



Short-term prediction of wind power based on BiLSTM–CNN–WGAN-GP

Ling Huang^{1,2,3} · Linxia Li¹ · Xiaoyuan Wei^{1,2,3} · Dongsheng Zhang¹

Accepted: 26 December 2021 / Published online: 17 January 2022
© The Author(s) 2022

Abstract

A short-term wind power prediction model based on BiLSTM–CNN–WGAN-GP (LCWGAN-GP) is proposed in this paper, aiming at the problems of instability and low prediction accuracy of short-term wind power prediction. Firstly, the original wind energy data are decomposed into subsequences of natural mode functions with different frequencies by using the variational mode decomposition (VMD) algorithm. The VMD algorithm relies on a decision support system for the decomposition of the data into natural mode functions. Once the decomposition is performed, the nonlinear and dynamic behavior are extracted from each natural mode function. Next, the BiLSTM network is chosen as the generation model of the generative adversarial network (WGAN-GP) to obtain the data distribution characteristics of wind power's output. The convolutional neural network (CNN) is chosen as the discrimination model, and the semi-supervised regression layer is utilized to design the discrimination model to predict wind power. The minimum–maximum game is formed by the BiLSTM and CNN network models to improve the quality of sample generation and further improve the prediction accuracy. Finally, the actual data of a wind farm in Jiuquan City, Gansu Province, China is taken as an example to prove that the proposed method has superior performance compared with other prediction algorithms.

Keywords Wind power forecasting · Generative adversarial network · Variational modal decomposition · Bidirectional long-term and short-term memory network · Convolutional neural network

1 Introduction

As an environment-friendly renewable energy, wind energy has the characteristics of no pollution and a wide range occurrence. It has been widely developed and utilized all over the world. The penetration of wind energy poses a great challenge to the economic operation, reliability, and real-time control of power energy systems.

According to the Global Wind Energy Council, it is estimated that the cumulative installed capacity of global wind power will reach 756 GW and that of China will reach 2145 GW in 2021. However, because of intermittency and randomness, a large number of wind power generation systems are limited in their ability to integrate into the power system (Wang and Yang 2021). Short-term wind power prediction is one of the key technologies to alleviate this problem, and provides a more accurate prediction value for power system dispatching. Accurate wind power forecasting can alleviate the pressure of power system dispatching, reduce power generation costs, and enhance the competitiveness of wind power in the power market (Duan et al. 2021). Therefore, effective short-term wind power prediction technology is very important for wind farm management and the economic and safe operation of wind power systems.

Wind power forecasting methods can be roughly divided into three categories: physical methods, statistical methods and machine learning methods (Chuang et al. 2019). The physical method relies on numerical weather forecast

Communicated by Tiancheng Yang.

✉ Ling Huang
huangling@lut.edu.cn

¹ College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China

² Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou University of Technology, Lanzhou 730050, China

³ National Demonstration Center for Experimental Electrical and Control Engineering Education, Lanzhou University of Technology, Lanzhou 730050, China

(NWP) to establish the model, which limits the effectiveness of the forecast accuracy of NWP, resulting in poor stability of the model (Jung and Broadwater 2014). Compared with physical methods, statistical methods are widely used, including auto-regressive moving average (ARMA) model, support vector regression (SVR), least squares support vector machine (LSSVM) (Hao and Tian 2019). Kavasseri et al. proposed fractional ARIMA model to predict the wind speed of the previous day and the first 2 days. The results show that compared with other benchmark methods, the improved ARIMA model can obtain more accurate and stable prediction values (Kavasseri and Seetharaman 2009). Ren et al. used the improved auto-regressive integrated moving average model to predict the short-term load (Ren et al. 2016). The prediction results show that the improved model is better than the back-propagation network algorithm and the traditional ARIMA model. Based on VMD, ARIMA and depth neural network, an adaptive hybrid model for wind speed prediction was designed (Zhang et al. 2020). However, most statistical methods are often modeled as simplified linear functions, so the prediction performance of statistical models may be limited to a great extent.

With the rapid development of artificial intelligence, many prediction technologies based on machine learning have been developed to realize high-precision predictions for wind power series with nonlinear and non-stationary characteristics (Zeng and Qiao 2012). Based on machine learning, it is divided into shallow network model and deep network model (Hu and Chen 2018). Shallow networks such as extreme learning machine (ELM) and SVM have the ability to adaptively learn nonlinear features, which are often reported in the literature. However, due to structural limitations, shallow networks have the drawbacks of being easily trapped in local optimality, overfitting, and poor convergence (Khodayar et al. 2018). Deep model has attracted much attention because of its potential to improve the prediction accuracy of wind power generation. Including artificial neural network (ANN), CNN, Short and long term memory network (LSTM), threshold recursive network (GRU), which have been successfully applied to wind power generation prediction (Kong et al. 2017). For example, Hong et al. utilized CNN to construct a novel spatiotemporal model to perform day-ahead wind speed prediction (Hong and Satriani 2020). He et al. developed a novel wind power hybrid forecasting model by applying light gradient enhancer (LGBM) and random forest (RF) to replace the BP network of the traditional DBN regression layer and combining with wavelet decomposition (Jiajun et al. 2020). A combination of deep learning and ensemble learning is developed to predict (Jiajun et al. 2020). To adjust the super parameters of LSTM, the LSTM, ensemble empirical mode decomposition (EEMD) and genetic

algorithm (GA) were used in Chen et al. (2021). The prediction effect is better than that of a single network, but the prediction time is longer when the amount of data is large. To solve the problem. Shi et al. A recursive and direct variational model is proposed to decompose the LSTM network for hourly and daily advance wind power prediction (Shi et al. 2018). Cali and Sharma (2019) also use the historical power and NWP data to predict the wind power using LSTM model along with sensitive analysis to identify the most informative input parameters from NWP data. Later, Yu et al. propose an LSTM enhanced forget-gate network model in which two peep holes are added to standard LSTM forget and output gates to improve the performance of the model (Yu et al. 2019). Generally speaking, these machine learning based models usually have better prediction effect than physical and statistical methods because of their strong feature learning ability. In addition, various signal decomposition technologies, such as wavelet transform (WT), empirical mode decomposition (EMD) (Jiang et al. 2020), integrated empirical mode decomposition (EEMD) (Sun and Wang 2018), have been used for wind power prediction together with RNN, CNN, LSTM and other machine learning algorithms. However, due to the volatility and discontinuity of wind power data, the prediction accuracy of wind power still needs to be improved.

Recently, a new network generative adversarial network (GAN) has been developed. Inspired by the two-player zero-sum game, GAN simultaneously trains two adversarial models: the generator network produces artificial data from noise through capturing the original data distribution, while the discriminator network is trained to distinguish generated data from original data (Creswell et al. 2018). Since its proposal, GAN has been applied to various research fields (Wang et al. 2017), including image processing, visual computing, speech, video and language processing. Due to its strong modeling and generation capabilities, GAN can create new data with similar distribution to the original data while maintaining the diversity of generated data (Goodfellow et al. 2020). However, few studies have applied GAN to the prediction of wind power. GAN is naturally suitable for the prediction of long-time wind power time series.

This paper focuses on the short-term wind power prediction. The data set studied is the measured data of the wind farm. However, due to the irregularity of the data, then considering the superiority of the improved Wasserstein-GAN (WGAN) over the original GAN, Wasserstein generative adversarial network with gradient penalty (WGAN-GP) is introduced to characterize the data set. The goal of WGAN-GP is to generate new and unique data, which capture the internal characteristics and distribution of historical wind power data, rather than simply remember

the input data. Therefore, this paper proposes a hybrid wind power prediction model based on BiLSTM–CNN–WGAN-GP (LCWGAN-GP) network and semi-supervised regression, integrates the semi-supervised regression with label learning into the LCWGAN-GP framework, extracts the internal nonlinear and dynamic behavior from the wind power time series, and improves the prediction performance of wind power. The main contributions of this paper are as follows:

- To solve the problem of gradient disappearance in the training process of traditional GAN network. Wasserstein generative adversarial network with gradient penalty is introduced to extract the data distribution characteristics of wind power time series from real samples and virtual samples.
- Aiming at the problem of wind power prediction, a prediction method based on LCWGAN-GP network and semi-supervised regression is proposed. The improved LCWGAN-GP network uses the generated model to learn training samples, and designs a discrimination model with semi-supervised regression layer to minimize the dual objective LCWGAN-GP network function, thereby improving the prediction accuracy of wind power.
- The prediction of wind power is performed by optimizing the LCWGAN-GP network, and then compared with GAN, BiLSTM, CNN, GRU, SVM and ARIMA neural networks to verify its prediction performance.

