + W_{ep}p_{t-1} + b_e)\\
&\text{cell state: } e_t = i_t \odot \tilde{e}_t + f_t \odot e_{t-1}\\
&\text{output: } p_t = o_t \odot h(e_t)\\
&\text{output layer: } y_t = \varphi(W_{yp}\cdot p_t + b_y)
\end{aligned} \tag{1}
$$

where x_t is the input vector, y_t is the output vector, i_t, o_t and f_t is the output of input gate, output gate, and forget gate respectively, $\tilde{e}_t$ and e_t is the temporary and finishing state of the memory cell in the memory block, p_t is the output of the memory block. σ denotes the gate activation function (generally the logistic sigmoid function), g and h are respectively the input and output activation function (usually the tanh function), $\odot$ is the element-wise multiplication between two vectors (Hadamard product), φ is the output activation function of LSTMs, linear function in this paper for time series prediction while W_{ix}, W_{ip}, W_{ie}, W_{ox}, W_{op}, W_{oe}, W_{fx}, W_{fp}, W_{fe}, W_{ex}, W_{ep} and W_{yp} represent the corresponding weight matrices, b_i, b_o, b_f, b_e and b_y are the related bias vectors.

2.2. The nonlinear-learning ensemble of deep learning time series prediction for wind speed forecasting

To achieve a better wind speed forecasting, nonlinear-learning ensemble of deep learning time series prediction based on LSTMs, SVRM and EO is developed in this paper. In the structure of nonlinear-learning ensemble learning, predictions of a cluster of LSTMs are input into a nonlinear-learning regression top-layer to produce the final forecasting. Considering the superiority in solving complex regression problems and extensive application and remarkable success in forecasting field [29,30], SVRM is introduced as the nonlinear-learning top-layer. However, the performance of ensemble learning depends vastly on the parameters of top-layer SVRM i.e. the punishment coefficient C and kernel parameter σ. To address this problem, parameters optimization of SVRM using real-coded EO is successfully developed in this paper. Before giving the optimization problem formulation for wind speed forecasting, the basic conceptions of SVRM and EO are introduced firstly.

2.2.1. Support vector regression machine

Given a set of samples $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$, $i = 1, 2, 3, \ldots, N$, with input vector $\boldsymbol{x}_i \in R^m$ and output vector $\boldsymbol{y}_i \in R$. The task of regression problems is to find a function $f(\boldsymbol{x})$ to reveal the relationship of inputs and outputs. The motivation of SVR is to achieve a linear regression in the high-dimensional feature space obtained by mapping the original input set through a predefined function $\phi(x)$ and to minimize the structure risk R[f] [28–30]. And the above process can be expressed as follow.

$$
\begin{aligned}
&f(x) = W^{\mathrm{T}}\phi(x) + b\\
&\mathrm{R}[f] = \tfrac{1}{2}\|W\|^2 + C\sum_{i=1}^{N} L(x_i, y_i, f(x_i))
\end{aligned} \tag{2}
$$

where W, b and C is respectively the regression coefficients vector, bias term and punishment coefficient, $L(\boldsymbol{x}_i, \boldsymbol{y}_i, f(\boldsymbol{x}_i))$ denotes the ε-insensitive loss function. The regression problem can be tackled by the following constrained optimization problem:

$$
\begin{aligned}
&\min \quad \tfrac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}(\zeta_i + \zeta_i^*)\\
&\text{s.t.} \quad y_i - (W^{\mathrm{T}}\phi(x) + b) \leqslant \varepsilon + \zeta_i\\
&\qquad (W^{\mathrm{T}}\phi(x) + b) - y_i \leqslant \varepsilon + \zeta_i^*\\
&\qquad \zeta_i, \zeta_i^* \geqslant 0 \ \ i = 1,2,3,\ldots,N
\end{aligned} \tag{3}
$$

where ζ_i and ζ_i^* denote slack variables to make constraints feasible. By introducing the Lagrange multipliers, the regression function can be given as follows

$$
f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x_i, x_j) + b \tag{4}
$$

where α_i and α_i^* are the Lagrange multipliers which satisfy the conditions $\alpha_i \geqslant 0, \alpha_i^* \geqslant 0$ and $\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0$. $K(x_i, x_j)$ is the kernel function and the commonly used radial basis function (RBF) is chosen as the kernel function in this paper which is defined as

$$
K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{5}
$$

where σ represents the RBF kernel width.

2.2.2. Extremal optimization

EO [33,34], is a novel promising intelligent optimization algorithm, which stimulated by the self-organized criticality from the statistical physics field [35,36]. In the last decade, the EO has been successfully applied to a variety of benchmark and real-world engineering optimization problems [37–41]. The relevant research has shown that the EO with simpler evolutionary operations can outperform the commonly-used optimization algorithms like GA and PSO. The algorithm steps of the real-coded τ-EO with adopting PLM are displayed below.

Input: The total number of variables N, the maximum number of iterations $I_{\max}$, the control parameter τ of probability distribution $P(k)$.
Output: The best solution optimized by EO.
Step 1: Generate an initial solution S randomly, where S is a combination of variables with count $L = N$. Set $S_{\text{best}} = S$ and compute the fitness $C(S_{\text{best}}) = C(S)$ based on the predefined fitness function.
Step 2: For the current solution S,

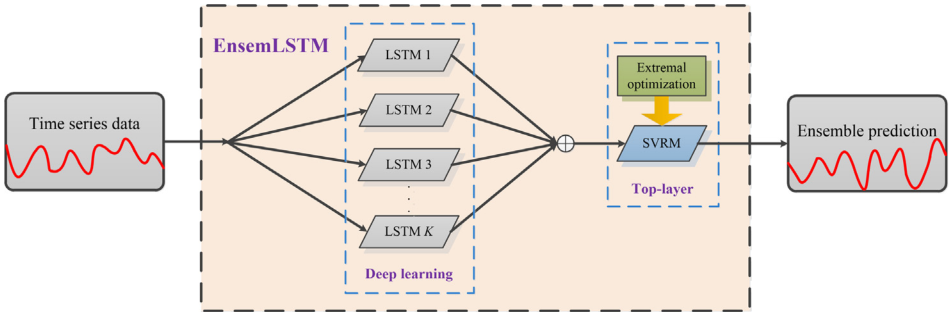

Fig. 2. The structure of this proposed EnsemLSTM with K LSTMs.

(a) Generate the solution S_i by mutating the component i ($1 \le i \le N$) (i.e. the corresponding variable is mutated by PLM) and keeping the others unchanged, then compute the fitness $C(S_i)$;

PLM can be explained by the following equations:

$$x' = x + \alpha \cdot \beta_{\max}$$
$$\alpha = \begin{cases} (2r)^{1/(q+1)} - 1, & \text{if } r \leqslant 0.5 \\ 1 - [2(1-r)]^{1/(q+1)}, & \text{otherwise} \end{cases} \quad (6)$$
$$\beta_{\max} = \max[x - l, u - x]$$

where x denotes the current value of the variable, x' is the mutated value, q is the PLM parameter, r is a random number belongs to [0,1], and l, u is respectively the lower and upper bound of the variable.

(b) Evaluate the local fitness $\lambda_i = C(S_i)\text{-}C_{\text{best}}$ for each component i and rank all the components according to λ_i, i.e., find a permutation Π_1 of the labels i such that $\lambda_{\Pi 1\ (1)} \le \lambda_{\Pi 1\ (2)} \le \cdots \le \lambda_{\Pi 1\ (N)}$;

(c) Select a rank $\Pi_1(k)$ according to a probability distribution $P(k) \propto k^{-\tau}, 1 \le k \le n$, where τ is a positive parameter, and denote the corresponding component as x_j;

(d) Mutate the value of x_j and set $S_{\text{new}} = S$ in which only x_j value is mutated;

(e) If $C(S_{\text{new}}) < C(S_{\text{best}})$, then $S_{\text{best}} = S_{\text{new}}$, $C(S_{\text{best}}) = C(S_{\text{new}})$;

(f) Accept S_{new} unconditionally;

Step 3: Repeat Step 2 until some predefined stopping criteria (i.e., the maximum number of iterations $I_{\max}$) is satisfied.

Step 4: Obtain the best solution S_{best} representing the optimal variables and the corresponding best fitness C_{best}.

2.2.3. Optimization of nonlinear-learning ensemble of deep learning time series prediction

From the perspective of optimization, how to choose the best parameters of the nonlinear-learning top-layer for ensemble of deep learning time series prediction for wind speed forecasting can be seen a typical optimization problem, which is described as follows:

$$\begin{aligned} \min \quad & f = \mathbf{Fitness}(C, \sigma) \\ \text{s.t.} \quad & l_C \leqslant C \leqslant u_C \\ & l_\sigma \leqslant \sigma \leqslant u_\sigma \end{aligned} \quad (7)$$

where C and σ represent the punishment coefficient and kernel parameter of SVRM, respectively. The l_C and u_C are the lower and upper bounds of C and the l_σ and u_σ are the lower and upper bounds of σ. In this paper, the predefined **Fitness** function is chosen as the MSE (mean square error) of 3-cross-validation on the train dataset. For the aim of simplification, the variable to be mutated is selected with the worst fitness instead of through the probability distribution $P(k)$.

