Data-driven Anomaly Detection Method Based on Similarities of Multiple Wind Turbines

Xiangjun Zeng, Ming Yang, Chen Feng, Mingqiang Wang, and Lingqin Xia

***Abstract*—The operating conditions of wind turbines (WTs) in the same wind farm (WF) may share similarities due to their shared manufacturing process, control strategy, and operating environment. However, the similarities of WTs are seldom considered in WT anomaly detection, resulting in the disregard of useful information. This paper proposes a method to improve the reliability and accuracy of WT anomaly detection using the supervisory control and data acquisition (SCADA) data of multiple WTs in the same WF. First, a similarity assessment method based on a comparison of different observation time series is proposed, which objectively quantifies the similarities of WT operating conditions. Then, the SCADA data of the target WT and selected WTs that are similar are used to establish several estimation models through a long short-term memory (LSTM) algorithm. LSTM models that exhibit good estimation performance are used to construct a combined estimation model that estimates the variations in the monitored variables of the target WT. Finally, an anomaly detection method that jointly compares the effective value and information entropy of the residuals is proposed to identify anomalies. The effectiveness and accuracy of the proposed method are verified using the data of two actual WFs.**

***Index Terms*—Anomaly detection, information entropy, long short-term memory, similarity assessment, wind farm, wind turbines.**

NOMENCLATURE

β	Comprehensive score of each long short-term memory (LSTM) submodel
γ	Similarity quantified value of two wind turbine (WT) operating conditions
d_i	Number of residual samples within the i^{th} interval
E_{mae}	Average error of estimation result
E_n	Information entropy sample residual
E_{rmse}	Root-mean-square error of estimation result
f_{loss}	Loss function of two-level LSTM algorithm
H_{en}	Threshold of residual information entropy
H_{rm}	Threshold of residual effect value
L	Number of linear segments of time series based on point-by-point segmentation
M_C, M_i	Combined state estimation model (CSEM) and the i^{th} LSTM submodel
m	Number of submodels used to build CSEM
Δm_n	Change in distance of the n^{th} broken line
N	Number of training samples
N_s	Number of statistical intervals of the residual
N_T	Number of samples in each residual statistical interval
n	Number of broken lines in a time series linearization
p_i	Proportion of residual samples in the i^{th} interval
R^2	Determination coefficient
R_m	Effective value of a sample residual
S_0, S_a	Area of a regular polygon and shaded part therein
S_{TS}	Similarity of time series
S_{TS}^i	Similarity assessment value of the i^{th} time series
T_{inv}	Sampling interval of supervisory control and data acquisition (SCADA) system
T_r	Time scale of each detection
t_{n-1}, t_n	Start and end time of the n^{th} broken line
V	Number of variables used for WT similarity assessment
y_i, y_i^*	Real and estimated values of the i^{th} target variable
$\bar{y}$	Average value of target variable
Δy_n	Amplitude change distance of the n^{th} broken line

I. INTRODUCTION

WIND turbine (WT) is a complex electromechanical hybrid system designed to convert wind energy into electrical energy. However, because of challenging operating environments, WTs are prone to failures. These frequent failures not only compromise operational reliability but also lead to increased operation and maintenance costs. Consequently, it is of great significance to explore anomaly detection methods that can adapt to diverse complex conditions and detect abnormal operational states of WTs with precision

Manuscript received: November 22, 2022; revised: March 6, 2023; accepted: June 13, 2023. Date of CrossCheck: June 13, 2023. Date of online publication: July 26, 2023.

X. Zeng and L. Xia are with the College of Electrical Engineering & New Energy, China Three Gorges University, Yichang 443002, China (e-mail: sxdxzxj@ctgu.edu.cn; xialingqin@ctgu.edu.cn).

M. Yang (corresponding author), C. Feng, and M. Wang are with the Key Laboratory of Power System Intelligent Dispatch and Control, Shandong University, Jinan 250061, China (e-mail: myang@sdu.edu.cn; fc5599@163.com; wang0367@sdu.edu.cn).

DOI: 10.35833/MPCE.2022.000769

and at the earliest possible stage.

Condition-monitoring-based anomaly detection methods can identify the abnormal states of WTs by observing various measurements such as vibration, electrical, and temperature signals. Depending on the analytical methods used, these detection methods can be classified into three categories: ① knowledge-based, ② signal-processing-based, and ③ data-driven. Knowledge-based methods rely on expert knowledge or experience. In [1], an offshore WT tower fault detection method based on the fuzzy set theory was proposed. In [2], a WT fault detection method based on expert experience and non-singleton fuzzy-logic inference was proposed. However, these methods often suffer from subjectivity and limited accessibility, as they rely heavily on individual expertise and may lack comprehensive knowledge regarding relevant faults.

Signal-processing-based methods primarily rely on time-frequency analyses such as Fourier and wavelet transforms as well as envelope and power spectrum analyses to extract fault characteristics from the monitored signals and identify any anomalies. In [3], a wavelet energy transmissibility function was proposed for main-bearing fault detection. In [4], a sparse coefficient spectrum analytical method was proposed for bearing fault characteristic order identification. In [5], a method combining mean-shift clustering and short-time Fourier transform was proposed for rotating mechanical fault detection. In [6], a multiscale filter reconstruction method for gearbox fault detection was proposed. Although signal-processing-based methods generally yield reliable detection results, these methods are highly specialized, which means that different monitoring signals may require different signal analytical techniques. In addition, these methods impose higher demands on signal quality and sampling frequency, thus requiring more expensive installations of dedicated sensors.

Data-driven anomaly detection methods estimate the values of state variables through the learning and inference of extensive data. Then, whether the operating status of the components or subsystems of WTs is normal must be determined by comparing actual and estimated values. In general, the data collected by the supervisory control and data acquisition (SCADA) system of a WT can support anomaly detection in many imported components or subsystems. In [7], a sparse Bayesian learning probability estimation model was proposed for detecting anomalies in gearbox oil temperature. Using probability estimation, this model considers the effects of uncertain factors on the detection results. In [8], a multi-fault detection method using an improved triplet-loss algorithm was proposed. This method enhances fault detection accuracy by employing a new and enhanced mapping function. In [9], a deep joint variational autoencoder (JVAE) based method was proposed to detect WT gearbox failures. This method uses a JVAE to reconstruct errors and incorporates a statistical process control chart for gearbox anomaly detection. In [10], a two-layer sparse filtering algorithm was employed to extract robust features from the current signal for WT gearbox diagnosis. In [11], cloud computing based on a hierarchical extreme learning machine algorithm was proposed for gearbox fault detection, which has the advantages of fast computing speed and nearly unlimited storage.

Compared with knowledge-based and signal-processing-based methods, data-driven methods provide greater flexibility and practicality. However, these methods have certain limitations that can affect the reliability and accuracy of anomaly detection results. One of these limitations is the neglect of the temporal dependence of variables when constructing state estimation models. In other words, these methods assume that the current sampling values of the WT state variables such as temperature, pressure, and rotor speed, are independent of their previous sampling values. In fact, the states of these variables are often highly dependent on their previous states and may evolve over time [12]. In addition, current studies seldom pay attention to the quality and quantity of training data, which significantly affect the estimation results. It is generally accepted that training data should be collected during the normal operating period of WTs, and the greater the amount of training data, the better the detection results will be [13]. However, ensuring that the collected training data consistently represent normal operating conditions poses challenges. The latent wear and deterioration of components may have commenced long before. In addition, unforeseen factors such as communication failures, accidental data loss, and regular memory cleaning can render normal operational data unavailable. These factors can result in insufficient or unreliable training data, which will ultimately affect the performance of the estimation model [14]. Addressing these issues is essential in ensuring the accuracy and effectiveness of WT anomaly detection.

A common method for incorporating the temporal dependence of the target variable in the estimation model is to include the previous actual values of the target variable as input [15]. An alternative method that may be more effective utilizes an algorithm with memory functions such as a long short-term memory (LSTM) algorithm. As a deep-learning algorithm, LSTM has excellent feature extraction and nonlinear expression capabilities, which can mine hidden dependencies from the nonstationary and nonlinear state variables of WTs. In addition, the unique control gate structure of LSTM provides inherent advantages in capturing the temporal dependency of time series data. In [16], LSTM-based anomaly detection methods were used for back-to-back converter open-circuit switch fault detection. In [17] and [18], an LSTM algorithm was adopted for WT multiclass fault diagnosis and bearing remaining life prediction. In [19], a convolutional bidirectional LSTM network was designed for equipment health monitoring, which leverages a convolutional neural network to extract local features and uses a bidirectional LSTM to capture long-term dependencies from sequential data. Numerous studies have confirmed that LSTM-based estimation models are highly effective in improving the accuracy of estimation results by mining the temporal dependence of the target variable. This provides an encouraging reference for the proposed method.

Some studies have attempted to use the SCADA data of other WTs in the same WF to alleviate the effects of insufficient training data on anomaly detection. Researchers have observed that WTs in the same WF generally share similar manufacturing processes, control strategies, and operating en-

vironments. Therefore, it is reasonable to assume that similarities exist in terms of their operating conditions. In [20], a method that compared the observations of a target WT with the average observations of all remaining WTs was proposed to realize anomaly detection. Although the proposed method is feasible and straightforward, its reliability is unsatisfactory. This is because some defective WTs may affect the overall average observation, thereby reducing the sensitivity and reliability of anomaly detection [21]. In addition, the operating status of WTs is influenced by multiple factors, and their similarities should be considered from both temporal and spatial perspectives rather than relying solely on the statistical similarities of individual variables.

This paper presents a novel method to enhance the reliability and accuracy of anomaly detection using SCADA data of WTs in the same WF. First, a similarity assessment method for WT operating conditions is proposed based on a time series comparison. This method comprehensively considers the similarities of multiple state variables of WTs and objectively quantifies WT similarities. Then, multiple LSTM models are trained using the SCADA data of different WTs including the target WT, and similar WTs are then identified. Those models showing strong performance are then utilized to establish a combined state estimation model (CSEM) in which the weights are related to the similarities of the WTs. When the estimation results from different LSTM models are integrated at the decision level, the CSEM significantly improves the estimation accuracy of the operating states of WTs. In addition, based on leveraged multiple LSTM submodels trained on historical data from different WTs, the CSEM demonstrates remarkable generalization and robustness in estimating the state variables of the target WT, even when abnormal or missing historical data exist. Finally, an anomaly detection method that does a joint comparison of the effective value and information entropy of the residuals is proposed. This method can reliably differentiate between noisy and anomalous states. The feasibility of the proposed method is verified based on actual SCADA data of two WFs. The contributions of this paper are as follows.

1) A CSEM that leverages the historical data of multiple WTs within the same WF is proposed. The CSEM exhibits excellent robustness and generalization performance and can estimate the state of a target WT accurately and reliably, even in the scenarios involving insufficient or abnormal historical data.

2) A WT operating condition similarity assessment method is proposed based on a time series similarity (TSS) comparison method. The TSS comparison method comprehensively considers the macro- and micro-similarities of the WT operating condition, enabling objective quantification of the similarities among different WTs.

