A Similarity Based Modeling Approach for Turbomachine Fault Prediction

Weijian Tang [1], Xiaomo Jiang [2], Haixin Zhao[3], Qing Chen [4], and Yunqing Gong[5]

[1,2,3,4]*Dalian University of Technology, Dalian, Liaoning, 116024, China*
tangweijian@mail.dlut.edu.cn
xiaomojiang2019@dlut.edu.cn
zhxstc@mail.dlut.edu.cn
chenqingdut@mail.dlut.edu.cn
[5]*Shenyang Blower Group Co., Ltd, Shenyang, Liaoning, 110022, China*
gongyunqing@shenguyun.com

ABSTRACT

Faults in the critical components of a turbomachine usually result in unplanned outage, leading to huge loss of properties and life. Condition monitoring becomes a promising tool to provide automatic early alerting of potential damage in critical components thus ensuring the system safety and reliability while lowering its maintenance cost. This is still a challenging hot topic due to the data imperfection and multivariate correlation, as well as the variation of faults and components in different turbomachines. In this paper, a condition monitoring method based on similarity-based model is proposed to solve these problems in fault prediction of large turbine machinery. Bayesian wavelet multi-scale reconstruction is proposed to address the potential noise in the sensed multivariate time historical data. The advanced signal processing balances the over-denoising and under-denoising of raw multivariate signals. An optimized auto-associative kernel regression (OAKR) approach is developed to represent the healthy status of the turbomachine system and further predict its responses under unknown status. The residual error between the estimated and measured values of the OAKR model will become the larger when the turbine machinery has an early fault. The statistical method of moving window is used to detect the change of mean square error of residuals over the time. When the mean square error exceeds a preset threshold, a fault mark will be given. A comparison study is conducted to demonstrate the effectiveness and feasibility of the proposed methodology by using the real-world data and events collected from a centrifugal compressor.

Keywords: Bayesian wavelets, OAKR, turbomachine, fault prediction

1. INTRODUCTION

Faults in a large-scale turbomachine such as gas turbine, steam turbine, or compressor usually result in unplanned outage and even huge loss of properties and life in the fields of power generation, oil & gas, and petrochemistry industries. As the quick development of high-performance computing capacity and artificial intelligence (AI) algorithms in the past decade, real-time condition monitoring system (CMS) has become an increasingly important tool in improving the safety, reliability and performance of a turbomachine, while reducing its unplanned breakdown, and lowering its maintenance costs. In the past decades a wide spectrum of data-driven predictive analytics methods like time series forecasting, machine learning and artificial neural network models have been developed to predict faults in condition monitoring of a turbomachine (Caselitz 2015, Bennouna 2005, Zaher 2009). These methods are generally composed of two main steps, 1) establishment of a high-fidelity predictive model to produce the system response and then 2) determination of a decision threshold to produce alarms when the system response is deviated too much from the actual measurement. Besides existing uncertainties in sensor data, both model establishment and threshold determination contain uncertainties, which would impact the fault prediction accuracy of a turbomachine to some degree. Therefore, it has become of key importance to accurately predict the fault for condition monitoring of a turbomachine considering various uncertainties.

Recently the auto-associative kernel regression (AAKR) method has been developed as a similarity-based model (SBM) for condition monitoring and fault alerting in large-scale turbomachines (Garvey 2007, Di Maio 2013, Fei 2015, Sairam 2016, Yu 2017, Guo 2011, Brandsæter 2017, Qian 2018, Baraldi 2015). This approach utilizes multivariate historical data collected at normal conditions to establish a system identification model representing the

physical system at the healthy status. The yielded model is then used to predict multivariate responses of the system under unknown conditions. A health index calculated from the difference of predicted responses and actual measurements is employed to quantitively assess the status of the system. This approach has demonstrated its high accuracy, flexibility and generalization with a broad spectrum of applications in various types of equipment (Garvey 2007, Di Maio 2013, Fei 2015, Sairam 2016, Yu 2017, Guo 2011, Brandsæter 2017, Qian 2018, Baraldi 2015). It can be effectively applied not only for non-rotating equipment such as boiler in a coal plant (Yu 2017), and gear box in a wind turbine (Guo 2011), but also for rotating machines such as bearings in engine (Brandsæter 2017), and blades in compressor or steam turbines (Qian 2018).

