Article:

Liu, Z.-H. orcid.org/0000-0002-6597-4741, Wang, C.-T. orcid.org/0000-0002-0772-1623, Wei, H.-L. orcid.org/0000-0002-4704-7346 et al. (3 more authors) (2024) A wavelet-LSTM model for short-term wind power forecasting using wind farm SCADA data. Expert Systems with Applications, 247. 123237. ISSN 0957-4174

https://doi.org/10.1016/j.eswa.2024.123237

Revised manuscript (with changes marked)

A Wavelet-LSTM Model for Short-Term Wind Power Forecasting Using Wind Farm SCADA Data

Zhao-Hua Liu[a,*], Chang-Tong Wang[a, b], Hua-Liang Wei[c], Bing Zeng[b], Ming Li[a], Xiao-Ping Song[b]

[a] School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, 411201, China

[b] Harbin Electric Corporation Wind Power Co., Ltd., Xiangtan, 411102, China

[c] Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK

Fina Accepted Manuscript. Published:
Expert Systems with Applications. Volume 247, 1 August 2024, 123237

* Corresponding author. Tel: +8615874160936

E-mail addresses: zhaohualiu2009@hotmail.com (Z.-H. Liu), ctwang@mail.hnust.edu.cn (C.-T. Wang), w.hualiang@sheffield.ac.uk (H.-L. Wei), zengbing2016@126.com (B. Zeng), minglee@hnu.edu.cn (M. Li), 48078841@qq.com (X.-P. Song).

Abstract

Supervisory Control and Data Acquisition (SCADA) system collects massive operation and environment information which directly or indirectly affects the output power in wind farms. Therefore, it becomes an imperious demand to analyze the underlying information from SCADA data for improving the performance of short-term wind power prediction. In this paper, an effective deep learning framework for short-term wind power forecasting based on SCADA data analysis is proposed. A data denoising scheme is designed based on wavelet decomposition. In this method, all SCADA signals (except the wind power signal itself) are decomposed into low-frequency component A and high-frequency component D respectively by the wavelet transform. Then, the maximum information coefficient (MIC) method is applied to choose features that have strong correlation with wind power. Finally, all the selected features and wind power are defined as input vector that are used to train long short-term memory networks. The simulation results based on real data extracted from a SCADA system installed in wind farm indicate that the designed deep learning framework can significantly improve the accuracy of short-term wind power prediction.

Keywords: Deep learning, long short-term memory networks, maximum information coefficient, SCADA, short-term wind power prediction, wavelet transform.

1. Introduction

The installed capacity of wind turbines and the grid connected capacity of wind power have increased year by year. Due to the characteristics of fluctuation and uncertainty of wind, the power grid fluctuates when large-scale wind power is connected to the grid, which may lead to safety accidents and affect the operation stability of the power grid (Pujari, Miriyala, Mittal, & Mitra, 2023). Short-term wind power forecasting is regard as a significant and effective technique for better carrying out power grid dispatching and decision-making, stabilizing the power grid, and applying wind energy efficiently (Guan, Wang, Liu, Gao, Xu, & Kan, 2024). Short-term wind power prediction with prediction horizons of less than 6 hours is usually considered by analyzing the historical data related to a wind power (Safari, Chung, & Price, 2018). Data preprocessing (e.g., feature selection, denosing, feature extraction) plays an important role in improving short-term wind power forecasting performance (Cheng, Zang, Xu, Wei, & Sun, 2021, Du, 2019, Wang, Zhu, Cheng, Zhou, Zhang, Xu, & Liu., 2023).

Many wind power prediction methods have been developed recently, which can be roughly divided into two categories: 1) physical methods, and 2) data-driven methods (Akhtar, Kirmani, Ahmad, & Ahmad, 2021). Physical methods, such as numerical weather prediction, predict the meteorological change trend of a region by building the corresponding physical model, adopting the information of topography environment and meteorological phenomena as the modeling data (Zjavka & Mišák, 2018). Ye *et al.* (Ye, Zhao, Zeng,& Zhang, 2017) developed a physical model based on the spatial correlation for wind turbines of wind farm and calculated the wind speed by solving the constructed differential equation. Finally, the wind power was estimated according to a power curve model. Commonly, the physical methods generate the rough prediction and are applicative for medium and long-term prediction (Miao, Li, Wang, & Li, 2021).

With the development of data acquisition equipment and the operation of wind farm, a wealth of wind power data have been collected. Therefore, data-driven methods provide an effective alternative solution. The typical data-driven methods mainly include traditional time series modelling methods, Kalman filtering method, and artificial intelligence methods. Time series methods use statistical analysis to extract the development trend of time series with time and extrapolate the possible future values (Yu, Ma, Ma, & Zhang, 2022). Typical time series methods include autoregressive (AR) model, AR moving average (ARMA) model, and AR integrated moving average (ARIMA) model (Meng, Chen, Ou, Ding, Zhou, Fan, & Yin, 2022). Jiang *et al.* (Jiang, Chen, Yu, & Liao, 2017) incorporated a boosting algorithm into the ARMA model to improve the ability for multi-step ahead wind power prediction. However, these methods require to assume that the time series to be modelled are linear and stationary, which are not suitable for nonlinear and nonstationary processes, such as wind power data (Hossain, Chakrabortty, Elsawah, Gray, & Ryan, 2021). Kalman filtering methods are a recursive filtering approach, employing a two stage scheme, that is, prediction and correction, to estimate the variation of wind power. Zhang *et al.* (Zhang, Yue, Dou, Li, & Hancke, 2022) utilized a Kalman filtering algorithm to perform the rough wind power prediction in long time scale, and used a reinforcement learning algorithm to correct the prediction results in short time scale. However, it was observed that it is difficult to estimate the noise of wind power series using Kalman filtering methods (Zhang, Liu, Zhang, Yan, Zang, Wu, Hua, & Wang, 2019).

With the in-depth development of artificial intelligence technology, many machine learning methods have now been applied to perform the wind power prediction tasks (Yin et al., 2021, Li et al., 2022). Li *et al.* (Li, Lin, Xu, Liu, & Liu, 2018) presented a wavelet decomposition and support vector machine (SVM) method based on an atomic search algorithm to predict the decomposed components of wind power. Xue *et al.* (Xue, Jia, Wen, & Farkoush, 2020) tested the prediction ability of an improved Gaussian process (GP) approach for both probabilistic and deterministic prediction of wind power from different perspectives. Different from SVM and GP, artificial neural networks (ANN) have adaptive and self-organizing ability to learn latent features from the input and output data (Xiong, Lou, Meng, Wang, Ma, & Wang, 2022). Tan *et al.* (Tan, Han, & Zhang, 2020) built an extreme learning machine (ELM) to predict the wind power, and Salp swarm algorithm was introduced to search the parameters of ELM. An *et al.* (An, Yin, Wu, She, & Chen, 2021) adopted the weighted naive Bayes technique to analyze the wind speeds from three different numerical weather prediction (NWP) organizations, and build a back propagation neural network (BPNN) to learn the relationship of wind power and wind speed. Abedinia *et al.* (Abedinia, Lotfi, Bagheri, Sobhani, Shafie-khah, & Catalão, 2020) used an improved empirical mode decomposition (EMD) algorithm to process the wind data and a bagging neural network was built to perform the prediction task, where chaotic binary shark smell optimization was utilized to optimize the bagging neural network and obtained more accurate wind power prediction results. Although ANNs have been applied to solve wind power prediction problem, simple ANNs cannot adequately mine useful information from complex wind power data (Zhou, Wang, Lu, & Zhao, 2022).

Deep learning due to its powerful self-learning ability, has been developed to perform wind power prediction tasks recently (Wang, Tao, Hu, & Zeng, 2021). For example, Peng *et al.* (Peng et al., 2021) designed a wind power prediction model based on convolutional neural network (CNN), where the NWP and wind power data were decomposed by wavelet transform (WT) first, and then the decomposed components were rearranged according to their frequencies. Chen *et al.* (Chen, Zhu, Li, Zhu, Shi, Li, Duan, & Liu, 2020) designed a complex stacked autoencoder framework to learn the heterogeneous features related to wind power and achieve higher prediction accuracy than other compared methods. Compared with other deep learning techniques, long short-term memory (LSTM) network can remember information of a time series for a period of time, and has been extensively applied to predict wind power (Wang, Zhu, Zhang, Cheng, & Zhou, 2023). Ye *et al.* (Ye, Dai, Pei, Lu, Zhao, Chen, & Wang, 2022) designed a sequence prediction model LSTM to perform wind power prediction based on intraday NWP data. In their model, a wave-oriented scheme and a modified fuzzy-C means algorithm were used to select the appropriate input data. Ko *et al.* (Ko, Lee, Kim, Hong, Dong, & Hur, 2021) introduced a multi-level residual network into the bi-directional LSTM networks (BLSTM) to void the over-fitting problem in wind power prediction task, which fused the long and short BLSTM networks to learn the change of peak values. As a version of LSTM, the gate recurrent unit (GRU) network has more simplified gate units (Wang, Li, Fu, & Tang, 2020). Han and Wu (Yan & Wu, 2020) used empirical WT to decompose the raw wind power data. Then, a GRU network model was used to predict the components with low sample entropy values, and kernel ELM was adopted to process the components with high sample entropy values.

