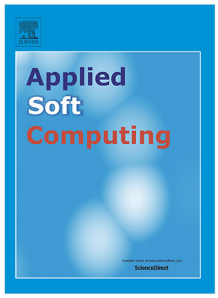
Accepted Manuscript



A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines

Xiaochuan Li, Xiaoyu Yang, Yingjie Yang, Ian Bennett, David Mba

PII: S1568-4946(19)30344-8
DOI: https://doi.org/10.1016/j.asoc.2019.105564
Article number: 105564
Reference: ASOC 105564

To appear in: *Applied Soft Computing Journal*

Received date : 15 October 2018
Revised date : 20 March 2019
Accepted date : 7 June 2019

Please cite this article as: X. Li, X. Yang, Y. Yang et al., A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines, *Applied Soft Computing Journal* (2019), https://doi.org/10.1016/j.asoc.2019.105564

A Novel Diagnostic and Prognostic Framework for Incipient Fault Detection and Remaining Service Life Prediction with Application to Industrial Rotating Machines

Xiaochuan Li[a], Xiaoyu Yang[b,a*], Yingjie Yang[a], Ian Bennett[c], David Mba[a,d]

[a]Faculty of Technology, De Montfort University, Leicester, LE1 9BH, UK
[b]Nanjing University of Aeronautics and Astronautics, Nanjing, China
[c]Shell Centre, London, SE1 7LZ, UK
[d]University of Nigeria, Nsukka, Nigeria

Abstract

Data-driven machine health monitoring systems (MHMS) have been widely investigated and applied in the field of machine diagnostics and prognostics with the aim of realizing predictive maintenance. It involves using data to identify early warnings that indicate potential system malfunctioning, predict when system failure might occur, and pre-emptively service equipment to avoid unscheduled downtime. One of the most critical aspects of data-driven MHMS is the provision of incipient fault diagnosis and prognosis regarding the system's future working conditions. In this work, a novel diagnostic and prognostic framework is proposed to detect incipient faults and estimate remaining service life (RSL) of rotating machinery. In the proposed framework, a novel canonical variate analysis (CVA)-based monitoring index, which takes into account the distinctions between past and future canonical variables, is employed for carrying out incipient fault diagnosis. By incorporating the exponentially weighted moving average (EWMA) technique, a novel fault identification approach based on Pearson correlation analysis is presented and utilized to identify the influential variables that are most likely associated with the fault. Moreover, an enhanced metabolism grey forecasting model (MGFM) approach is developed for RSL prediction. Particle filter (PF) is employed to modify the traditional grey forecasting model for improving its prediction performance. The enhanced MGFM approach is designed to address two generic issues namely dealing with scarce data and quantifying the uncertainty of RSL in a probabilistic form, which are often encountered in the prognostics of safety-critical and complex assets. The proposed CVA-based index is validated on slowly evolving faults in a continuous stirred tank reactor (CSTR) system, and the effectiveness of the proposed integrated diagnostic and prognostic method for the monitoring of rotating machinery is demonstrated for slow involving faults in two case studies of an operational industrial centrifugal pump and one case study of an operational centrifugal compressor.

---

* Corresponding author
E-mail address: yangxy@nuaa.edu.cn.

Keywords: Condition monitoring; Diagnosis and prognosis; Canonical variable analysis; Grey forecasting model; Particle filter

1. Introduction

Rotating machines, such as centrifugal pumps, are widely used due to their high performance and robustness. These machines typically operate under adverse conditions, such as frequent load changes and high speeds, and are thus subject to performance degradation and mechanical failure. In an effort to solve this problem, data-driven machine health monitoring systems (MHMS) [1] were introduced to realize predictive maintenance. Data-driven MHMS aim to monitor the degradation rather than just detecting the faults. It comprises four main steps: extracting features from collected data, detecting an incipient fault, determining the variables mostly affected by the fault, applying a prediction model on online measurements to predict machine deterioration. It is clear that these procedures are crucial for the safe, reliable, efficient and sustainable operation of any industrial system. Therefore, it is not surprising that automated data-driven machine health monitoring has been an active research area for decades now.

Industries can face large economic losses and security threats when faults occur. Condition monitoring and diagnosis are effective means to reduce the unplanned downtime and economic losses, and to sustain reliable system performance. The growing interest in the reliability of industrial processes and continuing progress in developing new signal processing techniques have motivated the development of advanced diagnostic approaches for complex industrial systems [2]. Multivariate statistical process monitoring (MSPM) techniques have recently seen improvements in diagnosing process abnormalities. Multivariate statistical analysis techniques such as principal component analysis (PCA) [3], independent component analysis (ICA) [4] and canonical variate analysis CVA [5] have been widely applied for the detection of process abnormalities in industrial plants and systems. In addition alternatives to the standard multivariate monitoring methods [6–9], which take into consideration the correlations between timestamps in the past and the future, have also been put forward for dynamic processes monitoring. Amongst the aforementioned MSPM techniques, CVA-based approaches were shown to be superior to other monitoring methods in terms of lead time and false positive rates [8]. Demand for facilitating fault prognosis has driven increased attention towards the development of incipient fault detection techniques, and great efforts have been made to improve the detectability of slow evolving faults [10-12]. However, these techniques require additional steps to select various model parameters so as to achieve reliable fault detection rate. In this study, traditional CVA approach is extended to form a novel monitoring index based on the distinctions between past and future canonical variables. Compared to traditional CVA-based indices, the new index relies on the dissimilarity between past and future measurements and is therefore more sensitive to incipient faults. For the present study, fault diagnosis is implemented by comparing the values of the new monitoring index with pre-defined thresholds calculated from kernel density estimation (KDE) [13]. It should be pointed out that although monitoring indices derived from CVA approach have been successfully applied for fault detection of engineering systems, their applicability for prognostics of rotating machines has not been fully studied. In fact, a monitoring index provides valuable information about the health status of an equipment, and therefore also has great potential in indicating the future behavior of the degrading system. Consequently, we will explore in this paper the use of the distinction-based monitoring index not only for detecting machine abnormalities but also for remaining service life (RSL) prediction.

Another major task of data-driven MHMS is to identify the influential variables that are most likely associated with the detected fault. Considering the possible synergy between different process variables, a MSPM model may provide more accurate diagnoses in comparison to when a single source of information or only a part of the variables are used [2]. However, it may become challenging for the accurate identification and localization of a fault when a large number of variables are considered [14]. Including a fault identification module comes along with various benefits. First, the process variables that do not show a degradation trend were eliminated automatically. From a prognostics perspective, variables that do not show a significant trend are not suitable for RSL prediction [15]. Computationally, the inclusion of only the influential variables alleviates the curse of dimensionality, reducing the computational costs in relation to prognostic feature extraction. Thirdly, fault identification allows the location, type and severity of the fault to be determined at an early stage, thereby allowing the optimization of preventive maintenance schedule and spare parts supply to be carried out along with RSL prediction so as to prevent catastrophic failures. Thus, identifying variables related to a fault is valuable for preventive maintenance and essential for developing effective diagnostic and prognostic tools. In this investigation, exponentially weighted moving average (EWMA) [16] is used together with Pearson correlation analysis (this technique is hereafter referred to as EWMA-Pearson) to realize automatic identification of the most fault-related variables. Pearson correlation analysis [17], as one the of most commonly used measures of correlation, has been widely applied in the field of medical research [18,19], time series analysis [20], pattern recognition [21], to name a few. In this study, we explore the use of Pearson correlation analysis for fault identification of mechanical systems, which has not been addressed before. EWMA, which is a well-known irregular fluctuations smoothing technique for time series analysis, was shown to be sensitive to small shifts [22] and therefore has gained popularity for providing incipient fault diagnosis recently [10,23]. However, most existing works related to EWMA propose solutions for incipient fault detection rather than incipient fault identification. To address this issue, EWMA is utilized to modify the traditional Pearson correlation analysis to improve its fault identification ability at early stages of degradation. The proposed fault identification method operates with various slowly developing faults but not restricted to mechanical system faults that evolve slowly over time.

The main aim of prognostics is to provide practitioners with warnings by predicting the deterioration of an incipient fault, thereby allowing engineers to control the progression of the fault and schedule repairs and maintenance. Typical procedures in data-driven MHMS involve a prognostic step where long-term predictions of continuous observations are carried out with the aim of estimating the RSL of the system. Various data-driven techniques are available for the long-term prediction of continuous observations, including statistical models and neural networks. In order to undertake training these models depend on large amounts of failure data. However, field failure data is extremely difficult to obtain, and this prevents these models from being applied in real industrial facilities. Even in the era of big machinery data, companies and practitioners still have a limited pool of “useful” data resources to fulfill prognostic tasks, since safety-critical equipment are usually not allowed to run to failure. In order to deal with scarce data, we apply grey forecasting model, which was originally devised to tackle small sample problems [24], to implement RSL prediction with limited amount of failure data. Moreover, a single-valued forecasting model learnt from historical data will only be learning the system’s deterministic and stochastic properties in general, not specifically. For example, a neural network may be learnt from past values. However, for prediction purposes, it is not only a single manifestation of a neural network prediction that is possible. Therefore, multiple random seeds and thus multiple manifestations of the trajectory need to be incorporated into the forecasting model. To address the aforementioned issue, this paper proposes an enhanced grey forecasting model based on the metabolism grey forecasting model (MGFM) [25] and particle filter (PF) [26] approaches (hereafter the proposed method is referred to as MGFM-PF). By leveraging the strengths of the MGFM and PF models, the proposed MGFM-PF method can provide site engineers with reliable and robust RSL estimates of systems operating under faulty conditions. There are many benefits of using PF in this study, including but not limited to: a) quantifying the uncertainty of RSL in a probabilistic form. Uncertainties associated with measurements and process noise are not taken into consideration when a single-valued forecasting model is utilized, which regards damage deterioration as deterministic in nature [27]. Using PF, prediction of RSL along with confidence levels can be achieved, and this helps site engineers to gain an insight into the uncertainty of the RSL; b) Improving the predictive accuracy of MGFM. Prediction of the progression of machinery fault is a complex nonlinear problem, and in this context PF becomes a very suitable tool because it is particular suitable for nonlinear systems [27]. Moreover, a high-order hidden Markov model (HMM) may be more suitable for predicting fault growth than a first-order model in real world applications [28]. In this work, MGFM is utilized together with a high-order PF to realize a high-order HMM for prediction of RSL. Note that prediction errors always exist even with a well-established predictor, and system dynamics may change when the fault propagates. Therefore, in this context, PF technique which has the ability to update its parameters according to the changes in system dynamics is desirable and can improve the prediction accuracy.

