

Article

# An Online Hybrid Model for Temperature Prediction of Wind Turbine Gearbox Components

Qiang Zhao <sup>1</sup>, Kunkun Bao <sup>1,\*</sup>, Jia Wang <sup>1</sup>, Yinghua Han <sup>2</sup> and Jinkuan Wang <sup>3</sup>

<sup>1</sup> School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China; zhaoqiang@neuq.edu.cn (Q.Z.); 1901902@stu.neu.edu.cn (J.W.)

<sup>2</sup> School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China; yghan@neuq.edu.cn

<sup>3</sup> College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; wjk@neuq.edu.cn

\* Correspondence: 1801875@stu.neu.edu.cn; Tel.: +86-1553-998-6652

Received: 4 September 2019; Accepted: 14 October 2019; Published: 16 October 2019



**Abstract:** Condition monitoring can improve the reliability of wind turbines, which can effectively reduce operation and maintenance costs. The temperature prediction model of wind turbine gearbox components is of great significance for monitoring the operation status of the gearbox. However, the complex operating conditions of wind turbines pose grand challenges to predict the temperature of gearbox components. In this study, an online hybrid model based on a long short term memory (LSTM) neural network and adaptive error correction (LSTM-AEC) using simple-variable data is proposed. In the proposed model, a more suitable deep learning approach for time series, LSTM algorithm, is applied to realize the preliminary prediction of temperature, which has a stronger ability to capture the non-stationary and non-linear characteristics of gearbox components temperature series. In order to enhance the performance of the LSTM prediction model, the adaptive error correction model based on the variational mode decomposition (VMD) algorithm is developed, where the VMD algorithm can effectively solve the prediction difficulty issue caused by the non-stationary, high-frequency and chaotic characteristics of error series. To apply the hybrid model to the online prediction process, a real-time rolling data decomposition process based on VMD algorithm is proposed. With aims to validate the effectiveness of the hybrid model proposed in this paper, several traditional models are introduced for comparative analysis. The experimental results show that the hybrid model has better prediction performance than other comparative models.

**Keywords:** deep learning; time series; temperature prediction; adaptive error correction; wind turbines; VMD

## 1. Introduction

Wind energy, as a clean and renewable energy, now has been one of the major potential and practical renewable resources. In recent years, the installed capacity of wind turbines all over the world has increased rapidly [1,2]. With the increase of installed capacity and wind turbine complexity, frequent malfunctions result in low reliability and expensive maintenance costs of wind turbines. According to statistics, the cost of operation and maintenance of onshore wind farms and offshore wind farms account for about 15–20% and 30–35% of the total revenue, respectively [2,3]. To raise the availability and reliability of wind turbines, monitoring the operation status of wind turbines and detecting potential faults are increasingly significant. Gearbox, as a key component of wind turbines, often occurs various faults, which leads to high maintenance costs. Statistically, the maintenance cost caused by gearbox is as high as 13% of the total cost [4]. In recent years, monitoring the operation status of the gearbox has attracted wide attention.

With the development of the wind power industry, there are numerous studies on wind turbines fault diagnosis and condition monitoring. According to the methods adopted by these studies, they can be roughly classified into two types: model-based methods and data-driven methods [5]. In addition to classical methods such as state estimation and parameter estimation, many new model-based studies have been proposed in recent years [6–10]. In [8], a set-valued approach is proposed for wind turbine fault diagnosis. In order to ensure the performance of fault diagnosis, model-based methods need to establish accurate mathematical models of wind turbines system. However, due to the complexity of wind turbine systems, it is difficult to establish an accurate mathematical model, which leads to the difficulty of model-based in practical application [5]. In contrast, data-driven methods do not require accurate mathematical models, and most wind turbines are equipped with a supervisory control and data acquisition (SCADA) system, which makes it easy to obtain data. Therefore, the data-driven method is a very worthwhile aspect to be studied for wind turbine fault diagnosis and condition monitoring. The temperature of gearbox components is closely related to the operation state of the gearbox. Excessive temperature will cause the occurrence of faults. Similarly, the occurrence of faults in a component will also be accompanied by a significant change in temperature [11]. Therefore, high temperature warning of gearbox components is crucial for condition monitoring of wind turbines and reduction of operational and maintenance costs. The key of high-temperature warning is to improve the accuracy of the temperature prediction model as much as possible. In this paper, a data-driven method based on temperature prediction is studied to monitor the operation status of the gearbox.

Generally, according to the sources of data, the time series prediction models can be divided into two categories as the multi-variable models and single-variable models in the wind turbines system. At present, most temperature prediction models adopt multi-variable data based on SCADA system [12,13]. Huang et al. [12] put up with a hybrid method combining principal component analysis (PCA) and nonlinear autoregressive dynamic neural network to establish a gearbox oil temperature prediction model. Wang et al. [13] presented a condition monitoring method of wind turbine main bearing based on the deep belief network (DBN), where DBN is adopted to establish the normal temperature prediction model, so as to realize the condition monitoring of wind turbine main bearing. However, the use of multi-variate data may increase the complexity and uncertainty of the modeling process, which will reduce the performance of the prediction model. Compared to the multi-variable model, the single-variable model has lower computational complexity and easier data acquisition [14].

Although single-variable methods are seldom used in temperature prediction of gearbox components, many prediction methods have been proven to be effective in other aspects of wind energy systems, such as wind speed and wind power. The prediction methods can be roughly classified into three categories: the statistical methods [15,16], conventional machine learning methods [17,18] and deep learning methods [19,20]. Among the statistical methods, autoregressive integrated moving average (ARIMA) is the most classical and widely adopted model. However, most statistical methods are difficult to deal with the non-linear characteristics of the time series, which results in low prediction accuracy. In addition, the conventional machine learning methods are also widely chosen in time series prediction, which mainly include back propagation (BP) neural networks, radial basis function (RBF) neural network, extreme learning machine (ELM), support vector machine (SVM) methods and so on. Nevertheless, although the traditional machine learning method is an intelligent method, its ability of learning data nonlinearity and non-stationarity is not strong because of its shallow structure. In recent years, with the breakthrough of neural network technology, deep learning approaches have attracted wide attention because of its better performance in many tasks. Compared with the shallow methods, the deep learning methods can have a better ability of non-linear expression and data feature extraction [21]. Wang et al. [19] carried out a novel hybrid deep learning-based approach. The comparison results indicate that the hybrid model can better learn the non-linear and non-stationary characteristics.

The performance of gearbox condition monitoring depends on a high precision temperature prediction model, especially in the part of the high-temperature series. To this end, it is of great significance to develop optimization methods for promoting prediction performance. The existing optimization algorithms have three main aspects, including signal processing techniques [22–24], parameters optimization techniques [25,26] and error correction techniques [27,28]. As shown in Table 1, it is a summary of the above-mentioned and related algorithms.

In signal processing techniques, the signal decomposition method is widely used, such as empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), fast ensemble empirical mode decomposition (FEEMD) and complete ensemble empirical mode decomposition (CEEMDAN). Various literatures have proved the effectiveness of decomposition algorithms. However, these traditional decomposition methods have some shortcomings. For example, sometimes it is difficult to decompose multiple low-frequency components for wavelet decomposition (WD) and wavelet packet decomposition (WPD), while other decomposition algorithms, including EMD, EEMD, FEEMD and so on, currently lack the strict mathematical proof [29]. In order to overcome these drawbacks, some new decomposition algorithms are adopted in time series prediction, such as empirical wavelet transform (EWT) and variational mode decomposition (VMD). In [24], the VMD approach is chosen to decompose the corresponding time-series signals, which avoids the interaction between different modes. In addition to the decomposition algorithm mentioned above, error correction is also a method to improve the performance of the prediction model [30]. In [28], an error correction model based on ICEEMDAN and ARIMA algorithm is proposed to promote the prediction accuracy.

In addition, there are still some deficiencies in the field of research, which need to be further studied. First, many literatures decompose training data and testing data together [31,32], which is not feasible in the process of real-time prediction. Regretfully, other literature does not clearly explain the construction process of the modeling data. Second, different from the wind speed prediction, the temperature will drop dramatically due to shutdown and other factors in the operation of wind turbines, which will result in inaccurate prediction results.

