An intelligent diagnostic and prognostic framework for large-scale rotating machinery in the presence of scarce failure data

Xiaochuan Li[1] ⓘ, Xiaoyu Yang[1,2], Yingjie Yang[1], Ian Bennett[3] and David Mba[1,4]

Abstract

In this work, a novel diagnostic and prognostic framework is proposed to detect faults and predict remaining service life of large-scale rotating machinery in the presence of scarce failure data. In the proposed framework, a canonical variate residuals–based diagnostic method is developed to facilitate remaining service life prediction by continuously implementing detection of the prediction start time. A novel two-step prognostic feature exploring approach that involves fault identification, feature extraction, feature selection and multi-feature fusion is put forward. Most existing prognostic methods lack a fault-identification module to automatically identify the fault root-cause variables required in the subsequent prognostic analysis and decision-making process. The proposed prognostic feature exploring method overcomes this challenge by introducing a canonical variate residuals–based fault-identification method. With this method, the most representative degradation features are extracted from only the fault root-cause variables, thereby facilitating machinery prognostics by ensuring accurate estimates. Its effectiveness is demonstrated for slow involving faults in two case studies of an operational industrial centrifugal pump. Moreover, an enhanced grey model approach is developed for remaining useful life prediction. In particular, the empirical Bayesian algorithm is employed to improve the traditional grey forecasting model in terms of quantifying the uncertainty of remaining service life in a probabilistic form and improving its prediction accuracy. To demonstrate the superiority of empirical Bayesian–grey model, existing prognostic methods such as grey model, particle filter–grey model and empirical Bayesian–exponential regression are also utilized to realize machinery remaining service life prediction, and the results are compared with that of the proposed method. The achieved predictive accuracy shows that the proposed approach outperforms its counterparts and is highly applicable in fault prognostics of industrial rotating machinery. The use of in-service data in a practical scenario shows that the proposed prognostic approach is a promising tool for online health monitoring.

Keywords

Condition monitoring, fault detection, prognosis, canonical variable analysis, grey model

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 subjected 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 of a system rather than just detecting the faults. It comprises two steps: (1) detect process abnormalities after the occurrence of failures and (2)

[1]Faculty of Computing, Engineering and Media, De Montfort University, Leicester, UK
[2]Nanjing University of Aeronautics and Astronautics, Nanjing, China
[3]Shell Centre, London, UK
[4]University of Nigeria, Nsukka, Nigeria

Corresponding author:
Xiaochuan Li, Faculty of Computing, Engineering and Media, De Montfort University, The Gateway, Leicester LE1 9BH, UK.
Email: xiaochuan.li@dmu.ac.uk

estimate future working conditions and the remaining service life (RSL). It is clear that those procedures are crucial for the safe, reliable, efficient and sustainable operation of any industrial system. Therefore, it is no surprise that automated data-driven MHMS has been an active research area for decades now.

It is unnecessary and difficult to perform RSL prediction when machinery is operating under healthy conditions since little information about the degradation trend can be found during this stage. To make a prognostic framework suitable for online monitoring, a diagnostic module should be included to ensure that the RSL prediction module is triggered only after certain failures are detected. Multivariate statistical process monitoring (MSPM) techniques have recently seen improvements in diagnosing process abnormalities. MSPM techniques such as principal component analysis (PCA),[2] independent component analysis (ICA)[3] and canonical variate analysis (CVA)[4] have been widely applied for the detection of process abnormalities in industrial plants and systems. In addition, alternatives to the standard multivariate monitoring methods,[5–8] which take into consideration the correlations between timestamps in the past and the future, have also been put forward for dynamic processes monitoring. Among the aforementioned MSPM techniques, CVA-based approaches were shown to be superior to other monitoring methods in terms of applicability to nonlinear dynamic systems.[9] In this study, a CVA-based monitoring index, namely canonical variate residuals (CVR),[10] is adopted to determine the start time of the unhealthy stage.

Construction of prognostic features plays an important role in machinery prognostics. RSL is usually obtained by estimating when the prognostic features reach the pre-defined threshold. An appropriate prognostic feature is expected to facilitate machinery prognostics by ensuring accurate estimates and simplifying the prognostic modelling. The prognostic feature can be data from a single sensor,[11] a feature extracted from raw sensory signals using dimension reduction techniques,[12,13] or the integration of multiple features extracted from multiple sensory signals.[14–16] In the abovementioned studies, prognostic features were either constructed using sensor signals acquired in a specific component where a known fault occurs (e.g. vibration signals for detecting bearing faults, current signals for detecting wind turbine gearbox faults and thermal couples for detecting valve failures) or generated with the integration of features extracted from all the sensors mounted on a machine. One of the challenges in effectively applying these techniques to the prognostics of operational large-scale rotating machines is the lack of a fault-identification module to automatically identify the fault root-cause variables required in the subsequent prognostic analysis. Including a fault-identification module comes along with various benefits, such as excluding variables which do not show a degradation trend, thereby reducing the computational costs in relation to feature selection, alleviating the curse of dimensionality, improving predictive accuracy and providing useful information for spare parts and maintenance materials provisioning. In an effort to solve this problem, a CVR-based fault-identification technique under the residual generation framework is put forward in this article. Although CVA has been demonstrated to outperform traditional MSPM methods for the monitoring of nonlinear dynamic processes,[17] and CVR have been shown to be a more sensitive index compared to CVA-based T^2 and Q indices,[10] investigation on the application of CVR to fault identification is limited. Unlike in Jiang and Braatz,[18] Jiang et al.[19] and Lu et al.,[20] the proposed CVR-based fault-identification framework does not need to know the cause-and-effect relationships among process variables and/or historical data regarding different types of faults. Moreover, this method can be applied to real industrial processes involving many strongly correlated variables with a low computational cost.

After fault identification, time-domain degradation features are then extracted from the identified faulty variables to represent the degradation process of the system. The accuracy of the RSL predictions is largely dependent on the performance of the extracted degradation features. Thus, it is crucial to select suitable features before subsequent prognostic analysis is carried out. Many performance metrics are available for the evaluation of the suitability of the extracted features for RSL prediction, including the widely applied monotonicity,[16] robustness[21] and correlation.[15] The trend of a degradation feature is expected to present correlation with the machine's service time. Thus, a metric named trendability was developed later on and utilized in Javed et al.[22] to measure the correlation between the degradation feature and time. Since a single performance metric may not be sufficient for prognostic feature selection, a hybrid metric which considers the properties of trendability, monotonicity, robustness and correlation is put forward in this study for selecting degradation features. In summary, a two-step prognostic feature exploring approach which consists of a fault-identification and a feature extraction–selection–fusion module is proposed in this study for the purpose of improving the predictive accuracy.

Various data-driven techniques are available for the prediction of remaining life, including self-organizing map,[12] *K*-nearest neighbours,[12] support vector machine[23] and neural networks.[4] Those prognostic techniques offer a trade-off between reliability, speed and applicability. In order to ensure accurate prediction,

prognostic models usually require large amounts of failure data for training. However, field failure data are extremely difficult to obtain, and this prevents those 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 fulfil prognostic tasks, since safety-critical equipment are usually not allowed to run to failure. For instance, the same type of failure has only been observed twice previously in 3 years of continuous monitoring of the pump used in this study. In order to deal with the problem of scarce data and to ensure accurate prediction in the presence of scarce failure data, we apply grey model (GM), which was originally devised to tackle small sample problems,[24] to implement RSL prediction. Moreover, a single-valued forecasting model learnt from historical data can only provide estimates that are deterministic in nature and can only provide predicted values which have no probabilistic significance. To address this issue, this article proposes an enhanced GM, called EB-GM, based on the GM and empirical Bayesian (EB) approaches. EB may be thought of as a method that borrows strength from each predicted local RSLs to pull each local estimate towards the true failure time to improve the overall predictive accuracy. Another reason why EB is adopted in this study is that it is particularly suitable for small sample analyses,[25] making it a promising tool for remaining lifetime prediction where field failure data scarcity is incurred.

The major contributions of this article are as follows:

- A CVR-based fault-detection method was adopted to continuously monitor the status of the system, enabling alarms and prognostic module to be automatically triggered when fault occurs, making the proposed prognostic approach suitable for online monitoring.
- A two-step prognostic feature discovering approach was proposed. In the first step, a CVR-based fault-identification module was developed to automatically locate and identify the detected fault, thereby allowing the most fault representative variables to be selected as inputs to the feature extraction module to maximize the accuracy of RSL prediction. This is the first time that the CVR-based contributions have been derived and utilized to facilitate prognostic analysis. The importance of including the CVR-based fault identification in prognostics is three-fold. First, the process variables that do not show a degradation trend were eliminated automatically at this stage. From a prognostics perspective, variables that do not show a significant trend are not suitable for RSL prediction.[16] Computationally,

the inclusion of only the fault-related variables provides a far more efficient method of discovering prognostic features than does the conventional non-selective feature extraction approach. Third, 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. In the second step, four different metrics, namely monotonicity, robustness, correlation and trendability, were employed to evaluate the suitability of the time-domain features extracted from the identified faulty variables for RSL prediction.

