

Research Article

Integrated Machine Learning and Enhanced Statistical Approach-Based Wind Power Forecasting in Australian Tasmania Wind Farm

Fang Yao ^{1,2}, Wei Liu,¹ Xingyong Zhao,¹ and Li Song¹

¹School of Electric Power and Architecture, Shanxi University, Taiyuan 030013, China

²School of Electrical and Electronics and Computer Engineering, University of Western Australia, Perth, WA 6009, Australia

Correspondence should be addressed to Fang Yao; y98122@hotmail.com

Received 19 July 2020; Revised 23 August 2020; Accepted 7 September 2020; Published 16 September 2020

Academic Editor: Qiang Chen

Copyright © 2020 Fang Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper develops an integrated machine learning and enhanced statistical approach for wind power interval forecasting. A time-series wind power forecasting model is formulated as the theoretical basis of our method. The proposed model takes into account two important characteristics of wind speed: the nonlinearity and the time-changing distribution. Based on the proposed model, six machine learning regression algorithms are employed to forecast the prediction interval of the wind power output. The six methods are tested using real wind speed data collected at a wind station in Australia. For wind speed forecasting, the long short-term memory (LSTM) network algorithm outperforms other five algorithms. In terms of the prediction interval, the five nonlinear algorithms show superior performances. The case studies demonstrate that combined with an appropriate nonlinear machine learning regression algorithm, the proposed methodology is effective in wind power interval forecasting.

1. Introduction

Wind power is rapidly expanding its market share around the world. However, the intermittency and uncertainty of wind make it a challenge to integrate wind power into the power system. The wind power forecasting system can greatly help the integration process since system operators rely on accurate wind power forecasts to design operational plans and assess system security [1, 2]. Servo mechanism is the foundation of wind turbines, and precise wind power forecasting can improve the accuracy of parameter estimation and control of wind turbine servo systems [3–5]. Predictions of wind power outputs are traditionally provided in the form of point forecasts. The advantage of point forecasts is that they can be easily understood. The single value is expected to tell everything about future power generation. Nowadays, the majority of the research efforts on wind power forecasting is still focused on point forecasting. The reviews of the state of the art in wind power forecasting can be found in [6, 7]. A book on physical

approaches to short-term wind power forecasting also partly discusses the state of the art in wind power forecasting [8].

However, even if both meteorological and power conversion processes are well understood and modelled, there will always be inherent and inevitable uncertainty in wind power forecasts. The uncertainty comes from the incomplete knowledge of the physical processes that influence future events [9]. The uncertainty of wind power forecasts mainly depends on the predictability of the current meteorological status and the level of the predicted wind speed [10]. To assist with the management of the forecasting uncertainty, extensive research studies have been conducted to develop wind power forecasting methods. Different regression methods have been introduced in [10–12]. These approaches use probabilistic forecasts generated by different quantile regression methods to provide the complete information of future wind production. A multiscale reliable wind power forecasting (WPF) method was developed by Yan et al. in [13]. This method provides the expected future value and the associated uncertainty by a multi-to-multi mapping network

and stacked denoising autoencode. Wang et al. developed a short-term wind prediction method with the convolutional neural network (CNN) based on information of neighbouring wind farms [14]. One popular approach is to use ensemble-based probabilistic forecasting methodologies, which enable better wind power management and trading purposes [15, 16]. In [17, 18], statistical analysis has been conducted to study the distribution of wind power forecasting errors. Because wind power is stochastic in nature, errors will always exist in wind power forecasts. Therefore, besides predicting the expected value of the future wind power, it is also important to estimate its forecasting errors.

A key weakness of the above studies lies in that they fail to establish proper statistical models for interval forecasting of wind power and also fail to take into account the time-changing effect of the error distribution. Generally speaking, a prediction interval is a stochastic interval, which contains the true value of wind power with a preassigned probability. Because the prediction interval can quantify the uncertainty of the forecasted wind power, it can be employed to evaluate the risks of the decisions made by market participants. Existing methods discussed above cannot effectively handle wind power interval forecasting since they mainly focus on predicting the expected point value of wind power.

There are two main challenges for providing accurate interval forecasting of wind power: (i) the expected value of wind power should be accurately predicted. This is difficult since wind power is a nonlinear time series and is therefore highly volatile. The nonlinear research system has high complexity in the fundamental research, and lots of nonlinear control problems unceasingly emerge in real fact [19–23]; (ii) the probability distribution of forecasting errors should also be accurately estimated. This is even more difficult since the error distribution can be time-changing. In this paper, a novel approach is proposed to forecast the prediction interval of wind power. A statistical model is first formulated to properly model the time series of wind speed. Based on the proposed model, a number of different machine learning algorithms are introduced to predict the expected value of wind speed and the parameters of forecasting error distribution. Prediction intervals of wind speed are then constructed based on the predicted wind speed value and error distribution. The wind speed prediction interval is finally transformed into the wind power prediction interval with the wind turbine power curve. Comprehensive studies are performed to compare the performances of six machine learning algorithms in wind power interval forecasting.

The main contributions of this paper are as follows:

- (1) A comprehensive statistical model is introduced, which forms the theoretical basis for wind power interval forecasting.
- (2) Different machine learning regression methods are incorporated into the proposed model. The comparison of different regression algorithms in wind power forecasting is presented.
- (3) The proposed integrated statistical machine learning approach can highlight the essential information of the available data.

The rest of the paper is organized as follows: in Section 2, a statistical model for the wind speed time series is formulated. We also introduce the Lagrange multiplier (LM) test to verify that the forecasting errors of wind power have a time-changing distribution. In Section 3, the basic concept of machine learning and six machine learning algorithms for wind power forecasting are introduced. Afterwards, comprehensive case studies are performed in Section 4. Section 5 finally concludes the paper.