3) An abnormal data detection method that does a joint comparison of the effective value and information entropy of the residuals is improved. This method can reliably distinguish between noise and abnormal signals.

The remainder of this paper is organized as follows. Section II describes the SCADA data used for anomaly detection and introduces the anomaly detection framework. Section III presents the proposed WT similarity assessment method. The CSEM and anomaly detection method are proposed in Section IV. Case studies are presented in Section V, and a conclusion is presented in Section VI.

II. SCADA Data Description and Anomaly Detection Framework

A. SCADA Data Description

The SCADA system of a WT collects a large amount of operational data that can be categorized into three types: environmental monitoring, operating conditions, and behavior indication. Environmental monitoring data include wind speed, wind direction, ambient temperature, and humidity. The data related to operating conditions reflect the operating status of the WT, including various rotor speeds (e.g., wind wheel and generator), temperatures (e. g., bearings, shafts, and lubricating oil), and electrical signals (e.g., output power, voltage, and current). The behavior indication data reflect the actions of different equipment such as relay devices, capacitors, and drive motors. The majority of action indication data are represented in Boolean format, which limits their ability to provide detailed information about the operating conditions of equipment. Therefore, only the first two data types are used in this paper.

B. Anomaly Detection Framework

Figure 1 shows the framework of the proposed anomaly detection method, which consists of three parts.

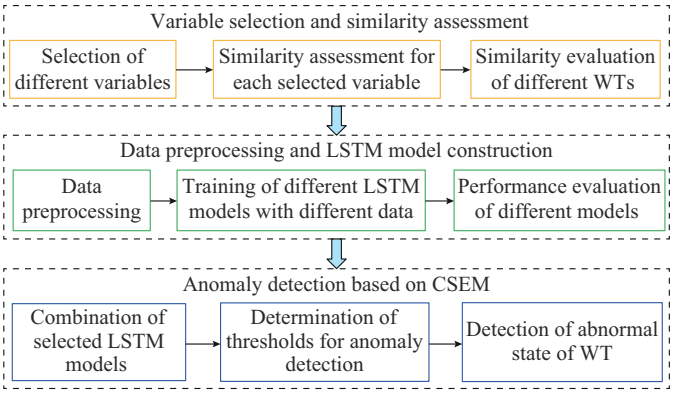


Fig. 1. Framework of proposed anomaly detection method.

The first part selects the state variables and evaluates the similarities between different WTs. The selection of variables must combine expert experience and data correlation analytical results. A similarity assessment of the WT operating conditions is introduced in the subsequent section. The second part preprocesses the SCADA data and trains different LSTM models with data from different WTs. Data preprocessing includes missing-value imputation, noise reduction, and normalization. As these are common data preprocessing methods, this paper does not introduce them in detail. The preprocessed data are divided into training data, validation data, and detection data. The training data and validation data are collected during the normal operating period of the WTs, whereas the detection data may contain fault information. The evaluation of the performance of each LSTM model is presented in this subsection. The third part identi-

fies the abnormal states of the detected object. First, a CSEM is constructed using multiple LSTM models through a weighted combination, and its performance is verified using a validation dataset. An abnormal state is then detected by evaluating two indices, namely the effective value and information entropy of the residuals of the target variable. Their thresholds are then set based on the maximum values obtained during the normal operating period. If the effective value of the residual exceeds its threshold and the information entropy is less than the corresponding threshold, the detection target may be in an abnormal operating state.

III. Proposed WT Similarity Assessment Method

The similarity assessment of individual state variables serves as the foundation for assessing the similarities of WT operating conditions. In this section, the limitations of existing similarity assessment methods are discussed, and a novel TSS comparison method and the corresponding WT similarity assessment method are developed.

A. Defects of Current Similarity Assessment

Some studies rely solely on statistical characteristics to determine the similarities between operating conditions and thus neglect the temporal properties of state variables. Figure 2 shows the main shaft temperature comparison of the two actual WTs for the same period, where the corresponding statistical characteristics of temperature are provided in Table I. It shows that the statistical characteristics of the temperature data of the two WTs are very similar, and the temperature distributions are nearly the same. However, the similarity between the actual main shaft temperature curves is not significant. This is because a similarity assessment based on statistical characteristics does not consider the time synchronization of the state variables. Therefore, when the similarities of the operating conditions of different WTs are evaluated, the synchronization of these variables in a time sequence must be considered.

In addition, some methods have assessed the similarities of different WT operating conditions by comparing only a specific variable [22]. SCADA data of one month from 33 WTs in the same WF were used to illustrate the defects of this type of method. The data distributions for the monitoring of wind speed and gearbox oil temperature for different WTs are presented using boxplots, as shown in Fig. 3. The figure shows that the external wind speeds captured by all WTs are very close, but the gearbox oil temperature distributions are quite different. The *k*-means clustering algorithm was also used to cluster the WTs based on the similarity between the wind speed and gearbox oil temperature. The clustering results are shown in Fig. 3. Classes A, B, and C are three classes obtained based on the mean, mode, and quartile of each variable as clustering conditions. Being classified into the same category indicates that their values are statistically similar. In addition, IQR stands for interquartile range, which is the range between the first quartile (Q1) and the third quartile (Q3) of the data. Classifying WTs based on the similarities of different variables yields varying results. Consequently, evaluating the overall similarity of WT operating conditions cannot rely solely on a specific variable. It is crucial to consider the similarities of multiple associated state variables to assess accurately the similarities between different WTs.

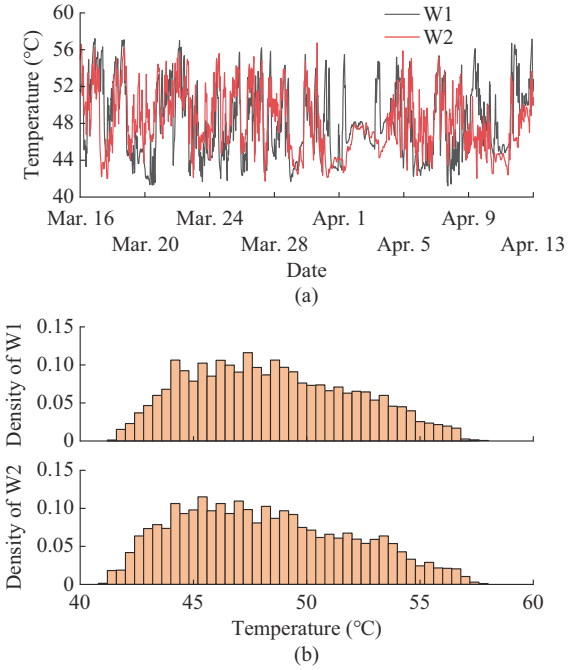


Fig. 2. Main shaft temperature comparison of two actual WTs for the same period. (a) Temperature curve and corresponding statistical characteristics. (b) Temperature distribution histograms.

TABLE I
Corresponding Statistical Characteristics of Temperature

WT	Temperature (°C)					Skewness	Kurtosis
	Max	Min	Mean	Median	Standard		
W1	57.1	41.2	48.3	48.1	3.7	0.24	–0.91
W2	56.7	41.5	48.3	48.1	3.3	0.22	–0.89

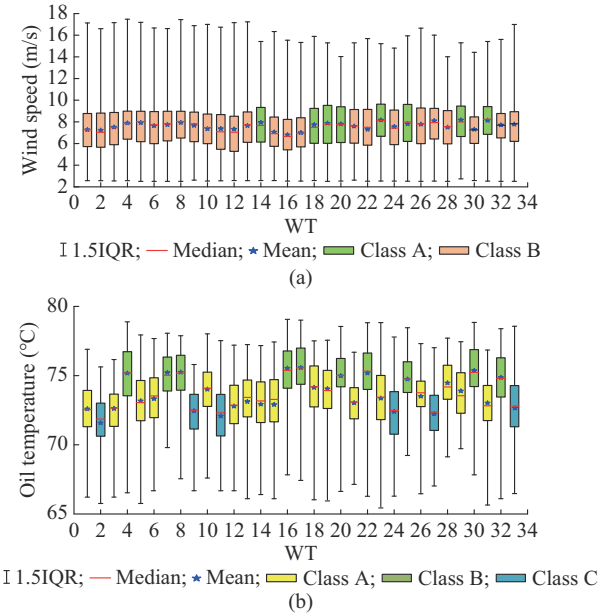


Fig. 3. Data distribution and rank correlation of state variables of different WTs in the same WF. (a) Wind speed. (b) Gearbox oil temperature.

B. Similarity Assessment of Time Series

Based on different measurement principles, the similarity assessment method for time series primarily includes distance and characteristic assessment methods [23]. The commonly used distance assessment methods are the Euclidean and Manhattan distances. The principle of these types of distance assessment method is simple. However, the assessment results are significantly affected by the dimensionality of the variables. The characteristic assessment method realizes similarity assessment by comparing the data features of the time series such as statistical indicators, distribution patterns, and change trends [24]. Although the method is efficient, it does not provide a quantification value for similarity. In addition, both distance and characteristic assessment methods have limitations in terms of capturing the temporal correlations between variables. To address these limitations, comprehensive methods that consider the synchronicity and similarity of time series have been proposed. The two most widely used methods are the dynamic time wrapping (DTW) [25] and shape-based distance (SBD) [26]. These methods offer enhanced accuracy and conciseness in evaluating the temporal correlations between the variables, where the time complexities of DTW and SBD are $O(N^2)$ and $O(N\lg(N))$, respectively. Although these two methods enable an objective assessment of the similarity of the time series, their computational efficiency is low, and considerable time is required to assess the similarity of long time series. Therefore, it is necessary to develop a TSS assessment method with high accuracy and computational efficiency.

