This is a repository copy of *Measuring data similarity in population-based structural health monitoring using distance metrics*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/206480/

Version: Published Version

Measuring data similarity in population-based structural health monitoring using distance metrics

Chandula T Wickramarachchi[1] ⓘ, **Eoghan Maguire**[2] ⓘ, **Elizabeth J Cross**[1] and **Keith Worden**[1]

Abstract
Population-based structural health monitoring (PBSHM) expands structural health monitoring (SHM) from a single structure to a group of structures, allowing inferences to be made within and between populations by transferring knowledge across them. Within the populations of interest, the similarity of structures, via their corresponding data, should be assessed to successfully implement PBSHM. This paper focusses on using distance metrics to assess similarity at the very start of the analysis chain, to discover information about a population for which there is little prior knowledge and before any analysis has taken place on individual structures. By doing so, it is possible to quickly and automatically identify abnormalities within the population, group similarly behaving structures together, and inform further decisions. The suitability of several candidate metrics that are not widely employed in SHM are tested using a number of commonly occurring feature behaviours, such as varying amplitudes and temporary mean shifts. The effect of data normalisation/standardisation on the metrics is also explored to identify interesting behaviours within the data. A case study is then presented where distance metrics are used to discover similarities and dissimilarities within temperature data from turbines in an offshore wind farm.

Keywords
Population-based structural health monitoring, similarity, distance metrics, wind turbines

Introduction

In civil and aerospace structures such as bridges and wind turbines, the presence of damage can be costly; damage can lead to failures that cause expensive downtimes, or in extreme cases, put lives at risk. Understandably, developing methods to identify damage and predict its progression is immensely important in order to avoid catastrophic failures. In general, current techniques address these tasks via regular manual inspections and expert engineering knowledge, which is time consuming and expensive to undertake. Structural health monitoring (SHM), seeks to detect and diagnose damage automatically from monitoring data collected from structures, reducing the manual burden on owners. The aim of SHM is to reduce unnecessary regular inspections and gain insights and predict behaviours using data and model-based methods. Ideally, this task would be formed as a supervised learning problem, where models are trained on damage-state labelled data. However, this necessitates measurements of a structure in different possible damage states, which

may include multiple-site and multiple types of damage. In reality, it is rare that monitoring data are available from structures in *any* of their damage states; typically, measurement systems that capture damage are rare, defects are usually repaired prior to damage to reduce risk to users, and, more importantly, data from a single structure is unlikely to contain all types of damage it can possibly experience – resulting in a training phase of the SHM model that will not generalise in the testing phase. Consequently, focussing on single structures in a standard SHM setting can give a limited view of the variety of damage that is possible.

[1]Dynamics Research Group, Department of Mechanical Engineering, The University of Sheffield, Sheffield, South Yorkshire, UK
[2]Vattenfall Research and Development, New Renewables, Edinburgh, Scotland

Corresponding author:
Chandula T Wickramarachchi, Dynamics Research Group, Department of Mechanical Engineering, The University of Sheffield, Mappin Street, Sheffield, South Yorkshire S1 3JD, UK.
Email: c.t.wickramarachchi@sheffield.ac.uk

Population-based structural health monitoring (PBSHM),[1–8] aims to address this shortcoming by studying entire groups of structures simultaneously, thereby increasing the knowledge-base of structural behaviours and damage states. Interestingly, the population of structures here can be *homogeneous* (structures within the population are nominally identical) or *heterogeneous*.

One of the forerunning technologies that aid PBSHM in achieving health-state inference across populations is *transfer learning*.[9] Transfer learning aims to improve the performance on target domains by leveraging information from related source domains. In the context of PBSHM, the target domain contains structures with incomplete data (such as missing/lack of damage labels), and the source domain contains structures with labelled data. For transfer learning, similarity between domains is a key factor. Dissimilar domains can lead to a phenomenon named *negative transfer*,[9] where the performance in the target is negatively affected by leveraging information from a source domain, that is, the performance is worse as a result of transfer than using just target data in the model. Therefore, to infer health states of one structure by studying a population of others, there is a clear need to determine similarity within populations, in order to identify and group similarly behaving structures.[6,10] Similarity, as it happens, is also an important consideration in standard supervised SHM methods, because unexpected disparities between training and testing data can lead to reduced model performances during testing.

In PBSHM, similarity can be measured at two levels: structural similarity and feature similarity. The hypothesis here is that feature data collected from structures is best suited for transfer learning if the structures are similar in their composition. Results from a recent study on PBSHM supports this hypothesis.[2,11] Structural similarity considers the dynamic behaviour of structures based on their construction and boundary conditions. By representing structures as attributed graphs, similarity of structures can be measured.[2,12] The converse hypothesis to the aforementioned 'similarity in structure implies similarity in data', is 'similarity in data implies similarity in structure',[13] and the wish in PBSHM is to enable assessment of similarity in both cases (and eventually simultaneously). In this paper, the focus is on measuring similarity of monitoring data and descriptive features collected from structures. It is important to understand the behaviour of data during normal operating conditions, under environmental and operational variations (EOVs), and during deterioration and/or damage, in order to identify the most suitable data for transfer.

It is well understood that humans are experts in identifying changes or anomalies and recognising patterns in their surroundings.[14] Unfortunately, as populations grow in size and variety (vast sensor networks, multiple structures within homogenous and heterogeneous populations), manual analysis of large monitoring datasets becomes an impossibility. There is a need, therefore, to develop preliminary methods to identify similar and anomalous behaviours/trends in data quickly and automatically in a principled way, that forgo the need for time-consuming and in-depth analysis of individual datasets in a population. The intention is that this information may then be initially used to group structures together, highlight structures or components of the data that are performing differently from the majority of the population, etc., and inform further decisions such as, establish whether standard machine-learning approaches are appropriate across a population (i.e. find suitable training and testing data), and assess whether transfer learning is required.

There are many stages in a monitoring campaign where measuring similarity in features may be helpful. This paper focusses on methods that can be used at the very start of the analysis chain, to discover information about a population for which there is little prior knowledge and before any analysis has taken place on individual structures. Therefore, the methods suggested for assessing similarity in this paper do not speculate, infer, or aim to identify the condition of the structure (from which the data were collected), or the cause of any identified dissimilarities – whether the data are affected by interesting behaviours such as EOVs and damage, or whether they are a result of issues in the measurement chain (sensor damage, for example). Consequently, the underlying structure of the data could take any form, and the suggested methods of assessing similarity should be flexible enough to accommodate this.

To assess the similarity and to discover information about the data, the idea of a metric is employed here by treating the feature space (The feature space here is the space in which all features that correspond to each structure in a population lie) F as a metric space. Specifically, distance metrics that obey the rules set by Dudley[15] will be explored, that is, the distances metrics are non-negative ($D(x, y) \geqslant 0$), give a value of zero if-and-only-if the two features are the same ($D(x, y) = 0$ if and only if $x = y$), symmetric ($D(x, y) = D(y, x)$), and obey the triangle inequality ($D(x, z) \leqslant D(x, y) + D(y, z)$). Here, D is the distance and x, y and z are quantities. Unlike in SHM, where a single metric might be used for assessment without much consideration of others,[16,17] in PBSHM there is no precursor or historical studies that evaluate the most suitable metric for assessing similarity in this new context. The approach in this

paper is to compare available metrics – that give an indication of significant variability in the data – across a population, for a given feature F_i. For simplicity, only time-domain features are considered here.

This paper also considers the behaviour of features and metrics as the data are normalised/standardised, because it is a common pre-processing step in SHM[18–20] and PBSHM.[1,21] Dissimilarities in standardised data can indicate the presence of interesting behaviours (such as, higher-order moments). In PBSHM, standardisation helps bring source and target domains into the same space for domain adaptation. Methods such as normal condition alignment (that standardises and aligns data when structures are under normal operating condition) has proven incredibly useful for PBSHM in transferring knowledge across structures.[22,23] Therefore, the behaviour of metrics as data are standardised is an important consideration for testing their effectiveness for PBSHM.

Original contributions

This is the first study that measures similarity of features across a population of structures to aid transfer learning in population-based SHM – a relatively new concept that aims to transfer information across groups of structures to increase the pool of available information, in order to increase performance of monitoring campaigns. Principled methods of assessing large amounts of engineering data quickly and automatically are necessary for PBSHM, although the methods suggested here can also help reduce analysis times for typical SHM steps such as identifying suitable training and testing data.

The study introduces distance metrics in the context of PBSHM, to automatically assess similarity between *distributions* of data which the authors argue are better suited to compare populations of structures (e.g. all wind turbines in a wind farm). Here, a survey of possibly useful distance metrics – that are not all widely used in SHM currently – is conducted to assess similarity of features in PBSHM.

The distance metrics are then tested against typical commonly occurring changes in data-sets and features that originate from SHM programmes. The responses of the metrics to the commonly occurring changes are mathematically formalised. The suitability and sensitivity of the chosen metrics to the feature changes in PBSHM are presented. The most appropriate candidate metric for a given feature behaviour is determined to aid fast and automatic similarity assessment.

The effect of data normalisation/standardisation on distance metrics is mathematically derived because it is a common analysis technique in domain adaptation for PBSHM, and a typical pre-processing step in SHM.

The metrics that are most suitable for determining similarity in standardised data are explored and presented.

The aforementioned methods are first tested on a simulated data-set, and then applied to a real, homogeneous population of wind turbines from the Lillgrund farm, to test their effectiveness in assessing similarity for PBSHM.

Related work

As the field of PBSHM is relatively new – and this is an early contribution for measuring feature similarity for PBSHM – there are only a few small number of examples of related work. Nevertheless, a number of studies that focus on PBSHM state that similarity is a key consideration in successful transfer within PBSHM. When transferring damage localisation information between aircraft wings, similarity between structures aided a more fruitful transfer.[11] When modelling a homogeneous offshore wind turbine population, the similarity was measured using the Fréchet number, a metric that takes the location and ordering of the points along the curves of two datasets into account.[24] This measure, however, is not studied in detail in a PBSHM context to understand its sensitivity to specific feature types. In bridge monitoring, the cross-modal assurance criterion was used to assess the similarity of bridges by studying their natural frequencies. In the current paper, a different approach is considered where time-domain features are explored as, typically, the data are available in the time domain at the start of the analysis chain.[25] Usually, many studies in PBSHM literature acknowledge that similarity measurement is vital, and conducts it visually[23,26,27] or using a specific metric.[24,25] This paper attempts to find a principled method of assessing feature similarity in PBSHM that provides a guideline on which metrics are most suitable for a given commonly occurring feature found in SHM and PBSHM.

Distance metrics

This paper addresses the initial analysis of population data. At this stage, the assumption is that there is little available information about the underlying structure of the data. As a result, it is not helpful to use a simple metric such as the Euclidean distance between means of two datasets to measure similarity. When assessing stochastic data from engineering structures, a more insightful metric would compare the density functions of the data, that is, it is possible to use distance metrics to assess whether two probability distributions ($p_i(\boldsymbol{x})$ and $p_j(\boldsymbol{x})$), or two probability density estimates ($\hat{p}_i(\boldsymbol{x})$ and $\hat{p}_j(\boldsymbol{x})$) are the same or different by measuring their overlap.[28] Many have studied the use of metrics based on probability densities because of their relevance for

information and probability theory as well as statistics.[29] In 2009, Sriperumbudur et al.[30] discussed the relationships between metrics based on probability densities, namely, ϕ-divergences and integral probability metrics (IPM), and presented a number of novel properties to help increase their applicability.

A commonly used metric in SHM[16,17,31,32] is the Kullback-Leibler (KL)-divergence,

$$KL(\hat{p}_i, \hat{p}_j) = \int \hat{p}_i(\boldsymbol{x}) \log\left(\frac{\hat{p}_i(\boldsymbol{x})}{\hat{p}_j(\boldsymbol{x})}\right) dx \quad (1)$$

which is a measure of information divergence[33] and relative entropy; it is a metric that calculates the distance between distributions of random variables and indicates the amount of extra information needed to model $\hat{p}_i(\boldsymbol{x})$ using $\hat{p}_j(\boldsymbol{x})$. However, the KL-divergence is not symmetric, and therefore does not constitute a distance metric. Instead, the Jenson-Shannon metric is investigated in this paper as it is the symmetric equivalent of the KL-divergence.

Other measures based on density estimates that are explored here are the Hellinger distance, the total variational distance, the Kolmogorov distance and the Area metric. Table 1 provides further details about these metrics and provides their equations. The measures based on density estimates considered in this paper fall into two categories: ϕ-divergences and IPMs. To put it simply, ϕ-divergences are concerned with ratios between probabilities,

$$\text{Distance}_\phi\left(\hat{p}_i, \hat{p}_j\right) = \int_{\mathcal{M}} \phi\left(\frac{d\hat{p}_i(\boldsymbol{x})}{d\hat{p}_j(\boldsymbol{x})}\right) d\hat{p}_i(\boldsymbol{x}) \quad (2)$$

and IPMs with the difference between probabilities,

$$\text{Distance}_{\mathcal{F}}\left(\hat{P}_i, \hat{P}_j\right) = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{M}} f \, d\hat{P}_i(\boldsymbol{x}) - \int_{\mathcal{M}} f \, d\hat{P}_j(\boldsymbol{x}) \right| \quad (3)$$

where $\hat{p}_i(\boldsymbol{x})$ and $\hat{p}_j(\boldsymbol{x})$ are either probability density functions (PDFs) or cumulative density functions (CDFs) – denoted in upper case, that is, $\hat{P}_i(\boldsymbol{x})$ and $\hat{P}_j(\boldsymbol{x})$. $\mathcal{M}$ is a measurable space, ϕ is a convex function, $\mathcal{F}$ is a class of real-valued bounded measurable functions on $\mathcal{M}$, f is the subset of functions that define the metric, and sup is the supremum: the least upper bound of point-wise difference.[30,32]

Another (true) metric that is considered here is the *maximum mean discrepancy* (MMD), which circumvents the need for density estimation (this is considered beneficial as there is no universally accepted methodology for conducting density estimations, and they can introduce unwanted variability into the metric formulations). The MMD uses a kernel function to obtain the maximum distance between the mean embeddings of two features or vectors that have been mapped into a reproducing kernel Hilbert space (RKHS). For a more comprehensive understanding, the reader is referred to.[34,35] In the past, the MMD has been used in a variety of disciplines, because of its strength in assessing similarity of a range of data types; from time series,[36,37] graphical data,[38] images,[39] to attribute matching.[34,40] The MMD has been applied successfully in the field of verification and validation in SHM,[32] for domain adaptation,[20,41,42] in computer sciences to distinguish between malicious and honest users,[43] and to evaluate the effectiveness of

Table 1. The distance metrics based on density estimation that are explored in this paper.

