Mechanical Systems and Signal Processing 199 (2023) 110446

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Latent indicators for temporal-preserving latent variable models in vibration-based condition monitoring under non-stationary conditions

Ryan Balshaw *, P. Stephan Heyns, Daniel N. Wilke, Stephan Schmidt

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa

ARTICLE INFO

Communicated by J.E. Mottershead

Keywords:
Unsupervised learning
Temporal preservation
Latent variable models
Latent health indicators
Condition monitoring

ABSTRACT

Condition-based monitoring for critical assets is reliant on the quality of the indicators used for condition inference. These indicators must be sensitive to the development of faults under constant and non-stationary operating conditions. Latent variable models in the time-preserving framework offer a powerful learning-based technique for the monitoring of critical assets as they only require healthy asset data, and provide indicators from the data space for asset condition inference. However, the latent manifold of latent variable models is often disregarded in favour of data space indicators, such as reconstruction errors, which are primarily focused on measuring the likelihood of the observed data. The latent space is often unintentionally unutilised. In this work, we highlight that the latent manifold is a powerful resource for condition inference and should be utilised for condition monitoring. We conceptually categorise and identify five classes of latent space health indicators that capture various manifold perspectives. These five classes introduce and allow for latent health indicator derivation and are useful for the condition inference task. Fifteen latent health indicators are considered in this work and are applied to the fault diagnostics task on two experimental datasets. An ensemble-based inference procedure is used, which produces a modular fault diagnosis framework. The indicators are shown to be informative for condition inference in both constant and variable operating conditions. Utilising the latent manifold in condition monitoring tasks is important to further develop the learning-based condition monitoring field.

1. Introduction

The demand for asset reliability and production consistency is ever-growing in industry. This demand is amplified for sectors whose clientèle satisfaction is contingent on consistent public utility delivery. Condition monitoring (CM) procedures are designed to ensure that assets have minimal downtime by using health indicators (HIs) to infer an asset's health state for effective asset maintenance. To provide an indication of the health state at any given time, data must be available that reflect an asset's internal state, as asset components can rarely be inspected visually [1]. The techniques used to obtain data is commonly vibration analysis or lubricant analysis, where the former is more prevalent in CM as richer diagnostic information can be inferred [1]. CM procedures rely on HIs, and this reliance has produced a significant research drive dedicated to the design of HIs that are rich in asset condition information [2]. However, non-stationary operating conditions (OCs) increase the complexity of the condition inference problem, which may hinder the performance of the HIs [3]. Stationary OCs imply that the properties of the observed data

* Corresponding author.
E-mail address: ryanbalshaw1@gmail.com (R. Balshaw).

https://doi.org/10.1016/j.ymssp.2023.110446
Received 15 December 2022; Received in revised form 20 March 2023; Accepted 10 May 2023
Available online 30 May 2023

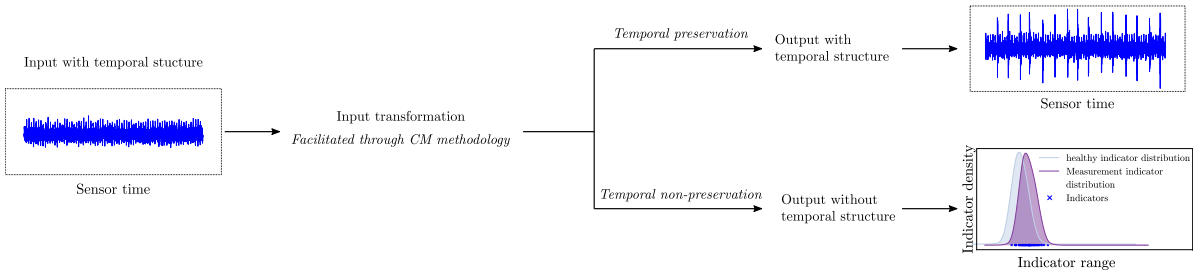


Fig. 1. The identification of vibration-based CM methodologies as temporally preserving or temporally non-preserving. The distinction, between temporal preservation and temporal non-preservation, irrespective of the methodology used, is based off the outputs of the methodology.

remain approximately constant through time. Non-stationary behaviour in the OCs, typically caused by variations in speed, load, or temperature, invoke a change in the observed vibration data through amplitude modulation and phase modulation. Faults in vibration-based CM applications are much the same and manifest through similar mechanisms. Hence, non-stationary OCs impede diagnosis as the instantaneous health state of an asset is entangled with information unrelated to the underlying asset condition. Two important aspects in CM are fault diagnostics and prognostics, and both aspects are dependent on the quality of the HIs used as representations of an asset's state [4,5].

There are two frameworks that CM techniques originate from, namely the *(i)* signal processing [6,7], and the *(ii)* learning-based frameworks [8,9]. Note that the learning-based framework encapsulates statistical learning, machine learning, and deep learning. While the dominant class of learning-based techniques are usually driven using classification or regression-based objectives, these objectives all rely on access to labelled fault data [9–11]. Latent variable models (LVMs) are a promising class of unsupervised learning-based techniques that model the physical representation of an asset. In the CM setting, LVMs learn a probabilistic representation of an asset to detect deviations from a reference asset state [12].

The temporal preservation approach for LVMs proposed by Balshaw et al. [13] ensures that the temporal characteristics of the vibration data are preserved in the HIs obtained from LVMs. The temporal preservation approach advocates that the sensor time, i.e., the time within a measured signal, must be preserved for informative LVM-based condition inference. Temporal preservation is the mechanism from which LVMs become applicable under constant and non-stationary OCs. LVMs facilitate a direct transformation of the observed vibration data into indicators, and temporal preservation provides temporal structure therein. Without temporal structure, extraction of fault feature information is limited as sensor time is effectively entangled within the latent manifold [13]. The fundamentals of temporal preservation are demonstrated in Fig. 1 for any methodology that performs a transformation to time-series data. If the output of the input data transformation contains temporal structure, i.e., sensor time is preserved in the output of the transformation, then the methodology, by design, performs temporal preservation. However, if the output is a single scalar or, equivalently, a collection of temporally non-preserving scalars, then the methodology removes the temporal structure present in the input data and is temporally non-preserving.

Fig. 2 compares signal processing and conventional learning-based approaches using the notion of temporal preservation and non-preservation. In Fig. 2, signal processing methods that are temporally non-preserving refer to methods that collapse sensor time and compute vibration data features at a record time level, e.g. Večeř et al. [14]. In contrast, signal processing methods with temporal preservation typically produce a transformed time-series signal with the fault covariates enhanced, e.g. Sawalhi et al. [15]. In the learning-based domain, temporally non-preserving methods produce indicator signals that are devoid of any temporal content, e.g. Booyse et al. [12], whereas methods that are temporally preserving produce indicator signals with temporal structure, e.g. Balshaw et al. [13]. Learning-based approaches are distinct as signal processing approaches rely on expert knowledge for signal transformation design, whereas learning-based approaches use data-driven models to facilitate the signal transformation. However, this does not imply that the two approaches are mutually exclusive, e.g. Wang et al. [16].

The focus of this study is orientated towards temporally structured learning-based approaches. Indicators from learning-based techniques require sensor time preservation to elevate the benefit they provide in vibration-based condition monitoring applications. However, unlike typical signal processing or learning-based research, which is focused on the development of state-of-the-art transformation methodologies, this work proposes a set of temporally preserved latent condition indicators. In this way, instead of limiting inference decisions to a single or sparse set of indicator signals, the information from a dense set of indicator signals is used for condition inference. Without preserving the temporal structure, the full potential of the LVM cannot be realised. The uniqueness of this work is that a temporally structured latent space enables a richer condition inference procedure as the best facets of LVMs can be used.

By preserving sensor time, learning-based indicator signals are obtained that capture the fault covariates in the vibration data. The temporal preservation approach not only preserves time within HIs, it also preserves and disentangles sensor time within the latent manifold [13]. In Fig. 3, the importance of using temporal preservation for the latent manifold is revealed. Without sensor time, as demonstrated through Fig. 3(a), there is no noticeable distinction between two faults, one localised and one distributed. By preserving sensor time, shown in Fig. 3(b), the distinctions between the two fault representations are clear and unique. The temporal characteristics of the manifold representation of the observed vibration data may then be exploited for condition inference.

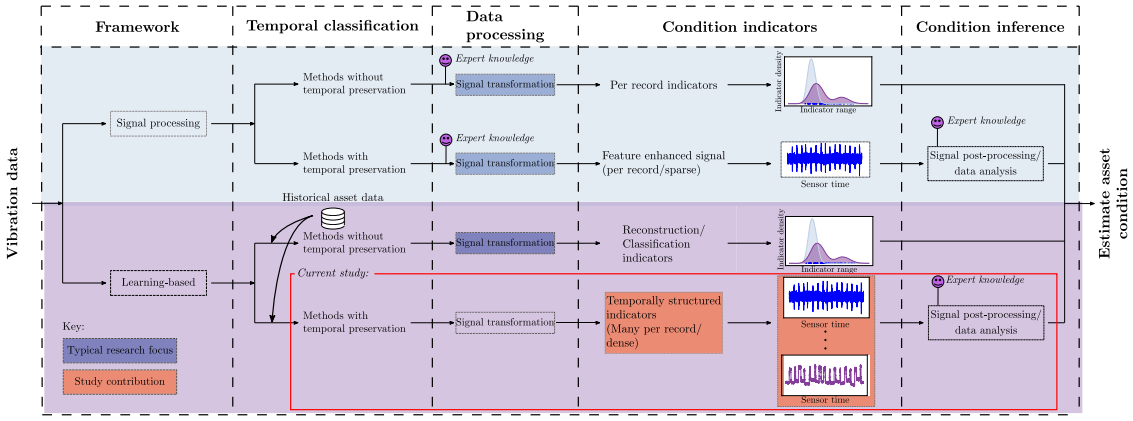
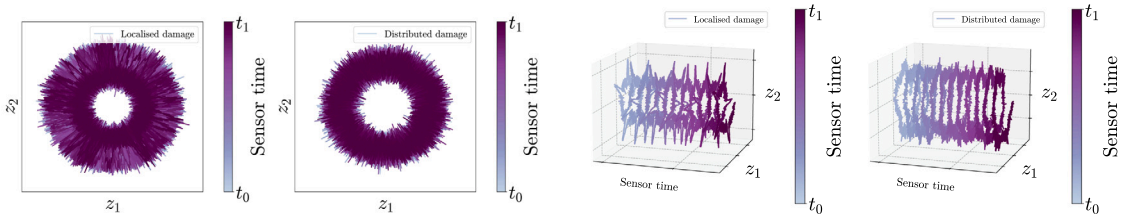


Fig. 2. The condition monitoring process for signal processing-based and learning-based approaches. The fundamental difference between techniques from the two approaches is driven through the notion of temporal preservation. This work uses temporally preserving LVMs and is focused on enriching the condition inference stage through an ensemble of latent HIs.



(a) Manifold without temporal preservation. (b) Manifold with temporal preservation.

Fig. 3. A pictorial demonstration of the latent manifold for two types of vibration data faults with and without sensor time preservation. In (a), the lack of temporal structure corrupts the inference process and there is little distinction between the fault types. In (b), the addition of temporal structure uncovers the clear fault distinctions in the manifold.
Source: Adapted from [13].

In the context of temporal preservation, the characteristics of signal processing-based or learning-based HI signals are used for condition inference. For the former, the methodology used to obtain a HI signal typically relies on the use of expert knowledge, e.g. Antoni [7], and physical information related to the fault mode of the asset of interest, whereas the latter relies on the characteristic of some statistical model [17]. Signal processing-based HIs use curated measures which enhance specific characteristics of vibration data, e.g. non-Gaussianity [18], and signal enhancement methodologies, e.g. demodulation-band selection [19], residual signal analysis [20], discrepancy analysis [21], or blind de-convolution [22], to obtain a fault-informative transformation of vibration data. This enhanced signal may then be transformed into a HI for condition inference. However, such practices require a profound understanding of the problem of interest, rely on the comprehension of many theoretical works, the final HI depends on the enhancement pipeline used, and indicators are typically processed one at a time. Learning-based HIs, in comparison, are identified using a statistical model to capture a representation of an assumed healthy reference state of the asset of interest. In the learning-based HI signal scenario, expert knowledge is removed, multiple representations of the asset reference state are identified, and HIs are used to detect deviations from the asset reference state [12,13]. In the temporally preserved setting, a dense set of HIs may be obtained for each representation of the asset reference state which enables an ensemble-based condition inference framework [13]. However, without temporal preservation, the HIs are limited to anomaly detection supported by a sparser set of HIs [12].

LVMs provide a data space and the latent space for condition inference, and the latent space is underutilised in condition monitoring. LVM-based condition inference techniques predominantly use the observed data space, and this can only provide two HI metrics for condition inference namely, *(i)* the reconstruction error, and *(ii)* the reconstruction likelihood [23]. Booyse et al. [12] used the discriminator of a generative adversarial network (GAN) as the data space's proxy to infer whether newly observed data was from the reference dataset. Gugulothu et al. [24] proposed an embedded latent health indicator (LHI) that uses an Euclidean distance measure to compare a latent representation of an observed data sample to a set of healthy latent representations. Balshaw et al. [13] proposed three LHIs that capture specific properties of the latent manifold. Two of these metrics are only accessible through the proposed temporal preservation approach, and they exploit the additional temporal structure induced in the latent manifold. Kim et al. [25] proposed a reconstruction along the projection pathway (RaPP) approach, which evaluates the Euclidean distance or the Mahalanobis distance between the hidden auto-encoder representations of an observed sample and the hidden auto-encoder representations of the reconstruction of the observed sample. González-Muñiz et al. [26] use the RaPP approach but propose using just the latent space of the auto-encoder for condition inference. Park et al. [27], independently from Kim et al. [25], proposed using

the Manhattan distance between the hidden auto-encoder representations of an observed sample and the reconstruction thereof. Additionally, [27] use the mean-centred cosine similarity for the hidden representations for condition inference. Kwon et al. [28] use a combination of the reconstruction error and a cosine similarity metric applied to the gradients of the auto-encoders. In the current literature, the latent manifold has been inadequately explored and its importance is understated. This is undesirable and must be rectified to fully enable latent condition inference.

In this work, we thoroughly explore the latent manifold of LVMs in a fault diagnostics context. This paper critically investigates the interaction of HIs on the latent space for a linear and nonlinear LVM. To ensure that the condition inference process is accurate, fifteen LHIs are proposed that are representative of the identified conceptual classes. As such, the inference process is no longer based on a single metric but rather an ensemble of metrics. This ensemble-based procedure allows for a modular fault diagnosis framework, whereby the indicators are interchangeable for any given dataset. The latent manifold is a subspace of the data manifold that captures the intrinsic structure of the data. This space can be used for condition inference, and a plethora of methodologies exist from which LHIs may be obtained. Exploration of the latent space is nontrivial, and thus scalar measures are required to characterise changes within the latent manifold. Previous research into LHIs has sidestepped any explicit measure formalisation and does not provide details on how to derive LHIs. Thus, we seek to make this formalisation explicit.

The main contributions of this research are summarised as follows:

1. Introduce and formalise conceptual categories that describe the latent manifold treatment perspectives for LHI derivation within the time preserved framework. These categories allow one to easily obtain scalar metrics and enrich LVM-based CM.
2. Consider and emphasise the importance of the LVM data space and the latent space for condition inference in a CM context using a time preservation approach.
3. Showcase the responsiveness of the latent manifold to anomalous data in the presence of constant and non-stationary OCs. Metrics from the proposed conceptual categories are used to clearly showcase how the latent manifold is indicative of damage, and that the manifold is useful to detect faults.
4. Using two experimental datasets, whereby each dataset captures a different operating condition class, the benefit of LVMs, as brought forward by the temporal preservation approach and the LHIs from the proposed conceptual classes, is made evident for CM applications.

To highlight the usefulness and responsiveness of the latent manifold to anomalous fault data, we consider two LVM formulations. The LVM formulations of interest are *(i)* probabilistic principal component analysis (PPCA) [29], and *(ii)* variational auto-encoders (VAEs) [30]. The choice of formulation is driven by methodologies that make explicit and tractable model density assumptions, as the focus of this work is on the latent manifold and how damage manifests therein, and thus these methods reflect different manifold considerations under a shared model density assumption. Both the PPCA and VAEs models are explicit density estimation methods which use Gaussian distributions [30,31]. The latent manifold in PPCA has a linear subspace, while VAEs utilise nonlinear parametric functions, often referred to as neural networks, to obtain a subspace that may be embedded nonlinearly within the observed data.

The layout of this paper is as follows: In Section 2 the important literature surrounding LVMs is discussed, the usage of LVMs for vibration data is formalised, and the conceptual categories for LHI metric derivation are introduced. In Section 3 the HIs and proposed LHIs are used in conjunction with PPCA and VAE models to evaluate the HI and LHI metric performance on experimental data. Lastly, Section 4 concludes the work and recommendations for future work are given.

2. Latent manifold analysis

In this work, LVMs act as a representation transformer that take vibration data and transform it to a set of HIs and LHIs. These HIs and LHIs serve as indicators from which the asset state can be inferred. The LVM itself is developed on a reference set of data, and this data is assumed as representative of a healthy asset state. The objective of this work is to improve the utilisation of the latent manifold for CM by developing LHI conceptual classes. These LHI classes introduce and capture different manifold perspectives, where each perspective can produce LHIs for condition inference. To detail the methods used to detect damage in the latent manifold, the concepts behind LVMs must first be introduced.