To address the shortcomings of the current TSS comparison methods, a new method for assessing the similarities of state variables in WTs is proposed in this paper. Consider two time series curves with equal lengths, denoted by l_1 and l_2 (as shown in Fig. 4), each composed of $L+1$ continuous sampling points. These two curves can be approximated by L broken lines using the point-by-point piecewise linearization method. Clearly, these broken lines have only three changing trends: rising, falling, and remaining. Let 1, −1, and 0 represent these three trends. The states of curves l_1 and l_2 can then be described by sets S_1 and S_2 as:

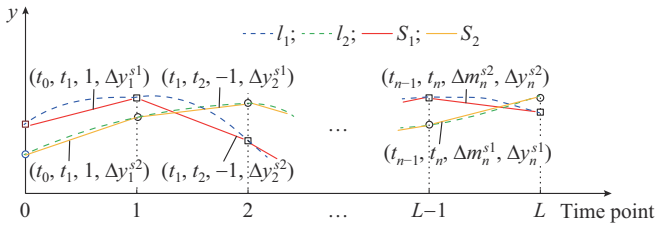


Fig. 4. Piecewise linear representation of time series.

$$\begin{cases} S_1 = \{(t_0, t_1, \Delta m_1^{s1}, \Delta y_1^{s1}),\ (t_1, t_2, \Delta m_2^{s1}, \Delta y_2^{s1}), \ldots, \\ \quad (t_{n-1}, t_n, \Delta m_n^{s1}, \Delta y_n^{s1})\} \\ S_2 = \{(t_0, t_1, \Delta m_1^{s2}, \Delta y_1^{s2}),\ (t_1, t_2, \Delta m_2^{s2}, \Delta y_2^{s2}), \ldots, \\ \quad (t_{n-1}, t_n, \Delta m_n^{s2}, \Delta y_n^{s2})\} \end{cases} \tag{1}$$

where superscripts $s1$ and $s2$ indicate different time series; and $(t_{n-1}, t_n, \Delta m_n^{s1}, \Delta y_n^{s1})$ reflects the state of the n^{th} broken line in S_1. The similarity S_{TS} of S_1 and S_2 can then be evaluated by:

$$S_{TS} = 1 - \frac{1}{2\sqrt{2}\,n} \sum_{i=0}^{n} \sqrt{(\Delta m_i^{s1} - \Delta m_i^{s2})^2 + (\Delta y_i^{s1} - \Delta y_i^{s2})^2} \tag{2}$$

The value range of S_{TS} is [0, 1], where the greater the value of S_{TS}, the greater the similarity. This method considers numerical differences and time synchronicity. It can also quickly and accurately evaluate the similarities between different time series. Through (2), this method comprehensively considers the similarities of the changing trend and changing amplitude distances of all broken lines. The computational complexity of the proposed method is $O(N)$, which is the same as those of the traditional European and Manhattan distance based similarity assessment methods. Thus, this method exhibits a higher computational efficiency than DTW and SBD. However, it should be noted that the proposed method requires that the sampling interval and total data length of the time series be compared for consistency, whereas DTW and SBD do not have such limitations.

C. Similarity Assessment of WTs

As previously discussed, assessing the similarities of WT operating conditions requires the simultaneous consideration of multiple associated variables. Accordingly, a comprehensive similarity assessment method for WT operating conditions is proposed based on the proposed TSS comparison method. The state variables used for the similarity assessment can be divided into two categories of macro- and micro-variables. The macro-variables indirectly reflect the similarities of the WTs in terms of their energy conversion efficiencies. For instance, if two WTs operate normally and have similar input wind speeds but significantly different rotor speeds, the efficiencies of converting wind energy into mechanical energy will differ. Micro-variables are used to compare the similarities in terms of the operating conditions of the detection component or subsystem, which includes multiple monitoring variables that can directly or indirectly reflect the operating conditions of the detection component or subsystem.

After the similarity of each selected variable of different WTs is evaluated, the regular polygonal radar chart shown in Fig. 5 is drawn with the selected variables as axes.

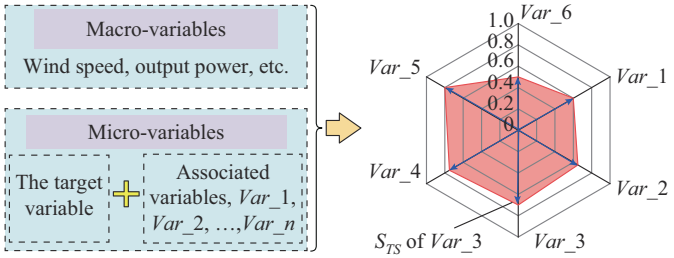


Fig. 5. Operating status similarity assessment of different WTs.

The distance from the center of the radar chart to any vertex is defined to be 1. The blue arrows point from the center to the vertices along different axes. The length of the blue arrow indicates the similarity between the corresponding variables evaluated using the TSS method. Next, connecting the endpoints of these blue arrows to form a closed image (shad-

ed area in the figure), and the percentage of this area in the entire radar chart is used to quantify the similarities of the WT operating conditions. The quantized value γ can be expressed as:

$$\gamma = \frac{S_a}{S_0} = \frac{1}{V}\left(\sum_{i=1}^{V-1} S_{TS}^{i} S_{TS}^{i+1} + S_{TS}^{1} S_{TS}^{n}\right) \quad n \geq 3 \tag{3}$$

where the value of γ is related to the number of selected variables V and similarity S_{ST}^{i} of each variable. If the criterion of γ is set to be 0.64, the mean of S_{ST}^{i} of all selected variables should be approximately 0.8.

IV. CSEM and Anomaly Detection Method

A. LSTM Unit

LSTM is an extension of the recurrent neural network (RNN) and is well suited for time series learning. Figure 6 shows a schematic of the RNN and LSTM. The RNN contains only a single neural layer, whereas the LSTM consists of four interacting layers. The three added layers constitute three special control gate structures: forget gate f_t, input gate i_t, and output gate O_t. In Fig. 6, C is the cell state; h is the hidden information; X is the input information; σ and tanh are the different activation functions; and subscripts t and $t-1$ represent different moments. The forget gate determines whether the information from the previous cell state C_{t-1} should be discarded or retained. The input gate i_t updates the cell state C_t, whereas the output gate outputs the new state C_t and the new hidden state of the cell h_t.

Stacking LSTM layers can improve the nonlinear learning ability without significantly increasing the memory size and learning cost [27]. This paper uses a two-level LSTM structure to build learning models. The construction and performance evaluation of the two-level LSTM are shown in Fig. 7. The purpose of the loss function of the two-level LSTM is to minimize the estimation error. The general loss function can be expressed as:

$$f_{loss} = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^*)^2 \tag{4}$$

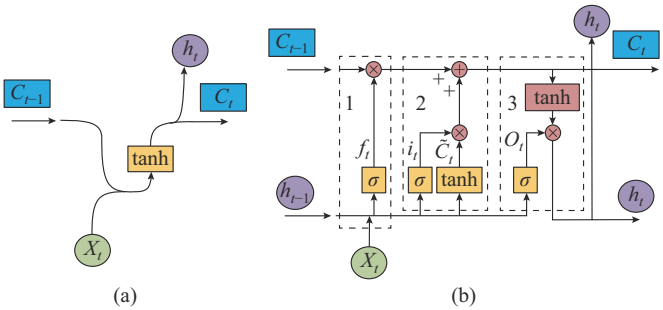


Fig. 6. Schematic of RNN and LSTM. (a) RNN. (b) LSTM.

Historical data of the target WT and selected similar WTs are used to train multiple LSTM models individually. The output of each LSTM model is the estimated value of the target variable for the corresponding WT. The input variables for each LSTM model derive from the corresponding WT, and these input variables are strongly correlated with the target variable. Based on the performance evaluations of all of the LSTM models, those with relatively good performance are used as sub-models to construct the CSEM.

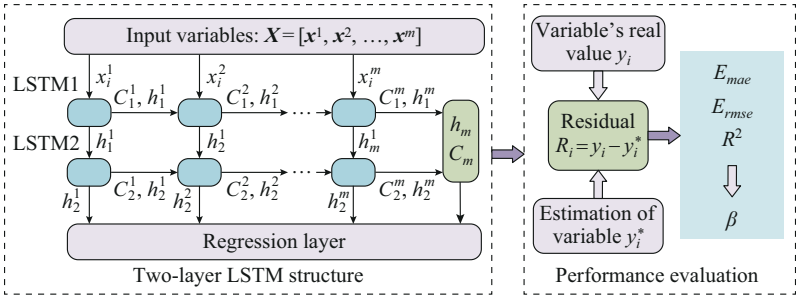


Fig. 7. Construction and performance evaluation of two-level LSTM.

B. CSEM Construction

To ensure the accuracy of CSEM estimation results, sub-models with good performance are selected to construct the CSEM. A comprehensive score β is designed to evaluate the performance of different models and is calculated by:

$$\begin{cases} \beta = R^2/(E_{mae} + E_{rmse}) \\ E_{mae} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{n}\left| y_i - y_i^* \right| \\ E_{rmse} = \sqrt{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(y_i^* - y_i)^2} \\ R^2 = \displaystyle\sum_{i=1}^{n}(y_i - y_i^*)^2 \Big/ \sum_{i=1}^{n}(y_i - \bar{y})^2 \end{cases} \tag{5}$$

where E_{mae} evaluates the average error between observations and expectations; and E_{rmse} measures the degree of dispersion between observations and expectations. The smaller the values of E_{mae} and E_{rmse}, the more accurately R^2 reflects the goodness of fit between the data and model. The range of R^2 is [0, 1], where the larger the value of R^2, the better the result. The value of β is closely related to the value of the aforementioned three indices, and the larger the value of β, the better the comprehensive performance of the model.

The LSTM models in which β has relatively large values are used to construct the CSEM by:

$$M_C = \frac{\sum_{i=0}^{m}\gamma_i M_i}{\sum_{i=0}^{m}\gamma_i} \tag{6}$$

Here, the contribution of each submodel to the CSEM is proportional to the similarity between the operating conditions of the selected and target WTs. In fact, CSEM is similar to an ensemble learning model, and its output is the weighted sum of the estimated results for each sub-model. When multiple LSTM models with good performance are integrated, the CSEM exhibits the advantages of diverse LSTM models, resulting in superior generalization and accuracy performance compared with the LSTM model trained solely on individual WT historical data. M_0 represents the LSTM model trained by the SCADA data of the target WT, and the corresponding $\gamma_0 = 1$. If significant data gaps or extensive abnormal data exist in the historical data of target WT, obtaining a reliable M_0 is impossible. The value of γ_0 is set to be 0, and thereafter, M_0 does not participate in the construction of the CSEM.

C. Anomaly Detection

An anomaly in the target variable can be determined by analyzing the residual, i.e., the difference between the actual and estimated values of target variables. An anomaly in the target variable considerably changes the value of the residual and the reverse information entropy of the residual sequence. Therefore, an anomaly detection method that does a joint comparison of the effective value and information entropy of the residuals is proposed. Two indices have been used to determine whether the residuals are abnormal. The first index R_m measures the overall degree of deviation in the residual sample. The greater the value of R_m, the higher the probability that the detection sample is abnormal. The other index E_n is used to evaluate the degree of uncertainty in which the detection sample is abnormal [28]. The lower the value of E_n, the greater the certainty that the detection sample is abnormal. To improve the detection accuracy, the proposed method divides the long time series of the target variable into many short periods for anomaly detection using sliding window sampling. The values of R_m and E_n for each short detection period can be calculated by:

$$\begin{cases} R_m = \sqrt{\dfrac{1}{N_T}\sum_{i=1}^{N_T}(y_i - y_i^*)^2} \\ E_n = -\sum_{i=1}^{N_s} p_i \ln p_i \\ p_i = \dfrac{d_i T_{inv}}{T_r} \quad T_r \geq T_{inv} \end{cases} \tag{7}$$

In this paper, three statistical intervals are set to calculate the information entropy for each detection period. These statistical intervals are determined using the boxplot of all the target variable residuals obtained based on the validation dataset, as shown in Fig. 8(a). The figure shows that the three statistical intervals are the normal, risk, and high-risk intervals, which are indicated by different colors. Four key values of the boxplot, given as Q_{low}, Q_1, Q_3, and Q_{up}, are used to divide the intervals, where Q_1 and Q_3 are the 0.25 and 0.75 quartiles of the residual sequence, respectively, and Q_{up} and Q_{low} are the upper and lower bounds of the boxplot, respectively, which can be calculated by:

$$\begin{cases} Q_{up} = Q_3 + 1.5(Q_3 - Q_1) \\ Q_{low} = Q_1 - 1.5(Q_3 - Q_1) \end{cases} \tag{8}$$