Unlike other parametric regression models in machine learning or nonparametric AI methods, the AAKR approach doesn't require any prior knowledge on the system or component under investigation. It provides a generic flexible nonparametric system identification approach to fault prediction in different multivariate scenarios of a complicated physical system. However, this approach has three main drawbacks. First, it is highly sensitive to the uncertainties existing in the sensor data such that it may produce a number of false alarms and even missing events. Second, the band width parameter in the kernel function of AAKR needs to be tuned from historical sensor data. An improper selection of the band width would result in inaccurate fault prediction. Third and the last, the subjectivity and variation in the general threshold determination play a key role on fault prediction of large-scale turbomachines. Recently some researchers have developed a hybrid model to partially address the above-mentioned issues by combining signal processing techniques with the AAKR method (Di Maio 2013, Yu 2017, Brandsæter 2017, Baraldi 2015). For instance, Di Maio (2013) employed the correlation analysis and generic algorithm to classify the signals and then applied AAKR for fault prediction using the classified signals. Yu (2017) averaged the prediction results obtained from multiple AAKR models to alleviate the effect of outcome variations on prediction accuracy. Brandsæter (2017) combined the k-means clustering and AAKR estimation approaches to improve the computation efficiency of the model. (Baraldi 2015) revised the weights of AAKR model to reduce the correlation of multiple variables. However, none of these has addressed its sensitivity to uncertainties in the data, modeling and thresholding.

It should be noted that determination of an alarming threshold also plays a critical role on fault prediction accuracy of the AAKR approach. Several techniques such as sequential probability ratio test (SPRT) (Di Maio 2013, Fei 2015, Qian 2018), likelihood ratio test (Sairam 2016), and empirical estimation (Guo 2011) are commonly used to determine the alarming threshold level. These methods are statistical methods, which take the statistical characteristics of the 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 time series signals into multi-resolution approximations and details through wavelet functions in the time-frequency domain. The Bayesian hypothesis testing is then applied to judge whether any noise exists in each decomposed coefficient series. The adept combination of multiresolution wavelet analysis and probabilistic Bayesian assessment (Jiang 2007) is able to trade-off the under-denoising and over-denoising of the raw signals, which outperforms the traditional soft or hard thresholding approach in the wavelet fashion. Furthermore, this paper presents a reinforced OAKR method by integrating advanced signal processing, probabilistic principal component analysis, AAKR kernel parameter optimization and adaptive thresholding to address the aforementioned issues of the traditional method for turbomachine fault prediction.

In the following sections, the data preprocessing of Bayesian wavelet denoising is introduced. The data preprocessing is employed to remove the potential noise from the raw data thus improving fault prediction accuracy. The simplex method is employed to adaptatively obtain the optimal band width of kernel function in the conventional AAKR, resulting in the OAKR model. A generic implementation procedure is developed to automate the application process of the proposed methodology for turbomachine fault prediction. The data and event collected from a real-world centrifugal compressor is used to demonstrate the effectiveness of the proposed Bayesian OAKR model. Finally, the conclusion is provided.

2. BAYESIAN OAKR METHOD

2.1. Bayesian multiresolution analysis thresholding

Wavelet multiresolution analysis works like a mathematic microscope broadly used to exploit details in the time series signals (Jiang 2010, Qibing 2013). Unlike the traditional Fourier transform where only the frequency features are extracted from the signals, the wavelet analysis decomposes the raw signals into different levels of coefficients in both 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 details for subsequent analysis.

In this paper, the wavelet packet decomposition is used to process the signal, and the signal whose wavelet coefficient is less than a certain threshold value is regarded as noise. After removing the noise, the inverse wavelet

transformation is carried out to reconstruct it to the time domain signal for subsequent fault identification.

Given a time series with *N* points for a sensor variable in the turbomachine, $x_i(i = 1,2, \dots, N)$, the DWPT approach decomposes it into a set of scaling coefficients $s_j(k)$ and wavelet coefficients $w_j(k)$ simultaneously. Note that the parameter *i* represents the time point in this formulation. The time series can be expressed in the summation of discrete decomposition coefficients as follows:

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

where *j* is the number of decomposition levels, *k* is the coefficient point, and $\varphi_{j,k}(i)$ and $\psi_{j,k}(i)$ are the *k*-th scaling and wavelet functions at time point *i* of the *j*-th level. Refer to Jiang (2007) for details of the DWPT decomposition.

Assuming the signal x_i is contaminated by the additive white Gaussian noise ε_i with the mean of zero and the standard deviation of σ^2, i.e., $\varepsilon_i \sim N(0, \sigma^2)$. The time series x_i can be expressed as:

$$x_i = g_i + \varepsilon_i, \quad i = 1,2, \dots, N \quad (2)$$

where the g_i is noiseless time series. The item ε_i is also split into a series of noise coefficients ε_{jk} together with the decomposition of signals x_i. For the sake of formulation, let d_{jk} represent the *j*-th level and *k*-th coefficient of $s_j(k)$ or $w_j(k)$, given by

$$d_{jk} = \hat{d}_{jk} + \sigma_j \varepsilon_{jk}, j = j_0, \dots, J-1; \quad k = 0,1, \dots, 2^j - 1 \quad (3)$$

where the noise item $\varepsilon_{jk} \sim N(0,1)$ is the independent random variable, $\hat{d}_{jk}$ is the decomposition coefficient after denoising and σ_j is the standard deviation of d_{jk}.