Description	Equation
The *Jenson-Shannon* distance is the symmetric equivalent of the KL-divergence.	$JS\left(\hat{p}_i, \hat{p}_j\right) = \sqrt{\frac{1}{2}KL(\hat{p}_i(\boldsymbol{x}), \hat{p}_k(\boldsymbol{x})) + \frac{1}{2}KL\left(\hat{p}_j(\boldsymbol{x}), \hat{p}_k(\boldsymbol{x})\right)}$ (4) where $\hat{p}_k(\boldsymbol{x}) = \frac{1}{2}\left(\hat{p}_i(\boldsymbol{x}) + \hat{p}_j(\boldsymbol{x})\right)$. $\hat{p}_i(\boldsymbol{x})$ and $\hat{p}_j(\boldsymbol{x})$ are absolutely continuous w.r.t $\hat{p}_k(\boldsymbol{x})$.
The *Hellinger* distance is the L_2-norm between two probabilities.	$H\left(\hat{p}_i, \hat{p}_j\right) = \sqrt{\frac{1}{2}\int \left(\sqrt{\hat{p}_i(\boldsymbol{x})} - \sqrt{\hat{p}_j(\boldsymbol{x})}\right)^2 dx}$ (5)
The *total variational* distance employs the L_1-norm: The ϕ-divergence. The IPM form.	$TV\left(\hat{p}_i, \hat{p}_j\right) = \frac{1}{2}\int \left|\hat{p}_i(\boldsymbol{x}) - \hat{p}_j(\boldsymbol{x})\right| dx$ (6) $TV\left(\hat{p}_i, \hat{p}_j\right) = \sup_{\|f\|_\infty \leq 1} \left|\hat{p}_i(\boldsymbol{x}) - \hat{p}_j(\boldsymbol{x})\right|$ (7)
The *Kolmogorov* distance is the largest vertical distance (or the maximum L_1-norm) between two CDFs.	$KOL\left(\hat{P}_i, \hat{P}_j\right) = \sup_{x \in \mathbb{R}}\left|\hat{P}_i(\boldsymbol{x}) - \hat{P}_j(\boldsymbol{x})\right|$ (8)
The *Area metric* is simply the difference between the area under two CDFs.	$A\left(\hat{P}_i, \hat{P}_j\right) = \int \left|\hat{P}_i(\boldsymbol{x}) - \hat{P}_j(\boldsymbol{x})\right| dx$ (9)

CDF: cumulative density function; HL: Hellinger; JS: Jenson-Shannon; KL: Kullback-Leibler KOL: Kolomogorov; TV: total variational.

treatment and diagnosis in cancer research[44] to name a few. The biased estimate of the MMD is given by,

$$\text{MMD}_b^2[\mathcal{F}, X, Y] = \frac{1}{m^2} \sum_{i,j=1}^{m} k(x_i, x_j)$$
$$+ \frac{1}{n^2} \sum_{i,j=1}^{n} k(y_i, y_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) \quad (10)$$

where $X := \{x_1, ..., x_m\}$, $Y := \{y_1, ..., y_n\}$ and m and n are the number of data points in X and Y respectively. $k(x, y)$ is a kernel between x and y. There also exists an unbiased version of the MMD that uses the population expectation, however, it is not utilised in this paper.

The Gaussian kernel $k(\cdot, \cdot)$,

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2l_1 l_2}\right) \quad (11)$$

used here, is universal and continuous in the RKHS.[34] The widths of the kernels $k(x_i, x_j)$ and $k(y_i, y_j)$ are l_1 and l_2 respectively. There is no universal method of choosing the length scale for the Gaussian kernel. Gretton et al. uses the median heuristic, as explained in Gretton A, et al.,[34] and defined in Garreau et al.[45] The same method is used in this paper, although here, the length scales are calculated separately for each kernel $k(x_i, x_j)$ and $k(y_i, y_j)$, and the resulting length scales are multiplied together for $k(x_i, y_j)$.

Kernel density estimation

Unlike the MMD, the IPM and ϕ-divergence calculations are based on the differences between estimated density functions of the data. Consequently, one of the setbacks of using these distance metrics is determining the underlying probability distributions of the random variables in question. There does not exist a universal method of obtaining these; instead, a number of different methods are available, each with their own merits and drawbacks. One method that is widely used in literature is *kernel density estimation* (KDE).[46–49]

To put it simply, the KDE places 'atoms' of probability at each observation. The width of these atoms is specified by a window length (or smoothing/bandwidth) parameter and the shape is determined by the choice of kernel type. Usually, the kernels used are symmetric and non-negative. The final density estimate is the sum of these atoms[50,51] over a sample set.

The equation for the PDF with the Gaussian kernel used in this work is,

$$\hat{f}(x; h) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h\sqrt{2\pi}} \exp{-\left(\frac{x - X_i}{h\sqrt{2}}\right)^2}, \quad x \in \mathbb{R} \quad (12)$$

where N is the number of samples, X_i are the datapoints, the width or bandwidth of the kernel functions is defined by h and can be calculated using the rule of thumb,[50]

$$h = \sigma \left\{ \frac{4}{(d+2)n} \right\}^{\frac{1}{(d+4)}} \quad (13)$$

where n is the number of observations, σ is the standard deviation of X, and d is the number of multivariate dimensions. In order to find the CDF using KDE, Equation (12) is integrated to obtain,

$$\hat{F}(x; h) = \int_{-\infty}^{x} \hat{f}(v) dv = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left[1 + \text{erf}\left(\frac{x - X_i}{h\sqrt{2}}\right) \right] \quad (14)$$

where erf is the error function.

All methods of density estimation are affected by the chosen parameters, such as the kernel type or the bandwidth of the kernel. As a result, care must be taken when choosing these values, especially in a PBSHM context. For example, when evaluating a metric based on integration of density functions, the integral requires a mesh – or a method of discretising the function – that fixes the x points at which the density functions are evaluated. These points therefore must sufficiently capture the trends seen in the data across the population, in order for the metrics to capture relevant information.

The structure of this paper is as follows. In the next section, a number of typical behaviours observed in SHM data-sets are introduced using a simulated dataset to investigate the response of the aforementioned distance metrics, and assess their suitability in identifying similarity in the PBSHM context. Then, in Section 'Data normalisation and distance metrics', using the same simulated dataset, the effects of data standardisation on the behaviour of distance metrics are explored and mathematically formalised; standardisation is an important step in transfer learning and domain adaptation within PBSHM. Furthermore, in Section 'Case study: similarity of turbines in the Lillgrund wind farm' the applicability of distance metrics to assess similarity within a population of real structures is investigated via a case study containing data from an offshore wind farm. Finally, concluding remarks are presented and future work is discussed in Section 'Conclusion'.

Distance metrics response to commonly occurring behaviours in SHM datasets

In order to gain insights and identify behaviours of interest, similarity can be assessed across populations as a first step, using distance metrics. Determining the

similarity of data can be useful for, and inform any following steps in the analysis chain. For example, it may be used to extract suitable data for training and testing, for discovering features that are non-representative of the population, for applications such as domain adaptation,[20] to name a few.

In this section, the aim is to determine the suitability of distance metrics suggested in Table 1 to measure similarity of time-domain features in the context of PBHSM. Consequently, the sensitivity of the distance metrics to the transient and permanent changes typically observed in SHM features is studied. By doing so, it may be possible to identify specific metrics that are highly sensitive to a given change in feature behaviour. These differences can stem from EOVs, structural deterioration and even contamination as a result of impediments to the measurement chain (e.g. the sensor network). A simulated dataset that mimics these variations is used in this section. Particularly, the effect of these behaviours on the kernels and density functions that drive the distance metrics are mathematically derived, with a focus on the parameters that underpin these functions. Later, in the case study, examples of these variations acting on a real-world PBSHM dataset are presented.

Typical variations in SHM features

Structures in operation undergo a range of EOVs in their lifetime, for example, temperature changes, wind loading, fluctuating traffic loading (across a bridge, say), etc. Typically, the influence of EOVs on monitoring data (and on any feature therefrom derived), manifests in a change in statistical moments (mean, variance, skewness, etc) in either the time or frequency domain (or both). Commonly in SHM, changes in the mean of the resonance frequencies are tracked for their sensitivity to damage. Naturally, these features are also sensitive to changing temperature and fluctuate accordingly[52,53] – giving rise to daily and seasonal trends. Other commonly occurring signatures of EOVs are changing variance,[54] here again attributed to temperature, temporary mean shifts[8,55] caused by daily changes in wind direction or seasonal changes in temperature, and spikes which can be attributed to, for example, a quick transition to the extremes of the environmental or operational envelope.[56] The presence of structural damage or damage in the measurement chain also affects measurements and features in a similar manner (changing EOVs are often referred to as *confounding influences* in an SHM assessment for this reason). In this section of the paper, the effect of these commonly observed signatures on a series of candidate metrics is studied. The intention is to assess metric sensitivity and determine how best one might approach an initial analysis that is looking to identify similarities and differences in a population, whether they be from EOVs, damage or other.

To mimic these behaviours of change in frequency, variance and mean, and also to observe their affect on distance metrics within the time domain, a simulated dataset is used where an 'original' X – data sampled from a normal distribution with a mean of zero and a standard deviation of one, that is, $\mathcal{N}(0, 1)$ – is altered according to α. Here, α represents the magnitude of alterations (change in frequency, temporary and permanent alterations in mean and variance). In order to change the stochastic feature to include commonly occurring SHM behaviours, a window function $\omega(t)$ is used. The window $\omega(t) = 1$ if $t \in [t_1, t_2]$ and $\omega(t) = 0$ otherwise, where t is time. Therefore, the specific behaviour explored only affects the stochastic feature X within the window. Now, α can represent a constant or an incremental change which affects the entire feature X, where $[t_1, t_2]$ encompass the entire feature. Alternatively, when α represents a transient change $[t_1, t_2]$ covers a only section of the feature X.

The presence of typical SHM behaviours are likely to affect the distance metrics. As the metrics are driven by the density functions of the features, the way in which the characteristics of the density functions are affected is explored here. Note that the integrals in Table 1 may be approximated by numerical summations as long as the integration mesh is invariant across all features that are being compared.

Change in amplitude/variance

Changes in variance and amplitude are often observed in SHM datasets as a result of EOVs or damage.

Effect on density estimates. To parameterise such changes in amplitude/variance by α, the features of the density estimations Equations (12) and (14) that drive the metrics in Table 1 are pointwise multiplied by α,

$$X(t) \mapsto \alpha\omega(t)X(t). \tag{15}$$

The σ in the width of the kernel in Equation (13) alters as,

$$\sigma(X(t)) \mapsto \sigma(\alpha\omega(t)X(t)) \tag{16}$$

to account for the change in variance. As seen in Figures 1 and 2, changes in α lead to changes in variance and kurtosis of the density functional, although the mean and the skewness are unaffected. The width of the curve can increase as a result of increased variance, because the data now lie further away from the mean. Consequently, all metrics in Table 1 will be
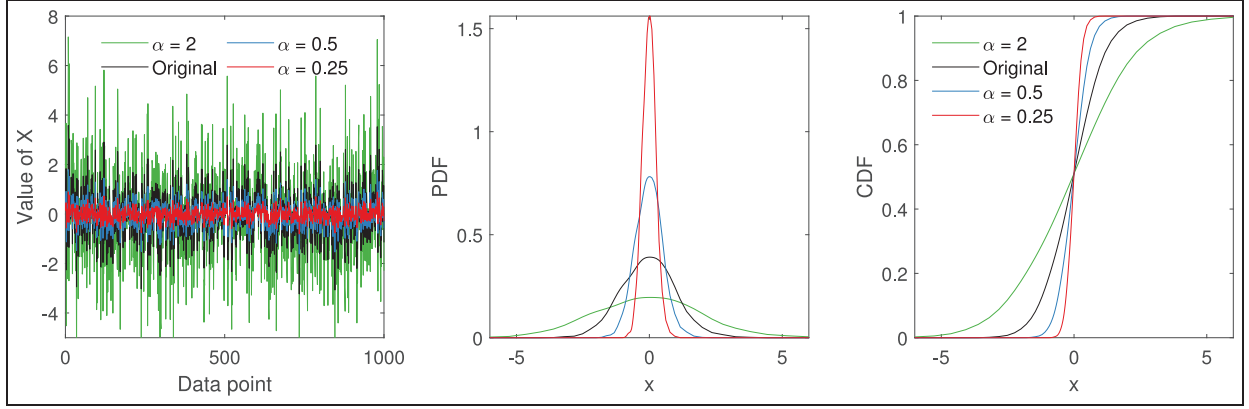
Figure 1. The effect of changing variance of a feature on probability density estimates. Here the amplitude of the original signal is changed to α.
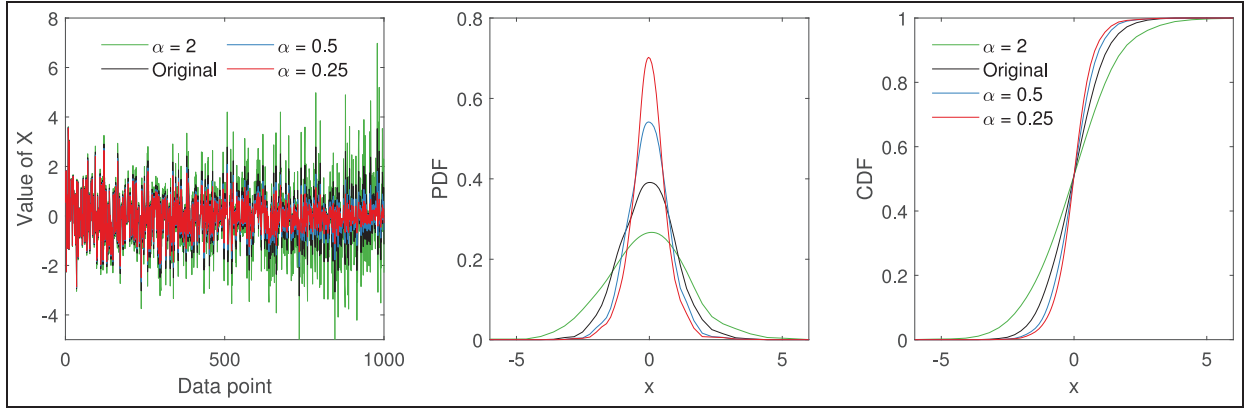
Figure 2. The effect of continuously changing amplitude of a feature on probability density estimates. Here the amplitude of the original signal is incrementally changed from 1 to α.

affected as they are driven by either the vertical distances or the area between functions.

If a temporary change in variance occurs (for a small time window where $[t_1, t_2]$ covers a section of the feature), as a result of EOVs, damage or an issue with the measurement chain, the resulting density functions take the form seen in Figure 3. As the effects are temporary, their influence on the density functions can be minimal; this influence is likely to increase with longer time windows or larger variances.

Effect on the Gaussian kernel used in the MMD. As the Gaussian kernel is employed differently in the MMD compared to KDE, it is useful to understand how the MMD formulation in Equation (10) is affected by a change in amplitude. When calculating the MMD, the Gaussian kernel is not evaluated over fixed intervals. Instead, the kernel gives a measure of similarity between two feature vectors, $X := \{x_1, ..., x_m\}$ and $Y := \{y_1, ..., y_n\}$. Here the kernel in Equation (11) is altered according to α. Only feature Y is changed by α

here. If the amplitude of signal Y changes by a constant value of α throughout, the kernel $k(y_i, y_j)$ is not affected,

$$
\begin{aligned}
k(\alpha y_i, \alpha y_j) &= \exp\left(-\frac{\|\alpha y_i - \alpha y_j\|^2}{\alpha^2 l_2 l_2}\right) \\
&= \exp\left(-\frac{\alpha^2 \|y_i - y_j\|^2}{\alpha^2 l_2 l_2}\right) = k(y_i, y_j)
\end{aligned}
\tag{17}
$$

because of the distributive property. It should be noted that the length scale should account for the new change in amplitude. When using the median heuristic to obtain the length scales from αy in this case, l_2 becomes αl_2.