2. The Statistical Model of the Wind Speed Time Series

To forecast the power output of a wind turbine, a widely used approach is to predict the wind speed first and then transform the predicted wind speed into wind power with the power curve. Therefore, in this section, a statistical model of wind speed is first formulated. We will also briefly explain how to integrate the proposed model with nonlinear regression techniques to forecast the prediction intervals of wind speed. The wind speed time series can usually be assumed to be generated by the following stochastic process:

$$Y_t = f\left(Y_{t-1}, Y_{t-2}, \dots, \vec{X}_t\right) + \varepsilon_t, \quad (1)$$

where Y_t denotes the random wind speed and y_t is the observed value of Y_t at time t . $\vec{X}_t \in R^m$ is an m -dimensional explanatory vector. Each element $X_{t,i}$ of \vec{X}_t represents an explanatory variable which can influence Y_t , for example, the temperature and humidity. The current value of Y_t can be determined by its lagged values Y_{t-1}, Y_{t-2}, \dots and the explanatory vector \vec{X}_t . Note that the mapping $f(\bullet)$ from $Y_{t-1}, Y_{t-2}, \dots, \vec{X}_t$ to Y_t can be any linear or nonlinear function. Most existing methods essentially forecast wind speed by estimating mapping $f(\bullet)$; the forecasted value $\hat{f}(\bullet)$ of $f(\bullet)$ can be called the point forecast of wind speed. According to (1), the wind speed Y_t contains two components: $f(\bullet)$ is a deterministic component, and ε_t is a random component, which is also known as noise. Statistical and engineering models are an approximation to reality, not reality, so they always have some degree of errors. Nowadays, there are lots of research studies about error tracking and control [24–29]. Precise prediction and reducing the error are the prerequisite for all further control works. Detailed statistical studies [30] show that ε_t can be assumed to follow a normal distribution. We therefore have

$$\varepsilon_t \sim N(\mu, \sigma^2). \quad (2)$$

Because $f(\bullet)$ is a deterministic function, we should be able to approximate it with arbitrary accuracy by employing a powerful nonlinear machine learning technique (e.g., neural network). Most existing wind speed forecasting methods mainly focus on estimating $f(\bullet)$ and selecting its estimated value as the predicted wind speed. On the contrary, because of the uncertainty introduced by noise ε_t , errors will always exist in wind speed forecasts. Therefore, estimating μ and σ^2 is essential for estimating the

uncertainty of Y_t . In models (1) and (2), parameters μ and σ^2 are assumed to be constant. In practice, the model parameters can usually be time-changing. We therefore introduce the following time-changing distribution model of wind speed:

$$\begin{aligned} Y_t &= f\left(Y_{t-1}, Y_{t-2}, \dots, \vec{X}_t\right) + \varepsilon_t, \\ \varepsilon_t &= \mu_t + \sigma_t \cdot \nu_t, \\ \nu_t &\sim N(0, 1), \\ \vec{X}_t' &= (X_{t,1}, X_{t,2}, \dots, X_{t,m}), \\ u_t &= g\left(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \vec{X}_t\right), \\ \sigma_t &= h\left(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \vec{X}_t\right). \end{aligned} \quad (3)$$

Similar to $f(\bullet)$, mappings $g(\bullet)$ and $h(\bullet)$ can also be either linear or nonlinear. According to model (3), the uncertainty of wind speed is time-changing. The mean and variance of noise ε_t are determined by the previous noises and the explanatory vector. Note that model (3) is a generalization to the traditional ARCH (AutoRegressive Conditional Heteroskedasticity) model; since by setting $u_t \equiv 0$ and assuming $f(\bullet)$ and $h(\bullet)$ are linear functions, model (3) will be identical to the ARCH model. To more strictly justify our model, the Lagrange multiplier (LM) test can be employed to verify that the wind speed has a time-changing distribution. In the case study, we will test whether the actual wind speed data of Australia have a time-changing distribution by performing the LM test.

Based on statistical model (3) of wind speed, we can construct the prediction interval, which contains the true value of wind speed with any preassigned probability. The definition of the prediction interval can be given as follows.

Definition 1. Given a time series $\{Y_t\}$ which is generated with model (3), an α -level prediction interval (PI) of Y_t is a stochastic interval $[L_t, U_t]$ calculated from $\{Y_t\}$ such that $P(Y_t \in [L_t, U_t]) = 1 - \alpha$.

Because noise ε_t is usually assumed to be normally distributed, the α -level prediction interval can therefore be calculated as

$$L_t = f_t(\bullet) + \mu_t - z_{(1-\alpha)/2} \times \sigma_t, \quad (4)$$

$$U_t = f_t(\bullet) + \mu_t + z_{(1-\alpha)/2} \times \sigma_t, \quad (5)$$

where $f_t(\bullet)$ represents the value of the deterministic component $f(\bullet)$ at time t , α is the confidence level, and $z_{(1-\alpha)/2}$ is the critical value of the standard normal distribution. Based on (4) and (5), to calculate the prediction interval, we should first obtain three quantities: the wind speed forecast $f_t(\bullet)$, the mean μ , and the variance σ^2 of the noise. In practice, traditional time-series models, such as ARIMA and GARCH, usually perform poorly on short-term wind speed forecasting since they are linear models and therefore cannot handle the complex nonlinear patterns of wind speed data. To give accurate wind speed

forecasts, the three mappings $f(\bullet)$, $g(\bullet)$, and $h(\bullet)$ in model (3) should be accurately estimated with nonlinear techniques. In this paper, we introduce six different machine learning methods to estimate $f(\bullet)$, $g(\bullet)$, and $h(\bullet)$. To apply machine learning methods to estimate $g(\bullet)$ and $h(\bullet)$, an unsolved problem is how to obtain the estimates of mean μ_t and variance σ_t^2 of the noise. In this paper, the moving window method is employed. Given the noise series $\{\varepsilon_t\}$, the estimates of μ_t and σ_t^2 can be calculated as

$$\hat{\mu}_t = \frac{1}{2n+1} \sum_{s=t-n}^{t+n} \varepsilon_s, \quad (6)$$

$$\hat{\sigma}_t^2 = \frac{1}{2n} \sum_{s=t-n}^{t+n} (\varepsilon_s - \hat{\mu}_s)^2. \quad (7)$$

By combining a machine learning method with proposed model (3), the main procedure of wind power interval forecasting is given as follows:

- (1) Given the historical wind speed data $\{Y_t\}$ and the explanatory vector data $\{\vec{X}_t\}$ for time period $[0, T]$, employ a machine learning technique to estimate function $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, \vec{X}_t)$. Denote the estimate of $f(\bullet)$ as $\hat{f}(\bullet)$.
- (2) Calculate the forecasting errors $e_t = Y_t - \hat{f}(Y_{t-1}, Y_{t-2}, \dots, \vec{X}_t)$ for period $[0, T]$. Note that e_t can be considered as the estimate of noise ε_t .
- (3) Based on error series $\{e_t\}$, calculate the estimates of μ_t and σ_t^2 with equations (6) and (7).
- (4) Based on error series $\{e_t\}$ and mean and variance estimate series $\{\hat{\mu}_t\}$ and $\{\hat{\sigma}_t^2\}$, employ a machine learning technique to estimate functions $\hat{\mu}_t = \hat{g}(e_{t-1}, e_{t-2}, \vec{X}_t)$ and $\hat{\sigma}_t = \hat{h}(e_{t-1}, e_{t-2}, \vec{X}_t)$, and use them as the estimates of $g(\bullet)$ and $h(\bullet)$.
- (5) To forecast the wind speed at t , first employ $\hat{f}(\bullet)$, $\hat{g}(\bullet)$, and $\hat{h}(\bullet)$ to calculate $\hat{f}_t(\bullet)$, $\hat{\mu}_t$, and $\hat{\sigma}_t^2$; then, calculate the wind speed prediction interval with equations (4) and (5).
- (6) Transform the wind speed prediction interval into the wind power prediction interval with the wind turbine power curve, which will be discussed in the following sections.