The rest of this paper is organized as follows: Sect. 2 introduces the algorithm framework and models of classical GANs networks. Section 3 proposes the prediction framework based on LCWGAN-GP and the general implementation steps. Then, the overall indices for evaluating the forecasting performance are introduced in Sect. 4. Section 5 provides the experimental results and comparative analysis. Finally, Sect. 6 summarizes some conclusions.

2 Generative adversarial network framework

GAN algorithm realizes its function by implicitly modeling the high-dimensional distribution of data (Wang and Li 2018). In the field of image processing, GAN can synthesize real images with better quality compared with other generation methods (Creswell et al. 2018). In terms of generating time series, GAN can better learn the distribution of time series, and the generative model and discriminative model play games with each other to produce the optimal prediction results.

2.1 Algorithm framework of GAN

As shown in Fig. 1, G and D correspond to the generative model and discriminative model of in the GAN, respectively. The real sample X and virtual sample $G(z)$ produced by the generative model can be used as the input to the discriminative model. The two models learn alternately and continuously optimize through the min–max game to improve the performance of the network (Husein et al. 2019). Finally, the trained generator can generate high-quality new sample data, but the discriminator cannot distinguish it from the real data.

The objective function of the discriminative model is defined as: when the input is X and $G(z)$, the output corresponds to 1 and 0, respectively. Its objective function is:

$$L_D = -E_{x \sim p_{\text{data}}}[\log D(x)] - E_{z \sim p_z}[\log(1 - D(G(z)))] \quad (1)$$

where x and z represent real samples and random noise, respectively; E stands for an expectation.

The goal of the generative model is to mimic the complex distribution in real data by minimizing the objective function in Eq. (2), make the value of $D(G(z))$ is close to 1:

$$L_G = E_{z \sim p_z} \log[1 - D(G(z))] \quad (2)$$

L_D and L_G constitutes min–max game, as follows:

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}}[\log D(x)] + E_{z \sim p_z}[\log(1 - D(G(z)))] \quad (3)$$

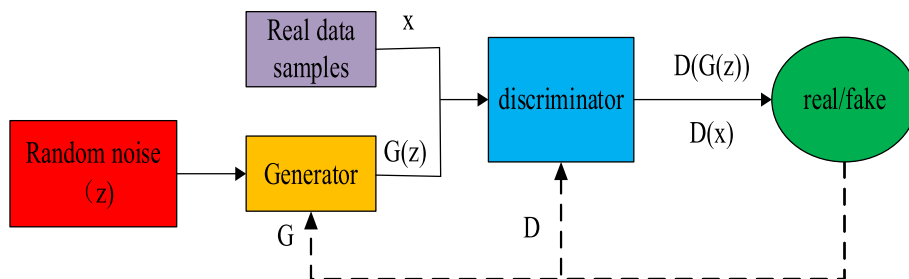
In the process of GAN training, firstly, the discriminative model D is trained to maximize the probability that the data comes from the real data, and the training generative model G is minimized. Through alternating operation, the generative model G is fixed to maximize the discriminative probability of D ; Fix the discriminative model D and optimize the generative model G to minimize the discriminative probability of D . Reach the stable state of Nash equilibrium (Mirza and Osindero 2014).

However, for the original GAN network, it is easy to produce two problems: one is the disappearance of the gradient in the training process, and the other is the confusion of the optimization objectives in the training process. The gradient becomes unstable and the diversity and accuracy of the generated data samples decrease, which in turn causes the model to collapse.

2.2 WGAN-GP model

In view of the problems in GAN model training, proposed a generative adversarial network (WGAN) based on Wasserstein distance (Arjovsky and Bottou 2017). The definition of Wasserstein distance is shown as follows:

Fig. 1 Basic framework of GAN



$$W(P_{\text{data}}, P_g) = \inf_{\gamma \in \Pi(P_{\text{data}}, P_g)} E_{(x,y) \sim \gamma} [\|x - y\|] \tag{4}$$

where $\Pi(P_{\text{data}}, P_g)$ denotes the set of all joint distributions $(x, y) \sim \gamma$ whose marginal distributions are respectively P_{data} and P_g . Intuitively, $(x, y) \sim \gamma$ calculates how much “mass” need to be transported from x to y in order to transform the distribution P_g into the distribution P_{data} . $W(P_{\text{data}}, P_g)$ all x and y satisfying the distribution, value of $\|x - y\|$ reaches the minimum, i.e., Wasserstein distance. Wasserstein distance reflects the distance between the two distributions, so it can effectively alleviate the gradient disappearance problem of the generator.

WGAN adopts the method of weight constraint on the Lipschitz constant problem, which improves the training process of GAN to a certain extent. Due to the interaction between WGAN weight constraint and loss function, the network training has the problems of difficult convergence and gradient explosion, and there is still a gap between the generated data and the real data.

Therefore, the improved model WGAN-GP is adopted in this paper, and the gradient penalty is used instead of weight pruning. The generator loss function is:

$$L_G = -E_{x \sim p_g} [f_w(x)] \tag{5}$$

where $f_w(x)$ is neural network.

In order to solve the mode collapse of the network and improve the convergence speed of the network, the gradient penalty is added to the discriminator loss function, and the generated samples are limited by Lipschitz. The discriminator loss function is:

$$L_D = E_{x \sim p_g} [f_w(x)] - E_{x \sim p_{\text{data}}} [f_w(x)] + \lambda E_{\hat{x} \sim p_{\hat{x}}} [(\|\nabla_{\hat{x}} f_w(\hat{x})\|_2 - 1)^2] \tag{6}$$

where λ represents the gradient penalty coefficient; $\|\cdot\|_2$ second norm of matrix; ∇ function of gradient; $p_{\hat{x}}$ generate a straight-line uniform sampling between the sample distribution $p_g(x)$ and the real data sample $p_{\text{data}}(x)$.

WGAN-GP model applies gradient penalty to the training network, improves the network convergence performance and the quality of generated data samples, makes the network training more stable and more universal under different network architectures. Therefore, this model uses

WGAN-GP to solve the problems of difficult convergence, gradient explosion and poor quality of generated data samples in the process of short-term wind power prediction.

3 Wind power prediction method based on BiLSTM–CNN–WGAN-GP (LCWGAN-GP)

In this section, the framework based on LCWGAN-GP network is illustrated in detail. When a single GAN network predicts the time series of wind power, it is prone to slow prediction speed and gradient disappearance. To overcome this phenomenon, the BiLSTM network with good learning performance is used as the generator, the CNN network with good nonlinear function is extracted as the discriminator, and the semi-supervised regression method is used to adjust the parameters of LCWGAN-GP network to better predict short-term wind power.

3.1 Prediction structure of LCWGAN-GP

The proposed prediction structure mixes different technologies of signal decomposition, sample generation, feature extraction and alternative training, as shown in Fig. 2. Firstly, variational mode decomposition (VMD) decomposes the original time series of wind energy data into subsequences of multiple intrinsic mode functions (IMFs) with different frequencies. Secondly, the generation model of LCWGAN-GP structure, BiLSTM network, is used to generate virtual wind power time series to obtain the data distribution characteristics of wind power. Then, the discrimination model of LCWGAN-GP structure is used to extract the CNN network with good nonlinear characteristics hidden in the wind power time series, and the semi-supervised regression is used to predict the wind power of subsequent layers. Finally, the model and discriminant model are generated by alternating iterative training, and the parameters of LCWGAN-GP are updated to the optimal state to minimize the prediction error.