2.1. Scope of work

For the succeeding subsections, a scope of work is included here to enable readers with sufficient knowledge of LVMs to bypass sections that contain literature related to the concepts and the use of LVMs in a condition monitoring setting. Section 2.2 details the general concepts and literature surrounding LVMs, Section 2.3 describes the parametrisation process for explicit (assumed Gaussian) LVMs in a linear and nonlinear setting, Section 2.4 describes how vibration data is pre-processed for LVM training and details how temporal preservation is enabled during model evaluation, Section 2.5 introduces and discusses the LHI classes for latent manifold condition inference, and Section 2.6 details the HIs and LHIs used in this work.

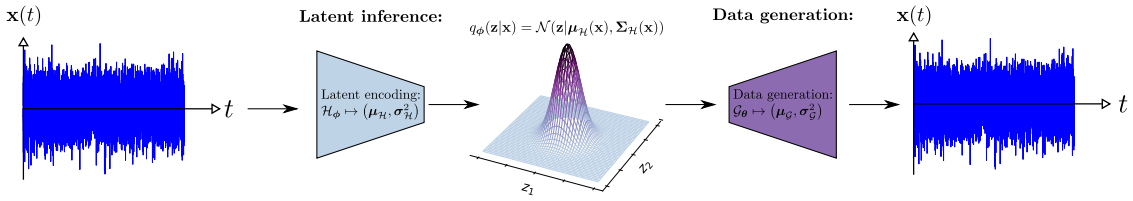


Fig. 4. A visual LVM illustration of how the posterior and generative distribution interact. This approach is simple, does not take any vibration data pre-processing into account, and primarily serves as visual demonstration of how observed data interacts with different aspects of the LVM.

2.2. Latent variable models

LVMs are the foundation of unsupervised learning-based techniques [30]. LVMs are generative modelling techniques and represent a class of statistical models concerned with capturing the distribution over some observed variable set $\mathbf{x}$, where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D$ is a D-dimensional observed random variable [31,32]. As the data distribution is assumed to be unknown, LVMs assume that a set of unobserved random variables $\mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^d$, where $d \leq D$, has a hidden relationship with the observed variables $\mathbf{x}$, and they describe the intrinsic nature of the data [31,32]. The latent variable set is referred to as the latent manifold [33]. This assumption gives rise to the generative process that describes how $\mathbf{z}$ forms part of the model [32]. In the CM setting, LVMs learn a probabilistic representation of an asset to detect deviations from a reference asset state [12]. Under the assumption that the reference state captures an asset in a healthy condition and under the assumption of no access to labelled vibration data [9], the detected deviations are attributed to anomalies in the asset due to a change in asset condition. A LVM defines a joint probability distribution

$$p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z}), \tag{1}$$

where $p(\mathbf{z})$ is the latent prior distribution and $p(\mathbf{x}|\mathbf{z})$ is the generative conditional distribution. Note that the factorisation of the joint distribution in Eq. (1) defines the generative process of LVMs [32]. The marginal distribution $p(\mathbf{x})$, the model evidence, over the observed variables is given by

$$p(\mathbf{x}) = \int_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}, \tag{2}$$

which offers a highly flexible model if $\mathbf{z}$ is continuous [34]. Inference is an important aspect of LVMs and the aforementioned generative process is only concerned with sample generation. To infer the latent variables for a data sample $\mathbf{x}$, we can use Bayes' theorem to compute the latent posterior distribution $p(\mathbf{z}|\mathbf{x})$ through

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}. \tag{3}$$

LVMs, in practice, utilise linear or nonlinear parametric functions to parametrise the distributions of interest, namely the conditional, posterior, and prior distributions [32]. Typically the prior distribution is assumed to be a simple, fixed distribution and the parametric distributions become $p(\mathbf{x}|\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$. The parametric functions of the aforementioned distributions are deterministic and commonly denoted as

$$\mathcal{G}_{\boldsymbol{\theta}} : \mathcal{Z} \mapsto \mathcal{X}, \tag{4}$$

$$\mathcal{H}_{\boldsymbol{\phi}} : \mathcal{X} \mapsto \mathcal{Z}, \tag{5}$$

where the transition functions $\mathcal{G}$ and $\mathcal{H}$ have parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ respectively. These functions are used to parametrise the generative distribution $p(\mathbf{x}|\mathbf{z})$ and posterior distribution $p(\mathbf{z}|\mathbf{x})$. In Fig. 4 an example of the transition functions is shown.

There are two main types of LVMs, namely *(i)* explicit (likelihood-based) models, and *(ii)* implicit models [35]. Explicit models perform maximum likelihood estimation using the marginal likelihood function to optimise the unknown parameters of the transition functions $\mathcal{G}$ and $\mathcal{H}$. Models in this category require explicit distribution parametrisation. The choice of transition function form is critical for explicit models, as linear functions may allow for exact posterior computation, whereas nonlinear functions, in general, require approximate inference techniques to optimise the model parameters [31,32,36]. Typical examples of explicit models include PPCA [29], VAEs [30], normalising flows [37,38], and energy-based models [39]. Both the PPCA and VAE LVM formulations maximise the marginal log-likelihood lower bound and consist of non-invertible transition functions, while normalising flows maximise the exact log-likelihood and use network architecture constraints to produce bijective transition functions. Implicit models, alternatively, do not parametrise the distributions of interest explicitly, and represent the data distribution of interest as a model of a latent sampling process [40]. This formulation requires the use of the density estimation by comparison techniques, such as the density ratio $\frac{p(\mathbf{x})}{q(\mathbf{x})}$, to effectively capture $p(\mathbf{x})$, where $q(\mathbf{x})$ is the distribution captured by the implicit LVM [40]. The most prominent implicit model is a GAN [41]. The caveat here is that most implicit models focus on sample generation through $\mathcal{G}_\theta$, and neglect the inference stage of LVMs. Bond-Taylor et al. [42] provide a succinct review on deep generative modelling techniques.

2.3. Explicit density estimation

Explicit LVMs are integral to this work to ensure that the latent manifold is accessible. The generative conditional and latent prior distributions are assumed to be Gaussian distributions of the form

$$p_\theta(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_\mathcal{G}(\mathbf{z}), \boldsymbol{\Sigma}_\mathcal{G}(\mathbf{z})), \tag{6}$$

where the covariance of the generative conditional distribution is given by

$$\boldsymbol{\Sigma}_\mathcal{G}(\mathbf{z}) = \text{diag}(\boldsymbol{\sigma}^2_\mathcal{G}(\mathbf{z})), \tag{7}$$

where $\text{diag}(\cdot)$ is the vector-to-matrix diag operator. The prior is a unit isotropic Gaussian of the form

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I}). \tag{8}$$

Finally, the generative transition function is expressed as

$$\left(\boldsymbol{\mu}_\mathcal{G}(\mathbf{z}), \boldsymbol{\sigma}^2_\mathcal{G}(\mathbf{z})\right) = \mathcal{G}_\theta(\mathbf{z}). \tag{9}$$

Using the parametrised distributions, maximum likelihood of the marginal distribution detailed in Eq. (2) can be used to optimise the parameters of $\mathcal{G}_\theta$. However, depending on whether $\mathcal{G}_\theta$ is linear or nonlinear, the integral in Eq. (2) may be intractable as it is a realisation of all possible values of $\mathbf{z}$, which may be overtly complicated to compute and possibly mathematically infeasible. PPCA is an LVM with a tractable solution, and thus the latent posterior distribution can be tractably computed [29]. The latent posterior distribution in PPCA is also a Gaussian distribution. However, VAEs commonly use neural networks for increased generative model flexibility. This decision causes Eq. (2) to be intractable, and thus the posterior distribution cannot be determined [30,34]. To overcome this issue, a parametric inference model $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ is introduced to approximate $p(\mathbf{z}|\mathbf{x})$ and a lower bound on the marginal distribution, commonly referred to as the evidence lower bound (ELBO) is maximised to find the optimal model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ [30]. The ELBO can be found through direct expansion of Eq. (2) or through variational inference techniques [34,36].

Regardless of the tractability of the process to obtain the posterior distribution, both the aforementioned LVMs use a latent encoding transition function $\mathcal{H}_{\boldsymbol{\phi}}$ to parametrise a Gaussian posterior distribution

$$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_\mathcal{H}(\mathbf{x}), \boldsymbol{\Sigma}_\mathcal{H}(\mathbf{x})), \tag{10}$$

where the latent covariance $\boldsymbol{\Sigma}_\mathcal{H}$ is given by

$$\boldsymbol{\Sigma}_\mathcal{H}(\mathbf{x}) = \text{diag}(\boldsymbol{\sigma}^2_\mathcal{H}(\mathbf{x})). \tag{11}$$

Finally, the latent transition function is expressed as

$$\left(\boldsymbol{\mu}_\mathcal{H}(\mathbf{x}), \boldsymbol{\sigma}^2_\mathcal{H}(\mathbf{x})\right) = \mathcal{H}_{\boldsymbol{\phi}}(\mathbf{x}). \tag{12}$$

PPCA and VAEs both operate on the same principle, whereby a posterior distribution is inferred from the observed data, and a sample can be generated by sampling the posterior $\mathbf{z} \sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and then sampling the generative distribution $\mathbf{x} \sim p_\theta(\mathbf{x}|\mathbf{z})$. However, the posterior covariance for a PPCA model is not a function of the input data (i.e., $\boldsymbol{\sigma}^2_\mathcal{H}$ is constant), while VAE models allow for this flexibility [29,30].

2.4. Data pre-processing

To use LVMs on asset vibration data, the data must be pre-processed to ensure that it is practical for computational usage. In this work, we reduce the dimensionality of the ith time-series signal $x_i[n]$ using a segmentation approach, where $n \subseteq \mathbb{Z}^+ = \{1, 2, \dots, L_{signal}\}$ is the discrete time variable representative of sensor time, i refers to record time, and L_{signal} is the dimensionality of $\mathbf{x}_i \in \mathbb{R}^{L_{signal}}$ [12,13]. The signal is transformed into a Hankel matrix $\mathbf{H}(\cdot)$ [43] by extracting signal segments from the raw vibration signals. This extraction process is given by

$$\mathbf{H}\left(\mathbf{x}_i\right) = \begin{bmatrix} x_i[1] & x_i[2] & \cdots & x_i[L_w] \\ x_i[L_{sft}] & x_i[L_{sft}+1] & \cdots & x_i[L_{sft}+L_w] \\ x_i[2\cdot L_{sft}] & x_i[2\cdot L_{sft}+1] & \cdots & x_i[2\cdot L_{sft}+L_w] \\ \vdots & \vdots & \ddots & \vdots \\ x_i[L_{sft}\cdot(L_H-1)] & x_i[L_{sft}\cdot(L_H-1)+1] & \cdots & x_i[L_{sft}\cdot(L_H-1)+L_w] \end{bmatrix} \tag{13}$$

where L_w is the window length, L_{sft} is the shift parameter, and $L_H = \lfloor \frac{L_{signal}-L_w}{L_{sft}} \rfloor + 1$ represents the number of rows in $H(\mathbf{x}_i)$. Note that the rows of Eq. (13) now represent the feature vectors that are fed into the LVMs. The data matrix $\overline{\mathbf{X}}_{data} \in \mathbb{R}^{(N\cdot L_H)\times L_w}$ is then developed using N healthy vibration signals through

$$\overline{\mathbf{X}}_{data} = \begin{bmatrix} \mathbf{H}\left(\mathbf{x}_1\right) \\ \vdots \\ \mathbf{H}\left(\mathbf{x}_N\right) \end{bmatrix}. \tag{14}$$

The data matrix $\overline{\mathbf{X}}_{data}$ is then partitioned into a training and validation set that is used for model training. In this study, the shift parameter is selected as $L_{sft} = 0.5L_w$, $\overline{\mathbf{X}}_{data}$ is split using an $80\% - 20\%$ training-validation split process, and the data is standardised using

$$\mathbf{X}_{data} = \left(\overline{\mathbf{X}}_{data} - \mathbf{1}\boldsymbol{\mu}_{train}^{T}\right) \text{diag}\left(\boldsymbol{\sigma}_{train}^{-1}\right), \tag{15}$$

which may be given in index notation as

$$X_{data,ij} = \frac{\overline{X}_{data,ij} - \mu_{train,j}}{\sigma_{train,j}}, \tag{16}$$

where $\boldsymbol{\mu}_{train} \in \mathbb{R}^{L_w}$ and $\boldsymbol{\sigma}_{train} \in \mathbb{R}^{L_w}$ are column vectors of the feature-wise means and standard deviations respectively of $\overline{\mathbf{X}}_{data}$, $\mathbf{1} \in \mathbb{R}^{L_w}$ is a constant vector with elements 1, and $\boldsymbol{\sigma}_{train}^{-1} = [\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_{L_w}}]^T \in \mathbb{R}^{L_w}$. For model evaluation, the temporal preservation approach is used, which effectively re-processes the available data using Eq. (13) with $L_{sft} = 1$ [13].

2.5. Latent manifold indicators

In this work, the latent manifold is used for condition inference. To enable LVMs to be applicable to all facets of fault diagnostics, we incorporate the benefits of the temporal preservation approach proposed in Balshaw et al. [13]. To detect deviations from the learnt healthy state in the latent space, the following five conceptual LHI classes are proposed, namely *(i)* likelihood-based, *(ii)* discriminative, *(iii)* statistical distance, *(iv)* prototypical approaches, and *(v)* multivariate signal processing approaches. These five classes capture different manifold perspectives, provide LHIs for condition inference, and are all methods that detect deviations in the latent manifold. The five classes are discussed in turn. Examples of indicators from each class are given in Table 1.

1. The likelihood-based class refers to techniques which model the latent space through a reference model $p_{ref}(\mathbf{z})$. For latent anomaly detection, the reference model negative log-likelihood is used to provide an indication of whether new data is from the reference latent manifold.
2. The discriminative class of approaches refers to techniques that adopt a discriminative approach to capturing the reference latent manifold. The distribution of interest is the posterior distribution $p(\mathbf{y}|\mathbf{z})$, where $\mathbf{y}$ is a discrete, class variable. However, the labels given to the observed reference latent data only represent one asset state, which is healthy.
3. Statistical distance-based LHIs are obtained whereby each input $\mathbf{x}$ passed through the transition function $\mathcal{H}_{\boldsymbol{\phi}}$ supplies an output that parametrises a latent posterior distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_{\mathcal{H}}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathcal{H}}(\mathbf{x}))$. We then interrogate the properties of this posterior distribution to determine whether it exhibits sensitivity to faulty data. There are two avenues of analysis within this approach, namely *(a)* dissimilarity-based comparisons, and *(b)* distribution similarity comparisons. Each of these avenues will now be discussed.
   (a) The dissimilarity-based sub-class inspects how the mean $\boldsymbol{\mu}_{\mathcal{H}}(\mathbf{x})$ of the posterior distribution changes. This comparison can occur in two ways, one through changes in the mean with respect to the latent prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, or through changes in the mean within a localised time-frame. The latter approach is analogous to a first-order differences approach, with a step size Δt given by $\Delta t = \frac{1}{F_s}$, where F_s is the signal sampling frequency.
   (b) In the distribution similarity sub-class, metrics that measure the similarity between two distributions are used. As with the mean-based sub-class, the distribution comparison may be between the sample posterior $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ and the latent prior $p(\mathbf{z})$, or within a localised time-frame. For the latter, we compare between two posterior distributions $q_{\boldsymbol{\phi}}(\mathbf{z}_t|\mathbf{x}_t)$ and $q_{\boldsymbol{\phi}}(\mathbf{z}_{t+1}|\mathbf{x}_{t+1})$. Note that t indicates the temporally-preserved latent or data representation at any time-point $t \in [0, L_H - 1]$, where $L_H = L_{signal} - L_w$ when temporal preservation is applied.
4. The prototypical class refers to methods that assume that there are a finite number of prototypical elements in the latent space that characterise the data, which defines a prototype manifold [44]. Under the prototype manifold, we can measure the distance between the prototype vectors and $\boldsymbol{\mu}_{\mathcal{H}}(\mathbf{x})$ to obtain a health indicator.
5. In the multivariate signal processing class, each latent dimension of $\mathbf{z}_t = [z_1, z_2, \ldots, z_d]^T$ is treated as an independent component. By collating each latent representation $\mathbf{z}_t$ for a given signal $x_i[n]$ processed into a Hankel matrix using Eq. (13), a per-dimension latent signal is obtained.

Henceforth, metrics that conduct comparisons to aspects of the manifold that are time independent, such as the latent prior or static representations of the manifold, are referred to as path independent (PI) metrics, and metrics that compare within sensor time are referred to as path dependent (PD) metrics. For notational simplicity, subscripts $PD|PI$ are used to describe HIs and LHIs that are PD or PI. PD metrics require sensor time, as they evaluate and quantify latent manifold changes within a temporally preserved setting. In Fig. 5, a visual representation of the LVM HI and LHI classes is shown. Each of the conceptual classes detailed in Fig. 5 capture a unique perspective of the latent manifold.

In Table 1, a list of potential LHI methods identified by the authors for the proposed conceptual categories is given. Table 1 demonstrates that the number of possible LHIs is numerous, and that the classes are not limited to only one metric. In this work, it is neither feasible nor useful to use all the methods identified in Table 1. Thus, an indicator subset is used in this work to demonstrate the effectiveness of the latent manifold for vibration-based condition inference in the presence of non-stationary OCs.