3. Proposed nonlinear-learning ensemble of deep learning time series prediction method

As an excellent deep learning time series prediction method, LSTMs can explore and exploit the hidden information of dynamic time series efficiently, but the forecasting ability of LSTMs could be influenced by the number of hidden layers in LSTMs and neurons count in each hidden layer. Inspired by the great performance of ensemble learning, nonlinear-learning ensemble of deep learning time series prediction method based on LSTMs, SVRM and EO named EnsemLSTM is proposed in this article.

In the proposed EnsemLSTM, the time series i.e. the wind speed time series data is predicted separately by a cluster of LSTMs with diverse numbers of hidden layers and neurons in each hidden layer firstly, to explore and exploit the implicit information of wind speed time series. Then to overcome the defects of linear representation of traditional combined models, one nonlinear-learning regression top-layer is applied for ensemble forecasting, which is feed and trained by these forecasting results of LSTMs. Considering the extensive application and remarkable success in forecasting field, the nonlinear-learning top-layer used is composed of SVRM and the real-coded EO adopting PLM is introduced to optimize the parameters of top-layer. Lastly, the final ensemble prediction for wind speed is output by the fine-turning top-layer. The overall structure of this proposed EnsemLSTM with K LSTMs is shown in Fig. 2. It should be pointed that there is no any theoretical knowledge to predefine the network structure of LSTMs for specific data and the practical solution is to select the hyper-parameters by trial-and-error experiments [42,43]. To make a trade-off between learning performance and model complexity, in the proposed EnsemLSTM, six diverse LSTMs are adopted based on trial and error, which are respectively LSTM1 with 1 hidden layer and 50 neurons in the hidden layer, LSTM2 with 1 hidden layer and 100 neurons in the hidden layer, LSTM3 with 1 hidden layer and 150 neurons in the hidden layer, LSTM4 with 2 hidden layers and 50,50 neurons in the hidden layers, LSTM5 with 2 hidden layers and 50,100 neurons in the hidden layers and LSTM6 with 2 hidden layers and 50,150 neurons in the hidden layers. Like traditional combined models, six diverse LSTMs (LSTM1-LSTM6) are built, as single prediction models to forecast the wind speed on train and test dataset. The prediction of six LSTMs models on train dataset is the input as train-feature into SVRM to learn. The task of SVRM is to learn the nonlinear relationship of six LSTMs predictors like solving multivariate regression problems. Then, the predicted results of the six LSTMs models on test dataset are the input as test-feature into the well-trained SVRM to produce the ensemble forecasting. And the output of proposed EnsemLSTM is the ensemble forecasting of SVRM optimized with EO as the final wind speed prediction. The searching range of the punishment coefficient C and kernel parameter σ optimized by real-coded EO adopting PLM are respectively [0, 1000] and [0, 1]. The PLM parameter q in EO is set 30 and the maximum number of iterations $I_{\max}$ is 1000.

The flowchart of the EnsemLSTM is shown in Fig. 3 and the detailed implementation of this proposed EnsemLSTM is listed in the Appendix.

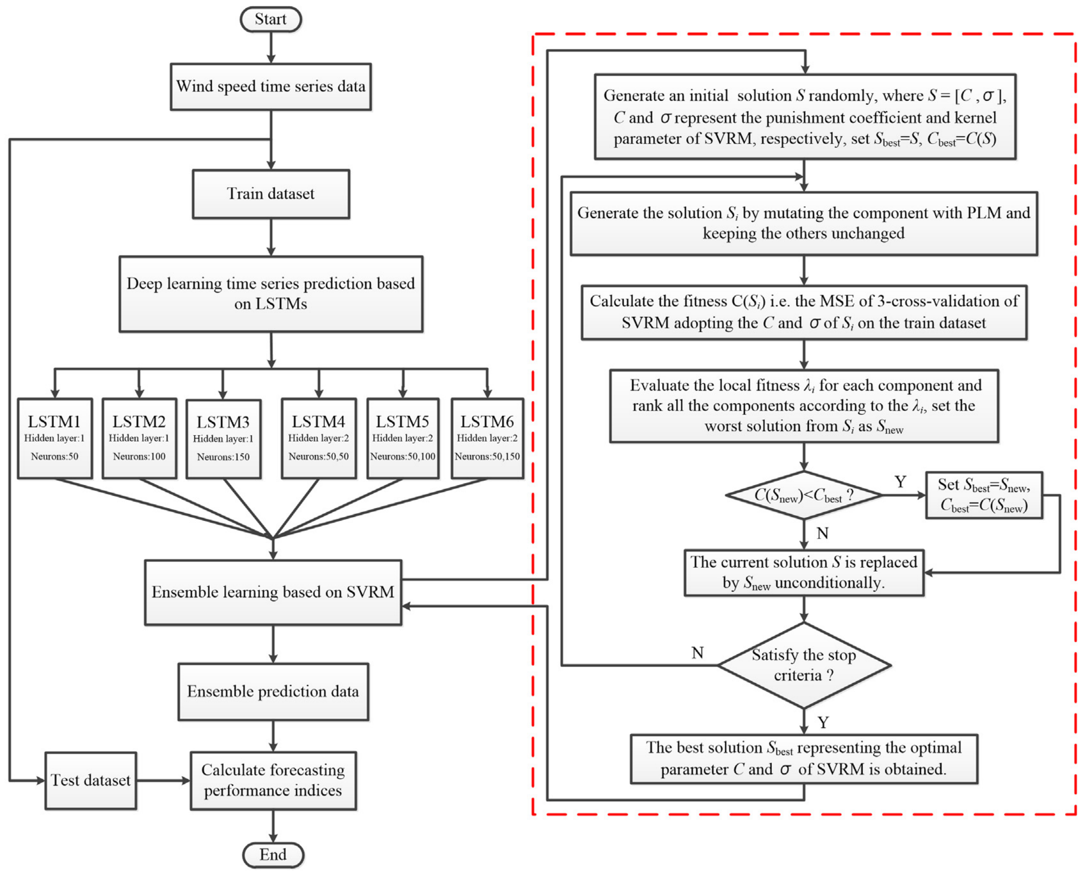

Fig. 3. The flowchart of this proposed EnsemLSTM.

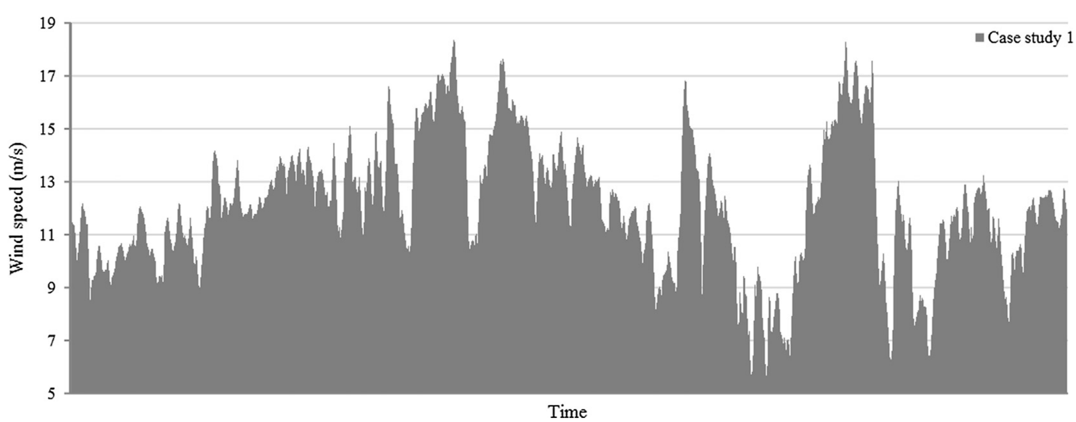

Fig. 4. The wind speed collected in case study 1.

Table 1
The statistical information of wind speed in case study 1.

Case study 1: ten-minute wind speed data from November 23, 2012 to November 28, 2012 (m/s)					
Dataset	Max	Median	Min	Mean	St.d
Entire dataset	18.500	12.100	5.500	12.166	2.460
Train dataset	18.500	12.300	5.500	12.438	2.269
Test dataset	18.300	11.600	6.300	11.531	2.761

St.d: standard deviation.

Table 2
Sets of parameters of different compared forecasting methods in case study 1.

Forecasting methods	Sets of parameters
EnsemLSTM	q = 30, I_{max} = 1000
ARIMA	(p, d, q) = (2, 0, 1)
SVR	C = 13.00, σ^2 = 0.25
ANN	1 hidden layers with 15 neurons
KNN	K = 5
GBRT	The maximum tree depth was set as 3, the number of the decision tree was set as 300

Table 3
The forecasting results of prediction models in case study 1.

