

Article

# LSTM Short-Term Wind Power Prediction Method Based on Data Preprocessing and Variational Modal Decomposition for Soft Sensors

Peng Lei <sup>1,2</sup>, Fanglan Ma <sup>2,3</sup>, Changsheng Zhu <sup>2,\*</sup> and Tianyu Li <sup>2</sup><sup>1</sup> Network & Information Center, Lanzhou University of Technology, Lanzhou 730050, China<sup>2</sup> School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China<sup>3</sup> Institute of Sensing Technology, Gansu Academy of Sciences, Lanzhou 730000, China

\* Correspondence: zhucs@lut.edu.cn

**Abstract:** Soft sensors have been extensively utilized to approximate real-time power prediction in wind power generation, which is challenging to measure instantaneously. The short-term forecast of wind power aims at providing a reference for the dispatch of the intraday power grid. This study proposes a soft sensor model based on the Long Short-Term Memory (LSTM) network by combining data preprocessing with Variational Modal Decomposition (VMD) to improve wind power prediction accuracy. It does so by adopting the isolation forest algorithm for anomaly detection of the original wind power series and processing the missing data by multiple imputation. Based on the process data samples, VMD technology is used to achieve power data decomposition and noise reduction. The LSTM network is introduced to predict each modal component separately, and further sum reconstructs the prediction results of each component to complete the wind power prediction. From the experimental results, it can be seen that the LSTM network which uses an Adam optimizing algorithm has better convergence accuracy. The VMD method exhibited superior decomposition outcomes due to its inherent Wiener filter capabilities, which effectively mitigate noise and forestall modal aliasing. The Mean Absolute Percentage Error (MAPE) was reduced by 9.3508%, which indicates that the LSTM network combined with the VMD method has better prediction accuracy.

**Keywords:** short-term wind power prediction; LSTM; VMD; data preprocessing; isolation forest; soft sensor



**Citation:** Lei, P.; Ma, F.; Zhu, C.; Li, T. LSTM Short-Term Wind Power Prediction Method Based on Data Preprocessing and Variational Modal Decomposition for Soft Sensors. *Sensors* **2024**, *24*, 2521. <https://doi.org/10.3390/s24082521>

Academic Editor: Chun Sing Lai

Received: 23 January 2024

Revised: 1 April 2024

Accepted: 5 April 2024

Published: 15 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The power output of a wind power generation system has great randomness, volatility and uncertainty, so its large-scale integration into the power grid brings great challenges to the safe and stable operation of the power system [1,2]. Soft sensor technology establishes a mathematical model between the auxiliary and target variables to predict the target variables, which has the characteristics of low cost and high accuracy [3]. However, strongly nonlinear, dynamic time-varying and multi-rate data characteristics are caused by poor soft sensor performance. Short-term forecasting of wind power using deep learning can provide a reference for the dispatch of the intraday power grid and is one of the key technologies in solving the above-mentioned problems [4].

At present, the related research on short-term wind power forecasting mainly focuses on establishing more accurate forecasting models, data preprocessing and signal decomposition. Many existing prediction models treat input samples as different classes and classify them, which may lead to long-term dependencies between samples that cannot be fully utilized [5]. In addition, human factors, or the failure of wind turbine equipment, may lead to abnormal or missing parts of the power data, which affects the accuracy of wind power prediction. Therefore, in order to ensure the reliability of the data, detecting abnormal values in power data is essential. The use of an unsupervised isolated forest algorithm

based on an ensemble learning strategy to detect abnormal data can effectively improve data processing efficiency and improve the accuracy of abnormality detection [6]. VMD is a new type of signal decomposition method which can decompose the signal into several modal components according to the potential characteristics of wind power, and VMD has less parameter settings, good robustness, high computational efficiency and a rich theoretical basis [7–9]. Using the VMD method can make the input data of the prediction model reflect the characteristics of the wind power signal more clearly. A new hybrid wind speed forecasting model using Variational Modal Decomposition (VMD), the partial autocorrelation function (PACF), and a weighted regularized extreme learning machine (WRELM) is proposed to improve the accuracy of wind speed forecasting [10].

In recent years, deep learning has dramatically improved the learning ability of neural networks via residual connection and shared weights, which are widely used in dynamic soft sensor modeling. A Recurrent Neural Network (RNN) [11] has recursive links in its network structure, and the relationship between samples can be considered in the learning process, so it is especially suitable for processing time series signals. But if there are long-term dependencies between samples, RNNs will suffer from vanishing gradients and exploding gradients. A Long Short-Term Memory (LSTM) network is an improved method for addressing this problem. In recent years, LSTM has been used more and more in wind power [12–15]. Yu et al. [16] combined wavelet transforms to establish a new hybrid model based on three recurrent neural networks. The results demonstrate that the three new hybrid models produce more accurate prediction results. Nevertheless, it is poor adaptability that the decomposition effect of the wavelet transform depends on the choice of threshold and the basis function. Curreri et al. [17] compared the recurrent neural networks and long short-term memory architectures in regard to their transferability. The obtained results demonstrate the suitability of the proposed transfer learning methods in the design of nonlinear dynamical models for industrial systems. Zhang et al. [18] proposed a strategy of building a soft sensor model based on local semi-supervised ensemble learning of least squares support vector regression, which is used to deal with nonlinear, dynamic time-varying and multi-rate data regression problems in wind power generation processes. Han et al. [19] used the VMD technique to decompose the original wind power signal and used the decomposed components as the input of the improved LSTM prediction model to predict the wind power. However, data preprocessing operations, such as outlier detection of the collected raw data, are ignored. Aiming at the above problems, this paper combines the forecasting model and wind power data processing to improve forecasting accuracy and proposes a LSTM short-term wind power forecasting model based on data preprocessing and VMD [20] for the soft sensor. We use the isolation forest and multiple imputation methods to deal with outliers and missing values of wind power data [21].

