

# An improved OAKR approach to condition monitoring of rotating machinery

Kexin Zhang<sup>1</sup>, Xiaomo Jiang<sup>2</sup>, Huiyu Hui<sup>3</sup> and Manman Wei<sup>4</sup>

<sup>1,2,3,4</sup> *State Key Lab of Structural Analysis, Optimization and CAE Software for Industrial Equipment, Provincial Key Lab of Digital Twin for Industrial Equipment, School of Energy and Power Engineering, Dalian University of Technology, Dalian, 116024, CN*

[zhangkexin@mail.dlut.edu.cn](mailto:zhangkexin@mail.dlut.edu.cn)

[xiaomojiang2019@dlut.edu.cn](mailto:xiaomojiang2019@dlut.edu.cn)

[huihuaiyu@mail.dlut.edu.cn](mailto:huihuaiyu@mail.dlut.edu.cn)

[wei\\_m\\_m@mail.dlut.edu.cn](mailto:wei_m_m@mail.dlut.edu.cn)

## ABSTRACT

Faults in main subsystems or components of a rotating machine often causes unscheduled shutdown, which may lead to not only huge economic losses, but also safety accidents. As an important part of intelligent maintenance, condition monitoring becomes a powerful tool in reducing maintenance costs through automatic fault alarming, thereby reducing potential downtime while improving system safety and reliability. An optimized auto-associative kernel regression (OAKR) model has been proposed recently and demonstrated as a promising tool for condition monitoring of various turbomachines, which is independent of fault mode and machine type. However, the fault identification accuracy of this approach largely relies on data quality in practical applications. Data incompleteness, parameter variation and system complexity often result in the inaccuracy of fault alarming for complicated rotating machinery. This paper proposes an improved OAKR method to address these issues, including utilizing wavelet packet Bayesian thresholding method (WPB) to reduce noise in the raw multivariate data, developing the Manhattan distance to calculate the sample similarity, and constructing a multivariate health index based on Multivariate Permutation Entropy to identify potential faults in equipment condition monitoring. Parametric analysis and a comparison study with original AAKR and OAKR methods by using the actual data of a gas turbine are conducted to illustrate the effectiveness and feasibility of the proposed methodology.

**Keywords:** rotating machine; condition monitoring; OAKR; wavelet packet; Bayesian thresholding

## 1. INTRODUCTION

Large rotating machines such as gas turbines and compressors are widely used in energy and power fields and are essential industrial equipment. Failure of such large rotating machinery can easily cause major property damage or endanger lives. Therefore, effective condition monitoring of rotating machinery to ensure the long-term safe and reliable operation of the unit is a growing concern in modern industry.

Currently, statistical, machine learning and deep learning methods are used for condition monitoring or fault warning. Kämpjärvi, Sourander, Komulainen, Vatanski, Nikus, and Jämsä-Jounela (2008) combined multivariate statistical analysis with artificial neural network (ANN) methods and applied them to fault detection during machine operation with good results. However, the above methods rely on large amounts of good quality equipment operating data.

In recent years, a data-driven approach based on similarity models with self-associative kernel regression (AAKR) has been proposed. In contrast to statistical methods, machine learning, and deep learning methods, the AAKR approach does not require any prior knowledge about the system under study. The method can be applied not only to static equipment such as boilers in coal-fired power plants (Yu, Jang, Yoo, Park & Kim, 2017), but also to rotating machinery such as blades of compressors or turbines (Qian, Feng & Ling, 2018). Baraldi, Di Maio, Turati, and Zio (2015) et al. modified the weights of the AAKR model to reduce the correlation between multiple variables. However, this approach has three drawbacks. Firstly, it is very sensitive to uncertainties in the input data and therefore may generate many false positives or misses. Secondly, the choice of bandwidth affects the accuracy of fault alarms. Finally, the method relies on the similarity theory to make predictions, so the method to calculate similarity is particularly important. However, none

of these methods proposed so far are a good solution to the problems mentioned earlier, such as sensitivity to data uncertainty.

In this paper, based on the OAKR method, firstly, the WPB is chosen to reduce the noise in the pre-processing stage of the data, and its advantages and disadvantages are also compared with the local outlier factor (LOF) method and the isolated forest method (iForest); The improved OAKR method is then applied to equipment condition monitoring, and the residual series is constructed from the predicted and true values, and the multivariate ranking entropy index is used to effectively analyze the residual series and achieve early fault warning in advance; finally, the proposed method is effectively validated with the data of a certain type of gas turbine as an example.

## 2. IMPROVED OAKR METHOD

### 2.1. Three Denoising Methods

For noisy data, this paper compares the impact of three different denoising methods on the early warning results, namely WPB, iForest, and LOF denoising. The principles of the three denoising methods are described below.

#### 2.1.1. Bayesian Multiresolution Analysis Thresholding

The WPB (Jiang, Mahadevan & Adeli, 2007) is a noise reduction technique based on advanced signal analysis and reliability methods. This paper is only a brief introduction; see the original article for details of the derivation.

