Fault Detection of Wind Turbines by Subspace Reconstruction-Based Robust Kernel Principal Component Analysis

Kai Zhang, Baoping Tang, Lei Deng, and Xiaoxia Yu

***Abstract*— Data collected from the supervisory control and data acquisition (SCADA) system are widely used in wind farms to detect wind turbines' (WTs) faults. However, it is challenging to extract wind turbines' performance trends and provide accurate alarms based on SCADA data with large operating conditions fluctuations. This article presents a fault detection frame of subspace reconstruction-based robust kernel principal component analysis (SR-RKPCA) model for wind turbines SCADA data to extract nonlinear features under discontinuous interference. First, to improve the stability of the fault detection model of wind turbines, an RKPCA method is developed to decompose the original signal, decomposed into principal component subspace and residual subspace. Second, the permutation entropy is adopted to measure the residual matrix components' disturbance level and remove trendless noise components. Then, a combined index is developed for fault detection with the aid of the residual space matrix with trend information. Ultimately, the proposed SR-RKPCA method for wind turbine detection is verified by investigating wind turbine faults cases. The proposed method has outstanding robustness and a more vital ability to extract nonlinear features for fault detection than traditional principal component analysis (PCA)- or KPCA-based methods.**

***Index Terms*— Fault detection, RKPCA, subspace reconstruction, supervisory control and data acquisition (SCADA), wind turbines (WTs).**

Symbol	Description
$Y \in \mathbb{R}^{n\times m}$	Input SCADA data set.
$\boldsymbol{t} \in \mathbb{R}^{n}$	ith score vector.
$\boldsymbol{p} \in \mathbb{R}^{m}$	Corresponding loading vector.
$X \in \mathbb{R}^{n}$	Matrix reconstructed by the PCs.
$E \in \mathbb{R}^{n}$	Residual matrix.
C^F	Covariance matrix on feature space.
λ	Eigenvalues.
Λ	Diagonal matrix of nonzero eigenvalues.
V	Eigenvectors.
K	Kernel matrix.
CPV	Cumulative percent variance.
T^2	Hoteling's T^2 statistic index.
τ^2	Confidence limits of T^2.
SPE	Squared prediction error.
δ^2	Confidence limits of SPE.
Ψ	Combined index of T^2 and SPE.
ζ	Confidence limits of Ψ.
$\mathcal{L}$	Augmented Lagrangian function.
$\odot$	Hadamard product.
$\Theta(\cdot)$	Element-wise soft threshold operator.
E_l^p	Phase space reconstructed matrix.
H_{PE}	Permutation entropy.
H_{NPE}	Normalized permutation entropy.

Manuscript received January 22, 2021; revised April 2, 2021; accepted April 19, 2021. Date of publication April 26, 2021; date of current version May 10, 2021. This work was supported in part by the National Key Research and Development Project under Grant 2020YFB1709800, in part by the National Natural Science Foundation of China under Grant 51775065, and in part by the Science and the Technology Projects in Chongqing under Grant cstc2019jcyj-zdxmX0026. The Associate Editor coordinating the review process was Arunava Naha. *(Corresponding authors: Baoping Tang; Lei Deng.)*

The authors are with The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400030, China (e-mail: zhangkai.ac@hotmail.com; bptang@cqu.edu.cn; denglei@cqu.edu.cn).

Digital Object Identifier 10.1109/TIM.2021.3075742

I. Introduction

THE beginning of the 21st century was a rapid growth period of global wind power. Currently, wind energy has been the fastest growing renewable energy source in the world. However, with the increase in wind turbine installations, wind turbine (WT) failures are increasingly prominent and show an upward trend [1]. Given this, fault detection technologies capture extensive attention in the industry and have shown great application value in ensuring wind turbines' safe operation.

The data-driven fault detection method of wind turbines is becoming a hot topic because it does not rely on accurate physical models and rich signal processing experience [2]–[4]. In recent years, various wind farms have collected various historical and real-time data and stored it in the database. The supervisory control and data acquisition (SCADA) systems are comprehensive technical means for monitoring wind turbines or industrial production [5]. It gains data with an interval of a few seconds to 10 min, installed in a host of WTs produced by major manufacturers, such as Vestas, GE, and Siemens, to monitor the operational performance mentioned in [6]. Based on wind turbine SCADA signals, such as power output, wind speed, and bearing temperatures, previous researchers conducted a lot of research on failure detection and diagnosis by principal component analysis (PCA) [7], support vector regression (SVR) [8], artificial neural networks (ANNs) [6], [9], [10] and its improved deep neural network [11], and other condition monitoring algorithms [6], [12]–[14].

As one of the most popular data-based fault detection methods, the PCA and its extension theory have been widely studied by scholars and applied in engineering practice, which

shows the effectiveness and superiority of such methods. PCA is a classical linear data dimension reduction technique, which transforms the original variables into new uncorrelated variables arranged by their variances [15]. However, it is not effective in handling nonlinear trends and is also not robust to sparse corruptions and outliers. To avoid false alarms when PCA performs fault detection on nonlinear data, kernel PCA (KPCA) and its extended methods were proposed to solve the nonlinear problem by kernel trick and have been applied to many practical problems. For example, work [16] presented a dynamic kernel principal component analysis (DKPCA) that PCA-based fault detection enables to monitor an arbitrary process with severe nonlinearity and (or) dynamics. Work [17] defined two statistics representing each variable's contribution to the monitoring statistics, Hotelling's T^2 and SPE of KPCA, respectively, based on the KPCA detection framework. Work [18] adopts the measurements in block Hankel matrices to improve the detection results. What is more, work [19] proposes a new hybrid nonlinear statistical modeling approach for nonlinear process monitoring by closely integrating PCA and KPCA by a serial model structure, which considers both linear and nonlinear principal components (PCs) simultaneously.

Unfortunately, PCA and KPCA cannot be directly applied to monitoring signals with strong background noise because the PCs obtained from the feature space are mapped back into the data space to reconstruct the observed variables, where the reconstruction errors are regarded as noises. On the other hand, to solve the problem of poor robustness, several robust PCA methods were proposed. Robust kernel principal component analysis (RKPCA) was initially proposed for noise reduction and repair in video and image by the Geman–McClure function. The principle of this method is to decompose noise information by decomposing the input data matrix into two information matrices of low-rank background information and sparse noise information. Similarly, in equipment fault diagnosis, the input matrix X of the SCADA monitoring data generally contains background trend information and contains noise signals. If RPCA preprocesses the data, the extraction effect of trend features can be significantly improved, and noise information can be effectively reduced [20].

However, these are still some unsolved issues. KPCA is not robust to sparse noises [20], and RPCA, as well as its recent extensions, is unable to handle nonlinear data [21] and becomes prone to false positives and ineffective when the observations are grossly corrupted with strong outliers and noise [22]. Consequently, there is a need to derive a variant of PCA that can handle nonlinear SCADA data and robust to wind turbines' noises. To complete it, work [23] extends KPCA to a framework for treating noise, missing data, and outliers by training the model on a clean training dataset. To avoid this strict application premise, recently, work [21] proposes an actual practical, robust extensions method that solves nonconvex and nondifferentiable by nonconvex alternating direction method of multipliers with backtracking line search and nonconvex proximal linearized minimization with adaptive step size.

Unlike fault classification that focuses on extracting the detailed information of the discrimination in signals, fault monitoring needs to extract the trend information in the vibration signal. However, when adopting PCA and its extended methods for fault detection, researchers tend to use only the projection matrix on the PC subspace and treat the residual subspace's projection matrix as random noise and directly discard it. Although KPCA solves nonlinear feature extraction and robustness to a certain extent, the approach of directly discarding residual information is not wise, especially in the extraction of trend information. The reason is the trend information in the PC cannot be fully extracted by RKPCA [24] or the non-Gaussian noise itself contains useful trend information in the residual subspace [25], which motivates a further study.