- An enhanced GM-based prognostic approach (i.e. EB-GM) is developed to assure that the proposed prognostic framework is capable of estimating RSL with satisfactory accuracy in the presence of scarce failure data.
- Predicting RSL using real-world in-service data collected from an operational industrial centrifugal pump.

Methodology

Framework of the proposed diagnostic and prognostic approach

Figure 1 shows the flowchart of the implementation process of the proposed integrated framework. The approach comprises the following three main steps:

- *Fault detection (determine prediction start time)*: Once the CVR 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 routine.
- *Discovering prognostic features*: The CVR-based fault-identification method identifies root-cause variables after the occurrence of the fault. Eight time-domain features, namely mean, root mean square, standard deviation, variance, skewness, kurtosis, crest factor and peak value, are extracted from the identified root-cause variables. The suitability of those features for RSL prediction is evaluated. The selected degradation features are then fused into a single-dimensional health indicator.
- *RSL prediction*: The GM approach is utilized to propagate the trend of the prognostic feature until it reaches the pre-determined threshold, providing an observation of the local RSL. Then, EB is employed to calculate the global RSLs based on the predicted local RSLs. In this study, the failure threshold is chosen to be the averaged maximum
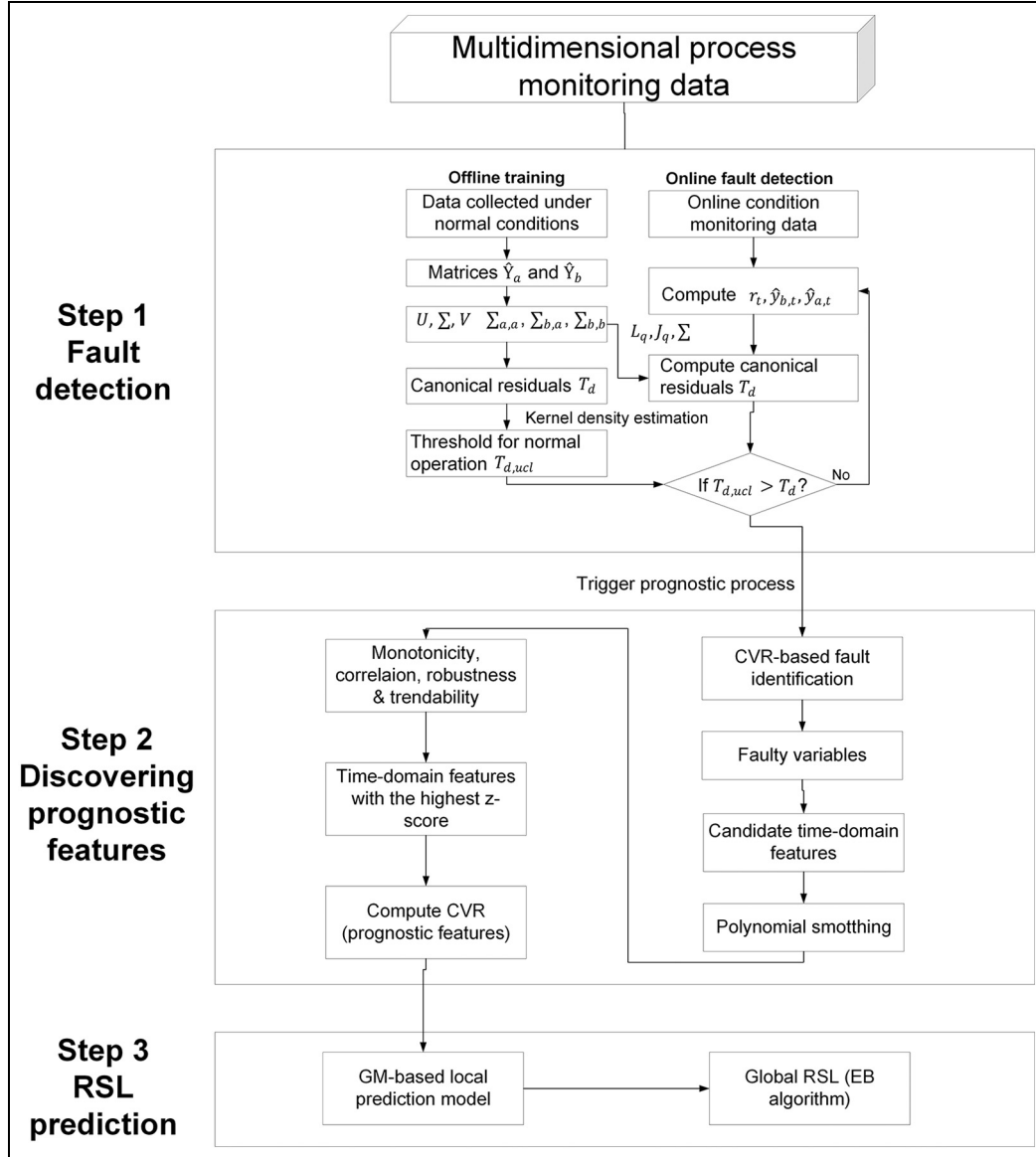
Figure 1. Flowchart of the proposed diagnostic and prognostic method.

value of the prognostic features (i.e. at the time when the equipment was forced to shut down) of all available fault cases.

Fault detection

CVR-based fault detection. Given two sets of variables $y_{1,t_k} \in \mathcal{R}^n$ and $y_{2,t_k} \in \mathcal{R}^n$, the purpose of CVA is to find matrices K and G that maximize the correlation between $z_{1,t_k} = K \cdot y_{1,t_k}$ and $z_{2,t_k} = G \cdot y_{2,t_k}$, where z_{1,t_k} and z_{2,t_k} are a linear combination of y_{1,t_k} and y_{2,t_k}, respectively. z_{1,t_k} and z_{2,t_k} are of reduced dimension compared with the original data since CVA is essentially a dimensionality reduction technique. Let

$[y_{t_1}, y_{t_2}, \ldots, y_{t_k}]$ $(y_{t_k} \in \mathcal{R}^n)$ denote a monitored time series captured from a system, where n is the number of process variables. The data can be augmented at every time instance, t_k, to generate two data sets as follows

$$y_{a,t_k} = \begin{bmatrix} y_{t_{k-1}} \\ y_{t_{k-2}} \\ \vdots \\ y_{t_{k-a}} \end{bmatrix} \in \mathcal{R}^{na} \qquad (1)$$

$$y_{b,t_k} = \begin{bmatrix} y_{t_k} \\ y_{t_{k+1}} \\ \vdots \\ y_{t_{k+b-1}} \end{bmatrix} \in \mathcal{R}^{nb} \qquad (2)$$

Apparently, at every time instance t_k, y_{a,t_k} and y_{b,t_k} are obtained by including the number of past measurements, a, and the number of future measurements, b. y_{a,t_k} and y_{b,t_k} are then normalized to the zero-mean vectors $\hat{y}_{a,t_k}$ and $\hat{y}_{b,t_k}$. Then, two augmented versions of the original data $[y_{t_1}, y_{t_2}, \ldots, y_{t_k}]$ can be obtained as follows

$$\hat{Y}_a = \left[\hat{y}_{a,t_k}, \hat{y}_{a,t_{k+1}}, \ldots, \hat{y}_{a,t_{k+N}}\right] \in \mathcal{R}^{na \times N} \quad (3)$$

$$\hat{Y}_b = \left[\hat{y}_{b,t_k}, \hat{y}_{b,t_{k+1}}, \ldots, \hat{y}_{b,t_{k+N}}\right] \in \mathcal{R}^{nb \times N} \quad (4)$$

The augmented time series $\hat{Y}_a$ and $\hat{Y}_b$ consist of the normalized $\hat{y}_{a,t_k} / \hat{y}_{b,t_k}$ obtained for the entire monitoring duration. N is the length of the augmented data and is calculated as $M - a - b + 1$, where M is the length of the original monitored data. In order to perform fault detection using CVA, a diagnostic observer need to be constructed from the augmented time series $\hat{Y}_a$ and $\hat{Y}_b$. To achieve this goal, one could perform singular value decomposition (SVD) on a matrix $\mathcal{H}$, which is composed of the covariance matrices $\sum_{a,a}$ and $\sum_{b,b}$ and cross-covariance matrix $\sum_{a,b}$ of $\hat{Y}_a$ and $\hat{Y}_b$ as follows

$$\mathcal{H} = \sum_{b,b}^{-1/2} \sum_{b,a} \sum_{a,a}^{-1/2} = U \sum V^T \quad (5)$$

$$\sum_{a,a} = \frac{\hat{Y}_a \hat{Y}_a^T}{(N-1)}, \quad \sum_{b,b} = \frac{\hat{Y}_b \hat{Y}_b^T}{(N-1)}, \quad \sum_{b,a} = \frac{\hat{Y}_b \hat{Y}_a^T}{(N-1)} \quad (6)$$

where $\sum = \text{diag}(\lambda_1, \ldots, \lambda_{na})$, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{na} > 0$ is a diagonal matrix and $\lambda_1, \ldots, \lambda_{na}$ are called canonical correlations.

