

Bayesian Stochastic Neural Network Model for Turbomachinery Damage Prediction

Shuhua Yang, Xiaomo Jiang, Shengli Xu and Xiaofang Wang

School of Energy and Power Engineering, Dalian University of Technology, Dalian, Liaoning, 116024, China

yshemail@sina.com

jiang.xiaomo@gmail.com

xusl@dlut.edu.cn

dlwxf@dlut.edu.cn

ABSTRACT

Turbomachinery often suffers various defects such as impeller cracking, resulting in forced outage, increased maintenance costs, and reduced productivity. Condition monitoring and damage prognostics has been widely used as an increasingly important and powerful tool to improve the system availability, reliability, performance, and maintainability, but still very challenging due to multiple sources of data uncertainties and the complexity of analytics modeling. This paper presents an intelligent probabilistic methodology for anomaly prediction of high-fidelity turbomachine, considering multiple data imperfections and multivariate correlation. The proposed method adeptly integrates several advanced state-of-the-art signal processing and artificial intelligence techniques: wavelet multi-resolution decomposition, Bayesian hypothesis testing, probabilistic principal component analysis, and fuzzy stochastic neural network modeling. The advanced signal processing is employed to reduce dimensionality and to address multivariate correlation and data uncertainty for damage prediction. Instead of conventionally using raw time series data, principal components are utilized in the establishment of stochastic neural network model and anomaly prediction. Bayesian interval hypothesis testing metric is then presented to quantitatively compare the predicted and measured data for model validation and anomaly evaluation, thus providing a confidence indicator to judge the model quality and evaluate the equipment status. Bayesian method is developed in this study for denoising the raw data with multiresolution wavelet decomposition, quantifying the model accuracy, and assessing the equipment status. The dynamic stochastic neural network model is established via the nonlinear autoregressive moving average with exogenous inputs approach. It seamlessly integrates the

fuzzy clustering and independent Bernoulli random function into radial basis function neural network. A natural gradient method based on Kullback-Leibler distance criterion is employed to maximize the log-likelihood loss function. The effectiveness of the proposed methodology and procedure is demonstrated with the 11-variable time series data and the forced outage event of a real-world centrifugal compressor.

1. INTRODUCTION

High-fidelity turbomachinery like centrifugal compressors and turbines often suffers various defects such as impeller cracking, resulting in forced outage or breakdown of the total plant, increased maintenance costs, and reduced productivity. The research conducted by Li, Zhang and Xu (2013), for instance, indicates that about 65% of issues in a centrifugal compressor are related to impellers or blades, and among them 40% related to blades have not been well understood due to complicated loading conditions. The replacement of cracked blade in a centrifugal compressor may cost the plant up to \$500,000 a day. Real-time condition monitoring of a mechanical system, along with anomaly diagnostics and damage prognostics, aims to monitor whether the system is being appropriately operated and maintained to yield optimal performance and maximal reliability. It proactively detects potential failure mode by triggering an alarm to avoid unplanned outages of the system, identifies root cause for the detected anomaly, and further provides actionable recommendations to advise the optimal operation or predictive maintenance of the plant. Condition monitoring and damage prognostics has therefore become an increasingly important tool to improve the system availability, reliability, performance, and maintainability, thus facilitating customers to increase operating revenues, reduce fuel costs, ensure safe and secure work environment, and enhance customer dispatch competitiveness. It usually involves automatically assessing the condition of a system and its components and predicting their remaining use time under current conditions through applying advanced signal

Shuhua Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

processing techniques and analytical algorithms on multivariate historical data. Analytics has attracted increasing attention in condition monitoring of turbomachinery to increase the efficiency of a plant, diagnose the root causes of faults in equipment, and enable a fleet-wide optimization of asset performance. Condition monitoring and damage prognostics of such a turbomachinery system, however, involve multiple sources of data uncertainties and complexity of modeling, making the problem very challenging.

The key to success of predictive monitoring includes two factors: reliable predictive model and quantitative alarming analytics. The predictive model needs to be established and trained from a set of historical operation data without any assumed anomaly, with the purpose of representing the system behavior under normal operation condition. The established model is next used to predict the system response under unknown conditions. The predicted response is compared with the sensed system response. A decision logic is then developed to judge the status of the system under the operation and trigger an alarm if the indicator representing the system status exceeds the predefined threshold. This paper attempts to develop an intelligent probabilistic modeling methodology for automatic damage prediction and anomaly alarming of turbomachinery, based on the adept integration of advanced signal processing, pattern recognition, intelligent algorithms, and probabilistic decision-making methods.

Over the past decade numerous researchers have developed conventional data analysis methods (e.g., Baydar, Chen, Ball & Kruger, 2001; Eftekharijad, Carrasco, Charnley & Mba, 2011; Galka & Tabaszewski, 2011; McFadden & Toozhy, 2000; Wang, Chen, Wu & Wu, 2001), signal processing techniques (e.g., Al-Badour, Sunar, Cheded, 2011; Liu, Guo, Wang, An, Guo & Lian, 2010; Rai & Mohanty, 2007) and artificial intelligence (AI) algorithms (e.g., Chen, Taniguchi, Toyota & He, 2005; Samanta & Al-Balushi, 2003; Wang, Golnaraghi & Ismail, 2004) for condition monitoring of turbomachinery. For instance, Baydar *et al.* (2001) applied the Hilbert–Huang Transform technique to analyze the vibration signal through intrinsic mode functions obtained from the empirical mode decomposition. The fast Fourier transform (FFT) technique is then applied on the intrinsic mode functions for bearing fault diagnosis.

Artificial intelligence-based methods have recently become increasingly popular in the damage diagnostics of rotating machinery. For example, Chen *et al.* (2005) combined the instantaneous power spectrum and genetic programming for fault diagnostics of plant machinery. Samanta and Al-Balushi (2003) presented the two-layer backpropagation-based artificial neural network for fault diagnosis of rolling element bearings. Wang *et al.* (2004) evaluated the recurrent neural networks and neuro-fuzzy approach for fault prognostics of rotating machinery. The neuro-fuzzy method was then

adopted to develop an online machine fault prognostic system. Recently, Edwards, Lee and Friswell (1998) provided a comprehensive literature review of the state of the art on the fault diagnosis of rotating machine mainly from the failure mechanism perspectives. It is recommended that more research on online condition monitoring of rotating machine is needed with emphasis on the data-driven expert system and neural network methods. Jardine, Lin and Banjevic (2006) provided another comprehensive literature review on various methods for machinery diagnostics prior to 2006, while most recently Lee, Wu, Zhao, Ghaffari, Liao and Siegel (2014) presented a state-of-the-art review on prognostics and health management for rotary machine. Both have emphasized that the signal processing techniques play a key role on the diagnostics and prognostics of a rotary machine.

These methods reviewed in the above-referenced literature have not fully addressed several important issues in the damage prognostics of large turbomachinery: 1) imperfection of the experimental data or field measurements, 2) impact of multivariate correlation on the damage prognostics, 3) nonlinear and stochastic behavior of rotating machine, 4) quantitative assessment of predictive models and 5) damage degree of the structural system under investigation. To the best knowledge of the authors, no other research has yet been reported on developing the stochastic intelligent model for damage prediction of turbomachinery under uncertainty, particularly addressing all the five challenges mentioned above. These challenges will be addressed through the intelligent probabilistic methodology presented in this study.

In the past decades, various data driven nonparametric methods have been widely pursued in health monitoring and damage detection of structural systems (e.g., Adeli & Jiang, 2006; Ghanem & Shinozuka, 1995; Hung, Huang, Wen & Hsu, 2003; Jiang & Adeli, 2005; Jiang, Mahadevan & Adeli, 2007; Lei, Jiang & Xu, 2012; Masri, Nakamura, Chassiakos & Caughey, 1996; Masri, Smyth, Chassiakos, Caughey & Hunter, 2000; Nakamura, Masri, Chassiakos & Caughey, 1998; Wu, Xu & Yokoyama, 2002). Ghanem and Shinozuka (1995) comprehensively presented the system identification theory and guidelines for various engineering applications. After then a number of authors (Adeli & Jiang, 2006; Hung *et al.*, 2003; Jiang & Adeli, 2005; Masri *et al.* 1996; Masri *et al.* 2000; Nakamura *et al.* 1998; Wu *et al.* 2002) developed various neural networks based artificial intelligence method for system identification and damage detection. Jiang *et al.* (2007) integrated the Bayesian theory with wavelet packet transform method for data denoising in structural system identification. Lei *et al.* (2012) applied an extended Kalman method for system identification using limited input and output measurement data. Unlike parametric methods, the identification model in the nonparametric approach does not represent any physical quantity directly, instead it is trained to approximate the physical structure and predict its response. In addition, the nonlinear autoregressive moving average with exogenous inputs (NARMAX) approach (Chatfield,

2004) is commonly used in the nonparametric methods for mapping the input-output relationship. The approach has flexibility in effectively representing system nonlinearity and has proven to be a powerful tool for mapping the nonlinear input-output relationship in system identification (e.g., Juang, 1994), therefore the NARMAX method has been employed in the nonparametric approach of this study.

Recently Jiang and his coauthors (e.g., Jiang & Adeli, 2007; Jiang & Mahadevan, 2008a&b; Jiang & Foster, 2013; Jiang & Foster, 2014; Xu, Jiang, Huang, Yang & Wang, 2016) developed various nonparametric methods for system identification, damage detection, and condition monitoring. In particular, Xu *et al.* (2016) proposed a probabilistic signal processing methodology for damage diagnosis of centrifugal compressors. The method adeptly integrates three advanced, state-of-the-art data mining techniques: discrete wavelet packet transform, Bayesian hypothesis testing, and probabilistic principal components analysis (PPCA). Bayesian wavelet approach (Jiang *et al.*, 2007) is employed to clean the raw data in order to improve the accuracy of anomaly diagnostics from multivariate time series data. Refer to (Jiang *et al.*, 2007) for the state-of-the-art review on the limitations of the conventional wavelet denoising approaches. In this study, the Bayesian wavelets approach is employed to provide more accurate data denoising particularly for the measured data even with uncertainties. The discrete wavelet packets decomposition provides more coefficients representing additional subtle details of the signal, thus avoiding under-denoising, and Bayesian hypothesis testing or the ratio of posterior odds approach provides a direct means to assess whether there is noise in the data, thus avoiding over-denoising. Refer to Abramovich, Sapatinas and Silverman (1998), Coifman and Donoho (1995) and Daubechies (1998) for more details about the wavelet decomposition, and Kass and Raftery (1995) and Vidakovic (1998) about Bayesian hypothesis testing. Besides, the PPCA technique is developed to quantify measurement errors (one type of imprecision), eliminate incoherence, and reduce dimension from multivariate sensor data for damage diagnostics. In this study, the preprocessed data is employed in the nonparametric modeling and damage prediction with the purpose of improving computing efficiency and prediction accuracy.

The *stochastic neural network* (SNN) has been demonstrate to be a powerful tool for effective approximation of complex nonlinear stochastic systems in financial engineering (Kamitsuji & Shibata, 2003; Lai & Wong, 2001) and mechanical engineering (Peeters & Roeck, 2001). More recently, Jiang, Mahadevan and Yuan (2017) developed a dynamic fuzzy stochastic neural network model for nonparametric system identification. The model is developed to handle two types of imprecision in the sensed data: fuzzy information and measurement uncertainties. A fuzzy C-means clustering algorithm is employed as a data mining tool to divide the sensed data into clusters with common features.

The fuzzy stochastic model is created by combining the fuzzy clusters of input vectors with the radial basis activation functions in the stochastic neural network. A natural gradient method is developed based on the Kullback-Leibler distance criterion for quick convergence of the model training. The model is validated using a Bayesian hypothesis testing-based metric and demonstrated with different numerical examples. This study is the first time to develop this fuzzy SNN model for damage prediction of turbomachinery.

