



Probabilistic power curve estimation based on meteorological factors and density LSTM

Peng Wang^a, Yanting Li^{a,*}, Guangyao Zhang^a

^a School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Handling Editor: Jesse L. The

Keywords:

Probabilistic power curve
Temporal effect
LSTM
Probability density function
Negative log-likelihood loss function
Heteroskedasticity

ABSTRACT

Wind power curve describes the relationship between wind speed and wind power output, which is useful for wind farm design and wind turbine condition monitoring. However, most research on power curves modeling neglected the effect of historical meteorological factors. Also, the research on probabilistic power curve is still limited. Therefore, this paper proposes a novel probabilistic power curve modeling approach. The effect of both current and historical meteorological variables, such as wind speed, wind direction, ambient temperature and turbulence intensity, on the power curve is studied by binning method and kernel density estimation. Next, instead of forecasting the wind power or its quantiles, a new probabilistic power curve model, Density LSTM with Negative Log-Likelihood Loss, is proposed to forecast the parameters of the probability density function of wind power directly. Finally, different probability distribution of wind power output is studied. Gamma, Laplace and Weibull distributions prove to be more suitable than Gaussian distribution. Based on the datasets of inland and offshore wind farms, it is verified that adding useful historical meteorological variables can improve the forecasting performance of the test dataset. Besides, the proposed probabilistic power curve can effectively improve the prediction performance in the probabilistic prediction task and can improve the prediction precision of the annual energy production.

1. Introduction

Wind energy is a renewable, pollution-free and widely distributed energy source and has received more and more attention. Wind turbine (WT) plays an important role in transforming wind power into electricity. The cumulative wind power capacity has continued to grow, from 198 GW in 2010 to 837 GW in 2021, and is expected to reach 3200 GW by 2030 [1]. However, the fluctuation and intermittency of wind speed are intense, which makes wind power present a certain uncertainty.

The wind turbine power curve describes the relationship between wind speed and wind power output. It is important to establish an accurate wind power curve because it is useful for wind power output forecasting [2–4], wind power monitoring [5–7] and wind power potential estimation [8]. Typically, the wind turbine power curve can be divided into four regions based on cut-in wind speed (V_i), rated wind speed (V_R) and cut-out wind speed (V_o) as shown in Fig. 1. When the wind speed is less than cut-in wind speed, the wind power is zero. When the wind speed is greater than rated wind speed and less than cut-out

wind speed, the wind power is approximately equal to the rated power P_R . When the wind speed is greater than cut-in wind speed and less than rated wind speed, the wind power curve presents a nonlinear relationship, which is the key to modeling. In this region, the theoretical power curve provided by the turbine manufacturer is expressed as $p = 0.5\pi c_p \rho R^2 V^3$, where c_p , ρ and R represent the wind turbine power coefficient, air density and wind rotor radius respectively [9]. However, the theoretical power curve cannot always capture the actual behavior of wind turbines due to the complex environment and operating conditions. Therefore, it is necessary to use data from actual wind turbines to construct more accurate wind power curves. SCADA (Supervisory Control and Data Acquisition) records relevant wind turbine data, such as wind speed, wind power output, ambient temperature, wind direction and wind turbine status, which is suitable for this task.

The existing research in the power curve model estimation field can be divided into two categories: deterministic method and probabilistic method. A large majority of existing wind power curve models are deterministic models which only provided a specific wind power estimation. Traditional deterministic models are parametric methods

* Corresponding author.

E-mail address: ytli@sjtu.edu.cn (Y. Li).

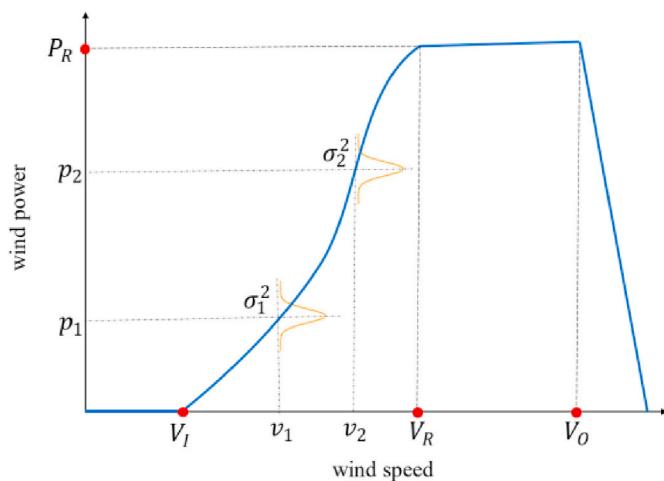


Fig. 1. Wind turbine power curve.

including piecewise linear regression, polynomial regression, exponential power curve, cubic power curve and nonparametric methods including artificial neural network, random forest, k-Nearest Neighbor, Copula model, etc. Detailed overviews of these models were given in Refs. [9–11]. There are some new methods proposed in recent years [12–17]. Lee and Ding et al. [12] proposed an additive multivariate kernel method that includes the other environmental factors, such as wind direction, air density, humidity, turbulence intensity and wind shears. It is worth noting that they found the important environmental factors which affect the power output and used operational data for wind turbines in inland and offshore wind farms to demonstrate the improvement achieved by their method. Saint-Drenan et al. [13] examined the impact of technical and environmental factors and proposed a parametric model which can generate the power curve of any wind turbine, adapted to specific conditions of any site. Mehrjoo et al. [14] described a monotone spline regression approach to solve the problem that the Nadaraya-Watson kernel estimator fails to preserve the monotonicity property of the power curve. Han et al. [15] divided wind speed into different bins and proposed an interval extreme probability density method to construct accurate power curves in the presence of abnormal data. Karamichailidou et al. [16] employed radial basis function neural networks to construct power curves and introduced the non-symmetric fuzzy means approach to training the network which is hybridized with the tabu search metaheuristic technique. Li et al. [17] integrated iForest, Nonsymmetric Fuzzy Means, Radial Basis Neural Network and metaheuristic algorithm to construct the power curve.

The error distribution in wind power curve modeling is usually asymmetric and complex, rather than a Gaussian distribution with constant variance [18]. However, most existing deterministic power curves failed to take the heteroscedasticity into consideration. In recent years, some researchers have tried to addressed error heteroscedasticity [18–21]. Wang et al. [18] developed two asymmetric spline regression models in which the errors are assumed to obey a mixture of asymmetric Gaussian distributions and a mixture of asymmetric exponential power distribution, respectively. Wei et al. [19] modeled the distribution of wind speed by a mixture of asymmetric Weibull distribution and the independent distribution of errors by a mixture of asymmetric Gaussian models and proposed a hybrid copula-based wind power curve model. Wang et al. [20] embedded the heteroscedasticity of WTPC modeling into multiple spline regression models (MSRM) based on Gaussian and Student's t-distributions and proposed two sparse heteroscedastic MSLR power curves. Zou et al. [21] proposed a novel loss function that considers the asymmetric error characteristic and a hybrid intelligent optimization method to construct a set of parametric WTPC models.

However, because of the complex climatic conditions and topography, it's difficult to get an accurate wind power curve based on

deterministic models because they cannot quantify the uncertainty of the wind power generation process. On the contrary, probabilistic power curve methods consider the wind power output as a random variable [22] and the uncertainty can be described by probability density functions, quantiles, or confidence intervals of the predicted values. Under the framework of stochastic optimization, probabilistic power curves are widely used for decision-making problems of power system operation, related to reserve requirement, trading strategy of wind power, unit commitment, energy storage sizing and optimal dispatch of wind-hydro power plants [23]. Fig. 1 shows an example of a probabilistic power curve. (v_1, p_1) and (v_2, p_2) are the wind speed and wind power output for two points in the power curve, the variances of wind power output p_1 and p_2 are σ_1^2 and σ_2^2 respectively.

Some researchers have done a lot of work to construct probabilistic power curves [24–28]. Xu et al. [24] proposed a quantile loss-based neural network algorithm to establish the quantile power curve which generates a series of power curves under any confidence level and quantile criteria to measure the wind turbine performance. Yun et al. [25] constructed wind speed bins and assumed that wind power output follows Weibull distribution. Then they utilized Monte-Carlo simulation to construct a probabilistic power curve which provides confidence intervals of wind power. Yan et al. [26] established probabilistic power curves based on Monte Carlo, Neural Network and Fuzzy Clustering. Virgolino et al. [27] combined Gaussian process and standard logistic functions to construct probabilistic power curves. Rogers et al. [28] proposed a heteroscedastic Gaussian Process model to construct the probabilistic model and used a Bayesian framework that exhibits built-in protection against over-fitting and robustness to noisy measurements.

To sum up, existing probabilistic power curves either assume wind power output to follow a specific probability distribution, for example, Weibull distribution and construct wind speed bins to solve corresponding parameters using distribution fitting techniques [25,26], or combine Gaussian process in which wind power output follows Gaussian distribution, with Bayesian framework which learns the parameters and hyperparameters through a variational sparse approximation [27,28]. However, binning method limits the extensibility of model input and Gaussian distribution may not describe the characteristics of wind power output accurately. There has not been a non-parametric probabilistic power curve that allows the wind power output to follow any other specific probability distribution except Gaussian distribution and learn model parameters without using wind speed bins. Besides, existing power curves haven't studied the impact of environmental factors' historical data on the current wind power output, which can be defined as the temporal effect. Because energy transfer is hysteresis, that is it takes a while to convert wind speed into wind power output, the temporal effect is worth studying. To fill the above two knowledge gaps, we aim to figure out the impact of historical meteorological factors and propose a nonparametric probabilistic power curve to forecast density functions of wind power output given specific probability distribution without using the binning method. Therefore, a novel probabilistic wind power curve estimation method based on multivariate historical meteorological factors and Density LSTM is proposed. Temporal effects of wind speed, wind direction and ambient temperature are validated. These factors are then used to construct Density LSTM which can forecast density functions of wind power output. The negative log-likelihood loss function is applied to train the model and learn parameters. The main contributions of this paper are as follows.

- 1) The effect of multivariate historical meteorological factors, including wind speed, wind direction and ambient temperature, on the wind power output, are important but neglected in previous studies. It is validated in this study and helps improve the forecasting performance of the test set.
- 2) Rather than providing the expected value or quantiles of wind power output, Density LSTM with density function as output is established.

All probabilistic representations (e.g., quantiles, confidence intervals) can be derived from the density function. The negative log-likelihood loss function is applied to train the model effectively. This approach also considers the existence of heteroscedasticity.

- 3) The distribution characteristic of wind power output in a specific wind speed bin is studied. It is found that wind power output does not follow the Gaussian distribution exactly in specific wind speed bins with width of 0.1 m/s. The other three probability distributions, Gamma, Laplace and Weibull distribution, fit wind power data better and outperform Gaussian distribution as the output of Density LSTM.

The rest of the paper is organized as follows. In Section 2, the datasets used in this paper and the data preprocessing method are described. In Section 3, the temporal effect is verified. In Section 4, the description of Density LSTM is presented, followed by the case study in Section 5. We finally conclude in Section 6.

2. Data description and preprocessing

2.1. Data description

In this paper, datasets of two wind farms are used. The inland and offshore wind farms of our research locate in Central and East China respectively. The inland wind farm contains 24 W2000N-105-80 wind turbines whose rated power is 2.0 MW. The offshore wind farm contains 26 W3600M-116-90 wind turbines whose rated power is 4.0 MW. Using two datasets to verify the performance of the proposed power curve can enhance the credibility of the results as their environmental conditions and wind turbine types are different.

The measurements of these two wind farms were recorded in the wind turbine SCADA system. The SCADA system recorded wind power

output, wind speed, wind direction, ambient temperature and operation state every 10 min from January 1, 2019 to December 31, 2019, resulting in 50126 and 49548 records for inland and offshore wind farms respectively. As the SCADA system may not record data during a specific period, the number of data samples for the inland and offshore wind farm is different. Characteristics of wind speed, wind direction, ambient temperature and turbulence intensity in two wind farms are as follows.

- 1) Wind speed. Wind speed affects the force on the wind turbine and largely determines the power curve. The boxen plots for wind speed in the second half year are shown in Fig. 2. It can be seen that the monthly wind speed distributions of inland and offshore wind turbines are different. The monthly average wind speed of offshore wind turbines is higher than that of the inland wind turbine. The peak wind speed of the offshore farm is mainly in summer, while that of the inland farm is in winter.
- 2) Wind direction. Wind turbines can adapt to the wind direction to increase power generation efficiency. When the wind direction changes, the turbulence intensity fluctuates due to a combination of topography and wake effects for particular wind turbines. Therefore, wind direction has a direct impact on wind power. As shown in Fig. 3, the wind direction variation of the offshore wind farm is larger than that of the inland wind farm.
- 3) Ambient temperature. Ambient temperature indirectly affects the temperature of the engine room and electrical generator, which may affect the operation of wind turbines. In the meantime, wind turbines are forced to operate at a limited power state in high temperatures. The boxplots of ambient temperature for two working states are shown in Fig. 4. It can be seen that the average ambient temperature of a limited working state is higher than that of a normal working state.

