Zhang, Y., Hutchinson, P., Lieven, N. A. J., & Nunez-Yanez, J. (2019). Adaptive event-triggered anomaly detection in compressed vibration data. *Mechanical Systems and Signal Processing*, *122*, 480-501. https://doi.org/10.1016/j.ymssp.2018.12.039

Adaptive event-triggered anomaly detection in compressed vibration data

Yang Zhang[a,b], Paul Hutchinson[c], Nicholas A J Lieven[a], Jose Nunez-Yanez[b]

[a] *Department of Aerospace Engineering, University of Bristol, Bristol, UK*
[b] *Department of Electrical & Electronic Engineering, University of Bristol, Bristol, UK*
[c] *Beran Instruments Limited, Hatchmoor Industrial Estate, Torrington, Devon, UK*

Abstract

Anomaly detection is a crucial task in Prognostics and Condition Monitoring (PCM) of machinery. In modern remote PCM systems, data compression techniques are regularly used to reduce the need for bandwidth and storage. In these systems the challenge arises of how the compressed (distorted) vibration data affects the condition monitoring algorithms. This paper introduces a novel algorithm that can adaptively establish normal bounds of operation from continuous noisy vibration profiles working with compressed vibration data. The proposed technique is based on four modules, including feature extraction, feature fusion, extreme value vibration modeling and adaptive thresholding for anomaly detection. The proposed method has been validated with experiments using three time-series datasets. The experimental results indicate that the proposed algorithm is able to perform detection of malfunctions in rotating machines effectively without faulty reference data. Moreover, the proposed method is able to produce accurate early warning and alarm indications from both the raw and compressed (distorted) datasets with equal veracity.

Keywords: Machine faulty detection, Health status modeling, Adaptive learning, Signal compression

Email address: yang.zhang@bristol.ac.uk (Yang Zhang)

1. Introduction

Prognostics and Condition Monitoring (PCM) of complex engineering systems has been proposed to evaluate the reliability of a machine within its operation conditions [1]. Failures in machine structural integrity can be a significant contributor to increased running costs, unplanned outages and even catastrophic events in plants (e.g. power generation plant). The stringent safety considerations together with the cost of such outages makes essential the condition monitoring (CM) of all machinery vital to plant operation. PCM needs continuous streams of raw noise and vibration data which are routinely captured by the CM systems in a constant high precision state. The storage of high precision floating point vibration data is costly, and a high bandwidth network is required to transfer the data from the sensor to the control room or monitoring station. This is currently addressed by applying algorithms that produce low bandwidth Condition Indicators (CI) in the process discarding raw data. However, the understanding of the failures resulting from transient events in retrospect becomes more difficult if the reduced sampling frequency means that valuable information has been lost. Another approach is to apply data compression algorithms [2] that can adapt its acquisition conditions and compression ratio without loss of critical information.

In this paper the vibration data is collected from multiple sensors and compressed using the techniques initially presented in [2] in real-time using a low precision base layer and high precision refinement layer. An Adaptive Event Trigger Mechanism (AETM) system is optimized to process only the base layer so no high precision decoding is required under nominal machine operation. This means that in the steady nominal state the refinement layer can be ignored safely. If a trend or event is detected that triggers a requirement for more detailed data analysis, the decompressor performs the reconstruction of the high precision signal on the basis of the base layer elements and the superimposed elements from the refinement layer. Event detection is performed by the AETM algorithm that detects a change in the noise and vibration signature properties

in the form of a trend or a transient event. The objective of this set-up is to obtain accurate monitoring data leading up to and during an event to enable effective data storage, bandwidth management and robust decision making.

The main contributions of this paper are as follows:

- A novel AETM algorithm that utilizes a hybrid structure composed of trend estimation and health status modeling based on Principal Component Analysis (PCA) and Multivariate Gaussian Distribution Modeling (MGDM).
- An in depth evaluation on how the performance of the AETM algorithm is influenced by compressed/distorted vibration signal using publicly available data sets and realistic machinery data.

The remainder of this paper is structured as follows. Section 2 gives an overview of the related work in PCM algorithms. Section 3 presents the proposed AETM system in detail. Section 4 presents descriptions of the experimental datasets, while Section 5 presents performance results with different testing conditions. Finally, Section 6 concludes this study and suggests potential future work.

2. Related work

A number of approaches have been investigated to create PCM systems in recent years. The approaches can be generally categorized into three types: Physics/model-based, data-driven and hybrid [3]. The physics/model-based approaches attempt to build mathematical or physical models based on the characteristic of the system physics, damage mechanics and/or human expert knowledge to describe the equipment damage or degradation process [4, 5, 6, 7]. The commonly used models include the Gaussian mixture model (GMM) [8], the Wiener process model [9] and Markov process model [10, 11, 12]. The data-driven method does not use any particular physical model and mainly relies

on measured data [13]. Hybrid approaches combine the above mentioned two methods together to improve the performance of the prediction [14].

An important consideration in model-based methods is how to estimate model parameters according to the characteristics of the monitored machine and the associated measured signals. In [15], a Bayesian-based Particle filtering (PF) is derived from traditional Kalman filtering, which is capable of combining measured signals and expert knowledge from empirical models to evaluate the state of the system. Zio et al. [16] employed the PF filtering algorithm in conjunction with a Monte Carlo estimation algorithm to perform Remaining Useful Life (RUL) predictions. In [14], a prognostics approach based on PF for estimation of the RUL of aluminum electrolytic capacitors has been proposed. The proposed approach assumes that the degradation model and the values of its parameters are known. Lei et al. [17] employed the Paris-Erdogan model to represent the degradation processes of rolling element bearings and employed maximum likelihood estimation for initialization of the model parameters. Yu [18] proposed a bearing health degradation monitoring method that was used to calculate the similarity between probability density functions described by two different hidden Markov models. The Wiener process has many properties, such as non-monotonic, infinite divisibility and a physical interpretations property, which can be employed to characterize the dynamic characteristic of a system. It has been widely used in various industrial applications [19, 20]. Duan et al. proposed a CM method which incorporation of the influence of both aging and degradation state in modeling the machine status in [21]. The proposed method utilized a semi-Markov decision process framework to obtain the thresholds by formulating and analyzing the health status of the machine as an optimization problem.

In practice, how to model the physics and principles of a system operation is not straight forward. The behaviors change differently among different components and machine status due to individual high dimensional diversity. The characteristics that define the nature and the origin of the degradation of the components/machines are usually unavailable. All of these factors make it

difficult to build an accurate and robust model of the machinery system.

The data-driven approach is a popular alternative for PCM proposed in recent years as it monitors the degradation process of a system from the collected data directly. The prognostic prediction of the failure progression using statistical and/or machine learning algorithms has been conducted in [22, 6, 23, 24, 25, 26, 27, 28, 29]. In general, the data-driven approaches follow a procedure composed of: vibration data acquisition, time/frequency domain feature extraction, feature fusion, prognostic modeling and prediction.

Data-driven approaches try to utilize the statistical features that are directly extracted from collected raw data as the indicators of PCM [30]. Each kind of feature is susceptible to a certain kind of failure in a certain degradation stage [31]. A robust indicator for PCM should take advantage of mutual information from multiple features of degradation assessment and health status prediction. The drawback of data-driven approach is that, the accuracy of the PCM model is highly dependent on the sensitivity and reliability of the extracted features. However, some of these features may present redundancy, irrelevance or have been distorted by a compression algorithm, and these kinds of features may lead to contradictory information and thus reduce the performance of the prognostic modeling and prediction analysis.

To avoid this problem a systematic data fusion algorithm is necessary to extract and select the most representative features from the statistical features. Previous research on multiple features selection and indicator construction with fusion have been proposed [32, 13, 24]. Feature selection and fusion algorithms aim at selecting a subset of the most relevant and informative features from the original feature space [33]. Feature selection based on principal component analysis (PCA) was investigated in [34, 35, 36]. Due to its ability to discriminate directions with the largest variance in a data set, the PCA algorithm was employed to identify the most representative features for defect classification in bearing condition monitoring. Mujica et al. proposed method based on multivariate statistical process control techniques, which employed PCA and partial least squares method to transform highly correlated redundant and noisy data

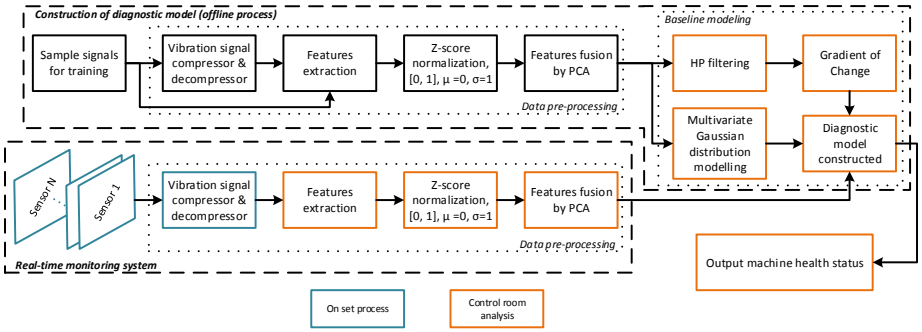


Figure 1: The flowchart of proposed Adaptive Event Trigger Mechanism (AETM) system.

to a model for structural damage detection [35]. In [36], the PCA algorithm has been employed in conjunction with T^2 and Q-statistic measures to detect and distinguish damages in structures. Kernel PCA (KPCA) which is a nonlinear extension of traditional PCA explores the nonlinear relationship among variables in high-dimensional feature spaces by means of integral operator and a nonlinear kernel function for nonlinear process monitoring [37, 38]. Kordestani et al. proposed a fault diagnosis and prognosis system in [39]. In order to remove re-dundancy in the multisensory data, the feature selection has been produced by Pearson product-moment rank correlation and SVM technique to capture redundant information in the multisensor data.

Nowadays, industrial CM systems use multiple sensors and generate continuous streams of raw noise and vibration data. Compression can be used to reduce storage requirements and data transfer bandwidth to the control room outlined in the authors' previous work [2]. The effects of compression on the trigger mechanism have not been previously investigated although it could have negative effects on the accuracy and robustness of the subsequent thresholding and decision making processes. To address this problem we create a AETM that follows a statistical approach for trend estimation and health degradation modeling using compressed vibration information.

3. Proposed Adaptive Event Trigger Mechanism (AETM) system

We create an adaptive monitoring method to quantify the operation status of machine health. The proposed health monitoring framework of the AETM is shown in Fig. 1, which contains two key parts: off-line diagnostic modeling and on-line health monitoring. The off-line diagnostic modeling employs samples of historic signals from a normal machine to construct a "normal" model. The samples have been compressed with the Vibration Signal Compressor (VSC) and directly imported to the feature extraction process. The Z-score method has been employed for normalization of the features. The PCA is utilized to fuse the extracted raw features from the sensor signals and compressed bit-stream to analyze the principal components (PCs), which are used as the inputs to the baseline modeling process. A gradient of change method in conjunction with a multivariate Gaussian distribution modeling constructs the diagnostic model. The diagnostic model performs the on-line analysis and quantification of the machine health status. The constructed diagnostic model adapts the trigger condition dynamically guided by data changes at the diagnostic level during machine operation.

The VSC algorithm proposed in [2] is part of a reliable and effective health monitoring technology for machines and civil infrastructure (e.g. power generation system), as shown in Fig. 2. The adaptive vibration signal compression scheme is composed of a lifting discrete wavelet transform (LDWT) with set-partitioning embedded blocks (SPECK) that sorts the wavelet coefficients by significance [40]. The algorithm exploits the clustering of energy found in the transformed domain and concentrated on those areas of the set which have high energy. This allows those signals with higher information to be encoded first based on their energy content. The output of the SPECK module is input to an optimized context-based arithmetic coder that generates the compressed bit-stream. The experimental results indicate that the VSC efficiently compresses the vibration signal with high quality reconstruction as it will be shown in Section 5.2.