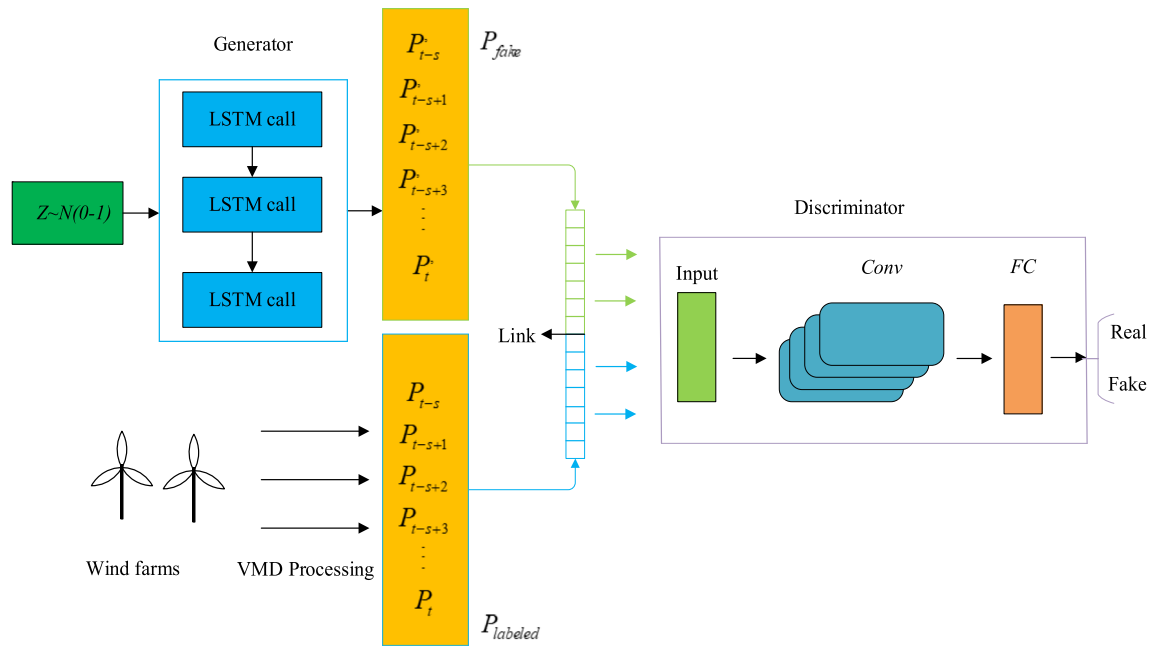


Fig. 2 Prediction structure of LCWGAN-GP

3.2 Variational modal decomposition

VMD is a new non-stationary signal adaptive decomposition method, which overcomes the problem of modal component aliasing in EMD method (Han et al. 2019). The original wind power data is nonlinear and unstable, which reduces the prediction results of wind power. Therefore, in this paper, VMD algorithm is used to decompose the collected nonlinear and non-stationary wind power data, and the original data are decomposed into different numbers of sub signals with limited bandwidth, eigenmode function. VMD decomposes a signal with strong fluctuation into K signals with obvious regularity, assuming that each modal component $u_k(t)$ has center frequency and limited bandwidth. The original wind signal $f(t)$ can be expressed as the sum of K IMF components, and the sum of the bandwidth of all modal components is the smallest. The IMF expression is:

$$u_k(t) = A_k(t) \cos(\phi_k(t)) \tag{7}$$

where $A_k(t)$ is instantaneous amplitude; $\phi_k(t)$ is Phase; k is modal component.

In VMD, different signal K values are different. In this paper, according to the K value setting rule in document (Ding et al. 2020), the K value is set to 3. Therefore, the original wind power generation sequence can be adaptively processed and decomposed into multiple subsequences based on VMD.

3.3 Construction of generation model

3.3.1 LSTM

In the traditional neural network, the historical data used in prediction is the information of N times before the prediction time. The historical information used in each prediction is rolling forward, which will cause the neglect of earlier information during prediction. The design of RNN can avoid this problem. LSTM is a special RNN, due to its unique design structure, LSTM network is very suitable for processing and prediction based on time series data. It also has strong fitting ability and feature extraction ability, and has good prediction effect. The storage unit of LSTM is equipped with forgetting gate, input gate and output gate to manage the removal or addition of storage unit. The three thresholds are composed of sigmoid activation function and point-by-point multiplication (Zhao et al. 2017), From Fig. 3.

The gate of LSTM network is represented by the follow:

$$i_t = \sigma(w_{ri}s_{t-1} + w_{ri}x_t + w_{ci}c_{t-1} + b_i) \tag{8}$$

$$f_t = \sigma(w_{rf}s_{t-1} + w_{rf}x_t + w_{cf}c_{t-1} + b_f) \tag{9}$$

$$c_t = f_t \times c_{t-1} + i_t \times \sigma(w_{rc}s_{t-1} + w_{xc}x_t + b_c) \tag{10}$$

$$o_t = \sigma(w_{ro}s_{t-1} + w_{xo}x_t + w_{co}c_{t-1} + b_o) \tag{11}$$

$$s_t = o_t \times \tanh(c_t) \tag{12}$$

At the current t time, x_t means input, s_{t-1} indicates the output of the previous moment, c_{t-1} indicates the hidden state of the previous moment, the value of the LSTM

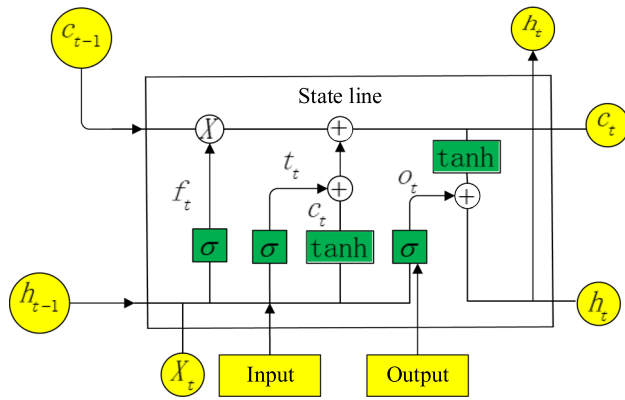


Fig. 3 Structure of LSTM

memory unit at the current time is c_t and the output value is s_t . Where, i_t, o_t, f_t , are the values of input gate, output gate and forgetting gate at time t respectively. w is the weight, b is the offset term, and σ is the activation function.

3.3.2 BiLSTM

Deep learning usually has ANN network structure, many model parameters, and requires a large number of training samples to make the algorithm converge. However, collecting a large number of samples is time-consuming and expensive. On the other hand, the traditional sample generation technology usually uses uniform sampling, which is basically the same as the original sample. Therefore, a new sample generation technology is needed to expand the training samples to enhance the generalization ability of deep learning algorithm.

Unidirectional LSTM has the important short time characteristics of very long interval and delay in processing and prediction time series, but it can only consider the information of past data for prediction. If the data of past several times are used for wind power prediction, it will only use several historical data closest to the prediction time and ignore the information contained in earlier historical data. BiLSTM has the ability to learn by using the information of past and future data, that is, to predict wind power by using past and future prediction data. It not only improves the prediction accuracy, but also shortens the prediction time. In order to effectively obtain the time variation characteristics of the input time series, the purpose of generator g is to learn the characteristic distribution of wind power. Therefore, this paper selects the bidirectional long-term and short-term memory network (BiLSTM) as the generator model to realize the long-term learning of data, including two LSTM layers (Kim and Moon 2019). The basic idea of BiLSTM is that the forward and backward of each training sequence are LSTM respectively, and these two layers are connected with the

input layer and the output layer, and the output layer integrates the past (forward) and future (reverse) information. The generator structure is shown in Fig. 4. Due to the complexity of LSTM training, the noise data of a series of fixed length sequences subject to Gaussian distribution are input into the generator. Each noise point is represented as a d -dimensional vector, the length of the sequence is T , the size of the input matrix is $T \times d$, and each layer has 100 units. Add a dropout layer after each layer to combine with the full connection layer.

The current hidden state depends on two hidden states, forward LSTM and backward LSTM. The output of the first BiLSTM layer at time t , as follows:

$$\vec{h}_t^1 = \tanh(w_{ih}^1 x_t + w_{hh}^1 \vec{h}_{t+1}^1 + b_h^1) \quad (13)$$

$$\overleftarrow{h}_t^1 = \tanh(w_{ih}^1 x_t + w_{hh}^1 \overleftarrow{h}_{t+1}^1 + b_h^1) \quad (14)$$

$$y_t^1 = \tanh(w_{ho}^1 \vec{h}_t^1 + w_{ho}^1 \overleftarrow{h}_t^1 + b_o^1) \quad (15)$$

The output depends on \vec{h}_t and \overleftarrow{h}_t , and h_o is initialized to a zero vector.