Note that this paper does not focus on the characteristics of signals in a certain wavelet band, but removes the low energy noise signals in each frequency band according to the threshold value. In addition, considering that this paper is focused on the mechanical fault, it is mainly manifested as the overall persistent or periodic deviation from normal operation of response characteristics. Therefore, the actual fault characteristic signals will show higher wavelet energy. The signal with low wavelet energy is regarded as noise, which will not affect the actual fault diagnosis.

Equation (3) can be expressed as the conditional probability distribution given by

$$d_{jk} | \hat{d}_{jk}, \sigma_j^2 \sim N(\hat{d}_{jk}, \sigma_j^2) \quad (4)$$

Bayes is introduced here because it can deal with the uncertainty of the data (Fenton 2011), but Bayes requires prior information, and we usually don't have any prior information about noise.

Assuming a random binary variable γ_{jk} with the Bernouli distribution π_j, we have

$$P(\gamma_{jk} = 1) = 1 - P(\gamma_{jk} = 0) = \pi_j \quad (5)$$

where π_j is usually chosen as 0.5 (Jiang 2007). Thus, an uninformative prior on the coefficient is given as

$$\hat{d}_{jk} | \gamma_{jk} \sim N(0, \gamma_{jk}\tau_j^2) \quad (6)$$

According to Bayes theorem, the posterior distribution of $\hat{d}_{jk}$ can be obtained as follows (Jiang 2007)

$$\bar{d}_{jk} | \gamma_{jk}, d_{jk}, \sigma_j^2 \sim N\left(\gamma_{jk} \frac{\tau_j^2}{\sigma_j^2 + \tau_j^2} d_{jk}, \gamma_{jk} \frac{\sigma_j^2 \tau_j^2}{\sigma_j^2 + \tau_j^2}\right) \quad (7)$$

Based on Eq. (7), the posterior likelihood ratio of $\gamma_{jk} = 0$ and $\gamma_{jk} = 1$, i.e., so-called Bayes factor, is calculated by

$$\eta_{jk} = \frac{1 - \pi_j}{\pi_j} \frac{(\sigma_j^2 + \tau_j^2)^{\frac{1}{2}}}{\sigma_j^2} exp\left(-\frac{\tau_j^2 d^2{}_{jk}}{2\sigma_j^2(\sigma_j^2 + \tau_j^2)}\right) \quad (8)$$

Refer to Jiang (2007) for details of the derivative of Eq. (8). For each d_{jk}, the Bayes factor η_{jk} is calculated to evaluate whether d_{jk} equals to zero or not. The value $\eta_{jk} > 1$ is in support of $d_{jk} = 0$. Otherwise, we have $d_{jk} = \hat{d}_{jk}$. The denoised coefficients are then used to reconstruct a clean time series $\hat{x}_i$ from Eq. (1) for subsequent analysis.

2.2. Optimized AAKR Approach

Auto-Associative Kernel Regression (AAKR) approach is a kind of similarity-based nonparametric modeling technique, which estimates the system response based on the similarity between measurement and memory vectors (Garvey 2007, Guo 2011). In comparison to other nonparametric approaches such as neural network, this method can be trained easily with the historical measurement, independent of the fault modes and equipment types, therefore, it provides a generic multivariate system identification approach for condition monitoring and fault prediction of industrial equipment.

In this study, the data obtained after denoising was used to train AAKR. The memory vector for the model training is stored in a matrix **Y**, in which $Y_{i,j}$ is the *i* vector of the *j* variable, and Y_j corresponds to the variable $\hat{x}_i$ obtained after denoising as described in the previous section. For the n_m-th memory vector, the matrix **Y** can be expressed as

$$\mathbf{Y} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,p} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n_m,1} & Y_{n_m,2} & \cdots & Y_{n_m,p} \end{bmatrix} \quad (9)$$

The $1 \times p$ monitoring response vector **y** of the turbomachine is expressed as

$$\mathbf{y} = [\mathrm{y}_1 \quad \mathrm{y}_2 \quad \cdots \quad \mathrm{y}_p] \tag{10}$$

The model prediction can be obtained by averaging each memory vector in the matrix $\mathbf{Y}_{\boldsymbol{n_m,p}}$ whose weights are estimated by using the healthy data. This paper proposes the optimized AAKR approach (so-called OAKR) by consisting of the following four steps.

First, calculate the commonly used Eurlean distance between the monitoring vector **y** and each memory vector $\mathbf{y}_i$ to produce a $n_m \times 1$ distance vector **d**, given by

$$d_i(\mathbf{Y}_i, \mathbf{y}) = \sqrt{\left(\Delta \mathrm{Y}_{i,1}\right)^2 + \left(\Delta \mathrm{Y}_{i,2}\right)^2 + \cdots + \left(\Delta \mathrm{Y}_{i,p}\right)^2} \tag{11}$$

Second, obtain the $n_m \times 1$ weight vector **w** through fitting the Gaussian kernel function as follows

$$\mathbf{w} = K_h(\mathbf{d}) = \frac{1}{\sqrt{2\pi h^2}} e^{-\boldsymbol{d}^2/h^2} \tag{12}$$

where the parameter h is the band width of the kernel function, which determines the smoothing level of the function. A smaller h can display more details in the data but fails to yield smooth tail of the function, while a larger h can produce relatively smoother transaction but lose details in the system identification, resulting in more false alarms in the condition monitoring of turbomachines. Therefore, the selection of the band width h plays a critical role on the fault prediction accuracy in condition monitoring of industrial equipment.