This study presents a Bayesian stochastic neural network method for automatic damage prediction and anomaly alarming in turbomachinery. To the best knowledge of the authors, this is the first attempt to adeptly integrate advanced signal processing, AI algorithm and probabilistic assessment methods for automatic damage prediction of turbomachinery. In this paper the stochastic neural network model is trained to predict the system response. Instead of directly using raw time series data as used in Jiang *et al.* (2017) for structural system identification, principal components obtained from wavelet PPCA approach are employed in the model establishment and response prediction, thus improving the damage prediction accuracy and efficiency. Bayesian interval hypothesis testing metric is then presented to quantitatively compare the predicted and measured time series and provide a confidence indicator to evaluate the status of the system. Bayesian probabilistic method is developed in this study for three purposes: thresholding the raw data with multiresolution wavelet decomposition, quantifying the model accuracy, and assessing the system status. A generalized procedure is provided to implement the proposed methodology. The method and procedure are demonstrated with the time series data and a failure event collected from a real-world centrifugal compressor.

2. STOCHASTIC NEURAL NETWORK PREDICTION MODEL

The mathematical model embedded in a neural network method is generally established from the input-output data sets collected from an underlying system, expressed as follows:

$$y_k = f(\mathbf{X}_k) + e_k \quad (1)$$

where the variable \mathbf{X}_k represents the k -th input vector (time t) in the NARMAX approach. In this study, rather than directly using multivariate time series data in Jiang *et al.* (2017), the multi-dimension principal components obtained from wavelet PPCA approach as described in Xu *et al.* (2016) are employed as the input-output set $\{\mathbf{X}_k, y_k\}$ in Eq. (1). In addition, rather than single output in Jiang *et al.* (2017), multiple output variables are predicted simultaneously from Eq. (1).

In Eq. (1), the symbol $f(\cdot)$ represents a scalar nonlinear stochastic neural network mapping or approximation function in this study, and e_k is the error between the actual and estimated values of the k -th future output y_k (i.e., at time

t). The error is an independent and identically-distributed (iid) zero-mean random variable, i.e., $e_k \sim N(0, \sigma^2)$, in which σ is the deviation of e_k . Figure 1 shows the proposed hybrid stochastic neural network model for damage prediction. The establishment of the predictive model represented by Eq. (1) consists of five steps: (1) preprocessing the raw data, (2) extracting the PCA features, (3) constructing the SNN model, (4) training the model, and (5) validating the model, as described subsequently.

2.1. Bayesian Wavelet Multiscale Analysis

Sensor data continuously collected from an operational turbomachinery is usually stored in a form of time series at a certain interval (e.g., 5 minutes). Sensor data often contains a wide spectrum of imperfections, such as noise, error and incoherence. In past decades, various methods have been presented to remove noise from sensor data, however, little study has been conducted to assess multiple imperfections simultaneously for anomaly detection. In this study, the comprehensive data mining methodology presented in Xu *et al.* (2016) is employed as part of the damage prediction framework to handle multiple imperfections in the raw signals and reduce data dimension for subsequent predictive modeling, thus improving the prediction accuracy. Bayesian

wavelet method is employed to remove noise, while multi-scale wavelet analysis is used to analyze the contexts in the signal.

Wavelets provide an effective hierarchical framework to interpret the information context in a signal through the multi-resolution time-frequency decomposition. The transient changes in the signal context are usually represented in the details at the fine resolution of a wavelet decomposition. Coifman and Wickerhauser (1992) proposed the *wavelet packet transform* (WPT) analysis to allow for a finer and adjustable resolution in the high frequencies (details). Compared with conventional wavelet transform methods, the discrete wavelet packet transform (DWPT) method is a more effective approach to extract features from either stationary or non-stationary signals to represent the underlying dynamic systems (Mallat, 1989). In this study, the DWPT approach is utilized for denoising and multiscale analysis of the i -th original time series $\mathbf{S}_{i \times N}$ ($i = 1, \dots, q$), each containing N raw data points. Each time series is decomposed into p sets of coefficients $\mathbf{A}_{p \times N}$, in which $p = 2^j$ is the number of wavelet coefficients for j -th level decomposition (e.g., $j = 1, 2$, and 3 for the 3-level decomposition). A Bayesian thresholding approach is then applied on each decomposed coefficient series to remove the possible noise. A cleansed time series is

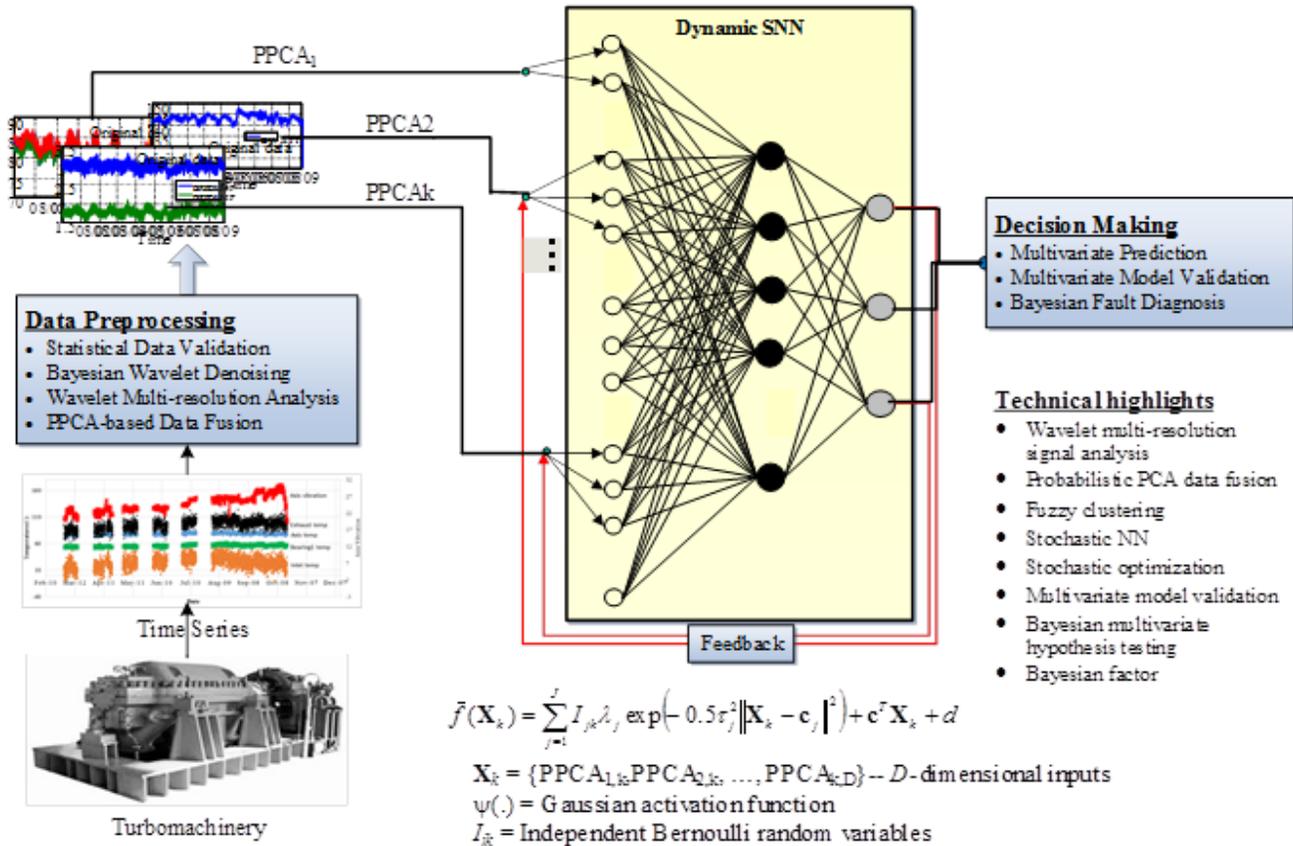


Figure 1. Illustration of hybrid SNN damage prediction method

reconstructed from the denoised coefficients for subsequent principal component analysis. The Bayesian wavelet denoising method is briefly described below. Refer to Jiang *et al.* (2007) for more details about this method.

In the DWPT approach, a given time series with N data points, $f(t_i)$ ($i = 1, 2, \dots, N$), is simultaneously decomposed into a series of scaling coefficients, $s_j(k)$, and wavelet coefficients, $w_j(k)$. The time series is then represented by the inverse wavelet transform and the DWPT coefficients, i.e., $s_j(k)$ and $w_j(k)$, as follows:

$$\bar{f}(t_i) = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} [s_j(k) \varphi_{j,k}(t_i) + w_j(k) \psi_{j,k}(t_i)] \quad (2)$$

where $\bar{f}(t_i)$ is the reconstructed time series after removing possibly noisy components from the decomposed coefficients described below, and the double summation indicates that the scaling and wavelet subspaces are simultaneously split into second-level subspaces to provide the frequency and time breakdown of the signal. The total number of decomposed coefficients at the j -th level, $L = 2^j$, is obtained from the DWPT approach. The discrete values of the original signal are used for initial values of the scaling coefficients. A recursive way is utilized to compute the DWPT coefficients of a given time series. Refer to Burrus, Gopinath and Guo (1998) for details regarding the computation of the coefficients.

For the sake of generality, let d_{jk} denote the k -th decomposed coefficients at the j -th level ($j = j_0, \dots, J-1; k = 0, 1, \dots, 2^j-1$), e.g., $s_j(k)$ and $w_j(k)$ in Eq. (2), obtained by the DWPT approach. Based on the Bayes' rule, the posterior distribution of the cleaned coefficients, \bar{d}_{jk} , is obtained as follows (Jiang *et al.*, 2007):

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

where the symbol \sim denotes a distribution; $N(a, b)$ is a normal distribution with the mean of a and the standard deviation of b ; \hat{d}_{jk} is the distribution of the decomposed coefficients; γ_{jk} is a binary random variable with independent Bernoulli distribution π_j , i.e., $P(\gamma_{jk}=1) = 1 - P(\gamma_{jk}=0) = \pi_j$. It determines whether the coefficients \bar{d}_{jk} are zero ($\gamma_{jk}=0$) or nonzero ($\gamma_{jk}=1$). The variance τ_j^2 represents the magnitude of \bar{d}_{jk} at the j -th level decomposition level. In practical applications, the same values of τ_j^2 and π_j may be assigned for all coefficients in the j -th level. The values of $\pi_j = 0.5$ and $\tau_j = 1$ are chosen for the non-informative priors in the example presented in this paper. The standard deviation σ_j is estimated from the wavelet coefficients of the j -th level DWPT decomposition through dividing the median of the wavelet

coefficients by a factor. The value of 0.6745 suggested by Percival and Walden (2000) for the factor is chosen in the examples presented in this paper.

Bayesian hypothesis test is then employed to threshold the decomposed coefficients by determining whether accepting the null hypothesis, $H_0: \bar{d}_{jk} = 0$. The thresholding rule is defined as follows:

$$\tilde{d}_{jk} = \hat{d}_{jk} I(\eta_{jk} < 1) \quad (4)$$

where η_{jk} is the posterior odds ratio of $\gamma_{jk}=0$ versus $\gamma_{jk}=1$, and $I(\cdot)$ is an indicator function. $I(\cdot)$ is equal to unity when $\eta_{jk} < 1$, which indicates that H_0 is rejected and the coefficient \bar{d}_{jk} is estimated by \hat{d}_{jk} . Otherwise, $I(\cdot)$ is equal to zero and \hat{d}_{jk} is removed. The variable η_{jk} is derived from the conditional probability of γ_{jk} , representing the Bayes factor. The thresholded coefficients \tilde{d}_{jk} obtained from Eq. (4) are used to reconstruct the cleaned time series based on Eq. (2), i.e., $\tilde{f}(t_i)$.