In the decomposition of discrete wavelet packets, given a time series of  $N$  measurement data points  $f(t_i)$  ( $i = 1, 2, \dots, N$ ), it is simultaneously decomposed into a series of scale coefficients  $s_j(k)$  wavelet coefficients  $w_j(k)$ . The time series can be expressed as

$$f(t_i) = \sum_{k \in Z} \sum_{j \in Z} [s_j(k)\varphi_{j,k}(t_i) + w_j(k)\psi_{j,k}(t_i)] \quad (1)$$

where  $j$  is the number of layers in the wavelet packet decomposition,  $k$  is the decomposition coefficient point, the  $\varphi_{j,k}(t_i)$  and  $\psi_{j,k}(t_i)$  denote respectively the  $k$ -th of the  $j$ -th layer at  $t_i$  wavelet packet decomposition scale function and wavelet function at the time.

Assuming that a time sequence  $f(t_i)$  contains an average of 0 and a square difference of  $\sigma^2$  with an independent co-distributed white noise  $\varepsilon(t_i) \sim N(0, \sigma^2)$ , the timescale  $f(t_i)$  can be expressed as  $f(t_i) = g(t_i) + \varepsilon(t_i)$ , where  $g(t_i)$  is the theoretical noise-free time series. For the  $f(t_i)$  wavelet packet decomposition, the noise term  $\varepsilon(t_i)$  is also decomposed into a series of corresponding noise coefficients  $\varepsilon_{jk}$ . Let  $d_{jk}$  represent the Eq. (1) in  $s_j(k)$  or  $w_j(k)$  the  $k$ -th decomposition coefficient of the  $j$ -th layer of and denote as

$$d_{jk} = \hat{d}_{jk} + \sigma_j \varepsilon_{jk}, j = j_0, \dots, J-1; k = 0, 1, \dots, 2^j - 1 \quad (2)$$

where  $\varepsilon_{jk} \sim N(0, 1)$  is the independent random variable and  $\hat{d}_{jk}$  is the decomposition coefficient after noise reduction.

Note that the actual fault signature signal has a higher wavelet energy. Signals with lower wavelet energy are considered as noise and do not affect the actual fault warning.

Eq. (2) can be expressed as the following conditional probability distribution:  $d_{jk} | \hat{d}_{jk}, \sigma_j^2 \sim N(\hat{d}_{jk}, \sigma_j^2)$ .

Assume that the prior distribution of  $\hat{d}_{jk}$  can be expressed as follows  $\hat{d}_{jk} | \gamma_{jk} \sim N(0, \gamma_{jk} \tau_j^2)$ , where  $\gamma_{jk}$  is a binary random variable that follows a Bernoulli distribution with  $P(\gamma_{jk}) = 1 = 1 - P(\gamma_{jk} = 0) = \pi_j$ . In practical application the parameters  $\tau_j^2$  and  $\pi_j$  are parameters of constant size at the  $j$ -th level of decomposition and are usually taken as  $\pi_j = 0.5$  and  $\tau_j^2 = 1$ , indicating that for each  $\gamma_{jk}$  is the unbiased assumption (Jiang & Mahadevan, 2008).

According to Bayesian inference, the posterior distribution of the  $\hat{d}_{jk}$  can be obtained from the following equation (Jiang et al. 2008):

$$\hat{d}_{jk} | \gamma_{jk}, d_{jk}, \sigma_j^2 \sim N\left(\gamma_{jk} \frac{\tau_j^2}{\sigma_j^2 + \tau_j^2} d_{jk}, \gamma_{jk} \frac{\sigma_j^2 \tau_j^2}{\sigma_j^2 + \tau_j^2}\right) \quad (3)$$

Based on Eq. (3), the Bayes factor, can be calculated from the following equation:

$$\eta_{jk} = \frac{1 - \pi_j}{\pi_j} \frac{(\sigma_j^2 + \tau_j^2)^{1/2}}{\sigma_j^2} \exp\left(-\frac{\tau_j^2 d_{jk}^2}{2\sigma_j^2(\sigma_j^2 + \tau_j^2)}\right) \quad (4)$$

When  $\eta_{jk} < 1$ , indicating that the corresponding coefficient is retained; when  $\eta_{jk} > 1$ , indicating that the corresponding coefficient is judged as noise. The noise-reduced time series  $x(t_i)$  can be obtained by reconstructing the wavelet inverse transform and the noise-reduced decomposition coefficients  $d_{jk}$  brought into Eq. (1).

#### 2.1.2. Isolation Forest

The core of the Isolation Forest algorithm is the construction of a forest (iForest) consisting of iTrees (isolation trees) (Liu, Ting & Zhou, 2008). To facilitate description and computation, the Isolation Forest algorithm introduces the definition of isolated trees and path lengths.

Isolated Tree: Let  $T$  be a binary tree and  $N$  be a node of  $T$ . When  $N$  is a leaf node,  $N$  is said to be an external node, and when  $N$  has two child nodes, it is said to be an internal node. Given a data set  $X = \{x_1, \dots, x_n\}$ , in order to construct the isolation tree, the randomly selected attribute  $q$  and the separator value  $p$  are recursively used to divide  $X$  until: 1) the tree reaches its maximum height; 2) there is only one piece of data left in the  $X$ ; 3) there are multiple identical pieces of data left in the  $X$ .

