RESEARCH ARTICLE

Remaining service life prediction based on gray model and empirical Bayesian with applications to compressors and pumps

Xiaochuan Li[1] | David Mba[1] | Edmund Okoroigwe[2] | Tianran Lin[3]

[1] Faculty of Computer, Engineering and Media, De Montfort University, Leicester, UK

[2] Department of Mechanical Engineering, University of Nigeria, Nsukka, Nigeria

[3] School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao, China

Correspondence
Xiaochuan Li, Faculty of Computer, Engineering and Media, De Montfort University, Leicester, LE1 9BH, UK.
Email: xiaochuan.li@dmu.ac.uk

Abstract

In this study, a three-step remaining service life (RSL) prediction method, which involves feature extraction, feature selection, and fusion and prognostics, is proposed for large-scale rotating machinery in the presence of scarce failure data. In the feature extraction step, eight time-domain degradation features are extracted from the faulty variables. A fitness function as a weighted linear combination of the monotonicity, robustness, correlation, and trendability metrics is defined and used to evaluate the suitability of the features for RSL prediction. The selected features are merged using a canonical variate residuals-based method. In the prognostic step, gray model is used in combination with empirical Bayesian algorithm for RSL prediction in the presence of scarce failure data. The proposed approach is validated on failure data collected from an operational industrial centrifugal pump and a compressor.

KEYWORDS
prognostics, remaining useful life, rotating machinery, scarce failure data

1 | INTRODUCTION

Prognostic feature construction plays an important role in machinery prognostics. A suitable prognostic feature is capable of facilitating machinery prognostics by ensuring accurate estimates and simplifying the prognostic modeling. The prognostic feature can be data from a single sensor,[1] a feature extracted from raw sensory signals using dimension reduction techniques,[2,3] or the integration of multiple features extracted from multiple sensory signals.[4–6] In this work, the prognostic feature is generated by merging multiple features extracted from all the fault-related variables. It is crucial to select suitable degradation features before subsequent prognostic analysis is carried out because the predictive accuracy and the complexity of prognostic modeling are largely affected by the performance of the extracted degradation features. Various metrics are available in the literature for the evaluation of degradation features, including monotonicity,[6] robustness,[7] correlation,[5] and so on. Considering that a component is more likely to degrade with the increase of the service time, a metric named trendability was developed later in Ref. 8 to measure the correlation between the degradation feature and time. Given that a single criterion may not be sufficient for prognostic feature selection, a hybrid criterion, which considers

the properties of trendability, monotonicity, robustness, and correlation, is proposed in this work for selecting degradation features. After feature selection, a feature-fusion technique is required to fuse the selected features into a single prognostic feature. In this paper, we explore the capability of canonical variate analysis (CVA) for representing the evolving trend and calculate the canonical variate residual (CVR) based on the selected time-domain features as a prognostic feature.

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 condition monitoring involve a prognostic step where long-term predictions of continuous observations are carried out with the aim of estimating the remaining service life (RSL) of the system. Various data-driven techniques are available for this task, including self-organizing map,[2] K nearest neighbors,[2] support vector machine,[9] and neural networks.[10] Those techniques offer a tradeoff between reliability, speed, and applicability. Prediction models usually require large amounts of historical 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 fulfill prognostic tasks, since safety-critical equipment are usually not allowed to run to failure. In this study, gray model[11] is used in combination with empirical Bayesian (EB) algorithm to realize accurate RSL prediction in the presence of scarce failure data.

The proposed prognostic framework comprises the following three main steps:

- Feature extraction: eight time-domain degradation features, namely, mean, root mean square, standard deviation, variance, skewness, kurtosis, crest factor, and peak value are extracted from the fault-related variables.
- Feature selection and fusion: The suitability of the extracted features for RSL prediction is evaluated based on four performance metrics, namely, monotonicity, robustness, correlation, and trendability. Then, the optimal features that are suitable for RSL estimation are fused into a prognostic feature using the CVA-based method.
- Prognostics: gray model is employed to generate an observation of the local RSL. Then, EB is applied to estimate the global RSL based on the local RSL estimates.

The contribution of this paper is summarized as follows:

- A feature extraction-fusion framework, which is able to find features that are suitable for prognostics from the raw measurements, is proposed. This framework consists of three main steps, namely, time-domain feature extraction, feature selection, and feature fusion.
- A gray-EB prognostic framework, which is able to continuously correct RUL predictions based on past information while providing probabilistic output, is proposed.

2 | METHODOLOGY

Figure 1 illustrates a flowchart of the implementation process of the proposed prognostic framework.