The second BiLSTM layer is used to obtain the output of time t , as follows:

$$y_t = \tanh(w_{ho}^2 \vec{h}_t^2 + w_{ho}^2 \overleftarrow{h}_t^2 + b_o^2) \quad (16)$$

The primary objective of generative model is to mimic the complex distribution of real samples in an unsupervised manner. Therefore, the GAN generative model can generate a deceptive virtual sample according to the random noise. The objective function of the unsupervised learning is expressed as:

$$L_G = \|E \cdot f(P_g) - E \cdot f(P_{\text{labeled}})\|^2 \quad (17)$$

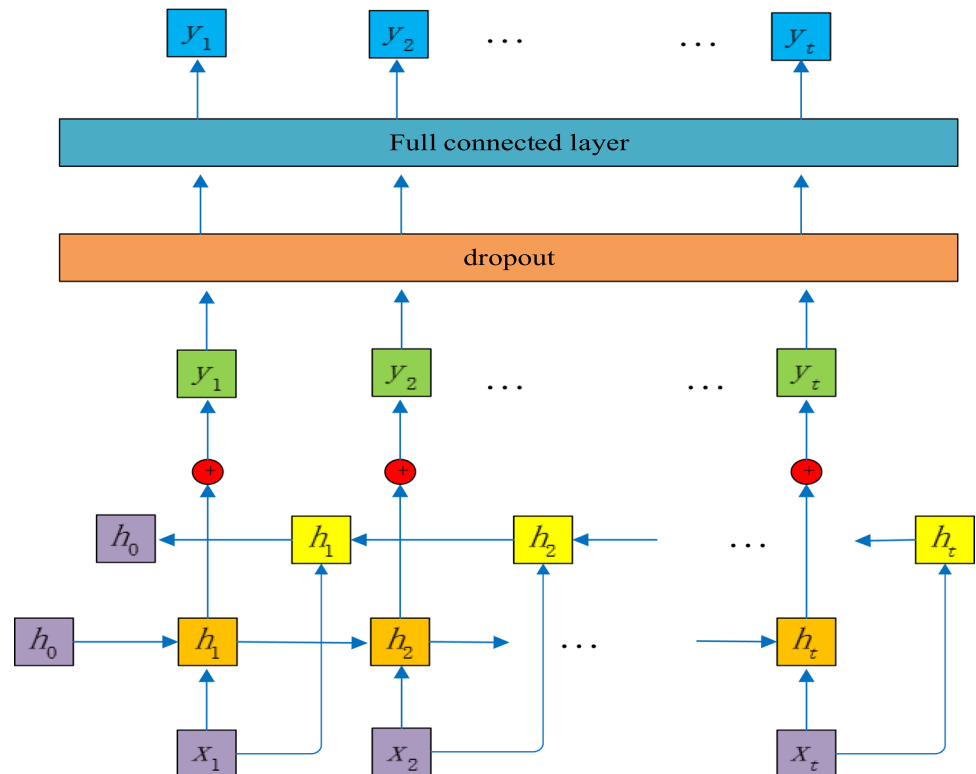
where $f(P_g)$ and $f(P_{\text{labeled}})$ denote the output of an intermediate layer of discriminative.

In the generation process of wind power time series, the input variables are the wind power series before the prediction point, wind speed, air density and roughness data. The wind power of the prediction point in the next 48 h is generated. By iteratively minimizing Eq. (17), the statistical features of the virtual samples and original samples tend to be more and more similar, and the generated virtual samples are regarded as real samples.

3.4 Construction of discriminant model

CNN is a commonly used neural network model in the field of deep learning, due to it has strong feature learning ability and can greatly reduce the number of parameters in the model, it is widely used in image recognition and other fields. CNN is a feedforward neural network, which is composed of input layer, convolution layer, pooling layer,

Fig. 4 Generation model framework of LCWGAN-GP



full connection layer and output layer. Because CNN adopts convolution operation in calculation, the operation speed is greatly improved compared with general matrix operation. The alternating use of convolution layer and pooling layer of CNN can effectively extract local features of data and reduce the dimension of local features; Due to weight sharing, the number of weights can be reduced and the complexity of the model can be reduced (Solas et al. 2019). In order to ensure the prediction accuracy of LCWGAN-GP method, convolutional neural network (CNN) with good fitting performance to nonlinear function is selected as the discrimination model to form the nonlinear mathematical relationship function between historical time series data and future wind power prediction value (Wu et al. 2021), as shown in Fig. 5. Due to the output of the generated model BiLSTM is one-dimensional data, CNN can not be directly used for signal processing of the time series of wind power data, so the data dimension conversion technique is adopted in this paper to solve this problem. The basic structure of LCWGAN-GP discrimination model can be expressed as:

- (1) The generated wind data and real wind samples are used as the input of the discrimination model;
- (2) Converting one-dimensional data samples (generated wind power data and real wind power data) into two-dimensional image samples as the input of convolution layer;

- (3) Multiple transpose convolution layers are used to map multiple low-dimensional images to high-dimensional data space, and the high-dimensional feature representation of input samples is obtained through a series of convolution operations;
- (4) The two-dimensional image features are converted into one-dimensional data, and the regression layer is connected to the full connection layer at the end of the discrimination model to realize the nonlinear regression of wind power data and obtain the results of wind power prediction.

The discriminant model includes feature extraction layer and regression layer. The feature extraction layer uses unsupervised learning mechanism, and the regression layer uses supervised learning mechanism. In order to enhance the regression ability of the designed discriminant model, the double objective optimization model is realized through alternating training, as follows:

$$L_D = L_{\text{unsup}} + L_{\text{sup}} \tag{18}$$

$$L_{\text{unsup}} = -\|E \cdot f(P_g) - E \cdot f(P_{\text{labeled}})\|^2 \tag{19}$$

$$L_{\text{sup}} = E \cdot (P_{t+1}^F - P_{t+1}^R)^2 \tag{20}$$

Here, P_{t+1}^F and P_{t+1}^R represent the wind power prediction result and the actual value at the next time. P_g and P_{labeled} represent the marked wind power sequence and the virtual wind power sequence.

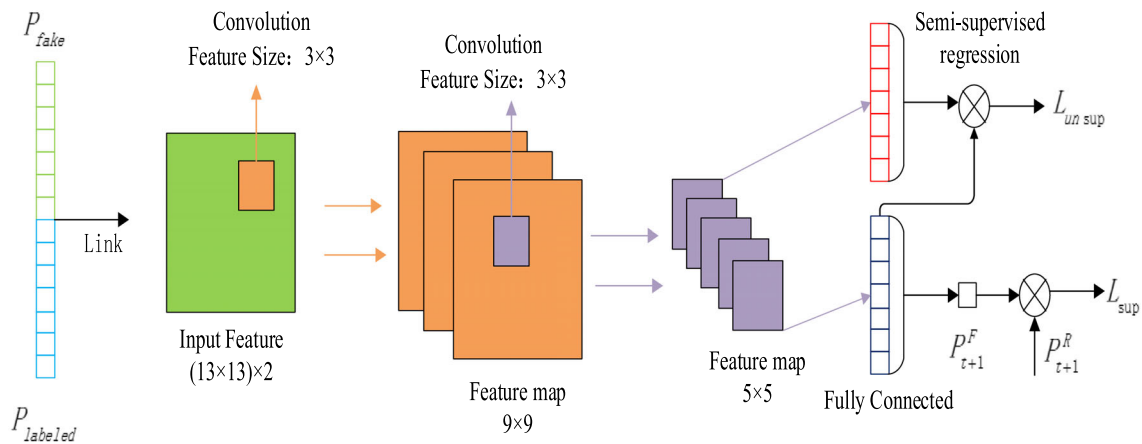


Fig. 5 Discriminant model framework of LCWGAN-GP

It can be found that Eq. (19) is an unsupervised loss function to make the statistical distribution of virtual samples and real samples as dissimilar as possible. Equation (20) is the supervised loss function, which is devoted to quantitatively assess the statistical deviations of the real wind power data as well as its predicted values. LCWGAN-GP generative and discriminative models are competing to form a min–max game. By introducing virtual samples into the LCWGAN-GP generation model, the discrimination model can more easily identify the potential characteristics in the original wind power samples, and make the prediction results of LCWGAN-GP closer to the real value.

3.5 Implementation of prediction method

This paper presents a hybrid prediction model based on VMD, BiLSTM, CNN and WGAN-GP for Short-term prediction of wind power. The overall prediction framework of the proposed prediction method is shown in Fig. 6.

Firstly, VMD algorithm is used to decompose the original wind power time series into multiple IMF signals and a residual signal. Then, the original signal is transformed into virtual samples through the generation model of LCWGAN-GP. Then, the virtual samples and IMF subsequences are input into the discriminant model, and the WPF results and prediction errors of each subsequence are obtained. Further, the prediction error is fed back to the LCWGAN-GP generation model and the generation parameters are updated, and each IMF subsequence with residual signal corresponds to an independent discrimination model. Alternate training to complete the LCWGAN-GP algorithm.