Motivated by the fact that CVA is able to find maximum correlations between past and future data $\hat{Y}_a$ and $\hat{Y}_b$, practitioners can detect subtle changes by examining how far away future canonical variates are deviated from past canonical variates. This leads to a diagnostic observer called CVR that quantifies the distinctions between the past and the future measurements. CVR are generated as

$$r_{t_k} = L_q^T \hat{y}_{b,t_k} - \sum_q J_q^T \hat{y}_{a,t_k} \quad (7)$$

where L_q^T denotes the first q rows of the projection matrix L^T, and $L^T = U_q^T \sum_{b,b}^{-1/2}$. Similarly, J_q^T is the first q rows of the projection matrix J^T, and $J^T = V_q^T \sum_{a,a}^{-1/2}$. $\sum_q = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$ is a diagonal matrix with its diagonal elements being the first q canonical correlations calculated as equation (5). Then, a diagnostic test statistics T_d can be formed as the multivariate standard distance of the discrepancy features from zero[26]

$$T_d = f\left(c(r_{t_k} - 0)^T S^{-1}(r_{t_k} - 0)\right)$$

$$= \frac{\left|c(r_{t_k} - 0)^T S^{-1}(r_{t_k} - 0)\right|}{|c|\left[(r_{t_k} - 0)^T S^{-1} S S^{-1}(r_{t_k} - 0)\right]} \quad (8)$$

$$= \left[(r_{t_k})^T S^{-1}(r_{t_k})\right]^{1/2} = \left[r_{t_k}^T \left(I - \sum \sum^T\right) r_{t_k}\right]^{1/2}$$

where c is a normalizing constant. The roots of the multivariate standard distance between two random vectors can be traced back to the results presented in Flury and Riedwyl.[26]

Exploring prognostic features

Identification of faulty variables using CVR-based contributions. The main purpose of this step is to identify the root-cause variables so that fault irrelevant variables can be excluded from the subsequent prognostic analysis. The importance of including the CVR-based fault identification in prognostics has been clarified in section 'Introduction'. In this article, we propose to identify the root-cause variables using a contribution plot, which can be obtained as follows

$$C_{T_d} = T_d = r_{t_k}^T \left(I - \sum_q^2\right)^{-1} r_{t_k}$$

$$= r_{t_k}^T \left(I - \sum_q^2\right)^{-1} \left(L_q \hat{y}_{b,t_k} - \sum_q J_q \hat{y}_{a,t_k}\right) \quad (9)$$

$$= \sum_{j=1}^{n} i, T_d$$

where C_{i,T_d} is a measure of the contribution of a variable $\hat{y}_i$ to the detected fault.

Similarly, the residual space contributions is defined as[27]

$$C_Q = Q = e^T e = e^T G \hat{y}_{a,t_k} = \sum_{i=1}^{n} \sum_{j=1}^{na-q} e_j G_{j,i} \hat{y}_{a,i} = \sum_{i=1}^{n} C_{i,Q} \quad (10)$$

Finally, the variable contribution based on the state and residual space can be obtained as

$$C_{T_d,Q} = 0.5 C_{T_d} + 0.5 C_Q \quad (11)$$

where $C_{T_d,Q}$ is a measure of the contribution of a process variable to the detected fault. Root-cause variables that incur large variations during faulty conditions will have a large $C_{T_d,Q}$ value.

Discovering prognostic features from the identified faulty variables. The main purpose of this step is to construct

a prognostic feature which is suitable for RSL prediction using the root-cause variables identified previously. This step involves extracting time-domain features from the root-cause variables, selecting the most appropriate features and constructing a prognostic feature using CVR-based dimension reduction technique.

Extracting time-domain features. After the fault root-cause variables are identified among all process variables, multiple time-domain features will be extracted so as to better reveal the deterioration process. Time-domain analysis is employed to extract mean, root mean square, standard deviation, variance, skewness, kurtosis, crest factor and peak value (see Appendix 1) from the root-cause variables. Normally, vibration signals/current measurements require a much higher sampling rate than conventional process measurements (e.g. pressure, flow rate or temperature) due to the fast dynamics of vibration and current measurements. With such high sampling rates, frequency domain features could be extracted after windowing the vibration/current signals and performing a short Fourier transform.[17] Then, 'signature' frequency domain features that correspond to different mechanical faults can be used to diagnose fault conditions and predict fault evolution.[28] Similarly, time-frequency domain features can be extracted over selected frequency bands across the samples.[29] However, the data used in this study were process measurements captured by the monitoring system of the plant. The sampling rate for all measurements was one sample per hour (e.g. 1/3600 Hz). Considering the low sampling rate of the data,

time-domain characteristics were chosen to represent the condition of the machine.

Construction of prognostic features. Since some extracted time-domain features may not be able to fully capture the dynamics of the system during degradation, the main purpose of this section is to find the features that are mostly suitable for RSL prediction and to fuse the selected features into a single-digit prognostic feature. Since one criterion may not be sufficient to select a suitable degradation feature for the task of RSL prediction, a hybrid metric which considers the properties of trendability, monotonicity, robustness and correlation is proposed in this study. The feature construction procedure comprises the following four main steps:

- Similar to Liao[14] and Liu et al.,[30] a polynomial function is applied to fit a trajectory of the time-domain features to accommodate the effect of noise. Polynomial smoothing is a flexible and robust curve fitting tool and has been widely used in practice. Polynomial curve fitting is conducted to decompose a time-domain feature into smooth trend and random noise as follows

$$F(t_k) = F_T(t_k) + F_N(t_k) \tag{12}$$

where $F(t_k)$ is the extracted time-domain feature, $F_T(t_k)$ is the smooth trend and $F_N(t_k)$ is the noise component.

- The metrics for evaluating the trendability, monotonicity, robustness and correlation of a degradation feature are as follows

$$
\begin{aligned}
Trend(F) &= \frac{K\left(\sum\limits_{k=1}^{K} \tilde{F}_T(t_k)\tilde{t}_k\right) - \left(\sum\limits_{k=1}^{K} \tilde{F}_T(t_k)\right)\left(\sum\limits_{k=1}^{K} \tilde{t}_k\right)}{\sqrt{\left[K\sum\limits_{k=1}^{K} \tilde{F}^2_T(t_k) - \left(\sum\limits_{k=1}^{K} \tilde{F}_T(t_k)\right)^2\right]\left[K\sum\limits_{k=1}^{K} \tilde{t}_k^2 - \left(\sum\limits_{k=1}^{K} \tilde{t}_k\right)^2\right]}} \\[2mm]
Mon(F) &= \frac{1}{K-1}\left|\sum\limits_{k=1}^{K} \delta(F_T(t_{k+1}) - F_T(t_k)) - \sum\limits_{k=1}^{K} \delta(F_T(t_k) - F_T(t_{k+1}))\right| \\[2mm]
Rob(F) &= \frac{1}{K}\sum\limits_{k=1}^{K} \exp\left(-\left|\frac{F_N(t_k)}{F(t_k)}\right|\right) \\[2mm]
Corr(F) &= \frac{\left|K\left(\sum\limits_{k=1}^{K} F_T(t_k)t_k\right) - K\sum\limits_{k=1}^{K} F_T(t_k)\sum\limits_{k=1}^{K} t_k\right|}{\sqrt{\left[K\sum\limits_{k=1}^{K} F_T(t_k)^2 - \left(\sum\limits_{k=1}^{K} F_T(t_k)\right)^2\right]\left[K\sum\limits_{k=1}^{K} t_k^2 - \left(\sum\limits_{k=1}^{K} t_k\right)^2\right]}}
\end{aligned}
\tag{13}
$$

where $\tilde{F}_T(t_k)$ and $\tilde{t}_k$ are the rank sequence of the feature $F_T(t_k)$ and time t_k, respectively.

- With the criteria of trendability, monotonicity, robustness and correlation, feature selection is achieved by calculating a weighted linear combination of the performance criteria. The optimum weighted coefficients are obtained by maximizing the objective function described as follows

$$I(F) = w_1 Trend(F) + w_2 Mono(F) + w_3 Rob(F) + w_4 Corr(F)$$

$$\text{s.t.} \sum_{j=1}^{4} w_j = 1, w_j > 0$$

$$(14)$$

where $I(F)$ is the objective function to be optimized and $w_j(j = 1, 2, 3, 4)$ are the weighting coefficients. Since I is positively related to the performance criteria, $I(F)$ is positively related to the time-domain features. The higher the I value, the more suitable the feature for RSL prediction.

- Due to the fact that the selected degradation features may still be high dimensional, a dimension reduction technique is required to fuse the selected features into a one-dimensional prognostic feature. In this article, we explore the capability of CVR for representing the underlying degradation process and calculate the CVR index based on the selected features as a prognostic feature.