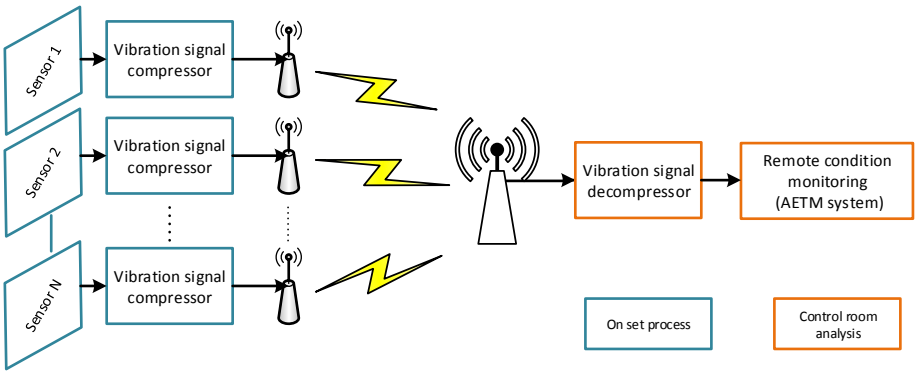


Figure 2: Example diagram of the signal acquisition, compression and transmission. For more details of the operation procedure of "On set precess"(blue boxes) and "Control room analysis" (orange boxes) can be found in Fig. 1.

3.1. Vibration Signal Compressor (VSC)

The benefit of the VSC algorithm is that a single coded bitstream can be used to decode the signal at any rate less than or equal to the coded rate, to provide the best reconstruction quality [2]. The algorithm is very useful for remote CM systems that can perform an initial base layer decoding so the AETM can quickly inspect the vibration signal and decide if it should be fully decoded (adding more detailed information to the base layer) for further analysis, or it is good enough to be categorized as normal health status signal.

In particular, the VSC algorithm [2] employs Lift Discrete Wavelet Transform (LDWT) for decomposing the complex vibration signal into basis signals of finite bandwidth, followed by Set-partitioning Embedded Blocks (SPECK) sorting algorithm to increase the coding performance by reducing the number of wavelet coefcients that needs to be coded, as well as exploiting the higher order statistical dependencies with context-based modeling and arithmetic coding. The experimental results indicate that the proposed VSC algorithm could achieves up to 59.41% bitrate reduction with same reconstruction quality level compare to the state-of-the-art JPEG2000 coder, while retaining the characteristics in the compressed vibration signals to ensure accurate event analysis [2].

3.2. Feature extraction

The vibration signal is composed of random noise and periodic fluctuations. Due to the low signal to noise ratio of the sensing measurement, it is challenging to model the relationship between the captured raw vibration signal and the health states. To tackle this problem, feature extraction methods aim to denoise the raw signal and refine the data.

In this paper, the feature set recommended by [13, 24, 41] includes ten features in time-domain to capture the degradation characteristics, including Mean, Root Mean Square (RMS), Peak-to-Peak, Square Mean Root, Standard Deviation, Kurtosis, Skewness, Crest Factor and Entropy. RMS is a measure for the magnitude of a varying quantity. It is also related to the energy of a signal. Kurtosis indicates the spikiness of a signal. Additionally, features from the time-frequency domain provide another perspective of the machinery conditions and reveal information that are not found in the statistical time-domain. Eleven features in time-frequency domain have been extracted. In particular, the signal is decomposed by a three-level lifting-based discrete wavelet transform. The energy of the wavelet coefficients and energy ratios of nine frequency sub-bands are also added into the feature set. The Kullback-Leibler Divergence (Div) has been employed as a condition feature, where φ denotes the sub-band of wavelet decomposition, E_i^c denotes the energy of the wavelet signal in the current state, and E_i^b is the energy of the signal in the baseline state (normal health state). The Div represents a measure of how the energy distribution of the current state diverges from the energy distribution of the baseline state. The definition of these features is listed in Table 1.

3.3. Indicator matrix construction

In order to process the features extracted from the vibration signal, an indicator matrix F corresponding to the monitored machine operational state is constructed as:

$$F = [f_1, f_2, ..., f_m]^T \tag{1}$$

where m denotes the number of features f (set to 21 in this study). The indicator condition is organized with signal features in rows. However, the different feature spaces are in different units and magnitudes, which causes that these features do not have equal contributions to the PCM indicators. To solve this problem, the vibration signal features in the indicator matrix are normalized by the Z-score standardizing method [42], converting all of the statistical features to a normalized space. The standardized indicator matrix, Z, retains the signature properties (e.g. skewness and kurtosis) of the original matrix, and has mean $\mu = 0$ and standard deviation $\sigma = 1$.

3.4. Feature fusion

In a multiple sensor condition monitoring system, the amount of feature data extracted from the raw data is overwhelming. Practically, the features can be treated as a high-dimensional multivariate indicator matrix composed of several vectors formed by the different features. It is impracticable to import the raw feature matrix to construct a model due to the high dimensionality and the high correlation between feature vectors. In order to improve the computational efficiency without compromising the signature of the extracted features, Principal component analysis (PCA) has been employed for feature fusion and eliminate correlations between variables [43].

The key idea of the PCA algorithm is that, by calculating the eigenvectors of the covariance matrix of the original inputs, the PCA algorithm transforms a high-dimensional input vector into a low-dimensional one whose components are uncorrelated. More specifically, given the normalized indicator matrix $Z = \{z_1, z_2, ..., z_m\}$, where $z_i^T = \{F_{i,j}, j \in [1, d]\}$, can be constructed. d denotes the number of features vectors of m samples. The corresponding covariance matrix R_Z can be obtained as,

$$R_Z = \sum_{j=1}^{m} (z_{i,j} - \bar{z})(z_{i,j} - \bar{z})^T \tag{2}$$

where $\bar{z}$ is the mean feature vector of all samples of Z and superscript T denotes the transportation. Accordingly, the corresponding eigenvalues of the matrix Z

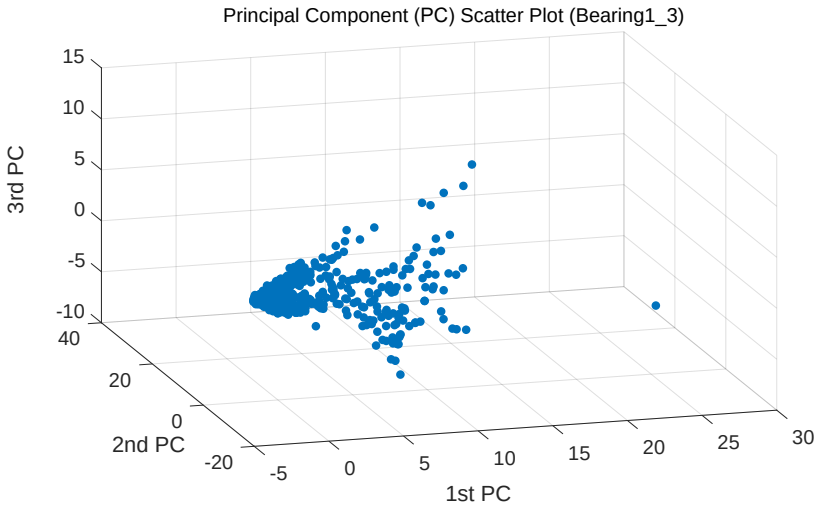


Figure 3: Example results of PCA algorithm, where only the most three representative PCs have been plotted for visualization

can be obtained by solving the following equation:

$$\lambda v = R_Z v \tag{3}$$

where λ and v denotes the eigenvalue and the eigenvector of R_Z, respectively. Consequently, by solving (3), the largest d eigenvalues can be obtained. The eigenvalues are usually arranged in the descending order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. The corresponding eigenvectors are represented by, $v_j, j \in [1, d]$. The jth principal component feature of a sample z_i, denoted as $\mathbb{P}_{i,j}$, can be obtained by projecting z_i onto the direction of the eigenvector v_j:

$$\mathbb{P}_{i,j} = v_j^T (z_i - \bar{z}) \tag{4}$$

The eigenvectors span a d-dimensional orthogonal space and the principal component (PC) representations can be obtained by projecting Z onto the space. For visualization, Fig. 3 shows the three most representative PCs.

The representativeness of each PC can be calculated based on its contribution to the machine health condition. In order to represent and detect machine health conditions effectively, the most representative PCs are selected by re-

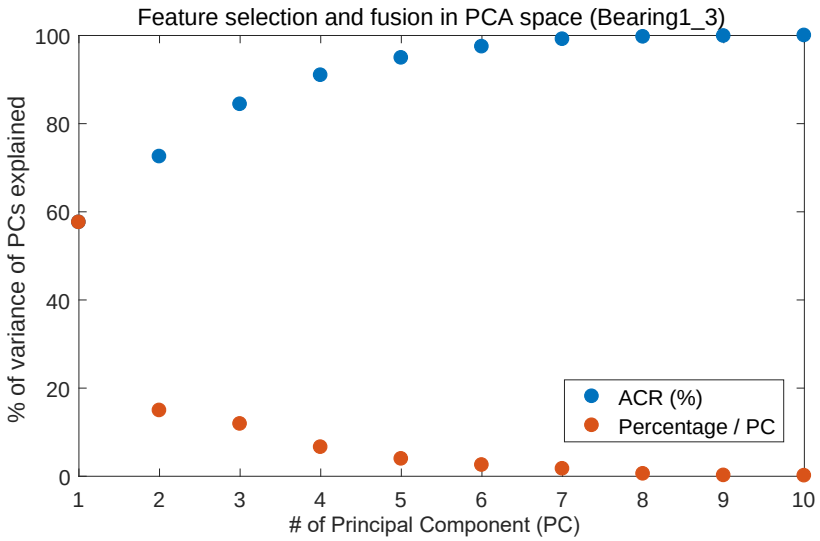


Figure 4: Example result of the Accumulative Contribution Rate (ACR) method.

ducing the computational load without compromising the characteristic of the features. The accumulative contribution rate (ACR) method has been employed to selects the first m PCs based on the eigenvalues. The ACR method can be defined as,

$$R_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \tag{5}$$

where R_m denotes the percentage of the total variance in the original feature space F, which can been discriminated by the m PCs. A specific threshold for the ACR method needs to be chosen depending on the specific machine monitoring applications and dataset. Fig. 4 shows an example result of the ACR method, which indicates that more than 85% of the information of the dataset can be represented by the first three principle components. Considering the increment of the computational complexity demands and diminishing returns, only the first three principle components have been employed to produce all the experimental results presented in this paper.

3.5. Data cleaning and trend detection

A fundamental challenge in machine condition monitoring and prognostics is the trend detection of the vibration signal. The PCA algorithm is employed to fuse and select the most representative information from the original feature space. However, the selected low-dimensional PCs contain noise components which is problematic for trend prediction, as the blue line shows in Fig. 5. The next subsections propose filtering based on Hodrick-Prescott (HP) filter [44] to smooth out the irregular roughness and trend estimation based on Mahalanobis distance adaptive thresholding method in conjunction with the Gradient of Change method.

3.5.1. Hodrick-Prescott (HP) filtering

The Hodrick-Prescott (HP) filter [44] is employed to obtain a smoothed variation in the principle component feature space. The HP filter is a mathematical tool used in macroeconomics for time series data analysis. The advantage of the HP filter is that it is suitable in filtering terms for estimating long term trends in time series without time delay [45]. Additionally, HP filter that provide results which are approximately consistent under aggregation are considerably robust with respect to the Auto Regressive Integrated Moving Average (ARIMA) model of the time series signal and perform better for the case of temporal aggregation than for systematic sampling [45, 46, 47]. The HP filter decomposition is normally used for estimating the up-/downward drift or trend in the non-stationary macroeconomic time series raw data.