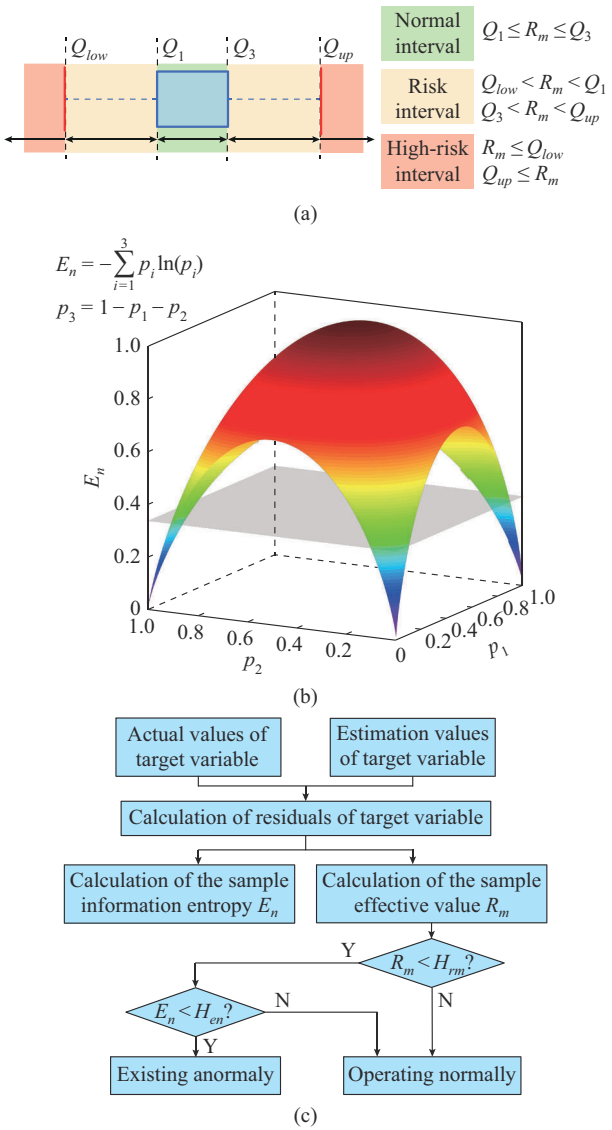


Fig. 8. Anomaly detection method. (a) Segment of detection sample. (b) Value analysis of information entropy. (c) Anomaly detection flow.

Figure 8(b) shows the relationship between the proportion p_i $(i = 1, 2, 3)$ contained in the three intervals and the corresponding E_n. The figure shows that when a large proportion of residuals fall in the same area, E_n is small, reflecting the high probability of the occurrence of an event. Figure 8(c) shows the anomaly detection flow, which includes two judgment processes. The first process determines whether R_m of the detection period exceeds the given threshold H_{rm}. If it does not exceed it, no anomaly occurs. Otherwise, another judgment based on E_n is required to determine whether R_m exceeding H_{rm} is caused by sporadic noise or abnormal operation. If E_n is less than the given threshold H_{en}, R_m exceeding H_{rm} is likely caused by an abnormal operation. The values of H_{rm} and H_{en} are obtained based on the normal operation data. First, the residuals of the target variable are calculated based on the validation data. Next, the target variable time series

are divided into different detection segments using sliding-window sampling. The values of R_m and E_n of each detection segment are then calculated. Finally, the maximum values of R_m and E_n in all detection segments are taken as the values of H_{rm} and H_{en}, respectively.

V. Case Studies

Case studies were conducted based on the SCADA data of two actual onshore WFs, labeled as WF1 and WF2, to verify the effectiveness of the proposed method. WF1 is located in the hilly area of the Shandong Peninsula, China, and WF2 is located in the Gobi of Gansu Province, China. Each WF contains 33 WTs, labeled as WT0-WT32, where WT0 is the target WT. Table II lists the data information of different WFs. Data may have been deleted under certain conditions such as during shutdowns, maintenance periods, or periods of very low wind speeds (i.e., less than the cut-in wind speed). One important consideration is that if the data from a specific period of one WT are deleted, the data from other WTs during the same period will also be removed. T_r was set to be 1 hour. In this paper, Python 3.8.3 programming language and Pytorch 1.7.1 architecture were used to build all the models. The computer was configured with an Intel i9-9900KF CPU with 32 GB of RAM and a NVIDIA GeForce RTX 2060 GPU.

TABLE II
Data Information of Different WFs

WF	Start and end dates	Sampling interval (min)	Fault alarm time of WT0	Fault alarm information
WF1	November 1, 2018-June 15, 2019	1	12:40, June 12, 2019	Gearbox front-bearing temperature abnormal
WF2	February 1, 2018-April 15, 2018	5	21:20, April 9, 2018	Gearbox oil temperature over-limited

A. Case 1

Case 1 was conducted based on the data of WF1, where those from November 1, 2018 to April 17, 2019 were used as the training data. These data were also used to evaluate the similarities of different WTs. The data from April 18, 2019 to May 15, 2019 were used as validation data to test the combined estimation model. The data from May 16, 2019 to June 15, 2019 were used to evaluate the anomaly detection method. The number of training data points was 246240, 224273 of which were available for similarity assessment and model training after screening. Eight variables were selected to evaluate the similarities among the WTs, including three macro-variables and five micro-variables. The three macro-variables were wind speed, main shaft rotation speed, and output power. They can reflect the energy conversion efficiencies of WTs from wind energy to mechanical energy and from mechanical energy to electrical energy. The five micro-variables were the target variable gearbox front-bearing temperature and its four closely related variables.

Figure 9 shows the importance coefficients between gearbox front-bearing temperature and different variables, obtained using a random forest algorithm. Based on the results shown in Fig. 9, the other four selected micro-variables were the gearbox oil temperature (Gearbox_oil_T), gearbox rear-bearing temperature (Gearbox_rear_bearing_T), generator front-bearing temperature (Gen_front_bearing_T), and nacelle temperature (Nacelle_T). The wind speed was not selected and instead used as a macro-variable.

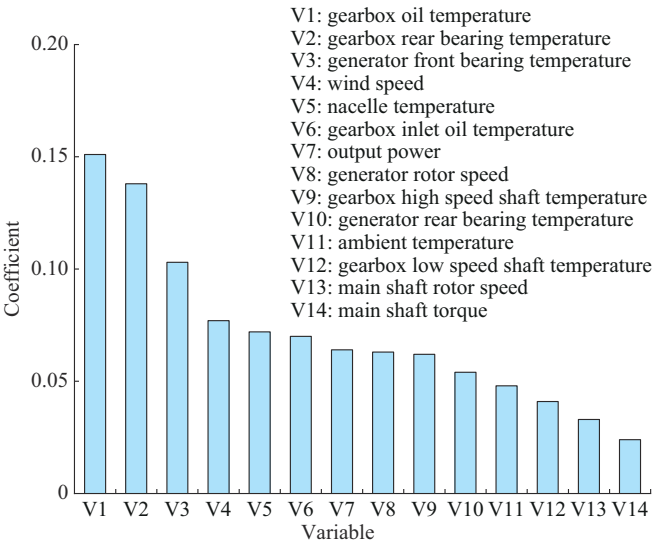


Fig. 9. Importance coefficients between gearbox front-bearing temperature and different variables.

Following the selection of variables, the similarity between each WT in WF1 and WT0 could be quantified according to the proposed similarity assessment method, with the results shown in Fig. 10. Five WTs, i.e., WT6, WT11, WT17, WT23, and WT32 (in orange color), with relatively high similarities to WT0 were preliminarily selected to provide their SCADA data for sub-model training, where their similarity quantization values are listed in Table III. Figure 11 shows a similarity radar chart between WT0 and the five WTs in WF1. The figure shows that the similarity between each variable of the selected WTs and the corresponding variable of WT0 exceedes 0.8.

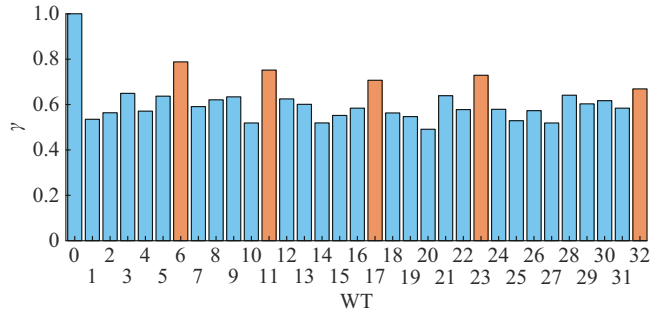


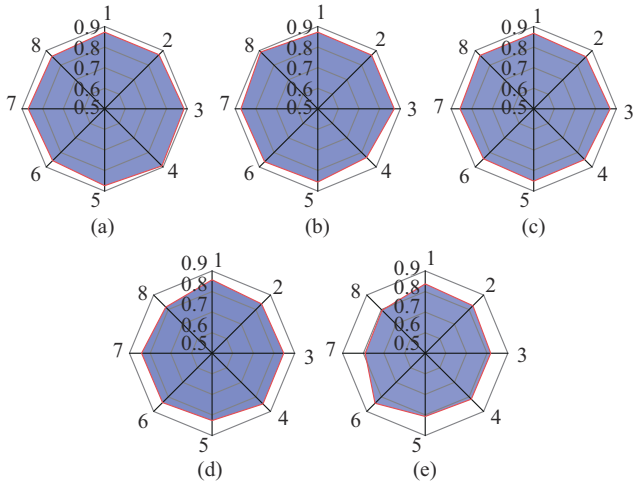
Fig. 10. Similarity assessment results between WT0 and other WTs in WF1.

The SCADA data of WT0 and five similar WTs were used to train six LSTM models, respectively. The output variable for each model is the expected gearbox front-bearing temperature of the corresponding WT. The input vari-

ables for each model include the gearbox oil temperature, gearbox rear-bearing temperature, generator front-bearing temperature, wind speed, and nacelle temperature of the corresponding WT. The trained models were tested to assess their suitability for constructing a CSEM. The performance evaluation of these models, trained using the SCADA data of similar WTs, consists of two parts: testing the estimation accuracy of the model using its own validation data, and testing the adaptability of the model using the validation data of WT0. Table III lists the performance assessment results of the different LSTM models in WF1. These results demonstrates that all of the models exhibited favorable estimation accuracy when being tested using their respective validation data. However, when the models trained with the data from similar WTs were tested using the validation data of WT0, the comprehensive score β of all the models decreased. This indicates that the different models exhibited different adaptabilities to the WT0 SCADA data. Notably, the sub-models corresponding to the WT with high similarity to WT0 exhibited good adaptability. In this paper, the model trained according to the data of WT0 and the three models with relatively good performances in both tests (i.e., models trained using the data of WT6, WT11, and WT23) were selected to construct the CSEM.

