Towards Smart Condition Monitoring of Rotary Machines: An Optimized Probabilistic Signal Reconstruction Methodology for Fault Prediction with Multi-source Uncertainties

Xiaomo Jiang¹, Member, IEEE, Weijian Tang¹, Haixin Zhao¹, and Xueyu Cheng²

¹Provincial Key Lab of Digital Twin for Industrial Equipment, State Key Lab of Structural Analysis for Industrial Equipment, Research Institute of Carbon Neutrality and School of Energy and Power Engineering, Dalian University of Technology, Dalian 116024, China
²College of Arts and Sciences, Clayton State University, Morrow, GA 30260, USA

Corresponding author: Xiaomo Jiang (e-mail: xiaomojiang2019@dlut.edu.cn).

This work was supported in part by the Department of Science and Technology of Liaoning Province under Grant No. 2020H1/10100018 and XLYC1902096.

ABSTRACT Faults in the critical components of a large rotatory machine often result in unplanned breakdowns, leading to a significant loss of property and life. Condition monitoring, as a key component in the smart maintenance of industrial equipment, has become a promising tool for providing automatic early alerting of potential damage to critical components, thus reducing potential outages and improving system safety and reliability while lowering maintenance costs. This is still a very challenging topic in various industrial fields because of data imperfection and multivariate correlation, as well as the variation in faults and components in different machines. This paper presents an optimized probabilistic signal reconstruction methodology to address these challenges in the fault prediction of rotatory machines using multivariate vibration signals. Three signal reconstruction methods, that is, Bayesian wavelet multiscale decomposition, probabilistic principal component analysis, and auto-associative kernel regression, were seamlessly integrated to address the noise, high dimensionality, and correlation in the sensed multivariate vibration data for accurate fault prediction. The bandwidth parameter in the auto-associative kernel regression approach was optimized to represent the health status of the rotatory machine. The obtained model was further utilized to predict the responses under unknown conditions. The alerting threshold, based on the squared mean errors of the predicted and measured time series, was automatically adjusted using a rolling window strategy and then employed to predict the possible fault. Finally, the validity, feasibility and generalization of the proposed methodology are illustrated by applying two cases of different rotating machines: centrifugal compressor and gas turbine.

INDEX TERMS Bayesian wavelets, similarity-based learning, OAKR, signal reconstruction, PPCA, fault prediction, smart maintenance

I. INTRODUCTION

Faults in large-scale rotatory machines such as gas turbines, steam turbines, or compressors usually result in unplanned outages and even a huge loss of property and life in the fields of power generation, oil and gas, and petrochemical industries. Effectively monitoring the operation state of a rotating machine and ensuring its long-term safety and reliability have recently attracted increasing attention in related industrial fields. Real-time condition monitoring has become a powerful tool to provide an early warning of a potential equipment fault and effectively prevent it from turning into a disastrous event leading to huge loss of properties and even life, thus improving the safety and reliability of the rotating machinery, avoiding unplanned breakdowns due to potential faults in the equipment, and lowering its maintenance costs. Frequently used condition monitoring methods for rotating machines mainly include two types: physics model-based [1] and signal processing-driven [2]. The former intends to stimulate the equipment state by modelling the failure mechanism, which is particularly feasible for relatively simple linear systems. The latter takes
full advantage of emerging digital technologies such as the Internet of Things (IoT), big data, artificial intelligence (AI), and cloud computing to capture equipment behaviors from the numerous operation data collected from a rotating machine. It has been broadly applied to various complex equipment such as gas turbines and steam turbines. In the past decades, a wide spectrum of data-driven predictive analytics methods, such as time series forecasting, machine learning, and artificial neural network models, have been developed to predict faults in machine condition monitoring [3]-[5]. These methods are generally composed of two main steps: 1) establishment of a high-fidelity predictive model to produce the system response, and 2) determination of a decision threshold to produce alarms when the system response deviates from the actual vibration measurement to a certain degree. In addition to the existing uncertainties in sensor data, both model establishment and threshold determination contain uncertainties, which would further impact the fault prediction accuracy of a rotatory machine. Therefore, it is of paramount importance to accurately predict faults for machine condition monitoring, considering various uncertainties.

Recently, the auto-associative kernel regression (AAKR) method was developed as a similarity-based model (SBM) for condition monitoring and fault alerting in large-scale industrial equipment [6]-[13]. This approach uses multivariate historical data collected under normal conditions to establish a system identification model that represents the health status of a physical system. The obtained model was then used to predict the multivariate responses of the system under unknown conditions. A health index calculated from the difference between the predicted responses and actual measurements was then used to quantitatively assess the status of the system. This approach has demonstrated high accuracy, flexibility, and generalization with broad applications in various types of industrial equipment [6]-[13]. It can be effectively applied not only for non-rotating equipment, such as boilers in a coal plant [9] and gear boxes in a wind turbine [10], but also for rotating machines, such as bearings in engines [11] and blades in compressors or steam turbines [12].

Unlike other parametric regression models in machine learning or nonparametric AI methods, the AAKR approach does not require any prior knowledge of the system or components under investigation. It provides a generalized flexible nonparametric system identification approach for fault prediction in different multivariate scenarios of a complicated physical system. However, this approach has three main limitations. First, it is highly sensitive to the uncertainties existing in sensor data, such that it may produce a number of false alarms and even missing events. Second, the bandwidth parameter in the kernel function of the AAKR must be tuned from historical sensor data. Improper selection of the bandwidth value may result in inaccurate fault predictions. Finally, subjectivity and variation in the general threshold determination play a key role in the fault prediction accuracy of large-scale rotatory machines. Recently, some researchers have developed a hybrid model to partially address the above-mentioned issues by combining signal-processing techniques with the AAKR method [7][9][11][13]. For instance, Maio et al. [7] combined a correlation analysis and a generic algorithm to classify the signals and then applied AAKR for fault prediction using the classified signals. Yu et al. [9] averaged the prediction results obtained from multiple AAKR models to alleviate the effect of outcome variations on the prediction accuracy.

Brandsæter et al. [11] combined the k-means clustering and AAKR estimation approaches to improve the computational efficiency of the method. Baraldi et al. [13] revised the weights of the AAKR model to reduce the correlations between multiple variables. However, none of the existing methods have comprehensively addressed their sensitivity to uncertainties in data, modeling, and thresholding. It should be noted that the determination of the alarm threshold also plays a critical role in the fault prediction accuracy of the AAKR approach. Several techniques such as the sequential probability ratio test (SPRT) [7][12], likelihood ratio test [8], and empirical estimation [8] are commonly used to determine the alarm threshold level. These methods use the statistical characteristics of a unit during normal operation as health indicators to monitor the unit operation status in real time.

In this study, a discrete wavelet packet transform is first employed to decompose a set of raw vibration signals into multiresolution approximations and details through wavelet functions in the time-frequency domain. Bayesian hypothesis testing was then applied to determine whether noise existed in each decomposed coefficient series. The adept combination of multiresolution wavelet analysis and probabilistic Bayesian assessment [14] can trade-off the under-denoising and over-denoising of raw signals, outperforming the traditional soft or hard thresholding approach in the wavelet fashion. Furthermore, this paper presents an enhanced optimized AAKR (OAKR) method that integrates advanced signal processing, probabilistic principal component analysis, kernel parameter optimization, and adaptive thresholding to address the aforementioned issues of the traditional method for machine fault prediction using multivariate vibration signals for the smart maintenance of rotatory machines.