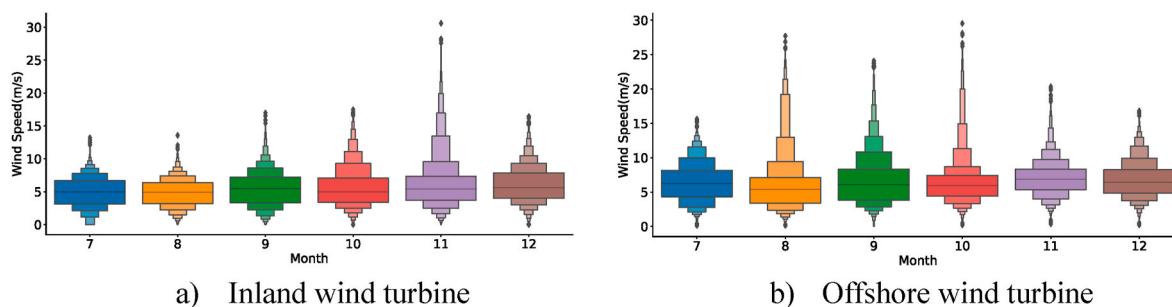


Fig. 2. Boxen plots of monthly wind speed in the second half of 2019.

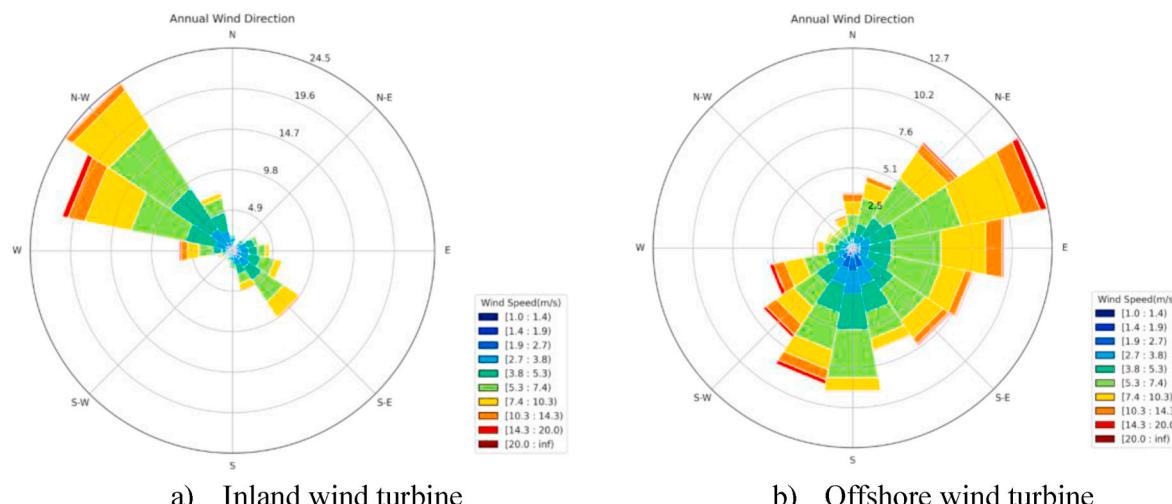


Fig. 3. Annual wind direction rose diagrams for two wind farms.

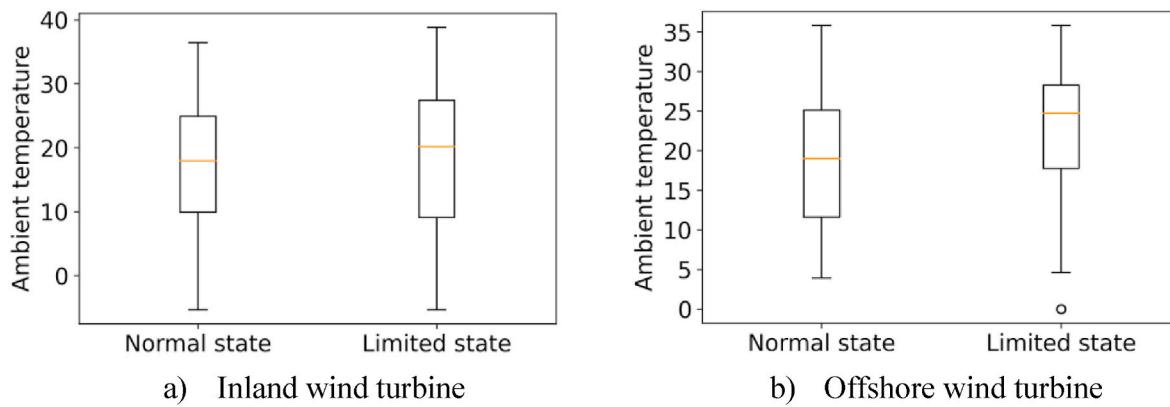


Fig. 4. Boxplots of ambient temperature for two working states.

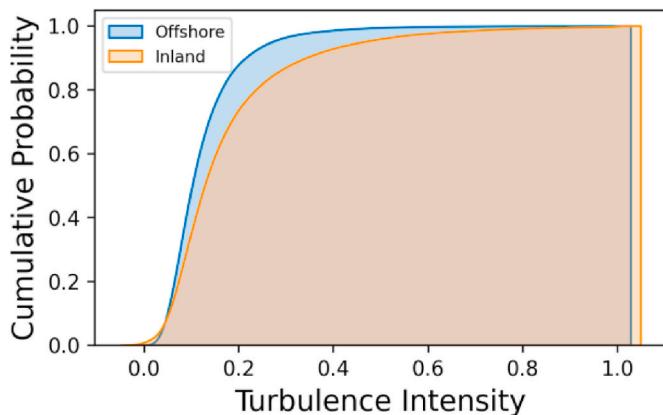


Fig. 5. Cumulative probability plot of turbulence intensity for two wind farms.

- 4) Turbulence intensity. Turbulence intensity describes the wind speed changes over spatial-temporal space. Turbulence intensity can reduce the wind energy utilization rate of wind turbine, and may affect the shape of power curve. In this paper, turbulence intensity is defined as the ratio of $\sigma(V)$ to $\mu(V)$ in the most recent 1 h, where $\sigma(V)$ and $\mu(V)$ are the standard deviation and mean of wind speed respectively. The cumulative probability plot of turbulence intensity for two wind farms is shown in Fig. 5. It can be seen in Fig. 5 that the

average turbulence intensity of inland wind farm is higher than that of offshore wind farm.

2.2. Data preprocessing

Due to the existence of outliers and missing values, before establishing the probabilistic power curve, the data should be preprocessed. Data preprocessing consists of outlier removal and data normalization.

There are 3 types of outliers in the SCADA system which are summarized by technicists and should be identified and removed. Type I outliers are the negative outliers whose values are negative or close to zero when the wind speed is larger than cut-in wind speed. Type II outliers randomly distributed around the normal data generated by control system sensor failure, wind direction fault and so on. Type III outliers are stacked points gathering in a line on one side of the power curve caused by curtailment command, communication failures and so on [29]. In recent years, outlier removal methods have mainly focused on numerical data-based approaches [29,30] and image processing technology [31–33]. For example, Luo et al. [29] proposed an improved density clustering method to remove type 3 outliers and a combined method including boundary extraction and regularization to remove type 2 outliers. Morrison et al. [30] explored the filtering impact by comparing the performances of four anomaly detection methods that are the Isolation Forest (iForest), Local Outlier Factor, Gaussian Mixture Models and k-Nearest Neighbor with/without filtering explicit and obvious anomalies from the SCADA in advance. Long et al. [31] constructed a three-dimensional WPC image considering wind speed, wind power and data frequency and introduced the Canny edge detection and Hough transform to extract features of outliers. Liang et al. [32]

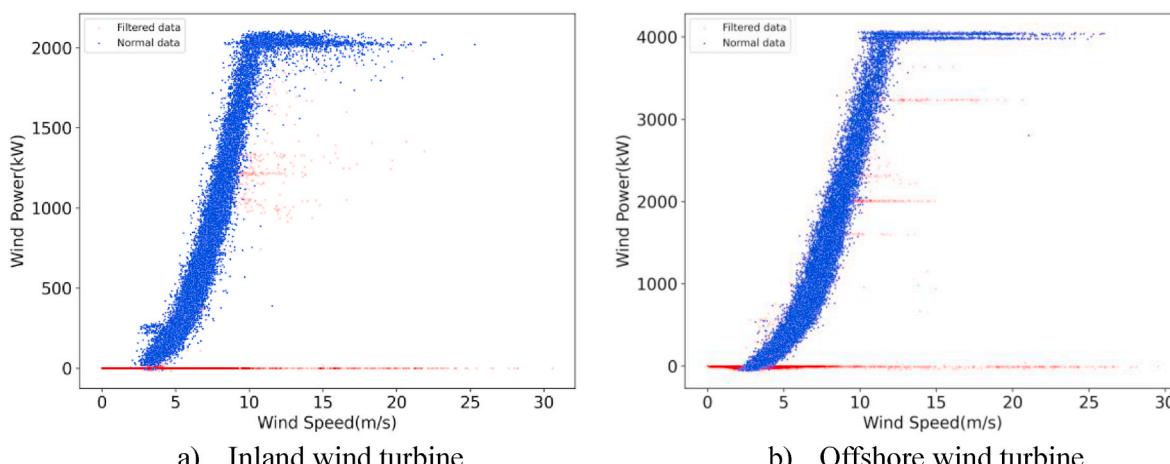


Fig. 6. Scatter plots of wind speed and wind power.

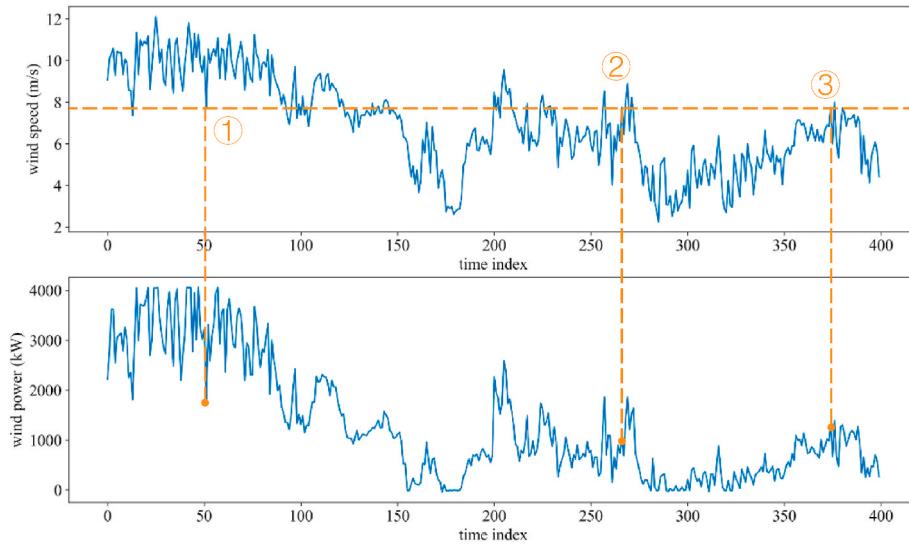


Fig. 7. High historical average wind speed leading to bigger wind power statistically.

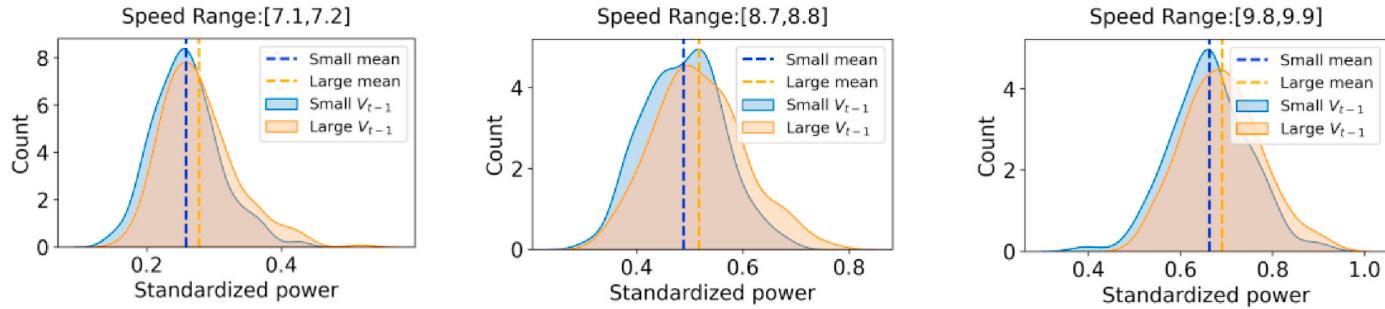


Fig. 8. Wind power density functions of B_{i1} and B_{i2} divided by v_{t-1}

proposed a WPC data cleaning method by transforming the scattered data into a digital image and using image thresholding based on the minimization of dissimilarity-and-uncertainty-based energy.

