ORIGINAL PAPER

Bridge health monitoring in environmental variability by new clustering and threshold estimation methods

Hassan Sarmadi[1,2] · Alireza Entezami[1,3] · Masoud Salar[3] · Carlo De Michele[3]

Abstract

Environmental variability is a major challenging issue in bridge health monitoring because bridges are more prone to such variability than other civil structures. To deal with this challenge, this article proposes a new machine-learning method for early damage detection under environmental variability by means of the k-medoids clustering, a new damage indicator, and an innovative approach for selecting a proper cluster number. Estimation of a reliable alarming threshold is another important challenge for early damage detection via most of the machine-learning methods. On this basis, a novel probabilistic approach using the theory of extreme value and a goodness-of-fit measure is proposed to estimate an alarming threshold. The major contributions of this article include proposing a new damage indicator suitable for decision making by clustering-based algorithms, an innovative cluster selection algorithm for dealing with the problem of environmental variability and increasing damage detectability, and a novel probabilistic method for threshold estimation. Modal-based features of the well-known Z24 Bridge are considered to verify the accuracy and effectiveness of the proposed approaches along with several comparative studies. Results show that the methods presented here are highly able to detect early damage even under strong environmental variations and estimate a reliable threshold.

Keywords Structural health monitoring · Environmental variability · Clustering · Statistical distance measure · Extreme value theory · Threshold estimation

1 Introduction

Bridges are prominent civil engineering structures due to their significant roles in social life, transportation networks, commerce, etc. Such structural systems are subjected to dead and live loads, natural disasters, environmental actions, and man-made hazards that may affect their performances and serviceability. Aging and material deterioration are other important issues that may threaten their safety and integrity. All the aforementioned cases may cause serious damage scenarios to these systems as cracks in concrete elements, losing bolts, weld failure and cracking in steel connections and elements, fatigue, failure, and even collapse. To prevent human and economic losses stemming from the occurrence of damage, structural health monitoring (SHM) is a great necessity for every society with any culture, social life, and economic conditions [1–3]. In addition, health monitoring of bridges is of more significance than the other kinds of civil structures because those are often lighter and slenderer leading to increases in their vibration levels under ambient excitations. Due to the importance of bridge health monitoring, the reader can find some valuable review articles in [4–8].

The process of SHM is often carried out in three main levels including early damage detection (Level 1), damage localization (Level 2), and damage quantification (Level 3). Early damage detection is an optimal monitoring procedure since small and frequent repairs are much less costly than major repairs, rehabilitation, retrofitting, and even rebuilding. This level is also important before taking any decision

✉ Hassan Sarmadi
hassan.sarmadi@mail.um.ac.ir; ipesfp.co@gmail.com

1 Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, P. Box: 9177948974, Mashhad, Iran

2 Head of Research and Development, Ideh Pardazan Etebar Sazeh Fanavar Pooya (IPESFP) Knowledge-Based Company, 29th Reza St., Reza Blvd, P. Box: 9176768544, Mashhad, Razavi Khorasan, Iran

3 Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milan, Italy

on locating and quantify damage, which are more difficult and complex than early damage detection. Data-based SHM is a relatively new and practical strategy for evaluating the safety and integrity of civil structures and detecting any potential damage [9]. This strategy is based on the statistical pattern recognition paradigm [10]. In the context of SHM, this paradigm includes sensing and data acquisition [11, 12], feature extraction [13], and decision making or feature classification by various machine-learning algorithms [9, 14–21]. The great advantage of data-based methods against model-based techniques is that those do not need any elaborate finite element modeling and model updating procedures. The feasibility of long-term SHM is the other benefit of data-based methods.

The basic premise of SHM is that the occurrence of damage changes the inherent structural parameters (i.e., often stiffness) as well as the vibration responses and characteristics. Modal frequencies are popular and widely used dynamic features for bridge health monitoring due to some great merits such as sensitivity to damage, simple identification via various techniques of operational modal analysis, and provision of global information for early damage detection. However, the main drawback of modal frequencies is their high sensitivity to environmental and/or operational variability conditions [22–24]. These conditions may arise from temperature fluctuations, humidity and moisture changes, wind speed and excitation amplitude variations, traffic, etc. [25]. Since the variations in structural responses caused by the environmental and/or operational variability are similar to damage, false alarm or *Type I* (i.e., the structure is undamaged but the method of damage detection mistakenly alerts the occurrence of damage) and false detection or *Type II* (i.e., the structure suffers from damage but the method of damage detection incorrectly declares the normal condition) are common errors in most of the modal-based SHM methods.

Machine learning algorithms present effective and tried-and-tested approaches to analyze features extracted from vibration data (e.g., modal frequencies) and making decisions about the current state of the structure, thereby finding its normal or damaged status [9]. These algorithms are usually divided into two main classes called supervised learning and unsupervised learning [18]. Both algorithms are intended to learn a statistical model (classifier or detector) by training data and make a decision via testing data. The main difference between supervised and unsupervised learning is that the former needs the information (features) of both undamaged and damaged (current) states to learn a model, while the latter requires the only information or features of the undamaged condition. In most cases of SHM applications, the current state of the structure is unknown. Under such circumstances, it is not practical and economical to impose intentional damage patterns on complex and

expensive civil structures in an effort to obtain information about the damaged condition. Therefore, one can conclude that unsupervised learning is more beneficial than supervised learning for health monitoring of civil structures.

Cluster analysis is a popular unsupervised learning method that aims at dividing similar objects into subsets or clusters. Regardless of the type of vibration data, damage-sensitive features, and structural systems, some well-known clustering techniques such as *k*-means [26, 27], *k*-medoids [28, 29], fuzzy clustering [30], and Gaussian mixture model [31, 32] have been utilized to detect damage. Although the utilization of cluster analysis is simple, the environmental and/or operational variability conditions seriously affect the performances of clustering algorithms (and the other unsupervised learning methods). Therefore, this problem is still a major challenge in SHM and it may become worse if the distance metric used in the algorithm of clustering has low damage detectability.

On the other hand, the decision making for early damage detection in most of the unsupervised learning methods requires an alarming threshold (i.e., a threshold limit) that enables them to alarm adverse changes in the structure caused by damage and correctly distinguish the damaged state from the undamaged condition. To put it another way, the estimate of a reliable threshold is critical because the final decision about the occurrence of damage depends strongly on it. In most cases, this limit is obtained by the probabilistic properties of the outputs of the model learned by the training data [16]. One of the powerful and effective ways for threshold estimation is based on the extreme value theory (EVT) [33, 34]. Under this theory, it is only necessary to select an extreme value distribution among Gumbel, Fréchet, and Weibull distribution models and use a technique for modeling that distribution. The threshold limit is estimated using the extreme quantile of the cumulative density function of the distribution under a significance level [35]. Nonetheless, this approach suffers from two main limitations. First, it is important to apply an analytical technique to select the most appropriate extreme value distribution among Gumbel, Fréchet, and Weibull models. Second, one needs to use an alternative technique so as to verify this choice [33].

To deal with these limitations, the best solution is to consider the generalized extreme value theory and utilize generalized extreme value (GEV) and generalized Pareto (GP) distribution models [33]. The great merit of these distributions is that each of them is a single distribution for modeling extreme quantities or rare events without any requirement of applying additional techniques for choosing and verifying the distribution model. In this regard, Block maxima (BM) and peak-over-threshold (POT) are two well-known approaches to modeling the GEV and GP distributions, respectively. Despite the applicability and effectiveness of

these techniques to the threshold estimation, choosing an optimal block number for the BM and determining a threshold value for the POT are their limitations. Any inappropriate choices of these parameters cause inaccurate alarming thresholds for damage detection along with increases in the false alarm and false detection errors.

Due to the importance of bridge health monitoring under varying environmental conditions, this article proposes a new machine-learning method in an unsupervised learning manner using the *k*-medoids clustering algorithm. The proposed clustering-based method aims to remove the deceptive effects of environmental variability and increase the detectability of damage. For these purposes, an $L_{p,r}$-distance measure is proposed to define a new damage indicator that can provide accurate results of damage detection with high damage detectability. In the proposed clustering-based method, the unfavorable effects of the environmental variability are removed by choosing an adequate cluster number among a wide range of sample clusters based on analyzing the variances regarding the outputs of the damage indicator obtained from the normal condition. For the first time, this article proposes a novel approach to model the GEV distribution and address the drawbacks of the BM and POT techniques for threshold estimation. The central idea behind the proposed approach is to utilize a goodness-of-fit (GOF) test based on Kullback–Leibler information for choosing adequate extreme values without selecting any block number or determining any threshold amount. Proposing a new clustering-based method using the *k*-medoids algorithm and an innovative cluster selection approach, a new damage indicator by the $L_{p,r}$-distance measure, a novel method for modeling the GEV distribution and threshold estimation are the main contributions of this article. The great advantages of these approaches include dealing with the problem of environmental variability, increasing damage detectability, determining a reliable threshold limit, and facilitating the process of threshold estimation by GOF without obtaining some requirements of the conventional techniques such as the number of blocks for BM and a threshold value for POT. The performance and accuracy of the proposed methods are validated by the well-known Z24 Bridge along with several comparative studies. Results demonstrate that the proposed clustering-based method in conjunction with the proposed $L_{p,r}$-distance measure and GOF highly succeeds in detecting damage, addressing the environmental variability, and providing high damage detectability.