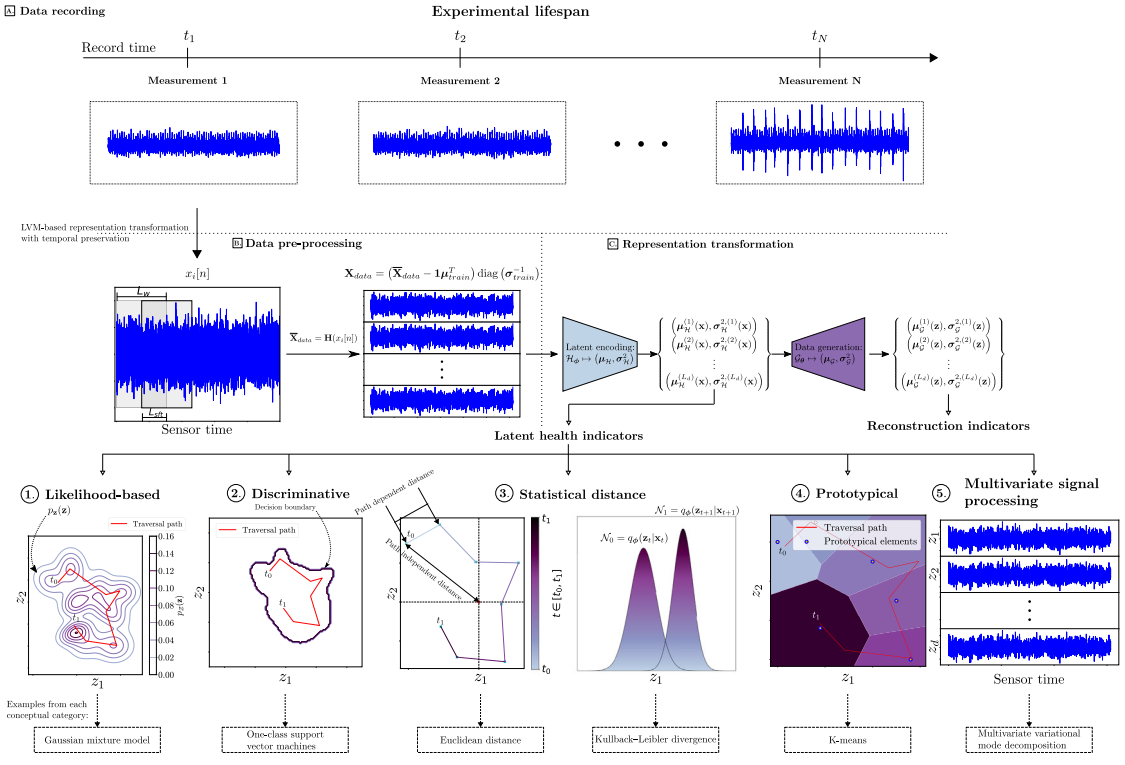


Fig. 5. A generalised overview of how vibration data interfaces with LVMs and how the data space and latent space is used to obtain condition indicators. The states A., B., and C. indicate the LVM process stages, while the numbers 1–5 indicate the five proposed conceptual categories detailed in Section 2.5. Note the latent dimensionality d is not specified, and the pictorial perspectives represent an intuitive illustration of the proposed LHI categories.

2.6. Proposed indicators

In Section 2.5, the five conceptual categories for LHI sampling were detailed. However, the conceptual classes provide several LHI methods, and it is infeasible for this study to use and compare every potential method available. Thus, a small representative subset of LHI methods will be used. These metrics are used to show that the latent manifold is indicative of damage, and that the conceptual classes each capture a unique manifold perspective. We do not use any indicators from the discriminative category as these methods are often computationally demanding due to convex optimisation techniques [48]. To use the proposed indicators, the assumption is made that the latent manifold is accessible. As some of the LHI metrics are dependent on models that capture the latent manifold, we use the latent representations of the training and validation data for reference latent model optimisation. All the proposed indicators exploit the temporal structure induced from the temporal preservation approach. Hence, metrics that are formulated in the PD setting will include reference to t, whereas t is omitted for PI metrics for notational simplicity.

2.6.1. Reconstruction indicators

The first metric is the reconstruction HI. This metric is derived through the reconstruction negative log-likelihood of $p(\mathbf{x}|\mathbf{z})$ and is given by

$$HI^{(1)}(\mathbf{x}) = \frac{1}{D}\sum_{l=1}^{D}\frac{(x_l - \overline{x}_l)^2}{\sigma_l^2}, \tag{17}$$

where $\overline{\mathbf{x}} = \boldsymbol{\mu}_{\mathcal{G}}$ is the reconstruction mean of $\mathbf{x}$, and $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^2_{\mathcal{G}}$ is the reconstruction variance.

2.6.2. Likelihood-based indicators

For metrics from the likelihood-based category, each metric evaluates the negative log-likelihood (NLL) function

$$LHI_{likelihood}(\mathbf{z}) = -\log p_{ref}(\mathbf{z}), \tag{18}$$

where $p_{ref}(\mathbf{z})$ is a reference latent model. This evaluation is PI as the NLL is a proxy metric for an observed sample $\mathbf{z}$ contrasted to the healthy latent manifold. In this work, three reference models are considered, namely, a Gaussian mixture model (GMM), a kernel

Table 1
Potential latent manifold indicator methods for each of the proposed conceptual categories. This table demonstrates the types of methods that may be derived from each class, and indicates whether a given method requires some external hyper-parameter selection and/or optimisation intervention. The metrics used in this study are highlighted in bold.

Conceptual class	Method	Requires hyper-parameter/ model parameter optimisation
Likelihood-based	Gaussian model - maximum likelihood estimate for covariance [31]	✓
	Gaussian model - minimum covariance determinant estimate for covariance [45]	✓
	Gaussian mixture model ($LHI^{(1)}$) [31]	✓
	Kernel density estimation ($LHI^{(2)}$) [31]	✓
	Generative topographic map ($LHI^{(3)}$) [46,47]	✓
Discriminative	Support vector data description [48]	✓
	One-class support vector machine [49]	✓
	One-class neural network [50]	✓
	Deep support vector data description [51]	✓
Statistical distance: dissimilarity-based	Manhattan distance (Minowski distance with $p = 1$)	✗
	Euclidean distance (Minowski distance with p=2) ($LHI^{(4)}_{PD\|PI}$)	✗
	Chebychev distance (Minowski distance with $p = \infty$)	✗
	Tanimoto distance (continuous form)	✗
	Kernel distance	✓
	Cosine distance ($LHI^{(5)}_{PD}$)	✗
	Canberra distance	✗
Statistical distance: distribution similarity	**Bhattacharyya distance** ($LHI^{(6)}_{PD\|PI}$)	✗
	Forward KL divergence	✗
	Reverse KL divergence	✗
	Symmetric KL divergence ($LHI^{(7)}_{PD\|PI}$)	✗
	Hellinger distance	✗
	Differential entropy ($LHI^{(8)}_{PD\|PI}$)	✗
	Jensen–Shannon divergence	✗
Prototypical	k-means [31]	✓
	k-medians ($LHI^{(9)}$) [31]	✓
	k-mediods [52]	✓
	Self-organising maps [53]	✓
Multivariate signal processing and analysis	Univariate empirical/variation mode decomposition [54–56]	✓
	Multivariate empirical/variational mode decomposition [57,58]	✓
	Minimum entropy deconvolution [15,59]	✓
	Auto-regressive (Box-Jenkins) models [60]	✓
	Moving statistics ($LHI^{(9)}$ and $LHI^{(10)}$)	✓

density estimation (KDE) model, and a generative topographic map (GTM) model [31]. These models use the latent representations of the data in the training set for model optimisation. The GMM LHI is given as

$$LHI^{(1)}(\mathbf{z}) = -\log\left(\sum_{k=1}^{K} \pi_k^{(GMM)} \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_k^{(GMM)}, \boldsymbol{\Sigma}_k^{(GMM)})\right), \tag{19}$$

where K is the number of Gaussian distribution components used to model the latent manifold. The expectation maximisation (EM) algorithm [61] is used to estimate the GMM parameters $\pi_k^{(GMM)}, \boldsymbol{\mu}_k^{(GMM)}$, and $\boldsymbol{\Sigma}_k^{(GMM)}$. The mixture number K is estimated using the silhouette coefficient [62] and Akaike information criterion (AIC) [31]. The KDE model LHI is given as

$$LHI^{(2)}(\mathbf{z}) = -\log\left(\frac{1}{K}\sum_{k=1}^{K} \frac{1}{(2\pi h^2)^{(d/2)}} \exp\left\{-\frac{\|\mathbf{z}-\mathbf{z}_k\|_2^2}{2h^2}\right\}\right), \tag{20}$$

where h is the KDE model bandwidth, $\|\cdot\|_2$ is the Euclidean norm, and K is the number of latent points used as reference centres. The hyper-parameter h is estimated using an exhaustive hyper-parameter grid search and k-fold cross validation on the model log-likelihood. The GTM LHI is given as

$$LHI^{(3)}(\mathbf{z}) = -\log\left(\frac{1}{K}\sum_{k=1}^{K} \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_k^{(GTM)}, \frac{1}{\beta}\mathbf{I})\right), \tag{21}$$

where $\boldsymbol{\mu}_k^{(GTM)}$ is a kernel-based regression mapping from a two-dimensional grid of points $\mathbf{u} \in \mathbb{R}^2$ to the latent space and β^{-1} is the local Gaussian variance. The regression mapping $\boldsymbol{\mu}_k$ is defined as

$$\boldsymbol{\mu}_k^{(GTM)} = \mathbf{W}\boldsymbol{\phi}(\mathbf{u}_k), \tag{22}$$

where $\mathbf{W}$ is a mapping matrix, $\boldsymbol{\phi}(\cdot)$ is a basis function and $\mathbf{u}_k$ is the kth node in the two-dimensional grid. For the GTM model, we set $K = 625$ to define a 25×25 two-dimensional grid to capture the latent space. The GTM model is a statistical alternative to the

self-organising map algorithm [46,53]. The EM algorithm is used to estimate the model parameters and the basis function of choice is a radially symmetric Gaussian [46,47].

2.6.3. Dissimilarity-based indicators

The dissimilarity metrics of interest here are the Euclidean and cosine distances. Dissimilarity metrics are easily formulated in both the PD and the PI setting. The Euclidean distance is given as

$$LHI^{(4)}_{PD|PI}(\mathbf{z}_{t+1}, \mathbf{z}_t) = \|\mathbf{z}_{t+1} - \mathbf{z}_t\|_2^2. \tag{23}$$

In the PI formulation, $\mathbf{z}_t$ is set to the zero vector $\mathbf{z}_t = \mathbf{0}$. The cosine distance is given as

$$LHI^{(5)}_{PD}(\mathbf{z}_{t+1}, \mathbf{z}_t) = \cos^{-1}\left(\frac{\mathbf{z}_{t+1}^T \mathbf{z}_t}{\|\mathbf{z}_{t+1}\|_2 \cdot \|\mathbf{z}_t\|_2}\right). \tag{24}$$

The cosine distance metric can only be used in the path-dependent setting, as setting $\mathbf{z}_t = \mathbf{0}$ produces an undefined metric.

2.6.4. Distribution indicators

To compute statistical distribution differences, an assumption is made that the latent space is Gaussian distributed, which means that closed form solutions are available. This exploitation of the explicit assumption ensures that the statistical distance metrics have closed-form solutions. Metrics from this category are designed to ensure that both the PD and PI settings are accessible. The Bhattacharrya distance, symmetric KL divergence and differential entropy are used as representative metrics from this LHI category. The Bhattacharrya distance is given by

$$LHI^{(6)}_{PD|PI}(\mathcal{N}_t \parallel \mathcal{N}_{t+1}) = \frac{1}{8}\left(\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\right)^T \boldsymbol{\Sigma}_{t+1}^{-1}\left(\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\right) + \frac{1}{2}\ln\left(\frac{\det(\boldsymbol{\Sigma})}{\sqrt{\det(\boldsymbol{\Sigma}_t)\det(\boldsymbol{\Sigma}_{t+1})}}\right), \tag{25}$$

where $\boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_{t+1}}{2}$, $\mathcal{N}_t = q_{\boldsymbol{\phi}}(\mathbf{z}_t|\mathbf{x}_t)$, $\mathcal{N}_{t+1} = q_{\boldsymbol{\phi}}(\mathbf{z}_{t+1}|\mathbf{x}_{t+1})$, and $\det(\cdot)$ is the determinant function. In the PI setting, $\mathcal{N}_t$ is set to the latent prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$. The symmetric KL divergence metric is the summation of the forward and reverse KL divergence terms and is given by

$$\begin{aligned} LHI^{(7)}_{PD|PI}(\mathcal{N}_t \parallel \mathcal{N}_{t+1}) &= KL(\mathcal{N}_t \parallel \mathcal{N}_{t+1}) + KL(\mathcal{N}_{t+1} \parallel \mathcal{N}_t) \\ &= \frac{1}{2}\left[\mathrm{tr}(\boldsymbol{\Sigma}_{t+1}^{-1}\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Sigma}_{t+1}) + \left(\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\right)^T\left(\boldsymbol{\Sigma}_t^{-1} + \boldsymbol{\Sigma}_{t+1}^{-1}\right)\left(\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t\right) - 2d\right], \end{aligned} \tag{26}$$

where $\mathrm{tr}(\cdot)$ is the trace operator. Finally, the differential entropy of a multivariate Gaussian distribution $\mathcal{N}(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given as

$$H(\mathcal{N}) = \frac{d}{2}\left(1 + \log(2\pi)\right) + \frac{1}{2}\log\left(\det(\boldsymbol{\Sigma})\right), \tag{27}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{(d\times d)}$. The differential entropy is used to produce a LHI of the form

$$LHI^{(8)}_{PD|PI}(\mathcal{N}_t \parallel \mathcal{N}_{t+1}) = \frac{1}{2}\left[\log\left(\det(\boldsymbol{\Sigma}_{t+1})\right) - \log\left(\det(\boldsymbol{\Sigma}_t)\right)\right]. \tag{28}$$

The decision to include $LHI^{(8)}_{PD|PI}(\mathcal{N}_t \parallel \mathcal{N}_{t+1})$ in this investigation was made to ascertain whether the LVMs exhibit sensitivity through $\boldsymbol{\mu}_{\mathcal{H}}$ alone, or through both $\boldsymbol{\mu}_{\mathcal{H}}$ and $\boldsymbol{\sigma}^2_{\mathcal{H}}$. The outputs $\boldsymbol{\mu}_{\mathcal{H}}$ and $\boldsymbol{\sigma}^2_{\mathcal{H}}$ of the latent transition function $\mathcal{H}_{\boldsymbol{\phi}}$ are both dependent on $\mathbf{x}$, and thus we wish to determine the sensitivity of $\boldsymbol{\sigma}^2_{\mathcal{H}}$ to damage in vibration data.

2.6.5. Prototypical indicators

LHI metrics from this category assume that there are a fixed number of prototypical elements N_p that characterise the latent space $\mathcal{Z}$ [44]. A k-medians algorithm is used to determine the prototypical elements in the latent space [31]. The k-medians LHI is given as

$$LHI^{(9)}(\mathbf{z}) = \min_{k=1,\ldots,K}\left\{\|\mathbf{z} - \mathbf{P}_k\|_2\right\}, \tag{29}$$

where $\mathbf{P}_k$ is the k^{th} prototypical element, $k = 1, \ldots, K$. This metric uses the Euclidean distance between $\mathbf{z}$ and the nearest prototypical element for condition inference. The number of prototypical elements K is estimated using the silhouette coefficient [62]. The k-medians metric is PI, and the prototypical elements serve as beacons from which we can measure Euclidean manifold dissimilarity within the latent manifold.

2.6.6. Multivariate signal indicators

By treating each latent dimension as an independent data channel, signal processing techniques may be utilised to detect deviations within the latent space. In this work, the moving RMS and the moving kurtosis are used for condition inference. The moving RMS, in a discrete-time setting, is given as

$$LHI^{(10)}[t] = \sum_{l=1}^{d}\sqrt{\frac{1}{N_w}\sum_{t}^{t+N_w} z_{l,t}^2}, \tag{30}$$

where the summation $\sum_{l=1}^{d}$ is the summation of the windowed RMS terms for each latent dimension, N_w represents the moving window size, and $z_{l,t}$ refers to the l^{th} index of the latent vector $\mathbf{z}_t$. In order to use the moving RMS, the latent representations of vibration data processed using Eq. (13) is required. The moving kurtosis is given as

$$LHI^{(11)}[t] = \sum_{l=1}^{d}\left[\frac{\frac{1}{N_w}\sum_{t}^{t+N_w}\left(z_{l,t}-\mu_l[t]\right)^4}{\left(\frac{1}{N_w}\sum_{t}^{t+N_w}\left(z_{l,t}-\mu_l[t]\right)^2\right)^2}-3\right], \tag{31}$$

where $\mu_l[t]$ is the window average

$$\mu_l[t] = \frac{1}{N_w}\sum_{t}^{t+N_w} z_{l,t}. \tag{32}$$

In this work, the window size N_w is set to $N_w = 512$ for the moving RMS and the moving kurtosis. The moving RMS and moving kurtosis are both PI metrics.

2.6.7. Final reflections

The proposed LHIs are imperative to ensure that all manifolds' perspectives are captured and inspected. Through the five proposed LHI categories, the objective is to showcase that the latent manifold is responsive to damage and that the condition of an asset can be inferred. This ensemble-based approach to latent manifold condition inference can potentially lead to improved detection and interpretation. By investigating the PD and PI characteristics of the latent manifold, manifold deviations through the path travelled or the rate of traversal are accessible. In this way, the proposed conceptual categories allow for different aspects of the latent space to be highlighted.

3. Experimental evaluation

In this section, the performance of the proposed LHIs will be demonstrated on two experimental datasets. This is imperative to emphasise the responsiveness of the latent manifold to anomalous data. The LVMs considered for this investigation are the PPCA and VAE models [29,30]. The network architectures, hyper-parameters and all supplementary information required for work reproducibility are given in the Appendix.