A number of studies addressing different aspects of MHMS have already been presented [5,29,30]. However, individual components of MHMS are usually separately investigated in traditional condition monitoring analyses, and this prevents these methods from being applied to real-industries where industry users usually wish to find a complete solution to predictive maintenance. Therefore, in this paper, we propose an integrated framework that covers all aspects of MHMS, from fault detection to fault identification to prognostics. The proposed framework was validated using historical failure data from an industrial centrifugal pump, retrieved from a server rather than raw sensor measurements commonly used. Use of actual information addresses a major prognostics challenge: limited works utilizing real-life data to demonstrate data-driven MHMS' applicability and benefits in industry.

The major contributions of this paper are as follows:

- The development of a novel RSL prediction model (e.g. MGFM-PF) able to address the challenge of failure data scarcity in the era of big machinery data to enable accurate prediction of the RSL of a system.
- The development of a novel integrated framework that covers all MHMS, from detection of incipient faults to determination of fault related variables to prediction of remnant life.
- Incipient fault diagnosis using a novel CVA distinction-based index.
- The development of the EWMA-Pearson method for early and accurate identification of fault related variables.
- The use of degradation data captured from an operational industrial centrifugal pump and a compressor.

2. Methodology

2.1 Existing condition monitoring methods

2.1.1 CVA revisited

Given two sets of zero-mean variables $y_{1,t} \in \mathcal{R}^n$ and $y_{2,t} \in \mathcal{R}^n$, CVA finds pairs of projection matrices K and G that maximize the correlation between the projections $z_{1,t} = K \cdot y_{1,t}$ and $z_{2,t} = G \cdot y_{2,t}$. The projections $z_{1,t}$ and $z_{2,t}$ are also referred to as canonical variates. To generate two data sets from the measured data

$y_t \in \mathcal{R}^n$, where n is the number of variables, the data were expanded at each sampling instance by including a, the number of previous samples, and b, the number of future samples, to generate the past and future vectors $y_{a,t} \in \mathcal{R}^{na}$ and $y_{b,t} \in \mathcal{R}^{nb}$.

$$y_{a,t} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-a} \end{bmatrix} \in \mathcal{R}^{na} \tag{1}$$

$$y_{b,t} = \begin{bmatrix} y_{t} \\ y_{t+1} \\ \vdots \\ y_{t+b-1} \end{bmatrix} \in \mathcal{R}^{nb} \tag{2}$$

To avoid domination of variables with large absolute values, $y_{a,t}$ and $y_{b,t}$ are then normalized to the zero-mean vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$. Then, the normalized future and past vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$ are rearranged as per Eqs. (3) and (4) to generate the reshaped matrices $\hat{Y}_a$ and $\hat{Y}_b$:

$$\hat{Y}_a = [\,\hat{y}_{a,t+1}, \hat{y}_{a,t+2}, \dots, \hat{y}_{a,t+N}] \in \mathcal{R}^{na\times N} \tag{3}$$

$$\hat{Y}_b = [\hat{y}_{b,t+1}, \hat{y}_{b,t+2}, \dots, \hat{y}_{b,t+N}] \in \mathcal{R}^{nb\times N} \tag{4}$$

where $N = M - a - b + 1$ and M denotes the length of y_t. Then the covariance matrices $\Sigma_{a,a}$ and $\Sigma_{b,b}$ and cross-covariance matrix $\Sigma_{a,b}$ can be computed as per (5):

$$\Sigma_{a,a} = \hat{Y}_a \hat{Y}_a^{\,T}/(N-1);\ \Sigma_{b,b} = \hat{Y}_b \hat{Y}_b^{\,T}/(N-1);\ \Sigma_{b,a} = \hat{Y}_b \hat{Y}_a^{\,T}/(N-1) \tag{5}$$

The vector of canonical correlations $D = \mathrm{diag}(\lambda_1, \dots, \lambda_k)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$ can be obtained by performing singular value decomposition (SVD) [31] on the matrix $\mathcal{H}$:

$$\mathcal{H} = \Sigma_{b,b}^{\ -1/2} \Sigma_{b,a} \Sigma_{a,a}^{\ -1/2} = U \Sigma V^T \tag{6}$$

2.1.2 Pearson correlation analysis

In order to identify the most fault affected variables, we use Pearson correlation analysis that results in a numerical value (i.e. Pearson correlation coefficient (PCC)) for how well variations in expression degrees of two variables correlates. PCC is one of the most commonly used statistical metrics in statistics that measures the direction and strength of a linear relationship between two random variables [32,33].

Given two sets of zero-mean random variables x and y, the PCC is defined as [17]:

$$\rho(x,y) = \frac{E[x,y]}{\sigma(x)\sigma(y)} \tag{7}$$

where $E[x,y]$ denotes the cross-correlation between the variable x and y, $\sigma(x)$ and $\sigma(y)$ denote the standard deviation of x and y, respectively. Standard PCCs were determined for each variable versus the fault detection index using data collected from early stages of deterioration. The PCCs give an indication of the contributions of different variables during the monitoring process. The higher the PCC of a performance variable, the larger the contribution of the specific variable to the detected fault.

2.1.3 Metabolism grey forecasting model

Metabolism grey forecasting model (MGFM) [25] is the basic model of grey theory and has been used widely since its development in the early 1980s. Grey system theory is a novel methodology that focuses on problems involving small data and poor information. It addresses uncertain systems with partially known information through generating, excavating, and extracting useful information from what is available. The theory enables a correct description of a system’s running behaviour and its evolution law, and thus generates quantitative predictions of future system changes. By updating the modelling data and introducing new information, MGFM can reflect the characteristics of the current situation. Grey forecasting model is suitable for real-time prediction with limited availability of degradation data. MGFM uses operations of accumulated generations to build different equations. The general procedure of MGFM is described as follows.

Consider the non-negative sequence of the original data $X^{(0)}$

$$X^{(0)} = \left(X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(n)\right) \quad (8)$$

Then $X^{(1)} = \left(X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(n)\right)$ is called the first order accumulative generation sequence of the sequence $X^{(0)}$, where

$$X^{(1)}(t) = \sum_{i=1}^{t} X^{(0)}(i) \quad t = 1,2,\cdots,n \quad (9)$$

A new sequence $Z^{(1)}$ can be extracted from $X^{(1)}$ as per the following:

$$Z^{(1)} = \left(Z^{(1)}(2), Z^{(1)}(3), \cdots, Z^{(1)}(n)\right) \quad (10)$$

$$Z^{(1)}(t) = 0.5\left(x^{(1)}(t-1) + x^{(1)}(t)\right), t = 2,3,\cdots,n \quad (11)$$

Then, the least square sequence estimation of the grey difference equation of MGFM is defined as follows:

$$x^{(0)}(t) + cz(t) = d \quad (12)$$

And the whitenization equation is as follows [25]:

$$\frac{dx^{(1)}(t)}{dt} + cx^{(1)}(t) = d \quad (13)$$

Where $[c,d]^T$ is the parameter vector of MGFM, which can be obtained by the least square estimation $[c,d]^T = (B^T B)^{-1} B^T Y$, in which

$$B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(\mathrm{n}) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(\mathrm{n}) \end{bmatrix}$$

According to Eq. (13), the solution of $x^{(1)}$ at time t is:

$$\hat{x}^{(1)}(t) = \left(x^{(0)}(1) - \frac{d}{c}\right) e^{-c(t-1)} + \frac{d}{c}, t = 1,2,\cdots,n \tag{14}$$

where $x^{(1)}(1) = x^{(0)}(1)$.

To obtain the predicted value of the primitive data at time t, the inverse accumulated generating operation is used to establish the following grey model:

$$\hat{x}^{(0)}(t) = x^{(1)}(t) - x^{(1)}(t-1) = (1 - e^{c})\left(x^{(0)}(1) - \frac{d}{c}\right) e^{-c(t-1)} \tag{15}$$

2.2 Enhancement of existing condition monitoring methods

2.2.1 Modified CVA

Suppose that N samples of the process data $\hat{Y}_a \in \mathcal{R}^{na\times N}$ and $\hat{Y}_b \in \mathcal{R}^{nb\times N}$ are available for diagnosing possible anomalies in the system under study, the remaining issue is to find the diagnostic observers that can achieve optimal fault detection with a given threshold. In conventional CVA-based approaches, only past data vectors $\hat{y}_{a,t}$ are used to construct test statistics:

$$z_t = K \cdot \hat{y}_{a,t} = V_q^T \Sigma_{a,a}^{-1/2} \hat{y}_{a,t} \tag{16}$$

$$e_t = G \cdot \hat{y}_{a,t} = V_{na-q}^T \Sigma_{a,a}^{-1/2} \hat{y}_{a,t} \tag{17}$$

Motivated by the fact that CVA is able to find maximum correlations between past and future data, one can detect subtle changes by examining how far away future canonical variates are deviated from past canonical variates (e.g. by examining the usual correlation between past and future). This leads to a diagnostic observer called canonical residuals that quantifies the distinctions between the past and future measurements. Canonical residuals are generated as follows:

$$r_t = L_q^T \hat{y}_{b,t} - \Sigma_q J_q^T \hat{y}_{a,t} \tag{18}$$

Where L_q^T denotes the first q rows of the projection matrix L^T, and $L^T = \Sigma_{b,b}^{-1/2} U_q^T$. Similarly, J_q^T is the first q rows of the projection matrix J^T, and $J^T = \Sigma_{a,a}^{-1/2} V_q^T$. $\Sigma_q = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$ is a diagonal matrix with its diagonal elements being the first q canonical correlations calculated as (6). Canonical residuals are measures of the discrepancies between the past and future measurements and are able to provide more effective feature representation of small shifts in the early stage of emerging faults compared with diagnostic statistics derived from traditional CVA approach [34].