2 Backgrounds

Cluster analysis is an unsupervised learning method based on dividing similar points with the small discrepancies or distances into a group or cluster. This concept provides an appropriate opportunity to utilize clustering techniques in SHM. Based on the general definition of the cluster analysis, one can exploit several clustering methods via the prototype-, density-, graph-based, and hybrid algorithms [36]. The major merit of prototype-based clustering approaches is that they are suitable for large and high-dimensional samples [37]. Therefore, the main focus of this article is to present a new application of one of the prototype-based clustering methods called the *k*-medoids clustering to the SHM problem.

2.1 The *k*-medoids clustering

The *k*-medoids clustering is a prototype-based partitioning method commonly used in domains that require robustness to outlier data, arbitrary distance metrics, and conditions that the mean and/or median do not have clear definitions [38]. This method is similar to the *k*-means clustering and the objective of both methods is to divide a set of observations or data points into *k* subsets (clusters) so that the subsets minimize the sum of distances between an observation and a center of the observation's cluster. In fact, both methods attempt to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster by a predefined objective function. For the *k*-means clustering, the prototype of interest is the centroid of data (the average between the points) in a cluster. This method employs the Euclidean distance as the usual distance metric and assigns a data point into the cluster that has the minimum distance from its centroid. In contrast, the *k*-medoids algorithm selects data points as centers (medoids) that can be chosen by arbitrary distances. Unlike the *k*-means clustering, the center of data in a cluster is a prototype in the *k*-medoids clustering. For this reason, this method is more resilient to noise and outliers in sampling data compared to the *k*-means clustering [38].

Assume that $\mathbf{X} = [\mathbf{x}_1,\ldots,\mathbf{x}_n] \in \mathfrak{R}^{q \times n}$ is a matrix of *n* observations and *q* variables. The algorithm of the *k*-medoids clustering divides the data set $\mathbf{X}$ into *k* clusters, which the number of clusters (*k*) is known as a priori. This method implements the clustering process iteratively by a predefined objective function until each representative observation is actually the medoid of the cluster. The objective function of interest is expressed as:

$$J(\mathbf{c}_1, \ldots, \mathbf{c}_k) = \min \sum d(\mathbf{x}_i, \mathbf{c}_j), \tag{1}$$

where $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,k$. Moreover, in Eq. (1), $\mathbf{c}_1 \ldots \mathbf{c}_k$ are the cluster centers (medoids) and *d* denotes a dissimilarity measure, which is no need for this measure to be symmetric or even metric. For the *k*-medoids clustering, the medoids or centers are obtained from an iterative algorithm when the objective function *J* reaches its minimum. One of

the algorithms is called partitioning around medoids (PAM), which proceeds in two steps including *build-step* and *swap-step*. The algorithm iterates the build- and swap-steps until the medoids do not change, or termination criteria are satisfied. The PAM algorithm minimizes the objective function by swapping all the non-medoid points and medoids iteratively until convergence.

2.2 Clustering in SHM

To detect damage by clustering techniques, it is necessary to define a strategy based on a clustering algorithm and a damage indicator. Generally, the process of clustering for SHM is carried out in the baseline (training) and monitoring (inspection) phases. During the baseline period, the damage-sensitive features (e.g., modal frequencies) of the undamaged states of the structure under different sources of environmental and/or operational variations are used to produce a training dataset. The main goal of the clustering algorithm used in the damage detection framework in the baseline phase is to determine the number of clusters. For the monitoring stage, the damage-sensitive features of the current state of the structure are applied to make a testing dataset. Since this state is unknown, which means that it can be normal or damaged, the use of testing data in the clustering algorithm indicates whether the structure is undamaged or damaged.

Here, it is supposed that $\mathbf{X} \in \mathfrak{R}^{q \times n}$ and $\mathbf{Z} \in \mathfrak{R}^{q \times m}$ are the training and testing matrices with the same variables and different observations. To employ the k-medoids clustering in the damage detection framework, the main requirements in the baseline phase are the number of clusters (k) and the cluster medoids ($\mathbf{c}_1,\ldots,\mathbf{c}_k$). In most cases, the damage indicator used in the framework of interest is defined as a distance measure so as to determine the dissimilarity between a feature vector and the cluster medoids. For a given q-dimensional feature vector of the testing data ($\mathbf{z}$), the damage indicator is the smallest distance between $\mathbf{z}$ and all cluster medoids. Using the Euclidean distance as a popular and widely used distance measure, the damage indicator DI^* is given by

$$DI^* = \min \left(\|\mathbf{z} - \mathbf{c}_1\|_2, \|\mathbf{z} - \mathbf{c}_2\|_2, ..., \|\mathbf{z} - \mathbf{c}_k\|_2 \right). \tag{2}$$

For each feature vector of the testing data, the calculation of DI continues to obtain a vector of the smallest distance values of DI^* regarding all observations. In the monitoring phase, this vector is designated as $\mathbf{d_m} = [DI_1^*, \ldots, DI_m^*]$, where m denotes the number of observations (the feature vectors) of the testing data. The decision about the occurrence of damage needs an alarming threshold. For this purpose, the feature vectors of the training data ($\mathbf{x}_1,\ldots,\mathbf{x}_n$) are used in Eq. (2) to define a vector of the smallest distance values of DI^* in the baseline

phase, which is designated as $\mathbf{d_b} = [DI_1^*,\ldots,DI_n^*]$. This vector is the output of the clustering-based method for estimating an alarming threshold. Any deviation of the values in $\mathbf{d_m}$ from the threshold is indicative of damage occurrence.

3 Proposed clustering-based method

The proposed clustering-based method is based on the k-medoids algorithm. The main objective is to obtain a set of medoids using an optimal cluster number that enables the proposed method to deal with the effects of environmental and/or operational variability. This method also presents a new damage indicator on the basis of an $L_{p,r}$-distance metric to increase damage detectability so that p and r are scalar values implying the powers of the $L_{p,r}$-distance.

3.1 $L_{p,r}$-distance measure

In most cases, the L_2-norm or Euclidean distance is a popular and widely used measure for defining an indicator for damage detection. The $L_{p,r}$-distance measure is a general form of the Euclidean distance. This is a kind of power distance measure that uses a formula mathematically equivalent to the power (p,r)-distance [39]. High damage detectability should be the main characteristic of an appropriate damage indicator. If the current state of the structure suffered from damage, the damage indicator with high detectability is able to effectively indicate this situation. Given the two arbitrary vectors $\mathbf{x}$ and $\mathbf{z}$, the $L_{p,r}$-distance measure is defined as follows [40]:

$$L_{p,r} = \left(\sum |\mathbf{x} - \mathbf{z}|^p \right)^{\frac{1}{r}}. \tag{3}$$

Depending on different values of p and r, it is feasible to express several distance metrics. For $p = r \geq 1$, the $L_{p,r}$-distance measure can be called the Euclidean, Manhattan, and Chebyshev metrics when the powers p and r are identically set as 2, 1, and ∞, respectively. In mathematics, the Chebyshev distance or maximum metric is a measure for calculating dissimilarity between two vectors, where the distance between them is the greatest of their differences along any coordinate dimension [41]. In the case of $p = 2$ and $r = 1$, the $L_{p,r}$-distance measure is equivalent to the squared Euclidean distance. For $0 < p = r < 1$, furthermore, this measure is called the fractional L_p-distance [40]. Based on the definition of the $L_{p,r}$-distance, one can define a new damage indicator using the feature vector of the testing data and the cluster medoids in the following form:

$$\mathrm{DI} = \min \left(L_{p,r}(\mathbf{z}, \mathbf{c}_1), L_{p,r}(\mathbf{z}, \mathbf{c}_2), ..., L_{p,r}(\mathbf{z}, \mathbf{c}_k) \right), \tag{4}$$

where

$$L_{p,r}(\mathbf{z}, \mathbf{c}_j) = \left(\sum \left| \mathbf{z} - \mathbf{c}_j \right|^p \right)^{\frac{1}{r}}, \tag{5}$$

where $j = 1, 2, \ldots, k$. Having considered all feature vectors of the testing matrix, it is possible to determine the vector $\mathbf{d_m} = [DI_1, \ldots, DI_m]$. The same procedure is performed to obtain the vector $\mathbf{d_b} = [DI_1, \ldots, DI_n]$ using the feature vectors of the training data. In other words, each of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is incorporated into Eq. (4) instead of $\mathbf{z}$. Finally, the vector $\mathbf{d_b}$ is applied to estimate an alarming threshold for damage detection.

3.2 Selection of an appropriate cluster number for SHM

The selection of a proper and optimal cluster number is an important subject in the prototype-based clustering algorithms such as the k-means, fuzzy c-means, and k-medoids. Since the final results of clustering depend on the number of clusters, it is essential to specify it in advance. Generally, the use of a few clusters may increase the errors in results, while relatively large clusters enable the clustering algorithm to decrease the errors. For SHM applications, the number of clusters is determined using the training data concerning the normal condition of the structure in the baseline phase. The common approach to choosing the cluster number for prototype-based clustering algorithms is to employ the Silhouette value technique [38]. However, it will be indicated that this technique is not resilient to SHM due to the presence of outliers, noise, or environmental and/or operational variability.

On this basis, it will be proved in this article that the poor performance of the k-medoids clustering in SHM due to high rates of Type I and Type II errors, as well as low damage detectability, arises from using an improper and relatively small cluster number. Since the effects of environmental and/or operational variations on the clustering results lead to false alarms and false detection errors, the choice of an appropriate cluster number with the emphasis on dealing with these effects highly enhances the performance of the clustering-based SHM method.