The aforementioned methods have achieved good results by digging up the latent features of the signals used at a certain extent. Signals used in these methods mainly include wind power, wind speed, wind direction, temperature, atmospheric pressure, NWP data, etc. However, the output power of wind turbine depends not only on environmental factors, but also on the historical operation state of wind turbine (Lin & Liu, 2020). Supervisory Control and Data Acquisition (SCADA) system is a widely used real-time monitoring system in wind farms; it collects a large number of environmental and operation data of wind turbine (Morrison, Liu, & Lin, 2022). Therefore, it is essential to dig up SCADA data effectively for improving the prediction performance of wind power. Liu *et al.* (Liu, Yang, & Zhang, 2021) analyzed the SCADA data from three dimensions (i.e., space, time, and physical meaning), where the spatial and physical information was extracted by combining K-means, K-shape, and CNN, and the temporal information was learned by GRU networks. In addition, the preprocessing for SCADA data, such as outlier detection and resampling, has also been investigated in the literature. For example, Lin *et al.* (Lin, Liu, & Collu, 2020) introduced an isolation forest algorithm to detect outliers of SCADA data first and deep learning neural networks were then established to mapping corrected SCADA data to wind power. Kisvari *et al.* (Kisvari, Lin, & Liu, 2021) preprocessed the SCADA data to remove the outliers by using an isolation forest algorithm, and then compared the wind power prediction performance of GRU and LSTM after performing feature engineering.

Although there have been many studies on wind power prediction at present, achieving more accurate wind power prediction still faces the following challenges: 1) Raw SCADA data contains plenty of complex information, even noise information. Many methods mentioned previously directly learn latent features from SCADA data to train a prediction model, and do not fully consider the effect of the 'hidden' information (e.g., noise) when design the prediction model, which may affect the prediction performance of the model; 2) SCADA data contains a large number of monitoring parameters, which are not all helpful in improving the accuracy of wind power prediction. Using these parameters without filtering for wind power prediction may even reduce the accuracy of the prediction. Therefore, it is necessary to design a more efficient and practical correlation analysis method to screen parameters that contribute to prediction tasks.

To further exploit the hidden information in SCADA data and the strong learning ability of LSTM for time series modelling and wind power forecasting, this paper proposes an effective deep learning framework for short-term wind power forecasting based on LSTM networks and SCADA data denoising. In many methods, raw SCADA data are directly used to drive the prediction model. However, as mentioned earlier, SCADA system records plenty of complex information from wind turbine, even the information about the SCADA system itself. These SCADA data contain a wealth of noise information, which makes it difficult to well capture the real relationships between wind power and other related signals using whatever prediction models. Therefore, in this proposed framework, a SCADA data denoising method based on WT is designed, where the other features potentially related to wind power are decomposed and reconstructed by WT, so that the SCADA data are holistically denoised. Then, a feature selection process based on the maximum information coefficient (MIC) method is performed to choose more relevant features. Finally, a multi-feature prediction model based on LSTM networks are established to represent the relationship between the other features and wind power; the model is then used to realize the wind power forecasting. The main contributions of this study are as follows:

1) An effective deep learning prediction framework is proposed. The framework integrates data denoising and feature selection, and adopts the deep learning algorithm for LSTM networks to autonomously dig up the latent features and perform wind power prediction task.

2) A reliable SCADA data denoising method based on WT is designed. The method denoises the SCADA data as a whole, making them easier to learn by the prediction model.

3) A MIC based method is integrated to calculate the correlation between the relevant features and wind power, and then the multi-feature LSTM networks are built to capture the latent relationship between the selected features and wind power.

The rest of this study is organized as follows. The characteristics of SCADA data are analyzed in Section 2. The proposed deep learning framework based on LSTM networks and SCADA data denoising for wind power forecasting are introduced in Section 3. The case studies and experimental results are presented in Section 4, and the work is concluded in Section 5.

2. Analysis for the characteristics of SCADA data

2.1. Basic framework of SCADA system

As an important module of wind farms, SCADA system aims to monitor the operating status of wind turbines in real time by using sensors installed in various systems of wind turbines, ensuring that these wind turbines can operate normally under harsh conditions. As shown in Fig. 1, SCADA system mainly consists of on-site monitoring system, central monitoring center, and remote monitoring center. The on-site monitoring system of wind turbines is generally deployed in the tower control cabinet of each wind turbine to achieve the wind turbine monitoring and control and real-time collection and transmission of operating data for wind turbine. The central monitoring center is arranged in control room of wind farm. The on-site personnel can understand the operation of wind turbines at any time and carry out control based on the operation data of wind turbines. The remote monitoring center can communicate with the monitoring center of wind farm through the remote communication network, and implement unified management and data analysis for all wind farms. SCADA system can collect a large amount of operating data from wind turbines, and the SCADA data mainly includes the following types of information: 1) Environmental information, such as wind speed, wind direction, and ambient temperature; 2) Transmission information, such as rotor speed and generator speed; 3) Electrical information, such as currents, voltages, grid frequency, power, and power factor; 4) Perceived information, such as temperature and vibration signals of various components, and 5) other information, such as yaw position and hydraulic system pressure (Udo & Muhammad, 2021).

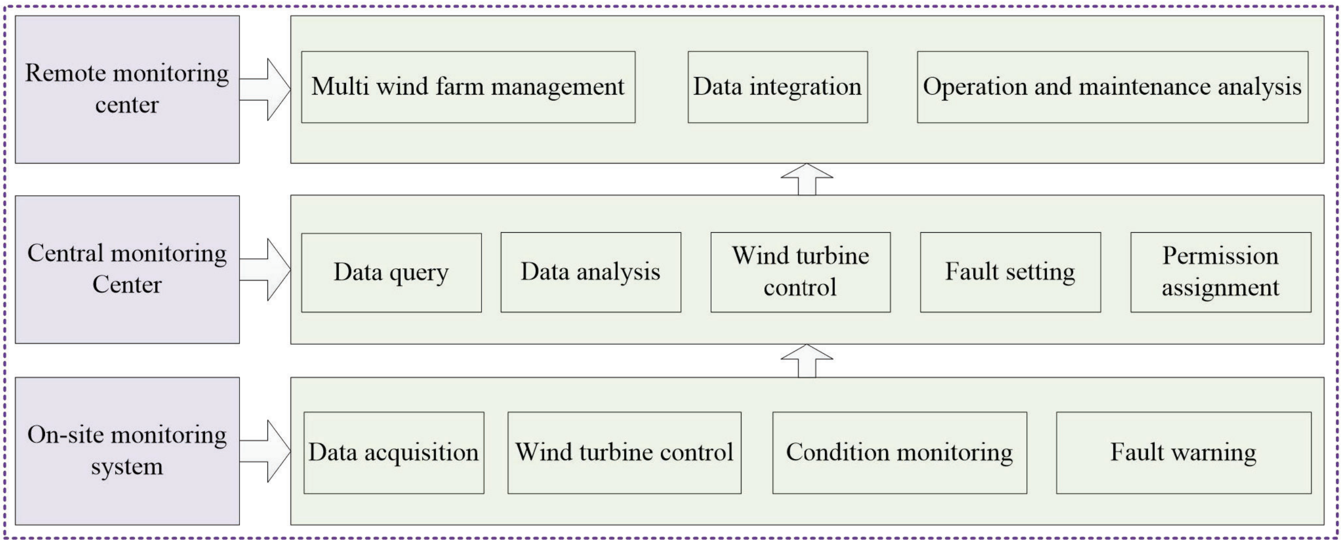


Fig. 1. Basic framework of the SCADA system in wind farm.

2.2. Characteristics of SCADA data

The environment in which wind turbines operate is constantly changing, resulting in varying operating conditions of wind turbines and increased sensor errors. In addition, the wind turbines themselves have complex structures, leading to the SCADA system needing to monitor a large number of parameters. Therefore, SCADA data also presents the following characteristics: 1) Strong randomness. As a major source of energy for wind power generation, wind resource has a certain degree of randomness and volatility, and changes in wind resources directly lead to changes in the parameter information of the internal components of the entire wind turbine, and the parameter information of each component in the system also has randomness and volatility. 2) Excessive noise. The SCADA system records signals such as temperature and vibration signals of wind turbines, which are easily affected by external environments. Additionally, the performance of the sensors themselves is also easy to change, leading to the integration of various noise information into the monitored data. 3) There are differences in the correlation between different parameters. The SCADA system measures dozens to hundreds of parameters, but wind turbines are a huge nonlinear operating system built from multiple subsystems. The correlation between various system parameters is not completely consistent. 4) Temporal correlation: When the wind turbine operates continuously, the actions of the wind turbines are constantly changing, so the generated operating data has strong temporal correlation (Zhang, Li, & Zhao, 2023).