Third, obtain the optimal band width through automatically minimizing the mean squared error of the model prediction using a set of new health data and Nelder-Mead optimization algorithm, thus reduce the possibility of false alarms of equipment condition monitoring, as described in next section.

Forth and the last, estimate the response $\hat{\mathbf{y}}$ by using the weights matrix **w** and the monitoring vector $\mathbf{y}_i$, as follows

$$\hat{\mathbf{y}} = \frac{\sum_{i=1}^{n_m}(w_i \mathbf{y}_i)}{\sum_{i=1}^{n_m} w_i} \tag{13}$$

It is observed from the OAKR formulation that this SBM approach is related to the measured data only, independent of the fault and equipment under monitoring. As a result, this approach provides a generic powerful tool for condition monitoring of the industrial equipment. During the model establishment the health data representing as many scenarios as possible are usually used to make the model robust.

2.3. Nelder-Mead Optimization

In this study, the AAKR method is improved by automatically selecting the optimal band width to minimize the model prediction error. The Nelder-Mead (NM) simplex algorithm is proposed to achieve this purpose due to its simplicity and effectiveness for online implementation (Lagarias 1998, Liao 2015). The NM algorithm is one of the heuristic optimization methods proposed by Nelder and Mead (1965), which does not require any mathematical derivative of the objective function but very effective in dealing with nonlinear and multi-dimensional problems (Liao 2015). The algorithm doesn't require derivative calculation but a few function values at each iteration, which makes it computationally very efficient and applicable for online implementation. It generally converges in a few iterations. Therefore, it is especially popular for the optimization problem where the derivative of an objective function is hardly obtained explicitly.

According to Lagarias et al. (1998), the NM algorithms requires the predefined coefficients for four operations: reflection (ρ), expansion (χ), contraction (γ) and shrinking (α). In this study the commonly used values $\rho = 1$, χ=2, and $\gamma = \alpha = 0.5$ are taken, and five steps in one iteration is described below for illustration of the NM simplex algorithm.

Step 1 *Order*: determine the order of the objective function $f(h)$ for the band width h in vertices n+1 of the simplex algorithm, such that

$$f(h_1) \le f(h_2) \le \cdots \le f(h_{n+1}) \tag{14}$$

where h_{n+1} represents the simplex value for next iteration, n is the number of simplexes, and $f(h)$ is the mean square error (MSE) between the prediction output and measurement data, calculated by

$$f(h) = \frac{1}{N}\frac{1}{M}\sum_{n=1}^{N}\sum_{m=1}^{M}(y_m - \hat{y}_m(h))^2 \tag{15}$$

where N is the number of samples, M is the number of model variables, and y_m and $\hat{y}_m$ are the measurement data and model output of the m-th variable, respectively.

Step 2 *Reflection*: obtain the reflection value h_r by

$$h_r = (1+\rho)\bar{h} - h_{n+1} \tag{16}$$

where $\bar{h} = \frac{1}{K}\sum_{i=1}^{K} h_i$ is the mean of the K best points. If $f(h_1) \le f(h_r) < f(h_n)$, then obtain the optimal band width $h_{opt} = h_r$ and terminate the iteration. Otherwise, go to next step.

Step 3 *Expansion*: obtain the expansion value h_e if $f(h_r) < f(h_1)$, by

$$h_e = \bar{h} + \chi(h_r - \bar{h}) \tag{17}$$

If $f(h_e) < f(h_r)$ then obtain the optimal band width $h_{opt} = h_e$ and terminate the iteration. Otherwise, update the value $h_{new} = h_r$ and go to next step.

Step 4 *Contraction*: if $f(h_n) \le f(h_r) < f(h_{n+1})$ then obtain the contraction value h_c by

$$h_c = \bar{h} + \gamma(h_r - \bar{h}) \tag{18}$$

If $f(h_c) < f(h_r)$ then obtain the optimal band width $h_{opt} = h_c$ and terminate the iteration. Otherwise, if $f(h_r) \geq f(h_{n+1})$ then obtain the contraction value h_c by

$$h_c = \bar{h} + \gamma(h_{n+1} - \bar{h}) \tag{19}$$

If $f(h_c) < f(h_{n+1})$ then obtain the optimal band width $h_{opt} = h_c$ and terminate the iteration. Otherwise, go to next step.