The central idea behind the proposed approach to selecting an appropriate cluster number is to find a number among a relatively wide range of sample clusters K_{max}, which is a large scalar value (e.g., 1000). The main criterion for this selection is based on evaluating the variances of different DI amounts of the vector $\mathbf{d_b}$ and choosing a cluster number with the smallest variance value. This approach aims to select a proper value of k to decrease or remove the variations in DI quantities resulting from the environmental variability. To obtain this amount, the k-medoids clustering is initially implemented by considering various cluster numbers. Under such circumstances, one can obtain different sets of $\mathbf{d_b}$ in

the baseline phase. Subsequently, the variances of all distance vectors are calculated to choose an appropriate cluster number, which possesses the smallest variance value. For the sake of simplicity, Fig. 1 depicts the flowchart of the proposed cluster selection approach.
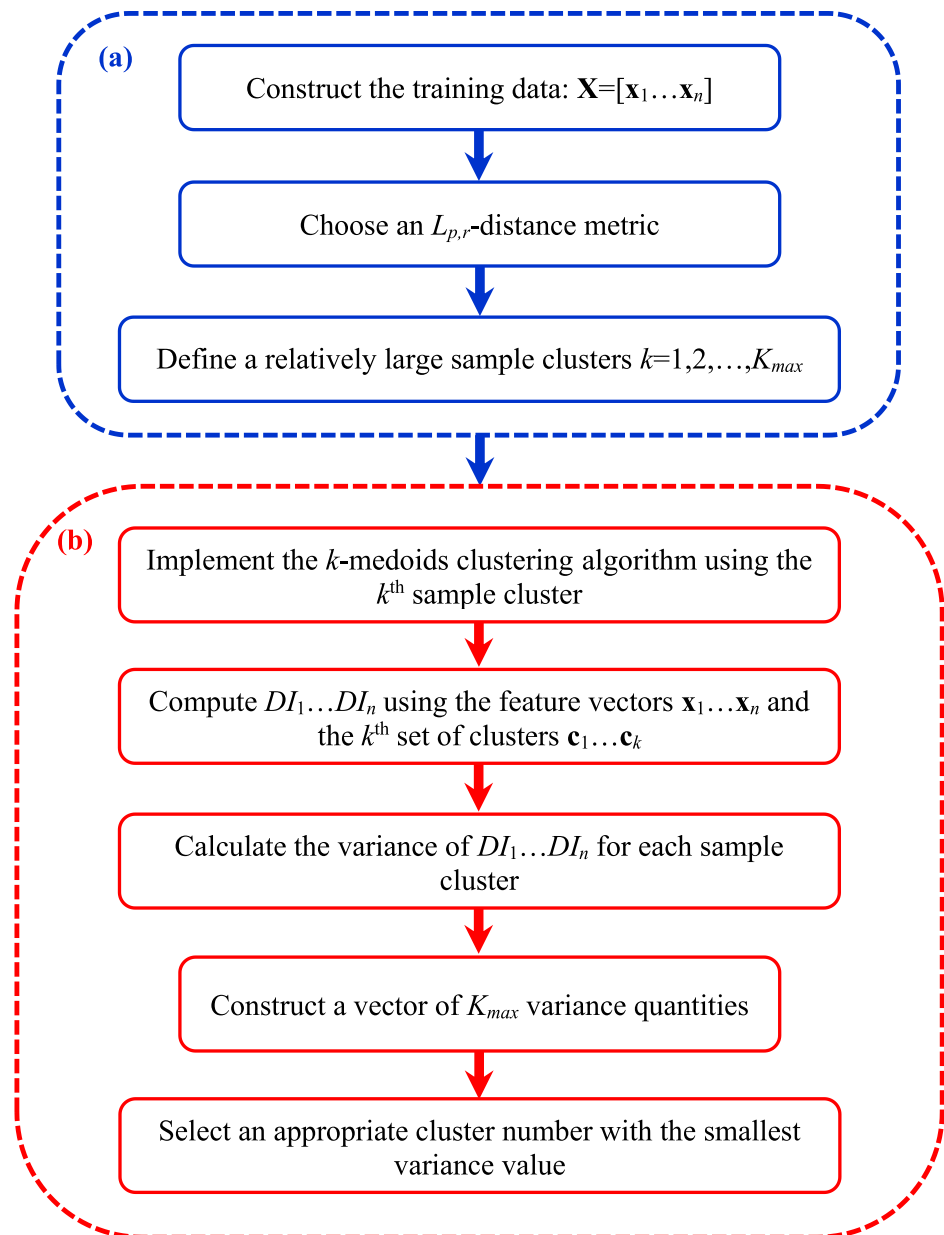
4 Proposed threshold estimation method

4.1 Extreme value theory

In statistics, the EVT is an approach to modeling the tails of a distribution by considering extreme quantities of sampling data or rare events [33]. The great advantage of this approach is to only focus on a few assumptions about the distribution of data rather than the modeling of whole data distribution. Furthermore, the EVT presents a robust method for the threshold limit determination [16]. Considering a large number of random data points, one can utilize three extreme value distributions including Gumbel, Fréchet, and Weibull [33, 34]. To determine a threshold value, it is necessary to select one of the above-mentioned distributions and then estimate the unknown parameters of the selected distribution. Finally, the alarming threshold is obtained from the extreme quantile of that distribution under a significance level. Due to the limitations of this approach, which have been explained earlier, it is possible to use the GEV or GP distribution models [33]. The main difference between these distributions originates from the methodology of modeling extreme values. More precisely, the extreme value modeling via the GEV distribution is based on the BM method [16], while the same modeling procedure by the GP distribution is carried out via the POT method [42]. The BM method relies upon dividing a set of data samples into non-overlapping blocks with equal size, extracting the maximum amount of each block, and fitting the GEV distribution model to a set of the maximum quantities extracted from all blocks. In contrast, the POT method is based on defining a threshold, choosing all extreme values (exceedances) above the threshold of interest, and fitting the GP distribution model to the exceedances. It should be clarified that the threshold used in the POT method and the threshold limit regarding the decision making and damage detection are two distinct subjects. For both the BM and POT techniques, the maximum likelihood estimation (MLE) is usually applied to estimate the unknown parameters of the GEV and GP distributions including the shape, scale, and location [33, 34].

4.2 Proposed GOF method

The strategy for modeling the tails or extremes of a distribution by the proposed GOF method differs from the conventional BM and POT techniques. The great advantage of this

(a)

Construct the training data: $\mathbf{X}=[\mathbf{x}_1\ldots\mathbf{x}_n]$

Choose an $L_{p,r}$-distance metric

Define a relatively large sample clusters $k=1,2,\ldots,K_{max}$

(b)

Implement the k-medoids clustering algorithm using the kth sample cluster

Compute $DI_1\ldots DI_n$ using the feature vectors $\mathbf{x}_1\ldots\mathbf{x}_n$ and the kth set of clusters $\mathbf{c}_1\ldots\mathbf{c}_k$

Calculate the variance of $DI_1\ldots DI_n$ for each sample cluster

Construct a vector of K_{max} variance quantities

Select an appropriate cluster number with the smallest variance value

method against the mentioned classical approaches is that it allows modeling an extreme value distribution without choosing any block number or determining any threshold level. It needs to mention that the GEV distribution model is considered to estimate an alarming threshold via the proposed GOF method. The fundamental principle of the proposed GOF method is to arrange data samples in descending order and find adequate maximum quantities from the first arranged samples via a GOF measure. In statistics, this measure is a statistical test for assessing the accuracy and adequacy of a fitted model. To put it another way, this test is intended to evaluate the acceptance or rejection of a theory or an idea. As a result, in the theory of interest, the test conforms to a null hypothesis ($\mathbb{H}_0$) in the case of acceptance,

and it is an alternative hypothesis ($\mathbb{H}_1$) in the case of rejection. Generally, a GOF test relies on a statistic (Q) that is a measure of the comparison between theoretical and empirical quantities. In most cases, the null hypothesis is rejected when the statistic Q is very large [43].

For the EVT, there are several GOF tests by considering their properties and ideas based on probability plots, an empirical distribution function, a log-likelihood function, Akaike or Bayesian information criteria, and Shapiro–Wilk's approach [43]. Recently, Pérez-Rodríguez et al. [44] proposed a new GOF test for the extreme value distribution based on the Kullback–Leibler information. Suppose that $y_1,\ldots,y_n$ are n random data samples, which are equivalent to $DI_1,\ldots,DI_n$ of the vector $\mathbf{d_b}$ regarding the normal condition.

