Renewable Energy 205 (2023) 598–607

Contents lists available at ScienceDirect

Renewable Energy

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

# Ultra-short-term wind power forecasting based on deep Bayesian model with uncertainty

Lei Liu [a,b], Jicheng Liu [a], Yu Ye [a], Hui Liu [a], Kun Chen [a], Dong Li [a,b], Xue Dong [c], Mingzhai Sun [a,*]

[a] School of Biomedical Engineering & Suzhou Institute For Advanced Research, University of Science and Technology of China, Suzhou, 215123, China
[b] School of Information Science and Technology, University of Science and Technology of China, Hefei, 230026, China
[c] Key Laboratory of Far-Shore Wind Power Technology of Zhejiang Province, Hangzhou, 311122, China

ARTICLE INFO

*Keywords:*
Wind power forecasting
Deep learning
Bayesian
Uncertainty
BiGRU
Attention

ABSTRACT

Wind energy is an important renewable clean energy resource. However, the stochastic and volatile nature of wind power brings significant challenges to the power system's reliable and secure operation. Accurate and reliable wind power prediction is critical for the integration of wind power into the grid. The existing wind power forecasting (WPF) methods lack an assessment of the reliability of the predicted results, which may result in a financial penalty for the wind energy producers. An accurate prediction with reliability measurement is urgently needed to encounter the intricate nature of the problem. In this paper, a Bayesian framework-based bidirectional gated logic unit (BiGRU) method was proposed for ultra-short-term wind power forecasting. First, an encoder-decoder (ED) architecture was combined with a BiGRU time series modeling and feature–temporal attention (FT-Attention) to improve the accuracy of wind power prediction. Then, two uncertainty losses were applied to improve the model's performance further. The proposed method obtains the uncertainty of forecast results, which effectively eliminates the untrusted results. Our proposed method demonstrated promising results for ultra-short-term wind power forecasting due to its competitive performance compared with traditional forecasting methods.

## 1. Introduction

In recent years, as renewable energy, wind power increased rapidly [1–3] due to its outstanding advantages in environmental protection. With the large-scale adoption of wind power generation systems, its intermittency and fluctuation bring significant challenges to the safety and stability of the power grid [4,5]. Therefore, accurate, reliable, and timely wind power forecasting is of great significance to the power planning and safe operation of the power grid [6,7].

There are mainly two categories of wind power prediction methods: physical and statistical methods [8,9]. Based on the numerical weather forecast (NWP), the physical method describes the physical process of converting wind energy to electric energy. It uses the wind farm's meteorological information and geographical information, together with its surrounding conditions, to forecast wind power [10,11]. The physical model requires rich physical information, and the calculation is complex and time-consuming, making it unsuitable for short-term wind power prediction [12,13]. Using a data-driven approach, statistical methods fit the mapping relationship between historical data and wind power by establishing mathematical functions [14]. Compared with solving physical equations, statistical methods are generally faster and more accurate [15], becoming more and more popular [11].

Traditional machine learning and deep learning methods are also applied in wind power prediction. Demolli et al. [16] used daily wind speed data for wind power prediction and comprehensively compared five machine learning algorithms (Lasso, KNN, XGBoost, SVM, and RF). Chen et al. [17] carried out wind power prediction based on a hybrid model of multi-resolution multi-learning integration (MRMLE) and adaptive model selection (AMS), and combined the selected sub-models with support vector machine (SVM) to get the final prediction results. Yu et al. [18] proposed a hybrid wind power prediction model combining improved wavelet transform and Elman neural network. Wang et al. [19] used the wavelet packet decomposition technology to process the original wind speed sequence. They combined it with the support vector machine (SVM) to improve the model's prediction ability.

As an essential branch of statistical methods, deep learning methods have become popular in wind power prediction due to their powerful feature extraction and nonlinear mapping capabilities. In [20], the wind power prediction was carried out by combining the secondary decomposition (SD) and bidirectional gated logic unit (BIGRU). The chicken colony optimization algorithm was further adopted for the

* Corresponding author.
*E-mail address:* mingzhai@ustc.edu.cn (M. Sun).

https://doi.org/10.1016/j.renene.2023.01.038
Received 26 February 2022; Received in revised form 8 January 2023; Accepted 9 January 2023
Available online 11 January 2023

joint optimization of network parameters. A new hybrid deep learning model was proposed by Yan et al. [21], which used Long Short-Term Memory (LSTM) to predict the power of low-frequency subsequences and built a deep confidence network to predict the power of high-frequency subsequences. Ju et al. [6] proposed a prediction model based on Convolutional Neural Network and LightgBM (Light Gradient Boosting Machine). A wind power interval prediction model based on LSTM was proposed in [22]. The model can obtain the upper and lower boundary estimates of power while predicting the wind power output and support the diversification decision through the predicted power interval errors. Ishii et al. [23] compared deep learning methods with various traditional methods and found that deep learning methods have better learning and expression ability than conventional statistical methods and shallow neural networks.

Although much research on wind power prediction has been conducted, there are still two important problems: (1) insufficient accuracy caused by the randomness and volatility of wind speed. (2) As far as we know, the existing methods only provide wind power estimates or its forecast intervals without the uncertainty measure of the forecast results. Researchers have paid uncertainty more attention since it can provide reliable support for comprehensive decision-making. Eaton-Rosen et al. [24] used a Bayesian neural network to quantify the uncertainty of semantic segmentation results, improving the model's safety for application in the medical field. In [25], the uncertainty was used to evaluate the reliability of liver lesion detection results, and 90% of false positives were effectively filtered out. Ghoshal et al. [26] showed that uncertainty could improve the model's performance using the chest X-ray data set training model for COVID-19 prediction. The prediction uncertainty has a strong correlation with the model's accuracy. In wind power forecasting, the introduction of uncertainty is beneficial to eliminating inaccurate power prediction, reducing the risk of grid connection and economic losses. It has crucial strategic and financial significance for power grid planning, comprehensive dispatching, and the further development of the wind power industry.

To solve the two problems above, we proposed a Bayesian framework-based BIGRU ultra-short-term wind power forecast. Encoder-Decoder (ED) architecture is combined with BiGRU time series modeling and feature-temporal attention (FT-Attention) to improve the accuracy of wind power prediction. At the same time, Bayesian theory is integrated into the deep learning model to obtain the uncertainty of forecast results, which effectively eliminates the untrusted results. Furthermore, two kinds of uncertainty losses were applied for training to further improve the model's performance. The experimental results show that the proposed methods improve the performance of wind power prediction and reject the unreliable results. Combining with uncertainty will improve the qualified rate of wind farm power reporting, which is important in enhancing wind power's economy and competitiveness in the new energy market. The main contributions of this paper are summarized as follows:

(1) Combining Bayesian theory with ultra-short-term wind power prediction to obtain the wind power prediction and its uncertainty. To the best of our knowledge, this is the first time that uncertainty has been used in wind power forecasting to measure the reliability of forecast results. The uncertainty effectively eliminates untrusted power and improves the qualified rate of wind farm power reporting.
(2) Introducing the aleatoric uncertainty loss and epistemic (model) uncertainty loss during training to reduce the adverse effects on the model's performance caused by the inherent noise of observation data, incomplete training samples, and an imperfect model.
(3) The combination of Encoder-Decoder architecture, BiGRU time series modeling, and feature-temporal attention (FT-Attention) mechanism improves the capability of the deep learning model on feature selection and representation.
(4) We proposed a comprehensive solution for ultra-short-term wind power prediction, effectively improving the model's accuracy. Obtaining the prediction results and uncertainty simultaneously is of great significance to ensure the safety of grid connection, the economic benefit of the wind farm, and the integration of power dispatching.

The remaining part of this study is organized as follows. Section 2 presents the details of the proposed method. Section 3 presents the dataset, feature selection, and preprocessing methods. Section 4 presents the main results with discussions. Section 5 presents the main conclusions of the study.

## 2. Methods

In this paper, we combined the encoder-decoder architecture with Bayesian Neural Network. The BiGRU was used to build an encoder and decoder and model time-series signals, which can effectively extract information dependencies and correlations between time-series data. The Feature-Temporal Attention module was added between the encoder and decoder modules to strengthen vital information and weaken the redundant information in terms of temporal and feature dimensions. The feature combination selected by the attention module was used as the input of the decoder to improve the prediction performance of the decoder. The Bayesian neural network was applied to obtain both the prediction and the uncertainty of prediction simultaneously. In addition, aleatoric uncertainty loss and epistemic uncertainty loss were added to the loss function to increase the performance of the model.