Step 5 *Shrinkage*: if $f(h_{new}) \geq f(h_{n+1})$ then update the vertices by

$$v_i = h_1 + \alpha(h_i - h_1), i = 2, \dots, n+1 \tag{20}$$

The vertices of the simplex at the next iteration consist of $x_1, v_2, v_3, \dots, v_{n+1}$, and repeat **Steps** 1 to 5 until find the optimal band width. Usually, it takes just fewer than 10 iterations to obtain the optimal value.

2.4. Model Performance Indicator

2.4.1. Mean Square Error

The MSE value obtained by Eq. (15) is employed in this study to quantitively evaluate the model validity and system healthy status. In the model training, the MSE value approaching to 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 yielded to alert the large deviation between the measurement data and the model prediction. In this study the alerting threshold is determined by using the model prediction under the healthy status.

2.4.2. R-square Metric

The *R*-square is defined as follows

$$R\text{-square} = \frac{SSR}{SST} \tag{21}$$

where $SSR = \sum_{m=1}^{M}(\hat{y}_m - \overline{y})^2$ and $SST = \sum_{m=1}^{M}(y_m - \overline{y})^2$, $\overline{y}$ is the mean value of the prediction data. The *R*-square is used to evaluate the performance of the model fitting. A value approaching one indicates a well-fitting model.

2.4.3. Denoising Performance Index

The performance of the denoising signal is evaluated quantitatively by 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) via taking the logarithm at the basis of 10, given by

$$\text{SNR} = 10\log_{10}\left\{\frac{\sum_{i=1}^{N}[x_i]^2}{\sum_{i=1}^{N}[\hat{x}_i - x_i]^2}\right\} \tag{22}$$

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

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2} \tag{23}$$

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

2.4.4. Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC)

ROC curve is an index to evaluate generalization performance of classification models in machine learning field. The horizontal axis of ROC curve is the rate of false positive samples, while the vertical axis is the rate of true positive samples. The area under the ROC curve, or AUC, is usually used to compare the generalization performance of the two classification models. The larger the AUC, the better the generalization performance of the model. In this paper, the ROC curve is used to compare the prediction performance of models under different working conditions. The horizontal axis represents the ratio of true alarms and the vertical axis represents the ratio of false alarms.

2.5. Fault alarming strategy

A rolling window strategy is employed to automatically judge whether the turbomachine has fault or not by evaluating the MSE between the model output and the actual measured values. 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 keeps changing in five continuous windows and exceeding the preset threshold, the turbomachine is judged to be anomaly with fault, thus yielding an alarm. In this study the threshold of $k\sqrt{MSE}$ obtained from the healthy status is used in the fault detection, where K was determined by experience. In this paper, K =1.3 was taken as the threshold value.

3. ILLUSTRATIVE EXAMPLES

The data and events with broken impellers in a real-world large centrifugal compressor are employed in this example to illustrate the effectiveness and feasibility of the proposed methodology.

3.1. Data

The data used in this example is collected from a centrifugal compressor in a petrochemistry factory. The impellers were broken due to the wheel damage resulting from the broken filtering net on March 7, 2019. The operating data from December 28th, 2018 to April 1st, 2019 are used in this example to demonstrate the effectiveness of the proposed methodology.

Figure 1 shows the rotating speed of the compressor over the time, which indicates a varying speed of the machine with the maximum speed up to 6000RPM, producing the high-frequency time series data. Note that in this study the data prior to March 7, 2019 is considered to represent the healthy or nearly healthy conditions of the machine. The sensed time series data of 14 vibration variables are used in this study.