Suppose the magnitude of each PC, $\mathbb{P}_t$, is composed of a trend component (τ_t) and a cyclical component (c_t), where t denotes PC number. The magnitude variation of the PCs can be defined as:

$$\mathbb{P}_t = \tau_t + c_t, \quad t = 1, 2, ..., M, \tag{6}$$

Hodrick and Prescott in [44] suggest a way to isolate c_t from L_t by following

minimization problem.

$$\mathrm{Min}_{\{\tau_t\}_{t=1}^{T}} \left\{ \sum_{t=1}^{T} (\mathbb{P}_t - \tau_t)^2 + \xi \sum_{t=1}^{T} \left[(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2}) \right]^2 \right\} \tag{7}$$

where the parameter ξ is a positive number which penalizes variability and controls the smoothness in the trend component. The solution series becomes smoother as ξ increases. The first term in the loss function (7) penalizes the variance of the cyclical component c_t, while the second term puts a penalty to the lack of smoothness in the trend component τ_t. The HP filter identifies the cyclical component c_t from $\mathbb{P}_t$ by the trade-off to the extent to which the trend component keeps track of the variation of the $\mathbb{P}_t$ for good fitting against the prescribed smoothness in τ_t.

Note that, as ξ approaches 0 in (7), the trend component becomes equivalent to the original series, while as ξ diverges to infinity, τ_t approaches a linear trend. In this paper, we set $\xi = 1600$ based on an empirical investigation. Fig. 5 shows the magnitude $\mathbb{P}_1$ of each PC with and without HP filtering, where the blue curve represents the raw $\mathbb{P}_1$ values and the red curve represents the values of $\mathbb{P}_1$ after the temporal filtering processing.

3.5.2. Gradient of Change (GoC) for trend detection

In this paper, we propose a Cumulative Gradient of Change (CGoC) method to calculate and specify how the trend of the machine health state changes. The benefit of the CGoC is the attribute that takes into consideration the short and long term trend estimation. In practice, in order to estimate the trend in a short time period, the time series data has been segmented to short non-overlapping time windows. The local trend can been estimated based on the calculation of the derivatives of each short segment. Following the output of the H-P filtering, suppose that the polynomial expressions, $f(x)$, for each short time window, shown as the colored segment lines in Fig. 6. Then, the rate at which x changes

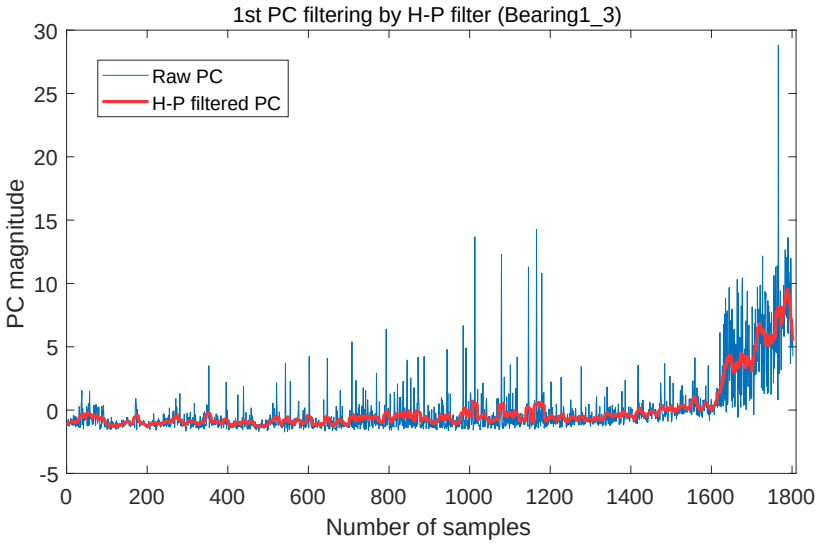


Figure 5: Example result of the raw Principal Component (PC) magnitude (blue curve) and the Hodrick-Prescott (H-P) filtered results (red curve).

can be denoted as, $\frac{\mathrm{d}}{\mathrm{dx}} f(x)$, where

$$\begin{cases} \text{if } \frac{\mathrm{d}}{\mathrm{dx}} f(x) \leq 0, & \text{then } f(x) \text{ is decreasing} \\ \text{if } \frac{\mathrm{d}}{\mathrm{dx}} f(x) = 0, & \text{then } f(x) \text{ is at a relative maximum or minimum} \\ \text{if } \frac{\mathrm{d}}{\mathrm{dx}} f(x) \geq 0, & \text{then } f(x) \text{ is increasing} \end{cases}$$

The instantaneous GoC (tangent line) of the expression is the first derivative at the point. So, the trend of the segment time window can be obtained by

$$\begin{aligned} T_w &= \sum_{x=1}^{W} \frac{\mathrm{d}}{\mathrm{dx}} f(x) \\ &= \sum_{x=1}^{W} \lim_{h \to \infty} \frac{f(x+h) - f(x)}{h} \end{aligned} \tag{8}$$

where T_w denotes the CGoC, which represents the estimated trend of the specific segment time window. W denotes the number of samples in the segment window. We cannot really take the limit $h \to \infty$, as we have only a quantized set of values to work with, as the smallest possible value is $h = 1$ as a division by

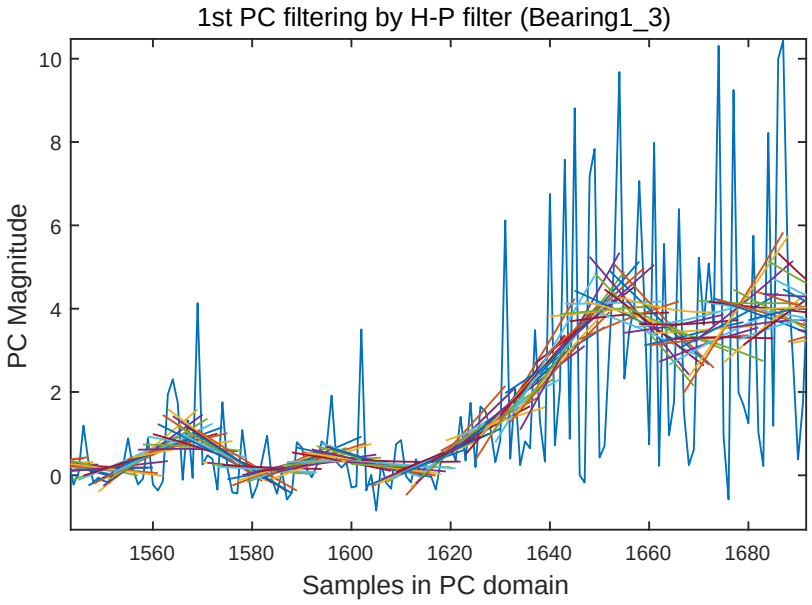


Figure 6: Example result of the proposed Gradient of Change method in Principal Component (PC) domain (blue curve) with Hodrick-Prescott (H-P) filter (colored segments).

zero occurs at the next smallest possible value $h = 0$. Therefore, the derivatives are approximated by,

$$\frac{\mathrm{d}}{\mathrm{dx}} f(x) \doteq (f(x+1) - f(x)) \tag{9}$$

Fig. 6 is an enlarged visualization shows that the local trend (colored line segments) of each segment time window. The uphill trend between the samples [1620 1650] can be clearly seen. The middle point value of each segment time window of the output of the HP filter has been utilized as magnitude value of the smoothed principal component, as shown in Fig. 5 (red curve).

The Mahalanobis distance has been employed to determine the threshold of the health condition states. The benefit of the Mahalanobis distance are the properties of unitless and scale-invariant, and takes into account the correlations of the dataset. The Mahalanobis distance (d) of the dataset, T_w, can be defined as,

$$d(T_w) = \sqrt{(T_w - \mu)^T \ \Sigma^{-1} (T_w - \mu)}, \tag{10}$$

where T_w is the vector of the CGoC with the corresponding mean μ, and Σ denotes the covariance matrix.

3.6. Multivariate Gaussian distribution modeling for anomaly detection

In [36], the Q-statistic and the T^2-statistic indices have been employed for analyzing the variability of the data projection in residual- and PCA- subspace respectively for damage detection indices. This method based on the assumption that the underlying process follows approximately a multivariate normal distribution where the first moment vector is zero.

In this section, we present the next stage of the proposed method which is to construct a model of normality using "healthy" training data. This is achieved via a Multivariate Gaussian Distribution Modeling (MGDM) in conjunction with extreme value detection with the Mahalanobis distance algorithm. The statistical decision process based on Gaussian probability models have low computational intensity and often lead to good performance in the detection task previously observed in the machine learning literature [48]. We shall model the feature signals as multivariate normal vectors $\mathbb{P}$ that follow a multivariate normal distribution with mean vector μ and covariance matrix Σ. The probability density function (PDF) using a multivariate Gaussian distribution is given by,

$$p(\mathbb{P};\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} exp(-\frac{1}{2}(\mathbb{P}-\mu)^T\Sigma^{-1}(\mathbb{P}-\mu)) \tag{11}$$

where $|\Sigma|$ represents the determinant of matrix Σ, n is the dimension of the data space $\mathbb{P}$, μ and Σ can be defined as follows,

$$\mu = \frac{1}{m}\sum_{i=1}^{m}\mathbb{P}^{(i)} \tag{12}$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(\mathbb{P}^{(i)}-\mu)\ (\mathbb{P}^{(i)}-\mu)^T. \tag{13}$$

Anomaly detection in the health status of the PCM system can been considered as outlier analysis for statistical purposes. The statistic analysis computes discordance measures for the signal and then compares the discordance with

a threshold. The discordance measures are based on statistics extracted from healthy/normal condition signal, while the health status thresholds are derived based on the optimal confidence region in the PDF. The probability densities are specified by the hypotheses in (14), since we can design an adaptive health status threshold based on the discordance between the hypotheses.

$$\begin{aligned} H_0 :&\ \mathbb{P} \sim \mathcal{N}(\mu_{\text{norm}}, \Sigma_{\text{norm}}) \rightarrow \text{Normal condition} \\ H_1 :&\ \mathbb{P} \sim \mathcal{N}(\mu_{\text{abnorm}}, \Sigma_{\text{abnorm}}) \rightarrow \text{Alarm condition} \end{aligned} \tag{14}$$

where the $\mathbb{P} \sim \mathcal{N}(\mu, \Sigma)$ represents the *Health status* classes follow $n-$dimensional multivariate Gaussian distribution with different mean vectors $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ in multivariate PC spaces.