RSL prediction

After a suitable prognostic feature is constructed following the procedures described in the previous sections, the GM-EB prognostic technique can be applied to predict the RSL of the machine. The GM-EB technique enhances the traditional GM method in a number of ways. First, GM-EB improves the predictive accuracy of GM. In order to improve the prediction made at each time instance, a number of GMs are constructed for each prediction starting point to generate a number of local RSL estimates. These local RSL estimates obtained at every time instance can be used to derive a local RSL observation (e.g. a weighted sum of all local RSL estimates). The EB algorithm assumes that the local RSL observations obtained at different time are related in that they can be treated as having a common distribution. Then, the final RSLs are calculated as a comprise between the local RSL observation and the distribution mean. These RSLs are called global RSLs. EB-GM enhances GM by allowing probabilistic RSL estimates and improved accuracy. Section 'GM' gives some basic knowledge about GM. Section

'Predicting local RSLs based on GM' details how the local RSLs are predicted. Section 'Predicting global RSLs based on EB' discusses how the prior and posterior distributions of the global RSLs are estimated from the observed local RSLs.

GM. GM 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. Grey forecasting model is suitable for real-time prediction with limited availability of degradation data. GM uses operations of accumulated generations to build different equations. The general procedure for GM is described in Appendix 1.

Predicting local RSLs based on GM. A single-valued grey forecasting model learnt from historical data may not be able to accurately predict the degradation trajectory because its outputs are deterministic in nature and do not consider measurement and process noise. To overcome this problem, several GMs can be trained for the same prediction start time. In this way, a number of local RSL estimates can be generated, thereby providing more reliable predictions than a single-valued estimate made by a single GM. As illustrated in Figure 2, a sequence of prognostic features $x_{t_i:t_k}$ $(i = 1, 2, \ldots, k - 3)$ can be obtained at time t_k. x_{t_1} is the value of the prognostic feature at the prediction start time which is determined by the fault-detection module, and x_{t_k} is the
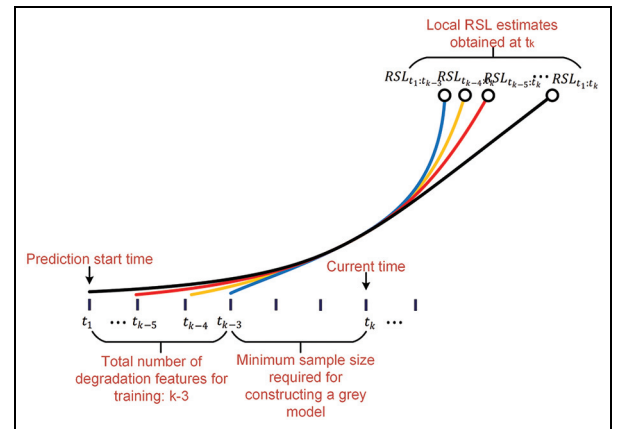

Figure 2. Illustration of how local RSL estimates, $RSL_{j:t_k}$ $(j = t_1, t_2, \ldots, t_{k-3})$, are obtained at time t_k.

value of the prognostic feature at time t_k. $x_{t_i:t_k}$ $(i = 1, 2, \ldots, k-3)$ contains $k-3$ degradation trajectories. Since GM requires at least four consecutive inputs, $x_{t_{k-3}:t_k}$ is the shortest prognostic feature that can be used for training a GM. Therefore, a total of $k-3$ local GMs are trained at time t_k. These trained models will be used to perform RSL prediction by propagating the prognostic features from t_k and onwards until it reaches the pre-defined threshold. These trained models will generate $k-3$ number of local RSL estimates. Each local RSL estimate can be assigned a weight w_j $(j = t_1, t_2, \ldots, t_{k-3})$ which is calculated based on the model fitting error $e_f(j)$, $j = t_1, t_2, \ldots, t_{k-3}$, and one-step prediction error, $e_p(j)$, $j = t_1, t_2, \ldots, t_{k-3}$, as follows

$$w_j = \frac{\dfrac{1}{e_f(j) + e_p(j)}}{\displaystyle\sum_{i=t_1}^{t_k} \dfrac{1}{e_f(i) + e_p(i)}}, \quad j = t_1, t_2, \ldots, t_{k-3} \tag{15}$$

The Euclidean distance was employed as a metric to measure the model fitting and one-step prediction errors. Based on the calculated weights, the observation of the local RSL at time t_k and the associated variance can be computed as

$$\begin{cases} \widehat{RSL}_{t_k} = \displaystyle\sum_{j=t_1}^{t_{k-3}} w_j RSL_{j:t_k} \\ \sigma_{t_k}^2 = \dfrac{1}{k-3} \displaystyle\sum_{j=t_1}^{t_{k-3}} w_j \left(RSL_{j:t_k} - \widehat{RSL}_{t_k} \right)^2 \end{cases} \tag{16}$$

where $\widehat{RSL}_{t_k}$ denotes the observation of the local RSL at time t_k and $\sigma_{t_k}^2$ is the associated variance. $RSL_{j:t_k}$ $(j = t_1, t_2, \ldots, t_{k-3})$ indicates the RSL estimate obtained at time t_k using a GM that is trained using data $x_{j:t_k}$.

Predicting global RSLs based on EB. The second step of the proposed EB-GM prognostic method involves applying the EB algorithm to the local RSL observations, $\widehat{RSL}_{t_k}$, to realize the global RSL prediction. Assume the local RSL observations follow a common prior distribution $\pi(\theta_{t_k}) \sim N(\mu, \tau^2)$, where μ and τ^2 are unknown parameters. Then, this case becomes a parametric EB problem. The following parts of this section discuss how the prior and posterior distributions of the global RSLs are estimated from the observed local RSLs.

$\widehat{RSL}_{t_k}$ is an observation of θ_{t_k} at time t_k, and $\sigma_{t_k}^2$ is a measure of the error for each RSL observation. To account for the influence of system dynamics, we set the sample variance σ^2 to be the mean of all local $\sigma_{t_k}^2$ and σ^2 calculated as follows

$$\sigma^2 = \frac{\sum_{i=t_1}^{t_k} \sigma_i^2}{k} \tag{17}$$

where t_1 and t_k denote the starting and ending points of the prediction, respectively. According to Casella[31] and Berger,[32] the joint distribution of $(\widehat{RSL}_{t_k}, \theta_{t_k})$ can be computed as

$$\begin{aligned} p\left(\widehat{RSL}_{t_k}, \theta_{t_k} \right) &= p\left(\widehat{RSL}_{t_k} | \theta_{t_k} \right) p(\theta_{t_k}) \\ &= \frac{1}{2\pi\sigma\tau} \exp\left\{ -\frac{\rho}{2} \left[\theta_{t_k} - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{\widehat{RSL}_{t_k}}{\sigma^2} \right) \right]^2 \right\} \\ &\quad \exp\left\{ -\frac{\left(\mu - \widehat{RSL}_{t_k} \right)^2}{2(\sigma^2 + \tau^2)} \right\} \end{aligned} \tag{18}$$

where $\rho = (1/\tau^2) + (1/\sigma^2)$.

We need to know the statistic characteristics of μ and τ^2, and all the information about the unknown hyper parameters μ and τ^2 is contained in the marginal distribution of $\widehat{RSL}_{t_k}$, $p(\widehat{RSL}_{t_k})$, which is computed as[31]

$$\begin{aligned} p\left(\widehat{RSL}_{t_k} \right) &= \int_{-\infty}^{+\infty} p\left(\widehat{RSL}_{t_k}, \theta_{t_k} \right) d\theta_{t_k} \\ &= \int_{-\infty}^{+\infty} p\left(\widehat{RSL}_{t_k} | \theta_{t_k} \right) p(\theta_{t_k}) d\theta_{t_k} \\ &= \frac{1}{\sqrt{2\pi}\rho\sigma\tau} \exp\left\{ -\frac{\left(\widehat{RSL}_{t_k} - \mu \right)^2}{2(\sigma^2 + \tau^2)} \right\} \sim N\left(\mu, (\sigma^2 + \tau^2) \right) \end{aligned} \tag{19}$$

The likelihood function of all estimated local RSLs $\widehat{RSL}_{t_1:t_k}$ is computed as

$$\begin{aligned} m&\left(\widehat{RSL}_{t_1}, \ldots, \widehat{RSL}_{t_k} \right) \\ \prod_{j=t_1}^{t_k} m\left(\widehat{RSL}_j \right) &= \prod_{j=t_1}^{t_k} \frac{1}{[2\pi(\sigma^2 + \tau^2)]^{\frac{1}{2}}} \exp\left\{ -\frac{\left(\widehat{RSL}_j - \mu \right)^2}{2(\sigma^2 + \tau^2)} \right\} \\ &= \left[2\pi(\sigma^2 + \tau^2) \right]^{-\frac{k}{2}} \exp\left\{ \frac{-ks^2}{2(\sigma^2 + \tau^2)} \right\} \\ &\quad \exp\left\{ \frac{-k(\hat{\mu}_k - \mu)^2}{2(\sigma^2 + \tau^2)} \right\} \end{aligned} \tag{20}$$

where $\hat{\mu}_k = \sum_{j=t_1}^{t_k} \widehat{RSL}_j / k$ and $s^2 = \sum_{j=t_1}^{t_k} (\widehat{RSL}_j - \hat{\mu}_k)^2 / k$.