2.2. Probabilistic Principal Component Analysis

After cleaning the multivariate time series data, the probabilistic principal component analysis (PPCA) approach (Tipping & Bishop, 1999) is utilized to (1) reduce data dimensionality, (2) address the multivariate correlation, and (3) handle data uncertainty. The PPCA is derived from a Gaussian latent variable model which is closely related to statistical factor analysis. The factor analysis is a mathematical technique used to reduce the number of variables (dimensionality reduction), while identifying the underlying factors that explain the correlations among multiple variables (Bezdek, 1981). For the convenience of formulation, let $\mathbf{y}_i = \tilde{f}(t_i) \in \mathfrak{R}^q$ represent the q -dimension real numbers, and $\mathbf{Y} = [y_1, \dots, y_N]^T$ denote the $N \times q$ data matrix representing the q variables, each containing N cleaned time series data points, $\tilde{f}(t_i)$. Let $\Phi = [\theta_1, \dots, \theta_N]^T$ be the $N \times d$ data matrix with $\theta_i \in \mathfrak{R}^d$ ($d \leq q$) representing d latent variables (factors) that cannot be observed, each containing the corresponding N positions in the latent space. The latent variables θ_i are conventionally defined to be independently distributed Gaussian variables with zero mean and unit variance, i.e. $\theta_i \sim N(0, \mathbf{I})$. The latent variable model relates the observable variable \mathbf{y}_i in the correlated data matrix \mathbf{Y} to the corresponding uncorrelated latent variable matrix Φ , expressed in the Gaussian distribution form as

$$\mathbf{y}_i | (\theta_i, \mathbf{W}, \boldsymbol{\psi}) \sim N(\mathbf{W}\theta_i + \boldsymbol{\mu}, \boldsymbol{\psi}), \quad i = 1, 2, \dots, N \quad (5)$$

where $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}$ is the isotropic noise covariance, the $q \times d$ weight matrix \mathbf{W} describes the relationship between the two sets of variables \mathbf{y}_i and $\boldsymbol{\theta}_i$, and the parameter vector $\boldsymbol{\mu}$ consists of q mean values obtained from the data matrix \mathbf{Y} , i.e. $\boldsymbol{\mu} = (1/N) \sum_{i=1}^N \mathbf{y}_i$. The unadjusted data in the lower dimensional latent space, $\boldsymbol{\Phi}^* = [\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_N^*]^T$, is obtained as:

$$\boldsymbol{\Phi}^* = \mathbf{M}_{ML}^{-1} \mathbf{W}_{ML}^T \mathbf{Y} \quad (6)$$

where $\mathbf{M}_{ML} = \sigma_{ML}^2 \mathbf{I} + \mathbf{W}_{ML}^T \mathbf{W}$. Equation (6) has the mean of $\mathbf{M}_{ML}^{-1} \mathbf{W}_{ML}^T \boldsymbol{\mu}$. The d -dimensional Principal Components (PC) matrix $\boldsymbol{\Phi}^*$ is applied as inputs to data-driven predictive modeling for damage prediction of the centrifugal compressor.

2.3. Constructing Fuzzy SNN Model

Different types of neural network models may be employed to approximate a mechanical system. The nonlinear fuzzy stochastic neural network model, originally developed by Jiang *et al.* (2017) for system identification, is developed for damage prediction in this study. The general dynamic input-output mapping in the model is expressed as follows:

$$\hat{y}_k = f(\mathbf{X}_k) = \sum_{j=1}^J I_{jk} \lambda_j \exp\left(-0.5\tau_j^2 \|\mathbf{X}_k - \mathbf{c}_j\|^2\right) + \mathbf{c}^T \mathbf{X}_k + d \quad (7)$$

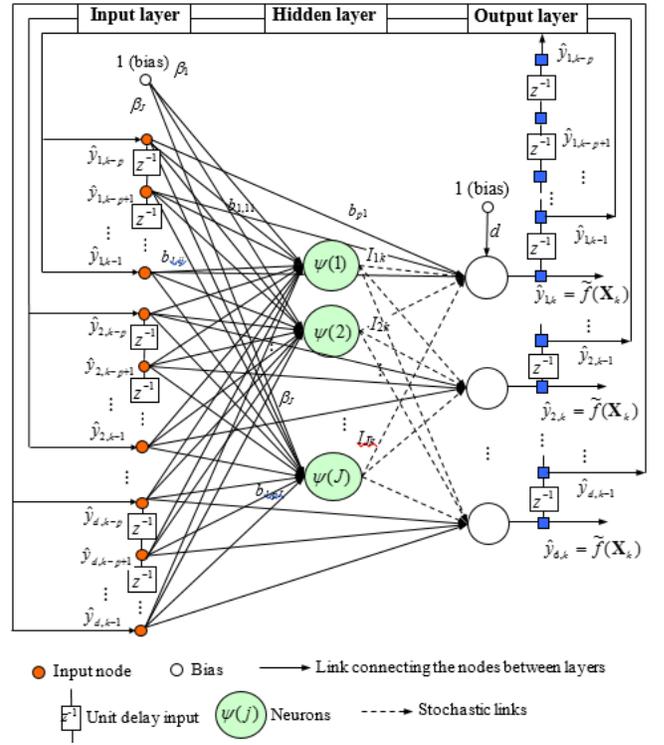
$k = 1, 2, \dots, N_{\max}$

where $\hat{y}_k = f(\mathbf{X}_k)$ is the k -th estimated value of the future output (principal component), \mathbf{X}_k is the PCs obtained from the PPCA of historical time series, and I_{jk} are independent Bernoulli random variables with $P\{I_{jk} = 1 | \mathbf{X}_k\} = 1 - P\{\delta_{jk} = 0 | \mathbf{X}_k\} = \psi(\alpha_j + \mathbf{a}_j^T \mathbf{X}_k)$, in which $\psi(u) = 1/(1 + e^{-u})$ is the logistic function with $\psi(-\infty) = 0$ and $\psi(\infty) = 1$. The symbol $\|\cdot\|$ denotes the Euclidean distance between two vectors, the parameter λ_j is the deterministic weight of the link of the j -th hidden node to the output; the Gaussian (bell-shaped) transfer function $\psi(j) = \exp\left(-0.5\tau_j^2 \|\mathbf{X}_k - \mathbf{c}_j\|^2\right)$ is used as the nonlinear activation function to form a radial basis function neural network (RBFNN), in which τ_j is the j -th spread parameter and \mathbf{c}_j is the j -th center of the Gaussian function, each having p values. The parameter \mathbf{c}_j is also referred to as the cluster of the multidimensional input vector to be obtained by using the fuzzy C-means (FCM) clustering approach (Moody & Darken, 1989) in this study. In this model, the multiplicity $I_{jk}\psi(j)$ is treated as a stochastic neuron. The RBFNN is chosen to construct the SNN model because it has been demonstrated to be an effective approach to model a nonlinear dynamic system with the least computational effort (Kennel, Brown & Abarbanel, 1992).

Figure 2 shows the feedback architecture of the three-layer dynamic SNN model with multiple-PC inputs multiple-PC

outputs, consisting of an input layer, a hidden layer, and an output layer. There are p input nodes in the input layer for each principal component. The ellipses in the hidden layer represent the activation functions $\psi(j)$ in Eq. (7). The dashed arrows represent the stochastic weights (i.e., I_{jk}) of the links connecting the neurons in the hidden layer to the PC output, which is different from traditional neural network models with deterministic weights for all links. Note that the inputs consist of the original PC series while the outputs are the approximated PCs. The previous p outputs for each PC series are fed back to the input layer to predict the current output.

The performance of the fuzzy SNN model is affected by three important factors. First, the input dimension for each PC series, p , for the NARMAX approach, is chosen by the attractor concept in the chaos theory, i.e., false nearest



$\mathbf{X}_k =$ Input vector $= [y_{k1}, y_{k2}, \dots, y_{kp}]^T$, ($k = 1, 2, \dots, N_{\max}$);

N_{\max} = number of input vectors; p = Dimension of input vector; J = Number of neurons; b_{ji} = Weight of the link connecting the i -th input node to the j -th neuron in the hidden layer ($i = 1, 2, \dots, p; j = 1, 2, \dots, J$); I_{jk} = k -th output of stochastic neural network of i -th principal component; $\psi(j)$ = j -th neuron in the hidden layer; β_j = Weight of the link connecting the bias to the j -th neuron. I_{jk} = Stochastic weight of the link connecting the j -th node in the hidden layer to the output node.

Figure 2. Architecture of the proposed MIMO dynamic stochastic neural network

neighbor (FNN) method. The FNN approach assumes that a small embedding dimension results in state space points that are far apart in the original state space to be considered neighboring points in the reconstructed state space (Moody & Darken, 1989). Second, the centers for the RBFNN model are determined by using the FCM clustering algorithm. Given the centers \mathbf{c}_j , the total number of parameters to be modified in the fuzzy SNN model is obtained by $n_p = 3J + J \times d \times p + d \times p + 1$, which consists of $J \times d \times p$ parameters a_{ij} ($i = 1, \dots, d \times p, j = 1, \dots, J$), J parameters α_j , hidden weights λ_j , and spread parameters τ_j ($j = 1, \dots, J$), $d \times p$ input weights b_i , and one bias d . The third influencing factor is the number of neurons, J . In this study, the Bayesian information criterion (BIC) [referred to as the Schwarz criterion by Kass and Raftery (1995)] is applied to select the optimum number of neurons to be used in the Fuzzy SNN model. The approach selects a simpler model (in terms of the number of parameters) with a good fit of the data based on Bayes' theorem.

2.4. Training the Stochastic Model

A natural gradient method based on the Kullback-Leibler distance criterion is developed to effectively avoid being entrapped in a local minimum. This approach has been observed to converge very quickly in training multilayer fuzzy SSN models (Jiang *et al.* 2017). The loss or objective function, $D(\boldsymbol{\theta})$, is defined as the complete-data log-likelihood function expressed as follows:

$$\begin{aligned} D(\boldsymbol{\theta}) &= -\log l(\mathbf{X}, \mathbf{y}, \boldsymbol{\theta}, \mathbf{I}) = -\log g(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \mathbf{I})h(\mathbf{X}) \\ &= -\log g(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \mathbf{I}) - \log h(\mathbf{X}) \end{aligned} \quad (8)$$

where \mathbf{y} represents a d -dimension PC output vector in the model, \mathbf{X} represents all input PC vectors corresponding to \mathbf{y} , $l(\mathbf{X}, \mathbf{y}, \boldsymbol{\theta}, \mathbf{I})$ is the likelihood function of \mathbf{y} , and $h(\mathbf{X})$ is the probability density function (pdf) of the input vector \mathbf{X} (representing the past PC series) which is independent of parameters $\boldsymbol{\theta}$. In training the model, the input vector \mathbf{X} consists of the previous prediction PC data. Therefore, Equation (8) represents the Kullback-Leibler distance that measures the similarity between the probability distribution of model prediction $h(\mathbf{X})$ and that of actual PCs $g(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \mathbf{I})$, in terms of entropy information of the data. Refer to Jiang *et al.* (2017) for more details about the model training.