In univariate space, in order to evaluate the discordance between two populations, the standardized or statistical distance is examined, such as, $d = \frac{|x_\zeta - \mu|}{\sigma_x}$, where x_ζ is the candidate outlier and μ and σ_x are the mean and standard deviation of the data sample respectively. In the multivariate space, not only the variances of the variables but also their covariances or correlations must be considered. Thus, the Mahalanobis distance d_{Mah} can be obtained according to the weighting matrix $\mathcal{S}$ as follow,

$$d_{\text{Mah}}(x_\zeta, \mu) := \|x_\zeta - \mu\|_{\mathcal{S}} = \sqrt{(x_\zeta - \mu)^T \mathcal{S}^{-1} (x_\zeta - \mu)} \tag{15}$$

In the multivariate (n-D) case, the Mahalanobis distance d_{Mah} provides such a measure. The equidensity contours of the PDF, p, characterized by constant d_{Mah}, are ellipsoids centered at μ whose axes are aligned with the eigenvectors of $\mathcal{S}$. The specific equidensity contour that accumulates a target probability can be found using the property that the squared Mahalanobis distance follows a $\mathcal{X}^2$ distribution, or Chi-squared distribution [49]. For the distribution $\mathcal{N}(\mu, \mathcal{S})$, whose PDF is (11), the probability accumulated in the health region R can be defined as follows,

$$R := \{x_\zeta \in \mathbb{R}^n, \text{ such that } d_{\text{Mah}}(x_\zeta, \mu) \leq d\} \tag{16}$$

$$\mathcal{P}(d) := \mathcal{P}(\|x_\zeta - \mu\|_{\mathcal{S}} \leq d) = \frac{\gamma(\frac{n}{2}, \frac{d^2}{2})}{\Gamma(\frac{n}{2})}, \tag{17}$$

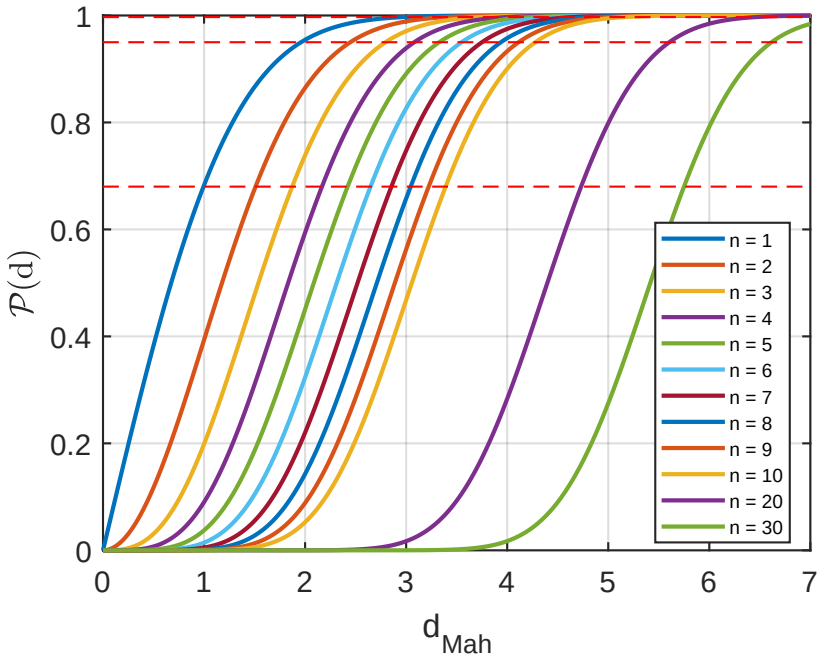


Figure 7: The CDF of (17) representing the relationship between the $\mathcal{P}$ and the d_{Mah} with n dimensions. The three red dash lines illustrate the 68-95-99.7 rule on the curve.

where $\Gamma(\alpha)$ is the Gamma function and $\gamma(\alpha, \beta)$ is the lower incomplete Gamma function. The equation (17) represents the values of the cumulative distribution function (CDF) of a centered chi-squared distribution with n degrees of freedom, $\mathcal{X}_n^2$, for a squared Mahalanobis distance d^2, as $\mathcal{P}_{\mathcal{X}_n^2(d^2)}$. Thus, given a probability value $0 < \mathcal{P} < 1$, the threshold value d_{Mah} that specifies the equidensity contour bounding the confidence region R for which $\mathcal{P}_{\text{health}} = \mathcal{P}(R)$ can be obtained by inverting (17). Fig. 7 illustrates that the correspondence between the cumulative probability $\mathcal{P}$ and the probability obtained with different dimensions $n = \{1, 2, ..., 10, 20, 30\}$ for the optimal threshold of the Mahalanobis distance d_{Mah}.

4. Experimental Datasets

In this paper, two public and one internal datasets have been employed for evaluating performance of the proposed method. The data sets include the NASA bearing dataset (*NASA_Bearing*) [50], PRONOSTIA (*PHM_Bearing*) [51] and a Power Plant Turbine dataset from an industrial installation (*Turbine*). In particular, the *NASA_Bearing* dataset contains three sets of data each consisting of four bearings. The vibration measurement signals provided for the dataset correspond to a system that was run to failure (RoF) under constant load and running conditions. Two high sensitivity accelerometers were installed on the bearing housing, where for each bearing [x- and y-axes] for data set 1, one accelerometer for each bearing for data sets 2 and 3. Vibration signals were sampled every 10 minutes, and duration of the sampling lasted 1 second with a sampling frequency at 20kHz. The *PHM_Bearing* dataset employed two accelerometers that were horizontally and vertically mounted on the bearing to monitor its vibration in x- and y-axes respectively. The vibration was measured under three different operating conditions (rpm and load), where the first set (*PHM_Bearing1*, 7 sets) with 1800rpm and 4000N, the second set (*PHM_Bearing2*, 7 sets) with 1650rpm and 4200N and the third set (*PHM_Bearing3*, 3 sets) with 1500rpm and 5000N. The vibration signals were sampled every 10s, and the duration of the sampling lasted 0.1s with a sampling frequency 25.6 kHz. The detailed information about the experimental setup of the datasets can be found in [50, 51].

The real vibration signals of the *Turbine* dataset have been collected from a power station main turbine. The dataset includes two sets, named *Turbine_1 and Turbine_2* respectively. The *Turbine_1* includes 28 channels of vibration data and for each channel are five derived 'condition indicators', including *RMS*, *Peak-to-peak*, *Order 1* (The vibration energy present at the turbines rotational frequency e.g. 50Hz or 3000rpm), *Sub Sync* (Vibration energy between 7% and 75% below Order 1 e.g. 3.5Hz to 37.5Hz) and *Non Sync* (Vibration energy above Order 1 but minus energy at orders 2, 3 and 4.), where 1-16 channels are

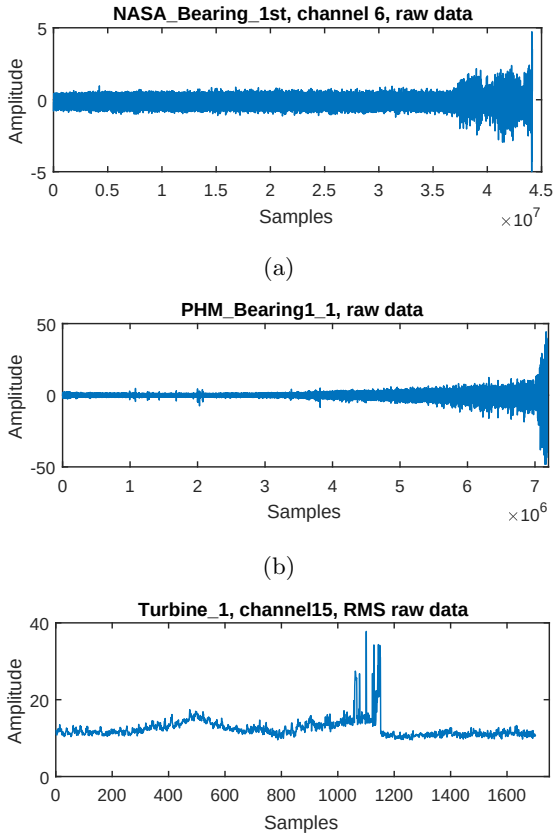


Figure 8: The example vibration signal raw data of (a) *NASA_Bearing*, (b) *PHM_Bearing* and (c) *Turbine* dataset.

in velocity units ([mm/s]) and 17-28 channels are in displacement unit ([um]). The *Turbine_2* has 16 channels of vibration data, and with same 'condition indicators' as in *Turbine_1*. The vibration signals have been measured as the main turbine was operating with range of 2950 - 3050rpm and different loads. Fig. 8 shows an example of the raw vibration signal contained in the tested datasets.

5. Experimental results

5.1. Health condition indicator construction

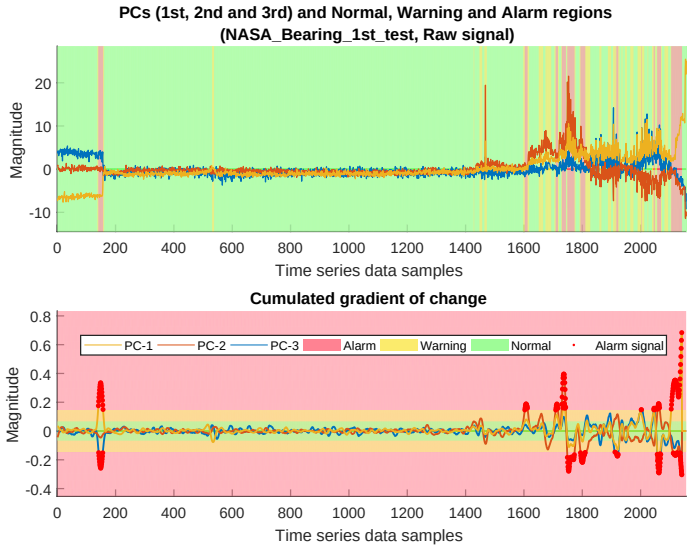
According to the construction procedure of the proposed PCM system, 21 features are extracted from the raw vibration signal to construct a raw feature matrix for the *NASA_Bearing* and *PHM_Bearing* dataset. For the *Turbine* dataset, a total of 5 condition indicators are available to construct the feature set in this experiment, which have been mentioned in Sec. 4. Then, the constructed feature matrix is used as the input into the features selection and fusion procedure using the PCA algorithm. Finally, the H-P filtered PCs are used as the input to the proposed CGoC and the MGDM algorithm.

5.2. Performance evaluation

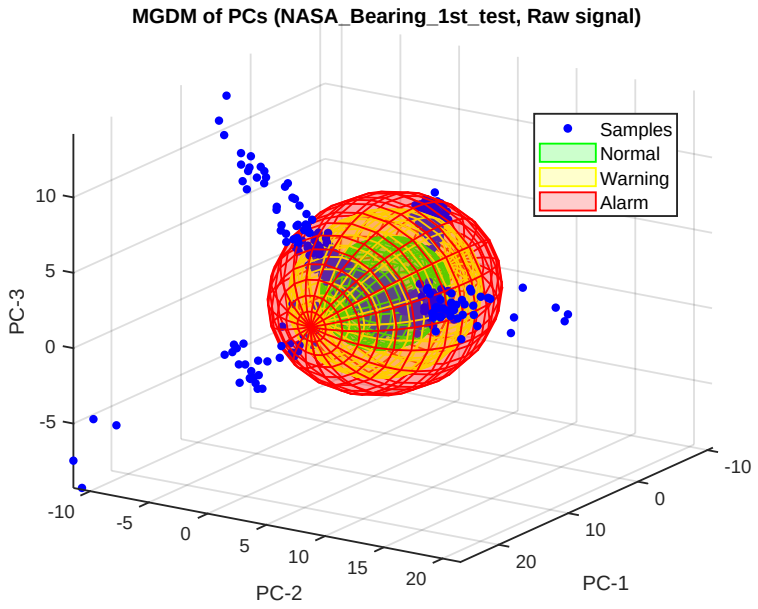
5.2.1. Evaluations on Raw signal

To evaluate the performance of the proposed PCM algorithm, the experiments use the raw vibration signal. In Fig. 9(a), (c) and (e), the first three principle components of each dataset and the results of the CGoC algorithm are shown, where green, yellow and red colored region indicates the *Normal*, *Warning* and *Alarm* operating status respectively. The results of the proposed MGDM algorithm are shown in Fig. 9(b), (d) and (f). For visualization, the first three PCs are plotted in 3D space, and the thresholds (d_{Mah}) for each status are chosen based on the equation (17) with $n = 3$ in this case. For example, during the estimation of the changes and the trend of the health status, the proposed CGoC algorithm is able to detect the vibration signal changes that occurred momentarily (e.g. the samples of *NASA_Bearing_1st_test* set ranging between samples {140, 157} in Fig. 9 (a)) or continuously for a period of the time (e.g. the samples of *NASA_Bearing_1st_test* set ranging between {1604, 2156}).