With the likelihood function $m(\widehat{RSL}_{t_1}, \ldots, \widehat{RSL}_{t_k})$, the next step is to seek maximum $m(\widehat{RSL}_{t_1}, \ldots, \widehat{RSL}_{t_k})$

in order to calculate the unknown parameters μ and τ^2 of the prior distribution $\pi(\theta_{t_k}) \sim N(\mu, \tau^2)$. Clearly, $\hat{\mu}_k$ is always the maximum likelihood estimation (MLE) of $\widehat{RSL}_{t_k}$, regardless of the value of τ^2. Hence it will always be true that $\hat{\mu} = \hat{\mu}_k$, being the MLE choice of μ. Therefore, the problem of maximum likelihood is then simplified to the maximization of the following function

$$\psi(\tau^2) = \left[(\sigma^2 + \tau^2) \right]^{-\frac{k}{2}} \exp\left\{ -\frac{ks^2}{2(\sigma^2 + \tau^2)} \right\} \quad (21)$$

over τ^2. It is easier to maximize $\log \varphi(\tau^2)$

$$\frac{d}{d\tau^2} \log \varphi(\tau^2) = \frac{-k}{2(\sigma^2 + \tau^2)} + \frac{ks^2}{2(\sigma^2 + \tau^2)^2} \quad (22)$$

According to Berger,[32] if $s^2 < \sigma^2$, then the maximum value is achieved when $\tau^2 = 0$. If $s^2 \geq \sigma^2$, equation (22) achieves maximum at $\tau^2 = (((1/k)s^2) - \sigma^2)$.

Therefore, the MLE estimates of the unknown parameters μ and τ^2 are

$$\begin{cases} \hat{\mu} = \hat{\mu}_k = \dfrac{\sum\limits_{j=t_1}^{t_k} \widehat{RSL}_j}{p} \\ \hat{\tau}^2 = \max\left(0, \frac{1}{k}s^2 - \sigma^2\right) \end{cases} \quad (23)$$

After MLE estimates of the unknown parameters μ and τ^2 are calculated, the posterior distribution of the global RSLs can be computed as follows

$$p\left(\theta_{t_k} | \widehat{RSL}_{t_k}\right) = \frac{p\left(\widehat{RSL}_{t_k}, \theta_{t_k}\right)}{m\left(\widehat{RSL}_{t_k}\right)}$$
$$= \left(\frac{\rho}{2\pi}\right)^{1/2} \exp\left\{ -\frac{\rho}{2} \left[\theta_{t_k} - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{\widehat{RSL}_{t_k}}{\sigma^2}\right) \right]^2 \right\} \quad (24)$$

Berger[32] suggested the following modified posterior distribution, $N(\mu^{EB}(\widehat{RSL}_{t_k}), V^{EB}(\widehat{RSL}_{t_k}))$, when the number of local RSL observations are moderate or small, where

$$\begin{cases} \mu^{EB}\left(\widehat{RSL}_{t_k}\right) = \widehat{RSL}_{t_k} - \hat{B}\left(\widehat{RSL}_{t_k} - \hat{\mu}_k\right) \\ V^{EB}\left(\widehat{RSL}_{t_k}\right) = \sigma^2\left(1 - \frac{k-1}{k}\hat{B}\right) + \frac{2}{k-3}\hat{B}^2\left(\widehat{RSL}_{t_k} - \hat{\mu}_k\right)^2 \end{cases} \quad (25)$$

where

$$\hat{B} = \left(\frac{k-3}{k-1}\right)\frac{\sigma^2}{\sigma^2 + \hat{\tau}^2}, \quad \hat{\tau}^2 = \frac{\sum_{j=t_{k1}}^{t_k}\left(\widehat{RSL}_j - \hat{\mu}_k\right)^2}{k-1} - \sigma^2$$

Therefore, the posterior distribution $p(\theta_{t_k} | \widehat{RSL}_{t_k})$ which describes the density of the global RSLs given the observations $\widehat{RUL}_{t_1:t_k}$ is obtained.

Case studies

Fault description

In order to assess the ability of the proposed diagnostic and prognostic framework to effectively detect faults and predict system RSL, the model was tested using data captured from an operational industrial centrifugal pump. This pump is a high-pressure centrifugal pump in real-life service conditions at a large refinery in Europe (hereafter referred to as pump A). It is worth noting that the data used in this article were operational data rather than experimental data. The data were recorded using the same monitoring system as used in Loutas et al.[33] and Loukopoulos et al.[34] For confidentiality purposes, we cannot detail the exact machine type. There are two case studies as the same type of failure has been observed only twice previously in 3 years of continuous monitoring of the pump. The measured time series consisted of 13 variables (Table 1 shows all measured variables). For this study, all data were captured at a sampling rate of one sample per hour. It is observed from Figure 3 that the readings of the four different bearing-temperature sensors start to increase near the end of the time series; the machine continued to run until the end of each time series. At that time, site engineers shut down the pump for inspection.

Table 1. Measured variables of pump A.

| Variable ID	Variable name	Units
1	Speed	r/min
2	Suction pressure	bar
3	Discharge pressure	bar
4	Discharge temperature	°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	°C
9	Radial vibration overall X 2	mm/s
10	Radial vibration overall Y 2	mm/s
11	Radial bearing temperature 2	°C
12	Active thrust bearing temperature	°C
13	Inactive thrust bearing temperature	°C
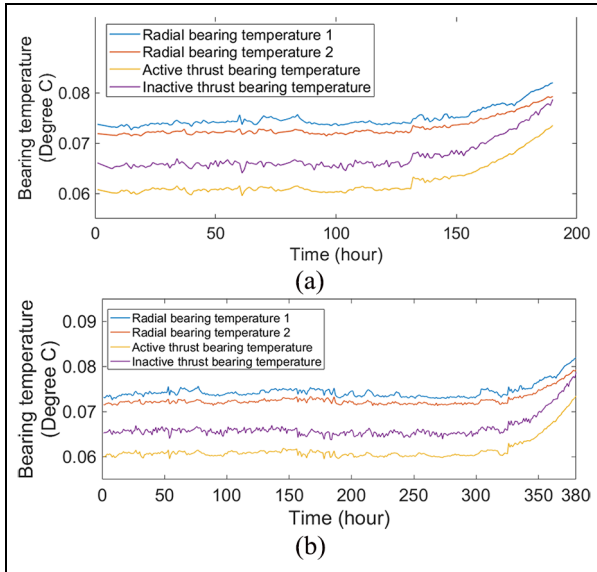
Figure 3. Trend of four different bearing-temperature sensor measurements of pump A (signals are normalized): (a) case study 1 and (b) case study 2.

Fault detection

To begin with, the CVR-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[4] of the root summed squares of all variables in the training data set. Here, the number of time lags a and b was set to 5. Since the underlying process data are non-stationary and nonlinear, and do not follow a Gaussian distribution, the kernel density estimation (KDE)[35] 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 pre-defined threshold is 99%). We set the β value equal to 0.99 in this study.

A key step in CVA is to determine the order of the reduction, that is, its number of retained states q. In this study, the optimal number of retained state q was selected such that the false alarm rate is minimized during cross-validation.[4] $q = 25$ was finally adopted in this work as it resulted in the lowest number of false positives. Figure 4 shows the results obtained in terms of fault detection for both case studies. The T_d statistics are sensitive to small shifts in the underlying process, resulting in a timely detection of the fault. The fault-detection results of the proposed method were compared with those of the PCA method (Figure 4(b) and (d)). It can be seen that although PCA offers similar fault-detection results in terms of fault-detection time in both case studies, it can result in large amounts of false alarms in the second case. These false alarms are caused by the changes in operational conditions, and this observation coincides with the findings of Ruiz-Cárcel et al.[9] in that CVA is more suitable than PCA
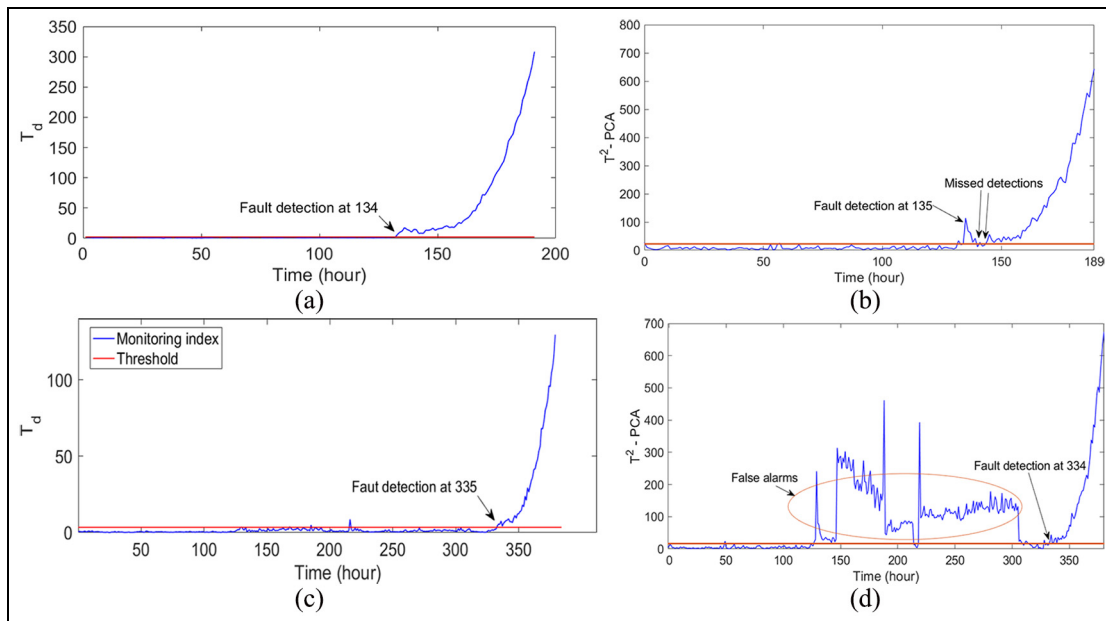