In the study, a new hybrid forecasting method is proposed, which consists of a preliminary temperature prediction model and an adaptive error correction model. The innovations and contributions of the proposed hybrid model are as follows: (a) with aims to avoid the complexity and uncertainty of multi-variable prediction model, a prediction model based on single-variable data is proposed. In this paper, a more suitable deep learning model for time series analysis, long short term memory (LSTM) model, is adopted, which can better learn the non-linear and non-stationary characteristics of temperature series; (b) in view of the problem of drastic temperature drop caused by the above mentioned downtime phenomenon, an adaptive error correction model is designed to improve the precision of prediction model; (c) to avoid the weakness of some decomposition algorithms mentioned above such as EMD, EEMD, FEEMD and CEEMDAN, the VMD decomposition algorithm is employed in this paper, which can effectively reduce the chaotic characteristics and non-stationary of error series; (d) in view of the above mentioned the modeling data construction problems, a rolling data decomposition process which can be applied in practice is proposed.

The organizational structure of the paper is as follows: (a) the framework and algorithms of the hybrid prediction model are explained in Section 2; (b) gearbox components temperature forecasting case studies are presented in Section 3; and (c) conclusions are drawn in Section 4.

**Table 1.** A summary of existing algorithms related to the proposed hybrid model.

	Methods	Article	Data	Model
prediction methods	statistical methods	Masseran et al. [15] Poggi et al. [16]	single-variable single-variable	ARIMA-ARCH AR
	conventional machine learning methods	Huang et al. [12] Li et al. [17] Abdoos et al. [18]	multi-variable single-variable single-variable	PCA, NARX ANN,RBF ELM
	deep learning methods	Wang et al. [13] Wang et al. [19] Wang et al. [20]	multi-variable single-variable single-variable	DBN DBN CNN
optimization methods	signal processing techniques	Liu et al. [22] Mi et al. [23] Naik et al. [24]	single-variable single-variable single-variable	WPD, FEEMD WPD, EMD VMD
	parameter optimization techniques	Meng et al. [25] Liu et al. [26]	single-variable multi-variable	crisscross optimization GA
	error correction techniques	Wang et al. [27] Wang et al. [28]	single-variable single-variable	Markov ICEEMDAN-ARIMA

## 2. Methodology

### 2.1. The Overall Framework of the Proposed Model

The overall framework of the hybrid model presented in this paper is shown in Figure 1. The general process of the proposed hybrid model is described as follows:

- The original temperature series was predicted by the LSTM model to generate preliminary prediction results. Meanwhile, error series was generated by comparing predicted values with actual values.
- Faced with the non-stationary, high-frequency and chaotic characteristics of error series, the VMD decomposition algorithm was employed to decompose it into sub-sequences of different frequencies. In order to apply the model to the online prediction process, as shown in Figure 2, a rolling data decomposition process was developed. In Figure 2,  $T_i$ ,  $U_i$  and  $R_{i,j}$  represent the original temperature series, the error series of the preliminary prediction and the frequency component of error series decomposed by the VMD algorithm respectively, where  $i$  is a time label and  $j$  stands for the labels of different frequency components.
- The prediction model of each frequency component was established by the error prediction model, and the final error prediction results were reconstructed based on the adaptive error correction algorithm.
- The final forecasting results were obtained by adding the error prediction results with the preliminary temperature prediction results. When the predicted temperature exceeds a certain threshold, a high-temperature warning should be carried out.

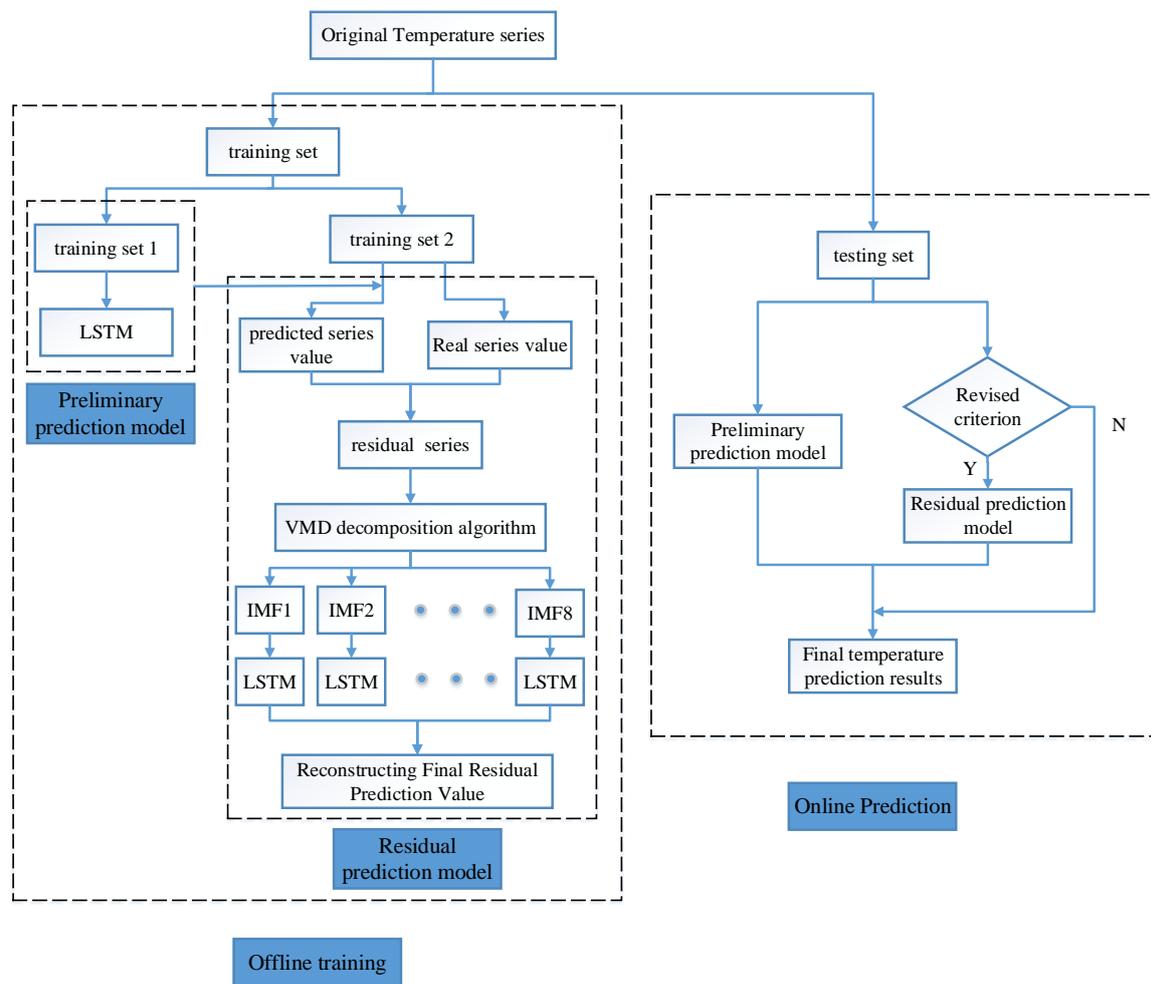


Figure 1. The overall framework of the proposed model.

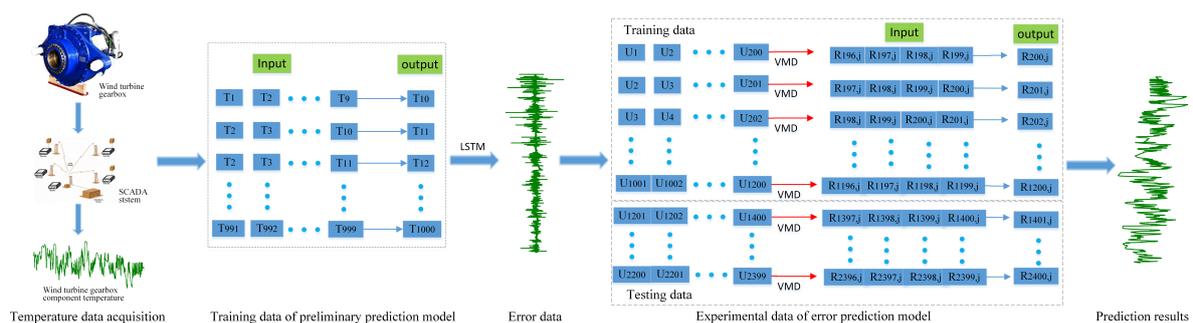


Figure 2. The construction process of modeling data for temperature prediction model.

