Mechanical Systems and Signal Processing xxx (xxxx) 110702

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Transmissibility-based damage detection with hierarchical clustering enhanced by multivariate probabilistic distance accommodating uncertainty and correlation

Lin-Feng Mei [a, b], Wang-Ji Yan [a, b, *], Ka-Veng Yuen [a, b], Wei-Xin Ren [c], Michael Beer [d, e, f]

[a] State Key Laboratory of Internet of Things for Smart City and Department of Civil and Environmental Engineering, University of Macau, China
[b] Guangdong–Hong Kong-Macau Joint Laboratory for Smart Cities, China
[c] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China
[d] Leibniz Universität Hannover, Institute for Risk and Reliability, Hannover, Germany
[e] University of Liverpool, Institute for Risk and Uncertainty, Peach Street, L69 7ZF Liverpool, United Kingdom
[f] Tongji University, International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Shanghai 200092, China

ARTICLE INFO

Keywords:
Transmissibility
Hierarchical clustering
Probabilistic distance
Multivariate distribution
Damage detection

ABSTRACT

This paper proposes a new damage detection method by integrating the advantage of transmissibility function (TF) as a health index sensitive to damage but robust to excitation and agglomerative hierarchical clustering (AHC) with intuitive explanation and visualization but avoiding specifying the number of clusters. Different from conventional AHC-based damage detection methods utilizing deterministic distance as a similarity metric and ignoring the distribution of structural features, a multivariate probabilistic distance-based similarity metric is proposed in this study to account for the uncertainty and correlation of multiple TFs following multivariate complex-valued Gaussian ratio distribution. To realize this, an analytically tractable approximation of the multivariate probabilistic distance is derived by Laplace's asymptotic expansion to avoid high-dimensional numerical integration. To accelerate the computation of probabilistic distances over a wide frequency band that are fused to formulate the similarity metric in AHC, a function vectorization scheme is proposed to avoid the time-consuming loop operation among different frequency points. A threshold is established via bootstrapped Monte Carlo simulation to cut the dendrogram produced by AHC. Two case studies are used to validate the performance of the proposed method, indicating that, compared to the damage detection methods based on the deterministic distance of the TF, the proposed method exhibits better performance due to improving the similarity metric based on multivariate probabilistic distance properly accommodating the correlation of different TFs.

1. Introduction

Engineering structures inevitably experience degradation during long-term operation, which, if not well-handled, could result in structural failure leading to severe casualties and monetary losses. To date, structural health monitoring (SHM) has become a promising candidate to prevent these catastrophic events and preserve the service life of structures. SHM usually refers to the entire process

* Corresponding author.
E-mail address: civilyanwj@gmail.com (W.-J. Yan).

https://doi.org/10.1016/j.ymssp.2023.110702
Received 8 December 2022; Received in revised form 24 June 2023; Accepted 17 August 2023
0888-3270/

of establishing a damage identification or condition analysis system, which mainly involves obtaining response measurements with periodical intervals, extracting damage-sensitive features, as well as conducting statistical analysis of these features to determine the current state of system health and predict the residual life [1]. Damage detection is one of the most widely applied SHM techniques [2,3], which provides a qualitative indication of the existence of damage as well as the basis for further damage identification tasks and can be divided into model-driven methods and data-driven methods [4,5]. Model-driven methods employ numerical models to detect the change in structural condition, but difficulty falls into obtaining a precise model, especially for large complex structures. In contrast, data-driven damage detection relies on statistical pattern recognition of damage-sensitive features based on machine learning algorithms, which mainly contain unsupervised learning and supervised learning methods [1,6]. Supervised damage detection can achieve high-level objectives including damage localization, classification, and quantification based on regression and classification algorithms, but its application is often restricted due to the unavailability of class labels corresponding to damage scenarios. Alternatively, unsupervised damage detection only utilizes data from the baseline condition to determine whether structural damage exists and thus is more applicable in many scenarios.

Currently, unsupervised damage detection approaches are dominated by novelty detection based on outlier analysis, where data from a baseline state is used to define the normal condition and deviations from this condition indicate the occurrence of damage [1]. A pioneering work in this field was proposed by Worden et al. [7], where they adopted the Mahalanobis squared distance (MSD) as the damage index and used a Monte Carlo simulation-based threshold to distinguish the normal condition and the damaged condition. This idea was subsequently followed and modified by many other researchers [8–12]. Inspired by the extreme value theory (EVT), Sarmadi and his colleagues [13–15] proposed a series of statistical methods to estimate the threshold value in damage detection with the presence of environmental and operational variabilities (EOVs). Despite the great achievements reached, most novelty detection methods fall into a two-class assumption, which fails to provide additional operating information on the structure that would help a decision-maker decide on further treatment methods for the detected structural anomaly [16]. Clustering is a popular approach to counteract the two-class assumption and has been widely applied in unsupervised damage detection, which performs a partition of data space into several clusters based on the intrinsic similarity of data points. Santos et al. [17] compared the performance of K-means, Gaussian Mixture Models (GMM), Support Vector Clustering (SVC), and Self-Organizing Maps (SOM) algorithms on benchmark datasets to validate the applicability of clustering-based methods for bridge damage detection. Figueiredo and Cross [18] proposed a GMM-based method for damage detection under unknown sources of variability, which exhibits better performance than some novelty detection methods on the Z24 bridge dataset. Alamdari et al. [19] proposed a spectral-based modified K-means clustering method to identify jack arches with abnormal responses. The feasibility of this method is validated via the monitoring data of the Sydney Harbour Bridge.

However, it is worth noting that a major challenge for these clustering-based damage detection methods is that most clustering algorithms require a predefined number of clusters, which could limit their applications as there is no standard method to determine the optimal number of clusters [19,20]. Conditioned on this frequently encountered issue, agglomerative hierarchical clustering (AHC) seems to be a promising candidate for unsupervised damage detection because its implementation only requires two characteristics, namely the similarity metric and the linkage criterion, while the number of clusters can be determined intuitively based on the similarity between clusters [21]. Meanwhile, AHC could present a clear visualization of multivariate data and identify the data's structure, which is suitable for damage detection with multiple damage scenarios [22]. Over the years, the feasibility of AHC for damage detection has been investigated and verified by many researchers [22–24]. However, these works adopted deterministic similarity metrics such as Euclidean distance and cosine distance, while ignoring the inherent probability distributions of the investigated structural features in the clustering process and thus cannot properly accommodate the uncertainties caused by measurement noise and EOVs, as well as the correlations among multivariate variables. This issue could limit the performance and robustness of AHC-based damage detection methods in real applications.

With the aforementioned issues and challenges, this work proposes an AHC framework enhanced by a multivariate probabilistic distance of TF vectors for damage detection based on the theoretical findings of circularly-symmetric complex Gaussian ratio distribution of TF vectors [25]. TF vector is used as the damage-sensitive feature to incorporate the correlations among dynamic responses at different degree-of-freedoms (DOFs) of the structure. The Bhattacharyya distance (D_B) is adopted to quantify the difference between the probability density functions (PDFs) of TF vectors to accommodate the uncertainties. Based on the Laplace's asymptotic expansion, an approximation method is developed to solve the intractable high-dimensional integration in the calculation of D_B. Then, a data fusion method is adopted to integrate information over a selected frequency band and formulate the similarity metric in AHC, while a function vectorization scheme is proposed to reduce the computational cost based on the independence of TF at different frequencies [26]. A threshold is constructed via bootstrapped Monte Carlo simulation combined with an empirical quantile estimator to cut the hierarchical dendrogram and obtain the damage detection result.

The major contributions of this work can be summarized as follows, and a schematic overview of the proposed damage detection method is shown in Fig. 1:

- A damage detection method is proposed based on TF vectors and a novel AHC framework enhanced by probabilistic distance, which overcomes the limitation of conventional deterministic distance-based AHC that is incapable of properly accommodating uncertainties and correlations among multivariate variables;
- A D_B-based similarity metric is introduced to accommodate the uncertainties and correlations of multiple TFs, and an analytically tractable approximation of D_B is derived based on Laplace's asymptotic expansion;
- A function vectorization scheme is developed to reduce the computational burden caused by integrating information (i.e., TFs) over a wide frequency band.

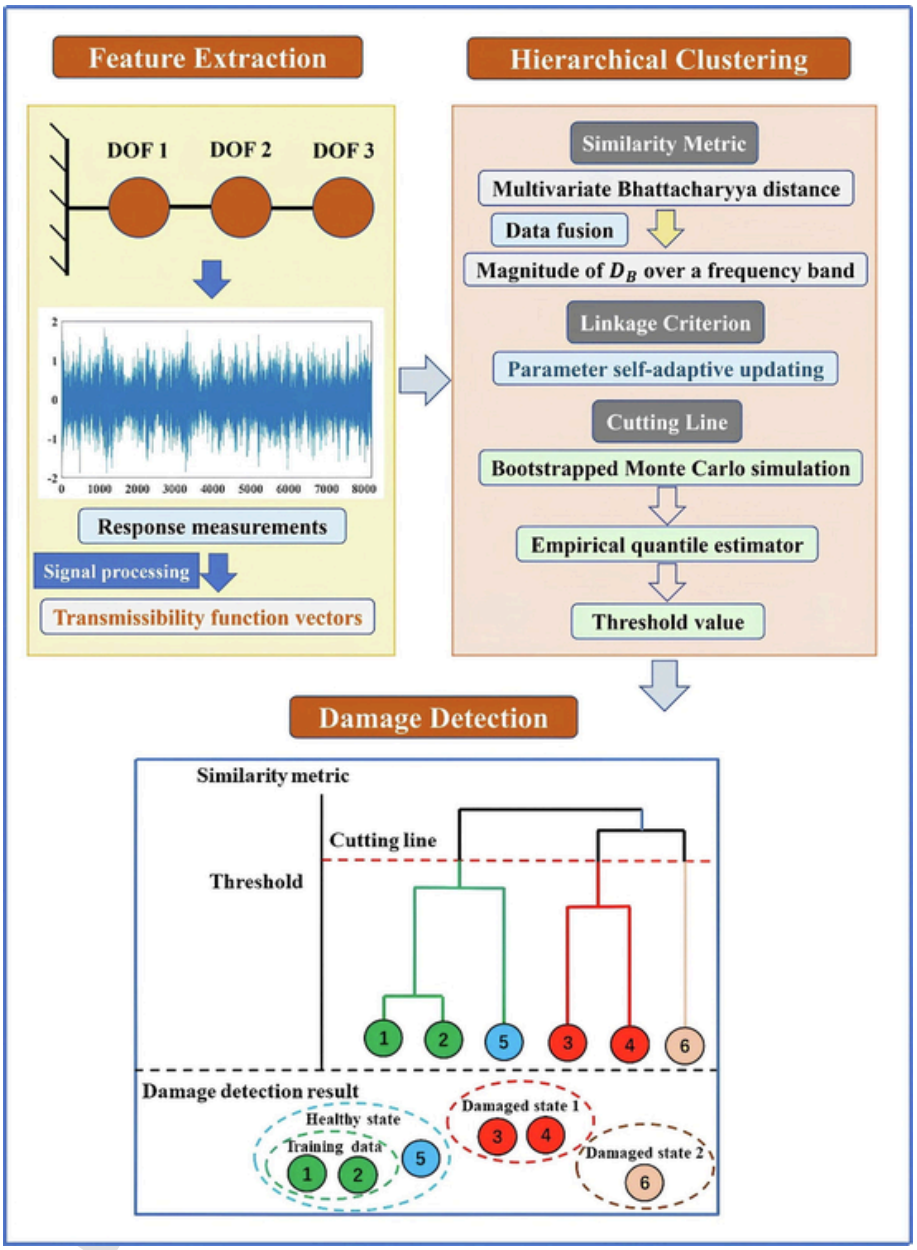


Fig. 1. Schematic overview of the proposed damage detection method.

2. Recap of TF-based damage detection and agglomerative hierarchical clustering

The aim of structural damage detection is to distinguish the data under a damaged state from the normal condition, which requires structure features extracted from raw response measurements that are sensitive to the change in structural condition [1]. Over the years, TF has become an increasingly popular damage-sensitive feature since this output-to-output relationship remains valid in the absence of information about ambient excitation [11,27] and is more sensitive to damage than dynamic properties [28]. A pioneering work of TF-based damage detection was proposed by a research group at NASA, where they adopted the integral of the difference of TFs under different states over a frequency band as the damage indicator [29]. This method has since been applied and further improved in [30,31]. Maia and his colleagues [23,32,33] proposed a series of TF-based damage detection and quantification methods, which demonstrate the applicability and efficiency of TF for damage detection. Considering the uncertainties arising from measurement noise and EOVs, Mao and Todd [34,35] proposed a power spectral density (PSD)-based statistical model to quantify the uncertainty of the magnitude of TFs for damage detection.

Agglomerative hierarchical clustering is an unsupervised technique that can be used to discriminate damage-sensitive features come from different structure states. This technique builds a binary tree called dendrogram initiating from leaf nodes containing individual data points to the root node that includes the whole dataset [21]. Each node represents an independent cluster, and the dendrogram is constructed by iteratively merging the two most similar clusters. Generally, AHC employs a certain deterministic distance to quantify the difference between samples, while the difference between clusters is determined via a specific linkage criterion [21]. However, a limitation of conventional AHC in damage detection is that the deterministic distance-based similarity metrics and weight assignment-based linkage criteria ignore the inherent probability distribution of the damage-sensitive features used for clustering and thus are incapable of properly accommodating the uncertainties and correlations among multivariate variables, which promotes the development of probabilistic distance-based AHC methods.

3. Agglomerative hierarchical clustering enhanced by multivariate Bhattacharyya distance of TFs

In this work, an agglomerative hierarchical clustering framework is proposed for damage detection based on the probabilistic model of TF vectors and the Bhattacharyya distance. The basic assumption is that the probability distribution of TF vectors obtained from the normal condition of the monitored structure diverges from the probability distributions of those drawn from potential damaged states. D_B is adopted to measure the difference between distributions of TF vectors, and the D_B over a frequency band are fused as the similarity metric to improve the robustness of the proposed method. Meanwhile, a function vectorization scheme is proposed to reduce the computational cost and accelerate the data fusion process. Additionally, a bootstrapped Monte Carlo simulation method combined with an empirical quantile estimator is adopted to establish a threshold to cut the dendrogram and obtain the damage detection result.