The Kullback–Leibler information between the empirical and estimated probability distribution functions, called $G(y)$ and $\hat{G}(y)$, is given by

$$\mathrm{KL}(G,\hat{G}) = \int_{-\infty}^{+\infty} G(y) \ln \frac{G(y)}{\hat{G}(y)}\, \mathrm{d}y$$
$$= \int_{-\infty}^{+\infty} G(y) \ln G(y)\, \mathrm{d}y - \int_{-\infty}^{+\infty} G(y) \ln \hat{G}(y)\, \mathrm{d}y. \qquad (6)$$

To obtain KL $(G,\hat{G})$, the first term of the right-hand side of Eq. (6) is estimated by the Vasicek estimator in the following form:

$$\int_{-\infty}^{+\infty} G(y) \ln G(y)\, \mathrm{d}y = \frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{n}{2h} \left(\ln\left(y_{i+h}\right) - \ln\left(y_{i-h}\right) \right) \right), \qquad (7)$$

where $h < n/2$, $y_{i-h} = y_1$ and $y_{i+h} = y_n$ if $i - h < 1$ and $i + h > n$. Regarding the variable h, it should be clarified that it is possible to choose any positive integer smaller than $n/2$. In this regard, the smallest (the lower bound) and largest (the upper bound) choices of h correspond to 1 and $(\frac{n}{2} - 1)$. In this article, the upper bound of h is considered to calculate the first term of the statistic of the Kullback–Leibler information. Another important note about the variable h is that it should be a positive integer. For a non-integer, therefore, one should round it to the lowest value. On the other hand, the second term of the right-hand side of Eq. (6) is estimated by

$$\int_{-\infty}^{+\infty} G(y) \ln \hat{G}(y)\, \mathrm{d}y = \frac{1}{n} \sum_{i=1}^{n} \ln G\left(\ln\left(y_i\right), \mu_y, \sigma_y \right). \qquad (8)$$

where μ_y and σ_y are the mean and standard deviation of $y_1, \dots, y_n$. Eventually, the statistic of the Kullback–Leibler information (Q_{KL}) is rewritten using the data samples $y_1, \dots, y_n$, and the amounts of n and h as follows:

$$Q_{KL} = -\frac{1}{n} \sum_{i=1}^{n} \ln \left(\frac{n}{2h} \left(y_{i+h} - y_{i-h}\right) \right) - \frac{1}{n} \sum_{i=1}^{n} y_i + \frac{1}{n} \sum_{i=1}^{n} \exp\left(y_i\right). \qquad (9)$$

The null hypothesis $\mathbb{H}_0$ is rejected for large values of Q_{KL}. Using the concept of the Kullback–Leibler information, the proposed GOF method arranges the samples $y_1, \dots, y_n$ in descending order so that the arranged data begins with $y_{\max}$ and ends with $y_{\min}$. Subsequently, an iterative algorithm is developed to obtain the number of adequate maximum quantities (the extreme values) for modeling by the GEV distribution. To determine this number, the iterative algorithm in the proposed GOF method measures different values of Q_{KL} from sample maximum numbers ($i = 1, 2, \dots, S$, where i is the number of iterations); that is, $Q_{\mathrm{KL}}^1, Q_{\mathrm{KL}}^2, \dots, Q_{\mathrm{KL}}^S$. A number ($s$) with the smallest Q_{KL} quantity presents the adequate maximum or extreme samples designated as $\hat{y}_1, \dots, \hat{y}_s$, where $\hat{y}_1 = y_{\max}$ and $\hat{y}_s > \hat{y}_{s+1}$. Eventually, the process of threshold

estimation is carried out using the mentioned maximum samples and modeling them via the GEV distribution in the following form [33]:

$$F(\hat{y}) = \exp \left\{ -\left[1 + \beta \left(\frac{\hat{y} - \mu}{\sigma} \right) \right]^{-\frac{1}{\beta}} \right\}, \qquad (10)$$

where F is a non-degenerate distribution function. Furthermore, β, σ, and μ are the unknown parameters of the GEV distribution known as the shape, scale, and location, respectively. The threshold value is then determined by inverting Eq. (10) and estimating the extreme quantile of the GEV distribution. On this basis, the alarming threshold under a significance level (α) is expressed as:

$$\tau_\alpha = \begin{cases} \mu - \dfrac{\sigma}{\beta} \left[1 - \{ -\log(1-\alpha) \}^{-\beta} \right], & \beta \neq 0 \\ \mu - \sigma \log \{ -\log(1-\alpha) \}, & \beta = 0 \end{cases}. \qquad (11)$$

Using the threshold τ_α, it is expected that no DI values of the vector $\mathbf{d_b}$ violate from τ_α. On the other hand, any deviation of the DI quantity of the vector $\mathbf{d_m}$ from the threshold limit is indicative of damage occurrence. For simplicity, Fig. 2 presents the flowchart of the threshold estimation by the proposed GOF method.

5 Application to the Z24 Bridge

5.1 Bridge description

In this section, the accuracy and performance of the proposed methods are validated by the modal features of the well-known Z24 Bridge [45]. Figure 3a, b shows the general and close view of this bridge. It was a classical post-tensioned concrete box-girder bridge located in Switzerland linking the villages of Koppigen and Utzenstorf as an A1 highway overpass between Bern and Zurich. The dimensions of the Z24 Bridge were the main span of 30 m and two side-spans of 14 m as shown in Fig. 3c. To construct a new bridge with a larger side span, the Z24 Bridge was demolished at the end of 1998. Before the complete demolition, a long-term continuous monitoring test was performed to quantify the environmental variability components (e.g., temperature, wind characteristics, and humidity) and acquire vibration data (acceleration time histories). Eventually, realistic damage patterns were gradually applied to the bridge in a controlled way during the month before complete demolition.

The modal features used in the process of early damage detection are a set of natural frequencies of four modes, which were obtained from the technique of frequency domain decomposition. This set consists of 3932

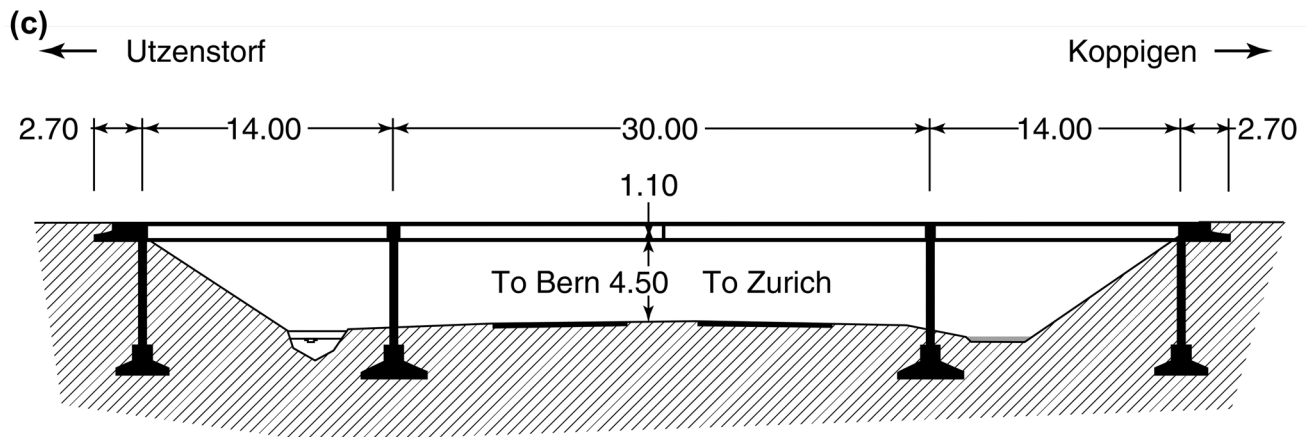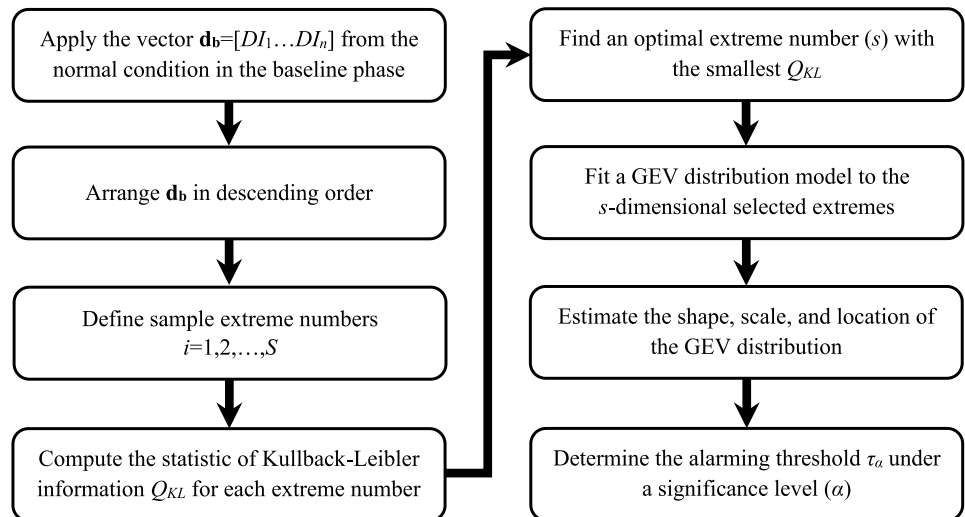**Fig. 2** The flowchart of the threshold estimation by the proposed GOF method

Apply the vector $\mathbf{d_b}=[DI_1 \ldots DI_n]$ from the normal condition in the baseline phase

↓

Arrange $\mathbf{d_b}$ in descending order

↓

Define sample extreme numbers $i=1,2,\ldots,S$

↓

Compute the statistic of Kullback-Leibler information Q_{KL} for each extreme number

→

Find an optimal extreme number (s) with the smallest Q_{KL}

↓

Fit a GEV distribution model to the s-dimensional selected extremes

↓

Estimate the shape, scale, and location of the GEV distribution

↓

Determine the alarming threshold τ_a under a significance level (α)

Fig. 3 **a** The general view of the Z24 Bridge, **b** the close view of the deck and one of the piers, **c** the longitudinal section

observations of the modal frequencies under varying actual environmental conditions. The first 3470 observations belong to the normal condition of the bridge and the last

462 observations are associated with the damaged state. Figure 4 illustrates the natural frequencies of the Z24 Bridge in four modes. As can be seen, the obvious jumps in the

modal frequencies of the normal condition are related to the changes in the asphalt layer in cold periods, which seriously affected the bridge stiffness and caused significant variations. In fact, these changes indicate high sensitivity of the measured natural frequencies to the environmental variability.