An LSTM network consists of an input layer, an output layer, and several recursive hidden layers between them. The recursive hidden layers are composed of several memory modules. Each module contains one or more self-connected memory cells and three gates that control the flow of information: memory gates, forgetting gates and output gates [19].

To minimize the sum of frequency bandwidths to complete noise reduction, VMD technology is adopted to decompose historical power data into several modal components. The LSTM method is imported to establish a prediction model for each modal component. The Dropout parameter regularization method is used to establish the model to prevent over-fitting. The Adam algorithm is adopted to optimize the effective training of the network parameters of the LSTM model. Ultimately, sum and reconstruct are used to predict the results of the components. Compared to the results of BP (Back Propagation), SVM (Support Vector Machine), LSTM and Complete Ensemble Empirical Mode Decomposition-LSTM (CEEMDAN-LSTM) models, the experimental results indicate that the VMD-LSTM model has higher prediction accuracy.

## 2. Method

### 2.1. VMD

VMD [22] decomposes the signal  $f(t)$  into discrete modal components  $u_k (k = 1, 2, 3, \dots, K)$ ; for each modal component  $u_k$ , the Hilbert transform is used to calculate the relevant analytic signal and obtain the unilateral spectrum, where  $\delta(t)$  is the pulse function.

$$U(t) = \left( \delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \quad (1)$$

The analytical signal obtained from Equation (1) above is mixed with the estimated center frequency  $e^{-j\omega_k t}$ , and the spectrum corresponding to each mode is transformed to the corresponding baseband.

$$U(\hat{t}) = U(t) \times e^{-j\omega_k t} \quad (2)$$

Calculate the square norm of the gradient of the demodulated signal, obtain the bandwidth of each mode, and then construct the constrained variational problem, where  $u_k = \{u_1, \dots, u_k\}$ ,  $\omega_k = \{\omega_1, \dots, \omega_k\}$ .

$$L2 = \begin{cases} \min_{(u_k), (\omega_k)} \left\{ \left\| \partial_t U(\hat{t}) \right\|^2 \right\} \\ \text{s.t. } \sum_{k=1}^K u_k = f \end{cases} \quad (3)$$

The augmented Lagrangian function is introduced to find the optimal solution for the constrained variational problem, where  $\alpha$  is the quadratic penalty factor, and  $\lambda(t)$  is the Lagrangian multiplication operator.

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t U(\hat{t}) \right\|^2 + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \quad (4)$$

The alternating direction multiplier method (ADMM) is used to update  $\hat{u}_k^{n+1}$  and  $\omega_k^{n+1}$  to find the optimal solution for Equation (3).

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (5)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (6)$$

where  $\hat{u}_k^{n+1}(\omega)$ ,  $\hat{f}(\omega)$ ,  $\hat{u}_i(\omega)$  and  $\hat{\lambda}(\omega)$  are Fourier transforms of  $u_k^{n+1}(t)$ ,  $f(t)$  and  $\lambda(t)$ , respectively.

### 2.2. LSTM

The LSTM memory unit is used to build the LSTM [23] network prediction model, and the Dropout regularization method [24] is used between the hidden layer and the Dense Layer to prevent over-fitting and improve the model generalization ability. The LSTM model network structure is shown in Figure 1. Among them, LSTM represents the LSTM memory unit, and Dense represents the full connection layer. When Dropout is applied to the LSTM layer, the input of the Dense layer is the output of the LSTM layer. This process is represented by dotted arrows and dotted circles to illustrate the principle of the Dropout method. The input in the  $t - 1$  time model is represented as the variable  $x_{t-1}$ . After the Dropout method, the LSTM layer outputs the results to the Dense layer, and the corresponding representation is  $h_t$ . If  $x_t$  and  $h_t$  are the inputs at the next time  $t + 1$ , the output can be represented as  $h_{t+1}$  after the Dropout method, where  $\alpha$  is the quadratic penalty factor, and  $\lambda(t)$  is the Lagrangian multiplication operator.

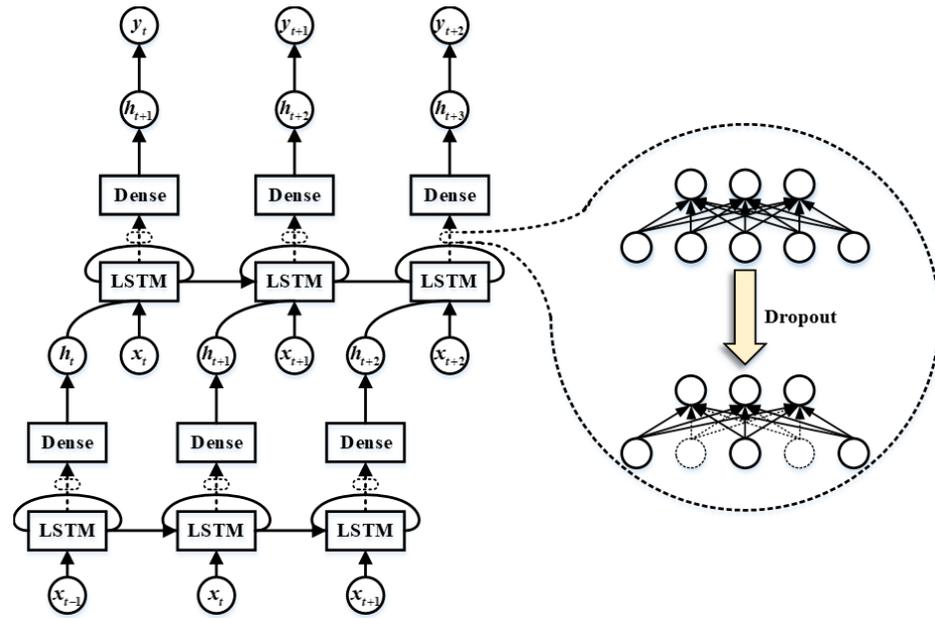


Figure 1. The structure of LSTM.