3.1. Multivariate probabilistic Distance-based similarity metric accommodating uncertainty and correlation

3.1.1. Probability distribution of TF vectors

Conditioned on a set of response measurements in time domain $\boldsymbol{y}(t)=\left\{y_0(t),y_1(t),...,y_{n_o-1}(t)\right\}$ for a linear system with n_o DOFs under a stationary excitation, the dynamic responses in the frequency domain can be obtained via Fast Fourier Transform (FFT) and are denoted by $\boldsymbol{Y}_k=\boldsymbol{Y}_k^{\Re}+\mathrm{i}\boldsymbol{Y}_k^{\Im}=\left\{Y_0^k,Y_1^k,...,Y_{n_o-1}^k\right\}$. A TF is defined as the ratio of an arbitrary response Y_i^k to a reference response Y_j^k and is denoted by U_{ij}^k:

$$U_{ij}^k=\frac{Y_i^k}{Y_j^k} \tag{1}$$

It is worth noting that all the subscripts and superscripts, k, in this work denote the frequency ω_k. Fixing the reference response, a TF vector at frequency ω_k can be formulated:

$$\boldsymbol{U}^k=\left\{U_{1,0}^k,U_{2,0}^k,...,U_{i,0}^k,...,U_{n_o-1,0}^k\right\}^T_{(n_o-1)\times 1}=\left\{\frac{Y_1^k}{Y_0^k},\frac{Y_2^k}{Y_0^k},...,\frac{Y_i^k}{Y_0^k},...,\frac{Y_{n_o-1}^k}{Y_0^k}\right\}^T_{(n_o-1)\times 1} \tag{2}$$

According to Ref. [25], the TF vector $\boldsymbol{U}=\boldsymbol{U}^{\Re}+\mathrm{i}\boldsymbol{U}^{\Im}$ follows a multivariate circularly-symmetric complex Gaussian ratio distribution, and the real-valued vector integrating the real and imaginary parts of the TF vector, denoted by $\gamma=\left[\left(\boldsymbol{U}^{\Re}\right)^T,\left(\boldsymbol{U}^{\Im}\right)^T\right]^T$, has a PDF in the following form:

$$p_\gamma(\boldsymbol{\eta})=\frac{\left(n_o-1\right)!}{\pi^{n_o-1}\left|\widetilde{\boldsymbol{\Sigma}}\right|^{\frac{1}{2}}\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}\right]^{n_o}} \tag{3}$$

where $\boldsymbol{\eta}=\left[1,\left(\boldsymbol{u}^{\Re}\right)^T,0,\left(\boldsymbol{u}^{\Im}\right)^T\right]_{2n_o\times 1}$; $\boldsymbol{u}=\boldsymbol{u}^{\Re}+\mathrm{i}\boldsymbol{u}^{\Im}$ composes of the values of the ratio vector $\boldsymbol{U}$; $\widetilde{\boldsymbol{\Sigma}}=\frac{1}{2}\begin{bmatrix}\boldsymbol{\Sigma}^{\Re} & -\boldsymbol{\Sigma}^{\Im}\\ \boldsymbol{\Sigma}^{\Im} & \boldsymbol{\Sigma}^{\Re}\end{bmatrix}$; $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}^{\Re}+\mathrm{i}\boldsymbol{\Sigma}^{\Im}$ denotes the covariance matrix of the frequency domain responses $\boldsymbol{Y}$. The statistical model of TF vectors will be used for damage detection in combination with an agglomerative hierarchical clustering framework in the following part.

3.1.2. Bhattacharyya distance between probability distributions of different TF vectors

In this work, the Bhattacharyya distance (D_B) is adopted to measure the difference between distributions of TF vectors as it has been proved in [36,37] that D_B could capture a higher level of statistical information from the investigated variables and is more comprehensive for uncertainty treatment than some traditional metrics. For two continuous probability distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ over $\boldsymbol{\Theta}$, D_B is defined as [38]:

$$D_B(p(\boldsymbol{x}), q(\boldsymbol{x})) = -\ln\left(\int \sqrt{p(\boldsymbol{x})\, q(\boldsymbol{x})}\, d\boldsymbol{x}\right) \tag{4}$$

Assume that the PDF of the real-valued TF vector under the baseline condition is denoted by $p_\gamma(\boldsymbol{\eta})$ and is given in Eq. (3). To avoid confusion, under a possibly damaged state, the PDF of TF vectors is denoted by:

$$p_\gamma^d(\boldsymbol{\eta}) = \frac{(n_o - 1)!}{\pi^{n_o - 1}\left|\widetilde{\boldsymbol{\Sigma}}_d\right|^{\frac{1}{2}}\left[\boldsymbol{\eta}^T \widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}\right]^{n_o}} \tag{5}$$

Therefore, according to Eqs. (3)-(5), D_B between $p_\gamma(\boldsymbol{\eta})$ and $p_\gamma^d(\boldsymbol{\eta})$ could be expressed as:

$$D_B\left(p_\gamma(\boldsymbol{\eta}), p_\gamma^d(\boldsymbol{\eta})\right) = -\ln\left(\int \sqrt{p_\gamma(\boldsymbol{\eta})\, p_\gamma^d(\boldsymbol{\eta})}\, d\boldsymbol{\eta}\right) \tag{6}$$

Equation (6) describes a probabilistic distance between two distributions of TF vectors under different states, which will be adopted to construct the similarity metric in the proposed AHC framework. It is worth noting that it is non-trivial for the Bhattacharyya distance to discern whether the difference between two distributions arises from different means, different covariances, or both, as D_B between two distributions only depends on the overlap area of their probability density functions [38]. As a result, Bhattacharyya distance may not explicitly indicate the specific type of change in TF vector distributions caused by structural damage. However, in the context of damage detection, this limitation would not be a significant concern because damage detection methods mainly focus on providing a qualitative indication of the occurrence of structural damage without delving into detailed analysis of the underlying damage mechanisms. Regardless of whether the deviation in the distribution of TF vectors from the normal condition stems from difference in means, covariances, or both, it will change the overlap of these two distributions, and thus can be detected using the Bhattacharyya distance.

3.1.3. Laplace's asymptotic approximation for Bhattacharyya distance

Despite the attractive properties of D_B, its application is significantly restricted by the stochastic feature and unaffordable computational burden resulting from integrating over the product of PDFs of multivariate variables. To overcome this limitation, the Laplace's asymptotic expansion is adopted to avoid numerical integration. Consider an integral in the following form:

$$I(\boldsymbol{\theta}) = \int_{\boldsymbol{\Omega}} p(\boldsymbol{\theta})\, q(\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{7}$$

where $p(\boldsymbol{\theta})$ and $q(\boldsymbol{\theta})$ are smooth functions for $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_n\}^T$ and $\boldsymbol{\Omega}$ is a subregion of $\mathfrak{R}^n$. By applying Laplace's method of asymptotic expansion to $I(\boldsymbol{\theta})$, an approximation of the integral can be obtained [39]:

$$I(\boldsymbol{\theta}) \approx (2\pi)^{\frac{n}{2}} p(\boldsymbol{\theta}^*)\, q(\boldsymbol{\theta}^*)\left|H(\boldsymbol{\theta}^*)\right|^{-\frac{1}{2}} \tag{8}$$

where n is the dimension of $\boldsymbol{\theta}$; $\boldsymbol{\theta}^*$ is the global maximum of the integrand $p(\boldsymbol{\theta})\, q(\boldsymbol{\theta})$ over $\boldsymbol{\Omega}$; $H(\boldsymbol{\theta}^*)$ refers to the Hessian matrix of $f(\boldsymbol{\theta}) = -\ln\left[p(\boldsymbol{\theta})\, q(\boldsymbol{\theta})\right]$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$; and $|\cdot|$ denotes the determinant. Based on the Laplace's approximation method, Eq. (6) can be rewritten as:

$$D_B\left(p_\gamma(\boldsymbol{\eta}), p_\gamma^d(\boldsymbol{\eta})\right) \approx -\ln\left((2\pi)^{\frac{n_o}{2}} \sqrt{p_\gamma(\boldsymbol{\eta}^*)\, p_\gamma^d(\boldsymbol{\eta}^*)}\left|H(\boldsymbol{\eta}^*)\right|^{-\frac{1}{2}}\right) \tag{9}$$

Assume $\boldsymbol{\tau} = \left[\left(\boldsymbol{u}^{\mathfrak{R}}\right)^T, \left(\boldsymbol{u}^{\mathfrak{I}}\right)^T\right]^T_{2(n_o-1)\times 1}$ contains the variables in $\boldsymbol{\eta}$, the global maximum $\boldsymbol{\tau}^*$ can be derived via matrix derivative and matrix blocking methods:

$$\left[\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\tau}}\right]^T \frac{\partial \left[p_\gamma(\boldsymbol{\eta}) p_\gamma^d(\boldsymbol{\eta})\right]}{\partial \boldsymbol{\eta}}\bigg|_{\boldsymbol{\tau}=\boldsymbol{\tau}^*} = \mathbf{0} \tag{10}$$

Then $\boldsymbol{\eta}^*$ can be obtained by back substitution accordingly. In terms of $H(\boldsymbol{\eta})$, it can be expressed as Eq. (11) and calculated via matrix derivative:

$$H(\boldsymbol{\eta}) = \left[-\frac{\partial^2 \ln\left[\sqrt{p_\gamma(\boldsymbol{\eta}) p_\gamma^d(\boldsymbol{\eta})}\right]}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\right] = \left[-\frac{1}{2}\frac{\partial^2 \ln\left[p_\gamma(\boldsymbol{\eta})\right]}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} - \frac{1}{2}\frac{\partial^2 \ln\left[p_\gamma^d(\boldsymbol{\eta})\right]}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\right] \tag{11}$$

The detailed derivation process of the global maximum $\boldsymbol{\eta}^*$ and the Hessian matrix $H(\boldsymbol{\eta})$ are described in Appendices A and B, respectively. This approximation avoids direct integration over multivariate variables and hence significantly enhances the feasibility of the proposed damage detection method. Nevertheless, it introduces additional computational costs, as well as potential approximation errors, compared to using alternative distances such as the Euclidean distance and Mahalanobis distance, which is another drawback of using the Bhattacharyya distance.

3.1.4. Function vectorization scheme for data fusion

Conditioned on the asymptotic independence of TF at different frequency points [26], TF vectors over a frequency band are incorporated for damage detection to improve the robustness of the proposed method. The frequency band is determined using the feature selection approach introduced in [40], aiming at obtaining features that are sensitive to damage while insensitive to EOVs. The advantage of this approach is that it can be directly implemented on the proposed damage detection method without any preprocessing and postprocessing. Besides, this feature selection approach can be conducted without data from the damaged states as damage usually causes linear structure exhibit nonlinear behavior [16,41]. Assume the selected frequency band is denoted by $\left[\omega_{k_1}, \omega_{k_2}\right] = \left[k_1 \Delta\omega, k_2 \Delta\omega\right]$ with $\Delta\omega$ denoting the frequency resolution, a D_B vector can be formulated using the D_B values at each individual frequency point and is denoted by $D_B^{[k_1,k_2]} = \left\{D_B^{k_1}, D_B^{k_1+1}, ..., D_B^{k_2}\right\}$. Then, the magnitude of $D_B^{[k_1,k_2]}$ is used as the similarity metric in AHC to quantify the difference between each pair of clusters, which is expressed as:

$$MD_B = \sqrt{\left(D_B^{[k_1,k_2]}\right) \cdot \left(D_B^{[k_1,k_2]}\right)^T} = \sqrt{\left(D_B^{k_1}\right)^2 + \left(D_B^{k_1+1}\right)^2 + ... + \left(D_B^{k_2}\right)^2} \tag{12}$$

Although the data fusion process improves the robustness of the proposed method since more information is incorporated, it would unavoidably lead to additional computational cost that could limit the application of the damage detection method. Conditioned on this issue, a function vectorization scheme in the environment of MATLAB is introduced to reduce the computational burden. There are two types of arithmetic operations in MATLAB, namely matrix operations based on linear algebra, and array operations conducted element-wisely. The basic principle of the vectorization scheme is extending element-wise operations to matrix-wise [42]:

- Assume $\mathbf{A}_k$ and $\mathbf{B}_k$ denote the k th slide of the three-dimensional matrices $\mathbf{A}$ and $\mathbf{B}$ with the same size, while c is a scalar value. The operation $\oplus$ is defined as the sum of a matrix slide with another one or with a scalar:

$$\mathbf{A}_k \oplus \mathbf{B}_k = \mathbf{A}(:,:,k) + \mathbf{B}(:,:,k) \tag{13a}$$
$$\mathbf{A}_k \oplus c = \mathbf{A}(:,:,k) + c \tag{13b}$$

- Correspondingly, simultaneous summation of a set of matrix slides with another set of matrix slides or with a scalar is defined as:

$$\mathbf{A}_{[k_1,k_2]} \oplus \mathbf{B}_{[k_1,k_2]} = \mathbf{A}\left(:,:,k_1 : k_2\right) + \mathbf{B}\left(:,:,k_1 : k_2\right) \tag{14a}$$
$$\mathbf{A}_{[k_1,k_2]} \oplus c = \mathbf{A}(:,:,k_1 : k_2) + c \tag{14b}$$

- On the other hand, the product of a matrix slide with another one or a scalar is denoted by the symbol $\otimes$:

$$\mathbf{A}_k \otimes \mathbf{B}_k = \mathbf{A}(:,:,k) \times \mathbf{B}(:,:,k) \tag{15a}$$
$$\mathbf{A}_k \otimes c = \mathbf{A}(:,:,k) \times c \tag{15b}$$

- Similarly, simultaneous multiplication operation of $\mathbf{A}_{[k_1,k_2]}$ with $\mathbf{B}_{[k_1,k_2]}$ or with c is defined as:

$$\mathbf{A}_{[k_1,k_2]} \otimes \mathbf{B}_{[k_1,k_2]} = \mathbf{A}\left(:,:,k_1:k_2\right) \times \mathbf{B}\left(:,:,k_1:k_2\right) \tag{16a}$$

$$\mathbf{A}_{[k_1,k_2]} \otimes c = \mathbf{A}(:,:,k_1:k_2) \times c \tag{16b}$$