### 2.1. BiGRU

Long short-term memory networks (LSTM) [27] and logic gated units (GRU) [28] have been proposed to overcome the problem of long-term memory loss in recurrent neural networks. GRU simplifies the gated unit of the LSTM logical unit, which is more efficient with a lower storage footprint.

Fig. 1 (a) shows the structure of the GRU, $x_t$ and $y_t$ are the input and output at time t, $h_{t-1}$ and $h_t$ represent the hidden state at time t-1 and t, respectively, and $\sigma$ refers to the activation function (sigmoid). W is the weight coefficients, and $\odot$ is an element-wise multiplication. Similar to the LSTM unit, GRU is only composed of an update gate and a reset gate, without having a separate memory cells. $z_t$ represents a set of reset gates, GRU uses the reset gate to choose which information to abandon at the previous moment, where $W_{xr}$ and $W_{hr}$ are weight information, and $h_{t-1}$ is the input of the previous moment, $b_r$ is the bias. $z_t$ is the update gate. GRU selects and updates which information at the current moment through the update gate, where $W_{xz}$ and $W_{hz}$ are weight information, and $h_{t-1}$ is the input of the previous moment, $b_z$ is the bias.

$$z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \tag{1}$$

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{2}$$

$$\tilde{h}_t = \tanh(W_{xh}x_t + W_{hh}(r_t \odot h_{t-1}) + b_r) \tag{3}$$

$$h_t = (1 - z_t)h_{t-1} + z_t\tilde{h}_t \tag{4}$$

The updated BiGRU network comprises two GRU networks in opposite directions as shown in Fig. 1(b), a typical BiGRU network with n GRU units. The input of BiGRU is a timing signal of length n ($X = [x_1, x_2, \ldots, x_n]$), and $Y = [y_1, y_2, \ldots, y_n]$ the corresponding output. Compared with the GRU, each time node of the BiGRU contains the information of the entire input sequence, which can perform better feature extraction over the whole input.

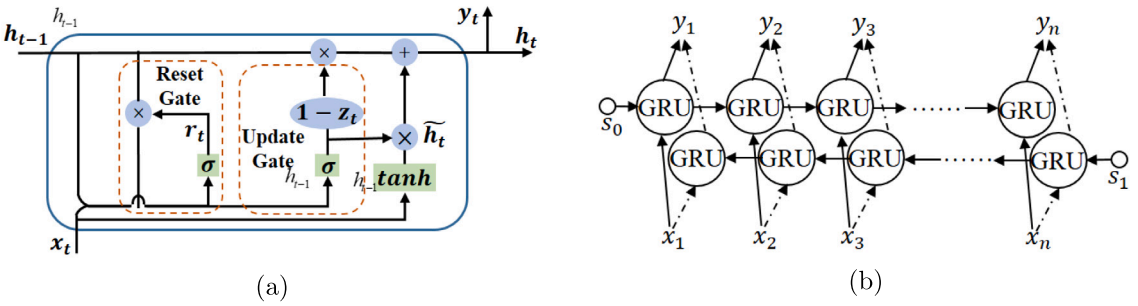


**Fig. 1.** The structure of the GRU and BiGRU networks. (a) The structure of the GRU. (b) The structure of the BiGRU network.

### 2.2. Bayesian neural network

During inference, with the neural network weights fixed, the prediction is deterministic without probabilistic or uncertainty. Bayesian neural network [29,30] can obtain both the prediction and the uncertainty of prediction simultaneously.

For a given data set $S = [(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)]$, the parameters $\omega$ of the Bayesian network are no longer fixed values, but from the posterior distribution $p(\omega|S)$. In this case, for any input $x$, the distribution of output $y$ is as follows:

$$p(y|x, S) = \int p(y|x, \omega)p(\omega|S)d\omega \tag{5}$$

Since $p(\omega|S)$ is hard to calculate, the researchers used the variational inference method to solve this problem. That is, the distribution function $q_\theta(\omega)$ is selected to approximate the true posterior distribution $p(\omega|S)$ by minimizing the Kullback-Leibler (KL) divergence. Gal et al. [31] proposed that dropout can be used as a Bayesian approximate inference method in the deep learning model and proved that the optimization of the model is equivalent to Expected Lower Bound (ELBO) in minimizing variational inference, which further simplifies the optimization process variational inference method. When training is finished, we obtain the best approximate posterior distribution $q^*_\theta(\omega)$, where $\theta$ is the variational parameter. In this case, the predicted distribution is as follows:

$$q^*_\theta(y|x) = \int p(y|x, \omega)q^*_\theta(\omega)d\omega \tag{6}$$

It is difficult to obtain the best approximate posterior distribution $q^*_\theta(\omega)$ due to the complexity of the deep learning model. Therefore, Monte Carlo sampling method is used to obtain the prediction distribution $\hat{q}(y|x)$ of input $x$ by multiple forward prediction in the prediction stage:

$$\hat{q}(y|x) = \frac{1}{N}\sum_{n=1}^{N} p(y|x, \hat{\omega}_n) \tag{7}$$

$\hat{\omega}_n$ is the parameter obtained by dropout, and $N$ represents the number of forwarding propagation of the neural network.

In summary, the Bayesian neural networks are more robust to over-fitting with dropout regularization. Furthermore, the Bayesian approach offers uncertainty estimates via its parameters in the form of probability distributions.

In wind power forecasting, $K$ times of forwarding propagation were required for the prediction of each input sequence. A set of wind power $Y = [y_1, y_2, \ldots, y_K]$ and noise prediction $\Sigma = [\sigma_1, \sigma_2, \ldots, \sigma_K]$ will be obtained. We took the average value of $Y$ and $\Sigma^2$ as the final prediction result $y$ and aleatoric uncertainty $U_A$. Aleatoric data uncertainty describes the variability in the observed data points [32]. Since the variance reflects the degree of dispersion of the output results, according to Gal et al. [32] proposed the definition of Epistemic uncertainty, we used the variance of predicted wind power as the epistemic uncertainty $U_E$. Aleatoric uncertainty captures data's inherent noise, and epistemic uncertainty measures the uncertainty of the results due to the imperfection of the model, the non-optimal parameters, or the incompleteness of the training samples [33]. Epistemic uncertainty will gradually decrease with the increase of observation data, while aleatoric uncertainty will not [34]. Aleatoric uncertainty reduces when less noise is within the data [35]. Introducing epistemic uncertainty and aleatoric uncertainty can effectively improve model performance [34] and reject incorrect predictions [36,37].

The comprehensive uncertainty $U$ was obtained through the weighted summation of aleatoric uncertainty and epistemic uncertainty, which was used to measure the reliability of the prediction results. The final prediction result $y$, aleatoric uncertainty $U_A$, epistemic uncertainty $U_E$ and comprehensive uncertainty $U$ are calculated as follows:

$$y = \frac{1}{K}\sum_{k=1}^{K} y_k \tag{8}$$

$$U_A = \frac{1}{K}\sum_{k=1}^{K} \sigma_k^2 \tag{9}$$

$$U_E = \frac{1}{K}\sum_{k=1}^{K}(y_k - \frac{1}{K}\sum_{k=1}^{K} y_k)^2 \tag{10}$$

$$U = \lambda_1 U_A + \lambda_2 U_E \tag{11}$$

The total number of forwarding predictions per sample is denoted $K$. $\lambda_1$ and $\lambda_2$ are the weight of aleatoric uncertainty and Epistemic uncertainty, respectively. The comprehensive uncertainty $U$ is used to measure the final uncertainty of the prediction results. The higher the comprehensive uncertainty $U$, the lower the reliability of the prediction result, and vice versa. In this paper, since the average value of $U_A$ was 1/100 of $U_E$, $\lambda_1$ and $\lambda_2$ were set to 100 and 1 in order to maintain the same order of magnitude of the two uncertainties.