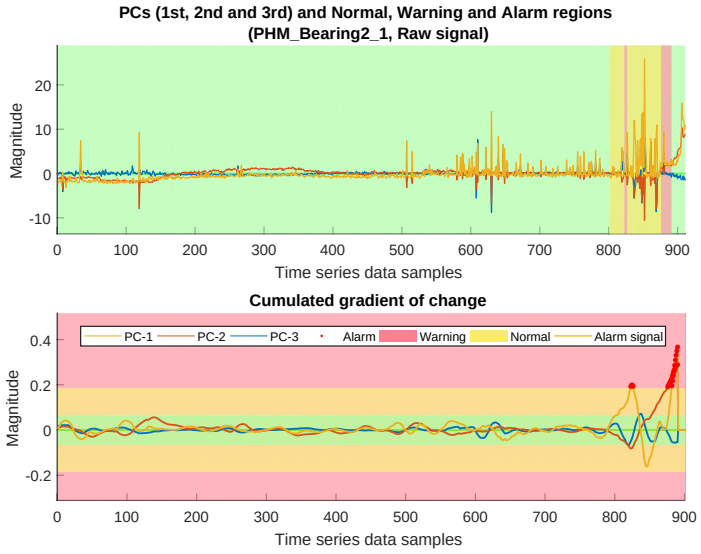
Additionally, the proposed MGDM algorithm models the health status with corresponding thresholds (d_{Mah}). The thresholds have been obtained based on the n-dimensional probability distribution model of the vibration data and the



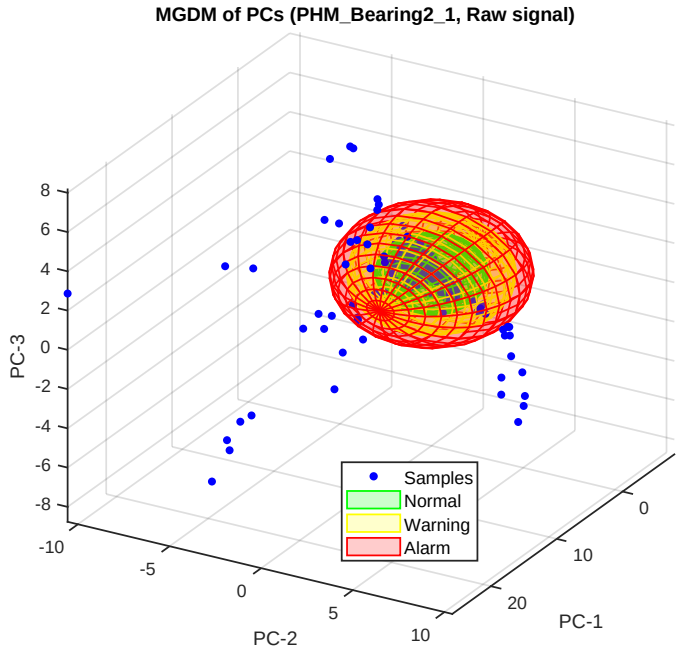
(a) CGoC algorithm results of "NASA_Bearing_1st_test", Raw signal.



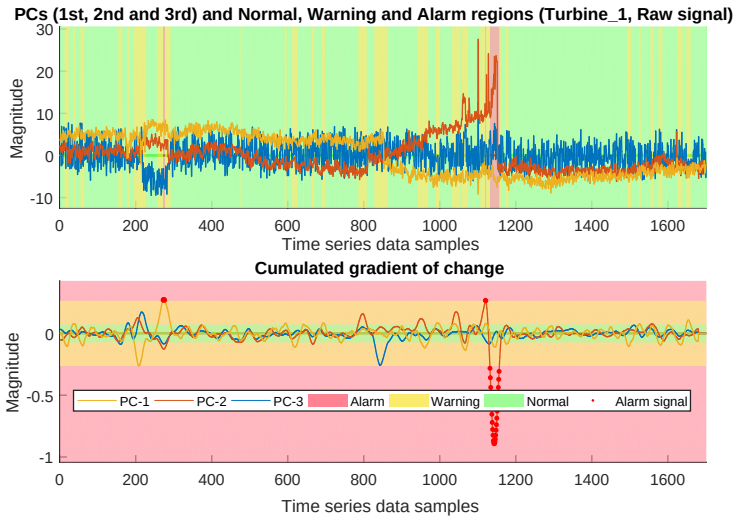
(b) MGDM algorithm results of "NASA_Bearing_1st_test", Raw signal.



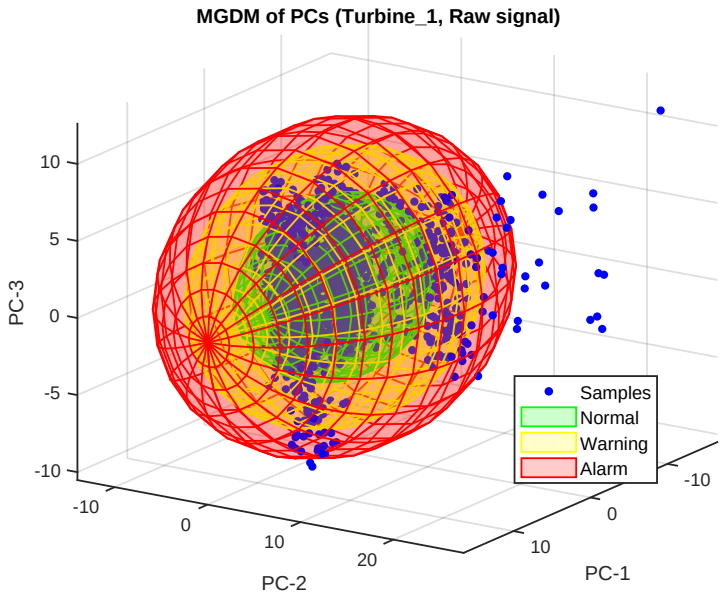
(c) CGoC algorithm results of "PHM_Bearing2_1", Raw signal.



(d) MGDM algorithm results of "PHM_Bearing2_1", Raw signal.



(e) CGoC algorithm results of "Turbine_1", Raw signal



(f) MGDM algorithm results of "Turbine_1", Raw signal

Figure 9: Experimental results of the Cumulative Gradient of Change (CGoC) algorithm ((a), (c) and (e)) and the Multivariate Gaussian Distribution Modeling (MGDM) algorithm ((b), (d) and (f)) on raw vibration signal of "NASA_Bearing_1st_test", "PHM_Bearing2_1" and "Turbine_1" datasets.

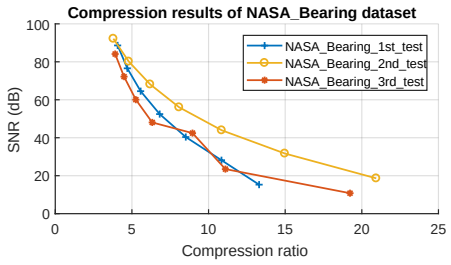
squared Mahalanobis distance that follows a $\mathcal{X}^2$ distribution (as discussed in Sec. 3.6). For this reason, the corresponding thresholds of the health conditions can be obtained adaptively based on the probability distribution of the new vibration signal and the dimensions of the data.

5.2.2. Evaluations on Compressed signal

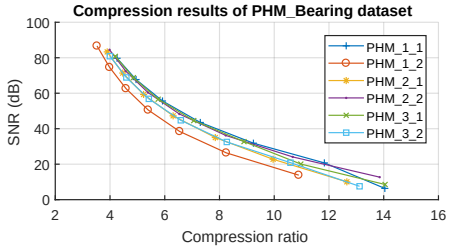
In order to evaluate the reliability and robustness of the proposed PCM algorithm in a modern practical scenario - the bandwidth and storage requirement are reduced using the state-of-the-art vibration signal encoder [2] compressing the raw vibration signal. Seven different distortion levels have been generated by compressing the vibration signal with different bitdepth representations, ranging from 4- to 16-bitdepth. All of the compression parameters and conditions have been set as same as which presented in [2]. For the *Turbine* dataset, 5 condition indicators raw signal have been compressed accordingly.

Fig. 10 shows the Signal to Noise Ratio corresponding to the compression ratio results of the three evaluated datasets respectively. As shown in the results, the higher bitdepth obtains better precision in the vibration signal, thus the 16 bitdepth compressor achieves more than 80 dB reconstruction accuracy, but with lower compression ratio. In contrast the 4 bitdepth compressor achieves less than 20 dB reconstruction accuracy, but compresses the vibration signal significantly, achieving more than 100:1 compression ratio on *Turbine_2* dataset.

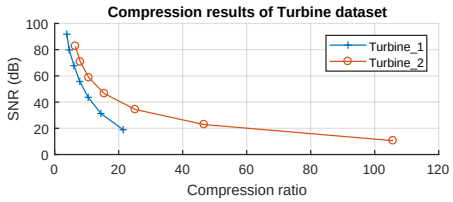
We have selected the graphs for raw signal (Fig. 9), 6 bitdepth (Fig. 11) and 4 bitdepth (Fig. 12) compressor, and one set from each test dataset. In particular, Fig. 11 shows the performance evaluation on the vibration signal compressed by the 6 bitdepth compressor. As shown in the figure, the proposed algorithm achieves very similar trend prediction and health status modeling results on the distorted dataset compare to the performance on the raw signal in Fig. 9. It is worth noting that, the 6 bitdepth compression result in Fig. 10 shows that, the reconstructed vibration signal has 26.5 dB signal to noise ratio (SNR) with 14.7:1 compression ratio on average. This means that the proposed method can estimate the trends and model the health status based on distorted



(a)



(b)



(c)

Figure 10: The rate-distortion performance (Compression Ratio vs. Signal to Noise Ratio (SNR)) of three test datasets: (a) *NASA_Bearing*, (b) *PHM_Bearing* and (c) *Turbine* dataset.

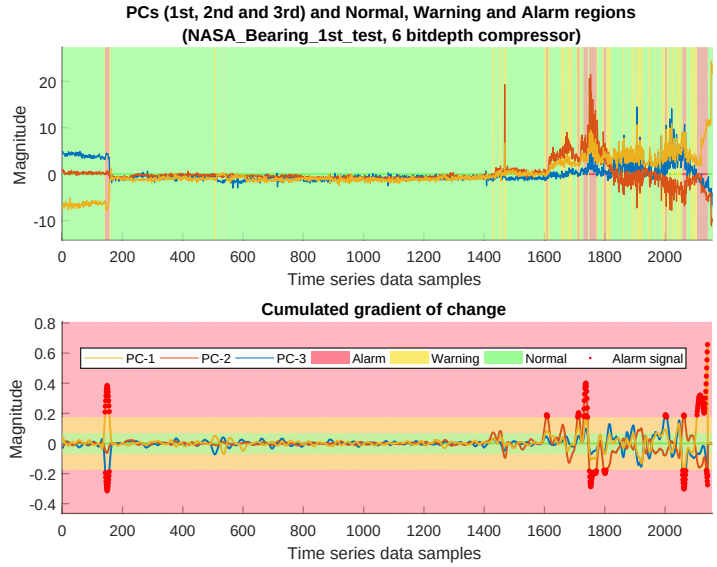
vibration accurately. This approach dramatically reduces the bandwidth and storage requirement for the condition monitoring system without compromising the condition monitoring performance.