Based on these symbols, D_B values over the selected frequency band can be calculated simultaneously via matrix-wise operations. According to Eq. (9), and the appendices, D_B at the frequency ω_k, denoted by D_B^k, is composed of three parts, namely $p_\gamma(\boldsymbol{\eta}^*)^k$, $p_\gamma^d(\boldsymbol{\eta}^*)^k$, and $H(\boldsymbol{\eta}^*)^k$. Use $(\boldsymbol{\eta}^*)^{[k_1,k_2]}$, $\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}$, and $\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}$ to denote the global maximums and the covariance matrices under different states over the selected frequency band, respectively. Then, it has:

$$p_\gamma(\boldsymbol{\eta}^*)^{[k_1,k_2]} = \frac{\left(n_o-1\right)!}{\pi^{n_o-1} \otimes \left|\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right|^{\frac{1}{2}} \otimes \left[\left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1} \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]}\right]^{n_o}} \tag{17a}$$

$$p_\gamma^d(\boldsymbol{\eta}^*)^{[k_1,k_2]} = \frac{\left(n_o-1\right)!}{\pi^{n_o-1} \otimes \left|\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right|^{\frac{1}{2}} \otimes \left[\left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1} \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]}\right]^{n_o}} \tag{17b}$$

$$\begin{aligned} H(\boldsymbol{\eta}^*)^{[k_1,k_2]} = -\frac{n_o}{2} \otimes \Bigg\{ & \left[-\frac{\left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1} \oplus \left(\left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1}\right)^T}{\left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1} \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]}} \right] \\ & \oplus \frac{\left[\left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1} \oplus \left(\left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1}\right)^T\right] \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]} \otimes \left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left[\left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1} \oplus \left(\left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1}\right)^T\right]}{\left[\left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left(\widetilde{\boldsymbol{\Sigma}}^{[k_1,k_2]}\right)^{-1} \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]}\right]^2} \\ & \oplus \frac{\left[\left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1} \oplus \left(\left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1}\right)^T\right] \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]} \otimes \left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left[\left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1} \oplus \left(\left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1}\right)^T\right]}{\left[\left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1} \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]}\right]^2} \\ & \oplus \left[-\frac{\left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1} \oplus \left(\left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1}\right)^T}{\left((\boldsymbol{\eta}^*)^{[k_1,k_2]}\right)^T \otimes \left(\left(\widetilde{\boldsymbol{\Sigma}}_d\right)^{[k_1,k_2]}\right)^{-1} \otimes (\boldsymbol{\eta}^*)^{[k_1,k_2]}} \right] \Bigg\} \end{aligned} \tag{17c}$$

where the superscripts $(\cdot)^{-1}$ and $(\cdot)^T$ denote the inverse and transpose of each element contained in the corresponding vectors, respectively. Then, $D_B^{[k_1,k_2]}$ can be derived using matrix-wise operations among $p_\gamma(\boldsymbol{\eta}^*)^{[k_1,k_2]}$, $p_\gamma^d(\boldsymbol{\eta}^*)^{[k_1,k_2]}$, and $H(\boldsymbol{\eta}^*)^{[k_1,k_2]}$ accordingly:

$$D_B^{[k_1,k_2]}\left(p_\gamma(\boldsymbol{\eta}), p_\gamma^d(\boldsymbol{\eta})\right) = -\ln\left((2\pi)^{\frac{n_o}{2}} \otimes \left[p_\gamma(\boldsymbol{\eta}^*)^{[k_1,k_2]}\right]^{\frac{1}{2}} \otimes \left[p_\gamma^d(\boldsymbol{\eta}^*)^{[k_1,k_2]}\right]^{\frac{1}{2}} \otimes \left|H(\boldsymbol{\eta}^*)^{[k_1,k_2]}\right|^{-\frac{1}{2}}\right) \tag{18}$$

The function vectorization scheme improves the computational efficiency and facilitates the data fusion process. It is worth noting that the function "bsxfun" in MATLAB can be used to achieve this vectorization scheme directly.

3.2. Linkage criteria using a parameter Self-adaptive updating scheme

In the context of AHC, the linkage criterion specifies the similarity between clusters based on that between individual samples. There are four types of commonly used linkage criteria, namely single-linkage, complete-linkage, average-linkage, and Ward's linkage [21]. However, these methods simply assign different weights to individual samples without considering the probability distributions represented by the clusters and thus could lead to misleading results in real applications. This motivates the authors to propose a novel linkage criterion accommodating the statistical model of TF vectors.

According to Section 3.1.3 and the appendices, the approximated D_B between $p_\gamma(\boldsymbol{\eta})$ and $p_\gamma^d(\boldsymbol{\eta})$ can be derived given the covariance matrices of FFT coefficients. Therefore, a cluster can be represented by the covariance matrix of FFT coefficients of the data contained in it, denoted by $\boldsymbol{\Sigma}$, in AHC. According to Ref. [43,44], the covariance matrix of FFT coefficients $\boldsymbol{Y}$ can be estimated via the expected value of the power spectral density (PSD) matrix $\mathbf{S}$, namely $\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{S}]$. Based on these theoretical findings, a parameter self-adaptive updating scheme is proposed as the linkage criterion in this work and is depicted in Fig. 2: Assume c_m denotes a certain cluster, $\mathscr{D}_m$ denotes the data contained in c_m and there are N_m sets samples contained in $\mathscr{D}_m$. At the k th frequency point, the covariance matrix of the FFT coefficients in c_m can be estimated by taking the average of the PSD matrices derived from each set of FFT coefficients:

$$\boldsymbol{\Sigma}_k^m = \mathbb{E}\left[\mathbf{S}_k^m\right] = \frac{1}{N_m}\sum_{n=1}^{N_m} \boldsymbol{Y}_k^{(n)}\left(\boldsymbol{Y}_k^{(n)}\right)^* \tag{19}$$

where $\boldsymbol{Y}^{(n)}$ denotes the n th FFT coefficients; $\boldsymbol{Y}^*$ denotes the complex conjugate of $\boldsymbol{Y}$. At a certain level in the AHC, two clusters c_i and c_j are merged to form a new cluster $c_{i,j}$, then the data contained in $c_{i,j}$ can be expressed as $\mathscr{D}_{i,j} = \mathscr{D}_i \cup \mathscr{D}_j$, and the corresponding covariance matrix $\boldsymbol{\Sigma}_k^{i,j}$ can be estimated via the expected value of PSD matrix of $\mathscr{D}_{i,j}$:

$$\boldsymbol{\Sigma}_k^{i,j} = \mathbb{E}\left[\mathbf{S}_k^{i,j}\right] = \frac{1}{N_i+N_j}\sum_{n=1}^{N_i+N_j} \boldsymbol{Y}_k^{(n)}\left(\boldsymbol{Y}_k^{(n)}\right)^* \tag{20}$$

where N_i and N_j are the number of samples contained in $\mathscr{D}_i$ and $\mathscr{D}_j$, respectively. During the AHC process, the covariance matrix of the newly merged cluster is reassessed at each level, and then the similarity metric matrix between clusters is reconstructed. This process terminates when all data are included in one cluster. It is worth noting that n_l continuous measurements are used to estimate the parameters of the clusters in the leaf nodes in this work, aiming at improving the robustness of the proposed damage detection method.

3.3. Threshold estimation based on bootstrapped Monte Carlo simulation

The AHC algorithm generates a dendrogram that depicts the cluster merging process from leaf nodes to the root node. Then, a cutting line is supposed to be established to obtain the clustering result, which is usually a threshold of the similarity metric [21]. In this work, the threshold value of MD_B is estimated via bootstrapped Monte Carlo simulation [45] combined with an empirical quantile estimator [15]. The rationale is that the training data from the normal condition are independent and identically distributed (i.i.d.) theoretically, and thus the difference of MD_B between each pair of clusters in the training set is caused by modelling error and measurement noise. In the proposed AHC algorithm, if the MD_B between the clusters to be merged is less than the threshold, data in these two clusters are supposed to follow the same distribution; otherwise, the two clusters are believed to represent different distributions. The advantage of this method is that the bootstrapped Monte Carlo simulation could generate samples based on limited response measurements without considering the underlying distribution of samples [45]. In contrast, generating TF vector samples directly from the multivariate Gaussian ratio distribution seems not to be computationally affordable, especially for large-scale infrastructures with many DOFs.

Fig. 3 presents the process of the threshold estimation method, which mainly contains the following steps:

(i) Conditioned on a $(p \times n)$ training set (number of dimensions × number of observations) obtained from the normal condition, the observations are denoted by an index vector $\boldsymbol{\Phi} = [1, 2, \ldots, n]$;

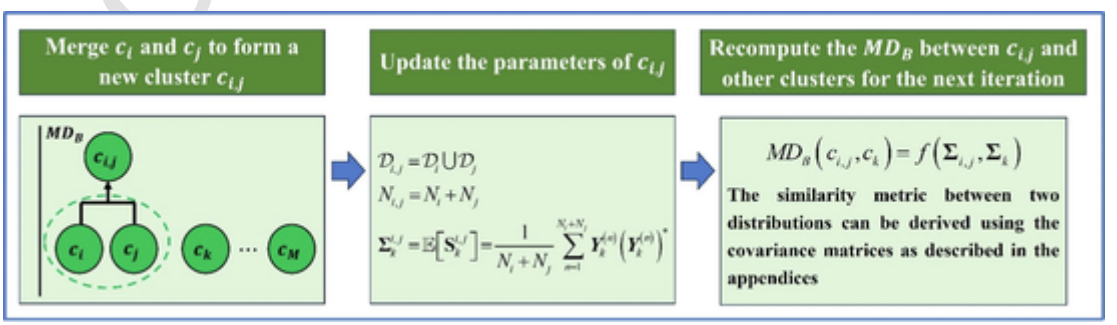


Fig. 2. Flowchart of the parameter self-adaptive updating scheme.

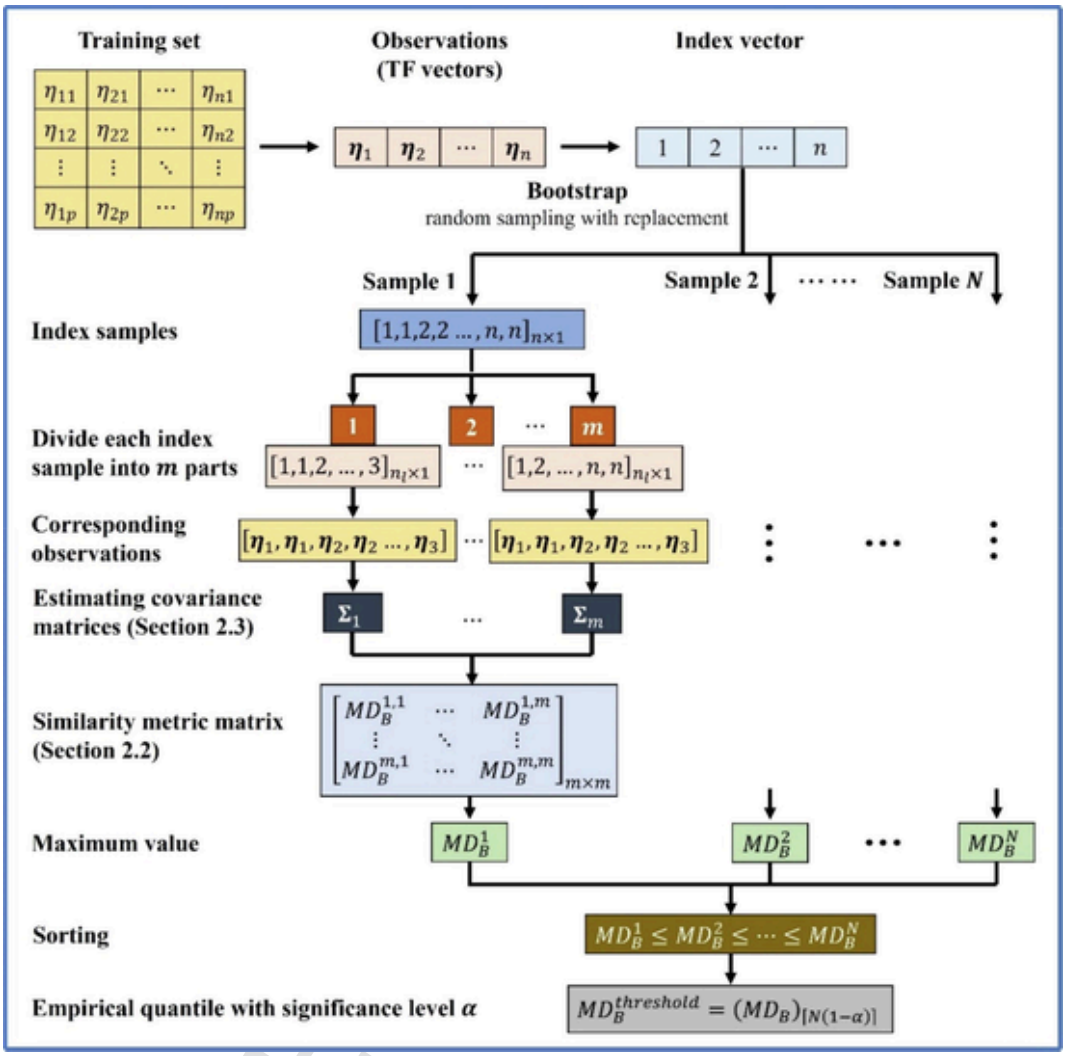


Fig. 3. Steps of estimating the threshold value of the similarity metric.

(ii) Conduct bootstrapped Monte Carlo simulation to generate n samples with replacement from the index vector $\boldsymbol{\Phi}$, and separate these samples into m parts with $m = \frac{n}{n_l}$, where n_l denotes the number of measurements used to estimate the parameters of clusters in the leaf nodes described in Section 3.2;

(iii) Construct m clusters using the corresponding observations of the m parts of the index samples in step (ii), denoted by $[c_1, c_2, ..., c_m]$, and estimate corresponding covariance matrices $[\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, ..., \boldsymbol{\Sigma}_m]$ according to Section 3.2;

(iv) Compute the MD_B between each pair of the m clusters according to Section 3.1 and the maximum value is stored;

(v) Repeat steps (ii)-(iv) for a large number (N) of times and sort the maximum values in ascending order, namely $MD_B^1 \le MD_B^2 \le ... \le MD_B^N$, the threshold value is defined as the empirical quantile with a significance level of α:

$$MD_B^{threshold} = MD_B^{\lceil N(1-\alpha) \rceil} \tag{21}$$

where $\lceil n \rceil$ denotes the smallest integer greater than or equal to n. For more details about empirical quantile estimator, one can refer to [46].