### 2.3. FT-Attention

The attention mechanism performed the selective learning of information by acquiring the importance of information. Inspired by Ref. [38], we designed a simple and efficient attention module for wind power prediction, called FT-Attention. FT-Attention consisted of two parts, learning the importance of different features and moments in different time sequences as shown in Fig. 2. The input of FT-Attention was a matrix $X \in R^{h\times w}$ with h features and w timing moments. During training, the vector $Z_f = [z_1, z_2, \ldots, z_h]$ was obtained by global pooling where $z_k = \frac{1}{W}\sum_i^W X(k, i)$. Then the weight vector Q was obtained by two one-dimensional convolution and nonlinear mapping. Finally, the weight vector Q was multiplied by the original input matrix X to obtain the feature-weighted matrix $\hat{U}_F SE$. Similarly, we got a time-weighted matrix $\hat{U}_T SE$. Our goal was to get the sum of the weighted feature and timing signal.

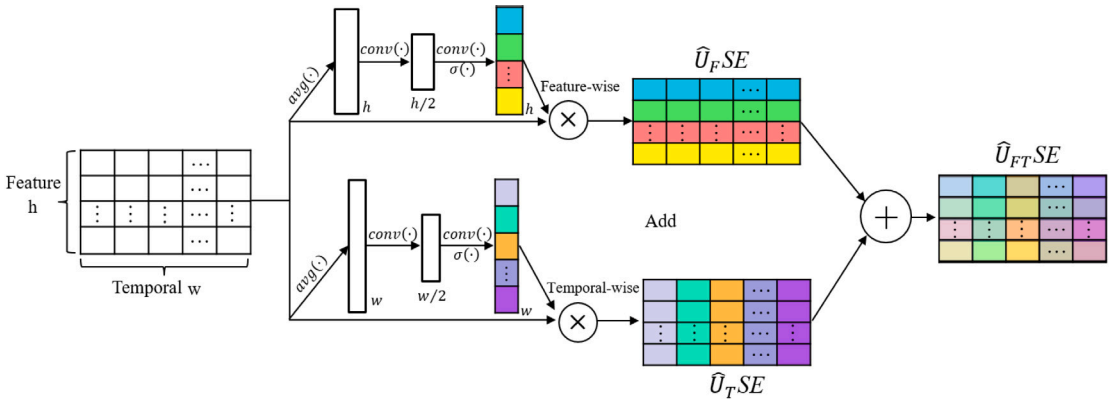


**Fig. 2.** The details of Feature Attention Module and Temporal Attention Module (FT-Attention).

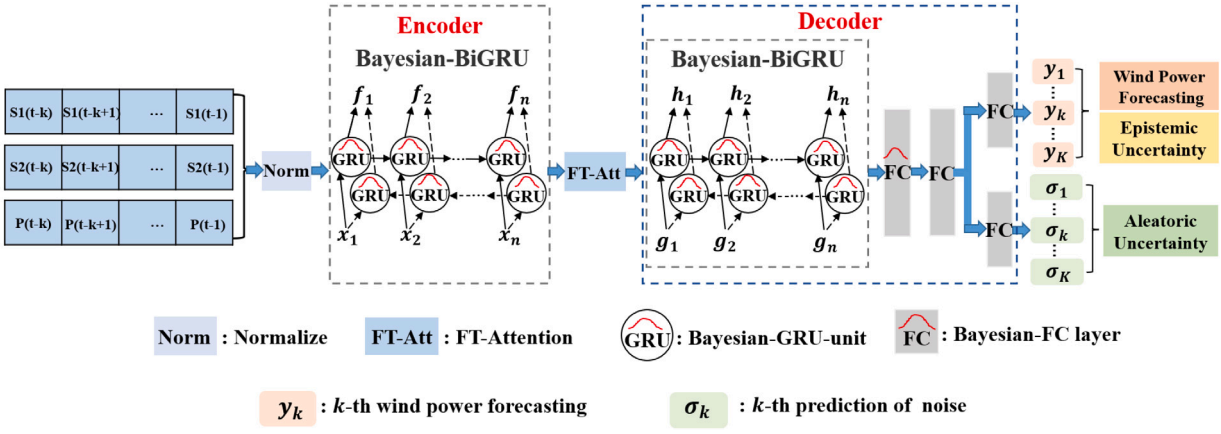


**Fig. 3.** Architecture of our proposed ultra-short-term wind power forecasting method.

### 2.4. Proposed method

The overall flow chart of the proposed algorithm is shown in Fig. 3. The input matrix $X \in R^{3\times 8}$ was composed of longitudinal wind speed, latitudinal wind speed, and actual wind power at 8 moments. In Fig. 3, where S1, S2, and P refer to longitudinal wind speed, latitudinal wind speed, and actual wind power, respectively. Since there are dimensional differences in longitudinal wind speed, latitudinal wind speed, and actual wind power, we first performed the normalization on the three input variables and got the normalized matrix $X_{norm} \in R^{3\times 8}$. Then, an encoder-decoder model (ED model) combined with an attention mechanism was used to forecast wind power and uncertainty simultaneously. The encoder was a Bayesian-BiGRU time series module composed of a series of Bayesian-GRU units, which was used to extract information of historical input to obtain the encoding feature $F_{enc} = [f_1, f_2, \ldots, f_n]$. Through attention learning, the attention module structure obtains the importance weight of the coding features in the time dimension and feature dimension and integrates the coding features according to the importance degree of information so that the network can focus more on the important features. A detailed description of the attention module was given in Section 2.3. The decoder is composed of a Bayesian-BiGRU time series module, a Bayesian-FC layer, and three common fully connected layers, in which the Bayesian-BiGRU time series module has the same structure as the encoder. The decoder further learns the attention-weighted features to obtain the decoded features $F_{dec} = [h_1, h_2, \ldots, h_n]$. Then, the predicted power value $Y = [y_1, y_2, \ldots, y_K]$ and noise prediction $\Sigma = [\sigma_1, \sigma_2, \ldots, \sigma_K]$ were obtained simultaneously after mapping the full connection layers. The final wind power forecasting result $y$, aleatoric uncertainty $U_A$ and the epistemic uncertainty $U_E$ were calculated by Eqs. (8), (9) and (10), respectively.

In our study, a dropout-based Bayesian neural network was adopted to construct the Bayesian-BiGRU module and Bayesian-FC layer. We used L1-loss as the basic loss function $L_{Basic}$ and introduced aleatoric uncertainty loss $L_{Aleatoric}$ and the epistemic uncertainty loss $L_{Epistemic}$. The comprehensive loss $L_{All}$ of the network is the weight of the above three losses. Each loss function is calculated as follows:

$$L_{Basis} = \frac{1}{N}\sum_{i=1}^{N} |y^i - \hat{y}^i| \tag{12}$$

$$L_{Aleatoric} = \frac{1}{N}\sum_{i}^{N}(U_A^i - 0)^2 = \frac{1}{N}\sum_{i}^{N}(U_A^i)^2 \tag{13}$$

$$L_{Epistemic} = \frac{1}{N}\sum_{i}^{N}(U_E^i - 0)^2 = \frac{1}{N}\sum_{i}^{N}(U_E^i)^2 \tag{14}$$

$$L_{All} = L_{Basic} + \alpha L_{Aleatoric} + \beta L_{Epistemic} \tag{15}$$

where $y^i$ and $\hat{y}^i$ are the predicted value and true value of wind power for the $i$th sample. The total number of a batch is denoted by $N$. $U_A^i$ and $U_E^i$ represent aleatoric uncertainty and epistemic uncertainty of the $i$th sample, respectively.