For the simplicity of formulation, let $\psi_{jk} = \psi(\alpha_j + \mathbf{a}_j^T \mathbf{X}_k)$ ($j = 1, \dots, J; k = 1, \dots, N_{\max}$) represents the logistic function at the j -th neuron for the k -th input vector, in which J and N_{\max} are the optimum number of hidden neurons and the maximum training data sets, respectively. For all features, the objective function in Eq. (8) can be expressed as

$$\begin{aligned} D(\boldsymbol{\theta}) &= \sum_{k=1}^{N_{\max}} \left\{ \frac{1}{2\sigma^2} [y_k - \hat{y}_k]^2 - \sum_{j=1}^J [\mathbf{I}_{jk} \log(\psi_{jk}) + \right. \\ &\quad \left. (1 - \mathbf{I}_{jk}) \log(1 - \psi_{jk})] - \log h(\mathbf{X}_k) \right\} + \frac{N_{\max} \log(2\pi\sigma^2)}{2} \end{aligned} \quad (9)$$

where $\hat{y}_k = f(\mathbf{X}_k, \boldsymbol{\theta})$ is the model prediction defined by Eq. (7). Next the natural gradient learning algorithm based on the Kullback-Leibler distance criterion is pursued to find the optimum $\boldsymbol{\theta}^*$ that maximizes the log likelihood function (Eq. 9). Note that the natural gradient learning has the same form as the online Gauss-Newton method for the sum-of-squares error function. The difference lies in that the former is based on the information theory considering the uncertainties in both feature data and predictive model, which is not considered in the latter based on the sum-of-squares error function. Refer to Jiang *et al.* (2017) for the details about the natural gradient learning method for model training. The established SNN model is validated by Bayesian metric and the features obtained from a set of new data, as described subsequently.

3. MODEL VALIDATION AND DAMAGE DETECTION

The validated model is used to predict the response features under unknown condition. The predicted data is then used to evaluate the system status. Both validating the predictive model and evaluating the system status involve comparing the predicted principal components and those obtained from measurement data. A Bayesian factor metric is employed to quantitatively judge the model and system status, i.e., whether the predicted features deviate from those obtained from actual measurements.

A Bayes factor-based assessment metric developed by Jiang and Mahadevan (2008b) for structural damage detection is employed in this study as a quantitative measure to evaluate predictive model and judge system status, by using different sets of data. In the model validation, the data collected from the system without any damage is used, while in the damage prediction, a set of data from the system with unknown condition is used. Within the context of binary hypothesis testing, consider two hypotheses H_0 and H_1 . The point null hypothesis ($H_0: y_{\text{exp}} = y_{\text{pred}}$) indicates that the predictive model is accurate or the system is healthy, while the alternative hypothesis ($H_1: y_{\text{exp}} \neq y_{\text{pred}}$) indicates that the model is inaccurate or the system is damaged. Their prior probabilities of acceptance are denoted by $\pi_0 = \Pr(H_0)$ and $\pi_1 = \Pr(H_1)$.

Assuming $\mathbf{y}_{\text{exp}} = \{y_{1,\text{exp}}, y_{2,\text{exp}}, \dots, y_{N_{\max},\text{exp}}\}$ and $\mathbf{y}_{\text{pred}} = \{y_{1,\text{pred}}, y_{2,\text{pred}}, \dots, y_{N_{\max},\text{pred}}\}$ to be the N_{\max} samples of PC features obtained from experimental data and model predictions, respectively. Let $e_i = y_{i,\text{exp}} - y_{i,\text{pred}}$ represent the difference between them, and $\boldsymbol{\varepsilon}_{\text{obs}} = \{e_1, e_2, \dots, e_{N_{\max}}\}$ represent N_{\max} values of the error, which is usually assumed to follow a normal distribution $N(\mu, \sigma^2)$ with σ known in this paper (estimated using the data set $\boldsymbol{\varepsilon}_{\text{obs}}$). The problem becomes testing $H_0: \mu = 0$ to in favor of the model or a healthy system, versus $H_1: \mu \neq 0$ to reject the model or in favor of a damaged system with $\mu | H_1 \sim N(\rho, \nu^2)$, in which ρ and τ are

two parameters of the prior density of μ under the alternative hypothesis. Per the suggestion in Migon and Gamerman (1999), the parameters $\rho = 0$ and $v^2 = \sigma^2$ are used if no information on $f(\mu|H_1)$ is available. The Bayes factor, the ratio of probability of observing the data given the null hypothesis (i.e., model is accepted or the system is healthy) to the probability of observing the data given the alternate hypothesis (i.e., model is rejected or the system is unhealthy), is calculated as follows (Jiang and Mahadevan 2008b):

$$B_{01} = \frac{\Pr(\text{data} | H_0 : \varepsilon_{\text{obs}} = 0)}{\Pr(\text{data} | H_1 : \varepsilon_{\text{obs}} \neq 0)} \\ = \sqrt{N_{\text{max}} + 1} \exp \left[\frac{-N_{\text{max}}^2 \bar{\varepsilon}_{\text{obs}}^2}{2(N_{\text{max}} + 1) \sigma^2} \right] \quad (10)$$

Refer to Jiang and Mahadevan (2008b) for details of the derivation of Eq. (10). Since B_{01} is non-negative, the value of B_{01} is often converted into the logarithm scale for the convenience of comparison among a large range of values, that is

$$b_{01} = \ln(B_{01}) = \frac{1}{2} \ln(N_{\text{max}} + 1) - \frac{N_{\text{max}}^2 \bar{\varepsilon}_{\text{obs}}^2}{2(N_{\text{max}} + 1) \sigma^2} \\ \leq \frac{1}{2} \ln(N_{\text{max}} + 1) \quad (11)$$

where $\ln(\cdot)$ is a natural logarithm operator, and the value of $0.5\ln(N_{\text{max}}+1)$ is the upper bound of b_{01} . A positive value of b_{01} indicates that the data supports the model or a healthy system (i.e., accepting H_0) while a negative value indicates that the data rejects the model or implies damage in the system (i.e., rejecting H_0). Kass and Raftery (1995) suggest interpreting b_{01} between 0 and 1 as weak evidence in favor of H_0 , between 3 and 5 as strong evidence, and $b_{01} > 5$ as very strong evidence. Negative b_{01} of the same magnitude is said to favor H_1 by the same amount.

The Bayesian measure of evidence that the predictive model is accurate (or the system is healthy) may also be quantified by the posterior probability of the null hypothesis $\Pr(H_0|\text{data})$, denoted by λ . For a binary hypothesis testing we have $\Pr(H_1|\text{data}) = 1 - \Pr(H_0|\text{data})$, and $\pi_1 = 1 - \pi_0$. Thus, using the Bayes theorem $\Pr(H_0|\text{data})$ can be obtained as follows:

$$\lambda = \Pr(H_0 | \text{data}) = \frac{B_{01}\pi_0}{1 - \pi_0 + B_{01}\pi_0} \quad (12)$$

Equation (12) is used to quantify the confidence in accepting the null hypothesis (i.e., accepting the model or a healthy system) based on the features extracted from sensor data and predicted output. Usually $\pi_0 = \pi_1 = 0.5$ is assumed due to the absence of any prior knowledge about the two hypotheses. In that case, Eq. (12) simplifies to $\lambda = B_{01}/(1+B_{01})$, where $B_{01} \rightarrow 0$ indicates 0% confidence in accepting the null hypothesis, and $B_{01} \rightarrow \infty$ indicates 100% confidence.

4. IMPLEMENTATION PROCEDURE

Figure 3 shows the generalized process for damage prediction using the proposed probabilistic methodology and the time series data collected from a turbomachine. The process consists of 21 steps (denoted by numbers in Fig. 3), as explained below.

1) Read raw time series data acquired from multiple sensors installed on a turbomachine. Figure 4 shows the example sensors used to acquire operational data from a centrifugal compressor. The sensor data includes various performance variables, such as compressor inlet and exhaust temperature, pressure, and axis displacement, temperature and vibration. These sensors measure the operational performance of the machine, generally in the form of time series. Assume the multiple time series data are obtained for q variables, each having M data points, yielding a raw data matrix $\mathbf{R}_{q \times M}$ for signal processing, damage diagnosis and damage prediction. The 1-hour interval data is used in the example presented in this paper for demonstration purpose.

2) Perform data validation on the acquired time series data for the unit under investigation. Techniques may include graphical plots, outlier analysis and data filtration. Outliers tend to pull the mean value towards themselves and inflate the variance in their direction, thus affecting the moment characteristics of the data. The outliers are usually inconsistent with the expectation of the majority elements of the series. They may be produced from measurement errors and equipment operation anomaly, which cannot be used to represent the normal operational condition of that unit. These outliers should therefore be removed for further analysis with proper justification. In this study, graphical analysis is first employed to identify the outliers in the time series data. Box plot is then applied to identify the outliers. The Box-plot is a type of exploratory data analysis technique which can visually show different types of populations without any assumption of statistical distribution. Note that the outliers are a kind of uncertainty and will be further processed through the advanced quantitative analysis by Bayesian wavelet signal processing and probabilistic principal component analysis subsequently.

3) Determine a validated data matrix $\mathbf{S}_{q \times N}$ for subsequent analysis, in which N is the number of validated data points ($N \leq M$). After the outlier analysis, data filtration may be performed to ensure that the sensor data represents the unit under normal operation via truncating sensor data in the reasonable operating range. Generally critical variables suggested by the subject matter expert should be used to define the normal operating status of the unit. In addition, the operational data during the outage downtime should be removed from the subsequent analysis. The proper range for each variable should be obtained from the manual of original equipment manufacturer, material standard, or design team.

4) Normalize each time series by using the mean value

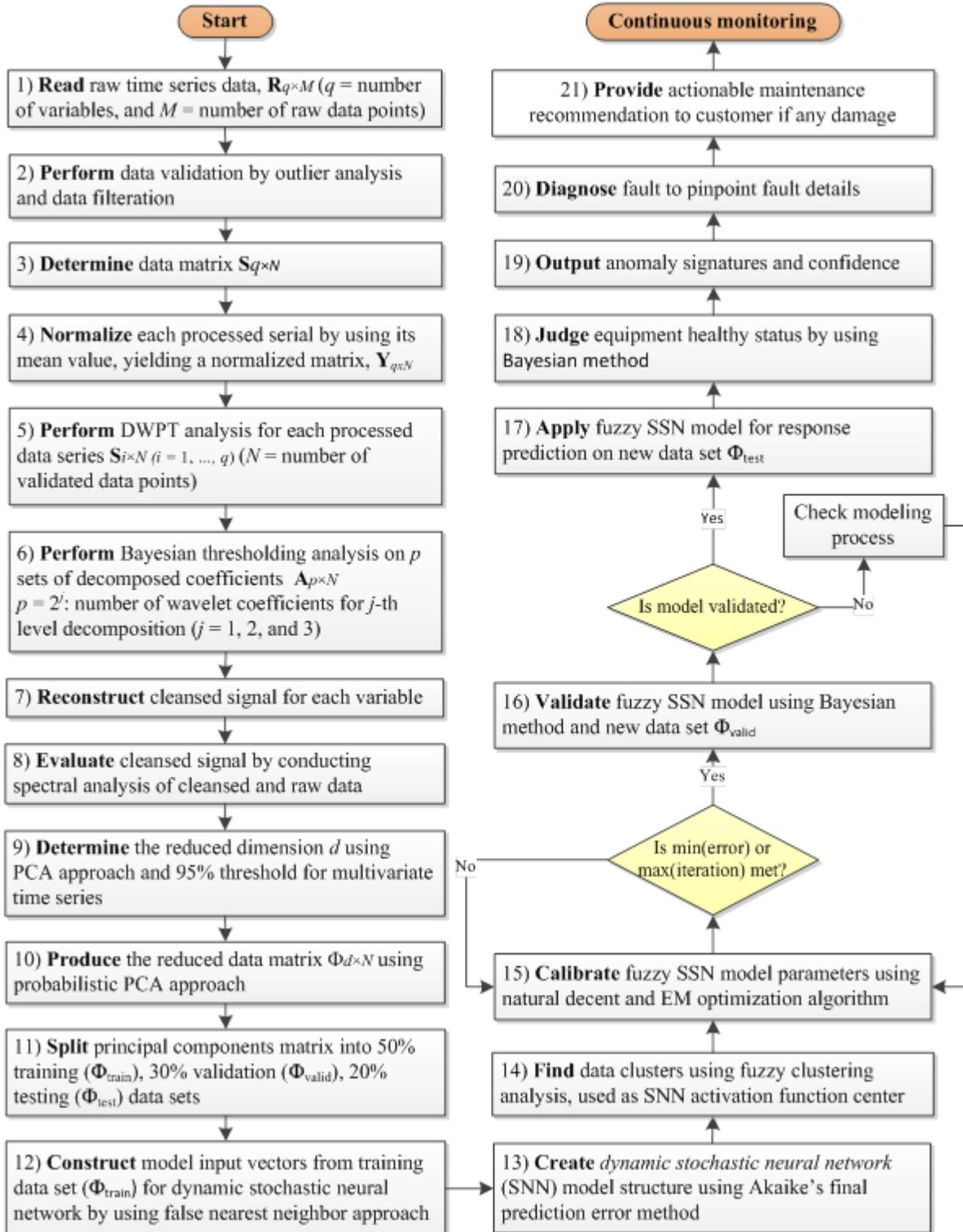


Figure 3. Flow chart of the proposed dynamic fuzzy SNN method for anomaly prediction

of each variable, yielding a normalized data matrix, $Y_{q \times N}$. Its purpose is to reduce the impact of different quantities of multivariate time series on the anomaly prediction. The normalization allows for the comparison of multiple time

series variables simultaneously. It avoids the duplicate contribution of the same response to damage detection and prevents the undue domination of variables with large numerical values over the variables with small numerical

values, thus improving the accuracy of damage detection using multivariate sensor time series.