### 2.2. Preliminary Prediction Model

This paper is devoted to the temperature prediction of wind turbines gearbox components so as to better realize the condition monitoring of gearbox. To avoid the complexity and uncertainty of the multi-variable prediction model, a single-variable prediction method is proposed. Due to the influence of complex operational conditions, it is difficult for conventional machine learning approaches to learn the nonlinear and non-stationary characteristics of gearbox components temperature data. Compared with traditional machine learning methods, deep learning methods have stronger non-linear

expression ability. LSTM is a deep learning model, which not only has stronger non-linear expression ability, but also is more suitable for the prediction model of time series because of its memory characteristics. Therefore, the LSTM algorithm was applied to the preliminary prediction model of the gearbox component temperature series in this paper.

The LSTM neural network is an improved model based on a recurrent neural network (RNN) [33]. The output of LSTM depends not only on the input and weight of the current neuron, but also on the input of the previous neuron. Therefore, the LSTM structure is usually more suitable for processing time-series data. The basic unit structure of the LSTM model is shown in Figure 3. Four elements, including state of each unit, input gate, forget gate and output gate, are the core of the LSTM model. The relationship of the LSTM unit states and the three gates are expressed as Equations (1)–(5) [34,35].

$$f_t = \sigma([w_{xf}, w_{hf}, e_f] \cdot [X_t, h_{t-1}, b_f]^T) \quad (1)$$

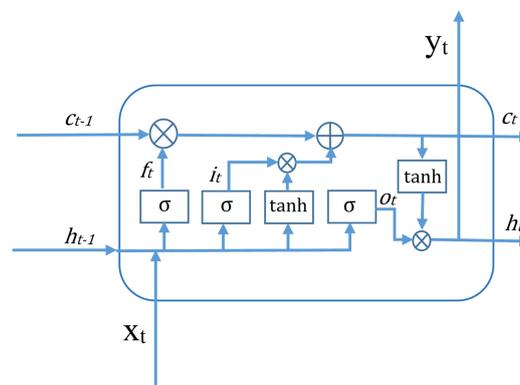
$$i_t = \sigma([w_{xi}, w_{hi}, e_i] \cdot [X_t, h_{t-1}, b_i]^T) \quad (2)$$

$$c_t = f_t * c_{t-1} + i_t * \tanh([w_{xc}, w_{hc}, e_c] \cdot [X_t, h_{t-1}, b_c]^T) \quad (3)$$

$$o_t = \sigma([w_{xo}, w_{ho}, e_o] \cdot [X_t, h_{t-1}, b_o]^T) \quad (4)$$

$$h_t = y_t = o_t * \tanh(c_t), \quad (5)$$

where  $X_t$  are input vectors;  $i_t$ ,  $o_t$  and  $f_t$  represent the output results of input gate, output gate and forget gate, respectively;  $c_t$  represents the activation status of each cell;  $h_t$  is the output results of memory unit. In addition,  $w_{xi}$ ,  $w_{hi}$ ,  $w_{xf}$ ,  $w_{hf}$ ,  $w_{xc}$ ,  $w_{hc}$ ,  $w_{xo}$  and  $w_{ho}$  are the corresponding weight vectors;  $\tanh$  and  $\sigma$  represent activation functions;  $b_i$ ,  $b_f$ ,  $b_c$  and  $b_o$  are the corresponding bias vectors;  $e_c$ ,  $e_o$ ,  $e_i$ ,  $e_f$  is the vectors of all 1 corresponding to  $b_c$ ,  $b_o$ ,  $b_i$ ,  $b_f$ .



**Figure 3.** The basic structure of long short term memory (LSTM) network.

### 2.3. Adaptive Error Correction Model

The temperature prediction accuracy of gearbox components greatly affects the high-temperature monitoring performance of gearbox. Therefore, an adaptive error correction model is presented in this paper, which can increase the accuracy of prediction by predicting error. However, due to the non-stationary and chaotic characteristics of the error series, it is difficult to predict the error series directly. Signal decomposition can effectively reduce the non-stationary and chaotic characteristics of time series, and many literatures have proved the effectiveness of the signal decomposition method. With aims to overcome the shortcomings of conventional decomposition, algorithms such as mode mixing problem and lack of mathematical proof, a kind of state-of-art VMD algorithm was applied in this paper. The final error value was reconstructed by predicting each decomposition component.

### 2.3.1. The VMD Algorithm

VMD, as a new signal decomposition method, has been widely used in recent years. Unlike EMD recursive solution, VMD transforms the solution problem into a variational problem. The purpose of the VMD algorithm is to find the inherent modal components of a specified number. To solve this variational problem, the alternate direction method of multipliers (ADMM) is selected to solve the modes and corresponding central frequencies. The specific algorithm process of the VMD algorithm is as follows [36].

#### (1) The constructive process of variational problems

To calculate the bandwidth of each mode component, the analytical signals of each mode component are obtained by Hilbert transform, and then the unilateral frequency spectrum is obtained as follows.

$$[\delta(t) + \frac{j}{\pi t}] * u_k(t), \quad (6)$$

where intrinsic mode function (IMF) is defined as an amplitude modulated frequency modulated signal. Its expression is  $u_k(t) = A_k(t)\cos(\phi_k(t))$ .  $\delta(t)$  is the Dirac distribution.

Then the corresponding baseband is obtained by spectrum conversion of analytic signal.

$$[(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t}. \quad (7)$$

By calculating the  $L^2$ -norm of the above analytical signal derivative and the bandwidth of each mode, the constrained variational problem is constructed as follows.

$$\min_{u_k, \omega_k} \sum_k \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t}\|^2 \quad s.t. \sum_k u_k = f, \quad (8)$$

where  $f$  is an input signal;  $\{u_k\} = \{u_1, \dots, u_K\}$  and  $\{\omega_k\} = \{\omega_1, \dots, \omega_K\}$  represent different modal components and corresponding frequency centers, respectively. In addition,  $\sum_k u_k = \sum_{k=1}^K u_k$ .  $\partial_t(\cdot)$  stands for differential symbols.  $K$  is the total number of sub-signals.

#### (2) The solution process of the variational problem

To solve this variational problem, the constrained variational problems of Equation (8) are transformed into unconstrained variational problems by using Lagrange multiplier method.

$$L(\{u_k\}, \{\omega_k\}, \lambda(t)) = \left( \eta \sum_k \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t}\|^2 + \|f(t) - \sum_k u_k(t)\|^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \right), \quad (9)$$

where  $\eta$  is a quadratic multiplication factor and  $\lambda(t)$  represents Lagrangian multipliers.

The ADMM algorithm is used to solve the above variational problems. In ADMM algorithm, the saddle point of the Lagrangian expression can be found by alternately updating  $u_k^{n+1}$ ,  $\omega_k^{n+1}$  and  $\lambda^{n+1}$ . Among them,  $u_k^{n+1}$  can be updated using the following equation.

$$u_k^{n+1} = \arg \min_{u_k \in X} \left\{ \eta \|\partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t}\|^2 + \|f(t) - \sum_i u_i(t) + \frac{\lambda(t)}{2}\|^2 \right\} \quad (10)$$

where the  $\omega_k$  and the  $\sum u_i(t)$  are equivalent to  $\omega_k^{n+1}$  and  $\sum u_i(t)^{n+1}$ , respectively.  $n$  is the number of iterations.

By using the Parseval/Plancherel Fourier isometry transformation, Equation (10) can be converted into a frequency domain form and solved in the frequency domain.

$$\hat{u}_k^{n+1} = \arg \min_{u_k \in X} \left\{ \eta \left\| j\omega [(1 + \text{sgn}(\omega + \omega_k)) \hat{u}(\omega + \omega_k)] \right\|^2 + \left\| \hat{f}(\omega) - \sum_i \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right\|^2 \right\} \quad (11)$$

where  $\hat{\cdot}$  is used to represent the frequency form of the corresponding signal.  $\text{sgn}$  is sign function.

Then the  $\omega$  can be updated as  $\omega - \omega_k$  in the first part.

$$\hat{u}_k^{n+1} = \arg \min_{u_k \in X} \left\{ \eta \left\| j(\omega - \omega_k) [(1 + \text{sgn}(\omega)) \hat{u}_k(\omega)] \right\|^2 + \left\| \hat{f}(\omega) - \sum_i \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right\|^2 \right\} \quad (12)$$

The problem can be changed into a non-negative frequency interval integral form.

$$\hat{u}_k^{n+1} = \arg \min_{u_k \in X} \left\{ \int_0^\infty 4\eta(\omega - \omega_k)^2 |\hat{u}_k|^2 + 2 \left| \hat{f}(\omega) - \sum_k \hat{u}_k(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right|^2 d\omega \right\} \quad (13)$$

Finally, the solution of the quadratic optimization problem can be obtained as follow.

$$\hat{u}_k^{n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\eta(\omega - \omega_k)^2} \quad (14)$$

where  $\hat{u}_k^{n+1}(\omega)$  can be regarded as the Wiener filtering of the current residual. Similarly, the central frequencies of the corresponding modes are updated as follows:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k|^2 d\omega}{\int_0^\infty |\hat{u}_k|^2 d\omega} \quad (15)$$

The  $\omega_k^{n+1}$  is the power spectrum center of the k-th modal component at the n+1 iteration.