## 3. Dataset and preprocessing

### 3.1. Dataset

The wind power data used in this study were collected from 47 wind turbines of a wind power station from August 1, 2017 to October 1, 2018. By recording every 15 min, 96 record points per day, a total of 40,896 data were acquired. The rated power of each wind turbine is 2000 kW, and the capacity of the wind farm is 94 MW. Each data entry

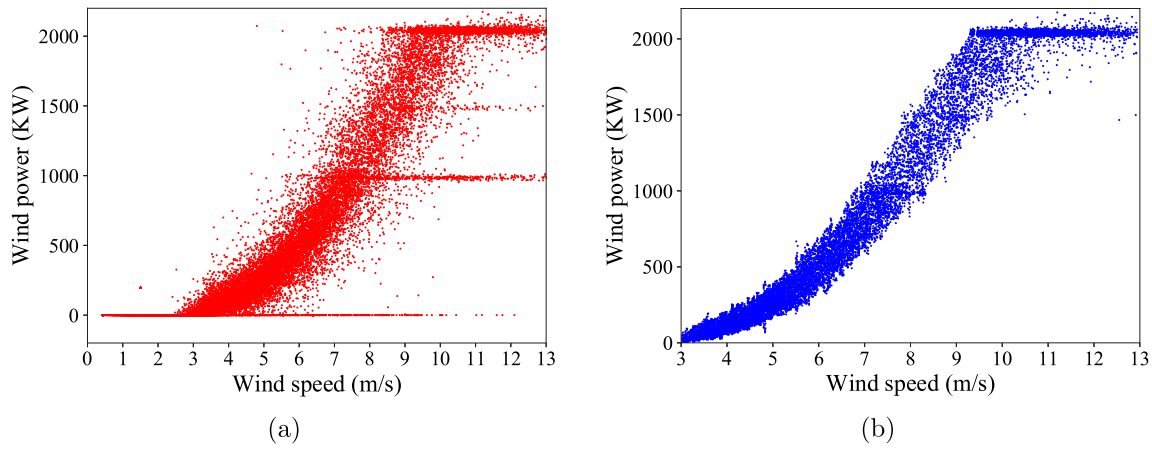


**Fig. 4.** The wind speed-power curve compare of the raw data after Gaussian fitting. (a) Wind speed-power curve of raw data. (b) Wind speed-power curve after Gaussian fitting.

**Table 1**
Correlation analysis of multiple variables.

| Variables | Range | Pearson coefficient |
|---|---|---|
| Environment temperature (°C) | −4.5~37.7 | 0.2141 |
| Pressure (hPa) | 1005.7~1035.4 | 0.0084 |
| Longitudinal wind speed (m/s) | 0~12.27 | 0.873 |
| Latitudinal wind speed (m/s) | 0~14.48 | 0.895 |
| Relative Humidity (%) | 0~69.9 | −0.3270 |
| Actual wind power (KW) | 0~2200.72 | – |

includes ambient temperature, air pressure, wind speed, wind direction, relative humidity, actual wind power, and fan working status (normal, shutdown, fault).

### 3.2. Data cleaning and normalization

The original wind speed-power data distribution was shown in Fig. 4 (a). To eliminate abnormal data and accelerate the convergence of the model, we performed a cleaning of the original data. According to the working state of the wind turbine, the data collected during the shutdown and fan fault need to be eliminated. Then, data below the cut-in wind speed (3 m/s), where the actual wind power is 0 kW, should also be excluded to prevent an unbalanced distribution of the data samples. The overall distribution of cleaned data is an S-shaped curve as shown in Fig. 4 (b). In addition, the Gaussian fitting (GF) method [4] was adopted to fit the data distribution to eliminate abnormal data with high wind speed and low power or low wind speed and high power. The GF performs the approximation of data in the form of the Gaussian function, and its formula is shown as follows:

$$G(i) = A\frac{e^{(x(i)-B)^2}}{C^2} \tag{16}$$

where $A$, $B$, and $C$ are constants, $i$, $x(i)$, and $G(i)$ represent the actual wind speed, the actual power, and the fitted power respectively.

To eliminate the magnitude influence between the data, we performed min-max normalization for all variables and mapped them to [0,1], as shown in Eq. (17).

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{17}$$

### 3.3. Feature selection

The input feature selection is particularly critical for improving the performance and reducing the computational cost of the prediction model [39]. Too few feature variables may not express enough correlation, resulting in reduced effectiveness of the prediction model. Similarly, too many features might overfit the model and increase the computational costs. Following the paper [40], the Pearson correlation coefficient (PCC) was employed to select effective inputs. PCC is a statistical metric that evaluates the direction and strength of a linear relationship between two variables, which has been widely used in many fields of science [41–43]. According to the guide that Evans (1996) [44]proposed, the evaluation criteria of PCC can be described as follows: $|\rho| \geq 0.8$: strong correlation; $0.6 \leq |\rho| < 0.8$: relatively strong correlation; $0.4 \leq |\rho| < 0.6$: moderate correlation; $0.2 \leq |\rho| < 0.4$: weak correlation; and when $0 \leq |\rho| < 0.2$, is irrelevant. The calculation formula of PCC is shown below:

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{18}$$

where $\bar{x}$, $\bar{y}$ are the average values of variables $x$ and y.

Inspired by [45], we used the sine value and cosine value of wind speed to construct longitudinal wind speed and latitudinal wind speed. The results of Pearson correlation analysis are shown in Table 1. Table 1 shows the PCC between each type of weather variable and wind power. From Table 1, we can see that the PCC between the longitudinal wind speed and the latitudinal wind speed and the wind power is 0.873 and 0.895, respectively, which indicates a strong correlation. In addition, the target power is highly autocorrelated with itself, so the composition of our input features was longitudinal wind speed, latitudinal wind speed, and wind power, which were consistent with previous studies [46,47] about effective inputs. Time-series information was adopted for modeling, and three variables, namely longitudinal wind speed $x_{lon}$, latitudinal wind speed $x_{lat}$ and wind power $x_{power}$ at historical moments, are selected to form the input feature vector as wind power at historical moments has a temporal correlation with wind power at predicted moments. The input feature vector is:

$$X_{input} = [x_{lon}, x_{lat}, x_{power}] \tag{19}$$

### 3.4. Selection of time series length

The length $l$ of the historical feature sequence of the input model is an important hyper-parameter in the prediction model, and its determination should take into account both model training complexity and prediction knowledge completeness [48]. When $l$ is too small, predictive knowledge is insufficient, and the model's prediction accuracy is limited; when $l$ is too large, the model training procedure is relatively complex, and parameter tuning may be challenging. To determine the value of $l$, we can consider analyzing the correlation between the historical feature and wind power. The PCC between the input feature vector and output power at 9 time lags were calculated as shown in Table 2. It can be seen from the results of correlation analysis that the correlation between the input feature vector and the output

**Table 2**
Pearson coefficient between time-delayed features and wind power.

| Variables | t−1 | t−2 | t−3 | t−4 | t−5 | t−6 | t−7 | t−8 | t−9 |
|---|---|---|---|---|---|---|---|---|---|
| Wind power | 0.945 | 0.925 | 0.907 | 0.891 | 0.875 | 0.862 | 0.849 | 0.835 | 0.821 |
| Longitudinal wind speed | 0.871 | 0.852 | 0.837 | 0.825 | 0.811 | 0.800 | 0.791 | 0.779 | 0.768 |
| Latitudinal wind speed | 0.893 | 0.873 | 0.859 | 0.845 | 0.832 | 0.820 | 0.809 | 0.797 | 0.785 |
| Average value | 0.903 | 0.883 | 0.868 | 0.854 | 0.839 | 0.827 | 0.816 | 0.804 | 0.791 |

**Table 3**
Comparisons of the GRU, BiGRU and Bayesian-BiGRU.

| Metrics | GRU | BiGRU | Bayesian-BiGRU | +FA | +TA | +FT |
|---|---|---|---|---|---|---|
| NMAE (%) | 9.13 | 7.81 | **7.41** | 7.40 | 6.98 | **6.66** |
| NRMSM (%) | 15.71 | 13.40 | **12.89** | 12.95 | 12.25 | **12.16** |
| MAPE (%) | 30.45 | 29.34 | **27.27** | 26.49 | 26.87 | **24.80** |
| $R^2$ | 0.6037 | 0.7118 | **0.7334** | 0.7308 | 0.7591 | **0.7627** |
| Pearson | 0.8184 | 0.8565 | **0.8638** | 0.8659 | 0.8765 | **0.8812** |

**Table 4**
Improvement of model brought by different uncertainty loss.

| Metrics | Basis (Bayesian-BiGRU+FT) | +Aleatoric | +Epistemic | +MA (All) |
|---|---|---|---|---|
| NMAE (%) | 6.66 | 6.62 | 6.33 | **5.74** |
| NRMSM (%) | 12.16 | 12.06 | 11.82 | **11.42** |
| MAPE (%) | 24.80 | 24.14 | 23.05 | **19.58** |
| $R^2$ | 0.7627 | 0.7664 | 0.7756 | **0.7907** |
| Pearson | 0.8812 | 0.8807 | 0.8875 | **0.8940** |

power weakens as the delay time increases. Follow [49], we selected the first 8 moments with a relatively strong correlation to the output power ($|\rho| > 0.8$) for time series modeling.