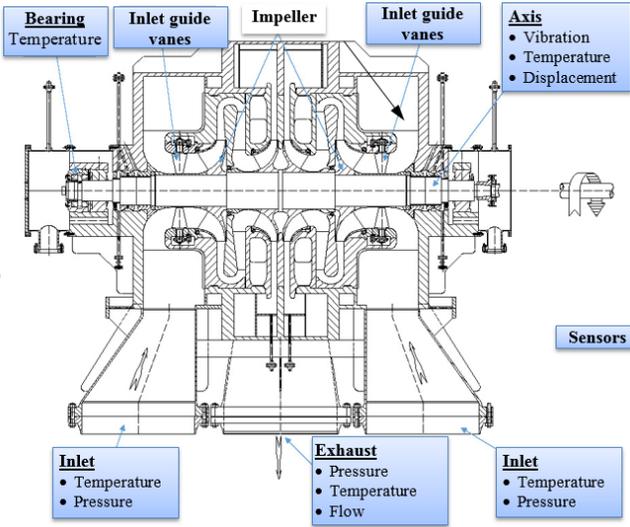


Figure 4. Illustration of sensors layout used to collect time series data

5) Perform discrete wavelet packet transform of the i -th time series $\mathbf{S}_{i \times N}$ ($i = 1, \dots, q$), each containing N data points. The DWPT approach decomposes the raw time series into multi-resolution time-frequency domains for each variable. Each time series will be decomposed on p sets of coefficients $\mathbf{A}_{p \times N}$, in which $p = 2^j$ is the number of wavelet coefficients for j -th level decomposition (e.g., $j = 1, 2$, and 3 for the 3-level decomposition).

6) Perform Bayesian hypothesis testing on each set of decomposed wavelet coefficients. The integrated Bayesian wavelet signal processing captures subtle details of signals meanwhile assessing noise, error and incoherence context in a direct means, thus avoiding the conventional under-denosing and over-denosing issues.

7) Reconstruct each time series signal from the cleansed wavelet coefficients by using Eq. (2).

8) Evaluate the effectiveness of the Bayesian wavelet cleansing approach via power spectral analysis on the cleansed and raw time series. The Welch method (Stoica & Moses 1997) may be to produce the spectral density from a finite-length signal by averaging the periodograms of overlapped, windowed signal sections. The periodogram is obtained by a short time Fourier transform. The average enables to reduce the variance or noise in the estimated power spectra, which is different from other existing methods such as the multiple signal classification method or the eigenvector method.

9) Determine the reduced dimension d using the conventional PCA approach and predefined 95% threshold for multivariate time series. Its purpose is to yield a d -

dimension reduced data matrix containing 95% information of the raw data.

10) Produce the reduced data matrix, $\Phi_{d \times N}$, using probabilistic PCA approach. The multivariate statistical analysis method is employed to fuse multivariate data, considering data uncertainty and correlation. The method incorporates statistical factor analysis, matrix Eigenvalue decomposition, and maximum likelihood estimation. Its purpose is to integrate useful information from multiple sensors, considering uncertainty and correlation in the multivariate sensor data.

11) Split the principal components matrix into three parts: 50% training (Φ_{train}), 30% validation (Φ_{valid}), and 20% testing (Φ_{test}) data sets. The training data is used to establish a predictive stochastic neural network model. The validation and testing data are then used to evaluate and test the predictive model for damage detection, respectively.

12) Construct input state vectors from training data set (Φ_{train}) for dynamic stochastic neural network. The false nearest neighbor approach with the chaotic attractor concept is used in this study.

13) Determine the number of hidden nodes in the SNN model by using the Akaike's final prediction error method to achieve an optimal model structure.

14) Find data clusters by using fuzzy clustering analytics. The clusters will be integrated with stochastic radial basis neurons to create the fuzzy SNN model (Eq. 7). The fuzzy C-means algorithm is employed to find clusters. It assigns membership to each data point corresponding to each cluster center based on the distance between the cluster center and the data point.

15) Calibrate the fuzzy SNN model parameters by using the natural gradient approach with the Kullback-Leibler distance criterion, and the expectation-maximum algorithm. The model parameters calibration will stop if either the minimum error between the predicted PC response and actual data or the maximum of calibration iteration is met, whichever is reached earlier.

16) Validate the fuzzy SNN model with both qualitative approach (graphical comparison) and quantitative methods (Bayesian hypothesis testing) for a new data set (Φ_{valid}). The same input dimension as the model establishment is used in the model validation. The graphical comparison is conducted in both time and frequency domains. A power density spectrum method, called *pseudospectrum* and pursued by Jiang and Adeli (2007) for structural damage detection, is employed in this paper to visually identify the difference. The pseudospectrum provides a reliable solution for eigenvalues of a *non-normal* matrix (whose eigenvectors are not orthogonal), which cannot be solved easily by a standard eigenvalue solution method. The approach produces a higher

frequency resolution even for data with high uncertainty (or high noise) than the conventional fast Fourier transform method (Migon & Gamerman, 1999). The quantitative methods include root mean square error and Bayes factor obtained from Bayesian hypothesis testing as described previously. If the model doesn't meet the acceptable threshold, the model parameters need to be calibrated again.

17) Apply the validated fuzzy model for response prediction on the new data set (Φ_{test}). Similarly, the same input dimension as the model establishment is used in the model prediction. The difference between the obtained prediction output and the actual principal component data is then calculated.

18) Judge the equipment healthy status by using Bayesian hypothesis testing method and the difference data obtained previously. Bayes factor and confidence obtained by Eqs. (11) and (12), respectively, are used as the quantitative assessment indicators. Bayesian hypothesis testing is conducted on a continuously rolling window (e.g., 12 hours used in the example presented in this paper). In each step, the currently predicted value is added to replace the earliest point in the rolling window, thus yielding the Bayesian factor and confidence over the time for the health assessment based on the testing data.

19) Output the healthy status signatures such as Bayes factor and the confidence. The Bayes factor in the logarithm scale has the negative value to indicate the possible damage occurring to the equipment under investigation, with more than 50% confidence. In that case, an alarm would be generated to notify the user for damage diagnostics or root cause analysis of the possible fault.

20) Diagnose the fault of equipment to pinpoint fault details. The PCA coefficients obtained previously can be used to identify the critical variables contributing to the fault. Time series analysis may be further used to pinpoint the fault initiation and details for maintenance recommendation.

21) Provide actionable maintenance recommendation if any damage is identified. The diagnosis results will not only facilitate the decision-making in alarming logics development, performance monitoring, and condition evaluation for the equipment, but also assist the asset management of the equipment by scientific schedule of inspection and maintenance and effective management of parts procurement.

5. ILLUSTRATIVE EXAMPLE

The proposed methodology and its implementation process shown in Fig. 3 is demonstrated with a set of operational data collected from a real centrifugal compressor. The time series data of the $q = 11$ variable (two bearing temperature variables) shown in Figure 4 was recorded at 1-hour interval from March 1st to Oct 18th, right before the forced outage due

to the compressor failure at multiple impellers. As an example, Figure 5 shows the impeller of a centrifugal compressor with multiple blades damage. By removing the data points during the maintenance outage, $M = 3710$ of raw data points is obtained for this study at each variable.



Figure 5. Photograph of impeller cracks in a real-world centrifugal compressor

5.1. Data Preprocessing

All operational data used in the example are recorded manually. These data unavoidably contain human errors, outliers, missing values, and noise. As an example, Figure 6 shows the raw time series data for axis vibration, inlet temperature and exhaust temperature measured at the centrifugal compressor. Note that the left vertical axis indicates the vibration variable, while the right vertical axis indicates the temperature variable. The gaps in the time series of Fig. 6 show the unavailable data during that period. Obviously, it is difficult to visually identify any significant change from the raw data series over the time, indicating that advanced analytics techniques are needed for signal processing of the data to develop automatic anomaly detection algorithms. In addition, it is observed that the multiple time series data indicate possible correlation from their trend over the time. Multivariate analysis techniques are therefore needed to facilitate the decision making on the diagnostics of a complicated system with multivariate time series data.

Furthermore, it is observed that each time series shows large variation over the time, indicating that the data may be contaminated by human errors due to the manual reading or noise from machine operation. The variation may result from the varying operational modes of the machine in practice. These noises or variabilities in the raw would influence the accuracy of diagnostics as shown later. Also, possible outlier points are observed in different time series. These obvious outliers are removed for subsequent data analysis.

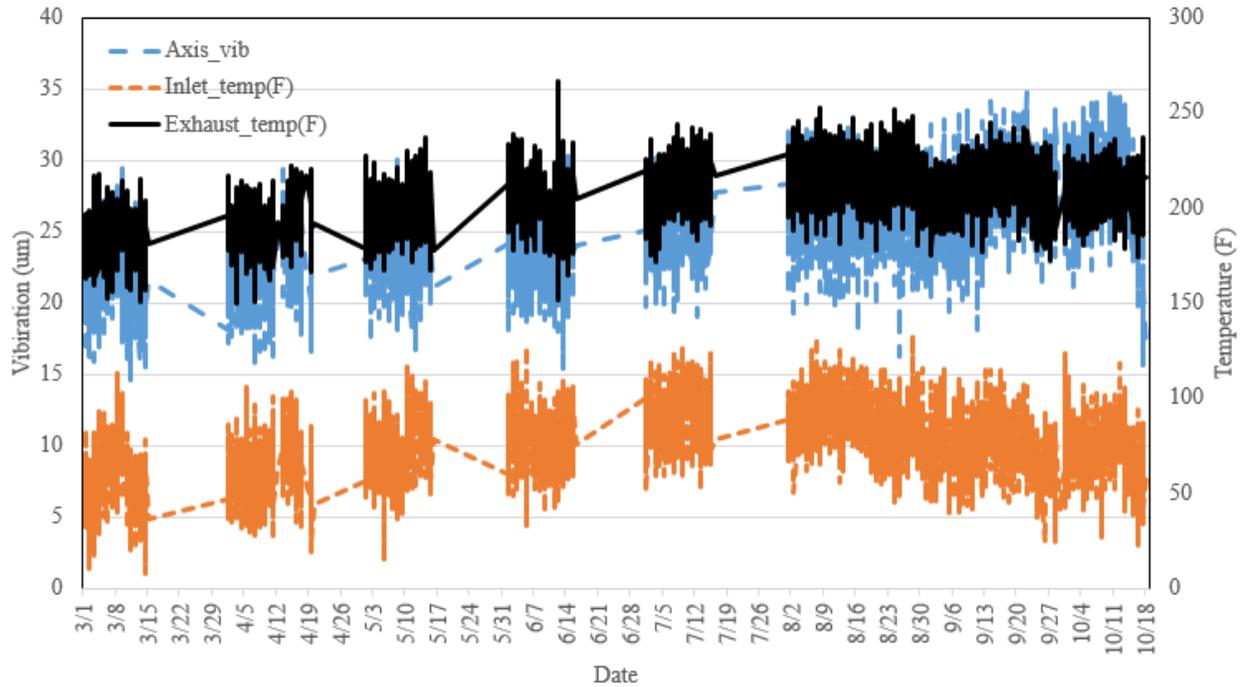


Figure 6. Data sample used in the demonstration example

A data matrix with the same dimension $q \times N$, where $q = 11$ and $N = 3616$ (in comparison to $N = 3710$ in the raw data set) is obtained by removing the obvious outliers or unavailable points from each time history data. In addition, in order to prevent the undue domination of variables with large numerical values over the variables with small numerical values in the multivariate damage prediction, each time series is divided by its corresponding mean value, thus normalizing the variables into dimensionless vector.