Inspired by the work of serial PCA (SPCA) [19], which integrated information of PCA and KPCA subspace, and the research on image recovery based on RKPCA [21], this article presents a fault detection frame based on a novel subspace reconstruction-based robust kernel principal component analysis (SR-RKPCA) model for wind turbines SCADA data to extract nonlinear features under discontinuous interference. The main contributions of this article are given in the following.

1) To improve the stability of the fault detection model of wind turbines, a robust KPCA method is developed to decompose the original signal, which is decomposed into PC subspace and residual subspace.
2) Permutation entropy (PE) is adopted to measure the residual matrix components' disturbance level and remove noise components without trend information. A combined index is developed based on the subspace reconstruction matrix for fault detection with the aid of residual space matrix's trend information.

The structure of this article is as follows. Section III reviews the fault detection based on PCA and its kernel extension method and then introduces the RKPCA theory. A novel SR-RKPCA method is proposed. The case studies in Section IV validate the performance of the proposed method. Also, some conclusions are obtained in Section V.

II. Methodology

Considering that the fault detection method proposed in this article involved PCA-based fault detection and its extended methods, this section will briefly introduce these methods. On this basis, this section will elaborate the proposed fault detection method in detail.

A. Fault Detection Based on PCA and KPCA

PCA can effectively extract critical elements and structures in the SCADA data of wind turbines and, at the same time, remove noise and redundancy. This method reduces the complexity of the original data to reveal the hidden trend information. Also, the corresponding eigenvalues and eigenvectors can be obtained by calculating the covariance matrix of the input data. From the perspective of linear algebra, PCA's goal

is to use another set of bases to redefine the obtained data space, which is expected to reveal the relationship between the original data as much as possible, namely the PC of this dimension. Finally, the final matrix is obtained by selecting the eigenvectors corresponding to the k features with the largest eigenvalues (with the most significant variance). As can be shown, the original dataset $Y \in \mathbb{R}^{n\times m}$, which is composed of n samples of m variables, is decomposed into the following two parts:

$$Y = X + E = \sum_{i=1}^{k} \boldsymbol{t}_i \, \boldsymbol{p}_i^{\mathrm{T}} + E \tag{1}$$

where $\boldsymbol{t}_i \in \mathbb{R}^n$ is the ith PC, $\boldsymbol{p}_i \in \mathbb{R}^m$ is the corresponding loading vector, and k is the number of retained PCs, while $X \in \mathbb{R}^{n\times m}$ is the matrix reconstructed by the PCs, which represent the dominating data change. $E \in \mathbb{R}^{n\times m}$ is the residual matrix.

KPCA is a nonlinear extension of the PCA algorithm. The idea is to convert the original space into the hidden feature space through a nonlinear mapping, which is about extracting the nonlinear PCs and retaining the original data structure [26]. In KPCA, the covariance matrix C^F is the mapping matrix on the hidden feature space by the nonlinear mapping of Y

$$C^F = \frac{1}{M}\sum_{j=1}^{M} \varphi\left(\boldsymbol{y}_j\right)\varphi\left(\boldsymbol{y}_j\right)^{\mathrm{T}}. \tag{2}$$

Then, the above formula can be derived as follows:

$$\lambda V = C^F V \tag{3}$$

where λ is the eigenvalues ($\lambda \geq 0$) and V is the eigenvectors. The $n \times n$ kernel matrix is defined as

$$K_{ij} = K\left(\boldsymbol{y}_i, \boldsymbol{y}_j\right) = \varphi\left(\boldsymbol{y}_i\right)\varphi\left(\boldsymbol{y}_j\right). \tag{4}$$

Normalize the feature vector $\bar{V}$. After the original sample data Y mapped to the high-dimensional feature space F, the ith principal $\boldsymbol{t}_i$ of $\varphi(\boldsymbol{y})$ on the feature vector matrix $\bar{V}$ can be represented as follows:

$$\boldsymbol{t}_i = \left(V^m, \varphi(\boldsymbol{y})\right) = \sum_{i=1}^{M} \alpha_i^m K\left(\boldsymbol{y}, \boldsymbol{y}_i\right) \tag{5}$$

where α_i^m is the correlation coefficient of C^F. Since the mapping $\varphi(\boldsymbol{y})$ of the data Y in the feature space is not necessarily zero-averaged, the matrix K needs to be zero-averaged according to the following formula:

$$\widetilde{K} = K - L_M K - K L_M + L_M K L_M \tag{6}$$

where L_M is an $M \times M$ matrix with all elements $1/M$. Let λ_i be the eigenvalue of the matrix K. The CPV is a measure of the percent variance captured by the first l eigenvalues whose contribution rate satisfies the predetermined limit, which is selected as the PC eigenvalues

$$\mathrm{CPV} = 100 \times \frac{\sum_{i=1}^{l}\lambda_i}{\sum_{i=1}^{m}\lambda_j}. \tag{7}$$

The PCA-based fault detection method is to map the data vectors to the mutually orthogonal principal space and residual space and conduct hypothesis testing by establishing statistics. Hoteling's T^2 statistic reflects a variation in the model subspace and has been widely applied to detect faults in wind turbine gearboxes and pitch motors [27], [28]. The calculation formula is as follows:

$$T^2 = \widehat{\boldsymbol{t}}^{\mathrm{T}} \Lambda \widehat{\boldsymbol{t}} \tag{8}$$

where $\widehat{\boldsymbol{t}}$ is the features of PCs and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_l)$ denotes the diagonal matrix of nonzero eigenvalues. The confidence threshold of T^2 can be obtained by the F distribution

$$\tau^2 \sim \frac{d(m-1)}{m-d} F(d, m-1; \alpha) \tag{9}$$

where $F(d, m-1; \alpha)$ is an F-distribution with degrees of freedom d and $m-1$ with a level of significance α.

The SPE index [19] presents the deviation level of the real-time data from the residual components of trained data, which is calculated as the squared norm of the residual components

$$\mathrm{SPE} = \widetilde{\boldsymbol{t}}^{\mathrm{T}} \widetilde{\boldsymbol{t}} \tag{10}$$

where $\widetilde{\boldsymbol{t}}$ is the residual components and the confidence limits for the SPE are calculated from the χ^2 distribution

$$\delta^2 = \frac{\sigma(x)}{2\mu}\chi_\alpha^2\left(\frac{2\mu^2}{\sigma}\right) \tag{11}$$

where μ and σ are the estimated mean and variance, respectively, of the SPE from the training data. $\chi_\alpha^2((2\mu^2/\sigma))$ is the critical value of the chi-squared variable with $((2\mu^2/\sigma))$ at the significance level α.

The combined index [29] is used to reconstruct the feature of PCs subspace and residual components subspace, providing a complete measure of the entire space variability

$$\Psi = \frac{\mathrm{SPE}}{\delta^2} + \frac{T^2}{\tau^2}. \tag{12}$$

The combined index should be within the combined control limits [30], that is

$$\zeta = g^\zeta \chi_\alpha^2\left(h^\zeta\right) \tag{13}$$

where the parameters are defined as

$$g^\zeta = \frac{l/\tau^4 + \sum_{i=l+1}^{m}\lambda_i^2/\delta^4}{(m-1)\left(l/\tau^2 + \sum_{i=l+1}^{m}\lambda_i/\delta^2\right)} \tag{14}$$

$$h^\zeta = \frac{\left(l/\tau^2 + \sum_{i=l+1}^{m}\lambda_i/\delta^2\right)^2}{l/\tau^4 + \sum_{i=l+1}^{m}\lambda_i^2/\delta^4}. \tag{15}$$

B. RPCA and RKPCA

Like the classical PCA method, RPCA essentially finds the best mapping of data in a low-dimensional space. For the input matrix $Y \in \mathbb{R}^{n\times m}$ of low-rank data, when Y is affected by random (sparse) noise, the low rank of Y will be destroyed. Accordingly, it is significant to decompose Y into the sum of a low-rank matrix and a sparse noise matrix with a true structure. If the low-rank matrix is found, the essential low-dimensional space of the data is located.