Figure 4. Fault-detection results for pump A: (a) case study 1 – proposed CVA; (b) case study 1 – PCA; (c) case study 2 – proposed CVA; (d) case study 2 – PCA.
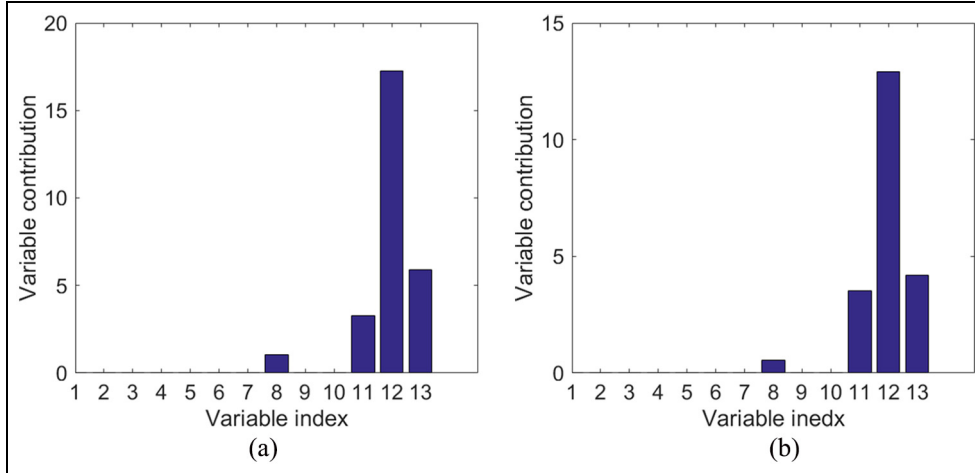Solid blue: test statistics; dashed red: upper control limit.

Figure 5. Contribution plots for identifying the detected fault: (a) case study 1 and (b) case study 2.
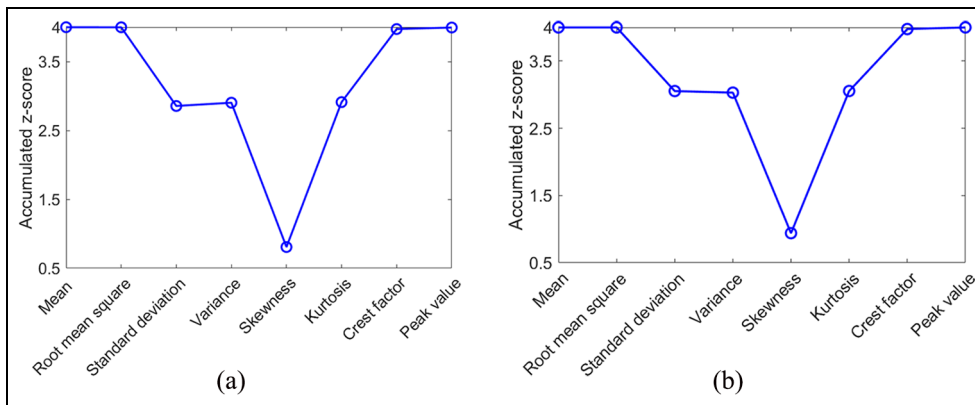

Figure 6. Performance metric $I(F)$ of the extracted features: (a) case study 1 and (b) case study 2.

for the monitoring of processes under changing operational conditions.

Discovering prognostic features

After fault detection, the proposed CVR-based fault-identification approach is first applied to identify the contribution from each process variable to the detected faults with the aim of identifying the faulty variables. The resultant contribution plots for fault cases 1 and 2 are displayed in Figure 5(a) and (b), respectively. The results indicate that variables 8, 11, 12 and 13 are the faulty variables in both cases because they have more dominant effects on the faults, which is coincident with the root cause of the fault as stated previously.

After the faulty variables were identified, eight time-domain statistical features are extracted. However, some features do not show a clear degradation trend and should not be considered for the subsequent prognostic analysis. Therefore, feature selection is implemented as

per equations (12)–(14) to select the features that are most suitable for RSL prediction. The accumulated performance scores $I(F)$ of the extracted features over all faulty variables (i.e. variables 8, 11, 12 and 13) are calculated and plotted in Figure 6. It can be seen from Figure 6(a) and (b) that the $I(F)$ scores of mean, root mean square, crest factor and peak value are much higher than those of the rest. Therefore, mean, root mean square, crest factor and peak value are selected. Then, a CVR-based prognostic feature was constructed based on the selected candidate features.

RSL prediction

The proposed EB-GM prognostic framework was employed to predict the RSL of the machine as per the procedures described in section 'RSL prediction'. It was mentioned in section 'Introduction' that EB allows the quantification of the uncertainties of RSL in a probabilistic form. In order to demonstrate the
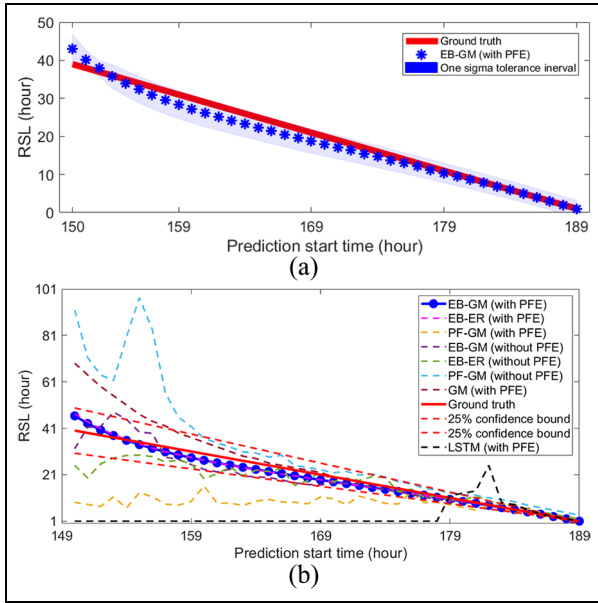
Figure 7. Predicted RSL of the pump (case study 1): (a) results obtained by empirical Bayesian–grey model with the proposed two-step PFE approach (hereafter referred to as EB-GM (with PFE)); (b) results obtained by EB-GM with PFE, EB-ER with PFE, PF-GM with PFE, EB-GM without PFE, EB-ER without PFE, PF-GM without PFE and grey model (GM) with PFE.

Figure 8. Predicted RSL of the pump (case study 2): (a) results obtained by empirical Bayesian–grey model with the proposed two-step PFE approach (hereafter referred to as EB-GM (with PFE)); (b) results obtained by EB-GM with PFE, EB-ER with PFE, PF-GM with PFE, EB-GM without PFE, EB-ER without PFE, PF-GM without PFE and grey model (GM) with PFE.
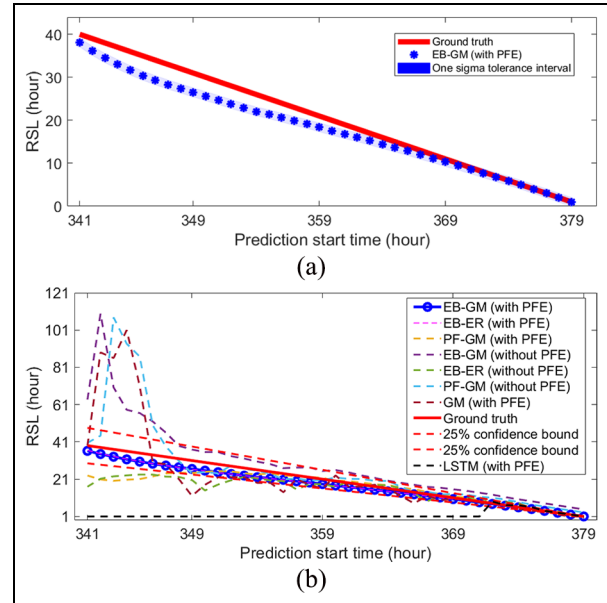
capability of EB to provide probabilistic RSL estimates, the predicted RSLs for different prediction starting points along with the associated one-sigma tolerance interval were depicted in Figures 7(a) and 8(a) for case studies 1 and 2, respectively. The blue shaded areas denote the one-sigma tolerance interval (covers around 68% of the RSL density) that was derived from the RSL densities.