SCADA data contains a large amount of feature information about the state of wind turbines, which can be used for various analysis tasks such as power prediction and fault diagnosis. However, due to some of the above characteristics and difficulties, the analysis of SCADA data lacks an efficient and reliable framework and system. Taking into account the characteristics of SCADA data, this paper proposes an effective deep learning framework for short term wind power forecasting based on LSTM networks and SCADA data denoising to effectively perform power prediction tasks. This method improves the accuracy of power prediction by globally denoising SCADA data, mining important information, and analyzing the correlation between parameters. The specific details are presented in section 3.

3. Proposed deep learning framework

3.1. The structure of the proposed method

The structure of the proposed method is depicted in Fig. 2, which includes three parts: 1) SCADA data denoising based on WT algorithm, 2) feature selection based on MIC, and 3) multi-feature prediction model based on LSTM networks. In the denoising process, one-level WT algorithm is performed to decompose each of the relevant features into low-frequency component A and high-frequency component D. Only the component A is used to reconstruct the feature data; the component D, which contains complex noise information, will be discarded. The MIC values between the relevant features and the wind power are calculated, and the features with high correlation with wind power signal are selected as the input data of the multi-feature LSTM networks. The details of the proposed framework are further explained in the following sections.

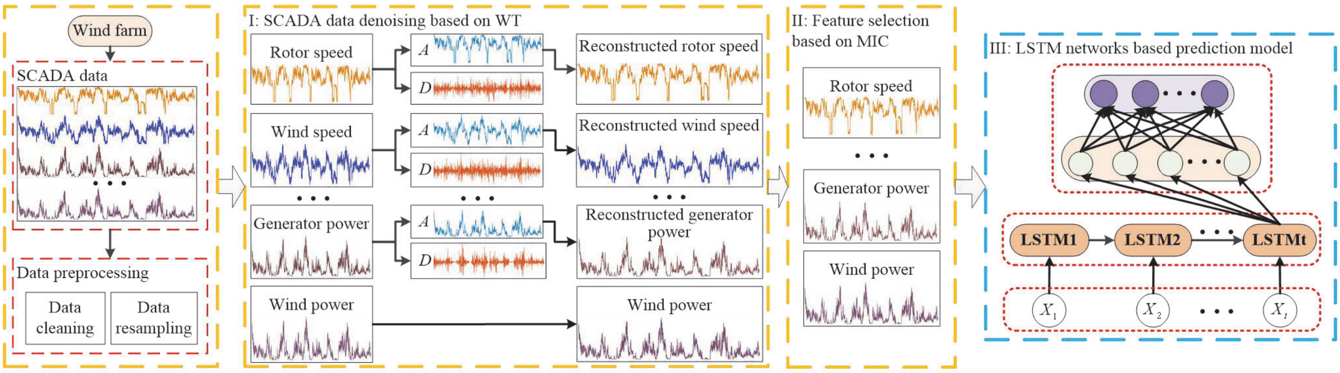


Fig. 2. Basic framework of the proposed method.

3.2. SCADA data denoising using WT

The SCADA system comprehensively records wind turbine signals and meteorological environment data. SCADA data are noisy and contain redundant information, which make it difficult to find best models for wind power prediction no matter what model structures and what training approaches are employed. Therefore, it is important to properly denoise SCADA data. This work proposes an effective SCADA data denoising method based on WT.

Compared to Fourier transform, WT provides an adaptive analysis method in both time and frequency domains. It decomposes the original time series to low-frequency and high-frequency components, and further decomposes the low-frequency component to another level of lower-frequency components; in this way, the signal is decomposed to many levels (Chui, 1992). The high-frequency components usually contain the noise information. In this way, WT can filter out noise at different scales.

For a wavelet basis function $\Phi(t)$ and an original signal $S(t)$, the continuous WT is defined as:

$$Coe_S(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} S(t)\Phi^*\left(\frac{t-b}{a}\right)dt, \qquad a > 0 \tag{1}$$

where $Coe_S(a,b)$ is the wavelet coefficients, and a and b are scale and translation parameters, respectively. $\Phi^*\left(\frac{t-b}{a}\right)$ is a conjugate function of $\Phi\left(\frac{t-b}{a}\right)$.

The reconstruction formula is:

$$S(t) = \frac{1}{P_\psi} \int_0^\infty \int_{-\infty}^{\infty} \frac{1}{a^2} \frac{1}{\sqrt{a}} Coe_S(a,b)\Phi\left(\frac{t-b}{a}\right)dadb \tag{2}$$

where the permissibility conditions are:

$$P_\psi = \int_{-\infty}^{\infty} \frac{|\Phi(\omega)|^2}{\omega} d\omega < \infty \tag{3}$$

In this designed denoising method, one-level WT algorithm is used to decompose all SCADA signals (except historical wind power data) to obtain low-frequency component A and high-frequency component D. The component D contains the noise information of original feature data, and thus only the component A is reconstructed to obtain new denoised feature data. The feature denoising process is shown in panel I of Fig. 2. It can be seen that the data of the features (e.g., rotor speed, wind speed, and generator power) are decomposed by the designed denoising method and the denoised data are obtained. Through this process, most high frequency noise can be removed. Note that the target, i.e., the wind power time series, is not preprocessed but directly used to train prediction models.

3.3. Feature selection based on MIC method

In the operation of wind turbines, a large number of parameters recorded by SCADA system change over time. However, not all these features are useful for wind power prediction, some of which are redundant and make no contribution to help improving wind power prediction performance. In addition, too many input features will also increase the running time of the program. Therefore, it is necessary to select important features, from SCADA data, that are highly correlated to wind power, so as to improve the prediction accuracy. In this paper, MIC is used for feature selection.

The MIC is developed from mutual information (MI) theory. When MI calculates the correlation between variables, it needs to discretize the variables, and it is deeply affected by the discretization effect. In addition, the results of MI has a lower limit of 0, but no upper limit, so it cannot be normalized. It is difficult to compare the dependence of between different variables according to a certain threshold. MIC, however, not only can characterize the linear and nonlinear relationships between variables, but also has good robustness and low complexity. The basic calculation process of MIC is as follows (Reshef et al., 2011):

1) Compose variables M and N into a two-dimensional data $\{(m_i, n_i)\}_{i=1}^T$, where T represents the total number of data points.

2) Select the I and J and form different grid schemes G, where I and J represent dividing the two-dimensional space into I rows and J columns, respectively. The data points in $\{(m_i, n_i)\}_{i=1}^T$ are located in different small grids.

3) Calculate the probability distribution of data in each small grid to obtain the corresponding MI values:

$$I_g(I,J) = \sum_{i\in I}\sum_{j\in J} P(i,j)\log_2 \frac{P(i,j)}{\sum_{i\in I} P(i,j)\sum_{j\in J} P(i,j)} \tag{4}$$

where $g \in G$, and $P(i, j)$ represents the probability of data points in the grid of row x and column y.

4) Calculate the MIC values in all grids G:

$$MI_G(I,J) = \max_{g \in G}(I_g(I,J)) \tag{5}$$

5) The maximum mutual information obtained from (5) is normalized into [0, 1]:

$$NMI_G(I,J) = \frac{MI_G(I,J)}{\log_2 \min\{I,J\}} \tag{6}$$

6) Calculate the MIC value for the grids when I and J are different:

$$MIC = \max_{IJ<B(T)}(NMI_G(I,J)) \tag{7}$$

where $B(T)$ is a function of T, usually taken as:

$$B(T) = (T)^{0.6} \tag{8}$$

By calculating the MIC values between other SCADA feature data and wind power data, the features with high MIC value are selected as the input features.

3.4. Prediction model based on LSTM networks with multi-features

Assuming that there are ρ samples in total in the SCADA data set after feature selection, and each sample includes f relevant features and a wind power value, the data set X is as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_\rho \end{bmatrix} = \begin{bmatrix} x_1, p_1 \\ x_2, p_2 \\ \cdots \\ x_\rho, p_\rho \end{bmatrix} = \begin{bmatrix} x_{11}, x_{12}, \cdots, x_{1f}, p_1 \\ x_{21}, x_{22}, \cdots, x_{2f}, p_2 \\ \cdots \\ x_{\rho 1}, x_{\rho 2}, \cdots, x_{\rho f}, p_\rho \end{bmatrix} \tag{9}$$

where X_i represents the ith sample, x_i denotes the relevant features of ith sample, p_i represents the ith wind power value, and x_{ij} represents the jth relevant feature value of ith sample.

The wind power prediction task is to forecast the value of wind power at the future time instant t+l by using the past t samples. The input and output data can be written as follows:

$$INPUT\_X = \begin{bmatrix} IX_1 \\ IX_2 \\ \cdots \\ IX_{\rho-t-l+1} \end{bmatrix} = \begin{bmatrix} X_1, X_2, \cdots, X_t \\ X_2, X_3, \cdots, X_{t+1} \\ \cdots \\ X_{\rho-t-l+1}, X_{\rho-t-l+2}, \cdots, X_{\rho-l} \end{bmatrix} \tag{10}$$

$$OUTPUT\_Y = \begin{bmatrix} OY_1 \\ OY_2 \\ \cdots \\ OY_{\rho-t-l+1} \end{bmatrix} = \begin{bmatrix} p_{t+1}, p_{t+2}, \cdots, p_{t+l} \\ p_{t+2}, p_{t+3}, \cdots, p_{t+l+1} \\ \cdots \\ p_{\rho-t-l+1}, p_{\rho-t-l+2}, \cdots, p_\rho \end{bmatrix} \tag{11}$$

According to (10) and (11), when $l = 1$, the prediction task is single-step prediction; when $l \geq 2$, the prediction task is multi-step prediction.