RPCA-based fault detection is the following optimization issue for the original monitoring data:

$$\min_{X,E} \mathrm{rank}(X) + \lambda \|E\|_0, \qquad \text{s.t. } Y = X + E. \tag{16}$$

Since rank(X) and $\|E\|_0$ norms are nonconvex and nonsmooth in optimization, it is further approximated by

$$\min_{X,E} \|X\|_* + \lambda\|E\|_1, \qquad \text{s.t. } Y = X + E \tag{17}$$

where $\|X\|_*$ is the matrix norm of X.

The RPCA convex optimization problem's solution is based on the following assumption [20]: 1) the optimal local solution of the convex optimization problem is the optimal global solution; 2) Lagrangian dual problems (nonconvex problems) can be equivalently converted into convex optimization problems or approximated to convex optimization problems; and 3) as a mature research topic, when a specific problem is considered a convex optimization problem, it can be solved. However, Y could be of high-rank or even full-rank in quite a few cases, which is beyond the assumption of RPCA. Nevertheless, when the data are seriously corrupted by sparse noise, RPCA's performance will suffer huge performance degradation. In fact, highly corrupted data may lead to significant deviations in sparse or low-level representations. The kernel can decompose nonlinear noise and clean data and then perform sparse representation or other methods on the clean data to improve the performances. The combined index RKPCA can be directly calculated to fault detection or be further enhanced, as discussed in Section II-C.

The radial basis function (RBF) kernel performs better in kernel methods than a polynomial kernel [21], [31]. Similarly, in RKPCA, the RBF kernel is found more effective. For this reason, this article pours particular attention into the RBF kernel. It can be written as

$$\begin{aligned} k(\boldsymbol{y}_i, \boldsymbol{y}_j) &= \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2\right) \\ &= C\exp\left(\frac{1}{\sigma^2}\langle \boldsymbol{y}_i, \boldsymbol{y}_j\rangle\right) \end{aligned} \tag{18}$$

$$\begin{aligned} \varphi(y) = [&c_0, c_1 y_1, \ldots, c_d y_d, \ldots, c_{ij} y_i y_j, \ldots, \\ &c_{i,\ldots,k} y_i, \ldots, y_k, \ldots,]. \end{aligned} \tag{19}$$

As to RKPCA, with the kernel matrix $K = [k(\boldsymbol{y}_i, \boldsymbol{y}_j)]_{n\times n} = [\varphi(\boldsymbol{y}_i)^{\mathrm{T}}, \varphi(\boldsymbol{y}_j)]_{n\times n}$, problem (17) can be rewritten as

$$\min_{X,E} \mathrm{Tr}\left(K^{1/2}\right) + \lambda\|E\|_1, \qquad \text{s.t. } Y = X + E. \tag{20}$$

In this formula, the two unknown variables X and E are separable. The former term $\mathrm{Tr}(K^{1/2})$ could be nonconvex if a nonlinear kernel such as RBF kernel is used. The latter $\lambda\|E\|_1$ and is convex. The augmented Lagrangian [32] for this nonconvex problem is given by

$$\begin{aligned} \mathcal{L}(X, E, Z) = &\mathrm{Tr}\left(K^{1/2}\right) + \lambda\|E\|_1 \\ &+ \langle X + E - M, Z\rangle + \frac{\mu}{2}\|X + E - M\|_F^2 \end{aligned} \tag{21}$$

where $Z \in R^{d\times n}$ is the matrix of Lagrange multipliers and $\mu > 0$ is a penalty parameter.

In general, when fault detection is performed, the points number of the input data $R \in R^{n\times m}$ is greater than 2000. Therefore, the kernel function needs to be decomposed by randomized singular value decomposition [33]. The specific algorithm is summarized in Table I (ω is a weight parameter and is set according to work [21]).

TABLE I
PROCEDURES OF RKPCA ALGORITHM

RKPCA solved by PLM+AdSS and RSVD

Input: $Y, k(.,.), \lambda, \omega = 0.1, c > 1, i_{max}$.
Initialize: $E(0) = 0, t = 0$
Repeat
Compute $K^{1/2} = U\Sigma^{1/2}V^T$ by Randomized Singular Value Decomposition.
$D = 1/2K^{-1/2} \bigodot K$
Compute $\partial J/\partial E^{(i-1)}$ by formula
$\frac{\partial J}{\partial E^{(i-1)}} = -\frac{2}{\sigma^2}\left(\left(X - E^{(i-1)}\right)D - \left(X - E^{(i-1)}\right) \odot (BD)\right)$
$J = Tr\left(K^{1/2}\right) + \frac{\mu}{2}\left\|\hat{X} + E - M + Q/\mu\right\|_F^2$
$\upsilon = \omega\left\|\frac{2}{\sigma^2}(D - I \odot BD)\right\|_2$
$E^{(i)} = \Theta_{\lambda/\upsilon}\left(E^{(i-1)} - \frac{\partial J}{\partial E^{(i-1)}}/\upsilon\right)$
If $J(E^{(i)}) > J(E^{(i-1)})$ then $\omega = c\omega$
End if
$i = i + 1$
Until converged or $i = i_{max}$
Output: $E = E^{(i)}, X = Y - E$

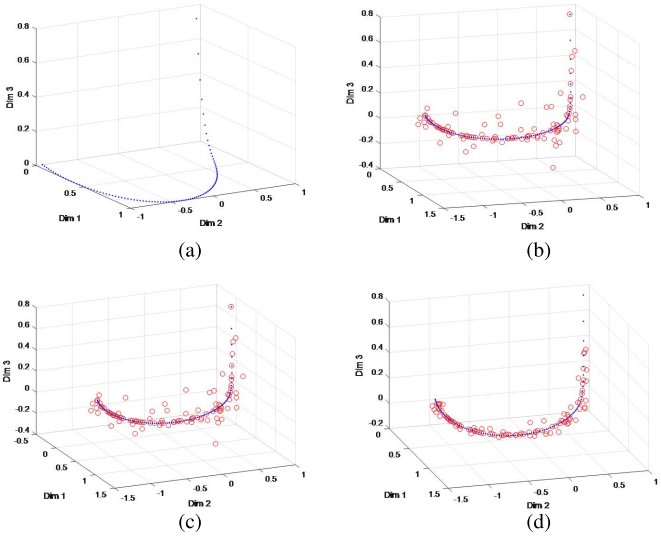

Fig. 1. Comparison of the recovery performance to nonlinear data. (a) Original data. (b) Data with noise. (c) RPCA. (d) RKPCA.

As an example, a nonlinear 3 × 100 matrix, shown in Fig. 1(a), is generated, and the Gaussian noise $N(0, 0.1)$ is added to one randomly chosen entry of each column as a corrupted matrix, shown in Fig. 1(b), where blue points mark the clean data, and the noise data are marked by red circles. Fig. 1(c) shows the matrix recovered by RPCA, whereas Fig. 1(d) shows the matrix recovered by RKPCA. As can be seen from the figure, RPCA failed in recovering the data, while RKPCA has an excellent performance, which has the ability of both nonlinear feature extraction and robustness.

C. Proposed Fault Detection Method Based on SR-RKPCA

In this section, a framework based on subspace restructuring RKPCA was submitted for the fault detection of wind turbines. Generally, wind turbines' fault detection is based on the sensitive SCADA indicators and an alarm threshold pretrained by normal historical data that includes various wind turbines'

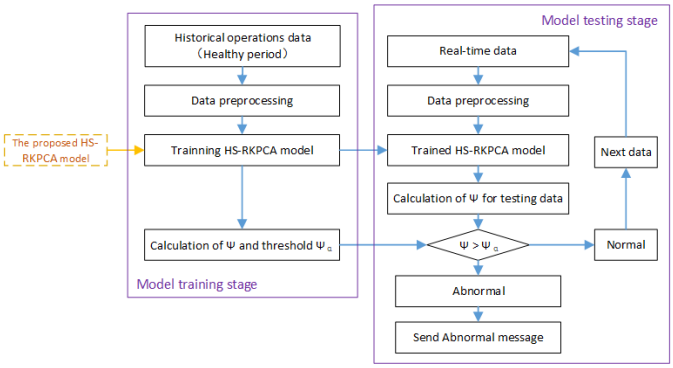

Fig. 2. General procedure of the proposed SR-RKPCA-based fault detection framework.