Path length: In an iTree, the number of edges from the root node to an external node is called the path length, denoted as  $h(x)$ .

The average of external node termination  $h(x)$  estimates are the same as those for unsuccessful searches in BST. We borrow the analysis from BST to estimate the average path length of iTree. The literature (Lee, 2017) gives the path lengths of failed queries in a binary lookup tree as

$$c(n) = 2H(n-1) - (2(n-1)/n) \quad (5)$$

where  $H(i)$  is the harmonic number, which can be estimated by  $\ln(i) + 0.5772156649$  (Euler's constant). Since  $c(n)$  is the average of  $h(x)$  for a given  $n$ , we use it to normalise  $h(x)$ . The anomaly score  $s$  of an instance  $x$  is defined as  $s(x, n) = 2^{-E(h(x))/c(n)}$ , where  $E(h(x))$  is the average of  $h(x)$  from the set of isolated trees. When  $s \rightarrow 1$ , they are outliers; when  $s \rightarrow 0$ , i.e. when the data return  $s$  much less than 0.5, they have a high probability of being normal.

### 2.1.3. Local Outlier Factor

The Local Outlier Factor (LOF) detection algorithm is an unsupervised density-based denoising method. It considers samples with a density much lower than that of their neighbors as noise (Zhang & Gai, 2020). Given a dataset:  $X = \{x^1, \dots, x^N\}$ , where  $N$  is the total number of points in  $X$ .

Any point  $x \in X$  of the LOF is:

$$LOF(x) = \left( \sum_{y \in P_k(x)} (\rho_k(y) / \rho_k(x)) \right) / |P_k(x)| \quad (6)$$

$P_k(x)$  is a  $k$ -distance neighbourhood point of  $x$ , defined as  $P_k(x) = \{y \in X \setminus \{x\} | d(x, y) \leq d_k(x)\}$ , where  $d_k(x)$  is called the  $k$ -distance of a point  $x$  and is defined as the distance  $d(x, x')$  between  $x$  and its  $k$ -th nearest neighbor  $x'$ . Since more than one point may exist at distance  $d(x, x')$ , the number of points in  $P_k(x)$ , i.e.  $|P_k(x)|$ , satisfies the relation:  $|P_k(x)| \geq k$ .

The density  $\rho_k(x)$  of  $x$  is defined as

$$\rho_k(x) = |P_k(x)| / \sum_{y \in P_k(x)} d_k(x, y) \quad (7)$$

where  $d_k(x, y)$  is called the reachable distance of point  $x$  with respect to point  $y$  and is defined as

$$d_k(x, y) = \begin{cases} d_k(y), & \text{if } d_k(y) > d(x, y) \\ d(x, y), & \text{else} \end{cases} \quad (8)$$

Therefore, substituting Eqs. (7) and (8) into Eq. (6) yields

$$LOF(x) = \frac{1}{\rho_k(x)} \frac{\sum_{y \in P_k(x)} \rho_k(y)}{|P_k(x)|} = \frac{\overline{\rho_k(x)}}{\rho_k(x)} \quad (9)$$

where  $\overline{\rho_k(x)}$  is defined as

$$\overline{\rho_k(x)} = \sum_{y \in P_k(x)} \rho_k(y) / |P_k(x)| \quad (10)$$

Based on Eq. (9), the LOF at any point  $x$  is the local average density at  $x$ , i.e. Eq. (10), divided by the density at  $x$ , i.e. Eq. (7). If the LOF is less than 1, it means that  $x$  is a dense point; if the LOF is greater than 1, it means that  $x$  may be an anomaly.

### 2.2. Optimized AAKR Approach

The self-associative kernel regression (AAKR) method is a similarity-based, non-parametric empirical modelling technique that uses historical fault-free data to predict the operation of a machine unit (Guo et al. 2011). The method is independent of equipment and fault type and is suitable for multivariate operation monitoring and fault warning for a wide range of equipment.

Extracting historical health data from the operation of the device is used to create the memory matrix  $X$  of the AAKR model, where  $X_{i,j}$  represents the  $i$ -th vector value of the  $j$ -th key variable. For  $n$  memory vectors, the memory matrix  $X$  for  $p$  variables can be expressed as

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,p} \end{bmatrix} \quad (11)$$

The monitoring vector is represented by a  $1 \times p$  matrix  $v$  as  $v = [v_1 \ v_2 \ \dots \ v_p]$ .

The AAKR method consists of four steps:

In the first step, the distance between the monitoring vector  $v$  and each memory vector  $X_i$  is calculated to obtain a  $n \times 1$  distance vector  $d_i$ , which is compared in this paper using the two methods mentioned subsequently.