In addition, the Adam algorithm is used to optimize LSTM network parameters and train the network model. The updated rules of the Adam algorithm optimization parameters are as follows:

Calculate the gradient when  $t$ , and initialize  $t_0 = 0$  where  $J(\theta)$  is the random objective function of  $\theta$ .

$$g_t = \nabla_{\theta} J(\theta_{t-1}) \quad (7)$$

Calculate the biased first moment estimate  $m_t$  of the gradient and initialize  $m_0 = 0$  where parameter  $\beta_1 = 0.9$ .

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (8)$$

Calculate the biased first moment estimate  $v_t$  of the gradient and initialize  $v_0 = 0$  where parameter  $\beta_2 = 0.999$ .

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (9)$$

Correct the deviation of the first order moment estimation and express the result as  $\widehat{m}_t$ .

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (10)$$

Correct the deviation of the first order moment estimation and express the result as  $\widehat{v}_t$ .

$$\widehat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (11)$$

Calculate parameter  $\theta$  at time  $t$ .

$$\theta_t = \theta_{t-1} - \frac{\alpha \widehat{m}_t}{\sqrt{\widehat{v}_t} + \epsilon} \quad (12)$$

### 3. Results and Analysis

In this section, we first describe the real-world dataset used in the experiments. Then, we discuss the experiments conducted on the dataset.

### 3.1. Experimental Data and Evaluation Indicators

To evaluate our method, we performed anomaly detection and missing value interpolation operations on 17,280 sets of data with a sampling interval of 1 min from 20 May to 31 May 2015 at a wind farm in Shanxi Province, China. We used the 12 day data as the experiment's dataset. Each datum in the dataset is a two-dimensional datum which includes the wind speed and the wind farm. Then, taking every 15 sets of data as a cycle, that is, a cycle of 15 min, the average value of the data within 15 min is recorded, and the 1152 sets of data recorded are used as wind power data research samples and used for experimental analysis. Following chronological order, we take the first 1056 sets of data as the training sets and forecast the wind power for 15 min in the future. After each step of forecasting is achieved, the predicted value is used instead of the actual value to perform the iterative calculation of the next forecast. The forecast sequence is 1058–1151, a total of 94 wind power data for each data point. The experimental data is processed by the min-max normalization (MMN) method for dimensionless data processing, and the data values are mapped to the [0, 1] interval. After the prediction is completed, the predicted value is converted to the original interval by inverse normalization.

In the experiment, we chose Root Mean Squared Error (*RMSE*), Mean Absolute Error (*MAE*), Mean Absolute Percentage Error (*MAPE*), and Relative Error (*RE*) as an evaluation index for the prediction results.  $y_i$  represents the actual value of wind power,  $\hat{y}_i$  represents the predicted value of wind power,  $N$  represents the wind power sequence data amount, and  $\delta_i$  represents the relative error.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

$$MAPE = \left( \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \quad (15)$$

$$\delta_i = \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (16)$$

### 3.2. Data Preprocessing and Result Analysis

In this work, the isolated forest algorithm is used to detect abnormal wind power data. After detection by the algorithm, the abnormal data are marked as one, and the non-abnormal data are marked as zero. All abnormal data are set to zero according to the index number of identifier one, and the complete data set is further obtained by the method of multiple imputation.

The columnar scatter diagram before and after processing the raw wind power data is shown in Figure 2. The horizontal axis represents wind speed (m/s); part A in the figure represents the collected 17,280 groups of original wind power data, and part B represents the wind power data after data preprocessing. The vertical axis is wind power (KW). It can be seen that the data distribution of the unprocessed part A is highly random and chaotic, and the data points located in the power value range above 1500 KW deviate from most of the data points. After processing, the data distribution of part B is concentrated and the data points at both ends of the cylinder center are compact and orderly, which effectively enhances the reliability and integrity of the data. The processed data are decomposed into  $K$  modal components using VMD signal decomposition technology, and the value of  $K$  is determined by observing the center frequency distribution corresponding to each component under different modal numbers. In Table 1, when the number of modes is five, the corresponding center frequencies of *IMF2*, *IMF3* and *IMF4* are 78.47 Hz, 123.54 Hz and

396.97 Hz, respectively, and the center frequencies are close to each other, which indicates that modal aliasing may occur, resulting in over-resolution of the wind power signal.

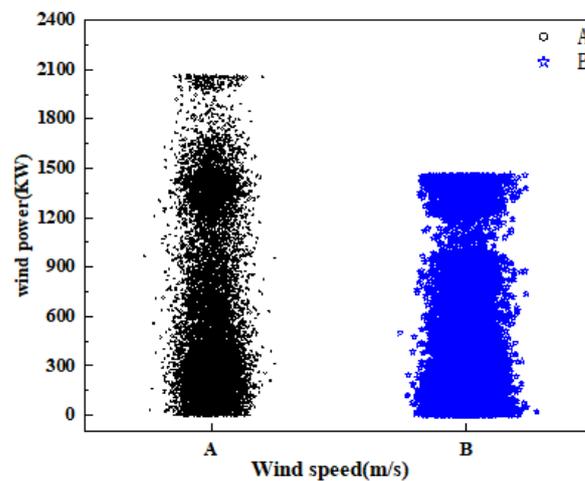


Figure 2. Comparison of histograms during data processing.

Table 1. Center frequency corresponding to different K.

