World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:15, No:8, 2021

Dimensionality Reduction in Modal Analysis for Structural Health Monitoring

Elia Favarelli, Enrico Testi, Andrea Giorgetti

Abstract—Autonomous structural health monitoring (SHM) of many structures and bridges became a topic of paramount importance for maintenance purposes and safety reasons. This paper proposes a set of machine learning (ML) tools to perform automatic feature selection and detection of anomalies in a bridge from vibrational data and compare different feature extraction schemes to increase the accuracy and reduce the amount of data collected. As a case study, the *Z-24* bridge is considered because of the extensive database of accelerometric data in both standard and damaged conditions. The proposed framework starts from the first four fundamental frequencies extracted through operational modal analysis (OMA) and clustering, followed by time-domain filtering (tracking). The fundamental frequencies extracted are then fed to a dimensionality reduction block implemented through two different approaches: feature selection (intelligent multiplexer) that tries to estimate the most reliable frequencies based on the evaluation of some statistical features (i.e., entropy, variance, kurtosis), and feature extraction (auto-associative neural network (ANN)) that combine the fundamental frequencies to extract new damage sensitive features in a low dimensional feature space. Finally, one-class classification (OCC) algorithms perform anomaly detection, trained with standard condition points, and tested with normal and anomaly ones. In particular, principal component analysis (PCA), kernel principal component analysis (KPCA), and autoassociative neural network (ANN) are presented and their performance are compared. It is also shown that, by evaluating the correct features, the anomaly can be detected with accuracy and an F_1 score greater than 95%.

Keywords—Anomaly detection, dimensionality reduction, frequencies selection, modal analysis, neural network, structural health monitoring, vibration measurement.

I. Introduction

NOWADAYS SHM represents a fundamental research field in a society where historical and modern infrastructures coexist harmoniously. In this scenario, monitoring infrastructures, buildings, and bridges became a topic of primary importance that aim to maintain and protect the existing structure instead of replacing it with functionally and economically costly solutions [1]. As far as bridges are concerned, some statistics highlight the relevance of the problem. For example, currently, in Italy there are almost 2000 bridges that require special monitoring; in France, 4000 bridges need to be restored, and 840 are considered in critical conditions; in Germany, 800 bridges are reputed critic; in the United States of America, among the $600,000$ bridges, according to a conservative estimate, at least 1% of them is considered deficient. Being able to detect anomalies is thus of paramount importance in this scenario, and structural health monitoring (SHM) offered numerous solutions [2]–[4].

Elia Favarelli is with Unibo - University of Bologna, Italy (e-mail: elia.favarelli2@unibo.it).

A vast overview of damage detection and localization strategies is present in literature [5], [6]. Over the years, several techniques have been developed to extract the most significant damage-sensitive features. Such techniques can be divided into model-free and model-based. In the first case, the only information is gathered by measurements (e.g., acceleration, temperature, position), while in model-based approaches, information comes from measurements and prior knowledge of a model of the structure [7]. An important research topic is represented by the sensing strategy that can be implemented with traditional sensing networks, mobile agents able to perform sensing, and manage the energy supply [8].

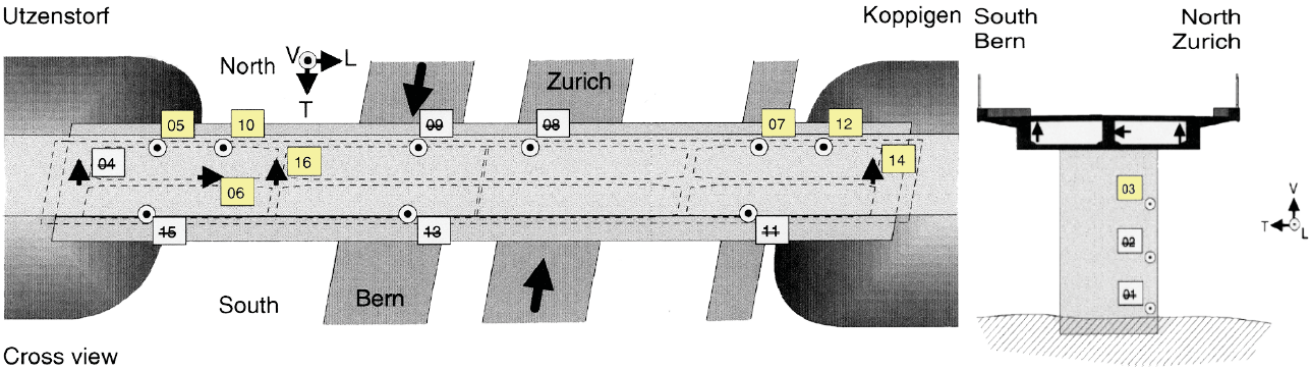
Since the whole monitoring procedure is quite complex and requires fine-tuning of several parameters, specific for the structure at hand, some works recently put forward the idea of adoption of ML techniques to detect changes in the damage sensitive features [9]–[13].