3.1. IMS dataset

The NSF I/UCR Center for Intelligent Maintenance Systems (IMS) has made available a dataset which is frequently used to benchmark CM techniques [63]. The available data was collected from three endurance tests in which bearing faults developed during the test's lifespan. The experimental setup was subjected to a constant load of 6000 lbs and a constant shaft speed of 2000 rpm for the entirety of the test's lifespan. To demonstrate the capabilities of the proposed HIs and LHIs, the vibration data related to channel one of dataset two was used in this investigation. This dataset consists of 984 vibration signals sampled for one second with a sampling frequency of 20.48 kHz [64].

3.1.1. LVM development

For analysis of the IMS dataset, the first 5% of the available IMS data is assumed as healthy. The healthy data is used for LVM training. The first 49 records are processed using Eq. (13), split into the training and validation datasets using a 80%–20% split procedure and standardised using Eq. (15). To perform model evaluation, we process all of the available data using the temporal preservation approach [13]. The properties of the PPCA and VAE models applied to the IMS dataset are given in the Appendix.

3.1.2. Analysis methodology

The HI and proposed LHI metrics used in this work produce indicator signals. These signals must be post-processed to quantify the responsiveness of the proposed indicators to the presence of a fault. It is emphasised that the analysis methodology is merely a means to facilitate quantitative metric comparison. Hence, we do not propose that this methodology is optimal for CM, only that it allows us to distinguish and quantify the performance of the metrics. For the IMS dataset, the indicator signal RMS and the amplitude of the spectral component corresponding to the ball pass frequency for the outer race (BPFO) is used for performance quantification purposes. The latter is used as the a priori for the considered IMS dataset is that a bearing fault occurred. As the indicators all have different orders of magnitude, the available indicators are standardised using the records used for LVM training. The standardisation process ensures that different indicators are non-dimensionalised, and removes any interpretation bias to the offset and the scale of the indicators [65,66]. The theoretical BPFO is 236 Hz or 7.092 shaft orders [64,67]. To allow for a quantitative performance comparison of the different metrics, the area under the receiver operating characteristic (ROC) curve (AUC) is used. As discussed in [13], the usefulness of the AUC in CM applications for rotating machinery is acutely sensitive to the shaft speed f_s, model window length L_w and the sampling rate F_s of the measured data. The AUC ignores sensor time and is used to quantitatively compare the performance of the considered HIs and LHIs, providing insight into the indicator's sensitivity to the fault present in the data. The reference class distribution is developed using the indicator signals corresponding to the measured vibration data used for LVM training.

To quantify the similarities and differences of the considered HIs and LHIs on an inter-method and intra-class level, the Pearson correlation coefficient, represented by $\rho_{X,Y}$ for two random variables X and Y, will be calculated for the RMS of the indicator signals through record time. The sample Pearson correlation coefficient, represented by r_{xy}, is a linear correlation measure that is formulated by normalising the covariance between two random variables and is calculated for N sets of paired data $\left[\left(x_1, y_1\right), \ldots, \left(x_N, y_N\right)\right]$ using

$$r_{xy} = \frac{\sum_{l=1}^{N} \left(x_l - \mathbb{E}\left[x\right]\right)\left(y_l - \mathbb{E}\left[y\right]\right)}{\sqrt{\sum_{l=1}^{N} \left(x_l - \mathbb{E}\left[x\right]\right)^2}\sqrt{\sum_{l=1}^{N} \left(y_l - \mathbb{E}\left[y\right]\right)^2}}, \tag{33}$$

where $\mathbb{E}\left[\cdot\right]$ is the expectation operator.

3.1.3. Results

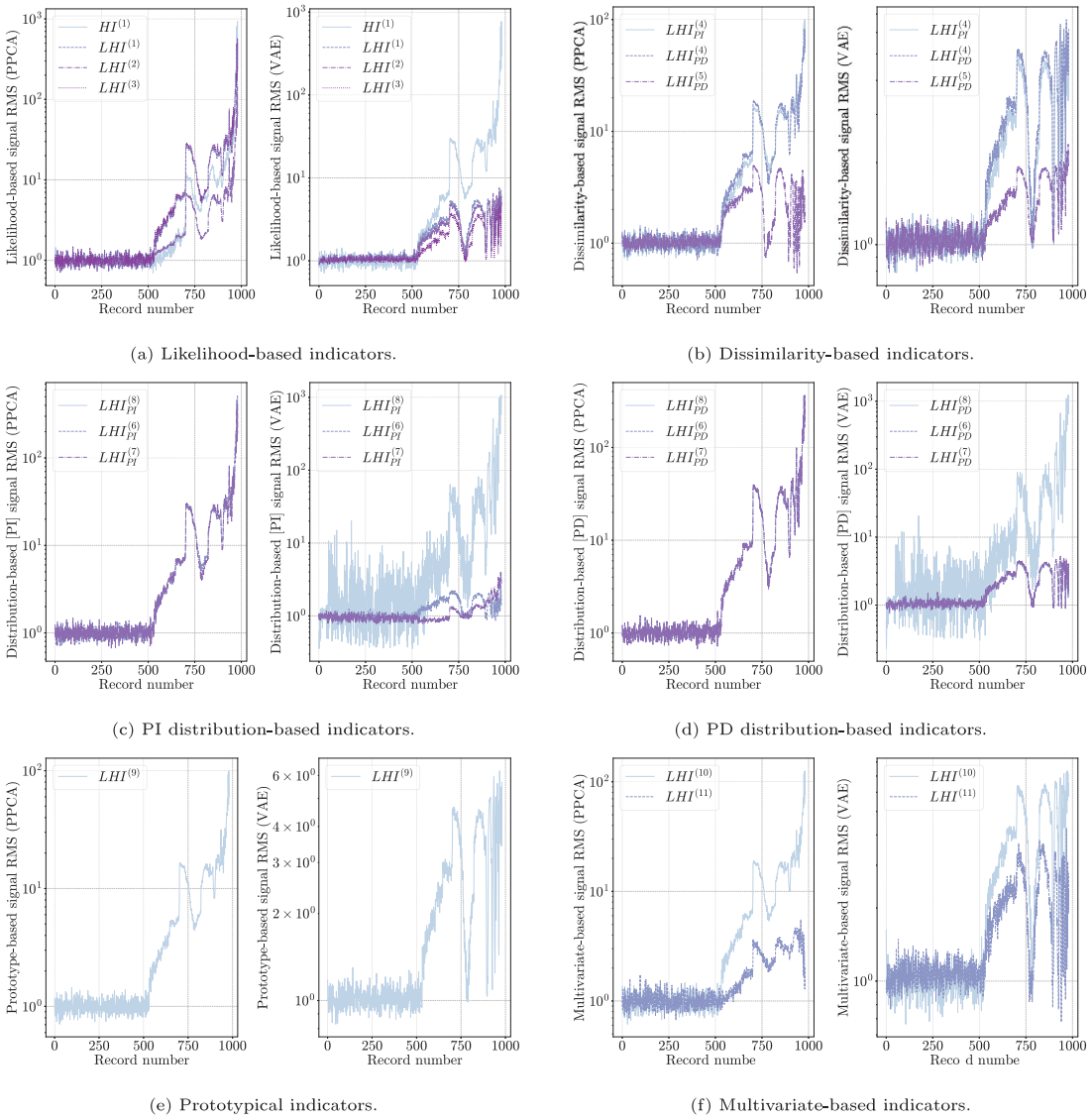
In Fig. 6 the HI and LHI indicator results are shown. Most metrics detect and respond to some change in the vibration data in the region around record number 500. This indicates that the response to anomalous data manifests in both the data space and the latent manifold. The dual nature of the responsiveness to anomalous data in both the data space and the latent manifold is useful for condition inference. As the number of latent manifold metrics surpasses the number of data space metrics, the damage detection process is enhanced by incorporating additional LHIs. Thus, the fusion of both data space and latent manifold indicators is powerful for condition inference. Each of the conceptual categories have metrics that respond to the anomalies present in the data, which emphasises the benefit of considering and using metrics from the conceptual categories proposed in Section 2.5. Furthermore, an explicit LVM with linear transition functions is responsive to the anomalous data present in the considered IMS dataset. Thus, both the linear and nonlinear explicit LVMs can capture the latent manifold of the reference IMS vibration data, and both manifolds are indicative of damage.

Due to the number of responsive metrics in Fig. 6, we will discuss metrics that appear to be unresponsive to the presence of the anomalous vibration data. These standardised metrics are $LHI_{PI}^{(8)}$ and $LHI_{PD}^{(8)}$ for the PPCA model seen in Figs. 6(c) and 6(d) respectively. The unresponsiveness of the $LHI_{PI|PD}^{(8)}$ metric is expected for the PPCA model as this metric represents differential entropy of a Gaussian distribution, and this is only a function of $\boldsymbol{\Sigma}_{\mathcal{H}}$. However, the covariance matrix of the PPCA latent posterior distribution is independent of the data [29,31]. Regardless of whether the PI or PD characteristics of the LHI are inspected, the resulting PPCA model can only respond to damage through changes in the posterior mean $\boldsymbol{\mu}_{\mathcal{H}}(\mathbf{x})$. In the PI and PD formulation, the VAE $LHI^{(8)}$ metric is responsive to the anomalies in the IMS vibration data. This result emphasises the need to consider both the PD and PI perspectives of the latent manifold when deriving LHIs, as the method in which the latent manifold responds to anomalous data is not guaranteed. Hence, metrics that monitor both the characteristics of the path traversed and the path traversal rate in the latent manifold are imperative for condition inference.

In an inter-method analysis of Fig. 6, the PPCA model produces stronger latent manifold deviations. The PPCA model metrics consistently have a larger scale change in comparison to the VAE metrics, for metrics that exhibit a response to the anomalous data. This scale change is quantifiable as the indicator signals are standardised. While the reconstruction indicator, $HI^{(1)}$, for the VAE model appears to be more sensitive to the anomalous data, the latent manifold, while responsive, is less sensitive to the anomalous data. This indicates that a linear latent manifold is sufficient for condition inference purposes, and indicates that the generative transition function, the VAE decoder, is more expressive than the latent posterior transition function.

To quantify the intra-class similarities for the considered metrics, the sample Pearson correlation coefficient was calculated using the RMS trends seen in Fig. 6 on a intra-class and inter-method level. The former refers to the correlation for metrics within different classes while the latter refers to the correlation between the metrics from a specific model class with all the metrics from another model. In Fig. 7, the intra-class correlation is shown. Evidently, from the intra-class correlation, most indicators from both models are highly correlated as the median of both methods is greater than 0.75. However, the tails in Fig. 7 suggest that there are unique responses in both methods, as the correlation distribution is spread over the r_{xy} domain. This suggests that, for the IMS dataset, most of the indicators presented in this work exhibit similar responses when inspecting the indicator signal RMS. In an inspection of Fig. 6, one would expect that the distribution-based metrics for the PPCA model have strong correlation and hence selecting either $LHI_{PD|PI}^{(6)}$ or $LHI_{PD|PI}^{(7)}$ would have been suitable. Additionally, the $LHI_{PI}^{(4)}$ and $LHI_{PD}^{(4)}$ metrics exhibit clear similarity, and hence either of these would have been suitable for the dissimilarity-based class. In an inter-method correlation comparison, shown in Fig. 8, the indicators from the different classes exhibit non-similar responses on an inter-method level. This indicates that the choice of method influences the sensitivity of the indicators to the fault present in the data. This indicates that the metrics used in this work are model formulation dependent. The prototype-based indicator from the PPCA model has a condensed r_{xy} range, which suggests that this indicator exhibits a similar response to the indicators obtained from the VAE model. The indicators from the VAE model, on an inter-method level, are unique to the PPCA model, as the r_{xy} distribution is spread over the domain for each of the indicator classes.

To infer the damage present in the IMS dataset, the spectral amplitude of the analytical BPFO component was monitored. In Fig. 9 the BPFO amplitudes are shown for the HIs and LHIs considered in this work. It is evident that many of the proposed LHI indicator signals contain this spectral component and that it grows over the experimental lifespan. The temporal preservation approach used in the model evaluation stage is critical to ensuring that we can access and trend this spectral component. Thus, it is possible to not only detect the anomalous data in the IMS dataset, but the cause of the anomalies may be attributed to damage on the outer race of the bearing. Fig. 10 depicts the AUC results for the considered HIs and LHIs. The AUC provides a quantitative perspective on the sensitivity of the considered HIs and LHIs for the PPCA and VAE models. In Fig. 10, sensitivity of the considered HIs and LHIs to the



(a) Likelihood-based indicators.

(b) Dissimilarity-based indicators.

(c) PI distribution-based indicators.

(d) PD distribution-based indicators.

(e) Prototypical indicators.

(f) Multivariate-based indicators.

Fig. 6. The RMS of the standardised HI and LHI indicator signals for the PPCA and VAE models. The results for each LHI category is shown for the PPCA and VAE models. In (d), $LHI^{(8)}$ in both the PI and PD formulation for the PPCA model cannot be seen as the presented figures are log-scaled and the standardised metric is zero throughout record number. Note that the model indicator is given in the *y*-axis label of each sub-figure.

bearing fault development during the experimental lifespan is clear. Furthermore, the variation in manifold sensitivity emphasises the requirement for an ensemble-based latent inference procedure, as different perspectives of the latent manifold produce a variation in the metric response. The VAE model HI and LHI indicators are more sensitive to the bearing self-healing phenomena present around record 750, which results in an AUC reduction. The strength of the PPCA AUC deviations indicates that facets of the PPCA latent manifold is more sensitive to the anomalous fault data, and the PPCA latent manifold response is significantly more noticeable than the VAE manifold response.

Hence, from the results in Figs. 6, 9 and 10, the proposed HIs and LHIs are useful for condition inference. The results in Figs. 6 indicate that the metrics are responsive to the fault present in the data, and that the RMS of the LHI is a suitable method for indicator trending. Fig. 9 serves as an indication of the usefulness of performing temporal preservation during the model analysis stage, as the fault type is isolated and identified. By performing an inter-method and intra-class analysis of the RMS trends of the indicators for the IMS dataset: *(i)* the metrics from the likelihood-based class that monitor the latent manifold offer similar insight, *(ii)* the nonlinear model has a more sensitive response in data space rather than in the latent manifold, *(iii)* the dissimilarity metrics exhibit similar responses across methods and only the $LHI^{(5)}$ metric has a unique response to the other metrics, *(iv)* the $LHI^{(6)}$ and $LHI^{(7)}$ metrics respond similarly, and *(v)* the uniqueness of the PI distribution-based metrics and the multivariate-based metrics is LVM methodology dependent.

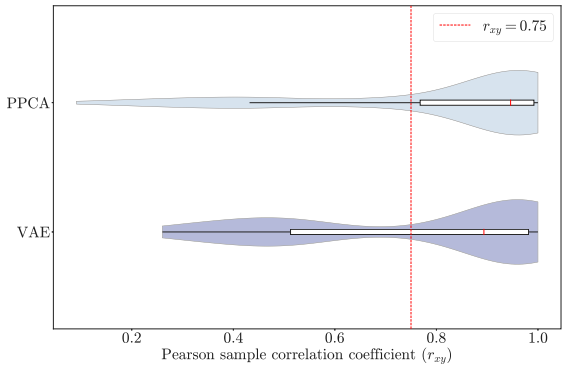
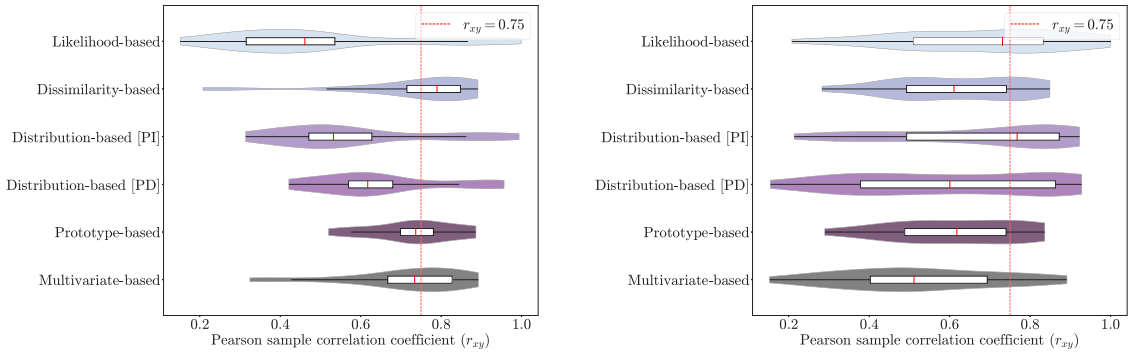


Fig. 7. Violin plots of the sample Pearson correlation coefficient on an intra-class level for the IMS dataset. Note that the intra-class comparison is visualised on a per-method level as we compute all of the r_{xy} terms for all metrics of a given LVM method to compare class-wide similarity.



(a) PPCA method-wide correlation.

(b) VAE method-wide correlation.