In the following sections, three signal reconstruction methods—Bayesian wavelet denoising, probabilistic principal component analysis, and auto-associative kernel regression (AAKR)—are introduced in the methodology for data preprocessing and modeling. The first two preprocessing methods were employed to remove potential noise from the raw vibration data and reduce the dimensions of the multivariate signals, thus improving fault prediction and diagnosis accuracy. The simplex method is employed to obtain the optimal bandwidth of the kernel function in conventional AAKR, resulting in an optimized similarity-based estimation model. A generalized implementation procedure was developed to automate the application of the proposed methodology for machine fault prediction and diagnosis. Vibration data and events collected from two real-
Principal component analysis (PCA) is a commonly used analysis based on eigenvalue decomposition. The yielding eigenvectors and eigenvalues represent the variations in the principal components and weights of the raw variables, respectively. The PCA approach converts high-dimensional correlated multivariate data into low-dimensional uncorrelated principal components, thus lowering the dimensionality and extracting the main features from multivariate data. The traditional PCA is a deterministic multivariate analysis approach that cannot address data uncertainties. Tipping and Bishop [17] proposed a probabilistic principal component analysis (PPCA) approach to handle uncertainties in multivariate data. Unlike factor analysis, which assumes varying noise variances, PPCA assumes the same noise variance on the raw signal, which largely simplifies the parameter estimation and signal reconstruction. This study developed a PPCA approach to further alleviate the impact of data uncertainties on the fault prediction accuracy of rotatory machines.

Given a d-dimensional multivariate observation matrix $X = \{x_1, x_2, ..., x_d\}$ where each variable has $N$ data points, that is, $x_p = [x_1, x_2, ..., x_d]^T \in \mathbb{R}^{N \times d}$, where $x_p$ denotes the cleansed time series obtained by the Bayesian wavelet thresholding method represented by Eq. (1). Let $X_i = \{\tilde{x}_{1,i}, \tilde{x}_{2,i}, ..., \tilde{x}_{d,i}\}$ be the i-th data point for the d-dimension variables, which can be expressed as

$$X_i = W Y_i + U + E_i \quad i = 1, 2, ..., N \quad (4)$$

where $Y_i$ are the latent random variables representing $q(q \leq d)$ principal components, which are assumed to be a multivariate Gaussian distribution, i.e., $Y_i \sim N(0, I)$, where the bold number $0$ represents a $d$-dimension vector of zeros, $I$ is the unit matrix, $W$ is the $d \times q$ matrix representing the relationship between $X_i$ and $Y_i$, $U$ is the mean vector of the $d$-dimension multivariate data, i.e., $U = \frac{1}{N} \sum_{i=1}^{N} X_i$, and $E_i$ is the white noise item in the data $X_i$, i.e., $E_i \sim N(0, \sigma^2 I)$, where $\sigma^2$ is the variation of the noise calculated from the raw data.

Equation (4) can be expressed in a distribution form as follows,

$$X_i \sim N(U, W W^T + \sigma^2 I) \quad (5)$$

Thus, the conditional distribution of $X_i$ on the latent variable $Y_i$ is obtained by

$$X_i|Y_i \sim N(W Y_i + U, \sigma^2 I) \quad (6)$$

According to the conditional probability formulation, the conditional distribution of $Y_i$ for data $X_i$ can be derived as follows:

$$Y_i|X_i \sim N(M^{-1}W^T(X_i - U), \sigma^2 M^{-1}) \quad (7)$$

where $M = W^T W + \sigma^2 I$. As $Y_i$ is a vector of $p$-dimensional random variables, its expectation conditional on data $X_i$, $M^{-1}W^T(X_i - U)$, is the desirable PPCA result, denoted by $\Theta$, that is,

$$\Theta = M^{-1}W^T(X_i - U) \quad (8)$$

world rotating machines, one centrifugal compressor, and one heavy-duty gas turbine, were employed to demonstrate the effectiveness, feasibility, and generalization of the proposed methodology and procedure. Finally, conclusions are presented.

II. Bayesian OAKR Method

A. Bayesian wavelet signal reconstruction

Wavelet multiresolution analysis works like a mathematical microscope that is broadly used to exploit details in time-series signals [15][16]. Unlike the traditional Fourier transform, where only frequency features are extracted from a signal, wavelet analysis decomposes the raw signal into different levels of coefficients in both the time and frequency domains through a series of wavelet functions. In particular, the discrete wavelet packet transform (DWPT) simultaneously splits the obtained approximation and detail coefficients from one level to the next, thus decomposing the raw signals into multiresolution levels of detail for subsequent analysis.

Given a time-series vibration signal with $N$ points for a sensor variable in a rotatory machine, $x_i(i = 1, 2, ..., N)$, the DWPT approach decomposes the signal into a set of scaling coefficients $s_j(k)$ and wavelet coefficients $w_j(k)$ simultaneously. Note that parameter $i$ represents the time point in this formulation. The time series can be reconstructed by summing the discrete decomposition coefficients, as follows:

$$x_i = \sum_{k \in Z} \sum_{j \in Z} s_j(k) \varphi_j(k) + \sum_{k \in Z} w_j(k) \psi_j(k)$$

where $j$ is the number of decomposition levels, $k$ is the coefficient point, and $\varphi_j(k)$ and $\psi_j(k)$ are the $k$-th scaling and wavelet functions, respectively, at time point $i$ of the $j$-th level. Jiang et al. [14] provided details on the DWPT decomposition and Bayesian thresholding for each decomposed coefficient. The denoising performance was evaluated quantitatively using the signal-to-noise ratio (SNR) and root mean squared error (RMSE) metrics. The SNR metric is calculated by the ratio of the summation of the squared signal to that of the squared noise and then converted into decibels (dB) by taking the logarithm on the basis of 10, given by

$$SNR = 10 \log_{10} \left( \frac{\sum_{i=1}^{N} |x_i|^2}{\sum_{i=1}^{N} |\tilde{x}_i - x_i|^2} \right) \quad (2)$$

A smaller SNR value indicates more noise in raw signals. The RMSE metric indicates the difference between the raw signal and denoised signal, expressed as

$$RMSE = \frac{1}{N} \sum_{i=1}^{N} (\tilde{x}_i - x_i)^2 \quad (3)$$

The larger RMSE value indicates the more noise in the raw signals.