2.1 | Feature extraction

Time-domain analysis is carried out to extract mean, root mean square, standard deviation, variance, skewness, kurtosis, crest factor, and peak value from the fault-related variables. Table 1 summarizes the time-domain features that are utilized in this study. The reason why only time domain features are considered in this paper is that the experimental data were collected from real-world operational machines. These machines were mounted with a variety of sensors when set up and most of them capture signals at a low sampling frequency. As a result, frequency/time-frequency domain features that were extracted form fast signals, such as vibration and acoustic emission, were not included in this study. The selected eight features were popular features used in the literature, and half of them have been experimentally verified to be effective for representing the fault evolution.
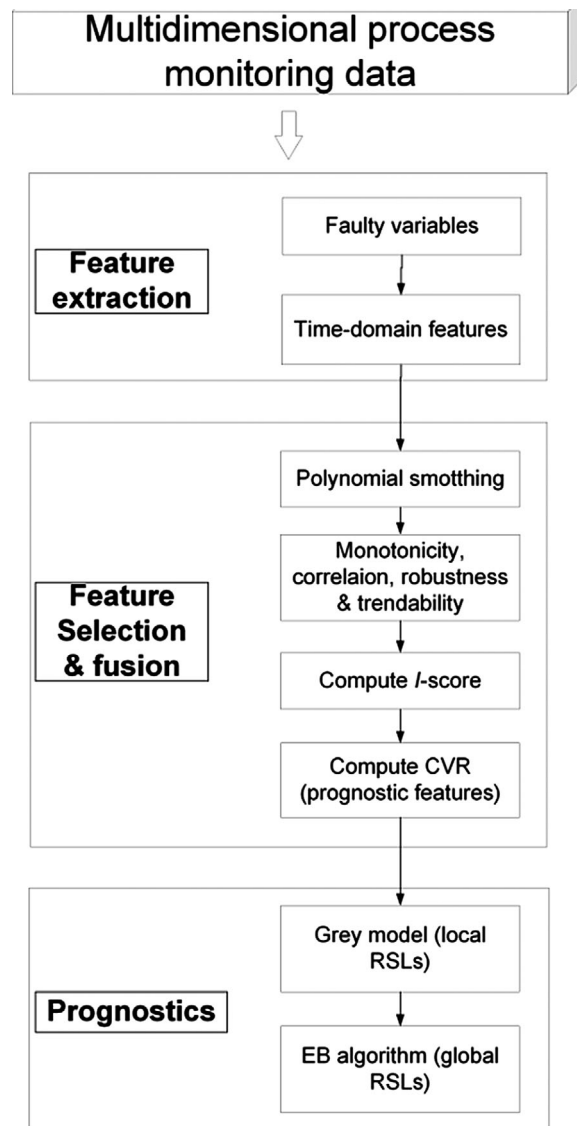
FIGURE 1 Proposed three-step prognostic framework for the RSL prediction of rotating machinery

TABLE 1 Degradation features

Index	Degradation features	Formula
1	Mean	$Y_{mean} = \frac{1}{D} \sum_{j=1}^{D} Y(j)$
2	Root mean square	$Y_{rms} = \sqrt{\frac{1}{D} \sum_{j=1}^{D} Y^2(j)}$
3	Standard deviation	$Y_{SD} = \sqrt{\frac{1}{D-1} \sum_{j=1}^{D} (Y(j) - \bar{Y})^2}$
4	Variance	$Y_{variance} = \frac{1}{D-1} \sum_{j=1}^{D} (Y(j) - \bar{Y})^2$
5	Skewness	$Y_{skewness} = \frac{1}{Y_{rms}^3} \sum_{j=1}^{D} (Y(j) - Y_{mean})^3$
6	Kurtosis	$Y_{kurtosis} = \frac{1}{Y_{rms}^4} \sum_{j=1}^{D} (Y(j) - Y_{mean})^4$
7	Crest factor	$Y_{CF} = \frac{Y_{peak}}{Y_{rms}}$
8	Peak value	$Y_{peak} = peak(Y(j)), j = 1, 2 \dots D$

D is the number of samples.

2.2 | Feature selection and fusion

The main aim of this section is to find a subset of suitable features, which can represent the fault evolution, and to fuse the subset of features into a single-dimensional prognostic feature. Given that one criterion is not sufficient to select a suitable degradation feature for the task of RSL estimation, a hybrid metric which can evaluate the suitability of features from different aspects is proposed in this study. The proposed metric is a weighted sum of the criteria, trendability, monotonicity, robustness, and correlation. The feature construction procedure comprises the following four main steps:

- A polynomial function is employed to fit a trajectory of each of the extracted feature to decompose the signal into a smooth trend and a noise component as follows:

$$X(t_k) = X_T(t_k) + X_N(t_k) \tag{1}$$

where $X(t_k)$ is the extracted feature, $X_T(t_k)$ is the smooth trend, and $X_N(t_k)$ is the noise component.

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

$$T(X) = \frac{K\left(\sum_{k=1}^{K} \tilde{X}_T(t_k)\tilde{t}_k\right) - \left(\sum_{k=1}^{K} \tilde{X}_T(t_k)\right)\left(\sum_{k=1}^{K} \tilde{t}_k\right)}{\sqrt{\left[K\sum_{k=1}^{K} \tilde{X}_T^2(t_k) - \left(\sum_{k=1}^{K} \tilde{X}_T(t_k)\right)^2\right]\left[K\sum_{k=1}^{K} \tilde{t}_k^2 - \left(\sum_{k=1}^{K} \tilde{t}_k\right)^2\right]}} \tag{2}$$

$$M(X) = \frac{1}{K-1}\left|\sum_{k=1}^{K} \delta(X_T(t_{k+1}) - X_T(t_k)) - \sum_{k=1}^{K} \delta(X_T(t_k) - X_T(t_{k+1}))\right| \tag{3}$$

$$R(X) = \frac{1}{K}\sum_{k=1}^{K} \exp\left(-\left|\frac{X_N(t_k)}{X(t_k)}\right|\right) \tag{4}$$

$$C(X) = \frac{\left|K\left(\sum_{k=1}^{K} X_T(t_k)t_k\right) - K\sum_{k=1}^{K} X_T(t_k)\sum_{k=1}^{K} t_k\right|}{\sqrt{\left[K\sum_{k=1}^{K} X_T(t_k)^2 - \left(\sum_{k=1}^{K} X_T(t_k)\right)^2\right]\left[K\sum_{k=1}^{K} t_k^2 - \left(\sum_{k=1}^{K} t_k\right)^2\right]}} \tag{5}$$

where $\tilde{X}_T(t_k)$ and $\tilde{t}_k$ are the rank sequences of the feature $X_T(t_k)$ and t_k, respectively.