Forecasting methods	MAE	RMSE	MAPE(%)	R
EnsemLSTM	**0.5746**	**0.7552**	**5.4167**	**0.9619**
ARIMA	0.6961	0.9257	6.5067	0.9420
SVR	0.5834	0.7729	5.4912	0.9599
ANN	0.6332	0.8397	6.1528	0.9545
KNN	0.6391	0.8360	6.0332	0.9530
GBRT	0.6296	0.8147	6.0733	0.9576

Best performance is highlighted in bold.

4. Evaluation of forecasting performance

Four commonly used statistical criteria are employed to evaluate the forecasting performance of wind speed prediction models. And they are defined as below.

Mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |f(i) - h(i)| \tag{8}$$

Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (f(i) - h(i))^2} \tag{9}$$

Mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \frac{|f(i) - h(i)|}{h(i)} \times 100\% \tag{10}$$

Correlation coefficient (R):

$$\mathrm{R} = \frac{\sum_{i=1}^{N} (f(i) - \bar{f})(h(i) - \bar{h})}{\sqrt{\sum_{i=1}^{N} (f(i) - \bar{f})^2} \cdot \sqrt{\sum_{i=1}^{N} (h(i) - \bar{h})^2}} \tag{11}$$

where $f(i)$ and $h(i)$ represent the predicted value and actual value at time i, respectively. And $\bar{f}$ and $\bar{h}$ denote the mean of the predicted values and the actual values, respectively. N is the total number of the data.

5. Experiments

5.1. Wind speed data description

Inner Mongolia, in China, is located in the monsoon region, and the average annual wind speed is about 3.7 m/s. And Inner Mongolia's wind energy reserves are very large about 270 million kW h, accounting for the total reserves of 1/5 and ranking first in China. In this paper, the proposed EnsemLSTM was applied to the wind speed data collected by a wind farm in Inner Mongolia, China. Two case studies with different prediction time horizons i.e. ten-minute ahead utmost short term wind speed forecasting and one-hour ahead short term wind speed forecasting were implemented to validate the effectiveness of EnsemLSTM and the results of experiments were compared with conventional forecasting methods including ARIMA, SVR, KNN, ANN and GBRT (gradient boosting regression tree). The parameters for ARIMA were determined based on the values of AIC and BIC. For SVR model, C was chosen as $y_{\max}$ - $y_{\min}$ according to [44] and the parameter σ was fixed by try-and-error experiments based on [45]. The predetermined K (i.e. the number of neighbors) of KNN model was set as 5. The ANN model was composed of one hidden layer and the number of neurons is decided by the trials. And for GBRT, the maximum tree depth was set as 3 and the number of the decision tree was set as 300. Without loss of generality, the look-back time lag was set as one in the following experiments. In other words, forecasting approaches to predict the next point $S(t + 1)$ were based on the last point data $S(t)$ i.e. the input was composed of the wind speed value of the previous time. The ARIMA and ANN models were available in the Econometrics toolbox and Neural Network toolbox in MATLAB respectively. The SVR, KNN and GBRT models were performed by using the scikit-learn machine learning package in Python 2.7 [46]. And the proposed EnsemLSTM was implemented by mixed-language programming based on MATLAB and Python 2.7 while the LSTMs algorithm was operated by using the "Keras" deep learning package [47]. All of models are run on a computer with Windows 10 operating system, Intel Core i5 CPU @ 2.30 GHz and RAM of 8.00 GB. Considering the impact of randomness, each experiment was run 30 times and then the statistical results were taken.

5.2. Case study 1: utmost short term wind speed forecasting

In this case study, the wind speed data sampled per ten minutes from November 23, 2012 to November 28, 2012 were utilized as the dataset to perform ten-minute ahead utmost short term wind speed

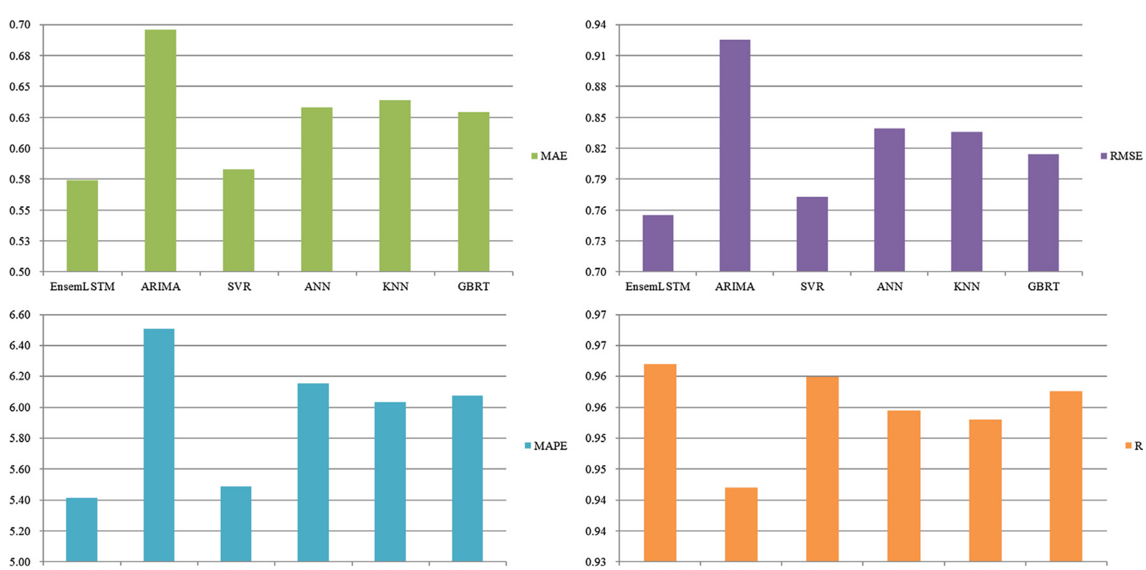

Fig. 5. The comparison of forecasting performance in case study 1.

Table 4
The statistical tests of forecasting performance comparisons in case study 1.

	Ranking tests	EnsemLSTM	ARIMA	SVR	ANN	KNN	GBRT	Statistic	p-value
MAE	Friedman	**1.000**	5.867	2.333	3.467	4.500	3.833	130.815	0.0000
	Friedman Aligned	**16.633**	162.400	52.567	92.667	116.83	101.90	120.118	0.0000
	Quade	**1.000**	5.759	2.357	3.639	4.421	3.824	42.405	0.0000
RMSE	Friedman	**1.000**	5.867	2.400	3.600	4.567	3.567	125.412	0.0000
	Friedman Aligned	**17.767**	162.267	58.267	95.833	117.87	91.000	115.456	0.0000
	Quade	**1.000**	5.755	2.404	3.768	4.548	3.525	42.968	0.0000
MAPE	Friedman	**1.000**	5.800	2.333	3.833	3.733	4.300	104.787	0.0000
	Friedman Aligned	**17.723**	158.400	53.167	102.53	101.90	109.77	112.599	0.0000
	Quade	**1.000**	5.645	2.316	4.131	3.724	4.183	39.013	0.0000
R	Friedman	**1.000**	5.900	2.200	3.633	4.733	3.533	195.558	0.0000
	Friedman Aligned	**20.233**	163.400	48.800	97.467	127.00	86.100	125.848	0.0000
	Quade	**1.000**	5.817	2.230	3.841	4.663	3.448	51.681	0.0000

Best performance is highlighted in bold.