In order to illustrate the performance limitation of the proposed method affected by the compressed and distorted vibration signal, Fig. 12 shows the experimental results of the proposed method evaluated on the vibration signal which have been compressed by the 4 bitdepth compressor. As shown in Fig. 10, the 4 bitdepth compressor achieves approximately 24.1:1 compression ratio

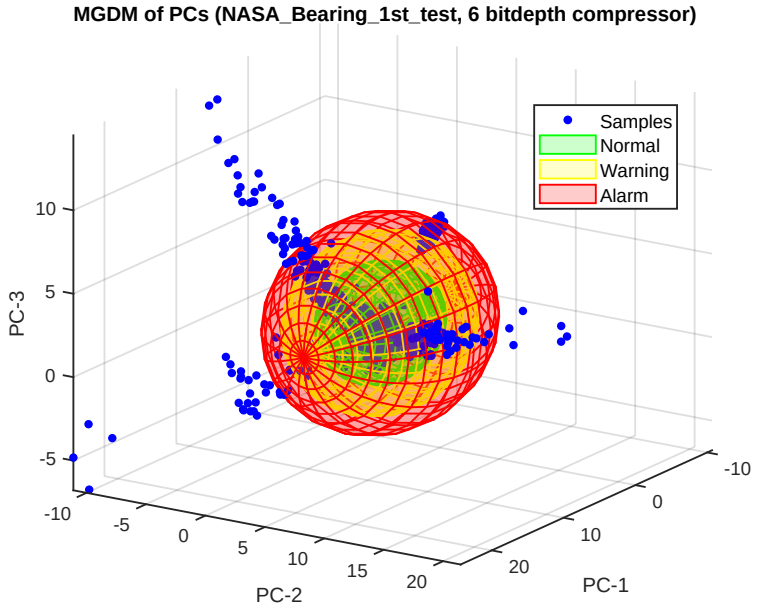
and 16.3 SNR on average. Under this high compression ratio scenario, the high frequency information contained by the principle components has been distorted significantly, as shown in Fig. 12 (a), (c) and (e). Compare to the results evaluated on the raw and 6 bitdepth signals, the proposed method apply to the 4 bitdepth compressed data shows an observable degradation in performance. In particular, the results of the CGoC algorithm indicate that the "*Warning*" signals has been triggered at the marginal time compare to the raw and 6 bitdepth results. However, the "*Alarm*" for the early faults with the 4 bitdepth compressed signal is not triggered correctly as can be seen during the samples between samples {200, 300} in Fig. 12 (e). Nevertheless, the "*Alarm*" signals have been triggered correctly for the actual faults. The MGDM algorithm achieves reasonable results, and the "Alarm" samples have been highlighted effectively.

6. Discussion and conclusion

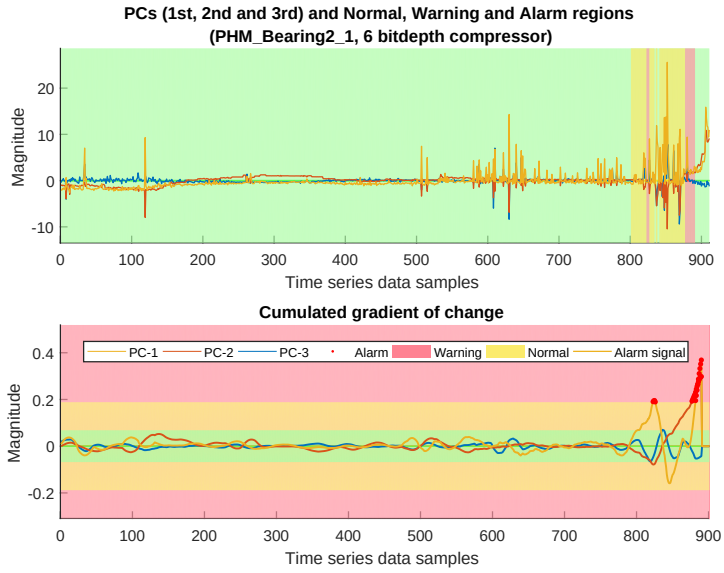
In this paper, we have proposed a hybrid data-driven approach for trend estimation and health status modeling to perform machine health prognostic and condition monitoring (PCM). The proposed method addresses the open challenges in the area of condition based maintenance of rotational machines. Firstly, as the vibration signal collected from the rotational machines contains noise this could result in false negative estimations. The proposed method is able to isolate and fuse the most representative features from the 21 time- and frequency-domain features extracted from the noisy vibration signal. Secondly, the proposed method only employs "normal" signals for the parameters' initialization, but does not require faulty reference data. Therefore, this method can be implemented in a wide range of rotational machine systems where fault data cannot be easily collected, such as a nuclear power plant. Thirdly, the experimental results indicate that the proposed method is capable of detecting and diagnosing intermittent and degradation faults. This capability enables early detection and trend prediction of the fault. Fourthly, the proposed method



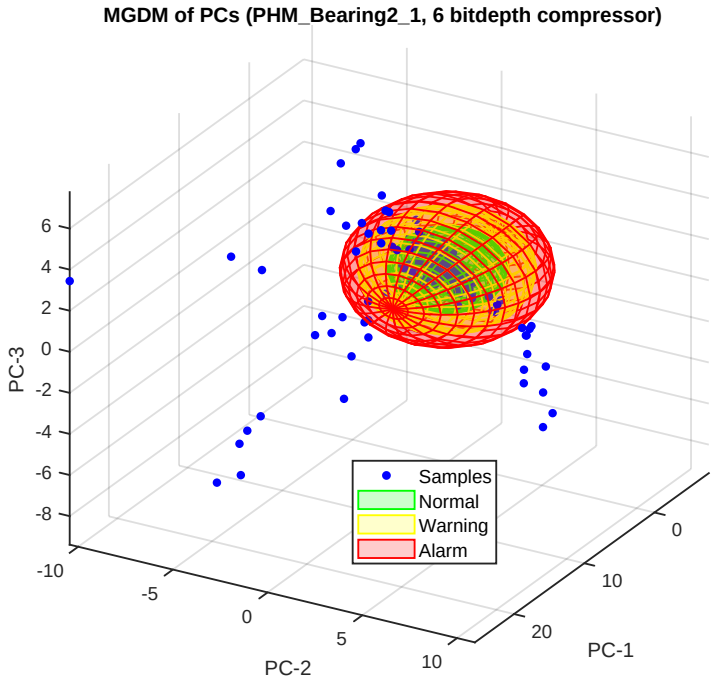
(a) CGoC algorithm results of "NASA_Bearing_1st_test", 6 bitdepth.



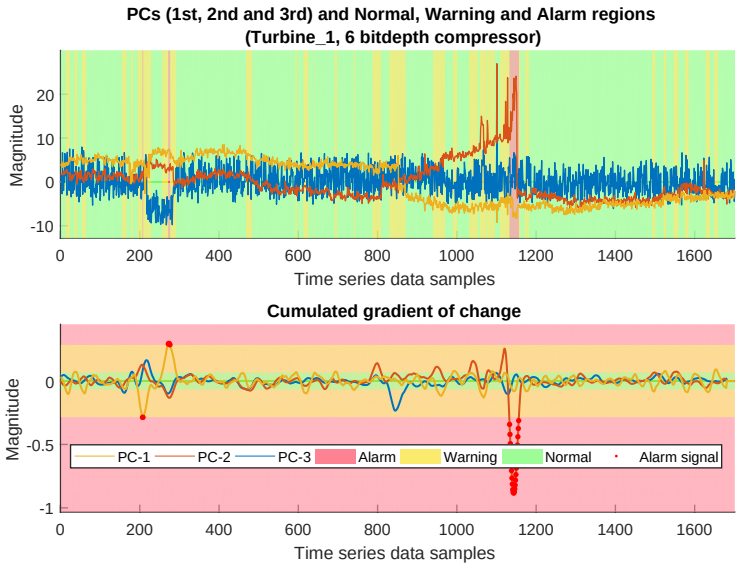
(b) MGDM algorithm results of "NASA_Bearing_1st_test", 6 bitdepth.



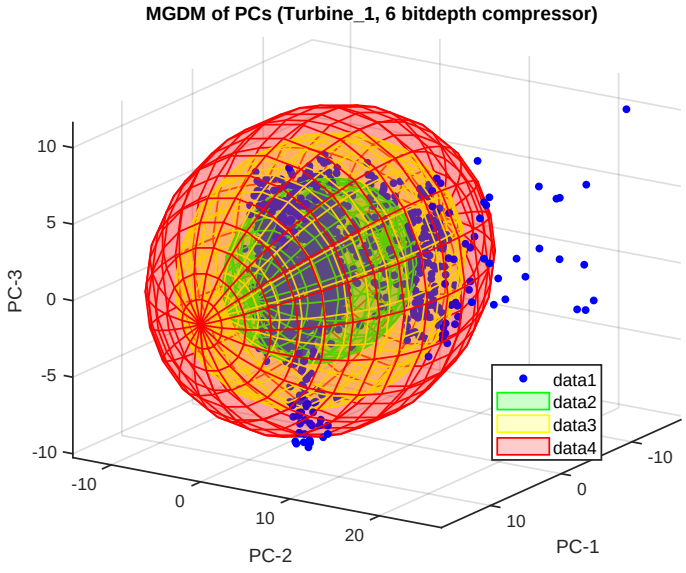
(c) CGoC algorithm results of "PHM_Bearing2_1", 6 bitdepth.



(d) MGDM algorithm results of "PHM_Bearing2_1", 6 bitdepth.

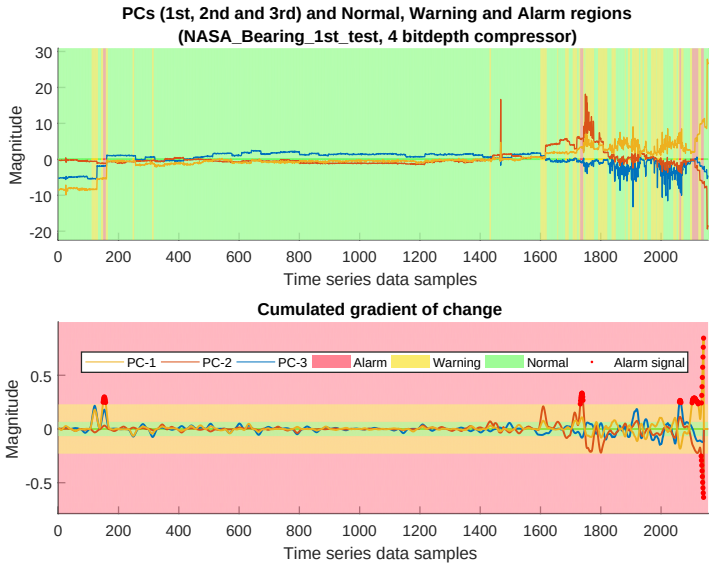


(e) CGoC algorithm results of "Turbine_1", 6 bitdepth.

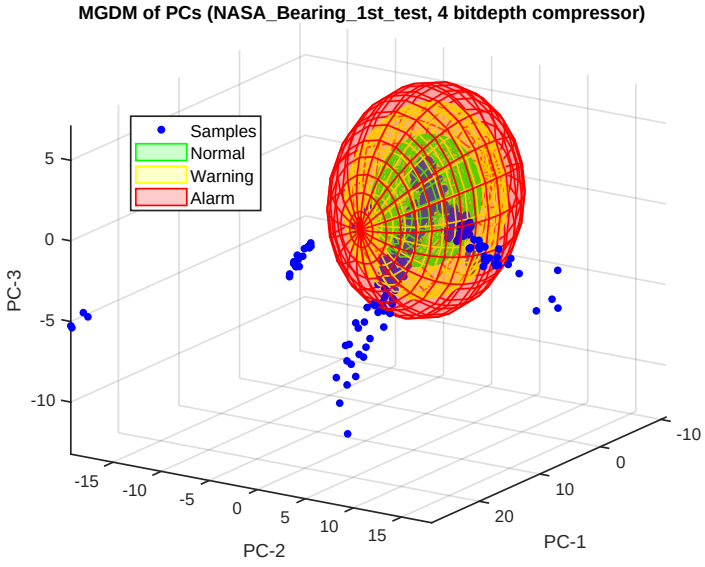


(f) MGDM algorithm results of "Turbine_1", 6 bitdepth.