The $k(x_i, y_j)$, however, affects the MMD as it changes to,

$$
k(x_i, \alpha y_j) = \exp\left(-\frac{\|x_i - \alpha y_j\|^2}{l_1 \alpha l_2}\right).
\tag{18}
$$
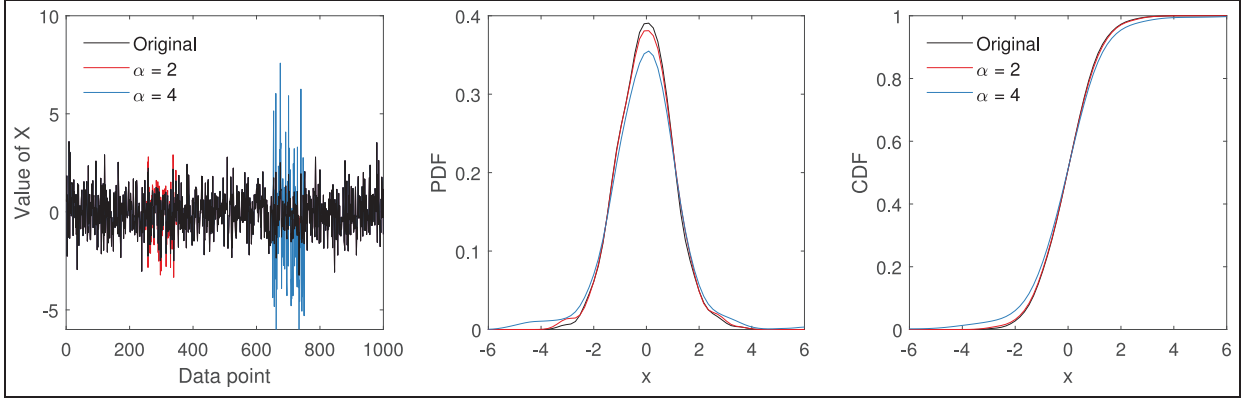
Figure 3. The effect of temporarily changing the amplitude of a feature on probability density estimates. Here the amplitude of the original signal is incrementally changed from 1 to α, within a small window.

If the change in amplitude varies across the feature, the distributive property no longer applies. If α is now a vector of the same length as Y, where each element changes the amplitude of each point in the feature, element-wise multiplication can be used to obtain αy_i and αy_j. The median heuristic used to calculate the length scales is also affected here, and results in,

$$k(\alpha y_i, \alpha y_j) = \exp\left(-\frac{\|\alpha y_i - \alpha y_j\|^2}{l_2^\alpha l_2^\alpha}\right) \neq k(y_i, y_j). \quad (19)$$

where l_2^α is the new length scale affected by α. Consequently, $k(x_i, y_j)$ becomes,

$$k(x_i, \alpha y_j) = \exp\left(-\frac{\| x_i - \alpha y_j\|^2}{l_1 l_2^\alpha}\right) \quad (20)$$

affecting the MMD.

Change in mean

The mean of features can alter as a result of EOVs and damage; this is an example of signal shift.

Effect on density estimates. The density functions (in Equations (12) and (14)) translate in mean as,

$$X(t) \mapsto \alpha\omega(t) + X(t) \quad (21)$$

Here, α can be used to shift the entire feature by a single value (Figure 4), incrementally (Figure 5), or shift parts of the feature temporarily (Figure 6). If the entire feature is shifted by a single value of α, then the width of the kernel h is unaffected because the variance does not change. However, if an incremental or temporary shift in the data is present, the change in variance will affect h as,

$$\sigma(X(t)) \mapsto \sigma(\alpha\omega(t) + X(t)) \quad (22)$$

When the entire feature is shifted in mean by a single value of α, that is, $[t_1, t_2]$ covers the entire feature, the mean of the density functions also shifts, whilst the variance and higher-order moments remain unaffected (Figure 4). Although in translation the area under the density functions will not change, the resulting distance metrics will be affected as the same integration matrix/discretisation of features is used for the original and translated functions.

Incremental changes in the mean shift the density functions and changes their shape, as the variance is also affected (Figure 5). The change in shape resembles an increase in variance, where peak heights are reduced and the width of the peaks are increased. The combination of change in variance and mean affects the skewness of the density functions.

When a temporary shift in mean is present, that is, $[t_1, t_2]$ covers a section of the feature, the resulting density functions resemble a mixture of Gaussians (that do not have the same mean) as seen in Figure 6, where multiple modes are present. As the skewness and kurtosis are affected by the variance of the data, higher moments of the functions appear during a temporary mean shift in data.

Effect on the Gaussian kernel used in the MMD. For the Gaussian kernel in the MMD formulation in Equation (11), if feature X is unaffected and the entire feature Y is translated by a single α, the length scale is not affected, as the distribution of Y does not change. The kernel $k(y_i, y_j)$ also doesn't change when altered by α,

$$\begin{aligned} k(y_i + \alpha, y_j + \alpha) &= \exp\left(-\frac{\| (y_i + \alpha) - (y_j + \alpha)\|^2}{l_2 l_2}\right) \\ &= \exp\left(-\frac{\| y_i - y_j\|^2}{l_2 l_2}\right) = k(y_i, y_j). \end{aligned} \quad (23)$$
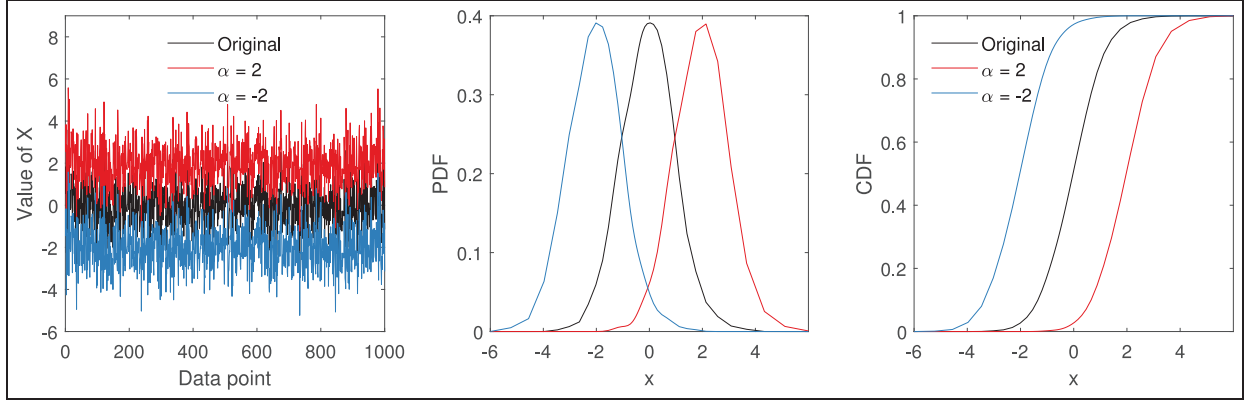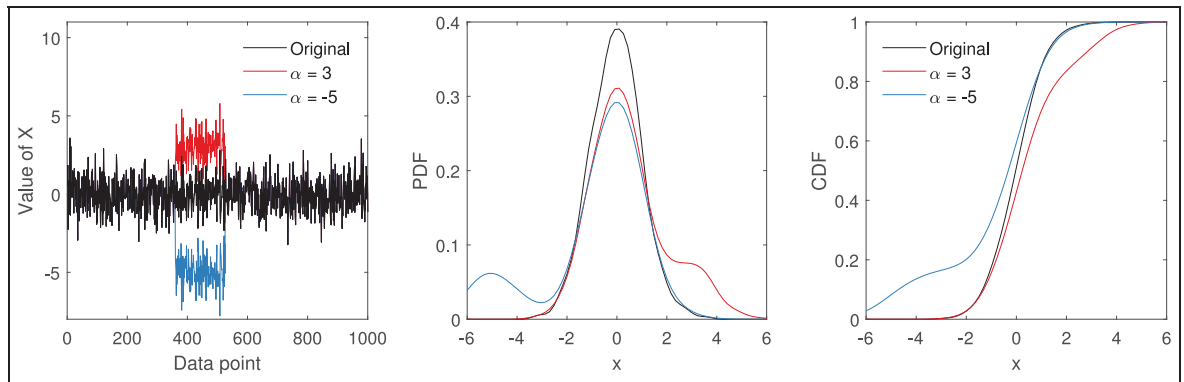
Figure 4. The effect of global mean shift on a feature by a factor α, and the corresponding density estimates.
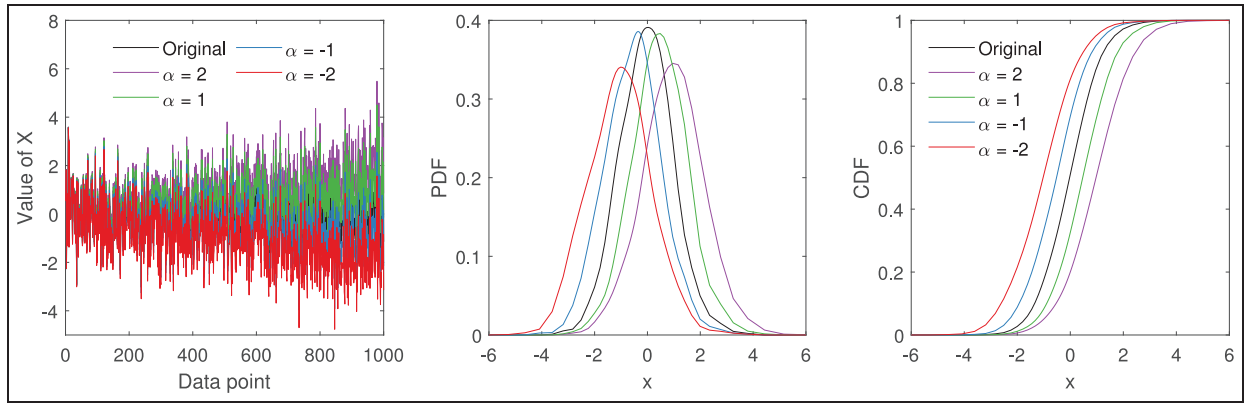
Figure 5. The effect of incremental mean shift on a feature from 0 to α, and the corresponding density estimates.

Figure 6. Temporary changes in mean to a feature by $\alpha\omega(t)$ and the resulting density functions.

However, $k\left(x_i, y_j\right)$ modifies to,

$$k\left(x_i, y_j + \alpha\right) = \exp\left(-\frac{\|x_i - (y_j + \alpha)\|^2}{l_1 l_2}\right) \qquad (24)$$

causing the MMD to change accordingly.

If, however, the mean is altered incrementally or if only a portion of the feature's mean is altered, the variance will be affected and therefore the length scales will also change. Consequently, $k\left(x_i, y_j\right)$ becomes,

$$k\left(x_i, y_j + \alpha\right) = \exp\left(-\frac{\|x_i - (y_j + \alpha)\|^2}{l_1 l_2^\alpha}\right) \qquad (25)$$

showing that changes in mean of one distribution will be detected by the MMD.
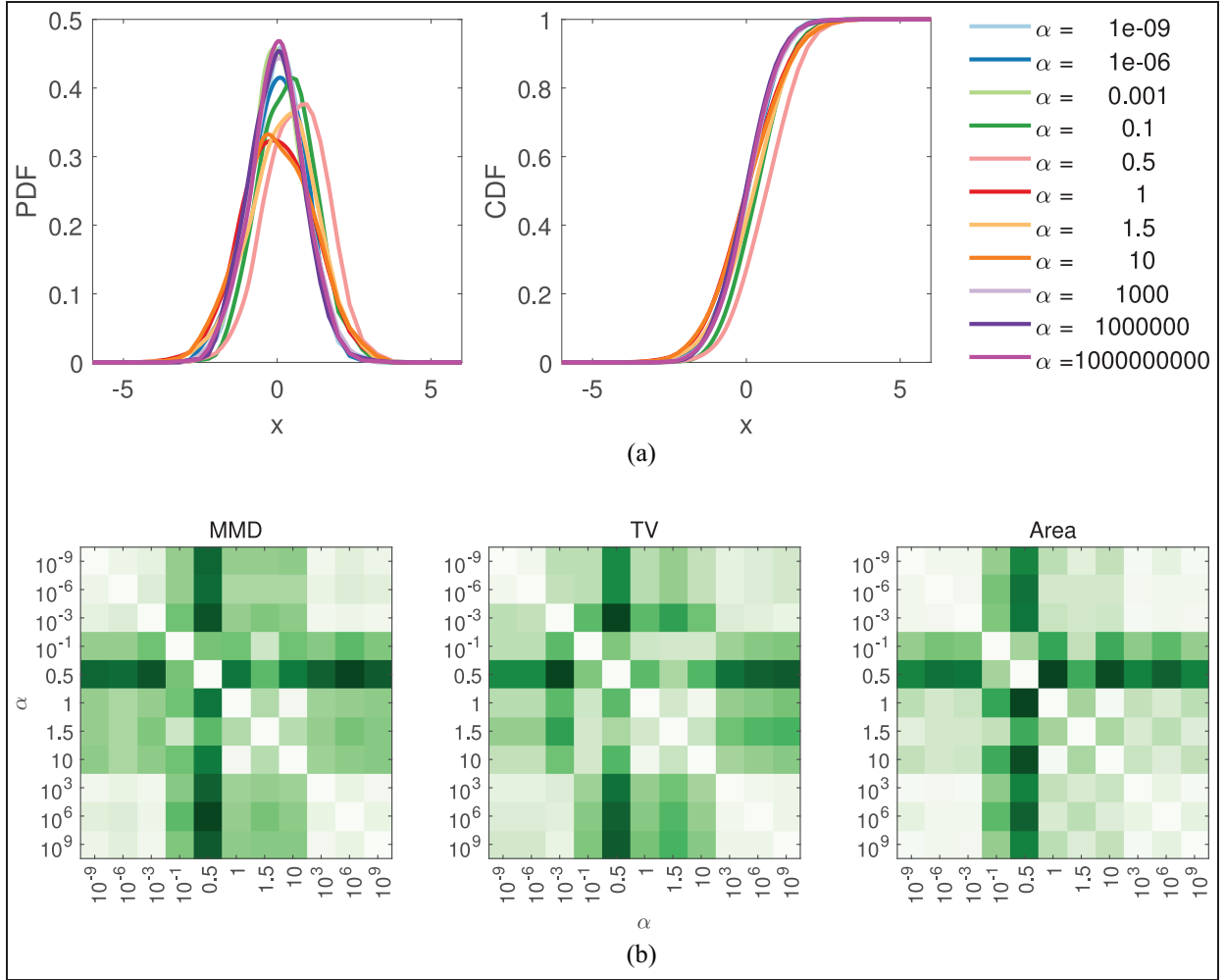
Figure 7. Change in frequency across a wide range and the resulting effect on density functions and distance metrics. One metric based on the PDF, CDF and the MMD are presented here. In (b) the lighter shades represent high similarity and the darker shades represent high dissimilarity. (a) The effect on density functions as the frequency of X is altered by α and (b) The effect on distance metrics as the frequency of X is altered by α.
CDF: cumulative density function; MMD: maximum mean discrepancy; PDF: probability density function.
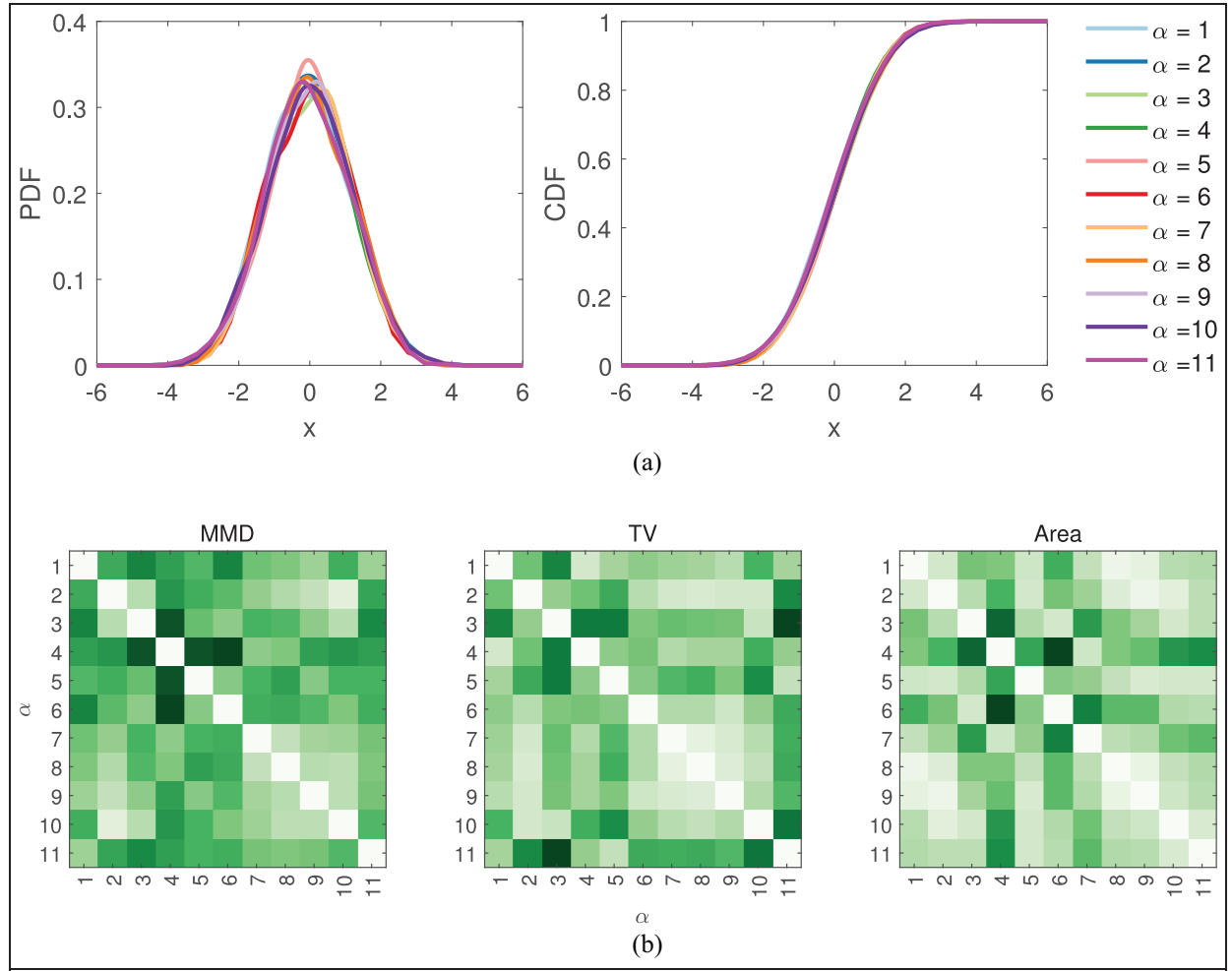
Change in frequency