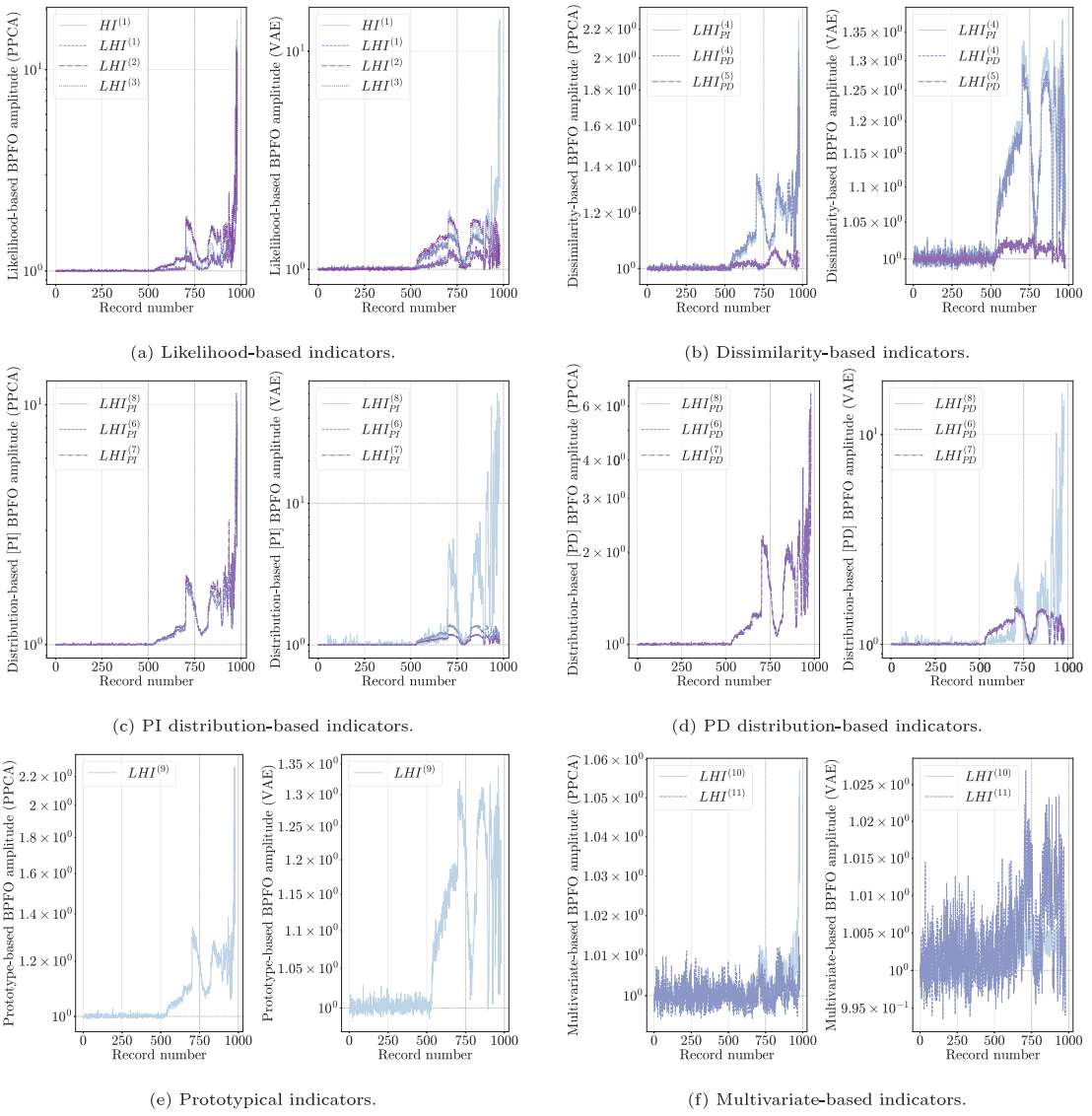
Fig. 8. Violin plots of the sample Pearson correlation coefficient on an inter-method level for the IMS dataset. Note that the inter-method comparison is visualised on a per-class level as we compute all of the r_{xy} terms for all metrics of a given indicator class and all metrics from another LVM method to compare method-wide similarity.

3.1.4. Comparison

To compare the LHIs used in this work to HIs from other works, two methods were used. For the first method, the Mahalanobis distance-based HI proposed in Wang et al. [68] is used. For the second method, a set of sparsity measures that provide HIs are used. These sparsity measures utilise the squared envelope spectrum and are designed to exhibit sensitivity to the impulsiveness of repetitive transients. The sparsity measures considered here are the spectral kurtosis [18], the spectral negentropy [69,70], the smoothness index (SI) [71], the Gini index (GI) [72], the ratio of the L_2 and L_1 norm ($\frac{L_2}{L_1}$) [73], the Hoyer measure (HM) [74], the Box–Cox (BC) measure [75], and the skewness. Note that we do not consider a passband $[l, h]$ filtered version of a signal $x_i[n]$ as our objective is not to obtain a transformed signal that maximises the sparsity measure. Methods focused on the optimal passband are well detailed in the literature, e.g. [19], but are beyond the scope of this work. The BC measure hyper-parameter λ is set to 0.5 as this represents a sparsity measure between the spectral kurtosis and the spectral negentropy.

In Fig. 11(a) the results of using the HI proposed in [68] for the IMS dataset is shown, and it is responsive to the bearing fault. However, its performance is inferior to the proposed HIs and LHIs as it is less responsive. Additionally, this HI is designed for bearing fault detection, uses hand-crafted time-domain features, and is temporally non-preserving, whereas the proposed LHIs require no prior knowledge of the fault mode and are trained on the observed vibration data alone. Thus, the responses from the proposed LHIs, in comparison with the methodology from [68], are useful and exhibit a clear response to the bearing fault present in this dataset.

The sparsity measures in Fig. 11(b) exhibit sensitivity to the bearing fault and are sensitive to the change in condition in the region surrounding record 500. Each of the considered sparsity metrics are non-monotonic, attributed to bearing self-healing, and are indicative of a change in asset state. The proposed LHIs, in comparison, exhibit similar sensitivity and are not designed to enhance any property of the observed data. Hence, they are useful to vibration-based CM. Additionally, the fault type can be inferred from the LHIs, which is only possible for the sparsity measures if they are used in a demodulation-band framework [19] or a blind-deconvolution framework [22,76].



(a) Likelihood-based indicators.

(b) Dissimilarity-based indicators.

(c) PI distribution-based indicators.

(d) PD distribution-based indicators.

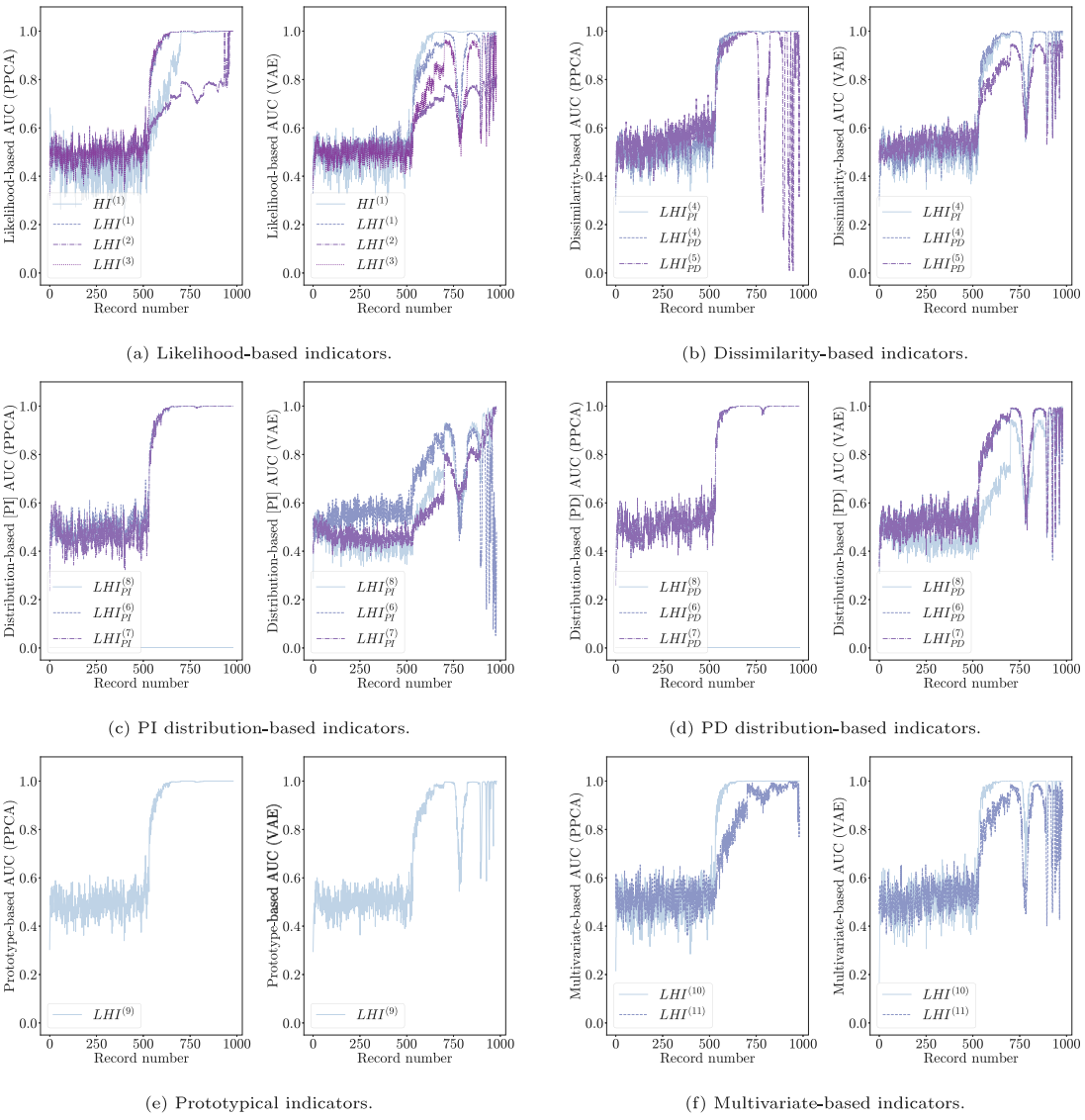
(e) Prototypical indicators.

(f) Multivariate-based indicators.

Fig. 9. The amplitude of the BPFO spectral component for the standardised HI and LHI indicator signals considered in this work. In (c) and (d), the amplitude of $LHI^{(8)}$ for the PPCA model cannot be seen as the metric is zero and the presented figures are log-scaled. Note that the spectral components for the considered HIs and LHIs were zero-mean adjusted using the values for the records corresponding to the training data and shifted to 1 for comparison purposes.

3.2. Gearbox dataset

The Centre for Asset Integrity Management at the University of Pretoria has made available an experimental gearbox dataset to benchmark the performance of the proposed HIs and LHIs [77]. The experimental gearbox dataset has representative data for a gearbox in a healthy and unhealthy condition with complex operating conditions. The available data was obtained through two distinguishable collection processes. In the first experiment, 100 vibration data samples were obtained for the gearbox in a reference (assumed healthy) condition. In the second experiment, the gearbox was disassembled to seed a localised gear tooth fault on the gear. An additional 200 vibration data samples were collected from the point of fault initialisation. The experimental setup, shown in Fig. 12, consists of an electrical motor, a step-down helical gearbox, two-step up helical gearboxes, and an alternator. The sampling rate of the optical probe attached to the input shaft of the centre gearbox was 51.2 kHz. Each of the available vibration samples from the centre gearbox was recorded for 20 s at a sampling rate of 25.6 kHz. In Fig. 13(a) the operating conditions for each vibration sample are shown. Following [13] and [78], we low-pass filter the available data with a third-order Butterworth filter around 3200 Hz. In Fig. 13(b), the statistical properties of the available vibration data samples are shown.



(a) Likelihood-based indicators.

(b) Dissimilarity-based indicators.

(c) PI distribution-based indicators.

(d) PD distribution-based indicators.

(e) Prototypical indicators.

(f) Multivariate-based indicators.

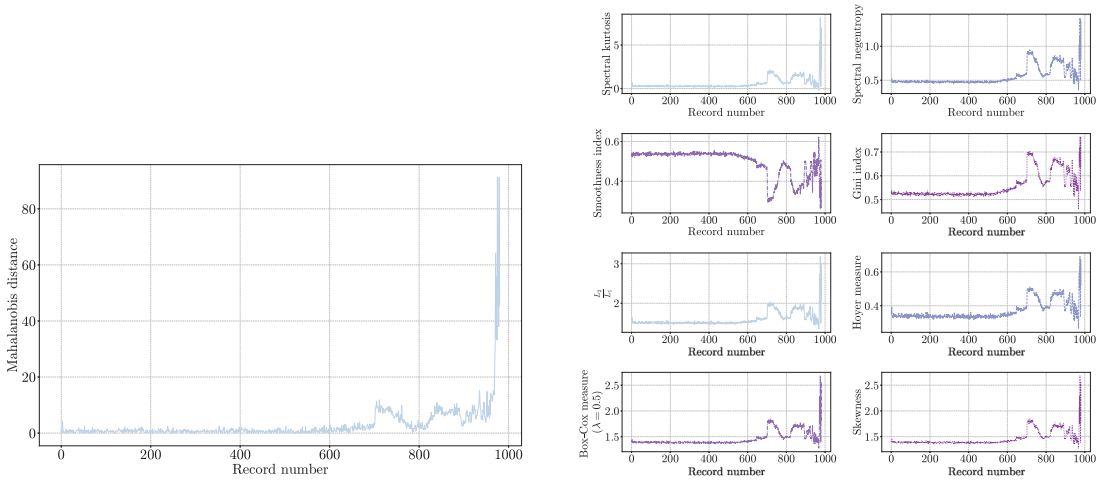
Fig. 10. The AUC for the standardised indicator signals considered in this work. In (c) and (d), the AUC is zero as the $LHI^{(8)}$ for the PPCA model is independent of the data and thus exhibits no deviations or changes in value. An AUC of 0.5 indicates that the indicator signal is indistinguishable from the indicators from the reference healthy data.

3.2.1. LVM development

For analysis of the gearbox dataset, 25% of the available healthy gearbox data is used for LVM training. The available healthy records are processed using Eq. (13), split into the training and validation datasets using a 80%−20% split procedure and standardised using Eq. (15). To perform model evaluation, we process all of the available data using the temporal preservation approach [13]. The use of the temporal preservation approach during model evaluation produces a set HI and LHI signals from which the state of the experimental state may be inferred. The Appendix details the properties of the PPCA and VAE models for the gearbox dataset.

3.2.2. Analysis methodology

The HIs and proposed LHI metrics used in this work produce indicator signals. As it is known that the fault characteristic present in the gearbox dataset is a gear tooth fault, an analysis procedure is followed to enable quantitative comparison of the metrics considered. The metrics, by design, are not insensitive to the non-stationary OCs present in the gearbox data. Hence, the analysis procedure used here is to quantify and, where applicable, highlight the usefulness of each indicator without sensitivity to the variation in the asset operating conditions. This analysis procedure is merely a means to enable metric comparison, and we do not propose that this procedure is state-of-the-art. To post-process the indicator signals, the time-synchronous average (TSA) of the



(a) Mahalanobis distance indicator. (b) Sparsity indicators.

Fig. 11. The Mahalanobis distance metric from [68] and the sparsity metrics applied to the IMS dataset.

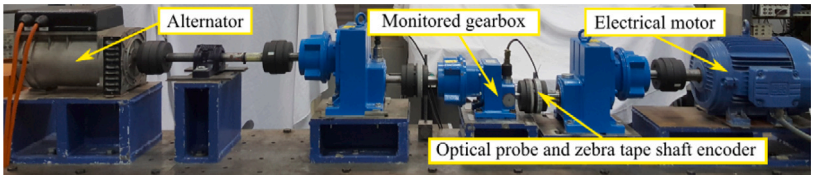
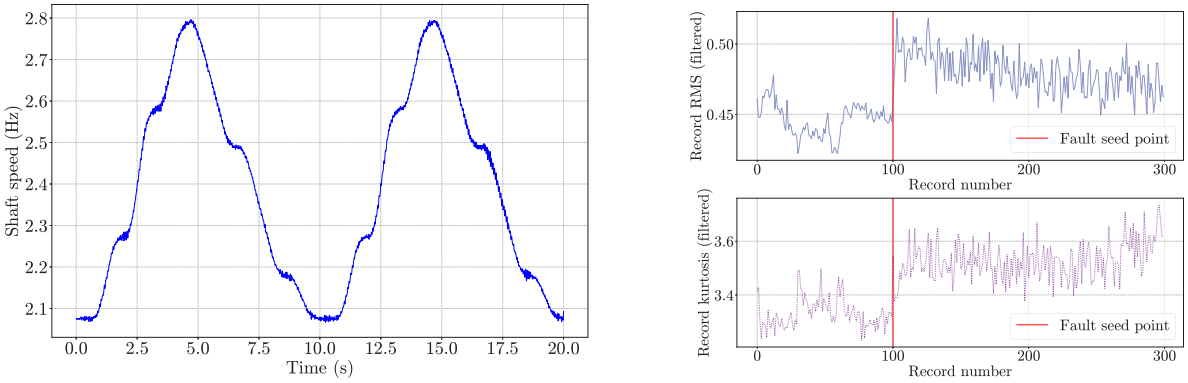


Fig. 12. A schematic of the experimental gearbox setup [77].