- With the metrics of trendability, monotonicity, robustness, and correlation, feature selection is achieved by computing a weighted sum of these criteria. The optimal weighted coefficients $w_j, j = 1, 2, 3, 4$ are obtained by maximizing the objective function as follows:

$$Z(X) = w_1 T(X) + w_2 M(X) + w_3 R(X) + w_4 C(X)$$

$$\text{s.t., } \sum_{j=1}^{4} w_j = 1, w_j > 0 \tag{6}$$

where Z is the objective function to be optimized, and $w_j, j = 1, 2, 3, 4$ are the weighted coefficients. The features with a high Z value should be selected for RSL prediction.

- A dimension reduction technique called canonical variate analysis (CVA) is utilized to fuse the selected degradation features into a one-dimensional prognostic feature. Interested readers are referred to Refs. 3 and 10 for further information about CVA. The detailed feature fusion procedure is given as follows.

Given the original measurements $y_t \in \mathcal{R}^n$ (where n is the number of process variables), the data were expanded by including a and b number of past and future samples, to generate the past and future vectors $y_{a,t} \in \mathcal{R}^{na}$ and $y_{b,t} \in \mathcal{R}^{nb}$

$$y_{a,t} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-a} \end{bmatrix} \in \mathcal{R}^{na}, y_{b,t} = \begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+b-1} \end{bmatrix} \in \mathcal{R}^{nb} \tag{7}$$

$y_{a,t}$ and $y_{b,t}$ are then normalized to generate the zero-mean vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$. Then, the rearranged data vectors $\hat{Y}_a$ and $\hat{Y}_b$ are generated as follows:

$$\hat{Y}_a = [\hat{y}_{a,t+1}, \hat{y}_{a,t+2}, \dots, \hat{y}_{a,t+N}], \quad \hat{Y}_b = [\hat{y}_{b,t+1}, \hat{y}_{b,t+2}, \dots, \hat{y}_{b,t+N}] \tag{8}$$

where $N = Q - a - b + 1$ and Q denotes the length of y_t. CVA searches for vectors J and L such that the maximum correlation between $L\hat{y}_{b,t}$ and $J\hat{y}_{a,t}$ is obtained. The solution to this problem can be achieved by performing singular value decomposition on the matrix $\mathcal{H}$

$$\mathcal{H} = \sum_{b,b}^{-1/2} \sum_{b,a} \sum_{a,a}^{-1/2} = U \sum V^T \tag{9}$$

where $\sum_{a,a}$ and $\sum_{b,b}$ and $\sum_{a,b}$ are the covariance and cross-covariance matrices of $\hat{Y}_b$ and $\hat{Y}_a$. The canonical correlation residuals r_t can be defined as:

$$r_t = L_q^T \hat{y}_{b,t} - \sum_q J_q^T \hat{y}_{a,t} \tag{10}$$

where L_q^T denotes the first q rows of matrix L^T, and $L_q^T = U_q^T \sum_{b,b}^{-1/2}$. Similarly, J_q^T is the first q rows of matrix J^T, and $J_q^T = V_q^T \sum_{a,a}^{-1/2}$. $\sum_q = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$ is a diagonal matrix with its diagonal elements being the first q components of $\sum$. The prognostic feature is obtained as:

$$T_d = [(r_t)^T S^{-1}(r_t)]^{1/2} \tag{11}$$

where $S = I - \sum \sum^T$.

These features were merged to form a one-dimensional HI, which will be further analyzed using gray model and EB. If the features were not merged, the computational load will increase at least by eight times given that a gray model is needed for each feature at every time instance. Moreover, the resultant RUL will involve a large confidence boundary given that each feature represents the fault evolution in their own characteristic way. Merging features together has been adopted extensively by researchers[4,12] and has been proven to be effective.