3. Machine Learning Methods for Wind Power Interval Forecasting

In this section, we first provide a brief introduction to machine learning, which is an important research area in forecasting. Six machine learning algorithms used in this paper are then presented. The power curve for converting wind speed into wind power is introduced. We finally discuss how to evaluate the performance of wind power interval forecasting methods.

3.1. Introduction to Machine Learning. Machine learning is science that studies how to use the computer to simulate or realize human learning activities. It is one of the most intelligent and leading-edge research fields in artificial intelligence. Machine learning techniques are essential to the renewable energy integration such as PV and wind power [31, 32].

Machine learning can be divided into supervised learning and unsupervised learning [33–35]. As can be seen from Figure 1, supervised learning can be classified into classification and regression, and unsupervised learning can be classified into clustering and correlation.

Regression [36] is a process to estimate a functional mapping between a data vector and a target variable. Regression aims at determining a continuous target variable, which is usually named as the dependent variable, while the data item itself is usually called independent variables, explanatory variables, or predictors. For example, in wind speed forecasting, the predictors can be historical wind speed, temperature, and humidity, while the independent variable is the future wind speed. Regression usually estimates the mapping based on a training dataset in which the independent variables of all data items have been given. Regression is therefore a supervised learning problem in the sense that the estimation of the mapping is supervised by the training data. Regression is also an important research area of statistics. The most important statistical method is linear regression, which assumes that the independent variable is determined by a linear function of predictors. In recent years, the machine learning society has proposed many other regression methods, such as deep learning. In this paper, we will introduce six different machine learning regression techniques and integrate them with the proposed statistical model to perform wind power interval forecasting.

3.2. Machine Learning Regression Algorithms Employed in This Paper

3.2.1. Linear Regression. Linear regression is a traditional and widely used statistical technique for regression. It is selected as the baseline technique in this paper and will be compared with five nonlinear techniques. Linear regression models the relationship between the dependent variable y_i and the vector of predictors x_i . Linear regression assumes that the independent variable y is linearly dependent on the predictors x plus a noise term ε_i . The model can be written as

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i' \beta^T + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (8)$$

where $x_i' \beta^T$ is the inner product between vectors x_i and β . And these n equations can be written in the vector form as

$$y = \beta^T X + \varepsilon, \quad (9)$$

where

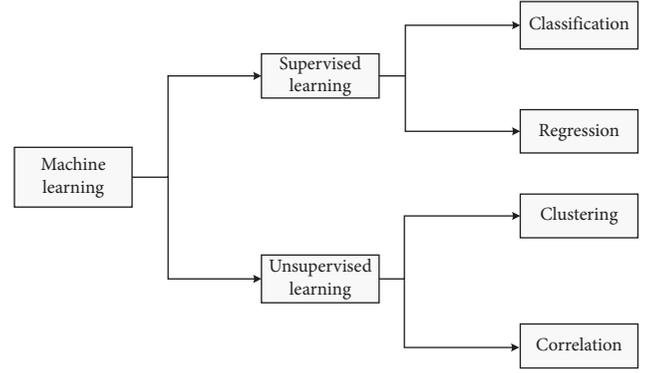


FIGURE 1: Composition structure of machine learning.

$$\begin{aligned} y &= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \\ X &= \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}, \\ \beta^T &= \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \\ \varepsilon &= \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \end{aligned} \quad (10)$$

ε is usually assumed to follow a normal distribution with a zero mean and variance σ^2 . We therefore have

$$\varepsilon \sim N(0, \sigma^2), \quad (11)$$

and β is a p -dimensional parameter vector, which specifies how much each component of X contributes to the output y [37].

3.2.2. Multilayer Perceptron Network. A multilayer perceptron (MLP) network is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate outputs. Based on the standard linear perceptron, MLP uses three or more layer nodes with nonlinear activation functions. An MLP network consists of a set of source nodes as the input layer, one or more hidden layers of computation nodes, and an output layer of nodes.

Figure 2 shows the signal flow process of a feedforward neural network. A MLP network has two stages: a forward pass and a backward pass. The forward pass includes presenting a sample input to the network and letting activations flow until they reach the output layer [38, 39].

3.2.3. Long Short-Term Memory (LSTM) Network-Based Deep Learning Method. The concept of deep learning was first proposed by Hinton et al. in 2006. Deep learning is a

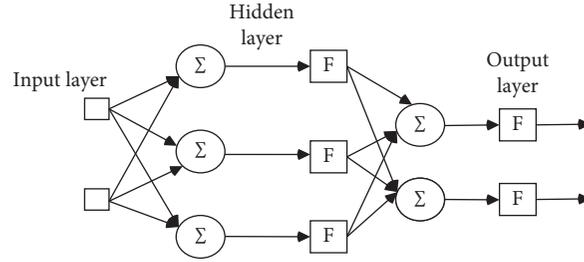


FIGURE 2: Diagram of a multilayer perceptron network.

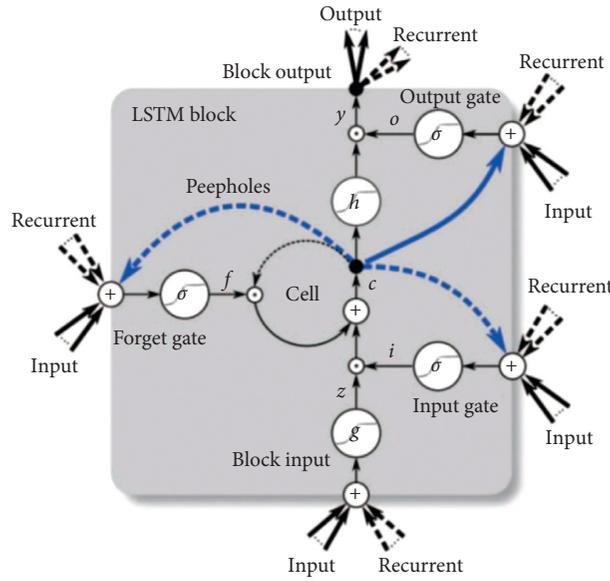


FIGURE 3: LSTM storage unit with door.

branch of machine learning. It is essentially a special artificial neural network. Deep learning utilizes a multilayer network structure and applies appropriate nonlinear transformation functions to each hidden node to achieve the purpose of high-level abstraction of the data. The traditional feedforward artificial neural network usually contains only one hidden layer, but there are many hidden layer structures in deep learning. Therefore, deep learning adopts a training mechanism completely different from the traditional artificial neural network in order to solve the problem of the deep neural network in training [40].