To demonstrate the superiority of the developed EB-GM fault prognostic method over the existing methods, the prognostic performance of the proposed EB-GM method and its counterparts, namely EB-ER method,[15] GM in combination with PF (PF-GM), the standard GM[36] and long short-term memory (LSTM),[37] for different fault cases is shown in Figures 7(b) and 8(b). In PF, the number of particles was set to 300 as a trade-off between accuracy and computational cost. 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. Furthermore, each prognostic method mentioned above was employed to predict the RSL of the machine using two different prognostic features: (1) prognostic features calculated using the proposed two-step prognostic feature exploring (PFE) approach and (2) prognostic features calculated without using the proposed PFE approach (i.e. the

prognostic features were computed directly from all process variables without sensor selection and feature extraction using the CVR-based method as per equations (1)–(14) without performing fault identification and feature enhancement). The purpose is to investigate whether the proposed PFE approach improves the predictive accuracy. It can be seen from Figure 7(b) that, every RSL predicted by the proposed EB-GM method (with PFE) is well located within the $\mp25\%$ confidence boundaries of the true RSL, indicating that the proposed prognostic framework has the ability to accurately predict the system's RSL. A large deviation from the actual RSL was generated by PF-GM and GM (with PFE), resulting in unsatisfied estimates. Although the LSTM method performs great in the end of the lifetime, its ability in predicting RSL at early stages seems to be hindered by the lack of training data. This coincides with the conclusion made in Wu et al.[38] that the LSTM deep learning method is not suitable for long-term predictions in the presence of scarce training data. It can also be observed from Figure 7(b) that all RSL estimation results show better performance with the features obtained using the proposed two-step PFE approach than those calculated without the PFE approach. Similar conclusions can be drawn from the results for case study 2, which are shown in Figure 8(b).

Table 2. Quantitative comparison between the proposed approach and its counterparts (case study 1).<mark>[AQ: 1]</mark>

With/without the proposed PFE method	Methods	MAD	MAPD	RMSD
Predict RSL with the proposed PFE method	**EB-GM**	**2.3251**	**0.0063**	**2.8822**
	EB-ER	2.352	0.0064	2.9073
	GM-PF	3.8148	0.0104	6.5241
	GM	2.8205	0.0077	2.9439
	LSTM	19.375	0.102	23.3351
Predict RSL without the proposed PFE method	EB-GM	49.7961	0.9547	100.4655
	EB-ER	52.7166	0.9548	116.9171
	GM-PF	5.2978	0.0144	7.7092

PFE: prognostic feature exploring; MAD: mean absolute deviation; MAPD: mean absolute percentage deviation; RMSD: root mean square deviation; RSL: remaining service life; EB-GM: empirical Bayesian–grey model; EB-ER: empirical Bayesian–exponential regression; GM-PF: empirical Bayesian–particle filter; LSTM: long short-term memory.

Table 3. Quantitative comparison between the proposed approach and its counterparts (case study 2).

With/without the proposed PFE method	Methods	MAD	MAPD	RMSD
Predict RSL with the proposed PFE method	**EB-GM**	**1.6097**	**0.0085**	**2.0831**
	EB-ER	1.6866	0.0089	2.2338
	GM-PF	12.1521	0.064	15.7813
	GM	4.375	0.023	8.1071
	LSTM	17.0571	0.0449	20.5002
Predict RSL without the proposed PFE method	EB-GM	3.3157	0.0175	4.3098
	EB-ER	10.9342	0.0575	19.0614
	GM-PF	4.4525	0.0234	6.385

PFE: prognostic feature exploring; MAD: mean absolute deviation; MAPD: mean absolute percentage deviation; RMSD: root mean square deviation; RSL: remaining service life; EB-GM: empirical Bayesian–grey model; EB-ER: empirical Bayesian–exponential regression; GM-PF: empirical Bayesian–particle filter; LSTM: long short-term memory.

The quantitative evaluation results of the aforementioned algorithms are detailed in Tables 2 and 3. Three performance metrics – (1) mean absolute deviation (MAD), (2) mean absolute percentage deviation (MAPD) and (3) root mean square deviation (RMSD) – were used to quantitatively evaluate the performance of the predictive models studied. Interested readers are referred to Celaya et al.[39] for further information about the three metrics. It can be seen from both case studies that the proposed intelligent prognostic framework has the following features: (1) It can be seen from Tables 2 and 3 that the prognostic performance can be greatly improved by applying the proposed two-step prognostic feature exploring method in that all tests on the pump show better performance with the proposed PFE approach. This is because the proposed PFE method produces a prognostic feature using only the most fault-related variables and the most fault representative time-domain features, which may lead to simple and accurate prognostics. Besides, the predicted RSL becomes smoother and more stable with the proposed PFE approach (see Figures 7(b) and 8(b)). These observations intuitively reflect the effectiveness of the proposed PFE method in polishing up the performance of EB-GM and other prognostic approaches studied; (2) It can be seen that the proposed EB-GM predictor performs a little better than EB-ER in terms of MAD, MAPD and RMSD. Besides, EB-GM outperforms standard GM and PF-GM in terms of MAD, MAPD and RMSD. It can be concluded that the proposed EB-GM predictor demonstrates comparable and superior prognostic performance than its counterparts. The forecast information provided by the proposed diagnostic and prognostic framework 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.

Conclusion

In this work, an integrated diagnostic and prognostic framework was proposed for fault diagnosis and RSL prediction of nonlinear dynamic systems in the presence of scarce failure data. The developed approach was validated through in-service condition monitoring

data acquired from an industrial centrifugal pump. The main findings of the study are as follows:

- Fault diagnosis was carried out by comparing the values of the CVR-based diagnostic index with predefined thresholds. The proposed diagnostic approach can distinguish normal operational conditions from slowly developing faults incurring system anomalies leading to continuous determination of prediction start time.
- The proposed two-step PFE method can effectively identify fault root-cause variables and explore suitable prognostic features for RSL prediction. The proposed PFE method helps to improve the predictive accuracy of EB-GM and other predictors. This method also allows the optimization of preventive maintenance schedule and spare parts supply to be carried out along with RSL prediction so as to prevent catastrophic failures.
- The EB-GM approach was introduced to learn the system's degradation from very limited amount of historical failure data and to predict the RSL with high accuracy. The EB-GM predictor demonstrates comparable and superior prognostic performance than its counterparts. Besides, EB-GM allows the quantification of the uncertainties of RSL in a probabilistic form.

The proposed diagnostic framework can be used to provide site engineers with reliable and accurate diagnosis of rotating machinery, and meanwhile, the prognostic framework can assist in the subsequent production planning and decision-making process and enhance profitability by eliminating unpredicted failures.

Declaration of conflicting interests

Funding

ORCID iD

Xiaochuan Li https://orcid.org/0000-0002-5494-1266

References

1. Zhao R, Wang D, Yan R, et al. Machine health monitoring using local feature-based gated recurrent unit networks. *IEEE Trans Ind Electron* 2018; 65: 1539–1548.

2. Wook S, Lee C, Lee J, et al. Fault detection and identification of nonlinear processes based on kernel PCA. *Chemom Intell Lab Syst* 2005; 75: 55–67.

3. Fan J and Wang Y. Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. *Inf Sci* 2014; 259: 369–379.

4. Li X, Duan F, Loukopoulos P, et al. Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor. *Control Eng Pract* 2018; 72: 177–191.

5. Li W and Qin SJ. Consistent dynamic PCA based on errors-in-variables subspace identification. *J Process Control* 2001; 11: 661–678.

6. Yin S, Zhu X, Member S, et al. Improved PLS focused on key-performance- indicator-related fault diagnosis. *IEEE Trans Ind Electron* 2015; 62: 1651–1658.

7. Russell EL, Chiang LH and Braatz RD. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemom Intell Lab Syst* 2000; 51: 81–93.

8. Stefatos G and Ben Hamza A. Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Syst Appl* 2010; 37: 8606–8617.

9. Ruiz-Cárcel C, Cao Y, Mba D, et al. Statistical process monitoring of a multiphase flow facility. *Control Eng Pract* 2015; 42: 74–88.

10. Pilario KES and Cao Y. Canonical variate dissimilarity analysis for process incipient fault detection. *IEEE Trans Ind Informatics* 2018; 14: 5308–5315.