2.3 | Prognostics

2.3.1 | Local RSL prediction

In this section, gray model is first utilized to propagate the trend of the prognostic feature until it reaches the predetermined threshold, providing an observation of the local RSL at time t_k. Then, EB is used to generate an estimate of the global RSL based on the obtained local RSL estimates obtained at different times. Since a single gray model may not be able to accurately predict the actual degradation trajectory of the machine, several successive prognostic features sorted in descending order of length are regarded as a training sequence in order to train several gray models so as to generate several local RSL estimates. In this way, a sequence of fitted local degradation features $\hat{X}_{j:t_k}, j = t_{k-s+1}, t_{k-s+2}, \dots, t_{k-n+1}$ with different lengths are obtained at every time instance. Each degradation feature is used to train a unique gray model, and a sequence of trained gray models are generated and utilized to predict the future values of the degradation feature
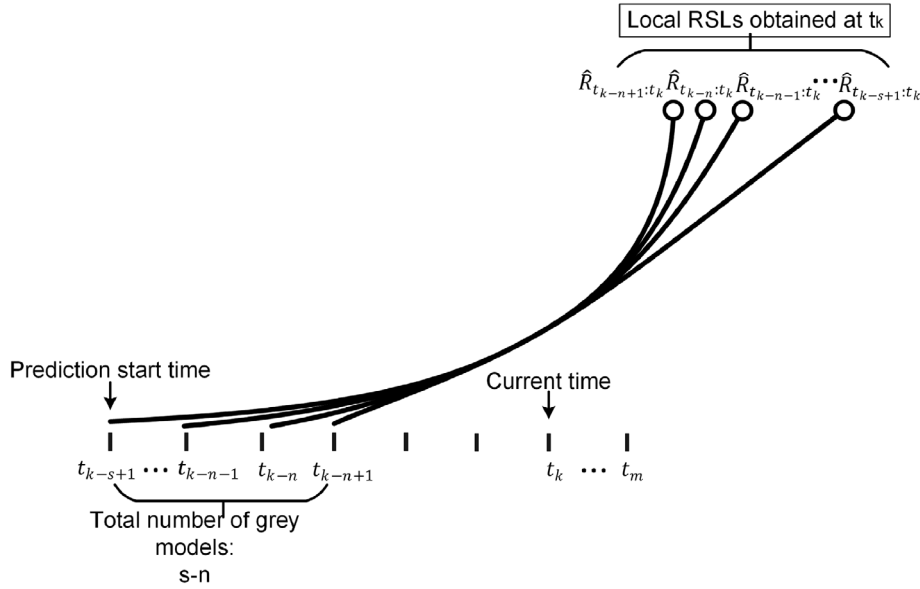
FIGURE 2   Illustration of how local RSLs are generated

until it reaches a predefined threshold. It is worth noting that these gray models are trained using data of different lengths, meaning that they cover a variety of fix-order Markov models. A fixed-order Markov model is commonly used to describe the fault evolution.[13] Here, instead of using a fixed-order model, a sequence of fixed-order Markov models is used since the fault progression may depend on the previous few steps.

As shown in Figure 2, at time t_k, the training sequence $\hat{X}_{j:t_k}$, $j = t_{k-s+1}, t_{k-s+2}, \dots, t_{k-n+1}$ hits the failure threshold at different times, yielding the local RSL estimates $\hat{R}_{t_{k-n+1}:t_k}, \dots, \hat{R}_{t_{k-s+1}:t_k}$. $\hat{R}_{t_{k-n+1}:t_k}$ denotes the local RSL estimated using the training feature $\hat{X}_{j:t_k}$, $j = t_{k-n+1}$. For predictions start at timet_k, t_{k-s+1} is the prediction start time, which can be determined by the machine's fault detection system. n is the minimum sample size required for generating a gray model ($n = 4$).

Fitting and prediction errors are calculated so as to evaluate the suitability of each local fitted trajectory for representing the fault progression:

$$\left\{ \begin{array}{l} e_f(j) = \dfrac{1}{t_k - j + 1} \displaystyle\sum_{i=j}^{t_k} \left(\hat{X}_{j:t_k}(i) - X_{j:t_k}(i) \right)^2, j = t_{k-s+1}, t_{k-s+2}, \dots, t_{k-n+1} \\ e_p(j) = \left(\hat{X}_{j:t_k}(t_{k+1}) - X_{j:t_k}(t_{k+1}) \right)^2, j = t_{k-s+1}, t_{k-s+2}, \dots, t_{k-n+1} \end{array} \right) \tag{12}$$

where $e_f(j)$ and $e_p(j)$ indicate the fitting and prediction error of the fitted feature $\hat{X}_{j:t_k}$, respectively. $\hat{X}_{j:t_k}(i)$ is the value of the fitted curve $\hat{X}_{j:t_k}$ at time i, and $\hat{X}_{j:t_k}(t_{k+1})$ is the predicted value of the degradation curve at time t_{k+1}. The fitting and prediction errors are utilized to calculate a weight coefficient for each of the local fitted feature as follows:

$$w_j = \frac{1}{e_f(j) + e_p(j)} \bigg/ \sum_{i=t_{k-s+1}}^{t_{k-n+1}} \frac{1}{e_f(i) + e_p(i)}, \quad j = t_{k-s+1}, \quad t_{k-s+2}, \dots, t_{k-n+1} \tag{13}$$

where w_j denotes the weight coefficient of the fitted degradation feature $\hat{X}_{j:t_k}$. Clearly, the weight coefficients are calculated based on the fitting and one-step prediction errors. The fitted gray models that can better represent the degradation trajectory will be given a higher weight, and vice versa. For details of gray model, authors are referred to Appendix 1. The predicted local RSL $\hat{R}_{t_k}$ at time t_k and the associated variance $\sigma_{t_k}^2$ are calculated as follows:

$$\left\{ \begin{array}{l} \hat{R}_{t_k} = \sum_{j=t_{k-s+1}}^{t_{k-n+1}} w_j \hat{R}_{j:t_k} \\ \sigma_{t_k}^2 = \frac{1}{s-n+1} \sum_{j=t_{k-s+1}}^{t_{k-n+1}} w_j \left(\hat{R}_{j:t_k} - \hat{R}_{t_k} \right)^2 \end{array} \right) \tag{14}$$