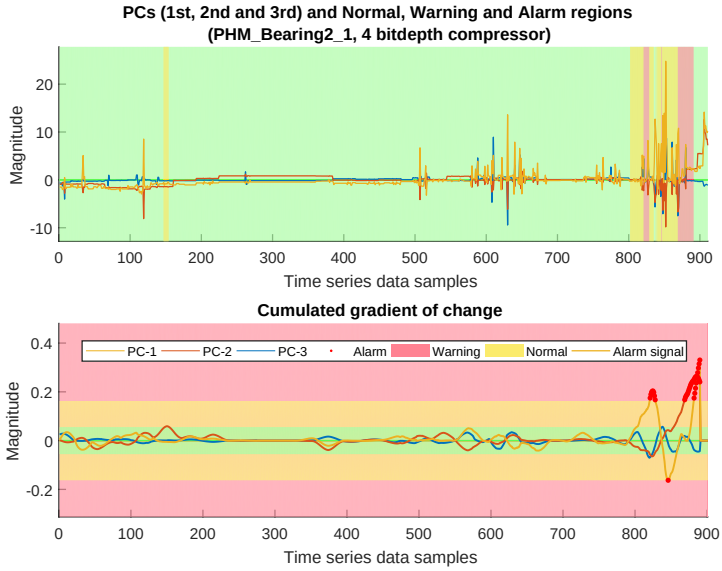
Figure 11: Experimental results of the CGoC ((a), (c) and (e)) and MGDM algorithm ((b), (d) and (f)) on compressed vibration signal of "NASA_Bearing_1st_test", "PHM_Bearing2_1" and "Turbine_1" datasets (6 bitdepth).



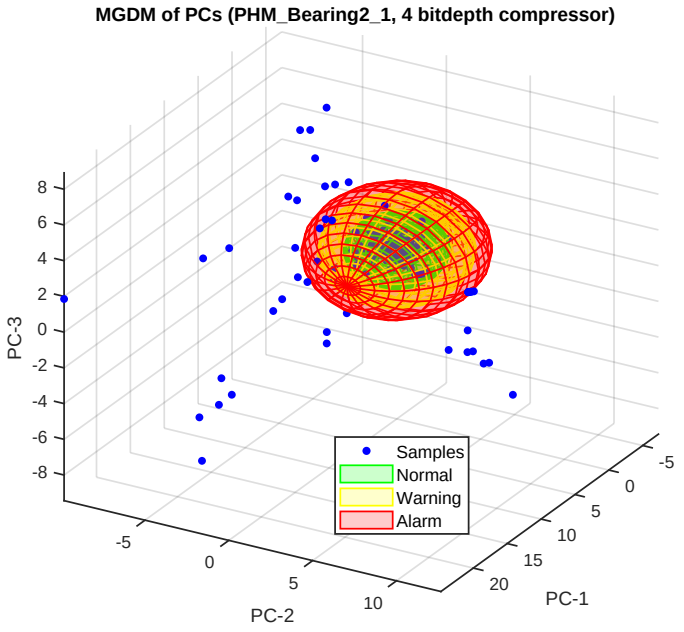
(a) CGoC algorithm results of "NASA_Bearing_1st_test", 4 bitdepth.



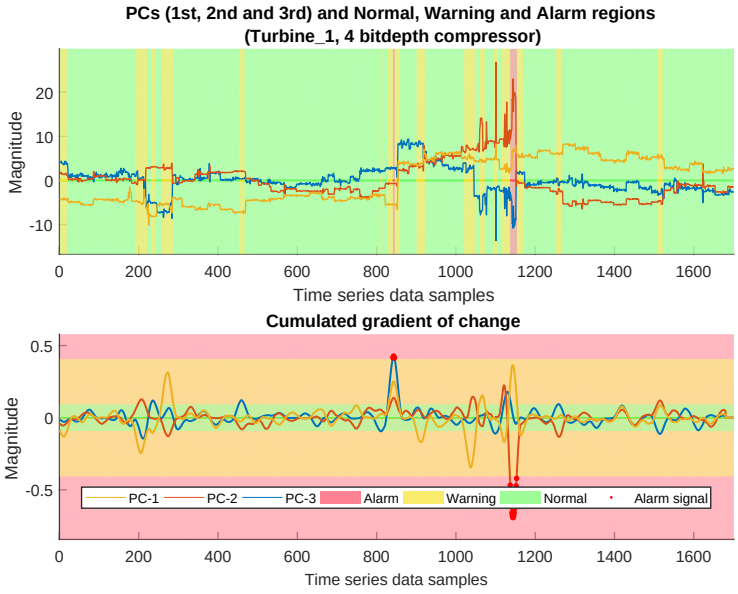
(b) MGDM algorithm results of "NASA_Bearing_1st_test", 4 bitdepth.



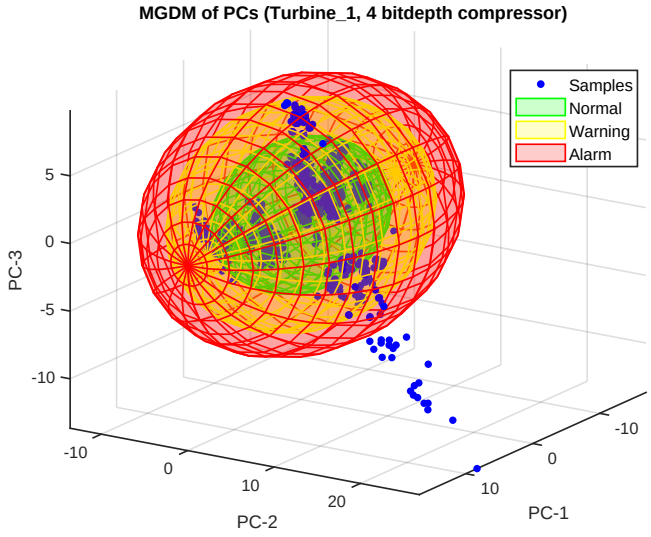
(c) CGoC algorithm results of "PHM_Bearing2_1", 4 bitdepth.



(d) MGDM algorithm results of "PHM_Bearing2_1", 4 bitdepth.



(e) CGoC algorithm results of "Turbine_1", 4 bitdepth.



(f) MGDM algorithm results of "Turbine_1", 4 bitdepth.

Figure 12: Experimental results of the CGoC ((a), (c) and (e)) and MGDM algorithm ((b), (d) and (f)) on compressed vibration signal of "NASA_Bearing_1st_test", "PHM_Bearing2_1" and "Turbine_1" datasets (4 bitdepth).

is capable of achieving robust estimation and modeling results on the raw and compressed vibration signal. Particularly, the experimental results indicate that the proposed method can estimate the trends and model the health status accurately with the distorted/compressed vibration signal which has 26.5 dB signal to noise ratio (SNR) with 14.7:1 compression ratio on average. This enables remote prognostic and condition monitoring applications to be deployed with lower bandwidth and storage requirements. Future work involves the integration of the proposed PCM algorithm in a power generation plant and the extension of the diagnostic algorithm to perform fault classification.

Acknowledgment

This work was supported by the Innovate UK funded project (Project Reference: 102063), Beran Instruments Ltd and EDF Energy.

References

[1] C. R. Farrar, K. Worden, An introduction to structural health monitoring, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 365 (1851) (2007) 303–315. doi:10.1098/rsta.2006.1928.
URL http://rsta.royalsocietypublishing.org/content/365/1851/303

[2] Y. Zhang, P. Hutchinson, N. Lieven, J. Nunez-Yanez, Optimal Compression of Vibration Data with Lifting Wavelet Transform and Context-based Arithmetic Coding, in: European Signal Processing Conference (EUSIPCO), 2017.

[3] G. J. Vachtsevanos, F. Lewis, A. Hess, B. Wu, Intelligent fault diagnosis and prognosis for engineering systems, Wiley Online Library, 2006, bibtex: vachtsevanos2006intelligent.

[4] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, J. Basic Eng 82 (1) (1960) 35–45. doi:10.1115/1.3662552.
URL http://dx.doi.org/10.1115/1.3662552

[5] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systemsReviews, methodology and applications, Mechanical Systems and Signal Processing 42 (1) (2014) 314–334. doi:10.1016/j.ymssp.2013.06.004.
URL http://www.sciencedirect.com/science/article/pii/S0888327013002860

[6] D. An, N. H. Kim, J.-H. Choi, Practical options for selecting data-driven or physics-based prognostics algorithms with reviews, Reliability Engineering & System Safety 133 (Supplement C) (2015) 223–236. doi:10.1016/j.ress.2014.09.014.
URL http://www.sciencedirect.com/science/article/pii/S0951832014002245

[7] H. Hanachi, J. Liu, A. Banerjee, Y. Chen, A. Koul, A Physics-Based Modeling Approach for Performance Monitoring in Gas Turbine Engines, IEEE Transactions on Reliability 64 (1) (2015) 197–205. doi:10.1109/TR.2014.2368872.

[8] J. Yu, Health Degradation Detection and Monitoring of Lithium-Ion Battery Based on Adaptive Learning Method, IEEE Transactions on Instrumentation and Measurement 63 (7) (2014) 1709–1721. doi:10.1109/TIM.2013.2293234.

[9] X.-S. Si, W. Wang, M.-Y. Chen, C.-H. Hu, D.-H. Zhou, A degradation path-dependent approach for remaining useful life estimation with an exact and closed-form solution, European Journal of Operational Research 226 (1) (2013) 53–66. doi:10.1016/j.ejor.2012.10.030.
URL http://www.sciencedirect.com/science/article/pii/S0377221712007953

[10] H. Dui, S. Si, M. J. Zuo, S. Sun, Semi-Markov Process-Based Integrated Importance Measure for Multi-State Systems, IEEE Transactions on Reliability 64 (2) (2015) 754–765. doi:10.1109/TR.2015.2413031.

[11] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, G. Tripot, A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models, IEEE Transactions on Reliability 61 (2) (2012) 491–503. doi:10.1109/TR.2012.2194177.

[12] A. Soualhi, H. Razik, G. Clerc, D. D. Doan, Prognosis of Bearing Failures Using Hidden Markov Models and the Adaptive Neuro-Fuzzy Inference System, IEEE Transactions on Industrial Electronics 61 (6) (2014) 2864–2874. doi:10.1109/TIE.2013.2274415.

[13] Y. Wang, Y. Peng, Y. Zi, X. Jin, K. L. Tsui, A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem, IEEE Transactions on Industrial Informatics 12 (3) (2016) 924–932. doi:10.1109/TII.2016.2535368.

[14] M. Rigamonti, P. Baraldi, E. Zio, D. Astigarraga, A. Galarza, Particle Filter-Based Prognostics for an Electrolytic Capacitor Working in Variable Operating Conditions, IEEE Transactions on Power Electronics 31 (2) (2016) 1567–1575. doi:10.1109/TPEL.2015.2418198.

[15] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing 50 (2) (2002) 174–188. doi:10.1109/78.978374.

[16] E. Zio, G. Peloni, Particle filtering prognostic estimation of the remaining useful life of nonlinear components, Reliability Engineering & System Safety 96 (3) (2011) 403–409. doi:10.1016/j.ress.2010.08.009.
URL http://www.sciencedirect.com/science/article/pii/S0951832010002152

[17] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A Model-Based Method for Remaining Useful Life Prediction of Machinery, IEEE Transactions on Reliability 65 (3) (2016) 1314–1326. doi:10.1109/TR.2016.2570568.