## 4. Results and discussion

### 4.1. Network structures and performance

Hyperparameter tuning direct impact on the performance of the WPF model. There are two types of parameter adjustment methods commonly used: Traditional methods, such as grid search (GS) [50]; Evolutionary methods, such as genetic algorithm (GA) and particle swarm optimization (PSO). GS methods are more suitable for deep learning networks with larger optimization spaces because the grid search method greatly reduces the computational cost [48]. In order to balance prediction accuracy and computational efficiency, an effective strategy combining manual experience and grid search was adopted to tune our hyper-parameter, which was consistent with previous studies [51–53] about hyperparameter tuning. In this study, there are six main hyperparameters that need to be determined: (1) the learning rate (denoted as $lr$, selected from {0.01, 0.005, 0.001, 0.0005}); (2) the number of Encoder and Decoder layers (denoted as $n$, selected from {1, 2, 3, 4, 5}, which determined the depth of Bayesian networks; (3) the number of neurons in each hidden layer (denoted as $h$, selected from {16, 32, 64, 128}); (4) the size of the epochs (denoted as $e$, selected from {100, 200, 500, 1000}); (5) the batch size (denoted as $b$, selected from {16, 32, 64, 128}); (6) the dropout rate (denoted as $p$, selected from {0.1, 0.3, 0.5, 0.7, 0.9}). The candidate models with various hyperparameter combinations were first trained on the training set and then evaluated on the test set. The model producing the lowest error was regarded as being governed by the optimal hyperparameters. The following parameters were determined by the grid search: $lr$ = 0.001, $n$ = 2, $h$ = 64, $e$ = 200, $b$ = 64 and $p$ = 0.5. In addition, the times of forwarding propagation $K$ were set to 50 and the optimizer was Adam. All experiments were performed on a workstation with a Dual Intel (R) Xeon (R) E5-2643 v4 CPU @3.4 GHz with 12 cores and ten NVIDIA GTX 1080Ti GPUs.

We compared the proposed Bayesian-BiGRU with GRU, BiGRU, and other previous methods; performed ablation experiments to study the effects of different attention modules. Furthermore, we tested the effects of uncertainty losses on the performance of the networks. To perform a comprehensive evaluation and comparison of the performance of the various algorithms, we applied the following metrics:

Normalized Mean Absolute Error (NMAE):

$$E_{NMAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - \hat{P}_i}{P_{cap}}\right| \times 100\% \tag{20}$$

Normalized Root Mean Square Error (NRMSE):

$$E_{NRMSE} = \frac{1}{P_{cap}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - \hat{P}_i)^2} \times 100\% \tag{21}$$

Mean Absolute Percentage Error (MAPE):

$$E_{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - \hat{P}_i}{P_i}\right| \times 100\% \tag{22}$$

where $P_i$ and $\hat{P}_i$ are the actual power value and the predicted power value of the $i$th sample. $N$ represents the number of reported forecasting samples. $P_{cap}$ is the rated turbine capacity.

Coefficient of Determination $R^2$:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}|P_i - \hat{P}_i|}{\sum_{i=1}^{N}|P_i - \bar{P}|} \tag{23}$$

where $\bar{P}$ represents the mean value of the actual power.

#### 4.1.1. Bayesian-BiGRU achieves the best performance

Comparisons of the GRU, BiGRU, and Bayesian-BiGRU were carried out to verify the effectiveness and superiority of the Bayesian-BiGRU. Table 3 shows that BiGRU achieves better performance than GRU in NMAE, NRMSE, MAPE, $R^2$ and Pearson coefficients, indicating that the BiGRU can better extract the features of the entire input sequence by learning the input information at all times. Based on the dropout regularization, the proposed Bayesian-BiGRU achieves significant improvement in power prediction by combining BiGRU and Bayesian neural networks. All the evaluation metrics in Table 3 show that the Bayesian-BiGRU outperforms both the GRU and BiGRU. In particular, Bayesian-BiGRU achieves a significant improvement of 2.07% and 3.18% in MAPE than BiGRU and GRU, respectively.

#### 4.1.2. FT-Attention improves model performance

We tested the effects of different attention modules on the performance of the Bayesian-BiGRU. Table 3 demonstrates that the introduction of either the Feature-Attention (FA) or the Temporal-Attention (TA) obtains a positive outcome, i.e., improves the model's performance. Furthermore, the combination of the two modules, i.e., the FT-Attention, further enhances the model's performance, as evaluated with NMAE, NRMSE, MAPE, and $R^2$. The above results demonstrate the effectiveness of the proposed attention modules.

#### 4.1.3. Uncertainty losses improve model performance

Based on the Bayesian-BiGRU with the FT-Attention module, we tested the effects of the aleatoric, epistemic, and a combination of the two loss (Eqs. (13), (14) and (15)). Table 4 demonstrates that introducing the two uncertainty losses brings significant improvement to the model, and a combination of the two further improves the

**Table 5**
Comparison of the performance of different algorithms.

| Method | SVM | BPNN | LSTM | BiLSTM | GRU | BiGRU | Ours |
|---|---|---|---|---|---|---|---|
| NMAE (%) | 10.02 | 8.68 | 9.26 | 9.11 | 9.13 | 7.81 | **5.74** |
| NRMSM (%) | 14.30 | 14.96 | 16.23 | 15.98 | 15.71 | 13.40 | **11.42** |
| MAPE (%) | 37.27 | 30.05 | 30.42 | 29.41 | 30.45 | 29.34 | **19.58** |
| $R^2$ | 0.6720 | 0.6406 | 0.5775 | 0.5904 | 0.6037 | 0.7118 | **0.7907** |
| Pearson | 0.8232 | 0.8314 | 0.8038 | 0.8158 | 0.8184 | 0.8565 | **0.8940** |

model's performance with substantial decreases of NMAE, NRMSE, MAPE, and increases of $R^2$ and Pearson by 0.92%, 0.74%, 5.22%, 0.028 and 0.0128, respectively.

#### 4.1.4. Comparisons of different algorithms

We compared the proposed algorithm with six classical methods: SVM [54], BPNN [40], LSTM [49], BiLSTM [55], GRU [52], BiGRU [56]. For SVM and BPNN, the setup was adopted as in studies [40, 54]. The BiGRU network was conducted by removing FT-Attention and Bayes modules. Further, each node in the BiGRU was replaced with the LSTM unit to form the BiLSTM network. We transformed the bidirectional structure in the BiGRU and BiLSTM into the one-way network to obtain the GRU and the LSTM model, respectively. The $L_{basic}$ was adopted as the loss function of the above model. For the training of the networks, the hyperparameter settings are completely consistent with our proposed method. In Table 5, it can be seen that the deep learning based models (LSTM, BiLSTM, GRU, BiGRU and ours) show significantly better performance than traditional machine learning methods as the excellent feature extraction and nonlinear mapping ability. The LSTM, BiLSTM and GRU deliver similar performance, and the BiGRU is significantly better than other classical models. Our proposed model achieves the best result as evaluated with all the listed metrics. The experimental results demonstrate the effectiveness and superiority of the proposed Bayesian-BiGRU structure with the uncertainty losses and FT-Attention modules. Meanwhile, we compared the power prediction results of our model with SVM, BPNN, LSTM, BiLSTM, GRU and BiGRU for two consecutive days as shown in Fig. 5. From the figure, we can see that the proposed method is significantly better than other classical models. The results showed that our model performs better for abrupt changes when the fan power changes significantly, while the power prediction results of other deep learning models are too smooth.

### 4.2. Explore the reliability of the forecasting with uncertainty

To ensure the safety of grid connection, the State Grid of China stipulates that the monthly average ultra-short-term wind power of the report qualified rate (QR) is not less than 85%. Wind farms will face severe financial penalties when the QR does not meet the requirements. To this end, we introduced uncertainty to eliminate the unreliable forecast power, which is beneficial to improve the qualified rate of wind power prediction and reduce the economic loss of wind farms.