2.3.2 | Global RSL prediction

This section discusses how the prior and posterior distributions of the global RSLs are estimated from the observed local RSLs. The degradation patterns of mechanical systems at different moments are diverse and the local failure time varies with progressive degradation. However, it would be rational for most applications that the extreme degradation patterns are rare, while the moderate ones are frequent (this paper is based on this assumption). Then, suppose the local RSLs follow a common prior distribution $\pi(\theta_{t_k}) \sim N(\mu, \sigma_z^2)$, where μ and σ_z^2 are unknown hyper parameters. Then, this case becomes an exchangeable EB problem.[14]

Recall that $\hat{R}_{t_k}$ is the observation of θ_{t_k} and $\sigma_{t_k}^2$ measures the error of each observation. In order to account for the impact of stochastic dynamics, the sample variance is set to the mean of $\sigma_{t_k}^2$ and is calculated as:

$$\sigma^2 = \frac{\sum_{i=t_{start}}^{t_m} \sigma_i^2}{t_m - t_{start} + 1} \tag{15}$$

where t_{start} is the prediction start time and t_m is the prediction end time. According to Refs. 14 and 15, the joint distribution of $(\hat{R}_{t_k}, \theta_{t_k})$ is computed as:

$$p(\hat{R}_{t_k}, \theta_{t_k}) = \frac{1}{2\pi\sigma\sigma_z} \exp\left\{ -\frac{\rho}{2} \left[\theta_{t_k} - \frac{1}{\rho} \left(\frac{\mu}{\sigma_z^2} + \frac{\hat{R}_{t_k}}{\sigma^2} \right) \right]^2 \right\} \exp\left\{ -\frac{(\mu - \hat{R}_{t_k})^2}{2(\sigma_z^2 + \sigma^2)} \right\} \tag{16}$$

where $\rho = \sigma_z^{-2} + \sigma^{-2}$.

In order to estimate the prior distribution of the unknown hyper parameters μ and τ^2, we need to first deduce the marginal distribution of $\hat{R}_{t_k}$

$$p\left(\hat{R}_{t_k}\right) = \int_{-\infty}^{+\infty} p\left(\hat{R}_{t_k}|\theta_{t_k}\right) p\left(\theta_{t_k}\right) d\theta_{t_k} = \frac{1}{\sqrt{2\pi\rho}\sigma\sigma_z} \exp\left\{ -\frac{\left(\hat{R}_{t_k} - \mu\right)^2}{2(\sigma_z^2 + \sigma^2)} \right\} \tag{17}$$

Apparently, $p(\hat{R}_{t_k})$ is a density of $N(\mu, (\sigma^2 + \tau^2))$. The likelihood function of all local RSLs is computed as:

$$\prod m(\hat{R}_j) = [2\pi(\sigma^2 + \tau^2)]^{-p/2} \exp\left\{ \frac{-ps^2}{2(\sigma^2 + \tau^2)} \right\} \exp\left\{ \frac{-p(\hat{\mu}_k - \mu)^2}{2(\sigma^2 + \tau^2)} \right\} \tag{18}$$

where $\hat{\mu}_k = \frac{\sum_{j=t_{start}}^{t_m} \hat{R}_j}{p}, s^2 = \frac{\sum_{j=t_{start}}^{t_m} (\hat{R}_j - \hat{\mu}_k)^2}{p}$. p is the total number of RSL estimations.

According to 14, the ML-II estimates of the unknown parameters μ and σ_z^2 are:

$$\begin{cases} \hat{\mu} = \hat{\mu}_k = \frac{\sum_{j=t_{start}}^{t_m} \hat{R}_j}{p} \\ \hat{\tau}^2 = \max\left(0, \frac{1}{p}s^2 - \sigma^2\right) \end{cases} \tag{19}$$

After the determination of the prior distribution, the posterior distribution $p(\theta_{t_k}|\hat{R}_{t_k})$ can be estimated. Berger[14] suggested the following modified posterior distribution $N(\mu^{EB}(\hat{R}_{t_k}), V^{EB}(\hat{R}_{t_k}))$, which is suitable for the cases when the number of local RSL observations is moderate or small.

$$\begin{cases} \mu^{EB}\left(\hat{R}_{t_k}\right) = \hat{R}_{t_k} - \hat{B}\left(\hat{R}_{t_k} - \hat{\mu}_k\right) \\ V^{EB}\left(\hat{R}_{t_k}\right) = \sigma^2\left(1 - \frac{p-1}{p}\hat{B}\right) + \frac{2}{p-3}\hat{B}^2\left(\hat{R}_{t_k} - \hat{\mu}_k\right)^2 \end{cases} \tag{20}$$

where $\hat{B} = \left(\frac{p-3}{p-1}\right)\frac{\sigma^2}{\sigma^2 + \hat{\tau}^2}, \hat{\tau}^2 = \frac{\sum_{j=t_k-s+4}^{t_m} (\hat{R}_j - \hat{\mu}_k)^2}{p-1} - \sigma^2$.
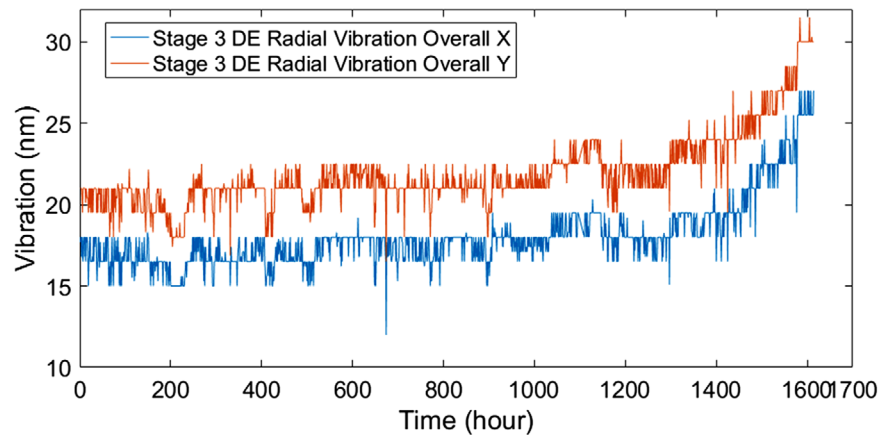
TABLE 2 Measured variables of pump A