Then, these data are formatted according to the input and output shapes of the prediction model. In the time series prediction task, each sample has f+1 features. This study designs multi-feature LSTM networks to process the time relationship in the SCADA data.

LSTM network (Fig. 3) is a classic recurrent neural network (RNN), which is typically composed of a memory unit consisting of a forgetting gate $F(t)$, an input gate $I(t)$, an output gate $O(t)$, and a state element $\alpha(t)$ to control information selection, thus achieving the function of forgetting or memory. It effectively solves the long-term dependency problem of RNN and avoids the problem of gradient vanishing or exploding (Lu, Liu, Wei, Chen, Zhang, & Li, 2021) The specific calculation process of LSTM network is as follows:

$$F(t) = \sigma(U_{FH}H(t-1) + U_{FX}X(t) + B_F) \tag{12}$$

$$I(t) = \sigma(U_{IH}H(t-1) + U_{IX}X(t) + B_I) \tag{13}$$

$$\tilde{\alpha}(t) = \tanh(U_{\alpha H}H(t-1) + U_{\alpha X}X(t) + B_\alpha) \tag{14}$$

$$\alpha(t) = F(t) \otimes \alpha(t-1) + I(t) \otimes \tilde{\alpha}(t) \tag{15}$$

$$O(t) = \sigma(U_{OH}H(t-1) + U_{OX}X(t) + B_O) \tag{16}$$

$$H(t) = O(t) \otimes \tanh(\alpha(t)) \tag{17}$$

where U and B are the corresponding weights and biases, respectively. $H(t)$ is the hidden state in the moment t, $X(t)$ is the input data, and $\tilde{\alpha}(t)$ and $\alpha(t)$ are the candidate state and the cell state, respectively. $\sigma()$ and $\tanh()$ are the activation functions.

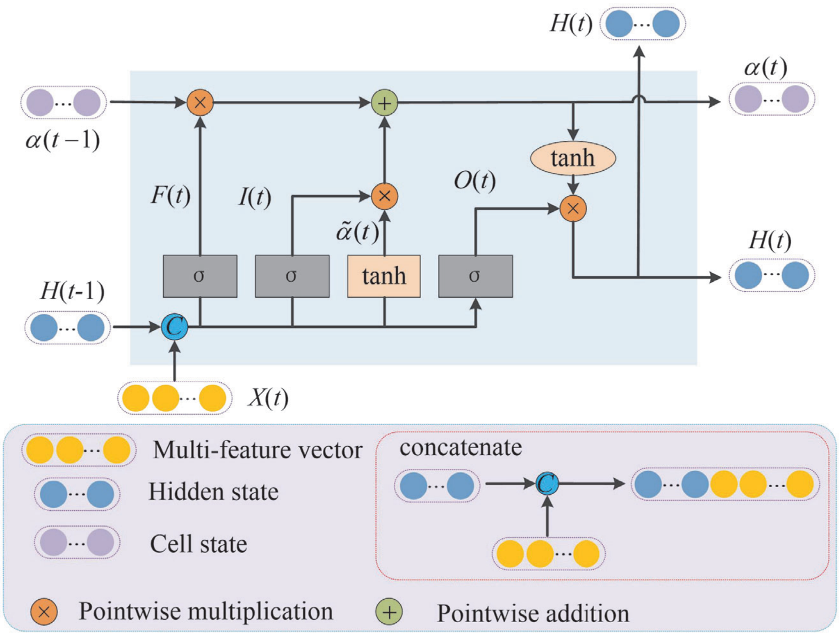


Fig. 3. LSTM network with multiple feature inputs.

From Fig. 3, $X(t)$ as the feature vector in the moment t, contains multiple features, and the concatenation operation concatenates the hidden state in the moment t-1 $H(t$-1$)$ with $X(t)$ to obtain a new vector. The new vector integrates the past information of multiple features and is fed into the LSTM network for further deep fusion, thereby obtaining deeper information of multiple features.

Take the first input IX_1 of *INPUT_X* as an example (see the panel III of Fig. 2). Through the LSTM layer, the latent features of IX_1 are mined, and the prediction results are calculated in the output layer. The multi-feature LSTM networks can integrate the input features at an abstract level and analyze the temporal relationship hidden in the time series.

3.5. Procedure of the proposed framework

According to the above descriptions, the procedure of the proposed wind power prediction method is summarized as follows:

Step 1: Data collection and preprocessing of SCADA data.

Step 2: Denoise the SCADA data. Use the proposed SCADA data denoising algorithm to decompose each feature of SCADA data (except wind power signal itself) to obtain two groups of components (low-frequency component A and high-frequency component D). The component A is reconstructed to obtain the denoised feature data.

Step 3: Feature selection. After denoising, the SCADA data are divided into training and test sets. The MIC values between relevant features and wind power are calculated on the train set, and then the required features are selected according to the correlation strength.

Step 4: Wind power prediction. The selected features and the historical wind power data are integrated as the input data to feed into multi-feature LSTM prediction model, and the prediction performance can be verified on the test set.

4. Case studies and discussions

4.1. Data description

To examine the performance of the proposed wind power prediction method, a real-world data set from a wind farm is considered. The wind farm is located in the south of China, where SCADA system is installed to collect data, with a sampling period of 15 minutes. In this study, the data for a total of 54 parameters from the real-world SCADA data set are selected to validate the proposed method. These parameters cover various key components of the turbines, including pitch system, nacelle system, generator system, yaw system, etc. All the parameters and their units are listed in Table 1. The active power is chosen to be the target signal for representing the wind power.

Table 1. Parameters selected from SCADA data set in this case studies

Parameter	Unit	Parameter	Unit	Parameter	Unit
Rotor speed	r/min	Generator stator temperature 1/2/3/4/5/6	°C	Twist angle 1/2	turns
Rotor angle	°	Generator air temperature 1/2	°C	Yaw brake hydraulic pressure	bar
Angle of blade 1/2/3	°	Generator current	A	Wind speed	m/s
Current of pitch motor 1/2/3	A	Temperature of INU	°C	Wind direction	°
Speed detection value of overspeed sensor	r/min	Temperature of ISU	°C	Air pressure	atm
Hub temperature	°C	Temperature of INU RMIO	°C	Current on grid side of the converter	A
Hub cabinet temperature	°C	Estimated power of pitch motor 1/2/3	W	Voltage on grid side of the converter	V
Acceleration in x/y direction	m/s^2	Temperature of converter box of blade 1/2/3	°C	Reactive power on grid side of the converter	kVar
Nacelle temperature	°C	Cooling liquid inlet/outlet temperature	°C	Power on generator side of the converter	kW
Main bearing Temperature 1/2	°C	Generator torque	kNm	Ambient temperature	°C
Nacelle cabinet temperature	°C	Cooling pump inlet/outlet pressure	bar	Generator frequency	Hz
Yaw alignment value	°	Active power	kW		

4.2. Parameter settings and benchmarking models

The determination of parameters of LSTM network models is important. In wind power forecasting task, the active power at the current moment is related to historical data of a period of time. In this study, the historical data from the past 24 hours were selected as the input of the model, i.e., the time steps of the LSTM model are 96. Then, the number of neurons in the hidden layers of the LSTM prediction model needs to be determined. According to the trial and error method, the number of neurons in the hidden layer is selected from 5 to 50, with an interval of 5.

For better evaluating the superiority of this proposed method, the comparison is carried out by choosing some existing prediction algorithms, including SVM, BPNN, CNN-LSTM (Zhang, Zhao, & Du, 2021), GRU-LSTM (Hossain, Chakrabortty, Elsawah, & Ryan, 2020), gradient boosting regressor (GBR) (Sobolewski, Tchakorom, & Couturier, 2023), and deep concatenated residual networks (DRNets) (Ko et al.,, 2021), and the parameters are determined by following the best prediction results. For SVM method, the radial basis function is selected as the kernel function, and the epsilon value is selected from 0.0001 to 0.1. For BPNN method, the number of neurons in the hidden layer is selected from 5 to 50, with an interval of 5. For CNN-LSTM method, one-dimensional CNN is used, and the number of neurons in the hidden layer of LSTM networks is selected from 5 to 50, with an interval of 5. For GRU-LSTM method, the output of the GRU networks is used as the input of the LSTM networks. For GBR method, the number of boosting stages to be performed is 50 and the learning rate is chosen to be 0.1. For DRNets method, the multi-level residual network is introduced in the bidirectional LSTM networks and the number of the layers in the bidirectional LSTM networks is 2. The number of iterations for all the deep learning methods is 50, and the batch size is 16.