However, existing data cleaning methods are computationally intensive and lack interpretability. To remove outliers effectively and simply, we use two variables that are “operating state” and “active power set value” in SCADA data to remove data samples under abnormal or limited working states. The former variable indicates whether the wind turbine is in a normal working state or in an abnormal state; the latter variable indicates whether the wind turbine is working under a limited working state. Fig. 6 shows the data before and after filtering: the blue points are the remaining data samples and the red points are removed data samples. After data filtering, inland wind turbine has 30186 data samples and offshore wind turbine has 37849 data samples.

After removing outliers, data preprocessing is performed. For offshore wind farm, wind power which is larger than 4000 kW is truncated as 4000 kW; for inland wind farm, wind power which is larger than 2000 kW is truncated as 2000 kW. The inland and offshore wind power output range from 0 to 2000 kW and 0–4000 kW respectively, which would cause computational burden in model training and inconvenience for model comparison. Therefore, wind power output y is transformed to [0, 1] as shown in Eq. (1). This approach refers to Lee and Ding’s method [12]. They normalized the power outputs by the rated power output to protect the identity of the turbine manufacturer.

$$y^* = \frac{y - y_{min}}{y_{max} - y_{min}} \quad (1)$$

where y_{max} and y_{min} are the maximum value and minimum value of wind power output.

Input meteorological factors (wind speed, wind direction, ambient temperature and turbulence intensity) are normalized according to Eq. (2).

$$x^* = \frac{x - \bar{x}}{s_x} \quad (2)$$

where \bar{x} and s_x are the mean and standard deviation of variable x respectively.

The pre-processed data will be used for subsequent analysis.

3. Temporal effect of historical meteorological variables on wind power

Given the same wind speed, wind power output varies greatly, especially when the wind speed is bigger than the cut-in wind speed and smaller than the rated wind speed, as shown in Fig. 6. This variance may be caused by the difference in wind turbines’ environmental conditions. Therefore, a promising research direction is the multivariate approach to wind turbine power curve: incorporating current environmental information besides wind speed as input variables for the data-driven model [34]. However, relative environmental variables are at the same time period with wind power in previous research. The effects of these variables’ historical data on wind power are neglected. For example, paper [12] studied the multivariate dependencies of different

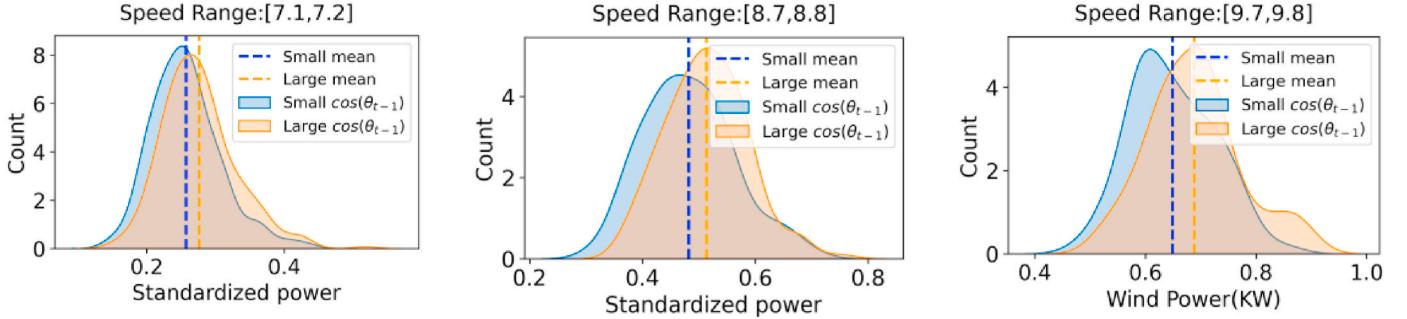


Fig. 9. Wind power density functions of B_{i1} and B_{i2} divided by $\cos(\theta_{t-1})$

current environmental variables, including wind speed, wind direction, ambient temperature, air pressure, humidity and turbulence intensity.

In addition to the influence of the current environmental variables, the difference in wind turbines' previous operation states may also affect wind power output. For example, if wind turbines have worked with high rotating speed for a long while, even if the current wind speed gets smaller, wind power output probably won't decrease. As shown in Fig. 7, wind speed and wind power values of the three selected data samples are (7.74 m/s, 1786 kW), (7.76 m/s, 1064 kW), (7.75 m/s, 1288 kW) respectively. The wind speeds of data samples 1, 2 and 3 are almost the same. However, because the previous wind speed of data sample 1 is bigger than that of data samples 2 and 3, the power output of data sample 1 is apparently bigger than that of data sample 2 and 3.

The above phenomenon can be defined as the temporal effect which means the historical working conditions have an influence on the current wind power output. The temporal effect may be caused by the hysteresis energy transfer mechanism as it takes a while to convert wind into power output.

Previous operation states can be measured by historical meteorological factors such as wind speed, wind direction, ambient temperature and turbulence intensity. To further demonstrate the existence of the temporal effect, the density functions of current wind power output given different historical meteorological factors are compared. SCADA data for a specific wind turbine at time t is represented as X_t , which contains information about wind speed, wind power, wind direction, ambient temperature and turbulence intensity. X_t is shown in Eq. (3),

$$X_t = [v_t, p_t, \theta_t, a_t, i_t] \quad (3)$$

where v_t is the wind speed at time t ; p_t is the wind power output at time t ; θ_t is the wind direction at time t ; a_t is the ambient temperature at time t ; i_t is the turbulence intensity at time t .

Motivated by Lee and Ding's method [12], we generated 80 wind speed bins by traversing wind speed from 5 m/s to 13 m/s with step 0.1 m/s. Wind speed bin B_i can be represented as Eq. (4).

$$B_i = \{X_t : v_t \in [v_i, v_i + 0.1]\} \quad v_i = 5 + 0.1 * (i - 1), i = 1, 2, \dots, 80 \quad (4)$$

where v_i is the traversed wind speed; v_t is the wind speed at time t . Then the influences of four historical meteorological factors, wind speed, wind direction, ambient temperature and turbulence intensity, on current wind power output are analyzed in each wind speed bin.

1) Wind speed

Based on the historical wind speed at previous n time steps, the data samples within wind speed bin B_i are further divided into two categories, B_{i1} and B_{i2} , as shown in Eq. (5).

$$B_{i1} = \{X_t \in B_i : v_{t-n} > \bar{v}_i\}; \quad B_{i2} = \{X_t \in B_i : v_{t-n} \leq \bar{v}_i\} \quad (5)$$

where \bar{v}_i is the average wind speed for data samples in B_i ; v_{t-n} is the historical wind speed for v_t with time step n . Then Kernel Density Estimation (KDE) [12] is used to calculate the density functions of wind power output p_t in B_{i1} and B_{i2} respectively.

KDE is a non-parametric estimation method that can estimate the density function of a random variable without the distribution hypotheses. If X_1, X_2, \dots, X_n are taken from unknown distribution samples, the KDE of the variable is defined by Eq. (6).

$$\hat{f}_d(x) = \frac{1}{nd} \sum_{i=1}^n K\left(\frac{X_i - x}{d}\right) \quad (6)$$

where $d > 0$ is the bandwidth which is determined by Scott method and $K(\cdot)$ is a non-negative kernel function which is Gaussian kernel in this paper.

After calculating the density functions of p_t in B_{i1} and B_{i2} , it is found that the average wind power output of data samples in B_{i1} is higher than that in B_{i2} significantly in all wind speed bins. Both the inland and offshore wind farms have this result, which indicates that the historical wind speed v_{t-n} has an influence on the current wind power output, where n is the time lag. This demonstrates the existence of temporal effects because if it does not exist, the mean of density function 1 and density function 2 should not differ. Fig. 8 shows density functions of wind power output in B_{i1} and B_{i2} divided by v_{t-1} for three wind speed bins whose v_i equals 7.1 m/s, 8.7 m/s and 10.9 m/s respectively. The

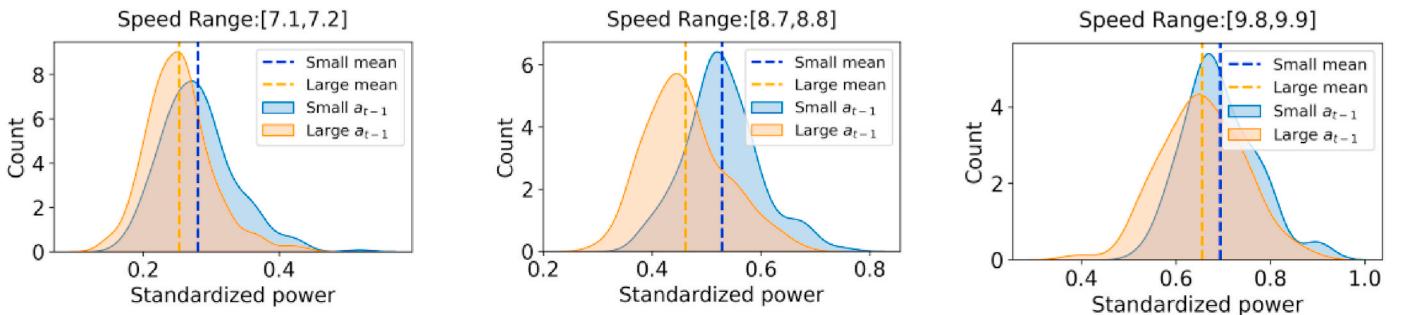


Fig. 10. Wind power density functions of B_{i1} and B_{i2} divided by a_{t-1}

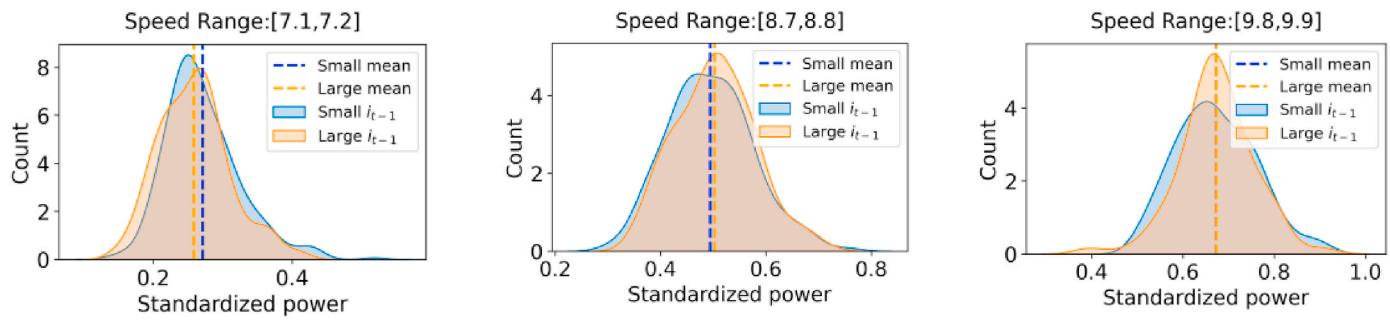


Fig. 11. Wind power density functions of B_{i1} and B_{i2} divided by i_{t-1}

orange curve represents the density of wind power output in B_{i1} , the blue curve represents the density of wind power output in B_{i2} , and the red curve represents the density of wind power output in B_i . “Large v_{t-1} ” means that v_{t-1} in B_{i1} is larger than that in B_{i2} . It can be seen that the mean of density function 1 is larger than that of density function 2. Meanwhile, as the wind speed increases, the variance of wind power output also increases, which indicates the existence of heteroskedasticity.