The formula of the report qualified rate (QR) and the retention rate (RR) of qualified predicted power are defined as:

$$QR = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} M_i} \tag{24}$$

$$RR = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} D_i} \tag{25}$$

where,

$$C_i = \begin{cases} 1, (1 - |\frac{P_i - \hat{P}_i}{P_{cap}}|) \geq r \quad if \quad U \leq U_{th} \\ 0, (1 - |\frac{P_i - \hat{P}_i}{P_{cap}}|) < r \quad if \quad U \leq U_{th} \end{cases} \tag{26}$$

$$M_i = \begin{cases} 1, U \leq U_{th} \\ 0, U > U_{th} \end{cases} \tag{27}$$

$$D_i = \begin{cases} 1, (1 - |\frac{P_i - \hat{P}_i}{P_{cap}}|) \geq r \\ 0, (1 - |\frac{P_i - \hat{P}_i}{P_{cap}}|) < r \end{cases} \tag{28}$$

where $P_i$, $\hat{P}_i$ and $P_{cap}$ are the actual power value, the predicted power, and the wind turbine capacity, respectively. $r$ denotes the qualification benchmark for predicted power. When $r$ is known, the corresponding QR can be obtained by the above formula. $C_i$ indicates whether the $i$th reported wind power meets the assessment conditions, after data selection using uncertainty. $U$ and $U_{th}$ represent the normalized uncertainty and the uncertainty threshold value respectively. $D_i$ indicates whether the $i$th reported wind power meets the assessment conditions for all wind turbine forecasting power whether to report or not. The predicted wind power is qualified only when $(1 - |\frac{P_i - \hat{P}_i}{P_{cap}}|) \geq r$ is met. $r \geq 0.9$ is needed according to the standard of conformity stipulated by the State Grid of China. The provincial power grids can formulate appropriate qualification benchmark for actual conditions based on the national power grid regulations.

#### 4.2.1. Exploring the impact of different uncertainty thresholds on QR and RR

Uncertainty is a metric used to assess the reliability of prediction results, and the formula for calculating it is as formula (11). The larger the uncertainty, the lower the prediction result's reliability; conversely, the lower the uncertainty, the higher the reliability of the prediction result. The uncertainty threshold $U_{th}$ is used to judge whether the forecast result is reliable. When the uncertainty $U$ is greater than the uncertainty threshold $U_{th}$ for each predicted power, the prediction result is deemed unreliable; when the uncertainty $U$ is less than the uncertainty threshold $U_{th}$, the forecast result is deemed reliable. The value of the uncertainty threshold impacts the indicators QR and RR, according to formulas (24)-(28). When the uncertainty threshold $U_{th}$ rises, QR decreases while RR increases; otherwise, QR falls while RR rises. As a result, the uncertainty threshold balances the QR and RR.

To eliminate the unreliable forecast power while ensuring the retention rate of qualified predicted power, we adopted the reported qualified rate (QR) and qualified forecast power retention rate (RR) as evaluation metrics to study the selection of uncertainty threshold. The following cases of $r = 0.9$ and $r = 0.95$ are provided below as examples. The same methods can be used to select the uncertainty threshold under other eligibility criteria.

Fig. 6(a) shows the variation curve of QR and RR with the uncertainty threshold $U_{th}$ when $r = 0.9$. It can be seen that the QR of the proposed prediction method is greater than 85% under any uncertainty threshold which meets the requirements of the State Grid. When $U_{th} = 1$, the minimum QR value obtained by the proposed method is 89.6%, which fully reflects the excellent performance of our method.

When $r = 0.95$ (Fig. 6(b)), QR gradually decreases while RR increases with the increase of the uncertainty threshold. When $U_{th} = 0.343$, the reported qualified rate QR = 85% (point A in Fig. 6(b)) and RR = 95.82% (point B in Fig. 6(b)). Compared with Fig. 6(a), increasing the conformity benchmark $r$ will lead to a significant decrease in the reported qualified rate. Selecting an appropriate uncertainty threshold can effectively eliminate unqualified samples. In this paper, we aim to make the qualified rate of wind farm meet the requirements of grid connection while ensuring a higher retention rate of qualified predicted power to reduce the economic loss of wind farm.

#### 4.2.2. Elimination of unreliable wind power prediction

Considering the power generation situation of the wind farm and the corresponding reward and punishment measures, we choose the qualification benchmark $r$ of 0.95 and uncertainty threshold $U_{th}$ of 0.343. Fig. 7(a) and Fig. 7(b) respectively show the results of wind power prediction and the corresponding elimination of untrusted prediction when the wind speed is stable and fluctuates greatly. It can

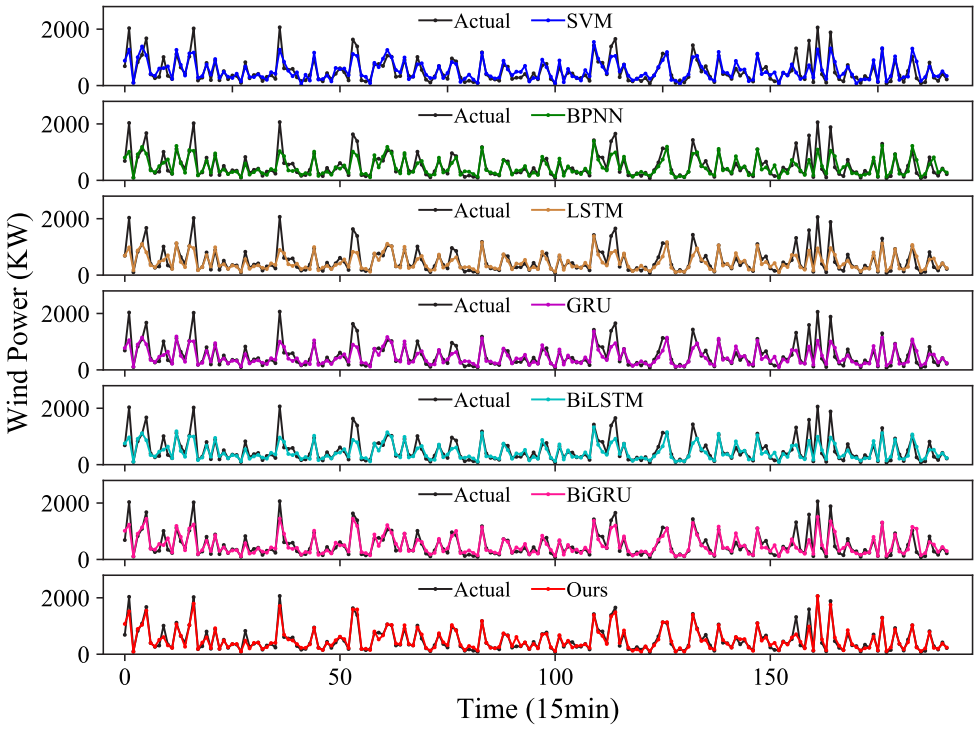


**Fig. 5.** Comparison of different forecasted wind power and actual wind power.

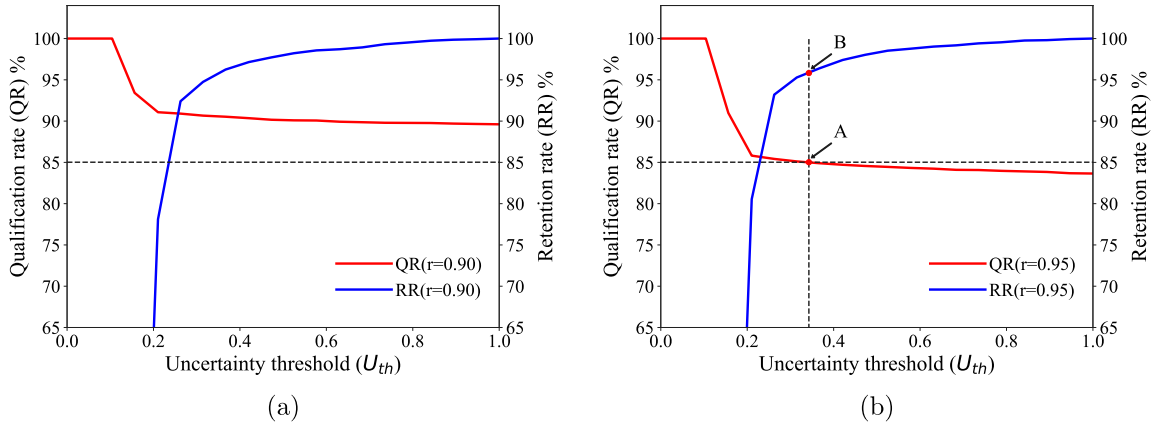


**Fig. 6.** To explore the effect of uncertainty threshold value on reported qualified rate (RR) and qualified sample retention rate (QR). (a) $r = 0.9$ (b) $r = 0.95$.

be seen that the reliability of wind power prediction results is high when the wind speed is stable, which meets the requirements of wind farms. The uncertainty of wind power prediction increases and unreliable prediction results appear when the wind speed fluctuates greatly (points marked by yellow triangles). The unqualified report rate of wind farms can be improved by using uncertainty to reject the report of unreliable prediction results (red points marked by ×). According to the power generation situation combined with the reward and punishment measures of each wind farm, the retention rate of qualified predicted power is ensured while the reported qualified rate is improved, to achieve the highest economic benefit of the wind farm.