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

FIGURE 3 Faulty variables of pump A (signals are normalized)
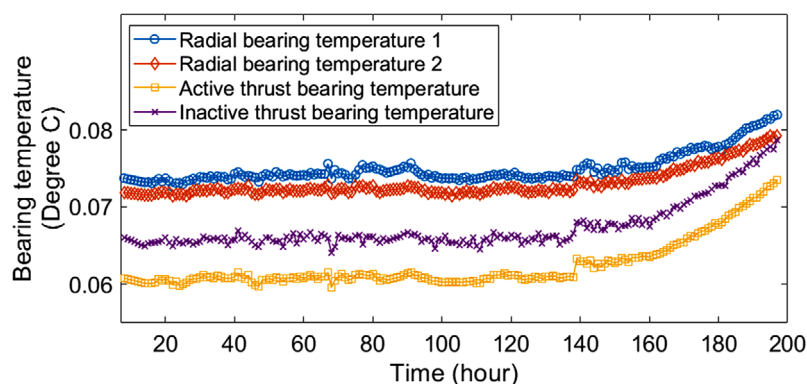
3 | RESULTS AND DISCUSSTION

The results of the proposed method for rotating machinery RSL prognosis are presented in this section, using the real-world condition monitoring data captured from an operational pump and a compressor. The measured time series from the pump consisted of 13 variables (Table 2 shows all measured variables for this pump), and the measured time series from the compressor consisted of 21 variables (Table 2). For this study, all data were captured at a sampling rate of one sample per hour. For the pump case, it is observable 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. For the compressor case, the readings from stage 3 radial vibration sensors showed an increase trend, which had led to machine trip (Figure 4). Information employed in this work came from compressors and pumps that have been used in a European refiner. The machines are instrumented with sensors collecting process-related measurements, that stream continuously, via internet, to a central location. They are stored, preprocessed, and analyzed for CBM purposes. Considering each sensor's sampling frequency, a large volume of data is created and thus a huge amount of storage is required. To mitigate this issue, a rule set was created deciding which values should be stored, creating nonuniformly sampled sets. Therefore, the data were collected from the data monitoring center via internet. The sampling frequency is 1 h per sample. The signals being used for the analysis are listed in Table 2 for pump A and Table 3 for compressor A, respectively.

FIGURE 4 Faulty variables of compressor A

TABLE 3 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		

3.1 | Feature extraction, selection, and fusion

As discussed in Section 2, eight time-domain features are extracted from the faulty variables. Feature selection is implemented to select the most representative features, thereby eliminating fault-irrelative variables. The accumulated z-score $Z(X)$ of the extracted features over all faulty variables is calculated and depicted in Figures 5 and 6, respectively. In order to select the most representative features, only the features with an average z-score higher than 0.9 are chosen for subsequent analysis. As a result, mean, root mean square, crest factor, and peak value are selected as the most representative features. It can also be seen from both figures that the z-scores of mean, root mean square, crest factor, and peak value are much higher than those of the rest. Then, CVR-based prognostic features were constructed based on the selected candidate features as per Equations (7)-(11).

Figures 7 and 8 show the prediction results for the pump and compressor cases obtained by the proposed prognostic method for different prediction starting points, respectively. It is observable from the figures that the predictive accuracy

TABLE 4 Prediction errors

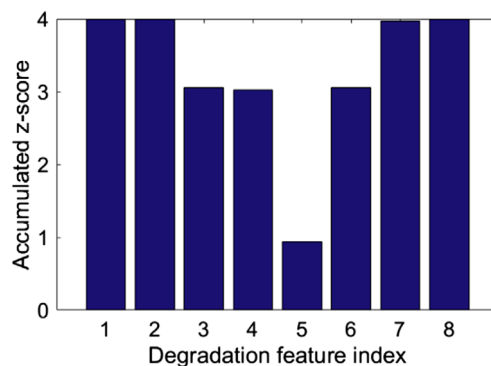|		MAD	MAPD	RMSD
Pump	Proposed	2.35	0.0064	2.91
	Gray model	2.82	0.077	2.94
	AR model	4.31	0.092	7.37
Compressor	Proposed	4.83	0.03	6.32
	Gray model	5.42	0.034	6.72
	AR model	5.65	0.057	6.87

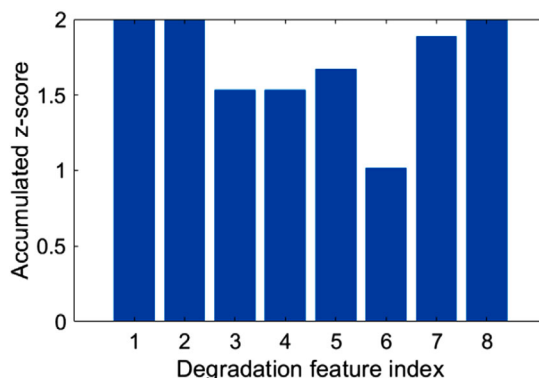**FIGURE 5**   Z-score of the extracted features (pump A)

FIGURE 6   Z-score of the extracted features (compressor A)

is lower at the beginning. This is because the data collected in a relatively early degradation stage may not be able to fully capture the fault evolution process for RSL prediction purposes. As the prediction starting point gets closer to the end-of-life, the predicted RSL gets closer to the true remaining useful life. Furthermore, it can be seen that the proposed prognostic framework has the ability to accurately predict the system's RSL since every predicted RSL is located well within the ∓25% confidence boundaries of the actual RSL.