Compared with the previous WPF model, the proposed semi supervised regression LCWGAN-GP method has the following advantages:

- (1) WGAN-GP is essentially a general framework, which can be combined with other time series prediction models and has strong compatibility;
- (2) The BiLSTM model is used to generate samples to enhance the data and improve the prediction accuracy;
- (3) The CNN used in the discriminant model has better time series feature extraction ability and reduces the prediction error. Therefore, these advantages can effectively improve the prediction accuracy of the proposed short-term wind power prediction method based on LCWGAN-GP and reduce the prediction error.

4 Prediction model error evaluation index

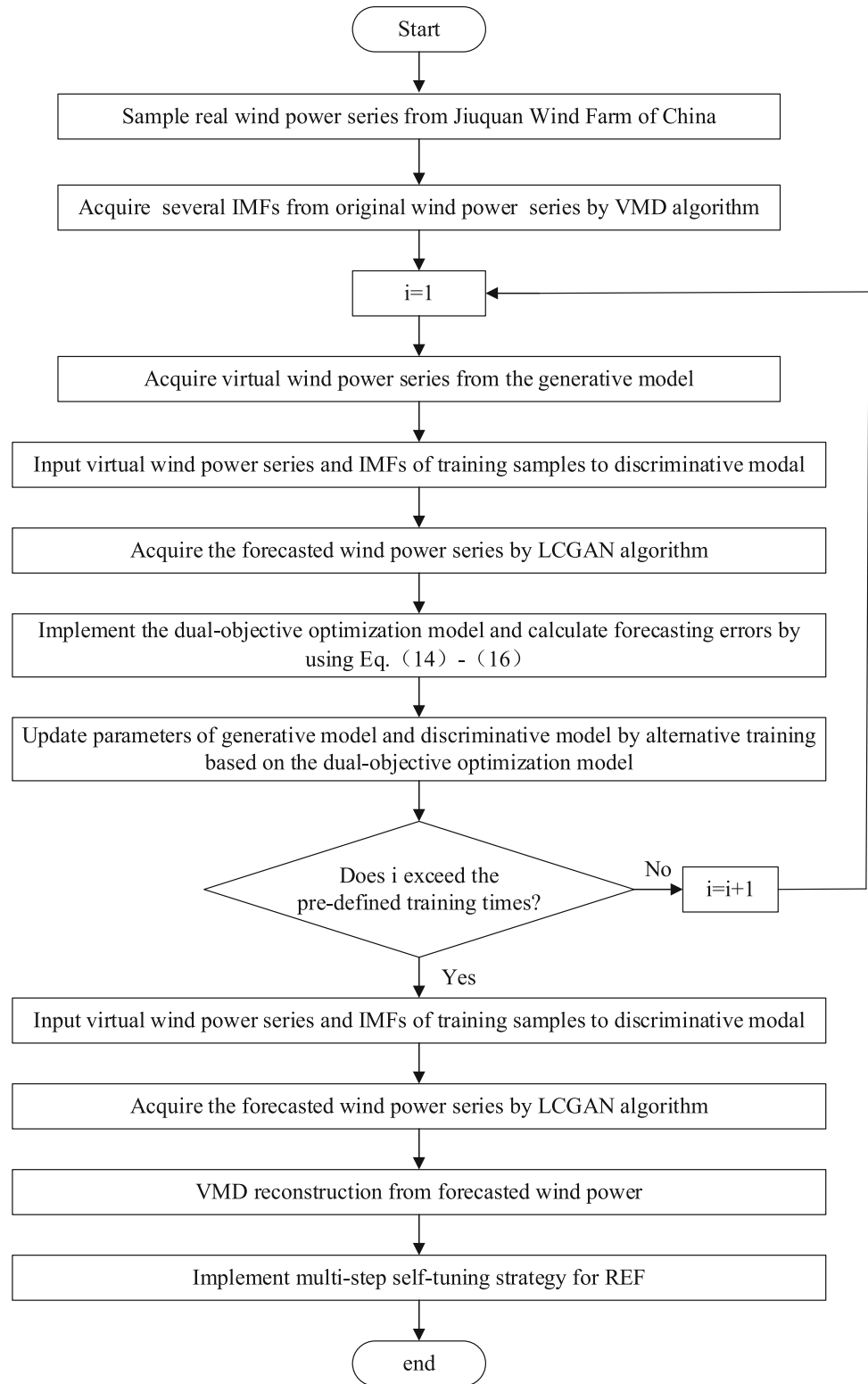
In order to comprehensively determine the superiority of the prediction model in this paper. The model evaluates the results of the prediction model by calculating the results of the error evaluation index (Li et al. 2021). In this paper, three indexes, mean absolute error (MAE), mean absolute percentage error (MAPE) and root-mean-square error (RMSE), are used to evaluate the performance of wind power prediction. Their definitions are given as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |P_i^F - P_i^R| \quad (21)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|P_i^F - P_i^R|}{P_i^R} \times 100\% \quad (22)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i^F - P_i^R)^2} \quad (23)$$

Fig. 6 Implementation framework of the proposed method



MAE and RMSE are used to evaluate the prediction ability and accuracy of the proposed prediction method. But MAPE is used to evaluate the deviation between the prediction results of relative points and the actual wind

output. The smaller the error is, the more accurate the prediction is and the better the prediction effect is.

5 Experimental results and comparative analysis

In this paper, the measured data set of a wind farm in Jiuquan, China is selected as the research object. The maximum power of a single unit in the wind farm is 2.5 MW and the time interval is 10 min. The seasonal wind power in spring, summer, autumn and winter is predicted. The data samples are four groups of data in March, June, September and December 2020. The real-time wind speed power curve after data exception processing is shown in Fig. 7. Each data set contains 1000 samples, the first 800 are training samples and the last 200 are test samples. The time series diagram of wind power historical data of wind farm is shown in Fig. 8 (taking September 2020 as an example). The prediction results of the proposed prediction method are compared with those of GAN, BiLSTM, CNN, GRU, SVM and ARIMA algorithms.

The input variables to generative adversarial network are the wind power sequence, wind speed and air density 24 h before the prediction point. The simulation platform used is as follows:

Processor and memory: Intel(R) Core(TM) i7-10750H CPU@2.6 GHz, 8 GB;

Operating system: 64-bit Windows 10;

Analysis software: Tensorflow2.5.0.

In order to improve the prediction accuracy, the power value is normalized in the range of $[-1, 1]$. Finally, the prediction results of GAN, BiLSTM, CNN, GRU, SVM and ARIMA are compared. The prediction results are shown in Figs. 9, 10, 11 and 12.

It can be seen from Figs. 9, 10, 11 and 12. That the prediction results based on LCWGAN-GP are very similar to the change trend of the actual wind power time series.

The proposed prediction method has the best prediction effect in spring, summer, autumn and winter, and has good adaptability and flexibility. According to the prediction results of different algorithms in each season, it can be seen concisely that the prediction effect based on improved LCWGAN-GP is the best, followed by the original GAN network, and the prediction effect of SVM and ARIMA is the worst. The error values of the proposed method and other algorithms in different seasons are compared. The error values of each model under different evaluation indexes are shown in Table 1.

As shown in Table 1, the error value of the proposed method in autumn is slightly smaller in the error comparison of four seasons, because the weather in autumn is relatively stable and there is no wind power climbing phenomenon. In general, the proposed method can be used to predict wind power all year round.