Before detecting damage by the proposed clustering-based method, it is necessary to define the training and testing sets. On this basis, the training data consists of 90% of the observations of the modal frequencies associated with the normal condition. In other words, it is a matrix with n observations and q variables, where $n = 3123$ and $q = 4$; that is, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_{3123}]$. On the other hand, the observations 3124–3470 (i.e., the remaining 10% of the modal frequencies of the normal condition), as well as the observations 3471–3932 regarding the damaged state, are applied to make the testing data, which is a matrix with 809 feature vectors ($m = 809$) and 4 variables; that is, $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_{809}]$.

5.2 Damage detection

The first step of the proposed clustering-based method is to choose an $L_{p,r}$-distance measure or specify the amounts of p and r. Based on the definition of the $L_{p,r}$-distance measure, four metrics as the various values of the p and r are utilized to investigate different distance measures for early damage detection. These metrics are the Chebyshev ($p = r = \infty$), Euclidean ($p = r = 2$), squared Euclidean ($p = 2$ and $r = 1$), and Manhattan ($p = r = 1$). However, the results of damage detection obtained from the Chebyshev metric are illustrated due to its better performance than the other metrics. As mentioned, the measured modal frequencies of the Z24 Bridge,

which are used as the main features for damage detection, are highly sensitive to the environmental variability. Based on the underlying idea of the proposed clustering-based method, one needs to deal with this problem by choosing an appropriate cluster number. According to the proposed cluster selection approach, 1500 sample clusters (K_{max}) are employed to incorporate into the algorithm of the k-medoids clustering and obtain 1500 variance amounts from 1500 sets of the DI values. Note that the calculated DI quantities are only related to the normal condition in the baseline phase. Figure 5a indicates the 1500 variances of the DI sets, where the smallest variance amount is found at the 992nd cluster. In other words, one can realize that the most proper number of clusters for addressing the influences of the environmental variations is identical to 992. Using this number, the k-medoids algorithm yields the cluster sets $\mathbf{c}_1, \dots, \mathbf{c}_{992}$, which are used to determine the DI values in the baseline and monitoring phases for obtaining the vectors $\mathbf{d_b}$ and $\mathbf{d_m}$.

Having considered the vector $\mathbf{d_b} = [DI_1, \dots, DI_{3123}]$, the proposed GOF method is applied to estimate an alarming threshold. Based on Fig. 2, the DI values of the vector $\mathbf{d_b}$ are arranged in descending order. The sample extreme number S and the variable h are set as 50 and 1560, respectively. Once again, it should be clarified here that the variable h is the upper bound of $n/2$; that is, $h = \frac{n}{2} - 1$. In the problem of the Z24 Bridge, the number of training samples (n) is equal to 3123, in which case the upper bound of h corresponds to 1560.5. As described previously, the variable h should be a positive integer. Since the calculated value of h is not an integer, one should round it to 1560. Using the statistic of the Kullback–Leibler information, subsequently, one can obtain a set of 50 amounts of Q_{KL} as can be observed in Fig. 5b.

Fig. 4 The modal frequencies of the Z24 Bridge: **a** mode 1, **b** mode 2, **c** mode 3, **d** mode 4
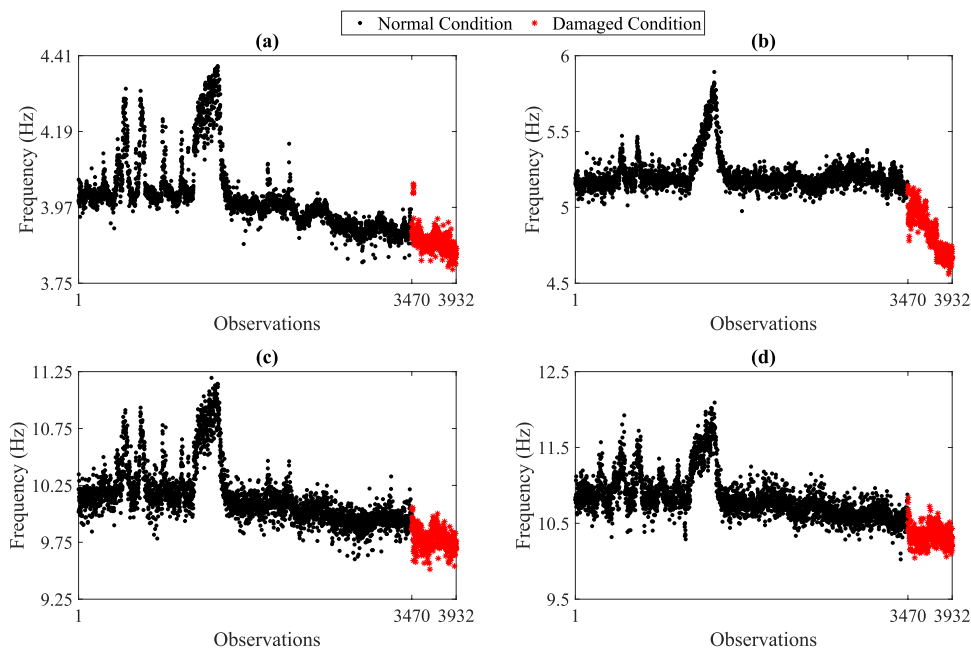
Fig. 5 **a** Determination of a proper cluster number for dealing with the environmental variability by the proposed cluster selection approach, **b** the values of Q_{KL} for choosing the most proper extreme samples
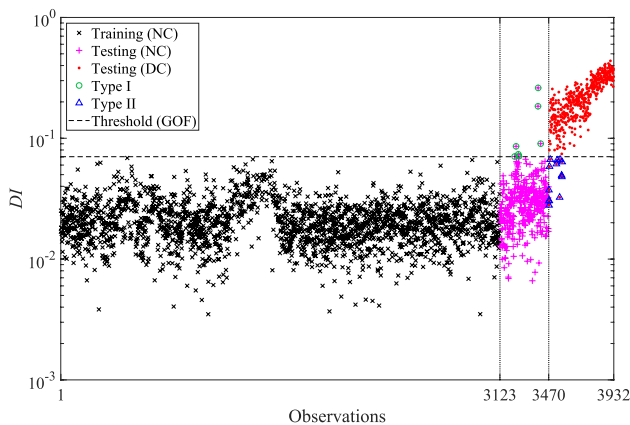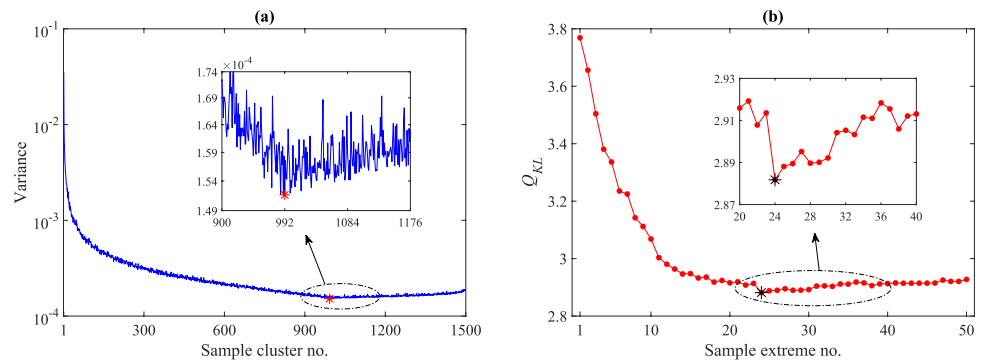

Fig. 6 Early damage detection by the proposed clustering-based method using the Chebyshev metric as the $L_{p,r}$-distance measure and GOF for the threshold estimation (*NC* Normal Condition, *DC* damaged condition)

Considering that the optimal extreme number (*s*) is one that provides the smallest Q_{KL} value, it is discerned in Fig. 5b that the proper value of *s* is equal to 24. Subsequently, the first 24 arranged *DI* values (i.e., $\hat{y}_1,...,\hat{y}_{24}$, where $\hat{y}_1 = DI_{max}$) are extracted to fit a GEV distribution to these extreme samples. The shape (β), scale (σ), and location (μ) of this distribution estimated by the MLE technique are equal to 0.2793, 0.0035, and 0.0542, respectively. Under the 5% significance level ($\alpha = 0.05$), the alarming threshold τ_α is computed by Eq. (11), which corresponds to 0.0704.

Figure 6 shows the result of early damage detection in the Z24 Bridge using the Chebyshev distance metric, where the horizontal dashed line is indicative of the threshold limit gained by the proposed GOF method. From Fig. 6, it is clear that no *DI* values of the observations 1–3123 (i.e., the training samples) exceed the threshold limit indicating the good performance of the proposed GOF method in estimating a proper threshold without any false alarm in the baseline phase. Moreover, most of the *DI* values of the observations 3124–3470, which are used in the monitoring stage and treat as the validation samples, fall below the threshold limit and

roughly have the same *DI* quantities as the training observations. In contrast, the majority of the *DI* values of the damaged state related to the observations 3471–3923 are over the threshold, which these outputs accurately imply the occurrence of damage. Therefore, one can conclude that the proposed clustering-based method in conjunction with the proposed GOF approach is successful in detecting damage even under the strong and nonlinear environmental variations. Regardless of the alarming threshold, it is perceived that there are clear differences between the *DI* values of the normal and damaged conditions. This conclusion also proves the high damaged detectability of the Chebyshev distance metric.