Since the condition monitoring data are mean-variance normalized, the mean of the canonical residuals r_t is:

$$E(r_t) = L_q^T E(\hat{y}_{b,t}) - \Sigma_q J_q^T E(\hat{y}_{a,t}) = 0 \tag{19}$$

The covariance of r_t can be calculated as:

$$\Sigma_r = E(rr^T) = J_q^T E(\hat{y}_{a,t}\hat{y}_{a,t}{}^T)J + \Sigma L_q^T E(\hat{y}_{b,t}\hat{y}_{b,t}{}^T)L_q^T\Sigma^T - J_q^T E(\hat{y}_{a,t}\hat{y}_{b,t}{}^T)L_q^T\Sigma^T - \Sigma L^T E(\hat{y}_{b,t}\hat{y}_{a,t}{}^T)J = I + \Sigma\Sigma^T - \Sigma\Sigma^T - \Sigma\Sigma^T = I - \Sigma\Sigma^T \tag{20}$$

The distinctions between the past and future measurements are centred around a zero mean under healthy conditions. Hence, a diagnostic test statistics can be formed as the multivariate standard distance of the discrepancy features from zero [35]:

$$T_d = f\left(c(r_t - 0)^T S^{-1}(r_t - 0)\right) = \frac{|c(r_t-0)^T S^{-1}(r_t-0)|}{|c|[(r_t-0)^T S^{-1} S S^{-1}(r_t-0)]} = [(r_t)^T S^{-1}(r_t)]^{1/2} = [r_t^T (I - \Sigma\Sigma^T) r_t]^{1/2} \tag{21}$$

where c is a normalizing constant, and $S = I - \Sigma\Sigma^T$ is the covariance matrix of the test and the healthy data. The roots of the multivariate standard distance between two random vectors can be traced back to the results presented in [35], which is described as follows:

Given two random vectors x_1 and x_2, the univariate standard distance between the two vectors is defined as follows:

$$f(a) = |a^T x_1 - a^T x_2| / (a^T S a)^{1/2} \tag{22}$$

where a is a vector of unit length and $a^T a = 1$. $a^T x_1$ and $a^T x_2$ are the orthogonal projections of the vectors x_1 and x_2 on the linear space spanned by a, respectively. S is the covariance matrix of x_1 and x_2. Thus, $f(a)$ denotes the univariate standard distance between vectors x_1 and x_2 in this subspace. According to [35], the multivariate standard distance between x_1 and x_2 is attained for $a = c(x_1 - x_2)^T S^{-1}(x_1 - x_2)$ and takes the value:

$$T_d = f\left(c(x_1 - x_2)^T S^{-1}(x_1 - x_2)\right) = [(x_1 - x_2)^T S^{-1}(x_1 - x_2)]^{1/2} \tag{23}$$

In this paper, fault detection is implemented using a novel CVA distinction-based index T_c which is defined as follows

$$T_c = \frac{T^2}{\sigma^{T^2}} + \frac{Q}{\sigma^Q} + \frac{T_d}{\sigma^T d} \tag{24}$$

$$T^2 = z_t^T z_t \tag{25}$$

$$Q = e_t^T e_t \tag{26}$$

where σ^{T^2}, σ^{Q} and σ^{T_d} are the threshold of Hotelling's T^2 and Q statistics [36], and, T_d index, respectively. The aforementioned fault thresholds are calculated using the KDE algorithm [13]. T_c extends the traditional T^2 and Q statistics to form a new index, and is developed and adopted for the first time as a diagnostic index for fault detection of rotating machinery.

2.2.2 EWMA-Pearson for fault identification

Motivated by the fact that EWMA smoothing is sensitive to small shifts in data [10,22], EWMA is incorporated into the contributions calculated from the data collected from early degradation states to enhance the fault identification performance of Pearson correlation analysis. Compared to Pearson correlation analysis [17], EWMA-Pearson incorporates historical information and has great potential for detecting small shifts in the mean change. In other words, EWMA-Pearson improves the fault identifiability of Pearson correlation analysis by enhancing the contributions from influential variables. To show that the proposed technique is superior to Pearson correlation analysis for fault identification, the fault identifiability is compared with that of Pearson correlation analysis in Section 3.2.2. The results show that the fault identification performance can be significantly improved by applying EWMA-Pearson.

The contribution of variable y_n based on the EWMA approach can be obtained as:

$$c_t = (1-\delta)c_{de,t} + \delta c_{t-1} \tag{27}$$

$$c_{t-1} = \frac{\sum_{k=t-W}^{t-1} c_{de,t}}{W} \tag{28}$$

where δ is the forgetting factor and W is the width of the moving window. The reconstructed contributions can provide information regarding the most strongly affected variables when a fault occurs. The most influential variables are spotted if they have the largest contributions during the early stages of degradation. The influential variables identified using the proposed method will be used subsequently to construct a new monitoring index using data collected from early stages of deterioration as per Eqs. (1)-(6) and (16)-(26). With this process, a refined health indicator is extracted and then fed into a MGFM-PF prognostic model.

2.2.3 Enhancement of metabolism grey forecasting model using particle filter

The particle filter is a recursive Bayesian filtering technique utilizing Monte Carlo simulations [37]. Based on Monte Carlo principles, particle filters are used to make approximations to the future status of the system dynamics. Particles with associated weightings determine the required posterior distribution of the health state. These particles change and respond recursively with the availability of new information [38]. The pseudo code of the PF algorithm is listed in Algorithm 1.

Algorithm 1: PF

Step 1 (state tracking)

for $i = 1,2,3, \ldots, N$

  Draw particle x_k^i as per Eq. (29)

End for

Implement prediction as per Eq. (37)

for $i = 1,2,3, \ldots, N$

  Compute weights w_k^i as per Eq. (36)

End for

Normalize weights $w_k^i = \frac{w_k^i}{\sum w_k^i}$, resmaple weights $\{w_{k-3:k}^i, w_k^i\}$

Step 2 (prediction)

for $i = 1,2,3, \ldots, N$

  Implement prediction as per Eqs. (38) − (39)

End for

Predict RSL as per Eq. (40)

In this study, the monitoring index that was reconstructed using data collected from early stages of deterioration is defined as the state x of the equipment under study. The system deterioration was assumed to be a fourth-order Markov process since the grey forecasting model uses four previous samples as the inputs. A high-order prediction model may be more appropriate to represent the fault progression than a first-order HMM [28]. The Markov process can be described as:

$$x_k = g(x_{k-1}, x_{k-2}, x_{k-3}, x_{k-4}) + v_k \tag{29}$$

where g is the state transition function, x_k is the system state at time k and v_k is the noise term. The noise term is assumed to follow a Gaussian distribution, and its statistical properties were initially determined by the MGFM's modelling errors. During the state tracking process, a sliding window containing 50 estimation residuals was adopted to update the process noise at every time instance. The estimation residual z_k is defined as the deviation between the estimated prognostic feature value and the true prognostic feature value at time k:

$$z_k = x_k - \tilde{x}_k \tag{30}$$

$$\mu_{w,k} = \sum_{i=0}^{d-1} z_{k-i} / d \tag{31}$$

$$\sigma_{w,k} = \frac{1}{d}\sqrt{\sum_{i=0}^{d-1} (z_{k-i} - \mu_{w,k})^2} \tag{32}$$

where $\tilde{x}_k$ is the estimated state at time k, $\mu_{w,k}$ and $\sigma_{w,k}$ are the mean and the standard deviation of the process noise at time k, and d is the length of the sliding window (which is set to 50 in this study). The estimation residuals covered by the window were utilized to compute an error density that is subsequently employed to update the model parameters in each iteration. In this way, the last estimation error which accounts for potential system dynamics change during the estimation process was included in the prediction model. With the updated model parameters, MGFM was utilized to propagate the trend of the system state as per Eq. (29), and details regarding this technique are discussed in section 2.4.1.

When a new measurement of the system state y_{k+1} becomes available, the state update step is implemented, and the posterior state probability transition density $p(x_{0:k}|y_{1:k})$ is estimated as:

$$p(x_{0:k}|y_{1:k}) = \frac{p(y_k|x_k)}{p(y_k|y_{1:k-1})}\, p(x_{0:k}|y_{1:k-1}) = \frac{p(y_k|x_k)p(x_k|x_{k-4:k-1})p(x_{0:k-1}|y_{1:k-1})}{p(y_k|y_{1:k-1})} \quad (33)$$

where $p(y_k|y_{1:k-1}) = \int p(y_k|x_k)p(x_k|y_{1:k-1})$. According to [28], since both the system state x_k and measurement y_k represent the monitoring index, the likelihood function $p(y_k|x_k)$ can be simply described as:

$$y_{k+1} = x_{k+1} + v_{k+1} \quad (34)$$

where v_{k+1} is the measurement noise.

Nevertheless, it is hard to compute the posterior probability transition density in real-world applications as per Eq. (30) since the integrals do not have an analytical solution in most cases. Hence, a four-order particle filtering approach is employed here to approximate the posterior state probability transition density $p(x_{0:k}|y_{1:k})$ by a set of particles with associated weightings.

$$p(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{N} w_k^i\, \delta(x_{0:k} - x_{0:k}^i) \quad (35)$$

where $x_{0:k}^i$ denotes a series of states estimated by the ith particle, and w_k^i denotes the weight of the ith particle in relation to $x_{0:k}^i$. N denotes the total number of particles, and in the present study the value of N was set to 1000 as the trade-off between accuracy and computational cost. The value of the weights are adjusted at every time instance when new measurements are available as follows

$$w_k^i \propto w_{k-1}^i p\left(y_k \middle| x_k^i\right), i = 1,2, \dots, N \quad (36)$$

To overcome the degeneracy problem present in the particle filter algorithm, systematic resampling [37] is implemented in every iteration with the aim of multiplying samples with high weights and suppressing samples with low weights. Intuitively, a particle that possesses a higher weight is more likely to be duplicated and vice versa. The resampled particles are than employed to estimate the system state as per Eq. (32), and which then makes up the prior density for the next state tracking (estimation) iteration.

At a given time k (when the actual system state is not available), the future value of the system state can be obtained by carrying out a one step-ahead prediction:

$$\tilde{x}_k = \sum_{i=1}^{N} w_{k-1}^i x_k^i \quad (37)$$

A multiple-step-ahead prediction can be obtained by iteratively propagating the particles as follows

$$\tilde{x}_{k+m} = \sum_{i=1}^{N} w_{k+m-1}^i x_{k+m}^i \quad (38)$$

$$x_{k+m}^{i} = g\left(x_{k+m-1}^{i}, x_{k+m-2}^{i}, x_{k+m-3}^{i}, x_{k+m-4}^{i}\right) + v_{k+m-1} \tag{39}$$

where $\tilde{x}_{k+m}$ denotes the multiple-step-ahead prediction at time $k+m$, and x_{k+m}^{i} denotes the value of the ith particle at time $k+m$. The process noise v_{k+m-1} remained unchanged during the prediction process and its value is determined the last time the actual measurement of the system state is available, namely, $v_{k+m-1} = v_{k+m-2} = \cdots = v_{k-1}$.

Finally, the RSL is calculated based on the particles propagated by the MGFM-PF method as follows

$$RSL = \sum_{i=1}^{N} w_{k}^{i} T_{TOF}^{i} \tag{40}$$

where w_{k}^{i} denotes the estimated weight of the ith particle during the prediction process. T_{TOF}^{i} denotes the time of failure predicted by the ith particle, which is defined as the time between now and the time point at which the propagated particle reaches the pre-defined failure threshold. In other words, the estimated RSL is a weighted sum of all particles starting from the point to commence prediction until the time instance at which the propagation of the particles approached the pre-determined threshold. The failure thresholds of the pump failure cases were chosen to be the average value of the monitoring indices (i.e. at the time when the equipment was forced to shut down) of all available fault cases. Due to the limited availability of failure data, we subjectively set the failure threshold of the compressor failure case to be the largest value of the prognostic feature as did Wu et al. [39].