The  $\hat{\lambda}^{n+1}(\omega)$  can be updated as:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \rho [\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)] \quad (16)$$

where  $\rho$  is the update coefficient of  $\hat{\lambda}^{n+1}(\omega)$ .

### 2.3.2. Adaptive Error Correction Algorithm

A prediction model is needed to predict each modal component after VMD decomposition. To simplify the complexity of the model, the LSTM model is employed to predict each component decomposed. The input in the error prediction model is the data of the past four moments. In the selection of the input number of the error prediction model, the grid search method is used to search the optimal parameters in the prediction performance of the model. Finally, each prediction component is reconstructed to get the final prediction value. However, the error prediction model has better prediction performance for the weak volatility part of the series than for the strong volatility part. Faced with highly volatile parts, the correction model may lead to deteriorating results. To reduce this situation, the following adaptive error correction algorithms are proposed to further improve the accuracy. The adaptive error correction algorithms are mainly considered in two aspects: effectiveness of correction model and amplitude analysis of primary error series. Given  $d$ ,  $m$  and  $c$  are the results of error prediction, error series after correction and error series before correction, respectively. When  $g > 0$ , the correction is defined as invalid. Where  $g$  is equivalent to the difference between  $|m|$  and  $|c|$ . When  $g$  exceeds a certain threshold for continuous moments, the correction of the next time may also be invalid. In addition, to effectively decrease the influence of worsening correction, it is necessary

to limit the amplitude of correction errors. The adaptive error correction algorithm is described in Algorithm 1.

---

**Algorithm 1** The adaptive error correction algorithm.

---

```

1: set  $g = |m| - |c|, k = [1, 2, 3, 4]$ 
2: for each  $i \in [2, \text{len}(g)]$  do
3:   if ( $g[i - 1] > \zeta$ ) and ( $g[i - 2] > \zeta$ ) and ( $|c| < 5$ ) then
4:      $d[i] = 0$ 
5:   end if
6: end for
7: for each  $i \in [2, \text{len}(g)]$  do
8:   if ( $c[i - 1] > 0$ ) then
9:     if ( $c[i - k] > \beta$ ) then
10:       $d[i] = d[i]$ 
11:    else if ( $c[i - k] > 0$ ) then
12:       $d[i] = \alpha$ 
13:    else
14:       $d[i] = 0$ 
15:    end if
16:  else
17:    if ( $c[i - k] < -\beta$ ) then
18:       $d[i] = d[i]$ 
19:    else if ( $c[i - k] < 0$ ) then
20:       $d[i] = -\alpha$ 
21:    else
22:       $d[i] = 0$ 
23:    end if
24:  end if
25: end for

```

---

where  $\zeta$ ,  $\alpha$  and  $\beta$  is defined as a threshold.

In the process of threshold setting of  $\zeta$ ,  $\alpha$  and  $\beta$ , the magnitude of error series is analyzed, and the threshold is set by grid search within a reasonable range. The effect of threshold setting on the accuracy of the hybrid temperature prediction model was studied by a grid search algorithm. In the experiment, the prediction performance of the model fluctuates slightly due to the influence of parameter initialization. Therefore, when choosing thresholds, we choose a group of thresholds whose prediction performance is in the middle, which can make the algorithm have better generalization ability and robustness.

#### 2.4. Model Performance Evaluation

In order to compare the prediction performance of different prediction models, three evaluation indexes, including the mean square error (MSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE), are exploited in this study. The equations of three evaluation indexes are explained as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(\hat{y}_i - y_i)| \quad (18)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{(\hat{y}_i - y_i)}{y_i} \right| \quad (19)$$

where  $\hat{y}_i$  and  $y_i$  are the predicted results of the model and the actual temperature values, respectively; and  $n$  is the length of the predicted temperature series.

### 3. Case Study and Contrast Analysis

#### 3.1. Data Description

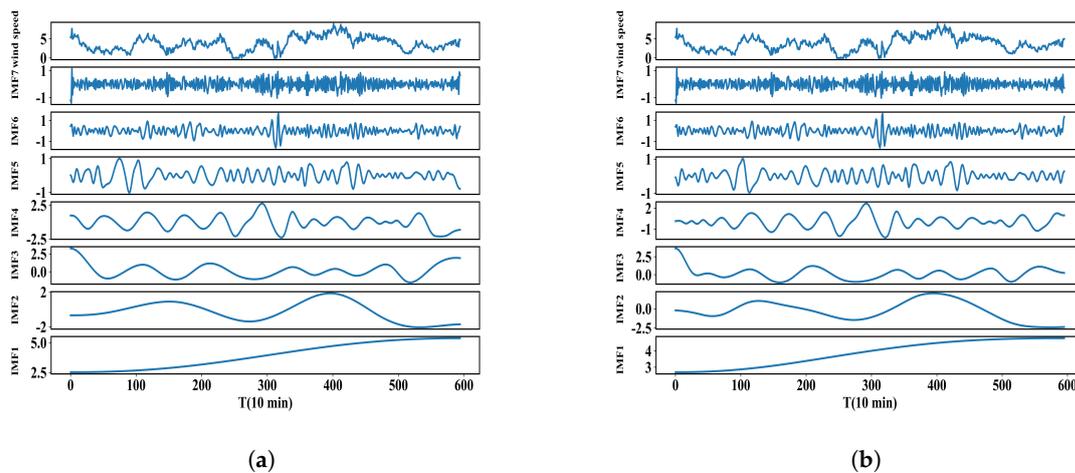
Almost all wind turbines are equipped with a SCADA system, which makes it very easy to obtain the temperature data of the gearbox components. In this study, the data is gathered from one wind farms in Shandong Province, China, which contains thirty-three wind turbines SCADA data at 10-min intervals from 1 February 2014 to 27 June 2014. In this study, three wind turbine prediction cases are provided to verify the superiority of the proposed hybrid model. The temperature data of gearbox components, including gearbox oil temperature, gearbox input shaft temperature, and gearbox output shaft temperature, are from SCADA system of #1, #2 and #3, where #1, #2 and #3 represent wind turbine 1, wind turbine 2 and wind turbine 3, respectively. Each data set contains 6400 series of 10-min data and is divided into two parts, including the first 5400 temperature series and the last 1000 series, which were used in the training process and the testing process, respectively. Generally speaking, in common types of wind turbines, the oil temperature early warning temperature threshold and alarm temperature threshold of the gearbox can be set to  $75^\circ$  and  $80^\circ$  respectively. The high temperature warning threshold of gearbox input and output shaft can be set to  $80^\circ$ . In addition, this paper also collects two sets of wind speed data of wind turbine 1 to analyze the influence of the decomposition process for the on-line prediction model. The two datasets contain 600 and 601 observations at 10-min intervals in time scale, respectively.

#### 3.2. Simulation Result

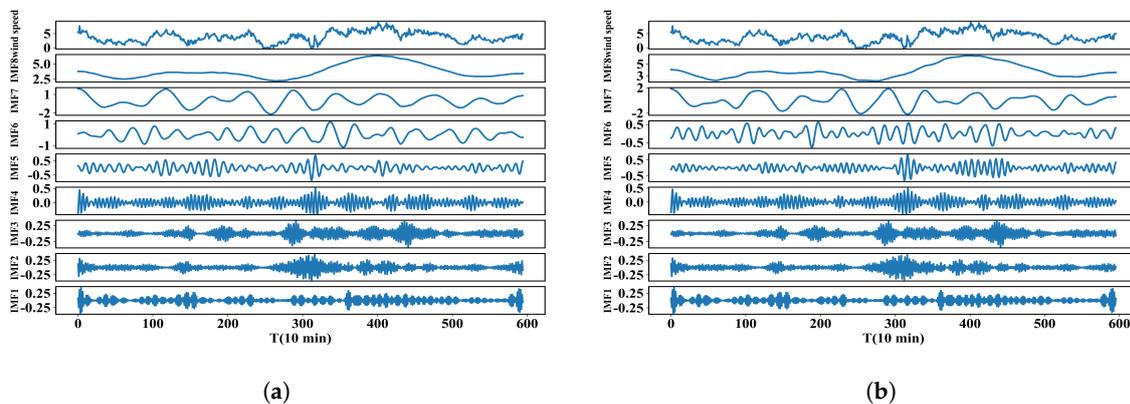
##### 3.2.1. The Case of Decompose Algorithm

To analyze the application of the decomposition algorithm in real-time time series prediction, the above two wind speed series are decomposed by EMD and VMD algorithm. Figures 4 and 5 show the decomposition results.