5.3 Comparisons

To demonstrate the superiority of the proposed clustering-based method, it is compared with the classical clustering-based technique as discussed in Sect. 2.2. Although both the classical and proposed methods utilize the *k*-medoids clustering algorithm, their differences pertain to the number of clusters used in this algorithm. For the classical technique, the number of clusters is determined by the conventional Silhouette value technique. In this regard, 30 sample clusters are considered to calculate their Silhouette values as shown in Fig. 7a. The most suitable cluster number is one that provides the largest Silhouette value. From this figure, it is apparent that the mentioned cluster number is identical to 2. Without considering any threshold limit, Fig. 7b indicates the result of early damage detection by the classical clustering-based technique using the Chebyshev metric as the $L_{p,r}$-distance measure. It can be observed that the *DI* amounts of many observations of the normal condition are larger than the maximum *DI* value of the damaged state. This result indicates the poor performance and low damage detectability of the classical clustering-based technique. Moreover, one can realize that although the proposed clustering-based method is able to address the effects of the environmental variability and increase the detectability of damage, the classical technique fails in providing reliable results. This

Fig. 7 a Determination of the number of clusters for the *k*-medoids clustering by the Silhouette value, **b** early damage detection by the classical clustering-based method using the Chebyshev metric as the $L_{p,r}$-distance measure (*NC* normal condition, *DC* damaged condition)
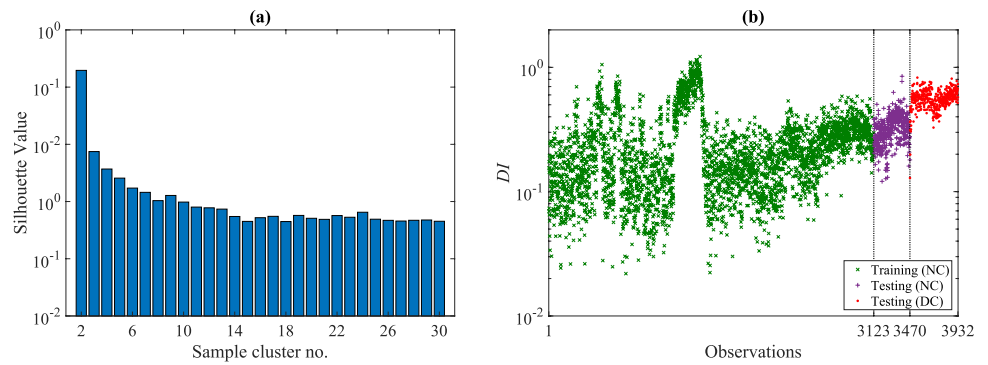

Table 1 Numbers and percentages of Type I, Type II, and total errors in detecting damage by different distance metrics using GOF for the threshold estimation

Distance metrics	Type I	Type II	Total
Chebyshev	7 (0.20%)	13 (2.81%)	20 (0.51%)
Euclidean	11 (0.31%)	19 (4.11%)	30 (0.76%)
Squared Euclidean	4 (0.11%)	24 (5.19%)	28 (0.71%)
Manhattan	6 (0.17%)	24 (5.19%)	30 (0.76%)

conclusion also confirms the accuracy of the idea behind the proposed approach to selecting an appropriate cluster number with the smallest variance.

In the following, the distance metrics Chebyshev, Euclidean, squared Euclidean, and Manhattan are compared to investigate their performances in detecting damage. This comparative study is based on evaluating the numbers and percentages of Type I, Type II, and total (misclassification) errors as presented in Table 1. As can be seen, the best performance in terms of the smallest misclassification rate is related to the Chebyshev distance metric. Conversely, both the Euclidean and the Manhattan metrics have the worst performances in the misclassification rate. Note that the Euclidean distance is a widely used metric in most of the clustering-based damage detection methods. Although the

squared Euclidean distance has the smallest Type I error, it suffers from a relatively large Type II error. On the other hand, the range of Type I errors in the distance metrics is approximately similar. Accordingly, it can be concluded that the Chebyshev distance outperforms the other metrics, particularly in the total error.

The other comparative study is concerned with the evaluation of the performances of the threshold estimation techniques. For this purpose, the misclassification rate (total error) in detecting damage based on the proposed clustering-based approach is applied to compare GOF, BM, and POT. Since the selections of an adequate block number for the BM method and an optimal threshold for choosing sufficient exceedances in the POT technique are critical issues, various block and exceedance numbers are utilized to compute different misclassification errors for these techniques. Figure 8 illustrates the rates of misclassification between GOF vs. BM and GOF vs. POT. In this comparison, the significance level is set at 0.05. An obvious indication in this figure is that the proposed GOF method possesses a smaller misclassification rate than the BM and POT techniques in all sample blocks and exceedances. This conclusion, thus, proves the superiority of the proposed GOF method over the mentioned conventional techniques.

In all the previous results, the process of early damage detection has been implemented using 90% of the modal

Fig. 8 Comparison of the threshold estimation methods in terms of the misclassification error **a** GOF vs. BM, **b** GOF vs. POT
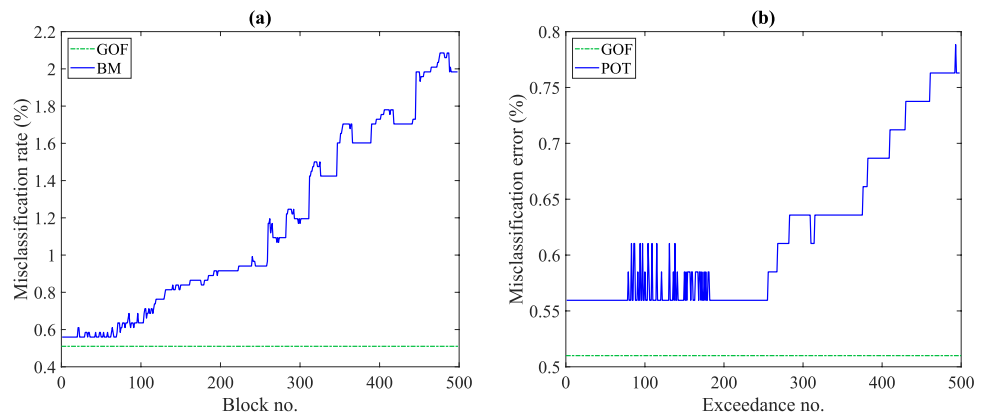
frequencies regarding the normal condition in an effort to make the feature vectors of the training data with 3123 observations. As the other comparison, the percentage for obtaining the training matrix is reduced to assess the effect of small training samples on the performance of the proposed clustering-based method. Accordingly, two different percentages including 75 and 60% are considered to define new training matrices with 2602 and 2082 observations. Under such circumstances, the remaining 25 and 40% of the modal frequencies of the normal condition and all modal frequencies of the damaged state are utilized to make two testing datasets. Having implemented all the steps of the proposed clustering-based method using the Chebyshev distance metric and GOF for the threshold estimation, Fig. 9 shows the results of damage detection under the reduced training samples. In addition, Table 2 lists the numbers and percentages of Type I, Type II, and total errors in detecting damage using the reduced training samples. As can be observed, all *DI* values of the training data are below the threshold limits. This confirms the reliability of the proposed GOF method in yielding an appropriate threshold limit even under reduced training samples. However, there are numerous false alarms in the validation samples (i.e., the remaining 25 and 40% of the observations of the modal frequencies associated with the normal condition). It can be seen that the rate of Type I error increases by decreasing the training samples.

From Table 2, it is obvious that the use of 60% training samples causes the worst performance in terms of Type I error. Although this percentage reduces Type II error, it suffers from a high misclassification rate. The same conclusion is observable when using 75% of the training samples. In fact, as the size (percentage) of training samples reduces, the misclassification rate and false alarm error increase as well. Despite the high damage detectability of the proposed clustering-based method, one can conclude that the use of adequate training samples is a significant issue and it is necessary to capture a wide range of environmental variations in the baseline phase.

Table 2 Numbers and percentages of Type I, Type II, and total errors in detecting damage by the proposed method using the Chebyshev metric and GOF under different training samples

Percentage of training samples	Type I	Type II	Total
90	7 (0.20%)	13 (2.81%)	20 (0.51%)
75	301 (8.67%)	3 (0.64%)	304 (7.73%)
60	748 (21.55%)	2 (0.4329)	750 (19.07%)

To evaluate the performance of the proposed clustering-based method without using all available data in the training phase, limited amounts of the modal frequencies regarding the normal condition are considered in two scenarios [16]. First, one supposes that there are smaller observations of the normal condition for the training procedure than the main problem. On this basis, it is assumed that the first 1735 observations of the modal frequencies are only available instead of utilizing all 3470 samples. Taking the 90% of 1735 observations of the normal condition, the new training matrix consists of 1561 feature vectors; that is, $\mathbf{X} \in \mathfrak{R}^{4 \times 1561}$. Moreover, the observations 1562–1735, which serve as the validation samples, and all modal frequencies of the damaged state (the same 462 observations) are used to make the new testing matrix $\mathbf{Z} \in \mathfrak{R}^{4 \times 636}$.