2.3 Overall framework of the integrated diagnostic and prognostic method

In summary, Fig. 1 shows a flowchart of the implementation process of the proposed integrated framework. Once the CVA deviation-based index exceeds the fault threshold, an alarm will be generated to give an indication of the occurrence of a fault, which will subsequently trigger the fault identification and prognostic modules. In other words, degradation data are identified by the CVA-based fault detection module. The EWMA-Pearson model performs fault identification after the occurrence of the fault, which will help in the reconstruction of the monitoring index. After the fault influential variables are successfully identified, the proposed MGFM-PF approach is utilized to propagate the trend of the reconstructed monitoring index until it reaches the pre-determined threshold. The RSL is calculated based on the probability density function of the propagated particles.

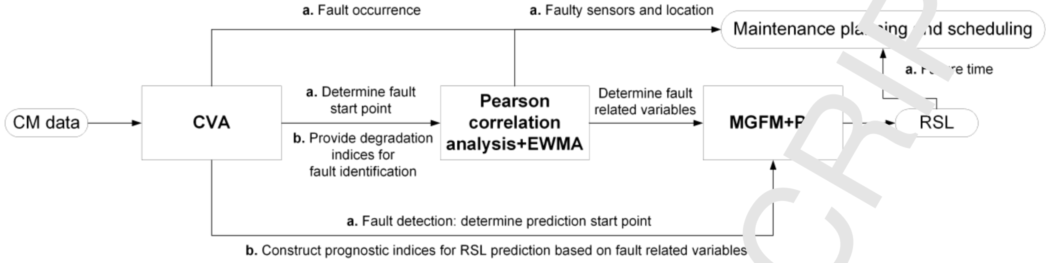


Figure 1. Flowchart of the proposed integrated diagnostic and prognostic method.

The detailed training and monitoring procedure of the proposed diagnostic and prognostic method is illustrated in Fig. 2. The extended CVA-based diagnostic method is trained using normal data. After fault detection, the EWMA-Pearson approach is applied to identify the most fault related variables using data obtained at early degradation stages. The same data set is also employed to train the MGFM-PF algorithm. The trained predictor is subsequently utilized to propagate particles forward to perform RSL prediction.

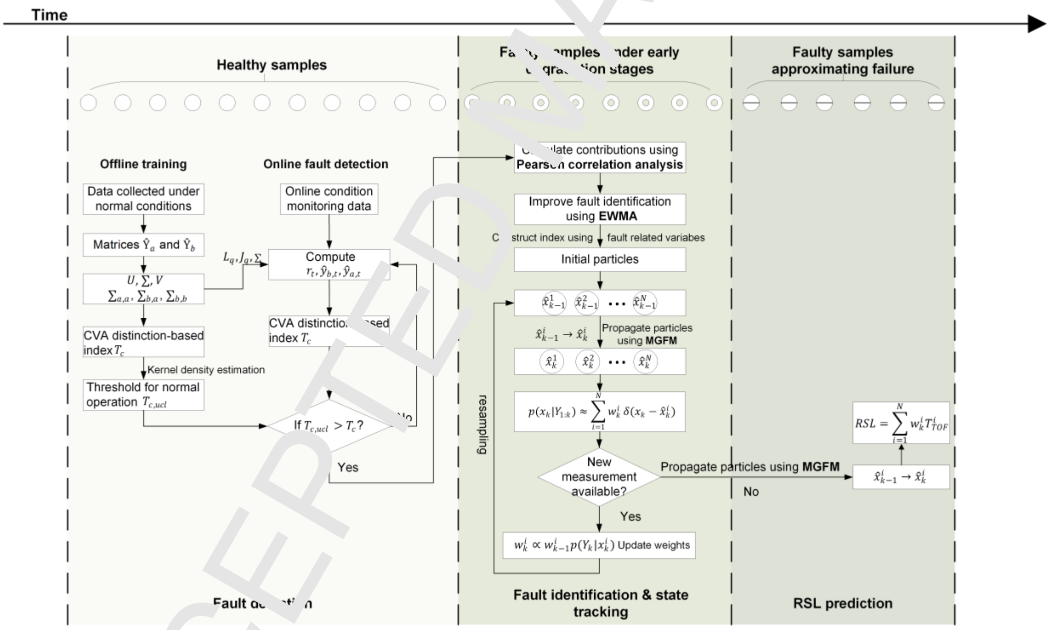


Figure 2. Detailed illustration of the proposed diagnostic and prognostic method.

3. Case studies

3.1 Continuous stirred tank reactor (CSTR) simulation cases

The purpose of this section is to evaluate the monitoring performance of the proposed CVA distinction-based index using a CSTR case study. The CSTR simulation model used in this paper is created by the authors of [40], which is specially designed for analysing slowly evolving faults. Fig. 3 illustrates the measurement locations and the control strategy of the reactor. The reactor temperature T is controlled by manipulating Q_c which denotes the coolant flow rate. The reader is referred to Karl and Li's work [40] for full details of the simulation program. Nine test sets are constructed to test the monitoring performance of the indices studied (see Table 1). Test sets 1-3 simulate sensor drifts on the measured variables T_i, T_c and T, respectively. During the simulation, a decaying component δt was added to the sensor measurements (i.e. $T_i = T_{i,0} + \delta t$; $T_c = T_{c,0} + \delta t$; $T = T_0 + \delta t$), where $T_{i,0}$, $T_{c,0}$ and T_0 denote the values of T_i, T_c and T under normal operating conditions, respectively. The values of the decaying rate δ for different test sets are detailed in Table 1. Test sets 4-6 simulate catalyst decay fault: $a_1 = a_0 \exp(-\delta t)$. In the model, a_1 is set to $a_0 = 1$ during normal operation. The decaying rate δ is varied among test sets in order to evaluate the effectiveness of the T_c index when the faults deteriorate at different rates. Test sets 7-10 simulate two simultaneous faults, catalyst decay and heat transfer fouling ($b_1 = b_0 \exp(-\delta t)$). $b_0 = 1$ denotes the value of b_1 during normal operation, and δ is the decaying rate. During faulty operating conditions, a_1 and b_1 decayed from their healthy values toward 0. In each test set, faults were introduced after 1400 min of normal operation. For illustrative purposes, the output variables under fault conditions 4 and 2 are plotted in Fig. 4 to demonstrate how the catalyst decay, and, simultaneous catalyst decay and heat transfer fouling affect the system outputs.

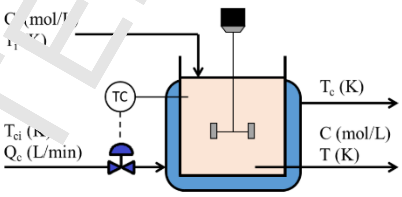


Figure 3. Schematic sketch of the CSTR process for collecting the test data [40]. C_i denotes the concentration in the reactor, T_i denotes the temperature of the ith reactor, T_{ci} denotes the outlet temperature of the ith cooling water and T_c denotes the inlet temperature of the cooling water.

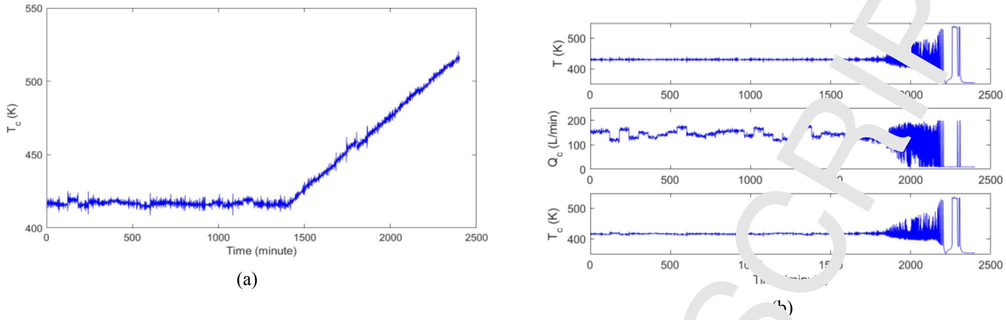
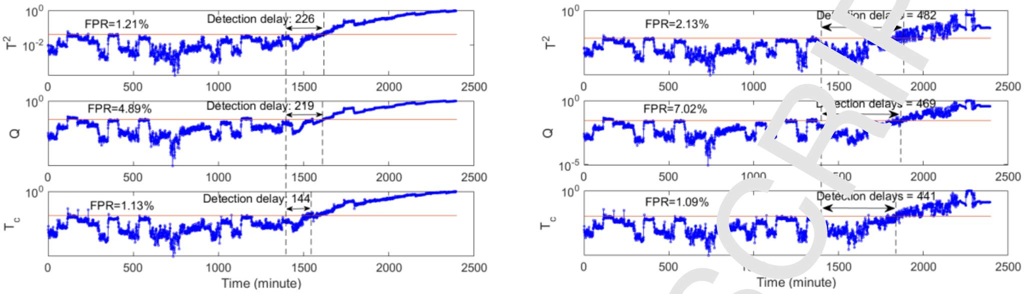


Figure 4. Output faulty variables under fault conditions 2 and 4. (a) fault case 2: T_c sensor drift, and (b) fault case 4: catalyst decay. Faults were introduced after 1400 minutes of normal operation.

Table 1. Monitoring performance of T_c, T^2 and Q

Test set	Fault description	Decaying rate setting	T_c	T^2	Q
1	Sensor drifts on the measured variable T_i	0.05	[illegible] [a]	237	272
			[illegible] [b]	7.23%	0.17%
2	Sensor drifts on the measured variable T_c	0.1	226	219	144
			1.21%	4.89%	1.13%
3	Sensor drifts on the measured variable T	0.12	96	96	96
			2.8%	7.15%	0.17%
4	Catalyst decay	0.0006	482	469	441
			0.67%	6.3%	0.71%
5	Catalyst decay	0.001	311	262	258
			0.5%	4.8%	0.5%
6	Catalyst decay	0.003	102	102	90
			4.05%	6.69%	2.76%
7	Catalyst decay + Heat transfer fouling	0.0008 for catalyst decay and 0.001[illegible] for heat transfer fouling	338	335	332
			5.39%	9.61%	0.25%
8	Catalyst decay + Heat transfer fouling	0.003 for catalyst decay and 0.004 for heat transfer fouling	110	112	106
			0.75%	7.2%	0.79%
9	Catalyst decay + Heat transfer fouling	0.0005 for catalyst decay and 0.0[illegible]1 for heat transfer fouling	507	510	506
			0.5%	4.76%	0.5%
	Averaged monitoring performance		277.89	260.22	**249.44**
			2.09%	6.51%	**0.78%**

[a] First row of each column: detection delays (minutes).
[b] Second row of each column: false positive rate (FPR).