Taking autumn as an example, the results of MAE, RMSE and MAPE of the proposed method are 2.0236 MW, 3.1936 MW and 1.03%, respectively, with the best prediction performance among all the comparative algorithms. In addition, the MAE, RMSE and MAPE of the statistical ARIMA method are 14.3621 MW, 16.3698 MW and 8.19%, respectively. It can also be observed that ARIMA is not good at handling wind power data with high randomness. Compared with ARIMA, SVM improved MAE and MAPE by 8.6667 MW and 2.4%, respectively. Compared with CNN and GRU, BiLSTM performs relatively well due to its strong time series processing ability. The improvement of MAPE results by this method is 0.98% and 4.15%, respectively. Obviously, the prediction performance of BiLSTM is not as good as that of the proposed method. In MAE evaluation index, LCWGAN-GP is 0.8066 MW lower than GAN. In the RMSE evaluation index, LCWGAN-GP is 0.3705 MW lower than

Fig. 7 After exception data processing

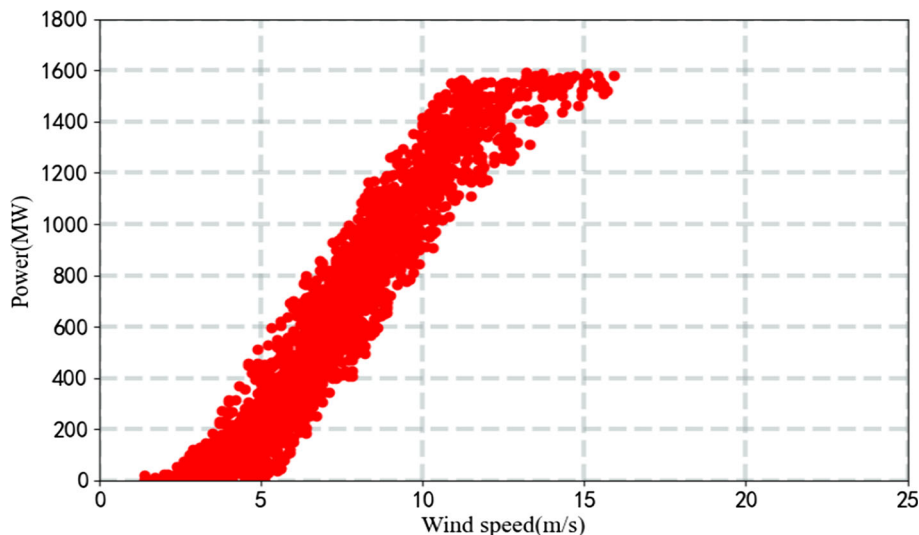


Fig. 8 Wind power time series

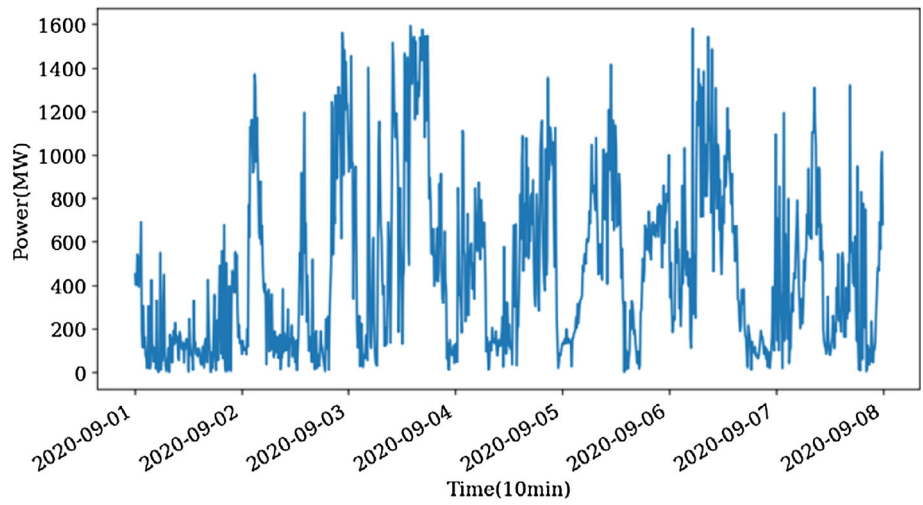


Fig. 9 Forecast results of wind power in spring

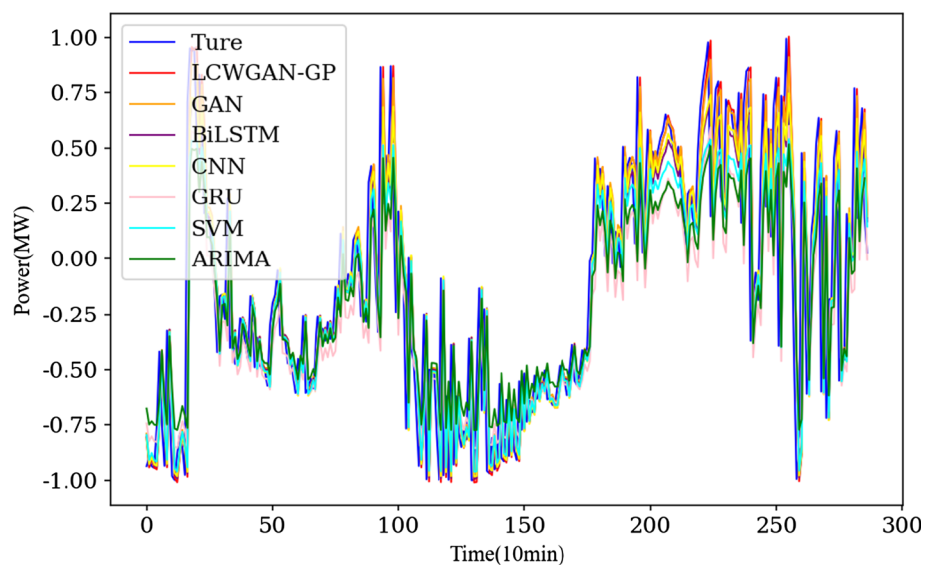


Fig. 10 Forecast results of wind power in summer

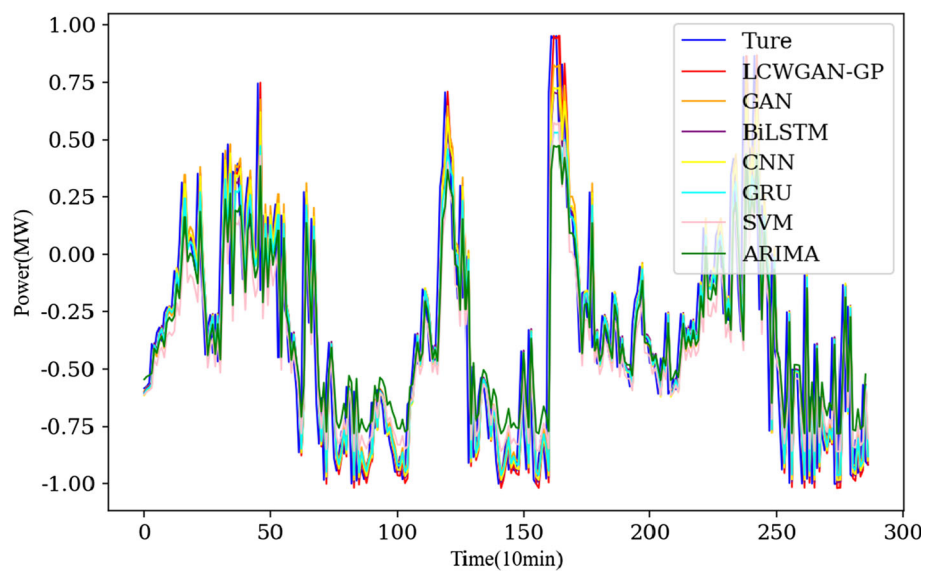


Fig. 11 Forecast results of wind power in autumn

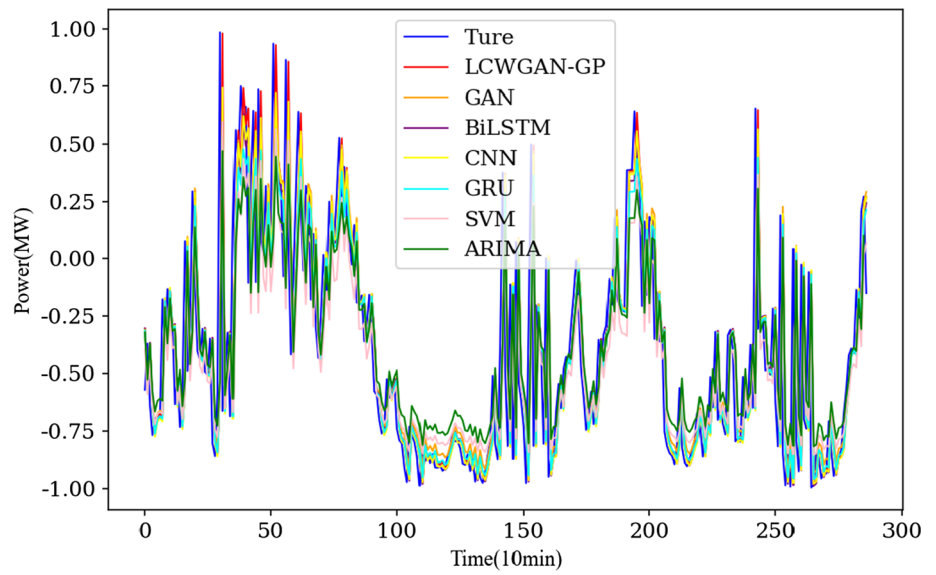


Fig. 12 Forecast results of wind power in winter

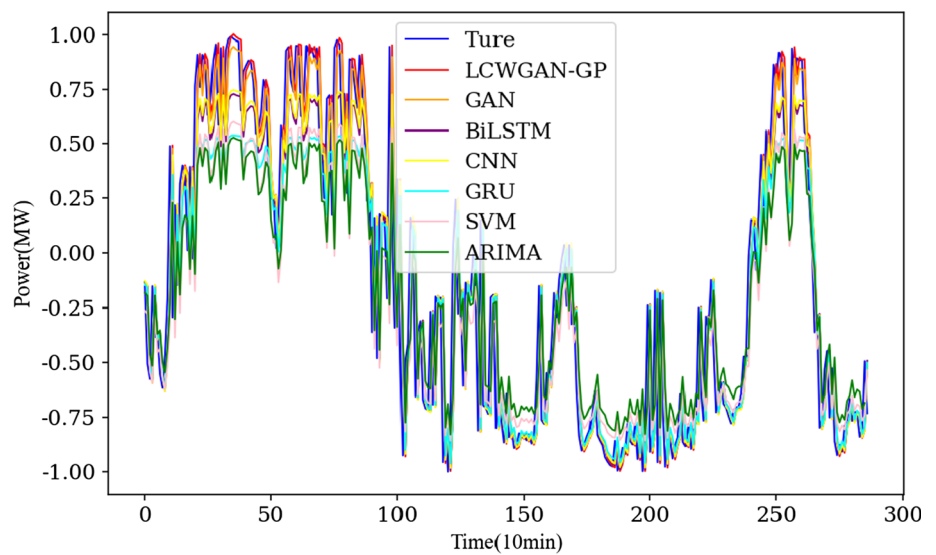


Table 1 Comparison of model prediction errors

Season	Index	LCWGAN-GP	GAN	BiLSTM	CNN	GRU	SVM	ARIMA
Spring	MAE	3.8102	3.8974	7.5692	8.2639	10.3415	8.0421	14.2369
	RMSE	4.1236	4.8631	12.2236	12.2327	14.3215	13.2469	17.1325
	MAPE	1.36%	1.74%	4.08%	4.28%	5.84%	4.56%	8.12%
Summer	MAE	3.3624	3.7614	7.8415	8.1214	9.7745	8.1126	15.7723
	RMSE	4.3524	4.7621	11.4732	12.0123	13.3619	13.3624	16.5874
	MAPE	1.54%	1.61%	3.98%	4.21%	5.77%	4.74%	8.25%
Autumn	MAE	2.0236	2.8302	5.7524	5.2301	6.5639	5.6954	14.3621
	RMSE	3.1936	3.5641	9.2365	9.4895	11.2369	9.7894	16.3698
	MAPE	1.03%	1.56%	3.13%	4.11%	7.28%	5.79%	8.19%
Winter	MAE	2.8225	3.2321	5.6321	5.4203	6.3654	5.8847	15.3654
	RMSE	3.0321	3.6645	9.2769	9.4412	11.5697	9.2635	15.4598
	MAPE	1.17%	1.59%	3.33%	4.18%	6.41%	6.26%	8.89%

GAN. In MAPE evaluation index, LCWGAN-GP decreased by 0.53% compared with GAN. Obviously, the prediction effect of LCWGAN-GP is better than GAN. Undoubtedly, the LCWGAN-GP based prediction method greatly outperforms other existing methods in all three evaluation metrics.