Through the analysis of Figures 4 and 5, it can be found that whether EMD or VMD decomposition algorithm, the new data may affect the results of previous data decomposition to a certain extent, which shows that the new data has a guiding effect on the results of the previous data decomposition. Therefore, it is not suitable for a real-time prediction model to decompose training data and testing data together. As shown in Figure 2, a real-time rolling data decomposition process based on VMD algorithm is proposed, which can be better applied to real-time prediction process. In the training data of the preliminary prediction model, the original temperature series is used to establish the preliminary prediction model (LSTM). In the training process of LSTM model, nine temperature values (such as  $T_1 \dots T_9$ ) are used as input vectors ( $X_i$ ) and the next temperature value (such as  $T_{10}$ ) is used as output ( $y_i$ ). Then, the error series generated by comparing the predicted result with actual value. In the training data of error prediction model, every 200 error series (such as  $U_1 \dots U_{200}$ ) as a group are decomposed by the VMD algorithm. Then the last data after decomposition (such as  $R_{200,j}$ ) is used as the predicted value, and the four data (such as  $R_{196,j} \dots R_{199,j}$ ) before the last data are used as input. In the testing data of the error prediction model, the last four decomposed data (such as  $R_{1397,j} \dots R_{1400,j}$ ) are taken as input vectors. The final error prediction results are reconstructed by predicting the value of each frequency component. The final prediction results are obtained by adding the adaptive error prediction results with the preliminary temperature prediction results.



**Figure 4.** The decomposition results of different wind speed series by the empirical mode decomposition (EMD) algorithm: (a) wind speed series 1; (b) wind speed series 2.



**Figure 5.** The decomposition results of different wind speed series by variational mode decomposition (VMD) algorithm: (a) wind speed series 1; (b) wind speed series 2.

### 3.2.2. The Case of Gearbox Components Temperature Prediction

The case uses gearbox components temperature data of #1, #2 and #3. Each experiment consists of seven prediction models, including the LSTM model, the BP neural network, the ELM model, the LSTM model with error correction (LSTM-EC), the ELM model with error correction (ELM-EC), the ELM model with adaptive error correction (ELM-AEC) and LSTM-AEC. In the experiment of comparing the hybrid model with other models, all models have similar parameter settings. All models are built and simulated under Windows 10 operating system, Inter-Core i5-7500 CPU @ 3.40 GHz and RAM of 8 GB. All the experiments are implemented through Python 3.6. The parameters  $\alpha$ ,  $\beta$ ,  $\zeta$  and the number of input data in the adaptive error correction algorithm are set to 0.5, 1, 1 and 4, respectively. In the preliminary temperature prediction model, nine temperature data of historical time were used as inputs of the model. The BP neural network, containing a hidden layer with 26 neurons, is used in the three experiments. Three experiments used ELM networks containing a hidden layer with 6, 10 and 10 neurons, respectively. In the training data selection of the LSTM model, in order to make the model more robust, 1000–4000 observations are selected from 4000 observations, which prevents special results from special training sets. The decomposition number of the VMD algorithm was set to 8. In addition, the learning rate was 0.6 and the Adagrad optimization algorithm was used in LSTM and BP models.

The temperature of three gearbox components, including gearbox oil temperature, gearbox input shaft temperature, and gearbox output shaft temperature, is predicted by the proposed hybrid

model. Tables 2–4 and Figures 6–14 show the prediction results of different models. In Figures 6–14, model A–H represents real value, LSTM-EC, LSTM-AEC, LSTM, ELM, ELM-AEC, ELM-EC and BP respectively. From the above prediction case, it can be concluded that:

- By comparing the predictive performance of LSTM, ELM, and BP, the forecasting accuracy of the LSTM model was higher than other prediction models under the same conditions. Take the prediction results of wind turbine one gearbox oil temperature as an example in Table 2, promoting of the MSE of the BP and ELM model by the LSTM model are 0.7129 and 0.4046, respectively. Thus, it can be seen that the LSTM model can learn more about the non-stationary and non-linear characteristics of temperature data to a certain extent.
- The prediction model with error correction has higher accuracy than the single prediction model in general. There are some prediction results, such as ELM and ELM-EC prediction results of the gearbox input shaft temperature in Table 3, which can prove this point. However, there are some special cases with opposite results, which contains three LSTM and LSTM-EC prediction results of the gearbox output shaft temperature and so on in Table 4. Therefore, it can be seen that some residual series will lead to worsening correction results.
- Whether with ELM or LSTM, the accuracy of the prediction model with adaptive error correction can be improved. For example, in the prediction results of gearbox oil temperature in Table 2, promoting of the MSE of the LSTM model by the LSTM-AEC model are 0.2317, 0.0654 and 0.0819, respectively.
- In all the prediction models involved, the proposed hybrid model has the best forecasting performance than other comparative models. From Figures 6–14, it can be seen that the predicted value of the proposed hybrid model in the high-temperature part is very accurate, which provides a guarantee for high-temperature warning of gearbox components. As shown in Figures 12 and 14, the temperature of the gearbox output shaft exceeds the high-temperature warning threshold at several points in #1 and #3 respectively, such as the high-temperature series starting from time points 45, 218, 629 and 938 in #1.

**Table 2.** Performance evaluations of different models for gearbox oil temperature prediction.

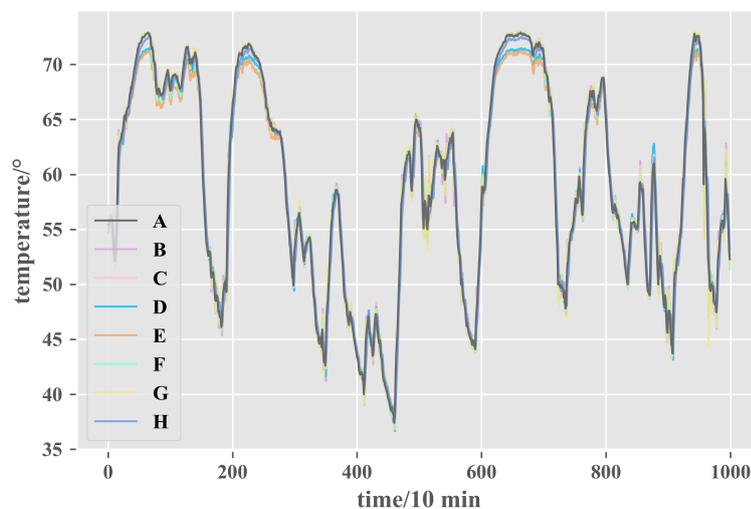
Model	Wind Turbine One			Wind Turbine Two			Wind Turbine Three		
	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE
BP	1.5008	0.9053	1.6086	1.5409	0.9932	1.8618	1.7483	1.0916	1.7954
ELM	1.1925	0.8580	1.4192	0.7639	0.6582	1.2515	1.8713	1.1128	1.7849
LSTM	0.7879	0.6476	1.0912	0.7425	0.6000	1.1564	0.7419	0.6194	1.0679
ELM-EC	0.8974	0.5929	1.0725	0.9790	0.6494	1.2663	0.9749	0.6270	1.1342
LSTM-EC	0.7373	0.5228	0.9471	0.8368	0.5815	1.1278	0.8096	0.5618	1.0225
ELM-AEC	0.6858	0.5438	0.9686	0.7066	0.5819	1.1310	0.7355	0.5911	1.0601
LSTM-AEC	0.5562	0.4902	0.8728	0.6771	0.5343	1.0401	0.6600	0.5426	0.9691

**Table 3.** Performance evaluations of different models for gearbox input shaft temperature prediction.

Model	Wind Turbine One			Wind Turbine Two			Wind Turbine Three		
	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE
BP	2.5913	1.0706	1.8015	1.9533	0.9144	1.6179	2.6841	1.0255	1.6514
ELM	1.6284	1.0104	1.5815	1.3447	0.7964	1.3750	2.3209	1.1940	1.8400
LSTM	1.1318	0.7194	1.2076	1.1029	0.5708	1.0103	1.6469	0.7458	1.1956
ELM-EC	1.4474	0.7071	1.1817	1.4839	0.7152	1.2518	2.3193	0.8454	1.3877
LSTM-EC	1.2412	0.6557	1.0925	1.2526	0.6527	1.1379	1.9056	0.7704	1.2582
ELM-AEC	1.1376	0.6430	1.0502	1.2871	0.6593	1.1627	1.6162	0.7207	1.1668
LSTM-AEC	1.0052	0.5630	0.9349	1.0691	0.5704	0.9981	1.5309	0.6614	1.0809