(a) (b)

Figure 5. Monitoring results of fault conditions 2 ((a). T_c sensor drift) and 4 ((b) catalyst decay).

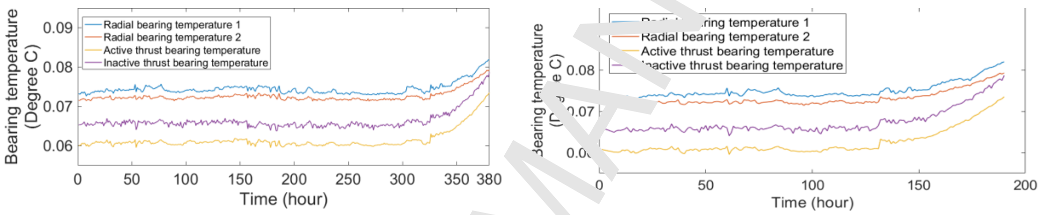
The proposed CVA distinction-based index was applied to monitor the CSTR numerical examples. All upper control limits for healthy operational conditions in this investigation were calculated at the 99% confidence level. Fault thresholds were calculated using the KDE method. The cross-validation method [5,41] was utilized to determine the optimal model order q, and a was set to 15 for all CSTR test sets. The detection delays and false positive rate (FPR) are presented in Table [illegible]. The FPR is calculated by dividing the number of false detections under normal conditions by the length of the testing data. The detection delay is defined as the period between the detection time and the start of fault. It is observable from the table that the averaged detection delays and FPR of the proposed method are smaller than those of the other methods, verifying that the proposed diagnostic method shows better monitoring performance than T^2 and Q. For illustrative purposes, the monitoring charts of test set 2 and 4 are presented in Fig. 5. In both cases, T^2 and Q struggle to cross the fault threshold, resulting in larger detection delays and larger FPRs, but the proposed distinction-based index achieves earlier detection times and lower FPRs.

The results in this section verify that the proposed CVA distinction-based index is new and outperforms traditional T^2 and Q indices when applied to slowly developing faults in that it is able to detect the faults earlier with a relative low FPR.

3.2 Centrifugal pump case studies 1 & 2

3.2.1 Fault description

In order to assess the ability of the proposed diagnostic and prognostic technique to effectively detect incipient faults and predict system RSL, the model was tested using two data sets captured from an operational industrial centrifugal pump. This pump is a high-pressure centrifugal pump running at a large refinery in Europe (hereafter referred to as pump A). The first measured time series consisted of 380 observations and 13 variables (Table 1 shows all measured variables). The second time series consisted of 197 observations. For this study, all data were captured at a sampling rate of one sample per hour. It is observable from Fig. 6 (a) that the unit is operated under healthy conditions between the 125th and the 334th point of the time series. The readings of the four different bearing-temperature sensors start to rise at around the 335th sampling point; the machine continued to run until the 380th sampling point. For fault case 2, it can be seen from Fig. 6 (b) that that the machine is operated under healthy conditions between the [illegible]th and the 137th point of the time series. The readings of the four different bearing-temperature sensors start to increase at around the 138th sampling point; the machine continued to run until the 197th sampling point. At that time, site engineers shut down the pump for inspection.



(a) (b)

Figure 6. Trend of four different bearing temperature sensor measurements of pump cases 1 and 2 (signals are normalized). (a) fault case 1; (b) fault case 2.

Table 2. Measured variables of pump A

Variable ID	Variable Name	Units
1	Speed	rpm
2	Suction pressure	bar
3	Discharge pressure	bar
4	Discharge temperature	degree C
5	Actual flow	kg/h
6	Radial vibration overall X 1	mm/s
7	Radial vibration overall Y 1	mm/s
8	Radial bearing temperature 1	degree C
9	Radial vibration overall X 2	mm/s
10	Radial vibration overall Y 2	mm/s
11	Radial bearing temperature 2	degree C
12	Active thrust bearing temperature	degree C

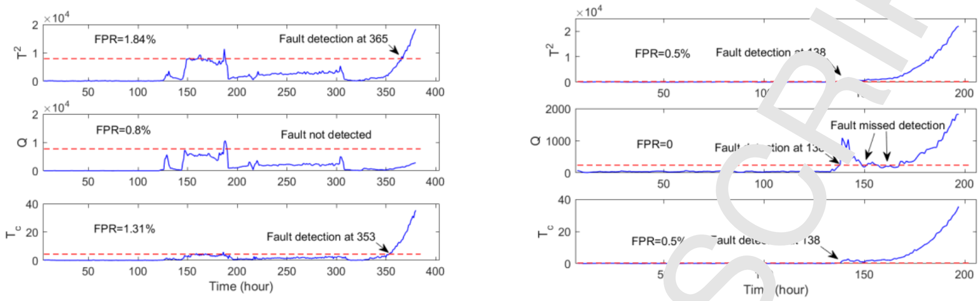
13	Inactive thrust bearing temperature	degree C

3.2.2 Results and discussion

3.2.2.1 Fault detection and identification

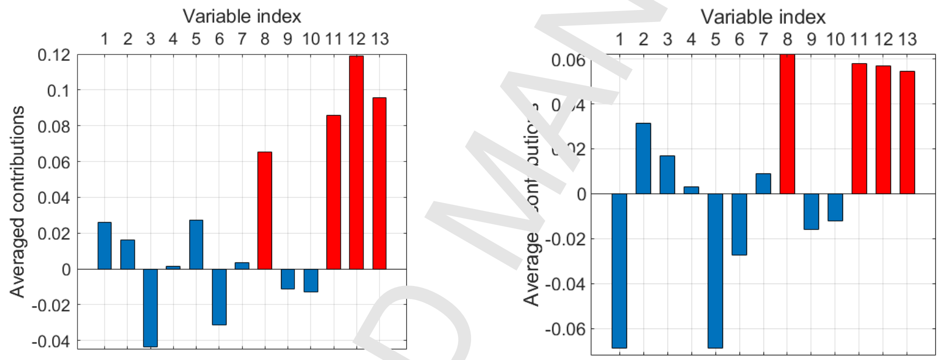
To begin with, the CVA-based diagnostic approach is trained using a data set collected from normal operating conditions. The scale of time lags a and b were estimated through the autocorrelation analysis [5] of the root summed squares of all variables in the training data set. Here the number of time lags a and b were set to 5. Since the underlying process data is non-stationary and non-linear, and does not follow a Gaussian distribution, the KDE method was adopted here to determine the upper control limits of the test statistics. All upper control limits for healthy operational conditions in this investigation were calculated at the 99% confidence level (i.e. the probability the test statistics are smaller than the predefined threshold is 99%). We set the value β equal to 0.99 in this section. Here, the optimal model order q in the CVA diagnostic model was set to 25 using the cross-validation method.

Fig. 7 (a) shows the results obtained in terms of fault detection for fault case 1. The T_c index is sensitive to small shifts in the underlying process, resulting in an earlier detection of the fault when compared to T^2 and Q statistics (12 and 27 hours earlier compared to T^2 and Q statistics, respectively). FPRs were also calculated for the tested statistics. The FPR of the T_c index is 1.31%, and the FPRs of the T^2 and Q statistics are 1.84% and 0.8%, respectively. Therefore, in this case the T_c statistics outperforms T^2 and Q in terms of fault detection time, providing ample time for subsequent product planning. Meanwhile, the T_c index has less FPR than T^2 and demonstrates similar performance as Q. It is worth noting that although the FPR of Q index is the lowest among the three test statistics, it failed to detect the fault during the entire operation of the machine and hence does not satisfy the practical application requirement. The fault detection results coincide with the conclusion made in the previous Section that the proposed index is able to detect the faults earlier with a relative low FPR. The fault detection result for fault case 2 is depicted in Fig. 7 (b). Although T_c demonstrates similar performance as T^2, it still performs better than Q in that it is able to avoid missed detections under faulty conditions. The reason why Q index incurs missed detections is that it is not as sensitive as T_c when applied to slowly developing faults. Collectively, T_c demonstrates superior or comparable performance than traditional T^2 statistics in terms of fault detection time and FPR. Moreover, T_c presents superior performance than Q statistics in terms of sensitivity to incipient faults under faulty conditions (i.e. the ability to avoid missed detections) in both cases.



(a) (b)

Figure 7. Fault-detection results for pump A case 1 & 2. (a) fault case 1; (b) fault case 2. T_c statistics obtained using CVA. Legend: solid blue – test statistics, dashed red – upper control limit.



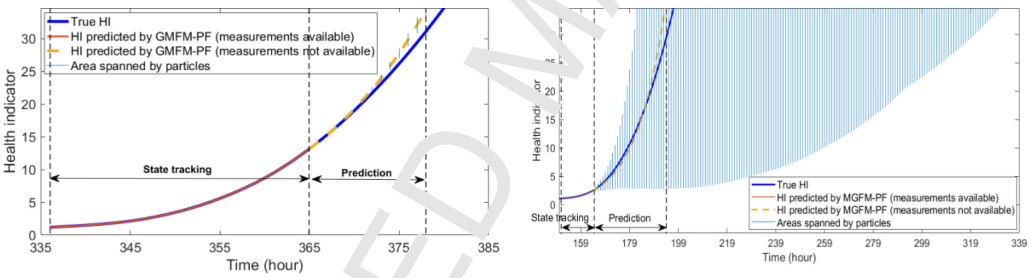
(a) (b)

Figure 8. Contribution plots for identifying the detected fault. (a) Fault case 1; (b) Fault case 2.