Modal Number	Center Frequency/Hz					
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
2	2.81	735.23				
3	1.43	382.21	738.86			
4	1.22	114.96	390.34	740.88		
5	1.08	78.47	123.54	396.97	757.22	
6	0.97	21.63	52.69	135.72	397.42	761.53

To determine the value of  $K$ , we analyzed the change curve of the center frequency of each component, which is shown in Figure 3. It can be seen from Figure 3 that when  $K = 2$ ,  $K = 3$  and  $K = 4$ , the curves have no obvious curvature, and the slope changes little. When  $K = 5$ , the curve shows an obvious downward bending phenomenon, and the change of curve bending indicates that when  $K$  is the critical value of the modal decomposition number, the critical value is considered to be an appropriate modal decomposition number.

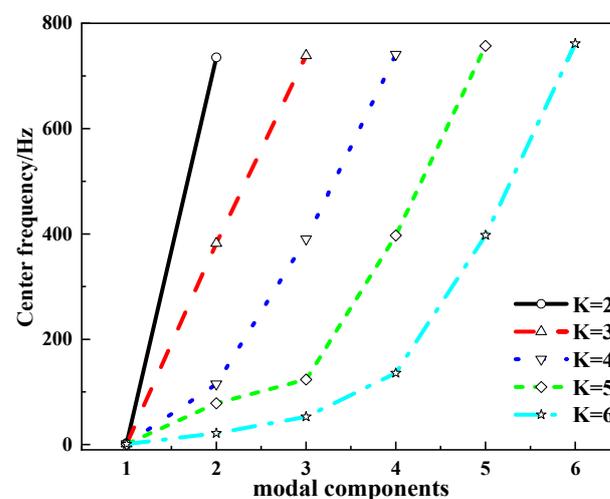


Figure 3. Change curve of center frequency of different modal components.

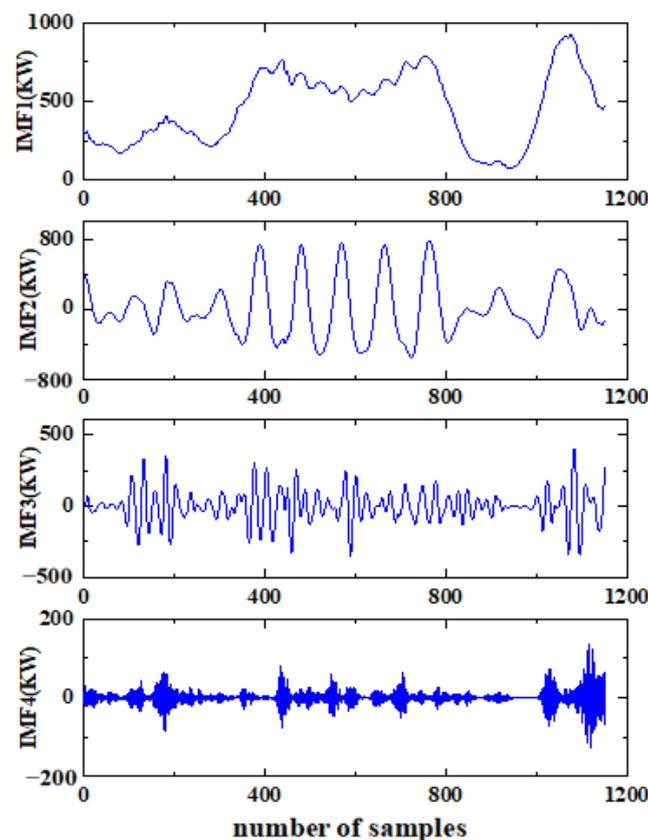
In order to further determine the value of  $K$ , the Pearson correlation coefficient of adjacent modal components is calculated, and Table 2 analyzes the correlation between adjacent modal components. Among them,  $C_{12}$  represents the Pearson correlation coefficient between  $IMF1$  and  $IMF2$ ,  $C_{23}$  represents the correlation coefficient between  $IMF2$  and  $IMF3$ , and the same is true for  $C_{34}$ ,  $C_{45}$  and  $C_{56}$ . When  $K < 5$ , the values of  $C_{12}$ ,  $C_{23}$  and  $C_{34}$  are all less than 0.1, indicating that the low- and high-frequency components obtained by signal decomposition have obvious  $C_{12}$  characteristics. When  $K = 5$ ,  $C_{12} = 0.3614$ ,  $C_{23} = 0.2906$  and the correlation coefficient is relatively large, which indicates that the low-frequency modal components obtained by decomposition are highly correlated, the signal decomposition is not sufficient, and modal aliasing is prone to occur.

**Table 2.** Pearson correlation coefficient of adjacent modes.

Modal Number	$C_{12}$	$C_{23}$	$C_{34}$	$C_{45}$	$C_{56}$
2	0.0915				
3	0.0618	0.0900			
4	0.0519	0.0901	0.0963		
5	0.3614	0.2906	0.0257	0.0810	
6	0.3501	0.2860	0.1284	0.0284	0.0601

In Table 2, the Pearson correlation coefficient of adjacent modes based on the above analysis determines  $K = 4$ , and sets parameters  $\alpha = 2000$  and  $\varepsilon = 0.000001$ , where  $\alpha$  is the quadratic penalty factor, and  $\varepsilon$  is the convergence accuracy.

The results of decomposing the wind power signal sequence by VMD are shown in Figure 4. Among them,  $IMF1$  to  $IMF4$  are the decomposed wind power sequence components from low frequency to high frequency.