The frequency content of sensor signals can vary during normal operating conditions as a result of EOVs and damage. However, the change in frequency of a feature,

$$X(t) = A \sin(2\pi\alpha t) + z(t) \qquad (26)$$

(where α is the frequency, A is the amplitude of the signal, and z is the added white Gaussian noise) is challenging to capture using the distance metrics suggested here. Of course, this is not to say that the frequency has no affect on the density functions – in fact, changes in frequency (across the entirety of X shown here for simplicity), can have a significant impact on the density functions. Take Figure 7(a), for example. In this figure,

α has been altered significantly to show the effect on density functions and the resulting distance metrics (Figure 7(b)). It is unlikely that the frequency of a given SHM feature would undergo a change in frequency by an order of 10^3, so this example is purely for demonstration purposes. The challenge here is that the alterations to the density functions are not necessarily easy to track; there does not seem to be a clear pattern or correlation between frequency changes and the resulting effect on the density functions or distance metrics.

When the change in α is much smaller, however, the effect on the density functions is harder to distinguish, as seen in Figure 8(a). Here, α has been altered between 1 and 11, a much more plausible change to be expected in features from operational civil structures. Again, the distance metrics are unable to capture that difference

Figure 8. Change in frequency across a relatively small range and the resulting effect on density functions and distance metrics. One metric based on the PDF, CDF and the MMD are presented here. In (b) the lighter shades represent high similarity and the darker shades represent high dissimilarity. (a) The effect on density functions as the frequency of X is altered by α and (b) The effect on distance metrics as the frequency of X is altered by α.
CDF: cumulative density function; MMD: maximum mean discrepancy; PDF: probability density function.

effectively, as the change to the density functions are seemingly random, because the pairwise distances between X and X_α (the altered feature) are very small.

It is clear that the method of employing distance metrics to identify feature changes at the top of the analysis chain suggested here, is unable to effectively capture changes in frequency in time-domain features. Consequently, change in frequency will not be explored further in this paper. However, SHM methods often track frequencies (such as natural frequencies of structures[56]), across time, where changes in frequencies can present themselves as changes in mean and/or variance. An example of such behaviour is presented in Figure 9, where freezing temperatures alter the natural frequencies of the Z24 bridge[56]– a widely explored case study in SHM. The distance metrics method suggested in this paper will be useful at identifying these changes in

frequencies, as they will alter the density functions as seen in Figure 6.

The response of the metrics to changes in amplitude and mean

For automated similarity/dissimilarity assessment within PBSHM, the method put forward in this paper is to use distance metrics to measure the difference between features collected from datasets. It is, therefore, important that the distance metrics suggested here are reponsive to typical behaviours observed in SHM features. The hope here is that the candidate metrics are able to detect changes in mean and variance of the features and provide insights into the underlying behaviour of the data at the very top of the analysis chain.
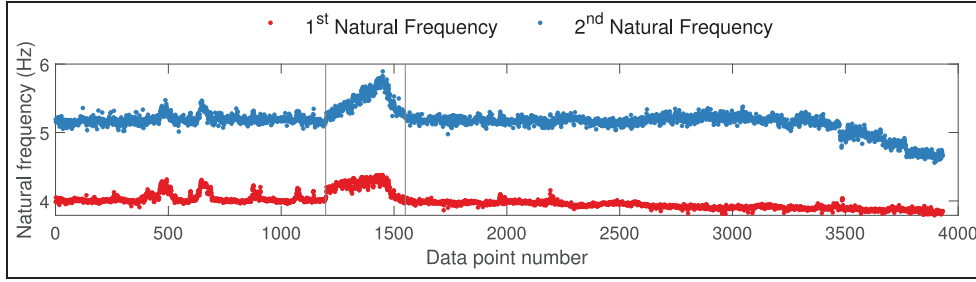
Figure 9. The natural frequencies of the Z24 bridge where changes in frequencies resulting from freezing conditions are highlighted between the two vertical black lines.[56]
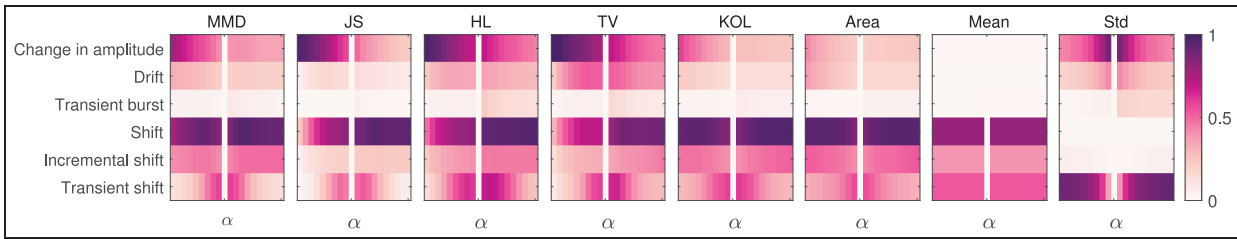
Figure 10. The sensitivity/response of each metric to varying α (the change in features as a result of commonly occurring SHM behaviours). All metric values are normalised between [0,1] to identify the most sensitive metric. The response of the first two statistical moments are also presented here for comparison purposes.
SHM: structural health monitoring.

Table 2. The α values used for altering each feature to produce Figure 10.

Change in amplitude/Drift	Transient burst	Shift/Incremental shift	Transient shift
0.1–2 in increments of 0.1	0.1–0.9 in increments of 0.1 then 1–4 in increments 0.3	−1.8 to 2 in increments of 0.2	−7.2 to 8 in increments of 0.8

Figure 10 presents how the distance metrics respond when the amplitude/variance of features are altered by α. Table 2 presents the α values used for altering each feature. Here, the discretisation mesh is set between -6 and 6 for all features based on density estimates. To demonstrate the effectiveness of distance metrics compared to simply evaluating summary statistics, the responses of the first and second statistical moments (referred to as 'Mean' and 'Std', respectively) to the features are also presented here. The data in this figure are normalised between [0,1] for comparison. Values closer to 1 in darker shades suggest a large response, whereas, values closer to 0 in lighter shades indicate little to no response. Additionally, Figure 11 provides the maximum metric value for each feature across all α values.

Key insights from Figures 10 and 11 are:

- When considering all feature changes ('Change in amplitude', 'Drift', 'Transient burst', 'Shift',

'Incremental shift', 'Transient shift') in Figure 10, the distance metrics show a higher sensitivity overall than simply using 'Mean' and 'Std'; for some features, the maximum values of 'Mean' and 'Std' are very small/negligible in Figure 11.

- The first statistical moment 'Mean' is unsuitable for detecting features based on an amplitude/variance change as those features do not drive a change in mean. Conversely, the 'Std' has a limited or null response when identifying global and incremental shifts in mean, as the change in standard deviations of the features are negligible. It is, however, able to detect transient shifts in mean as the standard deviation increases as a result of the multiple modes (as seen in Figure 6).

- Across the range of α values, all distance metrics and 'Mean' are most sensitive to 'Shift'/global mean shift. As the mean of the features is translated according to α, the distance between the original and altered density functions increases, the overlap
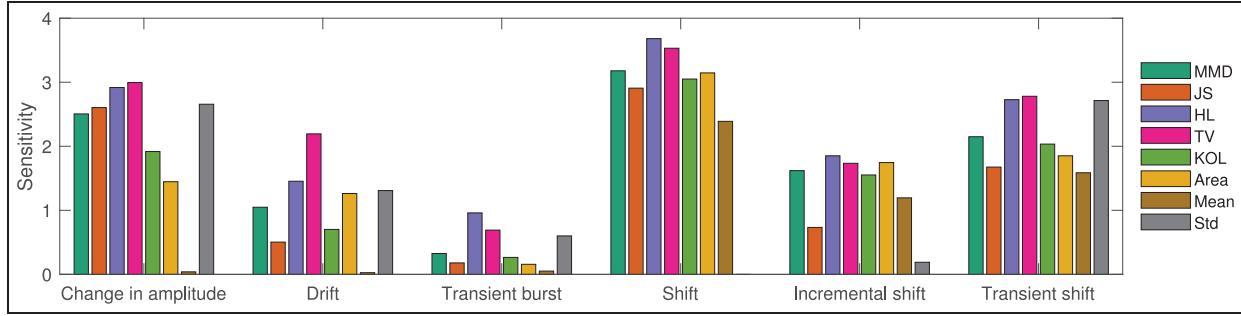
Figure 11. The maximum metric values from each SHM feature presented here to provide an overview of the metric response in Figure 10. The response of the first two statistical moments are also presented here for comparison purposes.
SHM: structural health monitoring.

Table 3. The sensitivity of each metric to each feature ordered from the most sensitive to least sensitive. This table gives an indication of which type of SHM feature each metric prioritises during automated similarity assessment.

Order of priority	MMD/JS/HL	TV	KOL	Area
1	Shift	Shift	Shift	Shift
2	Change in amplitude	Change in amplitude	Transient shift	Transient shift
3	Transient shift	Transient shift	Change in amplitude	Incremental shift
4	Incremental shift	Drift	Incremental shift	Change in amplitude
5	Drift	Incremental shift	Drift	Drift
6	Transient burst	Transient burst	Transient burst	Transient burst

MMD: maximum mean discrepancy; SHM: structural health monitoring.

decreases, leading to increases in distance-metric values. The 'Std' is unable to identify this significant change in features, as the standard deviation of the features do not vary here. As mean shifts are common in SHM data, the use of 'Std' in this scenario could be detrimental for anomaly detection.

- Variance changes for a short window ('Transient bursts') evoke the smallest response across all distance metrics. HL and TV metrics show the highest sensitivity where they are less sensitive to small variations compared to large variations. As short 'bursts' of variance increase can indicate the presence of damage or a change in conditions, the HL and TV metrics are preferred here.

- When considering a global change in amplitude, all distance metrics based on the PDF, the MMD and 'Std' present a large response. The peak height and area of the PDFs change readily with varying amplitude, as seen in Figure 1. The distance metrics are more sensitive to reduction in variance (or when $\alpha < 1$), that is, when the amplitude is decaying compared to growth, both of which may affect sensor signals as a result of damage or instability.[57,58] The 'Std' can be useful in detecting this feature where it is most sensitive to small changes in amplitude.

- The metrics are more sensitive to small transient shifts compared to large ones. The statistical

moments are also relatively sensitive to transient shifts.

The aforementioned changes in variance and mean clearly affect the density functions and their statistical moments, and in turn drives the changes in distance metrics. HL and TV metrics present the highest overall response across all features and are well suited for assessing similarity within SHM datasets. It should be noted that the MMD performs similarly to HL and TV (often presenting the third highest sensitivity within the distance metrics), and should be given serious consideration here as the most suitable candidate, because it negates the need for density estimation. This is a promising result, as it is preferable to avoid density estimation where possible.

Table 3 presents the order in which each metric prioritises a feature according to sensitivity. The MMD, JS and HL metrics show similar priorities here, whereas the TV, KOL and Area metrics have different preferences. The Area metric is unique in that it is more sensitive to changes in mean than variance. Although the difference between the sensitivity responses are relatively small across all metrics – with no obvious front-runners for any given feature (Figure 11) – the priorities of the metrics suggest that some are more sensitive to specific features over others. Therefore, the MMD, JS, and HL

metrics are expected to behave similarly when identifying features, whereas, TV, KOL, and Area metrics are expected to highlight different types of features. On this account, commonly occurring behaviours within new datasets may be identified by using a combination of distance metrics, allowing for fast and automatic information retrieval and similarity assessment.

By using a simulated dataset, in this section, the responses of candidate distance metrics to commonly occurring SHM features are explored. It is clear that the metrics are well suited for identifying features (especially over simple analysis using the first two statistical moments), where some metrics show a preference for identifying certain features over others. As a result, a given metric can give an indication of the type of features present in a signal without manual inspection. The results presented here using the simulated dataset are satisfactory pointers towards generalisation, because in a population-based setting – particularly within a homogeneous population where the behaviour of the data within a given window is assumed to be the same/very similar (Even in heterogenous population examples, feature similarity is important to avoid negative transfer[23,26]) – the metrics will not identify differences between general trends as $D(x, y) = 0$ when $x = y$. As a result, the metrics are expected to identify behaviours/features that are anomalous to the general population trends. Consequently, the underlying general behaviour of the simulated dataset (a Gaussian white noise signal), is less important in a population-based setting than the commonly occurring SHM features. Given that the features are based on typical behaviours observed in the SHM literature, this dataset is simple yet satisfactory for the purpose of metric analysis.

A note on real-time analysis of population data using distance metrics. Although not implemented in this paper, it is noted here that for PBSHM, it is possible to use these metrics to track the behaviour of structures over time, using a moving window. Care must be taken when discretising the density functions in this case, because an inadequate range could diminish the effect of unexpected SHM behaviours or highlight normal behaviours as abnormal. As the MMD formulation does not require the calculation of density functions, it is perhaps better suited for real-time analysis. In this case, a suitable moving window length can be chosen *a priori* that captures the general fleet-wide trends in historical data, offline. The data that are used to set the window length should be collected during normal operating condition in order to ensure that abnormalities (deviations from the normal condition) can be detected during real-time implementation.

The computational complexity of calculating the MMD is $\mathcal{O}(m + n)^2$ time, where m and n are the points sampled from distributions estimates $\hat{p}_i(\boldsymbol{x})$ and $\hat{p}_j(\boldsymbol{x})$ respectively. Whereas, the computational complexity of distance metrics based on KDE is $\mathcal{O}(n)$. Consequently, assessing the similarity of data in a PBSHM setting using distance metrics is relatively computationally inexpensive, especially when the dataset sizes are small. In real-world implementations, suitably small window lengths can ensure small dataset sizes. As a result, this framework can be compliant with edge processing for its implementation.