Three performance metrics—(1) mean absolute deviation (MAD); (2) mean absolute percentage deviation (MAPD); and (3) root mean square deviation (RMSD)—were used to evaluate the performance of the proposed approach. These performance metrics are frequently utilized in reliability analysis.[16,17] Interested readers are referred to Ref. 18 for further information about the three metrics. Table 4 summarizes the precision analysis for the prediction error of the proposed prognostic model. The proposed method was also compared with two different predictors, namely, gray model and AR model.[19] The comparison results show that the proposed method outperforms the other predictors. We compared our method with the traditional gray model to show the superiority of using EB on gray model. Gray model is often compared with mathematical models, such as linear regression, exponential regression, and AR models.[20] AR models are one of the most popular linear models in time series forecasting, which have been widely applied during the past decade.

FIGURE 7   RSL prediction performance of the proposed approach (pump A). Red: actual RSL. Red dashed: ∓25% confidence boundaries
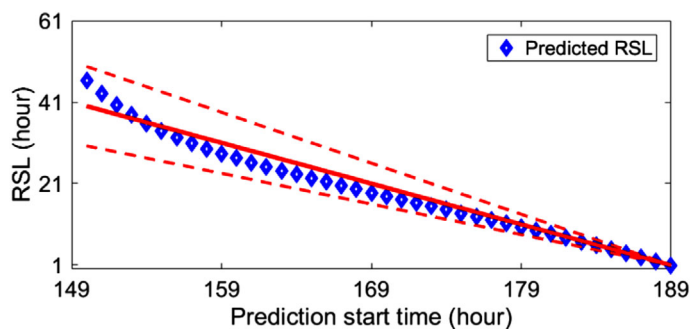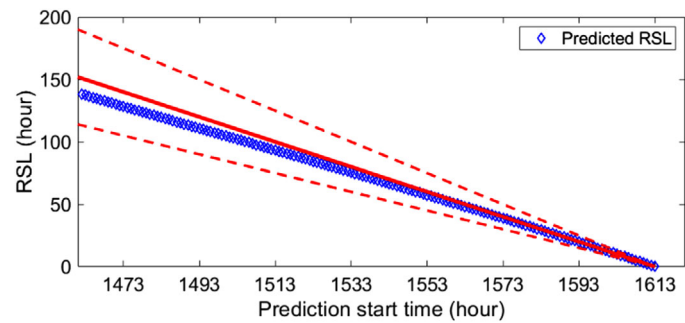
FIGURE 8 RSL prediction performance of the proposed approach (compressor A). Red: actual RSL. Red dashed: ∓25% confidence boundaries

4 | CONCLUSION

In this work, a novel prognostic framework was proposed for RSL prediction of nonlinear dynamic systems in the presence of scarce failure data. The developed approach was validated through condition monitoring data acquired from an industrial centrifugal pump and a compressor. The proposed prognostic feature exploring method can effectively find the most representative degradation features and generate suitable prognostic features for RSL prediction. The combination of EB and gray model was proposed to learn the system's degradation from very limited amount of historical failure data and to predict the RSL with high accuracy. The proposed approach can be used to provide site engineers with reliable and accurate estimates of RSL, thereby facilitating subsequent production planning and decision-making process.