2) Wind direction, ambient temperature and turbulence intensity

The analysis procedures of these three meteorological factors are similar to that of wind speed except for the dividing principle of B_{i1} and B_{i2} . The dividing principles of B_{i1} and B_{i2} for wind direction, ambient temperature and turbulence intensity are shown in Eqs. (7)–(9) respectively.

$$B_{i1} = \{X_t \in B_i : \cos(\theta_{t-n}) > \overline{\cos(\theta_i)}\}; B_{i2} = \{X_t \in B_i : \cos(\theta_{t-n}) \leq \overline{\cos(\theta_i)}\} \quad (7)$$

$$B_{i1} = \{X_t \in B_i : a_{t-n} > \bar{a}_i\}; B_{i2} = \{X_t \in B_i : a_{t-n} \leq \bar{a}_i\} \quad (8)$$

$$B_{i1} = \{X_t \in B_i : i_{t-n} > \bar{i}_i\}; B_{i2} = \{X_t \in B_i : i_{t-n} \leq \bar{i}_i\} \quad (9)$$

where $\overline{\cos(\theta_i)}$, \bar{a}_i , \bar{i}_i are the average cosine value of wind direction, average ambient temperature and mean turbulence intensity for data samples in B_i respectively; θ_{t-n} , a_{t-n} and i_{t-n} are historical wind direction, ambient temperature and turbulence intensity with time lag n respectively. After calculating the density functions of wind power output in B_{i1} and B_{i2} , it can be found that for wind direction, the average wind power output of data samples in B_{i1} is significantly higher than that in B_{i2} in all wind speed bins; for ambient temperature, the average wind power output of data samples in B_{i2} is higher than that in B_{i1} in all wind speed bins. However, for turbulence intensity, the average wind power output of data samples in B_{i1} and B_{i2} does not differ. Results for time lag $n = 1$ are shown in Fig. 9–11. It can be concluded that historical wind direction and ambient temperature have influence on wind power, but wind power is not affected by historical turbulence intensity. Large historical cosine value of wind direction and small ambient temperature may lead to larger wind power output.

Then the time lag i is changed and the results are analyzed. It is found that as the time lag increases, the difference between the wind power output in B_{i1} and B_{i2} gradually decreases. However, the relative numerical value of average wind power output in B_{i1} and B_{i2} for three meteorological factors, wind speed, wind direction and ambient temperature, maintains the result of time lag 1 when time lag is less than 20.

In conclusion, it is reasonable to consider historical environmental variables when constructing power curves for wind turbines. The historical data of wind speed, wind direction and ambient temperature will subsequently be used to establish the probabilistic power curve.

4. Density LSTM

In this section, a new probabilistic power curve named Density LSTM is introduced. The main idea is to combine the strength of LSTM with Bayesian framework: LSTM captures the long-and-short term dependencies within the SCADA data, and the Bayesian framework provides the future uncertainty information using predefined distributions. Two popular LSTM models, traditional LSTM and quantile LSTM, are introduced and compared with the new proposed Density LSTM with

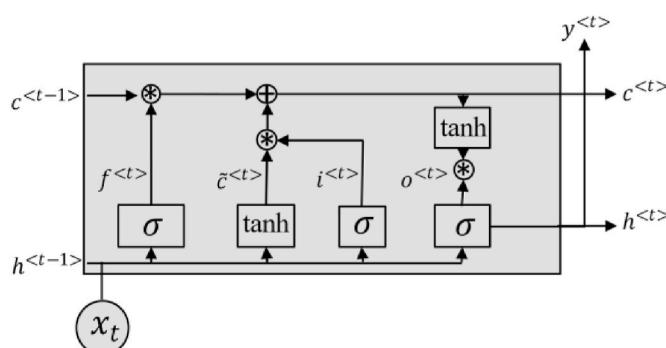


Fig. 12. The structure of LSTM.

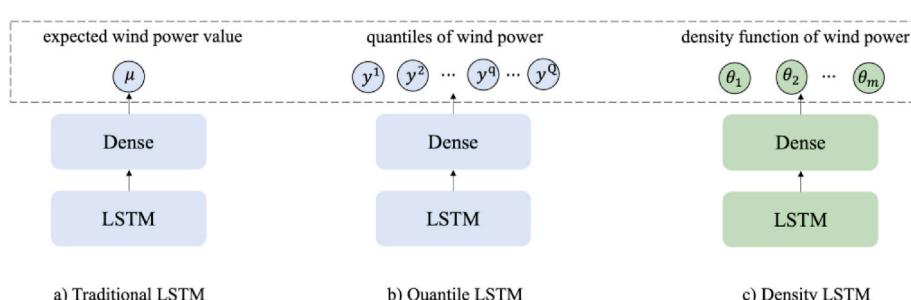


Fig. 13. Comparison of Density LSTM (c) and the other two LSTM (a, b).

Negative Log-Likelihood as loss function. Finally, the overall structure of Density LSTM is provided.

4.1. Traditional LSTM and quantile LSTM

Due to the complexity of input variables, a deep learning model with strong feature extraction ability is needed for time series and variable dimension information mining. LSTM [35] is a good choice for constructing the probabilistic power curve as the input variables are time series of meteorological factors. The structure of LSTM cell is shown in Fig. 12.

LSTM is widely used in time series prediction problems and is very effective because it adds the memory cell $c^{<t>}$ compared with Recursive Neural Network (RNN). LSTM can remember important historical information and forget unimportant information. LSTM has three gates: forget gate Γ_f , which deletes unnecessary information; update gate Γ_u , which is used to update $c^{<t>}$; output gate Γ_o , which is used to calculate the output $h^{<t>}$.

The formulas for Γ_f , Γ_u and Γ_o are shown in Eq. (10).

$$\begin{aligned} \Gamma_f &= \sigma(W_f[h^{<t-1>}, x^{<t>}] + b_f) \quad \Gamma_u = \sigma(W_u[h^{<t-1>}, x^{<t>}] + b_u) \Gamma_o \\ &= \sigma(W_o[h^{<t-1>}, x^{<t>}] + b_o) \end{aligned} \quad (10)$$

where W_f , W_u and W_o are corresponding weights of LSTM; b_f , b_u and b_o are corresponding biases of LSTM; $\sigma(\cdot)$ is the sigmoid activation function.

The updating formulas for memory cell $c^{<t>}$ are shown in Eq. (11).

$$\tilde{c}^{<t>} = \tanh(W_c[h^{<t-1>}, x^{<t>}] + b_c) \quad c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \quad (11)$$

where W_c and b_c are the weight and bias of $\tilde{c}^{<t>}$; the weights of $\tilde{c}^{<t>}$ and $c^{<t-1>}$ to calculate $c^{<t>}$ are controlled by Γ_u and Γ_f ; $\tanh(\cdot)$ is the activation function which equals to Eq. (12). $c^{<t>}$ records long term important information. $\tilde{c}^{<t>}$, Γ_u , and Γ_f are calculated when passing information over time t , which is used to determine whether to set $\tilde{c}^{<t>}$ as the newest value of $c^{<t>}$. This is the key of LSTM to solve gradient disappearance because if Γ_u is very close to 0, then $c^{<t>}$ is close to $c^{<t-1>}$. Even after many steps, the value of $c^{<t>}$ can be maintained well.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (12)$$

The final output $h^{<t>}$ of the LSTM cell is calculated as Eq. (13).

$$h^{<t>} = \Gamma_o * \tanh(c^{<t>}) \quad (13)$$

Traditional LSTM [35] and quantile LSTM [36] are two popular LSTM models with different loss function. The output of traditional LSTM is the expected wind power value as shown in Fig. 13(a). Quantile LSTM provides Q quantiles of wind power output and calculates a series of power and its corresponding probability according to the actual generation efficiency of the wind turbine, as shown in Fig. 13(b).

Loss function measures the difference between the forecast values and the actual data. It helps train the Neural Network and has a great influence on the prediction accuracy. A good loss function should fit the characteristics of the specific prediction task and the label data. For example, for regression problems, Mean Square Error (MSE) loss function is often used; for classification problems, cross entropy loss function is often used; for quantile prediction problems, quantile loss function is used.

The loss function of traditional LSTM is the MSE which is shown in Eq. (14). The implicit assumption of MSE is that wind power output follows a Gaussian distribution.

$$LMSE(y, f) = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2 \quad (14)$$

where y_i and \hat{y}_i represent the measured and predicted wind power

output at the time i respectively, and n is the number of data samples.

For quantile LSTM, pinball loss which is also called the quantile loss is used to guide the training process. The pinball loss is calculated as shown in Eq. (15). The pinball loss function imposes different degrees of penalties based on the quantile level and the fact that the estimated wind power is greater or less than the actual wind power.

$$L_{pinball}(y, f) = \frac{1}{nQ} \sum_i L_{q,i}(y_i - y_i^q) \quad L_{q,i}(y_i - y_i^q) = \begin{cases} (1-q)(y_i^q - y_i), & y_i^q \geq y_i \\ q(y_i - y_i^q), & y_i \geq y_i^q \end{cases} \quad (15)$$

where q represents the targeted quantile, y_i^q represents the estimated q -th quantile at the time t , and Q is the number of quantiles.

However, traditional LSTM only provides the expected value of wind power output and neglects the uncertainty of the wind power generation process due to the complex meteorological and operational conditions. Although quantile LSTM takes the future uncertainties into consideration and provides probabilistic predictions, it cannot give the concrete probability density and variance information of wind power output which is important for decision-making problems. To overcome the above problems, a new adaptive LSTM named Density LSTM is proposed in the next section.

4.2. Density LSTM with negative log-likelihood loss

Suppose given the current and historical environmental factors, the wind power output distribution is $F(\Theta)$ and $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$. The most common choice of F is Gaussian distribution with $\Theta = (\mu, \sigma)$, which are the mean and standard deviation of wind power output respectively. Besides Gaussian distribution, other probability distributions can also be considered. For example, paper [25] assumed that wind power output follows the Weibull distribution; paper [38] used Gaussian, Gamma and Laplace distribution to generate three individual probabilistic wind power forecasting models to form the ensemble forecasting framework. The probability distribution that fits the real scenario of the wind power generation process will lead to better prediction results. Their distribution parameters Θ and corresponding density functions are shown in Table 1.

Among the above 4 probability distributions, Gaussian and Laplace distributions are symmetric and Gamma and Weibull distributions are asymmetric. (1) Laplace distribution has a sharp peak and heavy tail compared with the Gaussian distribution. Its distribution parameters Θ contains μ and σ . The former characterizes the location (center) of the distribution and the latter characterizes the spread of the distribution. (2) Gamma distribution is the sum of independent and identically distributed exponential distributions. Its distribution parameters Θ contains α and β which are both positive values. The former is the concentration params of the distribution; the latter is the inverse scale params of the distribution. (3) Weibull distribution is often used to measure the life distribution of specific products. Its distribution parameters Θ contains β and η which are both positive values. The former is the concentration param of the distribution; the latter is the scale

Table 1
Distribution parameters and density functions of used probability distribution.

F	Distribution parameters Θ	Density function
Gaussian	μ, σ	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Laplace	μ, σ	$f(x) = \frac{1}{2\sigma} e^{-\frac{ x-\mu }{\sigma}}$
Gamma	α, β	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
Weibull	β, η	$f(x) = \frac{\beta}{\eta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^\beta}$

parameter of the distribution.

To provide more uncertainty information about wind power output and build a more precise probabilistic power curve, instead of forecasting the wind power or its quantiles, this paper proposes a new adaptive LSTM, Density LSTM, that forecasts the parameters of the probability density function of wind power directly. As shown in Fig. 13 (c). Unlike traditional LSTM or quantile LSTM, density LSTM predicted the distribution parameters of F , $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ directly. In this paper, four types of $F(\Theta)$ are considered: Gaussian, Gamma, Laplace and Weibull distributions.

Traditional loss functions, such as MSE and quantile loss, can only measure partial consistency of the predicted probability density functions and the actual wind power, and are not suitable to train the Density LSTM. Therefore, a more suitable loss function should be established to train the network. Negative log-likelihood measures the likelihood of raw wind power given a specific probability density function. Therefore, in this paper, negative log likelihood loss is used to evaluate the fitting performance of Density LSTM. The negative log likelihood loss function is defined in Eq. (16).

$$L_{NLL}(y, f) = -\frac{1}{n} \sum_i \log (P(y_i, f_i; \Theta_i)) \quad (16)$$

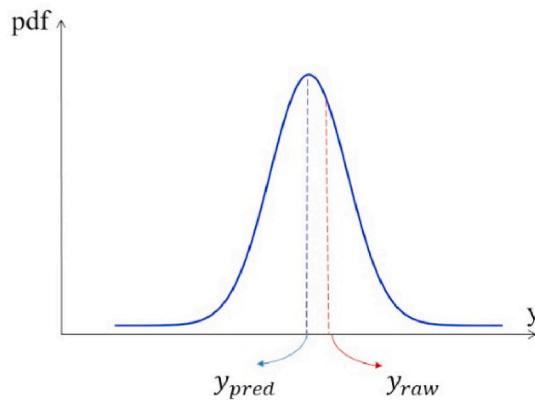
where y_i represents the actual wind power output; Θ_i is the predicted distribution parameter; f is the probability density distribution corre-

sponding to F ; and $P(y_i, f_i; \Theta_i)$ is the likelihood of y_i .