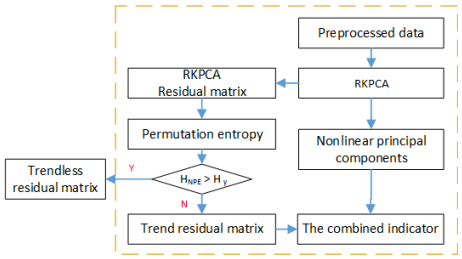

Fig. 3. Procedure of the proposed SR-RKPCA.

working conditions. If the index continuously exceeds the threshold when real-time data are input to the trained model, the wind turbine is considered abnormal. The general procedure of the proposed framework is shown in Fig. 2. The process of the proposed framework is divided into two stages: training and testing. During the training stage, after a series of preprocessing (it will be discussed in the case study), the training set is used to train the proposed SR-RKPCA model and to obtain the monitoring indicator and the alarm threshold. When real-time data are collected, the same preprocessing is required. Then, the preprocessed real-time data is input into the trained model to calculate the combined indicator Ψ. After that, the monitoring indicator is compared with the threshold of it. If the combined indicator index Ψ exceeds the threshold, an alarm message will be sent to the monitoring system. Otherwise, repeat the above steps. A summary of the proposed SR-RKPCA, as shown in Fig. 3, is presented as follows.

First, decomposition of preprocessed data, the RKPCA-based fault detection decomposes the PC space and residual space, intending to extract the trend information in them more effectively. The preprocessed SCADA data are mapped to the F space by the RKPCA method mentioned in Section II-B and the corresponding residual space signal $E = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$ can be obtained, and the residual space signal E can be further used for calculating the nonlinear PCs $X = Y - E = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$. In the traditional PCA-based model, this residual matrix is directly discarded, which may also give rise to part of the trend information being lost. Entropy is a measure of the disorder level of a nonlinear system. Among them, the PE can effectively magnify the weak changes of time sequences [33]. Gaussian noise is completely random noise, so its PE is larger than other non-Gaussian noise [25], which contains trend information of wind turbine status. Therefore, in this article, each component of the residual matrix's random distribution features is evaluated by PE. Also, the residual subspace components with lower values are retained for reconstruction for fault detection together with the PC components.

Second, phase space reconstruction and PE solution, we perform phase space reconstruction on the residual matrix to obtain the reconstructed matrix as follows:

$$e_i^p = \begin{bmatrix} e_i(1) & e_i(1+\eta) & \cdots & e_i(1+(\varepsilon-1)\eta) \\ e_i(2) & e_i(2+\eta) & \cdots & e_i(2+(\varepsilon-1)\eta) \\ \cdots & \cdots & \cdots & \cdots \\ e_i(\beta) & e_i(\beta+\eta) & \cdots & e_i(\beta+(\varepsilon-1)\eta) \end{bmatrix} \tag{22}$$

where $1 \le i \le n, \beta = m - (\varepsilon - 1)\eta$, ε is embedded dimension, and $\hat{1}$ is the delay time. The parameters are set as $\varepsilon = 5$ and the delay time $\eta = 1$ according to work [34], enabling PE to reflect the fluctuation tendency approximately. Each row in the matrix can be regarded as a reconstructed component, with a total of β reconstructed components. Then, the elements of each reconstructed component are rearranged in ascending order according to the magnitude of the value. Finally, the permutation of different symbol sequences, $m!$, mapped by the phase sequence, is obtained by extracting the column's index where each element is in the original reconstructed component. The following formula solves the probability of each combination:

$$p(i) = \frac{\mathrm{Num}\{e_i^p\}}{\beta} \tag{23}$$

where $\mathrm{num}(e_i^p)$ represents the number of any combination. Then, the PE of the residual matrix is solved as follows:

$$H_{\mathrm{PE}}(\eta) = \sum_{i=1}^{\beta} p(i)\ln(p(i)). \tag{24}$$

When $p(i) = 1/m!$, $H_{\mathrm{PE}}(\eta)$ reaches the maximum value $\ln(m!)$. To more intuitively observe and compare entropy values, $H_{\mathrm{PE}}(\eta)$ are normalized through

$$H_{\mathrm{NPE}}(\eta) = \frac{H_{\mathrm{PE}}(\eta)}{\ln(m!)}, \quad 0 \le H_{\mathrm{NPE}}(\eta) \le 1. \tag{25}$$

Third, subspace reconstruction, according to the principle of PE, the larger the value of PE, the stronger the time series's randomness and disorder; conversely, the smaller the value of PE, the stronger its regularity. Therefore, in this article, by zeroing the components whose PE value is higher than the threshold H_γ, the residual components with PE value lower than the threshold H_γ(when $H_\gamma = 0.5$, the white noise is obtained [25]) are selected to reconstruct with the PCs. As a result, this method retains trend information while eliminating interference information. The reconstruction formula is as follows:

$$R = \mathrm{X} + E\left(H_{\mathrm{NPE}}(\tau) < H_\gamma\right). \tag{26}$$

Finally, the reconstruction matrix's combined index is extracted as the final monitoring index through KPCA introduced in Section II-A.

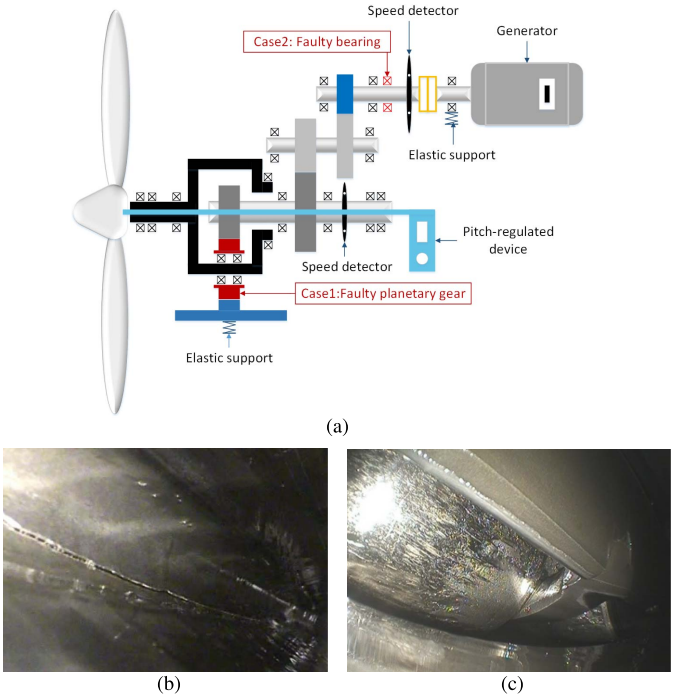

Fig. 4. Wind turbines with two fault scenarios. (a) Schematic of the doubly fed wind turbine transmission. (b) Penetrating crack in the first-stage planetary gear. (c) Large-area wear on the rear bearing of the gearbox.

III. Case Studies and Results Analysis

In this section, to demonstrate the effectiveness of the proposed fault detection framework based on SR-RKPCA, case studies are considered for two SCADA datasets, which come from gearboxes of 2 megawatt (MW) doubly fed wind turbines of a wind farm in China, as shown in Fig. 4(a). The first dataset was collected from December 1, 2015, to June 11, 2016, and planetary gear was found to have a penetrating crack, as shown in Fig. 4(b). While the second dataset covers 11 months from April 1, 2014, to April 25, 2015, the gearbox's rear bearing was found to have large-area wear, as shown in Fig. 4(c).