In the second step, the weight matrix  $w$  is calculated from the obtained distance matrix  $d$  and the Gaussian kernel function.  $w$  is also a  $n \times 1$  vector matrix with each element calculated by the following equation:

$$w_i = K_h(d_i) = \exp(-d_i^2/h^2) / \sqrt{2\pi h^2} \quad (12)$$

where  $h$  is the bandwidth of the kernel function, which determines how smooth the function is. A smaller  $h$  shows more detail but does not produce a smooth tail, while a larger  $h$  will lose detail in system identification and thus lead to more false alarms from the model. Therefore, the choice of bandwidth  $h$  plays a crucial role in the accuracy of fault warnings. To improve the accuracy of model prediction,  $h$  is optimized using the simplex method in the OAKR method.

In the third step, the bandwidth of the kernel function in the model is automatically optimized by a new set of health data and the Nelder-Mead optimization algorithm, and the optimal

bandwidth is obtained by minimizing the mean square error (MSE) during model training.

In the fourth step, the predicted values  $\hat{v}$  of the monitoring vector  $v$  are predicted by the calculated weights  $w$ , the  $\hat{v}$  from each memory vector  $X_i$  is calculated as the weighted average of the following equation:

$$\hat{v} = \sum_{i=1}^n (w_i X_i) / \sum_{i=1}^n w_i \quad (13)$$

During the OAKR model-building process, the memory matrix should contain all normal operating states of the unit as much as possible.

### 2.3. Two Ways to Calculate Distance

#### 2.3.1. Euclidean Distance

The Euclidean distance (Wang, Zhang & Feng, 2005) is defined in the representation of the distance between two or more points in Euclidean space, with two  $n$ -dimensional vectors denoted by  $a(x_1, \dots, x_n)$  and  $b(y_1, \dots, y_n)$ , the Euclidean distance between two  $n$ -dimensional vectors is calculated as  $d_{a,b} = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$ .

#### 2.3.2. Manhattan Distance

Manhattan distance (Melter, 1987) is used to measure the sum of the distance between two points on each axis in a standard coordinate system. The most prominent advantage of this method is that it is fast to compute and has two  $n$ -dimensional vectors  $a(x_1, \dots, x_n)$  with  $b(y_1, \dots, y_n)$ . The Manhattan distance between them is given by  $d_{a,b} = \sum_{k=1}^n |x_k - y_k|$ .

### 2.4. Model Performance Indicator

#### 2.4.1. Mean Square Error

The mean squared error (MSE) is used to calculate the residual between the predicted and true values of the model. The average mean squared error of the  $N$  test samples is calculated as follows:

$$MSE = \sum_{i=1}^N \sum_{j=1}^P (v_{i,j} - \hat{v}_{i,j})^2 / (NP) \quad (14)$$

where  $N$  is the total number of samples tested,  $P$  is the total number of model variables,  $v_{i,j}$  is the true value of the  $j$ -th variable for the  $i$ -th sample, and  $\hat{v}_{i,j}$  is the predicted value of the model corresponding to the  $j$ -th variable of the  $i$ -th sample.

#### 2.4.2. Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)

The ROC curve is a metric for assessing the generalization performance of classification models in the field of machine learning. The horizontal axis of the ROC curve is the ratio of false positive samples, while the vertical axis is the ratio of

true positive samples. The larger the AUC (the area under the ROC curve), the better the generalization performance of the model. In this paper, the ROC curve is used to compare the prediction performance of the models under different working conditions.

### 2.5. Fault Alarming Strategy

A moving window strategy is used to determine whether equipment operation is abnormal by monitoring fluctuations in the multivariate alignment entropy indicator between the OAKR model output and actual data. Entropy is a general measure of the complexity of a system. When monitoring machine condition, a continuous, significant change in system entropy indicates a possible failure.

The entropy used in this study is the multivariate multiscale permutation entropy (MMSPE) (Morabito et al. 2012). Assuming that the time series samples of actual monitoring data  $V_t = [v_{t-l+1}, \dots, v_t]$  and the time series sample  $\hat{V}_t = [\hat{v}_{t-l+1}, \dots, \hat{v}_t]$  of the OAKR model prediction data each containing  $l$  data points at moment  $t$ , such that  $\Delta = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T$  represent the residual series corresponding to the two sets of data, and  $\Delta$  is a  $p \times l$  matrix of  $p$  variables, each containing  $l$  observations, with multivariate multiscale alignment entropy achieved by the following two steps:

Firstly, the original time series  $\Delta$  is coarse-grained to define the scale factor  $s$ . Each element  $\tilde{\varepsilon}_{i,j}^s$  in the multivariate coarse-grained time series  $\tilde{\Delta}$  can be deduced as

$$\tilde{\varepsilon}_{i,j}^s = \frac{1}{s} \sum_{t=(j-1)+1}^{js} \varepsilon_{i,t}, i=1, \dots, p, 1 \leq j \leq \frac{l}{s} \quad (15)$$

Secondly, the multivariate multiscale permutation entropy of the coarse-grained multidimensional time series is calculated. For a multivariate time series  $\tilde{\Delta}$ , for each variable  $i \in [1, p]$  and alignment  $r \in [1, m!]$ , the number of occurrences  $q$  of the symbolic sequence  $S(r)$  for each variable after phase space reconstruction is counted, and the relative frequency  $P_{i,r}$  can be expressed as

$$P_{i,r} = q / (p \times K), \quad \sum_{i=1}^p \sum_{r=1}^{m!} P_{i,r} = 1 \quad (16)$$

The original time series is transformed into a time-correlated matrix from which the associated statistics and entropy can be calculated. The relative frequency of the edges of the sequence distribution can be expressed as  $P_r = \sum_{i=1}^p P_{i,r}$ ,  $r = 1, \dots, m!$ . The multivariate multiscale alignment entropy is calculated as  $H_{MMSPE}^s = - \sum_{r=1}^{m!} p_r \log_2 p_r$ .