4. Procedures of the novel damage detection method

The detailed process of the proposed damage detection method is summarized in Fig. 4 and involves the following steps:

(1) **Constructing TF vectors based on raw response measurements**

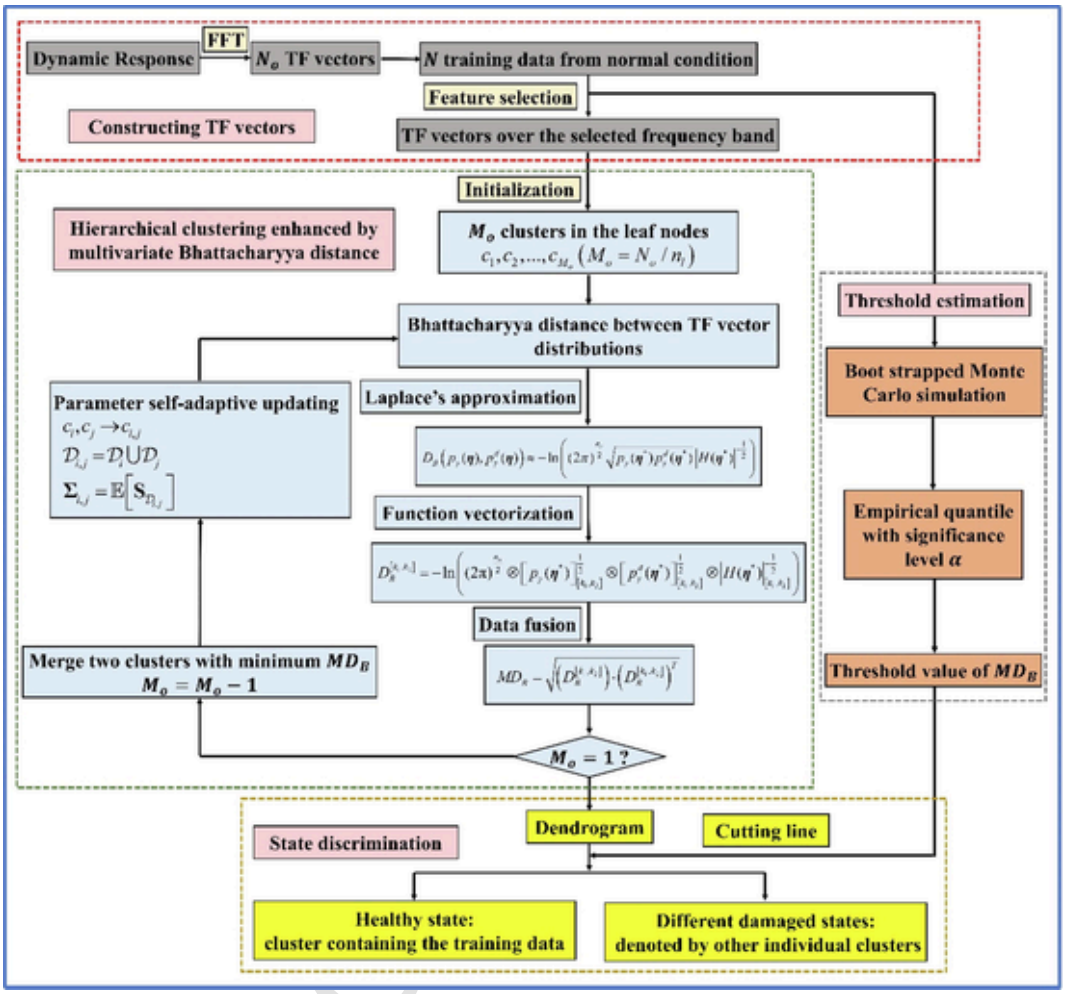


Fig. 4. Flowchart of the transmissibility-based damage detection method with agglomerative hierarchical clustering enhanced by multivariate probabilistic distance.

- Acquire N_o sets of response measurements of the monitored structure in the time domain;
- Formulate N_o TF vectors incorporating responses at all DOFs;
- Choose N TF vectors as the training set for selecting the frequency band (Section 3.1.4) and estimating the threshold (Section 3.3). It is worth noting that the training set only contains data from the normal condition in this work.

(2) **Agglomerative hierarchical clustering enhanced by multivariate probabilistic distance and function vectorization**

- Use n_l continuous measurements to form each cluster in the leaf nodes as described in Section 3.2, which will lead to M_o clusters in the leaf nodes with $M_o = \frac{N_o}{n_l}$, and estimate the covariance matrices of these clusters;
- Calculate the MD_B between each pair of clusters in the leaf nodes according to Section 3.1, and construct the initial MD_B matrix with the size of $M_o \times M_o$;
- Merge the two clusters with minimum MD_B and reassess the parameters of the merged cluster as described in Section 3.2, $M_o = M_o - 1$, then calculate the MD_B between each pair of clusters and construct a new MD_B matrix;
- Repeat the above two steps until $M_o = 1$ and construct the hierarchical dendrogram.

(3) **Threshold estimation based on bootstrapped Monte Carlo simulation and empirical quantile**

- Determine the threshold value of MD_B based on bootstrapped Monte Carlo simulation and empirical quantile estimator with significance level α using data in the training set as described in Section 3.3;
- Establish a constant cutting line to cut the dendrogram at the threshold value.

(4) **State discrimination based on the new agglomerative hierarchical clustering algorithm**

- The threshold value differentiates the variation in MD_B caused by modeling error and measurement noise from that caused by different damage scenarios. If the MD_B between two clusters exceeds the threshold value, they represent different structure states; otherwise, the MD_B between these clusters is caused by modeling error and measurement noise and they should be merged;

- The dendrogram and the cutting line will divide the TF vector samples into several clusters. The cluster containing the training data represents the normal condition, while other clusters represent different damaged states. It is worth noting that as an unsupervised method, the proposed method is only able to assign samples to different clusters to indicate that they come from different structure states. However, it cannot provide explicit definitions for these clusters except for the cluster representing the normal condition;
- When new TF vector samples are available, the process of constructing the dendrogram is repeated. Then, the dendrogram is cut using the predetermined threshold value to obtain the damage detection result. It is worth noting that as the bootstrapped Monte Carlo simulation is robust against data scarcity [45], threshold estimation does not need to be repeated when new data are observed.

5. Case studies

5.1. Laboratory experiment: A three-story building structure

The AHC-assisted damage detection method is validated via a benchmark experimental result of a three-story building structure (Fig. 5) conducted by the research group from Los Alamos National Laboratory [47]. This experiment is designed for damage detection with the presence of EOVs, which is one of the major factors that affect the accuracy of damage detection methods. 17 different structure states were included in this test and are summarized in Table 1. The linear effects (the alteration in mass and stiffness of certain floors) introduced in states 2–9 were used to simulate the effect of EOVs, while structural damage was represented by the nonlinearity of repeated impact introduced through a bumper mechanism shown in Fig. 5. Different gaps between the column and the bumper (states 10–17) denoted different damage extents. Four accelerometers were mounted at the center line of the base and each

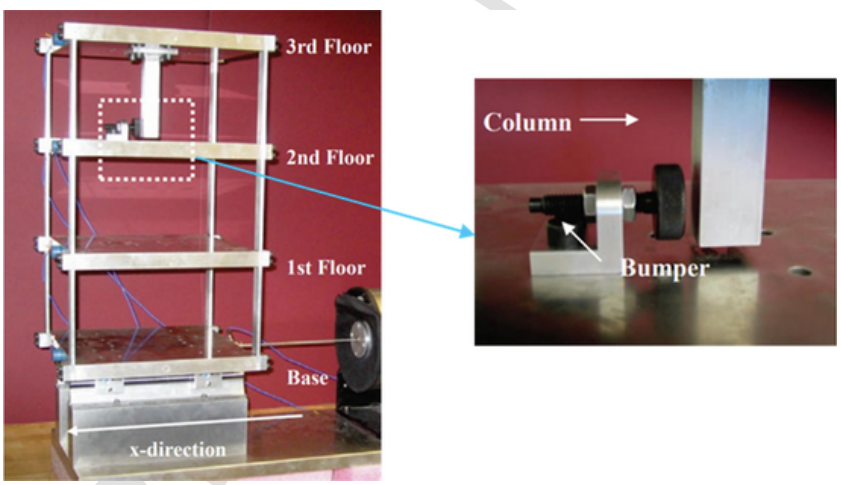


Fig. 5. The tested three-story building structure, reproduced from Ref. [47].

Table 1
Details of structure states in the experiment, reproduced from Ref. [47].

Label	State Condition	Description
State 1	Undamaged	Baseline condition
State 2	Undamaged	Added mass (1.2 kg) at the base
State 3	Undamaged	Added mass (1.2 kg) on the first floor
State 4	Undamaged	87.5% stiffness reduction in column 1BD
State 5	Undamaged	87.5% stiffness reduction in column 1AD and 1BD
State 6	Undamaged	87.5% stiffness reduction in column 2BD
State 7	Undamaged	87.5% stiffness reduction in column 2AD and 2BD
State 8	Undamaged	87.5% stiffness reduction in column 3BD
State 9	Undamaged	87.5% stiffness reduction in column 3AD and 3BD
State 10	Damaged	Gap = 0.20 mm
State 11	Damaged	Gap = 0.15 mm
State 12	Damaged	Gap = 0.13 mm
State 13	Damaged	Gap = 0.10 mm
State 14	Damaged	Gap = 0.05 mm
State 15	Damaged	Gap (0.20 mm) and mass (1.2 kg) at the base
State 16	Damaged	Gap (0.20 mm) and mass (1.2 kg) on the first floor
State 17	Damaged	Gap (0.10 mm) and mass (1.2 kg) on the first floor

floor to measure the responses while minimizing the torsional effect. 50 measurements were conducted in each state with a sampling frequency of 320 Hz and a duration of 25.6 s. Therefore, 8192 discrete data points were acquired in each channel in one measurement, and an overall response dataset with the size of $8192 \times 4 \times 850$ was obtained in the experiment. For more details about the laboratory experiment, one can refer to [47].

5.1.1. Correlations among TFs at different degree-of-freedoms

Choose the responses at the base as the reference responses, one can obtain four complex-valued TFs at each frequency point and thus formulate an eight-dimensional TF vector incorporating the real and imaginary parts of TFs. Fig. 6 presents the pairwise correlation plot of TFs at a certain frequency point, where the correlations among the components of TF vectors can be clearly observed, illustrating an attractive advantage of the proposed damage detection method: the correlations among TFs at different DOFs are incorporated in damage detection via the PDF of TF vectors as well as the multivariate Bhattacharyya distance.

5.1.2. Damage detection result

Data from the healthy states (states 1–9) is used as the training set, and the 50 measurements obtained from each structure state are used to estimate the covariance matrices of the clusters in the leaf nodes as mentioned in Section 3.2. The threshold value is estimated via bootstrapped Monte Carlo simulation [45] in combination with empirical quantile [46] to establish the cutting line of the hierarchical dendrogram as described in Section 3.3. Fig. 7 displays the histogram of MD_B samples obtained from 1000 times bootstrapped Monte Carlo simulation, where the threshold is defined using empirical quantile with a significance level $\alpha = 5\%$ as it is commonly used in statistical applications [46].

Fig. 8 presents the dendrogram as well as the cutting line of the proposed AHC framework, and the damage detection result of the proposed method is shown in Fig. 9. According to these figures, one can conclude that the proposed method creates five clusters to partition the TF vectors and each damage scenario introduced in this experiment is detected. Cluster 1 contains data from states 1–9, and thus denote the healthy state. Cluster 2 contains states 10, 15 and 16, which are damaged states with the lowest damage extent (0.20 mm Gap) according to Table 1. Meanwhile, there are three clusters accounting for other states with higher damage extent. States 12, 13 and 17 have similar damage extents and are assigned to one cluster, while state 11 (0.15 mm Gap) and state 14 (0.05 mm Gap) form individual clusters. Compared to traditional novelty detectors, the AHC-based damage detection method provides the similarity between each pair of structure states investigated and results in a number of clusters to account for different damage scenarios, which would help further analysis of the detected structural anomalies [16]. Moreover, the proposed method retains the flexibility to be conducted in a semi-supervised or supervised manner. For example, if the label of state 10 is known a priori, it could be concluded that states 10, 15, 16 are damaged states with 0.20 mm Gap.

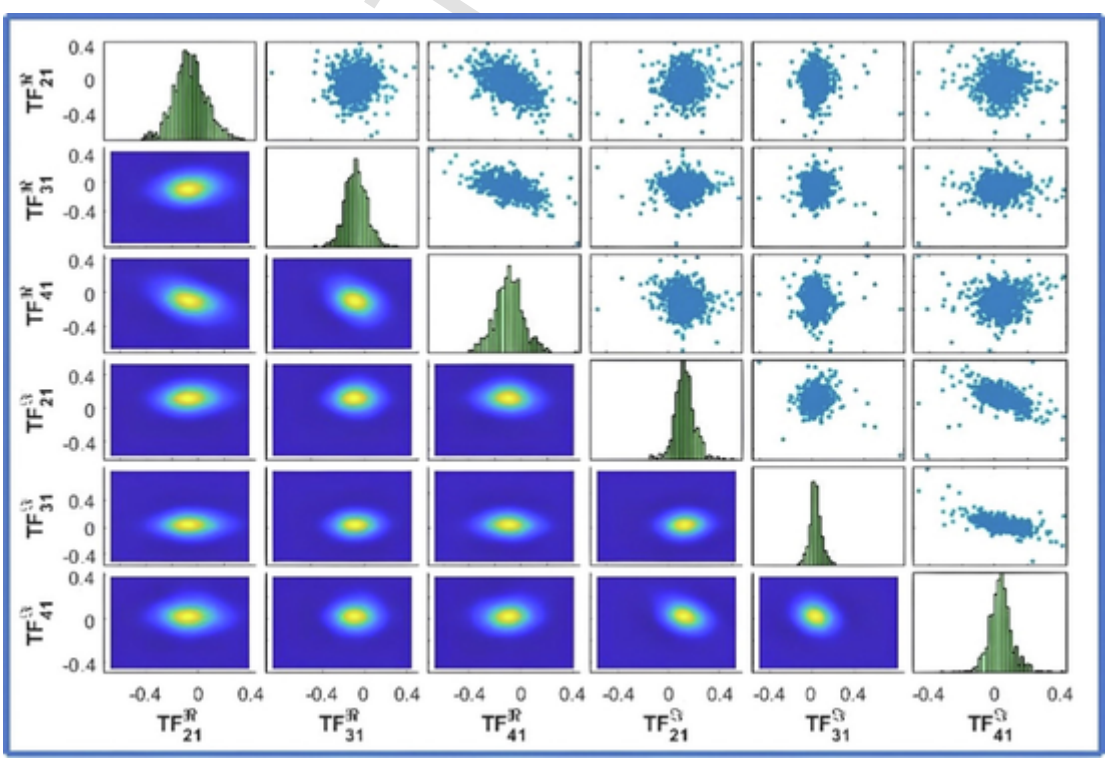


Fig. 6. Pairwise correlation plot of TFs at a certain frequency point.

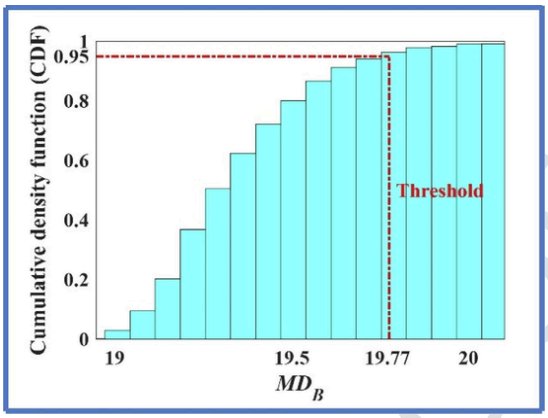


Fig. 7. Histogram of MD_B cootstrapped Monte Carlo simulation and the threshold value.