B. Probabilistic Principal Component Analysis

Principal component analysis (PCA) is a commonly used nonparametric statistical approach for multivariate data analysis based on eigenvalue decomposition. The yielding eigenvectors and eigenvalues represent the variations in the principal components and weights of the raw variables, respectively. The PCA approach converts high-dimensional correlated multivariate data into low-dimensional uncorrelated principal components, thus lowering the dimensionality and extracting the main features from multivariate data. The traditional PCA is a deterministic multivariate analysis approach that cannot address data uncertainties. Tipping and Bishop [17] proposed a probabilistic principal component analysis (PPCA) approach to handle uncertainties in multivariate data. Unlike factor analysis, which assumes varying noise variances, PPCA assumes the same noise variance on the raw signal, which largely simplifies the parameter estimation and signal reconstruction. This study developed a PPCA approach to further alleviate the impact of data uncertainties on the fault prediction accuracy of rotatory machines.
Note that in (7), the parameters $\mathbf{W}$ and $\sigma^2$ must be estimated. This study employed the expectation-maximization optimization approach to estimate the PPCA parameters. The likelihood function $\mathcal{L}$ can be expressed as

$$
\mathcal{L} = -\frac{N}{2} \{d \ln(2\pi) + \ln|\mathbf{M}| + \text{tr}(\mathbf{M}^{-1}\mathbf{S})\}
$$

where $\mathbf{S} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i - \mathbf{U})(\mathbf{X}_i - \mathbf{U})^T$ is the variance of the multivariate matrix $\mathbf{X}$ and $\text{tr}$ is the trace of the matrix. The optimal parameters $\mathbf{W}$ and $\sigma^2$ were obtained by maximizing the likelihood function $\mathcal{L}$ in (9). Refer to [17] for further details on the parameter estimation of the PPCA approach.

Similar to the classic PCA approach, the loadings of the PPCA, denoted by $\mathbf{L}$, are employed to identify the key factors contributing to a principal component, which are defined as the correlations between the original signal and principal components given by

$$
\mathbf{L} = \text{cov}(\mathbf{x}, \theta) = \frac{\mathbf{X} \mathbf{W}}{\sqrt{\text{tr}^2}}
$$

The loadings $\mathbf{L}$ are computed as the cross-variance matrix between the original signal and the principal components (PCs), that is, the product of the PPCA transfer coefficients $\mathbf{W}$ and the variance of the original signal divided by the square root of $N-1$.

### C. Optimized AAKR Approach

The auto-associative kernel regression (AAKR) approach is a similarity-based nonparametric modeling technique that estimates the system response based on the similarity between the measurement and memory vectors. Compared to other nonparametric approaches, such as neural networks, this method can be easily trained with historical measurements, independent of fault modes and equipment types. As a result, AAKR provides a robust multivariate system identification approach for condition monitoring and fault prediction of industrial equipment.

Instead of using a multivariate sensor time series, AAKR was trained using the principal components obtained from the PPCA. The memory vector for the model training is stored in matrix $\mathbf{\Theta}$ of (8), where $\theta_{ij}$ is the $i$-th vector of the $j$-th principal variable. For the $n_m$-th memory vector, matrix $\mathbf{\Theta}$ can be expressed as

$$
\mathbf{\Theta} = \begin{bmatrix}
\theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,p} \\
\theta_{2,1} & \theta_{2,2} & \cdots & \theta_{2,p} \\
\vdots & \vdots & \ddots & \vdots \\
\theta_{n_m,1} & \theta_{n_m,2} & \cdots & \theta_{n_m,p}
\end{bmatrix}
$$

The $1 \times p$ monitoring response vector $\mathbf{\Theta}$ of the machine is expressed as

$$
\mathbf{\Theta} = [\theta_1 \theta_2 \cdots \theta_p]
$$

The model prediction can be obtained by averaging each memory vector in the matrix $\mathbf{\Theta}_{n_m,p}$ the weights of which are estimated using healthy data. This study proposes an optimized AAKR approach (OAKR) that consists of the following four steps:

1. First, calculate the commonly used Euclidean distance between the monitoring vector $\mathbf{\Theta}$ and each memory vector $\mathbf{\Theta}_i$ to produce an $n_m \times 1$ distance vector $\mathbf{d}_i$, given by

$$
d_i(\mathbf{\Theta}_i, \mathbf{\Theta}) = \sqrt{(\Delta \theta_{1,i})^2 + (\Delta \theta_{2,i})^2 + \cdots + (\Delta \theta_{p,i})^2}
$$

Second, we obtained the $n_m \times 1$ weight vector $\mathbf{w}$ by fitting the Gaussian kernel function as follows:

$$
\mathbf{w} = K_h(\mathbf{d}) = \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{d^2}{2h^2}}
$$

where $h$ is the bandwidth of the kernel function that determines the smoothing level of the function. A smaller $h$ can display more details in the data but fails to yield a smooth tail of the function, whereas a larger $h$ can produce relatively smoother transactions but lose details in the system identification, resulting in more false alarms in the condition monitoring of rotatory machines. Therefore, the selection of bandwidth $h$ plays a critical role in the fault prediction accuracy in the condition monitoring of industrial equipment.

Third, the optimal bandwidth is obtained by automatically minimizing the mean squared error of the model prediction using a set of new data and the Nelder-Mead optimization algorithm, as described subsequently, thus reducing the false alarm rates of equipment condition monitoring.

Finally, we estimate the response $\hat{\mathbf{\Theta}}$ using the weight matrix $\mathbf{w}$ and monitoring vector $\mathbf{\Theta}_i$, as follows:

$$
\hat{\mathbf{\Theta}} = \frac{\sum_{i=1}^{n_m} w_i \mathbf{\Theta}_i}{\sum_{i=1}^{n_m} w_i}
$$

It should be mentioned that this OAKR approach is related to the measured data only, and is independent of the fault and equipment under monitoring. Consequently, this approach provides a powerful tool for monitoring the conditions of industrial equipment.

### D. Nelder-Mead Optimization

In this study, the AAKR method was improved by automatically selecting the optimal bandwidth to minimize the model prediction error. The Nelder-Mead (NM) simplex algorithm was proposed to achieve this purpose owing to its simplicity and effectiveness for online implementation [18][19]. The NM algorithm is a heuristic optimization method proposed by Nelder and Mead (1965) [20][21], which does not require any mathematical derivative of the objective function but is very effective in dealing with nonlinear and multidimensional problems [19]. The algorithm does not require mathematical derivatives, but a few function values at each iteration, which makes it computationally very efficient and broadly applicable for online implementation. In addition, the algorithm generally converges to a global optimal value within a few iterations. Therefore, it is especially popular for optimization problems, where the derivative of an objective function is hardly obtained explicitly.
According to Lagarias et al. (1998) [18], the NM algorithm requires predefined coefficients for four operations: reflection ($\rho$), expansion ($\chi$), contraction ($\gamma$), and shrinking ($\alpha$). In this study, the commonly used values $\rho = 1$, $\chi = 2$, and $\gamma = \alpha = 0.5$ are taken, and five steps in one iteration are employed to optimize the bandwidth parameter $h$. The objective function is the mean square error (MSE) between the prediction output and measurement data, calculated as follows:

$$f(h) = \frac{1}{N} \sum_{n=1}^{N} \left( \bar{\theta} - \bar{\theta}(h) \right)^2$$

where $N$ is the number of samples, $M$ is the number of model variables, and $\theta_m$ and $\bar{\theta}_m$ are the measurement data and model output of the $m$-th variable, respectively. For brevity, the details of the optimization method are omitted in this study. Refer to Tang et al. [22] for details regarding this algorithm applied to the OAKR method.