### 4.3. Discussion

Our study aims to improve the accuracy of ultra-short-term wind power forecasting by combining the Bayesian framework, BiGRU deep network, and uncertainty while eliminating unreliable forecasts. In the comparative experiment, the effect of different structures, attention mechanisms, and uncertainty loss on prediction accuracy was explored. Each node of the BiGRU can obtain the input information at all times, which improves the model's ability for feature extraction, thereby improving the accuracy of wind power prediction. We combined the Bayesian and BiGRU time-series networks through the dropout regularization to improve the model's accuracy. The proposed FT-Attention module obtains the weights in the dimension of feature and time simultaneously, which can assist the model in extracting important information and reducing the NMAE and NRMSE from 7.41% and 12.89% to 6.66% and 12.16%, respectively. Epistemic uncertainty loss and aleatoric uncertainty loss effectively suppress the adverse effect of the inherent noise of observation data, incomplete training samples, and imperfect models. The introduction of these two loss functions improved the model in NMAE, NRMSE, MAPE, $R^2$, and Pearson coefficient by 0.92%, 0.74%, 5.22%, 0.028, and 0.0128, respectively. The proposed algorithm achieves the best performance in various evaluation metrics compared with other models.

Uncertainty was adopted to measure the reliability of the predicted power. Choosing a suitable uncertainty threshold can eliminate unreliable forecasts while retaining qualified wind power forecasts to

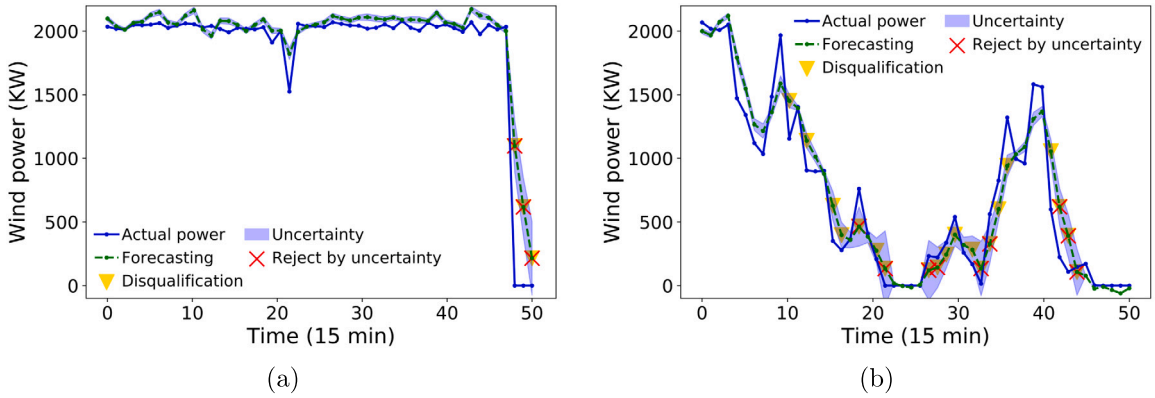


**Fig. 7.** Elimination of unreliable wind power prediction. (red × represents rejected prediction result by uncertainty, yellow triangles represents did not meet the quality requirements QR($r = 0.95$). (a) The wind speed is stable (b) The wind speed is fluctuates greatly.

get maximum economic benefits for wind farms. Through the exploration of the selection of uncertainty threshold, it is found that as the uncertainty threshold increases, the reported qualified rate gradually decreases, while the qualified sample retention rate gradually increases. Each wind farm can use our algorithm to obtain a suitable threshold according to its actual power generation situation. Furthermore, quantifying the uncertainty will benefit smart grid planning, comprehensive decision-making, and liberalized wind power market bidding strategies.

## 5. Conclusions

In our study, the wind power prediction method based on Bayesian network and uncertainty was proposed for the first time. To improve the capabilities of the model for feature selection and characterization, we designed the Encoder-Decoder architecture combined with BiGRU time series modeling and FT-Attention (Feature-Temporal Attention) mechanism which significantly improve the prediction accuracy of wind power. Further, aleatoric uncertainty loss and epistemic uncertainty loss were added to the loss function to improve the performance of the model. The experimental results show that these two uncertainty losses can effectively improve the accuracy and robustness of the model. Compared with BP, GRU, LSTM and other classical models, the proposed method shows a significantly better performance in NMAE, NRMSE, MAPE, $R^2$ and Pearson coefficient.

It is of great significance to use uncertainty to measure the reliability of prediction results. To improve the qualified rate of power forecast, we took the uncertainty of the model as the basis to reject unreliable predictions, and investigated the qualified rate and qualified sample retention rate of wind power forecast under different uncertainty thresholds. For wind farms in different regions, a reasonable uncertainty threshold was selected according to the power generation situation and reward and punishment measures. The overall goal is to improve the reporting rate while ensuring that qualified power forecasts maximize the economic benefits of wind farms.

The wind power forecasting method based on Bayesian with uncertainty proposed in this paper provides reliable and accurate results which are meant for ensuring the safety of grid connection, assisting the grid dispatching department in developing dispatching plans, and improving wind power consumption, which is of strategic significance for the further development of the wind power industry and the realization of “carbon neutrality” and “emission peak”. In the future, we will investigate the optimal bidding strategies based on the uncertainty.

## CRediT authorship contribution statement

**Lei Liu:** Conceptualization, Methodology, Data curation, Software, Visualization, Writing - original draft. **Jicheng Liu:** Conceptualization, Interpreted the results, Writing - review & editing. **Yu Ye:** Investigation, Validation. **Hui Liu:** Investigation, Validation. **Kun Chen:** Investigation, Validation. **Dong Li:** Investigation, Validation. **Xue Dong:** Data curation, Investigation, Validation, Supervision. **Mingzhai Sun:** Conceptualization, Interpreted the results, Writing - Review, Funding acquisition.

## 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.

## Acknowledgments

The authors are grateful for the financial supports from the National Natural Science Foundation of China under grand U19B2044, Natural Science Foundation of Anhui Province, China under Grant 2008085MH259 and Key Research and Development Program of Anhui Province under Grant 202004h07020015.

## References

[1] M. Lei, S. Luan, C. Jiang, H. Liu, Z. Yan, A review on the forecasting of wind speed and generated power, Renew. Sustain. Energy Rev. 13 (4) (2009) 915–920.
[2] A review of combined approaches for prediction of short-term wind speed and power, Renew. Sustain. Energy Rev. 34 (jun.) (2014) 243–254.
[3] A review on the selected applications of forecasting models in renewable power systems, Renew. Sustain. Energy Rev. 100 (2019) 9–21.
[4] G. Wang, R. Jia, J. Liu, H. Zhang, A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning, Renew. Energy 145 (2020) 2426–2434.
[5] Y. Zhang, K. Liu, L. Qin, X. An, Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods, Energy Convers. Manage. 112 (2016) 208–219.
[6] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, M.U. Rehman, A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting, IEEE Access (2019) 1.
[7] M. Marčiukaitis, I. Žutautaitė, L. Martišauskas, B. Jokšas, G. Gecevičius, A. Sfetsos, Non-linear regression model for wind turbine power curve, Renew. Energy 113 (2017) 732–741.
[8] E.B. Ssekulima, M.B. Anwar, A.A. Hinai, M. Moursi, Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review, IET Renew. Power Gener. 10 (7) (2016) 885–989.
[9] X. Zhao, J. Liu, D. Yu, J. Chang, One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data, Energy Convers. Manage. 164 (MAY) (2018) 560–569.
[10] X. Shi, X. Lei, Q. Huang, S. Huang, K. Ren, Y. Hu, Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short-term memory, Energies 11 (11) (2018).
[11] C. Wang, H. Zhang, P. Ma, Wind power forecasting based on singular spectrum analysis and a new hybrid Laguerre neural network, Appl. Energy 259 (2020).