Usually, the results of entropy fluctuate greatly, to better extract the result information, this paper uses Cumulative Summation (CUSUM) to analyze the results, the idea of CUSUM is to accumulate the sample data information, improving the sensitivity of the detection process. Whenever the CUSUM of the monitored quantity is significantly higher or lower than the normal value indicates that there is a change

in the system. For any multivariate alignment entropy time series  $H_{MMSP}^S(t)$  the CUSUM is calculated by the following formula:

$$\left\{ \begin{array}{l} S_{upper}(t) = \max(0, H_{MMSP}^S(t) - \mu_H - \alpha\sigma_H + S_{upper}(t-1)) \\ S_{lower}(t) = \min(0, H_{MMSP}^S(t) - \mu_H + \alpha\sigma_H + S_{lower}(t-1)) \end{array} \right\}, \quad (17)$$

$$S_{upper}(0) = 0, S_{lower}(0) = 0$$

where  $S_{upper}(t)$  denotes the upper cumulative sum and  $S_{lower}(t)$  denotes the lower cumulative sum.  $\mu_H$  and  $\sigma_H$  are the mean and variance of the unit's entropy values at the moment of health, respectively.  $\alpha$  is the parameter to detect the mean shift. In this study, when an abnormal condition occurs in the unit, it usually causes a decrease in the entropy value, so only the lower limit cumulative sum  $S_{lower}(t)$  is considered.

The CUSUM alarming threshold is determined by the  $3\sigma$  criterion. Since  $S_{lower}(t) \leq 0$  holds constantly, there is no need to set an upper alarm warning limit. When  $S_{lower}(t) \leq \mu_S - 3\sigma_S$  the system is considered to have failed, where  $\mu_H$  and  $\sigma_H$  are the mean and variance of the lower accumulation limit sum  $S_{lower}(t)$  of the unit's healthy operating hours, respectively.

### 3. ILLUSTRATIVE EXAMPLE

In this case, a gas turbine compressor blade damage failure unit is used as an example, and the process shown in Figure 1 compares three different denoising methods and two methods of calculating distances to arrive at the most effective method for OAKR improvement.

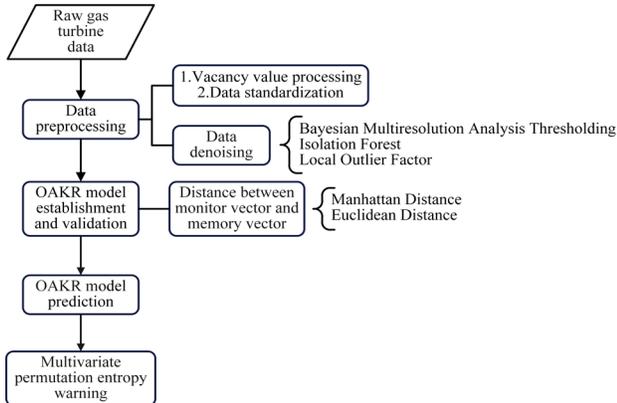


Figure 1 Flow chart of the improved OAKR method

#### 3.1. Data

The gas turbine suffered a failure of the compressor blades on September 5, 2006. This paper uses the data from August 1, 2006 to September 5, 2006 to demonstrate the validity of the method. The original data was sampled every 1 minute and down-sampled to 12 minutes due to the large amount of data. The down-sampled data are divided into four groups: training, optimization, testing and prediction sets. Table 1

shows the data set division. It is worth noting that for this paper the data before September 5, 2006 is considered to represent a healthy or near healthy condition of the unit.

#### 3.2. Data Preprocessing

The 35 variables of the gas turbine were used in the data preprocessing. Upward and downward data padding methods were used to deal with missing and infinite values in the characteristic data. The characteristic data for all variables were normalized.

Table 1 Data division of the OAKR model

	Period	Data points
Training	2006/08/01-2006/08/22	2409
Optimization	2006/08/22-2006/08/27	517
Testing	2006/08/27-2006/09/01	516
Prediction	2006/09/01-2006/09/05	531

#### 3.3. Comparison of Three Denoising Methods

The three methods mentioned previously were used to reduce the noise of the pre-processed data separately. Figure 2 shows a comparison of the data before and after denoising, and it can be seen near September 10 and August 19 that the data is obviously missing after denoising with iForest and LOF compared to WPB. Some important trend information is missing.