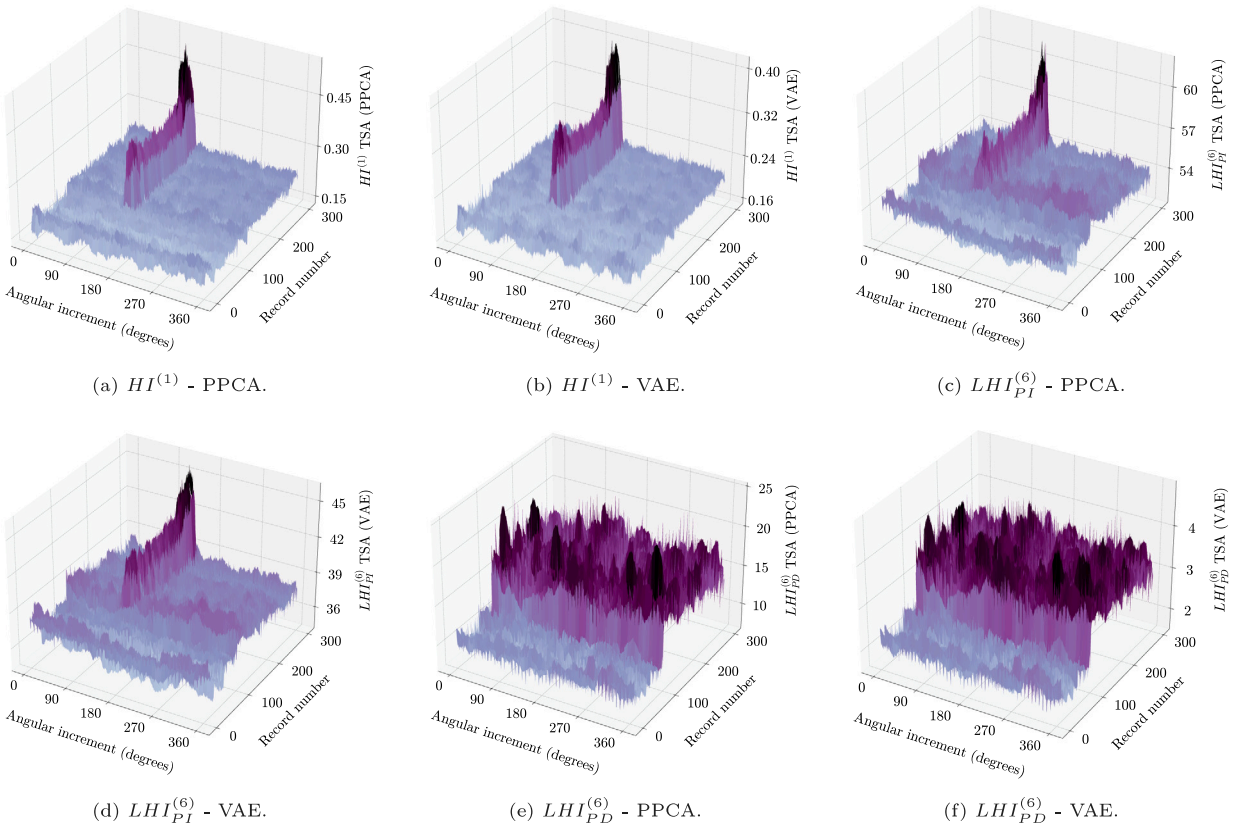
(a) Gearbox operating conditions. (b) Record statistics.

Fig. 13. The operating conditions and the per-record RMS and kurtosis for the vibration data available from the experimental gearbox dataset.

indicator signal is used to ascertain whether the gear tooth damage is detectable. The TSA of signal $x_i[n]$, $n \in \mathbb{Z}$, is given as

$$\overline{x}_{TSA}[n] = \frac{1}{N_r} \sum_{l=1}^{N_r-1} x_i[n + l \cdot N_s], \ 1 \leq n \leq N_s, \tag{34}$$

where N_r is the number of shaft rotations, and N_s is the number of points per rotation [79]. The resulting TSA signal is a sequence of N_s points and represents the deterministic component of $x_i[n]$ for one shaft rotation. Each TSA signal is realigned with respect to the tachometer butt joint. The location of the tachometer butt joint is found using the Bayesian geometry compensation algorithm proposed in Diamond et al. [80]. To qualitatively compare the LVMs considered in this work, and to assess the responsiveness of different indicators, a GMM-based algorithm is used to quantify the presence of an isolated fault [31]. In this algorithm, a GMM with two components is used to model the TSA. For the TSA of a healthy indicator signal, the components of the GMM should have



(a) $HI^{(1)}$ - PPCA. (b) $HI^{(1)}$ - VAE. (c) $LHI_{PI}^{(6)}$ - PPCA.

(d) $LHI_{PI}^{(6)}$ - VAE. (e) $LHI_{PD}^{(6)}$ - PPCA. (f) $LHI_{PD}^{(6)}$ - VAE.

Fig. 14. The TSAs for a selected subset of the HIs and LHIs considered in this work.

little to no observable statistical difference. However, for the TSA of an unhealthy indicator where localised damage is present, the components of the GMM should be statistically different. To quantify the statistical difference between the two local Gaussians of the GMM, the symmetric KL divergence is used. This divergence metric is useful as it offers an analytical solution, does not favour one of the local distributions over the other, and is non-negative. Given parameters of the two local GMM components (μ_0, σ_0^2) and (μ_1, σ_1^2), the symmetric KL divergence becomes

$$KL_{symmetric} = \frac{1}{2}\left(\frac{\sigma_0^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_0^2} + \frac{(\mu_1 - \mu_0)^2}{\sigma_1^2} + \frac{(\mu_0 - \mu_1)^2}{\sigma_0^2} - 2\right). \tag{35}$$

The use of the GMM-based algorithm allows for trending of the properties of the TSA over the experimental lifespan. As with the IMS dataset, we standardise the TSA of the indicators to remove any bias to the scale or offset of the indicators prior to employing the GMM-based algorithm. To quantify the similarities and differences of the considered HIs and LHIs on an inter-method and intra-class level, the sample Pearson correlation coefficient, given in Eq. (33), will be calculated for the symmetric KL divergence trends through record time. To allow for quantitative performance comparison of the different metrics, the AUC is used. However, as the fault in this dataset is localised around one gear tooth, we carefully select the synchronous average indicator values in the region of the tooth in which the fault was seeded as the test set for AUC evaluation. This ensures that we can quantify the responsiveness of different LHIs for this dataset without succumbing to AUC insensitivity issues. Note that the AUC procedure used is specific to the fault present in the gearbox dataset and cannot be readily applied to every fault case.

3.2.3. Results

In Fig. 14 the TSAs for a selected subset of the HI and LHI metrics considered in this work are shown as functions of record number. The selected subset of metrics showcases the TSAs of metrics from the data space and the latent manifold for both the PPCA and the VAE model. Evidently, through the inspection of Fig. 14, the gear tooth fault is detectable by both the PPCA and VAE models. In Figs. 14(e) and 14(f), the TSAs of the $LHI_{PD}^{(6)}$ metric for both LVMs are ineffective for condition inference. This result emphasises the necessity to consider several LHI metrics in the latent manifold, as the method of LVM latent deviation is not guaranteed.

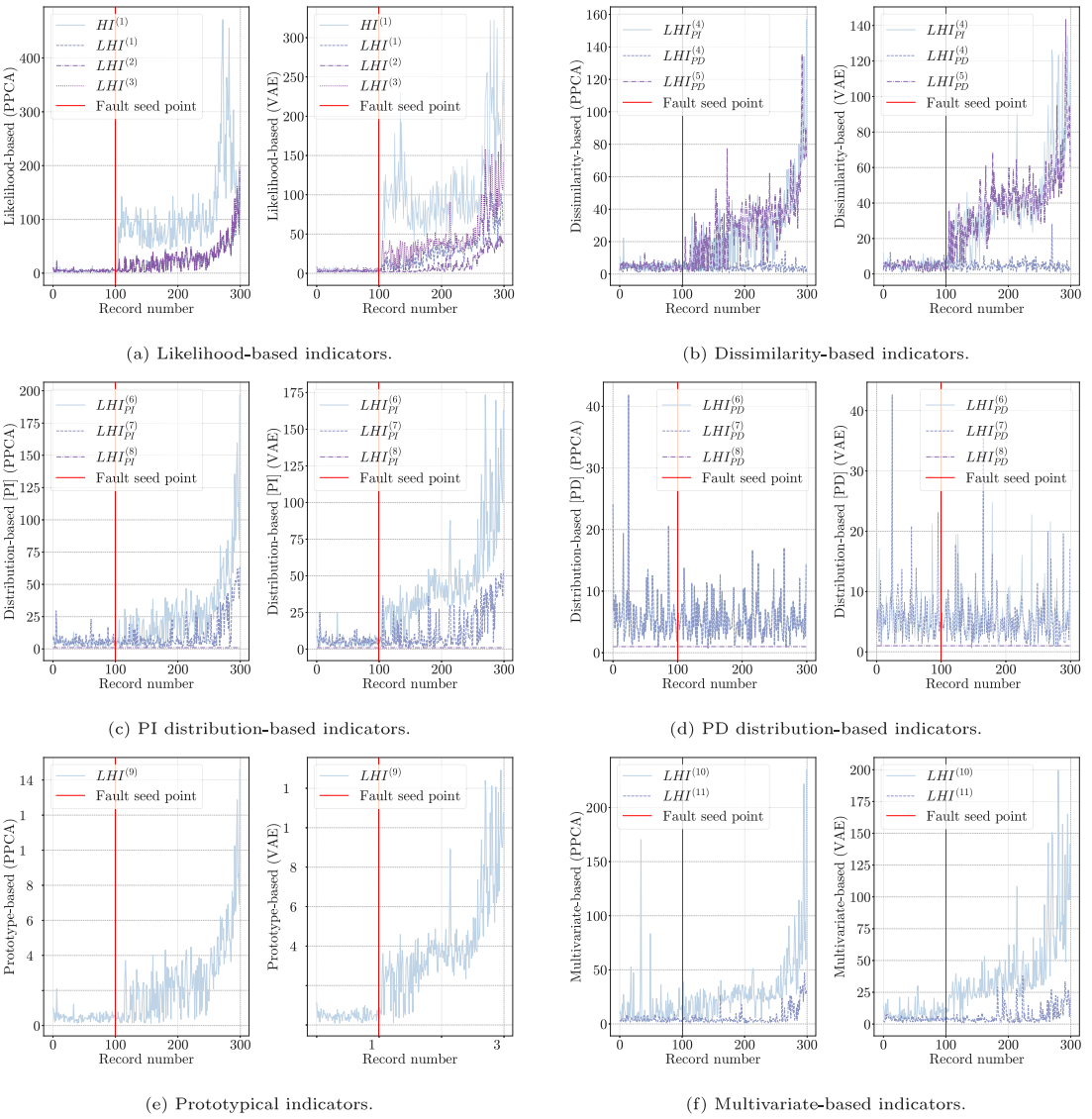


Fig. 15. The TSA GMM-based symmetric KL divergence trends for the HIs and LHIs considered in this work.

In Fig. 15 the GMM-based symmetric KL divergence trends using the post-processing methods discussed in Section 3.2.2 are shown. These trends are necessary to enable a quantitative metric comparison of the considered HIs and LHIs. Quantitative comparison metrics such as monotonicity and trendability, discussed in [81], are not considered here as we do not propose that the post-processing methodology used is a direct requirement for LVM-based condition inference. Fig. 15 shows that many of the metrics are responsive to the gear tooth fault. In Figs. 15(b) and 15(d), the Euclidean-based PD metrics are seen to be uninformative. However, as seen in Fig. 15(b), the angular metric $LHI^{(5)}_{PD}$, which is PD, is indicative of the damage present in the vibration data. The $LHI^{(8)}$ metric, which monitors the entropy of the latent posterior distribution, exhibits no response to the gear tooth fault. This result indicates that both the PPCA and VAE models respond to the damage in the data through the posterior mean $\boldsymbol{\mu}_{\mathcal{H}}$. While this is expected for the PPCA model, the VAE model, by design, had a posterior mean and a posterior covariance that were functions of the input data $\mathbf{x}$. This result emphasises the importance of considering metrics from the conceptual categories proposed in Section 2.5, as they ensure that all facets of the considered LVMs are understood. Fig. 15 indicates that both the PPCA and VAE models are useful for condition inference on the gearbox dataset. Thus, important considerations must be made when investigating alternative LVM formulations, and PPCA may serve as a useful baseline for performance comparison and quantification in future work.

In Fig. 16, the violin plots of the intra-class correlation at a per-method level are shown. Fig. 16 shows a clear correlation between indicators from different classes, and the metrics with weak correlation most likely belong to indicators evaluated in the PD setting,

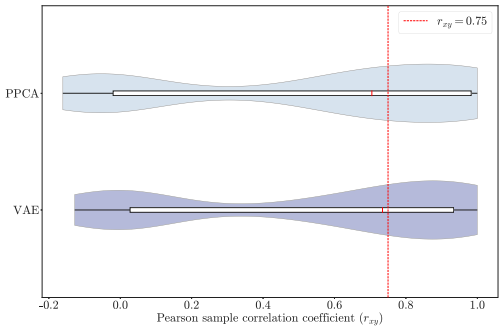
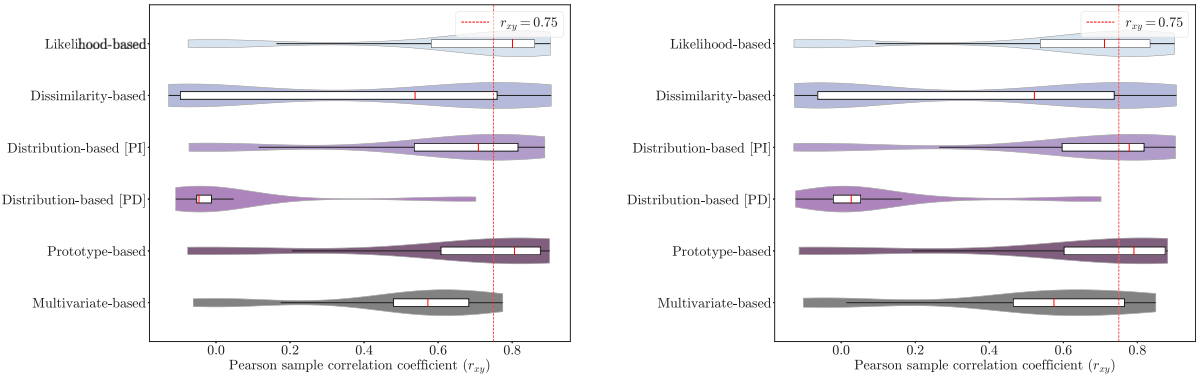


Fig. 16. Violin plots of the sample Pearson correlation coefficient on an intra-class level for the gearbox dataset. Note that the intra-class comparison is visualised on a per-method level as we compute all of the r_{xy} terms for all metrics of a given LVM method to compare class-wide similarity.



(a) PPCA method-wide correlation.

(b) VAE method-wide correlation.

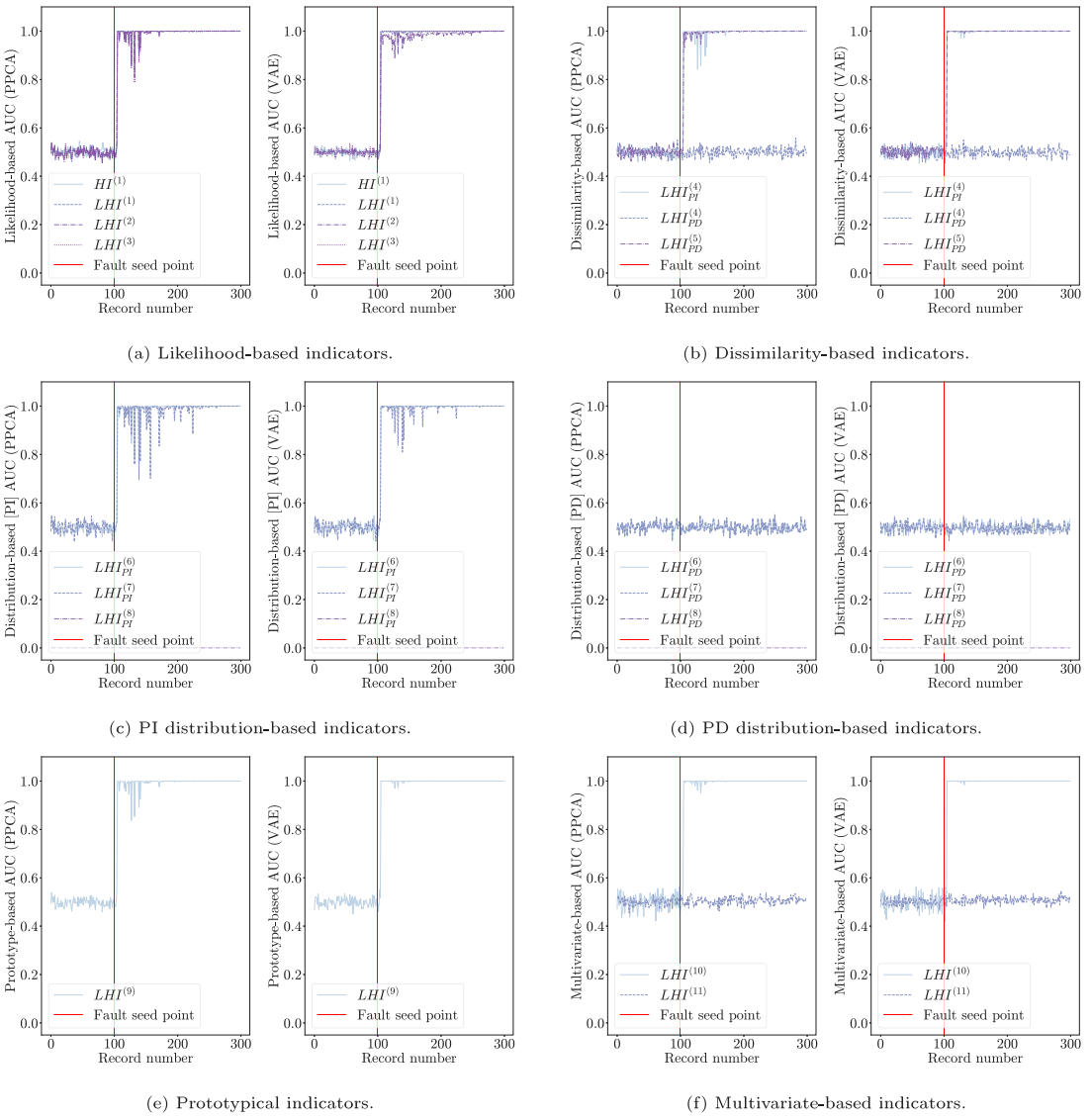
Fig. 17. Violin plots of the sample Pearson correlation coefficient on an inter-method level for the gearbox dataset. Note that the inter-method comparison is visualised on a per-class level as we compute all of the r_{xy} terms for all metrics of a given indicator class and all metrics from another LVM method to compare method-wide similarity.

namely $LHI_{PD}^{(4)}$, and $LHI_{PD}^{(6-8)}$. This highlights that there is some variety and uniqueness in the responsiveness of the indicators considered in this work.