LSTM is a time-cycle neural network, which can effectively solve the gradient explosion and gradient disappearance problems compared with the traditional cycle neural network. LSTM is composed of a set of cyclic subnets called memory blocks. Each memory block is composed of the input gate, forgetting gate, and output gate. Figure 3 shows the LSTM structure [41].

In general, the LSTM recursive neural network is composed of the following components: input gate i_t has the corresponding weight matrix w_{xi} , w_{hi} , w_{ci} , and b_i ; forgetting gate f_t has the corresponding weight matrix w_{xf} , w_{hf} , w_{cf} , and b_f ; and output door o_t , with the corresponding weight matrix w_{xo} , w_{ho} , w_{co} , and b_o . The function of the input gate is to

record the new information to the cell state selectively. The function of the forget gate is to selectively forget the status information in the cell; the function of the output gate is to export certain information from the cell. The detailed workflow of LSTM is shown in the following [42]:

$$\begin{aligned}
 i_t &= \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i), \\
 f_t &= \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f), \\
 o_t &= \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_{t-1} + b_o), \\
 c_t &= f_t c_{t-1} + i_t \tanh(w_{xc}x_t + w_{hc}h_{t-1} + w_{cc}c_{t-1} + b_c),
 \end{aligned} \tag{12}$$

where σ is the logistic sigmoid function with the output in $[0, 1]$ and \tanh represents the hyperbolic tangent function with the output in $[-1, 1]$.

3.2.4. Lazy IBK. Lazy IBK is one of the widely used lazy learning methods. Lazy learning methods defer the decision of how to assign the dependent variable until a new query explanatory vector is inputted. When the query explanatory vector is received, a set of similar data records is retrieved from the available training dataset and is used

to assign the dependent variable to the new instance [43]. In order to choose the similar data records, lazy methods employ a distance measure that will give nearby data records higher relevance. Lazy methods choose the k data records that are nearest to the query instance. The dependent variable of the new instance is determined based on the k -nearest instances.

Lazy learning algorithms have three basic steps:

- (i) *Defer*: lazy learning algorithms store all training data and defer processing until a new query is given.
- (ii) *Reply*: a local learning approach developed by Bottou and Vapnik in 1992 is a popular method to determine the dependent variables for news queries [44]. In the Bottou and Vapnik learning approach, instances are defined as points in the space, and a similarity function is defined on all pairs of these instances.
- (iii) *Flush*: after solving a query, the answer and any intermediate results are discarded.

3.2.5. Regression Tree. A regression tree is one of the widely used decision tree algorithms. A decision tree is a data-mining tool designed to extract useful information from large datasets and use the information to help decision-making processes. A regression tree consists of a set of nodes that can assign the value of the dependent variable to an explanatory vector. Regression tree constructs a tree style decision rule set and divides the training data into the leaf nodes of the decision tree according to the numerical or categorical values of explanatory variables. The regression rules of each leaf node are derived from a mathematical process that minimizes the regression errors of the leaf nodes [45].

3.2.6. Decision Table. Similar to the regression tree, decision table also determines the value of the dependent variable with a set of decision rules [46]. However, the decision table arranges decision rules as a table, rather than a tree. A decision table usually consists of a number of parallel decision rules. Similar to the regression tree, the training data will be divided into several groups, each of which will be represented by a decision rule. For a given explanatory vector (input), an appropriate decision rule will be first selected based on the values of its explanatory variables. The dependent variable for this input will be assigned as the average of the dependent variables of all training data vectors in the corresponding group. The dependent variable can also be determined by performing linear regression on the corresponding group of training data. Empirical studies show that the decision table has a similar performance to regression trees.

3.3. Converting Wind Speed to Wind Power. An elementary method is used in this paper to convert the predicted wind speed to the predicted wind power output of a wind turbine or wind farm. The predicted wind speed is provided by one of the six machine learning regression methods discussed

above. The wind speed is then input into the certified wind turbine power curve and transformed into the wind power.

The Vestas V90-3.0 MW wind turbine is selected for the case studies in this paper. Vestas V90-3.0 MW is a pitch regulated upwind wind turbine with active yaw and a three-blade rotor. It has a rotor diameter of 90 m with a generator rated at 3.0 MW. Vestas V90-3.0 MW is widely used in Australia wind power plants and has a proven high efficiency.

The typical power curve of Vestas V90-3.0 MW, 60 Hz, 106.7 dB(A) is shown in Figure 4. It can be clearly observed that the wind power output $p(u)$ is proportional to u^3 for small wind speed u . Moreover, the power curve is steep for medium wind speeds and flat for large wind speeds. The cut-in speed is 3.5 m/s, and the cut-out speed is 25 m/s [47].

3.4. Performance Evaluation. Before proposing the case study results, several criteria are introduced for performance evaluation. Given T historical wind power values p_t , $1 \leq t \leq T$, of a time series $\{p_t\}$ which are converted from T historical wind speed observations and the corresponding forecasted power values p_t^* , $1 \leq t \leq T$; mean absolute percentage error (MAPE) is defined as

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \frac{|p_t - p_t^*|}{p_t}. \quad (13)$$

MAPE is a widely used criterion for time-series forecasting. It will also be employed to evaluate the proposed method in the case studies.

Another two criteria are presented to evaluate interval forecasting. Given T wind power values p_t , $1 \leq t \leq T$, of a time series $\{P_y\}$ and the corresponding forecasted α -level prediction intervals $[l_t, u_t]$, $1 \leq t \leq T$, the empirical confidence $\hat{\alpha}$ [48] and the absolute coverage error (ACE) are defined as

$$\hat{\alpha} = \frac{\text{frequency}(p_t \in [l_t, u_t])}{T}, \quad (14)$$

$$\text{ACE} = |\alpha - \hat{\alpha}|,$$

where $\hat{\alpha}$ is the number of observations, which fall into the forecasted prediction interval (PI), divided by the sample size. It should be as close to α as possible.

4. Case Studies

4.1. The Setting of Case Studies. In the experiments, the wind power forecasting model has been evaluated using the wind speed data from the Devonport Airport Wind Station, Tasmania, Australia. The data were provided by the Australian Bureau of Meteorology. The training and testing data have the following four numerical features: wind speed, wind direction, humidity, and temperature. The training data are from 1st February 2018 to 1st March 2018, while the testing data are from 1st February 2019 to 1st March 2019.