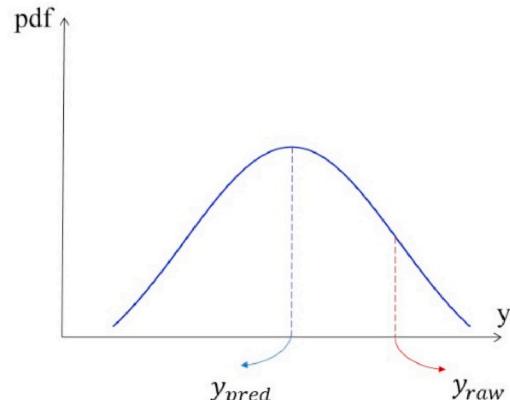
Based on the characteristics of the negative log-likelihood loss function, as the training process goes on, y_{pred} , the mean of probability density function, keeps approaching the actual wind power output y_{raw} . If the bias between y_{pred} and y_{raw} is small, to increase the likelihood of the actual wind power output, the variance of the probability density function must decrease. The output probability density function is more likely to have small bias and small variance, as shown in Fig. 14(a). On the contrary, if the bias between y_{pred} and y_{raw} is large, to increase the likelihood of the actual wind power output, the variance of the density function must increase. The output probability density function is more likely to have large bias and large variance, as shown in Fig. 14(b). Therefore, Density LSTM with negative log-likelihood loss function can generate density functions of wind power output with different variances which considers the existence of heteroscedasticity. In addition, the output density function contains the information of point prediction, quantile prediction and interval prediction. Therefore, it has more diverse application scenarios.

The structure of Density LSTM is a combination of LSTM, density function output and negative log-likelihood loss. The overall structure of Density LSTM is shown in Fig. 15.

Concretely, current and historical meteorological variables are inputs of Density LSTM. Input variables are firstly processed by stacked LSTM units according to Eq. 10–13, and the output is the hidden state



a) Small bias, small variance



b) Large bias, large variance

Fig. 14. Density function variance comparison of small and large bias.

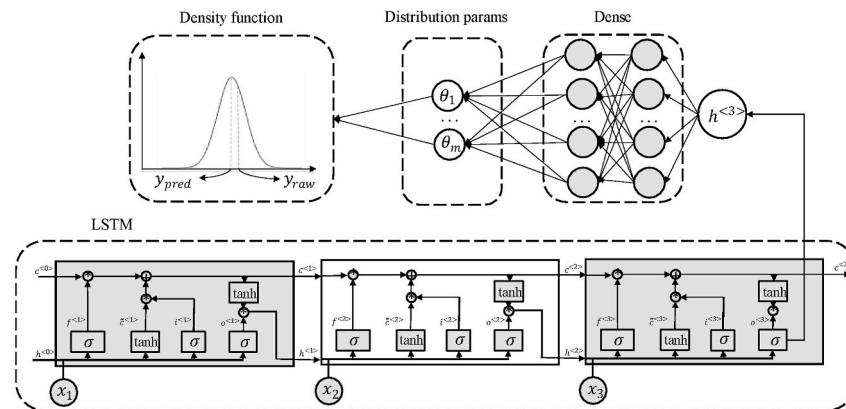


Fig. 15. Structure of overall networks.

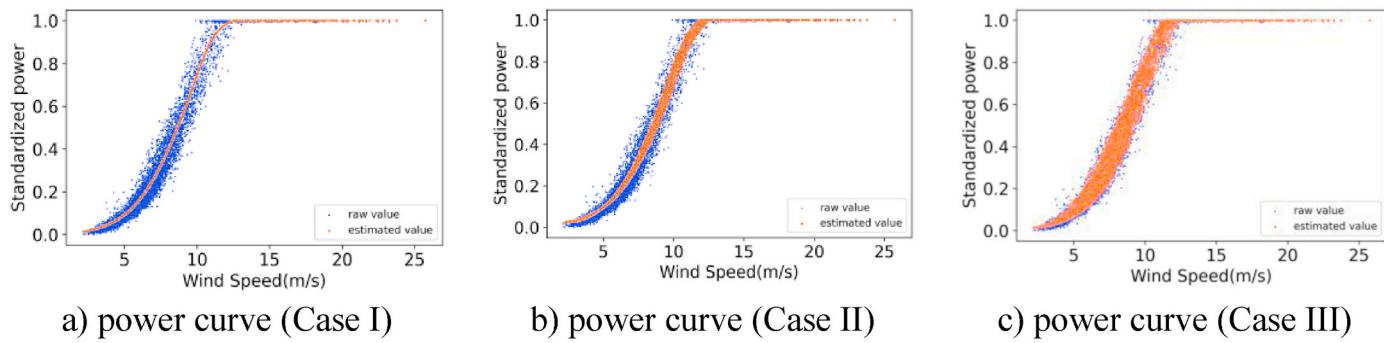


Fig. 16. Power curves with different inputs (F : Gaussian).

$h^{<\leftarrow}$. The output $h^{<\leftarrow}$ of LSTM is then passed to the Dense block to calculate the distribution parameters Θ . The dimension and physical meaning of Θ depend on the choice of the predefined probability distribution that fits the real characteristics of wind power output. The equations are shown in Eq. (17).

$$\theta_1 = g(W_1 z^{[n]} + b_1) \theta_m = g(W_m z^{[n]} + b_m) \quad (17)$$

where $z^{[n]}$ is the output of the last hidden layer in the Dense block; W_i is the weight matrix; b_i is the bias; $g(\cdot)$ is the activation function which helps satisfy the constraint that distribution parameters are positive values. After calculating Θ , the corresponding probability density function can be obtained according to the predefined probability distribution.

The negative log-likelihood loss function is then used to calculate the loss of the predicted density function according to Eq. (16). Then gradient descent algorithm is used to update the weights of the neural network based on the current loss value as shown in Eq. (18).

$$w_{t+1} = w_t - \alpha \frac{\partial \mathcal{R}_{\mathcal{D}}(w)}{\partial w} = w_t - \alpha \frac{1}{N} \sum_{i=1}^N \frac{\partial L_{NLL}(y_i, f_i; \theta)}{\partial w} \quad (18)$$

where \mathcal{D} is the training set; w_t is the weight value of t th iteration; α is the search step. The gradient descent algorithm requires calculating the partial derivative of the loss function concerning the current parameters. It's inefficient to calculate the partial derivative of each parameter one by one by the chain rule. Therefore, backpropagation is often used to compute gradients efficiently.

5. Performance evaluation

In this section, the performance of the newly proposed probabilistic power curve is evaluated. Firstly, evaluation criterions for probabilistic forecasts are introduced. Then the proposed Density LSTM and relevant baseline models are built on the python platform. Then, the impact of different inputs, different probability distributions F on the prediction performance of Density LSTM are evaluated. Finally, the forecasting performance of Density LSTM and relevant baseline models is compared.

5.1. Evaluation criteria for probabilistic forecasts

Reliability, sharpness, skill score and continuous ranked probability score (CRPS) are used to evaluate the probabilistic prediction results of wind power output. We randomly divide the inland and offshore wind farm dataset into a partition of 80% for training and 20% for testing, and use the test dataset to make an out-of-sample evaluation of the above four criteria. Given the wind power output P_t and $(1 - \beta)$ prediction interval $[\hat{q}_t^{(\beta/2)}, \hat{q}_t^{(1-\beta/2)}]$, the definitions of reliability, sharpness, skill score and CRPS are as follows.

Reliability is defined as Eq. (19). The closer the reliability index is to

zero, the closer the model is to the theoretical value of the algorithm. However, if the reliability index is greater than zero, the model produces a positive deviation, which is a favorable deviation.

$$Rel^{(1-\beta)} = \bar{\eta}^{(1-\beta)} - (1 - \beta) \quad (19)$$

where $\bar{\eta}^{(1-\beta)} = \frac{1}{T} \sum_{t=1}^T \eta_t^{(1-\beta)}$. When $\hat{q}_t^{(\beta/2)} < P_t < \hat{q}_t^{(1-\beta/2)}$, $\eta_t^{(1-\beta)} = 1$; otherwise, $\eta_t^{(1-\beta)} = 0$.

Sharpness is defined as Eq. (20). Sharpness is the average width of $(1 - \beta)$ predicted interval, which reflects the ability to predict interval to condense probability information. Under the same reliability, the smaller the sharpness, the better the prediction performance is.

$$Shp^{(1-\beta)} = \bar{\delta}^{(1-\beta)} = \frac{1}{T} \sum_{t=1}^T \delta_t^{(1-\beta)} \quad (20)$$

$$\text{where } \delta_t^{(1-\beta)} = \hat{q}_t^{(1-\beta/2)} - \hat{q}_t^{(\beta/2)}.$$

The mean skill score is defined as Eq. (21). SC_t reflects the prediction performance of all quantiles at time t . The larger SC_t is, the better the prediction interval is. The maximum value of MSC is 0.

$$MSC = \frac{1}{T} \sum_{t=1}^T SC_t = \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^m (\xi_t^{(\alpha_k)} - \alpha_k) (P_t - \hat{q}_t^{(\alpha_k)}) \quad (21)$$

where $\xi_t^{(\alpha_k)}$ is an indicator function: when $P_t \leq \hat{q}_t^{(\alpha_k)}$, $\xi_t^{(\alpha_k)} = 1$; otherwise, $\xi_t^{(\alpha_k)} = 0$.

CRPS is defined as Eq. (22). CRPS measures the difference between the predicted distribution and the real distribution which is zero when the predicted distribution is exactly the same as the real distribution. Prediction distributions that are too aggregated or too far away from the observed value will result in an increase in CRPS.

$$CRPS(F, x) = \int_{-\infty}^{\infty} [F(y) - \mathbf{1}\{y \geq x\}]^2 dy \quad (22)$$

where $\mathbf{1}\{y \geq x\}$ denotes a step function along the real line that attains the value 1 if $y \geq x$ and the value 0 otherwise.

5.2. Effect of different inputs and probability distributions

There are two main factors that influences the results of Density LSTM: different environmental inputs and the probability distribution F .

In this paper, to verify the importance of incorporating historical environmental data, the input of Density LSTM is varied in the following four ways:

Case I: current wind speed data; Case II: current wind speed, wind direction and ambient temperature data; Case III: current and historical wind speed, wind direction and ambient temperature data.

Case IV: adding historical turbulence intensity to the input variables in case III.

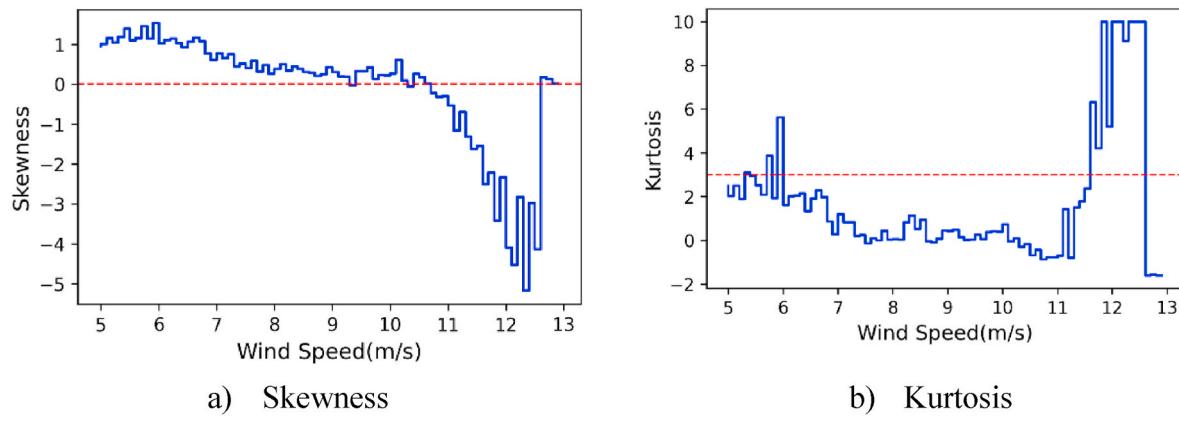


Fig. 17. Skewness and kurtosis of wind power output.

The prediction results of the first three different environmental inputs are analyzed by assuming wind power output to follow Gaussian distribution, as shown in Fig. 16. The blue points are the actual wind power values and the orange points are the estimated power values. The first power curve has a one-to-one mapping relationship between wind power generation and its measured wind speed; the second power curve shows a many-to-one mapping relationship between the three independent variables and wind power; the third power curve considers time information, which also shows a many-to-one relationship. It can be seen in Fig. 16 that,

- (i) Density LSTM with only wind speed as input has a good nonlinear fitting ability. However, the prediction performance is inferior to the other models, which shows that searching other appropriate variables as inputs except wind speed is important to improve the prediction ability and is particularly important for power curve modeling.
- (ii) The third power curve has wider and more practical power curve intervals than the second power curve, which reflects that the third power curve takes more possible wind turbine operating conditions into consideration.