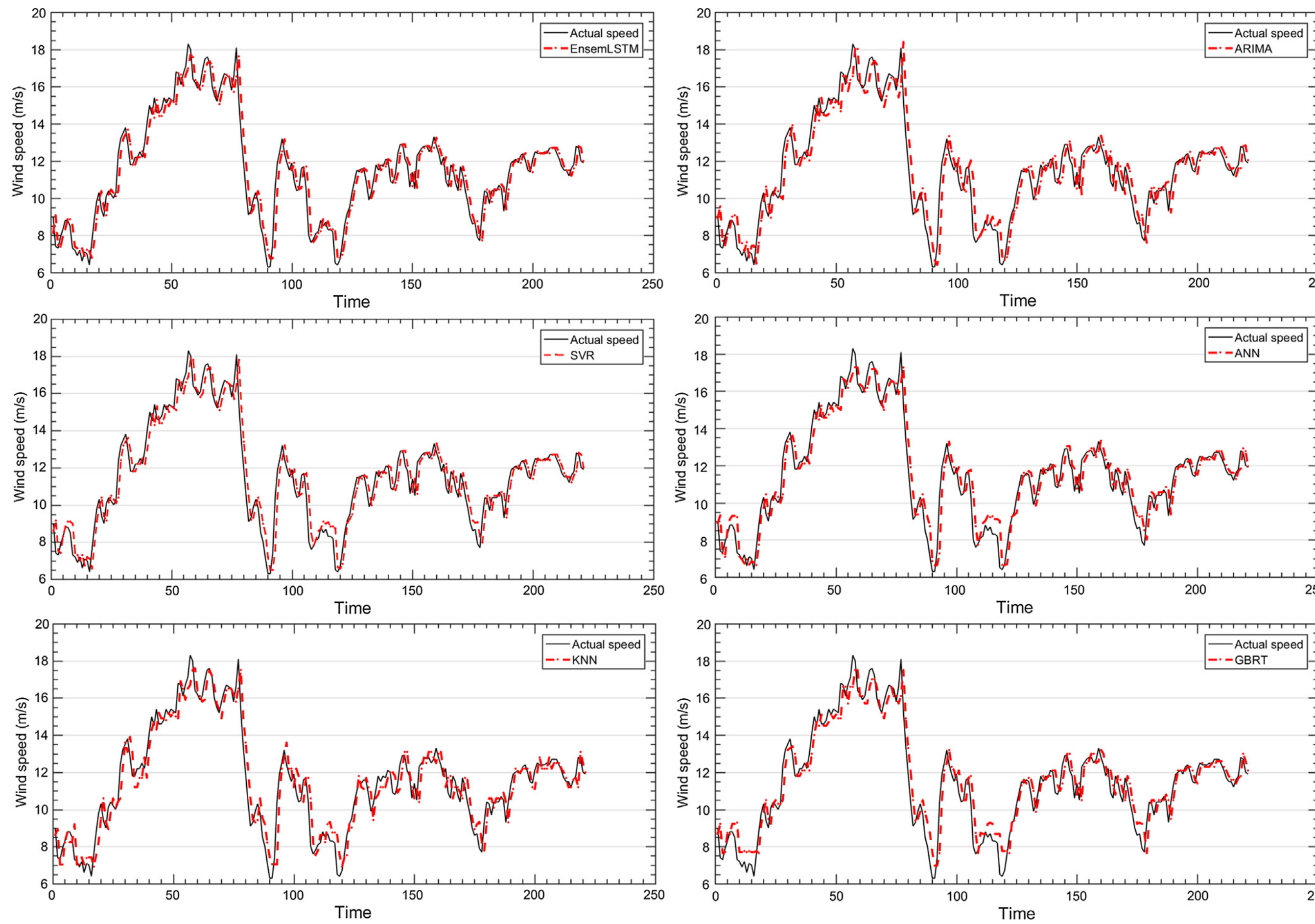

Fig. 6. The wind speed forecasting results in case study 1.

forecasting. The wind speed collected in case study 1 is displayed in Fig. 4. The total 738 obtained samples were divided into two parts (70% as the train set and 30% as the test set). The forecasting models were trained by the train dataset and validated on the test dataset. The statistical information of the above dataset is showed in Table 1. Table 2 displays the sets of parameters of different compared forecasting methods in case study 1. The forecasting results of different prediction models are listed in Table 3. Fig. 5 visualizes the comparison of forecasting performance. For the purpose of a comprehensive comparison of the proposed method and other prediction models, statistical tests including Friedman, Friedman Aligned and Quade tests were used to rank different methods with much more statistical reliability [48]. Table 4 shows the ranks, statistics and related p-values achieved by statistical tests in case study 1. And the forecast wind speeds are plotted in Figs. 6, 7 and 8 provide the bar graph and line chart of prediction residual errors for different forecast methods, respectively.

From Table 3 and Fig. 5, it can be seen that the proposed EnsemLSTM performs better than these compared widely used forecast approaches with the minimum value of MAE as 0.5746, RMSE as 0.7552 and MAPE as 5.4167% and the maximum value of R as 0.9619. And the best one of the compared prediction models is SVR with MAE as 0.5834, RMSE as 0.7729, MAPE as 5.4912% and R as 0.9599 while

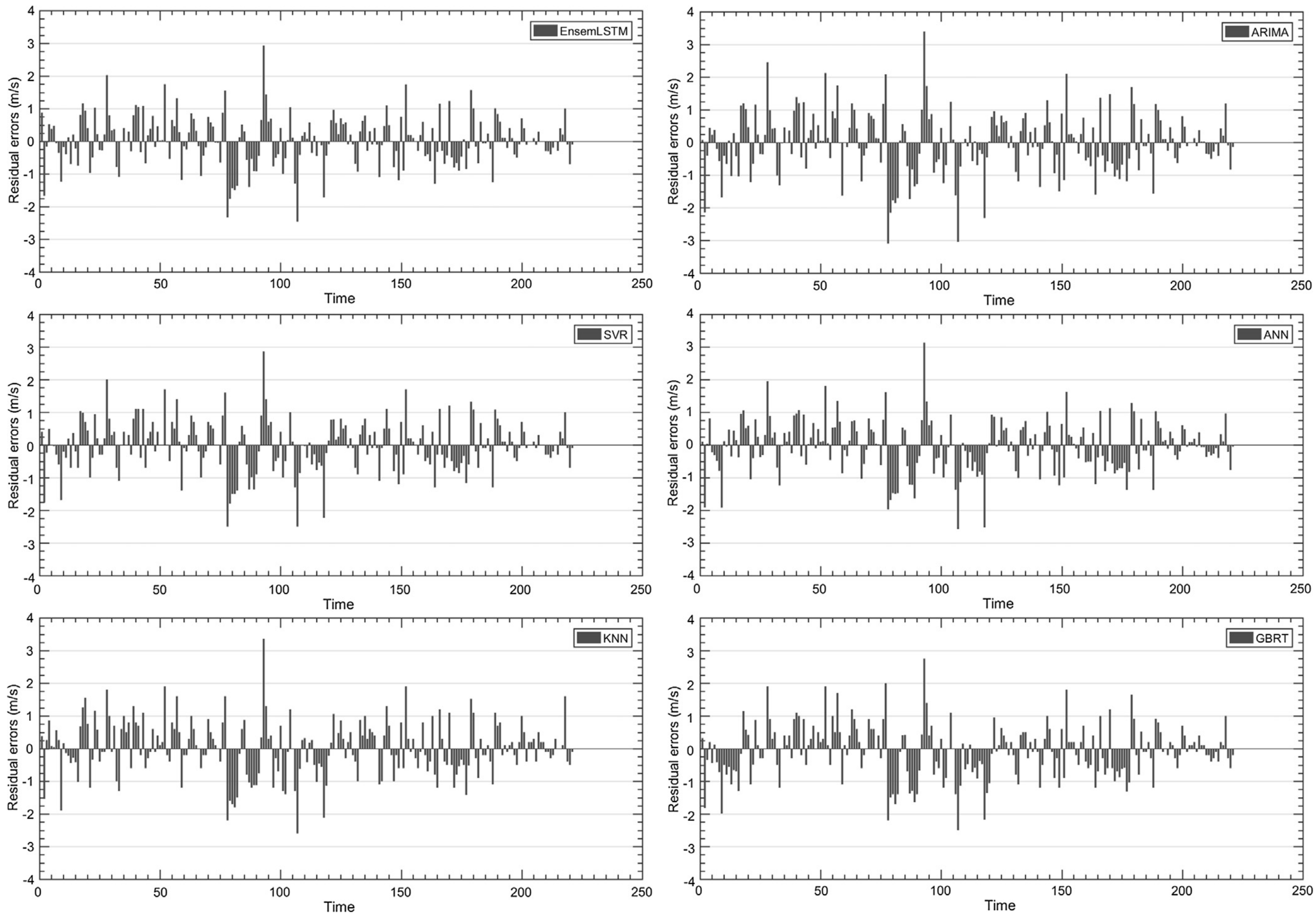

Fig. 7. The bar graphs of prediction residual errors for different forecast methods in case study 1.

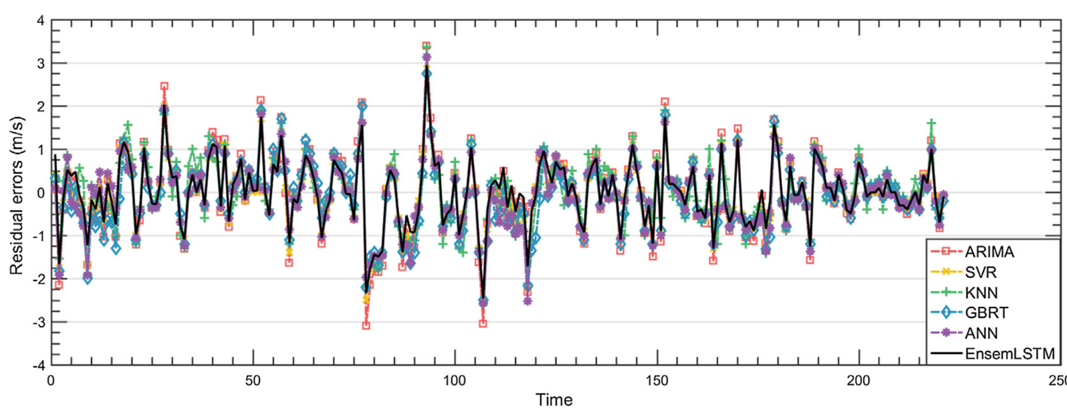

Fig. 8. The line chart of prediction residual errors for different forecast methods in case study 1.