In the next section, the effect of normalisation of the data on the metrics is discussed; normalising or standardising the data to have zero mean and unit standard deviation is a common pre-processing step in SHM analyses. Normalising removes the effects of mean and variance on the density estimations. As a result, it is possible to determine if a given metric is sensitive to higher-order moments that may be present as a result of variations within the data. In PBSHM applications that utilise transfer learning, normalisation can be vital for aligning domains for domain adaptation.[22] Techniques such as normal condition alignment can help move data and change the size of the space they occupy, in order to facilitate better decision boundaries that separate the data well, and allow easy comparison between domains. Foregoing alignment can sometimes even lead to negative transfer. The next section of this paper explores how normalisation can affect data and therefore the ability to find similarities using distance metrics, if the calculation of the distance metrics should be altered to aid normalisation, and how to analyse the results from normalised data.

Data normalisation and distance metrics

In conventional SHM, distance metrics can be used to identify whether structures are operating under normal (undamaged) conditions, or if damage has occurred, where a large distance between the normal and damage conditions is preferred. In PBSHM, it is desirable to reduce the distance between the data from normal conditions of different structures in a population to aid positive transfer in instances such as when attempting domain adaptation.[20,22] Here, distance metrics are used to identify the target label-set that is the closest match to the source domain. Normalising the source and target domains by their respective normal operating condition helps align the data by the normal condition. This form of normalisation is beneficial to transfer leaning, as learners are notoriously ineffective when translation has occurred between source and target data.[59,60] Normalisation can, however, mask the effects of temporal and spatial EOVs; the dynamic response of the structures can be altered by wind, thermal and induced vibrations, change of material properties and boundary conditions.[8,61]

In this paper, distance metrics are used to find similarity and information within a population when there is no knowledge of the underlying data structure. The assumption is that little to no information is available on the condition of the structure (from which the data were collected), or the data itself. In the previous section, distance metrics were tested against a number of typical behaviours in SHM features in order to determine their suitability at detecting changes and assessing similarity. These behaviours usually have an affect on the characteristics of feature distributions. In this section, normalisation is employed to eliminate the effects of lower-order moments from the data and highlight features containing higher-order moments such as skewness and kurtosis. Given that higher-order moments can indicate the presence of transient shifts/drifts in the data, normalising in this context may be helpful in identifying features that are affected by unexpected behaviours. Consequently, a single distance metric value used on absolute *and* normalised data may indicate useful information about the underlying structure of the data.

Normalisation in the form of standardisation is an alignment step that can be applied to data.[62–64] Standardisation translates and scales the data collected from structures by subtracting the mean and dividing by the standard deviation. When applied on a structure-by-structure basis – also referred to as self-standardisation – the resulting data from each structure has a mean of zero and a unit variance. Often it is used to move data to a similar space where comparisons between the data are meaningful. Standardisation can also reduce the condition numbers of matrices, allowing for numerical accuracy when computing their inverse. It is important to note that metrics cannot be used on self-standardised data to find differences between datasets because global transformations have not been conducted on the data; the mean and variance of each dataset are likely to be slightly different and, therefore, each dataset will be altered differently during self-standardisation. In this section, self-standardising is used to investigate the origin of the differences seen in PBSHM datasets.

Figure 12 visually presents the effect of self-standardisation on the density functions that drive the distance metrics. Here, the commonly occurring features in Figures 1 to 6 are used for demonstration. α values of 0.5, 2 and 5 are used to represent change in amplitude/variance, drift (time-varying amplitude) and transient bursts, respectively. An α value of 2 is used to shift the feature, and a value of 5 is used for the transient shift. After self-standardisation, the scaling and shifting effects from changing amplitude/variance and translation/shift are reduced. Density functions that contain higher-order moments are highlighted, as the

integration mesh essentially squeezes the functions into the same space. The next section explores how these transformations manipulate the density functions that drive the metrics, with a focus on the parameters used to obtain density estimates.

Data standardisation and the IPM and ϕ-divergences

The IPM and ϕ-divergences in Table 1 are driven by the probability and CDFs. These density functions (Equations (12) and (14)), are dependant on two variables: the width h and the points x. To track the effect of standardising on the distance metrics, the process in which these variables are altered is studied.

When the data are self-standardised, the bandwidth parameter, h (in Equation (13)) becomes,

$$h_N = 1 \left\{ \frac{4}{(d+2)n} \right\}^{\frac{1}{(d+4)}}$$
$$h = \sigma_X \times h_N \tag{27}$$

because the standard deviation of the self-standardised data (σ_{X_N}) is 1. Also, the values for x should now encompass the data from all structures. In the population-based setting, Equation (12) becomes,

$$\hat{f}(c, X_{Ni}; h_N) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h_N \sqrt{2\pi}} \exp - \left(\frac{c - X_{Ni}}{h_N \sqrt{2}} \right)^2$$
$$= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\frac{h}{\sigma_X} \sqrt{2\pi}} \exp - \left(\frac{c - \frac{X_i - \bar{X}}{\sigma_X}}{\frac{h}{\sigma_X} \sqrt{2}} \right)^2 \tag{28}$$

where c are the new points at which the density functions of the self-standardised data are evaluated. Similarly the CDF becomes,

$$\hat{F}(c, X_{Ni}; h_N) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left[1 + \mathrm{erf} \left(\frac{c - X_{Ni}}{h_N \sqrt{2}} \right) \right]$$
$$= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left[1 + \mathrm{erf} \left(\frac{c - \frac{X_i - \bar{X}}{\sigma_X}}{\frac{h}{\sigma_X} \sqrt{2}} \right) \right]. \tag{29}$$

Therefore, in a PBSHM setting, the discretisation and the smoothness of the density functions are affected when the data are self-standardised. As a result, these values should be re-evaluated when obtaining density functions for self-standardised data. Equations (28) and (29) can be fed back into the metric formulations where the integrals can be treated as sums, because the discretisations c are the same for all functions. Next, the way in which the MMD is affected during standardisation is explored.
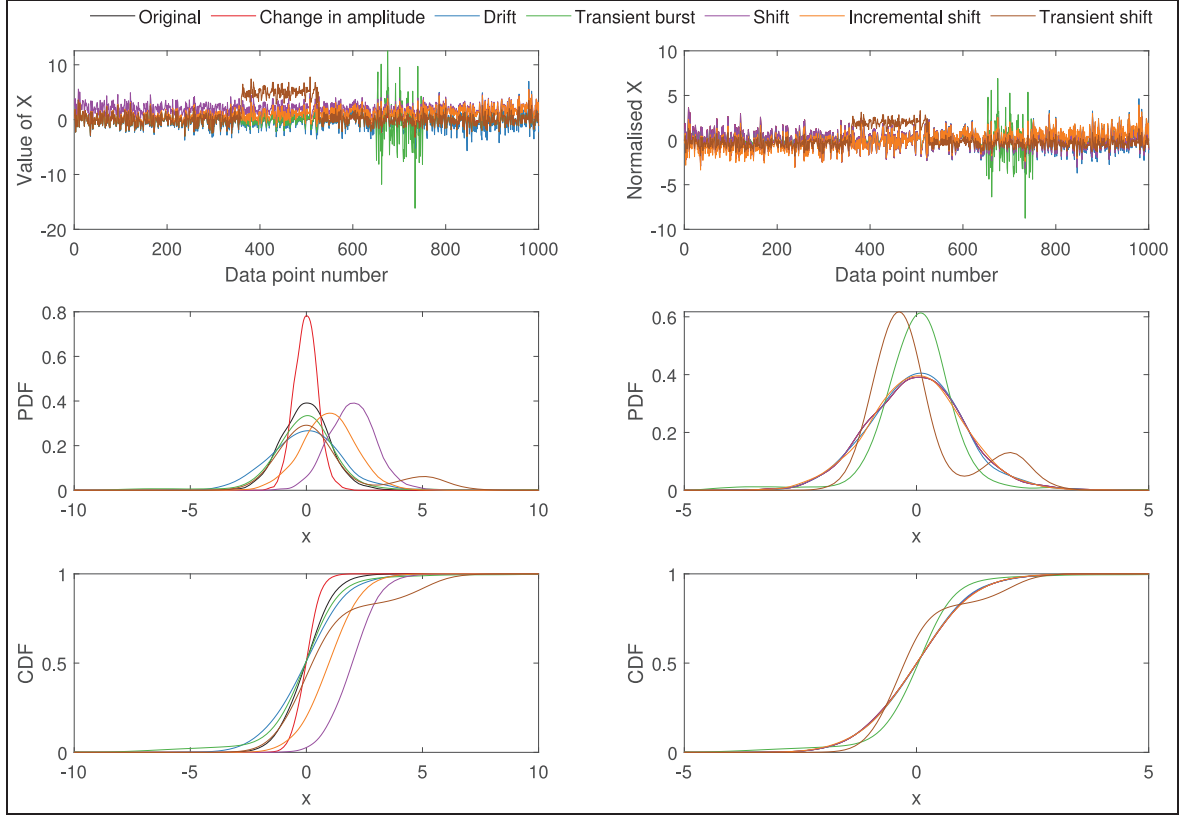
Figure 12. The effect of standardising on density functions is presented using the commonly occurring SHM behaviours. Standardising (figures on the left), leads to effects from changes in amplitude/variance and translation to reduce, and highlights transient changes.

SHM: structural health monitoring.

Data standardisation and the MMD

As the random variables x and y from Equation (10) are moved to a similar space by standardisation, the value of the MMD is expected to change because it is a measure of difference between the mean embedding of two distributions. As the Gaussian kernel for the MMD is calculated on the data from two distributions (Equation (11)), the length scales can be separated.

When both variables are from the same distribution, the normalised kernel is equal to the non-normalised kernel, that is,

$$k(x_N, x'_N) = \exp\left(-\frac{\left\|\left(\frac{x-\bar{x}}{\sigma_x}\right) - \left(\frac{x'-\bar{x}}{\sigma_x}\right)\right\|^2}{2\times\left(\frac{l_1}{\sigma_x}\right)^2}\right)$$

$$= \exp\left(-\frac{\|x - x'\|^2}{2\times(l_1)^2}\right) \quad (30)$$

where σ_x is the standard deviation of x. Similarly, $k(y, y') = k(y_N, y'_N)$. The same cannot be said for

$k(x_N, y_N)$ because the length scales are divided by the standard deviation of the corresponding variable σ_x and σ_y (the standard deviation of x and y respectively), that is,

$$k(x_N, y_N) = \exp\left(-\frac{\left\|\frac{x-\bar{x}}{\sigma_x} - \frac{y-\bar{y}}{\sigma_y}\right\|^2}{\frac{2\times l_1 l_2}{\sigma_x \sigma_y}}\right)$$

$$= \exp\left(-\frac{\sigma_x \sigma_y \left\|\frac{x-\bar{x}}{\sigma_x} - \frac{y-\bar{y}}{\sigma_y}\right\|^2}{2\times l_1 l_2}\right) \quad (31)$$

It is also interesting to note the effect of altering the length scales of the kernel when standardising, because obtaining the optimum choice of kernel size is an current research topic, and therefore the length scales are subject to change.[34] For example, if the relationship between $k(x, y)$ and $k(x_N, y_N)$ is $k(x, y) = \alpha \times k(x_N, y_N)$ and if the length scales are multiplied by a constant c, where $c \in \mathbb{R}$, the relationship

$$k(x_N, y_N) = \frac{k(x,y)}{\alpha} = \frac{\exp\left(-\frac{\|x-y\|^2}{2l_1 l_2}\right)}{\alpha} \quad (32)$$

becomes,

$$k_{N_c} = \frac{\exp\left(-\frac{\|x-y\|^2}{2l_1 l_2 \times c}\right)}{\alpha} = \frac{\exp\left(-\frac{\|x-y\|^2}{2l_1 l_2}\right)^{\frac{1}{c}}}{\alpha} = \frac{k(x,y)^{\frac{1}{c}}}{\alpha}. \quad (33)$$

Again, as the MMD is a true metric that circumvents the need for density estimation, it may be preferred over metrics based on density estimates when used for preliminary analysis; this is because, the discretisation step required for density estimates (on both absolute and self-standardised data), can be avoided when using the MMD.

The response of distance metrics to self-standardised data

In this section, the focus is on exploring the effect of self-standardised data on the response of distance metrics. This is an important consideration for PBSHM because, standardisation is a key alignment technique used to bring domains into a similar space for transfer learning. Furthermore, standardisation can be used to reduce the effects of the first two statistical moments on the data, thereby increasing the metric's sensitivity to higher-order moments. It is hoped that standardisation will help to identify interesting features in the data that were previously masked by the mean and variance. Earlier, in Figure 12 it was found that transient behaviours are highlighted after standardisation as they tend to contain higher-order moments. The aim here is to find metrics that are sensitive to those characteristics. Again, the commonly occurring behaviours introduced in Section 'Distance metrics response to commonly occurring behaviours in SHM datasets' are used here for demonstration.

When considering the commonly occurring SHM behaviours, Figure 13 show how each distance metric responds to self-standardised data. As with Figure 10, the response of 'Mean' and 'Std' are presented here for comparison between the distance metrics and the first two statistical moments. The results in 13 are combined with the absolute case in Figure 10 and normalised between [0,1] for ease of comparison. The corresponding self-standardised metric values for constant changes in amplitude, and constant mean shifts are not presented here because self-standardising these altered features leads to metric values of zero. In these figures, α can represent the increase/decrease in variance/amplitude, or a positive/negative shift in mean/translation. The ranges of α used in Figure 13 can be found in Table 2. Figure 14 summarises Figure 13 by providing the maximum response of each metric to each standardised feature.