Table 2
Partial AIC results of different probability distributions.

Speed Range	Gaussian	Gamma	Laplace	Weibull
[5.0, 5.1]	69.39	-57.21	-120.83	2.26
[6.0, 6.1]	998.71	354.84	281.99	599.12
[7.0, 7.1]	-99.55	-126.59	-117.24	-123.99
[8.0, 8.1]	-51.32	-55.09	-66.25	-52.20
[9.0, 9.1]	-18.16	-33.65	-41.54	-31.64
[10.0, 10.1]	-60.99	-54.65	-52.29	-68.72
[11.0, 11.1]	-167.86	-165.96	-148.69	-169.41
[12.0, 12.1]	-180.42	-216.39	-146.25	-227.99

Then, to provide the quantitative basis and reference for the selection of the probability distribution of Density LSTM's output, the distribution characteristics of wind power in specific wind speed bins are analyzed. The wind power distribution is irregular which makes it difficult to establish an accurate prediction model [39]. To the best of the authors' knowledge, most research on wind power distribution either directly specify a distribution [25] or use a multi-distribution ensemble model [38], or first give the distribution of wind speed and explore the distribution of wind power combined with the relationship between wind speed and wind power [40]. Few research directly studied the parameters of a specific distribution of wind power, like skewness and kurtosis, in wind speed bins. Although paper [26] divided wind speed bins and found that the practical distribution of power points within each small wind speed bin is similar to the Gaussian function, the width of their wind speed bin is 0.5 m/s, which is relatively crude. In this paper, as mentioned in Section 3, the width of the wind speed bin is 0.1 m/s, and the study of wind power distribution characteristics is more precise. Firstly, the skewness and kurtosis of wind power in different wind speed bins are calculated, as shown in Fig. 17. For Gaussian distribution, the skewness is 0 and the kurtosis is 3. It can be seen in Fig. 17 that as the wind speed increases, the distribution of wind power output has a thin tail with right skewness firstly, and then has a heavy tail with left skewness, which is not in accordance with Gaussian distribution. Secondly, the probability distribution F is varied among 4 commonly-used probability distributions: Gaussian, Gamma, Laplace and Weibull. Wind power data in different wind speed bins are fitted by Gaussian, Gamma, Laplace and Weibull distribution. Akaike Information Criterion (AIC) is used to measure the fitting performance. Partial fitting results are shown in Table 2 and Fig. 18. It can be seen that,

- (i) Variance of wind power output changes with wind speed.
- (ii) The other three probability distributions outperform Gaussian distribution, because wind power output in specific wind speed

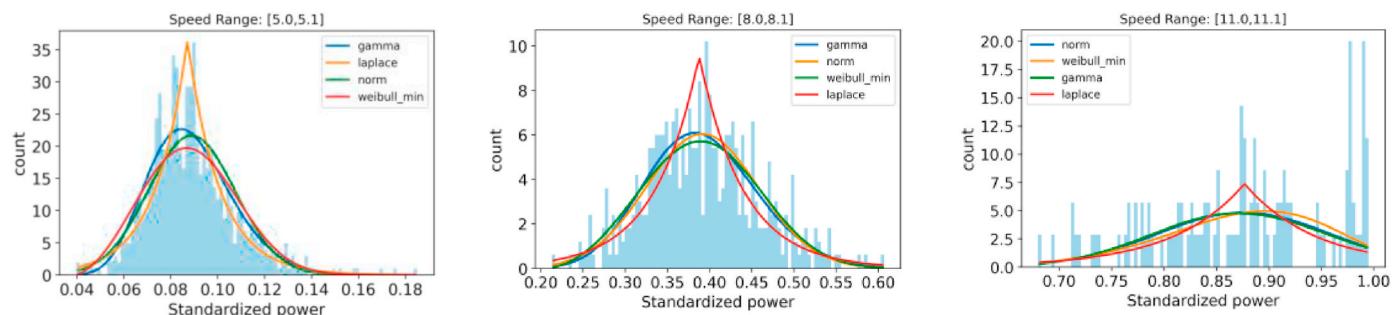


Fig. 18. Fitting results of different probability distributions.

Table 3Comparison of Density LSTM with different inputs and distribution F .

Input	Distribution F	Inland wind farm			Offshore wind farm		
		RMSE	MSC	CRPS	RMSE	MSC	CRPS
Case I	Gaussian	0.0572	-0.143	0.0292	0.0518	-0.130	0.0265
	Laplace	0.0547	-0.130	0.0266	0.0514	-0.124	0.0252
	Gamma	0.0546	-0.134	0.0273	0.0479	-0.116	0.237
	Weibull	0.0544	-0.131	0.0268	0.0477	-0.112	0.0228
Case II	Gaussian	0.0566	-0.141	0.0289	0.0462	-0.114	0.0233
	Laplace	0.0535	-0.127	0.0259	0.0448	-0.107	0.0218
	Gamma	0.0533	-0.129	0.0264	0.0451	-0.109	0.0224
	Weibull	0.0534	-0.129	0.0263	0.0450	-0.109	0.0221
Case III	Gaussian	0.0515	-0.125	0.0255	0.0396	-0.096	0.0194
	Laplace	0.0491	-0.118	0.0240	0.0366	-0.088	0.0179
	Gamma	0.0483	-0.114	0.0233	0.0367	-0.091	0.0185
	Weibull	0.0497	-0.121	0.0245	0.0381	-0.091	0.0184
Case IV	Gaussian	0.0509	-0.125	0.0254	0.0390	-0.097	0.0197
	Laplace	0.0502	-0.120	0.0245	0.0370	-0.089	0.0183
	Gamma	0.0483	-0.118	0.0239	0.0371	-0.093	0.0189
	Weibull	0.0490	-0.120	0.0245	0.0381	-0.091	0.0186

Table 4
Mean wind power for each turbine.

Turbine	T_1	T_2	T_3
Mean wind power (kW)	1193.2	1184	1272

bins is asymmetric and has a sharp peak and heavy tail, as shown in Fig. 17.

(iii) Wind power output is not exactly following Gaussian distribution given specific wind speed. It is necessary to compare the prediction performance of different probability distributions.

To simultaneously consider the effect of 4 different environmental inputs and 4 different probability distribution F , 16 different probabilistic power curves based on Density LSTM are established. The results of these power curves are compared in Table 3. Besides MSC and CRPS, Root Mean Square Error (RMSE) is used to evaluate the distance between raw wind power output and the corresponding median of predicted density functions. It can be seen that.

- (i) Given specific probability distribution, Density LSTM with input (2) has better prediction performance than that with input (1); that is because additional inputs differentiate wind turbine operation behaviors under different wind directions and ambient temperature conditions.
- (ii) Given specific probability distribution, Density LSTM with input (3) has the best prediction result for both the inland and offshore

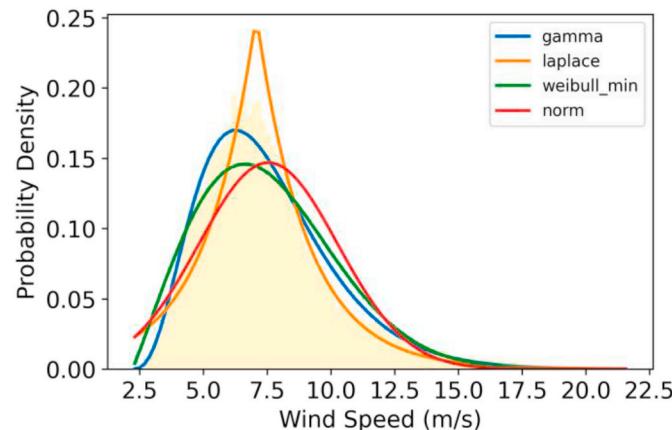
wind farms. That is because the wind turbine's historical operating status influences its current operation, which validates the necessity of considering temporal effect.

- (iii) Density LSTM with input (4) has inferior prediction performance compared with Density LSTM with input (3) for the test dataset. That is because historical turbulence intensity has no effect on wind power and only useful variables can improve the prediction performance for the test dataset.
- (iv) Given 3 different inputs, Laplace, Gamma and Weibull distributions all outperform Gaussian distribution in RMSE, MSC and CRPS. That indicates the wind power output does not follow the Gaussian distribution exactly and the above 3 distributions fits the wind power data better.
- (v) For inland wind farm, Gamma Density LSTM with input (3) has the best prediction accuracy. For offshore wind farm, Laplace Density LSTM with (3) has the best prediction performance.

5.3. Comparison with benchmark methods

As discussed in Section 5.2, using current and historical wind speed, wind direction and ambient temperature data as input is better than the other three inputs. Besides, Laplace, Gamma and Weibull distributions outperform Gaussian distribution. Therefore, for inland wind farm, Gamma Density LSTM with input (3) is used to compare with other methods; for offshore wind farm, Laplace Density LSTM with (3) is used to compare with other methods.

Popular probabilistic power curves proposed in recent years include

Fig. 19. Distribution fitting result of wind speed for T_1

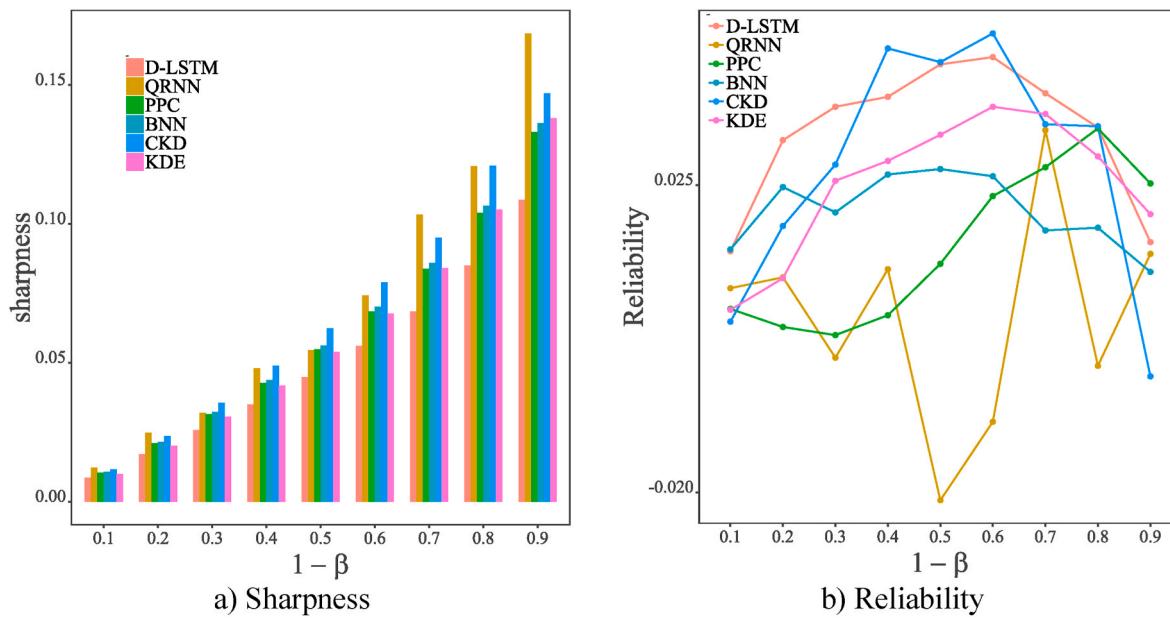


Fig. 20. Sharpness and reliability of relative methods.

KDE [12], Quantile Power Curve (QPC) [24], Probabilistic Power Curve (PPC) [25], Bayesian Neural Network (BNN) [37] and Conditional Kernel Density Estimation (CKD) [41] which is the basis of Lee and Ding's method [12]. Among them, KDE and PPC fit the probability density functions of wind power output given specific wind speed bins; QPC modified the output layer as Q quantiles based on the traditional neural network; BNN regularizes the weight of neural network by introducing uncertainty; CKD estimates the conditional density using multivariate input variables. The description of these methods and their specific parameter settings can be found in the appendix.

Firstly, these methods are used to predict the **annual energy production** (AEP) and the prediction precision is compared. Annual energy production (AEP) quantifies the wind energy potential of a given site. It is estimated by multiplying the total hours by the average wind turbine power [42], as shown in Eq. (23).

$$E_{AEP} = T \int_0^\infty P_{wt}(V)f_V(V)dV \quad (23)$$

in which T is the number of hours for one year; $P_{wt}(V)$ is the power curve

Table 5
Mean estimated AEP and corresponding estimation error.