Table 5
The impact of different ensemble learning top-layer on prediction performance in case study 1.

Different models	MAE	RMSE	MAPE(%)	R
EnsemLSTM	**0.5746**	**0.7552**	**5.4167**	**0.961938**
ANNLSTM	0.5850	0.7667	5.5145	0.961186
MeanLSTM	0.5782	0.7585	5.4822	0.961890

Best performance is highlighted in bold.

Table 6
The comparisons of forecasting results between the EnsemLSTM and single LSTMs in case study 1.

Different models	MAE	RMSE	MAPE (%)	R
EnsemLSTM	**0.5746**	**0.7552**	**5.4167**	**0.961938**
LSTM1	0.5834	0.7618	5.5520	0.961913
LSTM2	0.5818	0.7613	5.5393	0.961861
LSTM3	0.5794	0.7587	5.5017	0.961919
LSTM4	0.5778	0.7605	5.4598	0.961652
LSTM5	0.5776	0.7596	5.4494	0.961737
LSTM6	0.5802	0.7625	5.4595	0.961764

Best performance is highlighted in bold.

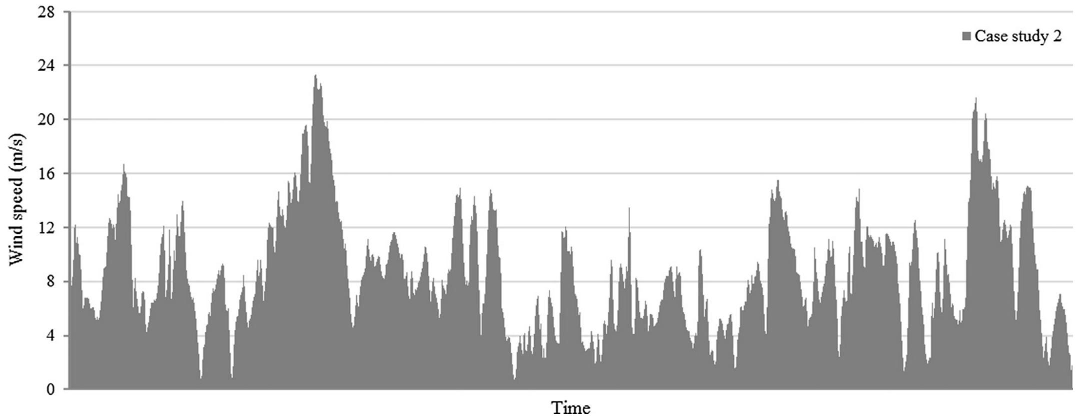

Fig. 9. The wind speed collected in case study 2.

Table 7
The statistical information of wind speed in case study2.

Case study 2: mean one-hour wind speed data collected from April 1, 2013 to April 30, 2013 (m/s)

Dataset	Max	Median	Min	Mean	St.d
Entire dataset	23.383	8.092	0.617	8.687	4.322
Train dataset	23.383	7.667	0.617	8.340	4.325
Test dataset	21.883	9.483	0.967	9.492	4.421

St.d: standard deviation.

Table 8
Sets of parameters of different compared forecasting methods in case study 2.

Forecasting methods	Sets of parameters
EnsemLSTM	$q = 30$, $I_{max} = 1000$
ARIMA	(p, d, q) = (2, 0, 1)
SVR	$C = 22.77$, $\sigma^2 = 1$
ANN	1 hidden layer with 10 neurons
KNN	$K = 5$
GBRT	The maximum tree depth was set as 3, the number of the decision tree was set as 300

Table 9
The forecasting results of prediction models in case study 2.

Forecasting methods	MAE	RMSE	MAPE	R
EnsemLSTM	**1.1410**	**1.5335**	**17.1076**	**0.9375**
ARIMA	1.3753	1.8337	20.7303	0.9098
SVR	1.1841	1.5766	17.7574	0.9338
ANN	1.1918	1.5784	18.1864	0.9340
KNN	1.2291	1.6223	17.9257	0.9297
GBRT	1.2143	1.5806	18.6117	0.9341

Best performance is highlighted in bold.

the worst one is ARIMA with MAE as 0.6961, RMSE as 0.9257, MAPE as 6.5067% and R as 0.9420. According to Table 4, the proposed EnsemLSTM also realizes the best ranks in Friedman tests, Friedman Aligned tests and Quade tests for the all of forecasting performance indices with a level of significance α = 0.0001. Moreover, from the analysis of Figs. 6, 7 and 8, the EnsemLSTM shows apparently a better curve fitting of the actual wind speed time series and smaller residual errors when contrasted with other forecasting results.

In order to verify the effectiveness of our proposed EnsemLSTM to improve the wind speed forecasting performance, the analysis of impacts of the ensemble learning top-layer on prediction performance and comparisons between the EnsemLSTM and single LSTMs are performed. Table 5 presents the impacts of different top-layers and the comparisons between ensemble and single models are shown in Table 6. In Table 5, the different models include the proposed EnsemLSTM, ANNLSTM (the nonlinear-learning top-layer is composed of an ANN model) and MeanLSTM (the ensemble learning is an average forecasting result). From Table 5, we can find that EnsemLSTM can achieve better forecasting performance, which indicates not only the SVRM's superiority to ANN but also the advantage of nonlinear-learning top-layer. Moreover, from Table 6, it can be clearly seen that the EnsemLSTM has improved the forecasting performance of single LSTMs, showing great strength of ensemble learning than single models.

Remark 1. In case study 1, ten-minute ahead utmost short term wind speed forecasting is performed and our proposed EnsemLSTM achieves a better forecasting performance than ARIMA, SVR, ANN, KNN and GBRT, which indicates the powerful learning ability of dynamic sequences (wind speed time series data). The Friedman tests, Friedman Aligned tests and Quade tests of MAE, RMSE, MAPE and R have also proved the superiority of EnsemLSTM from the statistical perspective. Moreover, the nonlinear-learning top-layer of SVRM in EnsemLSTM shows better ensemble learning performance when compared with ANNLSTM and MeanLSTM. And the comparisons among EnsemLSTM and six single prediction models LSTM1-LSTM6 also manifest the wonderful learning performance of ensemble learning.

5.3. Case study 2: short term wind speed forecasting

The one-hour ahead short term wind speed forecasting was investigated in this case study 2. And the mean one-hour wind speed data collected from April 1, 2013 to April 30, 2013 were used. The total 720 obtained data points were divided into two parts in the same way as discussed in the case study 1. The wind speed collected in case study 2 is shown in Fig. 9 and the statistical information of this dataset is shown in Table 7. Table 8 displays the sets of parameters of different compared forecasting methods in case study 2. The forecasting results and the statistical tests of forecasting performance comparisons between different prediction models are also displayed in Tables 9 and 10, respectively. Fig. 10 visualizes the comparison of forecasting results. It can be observed from Table 9 that the four forecasting performance indices in case study 2 become worse than the ones calculated in case study 1. It can be easily understood that short term wind speed forecasting is more complicated and difficult than utmost short term wind speed forecasting when the prediction time horizon became longer from ten minutes to one hour with the increase of the wind speed non-determinacy.

Similar to the results of case study 1, Table 9 and Fig. 10 show that the proposed EnsemLSTM outperforms these compared prediction models for short term wind speed forecasting with the minimum value

Table 10
The statistical tests of forecasting performance comparisons in case study 2.

	Ranking tests	EnsemLSTM	ARIMA	SVR	ANN	KNN	GBRT	Statistic	p-value
MAE	Friedman	**1.000**	6.000	2.400	2.933	4.767	3.867	238.732	0.0000
	Friedman Aligned	**16.033**	165.500	57.067	72.100	130.27	102.03	133.168	0.0000
	Quade	**1.065**	6.000	2.439	2.895	4.757	3.845	58.790	0.0000
RMSE	Friedman	**1.000**	6.000	2.567	2.900	4.867	3.667	241.907	0.0000
	Friedman Aligned	**15.500**	162.500	69.433	71.900	133.80	86.867	130.042	0.0000
	Quade	**1.000**	6.000	2.510	3.086	4.796	3.609	58.435	0.0000
MAPE	Friedman	**1.100**	5.967	2.567	3.167	3.533	4.667	124.272	0.0000
	Friedman Aligned	**17.733**	164.500	65.833	84.700	84.067	126.17	118.344	0.0000
	Quade	**1.161**	5.935	2.594	3.125	3.542	4.643	42.701	0.0000
R	Friedman	**1.033**	6.000	3.633	2.867	4.900	2.567	235.936	0.0000
	Friedman Aligned	**15.567**	165.500	86.433	71.033	134.53	69.933	130.832	0.0000
	Quade	**1.037**	6.000	3.596	3.026	4.845	2.497	58.732	0.0000

Best performance is highlighted in bold.