5.2. Bayesian Wavelet Denoising

Three decomposition levels of DWPT using the Daubechies wavelet of order 8 is first conducted to characterize the details of each time series data in this example. As an illustrative example, Figure 7 shows the three-level DWPT decomposition for the normalized exhaust flow data using the Daubechies wavelet of order 8. The normalized data shown in Fig. 7a are resolved into one approximation coefficients (Fig. 7b) and seven detail coefficients (Fig. 7c). Obviously, the approximation coefficients capture the principal trend of the raw data, while the high-frequency components in the raw signal are represented in the other decomposition detailed coefficients. It should be noted that the scale of the vertical axis in approximate coefficient is chosen differently from the detailed series in Fig. 7c for the sake of visibility of the illustration. Next, given a time series, the Bayesian thresholding is first applied to each decomposition coefficient produced from DWPT approach, as shown in Fig. 7. A cleaned time series is then reconstructed from the

thresholded wavelet coefficients. Refer to Xu *et al.* (2016) for the denoising result comparison.

5.3. Feature Extraction by Probabilistic PCA

The number of principal components of the cleansed data matrix \mathbf{Y} , $d = 3$, is obtained by predefining at least 95% information in the data to be considered, which result in the actual 95.5% information accounted for. The maximum likelihood estimates for the PPCA variability parameters σ^2 and \mathbf{W} are obtained for the cleaned 11×3616 data matrix. The resulting coefficients \mathbf{W}_{ML} are used for the data matrix to produce the 3×3616 reduced data matrices Φ^* using Eq. (6). Figure 8 shows the obtained weights of 11 variables contributing to the 3-dimension principal components. The weights can be used to effectively identify the critical variables which make significant contribution to the principal components. For instance, five variables, including the inlet temperature, exhaust temperature, inlet guide vane angle, axis vibration, and axis displacement, make main contribution to the first principal component, which accounts for 61% information of the original data. The information is useful for pinpointing variables which play a vital role on the damage identified from the predictive modeling and Bayesian assessment, further providing accurate actionable recommendation on maintenance activities.

5.4. Fuzzy SNN Modeling

Using the FNN method as described in Jiang *et al.* (2017), $p = 5$ is obtained for each principal component series (y_k) of the

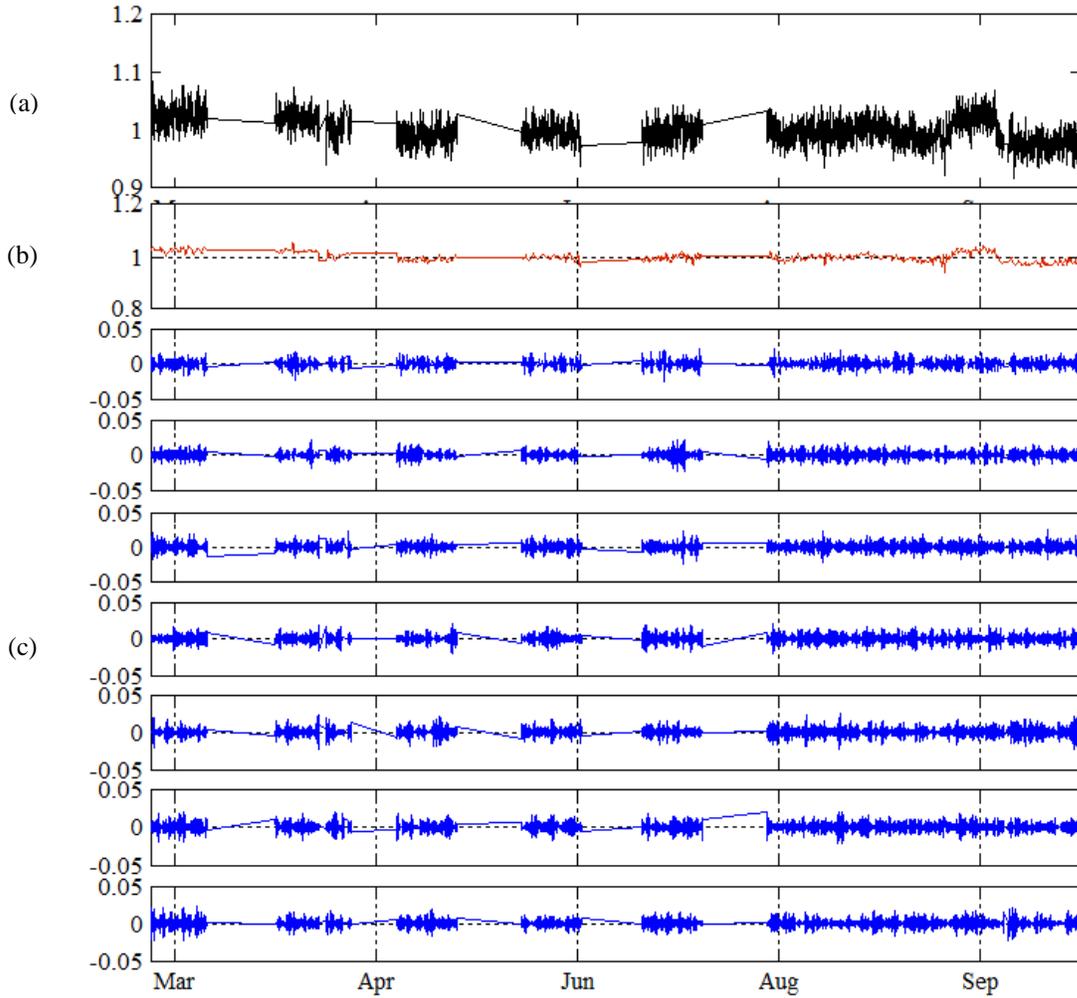


Figure 7. Three-level DWPT decomposition of normalized raw exhaust flow data using the Daubechies wavelet of order 8: a) Noisy time series, b) decomposition approximation coefficients, and c) Seven 3-level decomposition detail coefficients



Figure 8. Weights of 11 variables contributing to 3-dimension principal components

1073 data points, resulting in a 15-dimensional input vector and the maximum number of input-output data sets of $N_{max} = N - p = 1068$. Next, the BIC approach associated with the FCM method is applied for the constructed input vectors to

generate the optimum number of neurons $J = 2$. The model is then trained using the natural gradient method described previously. In the model training, the previous seven data points in each principal component are used as inputs, and the

current prediction y_k is used as the output of the model. The model converges quickly in only 4 iterations with a training error of $RMSE = 0.079$.

Figure 9 shows the comparison results of fuzzy SNN model in terms of both time (Fig. 9a) and frequency (Fig. 9c) domains for the first principal component. Figure 9b shows the difference of the predicted and actual principal components. It should be noted that the scale of the vertical axis is chosen differently for results comparison in Fig. 9a and error plots in Fig. 9b for the sake of visibility of the illustration. The Bayesian hypothesis testing is applied to the difference shown in Fig. 9b. The Bayes factor $b_{01} = 1.89$ is obtained, implying that the model is acceptable with the confidence of 86.9%. The pseudospectra of the predicted and actual values between the frequency range of 0 and π is shown in Fig.9c. The difference of the pseudospectra obtained from predicted and actual principal components (Fig. 9c) is insignificant, implying that the model shows high accuracy from the visual comparison.

The trained model is validated by applying to conduct one-step-ahead prediction on the validation data (350 points). In applying the trained model, the computed previous five values (i.e., $p = 5$) for each principal component are fed into

the input layer to calculate the current prediction. The comparison of the validation results for the three principal components is shown in Fig. 10a, which visually indicates the match between the predicted and actual results. The error of $RMSE = 0.0017$ and Bayes factor of 199 with the confidence of nearly 100% are obtained for the validation results, quantitatively demonstrates that the fuzzy SNN model provides an accurate prediction of the system response in terms of principal components.

5.5. Bayesian Damage Prediction

The trained model is further applied to conduct one-step-ahead prediction on the testing data (2171 points). Like model validation, the computed previous five values (i.e., $p = 5$) for each principal component are fed into the input layer to calculate the current prediction. The comparison of the prediction results and the error for the three principal components is shown in Fig. 10b and Fig.11a, respectively. Bayesian hypothesis testing is applied to the rolling window of 12 hours to yield the continuous quantitative indicators factors. Figure 11b shows the confidence over the time for both validation and testing data. It is observed that two alarms would be generated at 5:00AM of Sep 2nd and 2:00AM of Oct

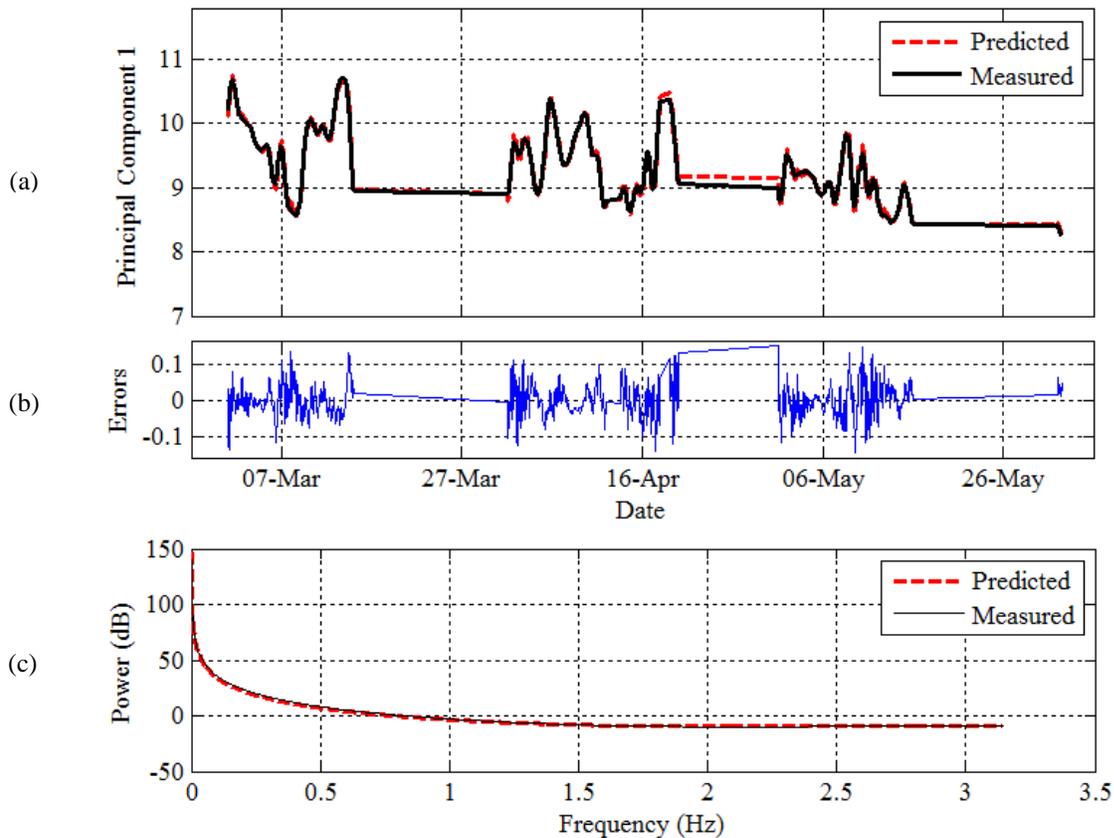


Figure 9. Modeling results of fuzzy SNN (1073 data points): (a) time domain comparison; (b) error ($RMSE = 0.079$ and $b_{01} = 1.89$), and (c) pseudospectra comparison