**Figure 4.** Results of VMD decomposition.

The RMSProp (Root Mean Square prop) algorithm, Adagrad (Adaptive gradient) algorithm, SGD Nesterov (Stochastic Gradient Descent Nesterov) algorithm and Adam (Adaptive Moment Estimation) algorithm are combined with the Dropout regularization method to train the LSTM network model. The input data is the preprocessed wind power sequence without VMD decomposition; the parameter  $p = 0.3$  in the Dropout method is set, and the Mean Square Error (*MSE*) function is constructed as the loss function. The different curves in Figure 5 represent the change of the loss function in each training model, respectively. It can be seen from the figure that Adam is significantly faster than the SGD Nesterov algorithm in terms of convergence speed. Adam, RMSProp and Adagrad have basically the same convergence speed, but in the 50th training cycle, Adam has the lowest training loss value, which is only 0.0078, indicating that the algorithm can achieve better convergence accuracy.

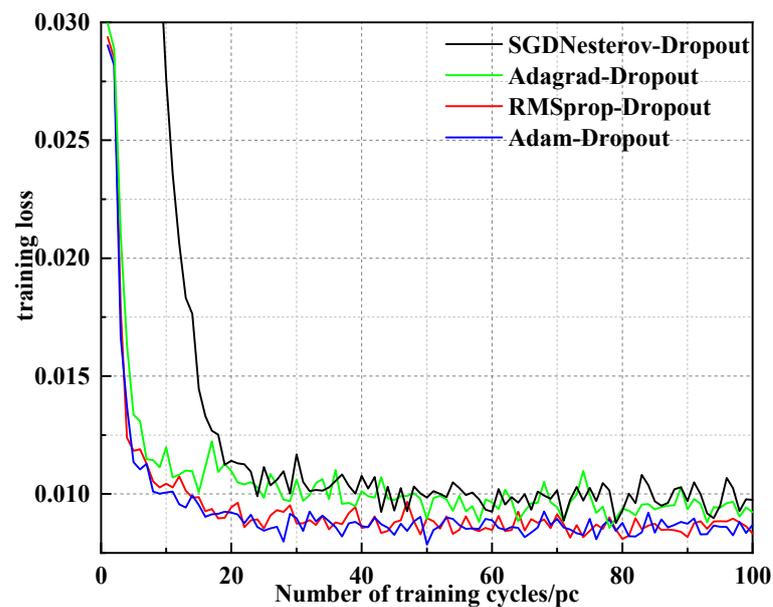


Figure 5. Training algorithm contrast curve.

The wind power sequence is decomposed by VMD, and then the LSTM network prediction model is established, respectively, to predict each IMF component and obtain the prediction result, as shown in Figure 5. In the LSTM network model, the input layer is one layer, the hidden layer is one layer, the hidden layer contains 16 LSTM neurons, the fully connected layer is one layer, the number of output layers is one, the loss function is *MSE*, and the optimization algorithm is Adam. Add the Dropout regularization method and set probability between the hidden layer and fully connected layer. The prediction results of each component are summed and reconstructed, and finally the prediction results of the VMD-LSTM [25–28] model are obtained. The LSTM model is used to predict the wind power sequence without VMD, and the prediction result of the LSTM model is obtained.

The comparison curve between the predicted value of the VMD-LSTM model and LSTM model and the actual wind power value is shown in Figure 6. It can be seen from Figure 7 that for the LSTM prediction model, the prediction results roughly follow the change trend of the actual wind power sequence, but when the actual power value suddenly changes, the prediction effect is poor. However, the error between the predicted value of the VMD-LSTM model and the actual value is small, and the trend change is completely consistent. Therefore, using VMD to decompose the original power signal can effectively extract data features and deal with fluctuations in wind power information.

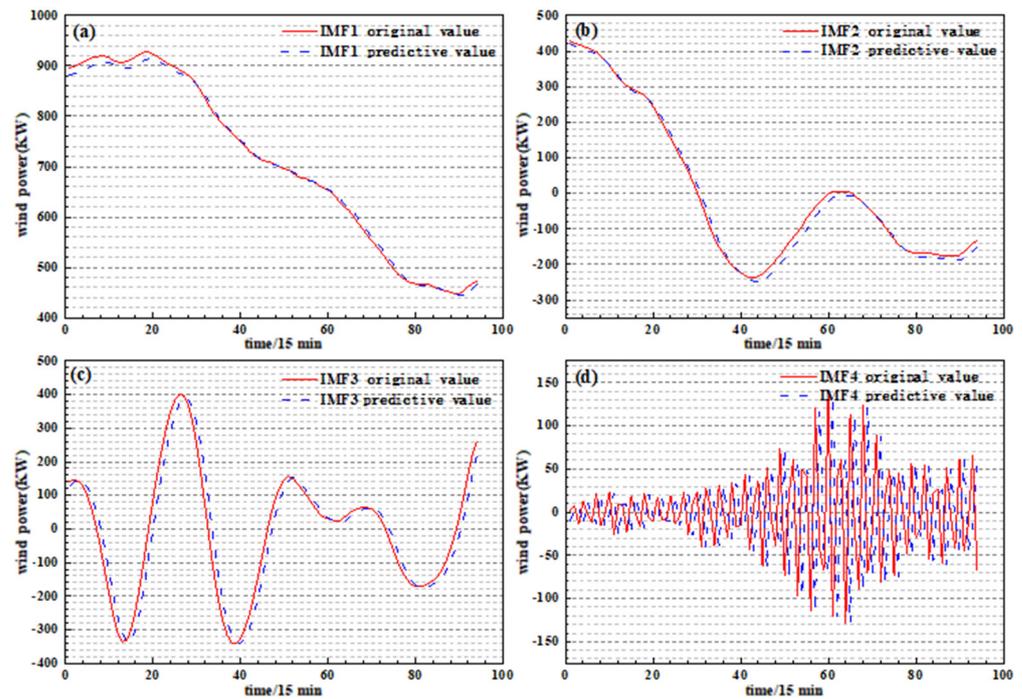


Figure 6. Prediction results of each IMF component: (a) *IMF1*; (b) *IMF2*; (c) *IMF3*; (d) *IMF4*.