**Table 4.** Performance evaluations of different models for gearbox output shaft temperature prediction.

Model	Wind Turbine One			Wind Turbine Two			Wind Turbine Three		
	MSE	MAE	MAPE	MSE	MAE	MAPE	MSE	MAE	MAPE
BP	3.9916	1.2172	2.0092	3.1086	1.1294	1.9873	3.5203	1.2143	1.9511
ELM	3.1493	1.3990	2.1770	2.0997	0.9992	1.7154	3.4253	1.4656	2.2725
LSTM	2.2322	0.9961	1.7227	1.7538	0.8150	1.4096	2.3699	0.9408	1.5193
ELM-EC	3.1834	1.0349	1.7282	2.4298	0.9219	1.5838	3.4413	1.0565	1.7332
LSTM-EC	2.7991	0.9691	1.6238	2.1177	0.8580	1.4753	3.0108	0.9840	1.6220
ELM-AEC	2.2232	0.8794	1.4491	2.0220	0.8294	1.4387	2.4348	0.9101	1.4894
LSTM-AEC	2.0654	0.8140	1.3872	1.6990	0.7347	1.2785	2.2739	0.8257	1.3676

**Figure 6.** The comparison of different models for gearbox oil temperature prediction in #1.

In practical application, the training process of the model is completed off-line. Once the model training is completed, the model can be used for real-time temperature prediction, which is guaranteed by the rolling data decomposition process proposed in this paper. In our forecasting case, it's like simulating the whole process, including model training and real-time forecasting. In addition, the experimental time was measured. The training time of this hybrid model was about 325.1167 s, but it should be noted that the training of the model was completed off-line. At the same time, 1000 temperature values were predicted, which took 49.8446 s. The average prediction time of each temperature was 0.0498 s, which fully satisfied the demand for 10-minute interval temperature prediction.

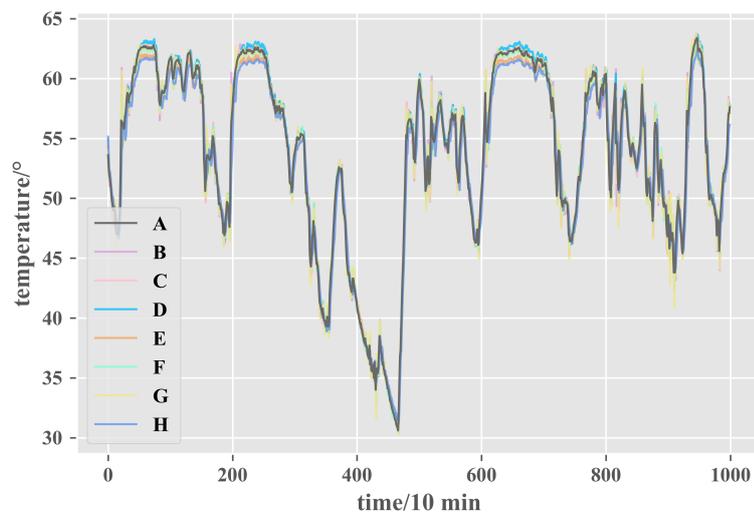


Figure 7. The comparison of different models for gearbox oil temperature prediction in wind #2.

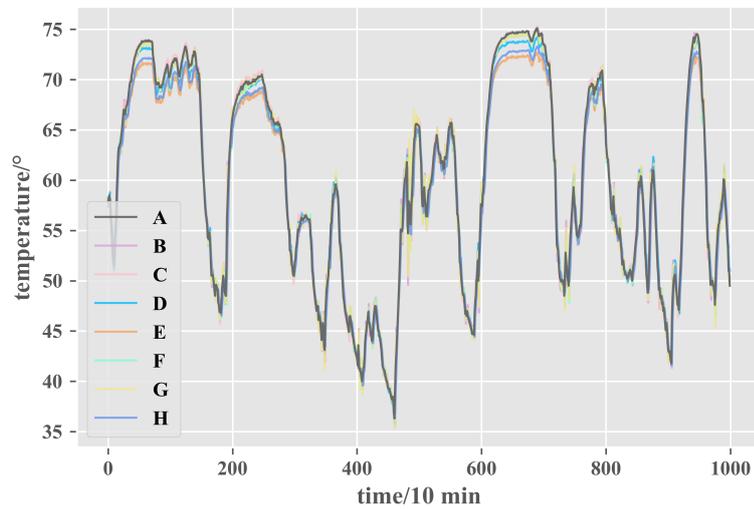


Figure 8. The comparison of different models for gearbox oil temperature prediction in #3.

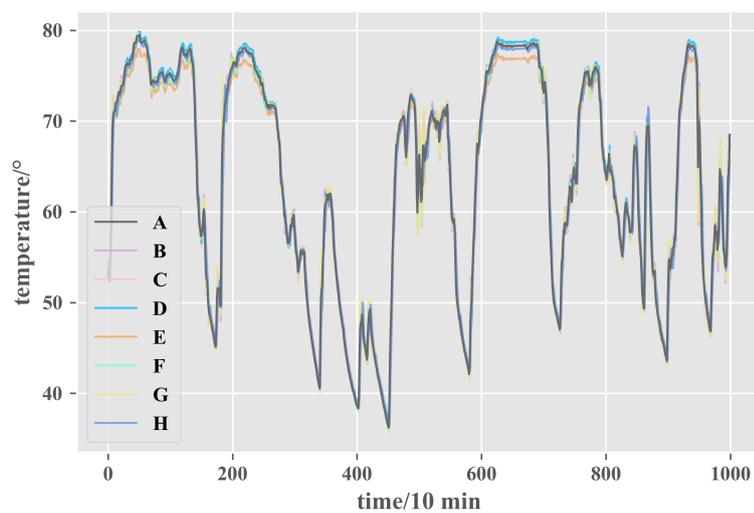


Figure 9. The comparison of different models for gearbox input shaft temperature prediction in #1.

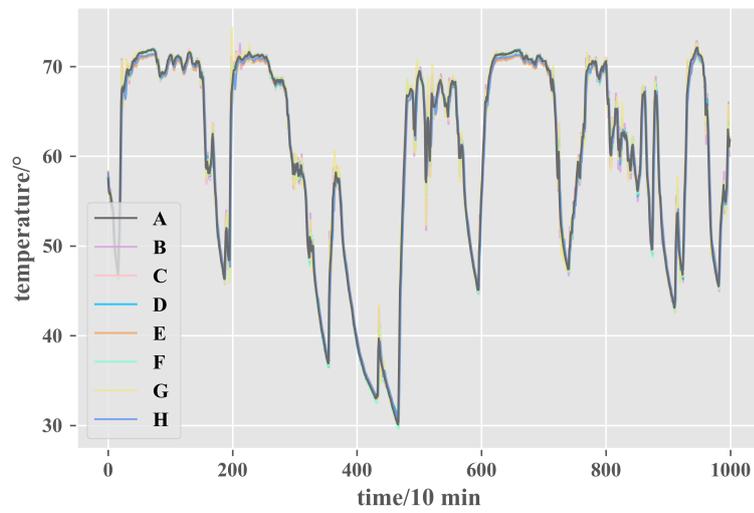


Figure 10. The comparison of different models for gearbox input shaft temperature prediction in #2.

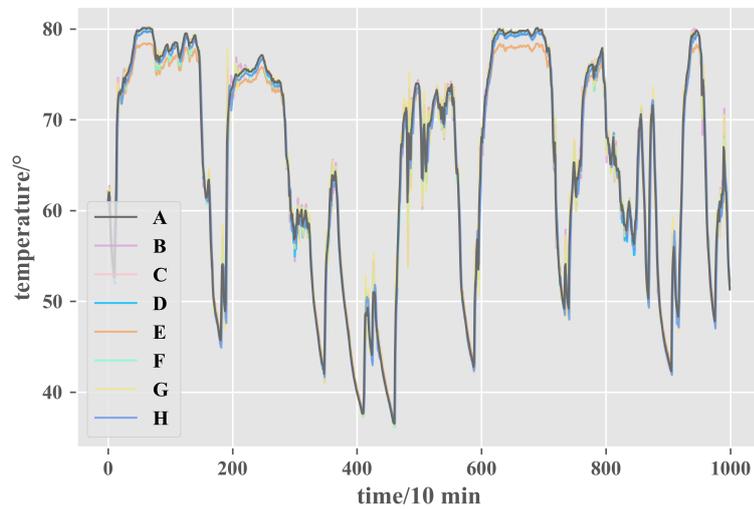


Figure 11. The comparison of different models for gearbox input shaft temperature prediction in #3.