E. Model Performance Indicator

1) Mean Square Error (MSE)
The MSE value was employed in this study to quantitatively evaluate the model validity and system health status. In model training, the MSE value obtained using Eq. (16), approaching zero indicates a well-trained model. The trained model is used to predict the response of the system or component under unknown conditions. When the obtained MSE value exceeds a preset threshold, an alarm is generated to alert the large deviation between the measurement data and the model prediction. In this study, the alert threshold was determined using model predictions under healthy conditions.

2) R-square Metric
The R-square is defined as follows

$$R\text{-square} = \frac{SSR}{SST}$$

where $SSR = \sum_{m=1}^{M} (\bar{\theta}_m - \bar{\theta})^2$ and $SST = \sum_{m=1}^{M} (\theta_m - \bar{\theta})^2$, and $\bar{\theta}$ is the mean value of the prediction data. The R-squared value was used to evaluate model-fitting performance. A value approaching one indicates a well-fitted model.

III. Implementation Procedure
The improved OAKR method combining Bayesian wavelet thresholding and probabilistic principal component analysis consists of 14 main steps, which is able to accurately predict the fault and identify the main variables contributing to the fault, as shown in Fig. 1 and briefly described as follows.

1) Sensor data collection. Multivariate sensor time series data collected from a rotatory machine are usually transmitted to local or cloud servers. High-frequency historical data of the vibration variables were extracted from the servers for offline modeling of the enhanced OAKR method for machine fault prediction.

2) Frequency feature extraction. The fast Fourier transform was conducted on the high-frequency vibration data to extract frequency features such as frequency, peak, angle, and magnitude values for subsequent data analysis and modeling.

3) Data imputation. The sensor data collected from a rotatory machine usually contain missing values due to sensor faults, data transmission, data storage, etc.. In this study, missing or infinite values for the critical variables were imputed using the forward or backward of the most possible neighbor points. This approach did not change the pattern of the raw data.

4) Outliers removing. Outliers in feature data may exist because of sensor faults, machine shutdowns, or startups. In this study, the feature data points during machine shutdown were removed by evaluating a machine rotating speed that was lower than a small positive value.

5) Data normalization. The different magnitudes and dimensions of various frequency features affect the fault prediction accuracy in the condition monitoring of a rotatory machine. In this study, the multivariate datasets were normalized to those with a mean of zero and variance of one, thus eliminating the magnitude effect of different variables.

6) Bayesian wavelet packet thresholding. The Bayesian discrete wavelet packet transform detailed in Jiang et al. [14] and Tang et al. [22] was employed to threshold the noisy feature data in this study.

7) Denoising evaluation. The denoising effectiveness is evaluated visually by a graphical comparison between the raw and cleansed signals and spectra plots, and quantitatively by using the SNR and RMSE metrics in (2) and (3). The Welch method [23] was employed in this study as a periodogram spectrum analysis approach to graphically evaluate denoising performance in the frequency domain.

8) Dimension reduction. The previously described PPCA approach was employed to extract the main information from the multivariate correlated data while reducing the data dimension in a vertical fashion. In this study, the lower dimension of the variables was obtained by.

FIGURE 1. Flowchart of implementing the proposed methodology for anomaly alarming and fault diagnosis of turbomachines
retaining the principal components with 95% variance in the raw data. The principal components were then obtained from the PPCA and used for subsequent OAKR modelling. It should be mentioned that the obtained PPCA coefficients in the model training were saved and used later for model prediction on the new dataset.

9) Similarity modeling. The principal components of the PCA were used to establish the AAKR model weights. In this study, 85% of the principal component data under a healthy status were used to train the model weights, of which 70% of the data were used to form the memory matrix and obtain the AAKR model to represent the system in a healthy state.

10) Band-width parameter optimization. The remaining 15% of the principal components for modelling were used to obtain the optimal bandwidth of the kernel function using the Nelder–Mead simplex algorithm, as described by Tang et al. [22], thus yielding an OAKR model.

11) Model validation. The remaining 15% of the principal component data at a healthy status was used to test the trained model. Both the MSE and R-square metrics were used to evaluate the model performance.

12) Model prediction. Given a set of time-series vibration data collected from the machine under unknown conditions, the methods described in steps 2) to 7) are applied to clean the data. The same PPCA coefficients obtained from the training data were used to reduce the dimensions of the multivariate data. The yielded principal components were used as inputs to the trained OAKR model to produce the response of the system under healthy conditions.

13) Fault alarming. A rolling window strategy is employed to automatically judge whether the machine has a fault by evaluating the mean square of error (MSE) between the model output and actual principal components in comparison to the preset threshold. Based on the data resolution, the rolling window size may be taken to be one minute, one hour or one day. When the MSE value continues to change in five continuous windows and exceeds the preset threshold, the machine is judged to have an anomaly with a fault, thus yielding an alarm. In this study, the threshold of $k\sqrt{MSE}$ obtained from the healthy data in model validation was used for fault detection. According to Baraldi et al. [24], $k=4$ was initiated in the examples presented in this study. Owing to the variation in different devices, the $k$ value may or may not be adjusted in practical applications to obtain the lowest false positive rate.

14) Identification of influencing factors. The PPCA loadings obtained by (10) are further employed to identify the sensor variables that mainly influence faults, which will facilitate the root cause analysis of faults in the smart maintenance of turbomachines.

IV. Illustrative Examples

Sensor data and events collected from two different real-world turbomachines, a centrifugal compressor and gas turbine, were employed in this study to illustrate the effectiveness and feasibility of the proposed methodology and implementation procedure shown in Fig. 1.

Example 1: Centrifugal compressor

1.A. Data

The data used in this example was collected from a centrifugal compressor in a petrochemical factory. Figure 2 shows a schematic of the compressor (Fig 2a) and its actual broken impeller (Fig 2b). The impellers were broken owing to wheel cracks resulting from the damaged filtering net on March 7, 2019. Operating data from December 28, 2018, to April 1, 2019, are used in this example to demonstrate the effectiveness of the proposed methodology.

![FIGURE 2. Illustrative graph of a centrifugal compressor (Example 1): a) sketch section with vibration sensor location and b) broken impellers](image)

![FIGURE 3. An example of time series for rotating speed (Example 1)](image)

Figure 3 shows an example plot for the rotating speed of the compressor over time, which indicates the varying speed of the machine with a maximum speed of up to 6000RPM, producing high-frequency time-series data. Note that in this study, the data prior to March 7, 2019, were considered to represent the healthy conditions of the machine. Sensed time-series data of 14 vibration variables were used in this study. For each variable, 11 frequency features were extracted from the 5-minute interval vibration sensor data, resulting in 154 high-dimensional features at each interval. The frequency features associated with the vibration variables, including the sensor location, direction, and features, are summarized in
Table I. Obviously, it is very difficult to directly apply the enhanced OAKR method to fault prediction using a superhigh-dimension feature data matrix.

| TABLE I. FEATURE VARIABLES (EXAMPLE 1) |
|----------------------------------------|
| Vibration sensor location | Direction | Features                                      |
| Compressor connection end | x, y       | GAP Electric Voltage, Effectiveness, Frequency, Peak, 0.5F, 1F Angle, 1F Magnitude, 2F Angle, 2F Magnitude, Residuals, Speed |
| Compressor non-connection end |                       |
| Steam turbine connection end |                       |
| Steam turbine non-connection end |                       |
| Compressor axis non-connection end | A, B, C               |
| Steam turbine axis non-connection end |                       |

1.B. Frequency feature extraction
A fast Fourier transform was applied to the wave samples to produce frequency and magnitude values, as shown in Fig. 4. The frequency features including one base frequency of 96.15 Hz and two times of base frequency (192.3 Hz), as well as the corresponding magnitude and angle values can be readily obtained from the frequency analysis.