[12] L. Li, Y. Li, B. Zhou, Q. Wu, Z. Gong, An adaptive time-resolution method for ultra-short-term wind power prediction, Int. J. Electr. Power Energy Syst. 118 (2020) 105814.
[13] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, A review of deep learning for renewable energy forecasting, Energy Convers. Manage. 198 (2019) 111799.
[14] Z. Sun, M. Zhao, Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis, IEEE Access PP (99) (2020) 1.
[15] L. Han, H. Jing, R. Zhang, Z. Gao, Wind power forecast based on improved long short term memory network, Energy 189 (2019) 116300.
[16] H. Demolli, A.S. Dokuz, A. Ecemis, M. GokceK, Wind power forecasting based on daily wind speed data using machine learning algorithms, Energy Convers. Manage. (2019).
[17] C. Chen, H. Liu, Medium-term wind power forecasting based on multi-resolution multi-learner ensemble and adaptive model selection – ScienceDirect, Energy Convers. Manage. 206.
[18] C. Yu, Y. Li, M. Zhang, An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on Elman neural network, Energy Convers. Manage. 148 (2017) 895–904.
[19] J.-Z. Wang, Y. Wang, P. Jiang, The study and application of a novel hybrid forecasting model — A case study of wind speed forecasting in China – ScienceDirect, Appl. Energy 143 (2015) 472–488.
[20] L. Xiang, J. Li, A. Hu, Y. Zhang, Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method, Energy Convers. Manage. 220 (2020) 113098.
[21] X. Yan, Y. Liu, Y. Xu, M. Jia, Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition, Energy Convers. Manage. 225 (4) (2020) 113456.
[22] Z. Min, W.B. Bo, B. Sg, C. Jw, Multi-objective prediction intervals for wind power forecast based on deep neural networks, Inform. Sci. (2020).
[23] T. Ishii, K. Otani, T. Takashima, Y. Xue, Solar spectral influence on the performance of photovoltaic (PV) modules under fine weather and cloudy weather conditions, Prog. Photovolt., Res. Appl. 21 (4) (2013).
[24] Z. Eaton-Rosen, F. Bragman, S. Bisdas, S. Ourselin, M.J. Cardoso, Towards safe deep learning: Accurately quantifying biomarker uncertainty in neural network predictions, 2018.
[25] I. Bhat, H.J. Kuif, V. Cheplygina, J. Pluim, Using uncertainty estimation to reduce false positives in liver lesion detection, 2021.
[26] B. Ghoshal, A. Tucker, Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection, 2020.
[27] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[28] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, 2014, arXiv preprint arXiv:1406.1078.
[29] J.S. Denker, Y. LeCun, Transforming neural-net output levels to probability distributions, Adv. Neural Inf. Process. (1998).
[30] R. Warner, Bayesian learning for neural networks (lecture notes in statistical Vol. 118) by Radford M. Neal, J. Amer. Statist. Assoc. 92 (438) (1997) 791–792.
[31] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep learning, 2015, JMLR.org.
[32] A. Kendall, Y. Gal, What uncertainties do we need in bayesian deep learning for computer vision? 2017, arXiv preprint arXiv:1703.04977.
[33] P. Seeböck, J.I. Orlando, T. Schlegl, S.M. Waldstein, H. Bogunović, S. Klimscha, G. Langs, U. Schmidt-Erfurth, Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT, IEEE Trans. Med. Imaging 39 (1) (2019) 87–98.
[34] R. Zheng, S. Zhang, L. Liu, Y. Luo, M. Sun, Uncertainty in bayesian deep label distribution learning, Appl. Soft Comput. 101 (2021) 107046.
[35] M. Rußwurm, M. Ali, X.X. Zhu, Y. Gal, M. Körner, Model and data uncertainty for satellite time series forecasting with deep recurrent models, in: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2020, pp. 7025–7028.
[36] I. Bhat, H.J. Kuijf, V. Cheplygina, J.P.W. Pluim, Using uncertainty estimation to reduce false positives in liver lesion detection, in: 2021 IEEE 18th International Symposium on Biomedical Imaging, ISBI, IEEE, 2021, pp. 663–667.
[37] K. Mason, J. Knights, M. Ramezani, P. Moghadam, D. Miller, Uncertainty-aware lidar place recognition in novel environments, 2022, arXiv preprint arXiv:2210.01361.
[38] H. Jie, S. Li, S. Gang, S. Albanie, Squeeze-and-excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. PP (99) (2017).
[39] H. Huang, R. Jia, X. Shi, J. Liang, J. Dang, Feature selection and hyper parameters optimization for short-term wind power forecast, Appl. Intell. 51 (10) (2021) 6752–6770.
[40] J. Liu, X. Wang, Y. Lu, A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system, Renew. Energy 103 (2017) 620–629.
[41] O.J. Dunn, V.A. Clark, Applied Statistics: Analysis of Variance and Regression, John Wiley & Sons, Inc., 1986.
[42] F. Bai, Y. Liu, Y. Liu, K. Sun, N. Bhatt, A. Del Rosso, E. Farantatos, X. Wang, Measurement-based correlation approach for power system dynamic response estimation, IET Gener. Transm. Distrib. 9 (12) (2015) 1474–1484.
[43] J. Benesty, J. Chen, Y. Huang, On the importance of the pearson correlation coefficient in noise reduction, IEEE Trans. Audio Speech Lang. Process. 16 (4) (2008) 757–765.
[44] Straightforward statistics for the behavioral sciences, J. Amer. Statist. Assoc. 91 (436) (1996) 1750.
[45] K. Wang, X. Qi, H. Liu, J. Song, Deep belief network based k-means cluster approach for short-term wind power forecasting, Energy 165 (2018) 840–852.
[46] S.-X. Wang, M. Li, L. Zhao, C. Jin, Short-term wind power prediction based on improved small-world neural network, Neural Comput. Appl. 31 (7) (2019) 3173–3185.
[47] T. Liu, Z. Huang, L. Tian, Y. Zhu, H. Wang, S. Feng, Enhancing wind turbine power forecast via convolutional neural network, Electronics 10 (3) (2021) 261.
[48] L. Wang, Y. He, L. Li, X. Liu, Y. Zhao, A novel approach to ultra-short-term multi-step wind power predictions based on encoder–decoder architecture in natural language processing, J. Clean. Prod. 354 (2022) 131723.
[49] J. Li, D. Geng, P. Zhang, X. Meng, Z. Liang, G. Fan, Ultra-short term wind power forecasting based on LSTM neural network, in: 2019 IEEE 3rd International Electrical and Energy Conference, CIEEC, IEEE, 2019, pp. 1815–1818.
[50] I. Syarif, A. Prugel-Bennett, G. Wills, SVM parameter optimization using grid search and genetic algorithm to improve classification performance, Telkomnika 14 (4) (2016) 1502.
[51] Z. Niu, Z. Yu, W. Tang, Q. Wu, M. Reformat, Wind power forecasting using attention-based gated recurrent unit network, Energy 196 (2020) 117081.
[52] A. Kisvari, Z. Lin, X. Liu, Wind power forecasting—A data-driven method along with gated recurrent neural network, Renew. Energy 163 (2021) 1895–1909.
[53] L.-L. Li, X. Zhao, M.-L. Tseng, R.R. Tan, Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm, J. Clean. Prod. 242 (2020) 118447.
[54] N.A. Treiber, J. Heinermann, O. Kramer, Wind power prediction with machine learning, in: Computational Sustainability, Springer, 2016, pp. 13–29.
[55] M. Neshat, M.M. Nezhad, E. Abbasnejad, S. Mirjalili, L.B. Tjernberg, D.A. Garcia, B. Alexander, M. Wagner, A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the lillgrund offshore wind farm, Energy Convers. Manage. 236 (2021) 114002.
[56] M. Ding, H. Zhou, H. Xie, M. Wu, Y. Nakanishi, R. Yokoyama, A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting, Neurocomputing 365 (2019) 54–61.