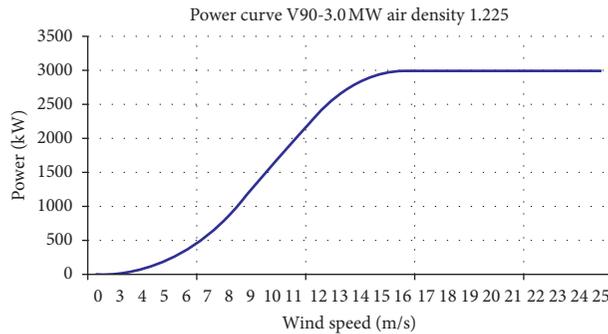


FIGURE 4: Power curve for Vestas V90-3.0MW, 60 Hz, 106.7.

TABLE 1: The results of the Lagrange multiplier test.

Dataset	Order	P value	LM statistics	Critical value
Feb 2008 to Mar 2008	1	0	1913.6	3.8415
Feb 2008 to Mar 2008	5	0	1964.6	11.0705
Feb 2008 to Mar 2008	10	0	1969.3	18.307
Feb 2009 to Mar 2009	1	0	2898.9	3.8415
Feb 2009 to Mar 2009	5	0	3057.2	11.0705
Feb 2009 to Mar 2009	10	0	3077	18.307

To empirically prove the validity of our model, we will first verify that the wind speed data exhibit time-changing distribution effect by performing the Lagrange multiplier test [49, 50]. The results of the LM test with 95% significance level on the data from 1st February 2019 to 1st March 2019 are given in Table 1.

As illustrated in Table 1, setting the significance level as 0.05, P value of the LM test is zero in all six cases. Moreover, the LM statistics are significantly greater than the critical value of the LM test in all occasions. These two facts strongly indicate that the wind speed data have strong effect of time-changing distribution. In the test, an order of 10 means that the variance σ_t^2 is correlated with its lagged values up to at least σ_{t-10}^2 . In other words, the wind speed at 10 time units before time t can still influence the uncertainty of the wind speed at time t .

4.2. Results of Wind Speed Forecasting. Wind speed forecasting is the first step of wind power forecasting. Six regression methods are first employed to perform one-hour-ahead wind speed forecasting in this paper. The performances of six algorithms are shown in Table 2.

As illustrated in Table 2, the MAPEs of LSTM and lazy IBK are smaller than other methods. Moreover, the MAPE of LSTM is under 10%, which is sufficiently good considering the very high volatility of wind speed. The results indicate that these two nonlinear machine learning regression methods perform well in wind speed forecasting.

The forecasting errors of three methods are graphically shown in Figure 5. In Figure 5, the visual inspection suggests that the forecasting errors of the three algorithms have a normal distribution. It is very important to know the type of the error distribution to ensure that the proposed statistical

TABLE 2: Prediction errors of different methods.

Regression methods	MAPE
Linear regression	12.81
Multilayer perceptron	12.32
LSTM	8.10
Lazy IBK	10.46
Decision table	15.10
Regression tree	11.26

model has a valid assumption. To empirically prove that the wind speed forecasting errors are normally distributed, the forecasting errors of all six methods are checked for normality by performing the Kolmogorov–Smirnov normality test. The test results also show that all the six forecasting methods have normally distributed errors. These results again verify the validity of the assumptions of our model.

4.3. Results of Wind Power Interval Forecasting. The wind speed forecasts given by the six machine learning regression algorithms are then converted into wind power forecasts as discussed in Section 3. Similarly, mean absolute percentage error (MAPE) is used to evaluate the performances of different methods. From Table 3, it is observed that, for wind power forecasting, the MAPE of LSTM is still lower than other five algorithms.

Based on Tables 2 and 3, the LSTM method is selected as the wind speed point forecasting method (the estimator of $f(\bullet)$). The procedure discussed in Section 2 is then employed to give the prediction intervals of wind power. We will employ all six regression methods to estimate $g(\bullet)$ and $h(\bullet)$ and then compare their performances in wind power interval forecasting.

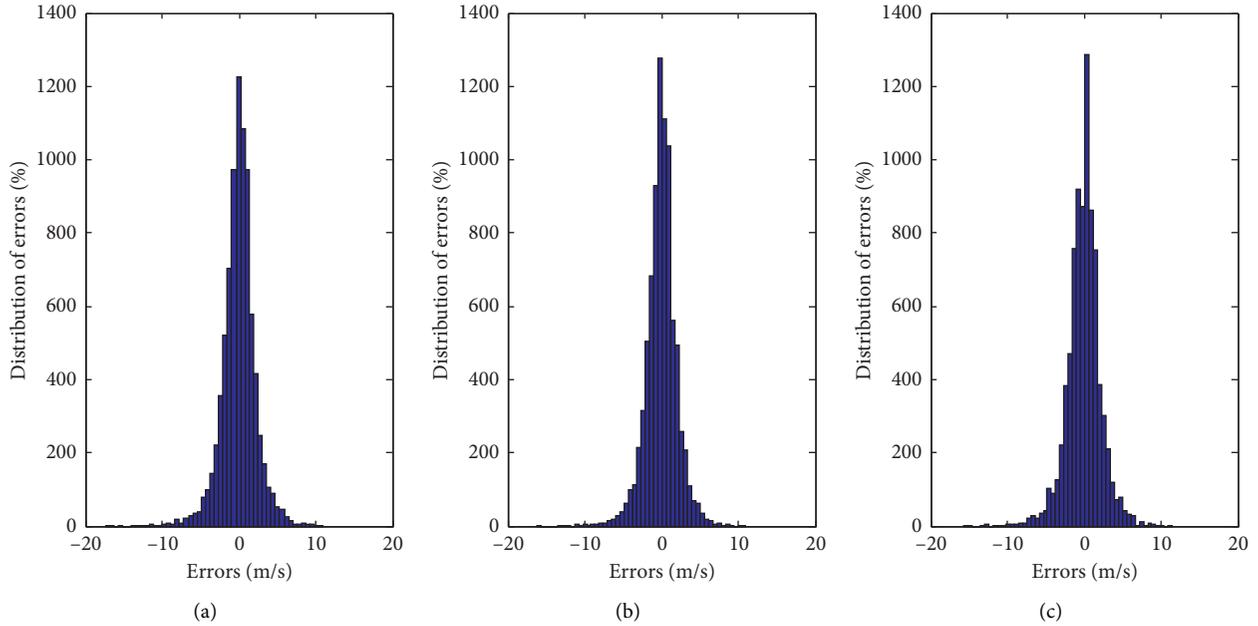


FIGURE 5: Distributions of the errors of (a) linear regression, (b) LSTM, and (c) regression tree.