The real data of Jiuquan wind farm in September 2021 are used to verify the overall prediction performance of the proposed method in one-step prediction and multi-step prediction. Multi-step prediction is an iterative point prediction process, which is carried out in rolling mode. In the process of each iteration, the prediction result of one-step in the forward time step is used as the input of multi-step prediction at a certain time point. The prediction structure and parameter settings are the same as those above. Figures 13 and 14 show one-step prediction and multi-step prediction results, respectively. The error values of the proposed method and other benchmark algorithms are compared in Tables 2 and 3. In Table 2, MAE, RMSE and MAPE of the proposed method are 0.2894 MW, 0.3356 MW and 1.26%, respectively. In Table 3, MAE, RMSE and MAPE of the proposed method are 0.4754 MW, 0.8632 MW and 2.35%, respectively.

Table 4 shows the comparison of calculation time of all algorithms. It can be seen from Table 4 that the proposed method requires more prediction time than GAN, BiLSTM, CNN, SVM and ARIMA. This is because the proposed method uses VMD for signal decomposition and needs training to generate model and discrimination model. However, the time of LCWGAN-GP prediction algorithm is shorter than that of DBN prediction algorithm. DBN has a complex training process, including hierarchical training and fine tuning. Therefore, from the perspective of

prediction performance and efficiency, the proposed prediction algorithm based on LCWGAN-GP is reliable.

6 Conclusions

A prediction method based on BiLSTM–CNN–WGAN-GP is proposed in this paper to improve the prediction accuracy of short-term wind power. The results show that compared with a single GAN network for prediction of wind power, this method improves the prediction accuracy to a certain extent, which solves the problem of low prediction accuracy. The original wind power time series is decomposed into sub-sequences with smooth outer contours by using the VMD decomposition method. Each sub-sequence is used as the input of the discriminant model to reduce the parameters of the neural network and improve the speed of wind power prediction. The improved BiLSTM can capture the characteristic information of time series from two directions, making full use of time series and improving the prediction accuracy. The combined generative confrontation network is constructed by combining BiLSTM, CNN, and WGAN-GP. The generation model adopts the BiLSTM network, the discrimination model adopts the CNN network, and the semi-supervised regression method. The generation model and discrimination model are trained repeatedly, and finally, the combined generation countermeasure network converges and the prediction results are obtained. Compared with the GAN, BiLSTM, CNN, GRU, SVM, and ARIMA networks, it has higher prediction accuracy and less time. The research at this stage is based on historical data on wind power. In the next step, the physical climate information from the wind