A. SCADA Data Interpretation

The SCADA system does not need to install an additional data acquisition system separately and is widely installed in wind turbines. Generally, the sampling interval of SCADA systems is 1–10 min, so a large amount of SCADA data is usually generated in the wind farm's operation. These SCADA data can be effectively used to monitor wind turbines and provide feasible fault alarms. Usually, operational wind farms collect generally various types of SCADA data, including data from output power, environmental parameters to system switching, and parameters related to wind speed, such as wind speed, wind direction, and yaw angle; parameters related to wind turbine performance, such as power, rotor speed, and blade pitch angle; parameters related to the temperature of wind turbines, such as the temperature of the oil sump of the gearbox, oil temperature at the inlet of the gearbox, and the temperature of the cooling water of the gearbox; and parameters related to the pressure of wind turbines, such as the gearbox outlet pressure, and gearbox inlet pressure. These parameters contain various status information of wind turbines' operation, and the wind turbines' operation status can be effectively monitored by analyzing the SCADA data.

However, because of the large types of SCADA monitoring variables for wind turbines, the datasets may contain irrelevant, missing, and redundant information to the wind turbine gearbox. SCADA monitoring variables need to be selected, cleaned, and preprocessed. In this article, 15 variables related to the wind turbine gearbox, listed in Table II, are selected from hundreds of SCADA parameters by previous works [10], [12] for monitoring the health status of the gearbox of wind turbines. Fig. 5 shows an example of the selected normalized wind turbine SCADA variables, which were acquired at 10-min intervals. These variables are related to the state of the gearbox and include the operating conditions of wind turbines.

TABLE II
Selected Wind Turbine Monitoring Variables

No.	SCADA variable	No.	SCADA variable
1	Active power	9	Gearbox output shaft No.2 temperature
2	Pitch angle	10	Gearbox cooling water temperature
3	Wind speed	11	Gearbox inlet oil temperature
4	Wind direction	12	Gearbox inlet pressure
5	Ambient temperature	13	Gearbox outlet pressure
6	Wind wheel speed	14	Nacelle X-direction vibration
7	Gearbox oil temperature	15	Nacelle Y-direction vibration
8	Gearbox input shaft No.1 temperature		

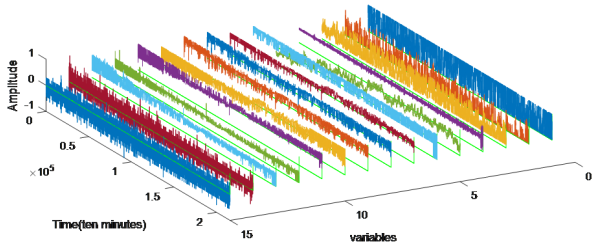

Fig. 5. Selected normalized variables of wind turbines SCADA data.

B. Data Preprocessing

Due to the variable operating conditions of wind turbines, there are quite a few missing data or invalid data in historical datasets that are arduous to use directly. At the same time, the kernel size of RKPCA increases exponentially with the increase of data. Therefore, it is indispensable to preprocess the original data before input the proposed model. First, in this article, the nonworking data, below the cut-in wind speed (3.5m/s) or exceed cut-out speed (25 m/s), is removed. Then, raw data (the green point clouds) and the power curve (the blue asterisks) of historical data are shown in Fig. 6, which is divided into the following three periods.

1) *Initial Period:* As shown in Fig. 6(a), the initial period's data should be avoided as a model training set for large-scale invalid data.
2) *Normal Period:* As shown in Fig. 6(b), the data in this period are collected under the healthy state of wind turbines, which occupies most of the time of wind turbine operation, and should be used as the training for detection model and alarm threshold.
3) *Abnormal Period:* Abnormal wind turbines are often accompanied by increased temperature. Fig. 6(c)

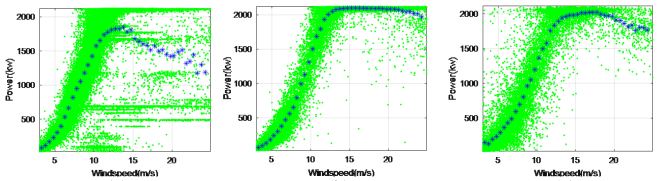

Fig. 6. Power curves of wind turbine operational data. (a) Initial period. (b) Normal period. (c) Abnormal period.

manifests the state of derating when the gearbox's temperature is higher than 85 °C for safety reasons.

The selection of SCADA datasets is determined by the relationship between active power and wind speed. The data obtained from the initial period that contains a large amount of invalid data or the abnormal period cannot be selected as training data from the above analysis. Therefore, during the training stage, the input data should be obtained from the noninitial period of healthy wind turbines or the normal period of abnormal wind turbines.

As for missing data, referring to work [10], local mean substitution is employed to complete this operation, which can repair missing data and improve computational efficiency and further increase the model's robustness. It can be expressed by

$$x_l = \frac{\sum_{l-k+1}^{l+k} x_i}{2k} \tag{27}$$

where x_l is the average value, k is the number of existing data values around the midpoint, and the span of local mean substitution is $2k$.

C. Results Analysis

The proposed method is validated against SCADA data of wind turbines with two different fault scenarios. The parameters of the proposed model vary with different data. The proposed model could successfully detect the fault by the recommended parameter settings for the datasets used in this article, as shown in Table III. The same data preprocessing performed, and then, the comprehensive indicator was extracted by PCA, KPCA, RPCA, and RKPCA methods for comparison. Their thresholds are not determined by the method, but by (13) of the combined indicator (including the statistical indicators' distribution and the confidence region of the statistical indicators together). Although the parameter α is taken as 0.05 for each method in this article and the $100(1-\alpha)\%$ control limit for the T^2 is calculated using the F distribution, this value needs to be applied in conjunction with more wind turbine data with necessary feedback adjustments. Currently, some scholars are studying the dynamic [35] or adaptive [36] adjustment of the fault detection threshold (not only for PCA-based methods), which is another fascinating but challenging research topic.

Besides, the deep learning method autoencoder (AE), thriving in anomaly detection, is also compared in this study. A network with a frame of "20-10-6-4-6-10-20" was trained with the batch size of 200 and epochs of 50 for comparison. The hyperparameters of activation function, optimizer, learning rate, and loss function are set to "rectified linear unit (ReLU)," "Adam," 0.001, and "mean squared error (MSE)." The rest are default parameters if not specified. The TensorFlow (1.15.4) and Keras (2.3.1) are installed to facilitate network construction and optimization. This article uses the Intel i7-6700 processor and NVIDIA's GeForce GTX 1080 GPU to speed up training and optimization time in hardware configuration.

TABLE III
PARAMETERS RECOMMENDED SETUP OF THE PROPOSED METHOD

Parameters	Values
Kernel type	RBF
Kernel width	10
Embedded dimension	5
Delay time	1
Principal contribution rate	0.85
The significance level	0.05
Weight	0.1
The permutation entropy threshold	0.5

1) Case 1: Wind Turbine With Gear Crack Failure: Generally speaking, gear cracks usually refer to fatigue cracks, which are mainly cracks that expand under the action of repeated alternating stresses or cyclic stresses that are significantly lower than the tensile strength of the material, usually appearing at the tooth roots. Also, sometimes, the initial cracks due to overload can also expand slowly like fatigue cracks. During operation, WTs gearboxs operating condition and the characteristics of the fault mainly determined by the vibration and noise of the gearbox, gear drive shaft torsional vibration, gear tooth root stress distribution, lubricating oil temperature and oil in the content of abrasive particles and their morphology and other components. At present, the commonly used condition monitoring technology oil analysis, vibration analysis, operating condition parametric analysis, and other technologies can achieve its monitoring. However, in most of the above technologies, oil analysis is carried out offline, while vibration data require the installation of additional sensors. The condition parameters obtained from the randomly installed SCADA system are much easier to obtain.