In Table 4, for 95% and 99% confidence levels, the ACEs of different regression methods are presented. As seen in Table 4, the ACEs of five nonlinear methods are similar regardless of the confidence level. On the contrary, all the five nonlinear regression algorithms outperform linear regression. This is a clear proof that strong nonlinearity exists in the wind power data.

The 95% level and 99% level prediction intervals given by different methods are illustrated in Figures 6 and 7. As illustrated, the prediction intervals given by all the five nonlinear machine learning algorithms perfectly contain the true values of wind power. These results clearly prove the effectiveness of the proposed statistical model. Moreover, the results also show that nonlinear machine learning regression methods are suitable candidates in wind power interval forecasting. Compared with other machine learning methods, LSTM performed best in wind power interval forecasting. LSTM is a deep learning neural network algorithm. The improvement of the structure level of the deep learning neural network will make the information abstraction ability of the deep learning model stronger. Therefore, its ability to extract and learn complex information from large amounts of data is also stronger. The accuracy of wind power interval forecasting will be improved accordingly. Multilayer perceptron (MLP) can be categorized as the feedforward neural network. In the traditional feedforward neural network such as MLP, the input layer, the hidden layer, and the output layer in the network are fully connected, but the nodes within each layer are disconnected. This structure results in the inability of the traditional feedforward

TABLE 3: The MAPE of different methods for wind power forecasting.

Regression methods	MAPE (%)
Linear regression	37.62
Multilayer perceptron	42.48
LSTM	19.24
Lazy IBK	28.09
Decision table	35.58
Regression tree	30.05

TABLE 4: Performances of different methods on wind power interval forecasting.

Regression methods	ACE for 95% confidence	ACE for 99% confidence
Linear regression	5.37	3.34
Multilayer perceptron	3.19	0.39
LSTM	3.02	0.16
Lazy IBK	3.16	0.38
Decision table	3.16	0.43
Regression tree	3.2	0.39

neural network to deal with the problem of correlation between inputs. Compared with the feedforward neural network, circular neural network introduces directional circulation. At this point, the nodes between hidden layers in the network are no longer disconnected but connected. And the input of the hidden layer includes not only the output of the input layer but also the output of the hidden

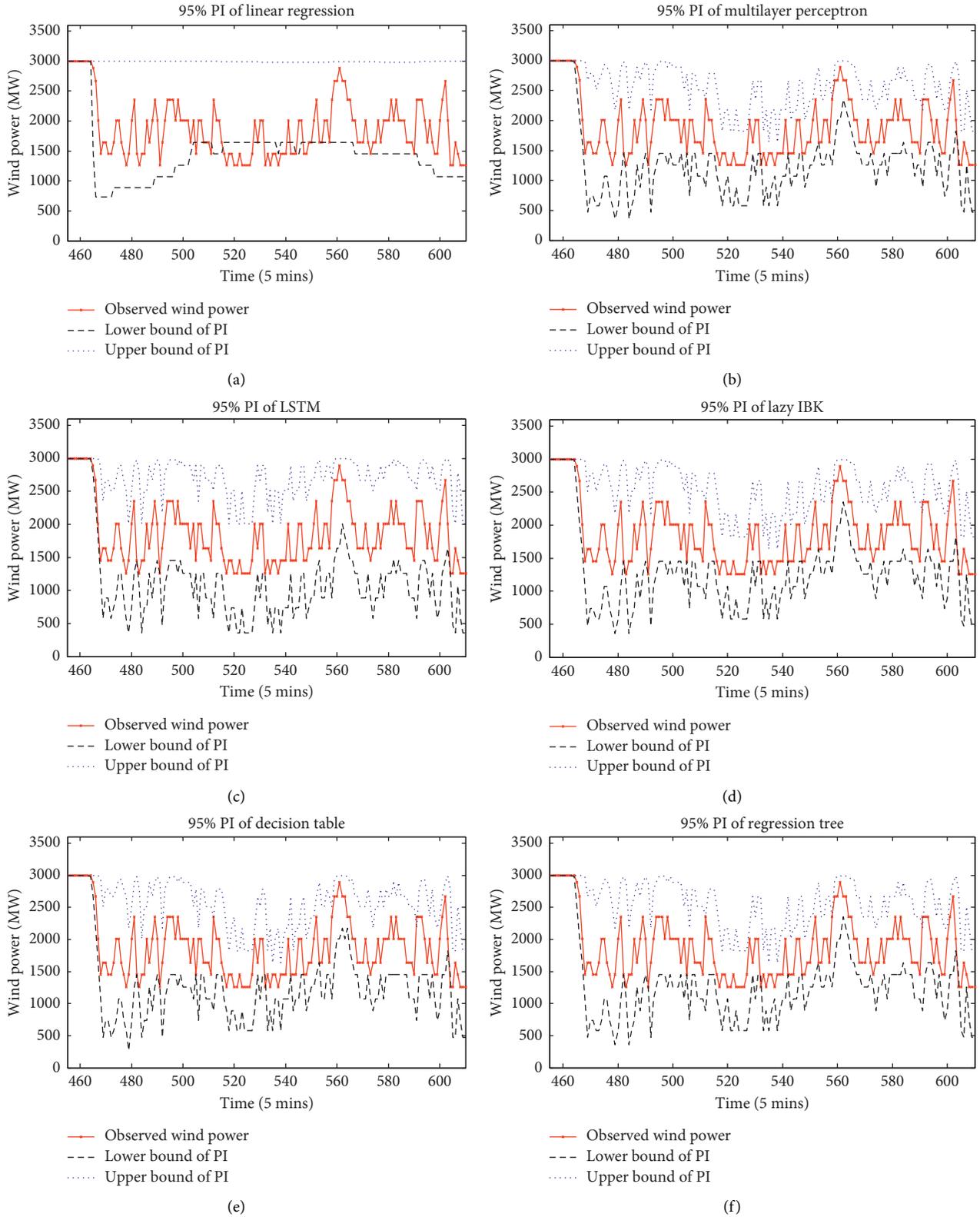


FIGURE 6: 95% level prediction intervals forecasted by six machine learning regression methods.