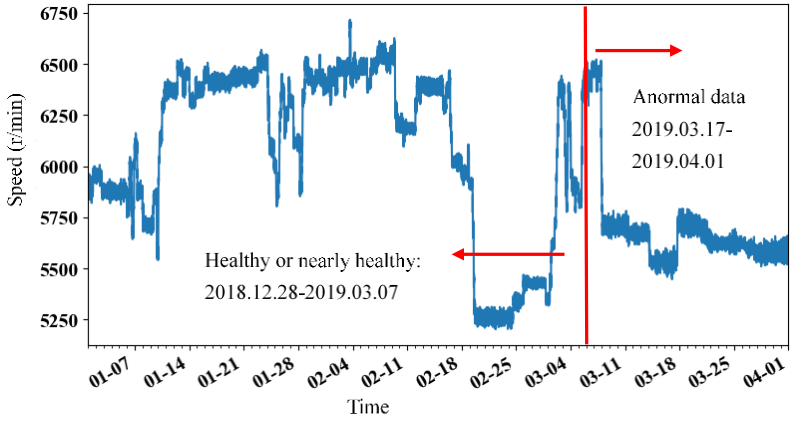


Figure 1 Speed time series

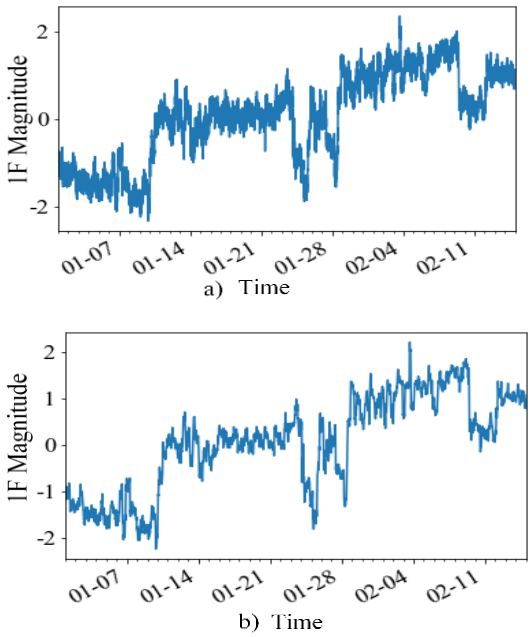


Figure 2 One frequency feature series plot: a) raw data, and b) denoised one

3.2. Data preprocessing

For demonstration purpose, 77 feature features obtained from 7 sensors in the compressor are used in the data preprocessing. The upward and downward data fill methods are used to process the missing values or infinite values in the feature data. The feature data for all variables are normalized to eliminate the amplitude effect of various variables.

3.3. Bayesian wavelet thresholding

The Bayesian wavelet packet denoising approach described previously is employed to remove the noise in each feature data. By a trial-and-error approach the three level DWPT decomposition with Daubechies 8 (Db8) mother wavelet or filters is enough to remove the noise from the signals. As an example, Figure 2 shows the raw time series (Fig. 2a) and the denoised one (Fig. 2b) for one frequency feature. It is clearly observed that the denoised series looks smoother than the raw one. The SNR value of 20db is obtained from Eq. (22), while the RMSE of 0.1 is obtained from Eq. (23), indicating a clean series with little noise.

Table 1. Feature data used for OAKR modeling and alarm prediction

	Period	Data points	MSE	R-square
Training	2018/12/28 -2019/2/1	850	0.022	0.977
Optimization	2019/2/1-2019/2/8	183	0.806	0.263
Testing	2019/2/8-2019/2/15	183	2.106	0.105
Prediction	2019/2/15-2019/4/1	1226	1.902	0.103

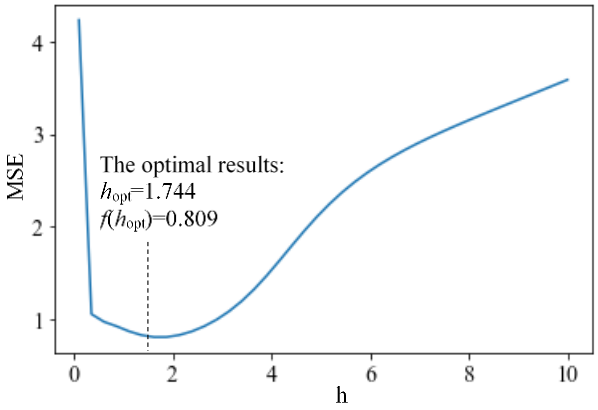


Figure 3 The relation curve between kernel bandwidth and MSE

3.4. OAKR model establishment and validation

The obtained Variables after denoising are used to establish the OAKR model, as summarized in Table 1. The three-month feature data at one-hour interval is divided into four groups: model training, band width optimization, model testing and prediction. The 850 training feature data sets are used to obtain the memory matrix, while the 183 data sets are used to optimize the band width in the kernel density function of the OAKR model by using the simplex method described previously. The optimal band width was obtained through 22 simplex iterations. Figure 3 shows the MSE curve versus the band width with the optimal value of h_{opt} = 1.744 and the minimum MSE of $f(h_{opt})$= 0.8, obtained by Eq. (15).

The obtained model is tested by a set of new data and then used to perform prediction. The MSE of 2.106 and the thresholding coefficient of 1.3 are obtained from the testing data. The threshold used for fault prediction is $1.3\sqrt{MSE}$ = 1.89.

3.5. Model prediction

The trained model is used to predict the system response for the data from Feb 15, 2019 to Apr 1st, 2019 as shown in Table 1. The raw data is preprocessed and cleansed by the Bayesian DWPT thresholding approach. The trained OAKR model is then used to estimate the response of the system. As an illustrative example, Figure 4 shows the comparison of the raw variable and the model prediction (Fig. 4a), and the averaged MSE of the difference values of all variables and their prediction (Fig. 4b).

It is observed from the predicted variable in Fig. 4a that the deviation between the true variable and the prediction result after the event on Mar 7th is significant, while the slight fluctuation is observed on the difference prior to the event. The same observation is made from the averaged MSE trend in Fig. 4b. Based on the threshold of 1.89 obtained previously, the first alarm would be triggered on Feb 20th to alert this event in advance of 15 days, then multiple alarms would be observed after the event of Mar 7th, that is, the MSE points exceeding the threshold (dash line in Fig. 4b).