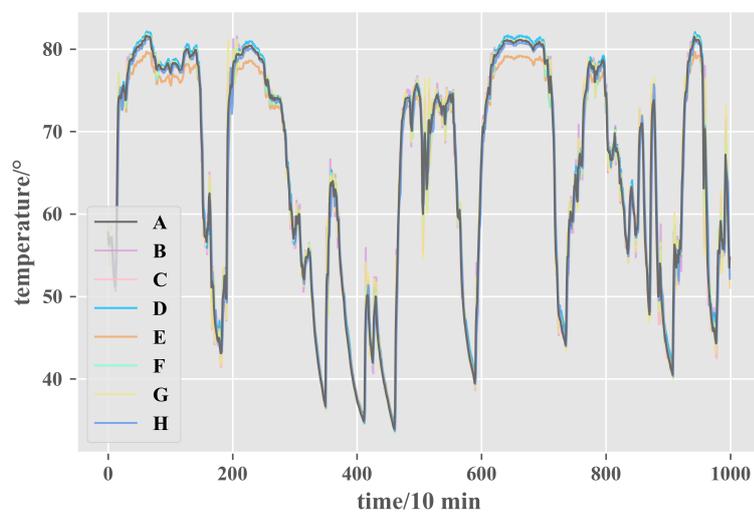
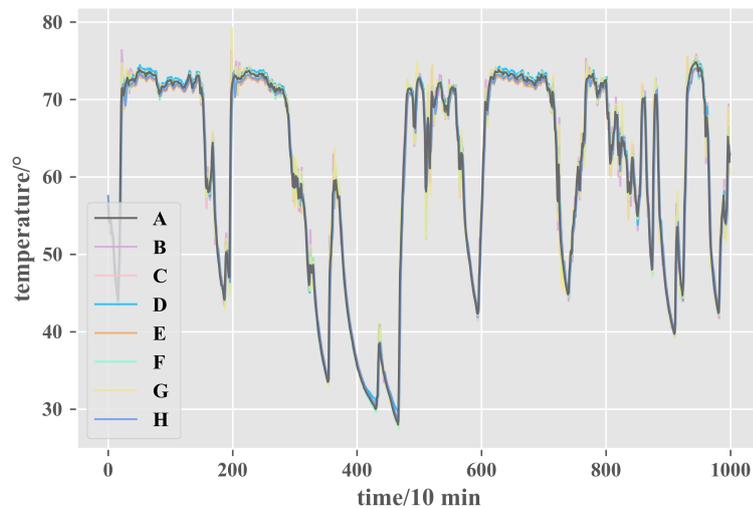
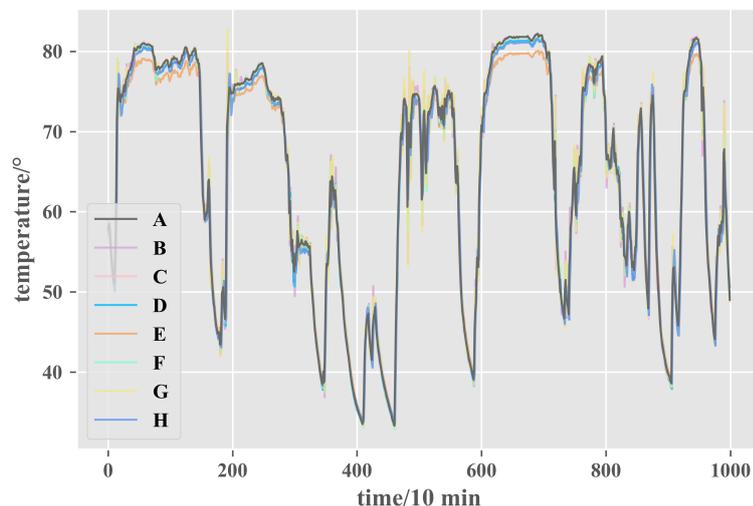


Figure 12. The comparison of different models for gearbox output shaft temperature prediction in #1.



**Figure 13.** The comparison of different models for gearbox output shaft temperature prediction in #2.



**Figure 14.** The comparison of different models for gearbox output shaft temperature prediction in #3.

#### 4. Conclusions

The accuracy of the prediction model directly affects the high-temperature warning performance of the wind turbines gearbox components. In order to achieve higher forecasting accuracy, a novel hybrid model, named the LSTM-AEC, is proposed in the study, which consists of the LSTM preliminary prediction model and adaptive error correction algorithm based on the VMD method. Besides, the dynamic and real-time data decomposition process of the VMD algorithm ensures that the proposed model can be used in the online process. To demonstrate the effectiveness and superiority of the proposed hybrid model, three wind turbine prediction experiments are given in this paper. The prediction models for performance comparison include the hybrid model (LSTM-AEC), BP, ELM, LSTM, ELM-EC, LSTM-EC, and ELM-AEC. Based on the comparative analysis of the prediction performance of different models, the following conclusions can be drawn. (a) By comparing LSTM with ELM and BP algorithms, it can be found that LSTM is superior to other models to some extent; (b) by comparing the two sets of models which contains ELM, ELM-EC, ELM-AEC, LSTM, LSTM-EC and LSTM-AEC, it is found that the adaptive error correction algorithm can optimize the preliminary prediction results to a certain extent; (c) according to the prediction results of three wind turbines, the proposed hybrid model has better performance than other comparative models. Moreover,

the prediction accuracy of the proposed hybrid model in the high-temperature series part is high, which lays a solid foundation for the high-temperature warning of the wind turbines gearbox components.

Although the current research shows that the hybrid model has better prediction performance in temperature prediction of gearbox components, there are still some limitations of the model which need further study. The influence of model parameter initialization results in the fluctuation of prediction performance. Although this fluctuation does not affect the conclusions drawn in this paper, it shows that the hybrid model proposed in this paper has the possibility of further improvement. In addition, the hybrid model proposed in this paper only predicts the temperature of gearbox components in one step, but in practical applications, the multi-step prediction is more greatly needed, which can provide more maintenance time. In future work, the problem of parameter initialization will be studied to further improve the performance and robustness of the prediction model, and the development of multi-step temperature prediction model is needed, which makes the prediction model more practical.

**Author Contributions:** Conceptualization, Q.Z., K.B. and Y.H.; Methodology, Q.Z. and K.B.; Software, Q.Z., J.W. (Jia Wang) and Y.H.; Validation, K.B., J.W. (Jia Wang) and Y.H.; Writing-Original Draft Preparation, Q.Z., K.B., J.W. (Jia Wang) and J.W. (Jinkuan Wang).

**Funding:** This work was supported by the National Key Research and Development Program of China (2016YFB0901900), the Natural Science Foundation of Hebei Province of China (F2017501107) and the Fundamental Research Funds for the Central Universities (N182303037).

**Acknowledgments:** The authors gratefully acknowledge the National Natural Science Foundation of China.

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

## Abbreviations

The following abbreviations are used in this manuscript:

LSTM	Long short term memory neural network
VMD	Variational mode decomposition
LSTM-EC	Combination of long short term memory neural network and Error correction
LSTM-AEC	Combination of long short term memory neural network and Adaptive error correction
SCADA	Supervisory control and data acquisition
PCA	Principal component analysis
DBN	Deep belief network
ARIMA	Autoregressive integrated moving average
ARIMA-ARCH	Combination of autoregressive integrated moving Average and autoregressive conditional heteroskedasticity
ANN	Artificial neural network
SVM	Support vector machine
ELM	Extreme learning machine
WT	Wavelet transform
AR	Autoregressive
WPD	Wavelet packet decomposition
FEEMD	Fast ensemble empirical mode decomposition
WD	Wavelet decomposition
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
CEEMDAN	Complete ensemble empirical mode decomposition
EWT	Empirical wavelet transform
RNN	Recurrent neural network
ADMM	Alternate direction method
IMF	Intrinsic mode function
MAPE	Mean absolute percentage error

MAE	Mean absolute error
MSE	Mean square error
BP	Back Propagation
ELM-EC	Combination of extreme learning machine and error correction
ELM-AEC	Combination of extreme learning machine and adaptive error correction
GA	Genetic algorithm
RBF	Radial basis function
ICEEMDAN	Improved complementary ensemble empirical mode decomposition with adaptive noise