Fig. 15 is used to further analyse the indicators that exhibit strong positive correlation. The $HI^{(1)}$ metric response is unique for the PPCA model, but all other metrics in the likelihood-based class respond similarly. Hence, using a single manifold-based metric from the three considered in this work would have been sufficient for condition inference. In the dissimilarity-based class, the $LHI_{PI}^{(4)}$ and $LHI_{PD}^{(5)}$ exhibit similar responses for the VAE model but not for the PPCA model. This indicates that the metrics in this class are unique and offer unique insight for this dataset. For metrics from the distribution-based class in the PI setting, the responsive metrics, $LHI_{PI}^{(6)}$ and $LHI_{PI}^{(7)}$, would exhibit strong positive correlation for both the PPCA and VAE models. This similarity indicates that choosing either of these metrics would have been suitable for condition inference on this dataset. However, Fig. 15(c) shows that $LHI_{PI}^{(6)}$ is more responsive to the development of the gear tooth fault. In Fig. 17 the violin plots are shown to enable an inter-method analysis. From Fig. 17, is clear that the classes with strong similarity are the dissimilarity-based metrics from the PPCA model, and the distribution-based metrics in the PI setting for the indicators from both models. The distribution-based metrics in the PD setting are outliers in the inter-method analysis, and this is attributed to their non-sensitivity to the presence of the gear tooth fault. This is distinct to the inter-method responses from the IMS dataset and highlights the uniqueness of the indicators proposed in this work and the requirement for an ensemble based latent inference procedure.

In Fig. 18 the AUC responses for the synchronous averaged indicator signals are shown using the approach described in Section 3.2.2. Fig. 18 shows that there are responsive and unresponsive LHIs in the conceptual categories. It is evident that the PD setting is a poor indicator framework for this dataset, but if we investigate the PD setting under a different coordinate system then indicator responsiveness can be obtained through $LHI_{PD}^{(5)}$. This highlights the necessity to consider metrics from different conceptual categories as each represent a different manifold perspective.

Hence, from the results in Figs. 14 and 15, it is possible to detect the gear tooth fault present in the data and establish the point of fault development using the proposed HIs and LHIs. The TSA results ensure that the tooth fault can be detected and isolated, and the analysis methodology detailed in Section 3.2.2 ensures that we can trend the TSAs for quantitative performance quantification. The Euclidean-based PD LHIs were found to respond inadequately to the gear tooth fault, which indicates that the manifold response to



(a) Likelihood-based indicators.

(b) Dissimilarity-based indicators.

(c) PI distribution-based indicators.

(d) PD distribution-based indicators.

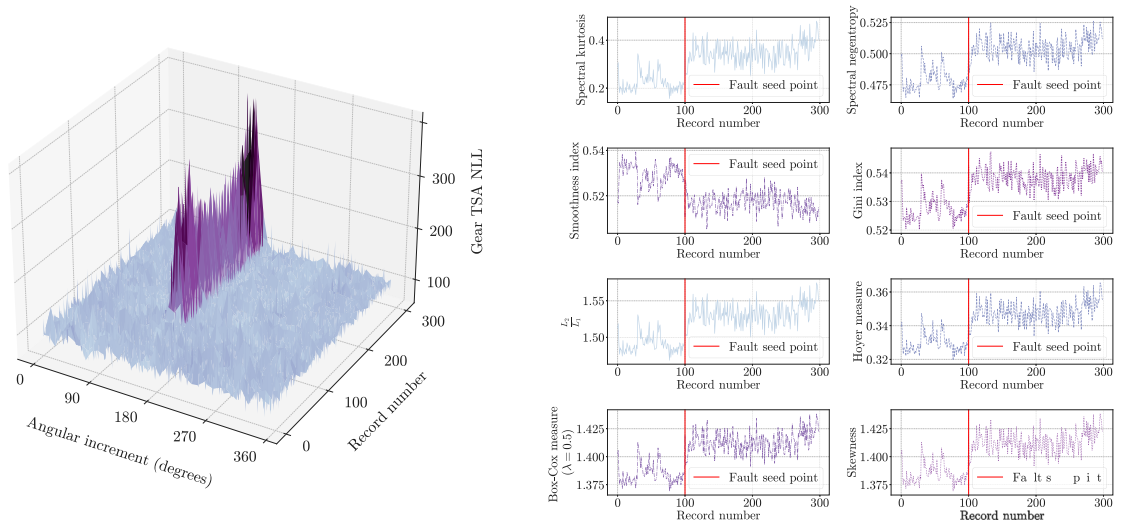
(e) Prototypical indicators.

(f) Multivariate-based indicators.

Fig. 18. The AUC for the synchronous average of the indicator signals around the gear tooth of interest. In (c) and (d) the AUC is zero as the $LHI^{(8)}$ metric for both models is insensitive to the tooth fault. An AUC of 0.5 indicates that the TSA of the HI and LHIs is indistinguishable from the indicators of the reference healthy set.

the tooth fault is to drive data away from the healthy manifold. By performing an inter-method and intra-class correlation analysis for the gearbox dataset: *(i)* the methods from the likelihood-based class that monitor the latent manifold offer similar insight, *(ii)* there is strong positive correlation for the responses for the metrics considered in this work between the PPCA and VAE models, *(iii)* the $LHI^{(6)}$ and $LHI^{(7)}$ metrics exhibit a similar response, and *(iv)* the PD setting is not useful for condition inference for the gearbox dataset. This dataset induces complexity through time-varying operating conditions, and the LVMs used in this work produce both a data space and a latent manifold that is practical for condition inference purposes. The proposed LHIs capture different perspectives of the latent manifold, and the quantification of the perspectives given in Section 2.5 ensures that the LHIs considered are sufficient for manifold condition inference.

Finally, in a comparison between the indicator responses exhibited for the IMS and gearbox dataset, detailed in Fig. 6 and Fig. 15 respectively, the uniqueness of the indicators presented in this work is exhibited. Not only do many indicators respond to the damage present in the data, but the responsiveness of the indicators is revealed to be dataset dependent. Furthermore, for the gearbox dataset, indicators are sensitive for vibration data with clear non-stationary OCs. These results advocate for the use of an ensemble-based LVM indicator approach as the latent manifold of LVMs is dependent on the fault conditions, the environment and operating conditions, and the asset of interest. By utilising indicators from each of the proposed conceptual categories, a richer information stream for condition inference becomes accessible and gives rise to an improved vibration-based CM methodology.



(a) TSA of the GMM NLL.

(b) Sparsity indicators.

Fig. 19. The GMM NLL metric from [82] and the sparsity metrics applied to the gearbox dataset.

3.2.4. *Comparison*

To compare the LHIs used in this work to HIs from other works, two HI methods were used. For the first method, the GMM-based NLL HI proposed in Heyns et al. [82] was used. Note that the GMM parameters were optimised using the same practices as for $LHI^{(1)}$. For the second method, the set of sparsity measures from Section 3.1.4 were calculated. In Fig. 19(a) the results using the methodology from [82] is shown. The proposed LHIs, in comparison with the methodology from [82], are useful and capable of exhibiting a clear response to the gear fault present in this dataset. While the methodology in [82] is designed for the detection of gear faults, the LHIs in our work exhibit similar responses without any directed pre-processing steps. Additionally, many of the LHIs exhibit a similar response to the GMM NLL, which further advocates for an ensemble-based condition inference approach as this allows for multiple diagnosis avenues. The sparsity measures in Fig. 19(b) are noticeably sensitive to non-stationary OCs not due to the gear tooth fault. While a change in condition is present around record 100, this may be attributed to the intervention step required to seed the tooth fault and the presence of the fault is not more pronounced in subsequent records. The lack of a clear degradation trend in the sparsity measures is attributed to the complexity of the gearbox dataset, and this highlights the necessity to consider focused post-processing practices to infer the fault type, as was done for the proposed LHIs. The considered sparsity metrics alone are not indicative enough to motivate asset maintenance.

4. Conclusion

The latent manifold of LVMs was shown in this work to be a powerful tool for condition inference. LVMs require no historical fault data, capture the intrinsic behaviour of an asset in a healthy state, and offer a data space and a latent space for condition inference. Not only does the LVM latent manifold exhibit a sensitivity to the presence of different faults, but the source of the damage can be accurately inferred in the presence of constant and non-stationary OCs. By considering two datasets with different fault sources, it is shown that the important facets of LVMs, the data space and the latent manifold, are useful for vibration-based CM. By considering several LHIs from the conceptual categories proposed in this work, it was shown that capturing different manifold perspectives is an important endeavour to understand all aspects of the considered models. The manifold perspectives are important as the latent manifold is unique to a given asset, and thus it is not possible to identify an optimal LHI or manifold perspective that generalises to all possible assets. Hence, the ensemble of LHIs considered in this work ensure that the condition inference process is robust and not dependent on the performance of a single indicator.

By considering the conclusions drawn from the per-dataset investigations and analysis procedures, a set of conclusions may be drawn on the indicators used in this work between the considered datasets. We summarise the main conclusions as follows:

1. The likelihood-based indicators that monitor the latent manifold ($LHI^{(1)}$, $LHI^{(2)}$, and $LHI^{(3)}$) did not offer unique responses on an inter-method and intra-class level. Hence, we recommend that future researchers start with the $HI^{(1)}$ and $LHI^{(1)}$ metrics as representative metrics from this class of indicators.
2. The $LHI^{(5)}_{PD}$ metric from the dissimilarity-based class offers unique insights into the latent manifold and the response therein to anomalous data. Hence, we recommend considering one dissimilarity metric in both the PI and PD setting and $LHI^{(5)}_{PD}$ as a starting point.

Table 2
The PPCA model properties used for model training.

Dataset type	L_w	L_{sft}	CCR
IMS	512	256	80%
Gearbox	512	256	80%

Table 3
The VAE model properties used for model training.

Dataset type	L_w	L_{sft}	Latent dimensionality	Number of training epochs	Learning rate	Batch size (B)
IMS	512	256	$d = 50$	2000	$1e^{-4}$	512
Gearbox	512	256	$d = 50$	2000	$1e^{-4}$	512

3. The distribution-based class is a useful class for LHI derivation. This work found that there was little noticeable difference between $LHI^{(6)}$ and $LHI^{(7)}$ on both datasets. Hence, we recommend using $LHI^{(6)}$ and $LHI^{(8)}$ as two representative metrics when starting off with this class.
4. The response of the PI and PD metrics are unique to the considered dataset. Hence, we recommend ensuring that both settings be investigated in future research into LVM-based CM methodologies.
5. There is clear benefit to an ensemble-based condition inference approach due to the intricacies associated with assets and the manifestation of faults in the observed vibration data. Utilising numerous metrics provides a modular fault diagnosis framework and can enrich the condition inference process.

In future investigations, focus must be given to improving the robustness of LVMs to complex operating environments and fault components that do not manifest through strong perturbations in the vibration data variance. Complex operating environments may induce extraneous impulsivity not related to the development of a fault, and such an environment impedes the condition inference task. LVM-based methodologies must be capable of overcoming this hurdle to remain a competitive alternative to state-of-the-art signal processing techniques. Faulty components that do not exhibit strong, detectable perturbations require LVM methodologies that are highly sensitive to changes in the asset state. Finally, it is also important to quantify how LVMs respond to assets that contain more complex phenomena.

CRediT authorship contribution statement

Ryan Balshaw: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **P. Stephan Heyns:** Conceptualization, Methodology, Supervision, Writing – review & editing, Project administration, Funding acquisition. **Daniel N. Wilke:** Conceptualization, Methodology, Supervision, Resources, Writing – review & editing. **Stephan Schmidt:** Conceptualization, Methodology, Supervision, Validation, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgement

The authors gratefully acknowledge the support that was received from AngloGold Ashanti, South Africa in the execution of this research.

Appendix. LVM model properties

It is pertinent that all applicable properties of the LVMs used in this work are detailed. Pytorch [83] was used to optimise the VAE model parameters. The system used to train the models had an Intel i7-8750H processor and a Geforce RTX 2070 graphics card. The properties of the PPCA and VAE models for the datasets considered are given in Tables 2 and 3 respectively. The cumulative contribution rate (CCR) was used to obtain the number of latent variables d for the PPCA model [31]. In Table 4 the network architectures of the VAE models used in this work are detailed. This work chose the latent encoding and data generation functions to be mirror images of each other. This was done by selecting the properties of the one-dimensional convolutional and de-convolutional layers such that an equivalent representation expansion or reduction is obtained from one layer to the next. The controllable properties of the convolutional and de-convolutional layers are the stride length L_{stride}, the kernel size L_{kernel}, and the padding length L_{pad}. For the VAE models, the validation model loss was monitored to ensure that no overfitting occurred. The model parameters for the training epoch with the lowest validation loss were retained and used for condition inference. The Adam algorithm [84,85] was used to optimise the VAE model parameters.

Table 4
The network architectures for both the VAE latent encoding networks and data generation networks. A depth order of 0–6 define the latent encoding network, while the generative network is the reverse of this order. Note that the posterior mean and covariance are both functions of the fourth network depth level. The parameter B is the batch size. The activation functions for the hidden layers of the encoding network are TanH, while for the generative network the hidden layer activation functions are ReLU. Note that ↓ denotes a convolutional layer, while ↑ denotes a de-convolutional layer.

Network depth	Layer operator	Output layer dimensionality	L_{stride}	L_{kernel}	L_{pad}	Activation function
0	Input layer	$\mathbb{R}^{B\times L_w}$	N/A	N/A	N/A	Linear
1	↓/↑	$\mathbb{R}^{B\times 32\times \frac{L_w}{4}}$	4	32	14	TanH (↓)/ReLU (↑)
2	↓/↑	$\mathbb{R}^{B\times 64\times \frac{L_w}{16}}$	4	32	14	TanH (↓)/ReLU (↑)
3	↓/↑	$\mathbb{R}^{B\times 128\times \frac{L_w}{64}}$	4	32	14	TanH (↓) /ReLU (↑)
4	Fully-connected	$\mathbb{R}^{B\times \frac{L_w}{2}}$	N/A	N/A	N/A	TanH (↓) /ReLU (↑)
5 (posterior mean)	Fully-connected	$\mathbb{R}^{B\times d}$	N/A	N/A	N/A	Linear
6 (posterior covariance)	Fully-connected	$\mathbb{R}^{B\times d}$	N/A	N/A	N/A	Softplus

References

[1] R.B. Randall, Vibration-based condition monitoring: Industrial, aerospace and automotive applications, Wiley, Chichester, UK, 2011, pp. 1–23, http://dx.doi.org/10.1002/9780470977668, URL https://onlinelibrary.wiley.com/doi/book/10.1002/9780470977668 (Chapter 1).

[2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to RUL prediction, Mech. Syst. Signal Process. 104 (2018) 799–834, http://dx.doi.org/10.1016/j.ymssp.2017.11.016, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327017305988.

[3] R. Zimroz, W. Bartelmus, T. Barszcz, J. Urbanek, Diagnostics of bearings in presence of strong operating conditions non-stationarity—A procedure of load-dependent features processing with application to wind turbine bearings, Mech. Syst. Signal Process. 46 (1) (2014) 16–27, http://dx.doi.org/10.1016/j.ymssp.2013.09.010, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327013004974.

[4] A.K. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (7) (2006) 1483–1510, http://dx.doi.org/10.1016/j.ymssp.2005.09.012, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327005001512.

[5] S. Khan, T. Yairi, A review on the application of deep learning in system health management, Mech. Syst. Signal Process. 107 (2018) 241–265, http://dx.doi.org/10.1016/j.ymssp.2017.11.024, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327017306064.

[6] R.B. Randall, J. Antoni, Rolling element bearing diagnostics - A tutorial, Mech. Syst. Signal Process. 25 (2) (2011) 485–520, http://dx.doi.org/10.1016/j.ymssp.2010.07.017, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327010002530.

[7] J. Antoni, Cyclostationarity by examples, Mech. Syst. Signal Process. 23 (4) (2009) 987–1036, http://dx.doi.org/10.1016/j.ymssp.2008.10.010, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327008002690.

[8] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, Mech. Syst. Signal Process. 108 (2018) 33–47, http://dx.doi.org/10.1016/j.ymssp.2018.02.016, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327018300748.

[9] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A.K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech. Syst. Signal Process. 138 (2020) 106587, http://dx.doi.org/10.1016/j.ymssp.2019.106587, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327019308088.

[10] S. Khan, T. Yairi, A review on the application of deep learning in system health management, Mech. Syst. Signal Process. 107 (2018) 241–265, http://dx.doi.org/10.1016/j.ymssp.2017.11.024, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327017306064.

[11] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237, http://dx.doi.org/10.1016/j.ymssp.2018.05.050, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327018303108.

[12] W. Booyse, D.N. Wilke, S. Heyns, Deep digital twins for detection, diagnostics and prognostics, Mech. Syst. Signal Process. 140 (2020) 106612, http://dx.doi.org/10.1016/j.ymssp.2019.106612, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327019308337.