TABLE III
PERFORMANCE ASSESSMENT RESULTS OF DIFFERENT LSTM MODELS IN WF1

WT	Γ	Accuracy verification				Adaptability verification			
		E_{mae}	E_{rmse}	R^2	β	E_{mae}	E_{rmse}	R^2	β
WT0	1.00	0.319	0.377	0.967	1.389				
WT6	0.79	0.325	0.382	0.969	1.371	0.458	0.461	0.937	1.020
WT11	0.75	0.271	0.357	0.981	1.562	0.428	0.552	0.925	0.944
WT23	0.73	0.289	0.375	0.972	1.464	0.477	0.536	0.922	0.910
WT17	0.71	0.333	0.392	0.979	1.350	0.527	0.632	0.899	0.776
WT32	0.67	0.314	0.386	0.969	1.384	0.614	0.731	0.871	0.648



1: output power; 2: wind speed; 3: rotor speed; 4: oil temperature
5: generator front bearing temperature; 6: nacelle temperature
7: gearbox rear bearing temperature; 8: gearbox front bearing temperature

Fig. 11. Similarity radar chart between WT0 and five selected WTs in WF1.

Next, three CSEM combination schemes were designed to verify the feasibility of the proposed method, as listed in Table IV. The performance of each CSEM in the three schemes was assessed based on the validation data of the WT0. Figure 12 shows the bearing temperature estimation results of WT0 under normal operation according to different schemes. Table V lists the model performance evaluation indices and the corresponding anomaly detection thresholds of the different schemes. The temperature curve estimated using the CSEM of Sch3 is the closest to the actual temperature curve. The performance evaluation indices of the CSEM in Sch3 are also better than those of the other two CSEMs, indicating that using the data of multiple similar WTs could improve the accuracy of the estimation results. In addition, the performance indices of the CSEM in Sch1 and Sch2 were very close, confirming that using only the data of similar WTs could also be used to construct an effective CSEM for estimating the state variables of WT0.

TABLE IV
DIFFERENT COMBINATION SCHEMES

Scheme	State estimation model
Sch1	CSEM is the LSTM model trained using only the data of WT0
Sch2	CSEM is combined with LSTM models and trained based on the data of selected similar WTs, excluding the LSTM trained with the data of WT0
Sch3	CSEM is combined with multiple LSTM models and trained with the data of WT0 and those of all selected similar WTs

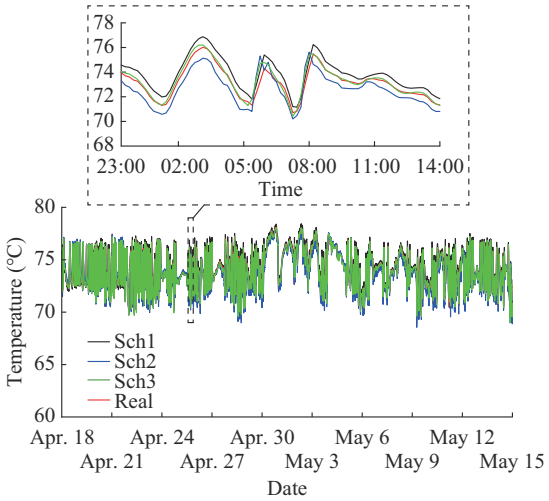


Fig. 12. Bearing temperature estimation results of WT0 under normal operation according to different schemes.

TABLE V
PERFORMANCE EVALUATION RESULTS OF DIFFERENT SCHEMES

Scheme	E_{mae}	E_{rmse}	R^2	β	H_{rm}	H_{en}
Sch1	0.319	0.377	0.967	1.389	0.865	0.328
Sch2	0.331	0.408	0.958	1.296	0.937	0.351
Sch3	0.291	0.323	0.978	1.593	0.543	0.296

Following the performance test, the anomaly detection effectiveness of the three schemes was evaluated using the de-

tection data of WT0. Figure 13(a) compares the actual and estimated gearbox front-bearing temperatures. Figure 13(b)-(d) show the anomaly detection results of schemes Sch1, Sch2, and Sch3, respectively.

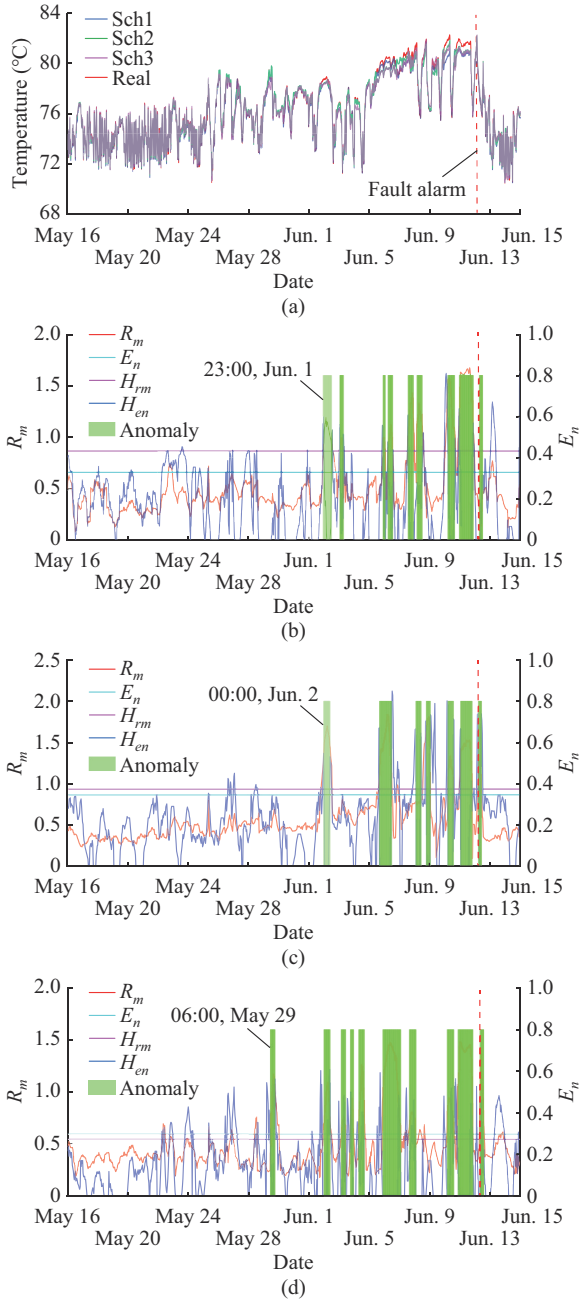


Fig. 13. Anomaly detection results of WT0 in WF1 by different schemes. (a) Comparison of estimation curves of bearing temperature based on different schemes and real curves. (b) Anomaly detection results of Sch1. (c) Anomaly detection results of Sch2. (d) Anomaly detection results of Sch3.

Each subfigure clearly shows that the recognized anomaly periods are consistent with the observed abnormal periods. This consistency verified the reliability of the proposed anomaly detection method. It should also be noted that Sch3 exhibited the earliest detection of abnormal temperatures, with the earliest time at 06:00 on May 29, 2019. This detection preceded the system fault alarm for 14 days, highlighting the early warning capability of Sch3. This result indicated that comprehensive utilization of the SCADA data of WT0 and other similar WTs could effectively improve the sensitivity of anomaly detection, and this is beneficial for early discovery of anomalies. In addition, the earliest time for Sch1 and Sch2 for determining an abnormal temperature is nearly the same. These schemes could detect abnormal temperatures 10 days prior to system failure, indicating that using the SCADA data of similar WTs to detect anomalies in WT0 was feasible.

B. Case 2

The data used in Case 2 were obtained from WF2. As the amount of available data is relatively small, using the SCADA data of similar WTs to improve the reliability of anomaly detection is more practical. The training, validation, and detection data were from the periods of February 1, 2018 to March 4, 2018, March 5, 2018 to March 23, 2018, and March 24, 2018 to April 14, 2018, respectively. In this case, the data in the total period of February 1, 2018 to March 23, 2018 were used to evaluate the similarities among the different WTs because the condition monitoring data of WTs in WF2 are insufficient. Of the total data available during this period, which amounted to 14688 records, only 12727 records passed the data screening process and are suitable for similarity assessment. The first 8436 records were used for model training. Similar to Case 1, eight variables were selected to evaluate the similarities under operating conditions of the WTs. Three macro-variables are wind speed, rotation speed, and output power. Five micro-variables considered included the target variable of the gearbox oil temperature and four other closely related variables. After the importance coefficients of the different variables to the gearbox oil temperature were compared, we selected the gearbox rear-bearing temperature, gearbox high-speed shaft temperature, gearbox inlet oil temperature, and gearbox front-bearing temperature as the remaining four micro-variables. The model output was the gearbox oil temperature during the subsequent training of the LSTM model. The model input included the four selected micro-variables as well as wind speed and nacelle temperature.

Figure 14 shows the similarity assessment results between WT0 and the other WTs in WF2. It shows that the operating conditions of WT1, WT3, WT5, WT8, and WT15 are relatively similar to WT0, where their similarity quantification values were obtained, as shown in Table VI. Based on a comparison of the performance evaluation index of each model under the estimation accuracy and adaptability test, the LSTM models trained by the data of WT0 and the data of WT1, WT5, and WT8 were selected to construct the CSEM.

The performances of the three CSEMs under the different schemes were tested using the validation data of WT0. Table VII lists the performance indices and anomaly detection thresholds of the different schemes. Figure 15 shows the actual and estimated gearbox oil temperature curves for the different schemes. A comparison of the performance evaluation indices and estimation curves revealed that the CSEM in Sch3 had the highest performance. The performances of the

CSEMs under Sch1 and Sch2 are very similar. The results also indicated that using data from similar WTs was beneficial in improving the estimation accuracy of the CSEM. When the SCADA data of multiple similar WTs were used, a CSEM could be constructed to estimate the state variables of WT0.

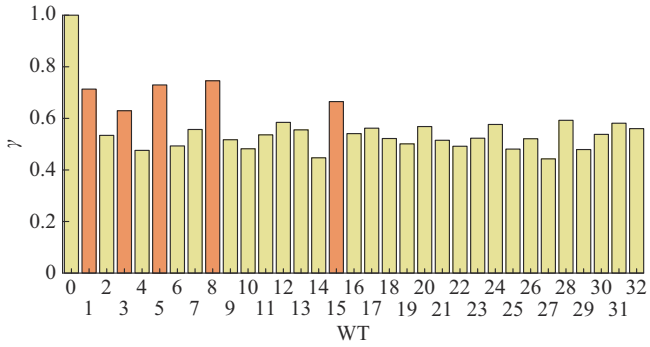


Fig. 14. Similarity assessment results between WT0 and other WTs in WF2.