Second, the daily observations of the normal and damaged states are incorporated to define new small training and testing matrices. Accordingly, the number of observations of the modal frequencies decreases to 235, where the observations 1–198 belong to the normal condition and the observations 199–235 are associated with the damaged state [16]. Hence, the training matrix is obtained from 90% of the daily observations of the modal frequencies concerning the normal condition; that is, $\mathbf{X} \in \mathfrak{R}^{4 \times 178}$. On the other hand, the remaining 10% of the daily observations of the normal condition, which serve as the validation samples, along with all daily observations of the damaged state are gathered to generate the testing matrix $\mathbf{Z} \in \mathfrak{R}^{4 \times 57}$. The results of early damage detection by the proposed clustering-based method, the Chebyshev distance metric, and the GOF approach under

Fig. 9 Early damage detection by the proposed clustering-based method using the Chebyshev metric as the $L_{p,r}$-distance measure and GOF for the threshold estimation under reduced training samples: **a** 75%, **b** 60%
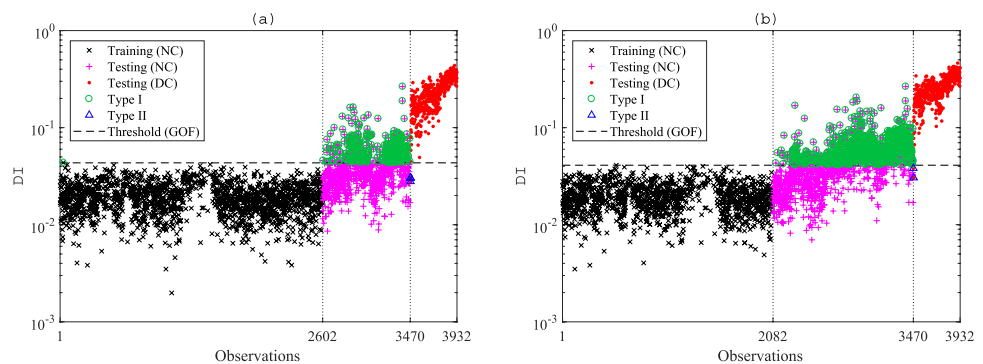
the above-mentioned scenarios are shown in Fig. 10. It needs to clarify that the optimal cluster numbers regarding the first and second scenarios are equal to 305 and 28, respectively.

As can be observed in Fig. 10a regarding the reduced version of the normal feature samples, most of the *DI* values in the observations 1–1735 fall below the threshold limit, except for only one point (Type I) among 174 validation samples. On the other hand, one can discern that the majority of the *DI* quantities of the damaged state exceed the threshold limit indicting accurate damage detection. However, the only three points (Type II) among 462 samples are under the threshold. The same conclusions with the different and small rates of Type I and Type II errors can be seen in Fig. 10b concerning the daily samples. Therefore, one can conclude that the proposed clustering-based method with the aids of the Chebyshev distance metric and the GOF approach is successful in accurately detecting damage using the small and daily feature samples.

To evaluate the performance of the Chebyshev distance and compare it with the other statistical measures, Table 3 lists the numbers and percentages of Type I, Type II, and total errors in the first and second scenarios. In the first scenario, one can realize that the Chebyshev distance still outperforms the other metrics in terms of all three errors. Moreover, the error rates of the Euclidean, Squared Euclidean, and Manhattan are close to each other. In the second scenario, the amounts in Table 3 reveal that all statistical divergence measures approximately have the same performances. The comparison between the error rates in Tables 1 and 3 demonstrates that the divergence measures roughly yield similar results when the number of feature samples is small (e.g., the second scenario regarding the daily samples). Nonetheless, the Chebyshev distance outperforms the other measures when there are relatively large samples.

All the previous results have been based on the *k*-medoids clustering. In the context of SHM by machine learning, there are other widely used techniques, which have broadly been applied to detect damage. On this basis, the final comparison is carried out by evaluating the performance of the proposed method by the well-known MSD [16, 19] and PCA [31] in terms of damage detectability without considering any alarming threshold. For this comparison, the training and testing data sets are based on the main problem (i.e., $\mathbf{X} \in \mathfrak{R}^{4 \times 3123}$ and $\mathbf{Z} \in \mathfrak{R}^{4 \times 809}$). Figure 11 illustrates the results of early damage detection by the above-mentioned conventional techniques, where DI_m and DI_p refer to the outputs of the MSD and PCA. Note that the number of principal components required for the PCA-based damage detection method is equal to 2. This number has been determined using the average eigenvalue criterion or Kaiser's criterion [19]. From Fig. 11a, b regarding the MSD and PCA methods, respectively, one can observe that the sudden jump in the distance values of the normal condition is still available. Moreover, some distance values of the normal condition are equal or larger than the corresponding values associated with the damaged state. These conclusions demonstrate the serious influence of environmental variability and poor performances of the conventional MSD and PCA methods for early damage detection.

Fig. 10 Early damage detection by the proposed clustering-based method and GOF: **a** the small samples of the normal condition, **b** the daily samples
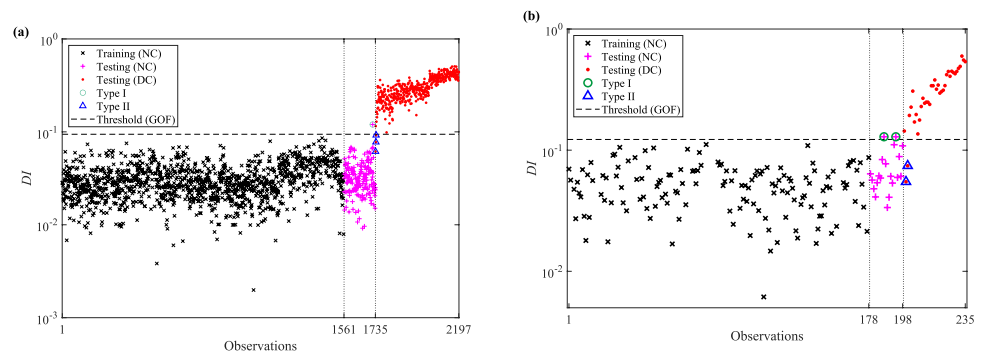
Table 3 Numbers and percentages of Type I, Type II, and total errors in detecting damage by different distance metrics and GOF using the small and daily feature samples

Distance metrics	Small samples			Daily samples		
	Type I	Type II	Total	Type I	Type II	Total
Chebyshev	1 (0.06%)	3 (0.64%)	4 (0.18%)	2 (1.01%)	2 (5.40%)	4 (1.70%)
Euclidean	4 (0.23%)	5 (1.08%)	9 (0.18%)	2 (1.01%)	3 (8.10%)	5 (2.12%)
Squared Euclidean	1 (0.06%)	8 (1.73%)	9 (0.18%)	2 (1.01%)	2 (5.40%)	4 (1.70%)
Manhattan	3 (0.23%)	7 (1.51%)	10 (2.16%)	3 (1.51%)	3 (8.10%)	6 (2.55%)

Fig. 11 Early damage detection without threshold values (*NC* normal condition, *DC* damaged condition): **a** MSD, **b** PCA
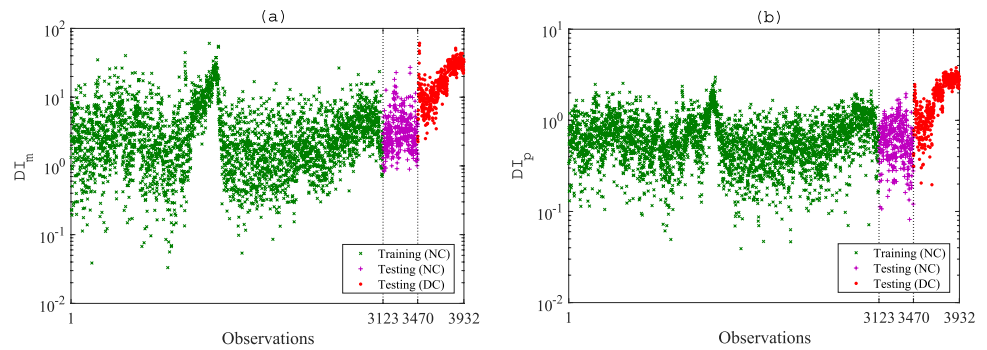
6 Conclusions

This article has proposed new clustering and threshold estimation methods for bridge health monitoring under environmental variability. The proposed clustering-based method has been developed from the *k*-medoids algorithm with a new approach to selecting an appropriate cluster number for dealing with the effects of environmental variations. To increase the detectability of damage, this article has presented the application of the $L_{p,r}$-distance metric to the algorithm of the *k*-medoids clustering. A novel threshold estimation method called GOF based on the EVT and GEV distribution modeling has also been proposed to define a reliable alarming threshold for early damage detection. The modal frequencies of the well-known Z24 Bridge have been utilized to verify the effectiveness and accuracy of the proposed methods along with several comparative studies.

The results have demonstrated that the proposed clustering-based method in conjunction with the proposed GOF approach highly succeeds in detecting damage under strong environmental variations. This conclusion is also valid for the scenarios of using the small and daily feature samples for early damage detection. The comparison among different $L_{p,r}$-distance measures has indicated that the Chebyshev metric outperforms the other distance measures, particularly the Euclidean and squared Euclidean distances, which are widely used in the clustering-based damage detection techniques. When the size of feature samples is small, it has been demonstrated that the statistical distances roughly have the same performances in terms of the rates of Type I, Type II, and total errors. It has been observed that the *k*-medoids clustering with a relatively large cluster number, which yields the smallest rate of variance, is able to deal with the negative effects of environmental variability conditions and increase the detectability of damage. The comparison between the proposed and classical clustering-based methods by considering the *k*-medoids algorithm and the Chebyshev distance metric has revealed that the former prevails over the latter. Furthermore, the proposed clustering-based method has been superior to the conventional MSD and PCA techniques in terms of damage detectability and SHM under strong environmental variability.