1.C. Data preprocessing
For demonstration purposes, 77 feature features obtained from seven sensors in the compressor were used for data preprocessing. Upward and downward data-fill methods were used to process the missing or infinite values in the feature series. The feature data for all variables were normalized using the max-min method to eliminate the amplitude effects of the various variables. As an example, Figure 5 shows the time series trends of the four feature variables for the raw data (Fig. 5a) and the normalized data (Fig. 5b). It can be observed that the magnitude of the original feature data (Fig. 5a) varies significantly, whereas the normalized data (Fig. 5b) display basically in a stable range.

1.D. Bayesian wavelet thresholding
The Bayesian wavelet packet denoising approach described previously was employed to remove noise from feature data. Using a trial-and-error approach, the three-level DWPT decomposition with the Daubechies function of eight mother wavelets was found to be sufficient for removing noise from the signals. For instance, given the feature sequence of variable CC_Xend_one_freq_y, the SNR and RMSE values obtained from (2) and (3) are 20 dB and 0.1, respectively, implying that the signal is contaminated to a certain degree. Figure 6 shows a comparison of the time series (Fig. 6a) and power spectral density (PSD) (Fig. 6b), where the solid and dashed lines represent the raw and denoised feature sequences, respectively. The data series became much smoother after denoising (Fig. 6a). This denoising method effectively removes the high-frequency noisy components via multiscale wavelet decomposition of the raw data while preserving its overall trend, as shown by the dashed dotted lines in Fig. 6a and Fig. 6b.

1.E. PPCA feature extraction
As an example, Figure 7 shows the relationship diagram of four features randomly selected from the 77 features after denoising. It can be observed that three of the four features have a highly linear relationship and are highly correlated with each other. If they are directly used in modeling, model complexity will increase, parameter estimation accuracy will be affected, and calculation cost will increase significantly. Therefore, principal component analysis was applied to all denoised feature data matrices to retain 95% of the original
value information, yielding a total of 11 principal components. PPCA with the same dimension reduction was then applied to obtain the weights and low-dimension principal components. Both the individual interpretation variance and the cumulative variance of PPCA are obtained by the ratio of the variance of each principal component to the total one after dimensionality reduction, as shown in Fig. 8. The first principal component represented over 61% of the raw data. The PPCA weights obtained from the training dataset were saved and subsequently applied to unused data for model prediction, whereas the obtained principal components were used to establish the enhanced OAKR model.

**FIGURE 6.** Data denoising results for variable CC_Xend_one_freq_y (Example 1): a) time series plots, and b) PSD graphs

**1.F. OAKR model establishment and validation**

The obtained 11 PPCA principal components (PCs) were used to establish the OAKR model, as summarized in Table II. The three-month feature data at one-hour intervals were divided into four groups: model training, bandwidth optimization, model testing, and prediction. The 850 training PC datasets were used to produce the memory matrix, while 183 PC sets were used to optimize the bandwidth in the kernel density function of the OAKR model using the simplex method described previously. The optimal bandwidth was obtained through 12 simple iterations. Figure 9 shows the MSE curve versus the bandwidth with the optimal value of $h_{opt} = 1.169$ and the minimum MSE of $f(h_{opt}) = 0.58$, obtained using (15). The obtained model was tested using a set of new data and then used for fault prediction. An MSE of 1.587 and a thresholding coefficient of 4 were obtained from the testing data. The threshold of $4\sqrt{MSE} = 5.038$ was then used for fault prediction in this example.

**FIGURE 7.** Diagram of relationships between features (Example 1)

**FIGURE 8.** Explained variance in PPCA (Example 1)

**FIGURE 9.** Curve of MSE versus kernel bandwidth $h$ (Example 1)

**1.G. Fault prediction**

The trained model was used to predict the system response of the centrifugal compressor for data from February 15, 2019 to
April 1, 2019, as shown in Table II. Raw data were preprocessed and cleansed using the Bayesian DWPT thresholding approach. The PPCA coefficient matrix was similarly applied to the cleaned data to yield 11 principal components, as previously described. The trained OAKR model was then used to estimate the response of the system based on principal components.

1.H. Critical factors identification

To further analyze the variables that directly affect component damage. Figure 12 shows the loadings of the two PCs (PC1 and PC2), which have a large impact on failure, obtained by Eq. (10), indicating the contribution of each variable to the PC. The results showed that the damage was represented by multiple features, as expected. Considering the high correlation between the original features, it is reasonable to assume that the component fault causes fluctuations in multiple principal features, which are identified as the main factors impacting component reliability.

As an example, Figure 10 shows the prediction results of four selected principal components, PC1 (Fig. 10a), PC2 (Fig. 10b), PC4 (Fig. 10c), and PC5 (Fig. 10d), which play key roles in fault detection. The predicted values of the four PCs are observed to deviate greatly from the actual values in the period of time before the occurrence of the fault, which captures the early trend of the fault and provides the theoretical basis for the anomaly alarm in the smart condition monitoring of the turbomachine.

The moving window strategy was used to calculate the mean square error (MSE) of the predicted value of each PC and generate an alarm if the MSE value exceeded a predefined threshold. As shown in Fig. 11, the obtained MSE value exceeded the threshold and triggered an alarm on February 19, 2019; thus, 16 days earlier than the actual time of failure (March 7, 2019), providing a powerful tool for scheduling outage maintenance of the machine ahead of its unplanned breakdown owing to significant damage.

| Period       | Data points | MSE   | R-square |
|--------------|-------------|-------|----------|
| Training     | 2018/12/28-2019/02/01 | 850   | 0.022    | 0.973    |
| Optimization | 2019/02/01-2019/02/08 | 183   | 0.580    | 0.346    |
| Testing      | 2019/02/08-2019/02/15 | 183   | 1.587    | 0.130    |
| Prediction   | 2019/02/15-2019/04/01 | 1226  | 1.902    | 0.103    |

![FIGURE 10. OAKR prediction result of PPCA components (Example 1): a)PC1, b)PC2, c)PC10, and d)PC11](image)

![FIGURE 11. Averaged MSE alarming curve for overall condition evaluation of the component (Example 1)](image)

![FIGURE 12. Contribution loadings for PC1 & PC2 (Example 1)](image)

Example 2: Gas turbine

2.A Data

The proposed methodology and procedure are further applied to predict the stage 1 bucket crack in a heavy-duty gas turbine to illustrate its generalization and robustness. Industrial heavy-duty gas turbines are commonly used to generate power by converting gas energy into mechanical work. Owing to the frequent start-up and long-term high-speed operation of a gas turbine, various failures are frequently observed in critical subsystems, such as turbines and compressors, some of which cause huge loss of life and properties. This example demonstrates the effectiveness of the proposed methodology for detecting stage 1 bucket cracks using relative monitoring data collected from a real-world heavy-duty gas turbine. The unit was shut down on January 20th, 2015, owing to bucket failure on the stage 1 turbine wheel. Time historical data of 37
sensors (tags) installed on the unit were extracted from the control system of the machine from January 2013 to January 2015 and used for demonstration purposes. The 21 variables shown in Table IV are employed in this example according to the inputs of the domain experts and preliminary data analysis. The sensor tag names are defined in Table IV, with example values for illustration purposes. The sensor data for each tag were extracted at intervals of 5 min for subsequent analysis.