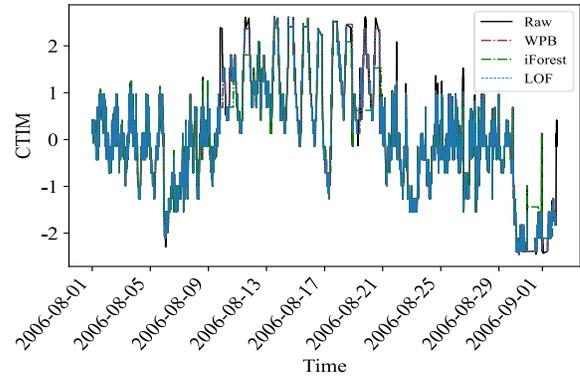


Figure 2 Comparison before and after denoising

#### 3.3.1. OAKR Model Establishment and Validation

The denoised data was used to build the OAKR model, as shown in Table 1. The training set is used to construct the memory matrix, and the optimization set is used to optimize the kernel function bandwidth  $h$  in the OAKR model. The optimal bandwidth  $h_{opt} = 0.713$  is obtained after optimizing the data with WPB noise reduction. The yielding minimum MSE is  $f(h_{opt}) = 1.190$  calculated from Eq. (14). The optimal bandwidth  $h_{opt} = 1.181$  is obtained after optimizing the data with the iForest noise reduction, and the minimum MSE is  $f(h_{opt}) = 0.8634$  calculated from Eq.

(14). The optimal bandwidth  $h_{opt} = 0.9188$  obtained by optimising the data after noise reduction with LOF, and the minimum MSE is  $f(h_{opt}) = 1.4198$  calculated by Eq. (14). Combining the denoising results and the results of optimising the bandwidth shows that WPB is better.

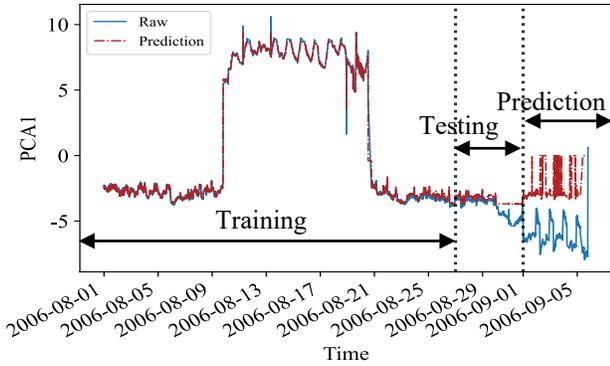


Figure 3 Comparison of prediction and raw variable

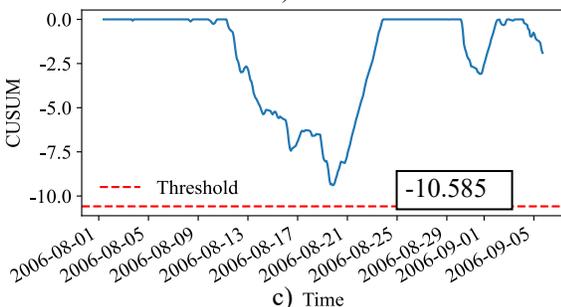
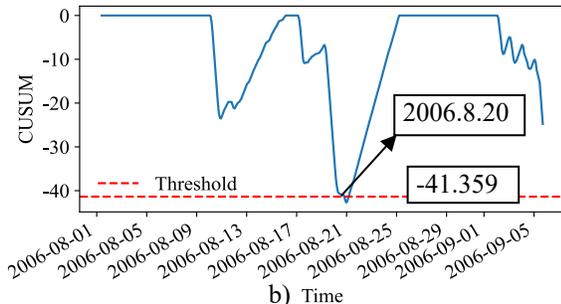
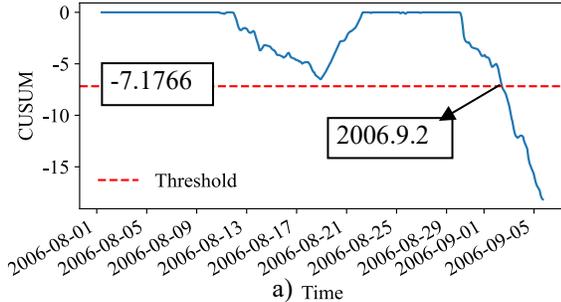


Figure 4 CUSUM indicator alarm time: a) WPB, b) iForest, and c) LOF

### 3.3.2. Model Prediction

The system response of the gas turbine is predicted by using the trained OAKR model. The comparison between the true and predicted values is shown in Figure 3, which shows that the residual between the predicted and true values increases before a fault occurs, indicating that a fault is imminent. Figure 4 shows the alarm times and thresholds for the various methods using the CUSUM metric. As can be seen from Fig.5 that the iForest treatment (Figure 4b) produces the earliest alarm time. The CUSUM value rises back to the normal immediately after exceeding the threshold until the failure occurs. Therefore, this alarm is not reliable due to the effect of noise. The denoised data with the LOF method (Figure 4c) does not obtain a valid alarm, but the denoised data by using WPB (Figure 4a) results in a 3-day early alarm.