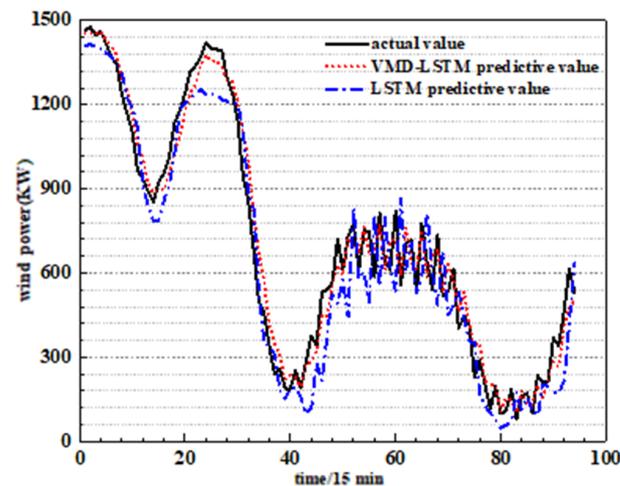
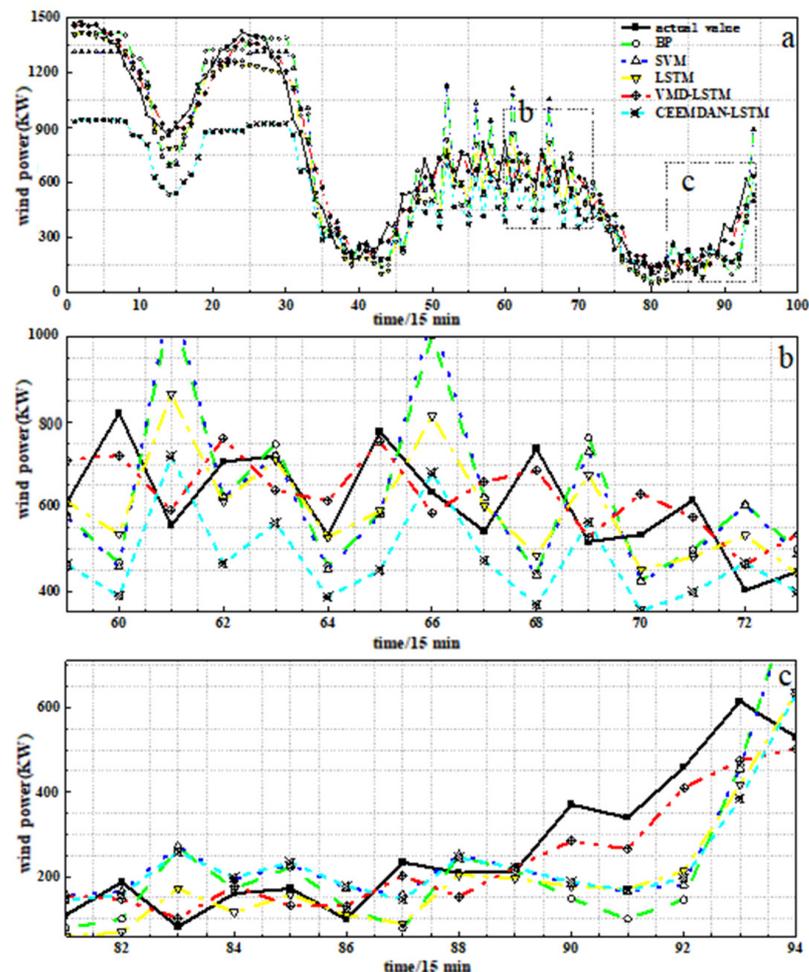


Figure 7. Prediction results of VMD-LSTM.

In this study, we use CEEMDAN-LSTM, BPNN and SVM to compare and analyze the performance of the VMD-LSTM model. In this experiment, it is assumed that there are 500 groups of white noise signals with a standard deviation of 0.2 in the CEEMDAN signal decomposition, and the maximum number of iterations is 5000. In the BPNN, the number of neurons in the input layer is one, the number of neurons in the hidden layer is 15, the number of neurons in the output layer is one, the learning rate is 0.01, and the maximum number of trainings is 100. Using grid method cross-validation in SVM, the parameters are 12.32 and 22.05.

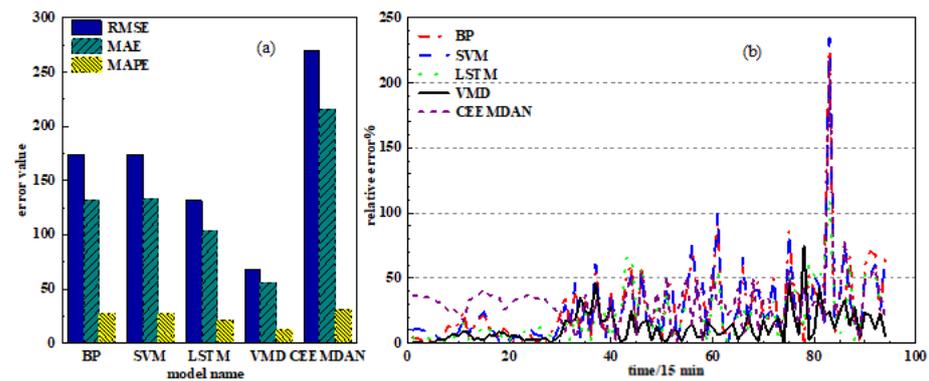
Figure 8a is a dotted line graph of the prediction results of each model; the actual value represents the actual wind power, and Figure 8b,c are partial enlarged comparisons of each model. It can be seen from Figure 8 that the wind power prediction curves of BPNN, SVM and LSTM are consistent with the trend of the actual wind power series, but the prediction results of the LSTM model are closer to the actual value. This proves that the LSTM model can effectively capture the long-term dependencies between data samples,

and the LSTM model is suitable for processing time series signals. Meanwhile, we can see that the CEEMDAN-LSTM model has a poor prediction effect and large prediction error. The VMD-LSTM model can not only accurately describe the dynamic changes of the original wind power sequence, but also the predicted value is closest to the actual wind power value. This shows that compared with CEEMDAN, the VMD method with a Wiener filter can effectively remove signal noise, can distinguish between effective signal information and noise signal significantly and has strong robustness. By decomposing the original wind power series through VMD, more accurate prediction results can be obtained, thereby improving the prediction performance of the model.