### TABLE III

| Tag name | Description                                      | Unit | Ex.value |
|----------|--------------------------------------------------|------|----------|
| CSGV     | Compressor IGV (inlet guide vanes) angle         | Degree | 85.99   |
| CTD      | Compressor discharge temperature                 | F     | 777.84   |
| CTIM     | Compressor inlet temperature                     | F     | 52.60    |
| DWATT    | Generator power watts                            | MW    | 430.66   |
| TNH      | Turbine HP shaft speed                           | %     | 100.07   |
| TTWS1AO1 | Turbine wheel space temperature – 1st stage after (Outer) #1 | F | 872.92 |
| TTWS1AO2 | Turbine wheel space temperature – 1st stage after (Outer) #2 | F | 865.67 |
| TTWS1FI1 | Turbine wheel space temperature – 1st stage forward (Inner) #1 | F | 878.05 |
| TTWS1FI2 | Turbine wheel space temperature – 1st stage forward (Inner) #2 | F | 886.52 |
| TTWS2AO1 | Turbine wheel space temperature – 2nd stage after (Outer) #1 | F | 918.57 |
| TTWS2AO2 | Turbine wheel space temperature – 2nd stage after (Outer) #2 | F | 899.31 |
| TTWS2FO1 | Turbine wheel space temperature – 2nd stage forward (Outer) #1 | F | 954.77 |
| TTWS2FO2 | Turbine wheel space temperature – 2nd stage forward (Outer) #2 | F | 964.14 |
| TTWS3AO1 | Turbine wheel space temperature – 3rd stage after (Outer) #1 | F | 568.75 |
| TTWS3AO2 | Turbine wheel space temperature – 3rd stage after (Outer) #2 | F | 610.01 |
| TTWS3FO1 | Turbine wheel space temperature – 3rd stage forward (Outer) #1 | F | 955.10 |
| TTWS3FO2 | Turbine wheel temperature – Stage 3 Rotor Forward-Outer 2 | F | 944.57 |
| TTXM     | Exhaust temperature                              | F     | 1176.80  |
| TTXS1    | Exhaust temperature spread #1                    | F     | 44.38    |
| TTXS2    | Exhaust temperature spread #2                    | F     | 36.02    |
| TTXS3    | Exhaust temperature spread #3                    | F     | 30.13    |

Figure 13 shows the basic information of the gas turbine event used in this example, including the approximate layout of sensors in the turbine (Fig. 13a) shown in Table IV, an illustrative photo of the bucket cracks at the stage 1 wheel (Fig. 13b), the crack locations in the stage 1 bucket (S1B) (Fig. 13c), and the numerical simulation of the temperature of one damage S1B (Fig. 13d). The 21 critical variables defined in Table III are used as examples for damage prediction. Multiple S1B parts were observed to have cracks during this event, as shown in Fig. 13b. A typical S1B specimen was observed to have three different locations with cracks (Fig. 13c). The stress concentration resulting from mechanical-thermal cycling was observed in the numerical simulation, as shown in Fig. 13d. In this study, operation data from Jan 1 to Nov 1, 2013, were used to train the proposed model, while those from Dec 2013 to Dec 2014 were employed to test the trained model.

#### 2.B Data preprocessing

As described in the implementation flow chart of Fig.1, the raw time series data of each sensor are preprocessed, including 1) eliminating shutdown data, 2) filling of null and infinite values to ensure data integrity, 3) data standardization to eliminate the dimensional influence of different variables, and 4) wavelet packet Bayesian threshold denoising to remove uncertain noise in the data. As an example, Figure 14 shows the time series plots of the CTD sensor data without (a) and with the removal of shutdown points (b). It can be clearly observed that the time series in Fig. 14b shows a distinct trend after removing the outliers in comparison to the raw data in Fig. 14a, implying the importance of data preprocessing.

For illustration purposes, Figure 15 shows the time series plots of the normalized data for four sensor tags, that is, CTD, CTIM, TTWS1AO1, and TTWS2FO1, as defined in Table III. It is observed that the trend of each variable indicates a large variation, which is rarely used to directly detect the possible damage in the component. However, multivariate time-series data collected from the system may be highly correlated with each other. In particular, various temperature sensors on the turbine wheel space show strong correlation, as indicated in the correlation heatmap in Fig. 16 for the variables listed in Table III, where either the very dark (negative) or very light (positive) block shows a strong correlation with a Pearson coefficient greater than 0.7. PPCA is employed to reduce the dimension and obtain the uncorrelated principal component series for anomaly detection using the proposed optimized similarity-based learning method.
FIGURE 14. Time series plot of CTD sensor data (Example 2): a) Raw data, and b) removing outliers

FIGURE 15. Time series plots of normalized sensor data (Example 2)

Similar to Example 1, prior to the PPCA dimension reduction, Bayesian wavelet thresholding is applied in this example to reduce possible noise in each set of time-series data. As an example, the SNR and RMSE of the CTD data, 58.0 and 0.99, were obtained from (2) and (3), respectively, indicating that the sensor data are lightly noisy as the RMSE is relatively small with regard to the averaged CTD values of 782F. Furthermore, Figure 17 shows a graphical comparison of the CTD data in terms of the time series (Fig. 17a) and power spectral density (PSD) plots (Fig. 17b), where the solid and dashed lines represent the raw and denoised signals, respectively. It can be observed that the difference between the raw and denoised data cannot be easily identified from the time series plot (Fig. 17a) but is clearly shown in the frequency graphs at high frequencies (Fig. 17b). This is reasonable because noise is typically contained in the high-frequency domain of the signal. This indicates that Bayesian wavelet packet threshold denoising retains the overall trend of the raw data while reducing high-frequency noise from the raw signal.

FIGURE 16. Correlation visualization of 21 critical variables (Example 2)

FIGURE 17. Graphical evaluation of data denoising for CTD variable (Example 2): a) time series plots, and b) PSD graphs

Subsequently, PPCA dimension reduction was performed on the preprocessed data, and more than 95% of the principal component information was retained. Finally, the 21 monitoring variables were reduced to the first 6 principal
components for the modeling analysis. The first principal component represented over 69% of the raw data. The individual explained variance and cumulative explained variance in PPCA dimension reduction are shown in Fig. 18.