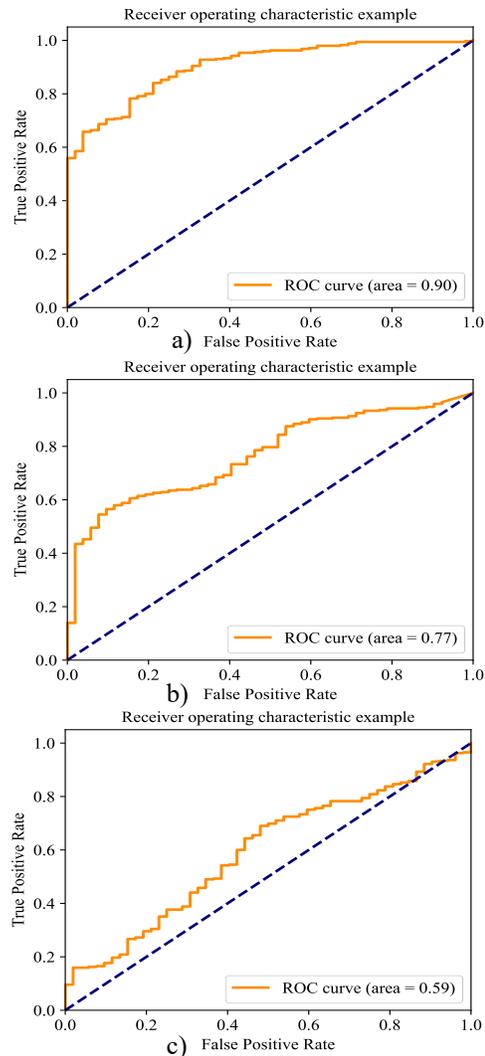


Figure 5 The results of ROC and AUC: a) WPB, b) iForest, and c) LOF

To compare the effect of denoising methods on fault prediction accuracy, Figure 5 shows the ROC curves and AUC values of the prediction results using different

denoising methods, where the AUC=0.90 for WPB (Figure 5a), AUC=0.77 for iForest (Figure 5b) and AUC=0.59 for LOF denoising (Figure 5c) are obtained. From the AUC values of the prediction results, it can be seen that the fault prediction performance using WPB is better. The running times of the models for the three methods were 122.37s, 93.85s and 138.65s respectively.

**3.4. Verification of Distance Method**

This section replaces the Manhattan distance with the Euclidean distance and the other parameters of the modelling process remain the same as the OAKR model in section 3.3. Only model establishment and validation are demonstrated in this section.

**3.4.1. OAKR model establishment and validation**

The optimal bandwidth  $h_{opt} = 0.3320$  is obtained by the optimization set, and the minimum MSE is calculated from Eq. (14) as  $f(h_{opt}) = 1.0326$ .

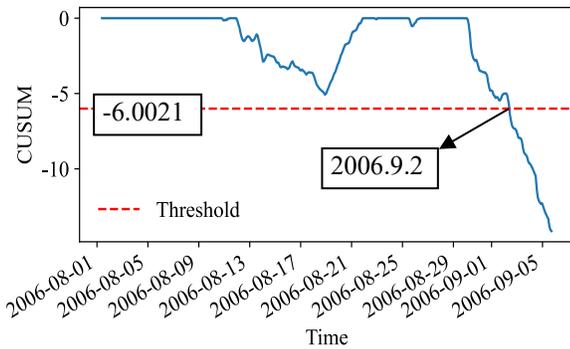


Figure 6 CUSUM indicator alarm time

**3.4.2. Model Prediction**

Again, the system response of the turbine is predicted by using the trained OAKR model. Figure 6 and Figure 7 show the alarm time as September 2, 2006, with an AUC value of 0.86. The run time of the OAKR model was 137.11 seconds.

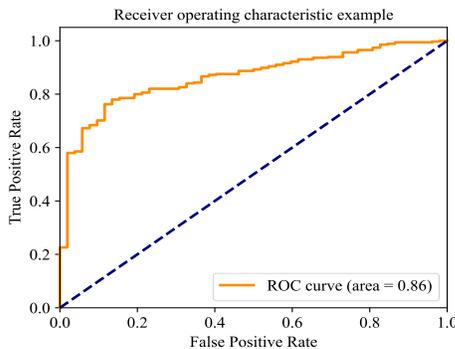


Figure 7 The results of ROC and AUC

**3.5. Summary**

From the results of the above gas turbine data runs, as can be seen from Table 2 (In the following tables, 'days' indicates the number of days of advance warning), using the AAKR method fails to alarm and has poor generalization performance. In the improved OAKR method, although the alarming time is similar by using the Manhattan distance and the Euclidean distance, the calculation is more efficient and the prediction performance is better by using the Manhattan distance. Regarding the denoising method (Table 3), combining the results analyzed in Section 3.3.2 with the AUC and CUSUM curves, WPB is recommended because it has a better warning effect and higher generalization performance.