In the first case, the SCADA data collected from December 1, 2015, to June 11, 2016, in a wind turbine, and the gear failure was detected by disassembling inspection on July 1, 2016. The data from December 1, 2015, to January 25, 2016, was set as the training data, which is considered as a normal period. Also, the rest of the data are tested as real-time monitoring data to validate the model's performance.

By the proposed model of the SCADA data, state statistics of wind turbines per day were recorded, and the combined index of T^2 and SPE was calculated based on time sequence, as shown in Fig. 7. Due to the huge fluctuation of the SCADA data of the wind turbine, the fault detection based on the original data is prone to false alarms, so this article introduces the exponentially weighted moving average (EWMA) [1] method to smooth the original data. To a certain extent, the EWMA can eliminate or reduce the random changes of the time series data caused by accidental factors' interference and reflect the time

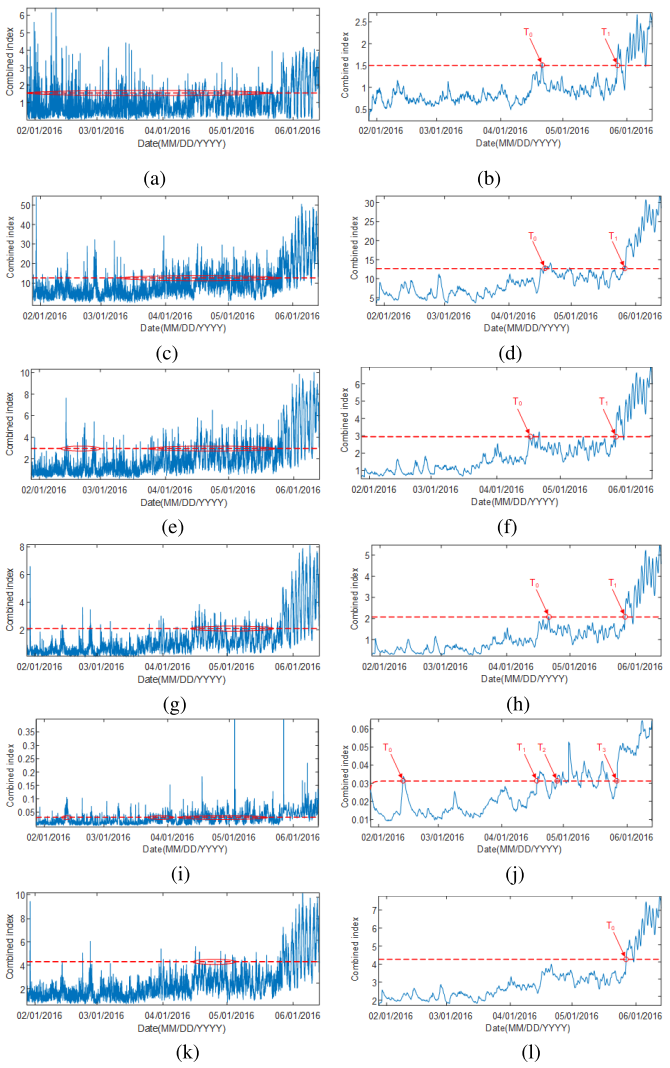

Fig. 7. Fault detection results of case 1 based on: (a) PCA, (b) EWMA of PCA, (c) KPCA, (d) EWMA of KPCA, (e) RPCA, (f) EWMA of RPCA, (g) RKPCA, (h) EWMA of RKPCA, (i) AE, and (j) EWMA of AE.

series's changing trend. However, the smoothed monitoring index has a certain lag, which will correspondingly delay the alarm of fault detection. Therefore, this article first triggered the alarm through the original indicator and then confirmed the fault through the smoothed alarm indicator, which combined the original indicator's detection advantages and the smoothed indicator.

Fig. 7(a)–(l) shows the fault detection results and the EWMA result of case 1 based on the SR-RKPCA model and the comparison models of PCA, KPCA, RPCA, RKPCA, and AE, where the left side is the original result and the right side is the EWMA result. As mentioned above, it can be seen from the results that the original combined indicator has great volatility, while the smoothed result has a certain degree of lag. It can be seen from Fig. 7(a) that the PCA model has little ability to extract nonlinear features as well as poor robustness, which leads to a large number of false alarms before the real failure occurs. Besides, even though the EWMA result, as shown in Fig. 7(b), was used to confirm the alarm, a false alarm was generated at T0. From the comparison experiment shown in Fig. 7(a), (c), (e), (g), and (k), it can be seen that the false alarms of the PCA, KPCA, RPCA, RKPCA, and SR-RKPCA models show a decreasing trend, which is clearly visible and shows that both nuclear improvements and robustness improvements reduce false alarms to some extent. In Fig. 7(b), (d), (f), and (h), false-positive confirmations were issued before the actual failure, and the results of the proposed model, according to Fig. 7(k), showed that the EWMA results of SR-RKPCA did not issue false-positive confirmations. The above results prove the effectiveness of the proposed method in suppressing false detection of false detection. According to Fig. 7(i) and (j), AE has high sensitivity and can also extract trend information effectively. However, compared to RPCA and its extended methods, it is prone to false positives due to its lower robustness.

2) Case 2: Wind Turbine With Bearing Wear Failure: This section introduces the second case to verify the robustness of the proposed SR-RKPCA model to the fault detection of strong interfering signals. The second dataset covers a period of 13 months from March 1, 2014, to April 25, 2015, and the failure was detected by disassembling inspection on November 27, 2014. The data from March 1, 2014, to April 30, 2014, was set as the training data, which is considered normal. Also, the rest of the data are tested as real-time monitoring data to validate the model's performance. In this specific case, the root cause was the large-area wear on the gearbox's rear bearing, which is a slower development rate than other sudden failures, such as cracks and fractures. However, due to the changing working conditions' interference, the collected signal changed abnormally on July 5, 2014, which is not a fault caused by the wind turbine's damage. This is the purpose of the proposed model, which is to obtain trend information while eliminating the interference of occasional anomalies. To implement that, SR-RKPCA and competing methods were trained and practically tested to get the following results.

Fig. 8(a)–(l) shows the fault detection results and the EWMA result of case 2 based on the SR-RKPCA model and the comparison models of PCA, KPCA, RPCA, RKPCA, and AE, where the left side is the original result and the right side is the EWMA result. It can be seen from Fig. 8(a), (c), (e), (g), and (k) that the original data showed abnormal data with large fluctuations on July 5, 2014. However, beginning of October 20, 2014, the combined index increases for a more extended period, demonstrating a trade-off between the prevention of false positives and the early detection of anomalies. In this case, PCA and KPCA showed a low anti-interference ability and issued false alarms. From Fig. 8(b) and (d), both models gave false alarm confirmations in the EWMA results. It can be seen that the EWMA results violate the anomaly limit before it returns to its normal range again. Also, Fig. 8(e) and (g) shows that after the improvement of robustness, the original indicator's detection results based on RPCA and RKPCA are more stable, which leads to fewer false positives. Although it also sent out false positives, no false-positive confirmations were issued in the EWMA results of Fig. 8(f) and (h). SR-RKPCA performed the best, as shown in Fig. 8(k) and (l), the original results rarely give false alarms before the actual alarm, and obviously,