4.3. Evaluation metrics

In order to reasonably evaluate the performance of the proposed method, normalized root mean square error (NRMSE) and normalized mean absolute error (NMAE) are used as the evaluation metrics, which can be expressed as

$$NRMSE = \frac{1}{P_{install}} \sqrt{\frac{1}{m}\sum_{i=1}^{m}(P_i - \hat{P}_i)^2} \times 100\% \tag{18}$$

$$NMAE = \frac{1}{P_{install}} \frac{1}{m}\sum_{i=1}^{m}|P_i - \hat{P}_i| \times 100\% \tag{19}$$

where $P_{install}$ represents the installed capacity of a wind turbine. P_i and $\hat{P}_i$ represent the ith actual and predicted values of wind power, respectively. m is the number of wind power data.

4.4. Case I: single-step prediction

1) Comparative experiments and analysis: To test the prediction ability of the proposed method, the single-step wind power prediction experiments are conducted based on the SCADA data, i.e., using the historical data to predict the value of the next moment. In this case, the original SCADA data contain a large amount of noise data. The proposed SCADA data denoising algorithm based on WT can effectively denoise the SCADA data. Following procedure described in the section 3.2, all the monitored parameters (except the active power signal itself) are decomposed using the one-level WT. Then the low-frequency components are reconstructed to new data. The denoised data set is divided into 12 data sets by month, and thus a total of 12 comparison tasks are divided based on the 12 data sets. In each task, the data of the last two days of the corresponding data set are used as the test set, and the other data of the month are used as the training set. Then, the feature selection process is performed to reduce the feature dimensions for each task. The MIC values between the active power and the monitored parameters are calculated. The monitored parameters with a MIC value greater than or equal to 0.5, along with the active power, are chosen to be the input features. As an example, the selected monitored parameters of the first task (1 month) are given in Table 2, and the number of the selected monitored parameters based on the 12 data sets is recorded in Table 3. It can be clearly seen that in different data sets, the number of the selected monitored parameters is different, which means that the correlation between these parameters and active power is not constant, and the feature selection is necessary in different prediction tasks.

Table 2. Selected monitored parameters of the first task (1 month)

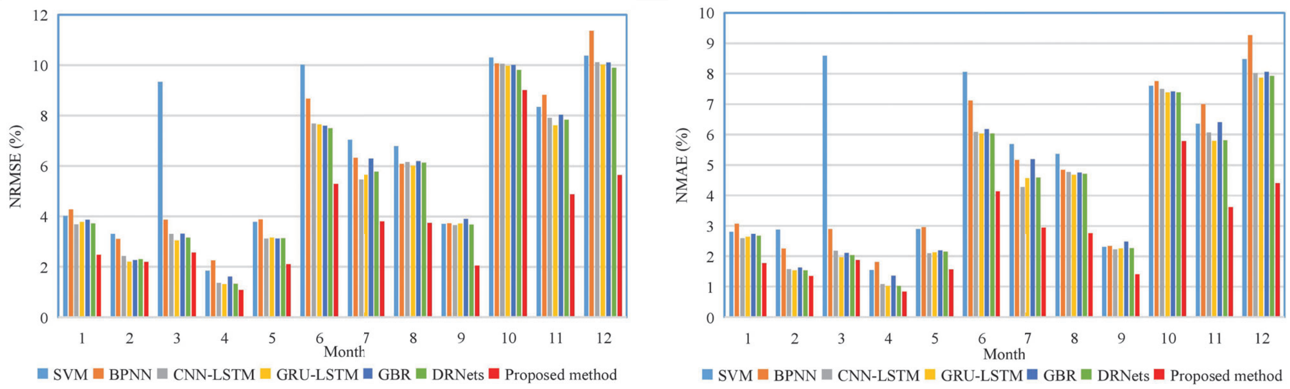
Monitored parameters	MIC	Monitored parameters	MIC
Rotor speed	0.9157	Current of pitch motor 2	0.8464
Speed detection value of overspeed sensor	0.9156	Current of pitch motor 1	0.8314
Generator frequency	0.9130	Estimated power of pitch motor 1	0.6934
Power on generator side of the converter	0.9096	Estimated power of pitch motor 3	0.6929
Generator current	0.9090	Estimated power of pitch motor 2	0.6721
Temperature of INU	0.9089	Angle of blade 1	0.6627
Generator torque	0.9031	Angle of blade 2	0.6621
Wind speed	0.8795	Angle of blade 3	0.6583
Current of pitch motor 3	0.8791	Main bearing temperature 2	0.6136
Temperature of ISU	0.8779	Main bearing temperature 1	0.6082

Table 3. Number of the selected monitored parameters based on the 12 monthly data sets

Month	1	2	3	4	5	6	7	8	9	10	11	12
Number of the selected monitored parameters	21	30	25	20	27	28	28	24	17	29	30	32

Table 4. Single-step prediction results of all the methods based on the data sets of 12 monthly (%)

Methods	Metrics	Month												Mean
		1	2	3	4	5	6	7	8	9	10	11	12	
SVM	NRMSE	4.0235	3.3029	9.3407	1.8508	3.7908	10.0172	7.0459	6.7849	3.7073	10.3091	8.3375	10.3836	6.5745
	NMAE	2.8050	2.8775	8.5910	1.5465	2.8978	8.0671	5.6923	5.3580	2.3101	7.6035	6.3492	8.4855	5.2153
BPNN	NRMSE	4.2794	3.1092	3.8685	2.2609	3.8888	8.6696	6.3281	6.0956	3.7390	10.0748	8.8198	11.3627	6.0414
	NMAE	3.0707	2.2522	2.8999	1.8149	2.9555	7.1183	5.1692	4.8462	2.3471	7.7577	6.9975	9.2664	4.7080
CNN-LSTM	NRMSE	3.6822	2.4239	3.2983	1.3612	3.1236	7.6914	5.4618	6.1697	3.6675	10.051	7.9031	10.1185	5.4127
	NMAE	2.5923	1.5755	2.1784	1.0836	2.1010	6.0905	4.2736	4.7779	2.2343	7.5010	6.0697	8.0261	4.0420
GRU-LSTM	NRMSE	3.7850	2.2095	3.0422	1.3205	3.1611	7.6548	5.6648	6.0259	3.7258	9.9692	7.6156	10.0256	5.3500
	NMAE	2.6442	1.5302	1.9680	1.0181	2.1312	6.0378	4.5717	4.6871	2.2640	7.3920	5.7926	7.8636	3.9917
GBR	NRMSE	3.8756	2.2703	3.3215	1.6109	3.1196	7.6040	6.2891	6.1964	3.9110	10.0085	8.0298	10.1038	5.5284
	NMAE	2.7361	1.6313	2.1097	1.3560	2.1864	6.1816	5.1831	4.7565	2.4854	7.4162	6.4025	8.0622	4.2089
DRNets	NRMSE	3.7247	2.3035	3.1526	1.3258	3.1346	7.5053	5.7820	6.1398	3.6863	9.8046	7.8352	9.8931	5.3573
	NMAE	2.6736	1.5370	2.0341	1.0240	2.1527	6.0390	4.5936	4.7135	2.2751	7.3846	5.8103	7.9347	4.0144
Proposed method	**NRMSE**	**2.4750**	**2.1901**	**2.5583**	**1.0852**	**2.1038**	**5.2969**	**3.8134**	**3.7438**	**2.0608**	**9.0128**	**4.8817**	**5.6471**	**3.7391**
	NMAE	**1.7737**	**1.3469**	**1.8755**	**0.8378**	**1.5697**	**4.1321**	**2.9486**	**2.7652**	**1.4071**	**5.7790**	**3.6237**	**4.4069**	**2.7055**



(a) (b)

Fig. 4. Prediction results of all the methods based on the 12 monthly data sets. (a) NRMSE values. (b) NMAE values.

All the methods are compared under the same conditions. The best results based on the metrics, NRMSE and NMAE, are given in Table 4, where the “Mean” column represents the mean values of the corresponding metric in 12 tasks. Fig. 4 shows a graphical illustration of the overall performance of the seven methods for the 12 tasks.

As can be seen from Table 4 and Fig. 4, the prediction performance of the proposed method outperforms all the six compared methods for all the 12 tasks. Compared with the simple machine learning methods, SVM and BPNN, the proposed method shows much better performance. For the 12 tasks, the NRMSE and NMAE values of the proposed method are around 12.57% and 24.00% (10 month) less than that of SVM, respectively. Similarly, the two metrics values of the proposed methods are around 10.54% and 25.51% (10 month) less than that of BP, respectively. From Fig. 4, it is obvious that in 3 month, SVM has the largest NRMSE and NMAE values, which are 265.11% and 358.06% larger than that of the proposed method. In most tasks, the proposed method can achieve significantly better results, such as 5 month, 9 month, and 12 month. This may be explained that for simple machine learning methods do not have time memory capabilities, and their ability to mine deep features of time series is relatively poor, making them difficult to store and transmit time information in SCADA data effectively. Different from the simple machine learning methods, the GBR method, as an integration algorithm, can integrate a bunch of poor prediction algorithms and obtain better prediction results. It can be seen from Table 4 that the mean values of the NRMSE and NMAE metrics for GBR method are 5.5284% and 4.2089%, respectively, which are smaller than those of SVM and BP.