**Figure 8.** Wind power prediction curve: (a) comparison of prediction results among different models; (b) partial amplification of each mode; (c) partial amplification of each mode.

In order to further reflect the prediction error at each moment, as shown in Figure 9, we calculate the relative error of each model and draw its relative error curve to visualize the error evaluation index of each model. Figure 9a is a columnar schematic diagram of the three error evaluation indicators of the model: *RMSE*, *MAE* and *MAPE*. It can be seen that the *RMSE*, *MAE* and *MAPE* values of the CEEMDAN-LSTM model are the largest, and the *RMSE*, *MAE* and *MAPE* values of the VMD-LSTM model are significantly lower than the error values of the other models. From the relative error curves of each model in Figure 9b, it can be seen that except for the relative error of the 78th prediction point, which is 75%, the relative errors of the remaining prediction points of the VMD-LSTM model are all below 50%. Especially, the prediction accuracy at the abrupt point of wind speed is significantly improved compared with other models, which indicates that the VMD-LSTM method can improve the accuracy of wind power prediction.



**Figure 9.** Comparison of prediction errors between models: (a) model error histogram; (b) model error curve.

Table 3 lists the specific values of the error evaluation indicators in each prediction model. It can be seen from Table 3 that the *RMSE*, *MAE* and *MAPE* of the VMD-LSTM model are 67.6993, 55.7662 and 12.0676, respectively, which has the smallest prediction error compared with other models. Compared with the single prediction models BPNN and SVM, the *RMSE* of VMD-LSTM is reduced by 106.1781 and 106.3299, respectively. After using the decomposition algorithm, the *MAPE* value of CEEMDAN-LSTM is increased by 10.04685% compared with LSTM, and the *MAPE* value of VMD-LSTM is decreased by 9.3508% compared with LSTM. Meanwhile, the *RMSE* and *MAE* values of VMD-LSTM are reduced by 63.9672 and 48.1798, respectively, compared with LSTM. It can be verified that the VMD-LSTM model has good prediction performance and can effectively improve the prediction accuracy of wind power, which is suitable for short-term prediction of actual wind power.

**Table 3.** Comparison of the prediction indexes of different prediction methods.

	BP	SVM	LSTM	VMD-LSTM	CEEMDAN-LSTM
<i>RMSE</i> (KW)	173.8774	174.0292	131.6665	67.6993	270.0046
<i>MAE</i> (KW)	131.4138	133.1894	103.9460	55.7662	215.4398
<i>MAPE</i> (%)	26.9278	26.8527	21.4184	12.0676	31.4652

#### 4. Conclusions

In this study, we proposed a LSTM short-term wind power prediction model based on isolated forest outlier detection and VMD for the soft sensor. In the model, the isolated forest algorithm is used to detect the outliers of the wind power series, and the missing values of the data are processed by the multiple imputation method. Denoising decomposition of VMD is performed on the research samples, and the number *K* of modal components is determined according to the center frequency and the correlation coefficient. The LSTM contains a memory unit that can store and update information for a long time, making the network more robust and accurate when processing long time sequences. Additionally, the VMD decomposes the wind power data into *K* model components, which are used as the inputs of the forecast model. And the LSTM short-term wind power prediction model is established which uses Dropout to prevent over-fitting, and uses the Adam algorithm to optimize the model. The experiment results verify that the model can improve the accuracy of short-term power prediction. In addition, through comparative experiments, we can draw the following conclusions:

1. Using the isolated forest algorithm to detect anomalies in the original wind power sequence and to perform multiple imputation processing on missing data.
2. In terms of data processing, the experimental data is processed using the minimum-maximum normalization (MMN) method for dimensionless data, and the data values are mapped to the  $[0, 1]$  interval, which improves the effectiveness of data processing.

3. Compared with the RMSProp algorithm, Adagrad algorithm and SGD Nesterov algorithm, using the Adam algorithm to optimize LSTM network parameters has better convergence accuracy.
4. The VMD method has better decomposition results than the CEEMDAN method because its own Wiener filter can effectively complete the noise reduction and prevent modal aliasing.
5. Compared with traditional BPNN and SVM, LSTM is suitable for short-term wind power prediction and has better prediction accuracy.