Method	T_1	T_2	T_3	Overall
D-LSTM	1232.3 (0.032)	1192.9 (0.007)	1258.0 (0.011)	0.0168
QPC	1188.1 (0.004)	1046.7 (0.116)	1232.2 (0.031)	0.0503
PPC	1264.3 (0.060)	1232.2 (0.041)	1331.4 (0.046)	0.0488
BNN	1126.7 (0.056)	1301.6 (0.099)	1236.9 (0.027)	0.0607
CKD	1254.2 (0.051)	1224.1 (0.034)	1309.1 (0.029)	0.0380
KDE	1249.5 (0.047)	1218.3 (0.029)	1322.5 (0.039)	0.0383

Table 6
Estimated 5% and 95% quantiles of AEP and corresponding skill score.

Method	T_1		T_2		T_3		Skill Score
	944.5	1520.0	879.0	1506.8	912.9	1604.9	
QPC	706.5	1918.9	611.3	1729.0	777.8	1793.3	-86.13
PPC	975.6	1552.9	922.0	1544.0	1003.9	1658.9	-60.68
BNN	787.7	1465.7	998.8	1604.4	808.2	1665.6	-72.21
CDK	1020.3	1520.3	980.7	1518.9	1043.5	1611.0	-49.80
KDE	986.7	1587.6	934.8	1581.8	1013.9	1712.0	-55.95

Secondly, the wind power forecasting performance of Density LSTM (D-LSTM) and baselines is compared as shown in Fig. 20 and Table 7.

and $f_V(V)$ is the wind speed probability density function (PDF).

In this paper, we proposed a probabilistic power curve. Besides the expected value of AEP, confidence intervals of AEP can also be obtained.

$$Q_\alpha = T \int_0^\infty P_{wt}^\alpha(V)f_V(V)dV \quad Q_{1-\alpha} = T \int_0^\infty P_{wt}^{1-\alpha}(V)f_V(V)dV \quad (24)$$

where α is the degree of confidence; $P_{wt}^{1-\alpha}(V)$ and $P_{wt}^\alpha(V)$ are the upper and lower boundary of the interval; α is set 0.05 in this paper.

Three wind turbines are selected to verify the performance of the proposed probabilistic power curve. The mean wind power for 10-min by each turbine is calculated as shown in Table 4.

Then we use Weibull, Gamma, Laplace and Gaussian distribution to fit the wind speed data. The fitting results for wind turbine T_1 is shown in Fig. 19. It can be seen that Gamma distribution fits the wind speed data better than the other three distribution. Therefore, Gamma distribution is used to construct $f_V(V)$ of three wind turbines.

After getting the probability density function of wind speed, the

Table 7
Comparison of probabilistic prediction methods.

Method	Inland wind farm			Offshore wind farm		
	RMSE	MSC	CRPS	RMSE	MSC	CRPS
D-LSTM	0.0487	-0.114	0.0233	0.0366	-0.087	0.0178
QPC	0.0757	-0.178	0.0371	0.0581	-0.135	0.0279
PPC	0.0559	-0.134	0.0272	0.0482	-0.112	0.0228
BNN	0.054	-0.136	0.0278	0.0433	-0.103	0.0208
CKD	0.074	-0.175	0.0360	0.0581	-0.129	0.0265
KDE	0.055	-0.132	0.0271	0.0466	-0.107	0.0221

mean and quantiles of AEP can be calculated through the monte carlo simulation method. The results can be seen in [Table 5](#) and [Table 6](#) respectively. It can be seen that the relative estimated error of Density LSTM is smaller than other probabilistic power curve and the skill score of Density LSTM is more closer to 0. Therefore, the proposed probabilistic power curve can improve the prediction precision of AEP compared with traditional power curve.

The sharpness and reliability of relative methods are shown in [Fig. 20](#). It can be seen that the prediction interval given by Density LSTM is narrower than that of other methods, as shown in [Fig. 20\(a\)](#). Meanwhile, the reliability of Density LSTM is higher than other methods except in the case of 90% coverage as shown in [Fig. 20\(b\)](#). Low sharpness and high reliability indicate that Density LSTM outperforms other methods.

RMSE, MSC and CRPS of the above models are shown in [Table 7](#). It can be seen that RMSE, MSC and CRPS of Density LSTM are greatly improved compared with other models in both the inland and offshore wind farm datasets, which illustrates the superior performance of Density LSTM for probabilistic wind power output forecasting. In terms of RMSE for median prediction results, Density LSTM is the lowest. For MSC, Density LSTM is closer to 0. Meanwhile, in terms of CRPS, Density LSTM is also superior to other models.

[Figs. 21 and 22](#) show the density prediction results of Gamma Density LSTM and Laplace Density LSTM for the inland wind farms respectively. The red dotted line denotes the actual wind power output and the blue curve denotes the density prediction functions. It can be seen that almost all the actual values are located in the middle of probability density curves, which means the raw wind power appears in

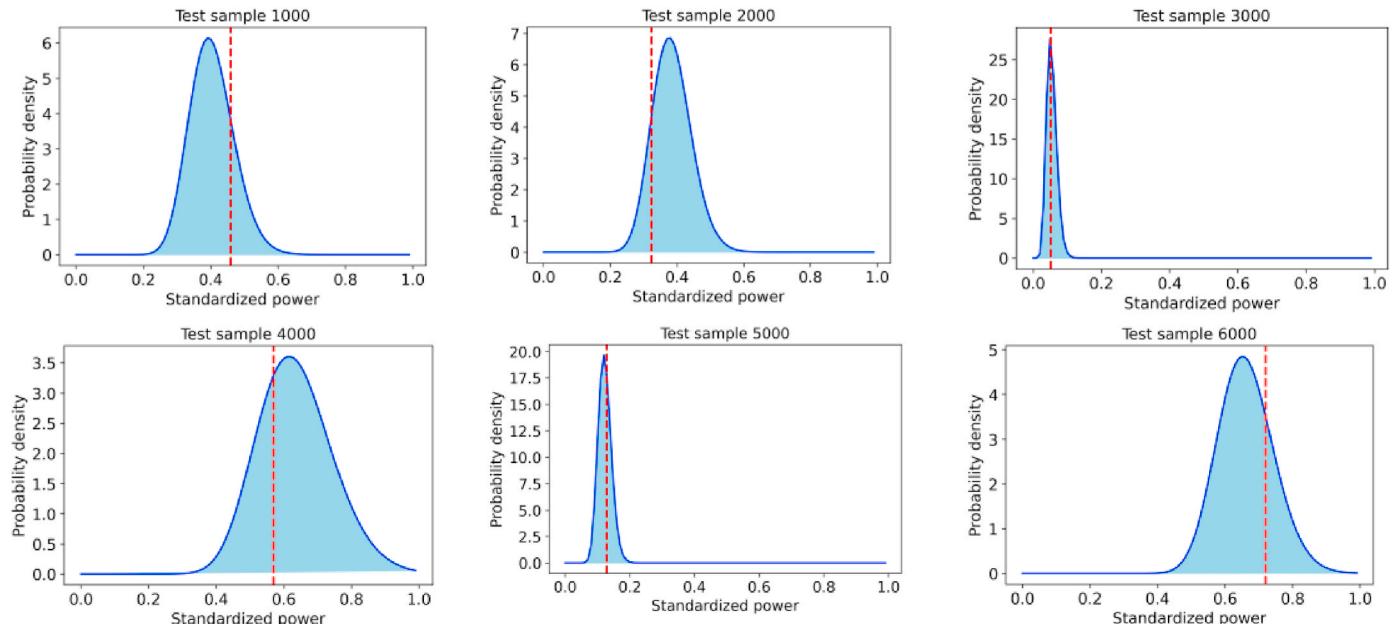


Fig. 21. Prediction results of Gamma Density LSTM.

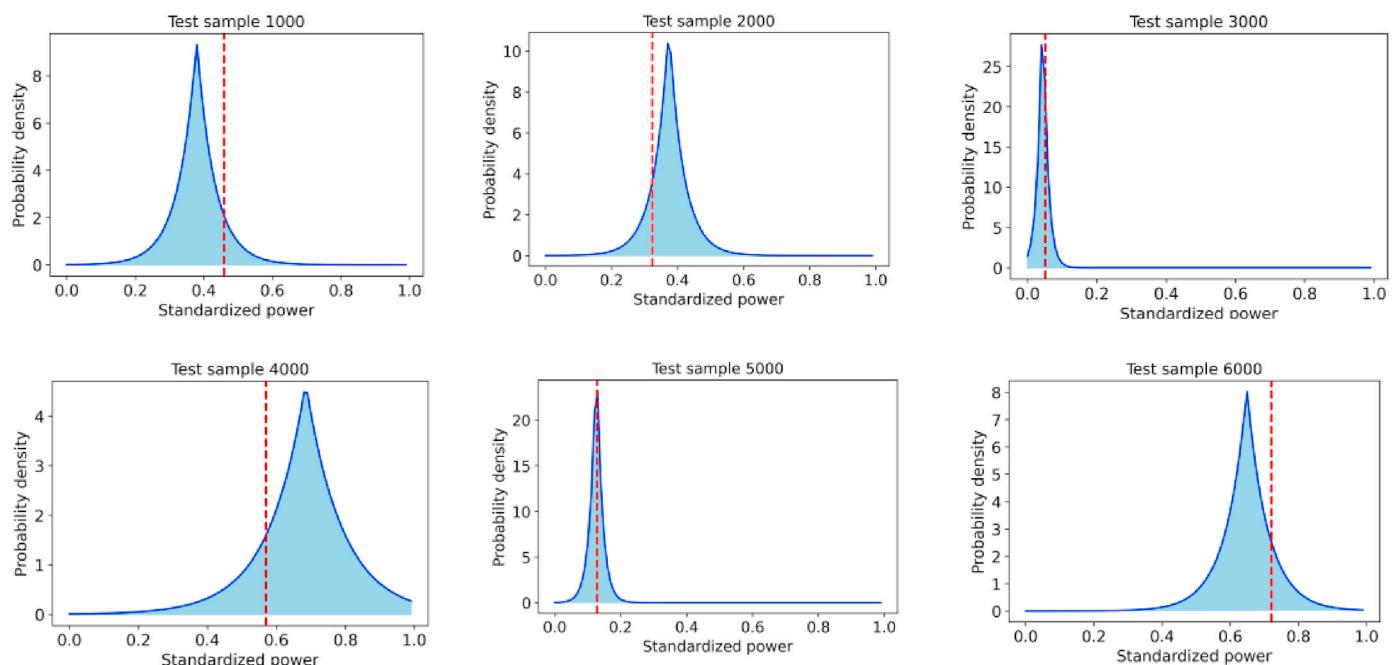


Fig. 22. Prediction results of Laplace Density LSTM.

the forecasting distributions with high probability. These figures provide complete probability descriptions of the future wind power output which show the advantages of the probability density forecasting method in quantifying the uncertainty and improving the prediction accuracy.

6. Conclusion

This paper proposes a novel probabilistic power curve, named Density LSTM, whose input variables are current and historical wind speed, wind direction and ambient temperature, and output is the probability density function of wind power output. The prediction results of relative models are compared in both inland and offshore wind farms. The main results are the followings.

- 1) Two variables, “operating state” and “active power set value”, are used to remove data samples that are under an abnormal working state or under a limited power state effectively.
- 2) Adding more useful meteorological factors and historical data can help improve the prediction performance of the power curve for the test dataset compared with only using wind speed as input.
- 3) The negative log-likelihood loss function is used to train Density LSTM. Compared with traditional methods, the probabilistic prediction results of the proposed method improve obviously.
- 4) The wind power output does not exactly follow Gaussian distribution given specific wind speed. Laplace and Gamma distribution outperform Gaussian distribution in probabilistic prediction tasks.

In the future, the proposed probabilistic power curve can be

Appendix

1. Quantile power curve

Quantile power curve (QPC) is a variant of traditional neural network. At the given input, QPC generates many power values according to the corresponding confidence level rather than one value in a deterministic power curve. The meaning of the quantile q is that at a certain wind speed, the ratio of the power below p_q to all the data points within this wind speed bin is q , as shown in Eq. (25).

$$\text{Prob}(p \leq p_q, v = v_q) = q \quad (25)$$

where Prob is the probability function; p is the wind power output generated by a wind turbine at a given wind speed v_q ; p_q represents the q quantile of the power distribution at wind speed v_q . The structure of the quantile power curve is shown in Fig. 23.