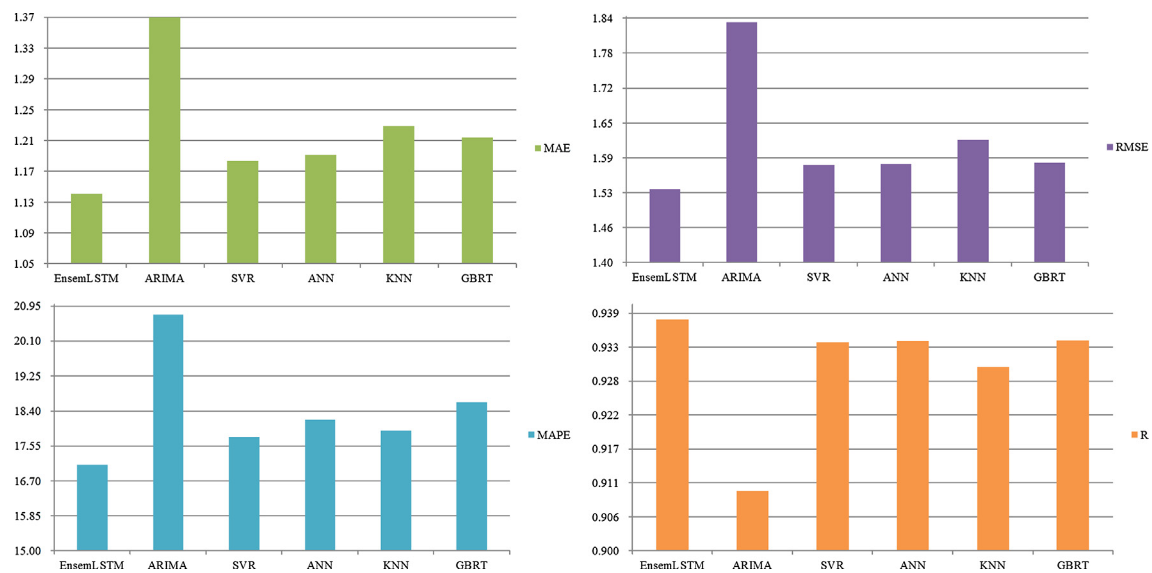

Fig. 10. The comparison of forecasting performance in case study 2.

of MAE as 1.1410, RMSE as 1.5335 and MAPE as 17.1076% and the maximum value of R as 0.9375 while the best forecasting performance indices of the compared prediction models are respectively MAE as 1.1841 for SVR, RMSE as 1.5766 for SVR, MAPE as 17.7574% for SVR and R as 0.9341 for GBRT. In addition, the EnsemLSTM also ranks first in statistic tests with a level of significance $\alpha = 0.0001$ for short term wind speed forecasting from Table 10. Fig. 11 plots the forecasting results in case study 2, Figs. 12 and 13 show the bar graph and line chart of prediction residual errors for different forecast methods respectively. Figs. 11, 12 and 13 demonstrate that the wind speed forecasted by the EnsemLSTM shows more similarities with the actual wind speed and preforms less residual errors in case study 2.

Table 11 displays the impact of different ensemble learning top-layers on forecasting performance in case study 2 and Table 12 shows the comparisons between EnsemLSTM and single LSTMs. Tables 11 and 12 have demonstrated that our proposed EnsemLSTM performs superior wind speed forecasting to ANNLSTM, MeanLSTM and single LSTMs, which is benefited from nonlinear-learning ensemble top-layer of SVRM.

Remark 2. One-hour ahead short term wind speed forecasting is performed in case study 2, which is more complicated and difficult than the ten-minute ahead utmost short term wind speed forecasting in case study 1. Although the wind speed forecasting performance become worse in case study 2, the statistical tests of experimental results have demonstrated that our proposed EnsemLSTM is still superior to ARIMA, SVR, ANN, KNN and GBRT. Furthermore, better forecasting performance of EnsemLSTM than ANNLSTM, MeanLSTM and single LSTMs is also verified in case study 2, which implies the great learning ability of the nonlinear-learning top-layer and ensemble learning. The above detailed comparisons in case study 1 and case study 2 have proven that the proposed EnsemLSTM using nonlinear-learning ensemble of deep learning can perform better wind speed forecasting than conventional prediction models.

5.4. Discussion and comparison

This section focuses on the discussion and comparison between the proposed EnsemLSTM and other conventional forecasting models. From the experimental results of utmost short term wind speed forecasting and short term wind speed forecasting, we can observe that EnsemLSTM performs the best in terms of forecasting metrics (MAE, RMSE, MAPE and R) among compared models, i.e., ARIMA, SVR, ANN, KNN and GBRT. And the related statistical tests in Tables 4 and 10 have proved the effectiveness of EnsemLSTM. Furthermore, from the structure of the proposed method, the forecasting ability of EnsemLSTM is depended upon LSTMs and the nonlinear-learning ensemble top-layer.

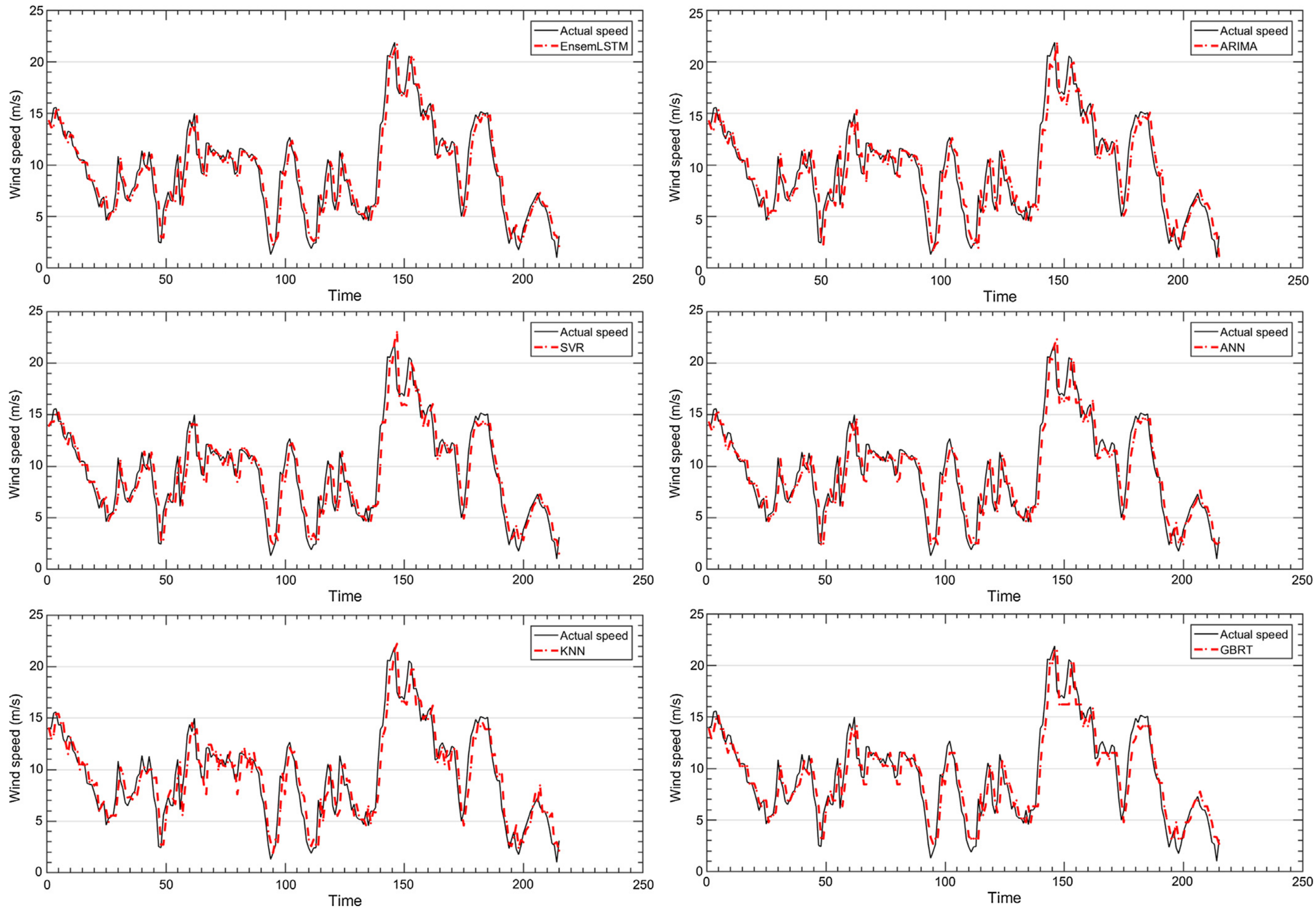

Fig. 11. The wind speed forecasting results in case study 2.