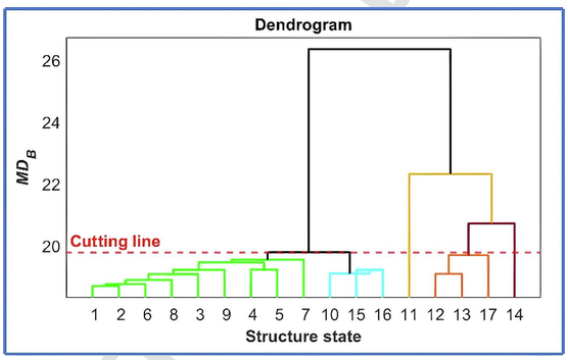


Fig. 8. Dendrogram and cutting line of the proposed agglomerative hierarchical clustering framework of the building structure.

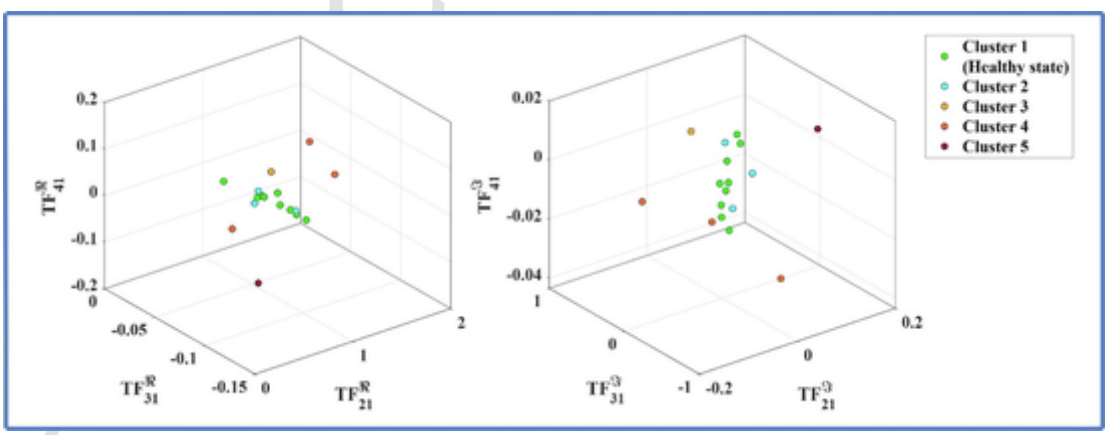


Fig. 9. Damage detection result of the building structure using the proposed method.

To investigate the sensitivity of the proposed method to measurement noise, the original dataset is artificially contaminated by normally distributed noise with different equivalent noise levels (1%-10%), following the method described in [48]. The performance of the proposed method under different noise levels is evaluated using the false positive rate (FPR), false negative rate (FNR), and accuracy (Acc), as presented in Table 2. It can be found that each damage state is successfully detected by the proposed method under a noise level of no more than 5%. When the noise level exceeds 6%, both the FPR and FNR increase with higher noise levels, resulting from misclassifying the states with the lowest damage extent (states 10, 15, 16) and states with the highest EOVs (states 5, 7, 9). Nevertheless, the proposed method maintains a high damage detection accuracy even when including large noise magnitudes in the response, which demonstrates its robustness against measurement noise.

Table 2
Damage detection results of the proposed method under different noise levels.

Noise Level (%)	FPR (%)	FNR (%)	Acc (%)
1	0	0	100
2	0	0	100
3	0	0	100
4	0	0	100
5	0	0	100
6	11.11	0	94.12
7	11.11	12.50	88.24
8	11.11	37.50	76.47
9	22.22	37.50	70.59
10	22.22	37.50	70.59

5.1.3. Effect of the function vectorization scheme

Apart from performance, the computational efficiency is also vitally significant for SHM system since methods requiring substantial computational resources would not be affordable in many applications. In this work, a function vectorization scheme is introduced to accelerate the data fusion process. To validate the effectiveness of the function vectorization scheme, the D_B values over the frequency band $\left[\omega_{k_1}, \omega_{k_2}\right]$ are incorporated to formulate $D_B^{[k_1,k_2]}$ as described in Eqs. (17)-(18). Then, the magnitude of $D_B^{[k_1,k_2]}$ is used as the similarity metric MD_B of the proposed method as described in Eq. (12), and the proposed method runs based on the vectorization scheme with matrix-wise operations. For comparison, the proposed method is repeated using element-wise operations based on the MD_B that integrates the D_B values over the same frequency band. The element-wise operations conduct $(k_2 - k_1)$ loops to calculate each D_B value in $D_B^{[k_1,k_2]}$ separately without vectorization. The proposed method runs multiple times based on MD_B that integrates D_B values over different frequency bands, both with and without vectorization, in order to evaluate the performance of the vectorization scheme across different frequency band widths. The computational costs for both scenarios are recorded and compared: With the vectorization scheme, the computational costs of the proposed method considering frequency bands containing 10, 100, 500, 1000, 2000, 3000, and 4000 frequency points are 1.22 s, 11.00 s, 55.77 s, 114.82 s, 247.81 s, 424.92 s, and 604.25 s, respectively. In contrast, without the vectorization scheme, the computational costs are 1.15 s, 11.60 s, 63.41 s, 137.99 s, 349.66 s, 697.24 s, and 1113.03 s, respectively. To present the effect of the function vectorization scheme more intuitively, the improvement percentages with respect to different frequency bands are calculated and shown in Fig. 10. Based on these results, it can be concluded that the effect of the vectorization scheme significantly improves with the increase of incorporated frequency points: When considering only 10 frequency points, the running time is slightly longer with the vectorization scheme than without vectorization (1.22 s with vectorization compared to 1.15 s without vectorization). However, when considering a broad frequency band containing 4000 frequency points, the running time decreases by approximately 500 s with the implementation of vectorization compared to without it (604.25 s with vectorization versus 1113.03 s without vectorization), which corresponds to a percentage increase in efficiency of over 45%. Therefore, the function vectorization scheme facilitates the application of the proposed damage detection method, especially when integrating information over a broad frequency band.

5.2. Field test: Z-24 bridge

5.2.1. Structure description

To further investigate the proposed method, another benchmark dataset obtained from a field test of Z24 bridge [49], which has become a test-bed for many damage detection methods, is adopted in this work for performance analysis. This dataset contains roughly one-year monitoring data of Z24 bridge under healthy conditions, whereas a progressive damage test was conducted in the

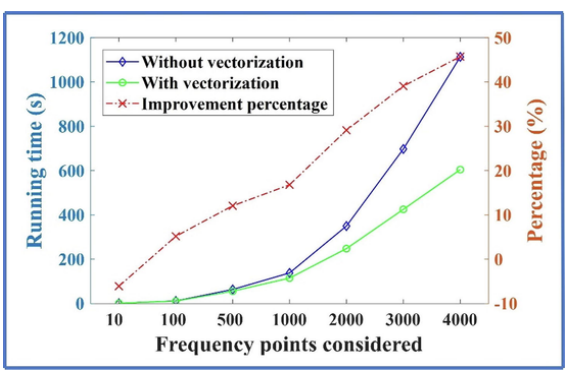


Fig. 10. Running times of the proposed method with respect to different widths of frequency bands.

last month of the monitoring period. The specific damage scenarios introduced in the damage test are summarized in Table 3. For more details about the monitoring process and the damage test, one can refer to [49–51]. The first five eigenfrequencies of the bridge deck are widely adopted as damage-sensitive features [52,53]. However, difficulty arises from distinguishing structural anomaly results from structural damage and that caused by changes in environmental conditions (primarily temperature), which would stiffen the deck asphalt and thus increase the natural frequencies. Another challenge is that certain damage scenarios, such as the spalling of concrete and landslides, have been reported to have negligible effect on bending stiffness and hence were not considered in the original paper [49], which illustrates the demand of more sensitive damage detection methods.

5.2.2. Damage detection result

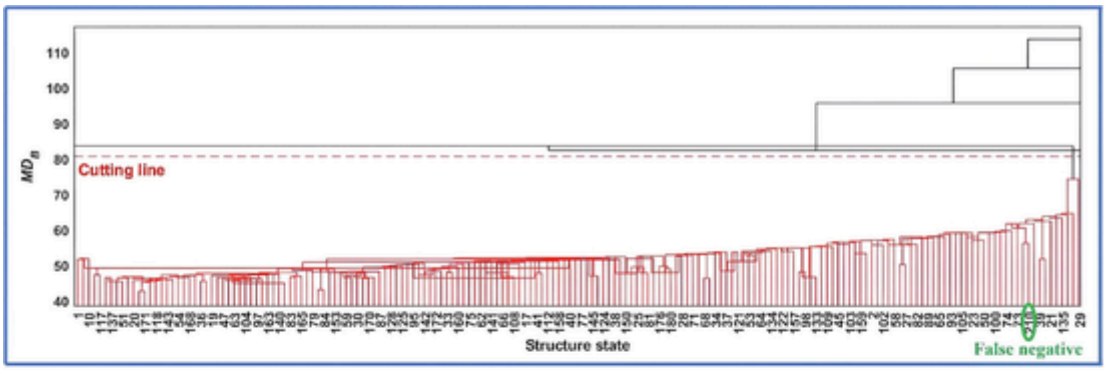
Since the Z24 bridge dataset contains sequential measurements collected every hour, each subset composed of one-day data is used to estimate the covariance matrices of clusters in leaf nodes, which results in 235 clusters in leaf nodes with the first 197 representing the normal condition, and the other 38 representing different damage scenarios. The data from the first 80 days (about 40% of normal condition data) are used as the training set. The dendrogram, as well as the cutting line, created via the proposed method is shown in Fig. 11. From these figures, one can find that nine clusters are created via the proposed method, including one cluster (red) containing all training data (Fig. 11a), two clusters (green and blue) containing the majority of damage data (Fig. 11b), and other six isolated clusters (Fig. 11c). From a damage detection perspective, one can conclude that the red cluster represents the normal condition, while the green and blue clusters denote two major damage scenarios, and each isolated cluster indicates an individual damaged state. Therefore, it can be found that there are 1 false negative indication and 18 false positive indications. All of the false positives come from the end of the monitoring period (Day 177 to Day 197), which could be caused by the long-term drift in the normal condition that leads to the alteration of the distribution of TF vectors [16]. All damage scenarios, including those with little effect on bending stiffness, are successfully detected, which demonstrates the sensitivity of the proposed method to structural damage. However, these introduced damage scenarios are not completely differentiated by the proposed method via distinct clusters, which could result from the lack of damage information in the training phase: As an unsupervised method, the proposed method only uses data from the normal condition to form the training set and estimate the threshold value. Then, the dendrogram is cut by a constant cutting line at the threshold value, which incorporates an implicit assumption that different damage scenarios can be differentiated using this threshold. However, as data from damaged states are not included in the training set, some factors that could affect the similarity between clusters and exist exclusively in certain damage scenarios but are absent in the training set, might not be properly accommodated by this threshold. As a result, data from these damage scenarios do not result in individual clusters. Another potential limitation of the proposed method is that the number of clusters is not determined in a fully automatic manner since it relies on an artificially introduced threshold value. This subjectivity is an inherent limitation of hierarchical clustering, which might limit the application of the proposed method. These problems mentioned in the above will be left as future endeavors.

5.2.3. Comparative study

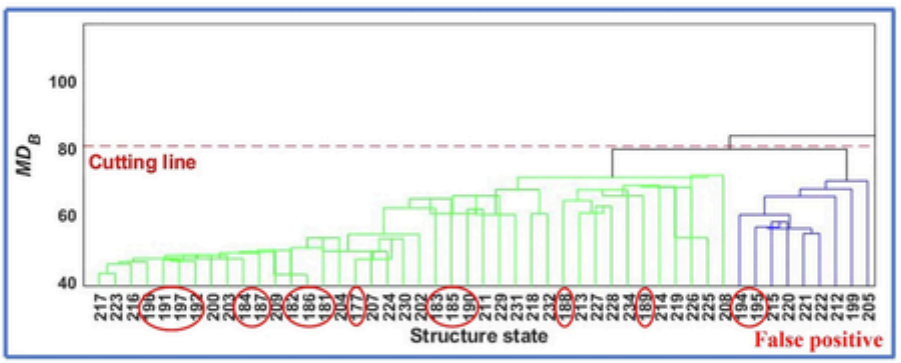
Despite the reliable and reasonable results of the proposed damage detection method, it is important to compare it with some state-of-the-art techniques. The GMM clustering is one of the most widely used probabilistic techniques for feature classification or damage detection [18,54]. Therefore, the GMM-based damage detection method introduced in [18] is included in the comparative study. In addition, three other well-known novelty detectors, namely the MSD-based method [7], the singular value decomposition (SVD)-based method [55], and the principal component analysis (PCA)-based method [18], are also included in the comparative study to illustrate the performance of the proposed method. For all the compared damage detection methods, the TF vectors are estimated using the measurements obtained each day. Then, the average of TF vectors over the selected frequency band is adopted as damage-sensitive features, which results in 235 samples totally for damage detection. The first 197 samples come from the normal

Table 3
Details of the introduced damage scenarios [49].

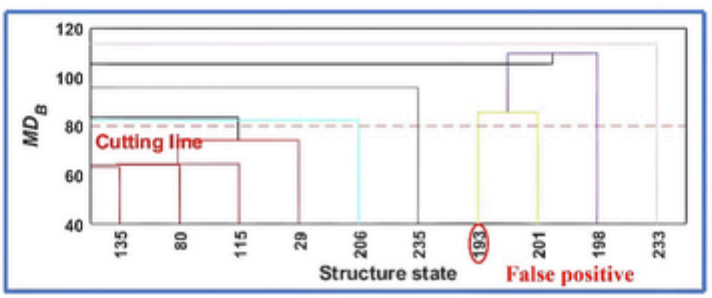
Date (1998)	Damage scenario
04–08	Reference condition I
09–08	Installation of pier settlement system
10–08	Lowering of pier, 20 mm
12–08	Lowering of pier, 40 mm
17–08	Lowering of pier, 80 mm
18–08	Lowering of pier, 95 mm
19–08	Lifting of pier, tilt of foundation
20–08	Reference condition II
25–08	Spalling of concrete at soffit, 12 m2
26–08	Spalling of concrete at soffit, 24 m2
27–08	Landslide of 1 m at abutment
31–08	Failure of concrete hinge
02–09	Failure of 2 anchor heads
03–09	Failure of 4 anchor heads
07–09	Rupture of 2 out of 16 tendons
08–09	Rupture of 4 out of 16 tendons
09–09	Rupture of 6 out of 16 tendons



(a)



(b)



(c)

Fig. 11. Dendrogram and cutting line of the Z24 bridge dataset: (a) the cluster denoting the healthy state; (b) two clusters containing the majority of damage data; (c) six isolated clusters.

condition, while the last 38 samples are from the damage test. The first 80 samples from the normal condition are used as the training set (same as the proposed method). For the GMM-based method, the number of components in GMM is determined using the Bayesian information criterion (BIC), while the threshold value for damage detection is defined at a significance level of 5%, as described in [18]. Fig. 12 presents the variation of BIC values with respect to different numbers of components, and the optimal number is determined to be 5. For the MSD-based method, the threshold is determined using the Monte Carlo method introduced in [7], while for the other two methods, the thresholds are defined by the 95% cut-off value over the training data as described in [55]. Besides, the number of principal components (PCs) in the PCA-based method is determined to make the cumulative percentage of the variance reach 80% as advised in [56]. The damage detection results of the four compared methods are shown in Fig. 13, from which one can conclude that false positive indications caused by long-term drift are also observed at the end of the monitoring period for these damage detection methods. All the compared methods exhibit lower sensitivity to damage than the proposed method, and thus result in more false negative indications. For the GMM-, MSD-, SVD-, and PCA-based methods, the number of false positive indications are 24, 12, 8, and 18, respectively, while the false negatives are 7, 13, 19, and 10, respectively.