TABLE VI
PERFORMANCE EVALUATION RESULTS OF DIFFERENT LSTM MODELS IN WF2

WT	γ	Accuracy verification				Adaptability verification			
		E_{mae}	E_{rmse}	R^2	β	E_{mae}	E_{rmse}	R^2	β
WT0	1.000	0.409	0.451	0.965	1.122				
WT8	0.750	0.422	0.461	0.964	1.092	0.572	0.695	0.936	0.739
WT5	0.729	0.411	0.457	0.965	1.112	0.584	0.653	0.925	0.748
WT1	0.692	0.379	0.441	0.971	1.184	0.616	0.659	0.919	0.721
WT15	0.650	0.431	0.484	0.958	1.047	0.672	0.721	0.907	0.651
WT3	0.638	0.383	0.438	0.973	1.185	0.813	0.857	0.875	0.524

TABLE VII
PERFORMANCE EVALUATION RESULTS OF DIFFERENT SCHEMES

Scheme	E_{mae}	E_{rmse}	R^2	β	H_{rm}	H_{en}
Sch1	0.409	0.451	0.961	1.117	0.827	0.393
Sch2	0.418	0.463	0.957	1.086	0.873	0.401
Sch3	0.364	0.396	0.970	1.276	0.741	0.384

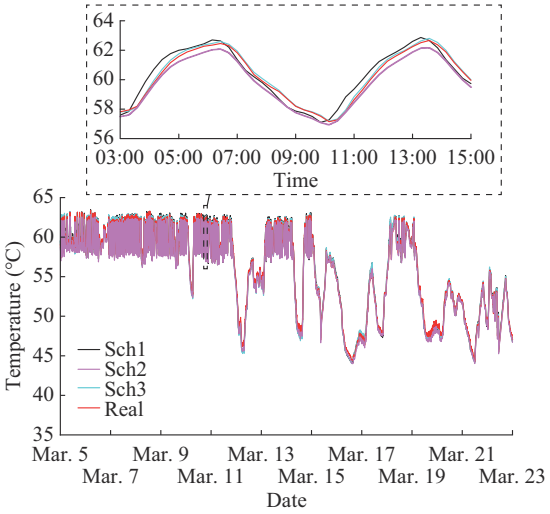


Fig. 15. Actual and estimated gearbox oil temperature curves for different schemes.

Figure 16 shows the anomaly detection results for WT0 in WF2 by different schemes. A comparison of the anomaly detection results of different schemes revealed that Sch3 is still the earliest scheme to discover the abnormal gearbox oil temperature, and the earliest time is 22:00 on March 30, 2018, 10 days ahead of the SCADA system fault alarm. In addition, the earliest time for Sch1 and Sch2 to detect the abnormal state of the oil temperature is 05:00 on April 1, 2018 and 03:00 on April 1, 2018, respectively. The anomaly detection results also proved the feasibility and effectiveness of the proposed anomaly detection method.

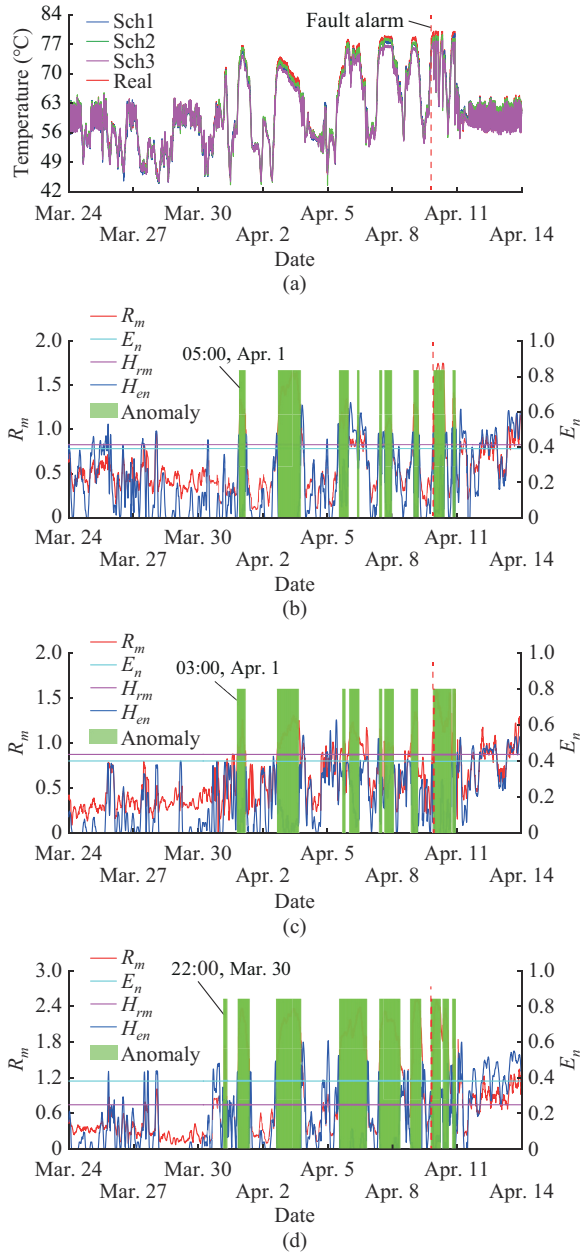


Fig. 16. Anomaly detection results of WT0 in WF2 by different schemes. (a) Comparison of estimation curves of gearbox oil temperature based on different schemes and real curves. (b) Anomaly detection results of Sch1. (c) Anomaly detection results of Sch2. (d) Anomaly detection results of Sch3.

Under the combined results of Cases 1 and 2, we found that the amount of data used to train the model significantly

affected its performance. The more training data the model used, the better the estimation accuracy. When the historical data of the target WT are insufficient, a CSEM could be constructed using data from similar WTs to improve the accuracy of the model estimation results. Also, it is noteworthy that the anomaly detection instances in both cases aligned with significant deviations between the estimated and actual values. This confirmed the efficacy of utilizing the effective value and information entropy of the residuals for precise anomaly identification.

C. Comparative Verification

1) Comparison of TSS Method with Other Methods

Two TSS measurement methods, namely, DTW [25] and SBD [26], were adopted for comparison to verify the effectiveness of the proposed TSS method. The wind speed and gearbox front-bearing temperature data of all WTs in WF1 from November 1, 2018 to April 17, 2019 were used for similarity comparison. Histograms and colored lines are used in Fig. 17 to show the similarity assessment results. The figure also shows the analytical time costs of the different methods. For comparison, the assessment results of the different methods were scaled to [0, 1]. The closer the assessment value was to be 0, the stronger the similarities of different time series.

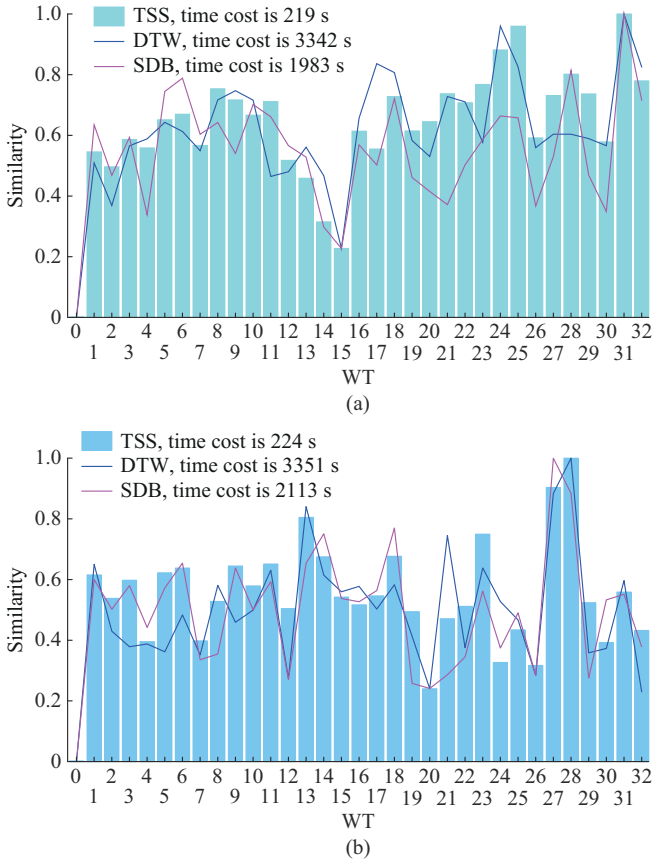


Fig. 17. Similarity assessment results of different methods. (a) Wind speed. (b) Bearing temperature.

As each similarity assessment method uses different calculation techniques, the method outcomes are also different, even when the similarities of the same time series are measured. As no unified quantitative standard exists for the similarity assessment of time series, accurately comparing the proposed method and DTW or SDB is challenging. However, based on the trend of the histogram and two line charts shown in Fig. 17, the assessment outcomes obtained by the three methods exhibited a consistent pattern despite different objectives. This demonstrated the effectiveness of the proposed TSS method. A comparison of time costs revealed that the TSS method significantly outperformed the SBD and DTW methods in terms of efficiency, which was consistent with theoretical analysis based on time complexity.

2) Anomaly Detection Comparison with Other Methods

Four data-driven WT anomaly detection methods were used for validation and comparison. The estimation models of these methods were constructed based on different algorithms, including the deep neural network (DNN) [29], gated recurrent unite (GRU) [30], convolutional neural network-LSTM (CNN-LSTM) [31], and multi-model combined (MMC) estimation [32]. All four benchmarking models used only the historical SCADA data of WT0 to train and test the corresponding estimation model. It should be noted that the MMC differs from the proposed CSEM, where the MMC was constructed based on different shallow learning sub-models, and each sub-model was trained using the same historical SCADA data from the WT0. In addition, the combined weights of the MMC were obtained using an optimization algorithm. The objective function minimizes the sum of the squares of the estimated residuals of all sub-models. To incorporate the temporal dependence of the target variable into the aforementioned benchmarking methods, the actual value of the target variable at the previous moment was included as model input at the current moment.

The performances of the four benchmark models and the CSEM of Sch3 were compared based on the data from the aforementioned two cases. Figure 18 presents the performance evaluation results of the different models when the validation data were used. It should be noted that the comprehensive scores β of all models shown in the figure are the normalized values, and the normalized cardinality is the maximum β of the five models.

Figure 18 shows that the performance of the CSEM model was better than those of the other models. The differences in performance among the different models may be related to the total amount of training data. When the training data of Case 1 are sufficient, the performance difference among the different models is relatively small, as shown in Fig. 18(a). However, for relatively insufficient training data for Case 2, the performances of all models except CSEM exhibited a significant decline, as shown in Fig. 18(b). The CSEM could maintain a relatively better performance because the construction of CSEM used the training data of multiple similar WTs, ensuring sufficient data for model training.

Next, different degrees of anomaly detection effectiveness of the different methods were compared based on the detection dataset of the two cases. For comparison, all methods adopted the abnormal recognition method proposed in this paper. Figure 19 shows the comparison of anomaly detection results of different methods.