Table 2 Comparison of different similarity methods

	Similarity methods	days	AUC	Running time(s)
AAKR	Euclidean		0.34	66.57
	Manhattan		0.41	65.51
OAKR	Euclidean	3	0.86	137.11
	Manhattan	3	0.90	122.37

Table 3 Comparison of different denoising methods

	WPB	iForest	LOF
days	3	16	
AUC	0.90	0.77	0.59
Running time(s)	122.37	93.85	138.65

**4. CONCLUSION**

In this paper, the OAKR method is improved in terms of noise reduction, similarity and alarming strategy for condition monitoring and fault detection of rotating machinery. After a comparison study, WPB is recommended for noise reduction of the original signal as it effectively avoid both under-de-noising due to multi-scale signal decomposition and over-de-noising due to unbiased judgement of the decomposition coefficients. The use of Manhattan distance is employed to calculate similarity instead of Euclidean distance in the OAKR model thus improving the model warning lead time and computational efficiency. Finally, the multivariate ranking entropy method is presented for early warning by the  $3\sigma$  criterion to determine the alarming threshold. The effectiveness and accuracy of the proposed method is verified with the data and event of a real-world gas turbine.

In future research, the improved OAKR would be developed to integrate the fault mechanism, in order to improve the efficiency, accuracy and reliability of condition monitoring and fault warning whenever the sample set is small in practical applications.

**REFERENCES**

Baraldi, P., Di Maio, F., Turati, P., & Zio, E. (2015). Robust signal reconstruction for condition monitoring of

industrial components via a modified Auto Associative Kernel Regression method. *Mechanical Systems and Signal Processing*, 60, 29-44.

- Guo, P., & Bai, N. (2011). Wind turbine gearbox condition monitoring with AAKR and moving window statistic methods. *Energies*, 4(11), 2077-2093.
- Jiang, X., Mahadevan, S., & Adeli, H. (2007). Bayesian wavelet packet denoising for structural system identification. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, 14(2), 333-356.
- Jiang, X., & Mahadevan, S. (2008). Bayesian wavelet methodology for structural damage detection. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, 15(7), 974-991.
- Kämpjärvi, P., Sourander, M., Komulainen, T., Vatanski, N., Nikus, M., & Jämsä-Jounela, S. L. (2008). Fault detection and isolation of an on-line analyzer for an ethylene cracking process. *Control engineering practice*, 16(1), 1-13.
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth IEEE international conference on data mining* (pp. 413-422). *IEEE*.
- Lee, S., & Mortari, D. (2017). Quasi-equal area subdivision algorithm for uniform points on a sphere with application to any geographical data distribution. *Computers & Geosciences*, 103, 142-151.
- Melter, R. A. (1987). Some characterizations of city block distance. *Pattern recognition letters*, 6(4), 235-240.
- Morabito, F. C., Labate, D., Foresta, F. L., Bramanti, A., Morabito, G., & Palamara, I. (2012). Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. *Entropy*, 14(7), 1186-1202.
- Qian, F., Feng, Y., & Ling, J. (2018, June). Condition monitoring of turbine generator using stator winding temperature. In *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 1-5). *IEEE*.
- Wang, L., Zhang, Y., & Feng, J. (2005). On the Euclidean distance of images. *IEEE transactions on pattern analysis and machine intelligence*, 27(8), 1334-1339.
- Yu, J., Jang, J., Yoo, J., Park, J. H., & Kim, S. (2017). Bagged auto-associative kernel regression-based fault detection and identification approach for steam boilers in thermal power plants. *Journal of Electrical Engineering and Technology*, 12(4), 1406-1416.
- Zhang, G., & Gai, Y. (2020). A study on the application of local outlier factor algorithm (LOF) in anomaly detection. *Network Security Technology & Application*, (11), 49-50.



**Kexin Zhang** received her Bachelor's degree in Energy and Power Engineering from Dalian Maritime University in 2020. She is currently pursuing a Master's degree in Power Machinery and Engineering at Dalian University of Technology, China. Her research interests include data-driven condition monitoring and fault diagnosis of large rotating machines, as well as the development and application of digital twins and virtual reality in the intelligent maintenance of industrial equipment.



**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, Optimization and CAE Software 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 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. His main research interests include digital twin, smart maintenance, AI4Science, and predictive analytics for rotatory machines, such as gas turbines, and compressors.



**Huaiyu Hui** was born in Dalian, Liaoning, China, in 1996. He received B.E. degrees in Engineering Mechanics from Xiangtan University, Hunan, China, in 2018. He is currently working toward the Ph.D degree in Engineering Mechanics from Dalian University of Technology, Liaoning, China. His research interests include convolutional neural network, machinery condition monitoring, and fault diagnosis.



**Manman Wei** received her Bachelor's degree in Energy and Power Engineering from Shandong University of Technology in 2022. She is currently pursuing a Master's degree in Power Machinery and Engineering at Dalian University of Technology, China. Her research interests include uncertainty analysis, as well as model verification and correction.