The CNN-LSTM, GRU-LSTM, DRNets, and the proposed method are designed based on deep learning techniques, and can obtain better performance than the simple machine learning methods. As can be seen from Table 4, the mean values of the NRMSE metric of the these methods in 12 tasks are 5.4127%, 5.3500%, 5.3573%, and 3.7391%, respectively, which are smaller than that of SVM (6.5745%) and BP (6.0414%). Correspondingly, the mean values of the NMAE metric of these methods in 12 tasks are 4.0420%, 3.9917%, 4.0144%, and 2.7055%, which are also smaller than that of SVM (5.2153%) and BP (4.7080%). In addition, compared to the other deep learning methods, the proposed method shows better performance in terms of the two evaluation metrics. For example, the NRMSE values of the proposed method are around 9.65% (2 month), 0.88% (2 month), and 4.92% (2 month) smaller than that of CNN-LSTM, GRU-LSTM, and DRNets methods, respectively. Correspondingly, the NMAE values of the proposed method are around 13.90% (3 month), 4.70% (3 month), and 7.80% (3 month) smaller than those of other three methods, respectively. The improvement in the prediction performance can be attributed to the introduction of the data denoising method and the capability of LSTM for processing and modelling multi-variate time series.

2) Result evaluation of the proposed SCADA data denoising algorithm based on WT: In order to more intuitively evaluate the impact of the proposed SCADA data denoising algorithm based on WT on the model training process, by considering whether to use this proposed SCADA data denoising algorithm, the loss changes of the proposed method during the training process based on four data sets (1 month, 4 month, 7 month, and 11 month) are plotted in Fig. 5, where “Without using the proposed data denoising algorithm” represents the proposed method without using the proposed data denoising algorithm, and “Using the proposed data denoising algorithm” represents the proposed method using the proposed data denoising algorithm. In these experiments, all other conditions are completely consistent. From Fig. 5, it can been clearly seen that the use of the proposed SCADA data denoising method based on WT can significantly accelerate the convergence of the proposed method. For example, from Fig. 5(a), the loss value of the proposed method using the proposed SCADA data denoising algorithm is significantly lower than that of the proposed method without using the proposed SCADA data denoising algorithm when the number of iterations reaches 10. When the number of iterations reaches 50, the gap is even greater. Similar changes can also be observed in Fig. 5(b)-(d). The proposed SCADA data denoising algorithm based on WT can effectively filter out noise information in SCADA data and reduce data complexity, which is beneficial for the model to learn more useful information and achieve better prediction performance.

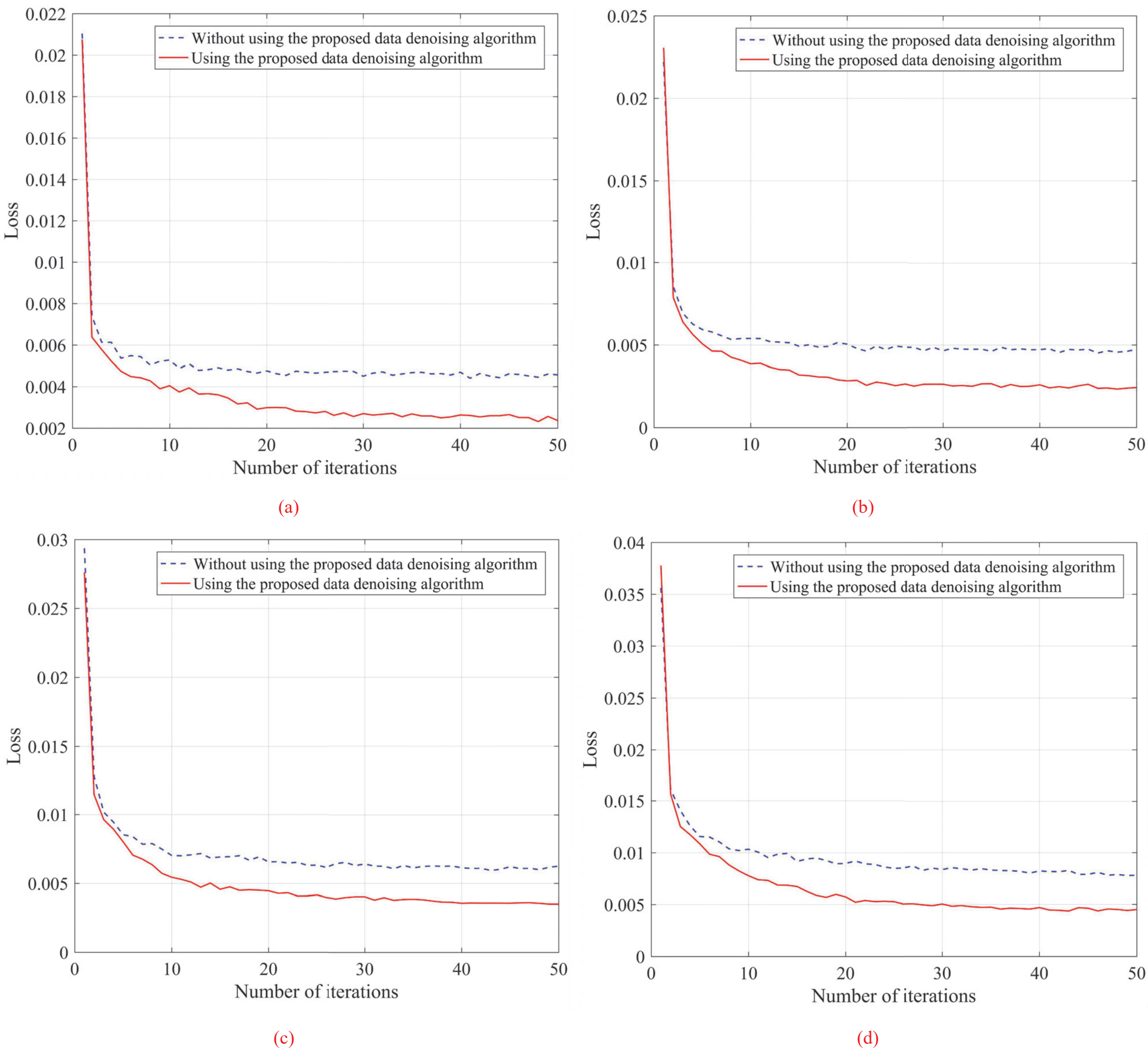


Fig. 5. Loss changes of the proposed method during the training process based on four monthly data sets by considering whether to use this proposed SCADA data denoising algorithm. (a) 1 month. (b) 4 month. (c) 7 month. (d) 11 month.

4.5. Case II: multi-step prediction

To further examine the performance of the proposed method, multi-step prediction experiments, with the prediction horizons of 15, 30, 45, and 60 minutes, are carried out. For comparison purposes, four methods, namely, CNN-LSTM, GRU-LSTM, GBR, and DRNets, are applied to solve the same prediction tasks on four data sets measured in the following four periods of 2020, that is, 01/01–31/03, 01/04–30/06, 01/07-30/09, and 01/10-31/12, respectively. The four data sets are named D1, D2, D3, and D4, respectively. In each data set, the data of the last seven days are used as for model test, and other data are used for model training. The denoising and the feature selection processes are the same as described for one-step ahead predictions. The NRMSE and NMAE values predicted by all methods are shown in Table 5, and the corresponding mean values are also calculated in this table.

From Table 5, it can be seen that the proposed method also outperforms other compared methods in most multi-step prediction tasks. Take data set D1 as an example. Compared with the DRNets method, the NRMSE values of the proposed method are around 38.04% (from 6.9341% to 4.2963%) smaller at the 15 minute prediction horizon, and are around 13.69%

Table 5. Multi-step prediction results of all the methods based on the data sets of four data sets (%)