Table 3. Averaged percentages of contributions from the influential variables under faulty conditions

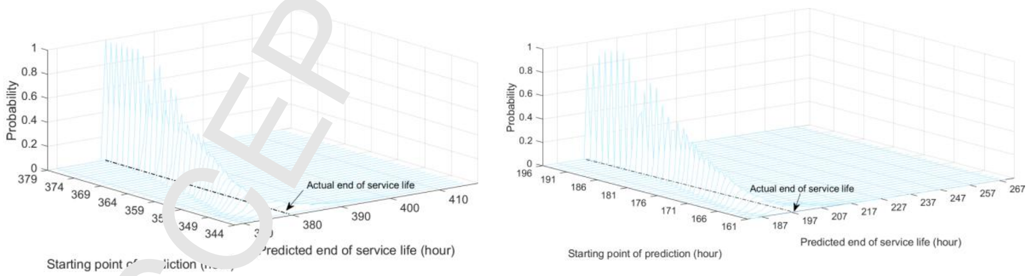
Averaged percentages of contributions	Pearson correlation analysis	EWMA-Pearson
Pump case 1	53.97%	**58.28%**
Pump case 2	66.83%	**72.28%**
Compressor	18.65%	**23.01%**

After fault detection, the proposed EWMA-Pearson fault identification approach is applied to the data obtained from early stages of deterioration. To be specific, PCCs were first determined for each process variable versus the diagnostic monitoring index. Then, the contributions (PCCs) were fed into a EWMA model to improve the fault identification performance. Here, the forgetting factor δ and width of the moving window W in Eqs. (27)-(28) were set to 0.9 and 5, respectively. Since PCCs at certain time instant only examines the contributions at one time point and therefore may not be able to accurately identify the process variables that are responsible for the detected fault. In an effort to solve this problem, the contributions calculated based on the EWMA-Pearson approach were averaged over a period of time (taken to be 25 hours starting from the 331th sampling point for fault case 1 and the 138th sampling point for fault case 2). With this process, contributions at multiple time stamps were stacked into one figure to clearly illustrate the contribution of different variables over the early degradation process. The resultant contribution plots is displayed in Fig. 8. The results indicate that variable no. 8, 11, 12 and 13 are influential variables in both cases, which is 100% accurate according to the root cause of the fault as stated previously. In order to show that EWMA-Pearson is superior to Pearson correlation analysis for fault identification, the percentages of contributions of faulty variables (i.e. the ratio of accumulated contributions from variables 8, 11, 12 and 13 to accumulated contributions from all variables) were calculated and the averaged percentages of contributions of faulty variables under faulty conditions were compared for EWMA-Pearson and Pearson correlation analysis (see Table 3). It can be seen from both case studies that the fault identification performance can be significantly improved by applying EWMA-Pearson.



(a) (b)

Figure 9. Predicted monitoring index for predictions starting at (a) the 365th sample and (b) 164th sample.



(a) (b)

Figure 10. Predicted posterior distributions and actual end of service life for (a) case 1 and (b) case 2. The black dashed line is parallel to the horizontal axis.

3.2.2.2 Fault prognostics

The MGFM-PF approach was employed to propagate the trend of the monitoring index until it reaches the pre-determined threshold. Similar to [42], a locally weighted regression filter (LOESS) with a span value of 0.3 was applied to the calculated degradation health indicator to smooth out the degradation trajectories. LOESS is a powerful smoothing technique based on a locally weighted regression function and a 2nd order polynomial function (reader is referred to [43] for more details about this technique). Fig. 9 (a) illustrates an exemplary result of the predicted monitoring index for predictions starting at the 365th sample for fault case 1. The red curve indicates the health indicator (HI) predicted by the proposed prognostic method when actual measurements of the system states are available. The light green curve indicates the mean values of the RSL probability density functions for each time instance. The shaded area shows the values within which the monitoring index was predicted as per Eqs. (8)-(15) and (18)-(3[illegible]). Similarly, Fig. 9 (b) shows the result of the predicted monitoring index for predictions starting at the 164th sample for fault case 2.

The predicted RSLs by MGFM-PF in terms of probability distributions and actual end of service life for case 1 and 2 are shown in Fig. 10. The results show that as the prediction start point gets closer to the actual failure time, the mean of the failure time distribution gets closer to the real end-of-life, and the variance gets smaller. It was mentioned in Section 1 that PF allows the quantification of the uncertainties of RSL in a probabilistic form. Fig. 10 has demonstrated that the prediction of the RSL can be obtained as probability distributions, and in order to further justify the necessity of using PF in this study, the predicted RSLs for different prediction starting points along with the associated tolerance intervals were depicted in Fig. 11. The dark blue shaded areas denote the one sigma tolerance interval (covers around 68% of the RSL density) that was derived from the RSL densities (shown in Fig. 10) with particles being assumed normally distributed. The light blue shaded areas denote the two sigma tolerance interval (covers around 95% of the RSL density). It can be seen from the figure that the confidence interval becomes narrower as the prediction start point moves toward the end-of-life. As a result, the RSL uncertainty is reduced considerably.

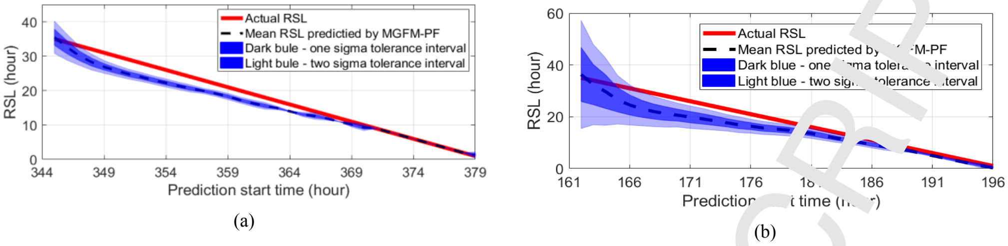


Figure 11. Predicted RSL and its confidence bounds (one and two sigma tolerance intervals) for (a) case 1 and (b) case 2.

The graphs in Fig. 12 demonstrate the prognostic performance of the proposed MGFM-PF method and MGFM [44], LSTM [45], ANFIS [46], and AR model [47] for different fault cases. The Akaike Information Criterion (AIC) was employed to determine the optimal parameters for AR models. For LSTM model, the network consists of one input layer, one hidden layer and one output layer. The number of hidden neurons in the hidden layer was set to 200. For ANFIS model, two membership functions (MFs) were chosen. It is observable from Fig. 11 that the predictive accuracy of MGFM-PF is lower at the beginning and the estimated RSLs lie well within the ∓25% confidence bounds, indicating that the proposed MGFM-PF model has the ability to accurately predict the system's remaining service life. As the prediction starting point gets closer to the actual end-of-life, the RSL predicted by MGFM-PF gets closer to the true remaining useful life, yielding more accurate estimations.

The predictive performance of the existing methods mentioned above are compared with the proposed MGFM-PF model, and the comparison results are detailed in Table 4.

Three performance metrics were used to quantitatively benchmark predictive performance of the models compared. The employed metrics are:

1. Root mean square deviation (RMSD):

$$RMSD = \sqrt{\sum_{i=1}^{N}(RSL_{pre,i} - RSL_{true,i})^2/N} \quad (41)$$

2. Mean absolute deviation (MAD):

$$MAD = \sum_{i=1}^{N}|RSL_{pre,i} - RSL_{true,i}|/N \quad (42)$$

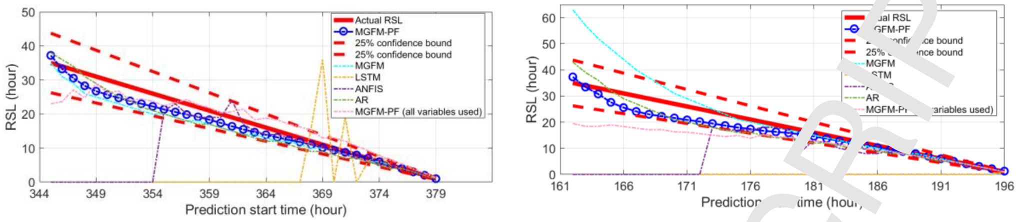
3. Cumulative relative accuracy (CRA):

$$CRA = \sum_{i=1}^{N}\left(1 - \frac{|RSL_{pre,i} - RSL_{true,i}|}{RSL_{true,i}}\right)/N \quad (43)$$

where $RSL_{pre,i}$ and $RSL_{true,i}$ denote the predicted and actual RSL, respectively, and N is the total number of predictions. The bold values in column 3 of Table 4 verify that the MGFM-PF method shows better prognostics performance than the other four methods. The RSLs were also calculated for the situations when all variables are utilized. It is clear that using only influential variables greatly improved the predictive accuracy. Furthermore, in order to obtain a general idea on the computational cost of the proposed MGFM-PF method, the processing times of MGFM-PF and its counterparts were recorded and listed in Table 4. All the methods considered were implemented with MATLAB R2018b and conducted on a desktop computer with a 3.2 GHz 8 core CPU and 16 GB memory. It is observable from Table 4 that the shortest processing time was achieved by MGFM in both cases, followed closely by AR model and MGFM-PF. LSTM and ANFIS were more time-consuming and could account for the relatively long processing times measured. However, compared with the long prediction timeframes (ranging from 1 hour to 3[illegible] hours), the processing times of all above mentioned methods are deemed to be relatively short, which confirms the applicability of the above mentioned methods for online monitoring. Collectively, the proposed MGFM-PF method outperforms its counterparts in that it greatly improves the predictive accuracy and has a computational speed that satisfies practical applications.

Table 4. Comparison of predictive accuracy and processing time using various models

Case studies	Algorithms	MGFM-PF	LSTM	AR	ANFIS	Grey	MGFM-PF (all variables used)
Pump case 1	**RMSD**	**2.44**	20.82	4	16.45	3.69	3.81
	MAD	**1.92**	17.49	[illegible]	9.69	2.97	2.8569
	CRA	**0.0051**	0.046	0.0094	0.025	0.0078	0.0075
	Processing time (s)	10.85	[illegible]	7.92	104.62	0.058	N/A
Pump case 2	**RMSD**	**3.14**	20.64	3.9	26.49	8.18	7.75
	MAD	**2.35**	17.87	3.2	16.51	4.2	5.81
	CRA	**0.012**	0.09	0.017	0.087	0.022	0.031
	Processing time (s)	10.[illegible]	[illegible]	42.66	140.25	0.032	
Compressor	**RMSD**	**1.91**	49.28	38.75	6.05	39.57	8.7
	MAD	**1.[illegible]**	41.79	33.8	3.46	34.44	4.53
	CRA	**0.0[illegible]084**	0.026	0.0209	0.0021	0.0213	0.0028
	Processing time (s)	177.04	667.88	41.11	629.42	2.26	N/A



(a) (b)

Figure 12. The comparison between the proposed MGFM-PF model and other existing methods: (a) fault case 1 and (b) fault case 2.

3.3 Centrifugal compressor case study

3.3.1 Fault description

In this section, a third industrial case study is utilized to further verify the effectiveness and generalizability of the proposed method, which involves predicting the RSL of a centrifugal compressor. This machine is a high-pressure centrifugal compressor (hereafter referred to as compressor A). The dataset consists of 1614 observations and 21 variables. Table 5 summarizes the name of different process variables. The sampling rate is one sample per hour. The degradation data are shown in Fig. 13, and the root-cause variables are the stage 3 drive-end radial vibration sensors overall X and Y. The compressor was turned off at the 1614th sampling point due to high levels of vibration.