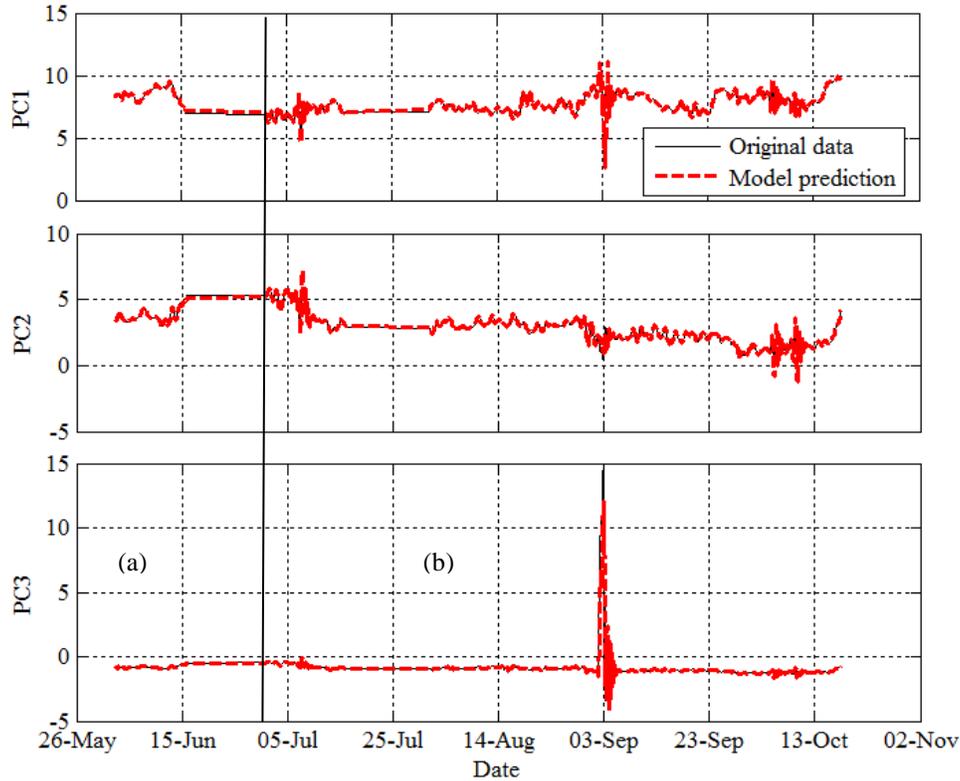


Figure 10. Validation and testing results of fuzzy SNN in time series: (a) Validation (350 data points) and (b) Testing (2171 data points)

9th to alert the possible damage on the equipment under operation, where the confidence of equipment healthy is less than 50%. The two alarms will last for 64 hours and 34 hours, respectively. The damage prediction results match well with the diagnostics results presented in Xu *et al.* (2016) about this compressor event. The damage on the impellers of this compressor might initiate in the end of July, which results in the principal components trend change, and then accumulated to a certain level to trigger the operation behavior change in the beginning of September when multiple alarms will be triggered. Finally, the compressor had a forced outage on Oct 23rd as the damage cumulates to the certain level that the unit cannot be operated any more. This indicates that the proposed intelligent algorithm associated advanced signal processing and data mining can quantitatively detect this damage in advance of about 2 weeks to prevent the forced outage, thus facilitating the decision making on the damage diagnostics and prediction with multivariate time series.

6. CONCLUDING REMARKS

This paper presents an innovative intelligent methodology for anomaly prediction of high-fidelity turbomachinery. The methodology and procedure seamless integrates the wavelet multi-resolution decomposition, Bayesian hypothesis testing, probabilistic principal component analysis (PPCA), and fuzzy stochastic neural network (SNN) modeling to provide

accurate anomaly prediction for a turbomachine, considering possible uncertainty and noise in the sensed multivariate time historical data. A generalized procedure is provided to implement the proposed methodology for damage prediction of a turbomachine. The proposed method and procedure is demonstrated with the 11-variable time series data and a forced outage event of a real-world centrifugal compressor.

The wavelet PPCA data preprocessing is employed in this study to reduce dimensionality and to address multivariate correlation and data uncertainty for damage prediction. Instead of directly using raw time series data as conventionally used in neural network model for structural damage detection, the yielded principal components are employed in the model establishment and anomaly prediction. Bayesian interval hypothesis testing metric is then presented to quantitatively compare the predicted and measured data for model validation and anomaly evaluation, thus providing a confidence indicator to judge the model quality and evaluate the equipment status. As such, Bayesian method is used in this study for three purposes: thresholding the raw data with multiresolution wavelet decomposition, quantifying the model accuracy, and assessing the system status.

The dynamic stochastic neural network model is established via the nonlinear autoregressive moving average with

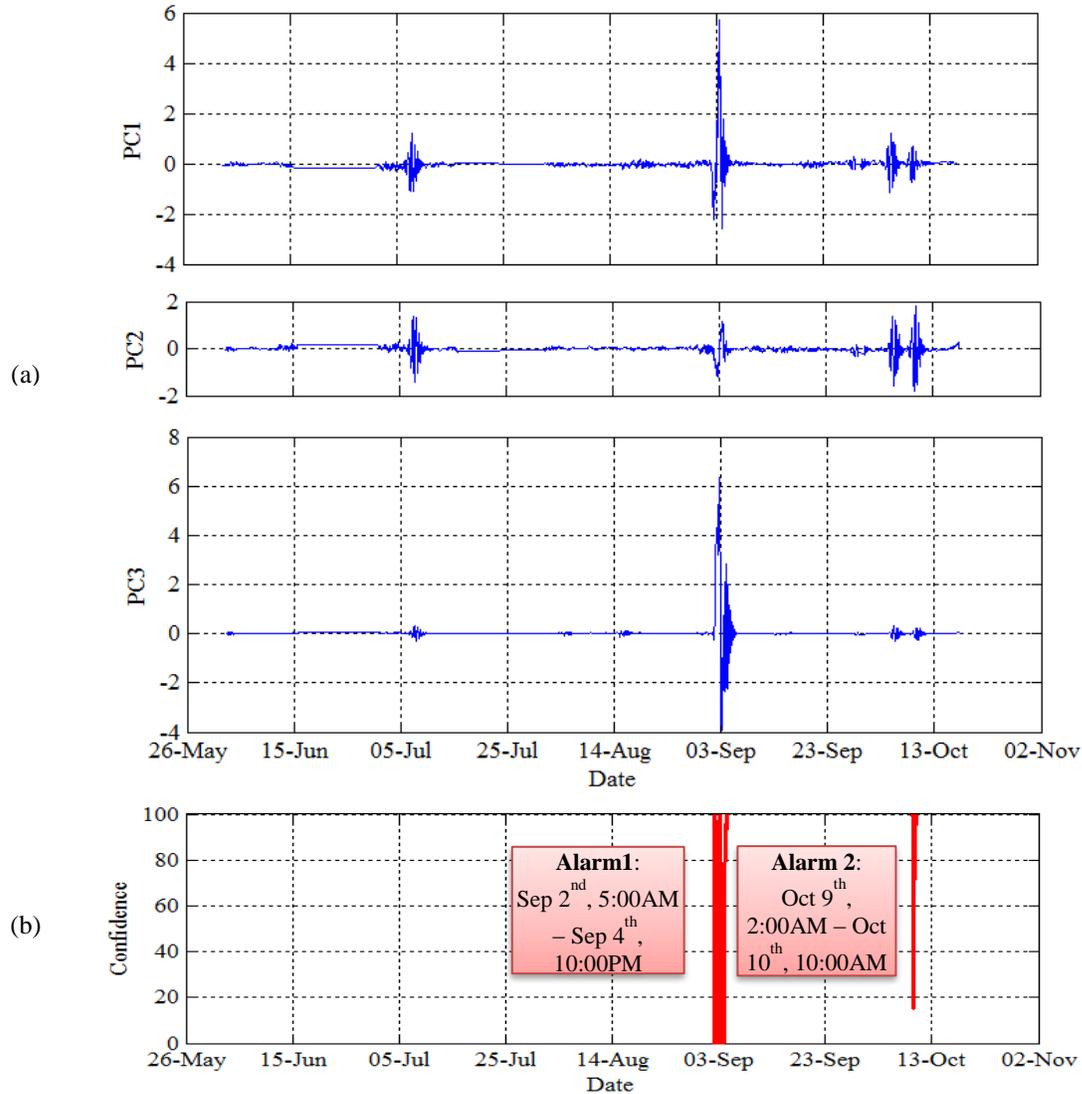


Figure 11. Validation and testing results of fuzzy SNN in time series: (a) Error and (b) Bayes confidence

exogenous inputs approach. It adeptly integrates the fuzzy clustering and independent Bernoulli random function into radial basis function neural network. A natural gradient method based on Kullback-Leibler (KL) distance criterion is employed to maximize the log-likelihood loss function. The KL distance measures the similarity between two probability density functions in terms of the entropy information of the data. The dynamic fuzzy SNN model is advantageous over the conventional static neural networks in two aspects. First, the former is a dynamic stochastic model that preserves the time sequence of the input vectors and memorizes the past of the time series data, while the latter is a deterministic model based on single-valued input nodes. Second, the fuzzy stochastic SNN can handle the fuzzy information and uncertainty in the sensed data through the fuzzy clustering

and stochastic neurons in the model, which cannot be dealt with by the latter.

Numerical results in the example indicate that the proposed intelligent probabilistic algorithm associated advanced signal processing and data mining can quantitatively detect this damage in advance of about 2 weeks to prevent the forced outage, thus facilitating the decision making on the damage diagnostics and prediction with multivariate time series, considering uncertainty and data correlation. The proposed methodology provides an advanced state-of-the-art approach and powerful tool to process multivariate time series data for damage prediction in order to evaluate the turbomachine condition. This study can facilitate cost-effective schedule for timely preventive maintenance and ensure product safety and increase product availability and customer satisfaction.

In future research, the proposed method will be extended to predict damage for more complicated scenarios in a centrifugal compressor, for example, many monitoring data sets, missing data in measurements and multiple-multivariate measurements with different boundary conditions. The application can also be extended to other turbomachine such as gas turbine and steam turbine, or any other asset damage prediction with measured time series data. In addition, further research may be conducted to optimally select the wavelet functions and decomposition level in the multiresolution analysis, as well as choose other stochastic neurons in the SNN model. The robust performance evaluation and comparison with other methods will be conducted in future research as well.