**FIGURE 18.** Explained variance in PPCA (Example 2)

Similarly, the principal component (PC) data obtained from the PPCA dimension reduction are divided into four groups, that is, training, optimization, validation, and prediction sets, as shown in Table V. For the sake of computing efficiency, the PC data were resampled from 5 minutes interval to 1 h for subsequent modeling.

| Data set | Period       | Data points |
|----------|--------------|-------------|
| Training | 2013/01/01-2013/08/15 | 3646        |
| Optimization | 2013/08/15-2013/09/19 | 782         |
| Validation | 2013/09/19-2014/01/01 | 782         |
| Prediction | 2014/01/01-2015/01/17 | 11759       |

2. C Fault prediction and diagnosis

The six PC variables obtained from the PPCA were used as the inputs of the OAKR model. Both the training and optimization datasets shown in Table V were used to construct an OAKR model, where the optimization dataset was used to yield the optimal bandwidth of the model kernel function (0.444) from the Nelder-Mead simplex algorithm. The validation set shown in Table V was then used to evaluate the accuracy of the yielded OAKR model, yielding an MSE of 0.559. The threshold for anomaly alarm in this example is predefined as $4\sqrt{MSE} = 2.990$, as explained earlier. The prediction set listed in Table V was used to test the OAKR model. The predicted values of all the six PC variables were obtained using the validated model.

Figure 19 shows the model prediction results for the six PC variables in comparison with the original ones. It can be clearly observed from Fig. 19 that the predicted values of PC2 (Fig. 19b), PC5 (Fig. 19e), and PC6 (Fig. 19f) deviated from the original values before the failure on Jan 20th, 2015, which was much larger than the other three. This implies that the failure initiated early in Feb 2014 and can be detected in advance to a large extent using the proposed methodology, and the sensor variables mainly contributing to the three PCs play a key role in the failure. Again, the moving window strategy was used to calculate the mean square error (MSE) of the predicted value of each PC and then generate an averaged MSE metric for the overall condition evaluation of the component over time, as shown in Fig. 20. Using the threshold value of 2.990 obtained from the validation data, an alarm was triggered when the MSE value exceeded the threshold on March 8, 2014, which was more than 10 months earlier than the actual failure time. It has been well recognized that blade cracks develop slowly in practice because of the constraints of buckets in the wheel, where they may exist for over one year of operation in some cases, as explained in the detailed failure analysis (e.g., Zdzislaw et al. [25] and Chang et al. [26]). Once again, this case demonstrates that the proposed methodology provides a powerful tool to trigger an early alert on component cracks for arranging condition-based maintenance in advance.

**FIGURE 19.** OAKR prediction result of PPCA components (Example 2): a)PC1, b)PC2, c)PC3, d)PC4, e) PC5, and f) PC6

**FIGURE 20.** Averaged MSE alarming curve for overall condition evaluation of the component (Example 2)
Based on the results shown in Figs. 19 and 20, it can be clearly observed that the fault was mainly attributed to PC2, PC5, and PC6. Figure 21 shows the loadings of the three PCs obtained using Eq. (10), which indicates the contribution of each variable to PCs. This indicates that the faults are primarily represented by six variables: CTD, CTIM, CSVG, DWATT, and two-stage one-wheel space temperature variables. This is consistent with the actual event of a stage-one bucket crack.

![Figure 21. Contribution loadings of variables for PC2, PC5 and PC6 (Example 2)](image)

V. Concluding remarks
This paper presents a novel probabilistic optimized similarity-based learning (SBL) and signal reconstruction methodology for the accurate fault prediction of rotatory machines by seamlessly integrating two advanced signal reconstruction techniques with the optimized auto-associative kernel regression (OAKR) approach. The Bayesian discrete wavelet packet thresholding is employed as an advanced time-frequency reconstruction approach to cleanse the raw signals by integrating Bayesian hypothesis testing and multiresolution wavelet analysis, which effectively avoids under-denoising due to multiscale signal decomposition or over-denoising due to unbiased judgment of the decomposition coefficients. Probabilistic principal components analysis is utilized to extract the main information from multivariate feature variables, considering data uncertainties, thus lowering the feature dimension and yielding uncorrelated principal components as inputs to improve the accuracy and efficiency of industrial equipment fault prediction.

The enhanced methodology with PPCA distinguishes it from the method presented by Tang et al. [22] in two aspects. First, it reduces the uncertainty and dimensions of multivariate sensor data, where the uncorrelated principal components are used as inputs for fault prediction. Second, PPCA is also utilized to identify the critical factors for facilitating the fault diagnosis of rotating machines, thus allowing for fault prediction and diagnosis simultaneously.

Instead of manually modifying the bandwidth parameter in the conventional auto-associative kernel regression method, the Nelder-Mead simplex algorithm was developed to automatically optimize the key parameters in the model, thus significantly improving the model prediction accuracy and efficiency. The alarm threshold is determined adaptively using a rolling window strategy based on the difference between the predicted and actual values. A generalized procedure was developed to automate the application of the proposed methodology for fault prediction of rotatory machines under data uncertainties.

The effectiveness and feasibility of the proposed probabilistic signal reconstruction methodology and procedure are demonstrated using data and events collected from two real-world large turbomachines, a centrifugal compressor, and a heavy-duty gas turbine. The numerical results indicate that the proposed methodology provides a powerful tool for early warning of machine anomalies, which facilitates real-time condition monitoring of rotatory machines for smart maintenance. The numerical implementations of two different rotating machines have also demonstrated the strong generalization and robustness of the proposed method.

In future research, the proposed optimal probabilistic similarity-based learning method along with multiple advanced signal reconstruction techniques will be investigated with different rotating machines such as wind turbines and steam turbines to validate its generality and robustness in general turbomachines, and integrated into a comprehensive digital twin-oriented software platform for smart monitoring and diagnosis of industrial equipment.

VI. References
[1] Ke, F. A., et al. "Vibration-based updating of wear prediction for spur gears." Wear, vol. 426–427, Part B, pp. 1410–1415, April. 2019.
[2] N., K. Wang and J. Zheng. "Rolling element bearings fault diagnosis based on a novel optimal frequency band selection scheme," in IEEE Access, vol. 7, pp. 80748-80766, 2019, doi: 10.1109/ACCESS.2019.2921409.
[3] P. Caselitz, J. Giebhardt. "Rotor condition monitoring for improved operational safety of offshore wind energy converters." J. Sol. Energy Eng, vol. 127, no. 2, pp. 253-261, May. 2005.
[4] Bennett O., et al. "Diagnosis of the doubly-fed induction generator of a wind turbine." Wind Engineering, vol. 29, no. 5, pp. 431-447, September, 2005.
[5] Zaber, A. S. A. E., et al. "Online wind turbine fault detection through automated SCADA data analysis." Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, vol. 12, no. 6, pp. 574-593, January. 2009.
[6] Garvey, J., et al. "Validation of on-line monitoring techniques to nuclear plant data." Nuclear Engineering and Technology, vol. 39, no. 2, pp. 133-142, 2007.
[7] F. D. Maio, P. Baraldi, E. Zio and R. Serenoi. "Fault Detection in Nuclear Power Plants Components by a Combination of Statistical Methods," in IEEE Transactions on Reliability, vol. 62, no. 4, pp. 833-845, Dec. 2013, doi: 10.1109/TR.2013.2285033.
[8] N. Sairam and S. Mandal, "Thermocouple sensor fault detection using Auto-Associative Kernel Regression and Generalized Likelihood Ratio Test," 2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2016, pp. 1-6, doi: 10.1109/ICCECE.2016.8009575.
[9] Yu, J., et al. "Bagged auto-associative kernel regression-based fault detection and identification approach for steam boilers in thermal power plants," Journal of Electrical Engineering and Technology, vol. 12, no. 4, pp. 1406-1416, 2017.
[10] P. Guo, and N. Bai. “Wind turbine gearbox condition monitoring with AAKR and moving window statistics methods.” Energies, vol. 4, no. 11, pp. 2077-2093, 2011.
A. Brandsater, E. Vanem and I. K. Glad, "Cluster Based Anomaly Detection with Applications in the Maritime Industry," 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), 2017, pp. 328-333, doi: 10.1109/SDPC.2017.69.