[18] J. Yu, Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring, Mechanical Systems and Signal Processing 83 (Supplement C) (2017) 149–162. doi:10.1016/j.ymssp.2016.06.004.
URL http://www.sciencedirect.com/science/article/pii/S0888327016301674

[19] Z. Xu, Y. Ji, D. Zhou, Real-time Reliability Prediction for a Dynamic System Based on the Hidden Degradation Process Identification, IEEE Transactions on Reliability 57 (2) (2008) 230–242. doi:10.1109/TR.2008.916882.

[20] J. Li, Z. Wang, Y. Zhang, H. Fu, C. Liu, S. Krishnaswamy, Degradation data analysis based on a generalized Wiener process subject to measurement error, Mechanical Systems and Signal Processing 94 (Supplement C) (2017) 57–72. doi:10.1016/j.ymssp.2017.02.031.
URL http://www.sciencedirect.com/science/article/pii/S088832701730095X

[21] C. Duan, V. Makis, C. Deng, An integrated framework for health measures prediction and optimal maintenance policy for mechanical systems using a proportional hazards model, Mechanical Systems and Signal Processing 111 (2018) 285–302. doi:10.1016/j.ymssp.2018.02.029.
URL http://www.sciencedirect.com/science/article/pii/S0888327018300876

[22] Y. Lei, J. Lin, M. J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: A review, Measurement 48 (Supplement C) (2014) 292–305. doi:10.1016/j.measurement.2013.11.012.

URL http://www.sciencedirect.com/science/article/pii/S0263224113005551

[23] Y. Wang, J. Xiang, R. Markert, M. Liang, Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications, Mechanical Systems and Signal Processing 66 (2016) 679–698. doi:10.1016/j.ymssp.2015.04.039.
URL http://www.sciencedirect.com/science/article/pii/S0888327015002897

[24] J. Wang, J. Xie, R. Zhao, K. Mao, L. Zhang, A New Probabilistic Kernel Factor Analysis for Multisensory Data Fusion: Application to Tool Condition Monitoring, IEEE Transactions on Instrumentation and Measurement 65 (11) (2016) 2527–2537. doi:10.1109/TIM.2016.2584238.

[25] M. Rbillat, O. Hmad, F. Kadri, N. Mechbal, Peaks Over Thresholdbased detector design for structural health monitoring: Application to aerospace structures, Structural Health Monitoring (2017) 1475921716685039doi:10.1177/1475921716685039.
URL http://journals.sagepub.com/doi/abs/10.1177/1475921716685039

[26] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals, Sensors 17 (2) (2017) 425. doi:10.3390/s17020425.
URL http://www.mdpi.com/1424-8220/17/2/425

[27] L. Ren, J. Cui, Y. Sun, X. Cheng, Multi-bearing remaining useful life collaborative prediction: A deep learning approach, Journal of Manufacturing Systems 43 (2017) 248–256. doi:10.1016/j.jmsy.2017.02.013.
URL http://www.sciencedirect.com/science/article/pii/S0278612517300262

[28] Y. Zhang, B. Tang, Y. Han, L. Deng, Bearing performance degradation assessment based on time-frequency code features and SOM network, Meas. Sci. Technol. 28 (4) (2017) 045601. doi:10.1088/1361-6501/aa56c9.
URL http://stacks.iop.org/0957-0233/28/i=4/a=045601

[29] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, Deep Learning and Its Applications to Machine Health Monitoring: A Survey, arXiv:1612.07640 [cs, stat]ArXiv: 1612.07640.
URL http://arxiv.org/abs/1612.07640

[30] J. Igba, K. Alemzadeh, C. Durugbo, E. T. Eiriksson, Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes, Renewable Energy 91 (2016) 90–106. doi:10.1016/j.renene.2016.01.006.
URL http://www.sciencedirect.com/science/article/pii/S0960148116300064

[31] H. Qiu, J. Lee, J. Lin, G. Yu, Robust performance degradation assessment methods for enhanced rolling element bearing prognostics, Advanced Engineering Informatics 17 (3) (2003) 127–140. doi:10.1016/j.aei.2004.08.001.
URL http://www.sciencedirect.com/science/article/pii/S1474034604000114

[32] A. Weckenmann, X. Jiang, K. D. Sommer, U. Neuschaefer-Rube, J. Seewig, L. Shaw, T. Estler, Multisensor data fusion in dimensional metrology, CIRP Annals 58 (2) (2009) 701–721. doi:10.1016/j.cirp.2009.09.008.
URL http://www.sciencedirect.com/science/article/pii/S0007850609001759

[33] V. Boln-Canedo, N. Snchez-Maroo, A. Alonso-Betanzos, Recent advances and emerging challenges of feature selection in the context of big data, Knowledge-Based Systems 86 (Supplement C) (2015) 33–45. doi:10.1016/j.knosys.2015.05.014.

URL http://www.sciencedirect.com/science/article/pii/S0950705115002002

[34] A. Malhi, R. X. Gao, PCA-based feature selection scheme for machine defect classification, IEEE Transactions on Instrumentation and Measurement 53 (6) (2004) 1517–1525. doi:10.1109/TIM.2004.834070.

[35] L. E. Mujica, J. Veh, M. Ruiz, M. Verleysen, W. Staszewski, K. Worden, Multivariate statistics process control for dimensionality reduction in structural assessment, Mechanical Systems and Signal Processing 22 (1) (2008) 155–171. doi:10.1016/j.ymssp.2007.05.001.
URL http://www.sciencedirect.com/science/article/pii/S0888327007000751

[36] L. Mujica, J. Rodellar, A. Fernndez, A. Gemes, Q-statistic and T2-statistic PCA-based measures for damage assessment in structures, Structural Health Monitoring: An International Journal 10 (5) (2011) 539–553. doi:10.1177/1475921710388972.
URL http://journals.sagepub.com/doi/10.1177/1475921710388972

[37] B. Schlkopf, A. Smola, K.-R. Mller, Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Neural Computation 10 (5) (1998) 1299–1319. doi:10.1162/089976698300017467.
URL https://doi.org/10.1162/089976698300017467

[38] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, I.-B. Lee, Nonlinear process monitoring using kernel principal component analysis, Chemical Engineering Science 59 (1) (2004) 223–234. doi:10.1016/j.ces.2003.09.012.
URL http://www.sciencedirect.com/science/article/pii/S0009250903004652

[39] M. Kordestani, A. Zanj, M. E. Orchard, M. Saif, A Modular Fault Diagnosis and Prognosis Method for Hydro-Control Valve System Based on Redun-

dancy in Multisensor Data Information, IEEE Transactions on Reliability (2018) 1–12doi:10.1109/TR.2018.2864706.

[40] W. A. Pearlman, A. Islam, N. Nagaraj, A. Said, Efficient, low-complexity image coding with a set-partitioning embedded block coder, IEEE Transactions on Circuits and Systems for Video Technology 14 (11) (2004) 1219–1235. doi:10.1109/TCSVT.2004.835150.

[41] Q. He, F. Kong, R. Yan, Subspace-based gearbox condition monitoring by kernel principal component analysis, Mechanical Systems and Signal Processing 21 (4) (2007) 1755–1772. doi:10.1016/j.ymssp.2006.07.014. URL http://www.sciencedirect.com/science/article/pii/S0888327006001695

[42] E. Sutrisno, H. Oh, A. S. S. Vasan, M. Pecht, Estimation of remaining useful life of ball bearings using data driven methodologies, in: 2012 IEEE Conference on Prognostics and Health Management, 2012, pp. 1–7. doi:10.1109/ICPHM.2012.6299548.

[43] A. Bellino, A. Fasana, L. Garibaldi, S. Marchesiello, PCA-based detection of damage in time-varying systems, Mechanical Systems and Signal Processing 24 (7) (2010) 2250–2260. doi:10.1016/j.ymssp.2010.04.009. URL http://www.sciencedirect.com/science/article/pii/S0888327010001159

[44] R. J. Hodrick, E. C. Prescott, Postwar U.S. Business Cycles: An Empirical Investigation, Journal of Money, Credit and Banking 29 (1) (1997) 1–16. doi:10.2307/2953682.
URL http://www.jstor.org/stable/2953682

[45] A. Maravall, A. del Ro, TIME AGGREGATION AND THE HODRICK-PRESCOTT FILTER, Banco de Espana Working Paper Series 08.

[46] M. O. Ravn, H. Uhlig, On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations, Review of Economics and Statistics 84 (2)

(2002) 371–376. doi:10.1162/003465302317411604.
URL http://www.mitpressjournals.org/doi/10.1162/003465302317411604

[47] A. Maravall, A. del Ro, Temporal aggregation, systematic sampling, and the HodrickPrescott filter, Computational Statistics & Data Analysis 52 (2) (2007) 975–998. doi:10.1016/j.csda.2007.08.001.
URL http://www.sciencedirect.com/science/article/pii/S0167947307002976

[48] V. J. Hodge, J. Austin, A Survey of Outlier Detection Methodologies, Artif Intell Rev 22 (2) (2004) 85–126. doi:10.1007/s10462-004-4304-y.
URL https://doi.org/10.1007/s10462-004-4304-y

[49] A. C. Rencher, Methods of Multivariate Analysis, John Wiley & Sons, 2003, google-Books-ID: SpvBd7IUCxkC.

[50] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of Sound and Vibration 289 (4) (2006) 1066–1090. doi:10.1016/j.jsv.2005.03.007.
URL http://www.sciencedirect.com/science/article/pii/S0022460X0500221X

[51] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, PRONOSTIA : An experimental platform for bearings accelerated degradation tests., in: IEEE International Conference on Prognostics and Health Management, PHM'12., Vol. sur CD ROM, IEEE Catalog Number : CPF12PHM-CDR, Denver, Colorado, United States, 2012, pp. 1–8, http://www.femto-st.fr/en/Research-departments/AS2M/Research-groups/PHM/IEEE-PHM-2012-Data-challenge.php.
URL https://hal.archives-ouvertes.fr/hal-00719503

Table 1: Features extraction

	Feature Name	Formula
Time-domain	Maximum	$x_{\max} = \max(x_t)$
	Mean	$\mu_x = \frac{1}{T}\sum_{t=0}^{T} x_t$
	RMS	$x_{\text{RMS}} = \left[\frac{1}{T}\sum_{t=0}^{T} x_t^2\right]^{1/2}$
	Peak-to-peak value	$x_p = \max(x_t) - \min(x_t)$
	Square mean root	$x_r = \left[\frac{1}{T}\sum_{t=0}^{T}[\lvert x_t\rvert]^{1/2}\right]^2$
	Standard deviation	$\sigma_x = \left[\frac{1}{T}\sum_{t=0}^{T}\left[x_t - \mu_x\right]^2\right]^{1/2}$
	Kurtosis	$\beta_x = \frac{1}{T}\sum_{t=0}^{T}\left[x_t - \mu_x\right]^4$
	Skewness	$S = E\left[(\frac{x_t - \mu_x}{\sigma_x})^3\right]$
	Crest factor	$C = \frac{x_p}{x_{\text{RMS}}}$
	Entropy	$H = -\sum_{t=0}^{T} p(x_t)\ \log_{10} p(x_t)$
Time-Frequency domain	Wavelet energy	$E_{\text{WT}} = \frac{1}{T}\sum_{i=0}^{T} wt_i^2(t)$
	Wavelet energy ratio	$R_{\text{WT}} = E_{\text{WT}_\varphi}/E_{\text{WT}_{LL}}$ (9 features)
	Relative entropy	$Div(E^c \parallel E^b) = \sum_{i=1}^{\varphi} E_i^c \log\frac{E_i^c}{E_i^b}$

where x_t, wt, φ, E_{WT} and Div denotes the sampled point in raw vibration signal series, the wavelet coefficients, the wavelet sub-band, wavelet energy and the Kullback-Leibler Divergence for representing the relative entropy of different wavelet sub-bands, respectively.