**Author Contributions:** Data curation, P.L. and F.M.; supervision, C.Z.; methodology, P.L. and F.M.; writing—original draft preparation, P.L. and T.L.; writing—review and editing, F.M. and T.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Applied Research and Development Project of Gansu Academy of Sciences (2021JK-04), the Science and Technology Planning Project of Chengguan District of Lanzhou (2023JSCX0004), Cooperation project between Lanzhou University of Technology and Longnan Power Supply Company of State Grid Gansu Electric Power Company (HX2023C5080001).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Due to the data sensitive, the data presented in this article are not available.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Wang, J.; Zhu, H.; Zhang, Y.; Cheng, F.; Zhou, C. A novel prediction model for wind power based on improved long short-term memory neural network. *Energy* **2023**, *265*, 126283. [[CrossRef](#)]
2. Zheng, H.; Hu, Z.; Wang, X.; Nie, J.; Cui, M. VMD-CAT: A hybrid model for short-term wind power prediction. *Energy Rep.* **2023**, *9*, 199–211. [[CrossRef](#)]
3. Hu, X.; Yu, Q.; Han, Y.; Chen, Z.; Geng, Z. Novel complex-valued long short-term memory network integrating variational mode decomposition for soft sensor. *J. Process Control.* **2023**, *129*, 103053. [[CrossRef](#)]
4. Ahmed, A.; Khalid, M. A review on the selected applications of forecasting models in renewable power systems. *Renew. Sustain. Energy Rev.* **2019**, *100*, 9–21. [[CrossRef](#)]
5. Zhang, Y.; Le, J.; Liao, X.; Feng, Z. A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing. *Energy* **2019**, *168*, 558–572. [[CrossRef](#)]
6. Puggini, L.; Seán, M. An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data. *Eng. Appl. Artif. Intell.* **2018**, *67*, 126–135. [[CrossRef](#)]
7. Parri, S.; Teeparthi, K.; Kosana, V. A hybrid methodology using VMD and disentangled features for wind speed forecasting. *Energy* **2024**, *288*, 0360–5442. [[CrossRef](#)]
8. Chen, H.; Wu, H.; Kan, T.; Zhang, J.; Li, H. Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction. *Int. J. Electr. Power Energy Syst.* **2023**, *154*, 109420. [[CrossRef](#)]
9. Zhao, Z.; Yun, S.; Jia, L.; Guo, J.; Meng, Y.; He, N.; Yang, L. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105982. [[CrossRef](#)]
10. Huang, N.; Yuan, C.; Cai, G.; Xing, E. Hybrid Short Term Wind Speed Forecasting Using Variational Mode Decomposition and a Weighted Regularized Extreme Learning Machine. *Energies* **2016**, *9*, 989. [[CrossRef](#)]
11. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training Recurrent Neural Networks. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1310–1318.
12. Guan, S.; Wang, Y.; Liu, L.; Gao, J.; Xu, Z.; Kan, S. Ultra-short-term wind power prediction method based on FTI-VACA-XGB model. *Expert Syst. Appl.* **2024**, *235*, 121185. [[CrossRef](#)]
13. Yin, L.; Zhao, M. Inception-embedded attention memory fully-connected network for short-term wind power prediction. *Appl. Soft Comput.* **2023**, *141*, 110279. [[CrossRef](#)]
14. Yang, S.; Yuan, A.; Yu, Z. A novel model based on CEEMDAN, IWOA, and LSTM for ultra-short-term wind power forecasting. *Environ. Sci. Pollut. Res.* **2023**, *235*, 11689–11705. [[CrossRef](#)] [[PubMed](#)]
15. Erick, L.; Carlos, V.; Héctor, A.; Esteban, G.; Henrik, M. Wind Power Forecasting Based on Echo State Networks and Long Short-Term Memory. *Energies* **2018**, *11*, 526. [[CrossRef](#)]

16. Yu, C.; Li, Y.; Bao, Y.; Tang, H.; Zhai, G. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. *Energy Convers. Manag.* **2018**, *178*, 137–145. [[CrossRef](#)]
17. Curreri, F.; Patanè, L.; Gabriella Xibilia, M. RNN- and LSTM-Based Soft Sensors Transferability for an Industrial Process. *Sensors* **2021**, *21*, 823. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, F.; Li, N.; Li, L.; Wang, S.; Du, C. A local semi-supervised ensemble learning strategy for the data-driven soft sensor of the power prediction in wind power generation. *Fuel* **2023**, *333*, 126435. [[CrossRef](#)]
19. Li, H.; Jing, H.; Zhang, R.; Gao, Z. Wind power forecast based on improved Long Short Term Memory Network. *Energy* **2019**, *189*, 116300.
20. Yu, M.; Niu, D.; Gao, T.; Wang, K.; Sun, L.; Li, M.; Xu, X. A novel framework for ultra-short-term interval wind power prediction based on RF-WOA-VMD and BiGRU optimized by the attention mechanism. *Energy* **2023**, *269*, 126738. [[CrossRef](#)]
21. Abou Houran, M.; Bukhari, S.M.S.; Zafar, M.H.; Mansoor, M.; Chen, W. COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications. *Appl. Energy* **2023**, *349*, 121638. [[CrossRef](#)]
22. Liu, T.; Ting, K.; Zhou, Z. Spectrum of variable-random trees. *J. Artif. Intell. Res.* **2008**, *32*, 355–384. [[CrossRef](#)]
23. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
24. Gal, Y.; Ghahramani, Z.B.A. Theoretically Grounded Application of Dropout in Recurrent Neural Networks. *Statistics* **2016**, *29*, 285–290.
25. Duan, J.; Wang, P.; Ma, W.; Tian, X.; Fang, S.; Chen, Y.; Chang, Y.; Liu, H. Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and Correntropy Long Short-term memory neural network. *Energy* **2021**, *214*, 118980. [[CrossRef](#)]
26. Liu, W.; Liu, Y.; Fu, L.; Yang, M.; Hu, R. Wind Power Forecasting Method Based on Bidirectional Long Short-Term Memory Neural Network and Error Correction. *Electr. Power Compon. Syst.* **2022**, *49*, 1169–1180. [[CrossRef](#)]
27. Hu, X.; Ma, L. Application of VMD-LSTM algorithm in short term load forecasting. *Electr. Power Sci. Eng.* **2018**, *34*, 9.
28. Wang, J.; Li, X.; Zhou, X.; Zhang, K. Ultra-short-term wind speed prediction based on VMD-LSTM. *Power Syst. Prot. Control.* **2020**, *34*, 45–52.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.