[13] R. Balshaw, P.S. Heyns, D.N. Wilke, S. Schmidt, Importance of temporal preserving latent analysis for latent variable models in fault diagnostics of rotating machinery, Mech. Syst. Signal Process. 168 (November 2021) (2022) 108663, http://dx.doi.org/10.1016/j.ymssp.2021.108663, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327021009870.

[14] P. Večeř, M. Kreidl, R. Šmíd, Condition indicators for gearbox condition monitoring systems, Acta Polytech. 45 (6) (2005) 35–43, http://dx.doi.org/10.14311/782, URL https://ojs.cvut.cz/ojs/index.php/ap/article/view/782.

[15] N. Sawalhi, R. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mech. Syst. Signal Process. 21 (6) (2007) 2616–2633, http://dx.doi.org/10.1016/j.ymssp.2006.12.002, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327006002603.

[16] D. Wang, Y. Chen, C. Shen, J. Zhong, Z. Peng, C. Li, Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring, Mech. Syst. Signal Process. 168 (November 2021) (2022) 108673, http://dx.doi.org/10.1016/j.ymssp.2021.108673, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327021009961.

[17] D. Wang, K.-L. Tsui, Q. Miao, Prognostics and health management: A review of vibration based bearing and gear health indicators, IEEE Access 6 (2018) 665–676, http://dx.doi.org/10.1109/ACCESS.2017.2774261, URL http://ieeexplore.ieee.org/document/8115325/.

[18] J. Antoni, R. Randall, The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process. 20 (2) (2006) 308–331, http://dx.doi.org/10.1016/j.ymssp.2004.09.002, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327004001529.

[19] W.A. Smith, P. Borghesani, Q. Ni, K. Wang, Z. Peng, Optimal demodulation-band selection for envelope-based diagnostics: A comparative study of traditional and novel tools, Mech. Syst. Signal Process. 134 (2019) 106303, http://dx.doi.org/10.1016/j.ymssp.2019.106303, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327019305242.

[20] R.B. Randall, N. Sawalhi, A new method for separating discrete components from a signal, in: Sound and Vibration, Vol. 45, No. 5, 2011, pp. 6–9, URL http://www.sandv.com/downloads/1105rand.pdf.

[21] S. Schmidt, P.S. Heyns, An open set recognition methodology utilising discrepancy analysis for gear diagnostics under varying operating conditions, Mech. Syst. Signal Process. 119 (2019) 1–22, http://dx.doi.org/10.1016/j.ymssp.2018.09.016, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327018306344.

[22] Y. Miao, B. Zhang, J. Lin, M. Zhao, H. Liu, Z. Liu, H. Li, A review on the application of blind deconvolution in machinery fault diagnosis, Mech. Syst. Signal Process. 163 (June 2021) (2022) 108202, http://dx.doi.org/10.1016/j.ymssp.2021.108202.

[23] J. An, S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, in: Special Lecture on IE, Vol. 2, No. 2, 2, (2) 2015, pp. 1—-18, URL http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf.

[24] N. Gugulothu, V. TV, P. Malhotra, L. Vig, P. Agarwal, G. Shroff, Predicting remaining useful life using time series embeddings based on recurrent neural networks, CEUR Workshop Proc. 2657 (2017) 1–9, arXiv:1709.01073.

[25] K.H. Kim, S.-H. Shim, Y. Lim, J. Jeon, J.-W. Choi, B. Kim, A.S. Yoon, RaPP: Novelty detection with reconstruction along projection pathway, in: 8th International Conference on Learning Representations, ICLR 2020, OpenReview.net, 2020, URL https://openreview.net/forum?id=HkgeGeBYDB.
[26] A. González-Muñiz, I. Díaz, A.A. Cuadrado, D. García-Pérez, Health indicator for machine condition monitoring built in the latent space of a deep autoencoder, Reliab. Eng. Syst. Saf. 224 (March) (2022) 108482, http://dx.doi.org/10.1016/j.ress.2022.108482, URL https://linkinghub.elsevier.com/retrieve/pii/S0951832022001417.
[27] J. Park, Y.G. Jung, A. Beng Jin Teoh, Discriminative multi-level reconstruction under compact latent space for one-class novelty detection, in: 2020 25th International Conference on Pattern Recognition, ICPR, IEEE, 2021, pp. 7095–7102, http://dx.doi.org/10.1109/ICPR48806.2021.9413248, arXiv:2003.01665, URL https://ieeexplore.ieee.org/document/9413248/.
[28] G. Kwon, M. Prabhushankar, D. Temel, G. AlRegib, Backpropagated gradient representations for anomaly detection, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16, Springer, 2020, pp. 206–226, http://dx.doi.org/10.1007/978-3-030-58589-1_13, URL https://link.springer.com/10.1007/978-3-030-58589-1_13.
[29] M.E. Tipping, C.M. Bishop, Probabilistic principal component analysis, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (3) (1999) 611–622, http://dx.doi.org/10.1111/1467-9868.00196, URL https://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00196.
[30] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, 2013, arXiv:1312.6114.
[31] C. Bishop, Pattern recognition and machine learning, first ed., Springer New York, NY, 2006, p. 738, URL https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/.
[32] D.M. Blei, Build, compute, critique, repeat: Data analysis with latent variable models, Annu. Rev. Stat. Appl. 1 (1) (2014) 203–232, http://dx.doi.org/10.1146/annurev-statistics-022513-115657, URL https://www.annualreviews.org/doi/10.1146/annurev-statistics-022513-115657.
[33] C. Fefferman, S. Mitter, H. Narayanan, Testing the manifold hypothesis, J. Amer. Math. Soc. 29 (4) (2016) 983–1049, http://dx.doi.org/10.1090/jams/852, URL https://www.ams.org/jams/2016-29-04/S0894-0347-2016-00852-4/.
[34] D.P. Kingma, M. Welling, An introduction to variational autoencoders, Found. Trends® Mach. Learn. 12 (4) (2019) 307–392, http://dx.doi.org/10.1561/2200000056, arXiv:1906.02691.
[35] P.J. Diggle, R.J. Gratton, Monte Carlo methods of inference for implicit statistical models, J. R. Stat. Soc. Ser. B Stat. Methodol. 46 (2) (1984) 193–212, http://dx.doi.org/10.1111/j.2517-6161.1984.tb01290.x, URL http://www.jstor.org/stable/2345504.
[36] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational inference: A review for statisticians, J. Amer. Statist. Assoc. 112 (518) (2017) 859–877, http://dx.doi.org/10.1080/01621459.2017.1285773, arXiv:arXiv:1601.00670v9, URL https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1285773.
[37] I. Kobyzev, S.J. Prince, M.A. Brubaker, Normalizing flows: An introduction and review of current methods, IEEE Trans. Pattern Anal. Mach. Intell. 43 (11) (2021) 3964–3979, http://dx.doi.org/10.1109/TPAMI.2020.2992934, URL https://ieeexplore.ieee.org/document/9089305/.
[38] G. Papamakarios, E. Nalisnick, D.J. Rezende, S. Mohamed, B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, 2019, arXiv:1912.02762.
[39] Y. Lecun, S. Chopra, R. Hadsell, M. Ranzato, F.J. Huang, A tutorial on energy-based learning, in: Predicting Structured Data, MIT Press, 2006, URL http://yann.lecun.com/exdb/publis/pdf/lecun-06.pdf.
[40] S. Mohamed, B. Lakshminarayanan, Learning in implicit generative models, 2016, arXiv:1610.03483.
[41] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT Press, 2016, URL http://www.deeplearningbook.org.
[42] S. Bond-Taylor, A. Leach, Y. Long, C.G. Willcocks, Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7327–7347, http://dx.doi.org/10.1109/TPAMI.2021.3116668, URL https://ieeexplore.ieee.org/document/9555209/.
[43] D. Eisenbud, Linear sections of determinantal varieties, Amer. J. Math. 110 (3) (1988) 541, http://dx.doi.org/10.2307/2374622, URL https://www.jstor.org/stable/2374622?origin=crossref.
[44] L. Ruff, J.R. Kauffmann, R.A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T.G. Dietterich, K.-R. Muller, A unifying review of deep and shallow anomaly detection, Proc. IEEE 109 (5) (2021) 756–795, http://dx.doi.org/10.1109/JPROC.2021.3052449, URL https://ieeexplore.ieee.org/document/9347460/.
[45] M. Hubert, M. Debruyne, P.J. Rousseeuw, Minimum covariance determinant and extensions, WIREs Comput. Stat. 10 (3) (2018) http://dx.doi.org/10.1002/wics.1421, arXiv:1709.07045, URL https://onlinelibrary.wiley.com/doi/10.1002/wics.1421.
[46] C.M. Bishop, M. Svensén, C.K.I. Williams, GTM: The generative topographic mapping, Neural Comput. 10 (1) (1998) 215–234, http://dx.doi.org/10.1162/089976698300017953, URL https://direct.mit.edu/neco/article/10/1/215-234/6127.
[47] C.M. Bishop, M. Svensén, C.K. Williams, Developments of the generative topographic mapping, Neurocomputing 21 (1–3) (1998) 203–224, http://dx.doi.org/10.1016/S0925-2312(98)00043-5, URL https://linkinghub.elsevier.com/retrieve/pii/S0925231298000435.
[48] D.M. Tax, R.P. Duin, Support vector domain description, Pattern Recognit. Lett. 20 (11–13) (1999) 1191–1199, http://dx.doi.org/10.1016/S0167-8655(99)00087-2, URL https://linkinghub.elsevier.com/retrieve/pii/S0167865599000872.
[49] B. Schölkopf, R.C. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, Adv. Neural Inf. Process. Syst. (1999) 582–588, URL https://proceedings.neurips.cc/paper/1999/file/8725fb777f25776ffa9076e44fcfd776-Paper.pdf.
[50] R. Chalapathy, A.K. Menon, S. Chawla, Anomaly detection using one-class neural networks, 2018, arXiv:1802.06360.
[51] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S.A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: Proceedings of the 35th International Conference on Machine Learning, Vol. 10, PMLR, 2018, pp. 4393—-4402, URL http://proceedings.mlr.press/v80/ruff18a/ruff18a.pdf.
[52] L. Kaufman, P.J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, USA, 1990, http://dx.doi.org/10.1002/9780470316801, URL http://doi.wiley.com/10.1002/9780470316801,
[53] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybernet. 43 (1) (1982) 59–69, http://dx.doi.org/10.1007/BF00337288, URL http://link.springer.com/10.1007/BF00337288.
[54] N.E. Huang, S.R. Long, Z. Shen, The mechanism for frequency downshift in nonlinear wave evolution, in: Advances in Applied Mechanics, Vol. 32, No. C, 1996, pp. 59–117C, http://dx.doi.org/10.1016/S0065-2156(08)70076-0, URL https://linkinghub.elsevier.com/retrieve/pii/S0065215608700760.
[55] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995, http://dx.doi.org/10.1098/rspa.1998.0193, URL https://royalsocietypublishing.org/doi/10.1098/rspa.1998.0193.
[56] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544, http://dx.doi.org/10.1109/TSP.2013.2288675, URL http://ieeexplore.ieee.org/document/6655981/.
[57] N. ur Rehman, D. Mandic, Empirical mode decomposition for trivariate signals, IEEE Trans. Signal Process. 58 (3) (2010) 1059–1068, http://dx.doi.org/10.1109/TSP.2009.2033730, URL http://ieeexplore.ieee.org/document/5280229/.
[58] N.U. Rehman, H. Aftab, Multivariate variational mode decomposition, IEEE Trans. Signal Process. 67 (23) (2019) 6039–6052, http://dx.doi.org/10.1109/TSP.2019.2951223, URL https://ieeexplore.ieee.org/document/8890883/.
[59] R.A. Wiggins, Minimum entropy deconvolution, Geoexploration 16 (1–2) (1978) 21–35, http://dx.doi.org/10.1016/0016-7142(78)90005-4, URL https://linkinghub.elsevier.com/retrieve/pii/0016714278900054.
[60] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: Forecasting and control, fifth ed., in: Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, New Jersey, 2015, URL https://www.proquest.com/scholarly-journals/time-series-analysis-forecasting-control-5th/docview/1949016286/se-2.

[61] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1) (1977) 1–22, http://dx.doi.org/10.1111/j.2517-6161.1977.tb01600.x, URL http://www.jstor.org/stable/2984875.
[62] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (C) (1987) 53–65, http://dx.doi.org/10.1016/0377-0427(87)90125-7, URL https://linkinghub.elsevier.com/retrieve/pii/0377042787901257.
[63] H. Qiu, J. Lee, J. Lin, G. Yu, R.T. Services (2007), IMS, University of Cincinnati. Bearing Data Set, NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007, URL http://ti.arc.nasa.gov/project/prognostic-data-repository.
[64] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, J. Sound Vib. 289 (4–5) (2006) 1066–1090, http://dx.doi.org/10.1016/j.jsv.2005.03.007, URL https://linkinghub.elsevier.com/retrieve/pii/S0022460X0500221X.
[65] D.W. Hogg, Data analysis recipes: Probability calculus for inference, 2012, pp. 1–19, arXiv:1205.4446.
[66] T.Y. Lee, J.V. Zidek, N. Heckman, Dimensional analysis in statistical modelling, 2020, pp. 1–41, arXiv:2002.11259.
[67] W. Gousseau, J. Antoni, F. Girardin, J. Griffaton, Analysis of the rolling element bearing data set of the center for intelligent maintenance systems of the university of cincinnati, in: 13th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2016/MFPT 2016, 2016, URL https://hal.science/hal-01715193/file/216-Gousseau.pdf.
[68] Y. Wang, Y. Peng, Y. Zi, X. Jin, K.-L. Tsui, A two-stage data-driven-based prognostic approach for bearing degradation problem, IEEE Trans. Ind. Inform. 12 (3) (2016) 924–932, http://dx.doi.org/10.1109/TII.2016.2535368, URL http://ieeexplore.ieee.org/document/7420685/.
[69] J. Antoni, The infogram: Entropic evidence of the signature of repetitive transients, Mech. Syst. Signal Process. 74 (2016) 73–94, http://dx.doi.org/10.1016/j.ymssp.2015.04.034, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327015002174.
[70] D. Wang, Z. Peng, L. Xi, The sum of weighted normalized square envelope: A unified framework for kurtosis, negative entropy, gini index and smoothness index for machine health monitoring, Mech. Syst. Signal Process. 140 (2020) 106725, http://dx.doi.org/10.1016/j.ymssp.2020.106725, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327020301114.
[71] I.S. Bozchalooi, M. Liang, A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection, J. Sound Vib. 308 (1–2) (2007) 246–267, http://dx.doi.org/10.1016/j.jsv.2007.07.038, URL https://linkinghub.elsevier.com/retrieve/pii/S0022460X07005871.
[72] Y. Miao, M. Zhao, J. Lin, Improvement of kurtosis-guided-grams via Gini index for bearing fault feature identification, Meas. Sci. Technol. 28 (12) (2017) 125001, http://dx.doi.org/10.1088/1361-6501/aa8a57, URL https://iopscience.iop.org/article/10.1088/1361-6501/aa8a57.
[73] D. Wang, Spectral L2/L1 norm: A new perspective for spectral kurtosis for characterizing non-stationary signals, Mech. Syst. Signal Process. 104 (2018) 290–293, http://dx.doi.org/10.1016/j.ymssp.2017.11.013, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327017305952.
[74] N. Hurley, S. Rickard, Comparing measures of sparsity, IEEE Trans. Inform. Theory 55 (10) (2009) 4723–4741, http://dx.doi.org/10.1109/TIT.2009.2027527, arXiv:0811.4706, URL http://ieeexplore.ieee.org/document/5238742/.
[75] D. Wang, J. Zhong, C. Li, Z. Peng, Box-cox sparse measures: A new family of sparse measures constructed from kurtosis and negative entropy, Mech. Syst. Signal Process. 160 (2021) 107930, http://dx.doi.org/10.1016/j.ymssp.2021.107930, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327021003253.
[76] C. Peeters, J. Antoni, J. Helsen, Blind filters based on envelope spectrum sparsity indicators for bearing and gear vibration-based condition monitoring, Mech. Syst. Signal Process. 138 (2020) 106556, http://dx.doi.org/10.1016/j.ymssp.2019.106556, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327019307770.
[77] S. Schmidt, P.S. Heyns, Localised gear anomaly detection without historical data for reference density estimation, Mech. Syst. Signal Process. 121 (2019) 615–635, http://dx.doi.org/10.1016/j.ymssp.2018.11.051, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327018307696.
[78] Y. Chen, S. Schmidt, P.S. Heyns, M.J. Zuo, A time series model-based method for gear tooth crack detection and severity assessment under random speed variation, Mech. Syst. Signal Process. 156 (2021) 107605, http://dx.doi.org/10.1016/j.ymssp.2020.107605, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327020309912.
[79] S. Braun, The synchronous (time domain) average revisited, Mech. Syst. Signal Process. 25 (4) (2011) 1087–1102, http://dx.doi.org/10.1016/j.ymssp.2010.07.016, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327010002529.
[80] D. Diamond, P. Heyns, A. Oberholster, Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression, Mech. Syst. Signal Process. 81 (2016) 402–418, http://dx.doi.org/10.1016/j.ymssp.2016.02.060, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327016001114.
[81] Y. Lei, J. Lin, Z. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 35 (1–2) (2013) 108–126, http://dx.doi.org/10.1016/j.ymssp.2012.09.015, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327012003731.
[82] T. Heyns, P. Heyns, J. de Villiers, Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox, Mech. Syst. Signal Process. 32 (2012) 200–215, http://dx.doi.org/10.1016/j.ymssp.2012.05.008, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327012002221.
[83] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019) arXiv:1912.01703.
[84] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv:1412.6980.
[85] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2017, arXiv:1711.05101.