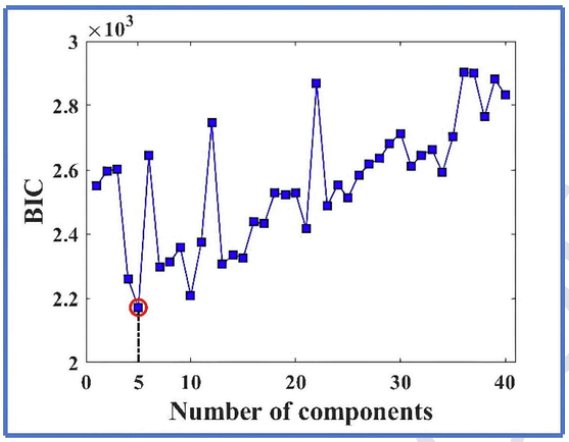


Fig. 12. BIC values with respect to different number of components.

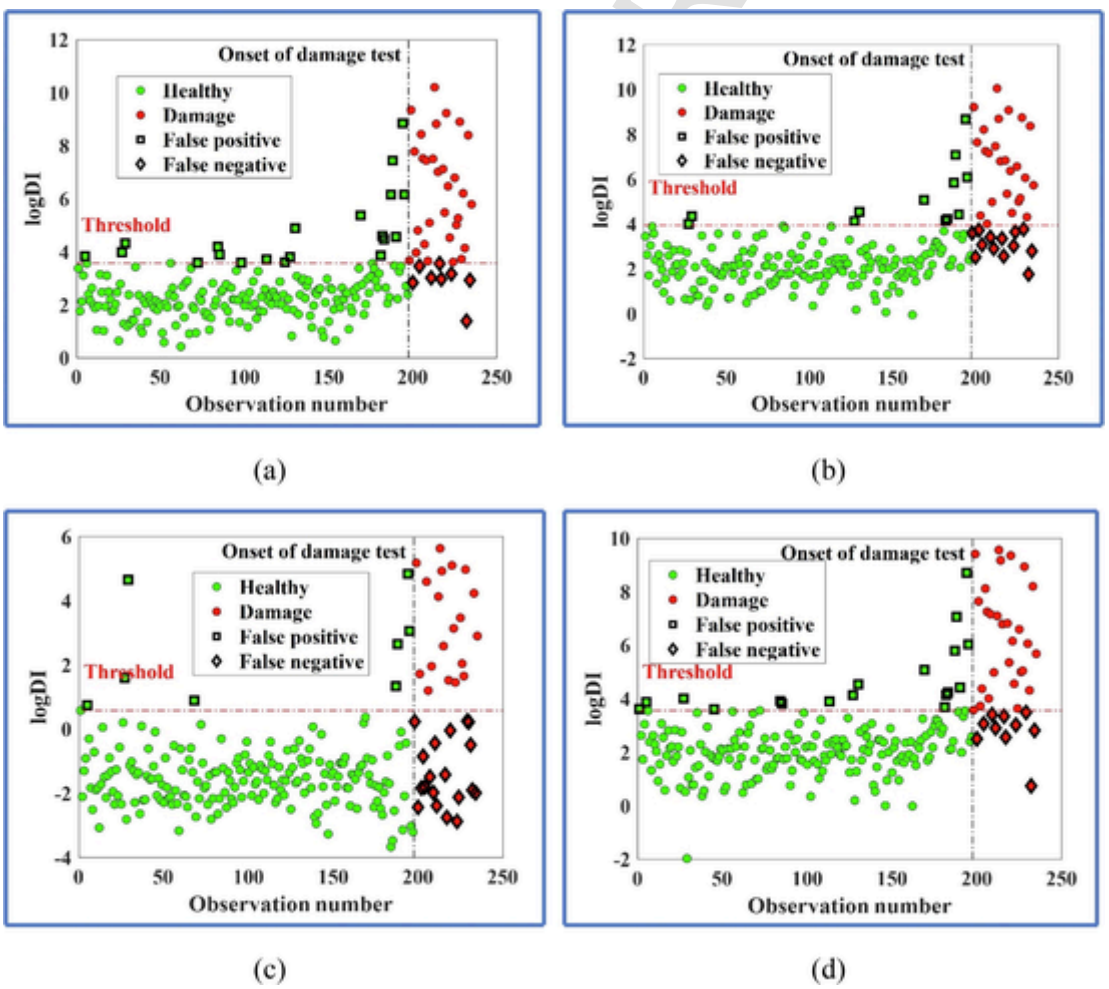


Fig. 13. Damage detection results of the compared methods (DI denotes the damage indicator): (a) the GMM-based method; (b) the MSD-based method; (c) the SVD-based method; (d) the PCA-based method.

To provide a more intuitive comparison of the performance of these methods, the true positive rate (TPR), FPR, true negative rate (TNR), FNR, and Acc corresponding to the damage detection results are calculated to construct the confusion matrices. Fig. 14 presents a schematic diagram of the confusion matrix containing the abovementioned parameters, while the confusion matrices corresponding to the damage detection results are shown in Fig. 15. Additionally, the balanced accuracy (BA) [57], which is defined as the average of TPR and TNR, is also computed for performance analysis. The BA values for the proposed method and the compared methods on the Z24 bridge dataset are as follows: 94.11% for the proposed method, 84.69% for the GMM-based method, 79.84% for the MSD-based method, 72.96% for the SVD-based method, and 82.27% for the PCA-based method. The rationale of using BA is that the Z24 bridge dataset is highly imbalanced (the amount of data from the normal condition is approximately five times that of data from damaged states) and thus damage detectors might take advantage of good prediction on the majority class (normal condition) to achieve a high conventional accuracy [57].

It can be observed in Fig. 15 that the proposed method exhibits an accuracy slightly higher than that of the four compared methods. However, the difference in BA is more significant than that in Acc, which demonstrates that these compared methods benefit from good prediction on the normal condition data. Therefore, BA is more appropriate to analyze the performance of damage detectors on this imbalanced dataset. The SVD-based method has the lowest FPR (4.06%) but the highest FNR (50%), which could result from the lack of sensitivity in high-dimensional spaces due to its reliance on Euclidean distance metric [16]. The MSD- and PCA-based methods exhibit better performance than the SVD-based method, but they also suffer from false negative indications, which could be caused by the implicit Gaussian assumption involved in MSD that fails to capture the statistical properties of TF vectors properly. The GMM-based method models the structural features using a mixture of Gaussian distributions to accommodate the uncertainties involved in damage detection in a probabilistic manner, and thus leads to a better damage detection result compared to others. However, it is worth noting that the GMM-based method is still influenced by the Gaussian assumption of TF vectors. Compared to these damage detection techniques, the proposed method adopts the complex Gaussian ratio distribution to capture the probabilistic properties of TF vectors in a more appropriate manner and employs the approximated D_B to quantify the similarity between distributions of TF vectors, which would be more robust against uncertainties than using deterministic distances. In addition, the proposed method incorporates bootstrapped Monte Carlo simulation to distinguish the modelling error and measurement noise from the changes in distributions of TF vectors, and thus lead to a more reasonable threshold. As a result, the proposed method exhibits better performance than these compared methods, especially in terms of false negatives.

6. Conclusion

This work introduces a methodology that incorporates TF vector and agglomerative hierarchical clustering for structural damage detection, aiming at utilizing the advantage of TF that is sensitive to damage and insensitive to excitation as well as the advantage of AHC that do not require prespecifying the number of clusters. Considering the limitation of conventional deterministic distance-based AHC that ignores the distribution of structural features, the multivariate Bhattacharyya distance is adopted to accounted for the uncertainty and correlation of multiple TFs based on the theoretical findings of multivariate complex-valued Gaussian ratio distribution of TF vectors. Based on Laplace's asymptotic expansion, an analytically tractable approximation of D_B is efficiently derived without assorting to high-dimensional numerical integration, and the D_B values over a selected frequency band are fused to formulate the similarity metric in the AHC framework. To accelerate the data fusion process, a function vectorization scheme is adopted to avoid the time-consuming loop operations. In addition, a parameter self-adaptive updating scheme is proposed as the linkage criterion, while a threshold is established via bootstrapped Monte Carlo simulation combined with an empirical quantile estimator to cut the dendrogram and obtain the damage detection result. Two case studies, including a laboratory experiment and a field test, are adopted to validate the performance of the proposed damage detection method, demonstrating that the proposed method exhibits better perfor-

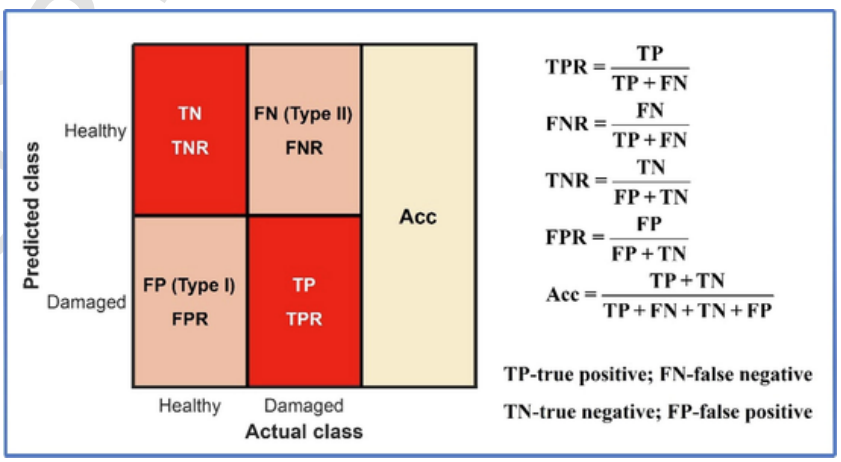
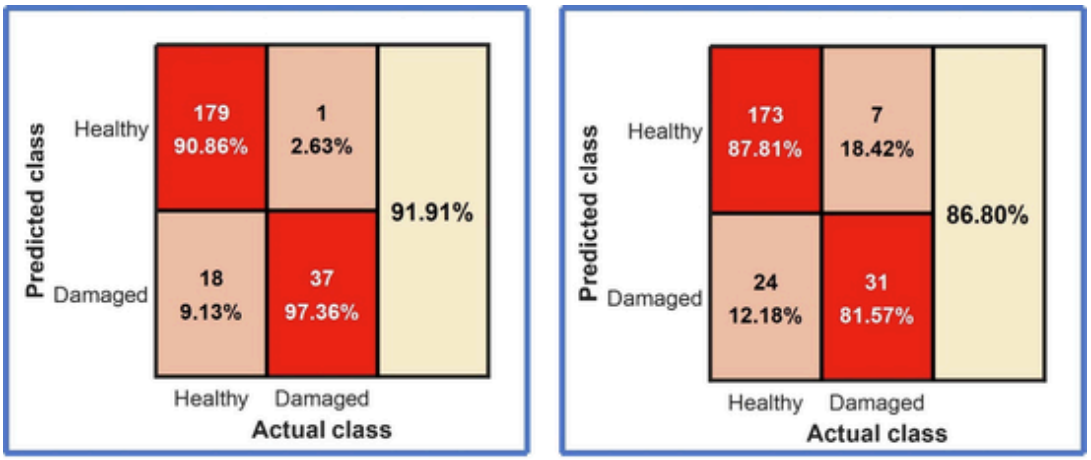


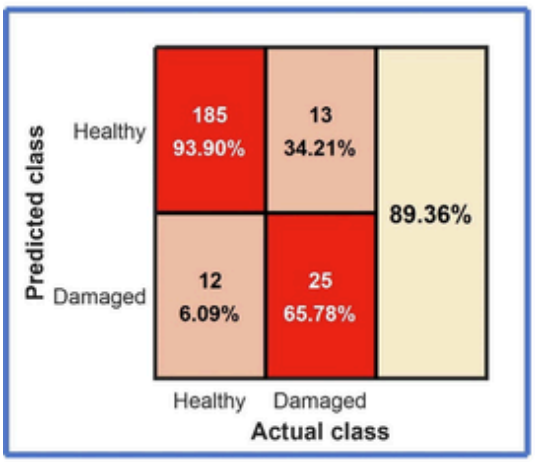
Fig. 14. Schematic diagram of the confusion matrix.