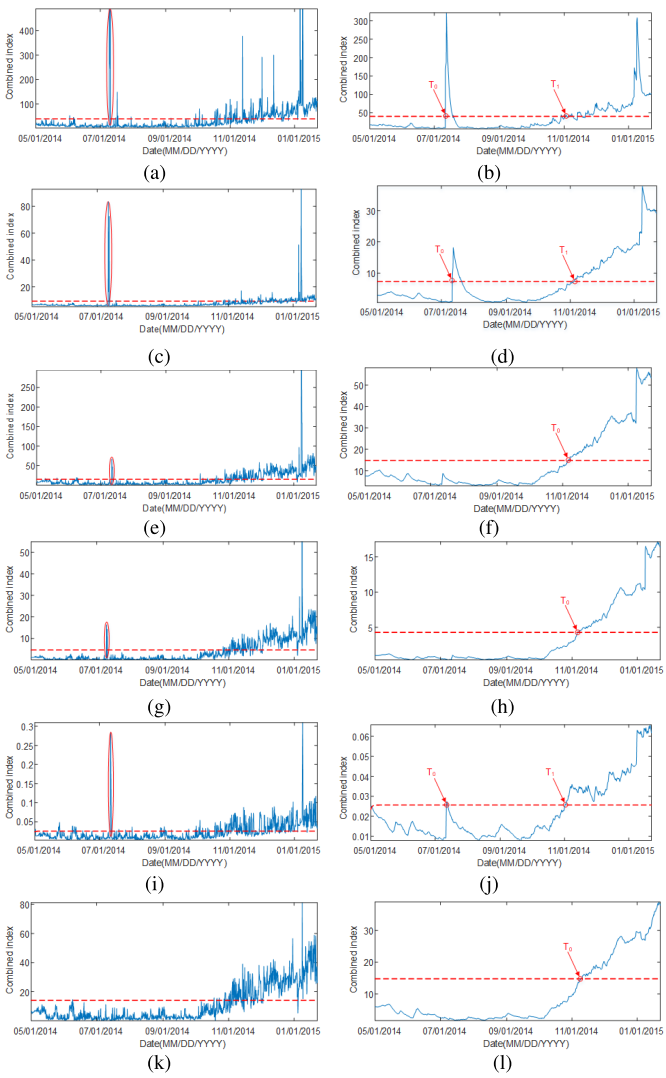

Fig. 8. Fault detection results of case 2 based on: (a) PCA, (b) EWMA of PCA, (c) KPCA, (d) EWMA of KPCA, (e) RPCA, (f) EWMA of RPCA, (g) RKPCA, (h) EWMA of RKPCA, (i) AE, and (j) EWMA of AE.

the results are not confirmed in the EWMA results. Besides, AE presents similar results to case 1, which is more fault-sensitive but less robust, as shown in Fig. 8(i) and (j).

3) SCADA-Based Wind Turbine Gearbox Online Alarm Analysis: The correct steps for wind turbine gearbox operations and maintenance based on this article's method should be as follows. First, online detection of gearbox failure through the comprehensive indicator based on the proposed method; when the alarm is issued for the failure, the cause of the failure can be further determined based on different variables' contribution. Therefore, the process involves the following concerns: the unified standard data interface, the accuracy and efficiency of the detection, and the physical connection between the variables.

Model online applications first need a unified standard data interface, of which the "JavaScript object notation (JSON)" format for data transfer is recommended. Then, the environment must install the model in the monitoring system to ensure the deployment of the saved model after training with offline data. To show each method's performance more

TABLE IV

FALSE ALARM RATE STATISTICS OF THE PROPOSED METHOD AND COMPETING METHODS

Methods		False alarm rate	
		Original data (false alarm points/total points)	EWMA (false alarms /total alarms)
Case 1	PCA	10.14%	1/2
	KPCA	5.69%	1/2
	RPCA	6.73%	1/2
	RKPCA	3.54%	1/2
	AE	9.26%	3/4
	SR-RKPCA	2.70%	0/1
Case 2	PCA	11.88%	1/2
	KPCA	9.73%	1/2
	RPCA	10.67%	0/1
	RKPCA	6.62%	0/1
	AE	10.53%	1/2
	SR-RKPCA	5.51%	0/1

visually, Table IV lists the false alarm rate of each method. The false alarm rate is defined as the number of false alarm points divided by the total number of points before the true alarm. Given the low number of alarms smoothed by EWMA, the false alarm rate of EWMA is directly listed with its number of false alarms and total alarms. From Table IV, the results show that failures of wind turbines can be detected by the SR-RKPCA robustly. Compared with PCA, KPCA-based method has a better effect on the extraction of nonlinear trends, while detection methods based on RPCA and RKPCA possess more advantages in the robustness of the combined index. The fault detection method based on SR-RKPCA proposed in this article considers the feature extraction of the nonlinear trend and robustness, which can detect faults effectively and reduce false alarms. Of course, the default alarm thresholds do not fit all wind turbines, and fine-tuning the thresholds based on the alarm results can further improve the model's performance.

Once the model is deployed, the SCADA system updates data every 10 min, during which the model can compute the detection result with historical data of last few months. To facilitate online applications, it is necessary to consider the computational efficiency of the model. Table V presents a comparison of the time consumption of each method. From Table V, the robustness improvement methods (RPCA, RKPCA, and SR-RKPCA) increase the time consumption, but their time consumption is much lower than that of the kernel improvement. Besides, the SCADA system's analysis interval is 10 min, while the proposed method in this article tests no more than 30 s for a single month of data (no more than 300 s for a year of data), which meets the requirement of an online system update.

When a fault is detected, the analysis of the physical connection between variables will help locate the source. For example, the variables' T^2 contribution [27] of case 1 can be calculated as shown in 9. From this histogram, active

TABLE V
COMPARISON OF COMPUTATIONAL TIME TAKEN BY METHODS

Methods		Time consumption(s)		
		Training	Test	Test time/month
Case 1	PCA	0.17	0.63	0.14
	KPCA	14.37	75.4	16.39
	RPCA	0.33	0.66	0.14
	RKPCA	18.96	76.36	16.6
	AE (GPU)	29.86	1.41	0.31
	SR-RKPCA	22.08	90.71	19.72
Case 2	PCA	0.64	1.39	0.12
	KPCA	38.53	245.56	20.52
	RPCA	0.85	1.66	0.14
	RKPCA	42.07	281.39	23.51
	AE (GPU)	59.69	2.9	0.24
	SR-RKPCA	54.83	296.04	24.74

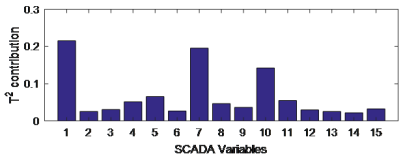

Fig. 9. T^2 contribution of the SCADA variables for the gearbox fault of case 1.

power, gearbox oil temperature, and gearbox cooling water temperature show a high contribution for the fault of case 1. This indicates that high temperatures have occurred in the wind turbine gearbox and have seriously affected the active power. Wind turbine gearboxes are usually cooled by a multistage system. In this study, when the gearbox lubricating oil temperature is below 30 °C, it is cooled using oil. When the oil temperature is between 30 °C and 45 °C, the control valve opens, and the gearbox forms a cooling water circuit through the oil–water radiator. When the oil temperature continues to rise to 60 °C, the cooling water fan is activated to increase the radiator's efficiency through the fan's forced air cooling. When the gearbox oil temperature continues to rise to the safety threshold (usually between 80 °C and 90 °C, the wind turbine will reduce the heat of the gearbox by power-limited operation or shutdown.

However, the influence between the SCADA variables is multiple. For example, wind speed and wind direction mainly reflect the external environmental conditions under which the wind turbine works. When wind speed and wind direction exceed a specific limit, wind wheel speed can be adjusted by pitch angle to stabilize the active power. Since the wind wheel speed and gearbox vibration are correlated, the wind wheel speed can reflect the gearbox condition. Besides, the gearbox oil level and oil pressure can reflect the gearbox oil leakage. Therefore, the analysis of the contribution of variables alone is not enough to locate the fault source. Combining the vibration data from the condition monitoring system (CMS), the proposed method can more accurately determine the fault location.

IV. CONCLUSION

A novel SCADA data-driven approach, referred to as SR-RKPCA, has been proposed for nonlinear process fault detection under discontinuity interference. Through the theoretical and experimental analysis of this study, the following conclusions are summarized.