The key observations from Figures 13 and 14 and comparisons between the absolute case in Figures 10 and 11 are:

- Unsurprisingly, the metric responses are markedly reduced when assessing self-standardised data because the effects of mean and standard deviation are removed.
- The first two statistical moments 'Mean' and 'Std' are unable to detect any changes in the features within self-standardised data, making them unsuitable for similarity assessment in PBSHM where the data are standardised.
- Unlike the first two statistical moments, the distance metrics show some response to the feature changes, with a relative high response towards transient shifts.
- The distance metrics based on the PDF and the MMD present a higher sensitivity to transient bursts in the self-standardised case in Figure 14 compared to the absolute case in Figure 11 because, the higher-order moments such as skewness and
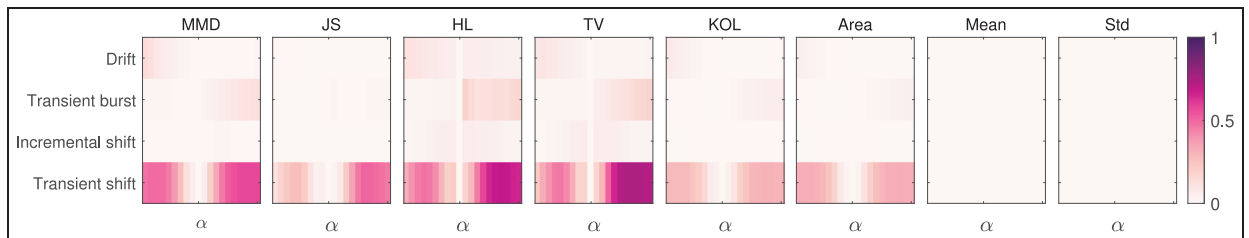

Figure 13. The sensitivity of each metric to varying α when the data containing the commonly occurring SHM features are self-standardised. The metric values for change in amplitude and feature shift are not presented as standardisation leads to metric values of zero. The response of the 'Mean' and 'Std' are also presented to represent the first two statistical moments. All metric values are normalised between [0,1] for comparison.
SHM: structural health monitoring.
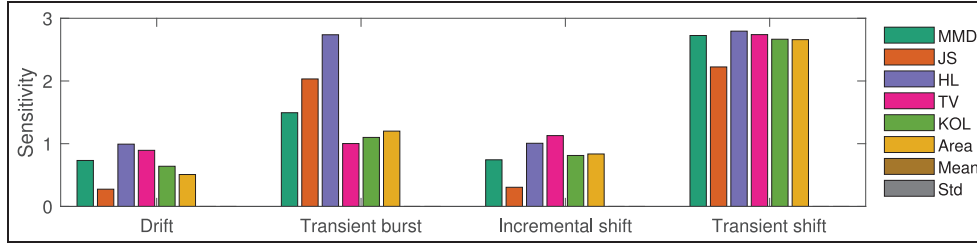
Figure 14. The maximum metric values from each SHM feature presented here to provide an overview of the metric response in Figure 13. The response of the first two statistical moments are also presented here for comparison purposes. SHM: structural health monitoring.
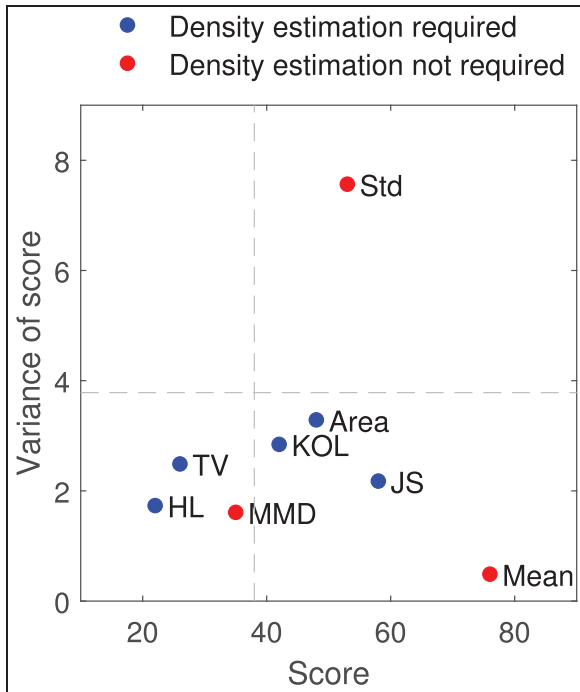
Figure 15. The overall performance of the metrics across all absolute and standardised features. The metric with the largest response has the lowest score. The metric with the lowest variance performs similarly across all features. The best overall metric should present a low variance and a low score. The dotted dashed lines indicate 50% of the score and variance.

kurtosis are highlighted as the data are standardised (as observed in Figure 12).

• As data are standardised, metrics are more responsive to transient behaviours over incremental behaviours.

These results show that there are advantages to undertaking data standardisation in the context of identifying commonly occurring SHM features using distance metrics. Transient effects are especially highlighted in the standardised case, when the effects of general trends (overall mean and variance changes) are diminished. As transient features can be indicative of damage and EOVs, it is useful to undertake data standardisation to identify them.

Table 4 presents the order of priority for each metric as the data are standardised, that is, the standardised features that each metric is sensitive to, ordered from most to least. The results here show that there is a high level of uniformity across the candidate metrics, except for the TV metric. The TV metric is more sensitive to incremental shifts over transient bursts; perhaps because incremental shifts not only shift the data but also reduces the area under the PDF, the TV metric (in the ϕ-divergent case), is more sensitive here as it responds to the largest and the smallest areas under the PDF. This difference in response by the TV metric is helpful, as using a combination of the TV and other metrics will be useful for capturing a wider variety of features from standardised data.

The recommendation here is therefore to assume that the metrics based on the PDF are likely to identify overall trends when using absolute data, and highlight transient behaviours when the data are standardised. When considering metrics based on the CDF, transient shifts can be identified from absolute data. As a result, when using absolute data, a combination of HL/TV and KOL/Area should detect a variety of SHM features from the dataset automatically. In the standardised case, the TV metric can be used alongside HL to identify a range of transient and incremental features.

Overall response of the distance metrics

Figure 15 presents the overall performance of the metrics when considering the SHM features in both absolute and standardised cases (six absolute and four standardised features). The x-axis of the figure represents the 'score' of each metric; the score is calculated by identifying the maximum response of each metric to each feature and giving a value of one for the most sensitive metric and a value of eight for the least sensitive

Table 4. The sensitivity of each metric to each standardised feature ordered from the most sensitive to least sensitive. This table gives an indication of which type of SHM feature each metric prioritises during automated similarity assessment of standardised data.

Order of priority	MMD/JS/HL/KOL/Area	TV
1	Transient shift	Transient shift
2	Transient burst	Incremental shift
3	Incremental shift	Transient burst
4	Drift	Drift

MMD: maximum mean discrepancy; SHM: structural health monitoring.

metric (as there are eight metrics assessed in total: six distance metrics, two statistical moments). A metric can have a total score between values of 10 and 80 where a value closer to 10 suggests that the metric has a high sensitivity to all SHM feature and a score of 80 suggests that the metric is insensitive to all SHM features. The scores are plotted against the variance of the score in the y-axis. A small variance suggests that the response of the metric was similar across all features, whereas a high variance of score suggests the response of the metric varied significantly from feature to feature. The best overall metric should, therefore, have a low score and a low variance. In Figure 15 two dashed lines separate the metrics at 50% of the score (vertical) and 50% of the variance (horizontal). Those in the bottom left corner are the best performing metrics from the chosen candidates. Within this quarter, the HL has the lowest score and the MMD has the lowest variance. Again, as the metrics based on density estimation require a discretisation mesh to be selected, the similar performance of the MMD is advantageous for feature assessment at the start of the analysis chain.

The results from the simulated dataset suggests that the distance metrics chosen in this paper are suitable candidates for detecting typical SHM behaviours. It is clear that using a combination of metrics across absolute and standardised data can give insights into dissimilarity across populations as well as provide an indication of what type of features are present within the data automatically. These findings are incredibly helpful for PBSHM, where manually analysing data to find suitable data for training and testing or transfer learning is an impossibility. In order to validate these findings and to understand whether these metrics are flexible in identifying SHM features from measured data from operational structures, a case study is presented next. The case study focuses on identifying similarity/dissimilarity of features collected during a year-long monitoring campaign of a wind turbine fleet.

Case study: similarity of turbines in the Lillgrund wind farm

This case study investigates the use of distance metrics to automatically identify similar/dissimilar behaviours within SHM features collected from a population of wind turbines. This is a homogeneous population, which is a sensible starting point for population-wide assessment. Modern wind turbines continuously collect data pertaining to environmental and operational conditions. Each turbine collects data from multiple channels, usually at ten-minute intervals. The available data for analysis, therefore, grows with population size. On that account, there is a clear need for fast and automated analysis for PBSHM.

The aim here is to gain an understanding of the effectiveness of the investigated distance metrics, at identifying interesting features from real, operational structures, in order to evaluate whether they can be used for automated data assessment in PBSHM. The metrics will be tested on absolute and standardised data as before, to determine whether normalisation is helpful and to discover transient behaviours. If successful, the hope is that these metrics will be used to group together structures with similar behaviours, and reduce the burden on transfer learners to achieve positive transfer, by establishing that the underlying distributions of the domains are similar. Furthermore, the assumption is that identifying dissimilarities within the turbines should indicate unexpected EOVs or damage.

To reiterate, the assumption in this paper, and indeed in this case study, is that the similarity assessment is undertaken at the very start of the analysis chain where little to no information about the data is known. The aim here is to automatically gain knowledge of the data collected from an entire population, reducing – or even alleviating – the need for time-consuming manual analysis.

The Lillgrund wind farm and the SCADA dataset

Located between the Swedish and Danish borders, the Lillgrund wind farm has 48, 2.3 MW wind turbines.[65] This is a *homogeneous* population, as all the structures are nominally identical, but with varying boundary conditions.[1,3,4,20] Each turbine is fitted with a Supervisor Control and Data Acquisition (SCADA) system that collects and saves data from a range of sensors at ten-minute intervals. Henceforth, the 48 turbines in the Lillgrund wind farm will be denoted T1 to T48.

In this work, the brake-gear temperature data collected during the year 2009 are studied (Figure 16).
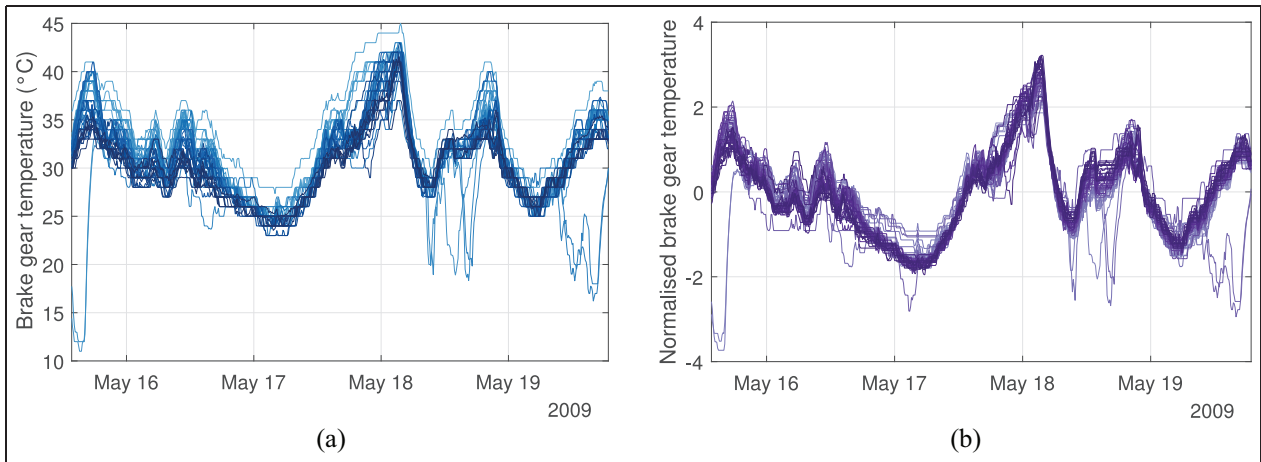
Figure 16. Brake-gear temperature of the 48 turbines in the chosen time window. Some anomalous behaviours can be seen here, where data appear outside of the general trend. Distance metrics may help in identifying these anomalous behaviours. (a) Absolute values and (b) normalised values.

This particular feature is chosen because of its sensitivity to daily temperature variation, and also because it includes behaviours such as shifts in mean, change in variance and transient mean shifts (such as those explored in Sections 'Distance metrics response to commonly occurring behaviours in SHM datasets' and 'Data normalisation and distance metrics'), that may be indicative of operational conditions or damage. To ease the computational burden, a window of data from the middle of May is chosen for analysis. The first step in the analysis process is to identify the most similar and dissimilar turbines within the fleet.

Results and discussions

The results calculated by the distance metrics are presented in Figure 17 in the form of confusion matrices where one-to-one comparisons are made across the population. The darker shades in these figures signify dissimilarity whilst the lighter shades indicate similarity. There are clear differences between the absolute and standardised cases here, suggesting that standardisation has helped identify different turbines to the absolute case, most likely as a result of transient behaviours. The metrics show that a significant number of turbines have high similarities across the population, suggesting that they can be used for transfer learning. This result is a strong indication that the underlying distribution of the general population is similar, which is evident in Figure 16.

One of the most interesting observations from the confusion matrices is the interaction between two of the most dissimilarly behaving turbines (T10 and T12), in the self-standardised case. It is clear that, even though the distances between data from T10 and T12 are consistently large compared to the rest of the population, the distances between each other are relatively small according to most metrics. This finding suggests that the anomalous data from these particular turbines actually behave similarly to each other and could be an indication that the underlying mechanisms or forces driving these behaviours are also the same. This result illustrates the purpose of PBSHM in a number of ways. For example, if presented with a dataset that contains healthy and damage-state data, it may be possible to identify other datasets that are undergoing the same mechanisms, and detect damage prior to failure using this method for guidance. If the dissimilar behaviour is a result of EOVs, it may be possible to group the relevant structures together and label them appropriately. On the other hand, T10 deviates from T12 around data points 400–450 (highlighted in Figure 18(c)), but that does not reduce the similarity between them, in comparison to the data from the rest of the population, across the window. This result shows that a temporary variation is not as significant for the metrics when the overall trend is very similar, and also demonstrates the importance of the window in which the metrics are calculated; temporary variations may be masked by more persistent trends, resulting in misleading conclusions.