To improve the generalization capability and robustness of single LSTMs, the ensemble learning of LSTMs with diverse hidden layers and neurons are introduced in this paper. And the analysis in Tables 5–6 and 11–12 have also manifested that the ensemble learning top-layer of SVRM optimized by the EO is superior to ANNLSTM, MeanLSTM and single LSTMs. On the other hand, due to the liner characteristic of ARIMA and simple calculation of SVR, ANN, KNN and GBRT, the proposed ensemble model based on deep learning is more complex than them. In addition, the forecasting ability of proposed method can be influenced by the construction of ensemble learning and could be enhanced by adopting more proper structure for specific data. Therefore, we conclude that the EnsemLSTM proposed in this paper is effective and promising, which can be seen as an alternative reliable technique for wind speed forecasting.

6. Conclusion and future work

Wind speed forecasting is an essential issue in wind energy generation, conversion and operation, which has been attracting a lot of attentions. This paper has introduced a novel method using nonlinear-learning ensemble of deep learning time series prediction based on LSTMs, SVRM and EO for wind speed forecasting. In the proposed EnsemLSTM, a cluster of LSTMs with diverse hidden layers and neurons are employed separately to learn the information of wind speed time series firstly. Then, predictions of LSTMs are aggregated into a nonlinear-learning regression top-layer composed of SVRM and the EO is introduced to optimize the parameters of top-layer. Lastly, the final ensemble wind speed forecasting is given by the fine-turning top-layer. To verify the effectiveness of the proposed EnsemLSTM, two case studies data collected from a wind farm in Inner Mongolia, China, are adopted to perform ten-minute ahead utmost short term wind speed forecasting and one-hour ahead short term wind speed forecasting. When compared with other popular prediction models including ARIMA, SVR, ANN, KNN and GBRT, the proposed EnsemLSTM can achieve a better forecasting performance with the minimum value of MAE, RMSE and MAPE and the maximum value of R. Moreover, EnsemLSTM also realizes the best ranks in the statistical tests of experimental results including Friedman, Friedman Aligned and Quade tests. Furthermore, the analysis of impact of the ensemble learning top-layer on forecasting performance and comparisons between the EnsemLSTM and single LSTMs present that the nonlinear-learning top-layer of SVRM optimized by the EO is superior over ANNLSTM, MeanLSTM and single LSTMs. Based on the nonlinear-learning ensemble of LSTMs, SVRM and EO, the proposed EnsemLSTM achieved the satisfactory wind speed forecasting performance.

Univariate time series prediction for wind speed forecasting is investigated in this paper. In the near future, multivariate time series prediction based on deep learning algorithms using more interrelated

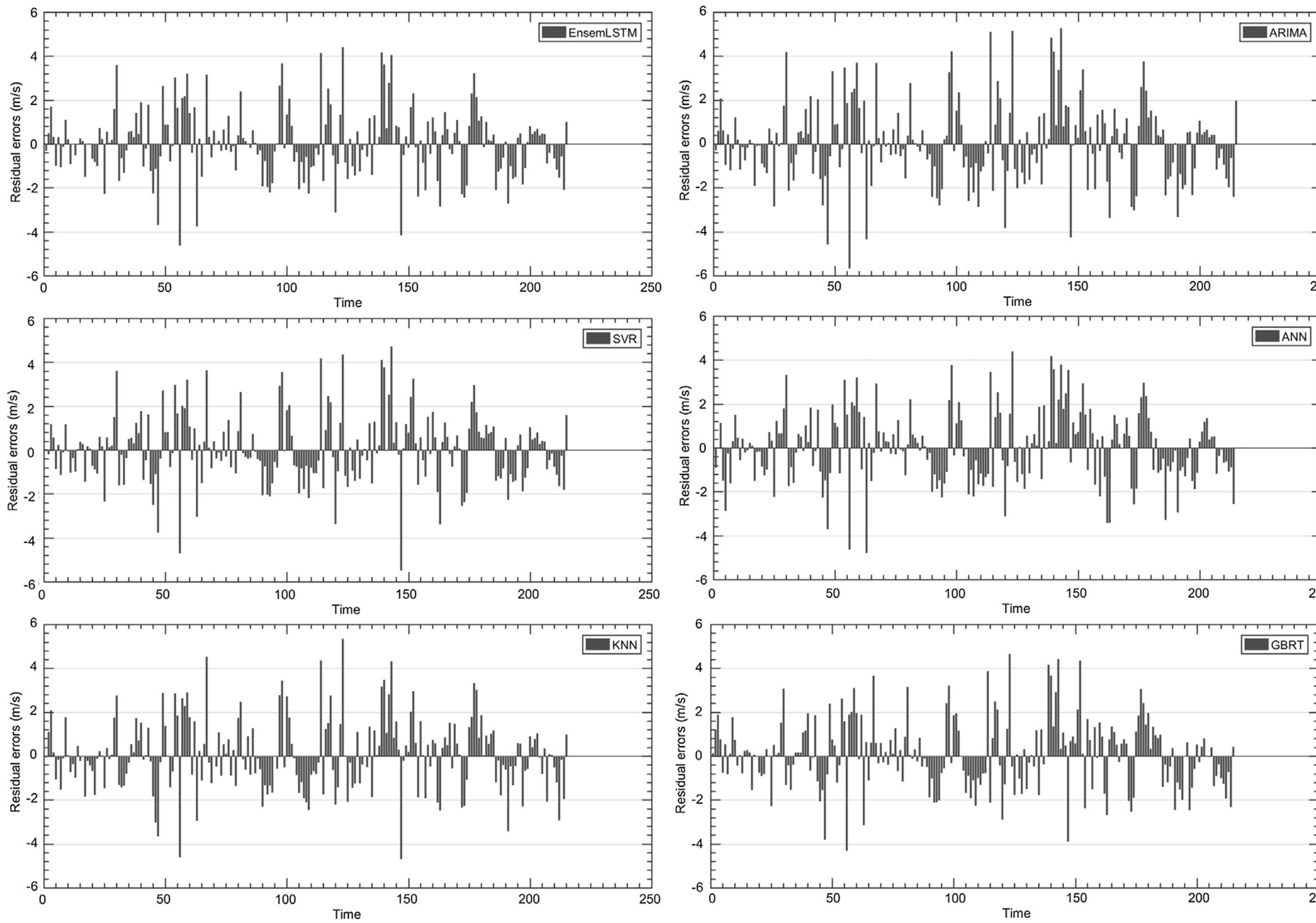

Fig. 12. The bar graphs of prediction residual errors for different forecast methods in case study 2.

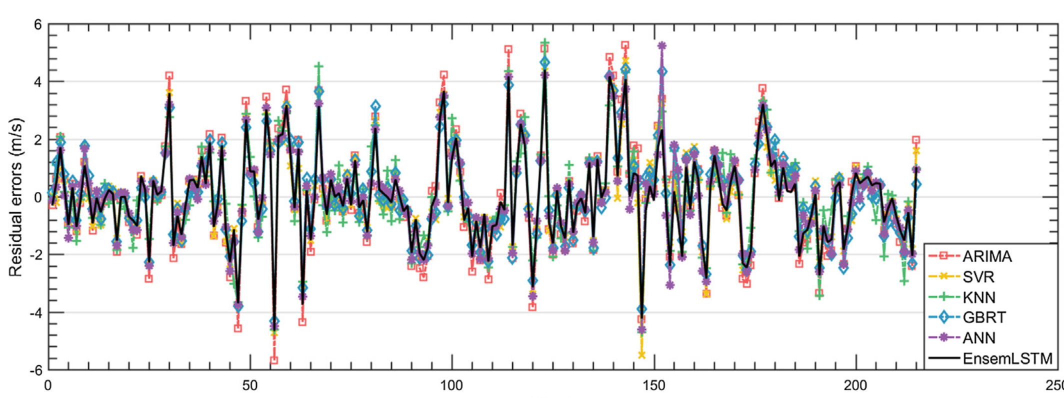

Fig. 13. The line chart of prediction residual errors for different forecast methods in case study 2.

Table 11
The impact of different ensemble learning top-layer on prediction performance in case study 2.

Different models	MAE	RMSE	MAPE (%)	R
EnsemLSTM	**1.1410**	**1.5335**	**17.1076**	**0.937498**
ANNLSTM	1.1660	1.5610	17.8027	0.936749
MeanLSTM	1.1446	1.5451	17.5856	0.937349

Best performance is highlighted in bold.

Table 12
The comparisons of forecasting results between the EnsemLSTM and single LSTMs in case study 2.