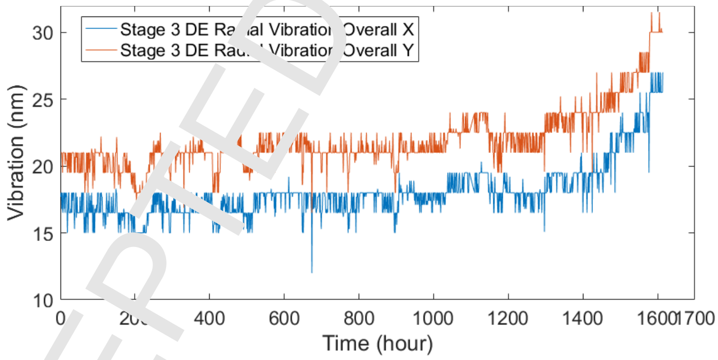


Figure 13. Trend of two different bearing vibration sensor measurements of compressor A

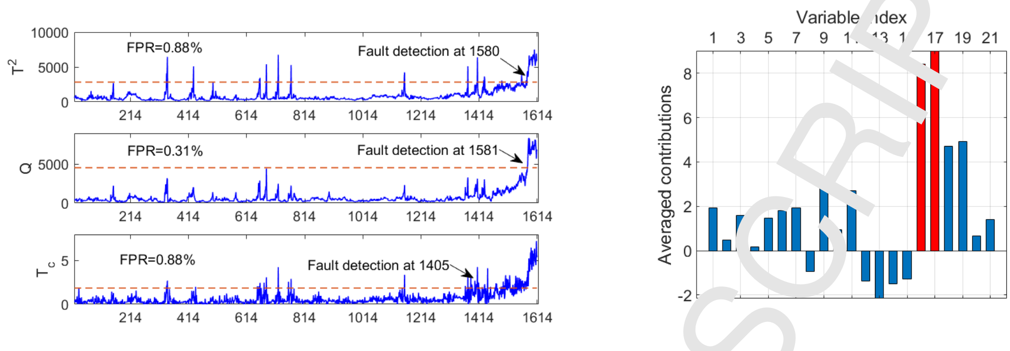
Table 5. Measured variables of compressor A

ID	Variable Name	ID	Variable Name
1	Stage 1 Suction Pressure	12	Stage 1-2 Non-drive-end (NDE) Radial Vibration Overall X
2	Stage 1 Discharge Pressure	13	Stage 1-2 Non-drive-end (NDE) Radial Vibration Overall Y
3	Stage 1 Suction Temperature	14	Stage 1-2 Thrust Position Axial Probe 1
4	Stage 1 Discharge Temperature	15	Stage 1-2 Thrust Position Axial Probe 2
5	Stage 2 Suction Pressure	16	Stage 3 Drive-end (DE) Radial Vibration Overall X
6	Stage 2 Discharge Pressure	17	Stage 3 Drive-end (DE) Radial Vibration Overall Y
7	Stage 2 Suction Temperature	18	Stage 3 Non-drive-end (NDE) Radial Vibration Overall X
8	Stage 2 Discharge Temperature	19	Stage 3 Non-drive-end (NDE) Radial Vibration Overall Y
9	Stage 3 Suction Pressure	20	Stage 3 Thrust Position Axial Probe 1
10	Stage 1-2 Drive-end (DE) Radial Vibration Overall X	21	Stage 3 Thrust Position Axial Probe 2
11	Stage 1-2 Drive-end (DE) Radial Vibration Overall Y		

3.3.2 Results and discussion

3.3.2.1 Fault detection and identification

Similar to the procedure described in Section 3.2, the scale of time lags a and b were estimated through the autocorrelation analysis [5] and were set to 10. The optimal model order q in Eq. (18) was set to 17 using the cross-validation method. Fig. 14 (a) and (b) demonstrate the results obtained in terms of fault detection and identification, respectively. In Fig. 14 (a), the T_c statistics is more sensitive than T^2 and Q statistics at the early stages of deterioration and crosses its control limits at 1405 hour with a FPR of 0.88%. In this study, fault detection is defined as the first time when 6 consecutive sampling points are above the control limits, which was suggested by the authors of [48,49]. The fault detection results verify that the proposed method maximizes the fault detectability under an acceptable FPR when compared to traditional CVA test statistics. Fig. 14 (b) indicates the most influential variables are variables 16 and 17, which is 100% accurate according to the root-cause of the fault as stated in Section 3.3.1. Moreover, the averaged percentages of contributions of influential variables under faulty conditions are compared for EWMA-Pearson and Pearson correlation analysis (see Table 3). It is clear that EWMA-Pearson can obtain a better contribution rate of influential variables comparing to the performance of Pearson correlation analysis.



(a) (b)

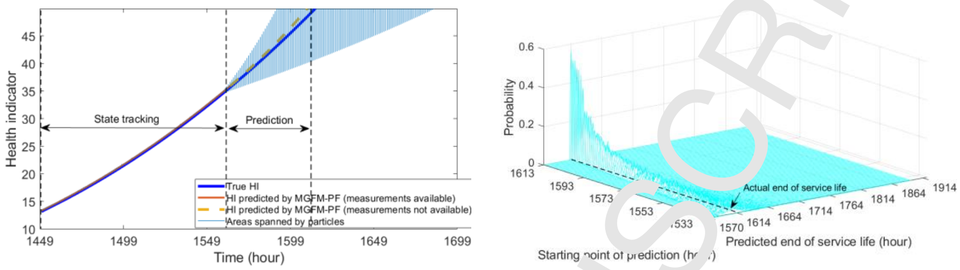
Figure 14. Fault detection and identification results of compressor A. (a) fault detection, (b) fault identification.

3.3.2.2 Fault prognostics

Fig. 15 (a) shows the predicted degradation feature at 1559 hour. Fig. 15 (b) shows the predicted posterior distributions of failure time at the different prediction start points for compressor A, which implicate the potential degradation mechanics at the corresponding time. It can be seen that as the prediction start point moves towards the actual failure time, the mean of the failure time distribution gets closer to the real failure time, and the variance gets smaller.

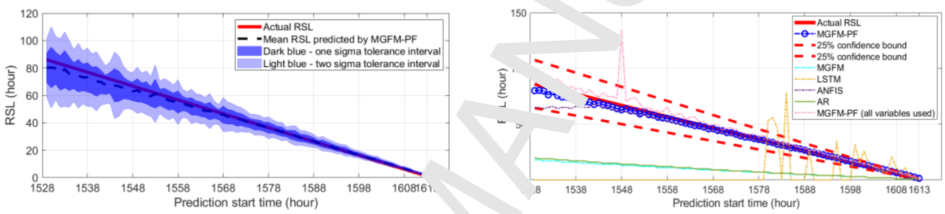
The predicted RSLs for different prediction starting points are illustrated in Fig. 16 (a). The data before 1527 hour are used for state tracking and the RSL forecast starts at 1528 hour. The black dashed line indicates the mean RSL that were derived from the RSL density functions shown in Fig. 15 (a). The dark and light blue areas represent the one and two sigma tolerance intervals containing 68% and 95% of the RSL probability densities at each cycle, respectively. The red straight line demonstrates the actual RSL. It is clear that the mean RSL curve of the proposed method coincides with the actual RSL.

Finally, the RSL of compressor A predicted using MGEM-PF was compared with MGFM, LSTM, ANFIS, and AR model. The comparison results are depicted in Fig. 16 (b) and the three performance metrics were calculated and listed in Table 4. RMSD, MAD and CRA of MGFM-PF are all minimal compared to the others algorithms. The evaluation metrics confirm that the MGFM-PF method outperforms its counterparts. It can also be observed from Table 4 that the prediction errors of MGFM-PF when only the influential variables were used are smaller than those when all process variables were utilized, which verifies the necessity of using only influential variables for prediction. The forecast information provided by the proposed MGFM-PF model can be used to develop production plans in advance and provide ample time for organizing spare repairs and scheduling maintenance so as to prevent serious abnormal conditions, catastrophic failures or even emergency situations.



(a) (b)

Figure 15. (a) Predicted health indicator at 1559 hour; (b) Predicted posterior distributions of compressor A.



(a) (b)

Figure 16. (a) Predicted RSL and its confidence bounds (one and two sigma tolerance intervals) for compressor A; (b) The comparison between the proposed MGFM-PF model and other existing methods.

4. Conclusion

In this work, an integrated diagnostic and prognostic framework was proposed for incipient fault diagnosis and remaining service life prediction in nonlinear dynamic processes. The developed approach was validated through a CSTR case study and condition monitoring data acquired from an industrial centrifugal pump and a compressor. Fault diagnosis was carried out by comparing the values of the canonical distinction-based diagnostic index with pre-defined thresholds. The proposed diagnostic approach can distinguish normal operational conditions from slowly developing faults incurring system anomalies leading to an early detection of faults. Moreover, the proposed EWMA-Pearson method can effectively identify fault influential variables in early degradation stages, and the fault identifiability is greatly improved when compared to Pearson correlation analysis. MGFM was introduced to learn the system evolution and to compensate for the lack of historical failure data. The procedure for propagating the particles and forecasting RSLs was performed in a stable and fast manner, due to the computational benefits of the MGFM. The novel MGFM-PF scheme leads to very good prediction results. The predictive accuracy of the MGFM-PF method was demonstrated to be superior to MGFM, LSTM, ANFIS and AR model. Through experiments, capability of using only fault related variables for RSL prediction has been shown. The associated processing time, although, larger than AR and MGFM, confirms the general applicability of the MGFM-PF technique for online condition monitoring. In three real-world machine health monitoring case studies, the robustness and effectiveness of the proposed condition monitoring approach were verified. The proposed hybrid diagnostic method can be used to provide site engineers with reliable diagnosis of rotating machinery and meanwhile the MGFM-PF technique can assist in the subsequent production planning and decision-making process and enhance profitability by eliminating unpredicted failures. While the proposed framework is tested using run-to-failure data captured from a centrifugal pump and a compressor and forms an initial step towards machinery prognostics in grey model framework, it can be applied to other industrial rotating machines, such as gas turbines and wind turbines. With the continuous improvement of machinery, data on faulty are increasingly limited. The needs for such a methodology can only increase.

A consideration for future study is to improve the model's performance for long-term RSL predictions. In addition, strategies for identifying fault influential variables at an earlier stage of degradation should be further studied in the future.

5. Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.71671091, 71701098), the China Scholarship fund, Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX18_0233), and the Leverhulme Trust International Research Network project (IN-2014-020).

References

[1] R. Zhao, D. Wang, R. Yan, S. Member, K. Mao, F. Shen, J. Wang, Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks, IEEE Trans. Ind. Electron. 65 (2018) 1539–1548.