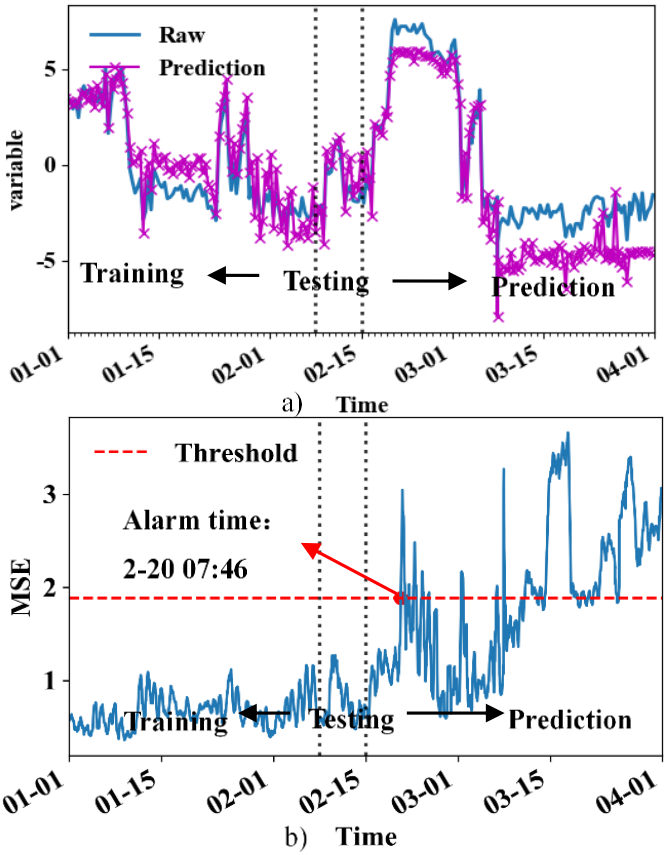


Figure 4 Model prediction result: a) comparison of prediction and raw variable, and b) averaged MSE trend of all results

3.6. Importance of denoising

In order to compare the influence of the denoising module on the accuracy of fault prediction, the undenoised data are used in this section to carry out training, testing and prediction of the OAKR model. Other parameters in the modeling process remain the same as the above OAKR model using the denoising module. Figure 5 shows the averaged MSE of the difference values of all variables and their prediction.

It can be observed that, compared with the prediction results of the denoised model, the MSE trend of the prediction results of the non-denoised model is not very obvious at the early stage of the fault, but only increases significantly when the fault occurs. Fig. 6a and Fig. 6b respectively show the ROC curve and AUC value of the MSE prediction results of the de-noising model (AUC=0.95) and the non-de-

noising model (AUC=0.92). According to the calculation results of AUC, the denoised model has better fault prediction performance. The use of denoising data in the OAKR model can better identify early failure trends.

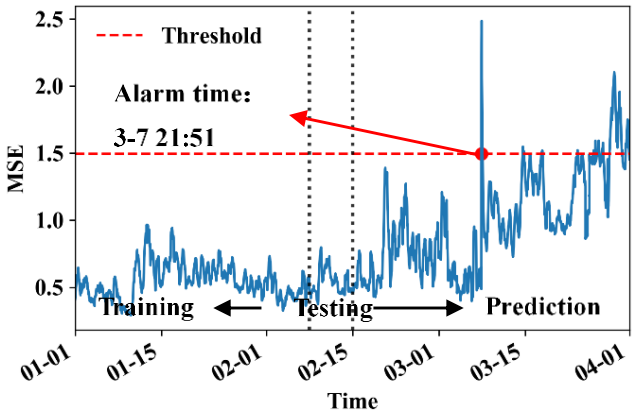
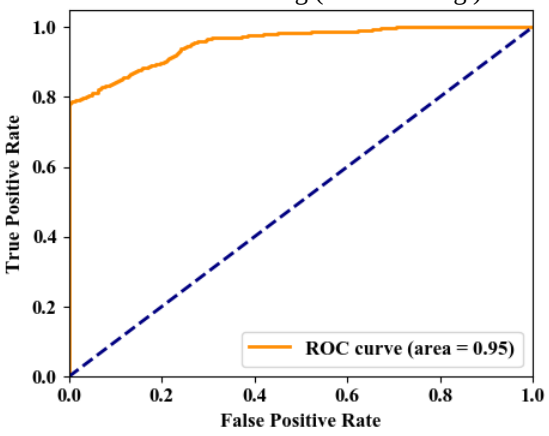
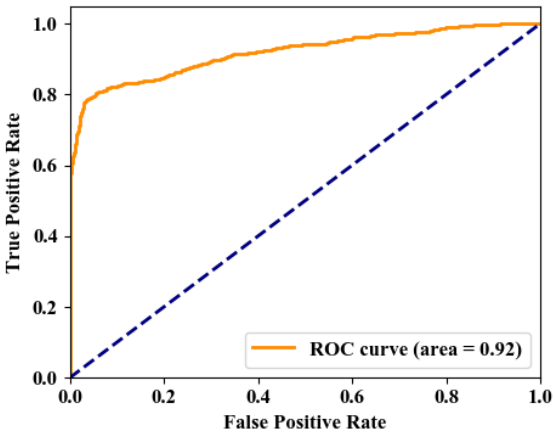