Identifying common SHM features. The two most dissimilar turbines from the confusion matrices when using absolute and normalised values are identified for each metric in Figure 19 by summing the difference values across all comparisons between turbines. Given the findings from earlier in this paper, these results should
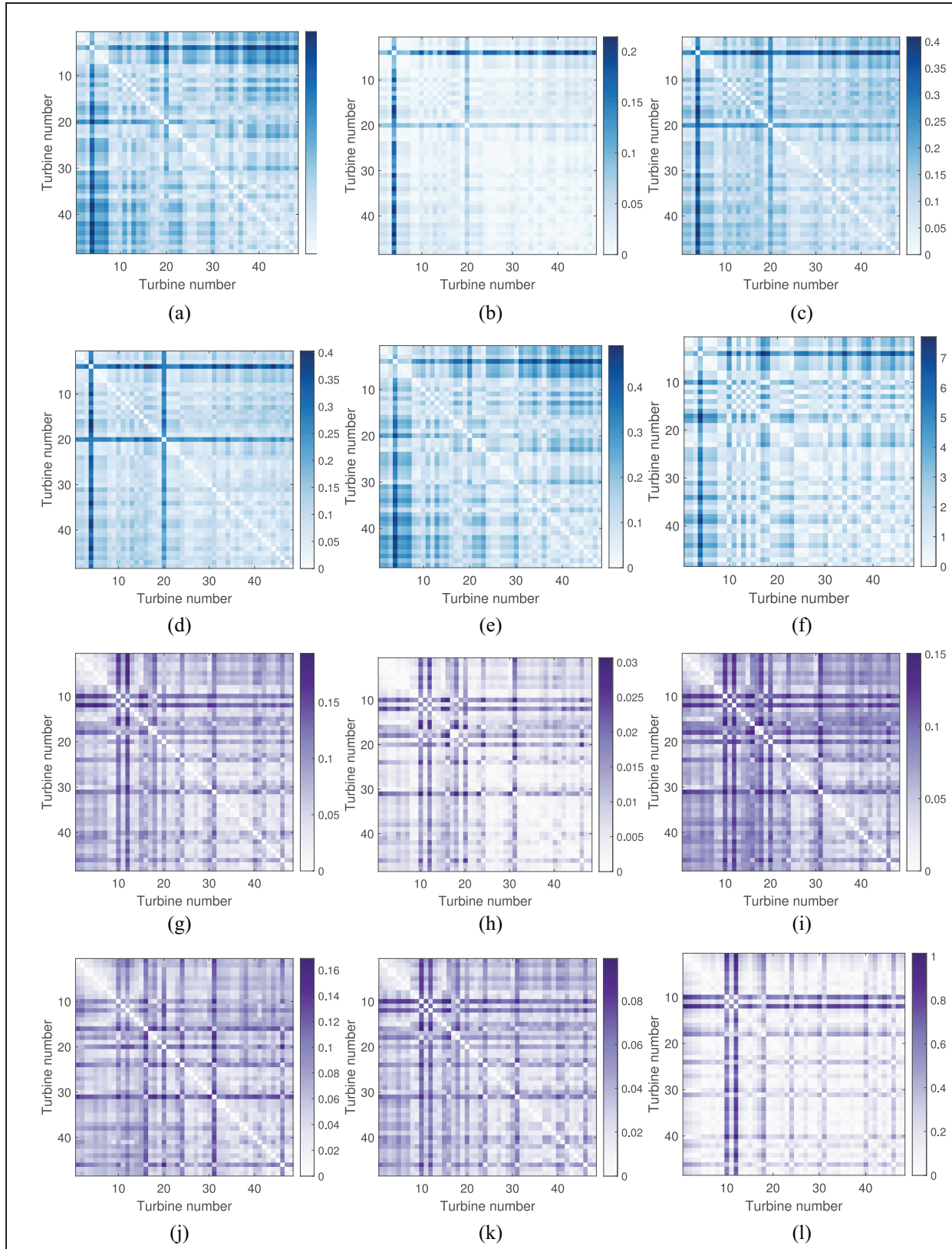
Figure 17. The confusion matrices showing distances calculated by metrics using absolute and standardised data. (a) MMD absolute, (b) JS absolute, (c) HL absolute, (d) TV absolute, (e) KOL absolute, (f) Area absolute, (g) MMD normalised, (h) JS normalised, (i) HL normalised, (j) TV normalised, (k) KOL normalised, (l) Area normalised.
MMD: maximum mean discrepancy.
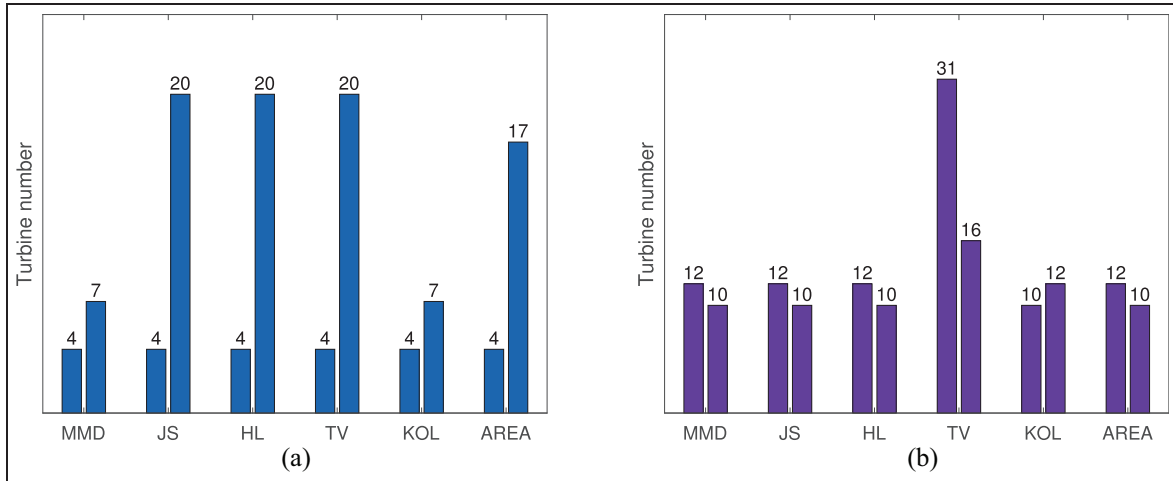
Figure 18. The absolute and normalised brake-gear temperatures. Highlighted here are the most dissimilar turbines identified by the distance metrics. (a) MMD and KOL absolute, (b) JS, HL, TV and Area absolute, (c) MMD, JS, HL and Area normalised, and (d) TV normalised.
MMD: maximum mean discrepancy.
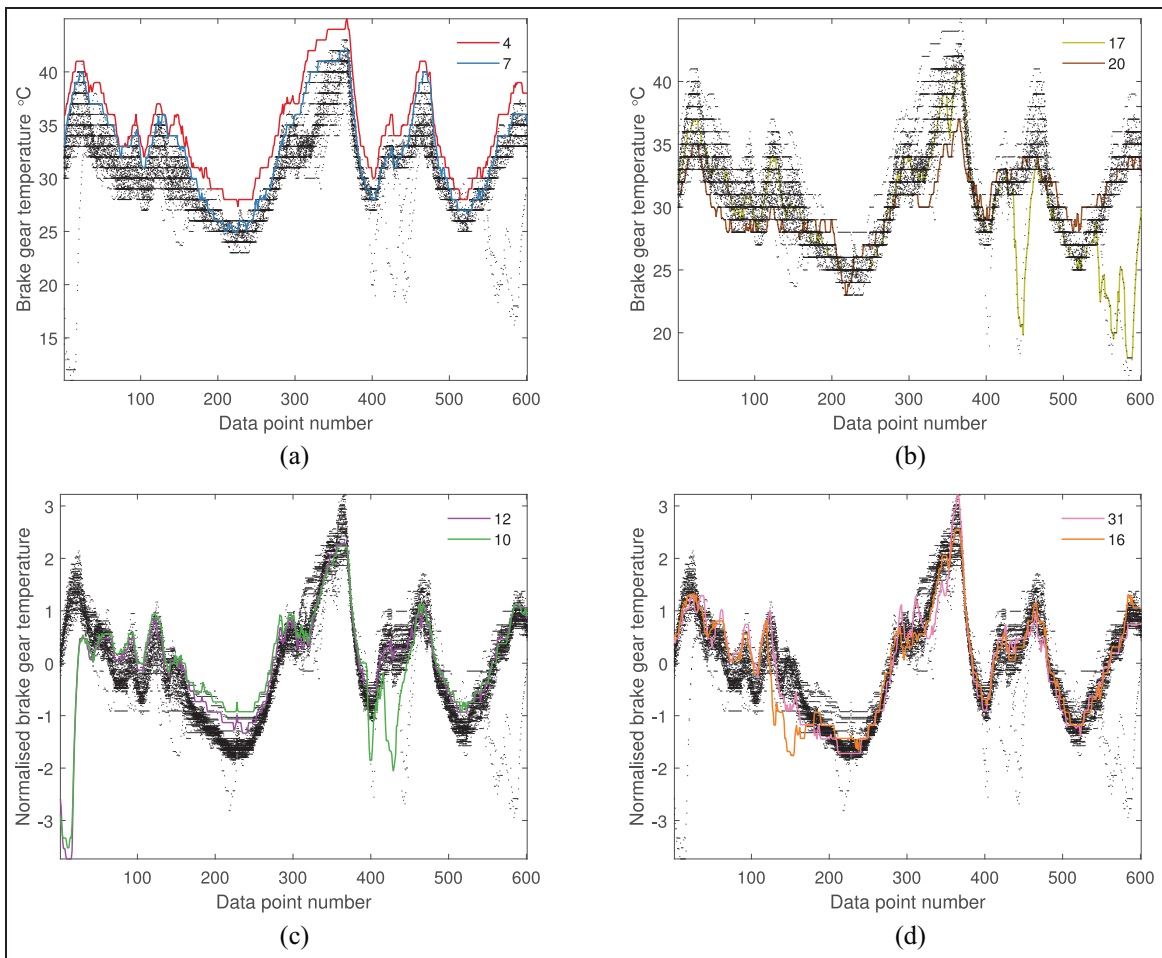
Figure 19. Most dissimilar turbines in the population when considering the brake-gear temperatures. (a) Absolute values and (b) self-standardised values.
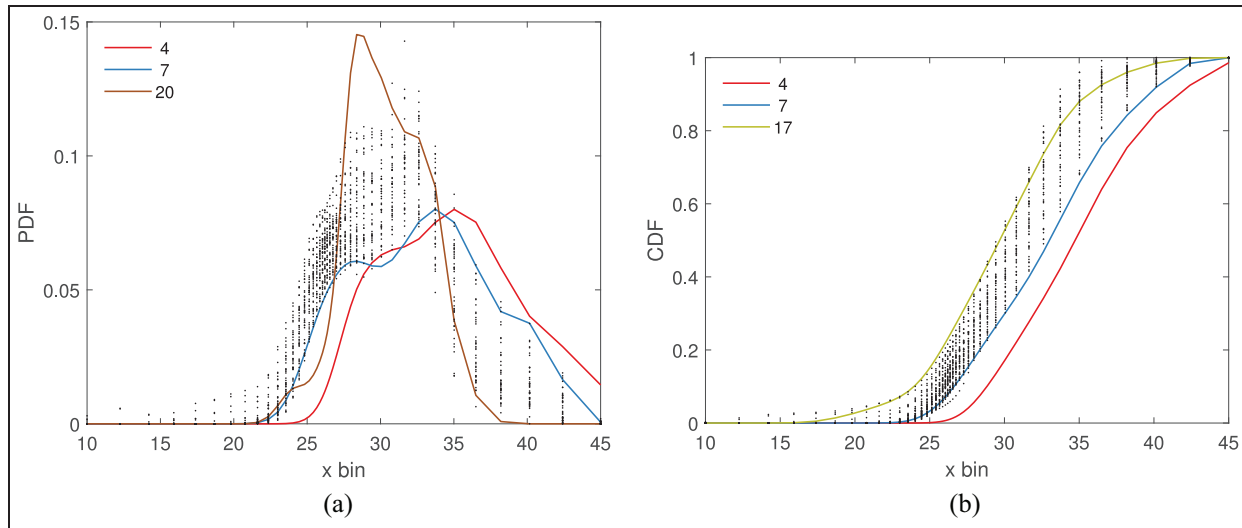
Figure 20. The density functions of the absolute data with the most dissimilar turbines highlighted. (a) Based on the PDF and (b) based on the CDF.
CDF: cumulative density function; PDF: probability density function

give an indication of the type of SHM features identified by the metric.

Findings from the absolute case are:

- According to Table 3, it is probable that T4, T7 and T20 stem from a global shift in mean, or a change in variance because the MMD, JS, HL and TV metrics prioritise those features over others. Indeed, Figures 18(a), 18(b) and 20(a) show that T4 and T7 are shifted in mean and T20 has a small variance.

    - These shifts in mean of T4 and T7 are a result of the turbine's position within the farm (T4 and T7 is located along one of the edges of the farm), and the speed and direction of the wind during this time period, suggesting that identifying signal shift via distance metrics may be useful for environmental mapping.
    - The density function of T20 (Figure 20(a)), that drives these metrics, resembles the effects of decrease in amplitude from Figure 1. The temperature variation of T20 across the time period is smaller than others in the population (Figure 18(b)), possibly because of an operational issue; for example, the sensor may be damaged, consequently reducing its sensitivity to temperature changes. Clearly, automatically identifying this decrease in variance within the population would not be a simple task for a human. Given that three out of six metrics have detected this feature, this finding further supports the use of distance metrics for automated anomaly

detection and suggests that they may be useful in detecting faults in the instrumentation.

- The Area metric identifies T17 as dissimilar, which could mean a shift in mean or a transient shift according to Table 3. Again, Figure 18(b) confirms that a transient shift has taken place in T17.

As expected from previous sections, when the data are standardised, T4 and T7 and T20 are no longer significantly dissimilar to the others, because the effects of translation and variance are diminished.

Insights from the standardised case are:

- All metrics except TV identify T10 and T12 as the most dissimilar, suggesting from Table 4 that incremental shifts may be the cause. Figure 18(c) confirms that this is the case.
- Surprisingly, the transient shifts of T17 are not highlighted by the metrics after self-standardisation, as expected from the results in Table 4. This result is because the transient shifts of other turbines (Figure 21(a)), are now more pronounced than that of T17.
- Interestingly, the TV metric identifies different turbines (T16 and T31), to the other metrics as suggested by the findings from Table 4, possibly owing to small incremental shifts observed in Figure 18(d).

The results presented in this section closely follow the findings from Sections 'Distance metrics response to commonly occurring behaviours in SHM datasets'
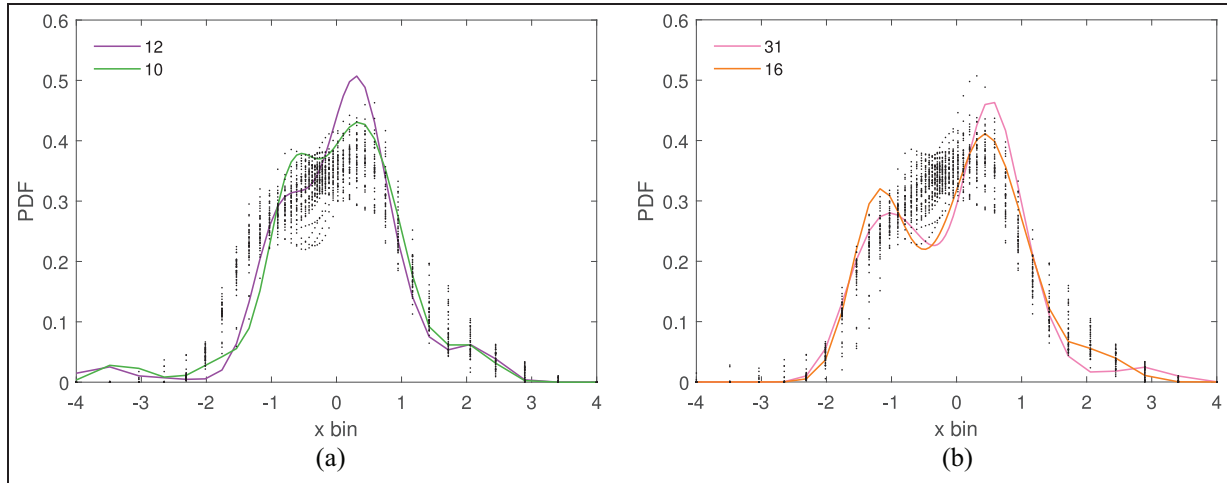
Figure 21. The density functions of the self-standardised data with the two most dissimilar turbines highlighted. (a) JS and HL normalised and (b) TV normalised.
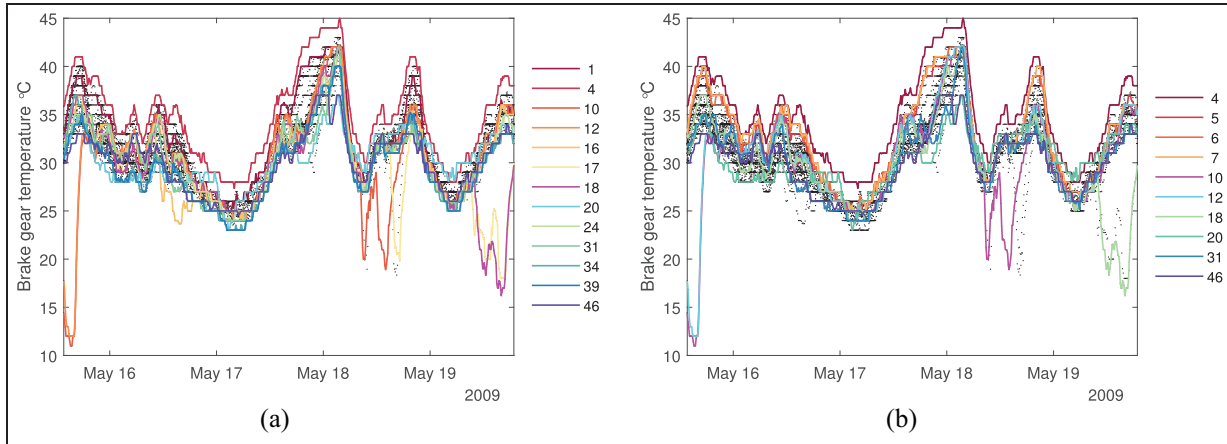
Figure 22. (a) The most dissimilar turbines identified by a combination of HL, TV, and Area giving a broad range of dissimilarities. When considering the the 10 most dissimilar turbines from the absolute and standardised cases, 13 turbines are identified by the three metrics and (b) the 10 most dissimilar turbines (from both absolute and standardised cases) identified by the MMD, a metric which has one of the best overall performances, without the need for density estimation. (a) TV, HL, Area and (b) MMD. MMD: maximum mean discrepancy.

and 'Data normalisation and distance metrics', validating the earlier findings and suggesting that the simulated dataset is suitable for evaluating the candidate distance metrics. As the general trend is similar across the population, only anomalous behaviours are identified easily and automatically.