11. Cheng F, Qu L and Qiao W. Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis. *IEEE Trans Sustain Energy* 2018; 9: 157–167.

12. Loukopoulos P, Zolkiewski G, Bennett I, et al. Abrupt fault remaining useful life estimation using measurements from a reciprocating compressor valve failure. *Mech Syst Signal Process* 2019; 121: 359–372.

13. Li X, Duan F, Mba D, et al. Canonical variate analysis, probability approach and support vector regression for fault identification and failure time prediction. *J Intell Fuzzy Syst* 2018; 34: 3771–3783.

14. Liao L. Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Trans Ind Electron* 2014; 61: 2464–2472.

15. Wu J, Wu C, Cao S, et al. Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines. *IEEE Trans Ind Electron* 2018; 66: 529–539.

16. Liu K, Gebraeel NZ and Shi J. A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. *IEEE Trans Autom Sci Eng* 2013; 10: 652–664.

17. Ruiz-Cárcel C, Jaramillo VH, Mba D, et al. Combination of process and vibration data for improved condition monitoring of industrial systems working under variable operating conditions. *Mech Syst Signal Process* 2016; 66–67: 699–714.

18. Jiang B and Braatz RD. Fault detection of process correlation structure using canonical variate analysis-based correlation features. *J Process Control* 2017; 58: 131–138.

19. Jiang B, Zhu X, Huang D, et al. A combined canonical variate analysis and Fisher discriminant analysis (CVA–FDA) approach for fault diagnosis. *Comput Chem Eng* 2015; 77: 1–9.

20. Lu Q, Jiang B, Gopaluni RB, et al. Locality preserving discriminative canonical variate analysis for fault diagnosis. *Comput Chem Eng* 2018; 117: 309–319.

21. Zhang B, Zhang L and Xu J. Degradation feature selection for remaining useful life prediction of rolling element bearings. *Qual Reliab Eng* 2015; 38: 547–554.

22. Javed K, Gouriveau R and Zerhouni N. Enabling health monitoring approach based on vibration data for accurate prognostics. *IEEE Trans Ind Electron* 2015; 62: 647–656.

23. Khelif R, Chebel-morello B, Malinowski S, et al. Direct remaining useful life estimation based on support vector regression. *IEEE Trans Ind Electron* 2017; 64: 2276–2285.

24. Wei B, Xie N and Hu A. Optimal solution for novel grey polynomial prediction model. *Appl Math Model* 2018; 62: 717–727.

25. Casella G. Illustrating empirical Bayes methods. *Chemom Intell Lab Syst* 1992; 16: 107–125.

26. Flury BK and Riedwyl H. Standard distance in univariate and multivariate analysis. *Am Stat* 1986; 40: 249–251.

27. Jiang B, Huang D, Zhu X, et al. Canonical variate analysis-based contributions for fault identification. *J Process Control* 2015; 26: 17–25.

28. Zhang R, Gu F, Mansaf H, et al. Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis. *Mech Syst Signal Process* 2017; 94: 202–213.

29. Singleton RK, Strangas EG and Aviyente S. Time-frequency complexity based remaining useful life (RUL) estimation for bearing faults. In: *Proceeding of the 2013 9th IEEE international symposium on diagnostics for electric machines, power electronics and drives (SDEMPED)*, Valencia, 27–30 August 2013, pp. 600–606. New York: IEEE.

30. Liu Y, Shuai Q, Zhou S, et al. Prognosis of structural damage growth via integration of physical model prediction and Bayesian estimation. *IEEE Trans Reliab* 2017; 66: 700–711.

31. Casella G. An introduction to empirical Bayes data analysis. *Am Stat* 1985; 39: 83–87.

32. Berger JO. *Statistical decision theory and Bayesian analysis*. Durham, NC: Springer, 2013.

33. Loutas T, Eleftheroglou N, Georgoulas G, et al. Valve failure prognostics in reciprocating compressors utilizing temperature measurements, PCA-based data fusion and probabilistic algorithms. *IEEE Trans Ind Electron*. Epub ahead of print 9 July 2019. DOI: 10.1109/TIE.2019. 2926048.

34. Loukopoulos P, Pilidis P, Bennett I, et al. Abrupt fault remaining useful life estimation using measurements from a reciprocating compressor valve failure. *Mech Syst Signal Process* 2019; 121: 359–372.

35. Odiowei PEP and Yi C. Nonlinear dynamic process monitoring using canonical variate analysis and Kernel density estimations. *IEEE Trans Ind Informatics* 2010; 6: 36–45.

36. Liu S, Forrest J and Yang Y. A brief introduction to grey systems theory. *Grey Syst Theory Appl* 2012; 2: 89–104.

37. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9: 1735–1780.

38. Wu Q, Ding K and Huang B. Approach for fault prognosis using recurrent neural network. *J Intell Manuf*. Epub ahead of print 2 January 2018. DOI: 10.1007/s10845-018-1428-5.

39. Celaya J, Saxena A, Celaya J, et al. Metrics for evaluating performance of prognostic techniques. In: *Proceeding of the 2008 international conference on prognostics and health management*, Denver, CO, 6–9 October 2008, pp. 1–17. New York: IEEE.

Appendix 1 <mark>[AQ: 2]</mark>

Grey model

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

$$X^{(0)} = \left[X_{t_1}^{(0)}, X_{t_2}^{(0)}, \ldots, X_{t_k}^{(0)} \right] \tag{26}$$

where $X^{(0)}$ is a single-digit sequence. Then, $X^{(1)} = [X_{t_1}^{(1)}, X_{t_2}^{(1)}, \ldots, X_{t_k}^{(1)}]$ is called the first-order accumulative generation sequence of the sequence $X^{(0)}$, where

$$X_{t_k}^{(1)} = \sum_{i=1}^{t_k} X_j^{(0)} \tag{27}$$

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

$$Z^{(1)} = \left[Z_{t_1}^{(1)}, Z_{t_2}^{(1)}, \ldots, Z_{t_k}^{(1)} \right] \tag{28}$$

$$Z_{t_k}^{(1)} = 0.5 \left(X_{t_{k-1}}^{(1)} + X_{t_k}^{(1)} \right) \tag{29}$$

Then, the least-squares sequence estimation of the grey difference equation of GM is defined as follows

$$X_{t_k}^{(0)} + cZ_{t_k}^{(1)} = d \tag{30}$$

And the whitenization equation is as follows

$$\frac{dX_{t_k}^{(1)}}{dt_k} + cX_{t_k}^{(1)} = d \tag{31}$$

where $[c, d]^T$ is the parameter vector of GM model, which can be obtained by the least-squares estimation $[c, d]^T = (B^T B)^{-1} B^T W$, in which

$$B = \begin{bmatrix} -Z^{(1)}_{t_2} & 1 \\ -Z^{(1)}_{t_3} & 1 \\ \vdots & \vdots \\ -Z^{(1)}_{t_k} & 1 \end{bmatrix}, \quad W = \begin{bmatrix} X^{(0)}_{t_2} \\ X^{(0)}_{t_3} \\ \vdots \\ X^{(0)}_{t_k} \end{bmatrix}$$

According to equation (31), the solution of $X^{(1)}$ at time t_k is

$$X^{(1)}_{t_k} = \left(X^{(0)}_{t_1} - \frac{d}{c}\right)e^{-c(t_k-1)} + \frac{d}{c} \tag{32}$$

where $X^{(1)}_{t_1} = X^{(0)}_{t_1}$.

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

$$X^{(0)}_{t_k} = X^{(1)}_{t_k} - X^{(1)}_{t_{k-1}} = (1 - e^c)\left(X^{(0)}_{t_1} - \frac{d}{c}\right)e^{-c(t_k-1)} \tag{33}$$

Degradation features

List of extracted degradation features

Index	Degradation features	Formula
1	Mean	$Y_{mean} = \dfrac{1}{N}\sum\limits_{t_k=1}^{S} Y(t_k)$
2	Root mean square	$Y_{rms} = \sqrt{\dfrac{1}{S}\sum\limits_{t_k=1}^{S} Y^2(t_k)}$
3	Standard deviation	$Y_{SD} = \sqrt{\dfrac{1}{S-1}\sum\limits_{t_k=1}^{S}(Y(t_k) - \bar{Y})^2}$
4	Variance	$Y_{variance} = \dfrac{1}{S-1}\sum\limits_{t_k=1}^{S}(Y(t_k) - \bar{Y})^2$
5	Skewness	$Y_{skewness} = \dfrac{1}{Y_{rms}^3}\sum\limits_{t_k=1}^{S}(Y(t_k) - Y_{mean})^3$
6	Kurtosis	$Y_{kurtosis} = \dfrac{1}{Y_{rms}^4}\sum\limits_{t_k=1}^{S}(Y(t_k) - Y_{mean})^4$
7	Crest factor	$Y_{CF} = \dfrac{Y_{peak}}{Y_{rms}}$
8	Peak value	$Y_{peak} = peak(Y(t_k)), \; t_k = 1, 2, \ldots, S$

S is the number of samples.