Figure 5 OAKR model prediction results:Averaged MSE and alarming (No denoising)



(a) ROC curve of denoising



(b) ROC curve without denoising

Figure 6 The results of ROC and AUC

4. CONCLUDING REMARKS

This paper presents a novel enhanced probabilistic similarity-based modeling (SBM) methodology for turbomachine fault prediction by seamlessly integrating advanced signal processing techniques with the optimized auto-associative kernel regression (OAKR) approach. The Bayesian discreate wavelet packets thresholding is employed as an advanced time-frequency decomposition approach to cleanse the raw signals by combining the Bayesian hypothesis testing and multiresolution wavelet analysis, thus effectively avoiding under-denoising due to the multiscale signal decomposition and over-denoising due to unbiased judgment on the decomposition coefficients. Instead of manual modification of the band width parameter in the conventional auto-associative kernel regression method, Nelder-Mead simplex algorithm is developed to automatically optimize the key parameter in the model, thus significantly improving the model prediction accuracy and efficiency. The alarming threshold is determined adaptively by a rolling window of the difference between the prediction outputs and the actual values. A generalized procedure is developed to automate the application of the proposed methodology for turbomachine fault prediction under data uncertainties. The effectiveness and feasibility of the proposed probabilistic SBM methodology and procedure is demonstrated with the data and event collected from a real-world large centrifugal compressor.

In future research the OAKR model will be investigated with other types of turbomachines such as gas turbines and steam turbines to illustrate its generalization. A probabilistic alarming strategy would be developed to automate the confidence of the fault prediction for turbomachines under multiple scenarios.

REFERENCES

Caselitz, P., & Giebhardt, J. (2005). Rotor condition monitoring for improved operational safety of offshore wind energy converters. J. *Sol. Energy Eng.*, 127(2), 253-261.

Bennouna, O., Héraud, N., Camblong, H., & Rodriguez, M. (2005). Diagnosis of the doubly-fed induction generator of a wind turbine. *Wind Engineering*, 29(5), 431-447.

Zaher, A. S. A. E., McArthur, S. D. J., Infield, D. G., & Patel, Y. (2009). Online wind turbine fault detection through automated SCADA data analysis. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 12(6), 574-593.

Garvey, J., Garvey, D., Seibert, R., & Hines, J. W. (2007). Validation of on-line monitoring techniques to nuclear plant data. *Nuclear Engineering and Technology*, 39(2), 133.

Di Maio, F., Baraldi, P., Zio, E., & Seraoui, R. (2013). Fault detection in nuclear power plants components by a combination of statistical methods. *IEEE Transactions on Reliability*, 62(4), 833-845.

Fei, X., Research on Distributed Condition Monitoring Method for Primary Circuit of Nuclear Power Plant.*MS Thesis. Harbin Engineering University*, 2015.

Sairam, N., & Mandal, S. (2016, December). Thermocouple sensor fault detection using auto-associative Kernel regression and generalized likelihood ratio test. *In 2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE) (pp. 1-6). IEEE.*

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.

Guo, P., & Bai, N. (2011). Wind turbine gearbox condition monitoring with AAKR and moving window statistic methods. *Energies*, 4(11), 2077-2093.

Brandsæter, A., Vanem, E., & Glad, I. K. (2017, August). Cluster based anomaly detection with applications in the maritime industry. *In 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC) (pp. 328-333). IEEE.*

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.*

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.

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.

Fenton, N., & Neil, M. (2011). The use of Bayes and causal modelling in decision making, uncertainty and risk. CEPIS Upgrade, 12(5), 10-21.

Jiang, S. X., & Liu, Y. Y. (2010, July). Injector waveform analysis and engine fault diagnosis based on frequency space subdivision in wavelet transform. *In 2010 International Conference on Wavelet Analysis and Pattern Recognition (pp. 318-323). IEEE.*

Qibing, J., & Khursheed, S. (2013, July). A wavelet theory about online wavelets denoising based on Moving Window and Principal Component Analysis (PCA). *In 2013 International Conference on Wavelet Analysis and Pattern Recognition (pp. 56-61). IEEE*.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4), 308-313.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder--Mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1), 112-147.

Liao, S. H., Hsieh, J. G., Chang, J. Y., & Lin, C. T. (2015). Training neural networks via simplified hybrid algorithm mixing Nelder–Mead and particle swarm optimization methods. *Soft Computing*, 19(3), 679-689.