[2] X. Li, F. Duan, D. Mba, I. Bennett, Multidimensional prognostics for rotating machinery: A review, Adv. Mech. Eng. 9 (2017) 1–20. doi:10.1177/1687814016685004.

[3] S. Wook, C. Lee, J. Lee, J. Hyun, I. Lee, Fault detection and identification of nonlinear processes based on kernel PCA, Chemom. Intell. Lab. Syst. 75 (2005) 55–67. doi:10.1016/j.chemolab.2004.05.001.

[4] J. Fan, Y. Wang, Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis, Inf. Sci. (Ny). 259 (2014) 369–379. doi:10.1016/j.ins.2013.06.021.

[5] X. Li, F. Duan, P. Loukopoulos, I. Bennett, D. Mba, Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor, Control Eng. Pract. 72 (2018) 177–191.

doi:10.1016/j.conengprac.2017.12.006.

[6] W. Li, S.J. Qin, Consistent dynamic PCA based on errors-in-variables subspace identification, J. Process Control. 11 (2001) 661–678.

[7] S. Yin, X. Zhu, S. Member, O. Kaynak, Improved PLS Focused on Key-Performance- Indicator-Related Fault Diagnosis, IEEE Trans. Ind. Electron. 62 (2015) 1651–1658.

[8] C. Ruiz-cárcel, Y. Cao, D. Mba, L. Lao, R.T. Samuel, Statistical process monitoring of a multiphase flow facility, Control Eng. Pract. 42 (2015) 74–88. doi:10.1016/j.conengprac.2015.04.012.

[9] G. Stefatos, A. Ben Hamza, Dynamic independent component analysis approach for fault detection and diagnosis, Expert Syst. Appl. 37 (2010) 8606–8617. doi:10.1016/j.eswa.2010.06.101.

[10] H. Ji, X. He, J. Shang, D. Zhou, Incipient fault detection with smoothing techniques in statistical process monitoring, Control Eng. Pract. 62 (2017) 11–21. doi:10.1016/j.conengprac.2017.03.001.

[11] B. Jiang, R.D. Braatz, Fault detection of process correlation structure using canonical variate analysis-based correlation features, J. Process Control. 58 (2017) 131–138. doi:10.1016/j.jprocont.2017.09.003.

[12] H. Chen, B. Jiang, S. Member, N. Lu, Z. Mao, Deep PCA Based Real-Time Incipient Fault Detection and Diagnosis Methodology for Electrical Drive in High-Speed Trains, IEEE Trans. Veh. Technol. 67 (2018) 4819–4830. doi:10.1109/TVT.2018.2818538.

[13] P.E.P. Odiowei, C. Yi, Nonlinear Dynamic Process Monitoring Using Canonical Variate Analysis and Kernel Density Estimations, IEEE Trans. Ind. Informatics. 6 (2010) 36–45. doi:10.1109/TII.2009.2032654.

[14] R. Tan, Y. Cao, Deviation Contribution Plots of Multivariate Statistics, IEEE Trans. Ind. Informatics. 15 (2019) 833–841. doi:10.1109/TII.2018.2841658.

[15] K. Liu, N.Z. Gebraeel, J. Shi, A Data-Level Fusion Model for Developing Composite Health Indices for Degradation Modeling and Prognostic Analysis, IEEE Trans. Autom. Sci. Eng. 10 (2013) 652–664. doi:10.1109/TASE.2013.2250282.

[16] J.M. Lucas, M.S. Saccucci, Exponentially Weighted Moving Average Control Schemes : Properties and Enhancements, Technometrics. 32 (1990) 1–12.

[17] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, London, Edinburgh, Dublin Philos. Mag. J. Sci. 2 (1901) 559–572. doi:10.1080/14786440109462720.

[18] J. Adler, I. Parmryd, Quantifying Colocalization by Correlation : The Pearson Correlation Coefficient is Superior to the Mander ' s Overlap Coefficient, Cytom. Part A. 77 (2010) 733–742. doi:10.1002/cyto.a.20896.

[19] N.S. Freedman, J. Gazendam, L. Levan, A.I. Pack, R.J. Schwab, Abnormal Sleep / Wake Cycles and the Effect of Environmental Noise on Sleep Disruption in the Intensive Care Unit, Am. J. Respir. Crit. Care Med. 163 (2001) 451–457.

[20] R.O. Duda, P.E. Hart., Estimation of time delay by coherence analysis, Proc. IEEE. 75 (1987) 236–255.

[21] R.O. Duda, P.E. Hart., Pattern classification and scene analysis, A Wiley-Interscience Publication, New York: Wiley, 1973.

[22] J. Chen, C.-M. Liao, F.R.J. Lin, M.-J. Lu, Principle component analysis based control charts with memory effect for process monitoring, Ind. Eng. Chem. Res. 40 (2001) 1516–1527.

[23] H. Ji, X. He, J. Shang, D. Zhou, Incipient Sensor Fault Diagnosis Using Moving Window Reconstruction-Based Contribution, Ind. Eng. Chem. Res. 55 (2016) 2746–2759.

[24] B. Wei, N. Xie, A. Hu, Optimal solution for novel grey polynomial prediction model, Appl. Math. Model. 62 (2018) 717–727. doi:10.1016/j.apm.2018.06.035.

[25] S. Liu, Y. Yang, J. Forrest, Grey data analysis, Springer Singapore, 2017.

[26] Z. Chen, Bayesian filtering: From Kalman filters to particle filters, and beyond, Statistics (Ber). 182 (2003) 1–69.

[27] M.S. Jha, G. Dauphin-Tanguy, B. Ould-Bouamama, Particle filter based hybrid prognostics for health monitoring of uncertain

systems in bond graph framework, Mech. Syst. Signal Process. 75 (2016) 301–329. doi:10.1016/j.ymssp.2016.0[illegible].010.

[28] C. Chen, B. Zhang, G. Vachtsevanos, M. Orchard, Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering, IEEE Trans. Ind. Electron. 58 (2011) 4353–4364. doi:10.1109/TIE.2010.2098369.

[29] C. Lu, Z. Wang, W. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identi fi cation, Signal Processing. 130 (2017) 377–388. doi:10.1016/j.sigpro.2016.07.028.

[30] L. Liao, F. Köttig, A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction, Appl. Soft Comput. J. 44 (2016) 191–199. doi:10.1016/j.asoc.2016.03.013.

[31] G.H. Golub, C. Reinsch, Singular value decomposition and least squares solutions, in: Linear Algebra, 1971: pp. 134–151.

[32] J. Benesty, S. Member, J. Chen, Y.A. Huang, On the Importance of the Pearson Correlation Coefficient in Noise Reduction, IEEE Trans. Audio. Speech. Lang. Processing. 16 (2008) 757–765.

[33] K. Pearson, Mathematical contributions to the theory of evolution.—III. Regression, heredity and panmixia, Philos. Trans. R. Soc. 187 (1895) 253–318.

[34] B.C. Juricek, D.E. Seborg, W.E. Larimore, Fault Detection Using Canonical Variate Analysis, Ind. Eng. Chem. Res. 43 (2004) 458–474.

[35] B.K. Flury, H. Riedwyl, Standard distance in univariate and multivariate analysis, Am. Stat. 40 (1986) 249–251.

[36] H. Hotelling, New Light on the Correlation Coefficient and its Transforms, J. R. Stat. Soc. Ser. B. 15 (1953) 193–232.

[37] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2002) 174–188. doi:10.1109/78.978374.

[38] D. An, J.H. Choi, N.H. Kim, Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab, Reliab. Eng. Syst. Saf. 115 (2013) 161–169. doi:10.1016/j.ress.2013.02.019.

[39] J. Wu, C. Wu, S. Cao, S.W. Or, C. Deng, X. Shao, Degradation Data-Driven Time-To-Failure Prognostics Approach for Rolling Element Bearings in Electrical Machines, IEEE Trans. Ind. Electron. 66 (2018) 529–539. doi:10.1109/TIE.2018.2811366.

[40] K.E.S. Pilario, Y. Cao, Canonical Variate Dissimilarity Analysis for Process Incipient Fault Detection, IEEE Trans. Ind. Informatics. 3203 (2018) 1–8. doi:10.1109/TII.2018.2810822.

[41] X. Li, F. Duan, I. Bennett, D. Mba, Canonical variate analysis, probability approach and support vector regression for fault identification and failure time prediction, J. Intell. Fuzzy Syst. 34 (2018). doi:10.3233/JIFS-169550.

[42] K. Javed, R. Gouriveau, N. Zerhouni, Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics, IEEE Trans. Ind. Electron. 62 (2015) 647–656. doi:10.1109/TIE.2014.2327917.

[43] J. Iwaniec, W. Lisowski, T. Uhl, Nonparametric approach to improvement of qualitt of modal parameters estimation, J. Theor. Appl. Mech. 43 (2005) 327–3[illegible]4.

[44] S. Liu, J. Forrest, Y. Yang, A brief introduction to grey systems theory, Grey Syst. Theory Appl. 2 (2012) 89–104. doi:10.1109/GSIS.2011.[illegible]4401[illegible].

[45] S. Hochreiter, J. Schmidhuber, Long Short-term Memory, Neural Comput. 9 (1997) 1735–1780.

[46] J.R. Jang, ANFIS : Adap tive-Network-Based Fuzzy Inference System, IEEE Trans. Syst. Man, Cybern. Syst. 23 (1993) 665–685. doi:10.1109/21.256541.

[47] Robert F. Engle, Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation, Econometrica. 50 (1982) 987–1008. doi:10.2307/1912773.

[48] Z. Chen, S.X. Ding, K. Zhang, Z. Li, Z. Hu, Canonical correlation analysis-based fault detection methods with application to alumina evaporation process, Control Eng. Pract. 46 (2016) 51–58. doi:10.1016/j.conengprac.2015.10.006.

[49] L.H. Chiang Russell, E. L. , and Braatz, R. D, Fault diagnosis in chemical processes using Fisher discriminant analysis,

discriminant partial least squares, and principal component analysis, Chemom. Intell. Lab. Syst. 50 (2000) 243–252.

- The development of a novel RSL prediction model (e.g. MGFM-PF) able to address the challenge of failure data scarcity in the era of big machinery data to enable accurate prediction of the RSL of a system.
- The development of a novel integrated framework that covers all MHMS, from detection of incipient faults to determination of fault related variables to prediction of remnant life.
- Incipient fault diagnosis using a novel CVA distinction-based index.
- The development of the EWMA-Pearson method for early and accurate identification of fault related variables.
- The use of degradation data captured from an operational industrial centrifugal pump and a compressor.