ACKNOWLEDGMENT

The research reported in this paper was supported by Haitian Scholar Fund from Dalian University of Technology awarded to Dr. Xiaomo Jiang, funds from National Natural Science Foundation of China (11402047) and Collaborative Innovation Center of Major Machine Manufacturing in Liaoning awarded to Dr. Shengli Xu, and funds from National Program on Key Basic Research Project (2015CB057301) and Advanced Design and Manufacturing Common Technology Innovation Platform Project of Liaoning (2015106016) awarded to Professor Xiaofang Wang and Shuhua Yang. The support is gratefully acknowledged.

REFERENCE:

- Abramovich, F., Sapatinas, T., & Silverman, B.W. (1998). Wavelet thresholding via Bayesian approach. *Journal of the Royal Statistical Society*, 60(3), 725 – 749. doi: 10.1111/1467-9868.00151.
- Adeli, H. & Jiang, X. (2006). Dynamic fuzzy wavelet neural network model for structural system identification, *ASCE Journal of Structural Engineering*, 132(1), 102–111. doi: 10.1061/(ASCE)0733-9445(2006)132:1(102).
- Al-Badour, F., Sunar, M., & Cheded, L. (2011). Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques, *Mechanical Systems and Signal Processing*, 25(6), 2083–2101. doi: 10.1016/j.ymsp.2011.01.017.
- Baydar, N., Chen, Q., Ball, A., & Kruger, U. (2001). Detection of incipient tooth defect in helical gears using multivariate statistics, *Mechanical Systems and Signal Processing*, 15(2), 303–321. doi: 10.1006/mssp.2000.1315.
- Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum.
- Burrus, C.S., Gopinath, R.A., & Guo, H. (1998). *Introduction to Wavelets and Wavelet Transforms: a Primer*, New Jersey: Prentice Hall.
- Chatfield, C. (2004). *The Analysis of Time Series: an Introduction*. 6th edition. Boca Raton, FL: Chapman & Hall/CRC.
- Chen, P., Taniguchi, M., Toyota, T., & He, Z. (2005). Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming, *Mechanical Systems and Signal Processing*, 19(1), 175–194. doi: 10.1016/j.ymsp.2003.11.004.
- Coifman, R.R. & Donoho, D.L. (1995). Translation-invariant de-noising. In Antoniadis, A. & Oppenheim, G. (Eds.), *Wavelets and Statistics*, Lecture Notes in Statistics, 103 (125 – 150), New York: Springer-Verlag.
- Coifman, R.R. & Wickerhauser, M.V. (1992). Entropy-based algorithms for best basis selection, *IEEE Transaction on Information Theory*, 38(2), 713–718. doi: 10.1109/18.119732.
- Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets, *Communication on Pure and Applied Mathematics*, 41(7), 909 – 996. doi: 10.1002/cpa.3160410705.
- Edwards, S., Lee, A.W., & Friswell, M.I. (1998). Fault Diagnosis of Rotating Machinery, *Shock and Vibration*, 30(1), 4-13. doi: 10.1177/058310249803000102.
- Eftekharijad, B., Carrasco, M.R., Charnley, B., & Mba, D. (2011). The application of spectral kurtosis on acoustic emission and vibrations from a defective bearing, *Mechanical Systems and Signal Processing*, 25(1), 266–284. doi: 10.1016/j.ymsp.2010.06.010.
- Galka, T. & Tabaszewski, M. (2011). An application of statistical symptoms in machine condition diagnostics, *Mechanical Systems and Signal Processing*, 25(1), 253–265. doi:10.1016/j.ymsp.2010.07.006.
- Ghanem, R. & Shinozuka, M. (1995). Structural system identification; parts I and II, *ASCE Journal of Engineering Mechanics*, 121(2), 255–273. doi: 10.1061/(ASCE)0733-9399(1995)121:2(255).
- Hung, S.L., Huang, C.S., Wen, C.M., & Hsu, Y.C. (2003). Nonparametric identification of a building structure from experimental data using wavelet neural network. *Computer-Aided Civil and Infrastructure Engineering*, 18(5), 358–370. doi: 10.1111/1467-8667.t01-1-00313.
- Jardine, A.K.S., Lin, D. & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. doi: 10.1016/j.ymsp.2005.09.012.

- Jiang, X. & Adeli, H. (2005). Dynamic wavelet neural network for nonlinear system identification, *Computer-Aided Civil and Infrastructure Engineering*, 20(4), 316–330. doi: 10.1111/j.1467-8667.2005.00399.x.
- Jiang, X. & Adeli, H. (2007). Psuedospectra, MUSIC, and dynamic wavelet neural network for damage detection of high-rise buildings, *International Journal for Numerical Methods in Engineering*, 71(5), 606–629. doi: 10.1002/nme.1964.
- Jiang, X. & Foster, C. (2013). Remote thermal performance monitoring: Turning data to solution, *ASME Power Conference (V002T13A004)*, Jul 29 – Aug 1, Boston, Massachusetts. doi:10.1115/POWER2013-98246.
- Jiang, X. & Foster C. (2014). Plant performance monitoring and diagnostics: remote, automated, and real-time, *ASME Turbo Expo (V006T06A034)*, Jun 16 – 20, Düsseldorf, Germany. doi: 10.1115/GT2014-27314.
- Jiang, X. & Mahadevan, S. (2008a). Bayesian wavelet methodology for structural damage detection, *Structural Control and Health Monitoring*, 15(7), 974–991. doi: 10.1002/stc.230.
- Jiang, X. & Mahadevan, S. (2008b). Bayesian probabilistic inference for nonparametric damage detection of structures, *ASCE Journal of Engineering Mechanics*, 134(10), 820–831. doi: 10.1061/(ASCE)0733-9399(2008)134:10(820).
- Jiang, X., Mahadevan, S., & Adeli, H. (2007). Bayesian wavelet packet denoising for structural system identification, *Structural Control and Health Monitoring*, 14(2), 333–356. doi: 10.1002/stc.161.
- Jiang, X., Mahadevan, S., & Yuan, Y. (2017). Fuzzy stochastic neural network model for structural system identification. *Mechanical Systems and Signal Processing*, 82(1), 394–411. doi: 10.1016/j.ymsp.2016.05.030.
- Juang, J.N. (1994). *Applied system identification*. NJ: Prentice Hall, Englewood Cliffs.
- Kamitsuji, S. & Shibata, R. (2003). Effectiveness of stochastic neural network for prediction of fall or rise of TOPIX, *Asia-Pacific Financial Markets*, 10(2), 187–204. doi: 10.1007/s10690-005-6010-4.
- Kass, R. & Raftery, A. (1995). Bayes factors, *Journal of the American Statistical Association*, 90(430), 773–795. doi: 10.1080/01621459.1995.10476572.
- Kennel, M.B., Brown, R., & Abarbanel, H.D.I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review A*, 45: 3403–3411. doi: 10.1103/PhysRevA.45.3403.
- Lai, T.L. & Wong, S.P.S. (2001). Stochastic neural networks with applications to nonlinear time series, *Journal of the American Statistical Association*, 96(455), 968–981. doi: 10.1198/016214501753208636.
- Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications, *Mechanical Systems and Signal Processing*, 42(1-2), 314–334. doi: 10.1016/j.ymsp.2013.06.004.
- Lei, Y., Jiang Y., & Xu Z. (2012), Structural damage detection with limited input and output measurement signals, *Mechanical Systems and Signal Processing*, 28, 229–243. doi: 10.1016/j.ymsp.2011.07.026.
- Li, H., Zhang, X., & Xu, F. (2013). Experimental investigation on centrifugal compressor blade crack classification using the squared envelope spectrum, *Sensors*, 13(9), 12548–63. doi: 10.3390/s130912548.
- Liu, Y., Guo, L., Wang, Q., An, G., Guo, M., & Lian, H. (2010). Application to induction motor faults diagnosis of the amplitude recovery method combined with FFT, *Mechanical Systems and Signal Processing*, 24(8), 2961–2971. doi: 10.1016/j.ymsp.2010.03.008.
- Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693. doi: 10.1109/34.192463.
- Masri, S.F., Nakamura, M., Chassiakos, A.G., & Caughey, T.K. (1996). Neural network approach to the detection of changes in structural parameters. *ASCE Journal of Engineering Mechanics*, 122(4), 350–360. doi: 10.1061/(ASCE)0733-9399(1996)122:4(350).
- Masri, S.F., Smyth, A.W., Chassiakos, A.G., Caughey, T.K., & Hunter, N.F. (2000). Application of neural networks for detection of changes in nonlinear systems, *ASCE Journal of Engineering Mechanics*, 126(7), 666–676. doi: 10.1061/(ASCE)0733-9399(2000)126:7(666).
- Migon, H.S. & Gamerman, D. (1999). *Statistical Inference-An Integrated Approach*. London, UK: Arnold, a Member of the Holder Headline Group.
- McFadden, P.D. & Toozhy, M.M. (2000). Application of synchronous averaging to vibration monitoring of rolling element bearings, *Mechanical Systems and Signal Processing*, 14(6), 891–906. doi: 10.1006/mssp.2000.1290.
- Moody, J. & Darken, C.J. (1989). Fast learning in networks of locally-tuned processing units, *Neural Computation*, 1, 281–294. doi: 10.1162/neco.1989.1.2.281.
- Nakamura, M., Masri, S.F., Chassiakos, A.G., & Caughey, T.K. (1998). A method for non-parametric damage

detection through the use of neural network, *Earthquake Engineering and Structural Dynamics*, 27(9), 997–1010. doi: 10.1002/(SICI)1096-9845(199809)27:9<997::AID-EQE771>3.0.CO;2-7.

- Peeters, B. & Roeck, G.D. (2001). Stochastic system identification for operational modal analysis: A review, *ASME Journal of Dynamic Systems, Measurement, and Control*, 10(4), 659–667. doi: 10.1115/1.1410370.
- Percival, D.B. & Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis*, New York: Cambridge University Press.
- Rai, V.K. & Mohanty, A.R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform, *Mechanical Systems and Signal Processing*, 21(6), 2607–2615. doi: 10.1016/j.ymssp.2006.12.004.
- Samanta, B. & Al-Balushi, K.R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, *Mechanical Systems and Signal Processing*, 17(2), 317–328. doi: 10.1006/mssp.2001.1462.
- Stoica, P. & Moses, R.L. (1997). *Introduction to Spectral Analysis*. New Jersey: Prentice-Hall, Englewood Cliffs.
- Tipping, M.E. & Bishop, C.M. (1999). Probabilistic principal component analysis, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 611–622. doi: 10.1111/1467-9868.00196.
- Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors, *Journal of the American Statistical Association*, 93(441), 173 – 179. doi: 10.1080/01621459.1998.10474099.
- Wang, W.J., Chen, J., Wu, X.K., & Wu, Z.T. (2001). The application of some non-linear methods in rotating machinery fault diagnosis, *Mechanical Systems and Signal Processing* 15(4), 697–705. doi: 10.1006/mssp.2000.1316.
- Wang, W.Q., Golnaraghi, M.F., & Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy systems, *Mechanical Systems and Signal Processing*, 18(4), 813–831. doi: 10.1016/S0888-3270(03)00079-7.
- Wu, Z.S., Xu, B., & Yokoyama, K. (2002). Decentralized parametric damage detection based on neural networks, *Computer-Aided Civil-Infrastructure Engineering*, 17(3), 175–184. doi: 10.1111/1467-8667.00265.
- Xu, S., Jiang, X., Huang, J., Yang, S., & Wang, X. (2016). Bayesian wavelet PCA methodology for turbomachinery damage diagnosis under uncertainty, *Mechanical Systems and Signal Processing*, 80(1), 1-18. doi: 10.1016/j.ymssp.2016.04.031.