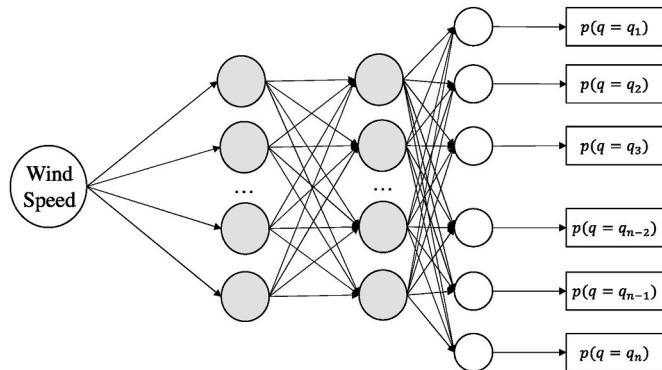


Fig. 23. Framework of quantile power curve

The model parameters of the compared quantile power curve are shown in Table 8.

improved in the following directions. Firstly, nonparametric methods can be used to obtain the empirical probability density functions of the wind power without specifying the distribution type in advance. Secondly, besides temporal effect, spatial effect of different wind turbines can be studied which may also improve the accuracy of wind power curve. Thirdly, besides negative log-likelihood loss function, other loss functions can be studied to train the power curve.

Author contribution

Peng Wang: Methodology, Writing – original draft, Software, Visualization.Yanting Li: Conceptualization, Writing – review & editing, fund acquisition Guangyao Zhang: Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This project is funded by the National Natural Science Foundation of China. (72072114).

Table 8
Model parameters of the compared quantile power curve

Number of input layer neurons	16
Number of hidden layers	5
Number of neurons in each layer of the hidden layer	256, 256, 128, 64, 32
Number of output layer neurons (quantile levels)	9
Loss function	Quantile loss
Activation function of hidden layers	RuLU
Activation function of output layer	Linear
Model optimizer	Adam

2. Probabilistic power curves

There are two steps to construct the probabilistic power curve (PPC). In the first step, a database for performing the probability distribution fit is constructed using the measured wind speed and wind power output. The wind speed was divided into regular intervals to filter the output data according to the wind speed interval. In each wind speed interval, the Weibull distribution is used to represent the output distribution of wind power. The probability density function of the Weibull distribution is as follows.

$$f(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} e^{-\left(\frac{v}{c}\right)^k} \quad (26)$$

where k is a shape parameter, and c is a scale parameter. Then, Maximum Likelihood Method (MLM) is used to estimate the parameters of the Weibull distribution.

In the second step, Monte Carlo Simulation (MSC) is performed to generate accidental values using random sampling through simulations after calculating the parameters from the Weibull distribution. The randomized data is used to construct the quantiles of the power curve.

The model parameters of compared PPC are as follows. The wind speed interval width is 0.1 m/s and the number of randomized data is 100,000.

3. Conditional kernel density estimation

Conditional kernel density (CKD) estimation can be used for estimating the conditional density, $p(y|x)$. Specifically, the density of y conditional on x can be expressed as

$$\hat{f}(y|x) = \sum_{i=1}^N \omega_i(x) \mathcal{K}_{h_y}(y - y_i) \quad (27)$$

where

$$\omega_i(x) = \frac{\mathcal{K}_{h_x}(\|x - x_i\|)}{\sum_{i=1}^N \mathcal{K}_{h_x}(\|x - x_i\|)} \quad (28)$$

where $h_x = (h_1, \dots, h_q)$ and h_y are bandwidth parameters controlling the smoothness in, respectively, the environmental factors, x , and the power output, y , and q is the number of explanatory variables in x . \mathcal{K}_{h_y} is a scaled kernel function and takes the form of $h_y^{-1} K(l/h_y)$, where $K(\bullet)$ is kernel function. $\mathcal{K}_{h_x}(\|l\|)$ is a multivariate kernel function and is composed of a product kernel that is a multiplication of univariate kernel functions, such as:

$$\mathcal{K}_{h_x}(\|l\|) = \mathcal{K}_{h_1}(l_1) \mathcal{K}_{h_2}(l_2) \dots \mathcal{K}_{h_q}(l_q) \quad (29)$$

In addition, in order to reduce the burden of computation, K-Nearest Neighbor (KNN) is used to select data points when estimating the conditional density.

The model parameters of compared CKD are as follows. The number of explanatory variables q is 16; Kernel function K is chosen to be a univariate Gaussian kernel, $K(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$; the number of neighbors k for KNN is 10.

4. Bayesian neural networks

Different from backpropagation neural network, The Bayesian neural network regards the weight as obeying the Gaussian distribution whose mean value is μ and variance is σ^2 , and each weight obeys a different Gaussian distribution. During training, the Bayesian neural network optimizes the mean value and variance of the weight.

Bayesian inference for neural networks calculates the posterior distribution of the weights given the training data, $p(w|\mathcal{D})$. This distribution answers predictive queries about unseen data by taking expectations: the predictive distribution of an unknown label \hat{y} of a test data item \hat{x} , is given by $P(\hat{y}|\hat{x}) = E_{p(w|\mathcal{D})}[P(\hat{y}|\hat{x}, w)]$. Variational learning is used to find the parameters θ of a distribution on the weight $q(w|\theta)$ that minimises the Kullback-Leibler (KL) divergence with the true Bayesian posterior on the weights:

$$\theta^* = \arg \min_{\theta} KL[q(w|\theta) \| P(w|\mathcal{D})] \quad (30)$$

The model parameters of compared BNN are shown in Table 9.

Table 9
Model parameters of the compared BNN

Number of input layer neurons	16
Number of hidden layers	2
Number of neurons in each layer of the hidden layer	8, 8
Output layer	Independent Normal
Loss function	Negative Log Likelihood
Model optimizer	RMSprop

5. Kernel density estimation

The procedure of KDE is similar to that of PPC. The difference is that kernel density is used to represent the output distribution of wind power rather than Weibull distribution.

The model parameters of compared PPC are as follows. The wind speed interval width is 0.1 m/s, the kernel function is Gaussian kernel, the bandwidth is 1.06 and the number of randomized data is 100,000.

References

- [1] Global Wind Energy Council. Global wind report. 2022.
- [2] Mehrjoo M, Jozani MJ, Pawlak M, et al. A multilevel modeling approach towards wind farm aggregated power curve[J]. IEEE Trans Sustain Energy 2021;12(4): 2230–7.
- [3] Yang M, Shi C, Liu H. Day-ahead wind power forecasting based on the clustering of equivalent power curves[J]. Energy 2021;218:119515.
- [4] Wacker B, Seebaß JV, Schlüter JC. A modular framework for estimating annual averaged power output generation of wind turbines[J]. Energy Convers Manag 2020;221:113149.
- [5] Guo P, Infield D. Wind turbine power curve modeling and monitoring with Gaussian process and SPRT[J]. IEEE Trans Sustain Energy 2018;11(1):107–15.
- [6] Helbing G, Ritter M. Improving wind turbine power curve monitoring with standardization [J]. Renew Energy 2020;145:1040–8.
- [7] Hu Y, Xi Y, Pan C, et al. Daily condition monitoring of grid-connected wind turbine via high-fidelity power curve and its comprehensive rating[J]. Renew Energy 2020;146:2095–111.
- [8] Cai Y, Breon FM. Wind power potential and intermittency issues in the context of climate change[J]. Energy Convers Manag 2021;240:114276.
- [9] Wang Y, Hu Q, Li L, et al. Approaches to wind power curve modeling: a review and discussion. Renew Sustain Energy Rev 2019;116:109422.
- [10] Lydia M, Kumar SS, Selvakumar, et al. A comprehensive review on wind turbine power curve modeling techniques. Renew Sustain Energy Rev 2014;30:452–60.
- [11] Carrillo C, Montaño AFO, Cidras J, et al. Review of power curve modelling for wind turbines. Renew Sustain Energy Rev 2013;21:572–81.
- [12] Lee G, Ding Y, Genton MG, et al. Power curve estimation with multivariate environmental factors for inland and offshore wind farms. J Am Stat Assoc 2015; 110(509):56–67.
- [13] Saint-Drenan YM, Besseau R, Jansen M, et al. A parametric model for wind turbine power curves incorporating environmental conditions[J]. Renew Energy 2020; 157:754–68.
- [14] Mehrjoo M, Jozani MJ, Pawlak M. Wind turbine power curve modeling for reliable power prediction using monotonic regression[J]. Renew Energy 2020;147:214–22.
- [15] Han S, Qiao Y, Yan P, et al. Wind turbine power curve modeling based on interval extreme probability density for the integration of renewable energies and electric vehicles[J]. Renew Energy 2020;157:190–203.
- [16] Karamichailidou D, Kaloutsis V, Alexandridis A. Wind turbine power curve modeling using radial basis function neural networks and tabu search[J]. Renew Energy 2021;163:2137–52.
- [17] Li T, Liu X, Lin Z, et al. Ensemble offshore wind turbine power curve modeling—an integration of isolation forest, fast radial basis function neural network, and metaheuristic algorithm. J Energy 2022;239:122340.
- [18] Wang Y, Hu Q, Pei S. Wind power curve modeling with asymmetric error distribution[J]. IEEE Trans Sustain Energy 2019;11(3):1199–209.
- [19] Wei D, Wang J, Li Z, et al. Wind power curve modeling with hybrid copula and grey wolf optimization[J]. IEEE Trans Sustain Energy 2021;13(1):265–76.
- [20] Wang Y, Li Y, Zou R, et al. Sparse heteroscedastic multiple spline regression models for wind turbine power curve modeling[J]. IEEE Trans Sustain Energy 2020;12(1): 191–201.
- [21] Zou R, Yang J, Wang Y, et al. Wind turbine power curve modeling using an asymmetric error characteristic-based loss function and a hybrid intelligent optimizer[J]. Appl Energy 2021;304:117707.
- [22] He Y, Li H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation[J]. Energy Convers Manag 2018; 164:374–84.
- [23] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renew Sustain Energy Rev 2014;32:255–70.
- [24] Xu K, Yan J, Zhang H, et al. Quantile based probabilistic wind turbine power curve model[J]. Appl Energy 2021;296:116913.
- [25] Yun E, Hur J. Probabilistic estimation model of power curve to enhance power output forecasting of wind generating resources. J Energy 2021;223:120000.
- [26] Yan J, Zhang H, Liu Q, et al. Uncertainty estimation for wind energy conversion by probabilistic wind turbine power curve modeling. Appl Energy 2019;239:1356–70.
- [27] Virgolino G, Mattos C, Magalhes J, et al. Gaussian processes with logistic mean function for modeling wind turbine power curves[J]. Renewable Energy; 2020. p. 162.
- [28] Rogers TJ, Gardner P, Dervilis N, et al. Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression[J]. Renew Energy 2020;148:1124–36.
- [29] Luo Z, Fang C, Liu C, et al. Method for cleaning abnormal data of wind turbine power curve based on density clustering and boundary extraction[J]. IEEE Trans Sustain Energy 2021;13(2):1147–59.
- [30] Morrison R, Liu X, Lin Z. Anomaly detection in wind turbine SCADA data for power curve cleaning[J]. Renew Energy 2022;184:473–86.
- [31] Long H, Xu S, Gu W. An abnormal wind turbine data cleaning algorithm based on color space conversion and image feature detection[J]. Appl Energy 2022;311: 118594.
- [32] Liang G, Su Y, Chen F, et al. Wind power curve data cleaning by image thresholding based on class uncertainty and shape dissimilarity[J]. IEEE Trans Sustain Energy 2020;12(2):1383–93.
- [33] Long H, Sang L, Wu Z, et al. Image-based abnormal data detection and cleaning algorithm via wind power curve[J]. IEEE Trans Sustain Energy 2019;11(2): 938–46.
- [34] Astolfi D. Perspectives on SCADA data analysis methods for multivariate wind turbine power curve modeling. J. Machines 2021;9(5):100.
- [35] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Comput 1997;9 (8):1735–80.
- [36] Yi W, Gan D, Sun M, et al. Probabilistic individual load forecasting using pinball loss guided LSTM[J]. Appl Energy 2019;235:10–20.
- [37] Blundell C, Cornebise J, Kavukcuoglu K, et al. Weight uncertainty in neural network[C]//International conference on machine learning. PMLR 2015:1613–22.
- [38] Sun M, Feng C, Zhang J. Multi-distribution ensemble probabilistic wind power forecasting[J]. Renewable Energy; 2020. p. 148.
- [39] Yuan K, Zhang K, Zheng Y, et al. Irregular distribution of wind power prediction [J]. Journal of Modern Power Systems and Clean Energy 2018;6(6):1172–80.
- [40] Villanueva D, Feijóo A. Wind power distributions: a review of their applications[J]. Renew Sustain Energy Rev 2010;14(5):1490–5.
- [41] Rosenblatt M. Conditional probability density and regression estimators[J]. Multivariate analysis II 1969;25:31.
- [42] Jung S, Vanli OA, Kwon SD. Wind energy potential assessment considering the uncertainties due to limited data[J]. Appl Energy 2013;102:1492–503.