1) The SR-RKPCA-based fault detection method for wind turbine gearbox has been derived in this article. The PE has been developed to measure the residual matrix components' disturbance level and remove noise components without trend information. The operational SCADA datasets containing nonlinear trends and interference have confirmed the superior performance of the proposed SR-RKPCA-based fault detection approach for the gearbox of wind turbines.
2) Deep learning methods can adaptively extract nonlinear trends. In the comparative experiments, AE shows good fault sensitivity but poor performance in noise robustness. Reducing this method's false positives by improving its robustness is a worthy research topic on which some scholars are working.
3) The brief process and essential technology for online fault detection based on the proposed method are discussed in this article; among them, the alarm threshold's dynamic or adaptive adjustment is an exciting but challenging research topic for a long time.
4) Based on the fault detection results, this article further summarizes the contribution of different SCADA variables to the fault and the interaction between the variables, providing a valuable reference for wind gearbox fault monitoring and diagnosis in the future.

REFERENCES

[1] B. Yang, R. Liu, and X. Chen, "Sparse time-frequency representation for incipient fault diagnosis of wind turbine drive train," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 11, pp. 2616–2627, May 2018.
[2] Y. Qin, S. Xiang, Y. Chai, and H. Chen, "Macroscopic–microscopic attention in LSTM networks based on fusion features for gear remaining life prediction," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10865–10875, Dec. 2020.
[3] R. Chen, X. Huang, L. Yang, X. Xu, X. Zhang, and Y. Zhang, "Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform," *Comput. Ind.*, vol. 106, pp. 48–59, Apr. 2019.
[4] K. Zhang, B. Tang, Y. Qin, and L. Deng, "Fault diagnosis of planetary gearbox using a novel semi-supervised method of multiple association layers networks," *Mech. Syst. Signal Process.*, vol. 131, pp. 243–260, Sep. 2019.
[5] K. Huang, H. Wen, C. Zhou, C. Yang, and W. Gui, "Transfer dictionary learning method for cross-domain multimode process monitoring and fault isolation," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8713–8724, Nov. 2020.
[6] P. Bangalore and M. Patriksson, "Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines," *Renew. Energy*, vol. 115, pp. 521–532, Jan. 2018.
[7] X. Jin, J. Fan, and T. W. S. Chow, "Fault detection for rolling-element bearings using multivariate statistical process control methods," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 9, pp. 3128–3136, Sep. 2019.
[8] M. Tang *et al.*, "Development of an SVR model for the fault diagnosis of large-scale doubly-fed wind turbines using SCADA data," *Energies*, vol. 12, no. 17, p. 3396, Sep. 2019.
[9] M. Schlechtingen and I. Ferreira Santos, "Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1849–1875, Jul. 2011.

[10] Z. Kong, B. Tang, L. Deng, W. Liu, and Y. Han, "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units," *Renew. Energy*, vol. 146, pp. 760–768, Feb. 2020.
[11] G. Helbing and M. Ritter, "Deep learning for fault detection in wind turbines," *Renew. Sustain. Energy Rev.*, vol. 98, pp. 189–198, Dec. 2018.
[12] W. Yang, R. Court, and J. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis," *Renew. Energy*, vol. 53, pp. 365–376, May 2013.
[13] P. B. Dao, W. J. Staszewski, T. Barszcz, and T. Uhl, "Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data," *Renew. Energy*, vol. 116, pp. 107–122, Feb. 2018.
[14] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009.
[15] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 377–386, Jan. 2016.
[16] S. W. Choi and I.-B. Lee, "Nonlinear dynamic process monitoring based on dynamic kernel PCA," *Chem. Eng. Sci.*, vol. 59, no. 24, pp. 5897–5908, Dec. 2004.
[17] J.-H. Cho, J.-M. Lee, S. Wook Choi, D. Lee, and I.-B. Lee, "Fault identification for process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 60, no. 1, pp. 279–288, Jan. 2005.
[18] V. H. Nguyen and J.-C. Golinval, "Fault detection based on kernel principal component analysis," *Eng. Struct.*, vol. 32, no. 11, pp. 3683–3691, Nov. 2010.
[19] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 560–572, Mar. 2018.
[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 1, pp. 1–37, 2009.
[21] J. Fan and T. W. S. Chow, "Exactly robust kernel principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 749–761, Mar. 2020.
[22] Y. Wei, D. Wu, and J. Terpenny, "Robust incipient fault detection of complex systems using data fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9526–9534, Dec. 2020.
[23] M. Debruyne and T. Verdonck, "Robust kernel principal component analysis and classification," *Adv. Data Anal. Classification*, vol. 4, nos. 2–3, pp. 151–167, Sep. 2010.
[24] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, Jul. 1997.
[25] D. M. Mateos, S. Zozor, and F. Olivares, "Contrasting stochasticity with chaos in a permutation Lempel–Ziv complexity-Shannon entropy plane," *Phys. A, Stat. Mech. Appl.*, vol. 554, Sep. 2020, Art. no. 124640.
[26] X. Deng and X. Tian, "A new fault isolation method based on unified contribution plots," in *Proc. 30th Chin. Control Conf.*, 2011, pp. 4280–4285.
[27] Y. Wang, X. Ma, and P. Qian, "Wind turbine fault detection and identification through PCA-based optimal variable selection," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1627–1635, Oct. 2018.
[28] H.-H. Yang, M.-L. Huang, and S.-W. Yang, "Integrating auto-associative neural networks with hotelling t2 control charts for wind turbine fault detection," *Energies*, vol. 8, no. 10, pp. 12100–12115, 2015.
[29] H. H. Yue and S. J. Qin, "Reconstruction-based fault identification using a combined index," *Ind. Eng. Chem. Res.*, vol. 40, no. 20, pp. 4403–4414, Oct. 2001.
[30] F. Chen, Y. Yang, B. Tang, B. Chen, W. Xiao, and X. Zhong, "Performance degradation prediction of mechanical equipment based on optimized multi-kernel relevant vector machine and fuzzy information granulation," *Meas. J. Int. Meas. Confed.*, vol. 151, p. 107116, 2020.
[31] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, Jan. 2016.
[32] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, Jan. 2011.
[33] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.
[34] Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, "A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy," *Mech. Syst. Signal Process.*, vol. 105, pp. 319–337, May 2018.
[35] D. Chakraborty and H. Elzarka, "Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold," *Energy Buildings*, vol. 185, pp. 326–344, Feb. 2019.
[36] A. Bakdi, A. Kouadri, and A. Bensmail, "Fault detection and diagnosis in a cement rotary kiln using PCA with EWMA-based adaptive threshold monitoring scheme," *Control Eng. Pract.*, vol. 66, pp. 64–75, Sep. 2017.

Kai Zhang received the B.E. degree in mechanical engineering from the Taiyuan University of Technology, Taiyuan, China, in 2014, and the M.S. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree in mechanical engineering with Chongqing University, Chongqing, China.

His research interests include deep learning and pattern recognition, semisupervised learning, intelligent fault diagnosis, and fault location of complex mechanical systems.

Baoping Tang received the M.S. and Ph.D. degrees in mechanical engineering from Chongqing University, Chongqing, China, in 1996 and 2003, respectively.

He is currently a Professor and a Ph.D. Supervisor with the College of Mechanical Engineering, Chongqing University. More than 150 papers have been published in his research career. His main research interests include wireless sensor networks, mechanical and electrical equipment security service and life prediction, measurement technology, and instruments.

Dr. Tang was a recipient of the National Scientific and Technological Progress 2nd Prize of China in 2004 and the National Invention 2nd Prize of China in 2015.

Lei Deng received the M.S. and Ph.D. degrees from Chongqing University, Chongqing, China, in 2001 and 2010, respectively.

She is currently a Vice Professor with the College of Mechanical Engineering, Chongqing University. Her main research interests include logistic and supply chain management, wireless sensor networks research and application, and equipment health management.

Xiaoxia Yu was born in Chongqing, China, in June 1993. He received the B.E. and M.S. degrees in vehicle engineering from the Chongqing University of Technology, Chongqing, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree in mechanical engineering, Chongqing University, Chongqing.

His research interests include deep learning, pattern recognition, and intelligent fault diagnosis.