F. Qian, Y. Feng and J. Ling, "Condition Monitoring of Turbine Generator Using Stator Winding Temperature," 2018 IEEE International Conference on Prognostics and Health Management (ICPHM), 2018, pp. 1-5, doi: 10.1109/ICPHM.2018.8448408.

Baraldi, P., et al. "Robust signal reconstruction for condition monitoring of industrial components via a modified Auto Associative Kernel Regression method." Mechanical Systems and Signal Processing, vol. 60, pp. 29-44, August, 2015.

X. Jiang, S. Mahadevan, and H. Adeli. "Bayesian wavelet packet denoising for structural system identification." Structural Control and Health Monitoring: The Official Journal of the American Society of Structural Control and Monitoring and of the European Association for the Control of Structures, vol. 14, no. 2, pp. 333-356, 2007.

S. Jiang and Y. Liu, "Injector waveform analysis and engine fault diagnosis based on frequency space subdivision in wavelet transform," 2010 International Conference on Wavelet Analysis and Pattern Recognition, 2010, pp. 318-323, doi: 10.1109/ICWAPR.2010.5576369.

Jin Q. and S. Khursheed, "A wavelet theory about online wavelets denoising based on Moving Window and Principal Component Analysis (PCA)," 2013 International Conference on Wavelet Analysis and Pattern Recognition, 2013, pp. 56-61, doi: 10.1109/ICWAPR.2013.6599292.

X. Jiang, S. Yang, F. Wang, S. Xu and X. Cheng, "OrbitNet: A New CNN Model for Automatic Fault Diagnostics of Turbomachines," Applied Soft Computing, vol.110, pp. 107702, 2021, doi:10.1016/j.asoc.2021.107702.

M. E. Tipping and C. M. Bishop. "Probabilistic principal component analysis." Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 61, no. 3, pp. 611-622, 1999.

Lagarias, J. C., et al. "Convergence properties of the Nelder–Mead simplex method in low dimensions." SIAM Journal on optimization, vol. 9, no. 1, pp. 112-147, 1998.

Liao, S., et al. "Training neural networks via simplified hybrid algorithm mixing Nelder–Mead and particle swarm optimization methods." Soft Computing, vol. 19, no. 3, pp. 679-689, 2015.

J.A. Nelder and R. Mead. "A simplex method for function minimization." The computer journal, vol. 7, no. 4, pp. 308-313, 1965.

W. Tang, X. Jiang, X. Zhao, Q. Chen, et al. (2021). "A probabilistic similarity based modeling approach for turbine machine fault prediction." PHM Society European Conference, vol.6, no.1, pp. 9. doi:10.36001/phme.2021.v6i1.2952.

P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," in IEEE Transactions on Audio and Electroacoustics, vol. 15, no. 2, pp. 70-73, June 1967, doi: 10.1109/TAU.1967.1161901.

Baraldi, F. Di Maio, D. Genini and E. Zio, "Comparison of data-driven reconstruction methods for fault detection," in IEEE Transactions on Reliability, vol. 64, no. 3, pp. 852-860, Sept. 2015, doi: 10.1109/TR.2015.2436384.

M., Zdzislaw, Kubiak, J. and M., Carlos. "Failure analysis of gas turbine last stage bucket made of udimet 500 superalloy." Practical Failure Analysis, vol. 2, pp. 51-57, 2002, doi: 10.1361/152981502770352184.

J. Chang, Y. Yun, C. Choi, and J. Kim, "Failure analysis of gas turbine buckets," Engineering Failure Analysis, vol. 10, no. 5, pp. 559-567, 2003, doi:10.1016/S1350-6307(03)00037-2.

Xiaomo Jiang received the M.S. degree in Structural Engineering from the National University of Singapore, Singapore, in 2000, and the Ph.D. degree in Intelligent Structures from The Ohio State University, Columbus, OH, USA, in 2005, respectively. He is currently a distinguished professor at the School of Energy and Power Engineering and State Key Laboratory of Structural Analysis for Industrial Equipment, Director of Provincial Key Laboratory of Digital Twin for Industrial Equipment, and Director of Research Institute of Carbon Neutrality, Dalian University of Technology, Dalian, China. He was a senior engineer and technical leader in the monitoring, diagnostics, and prognostics of gas turbines and steam turbines with General Electric Company, Atlanta, GA, USA, from 2008 to 2017, and a Postdoctoral Research Associate with Vanderbilt University, Nashville, TN, USA, from 2005 to 2007. He has coauthored one book, five book chapters, and more than 100 research articles, and had over ten Europe and US patents in the cross-disciplinary fields of computer science, engineering, artificial intelligence, and applied statistics. He has been listed as one of the highly cited Chinese researchers in Mechanical Engineering by Elsevier since 2015 for 6 years in a row. He has been invited to serve as a Peer Reviewer, Editorial Board, or Associate Editor for more than 40 different international journals and conferences. His main research interests include energy Internet of Things, digital twin, smart maintenance, and predictive analytics for rotatory machines, such as gas turbines, steam turbines, and compressors.

Weijian Tang received the B.S. degree in Engineering Mechanics from Hefei University of Technology, China in 2019. He is currently pursuing the M.S. degree in Engineering Mechanics at Dalian University of Technology, China. His research interests include data-driven condition monitoring and predictive maintenance for large rotatory machines and the development and application of artificial intelligence and machine learning algorithms for smart maintenance of industrial equipment.

Haixin Zhao received the B.S. degrees in Engineering Mechanics from the Dalian University of Technology, China, in 2015 and is currently a Ph.D. candidate in Computational Mechanics from the Dalian University of Technology. His research interests include fault diagnosis and prognosis of rotating machinery, predictive maintenance, signal processing, fault simulation and artificial intelligence for smart maintenance of industrial equipment.

Xueyu Cheng received the M.S. and Ph.D. degrees in Economics from the Ohio State University, Columbus, OH, USA in 2007. She is currently an associate professor at the College of Arts and Sciences at Clayton State University, Morrow, GA, USA. She was an assistant professor of economics at Alabama State University, Montgomery, AL. Her research focuses on the application of statistics to various fields. She has authored and co-authored one book chapter and many research articles. He has been invited to be a speaker and peer reviewer at many international conferences and journals.