The results from the case study indicate that a combination of HL, TV and Area metrics should help broaden the number of dissimilar turbines that can be identified from the dataset. Tables 3 and 4 also agree that these metrics prioritise a wide range of features. The most dissimilar turbines from each of these metrics across absolute and standardised data are highlighted in Figure 22(a). As the MMD negates the need for density estimation, and as it has a low overall score and

variance in Figure 15, the most dissimilar turbines identified by the MMD are also presented in Figure 22(b). The combination of HL, TV and Area metrics find a large number of dissimilar turbines spanning many different features. The MMD performs relatively well, where it captures many of the same turbines. The recommendation is, therefore, to use a combination of HL, TV and Area metrics for a comprehensive dissimilarity identification and use the MMD to avoid the drawbacks of density estimation.

Conclusion

This is the first exploration of automated similarity assessment of feature data in population-based SHM,

a new research field that aims to transfer knowledge across structures to address the shortcomings of SHM. PBSHM considers data from multiple structures and sensor networks, and therefore, new technologies are required to analyse vast amounts of data quickly and automatically, to aid positive transfer learning, and help identify suitable training and testing data. In this paper, distance metrics – that are not all widely used in SHM currently – were introduced in the context of PBSHM, and analysed for their ability to highlight similarity between members of the population at the top of the analysis chain, to aid the decision-making process.

Six distance metrics were tested on a simulated dataset, and on feature data collected from an operational wind farm, to evaluate their effectiveness at identifying commonly occurring features in PBSHM datasets. In both cases, the effect of data standardisation was also studied as it is a vital alignment step in transfer learning. The findings between the simulated and real datasets agreed and suggested that distance metrics are well suited for similarity assessment in PBSHM.

The results showed that most metrics were responsive to shifts in mean and variance when assessing absolute data, and transient effects when assessing standardised data. By assessing the response of the metrics to SHM features, it was found that a combination of metrics is required to identify a comprehensive variety of dissimilar behaviours within a population. The recommendations from the study are to use a combination of Hellinger, Total Variational (IPM form) and Area metrics to identify a broad range of dissimilarities in a population. If avoiding the drawbacks of density estimation is important, then the MMD is a useful metric, as it has a good response to all features evaluated.

These metrics are only suitable for use in large populations because the similarities (and, likewise, dissimilarities), are identified by comparison across the population. It is not possible to highlight dissimilarly behaving data from a structure using the distance metric value between itself and another structure. For these metrics to be used effectively, the population size should be large enough that it is possible to recognise the difference between data from a group of similarly behaving structures against a dissimilar structure.

In this paper, a homogeneous population of structures is studied, which reduces the complexity of assessing the similarity across the fleet, because the sensor network is also assumed to be similar across the population. However, it is entirely possible that small dissimilarities in the data will arise if some aspects of the sensor networks (sensor positions, types, manufacturers, models) are different across the population. This more complex case is the basis for transfer learning in

PBSHM where knowledge transfer between source and target structures takes place where the behaviours of structures (and therefore the collected data) are different but related.

The metrics in this study were able to automatically identify numerous dissimilarities within absolute and standardised data, and give an indication of what type of features are present in the dataset without requiring manual inspection. Interestingly, the metrics detected some structures that behaved dissimilarly to the main population but behaved similarly to each other. As a consequence, it may be possible to group dissimilarly behaving structures in order to learn the mechanisms that drive these changes, or quickly find associations when analysing new data that may have similar trends. Overall, the use of distance metrics was found to be extremely helpful for PBSHM, and is likely to be advantageous to other systems concerned with fleet management.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Chandula T Wickramarachchi (iD) https://orcid.org/0000-0003-2454-6668
Eoghan Maguire (iD) https://orcid.org/0000-0002-1004-1979

References

1. Bull L, Gardner P, Gosliga J, et al. Towards population-based structural health monitoring, Part I: homogeneous

populations and forms. In: *Proceedings of IMAC XXXVIII – the 38th international modal analysis conference*, Houston, TX, USA, 2020.

2. Gosliga J, Gardner P, Bull L, et al. Foundations of population-based SHM, Part II: heterogeneous populations – graphs, networks, and communities. *Mech Syst Signal Process* 2021; 148: 107144.

3. Gosliga J, Gardner P, Bull L, et al. Towards population-based structural health monitoring, Part II: heterogeneous populations and structures as graphs. In: *Proceedings of IMAC XXXVIII – the 38th international modal analysis conference*, Houston, TX, USA, 2020.

4. Gosliga J, Gardner P, Bull L, et al. Towards population-based structural health monitoring, Part III: graphs, networks and communities. In: *Proceedings of IMAC XXXVIII – the 38th international modal analysis conference*, Houston, TX, USA, 2020.

5. Gardner P and Worden K. Towards population-based structural health monitoring, Part IV: heterogeneous populations, matching and transfer. In: *Proceedings of IMAC XXXVIII – the 38th international modal analysis conference*, Houston, TX, USA, 2020.

6. Wickramarachchi C, Brennan D, Lin W, et al. Towards population-based structural health monitoring, Part V: networks and databases. In: *Submitted to Proceedings of IMACXXXIX – the 39th international modal analysis conference*, Orlando, FL, USA, 2021.

7. Worden K. Towards population-based structural health monitoring, Part VI: structures as geometry. In: *Proceedings of IMAC XXXVIII – the 38th international modal analysis conference*, Houston, TX, USA, 2020, pp. 221–236.

8. Lin W, Worden K and Cross E. Towards population-based structural health monitoring, Part VII: EoV fields: environmental mapping. In: *Proceedings of IMAC XXXVIII – the 38th international modal analysis conference*, Houston, TX, USA, 2020.

9. Pan SJ and Yang Q. A survey on transfer learning. *IEEE Trans Know Data Eng* 2009; 22: 1345–1359.

10. Wickramarachchi C, Maguire A, Worden K, et al. On assessing the similarity of wind turbines for population-based structural health monitoring. In: *Proceedings of the 10th international conference on structural health monitoring of intelligent infrastructure*, Porto, 2021.

11. Gardner P, Bull LA, Gosliga J, et al. A population-based SHM methodology for heterogeneous structures: transferring damage localisation knowledge between different aircraft wings. *Mech Syst Signal Process* 2022; 172: 108918.

12. Wickramarachchi C, Gosliga J, Bunce A, et al. Similarity assessment of structures for population-based structural health monitoring via graph kernels. *Struct Health Monit J* 2023.

13. Wickramarachchi C, Poole J, Cross E, et al. On aspects of geometry in SHM and population-based SHM. In: *Proceedings of IMAC-XL– the 40th international modal analysis conference*, Orlando, FL, USA, 2022.

14. Bishop CM and Nasrabadi NM. *Pattern recognition and machine learning*, volume 4. Berlin, Heidelberg: Springer, 2006.

15. Dudley R. *Real analysis and probability*. Cambridge: Cambridge University Press, 2002.

16. Entezami A, Sarmadi H, Behkamal B, et al. Big data analytics and structural health monitoring: a statistical pattern recognition-based approach. *Sensors* 2020; 20: 2328.

17. Kalimullah N, Shelke A and Habib A. Multiresolution dynamic mode decomposition (mrDMD) of elastic waves for damage localisation in piezoelectric ceramic. *IEEE Access* 2021; 9: 120512–120524.

18. Worden K and Cross E. On switching response surface models, with applications to the structural health monitoring of bridges. *Mech Syst Signal Process* 2018; 98: 139–156.

19. Bull L, Worden K, Cross E, et al. Outlier ensembles: an alternative robust method for inclusive outlier analysis with structural health monitoring data. In: *9th European workshop on structural health monitoring, EWSHM 2018*, Manchester, 2018.

20. Gardner P, Liu X and Worden K. On the application of domain adaptation in structural health monitoring. *Mech Syst Signal Process* 2020; 138: 1–24.

21. Gardner P, Bull L, Gosliga J, et al. Foundations of population-based SHM, Part III: heterogeneous populations – Mapping and transfer. *Mech Syst Signal Process* 2021; 149: 107142.

22. Poole J, Gardner P, Dervilis N, et al. On statistic alignment for domain adaptation in structural health monitoring. *Struct Health Monit* 2023; 22: 1581–1600.

23. Gardner P, Bull LA, Dervilis N, et al. Domain-adapted gaussian mixture models for population-based structural health monitoring. *J Civil Struct Health Monit* 2022; 12: 1343–1353.

24. Black IM, Häckell MW and Kolios A. Population based structural health monitoring: homogeneous offshore wind model development. *Wind Energy Sci Discussions* 2022; 1: 1–22.

25. Giglioni V, Poole J, Venanzi I, et al. A domain adaptation approach to damage classification with an application to bridge monitoring, https://ssrn.com/abstract = 4545602, 2023.

26. Bull L, Gardner P, Dervilis N, et al. On the transfer of damage detectors between structures: an experimental case study. *J Sound Vibration* 2021; 501: 116072.

27. Pan Q, Bao Y and Li H. Transfer learning-based data anomaly detection for structural health monitoring. *Struct Health Monit* 2023; 22: 3077–3091.

28. Cha SH. Comprehensive survey on distance/similarity measures between probability density functions. *Int J Math Models Methods Appl Sci* 2007; 1: 300–307.

29. Liese F and Vajda I. On divergences and informations in statistics and information theory. *IEEE Trans Inf Theory* 2006; 52: 4394–4412.

30. Sriperumbudur B, Fukumizu K, Gretton A, et al. On integral probability metrics, -divergences and binary classification. *arXiv preprint arXiv:09012698*, 2009.

31. Lee S, Yun G and Shang S. Reference-free damage detection for truss bridge structures by continuous relative wavelet entropy method. *Struct Health Monit* 2014; 13: 307–320.

32. Gardner P, Lord C and Barthorpe R. A unifying framework for probabilistic validation metrics. *J Verif Valid Uncert Quantif* 2019; 4: 1–11.

33. Wang Q, Kulkarni S and Verdú S. Divergence estimation of continuous distributions based on data-dependent partitions. *IEEE Trans Inf Theory* 2005; 51: 3064–3074.

34. Gretton A, Borgwardt K, Rasch M, et al. A kernel two-sample test. *J Mach Learn Res* 2012; 13: 723–773.

35. Gretton A. *Introduction to RKHS, and some simple kernel algorithms*. London: University College London, 2016.

36. Jia X, Zhao M, Di Y, et al. Assessment of data suitability for machine prognosis using maximum mean discrepancy. *IEEE Trans Ind Electron* 2018; 65: 5872–5881.

37. Wickramarachchi C, Leahy W, Worden K, et al. On metrics assessing the information content of datasets for population-based structural health monitoring. In: *Submitted to special collection of the 10th European workshop on structural health monitoring, EWSHM 2020*, Palermo, Italy, pp. 1–11.

38. Borgwardt K, Gretton A, Rasch M, et al. Integrating structured biological data by Kernel maximum mean discrepancy. *Bioinformatics* 2006; 22: 49–57.

39. Dziugaite G, Roy D and Ghahramani Z. Training generative neural networks via maximum mean discrepancy optimization. In: *Uncertainty in artificial intelligence - Proceedings of the 31st conference, UAI 2015*, Amsterdam, Netherlands, pp. 258–267.

40. Gretton A, Borgwardt K, Rasch M, et al. A kernel approach to comparing distributions. In: *Proceedings of the national conference on artificial intelligence*, Menlo Park, CA, 2007, vol. 2, pp. 1637–1641.

41. Yan H, Ding Y, Li P, et al. Mind the class weight bias: weighted maximum mean discrepancy for unsupervised domain adaptation. In: *Proceedings -30th IEEE conference on computer vision and pattern recognition, CVPR 2017*, Honolulu, Hawaii, January 2017, pp. 945–954.

42. Chen Y, Song S, Li S, et al. A graph embedding framework for maximum mean discrepancy-based domain adaptation algorithms. *IEEE Trans Image Process* 2020; 29: 199–213.

43. Yuan S, Li L and Chigan C. Maximum mean discrepancy based secure fusion strategy for robust cooperative spectrum sensing. In: *IEEE international conference on communications*, Kansas, USA, May 2018.

44. Gangeh M, Tadayyon H, Sannachi L, et al. Computer aided theragnosis using quantitative ultrasound spectroscopy and maximum mean discrepancy in locally advanced breast cancer. *IEEE Trans Medical Imaging* 2016; 35: 778–790.

45. Garreau D, Jitkrittum W and Kanagawa M. Large sample analysis of the median heuristic, *arXiv:1707.07269v3*, 2017.

46. Alonso H, Ribeiro P and Rocha P. An automatic system for on-line change detection with application to structural health monitoring. In: *IECON proceedings (industrial electronics conference)*, Porto, Portugal, 2009. pp. 3309–3314.

47. Gagar D, Foote P and Irving P. Effects of loading and sample geometry on acoustic emission generation during fatigue crack growth: implications for structural health monitoring. *Int J Fatigue* 2015; 81: 117–127.

48. Nichols J, Todd M, Seaver M, et al. Use of chaotic excitation and attractor property analysis in structural health monitoring. *Phys Rev E* 2003; 67: 16209.

49. Kaloop M and Hu J. Stayed-cable bridge damage detection and localization based on accelerometer health monitoring measurements. *Shock Vibr* 2015; 2015: 1–11.

50. Silverman B. Density estimation for statistics and data analysis. In: Cox DR (ed.) *Monographs on Statistics and Applied Probability*. London: Chapman and Hall, 1986.

51. Botev Z, Grotowski J and Kroese D. Kernel density estimation via diffusion. *Ann Stat* 2010; 38: 2916–2957.

52. Sohn H. Effects of environmental and operational variability on structural health monitoring. *Philos Trans Royal Soc A* 2007; 365: 539–560.

53. Girolami A, Brunelli D and Benini L. Low-cost and distributed health monitoring system for critical buildings. In: *2017 IEEE workshop on environmental, energy, and structural monitoring systems, EESMS 2017 - Proceedings*, Milan, Italy.

54. Worden K, Cross E and Barton E. Damage detection on the NPL footbridge under changing environmental conditions. In: *Proceedings of the 6th European workshop on structural health monitoring*, Dresden, Germany, pp. 1–8.

55. Dervilis N, Cross E, Barthorpe R, et al. Robust methods of inclusive outlier analysis for structural health monitoring. *J Sound Vibr* 2014; 333: 5181–5195.

56. Peeters B and De Roeck G. One-year monitoring of the Z24-bridge: environmental effects versus damage events. *Earthquake Eng Struct Dyn* 2001; 30: 149–171.

57. Zhang C, Yu X, Alexander L, et al. Piezoelectric active sensing system for crack detection in concrete structure. *J Civ Struct Health Monit* 2016; 6: 129–139.

58. Prasad B and Babu M. Correlation between vibration amplitude and tool wear in turning: numerical and experimental analysis. *Eng Sci Technol Int J* 2017; 20: 197–211.

59. Li Y, Wang N, Shi J, et al. Adaptive batch normalization for practical domain adaptation. *Pattern Recognit* 2018; 80: 109–117.

60. Tasar O, Tarabalka Y, Giros A, et al. Standardgan: multisource domain adaptation for semantic segmentation of very high resolution satellite images by data standardization. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Seattle, USA, 2020, pp. 192–193.

61. Hu W, Thöns S, Rohrmann RG, et al. Vibration-based structural health monitoring of a wind turbine system Part II: environmental/operational effects on dynamic properties. *Eng Struct* 2015; 89: 273–290.

62. Zhou D, Weston J, Gretton A, et al. Ranking on data manifolds. *Advances in neural information processing systems* 2003; 16.

63. Rasmussen C and Williams C. *Gaussian processes for machine learning*. Massachusetts, USA: MIT Press, 2006.

64. Chen Y, Garcia E, Gupta M, et al. Similarity-based classification: concepts and algorithms. *J Mach Learn Res* 2009; 10: 747–776.

65. Dahlberg JÅ. Assessment of the Lillgrund windfarm: power performance, wake effects. Technical report, Vattenfall, 2009.