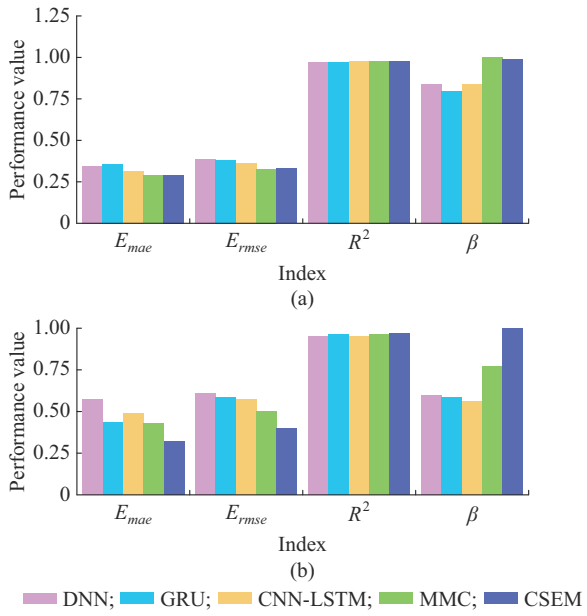
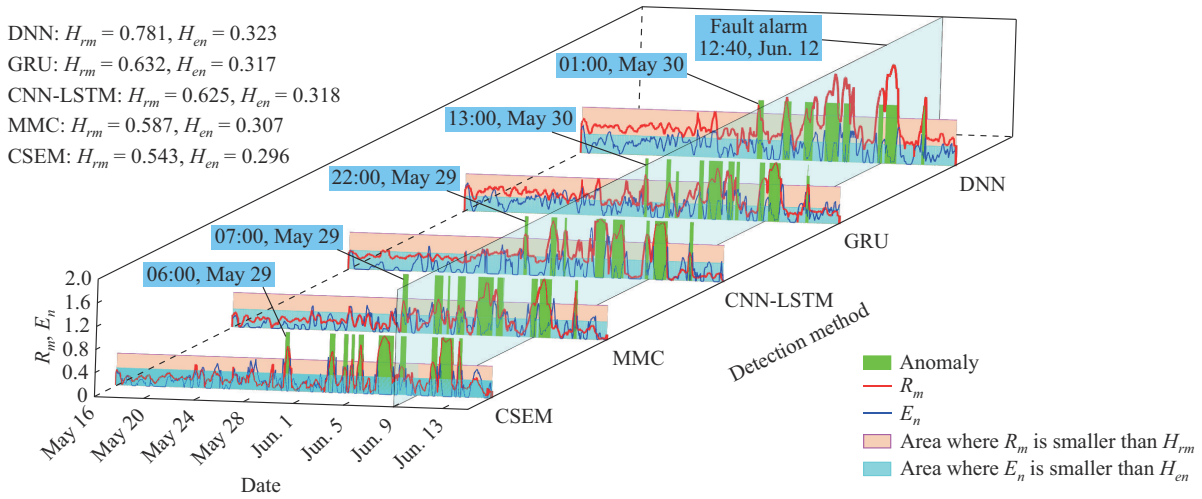


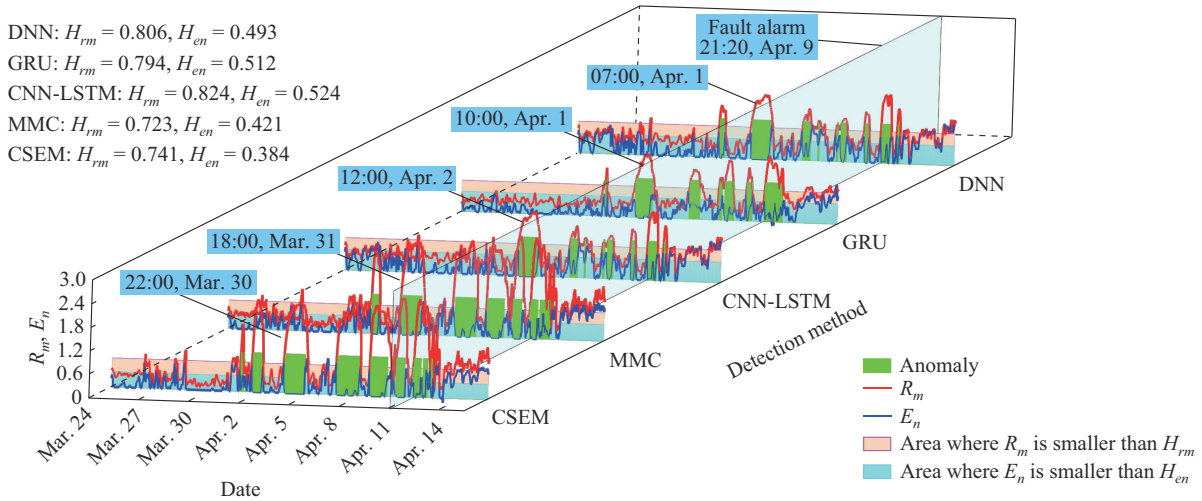
Fig. 18. Performance evaluation results of different models when validation data were used. (a) Case 1. (b) Case 2.

When the training data are sufficient, the detection results of CSEM and MMC are similar. Both CSEM and MMC could detect the abnormal temperature of the gearbox front-bearing temperature 14 days earlier than the system fault alarm. In addition, these two combined-model-based methods detected anomalies significantly earlier than the other single-model-based methods. However, when the training data of WT0 are insufficient for Case 2, the anomaly detection results of the different models as shown in Fig. 19(b) were different, and the proposed method could detect the abnormal state of the gearbox oil temperature earlier than the other methods. These comparison results revealed that the proposed method are more sensitive to anomalies, and the performance of the CSEM could be guaranteed even when the training data of the WT are insufficient.

The anomaly detection reliabilities of the different methods were compared when the training data contained abnormal data. The anomaly detection results shown in Fig. 19 reveal that the target variables of the two failed WTs were detected abnormally many times prior to the SCADA system fault alarm.



(a)



(b)

Fig. 19. Comparison of anomaly detection results of different methods. (a) Case 1. (b) Case 2.

Therefore, this contrast experiment used a portion of the detection data containing abnormal data to train the estimation models. Since the amount of detection data for Case 2 is small, this contrast experiment used only the detection data for Case 1. As new training data, 12774 records from May 25, 2019 to June 5, 2019 were used. Based on the anomaly detection results of Sch3 for Case 1, the abnormal data accounted for 26.4% of the new training data. In addition, the data from June 6, 2019 to June 15, 2019 were used as new detection data. The four comparison methods used only the historical data of WT0 to train their estimation models, whereas the proposed method used both the historical data of WT0 and three selected similar WTs in the same period to construct the CESM. The input and output variables for each model are the same as those for Case 1. In this contrast experiment, no validation data were used, and the anomaly detection thresholds H_{rm} and H_{en} were determined based on the new training data.

Figure 20 shows the anomaly detection results of the different methods. The figure shows that under abnormal data in the training data, the anomaly detection results of the different methods were different. All methods could detect the abnormal temperature of the gearbox front bearing prior to the SCADA system fault alarm. CSEM is the first to discover abnormal temperatures and could also detect multiple temperature anomalies prior to the system fault alarm. In addition, a comparison of the anomaly detection results of the different methods, as shown in Fig. 19(a) and Fig. 20(b), revealed that the anomaly detection results of the CSEM are roughly the same after June 6. However, the anomaly detection results of the DNN, GRU, CNN-LSTM, and MMC were significantly different after June 6. The GRU, CNN-LSTM, and MMC identified the normal data as abnormal, whereas the DNN did not detect the abnormal temperature at the time of the system fault alarm. The aforementioned results prove that the proposed method could effectively alleviate the effects of abnormal training data of the target WT on the anomaly detection results.

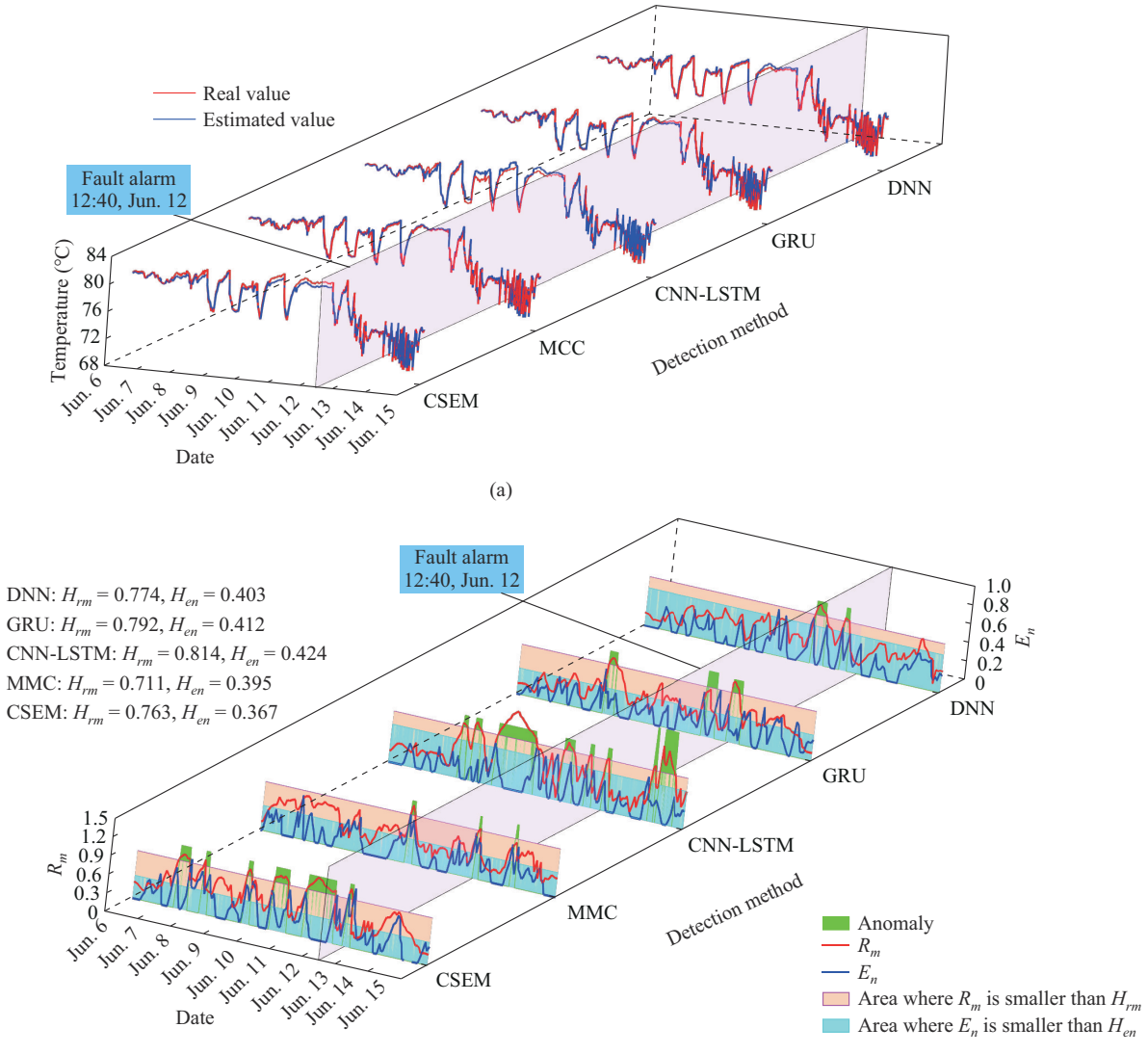


Fig. 20. Anomaly detection results of different methods. (a) Real and estimated values. (b) Anomaly detection results.