Predicted class
Healthy
Damaged
179
90.86%
1
2.63%
91.91%
18
9.13%
37
97.36%
Healthy
Damaged
Actual class

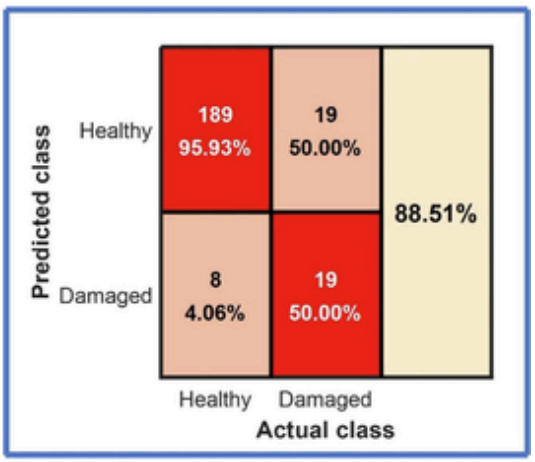
(a)

Predicted class
Healthy
Damaged
173
87.81%
7
18.42%
86.80%
24
12.18%
31
81.57%
Healthy
Damaged
Actual class

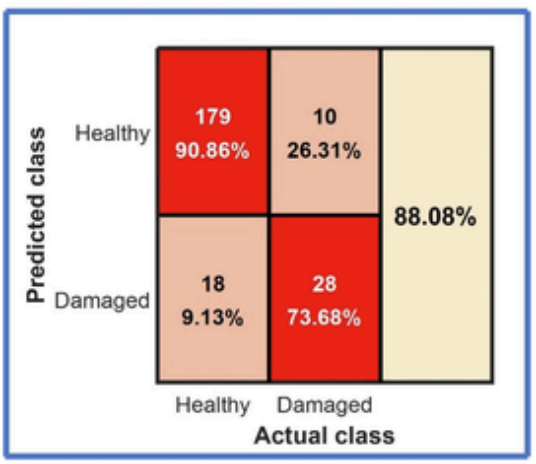
(b)


Predicted class
Healthy
Damaged
185
93.90%
13
34.21%
89.36%
12
6.09%
25
65.78%
Healthy
Damaged
Actual class

(c)


Predicted class
Healthy
Damaged
189
95.93%
19
50.00%
88.51%
8
4.06%
19
50.00%
Healthy
Damaged
Actual class

(d)


Predicted class
Healthy
Damaged
179
90.86%
10
26.31%
88.08%
18
9.13%
28
73.68%
Healthy
Damaged
Actual class

(e)

Fig. 15. Confusion matrices corresponding to the damage detection results of different methods: (a) the proposed method; (b) the GMM-based method; (c) the MSD-based method; (d) the SVD-based method; (e) the PCA-based method.

mance and robustness compared to damage detectors based on deterministic distance, which results from adopting probabilistic distance accommodating the uncertainty and correlation.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

This research has been supported by the Science and Technology Development Fund, Macau SAR (File no.: 017/2020/A1, 101/2021/A2, 0010/2021/AGJ, SKL-IOTSC(UM)-2021-2023), the Research Committee of University of Macau (File no.: MYRG2020-00073-IOTSC and MYRG2022-00096-IOTSC), Guangdong-Hong Kong-Macau Joint Laboratory Program (Project No.: 2020B1212030009) and Shenzhen Science and Technology Program (Grant Nos. JSGG20210802093207022, KQTD20180412181337494 and ZDSYS20201020162400001). Also, the authors are deeply appreciative to professors at the University of Leuven in Belgium for providing the experimental datasets of the benchmark structure.

Appendix A:. Global maximum $\boldsymbol{\eta}^*$

According to Eq. (9), the approximated Bhattacharyya distance is given by:

$$D_B\left(p_\gamma(\boldsymbol{\eta}), p_\gamma^d(\boldsymbol{\eta})\right) \approx -\ln\left((2\pi)^{\frac{n_o}{2}}\sqrt{p_\gamma(\boldsymbol{\eta}^*)p_\gamma^d(\boldsymbol{\eta}^*)}\left|H\left(\boldsymbol{\eta}^*\right)\right|^{-\frac{1}{2}}\right) \tag{A1}$$

where $\boldsymbol{\eta}^*$ is the global maximum of the integrand $\sqrt{p_\gamma(\boldsymbol{\eta})p_\gamma^d(\boldsymbol{\eta})}$. Assume $\boldsymbol{\tau} = \left[\left(\boldsymbol{u}^{\Re}\right)^T, \left(\boldsymbol{u}^{\Im}\right)^T\right]^T_{2(n_o-1)\times 1}$ and $L = p_\gamma(\boldsymbol{\eta})p_\gamma^d(\boldsymbol{\eta})$, then it has:

$$\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\tau}} = \begin{bmatrix} \frac{\partial\eta_1}{\partial\tau_1} & \frac{\partial\eta_1}{\partial\tau_2} & \cdots & \frac{\partial\eta_1}{\partial\tau_{2n_o-2}} \\ \frac{\partial\eta_2}{\partial\tau_1} & \frac{\partial\eta_2}{\partial\tau_2} & \cdots & \frac{\partial\eta_2}{\partial\tau_{2n_o-2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\eta_{2n_o}}{\partial\tau_1} & \frac{\partial\eta_{2n_o}}{\partial\tau_2} & \cdots & \frac{\partial\eta_{2n_o}}{\partial\tau_{2n_o-2}} \end{bmatrix}_{2n_o\times(2n_o-2)} = \begin{bmatrix} \mathbf{0}_{1\times(n_o-1)} & \mathbf{0}_{1\times(n_o-1)} \\ \mathbf{E}_{(n_o-1)\times(n_o-1)} & \mathbf{0}_{(n_o-1)\times(n_o-1)} \\ \mathbf{0}_{1\times(n_o-1)} & \mathbf{0}_{1\times(n_o-1)} \\ \mathbf{0}_{(n_o-1)\times(n_o-1)} & \mathbf{E}_{(n_o-1)\times(n_o-1)} \end{bmatrix} \tag{A2}$$

The global maximum $\boldsymbol{\tau}^*$ could be derived by solving $\frac{\partial L}{\partial\boldsymbol{\tau}} = 0$ as expressed in Eq. (A3). Accordingly, $\boldsymbol{\eta}^*$ could be obtained by back substitution:

$$\left[\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\tau}}\right]^T \frac{\partial\left[p_\gamma(\boldsymbol{\eta})p_\gamma^d(\boldsymbol{\eta})\right]}{\partial\boldsymbol{\eta}} = \mathbf{0} \tag{A3}$$

According to Eq. (3) and Eq. (5), Eq. (A3) could be expressed as:

$$\left[\frac{\partial\boldsymbol{\eta}}{\partial\boldsymbol{\tau}}\right]^T \frac{(n_o-1)!(n_o-1)!}{\pi^{2n_o-2}\left|\tilde{\Sigma}\right|^{\frac{1}{2}}\left|\tilde{\Sigma}_d\right|^{\frac{1}{2}}} \frac{\partial\left(\left[\eta^T\tilde{\Sigma}^{-1}\eta\right]^{-n_o}\left[\eta^T\tilde{\Sigma}_d^{-1}\eta\right]^{-n_o}\right)}{\partial\eta} = \mathbf{0} \tag{A4}$$

which is the same as:

$$\begin{bmatrix} \mathbf{0}_{1\times(n_o-1)} & \mathbf{E}_{(n_o-1)\times(n_o-1)} & \mathbf{0}_{1\times(n_o-1)} & \mathbf{0}_{(n_o-1)\times(n_o-1)} \\ \mathbf{0}_{1\times(n_o-1)} & \mathbf{0}_{(n_o-1)\times(n_o-1)} & \mathbf{0}_{1\times(n_o-1)} & \mathbf{E}_{(n_o-1)\times(n_o-1)} \end{bmatrix} \times \frac{\partial\left(\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}\right]^{-n_o}\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}\right]^{-n_o}\right)}{\partial\boldsymbol{\eta}} = \mathbf{0} \tag{A5}$$

Based on matrix calculus, it can be derived that:

$$\begin{aligned} &\frac{\partial\left(\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}\right]^{-n_o}\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}\right]^{-n_o}\right)}{\partial\boldsymbol{\eta}} = \\ &-n_o\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}\right]^{-n_o}\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}\right]^{-n_o}\left\{\frac{\left[\widetilde{\boldsymbol{\Sigma}}^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}^{-1}\right)^T\right]\boldsymbol{\eta}}{\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}\right]}+\frac{\left[\widetilde{\boldsymbol{\Sigma}}_d^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}_d^{-1}\right)^T\right]\boldsymbol{\eta}}{\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}\right]}\right\} \end{aligned} \tag{A6}$$

Therefore, Eq. (A5) could be rewritten as:

$$\begin{bmatrix} \mathbf{0}_{(n_o-1)\times1} & \mathbf{E}_{(n_o-1)\times(n_o-1)} & \mathbf{0}_{(n_o-1)\times1} & \mathbf{0}_{(n_o-1)\times(n_o-1)} \\ \mathbf{0}_{(n_o-1)\times1} & \mathbf{0}_{(n_o-1)\times(n_o-1)} & \mathbf{0}_{(n_o-1)\times1} & \mathbf{E}_{(n_o-1)\times(n_o-1)} \end{bmatrix} \left\{\frac{\left[\widetilde{\boldsymbol{\Sigma}}^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}^{-1}\right)^T\right]\boldsymbol{\eta}}{\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}\right]}+\frac{\left[\widetilde{\boldsymbol{\Sigma}}_d^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}_d^{-1}\right)^T\right]\boldsymbol{\eta}}{\left[\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}\right]}\right\} = \mathbf{0} \tag{A7}$$

Use matrix blocking method to divide $\left[\widetilde{\boldsymbol{\Sigma}}^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}^{-1}\right)^T\right]$ and $\left[\widetilde{\boldsymbol{\Sigma}}_d^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}_d^{-1}\right)^T\right]$ into submatrices:

$$\widetilde{\boldsymbol{\Sigma}}^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}^{-1}\right)^T=\begin{bmatrix} a_1 & \mathbf{a}_1 & a_2 & \mathbf{a}_2 \\ \mathbf{a}'_1 & \mathbf{A}_1 & \mathbf{a}'_2 & \mathbf{A}_2 \\ a_3 & \mathbf{a}_3 & a_4 & \mathbf{a}_4 \\ \mathbf{a}'_3 & \mathbf{A}_3 & \mathbf{a}'_4 & \mathbf{A}_4 \end{bmatrix}_{2n_o\times2n_o};\widetilde{\boldsymbol{\Sigma}}_d^{-1}+\left(\widetilde{\boldsymbol{\Sigma}}_d^{-1}\right)^T=\begin{bmatrix} b_1 & \mathbf{b}_1 & b_2 & \mathbf{b}_2 \\ \mathbf{b}'_1 & \mathbf{B}_1 & \mathbf{b}'_2 & \mathbf{B}_2 \\ b_3 & \mathbf{b}_3 & b_4 & \mathbf{b}_4 \\ \mathbf{b}'_3 & \mathbf{B}_3 & \mathbf{b}'_4 & \mathbf{B}_4 \end{bmatrix}_{2n_o\times2n_o} \tag{A8}$$

where a_n and b_n $(n=1,2,3,4)$ are scalars; $\mathbf{a}_n$ and $\mathbf{b}_n$ are column vectors with the size of $1\times(n_o-1)$; $\mathbf{a}'_n$ and $\mathbf{b}'_n$ are row vectors with the size of $(n_o-1)\times1$; $\mathbf{A}_n$ and $\mathbf{B}_n$ are matrices with the size of $(n_o-1)\times(n_o-1)$. Therefore, substituting Eq. (A8) and $\boldsymbol{\eta}=\left[1,\left(\boldsymbol{u}^{\Re}\right)^T,0,\left(\boldsymbol{u}^{\Im}\right)^T\right]^T$ into Eq. (A7), assume $p=\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\eta}$ and $q=\boldsymbol{\eta}^T\widetilde{\boldsymbol{\Sigma}}_d^{-1}\boldsymbol{\eta}$, then it has:

$$\frac{1}{p}\begin{bmatrix}\mathbf{a}'_1+\mathbf{A}_1\left(\mathbf{u}^{\Re}\right)^T+\mathbf{A}_2\left(\mathbf{u}^{\Im}\right)^T\\ \mathbf{a}'_3+\mathbf{A}_3\left(\mathbf{u}^{\Re}\right)^T+\mathbf{A}_4\left(\mathbf{u}^{\Im}\right)^T\end{bmatrix}+\frac{1}{q}\begin{bmatrix}\mathbf{b}'_1+\mathbf{B}_1\left(\mathbf{u}^{\Re}\right)^T+\mathbf{B}_2\left(\mathbf{u}^{\Im}\right)^T\\ \mathbf{b}'_3+\mathbf{B}_3\left(\mathbf{u}^{\Re}\right)^T+\mathbf{B}_4\left(\mathbf{u}^{\Im}\right)^T\end{bmatrix}=\begin{bmatrix}\mathbf{0}_{n_o-1}\\ \mathbf{0}_{n_o-1}\end{bmatrix} \tag{A9}$$

Noting that $\boldsymbol{\tau}=\left[\left(\boldsymbol{u}^{\Re}\right)^T,\left(\boldsymbol{u}^{\Im}\right)^T\right]^T$, hence Eq. (A9) could be rewritten as:

$$\left(\frac{1}{p}\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\\ \mathbf{A}_3 & \mathbf{A}_4\end{bmatrix}+\frac{1}{q}\begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_2\\ \mathbf{B}_3 & \mathbf{B}_4\end{bmatrix}\right)\boldsymbol{\tau}=-\frac{1}{p}\begin{bmatrix}\mathbf{a}'_1\\ \mathbf{a}'_3\end{bmatrix}-\frac{1}{q}\begin{bmatrix}\mathbf{b}'_1\\ \mathbf{b}'_3\end{bmatrix} \tag{A10}$$

Since p and q depend on $\boldsymbol{\tau}$, Eq. (A10) is a matrix equation in the form of $M(\boldsymbol{\tau})\cdot\boldsymbol{\tau}=N(\boldsymbol{\tau})$. The numerical solution of the equation, namely the global maximum $\boldsymbol{\tau}^*$, could be derived via an iterative algorithm, and thus $\boldsymbol{\eta}^*$ is obtained accordingly.