Different models	MAE	RMSE	MAPE (%)	R
EnsemLSTM	**1.1410**	**1.5335**	**17.1076**	**0.937498**
LSTM1	1.1503	1.5443	17.8296	0.937486
LSTM2	1.1437	1.5463	17.4906	0.937257
LSTM3	1.1451	1.5483	17.4886	0.937313
LSTM4	1.1461	1.5466	17.6182	0.937308
LSTM5	1.1471	1.5483	17.5832	0.937272
LSTM6	1.1494	1.5502	17.6746	0.937261

Best performance is highlighted in bold.

features like weather conditions, human factors and power system statuses will be researched for more complex wind speed prediction. On the other hand, the authors will also attempt to study more efficient ensemble learning structures to promote the model forecasting ability.

Acknowledgements

This work was supported by the Natural Science Foundation of China (Grant No. 61573095) and Zhejiang Provincial Natural Science Foundation (Nos. LY16F030011 and LZ16E050002).

Appendix A

Algorithm: EnsemLSTM

Input: The whole wind speed time series data

Output: The ensemble prediction wind speed data and forecasting performance indices

1: The whole wind speed time series data is split into two parts i.e. the train dataset and test dataset.

2: The parameters of EO: the total number of variables N is set as 2, the PLM parameter q is set as 30 and the maximum number of iterations $I_{\max}$ is set as 1000.

3: A cluster of LSTMs with diverse numbers of hidden layers and neurons in each hidden layer are trained on the train dataset and independent predictions of LSTMs for wind speed forecasting are made.

4: Predictions of LSTMs are aggregated into a nonlinear-learning regression top-layer composed of SVRM. // *Parameters of SVRM are optimized by EO*.

5: Generate an initial solution S with two components randomly, where $S = [C, \sigma]$, $C \in [0,1000]$, $\sigma \in [0,1]$, set $S_{\text{best}} = S$, compute the fitness $C(S_{\text{best}}) = C(S)$ i.e. the MSE of 3-cross-validation of SVRM adopting the C and σ of S on the train dataset. // *C and σ represent the punishment coefficient and kernel parameter of SVRM, respectively*.

6: **WHILE** *the maximum number of iterations $I_{\max}$ is not satisfied* **DO**

7: **FOR** $i \leftarrow 1$ **TO** N **DO**

8: Generate the solution S_i by mutating the component i ($1 \leqslant i \leqslant N$) with PLM and keeping the others unchanged.

9: Compute the fitness $C(S_i)$ i.e. the MSE of 3-cross-validation of SVRM adopting the C and σ of S_i on the train dataset.

Evaluate the local fitness $\lambda_i = C(S_i) - C_{\text{best}}$.
ENDFOR
Rank all the components according to λ_i.
Set the worst solution from S_i as S_{new}.
IF $C(S_{\text{new}}) < C(S_{\text{best}})$ **THEN**
$S_{\text{best}} = S_{\text{new}}$.
$C(S_{\text{best}}) = C(S_{\text{new}})$.
ENDIF
The current solution S is replaced by S_{new} unconditionally.
ENDWHILE
The best solution S_{best} representing the optimal parameter C and σ of SVRM is obtained.
The final ensemble prediction for wind speed is output by the fine-turning top-layer i.e. SVRM optimized by EO.
Forecasting performance indices are calculated based on the test dataset and ensemble prediction wind speed data.

References

[1] http://www.gwec.net/.
[2] Khare V, Nema S, Baredar P. Solar-wind hybrid renewable energy system: a review. Renew Sustain Energy Rev 2016;58:23–33.
[3] Al-falahi M, Jayasinghe S, Enshaei H. A review on recent size optimization methodologies for standalone solar and wind hybrid renewable energy system. Energy Convers Manag 2017;143:252–74.
[4] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34:243–54.
[5] Liu H, Tian H, Li Y. An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system. J Wind Eng Ind Aerodyn 2015;141:27–38.
[6] Kavasseri R, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34:1388–93.
[7] Ren C, An N, Wang J, Li L, Hu B, Shang D. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowledge-Based Syst 2014;56:226–39.
[8] Aghajani A, Kazemzadeh R, Ebrahimi A. A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm. Energy Convers Manag 2016;121:232–40.
[9] Zhang C, Wei H, Xie L, Shen Y, Zhang K. Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework. Neurocomputing 2016;205:53–63.
[10] Li G, Shi J. Applications of Bayesian methods in wind energy conversion systems. Renew Energy 2012;43:1–8.
[11] Zhang C, Zhou J, Li C, Fu W, Peng T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. Energy Convers Manag 2017;143:360–76.
[12] Chang G, Lu H, Chang Y, Lee Y. An improved neural network-based approach for short-term wind speed and power forecast. Renew Energy 2017;105:301–11.
[13] Noorollahi Y, Jokar MA, Kalhor A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. Energy Convers Manag 2016;115:17–25.
[14] Ma X, Jin Y, Dong Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. Appl Soft Comput J 2017;54:296–312.
[15] Jiang P, Wang Y, Wang J. Short-term wind speed forecasting using a hybrid model. Energy 2017;119:561–77.
[16] Chen K, Yu J. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. Appl Energy 2014;113:690–705.
[17] Xiao L, Shao W, Yu M, Ma J, Jin C. Research and application of a combined model based on multi-objective optimization for electrical load forecasting. Energy 2017;119:1057–74.
[18] Xiao L, Wang J, Dong Y, Wu J. Combined forecasting models for wind energy forecasting: a case study in China. Renew Sustain Energy Rev 2015;44:271–88.
[19] Wang J, Heng J, Xiao L, Wang C. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. Energy 2017;125:591–613.
[20] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis – combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. Energy 2015;93:41–56.
[21] Längkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognit Lett 2014;42:11–24.
[22] Lv Y, Duan Y, Kang W, Li Z, Wang F. Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 2014;16:865–73.
[23] Qiu X, Ren Y, Suganthan PN, Amaratunga G. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Appl Soft Comput 2017;54:246–55.
[24] Hu Q, Zhang R, Zhou Y. Transfer learning for short-term wind speed prediction with deep neural networks. Renew Energy 2016;85:83–95.
[25] Khodayar M, Kaynak O, Khodayar ME. Rough deep neural architecture for short-

term wind speed forecasting. IEEE Trans Ind Inform 2017;13:2770–9.
[26] Wang H, Wang G, Li G, Peng J, Liu Y. Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl Energy 2016;182:80–93.
[27] Ren Y, Zhang L, Suganthan PN. Ensemble classification and regression: Recent developments, applications and future directions. IEEE Comput Intell Mag 2016;11:41–53.
[28] Yang W, Wang J, Wang R. Research and application of a novel hybrid model based on data selection and artificial intelligence algorithm for short term load forecasting. Entropy 2017;19(2):52.
[29] Zhang X, Wang J, Zhang K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electr Power Syst Res 2017;146:270–85.
[30] Xu Y, Yang W, Wang J. Air quality early-warning system for cities in China. Atmos Environ 2017;148:239–57.
[31] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1–32.
[32] Sainath TN, Vinyals O, Senior A, Sak H. Convolutional, long short-term memory, fully connected deep neural networks. IEEE Int Conf Acoust Speech Signal Process 2015;2015:4580–4.
[33] Boettcher S, Percus A. Nature's way of optimizing. Artif Intell 2000;119:275–86.
[34] Boettcher S, Percus AG. Optimization with extremal dynamics. Phys Rev Lett 2001;86:5211–4.
[35] Bak P, Sneppen K. Punctuated equilibrium and criticality in a simple model of evolution. Phys Rev Lett 1993;71:4083–6.
[36] Lu Y, Chen M, Chen Y. Studies on extremal optimization and its applications in solving realworld optimization problems. In: Foci; 2007. p. 162–8.
[37] Lu Y, Chen Y, Chen M, Chen P, Zeng G. Extremal optimization: fundamentals, algorithms, and applications. CRC Press & Chemical Industry Press; 2016.
[38] Zeng G, Chen J, Li L, Chen M, Wu L, Dai Y, et al. An improved multi-objective population-based extremal optimization algorithm with polynomial mutation. Inf Sci 2016;330:49–73.
[39] Zeng G, Chen J, Dai Y, Li L, Zheng C, Chen M. Design of fractional order PID controller for automatic regulator voltage system based on multi-objective extremal optimization. Neurocomputing 2015;160:173–84.
[40] Zeng G, Chen J, Chen M, Dai Y, Li L, Lu K, et al. Design of multivariable PID controllers using real-coded population-based extremal optimization. Neurocomputing 2015;151:1343–53.
[41] Lu K, Zhou W, Zeng G, Du W. Design of PID controller based on a self-adaptive state-space predictive functional control using extremal optimization method. J Franklin Inst 2018;355:2197–220.
[42] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.
[43] Wu Y, Yuan M, Dong S, Lin L, Liu Y. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing 2018;275:167–79.
[44] Shi J, Guo J, Zheng S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renew Sustain Energy Rev 2012;16:3471–80.
[45] Zhou J, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers Manag 2011;52:1990–8.
[46] http://scikit-learn.org/.
[47] https://keras.io/.
[48] Derrac J, García S, Molina D, Herrera F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evol Comput 2011;1:3–18.