Data sets	Prediction horizons	Methods									
		CNN-LSTM		GRU-LSTM		GBR		DRNets		Proposed method	
		NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
D1	15 minutes	7.6558	5.1466	7.1132	4.4753	7.4091	4.6192	6.9341	4.6070	**4.2963**	**2.7541**
	30 minutes	9.0078	6.1180	8.4335	5.3050	10.5078	6.2659	9.6233	6.2354	**7.2061**	**4.5565**
	45 minutes	9.9199	6.7652	9.3415	5.8247	11.8902	6.8728	10.6833	7.0363	**8.9248**	**5.7993**
	60 minutes	10.6265	7.3437	10.0818	**6.2720**	11.0262	6.7109	11.2350	7.5120	**9.6967**	6.3888
D2	15 minutes	7.3995	5.8197	7.1194	5.4006	7.0935	5.2487	7.6759	5.0813	**4.9473**	**3.6473**
	30 minutes	8.9420	6.9667	8.9557	6.8247	9.1200	6.9314	10.5986	6.7854	**7.6804**	**5.8197**
	45 minutes	9.8180	7.6808	9.9368	7.7081	9.9682	7.8149	11.7003	7.5549	**9.5425**	**7.3554**
	60 minutes	10.3579	8.2165	10.5723	8.3303	10.5733	8.4042	13.1282	8.4371	**10.2430**	**7.9831**
D3	15 minutes	5.1211	3.2515	5.0583	3.1569	5.2745	3.4939	5.2801	3.3812	**3.2907**	**2.0346**
	30 minutes	5.8343	3.7937	5.8374	3.8248	5.9455	4.2059	6.2578	4.1922	**5.0231**	**3.1716**
	45 minutes	6.0201	3.9336	6.1000	3.9353	6.0327	4.0771	6.9198	4.7147	**5.8939**	**3.8481**
	60 minutes	6.2307	4.2382	6.3205	4.2324	6.6267	4.8812	7.4841	5.3792	**6.0935**	**4.0403**
D4	15 minutes	7.6582	5.2120	7.9189	5.3868	7.9663	5.4943	8.0404	5.4054	**4.9148**	**3.4286**
	30 minutes	9.3795	6.3679	9.6167	6.5789	9.9670	6.9612	9.8555	6.7676	**8.2126**	**5.6947**
	45 minutes	10.4428	**7.1343**	10.5141	7.2594	10.9013	7.6230	11.3317	7.7278	**10.2419**	7.1900
	60 minutes	11.6638	8.0336	11.6161	**7.9764**	12.0494	8.5058	12.4723	8.4311	**11.3155**	8.0415
Mean		8.5049	6.0014	8.4085	5.7807	8.8970	6.1319	9.3263	6.2030	**7.3452**	**5.1096**

(from 11.2350% to 9.6967%) smaller at the 60 minute prediction horizon; the NMAE values are around 40.22% (from 4.6070% to 2.7541%) smaller at the 15 minute prediction horizon, and are around 14.95% (from 7.5120% to 6.3888%) smaller at the 60 minute predicted horizon. Compared with the CNN-LSTM method, the NRMSE values of the proposed method are around 43.88% (from 7.6558% to 4.2963%) smaller at the 15 minute prediction horizon, and are around 8.75% (from 10.6265% to 9.6967%) smaller at the 60 minute prediction horizon; the NMAE values are around 46.49% (from 5.1466% to 2.7541%) smaller at the 15 minute predicted horizon, and are around 13.00% (from 7.3437% to 6.3888%) smaller at the 60 minute predicted horizon.

In all the prediction tasks, the metric, NRMSE, has smaller values for the proposed method. The rate of change of the NRMSE metric is calculated for each task, and the NRMSE values of the proposed method are at least 30.26% (15 minutes), 12.44% (30 minutes), 1.92% (45 minutes), and 1.11% (60 minutes) smaller than that of other methods in different prediction horizons, respectively. On the other hand, the metric, NMAE, has similar characteristics for the proposed method. The NMAE values of the proposed method are at least 8.22% (15 minutes) and 10.57% (30 minutes) smaller than that of other methods in prediction horizons, 15 and 30 minutes, respectively. In prediction horizons, 45 and 60 minutes, the NMAE values of the proposed method are 0.10%-1.86% larger than that of some methods in three tasks. However, overall, the proposed method performs well.

The mean values in Table 5 show that the proposed method has smaller prediction error in two evaluation metrics. For example, the NRMSE values of the proposed method are at least 12.65% smaller than that of other methods, and correspondingly, the NMAE values of the proposed method are at least 11.61% smaller than that of other methods. As can be seen from the above, the proposed method can achieve higher prediction accuracy than ordinary models. This is mainly because the proposed method denoises SCADA feature data through the designed SCADA data denoising algorithm, resulting in a more pronounced trend of new feature data with less noise information, which is more conducive to model learning.

Furthermore, it can be seen from this table that as the prediction horizon increases, the prediction performance of all methods decreases. This is because the longer the prediction horizon, the more uncertainty is involved in the prediction. However, the proposed method can maintain higher prediction accuracy than other models at various prediction horizons, indicating that using the proposed deep learning models, joined by the designed SCADA data noising algorithm to is applicable for wind power prediction tasks.

5. Conclusion

With the large-scale integration of wind power into the power grid, the prediction of wind power has increasingly become an important issue. In this paper, an effective deep learning model based on LSTM networks and data denoising scheme is proposed for improving the short-term wind power forecasting. The denoising method can effectively filter out noise from the SCADA data. Based on the denoised data, the feature selection process is performed by utilizing the MIC algorithm. According to the results of the feature selection, an input vector is defined and used to train the prediction model. Both single-step and multi-step prediction experiments are conducted, and the proposed method is compared with seven existing methods, namely, SVM, BPNN, CNN-LSTM, GRU-LSTM, GBR, and DRNets. The results show that the proposed method can improve the wind power prediction ability based on the SCADA data. In our work, only up to 60-minute ahead predictions are considered. In fact, it may be necessary to predict the changes in wind power over the next few hours in practical applications. Therefore, exploring the application of the proposed method in many hours ahead wind power predictions will be of great significance in the future work. In addition, this work focuses on the application of SCADA data in power prediction, without paying too much attention to the problem of gradient vanishing or exploding in LSTM networks. In future work, we will pay more attention to this issue to achieve better performance.

Credit Author Statement

Zhao-Hua Liu: Conceptualization, Methodology, Writing - Review & Editing, Funding acquisition, Resources, Supervision. **Chang-Tong Wang:** Methodology, Software, Validation, Writing - Original Draft, Investigation. **Hua-Liang Wei:** Writing - Review & Editing. **Bing Zeng:** Validation, Funding acquisition, Project administration, Data Curation. **Ming Li:** Visualization, Project administration. **Xiao-Ping Song:** Validation, Funding acquisition, Formal analysis.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61972443, in part by the National Key Research and Development Project of China under Grant 2019YFE0105300, and in part by the Hunan Provincial Key Research and Development Project of China under Grant 2022WK2006.

References

Abedinia, O., Lotfi, M., Bagheri, M., Sobhani, B., Shafie-khah, M., & Catalão, J. P. S. (2020). Improved EMD-based complex prediction model for wind power forecasting. IEEE Transactions on Sustainable Energy, 11(4), 2790-2802. https://doi.org/10.1109/TSTE.2020.2976038.

Akhtar, I., Kirmani, S., Ahmad, M., & Ahmad, S. (2021). Average monthly wind power forecasting using fuzzy approach. IEEE Access, 9, 30426-30440. https://doi.org/10.1109/ACCESS.2021.3056562.

An, J., Yin, F., Wu, M., She, J., & Chen, X. (2021). Multisource wind speed fusion method for short-term wind power prediction. IEEE Transactions on Industrial Informatics, 17(9), 5927-5937. https://doi.org/10.1109/TII.2020.3006928.

Chen, J., Zhu, Q., Li, H., Zhu, L., Shi, D., Li, Y., Duan, X., & Liu, Y. (2020). Learning heterogeneous features jointly: a deep end-to-end framework for multi-step short-term wind power prediction. IEEE Transactions on Sustainable Energy, 11(3), 1761-1772. https://doi.org/110.1109/TSTE.2019.2940590.

Cheng, L., Zang, H., Xu, Y., Wei, Z., & Sun, G. (2021). Augmented convolutional network for wind power prediction: a new recurrent architecture design with spatial-temporal image inputs. IEEE Transactions on Industrial Informatics, 17(10), 6981-6993. https://doi.org/10.1109/TII.2021.3063530.

Chui, C. K. (1992). An introduction to wavelets. Boston, MA: Academic Press.

Du, P. (2019). Ensemble machine learning-based wind forecasting to combine NWP output with data from weather station. IEEE Transactions on Sustainable Energy, 10(4), 2133-2141. https://doi.org/10.1109/TSTE.2018.2880615.

Guan, S., Wang, Y., Liu, L., Gao, J., Xu, Z., & Kan, S. (2024). Ultra-short-term wind power prediction method based on FTI-VACA-XGB model. Expert Systems with Applications, 35, Article 121185. https://doi.org/10.1016/j.eswa.2023.121185.

Hossain, M. A., Chakrabortty, R. K., Elsawah, S., Gray, E. M., & Ryan, M. J. (2021). Predicting wind power generation using hybrid deep learning with optimization. IEEE Transactions on Applied Superconductivity, 31(8), 1-5. https://doi.org/10.1109/TASC.2021.3091116.

Hossain, M. A., Chakrabortty, R. K., Elsawah, S., & Ryan, M. J. (2020). Hybrid deep learning model for ultra-short-term wind power forecasting. 2020 IEEE International Conference on Applied Superconductivity and Electromagnetic Devices (ASEMD), Tianjin, China. https://doi.org/10.1109/ASEMD49065.2020.9276090.

Jiang,Y., Chen, X., Yu, K., & Liao, Y. (2017). Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm. Journal of Modern Power Systems and Clean Energy, 5(1), 126-133. https://doi.org/10.1007/s40565-015-0171-6.