Appendix B:. Hessian matrix $H(\boldsymbol{\eta})$

$H(\boldsymbol{\eta})$ is the Hessian matrix of $f(\boldsymbol{\eta})=-\ln\left(\sqrt{p_\gamma(\boldsymbol{\eta})p_\gamma^d(\boldsymbol{\eta})}\right)$ which could be expressed as:

$$H(\boldsymbol{\eta})=\left[-\frac{\partial^2\ln\left(\sqrt{p_\gamma(\boldsymbol{\eta})p_\gamma^d(\boldsymbol{\eta})}\right)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\right]=\left[-\frac{1}{2}\frac{\partial^2\ln\left[p_\gamma(\boldsymbol{\eta})\right]}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}-\frac{1}{2}\frac{\partial^2\ln\left[p_\gamma^d(\boldsymbol{\eta})\right]}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\right] \tag{B1}$$

According to Eq. (3), for the healthy state, it has:

$$\frac{\partial^2 \ln\left[p_\gamma(\boldsymbol{\eta})\right]}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T} = \frac{\partial\left[\frac{1}{p_\gamma(\boldsymbol{\eta})}\frac{\partial p_\gamma(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right]}{\partial\boldsymbol{\eta}^T} = \left(\frac{\partial\left[\frac{1}{p_\gamma(\boldsymbol{\eta})}\frac{\partial p_\gamma(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right]^T}{\partial\boldsymbol{\eta}}\right)^T \tag{B2a}$$

$$\frac{1}{p_\gamma(\boldsymbol{\eta})}\frac{\partial p_\gamma(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = \frac{\pi^{n_o-1}\left|\tilde{\Sigma}\right|^{\frac{1}{2}}\left[\boldsymbol{\eta}^T\tilde{\Sigma}^{-1}\boldsymbol{\eta}\right]^{n_o}}{(n_o-1)!}\left\{-\frac{n_o!}{\pi^{n_o-1}\left|\tilde{\Sigma}\right|^{\frac{1}{2}}}\frac{\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]\boldsymbol{\eta}}{\left[\boldsymbol{\eta}^T\tilde{\Sigma}^{-1}\boldsymbol{\eta}\right]^{n_o+1}}\right\} = -\frac{n_o\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]\boldsymbol{\eta}}{\boldsymbol{\eta}^T\tilde{\Sigma}^{-1}\boldsymbol{\eta}} \tag{B2b}$$

$$\begin{aligned}\frac{\partial^2 \ln[p_\gamma(\eta)]}{\partial\eta\partial\eta^T} &= \left(\frac{\partial\left\{-n_o\left[\eta^T\tilde{\Sigma}^{-1}\eta\right]^{-1}\eta^T\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]\right\}}{\partial\eta}\right)^T \\ &= n_o\left\{\frac{\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]\eta\eta^T\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]}{\left[\eta^T\tilde{\Sigma}^{-1}\eta\right]^2} - \frac{\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T}{\eta^T\tilde{\Sigma}^{-1}\eta}\right\}\end{aligned} \tag{B2c}$$

Similarly, for the possibly damaged state, one has:

$$\frac{\partial^2 \ln\left[p_\gamma^d(\eta)\right]}{\partial\eta\partial\eta^T} = n_o\left\{\frac{\left[\tilde{\Sigma}_d^{-1}+\left(\tilde{\Sigma}_d^{-1}\right)^T\right]\eta\eta^T\left[\tilde{\Sigma}_d^{-1}+\left(\tilde{\Sigma}_d^{-1}\right)^T\right]}{\left[\eta^T\tilde{\Sigma}_d^{-1}\eta\right]^2} - \frac{\left[\tilde{\Sigma}_d^{-1}+\left(\tilde{\Sigma}_d^{-1}\right)^T\right]}{\eta^T\tilde{\Sigma}_d^{-1}\eta}\right\} \tag{B3}$$

Based on Eq. (B2) and Eq. (B3), Eq. (B1) could be rewritten as:

$$\begin{aligned}H(\eta) = &-\frac{n_o}{2}\left\{\frac{\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]\eta\eta^T\left[\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T\right]}{\left[\eta^T\tilde{\Sigma}^{-1}\eta\right]^2} - \frac{\tilde{\Sigma}^{-1}+\left(\tilde{\Sigma}^{-1}\right)^T}{\eta^T\tilde{\Sigma}^{-1}\eta}\right\} \\ &-\frac{n_o}{2}\left\{\frac{\left[\tilde{\Sigma}_d^{-1}+\left(\tilde{\Sigma}_d^{-1}\right)^T\right]\eta\eta^T\left[\tilde{\Sigma}_d^{-1}+\left(\tilde{\Sigma}_d^{-1}\right)^T\right]}{\left[\eta^T\tilde{\Sigma}_d^{-1}\eta\right]^2} - \frac{\left[\tilde{\Sigma}_d^{-1}+\left(\tilde{\Sigma}_d^{-1}\right)^T\right]}{\eta^T\tilde{\Sigma}_d^{-1}\eta}\right\}\end{aligned} \tag{B4}$$

Therefore, according to Eqs. (A1) to (A10) and Eqs. (B1) to (B4), the Laplacian approximation of D_B between distributions of TF vectors under the healthy state and a possibly damaged state can be derived analytically, which avoids high-dimensional numerical integration and thus significantly improves the efficiency and feasibility of the proposed damage detection algorithm.

References

[1] Farrar, C.R. and K. Worden, *Structural health monitoring: a machine learning perspective*. 2012: John Wiley & Sons.
[2] O. Avci, et al., A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications, Mechanical Systems and Signal Processing 147 (2021) 107077.
[3] Zhang, Y. and K.-V. Yuen, *Review of artificial intelligence-based bridge damage detection.* Advances in Mechanical Engineering, 2022. **14**(9): p. 16878132221122770.
[4] V.R. Gharehbaghi, et al., A Critical Review on Structural Health Monitoring: Definitions, Methods, and Perspectives, Archives of Computational Methods in Engineering (2021) 1–27.
[5] A. Malekloo, E. Ozer, M. AlHamaydeh, M. Girolami, Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights, Structural Health Monitoring 21 (4) (2022) 1906–1955.
[6] M. Flah, I. Nunez, W. Ben Chaabene, M.L. Nehdi, Machine learning algorithms in civil structural health monitoring: a systematic review, Archives of Computational Methods in Engineering 28 (4) (2021) 2621–2643.
[7] K. Worden, G. Manson, N.R. Fieller, Damage detection using outlier analysis, Journal of Sound and Vibration 229 (3) (2000) 647–667.
[8] K. Worden, A.J. Lane, Damage identification using support vector machines, Smart Materials and Structures 10 (3) (2001) 540–547.
[9] G. Park, A.C. Rutherford, H. Sohn, C.R. Farrar, An outlier analysis framework for impedance-based structural health monitoring, Journal of Sound and Vibration 286 (1–2) (2005) 229–250.
[10] M. Gul, F.N. Catbas, Structural health monitoring and damage assessment using a novel time series analysis methodology with sensor clustering, Journal of Sound and Vibration 330 (6) (2011) 1196–1210.
[11] W.-J. Yan, D. Chronopoulos, K.-V Yuen, Y.-C. Zhu, Structural anomaly detection based on probabilistic distance

measures of transmissibility function and statistical threshold selection scheme, Mechanical Systems and Signal Processing 162 (2022) 108009.
[12] B.T. Svendsen, G.T. Frøseth, O. Øiseth, A. Rønnquist, A data-based structural health monitoring approach for damage detection in steel bridges using experimental data, Journal of Civil Structural Health Monitoring 12 (1) (2022) 101–115.
[13] H. Sarmadi, A. Entezami, M. Salar, C. De Michele, Bridge health monitoring in environmental variability by new clustering and threshold estimation methods, Journal of Civil Structural Health Monitoring 11 (3) (2021) 629–644.
[14] H. Sarmadi, K.V. Yuen, Early damage detection by an innovative unsupervised learning method based on kernel null space and peak-over-threshold, Computer-Aided Civil and Infrastructure Engineering 36 (9) (2021) 1150–1167.
[15] H. Sarmadi, K.-V. Yuen, Structural health monitoring by a novel probabilistic machine learning method based on extreme value theory and mixture quantile modeling, Mechanical Systems and Signal Processing 173 (2022) 109049.
[16] T.J. Rogers, K. Worden, R. Fuentes, N. Dervilis, U.T. Tygesen, E.J. Cross, A Bayesian non-parametric clustering approach for semi-supervised structural health monitoring, Mechanical Systems and Signal Processing 119 (2019) 100–119.
[17] A. Santos, E. Figueiredo, J. Costa, Clustering studies for damage detection in bridges: A comparison study, Structural Health Monitoring 2015 (2015).
[18] E. Figueiredo, E. Cross, Linear approaches to modeling nonlinearities in long-term monitoring of bridges, Journal of Civil Structural Health Monitoring 3 (3) (2013) 187–194.
[19] M.M. Alamdari, T. Rakotoarivelo, N.L.D. Khoa, A spectral-based clustering for structural health monitoring of the Sydney Harbour Bridge, Mechanical Systems and Signal Processing 87 (2017) 384–400.
[20] X. Hu, L. Xu, Investigation on several model selection criteria for determining the number of cluster, Neural Information Processing-Letters and Reviews 4 (1) (2004) 1–10.
[21] Reddy, C.K. and B. Vinzamuri, *A Survey of Partitional and Hierarchical Clustering Algorithms*. Data Clustering: Algorithms and Applications, 2013. **87**.
[22] J.P. Santos, C. Crémona, A.D. Orcesi, P. Silveira, Multivariate statistical analysis for early damage detection, Engineering Structures 56 (2013) 273–285.
[23] Y.-L. Zhou, N.M.M. Maia, R.P.C. Sampaio, M.A. Wahab, Structural damage detection using transmissibility together with hierarchical clustering analysis and similarity measure, Structural Health Monitoring 16 (6) (2017) 711–731.
[24] J. Sokołowski, J. Obuchowski, A. Wyłomańska, P. Kruczek, R. Zimroz, Multiple local damage detection method based on time-frequency representation and agglomerative hierarchical clustering of temporary spectral content, Applied Acoustics 147 (2019) 44–55.
[25] W.-J. Yan, W.-X. Ren, Circularly-symmetric complex normal ratio distribution for scalar transmissibility functions. Part I: Fundamentals, Mechanical Systems and Signal Processing 80 (2016) 58–77.
[26] W.-J. Yan, W.-X. Ren, Circularly-symmetric complex normal ratio distribution for scalar transmissibility functions. Part II: Probabilistic model and validation, Mechanical Systems and Signal Processing 80 (2016) 78–98.
[27] L.-F. Mei, W.-J Yan, K.-V. Yuen, M. Beer, Structural Novelty Detection Based on Laplace Asymptotic Expansion of the Bhattacharyya Distance of Transmissibility Function and Bayesian Resampling Scheme, Journal of Sound and Vibration (2022) 117277.
[28] S. Chesné, A. Deraemaeker, Damage localization using transmissibility functions: a critical review, Mechanical Systems and Signal Processing 38 (2) (2013) 569–584.
[29] H. Zhang, M.J. Schulz, A. Naser, F. Ferguson, P.F. Pai, Structural health monitoring using transmittance functions, Mechanical Systems and Signal Processing 13 (5) (1999) 765–787.
[30] X.Y. Zhao, Z.-Q. Lang, G. Park, C.R. Farrar, M.D. Todd, Z. Mao, K. Worden, A new transmissibility analysis method for detection and location of damage via nonlinear features in MDOF structural systems, IEEE/ASME Transactions on Mechatronics 20 (4) (2015) 1933–1947.
[31] L. Feng, X. Yi, D. Zhu, X. Xie, Y. Wang, Damage detection of metro tunnel structure through transmissibility function and cross correlation analysis using local excitation and measurement, Mechanical Systems and Signal Processing 60–61 (2015) 59–74.
[32] N.M.M. Maia, R.A.B. Almeida, A.P.V. Urgueira, R.P.C. Sampaio, Damage detection and quantification using transmissibility, Mechanical Systems and Signal Processing 25 (7) (2011) 2475–2483.
[33] Y.-L. Zhou, N.M. Maia, M. Abdel Wahab, Damage detection using transmissibility compressed by principal component analysis enhanced with distance measure, Journal of Vibration and Control 24 (10) (2018) 2001–2019.
[34] Z. Mao, M. Todd, A model for quantifying uncertainty in the estimation of noise-contaminated measurements of transmissibility, Mechanical Systems and Signal Processing 28 (2012) 470–481.
[35] Z. Mao, M. Todd, Rapid structural condition assessment using transmissibility with quantified confidence for decision making, in: Topics in Model Validation and Uncertainty Quantification, Volume 4, Springer, 2012, pp. 133–140.
[36] S. Bi, M. Broggi, M. Beer, The role of the Bhattacharyya distance in stochastic model updating, Mechanical Systems and Signal Processing 117 (2019) 437–452.
[37] S. Bi, S. Prabhu, S. Cogan, S. Atamturktur, Uncertainty quantification metrics with varying statistical information in model calibration and validation, AIAA journal 55 (10) (2017) 3570–3583.
[38] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bulletin of the Calcutta Mathematical Society 35 (1943) 99–109.
[39] C. Papadimitriou, J.L. Beck, L.S. Katafygiotis, Asymptotic expansions for reliability and moments of uncertain systems, Journal of Engineering Mechanics 123 (12) (1997) 1219–1229.
[40] G. Manson, K. Worden, D. Allman, Experimental validation of a structural health monitoring methodology: Part II. Novelty detection on a Gnat aircraft, Journal of Sound and Vibration 259 (2) (2003) 345–363.
[41] K. Worden, C.R. Farrar, J. Haywood, M. Todd, A review of nonlinear dynamics applications to structural health monitoring, Structural Control and Health Monitoring 15 (4) (2008) 540–567.
[42] W.-J. Yan, et al., Vectorization and distributed parallelization of Bayesian model updating based on a multivariate complex-valued probabilistic model of frequency response functions, Mechanical Systems and Signal Processing 156 (2021) 107615.
[43] L.S. Katafygiotis, K.-V. Yuen, Bayesian spectral density approach for modal updating using ambient data, Earthquake Engineering & Structural Dynamics 30 (8) (2001) 1103–1123.
[44] K.-V. Yuen, L.S. Katafygiotis, Bayesian fast Fourier transform approach for modal updating using ambient data, Advances in Structural Engineering 6 (2) (2003) 81–95.
[45] Raychaudhuri, S. Introduction to monte carlo simulation. in 2008 Winter Simulation Conference. 2008. IEEE.
[46] Klemelä, J., *Nonparametric finance*. 2018: John Wiley & Sons.
[47] Figueiredo, E., et al., *Structural health monitoring algorithm comparisons using standard data sets*. 2009, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
[48] M.Z. Sarwar, D. Cantero, Deep autoencoder architecture for bridge damage assessment using responses from several vehicles, Engineering Structures 246 (2021) 113064.
[49] J. Maeck, G. De Roeck, Description of Z24 benchmark, Mechanical Systems and Signal Processing 17 (1) (2003) 127–131.
[50] B. Peeters, G. De Roeck, One-year monitoring of the Z24-Bridge: environmental effects versus damage events, Earthquake Engineering & Structural Dynamics 30 (2) (2001) 149–171.
[51] G.De Roeck, The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience, Journal of Structural Control 10 (2) (2003) 127–134.
[52] J. Maeck, G. De Roeck, Damage assessment using vibration analysis on the Z24-bridge, Mechanical Systems and Signal Processing 17 (1) (2003) 133–142.
[53] B. Peeters, J. Maeck, G.De Roeck, Vibration-based damage detection in civil engineering: excitation sources and temperature effects, Smart Materials and Structures 10 (3) (2001) 518–527.

[54] E. Figueiredo, L. Radu, K. Worden, C.R. Farrar, A Bayesian approach based on a Markov-chain Monte Carlo method for damage detection under unknown sources of variability, Engineering Structures 80 (2014) 1–10.
[55] E. Figueiredo, G. Park, C.R. Farrar, K. Worden, J. Figueiras, Machine learning algorithms for damage detection under operational and environmental variability, Structural Health Monitoring 10 (6) (2011) 559–572.
[56] A. Meixedo, et al., Online unsupervised detection of structural changes using train–induced dynamic responses, Mechanical Systems and Signal Processing 165 (2022) 108268.
[57] M. Bekkar, H.K. Djemaa, T.A. Alitouche, Evaluation measures for models assessment over imbalanced data sets, Journal of Information Engineering and Applications 3 (10) (2013).