D. Discussion

The comparison results of the three schemes in Cases 1 and 2 demonstrated that the proposed CSEM outperformed the LSTM model trained solely on historical data from a single WT in terms of estimation accuracy and model stability. This was primarily because the CSEM uses a decision-layer fusion method to combine the advantages of multiple LSTM models. This not only undermines the effects of anomalous and insufficient training data on the estimation results of a single LSTM model, it also enhances the reliability of the CSEM when dealing with various complex scenarios. In addition, the proposed method can reliably and efficiently assess the similarities of two time-series as follows. First, the TSS employs a piecewise linearization method for segment comparison, which ensures the reliability of time series analysis. Second, Euclidean distance is used to calculate the amplitude and trend change distances of each segment, making it more efficient than DTW and SBD. Finally, the CSEM demonstrated superior accuracy and reliability in detecting anomalies as compared with the other models. Its success can be attributed to its robustness and generalization capabilities, which can use historical data from similar WTs to mitigate the effects of abnormal or insufficient data on model training. The proposed method was demonstrated to be effective and feasible through case studies and comparative experiments. This offers a new perspective on leveraging valuable information from the condition-monitoring data of multiple WTs.

VI. Conclusion

This paper proposed a method to improve anomaly detection accuracy using the SCADA data of WTs under similar operating conditions in the same WF. An operating condition similarity assessment method was first presented that could objectively quantify the operating condition similarities of different WTs by considering the similarities of multiple variables. Different LSTM models were then trained using SCADA data from the target WT and selected similar WTs. A linear combination method was used to combine the LSTM models with superior performance to construct a CSEM for estimating the expected state of the detection target variable. In addition, an improved abnormal state identification method combining residuals effective value and information entropy comparisons was proposed to identify the abnormal state of the target variable. The effectiveness of the proposed method was verified using actual SCADA data from two onshore WFs. The results demonstrated that the sensitivity and reliability of the CSEM could be improved by using the SCADA data of similar WTs. Even when only the historical data of similar WTs are used, a CSEM could be constructed to conduct anomaly detection for the target WT. Experiments were conducted to compare and verify the detection accuracy of the proposed method with those of other anomaly detection methods. The experimental results showed that the proposed method had more advantages than the other methods and could maintain good detection accuracy when the historical training data of the target WT were insufficient or contained abnormal data.

References

[1] D. Agarwal and N. Kishor, "A fuzzy inference-based fault detection scheme using adaptive thresholds for health monitoring of offshore wind-farms," *IEEE Sensors Journal*, vol. 14, no. 11, pp. 3851-3861, Nov. 2014.

[2] F. Qu, J. Liu, H. Zhu *et al*., "Wind turbine fault detection based on expanded linguistic terms and rules using non-singleton fuzzy logic," *Applied Energy*, vol. 262, p. 114469, Mar. 2020.

[3] L. Zhang and Z. Lang, "Wavelet energy transmissibility function and its application to wind turbine bearing condition monitoring," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1833-1843, Oct. 2018.

[4] Q. Zhou, C. Wang, and G. Zhang, "Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems," *Applied Energy*, vol. 250, pp. 1559-1580, Sept. 2019.

[5] S. Fong, J. Harmouche, S. Narasimhan *et al.*, "Mean shift clustering-based analysis of nonstationary vibration signals for machinery diagnostics," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 4056-4066, Jan. 2020.

[6] J. Wang, F. Cheng, W. Qiao *et al.*, "Multiscale filtering reconstruction for wind turbine gearbox fault diagnosis under varying-speed and noisy conditions," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4268-4278, May 2018.

[7] X. Zeng, M. Yang, and Y. Bo, "Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation," *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106233, Dec. 2020.

[8] F. Qu, J. Liu, X. Liu *et al.*, "A multi-fault detection method with improved triplet loss based on hard sample mining," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 127-137, Jan. 2020.

[9] L. Yang and Z. Zhang, "Wind turbine gearbox failure detection based on SCADA data: a deep learning-based approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, p. 3507911, Apr. 2021.

[10] Q. He, J. Zhao, G. Jiang *et al.*, "An unsupervised multiview sparse filtering approach for current-based wind turbine gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5569-5578, Aug. 2020.

[11] P. Qian, D. Zhang, X. Tian *et al.*, "A novel wind turbine condition monitoring method based on cloud computing," *Renewable Energy*, vol. 135, pp. 390-398, May 2019.

[12] W. Lu, Y. Li, Y. Cheng *et al.*, "Early fault detection approach with deep architectures," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 7, pp. 1679-1689, Jul. 2018.

[13] M. Schlechtingen and I. Ferreira Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1849-1875, Jul. 2011.

[14] W. Liao, B. Bak-Jensen, J. R. Pillai *et al.*, "Data-driven missing data imputation for wind farms using context encoder," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 964-976, Jul. 2022.

[15] J. Chen, J. Li, W. Chen *et al.*, "Anomaly detection for wind turbines based on the reconstruction of condition parameters using stacked denoising autoencoders," *Renewable Energy*, vol. 147, pp. 1469-1480, Mar. 2020.

[16] Z. Xue, K. Xiahou, M. Li *et al.*, "Diagnosis of multiple open-circuit switch faults based on long short-term memory network for DFIG-based wind turbine systems," *IEEE Journal of Emerging & Selected Topics in Power Electronics*, vol. 8, no. 3, pp. 2600-2610, Sept. 2019.

[17] Y. Pang, Q. He, G. Jiang *et al.*, "Spatio-temporal fusion neural network for multi-class fault diagnosis of wind turbines based on SCADA data," *Renewable Energy*, vol. 161, pp. 510-524, Dec. 2020.

[18] M. Ma and Z. Mao, "Deep-convolution-based LSTM network for remaining useful life prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1658-1667, Mar. 2021.

[19] R. Zhao, R. Yan, J. Wang *et al.*, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, p. 273, Feb. 2017.

[20] P. Cambron, C. Masson, A. Tahan *et al.*, "Control chart monitoring of wind turbine generators using the statistical inertia of a wind farm average," *Renewable Energy*, vol. 116, pp. 88-98, Feb. 2018.

[21] R. Fang, R. Shang, M. Wu *et al.*, "Application of gray relational analysis to *k*-means clustering for dynamic equivalent modeling of wind farm," *International Journal of Hydrogen Energy*, vol. 42, no. 31, pp. 20154-20163, Aug. 2017.

[22] P. Sun, J. Li, C. Wang *et al.*, "A generalized model for wind turbine anomaly identification based on SCADA data," *Applied Energy*, vol. 168, pp. 550-567, Apr. 2016.
[23] Y. Zhang, "Research on time series pattern matching," Ph.D. dissertation, Huazhong University of Science and Technology, Wuhan, China, 2012.
[24] Z. Li, H. Zhang, S. Wu *et al.*, "Similarity measure of time series based on feature extraction," in *Proceedings of 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, Chengdu, China, pp. 13-16, Apr. 2020.
[25] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of 12th National Conference on Artificial Intelligence – Workshop on Knowledge Discovery in Databases*, Seatle, USA, Aug. 1994, pp. 359-370.
[26] J. Paparrizos, L. Gravano, and S. Acm, "*k*-shape: efficient and accurate clustering of time series," in *Proceedings of ACM SIGMOD International Conference on Management of Data*, Melbourne, Australia, Jun. 2015, pp. 1855-1870.
[27] M. Li, D. Yu, Z. Chen *et al.*, "A data-driven residual-based method for fault diagnosis and isolation in wind turbines," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 895-904, Apr. 2019.
[28] G. Leite, G. Cunha, J. Junior *et al.*, "Alternative fault detection and diagnostic using information theory quantifiers based on vibration time-waveforms from condition monitoring systems: application to operational wind turbines," *Renewable Energy*, vol. 164, pp. 1183-1194, Feb. 2021.
[29] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 980-987, Mar. 2015.
[30] X. Su, Y. Shan, C. Li *et al.*, "Spatial-temporal attention and GRU based interpretable condition monitoring of offshore wind turbine gearboxes," *IET Renewable Power Generation*, vol. 16, no. 2, pp. 402-415, Feb. 2022.
[31] Z. Li, N. Qu, X. Li *et al.*, "Partial discharge detection of insulated conductors based on CNN-LSTM of attention mechanisms," *Journal of Power Electronics*, vol. 21, no. 7, pp. 1030-1040, Jul. 2021.
[32] J. Chen, W. Chen, J. Li *et al.*, "A generalized model for wind turbine faulty condition detection using combination prediction approach and information entropy," *Journal of Environment Informatics*, vol. 32, no. 1, pp. 14-24, Sept. 2018.

Xiangjun Zeng received the B.S. and M.S. degrees in electrical engineering from Xinjiang University, Urumqi, China, in 2014 and 2017, respectively, and the Ph.D. degree in electrical engineering from Shandong University, Jinan, China, in 2022. He is currently a Lecturer in the College of Electrical Power & New Energy, China Three Gorges University, Yichang, China. His research interests include condition-based maintenance, intelligent anomaly detection and fault diagnosis for various power equipment.

Ming Yang received the B.Sc. and Ph.D. degrees in electrical engineering from Shandong University, Jinan, China, in 2003 and 2009, respectively. He was an exchange Ph.D. student with the Energy System Research Center, The University of Texas at Arlington, Arlington, USA, from October 2006 to October 2007. He conducted post doctoral research with the School of Mathematics of Shandong University from July 2009 to July 2011. He is currently working as a Professor in Shandong University. His research interests include power system optimal operation and control, renewable energy generation forecast, and condition-based anomaly detection and fault diagnosis of various power equipment.

Chen Feng received the B.S. degree from Shandong University of Science and Technology, Qingdao, China, in 2012, and the M.S. and Ph.D. degrees from China University of Mining and Technology, Beijing, China, in 2016 and 2020, respectively. From 2017 to 2018, he was a Visiting Scholar with the Ohio State University, Columbus, USA. He is currently a Postdoctoral Researcher in Shandong University, Jinan, China. His research interests include anomaly detection, machine learning-based maintenance, and intelligent fault diagnosis for various power equipment.

Mingqiang Wang received the B.Sc. and M.S. degrees in electrical engineering from Shandong University, Jinan, China, in 2004 and 2007, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2012. Currently, he holds the position of Associate Professor at the School of Electrical Engineering, Shandong University. His primary research interests include economic operation and control of power systems as well as safe and economical operation of microgrid.

Xialing Qin received the B.Sc. degree in electrical engineering and automation from Shandong University of Technology, Zibo, China, in 2013, and the M.S. degree in electrical engineering from Xinjiang University, Urumqi, China, in 2017. Then, she worked as an Electrical Engineer at the State Grid Corporation, Urumqi, China, from 2017 to 2022. She is currently an Experimental Teacher at the College of Electrical Power & New Energy, China Three Gorges University, Yichang, China. Her primary research interests include power system protection and fault diagnosis.