Furthermore, the comparisons between the proposed GOF method and the conventional BM and POT techniques have demonstrated that GOF not only facilitates the process of threshold estimation but also provides more reliable results owing to a smaller misclassification rate compared to the other techniques. Eventually, it has been seen that the use of small training samples considerably increases the false alarms and misclassification rates. This conveys that the proposed clustering-based method is sensitive to the number of samples needed for the training process. Therefore, it is preferable to apply a wide range of training samples that cover all possible environmental variability conditions in the baseline phase. For further research, it is recommended to develop an algorithm to determine the adequate number of training samples.

Compliance with ethical standards

References

1. Karbhari VM, Ansari F (2009) Structural health monitoring of civil infrastructure systems. Elsevier, London
2. Brownjohn JM, De Stefano A, Xu Y-L, Wenzel H, Aktan AE (2011) Vibration-based monitoring of civil infrastructure: challenges and successes. J Civ Struct Health Monit 1(3–4):79–95
3. Li H-N, Ren L, Jia Z-G, Yi T-H, Li D-S (2016) State-of-the-art in structural health monitoring of large and complex civil infrastructures. J Civ Struct Health Monit 6(1):3–16. https://doi.org/10.1007/s13349-015-0108-9
4. Li H, Ou J (2016) The state of the art in structural health monitoring of cable-stayed bridges. J Civ Struct Health Monit 6(1):43–67
5. An Y, Chatzi E, Sim S-H, Laflamme S, Blachowski B, Ou J (2019) Recent progress and future trends on damage identification

methods for bridge structures. Struct Contr Health Monit 26(10):e2416. https://doi.org/10.1002/stc.2416

6. Ye S, Lai X, Bartoli I, Aktan AE (2020) Technology for condition and performance evaluation of highway bridges. J Civ Struct Health Monit 10(4):573–594. https://doi.org/10.1007/s13349-020-00403-6

7. Xu Y, Brownjohn JMW (2018) Review of machine-vision based methodologies for displacement measurement in civil structures. J Civ Struct Health Monit 8(1):91–110. https://doi.org/10.1007/s13349-017-0261-4

8. Dorafshan S, Maguire M (2018) Bridge inspection: human performance, unmanned aerial systems and automation. J Civ Struct Health Monit 8(3):443–476. https://doi.org/10.1007/s13349-018-0285-4

9. Avci O, Abdeljaber O, Kiranyaz S, Hussein M, Gabbouj M, Inman DJ (2021) A review of vibration-based damage detection in civil structures: from traditional methods to Machine Learning and Deep Learning applications. Mech Syst Sig Process 147:107077. https://doi.org/10.1016/j.ymssp.2020.107077

10. Entezami A, Sarmadi H, Behkamal B, Mariani S (2020) Big data analytics and structural health monitoring: a statistical pattern recognition-based approach. Sensors 20(8):2328. https://doi.org/10.3390/s20082328

11. Wang ML, Lynch JP, Sohn H (2014) Sensor technologies for civil infrastructures: applications in structural health monitoring. Woodhead Publishing (Elsevier), Cambridge

12. Wu ZY, Zhou K, Shenton HW, Chajes MJ (2019) Development of sensor placement optimization tool and application to large-span cable-stayed bridge. J Civ Struct Health Monit 9(1):77–90. https://doi.org/10.1007/s13349-018-0320-5

13. Amezquita-Sanchez JP, Adeli H (2016) Signal processing techniques for vibration-based health monitoring of smart structures. Arch Comput Methods Eng 23(1):1–15. https://doi.org/10.1007/s11831-014-9135-7

14. Flah M, Nunez I, Ben Chaabene W, Nehdi ML (2020) Machine learning algorithms in civil structural health monitoring: a systematic review. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-020-09471-9

15. Azimi M, Eslamlou AD, Pekcan G (2020) Data-driven structural health monitoring and damage detection through deep learning: state-of-the-art review. Sensors 20(10):2778

16. Sarmadi H, Karamodin A (2020) A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects. Mech Syst Sig Process 140:106495. https://doi.org/10.1016/j.ymssp.2019.106495

17. Nazarian E, Taylor T, Weifeng T, Ansari F (2018) Machine-learning-based approach for post event assessment of damage in a turn-of-the-century building structure. J Civ Struct Health Monit 8(2):237–251. https://doi.org/10.1007/s13349-018-0275-6

18. Sarmadi H, Entezami A (2021) Application of supervised learning to validation of damage detection. Arch Appl Mech 91(1):393–410. https://doi.org/10.1007/s00419-020-01779-z

19. Sarmadi H, Entezami A, Saeedi Razavi B, Yuen K-V (2021) Ensemble learning-based structural health monitoring by Mahalanobis distance metrics. Struct Contr Health Monit 28(2):e2663. https://doi.org/10.1002/stc.2663

20. Daneshvar MH, Gharighoran A, Zareei SA, Karamodin A (2020) Early damage detection under massive data via innovative hybrid methods: application to a large-scale cable-stayed bridge. Struct Infrastruct Eng. https://doi.org/10.1080/15732479.2020.1777572

21. Entezami A, Sarmadi H, Mariani S (2020) An unsupervised learning approach for early damage detection by time series analysis and deep neural network to deal with output-only (big) data. Eng Proc 2(1):17. https://doi.org/10.3390/ecsa-7-08281

22. Xia Y, Chen B, Weng S, Ni Y-Q, Xu Y-L (2012) Temperature effect on vibration properties of civil structures: a literature review and case studies. J Civ Struct Health Monit 2(1):29–46. https://doi.org/10.1007/s13349-011-0015-7

23. Han Q, Ma Q, Xu J, Liu M (2020) Structural health monitoring research under varying temperature condition: a review. J Civ Struct Health Monit. https://doi.org/10.1007/s13349-020-00444-x

24. Mu H-Q, Yuen K-V (2018) Modal frequency-environmental condition relation development using long-term structural health monitoring measurement: uncertainty quantification, sparse feature selection and multivariate prediction. Meas 130:384–397

25. Sohn H (2007) Effects of environmental and operational variability on structural health monitoring. Philos Trans Royal Soc A 365(1851):539–560

26. Santos JP, Crémona C, Calado L, Silveira P, Orcesi AD (2016) On-line unsupervised detection of early damage. Struct Contr Health Monit 23(7):1047–1069

27. Diez A, Khoa NLD, Alamdari MM, Wang Y, Chen F, Runcie P (2016) A clustering approach for structural health monitoring on bridges. J Civ Struct Health Monit 6(3):429–445

28. de Almeida CR, Cury A, Barbosa F (2019) Automated real-time damage detection strategy using raw dynamic measurements. Eng Struct 196:109364. https://doi.org/10.1016/j.engstruct.2019.109364

29. Entezami A, Sarmadi H, Saeedi Razavi B (2020) An innovative hybrid strategy for structural health monitoring by modal flexibility and clustering methods. J Civ Struct Health Monit 10(5):845–859. https://doi.org/10.1007/s13349-020-00421-4

30. da Silva S, Junior MD, Junior VL, Brennan MJ (2008) Structural damage detection by fuzzy clustering. Mech Syst Sig Process 22(7):1636–1649

31. Figueiredo E, Cross E (2013) Linear approaches to modeling nonlinearities in long-term monitoring of bridges. J Civ Struct Health Monit 3(3):187–194

32. Yang C, Liu Y, Sun Y (2016) Damage detection of bridges considering environmental temperature effect by using cluster analysis. Procedia Engineering 161:577–582. https://doi.org/10.1016/j.proeng.2016.08.695

33. Coles S, Bawa J, Trenner L, Dorazio P (2001) An introduction to statistical modeling of extreme values, vol 208. Springer, Berlin

34. Salvadori G, De Michele C, Kottegoda NT, Rosso R (2007) Extremes in nature: an approach using copulas, vol 56. Springer Science and Business Media, Berlin

35. Deraemaeker A, Worden K (2018) A comparison of linear approaches to filter out environmental effects in structural health monitoring. Mech Syst Sig Process 105:1–15. https://doi.org/10.1016/j.ymssp.2017.11.045

36. Wu J (2012) Advances in k-means clustering: a data mining thinking. Springer Science and Business Media, Heidelberg

37. Gan G, Ma C, Wu J (2007) Data clustering: theory, algorithms, and applications, vol 20. University City, Siam

38. Aggarwal CC, Reddy CK (2016) Data clustering: algorithms and applications. CRC Press, Boca Raton

39. Weller-Fahy DJ, Borghetti BJ, Sodemann AA (2014) A survey of distance and similarity measures used within network intrusion anomaly detection. IEEE Commun Surv Tutor 17(1):70–91

40. Deza MM, Deza E (2014) Encyclopedia of distances, 3rd edn. Springer, Heidelberg

41. Abello J, Pardalos PM, Resende MGC (2013) Handbook of massive data sets. Springer US, New York

42. Rébillat M, Hmad O, Kadri F, Mechbal N (2018) Peaks over threshold-based detector design for structural health monitoring: application to aerospace structures. Struct Health Monit 17(1):91–107. https://doi.org/10.1177/1475921716685039

43. Krit M, Gaudoin O, Remy E (2019) Goodness-of-fit tests for the Weibull and extreme value distributions: a review and

comparative study. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2019.1594292

44. Perez-Rodriguez P, Vaquera-Huerta H, Villaseñor-Alva JA (2009) A goodness-of-fit test for the gumbel distribution based on Kullback–Leibler information. Commun Stat Theory Methods 38(6):842–855

45. Maeck J, De Roeck G (2003) Description of Z24 Bridge. Mech Syst Sig Process 17(1):127–131. https://doi.org/10.1006/mssp.2002.1548