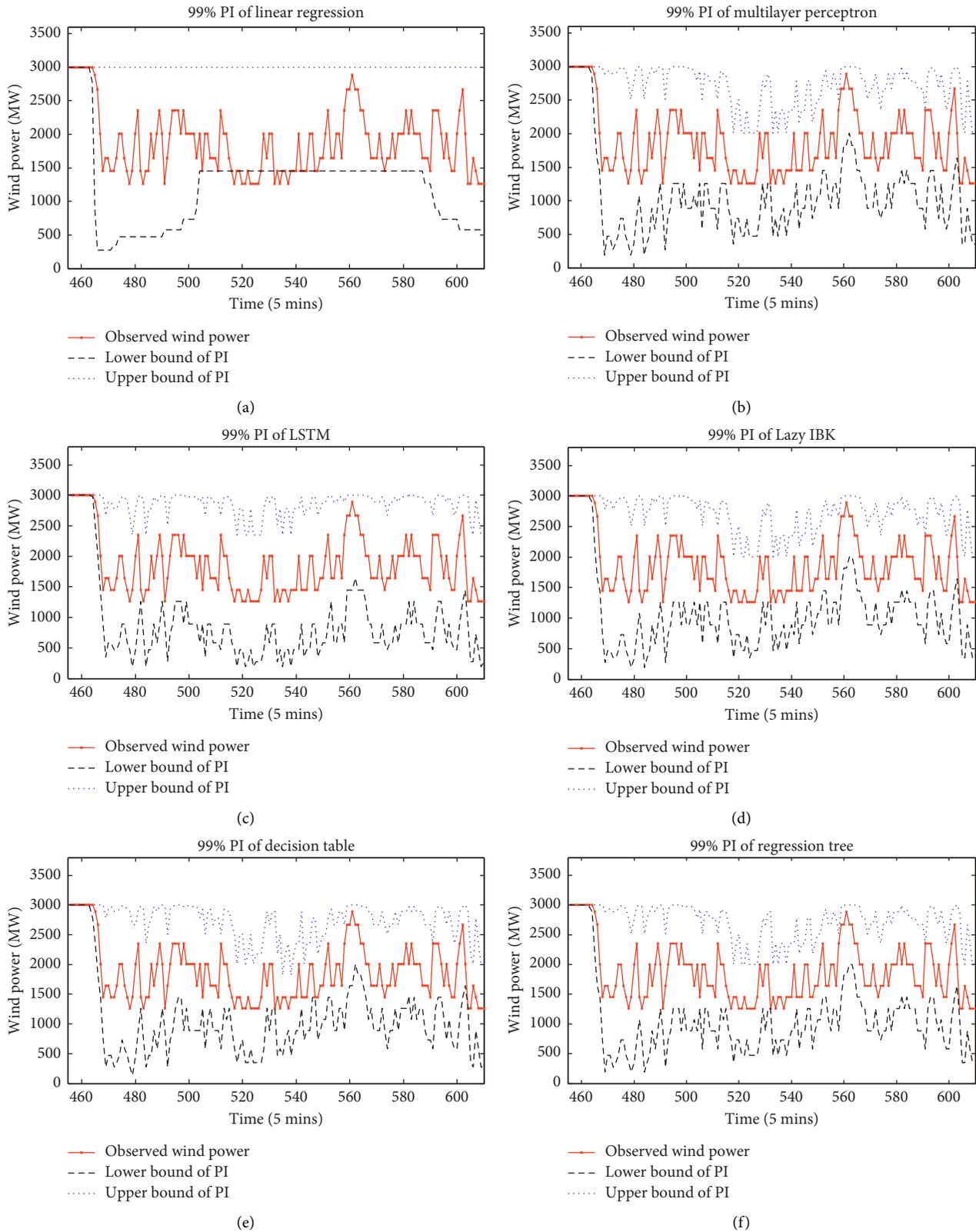


FIGURE 7: 99% level prediction intervals forecasted by six machine learning regression methods.

layer at the last moment. As a conclusion, LSTM can perform better than MLP.

5. Conclusion

This research work develops a novel comprehensive integrated statistical machine learning strategy for wind power forecasting in Australian wind farm, including exploration of the statistical characteristics of the data by statistical tools and developing the forecasting model by different statistical machine learning methods. Accurate wind power interval forecasting is essential for efficient planning and operation of power systems. Wind energy is characterised by its non-linearity and intermittency, which pose significant challenges for wind power forecasting. Traditional linear time-series models cannot appropriately handle these challenges and therefore cannot achieve satisfactory performances. In this paper, we propose a machine learning-based statistical approach, which can handle nonlinear time series with time-changing distributions, thus is suitable for wind power interval forecasting.

Compared with other relevant references, this research work shows that classical regression techniques are not suitable for complicated applications such as wind power interval forecasting. It is inappropriate simply to use linear assumptions for these problems. In addition, other research works only using complicated machine learning approaches failed to balance the important information of the historical data. Experimental results show that LSTM is the most suited candidate for wind power forecasting. Moreover, the effectiveness and accuracy of the proposed model in wind power interval forecasting are also proven with the case studies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] L. Chen, Z. Li, and Y. Zhang, "Multiperiod-ahead wind speed forecasting using deep neural architecture and ensemble learning," *Mathematical Problems in Engineering*, vol. 2019, Article ID 9240317, 14 pages, 2019.
- [2] Q. Wang, Y. Lei, and H. Cao, "Wind power prediction based on nonlinear partial least square," *Mathematical Problems in Engineering*, vol. 2019, Article ID 6829274, 9 pages, 2019.
- [3] S. Wang, J. Na, and Y. Xing, "Adaptive optimal parameter estimation and control of servo mechanisms: theory and experiment," *IEEE Transactions on Industrial Electronics*, *Early Access*, p. 1, 2020.
- [4] S. Wang and J. Na, "Parameter estimation and adaptive control for servo mechanisms with friction compensation," *IEEE Transactions on Industrial Informatics*, *Early Access*, vol. 16, 2020.
- [5] S. Wang, L. Tao, Q. Chen, J. Na, and X. Ren, "USDE-based sliding mode control for servo mechanisms with unknown system dynamics," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 2, pp. 1056–1066, 2020.
- [6] M. Santhosh and C. Venkaiah, "Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: a review," *Engineering Reports*, vol. 2, no. 6, pp. 1–20, 2020.
- [7] G. Gregor and K. George, *Renewable Energy Forecasting: From Models to Applications*, Elsevier, Amsterdam, Netherlands, 2017.
- [8] M. Lange and U. Focken, *Physical Approach to Short-Term Wind Power Prediction*, Springer, Berlin, Germany, 2006.
- [9] Z. Zhang, Y. Chen, X. Liu et al., "Two-stage robust security-constrained unit commitment model considering time autocorrelation of wind/load prediction error and outage contingency probability of units," *IEEE Access*, vol. 7, pp. 2169–3536, 2019.
- [10] Q. Hu, S. Zhang, M. Yu, and Z. Xie, "Short-term wind speed or power forecasting with Heteroscedastic support vector regression," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 241–249, 2016.
- [11] C. Wan, J. Lin, J. Wang et al., "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2767–2778, 2016.
- [12] Y. Ren, P. N. Suganthan, and N. Srikanth, "A novel empirical mode decomposition with support vector regression for wind speed forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1793–1798, 2016.
- [13] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3276–3284, 2018.
- [14] Z. Wang, J. Zhang, Y. Zhang, C. Huang, and L. Wang, "Short-term wind speed forecasting based on information of neighboring wind farms," *IEEE Access*, vol. 8, pp. 16760–16770, 2020.
- [15] L. Zhang, Y. Dong, and J. Wang, "Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy," *IEEE Access*, vol. 7, pp. 176000–176023, 2019.
- [16] Y.-K. Wu, P.-E. Su, T.-Y. Wu, J.-S. Hong, and M. Y. Hassan, "Probabilistic wind-power forecasting using weather ensemble models," *IEEE Transactions on Industry Applications*, vol. 54, no. 6, pp. 5609–5620, 2018.
- [17] F. Ge, Y. Ju, Z. Qi et al., "Parameter estimation of a Gaussian mixture model for wind power forecasting error by Riemann L-BFGS optimization," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 258–265, 2007.
- [18] C. Wang, Q. Teng, X. Liu et al., "Optimal sizing of energy storage considering the spatial-temporal correlation of wind power forecast errors," *IET Renewable Power Generation*, vol. 13, no. 4, pp. 530–538, 2019.
- [19] J. Na, B. Wang, G. Li, S. Zhan, and W. He, "Nonlinear constrained optimal control of wave energy converters with adaptive dynamic programming," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 7904–7915, 2019.
- [20] J. Na, Y. Huang, X. Wu et al., "Adaptive finite-time fuzzy control of nonlinear active suspension systems with input delay," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2639–2650, 2019.
- [21] J. Na, A. Chen, Y. Huang et al., "Air-fuel ratio control of internal combustion engines with unknown dynamics