Kisvari, A., Lin, Z., & Liu, X. (2021). Wind power forecasting – A data-driven method along with gated recurrent neural network. Renewable Energy, 163, 1895-1909. https://doi.org/10.1016/j.renene.2020.10.119.

Ko, M.-S., Lee, K., Kim, J.-K., Hong, C. W., Dong, Z. Y., & Hur, K. (2021). Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting. IEEE Transactions on Sustainable Energy, 12(2), 1321-1335. https://doi.org/10.1109/TSTE.2020.3043884.

Li, C., Lin, S., Xu, F., Liu, D., & Liu, J. (2018). Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in Northwest China. Journal of Cleaner Production, 205, 909-922. https://doi.org/10.1016/j.jclepro.2018.09.143.

Li, J., Zhang, S., & Yang, Z. (2022). A wind power forecasting method based on optimized decomposition prediction and error correction. Electric Power Systems Research, 208, Article 107886. https://doi.org/10.1016/j.epsr.2022.107886.

Lin, Z. & Liu, X. (2020). Wind power forecasting of an offshore wind turbine based on high frequency SCADA data and deep learning neural network. Energy, 201, Article117693. https://doi.org/10.1016/j.energy.2020.117693.

Lin, Z., Liu, X., & Collu, M. (2020). Wind power prediction based on high-frequency SCADA data along with isolation forest and deep learning neural networks. International Journal of Electrical Power & Energy Systems, 118, Article 105835. https://doi.org/10.1016/j.ijepes.2020.105835.

Liu, X., Yang, L., & Zhang, Z. (2021). Short-term multi-step ahead wind power predictions based on a novel deep convolutional recurrent network method. IEEE Transactions on Sustainable Energy, 12(3), 1820-1833. https://doi.org/10.1109/TSTE.2021.3067436.

Lu, B. -L., Liu, Z. -H., Wei, H. -L., Chen, L., Zhang, H., & Li, X. -H. (2021). A deep adversarial learning prognostics model for remaining useful life prediction of rolling bearing. IEEE Transactions on Artificial Intelligence, 2(4), 329-340. https://doi.org/10.1109/TAI.2021.3097311.

Meng, A., Chen,S., Ou, Z., Ding, W., Zhou, H., Fan, J., & Yin, H. (2022). A hybrid deep learning architecture for wind power prediction based on bi-attention mechanism and crisscross optimization. Energy, 238, Article 121795. https://doi.org/10.1016/j.energy.2021.121795.

Miao, C., Li, H., Wang, X., & Li, H. (2021). Ultra-short-term prediction of wind power based on sample similarity analysis. IEEE Access, 9, 72730-72742. https://doi.org/10.1109/ACCESS.2021.3080140.

Morrison, R., Liu, X., & Lin, Z. (2022). Anomaly detection in wind turbine SCADA data for power curve cleaning. Renewable Energy, 184, 473-486. https://doi.org/10.1016/j.renene.2021.11.118.

Peng, X., Li, Y., Dong, L., Cheng, K., Wang, H., Xu, Q., Wang, B., Liu, C., Che, J., Yang, F., & Li, W. (2021). Short-term wind power prediction based on wavelet feature arrangement and convolutional neural networks deep learning. IEEE Transactions on Industry Applications, 57(6), 6375-6384. https://doi.org/10.1109/TIA.2021.3106887.

Pujari, K. N., Miriyala, S. S., Mittal, P., & Mitra, K. (2023). Better wind forecasting using evolutionary neural architecture search driven green deep learning. Expert Systems with Applications, 214, Article 119063. https://doi.org/10.1016/j.eswa.2022.119063.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., Mcvean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., & Sabeti, P. C. (2011). Detecting novel associations in large data sets. Science, 334(6062), 1518-1524.

Safari, N., Chung, C. Y., & Price, G. C. D. (2018). Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis. IEEE Transactions on Power Systems, 33(1), 590-601. https://doi.org/10.1109/TPWRS.2017.2694705.

Sobolewski, R. A., Tchakorom, M., & Couturier, R. (2023). Gradient boosting-based approach for short- and medium-term wind turbine output power prediction. Renewable Energy, 203, 142-160. https://doi.org/10.1016/j.renene.2022.12.040.

Tan, L., Han, J., & Zhang, H. (2020). Ultra-short-term wind power prediction by salp swarm algorithm-based optimizing extreme learning machine. IEEE Access, 8, 44470-44484. https://doi.org/10.1109/ACCESS.2020.2978098.

Udo, W. & Muhammad, Y. (2021). Data-driven predictive maintenance of wind turbine based on SCADA data. IEEE Access, 9, 162370-162388.

Wang, J., Zhu, H., Cheng, F., Zhou, C., Zhang, Y., Xu, H., Liu, M. (2023). A novel wind power prediction model improved with feature enhancement and autoregressive error compensation. Journal of Cleaner Production, 420, Article 138386. https://doi.org/10.1016/j.jclepro.2023.138386.

Wang, J., Zhu, H., Zhang, Y., Cheng, F., Zhou, C. (2023). A novel prediction model for wind power based on improved long short-term memory neural network. Energy, 265, Article 126283. https://doi.org/10.1016/j.energy.2022.126283.

Wang, L., Tao, R., Hu, H., & Zeng, Y.-R. (2021). Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder. Renewable Energy, 164, 642-655. https://doi.org/10.1016/j.renene.2020.09.108.

Wang, R., Li, C., Fu, W., & Tang, G. (2020). Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction. IEEE Transactions on Neural Networks and Learning Systems, 31(10), 3814-3827. https://doi.org/10.1109/TNNLS.2019.2946414.

Xiong, B., Lou, L., Meng, X., Wang, X., Ma, H., & Wang, Z. (2022). Short-term wind power forecasting based on attention mechanism and deep learning. Electric Power Systems Research, 206, Article 107776. https://doi.org/10.1016/j.epsr.2022.107776.

Xue, H., Jia, Y., Wen, P., & Farkoush, S. G. (2020). Using of improved models of Gaussian processes in order to regional wind power forecasting. Journal of Cleaner Production, 262, Article 121391. https://doi.org/10.1016/j.jclepro.2020.121391.

Yan, H. & Wu, Z. (2020). A hybrid short-term wind power prediction model combining data processing, multiple parameters optimization and multi-intelligent models apportion strategy. IEEE Access, 8, 227126-227140. https://doi.org/10.1109/ACCESS.2020.3046001.

Ye, L., Zhao, Y., Zeng, C., & Zhang, C. (2017). Short-term wind power prediction based on spatial model. Renewable Energy, 101, 1067-1074. https://doi.org/10.1016/j.renene.2016.09.069.

Ye, L., Dai, B., Pei, M., Lu, P., Zhao, J., Chen, M., & Wang, B. (2022). Combined approach for short-term wind power forecasting based on wave division and seq2seq model using deep learning. IEEE Transactions on Industry Applications, 58(2), 2586-2596. https://doi.org/10.1109/TIA.2022.3146224.

Yin, H., Ou, Z., Fu, J., Cai, Y., Chen, S., & Meng, A. (2021). A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture. Energy, 234, Article 121271. https://doi.org/10.1016/j.energy.2021.121271.

Yu, L., Ma, Y., Ma, Y., & Zhang, G. (2022). A complexity-trait-driven rolling decomposition-reconstruction-ensemble model for short-term wind power forecasting. Sustainable Energy Technologies and Assessments, 49, Article 101794. https://doi.org/10.1016/j.seta.2021.101794.

Zhang, H., Yue, D., Dou, C., Li, K., & Hancke, G. P. (2022). Two-Step Wind Power Prediction Approach With Improved Complementary Ensemble Empirical Mode Decomposition and Reinforcement Learning. IEEE Systems Journal, 16(2), 2545-2555. https://doi.org/10.1109/JSYST.2021.3065566.

Zhang, H., Zhao, L., & Du, Z. (2021). Wind power prediction based on CNN-LSTM. 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), Taiyuan, China, 3097-3102. https://doi.org/10.1109/EI252483.2021.9713238.

Zhang, G., Liu, H., Zhang, J., Yan, Y., Zang, L., Wu, C., Hua, X., & Wang, Y. (2019). Wind power prediction based on variational mode decomposition multi-frequency combinations. Journal of Modern Power Systems and Clean Energy, 7(2), 281–288. https://doi.org/10.1007/s40565-018-0471-8.

Zhang, G., Li, Y., & Zhao, Y. (2023). A novel fault diagnosis method for wind turbine based on adaptive multivariate time-series convolutional network using SCADA data. Advanced Engineering Informatics, 57, Article 102031. https://doi.org/10.1016/j.aei.2023.102031.

Zhou, Y., Wang, J., Lu, H., & Zhao, W. (2022). Short-term wind power prediction optimized by multi-objective dragonfly algorithm based on variational mode decomposition. Chaos, Solitons and Fractals, 157, Article 111982. https://doi.org/10.1016/j.chaos.2022.111982.

Zjavka, L. & Mišák, S. (2018). Direct wind power forecasting using a polynomial decomposition of the general differential equation. IEEE Transactions on Sustainable Energy, 9(4), 1529-1539. https://doi.org/10.1109/TSTE.2018.2794515.