**Variables**

$X_t$	Input vectors of LSTM neural network
$y_t$	Output of LSTM neural network
$i_t$	The output results of LSTM input gate
$f_t$	The output results of LSTM forget gate
$o_t$	The output results of LSTM output gate
$c_t$	The activation state of each cell
$h_t$	The output results of memory unit
$w_{xi}, w_{hi}, w_{xf}, w_{hf}$	The corresponding weight vectors
$w_{xc}, w_{hc}, w_{xo}, w_{ho}$	The corresponding weight vectors
$b_i, b_f, b_c, b_o$	The corresponding bias vectors
$e_i, e_f, e_c, e_o$	Vectors of all 1 corresponding to $b_i, b_f, b_c, b_o$
$[\cdot]^T$	Transposition operation
$\delta(t)$	Dirac distribution
$u_k(t)$	Intrinsic mode function
$\omega_k$	Center frequencies of corresponding modes
$K$	Total number of modal components
$f$	The decomposed original signal
$\eta$	Quadratic multiplication factor
$\lambda(t)$	Lagrangian multipliers
$\ \cdot\ $	The $L^2$ -norm symbol
$n$	The number of iterations
$\hat{(\cdot)}$	The frequency form of the corresponding signal
$sgn$	Sign function
$\rho$	Update coefficient of $\lambda^{n+1}(\omega)$
$d$	Result of error prediction
$m$	Error series after correction
$c$	Error series before correction
$ \cdot $	The absolute value symbol
$g$	Difference between $ m $ and $ c $
$\xi, \alpha, \beta$	Threshold of Algorithm 1
$T_i$	The original temperature series
$U_i$	The error series of the preliminary prediction
$R_{i,j}$	The frequency component decomposed by the VMD algorithm
#1, #2, #3	wind turbine one, wind turbine two and wind turbine three

**Parameters****References**

1. Chen, J.; Zeng, G.Q.; Zhou, W.; Du, W.; Lu, K.D. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Convers. Manag.* **2018**, *165*, 681–695. [[CrossRef](#)]
2. Xiafei, L.; Ping, Y.; Hongxia, G.; Xiwen, W.U. Review of Fault Diagnosis Methods for Large Wind Turbines. *Power Syst. Technol.* **2017**, *41*, 3480–3491. [[CrossRef](#)]
3. Helbing, G.; Ritter, M. Deep Learning for fault detection in wind turbines. *Renew. Sustain. Energy Rev.* **2018**, *98*, 189–198. [[CrossRef](#)]

4. Jiang, G.; He, H.; Yan, J.; Xie, P. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3196–3207. [[CrossRef](#)]
5. Jiang, G.; Xie, P.; He, H.; Yan, J. Wind Turbine Fault Detection Using a Denoising Autoencoder With Temporal Information. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 89–100. [[CrossRef](#)]
6. Odgaard, P.F.; Stoustrup, J. A Benchmark Evaluation of Fault Tolerant Wind Turbine Control Concepts. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 1221–1228. [[CrossRef](#)]
7. Badihi, H.; Zhang, Y.; Hong, H. Wind Turbine Fault Diagnosis and Fault-Tolerant Torque Load Control Against Actuator Faults. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 1351–1372. [[CrossRef](#)]
8. Casau, P.; Rosa, P.; Tabatabaeipour, S.M.; Silvestre, C.; Stoustrup, J. A set-valued approach to FDI and FTC of wind turbines. *IEEE Trans. Control Syst. Technol.* **2014**, *23*, 245–263. [[CrossRef](#)]
9. Tabatabaeipour, S.M.; Odgaard, P.F.; Bak, T.; Stoustrup, J. Fault Detection of Wind Turbines with Uncertain Parameters: A Set-Membership Approach. *Energies* **2012**, *5*, 2424–2448. [[CrossRef](#)]
10. Badihi, H.; Zhang, Y.; Hong, H. Fault-tolerant cooperative control in an offshore wind farm using model-free and model-based fault detection and diagnosis approaches. *Appl. Energy* **2017**, *201*, 284–307. [[CrossRef](#)]
11. Peng, G.; Linlin, L.; Dengchang, M. Wind Turbine Gearbox Condition Monitoring with IPSO-BP. *Acta Energiae Solaris Sin.* **2012**, *33*, 439–445. [[CrossRef](#)]
12. Zhongshan, H.; Ling, T.; Dong, X.; Yaozhong, W. Prediction of oil temperature variations in a wind turbine gearbox based on PCA and an SPC-dynamic neural network hybrid. *J. Tsinghua Univ. Technol.* **2018**, *58*, 539. [[CrossRef](#)]
13. Hongbin, W.; Hong, W.; Qun, H.; Yueling, W.; Zhen, Z. Condition Monitoring Method for Wind Turbine Main Bearings Based on DBN. *China Mech. Eng.* **2018**, *29*, 948–953. [[CrossRef](#)]
14. Mi, X.; Liu, H.; Li, Y. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. *Energy Convers. Manag.* **2019**, *180*, 196–205. [[CrossRef](#)]
15. Masseran, N. Modeling the fluctuations of wind speed data by considering their mean and volatility effects. *Renew. Sustain. Energy Rev.* **2016**, *54*, 777–784. [[CrossRef](#)]
16. Poggi, P.; Muselli, M.; Notton, G.; Cristofari, C.; Louche, A. Forecasting and simulating wind speed in Corsica by using an autoregressive model. *Energy Convers. Manag.* **2003**, *44*, 3177–3196. [[CrossRef](#)]
17. Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320. [[CrossRef](#)]
18. Abdoos, A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **2016**, *203*, 111–120. [[CrossRef](#)]
19. Wang, H.; Wang, G.; Li, G.; Peng, J.; Liu, Y. Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Appl. Energy* **2016**, *182*, 80–93. [[CrossRef](#)]
20. Wang, H.Z.; Li, G.Q.; Wang, G.B.; Peng, J.C.; Jiang, H.; Liu, Y.T. Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl. Energy* **2017**, *188*, 56–70. [[CrossRef](#)]
21. Liu, H.; Mi, X.; Li, Y. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. *Energy Convers. Manag.* **2018**, *166*, 120–131. [[CrossRef](#)]
22. Liu, H.; Tian, H.; Liang, X.F.; Li, Y.F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. *Appl. Energy* **2015**, *157*, 183–194. [[CrossRef](#)]
23. Mi, X.W.; Liu, H.; Li, Y.F. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. *Energy Convers. Manag.* **2017**, *151*, 709–722. [[CrossRef](#)]
24. Naik, J.; Bisoi, R.; Dash, P. Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression. *Renew. Energy* **2018**, *129*, 357–383. [[CrossRef](#)]
25. Meng, A.; Ge, J.; Yin, H.; Chen, S. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Convers. Manag.* **2016**, *114*, 75–88. [[CrossRef](#)]
26. Liu, D.; Wang, J.; Wang, H. Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. *Renew. Energy* **2015**, *78*, 599–608. [[CrossRef](#)]
27. Wang, Y.; Wang, J.; Wei, X. A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: A case study of wind farms in northwest China. *Energy* **2015**, *91*, 556–572. [[CrossRef](#)]

28. Wang, L.; Li, X.; Bai, Y. Short-term wind speed prediction using an extreme learning machine model with error correction. *Energy Convers. Manag.* **2018**, *162*, 239–250. [[CrossRef](#)]
29. Liu, H.; Mi, X.; Li, Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Convers. Manag.* **2018**, *159*, 54–64. [[CrossRef](#)]
30. Liu, H.; Duan, Z.; Han, F.Z.; Li, Y.F. Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. *Energy Convers. Manag.* **2018**, *156*, 525–541. [[CrossRef](#)]
31. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [[CrossRef](#)]
32. Yin, H.; Dong, Z.; Chen, Y.; Ge, J.; Lai, L.L.; Vaccaro, A.; Meng, A. An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. *Energy Convers. Manag.* **2017**, *150*, 108–121. [[CrossRef](#)]
33. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
34. Yan, K.; Wang, X.; Du, Y.; Jin, N.; Huang, H.; Zhou, H. Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy. *Energies* **2018**, *11*, 89. [[CrossRef](#)]
35. Cao, Y.; Gui, L. Multi-Step wind power forecasting model Using LSTM networks, Similar Time Series and Light GBM. In Proceedings of the 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 10–12 November 2018; pp. 192–197. [[CrossRef](#)]
36. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).