- estimator: theory and experiments," *IEEE Transactions on Control Systems Technology*, vol. 8, p. 8, 2019.
- [22] C. Wu, Y. Zhao, and M. Sun, "Enhancing low-speed sensorless control of PMSM using phase voltage measurements and online multiple parameter identification," *IEEE Transactions on Power Electronics*, vol. 35, no. 10, pp. 10700–10710, 2020.
- [23] M. Sun, "Two-phase attractors for finite-duration consensus of multiagent systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1757–1765, 2020.
- [24] Q. Chen, H. Shi, and M. Sun, "Echo state network based backstepping adaptive iterative learning control for strict-feedback systems: an error-tracking approach," *IEEE Transactions on Cybernetics, Early Access*, vol. 50, no. 7, Article ID 2931877, 2019.
- [25] Q. Chen, S. Xie, M. Sun, and X. He, "Adaptive nonsingular fixed-time attitude stabilization of uncertain spacecraft," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 6, pp. 2937–2950, 2018.
- [26] M. Tao, Q. Chen, X. He et al., "Adaptive fixed-time fault-tolerant control for rigid spacecraft using a double power reaching law," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 12, pp. 4022–4040, 2019.
- [27] Q. Chen, X. Ren, J. Na et al., "Adaptive robust finite-time neural control of uncertain PMSM servo system with nonlinear dead zone," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3725–3736, 2017.
- [28] Q. Zheng, L. Shi, J. Na, X. Ren, and Y. Nan, "Adaptive echo state network control for a class of pure-feedback systems with input and output constraints," *Neurocomputing*, vol. 275, no. 1, pp. 1370–1382, 2018.
- [29] Q. Chen, X. Yu, M. Sun et al., "Adaptive repetitive learning control of PMSM servo systems with bounded nonparametric uncertainties: theory and experiments," *IEEE Transactions on Industrial Electronics*, 2020.
- [30] D. Tung and T. Le, "A statistical analysis of short-term wind power forecasting error distribution," *International Journal of Applied Engineering Research*, vol. 12, no. 10, pp. 2306–2311, 2017.
- [31] L. Wang, R. Yan, F. Bai et al., "A distributed inter-phase coordination algorithm for voltage control with unbalanced PV integration in LV systems," *IEEE Transactions on Sustainable Energy, Early Access*, Article ID 2970214, 2020.
- [32] L. Wang, R. Yan, and T. Saha, "Voltage regulation challenges with unbalanced PV integration in low voltage distribution systems and the corresponding solution," *Applied Energy*, vol. 256, no. 1, Article ID 113927, 2019.
- [33] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [34] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: a survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.
- [35] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [36] R. Thomas, "Modern Regression Methods," John Wiley & Sons, Hoboken, NJ, USA, 2009.
- [37] R. Maronna, R. Martin, V. Yohai, and M. S-Barrera, *Robust Statistics: Theory and Methods*, Wiley, New York, NY, USA, 2019.
- [38] J. Keller, D. Liu, and D. Fogel, "Multilayer Neural Networks and Backpropagation," Wiley, Hoboken, NJ, USA, 2016.
- [39] E. Alpaydin, "Introduction to Machine Learning," The MIT Press, Cambridge, MA, USA, 2014.
- [40] E. Alpaydin, "Machine Learning: The New AI," The MIT Press, Cambridge, MA, USA, 2016.
- [41] S. Kok and M. Simsek, "A Deep Learning Model for Air Quality Prediction in Smart Cities," pp. 1983–1990, IEEE International conference on Big Data, Boston, MS, USA, 2017.
- [42] P. Zhou, W. Shi, and J. Tian, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 207–212, Berlin, Germany, 2016.
- [43] Z. Hou, S. Liu, and T. Tian, "Lazy-learning-based data-driven model-free adaptive predictive control for a class of discrete-time nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 8, pp. 1914–1928, 2017.
- [44] L. Merschmann and A. Plastino, "A lazy data mining approach for protein classification," *IEEE Transactions on NanoBioscience*, vol. 6, no. 1, pp. 36–42, 2007.
- [45] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1978–1987, 2013.
- [46] M. Azad and M. Moshkov, "Minimization of decision tree depth for multi-label decision tables," in *Proceedings of the 2014 IEEE International Conference on Granular Computing (GrC)*, pp. 368–377, Noboribetsu, Japan, October 2014.
- [47] VESTAS Wind Power Solution Company Webpage, http://www.vestas.com/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2FEN%2FBrochures%2FVestas_V_90-3MW-11-2009-EN.pdf.
- [48] E. Mazloumi, G. Currie, and G. Rose, "Statistical confidence estimation measures for artificial neural networks: application in bus travel time prediction," in *proceedings of the 2010 Transportation Research Board 89th Annual Meeting*, pp. 1–13, Washington, DC, USA, 2010.
- [49] M. Wang, K. Ngan, and H. Li, "Low-delay arte control for consistent quality using distortion-based lagrange multiplier," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 2943–2955, 2016.
- [50] Y. Lin and A. Abur, "A new framework for detection and identification of network parameter errors," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1698–1706, 2018.