Fig. 13 One-step WPF results of the proposed approach

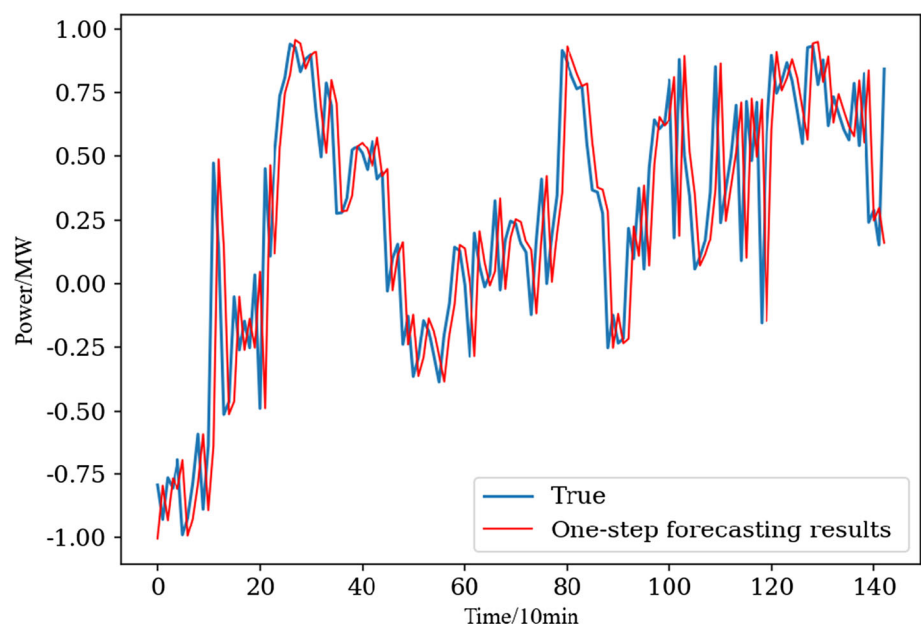


Fig. 14 Multi-step WPF results of the proposed approach

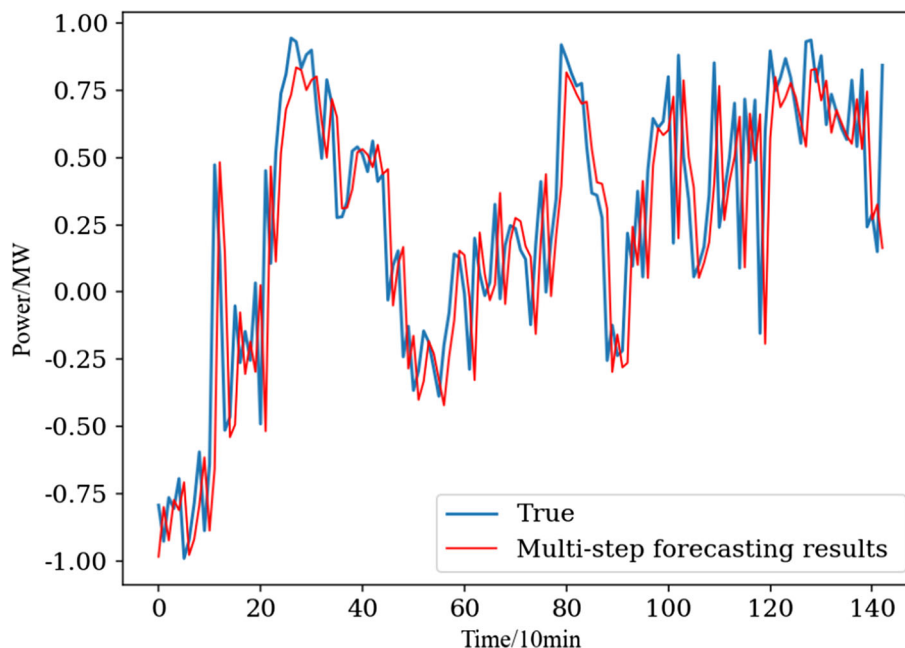


Table 2 One-step WPF indices with different forecasting methods

Index	LCWGAN-GP	GAN	BiLSTM	CNN	DBN	SVM	ARIMA
MAE	0.2894	0.7845	1.0236	1.1424	1.6515	1.8958	2.3695
RMSE	0.3356	0.8423	1.3625	1.8966	2.2536	2.7632	3.5469
MAPE	1.26%	2.65%	3.14%	5.59%	6.25%	8.23%	9.36%

Table 3 Multi-step WPF indices with different forecasting methods

Index	LCWGAN-GP	GAN	BiLSTM	CNN	DBN	SVM	ARIMA
MAE	0.4754	1.2568	2.3968	2.8714	3.1962	3.8854	5.5714
RMSE	0.8632	2.2354	2.9687	3.9864	4.2369	4.8751	6.6984
MAPE	2.35%	3.76%	4.54%	6.88%	8.24%	9.28%	10.37%

Table 4 Computational time of seven prediction method

	LCWGAN-GP	GAN	BiLSTM	CNN	DBN	SVM	ARIMA
Runtime/s	18.2574	14.2538	4.5213	10.5476	19.3254	4.5369	5.5548

farm will also be used as a characteristic input to establish a more perfect wind power prediction system.

Funding The Science and Technology Key Research and Development Projects in Gansu Province, ultra-wideband radar life detection system and key technology research, 20YF3GA018.

Declarations

Conflict of interest The authors declared that they have no conflicts of interest to this work.

Ethical approval The paper does not deal with any ethical problems.

Informed consent We declare that all authors have informed consent.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arjovsky M, Bottou L (2017) Towards principled methods for training generative adversarial networks. arXiv preprint [arXiv:1701.04862](https://arxiv.org/abs/1701.04862)
- Cali U, Sharma V (2019) Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. *Int J Smart Grid Clean Energy* 8(2):103–110
- Chen Y, Dong Z, Wang Y et al (2021) Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history. *Energy Convers Manag* 227:113559
- Chuang LI, Xiangyu K, Xingguo W et al (2019) Short-term wind power forecasting method based on mode decomposition and feature extraction. In: 2019 22nd International conference on electrical machines and systems (ICEMS). IEEE, pp 1–5
- Creswell A, White T, Dumoulin V et al (2018) Generative adversarial networks: an overview. *IEEE Signal Process Mag* 35(1):53–65
- Ding J, Xiao D, Li X (2020) Gear fault diagnosis based on genetic mutation particle swarm optimization VMD and probabilistic neural network algorithm. *IEEE Access* 8:18456–18474
- Duan J, Wang P, Ma W et al (2021) Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and correntropy long short-term memory neural network. *Energy* 214:118980
- Goodfellow I, Pouget-Abadie J, Mirza M et al (2020) Generative adversarial networks. *Commun ACM* 63(11):139–144
- Han L, Zhang R, Wang X et al (2019) Multi-step wind power forecast based on VMD-LSTM. *IET Renew Power Gener* 13(10):1690–1700
- Hao Y, Tian C (2019) A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Appl Energy* 238:368–383
- Hong YY, Satriani TRA (2020) Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network. *Energy* 209:118441
- Hu YL, Chen L (2018) A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm. *Energy Convers Manag* 173:123–142
- Husein AM, Arsyal M, Sinaga S et al (2019) Generative adversarial networks time series models to forecast medicine daily sales in hospital. *Sinkron: jurnal dan penelitian teknik informatika* 3(2):112–118
- Jiajun H, Chuanjin Y, Yongle L et al (2020) Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning. *Energy Convers Manag* 205:112418
- Jiang Y, Liu S, Zhao N et al (2020) Short-term wind speed prediction using time varying filter-based empirical mode decomposition and group method of data handling-based hybrid model. *Energy Convers Manag* 220:113076
- Jung J, Broadwater RP (2014) Current status and future advances for wind speed and power forecasting. *Renew Sustain Energy Rev* 31:762–777
- Kavasseri RG, Seetharaman K (2009) Day-ahead wind speed forecasting using f-ARIMA models. *Renew Energy* 34(5):1388–1393
- Khodayar M, Wang J, Manthouri M (2018) Interval deep generative neural network for wind speed forecasting. *IEEE Trans Smart Grid* 10(4):3974–3989
- Kim J, Moon N (2019) BiLSTM model based on multivariate time series data in multiple field for forecasting trading area. *J Ambient Intell Human Comput* 33:1–10
- Kong W, Dong ZY, Jia Y et al (2017) Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans Smart Grid* 10(1):841–851
- Li L, Cen ZY, Tseng ML et al (2021) Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic-support vector regression machine. *J Clean Prod* 279:123739
- Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint [arXiv:1411.1784](https://arxiv.org/abs/1411.1784)
- Ren Y, Suganthan PN, Srikanth N et al (2016) Random vector functional link network for short-term electricity load demand forecasting. *Inf Sci* 367:1078–1093
- Shi X, Lei X, Huang Q et al (2018) Hourly day-ahead wind power prediction using the hybrid model of variational mode decomposition and long short-term memory. *Energies* 11(11):3227
- Solas M, Cepeda N, Viegas JL (2019) Convolutional neural network for short-term wind power forecasting. In: 2019 IEEE PES Innovative smart grid technologies Europe (ISGT-Europe). IEEE, pp 1–5
- Sun W, Wang Y (2018) Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Convers Manag* 157:1–12
- Wang Y, Li H (2018) A novel intelligent modeling framework integrating convolutional neural network with an adaptive time-series window and its application to industrial process operational optimization. *Chemom Intell Lab Syst* 179:64–72
- Wang J, Yang Z (2021) Ultra-short-term wind speed forecasting using an optimized artificial intelligence algorithm. *Renew Energy* 171:1418–1435
- Wang K, Gou C, Duan Y et al (2017) Generative adversarial networks: introduction and outlook. *IEEE/CAA J Autom Sin* 4(4):588–598
- Wu Q et al (2021) Ultra-short-term multi-step wind power forecasting based on CNN-LSTM. *IET Renew Power Gener* 15(5):1019–1029
- Yu R, Gao J, Yu M et al (2019) LSTM-EFG for wind power forecasting based on sequential correlation features. *Futur Gener Comput Syst* 93:33–42
- Zeng J, Qiao W (2012) Short-term wind power prediction using a wavelet support vector machine. *IEEE Trans Sustain Energy* 3(2):255–264
- Zhang J, Wei Y, Tan Z (2020) An adaptive hybrid model for short term wind speed forecasting. *Energy* 190:115615
- Zhao Z, Chen W, Wu X et al (2017) LSTM network: a deep learning approach for short-term traffic forecast. *IET Intel Transport Syst* 11(2):68–75

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.