ORCID

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

REFERENCES

1. Cheng F, Qu L, Qiao W. Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis. *IEEE Trans Sustain Energy*. 2018;9(1):157-167.
2. Loukopoulos P, Zolkiewski G, Bennett I, et al. Abrupt fault remaining useful life estimation using measurements from a reciprocating compressor valve failure. *Mech Syst Sig Process*. 2019;121:359-372. https://doi.org/10.1016/j.ymssp.2018.09.033
3. Li X, Duan F, Mba D, Bennett I. Canonical variate analysis, probability approach and support vector regression for fault identification and failure time prediction. *J Intell Fuzzy Syst*. 2018;34:3771-3783. https://doi.org/10.1109/SDPC.2017.146
4. Liao L. Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Trans Ind Electron*. 2014;61(5):2464-2472.
5. Wu J, Wu C, Cao S, Or SW, Deng C, Shao X. Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines. *IEEE Trans Ind Electron*. 2018;66(1):529-539. https://doi.org/10.1109/TIE.2018.2811366
6. Liu K, Gebraeel NZ, 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(3):652-664. https://doi.org/10.1109/TASE.2013.2250282
7. Zhang B, Zhang L, Xu J. Degradation feature selection for remaining useful life prediction of rolling element bearings. *Qual Reliab Eng*. 2015;32:547-554. https://doi.org/10.1002/qre.1771
8. Javed K, Gouriveau R, Zerhouni N. Enabling health monitoring approach based on vibration data for accurate prognostics. *IEEE Trans Ind Electron*. 2015;62(1):647-656. https://doi.org/10.1109/TIE.2014.2327917
9. Khelif R, Chebel-morello B, Malinowski S, Laajili E. Direct remaining useful life estimation based on support vector regression. *IEEE Trans Ind Electron*. 2017;64(3):2276-2285. https://doi.org/10.1109/TIE.2016.2623260
10. Li X, Duan F, Loukopoulos P, Bennett I, Mba D. 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. https://doi.org/10.1016/j.conengprac.2017.12.006
11. Liu S, Yang Y, Forrest J. *Grey Data Analysis*. Singapore: Springer; 2017.
12. Loutas T, Eleftheroglou N, Georgoulas G, Loukopoulos P, Mba D, Bennett I. Valve failure prognostics in reciprocating compressors utilizing temperature measurements, PCA-based data fusion and probabilistic algorithms. *IEEE Trans Ind Electron*. 2019;67:5022-5029. https://doi.org/10.1109/TIE.2019.2926048
13. Chen C, Zhang B, Vachtsevanos G, Orchard M. Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering. *IEEE Trans Ind Electron*. 2011;58:4353-4364. https://doi.org/10.1109/TIE.2010.2098369
14. Berger JO. *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media; 2013.
15. Casella G. An introduction to empirical Bayes data analysis. *Am Statist*. 1985;39(2):83-87. https://doi.org/10.2307/2682801
16. 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(2):1539-1548.

17. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. *Mech Syst Sig Process*. 2019;115:213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

18. Celaya J, Saxena A, Celaya J, Balaban E. Metrics for evaluating performance of prognostic techniques. *International Conference on Prognostics and Health Management, 2008*. 2008. p. 1–17. https://doi.org/10.1109/PHM.2008.4711436

19. Engle RF. Autoregressive conditional heteroscedacity with estimates of variance of United Kingdom inflation. *Econometrica*. 1982;50(4):987-1008. https://doi.org/10.2307/1912773

20. Yang X, Fang Z, Yang Y, Mba D, Li X. A novel multi-information fusion grey model and its application in wear trend prediction of wind turbines. *Appl Math Model*. 2019;71:543-557. https://doi.org/10.1016/j.apm.2019.02.043

AUTHOR BIOGRAPHIES

Xiaochuan Li: Xiaochuan Li is a lecturer at De Montfort University, Leicester, United Kingdom. Xiaochuan Li's current research is focused on condition monitoring, multivariate fault detection and diagnosis, and prognostics.

David Mba: Professor David Mba is the Pro Vice-Chancellor Enterprise and Dean of Computing, Engineering and Media of De Montfort University, Leicester, United Kingdom. David's current research is focused on machine condition monitoring, machine fault diagnosis, model-based prognostics, and machine performance prediction. He has published over 250 journals and conference papers.

Edmund Okoroigwe: Edmund Okoroigwe is with the Department of Mechanical Engineering, University of Nigeria, Nsukka, Nigeria. His research interest lies in solar power technologies, thermal grid electricity, and so on.

Tianran Lin received the B. Eng. degree in mechanical engineering from Jiangnan University, Wuxi, China, in 1987, and the M. Eng. and Ph.D degrees in mechanical engineering from the University of Western Australia, Perth, Australia, in 2001 and 2006, respectively. He is currently a "Taishan Scholar" distinguished Professor in the School of Mechanical and Automatic Engineering at Qingdao University of Technology, Qingdao, China. His research interests include noise and vibration analysis and control, signal processing, machine fault diagnosis, and artificial intelligent. Pro. Lin is a fellow of the International Society of Engineering Asset Management.

APPENDIX 1

Gray model is the basic model of gray theory and has been used widely since its development in the early 1980s. Gray 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 behavior and its evolution law, and thus generates quantitative predictions of future system changes. By updating the modeling data and introducing new information, gray model can reflect the characteristics of the current situation. Gray forecasting model is suitable for real-time prediction with limited availability of degradation data. Gray model uses operations of accumulated generations to build different equations. The general procedure for gray model is described as follows.

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

$$X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)) \tag{A.1}$$

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

$$X^{(1)}(k) = \sum_{i=1}^{k} X^{(0)}(i) k = 1, 2, \dots, n \tag{A.2}$$

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

$$Z^{(1)} = (Z^{(1)}(2), Z^{(1)}(3), \dots, Z^{(1)}(n)) \tag{A.3}$$

$$Z^{(1)}(k) = 0.5(x^{(1)}(k-1) + x^{(1)}(k)), k = 2, 3, \dots, n \tag{A.4}$$

Then, the least square sequence estimation of the gray difference equation of gray model is defined as follows:

$$x^{(0)}(k) + az(k) = b \tag{A.5}$$

And the whitenization equation is as follows:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{A.6}$$

where $[a, b]^T$ is the parameter vector of gray model, which can be obtained by the least square estimation $[a, b]^T = (B^T B)^{-1} B^T Y$, in which

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

According to Equation (13), the solution of $x^{(1)}$ at time k is:

$$\hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-a(k-1)} + \frac{b}{a}, k = 1, 2, \dots, n \tag{A.7}$$

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

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

$$\hat{x}^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1) = (1 - e^a)\left(x^{(0)}(1) - \frac{b}{a}\right) e^{-a(k-1)} \tag{A.8}$$