In this work, we attempt to provide a methodology to detect anomalies in bridges starting from vibrational data automatically. The proposed framework starts from the fundamental frequencies extracted from the accelerometer measurements through stochastic subspace identification (SSI), cleaning, and clustering [7], [9], [15]–[19], and then performs modal frequencies tracking in the time domain. The modal frequencies are then used as starting feature space that will be reduced through feature selection and feature extraction techniques. Finally, one-class classifiers are implemented to perform anomaly detection. In particular, the main contributions are the following:

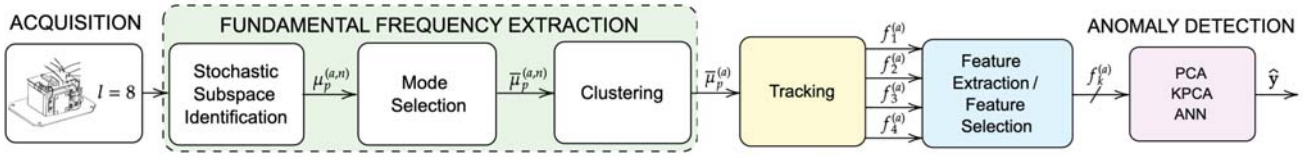
- We propose a time-domain tracking algorithm based on modal frequencies density.
- We investigate strategies to find the reliable frequencies extracted among all the extracted ones.
- We evaluate the effects of feature extraction and feature selection on the anomaly detectors' performance.
- We suggest techniques to increase the anomaly detection accuracy in SHM and reduce the amount of data.

The performance of the proposed solution is investigated on a real dataset using the accelerometric data available for the *Z-24* bridge [20], [21]. The proposed anomaly detection algorithms PCA, KPCA, and ANN are compared in terms of F_1 score, accuracy, recall, and precision. Several tests are performed to assess the impact of feature extraction and feature selection on the algorithms' performance.

Throughout the paper, capital boldface letters denote matrices and tensors, lowercase bold letters denote vectors, $(\cdot)^{\mathrm{T}}$ stands for transposition, $||\cdot||$ is the ℓ_2-norm of a vector,



(a)



(b)

Fig. 1 (a) Data acquisition setup along the Z24 bridge: the selected accelerometers, their positions, and the measured acceleration direction [14]; (b) Block diagram for signal acquisition, processing, feature extraction, and detection

and $\mathbb{1}\{a, b\}$ is the indicator function equal to 1 when $a = b$, and zero otherwise.

This paper is organized as follows. In Section II, a brief overview of the acquisition system, the accelerometers setup, and the monitoring scenario is presented. The fundamental frequencies extraction technique adopted is presented in Section III. The feature extraction and feature selection paradigms are presented in Section IV. A survey of anomaly detection techniques is reported in Section V. Numerical results are given in Section VI. Conclusions are drawn in Section VII.

II. System Configuration

The *Z-24* bridge was located in the Switzerland canton Bern. The bridge was a part of the road connection between Koppigen and Utzenstorf, overpassing the A1 highway between Bern and Zurich. It was a classical post-tensioned concrete two-cell box girder bridge with a main span of 30 m and two side spans of 14 m. The bridge was built as a freestanding frame with the approaches backfilled later. Both abutments consisted of triple concrete columns connected with concrete hinges to the girder. Both intermediate supports were concrete piers clamped into the girder. An extension of the bridge girder at the approaches provided a sliding slab. Al supports were rotated with respect to the longitudinal axis that yielded a skew bridge. The bridge was demolished at the end of 1998 [20]. Before complete demolition, the bridge was subjected to a long-term continuous monitoring test and several progressive damage tests:

- A long-term continuous monitoring test took place during the year before demolition. The aim was to quantify the environmental variability of the bridge dynamics.
- Progressive damage tests took place over a month, shortly before complete demolition. The aim was to prove experimentally that realistic damage has a measurable influence on the bridge dynamics. Progressive damage tests were alternated with short-term monitoring tests while the continuous monitoring system was still running during these tests.

The accelerometers position and their measurements axis are shown in Fig. 1 (a). In this work, we considered $l = 8$ accelerometers, identified as 03, 05, 06, 07, 10, 12, 14, and 16, which are present in both long-term continuous monitoring phase and in the progressive damage one.[1] Longitudinal acceleration is collected by sensors 03 and 06, transversal acceleration is measured by sensors 14 and 16, and all the remaining sensors gather vertical accelerations.

Every hour $N_s = 65536$ samples are acquired from each sensor with sampling frequency $f_{\text{samp}} = 100\,\text{Hz}$ which corresponds to an acquisition time $T_a = 655.36\,\text{s}$. Since the measurements are not always available, there are $N_a = 4107$ acquisitions collected in 44 weeks.

III. Fundamental Frequencies Extraction

The fundamental frequencies extraction chain is depicted in Fig. 1 (b); from the vibrational data the fundamental modes $\mu_p^{(a,n)}$ are extracted through the widely known SSI algorithm [7], where p represents the pth mode, a stays for the acquisition index, and n represents the model order varied in the range $n \in [2, 160]$ (with step 2) [12]. The resulting modes can be cleaned up by the spurious ones though classical mode selection methods (i.e., modal assurance criterion (MAC), mean phase deviation (MPD), complex conjugate poles check, and dumping ratios check) [16]–[19] and clustered with the K-means algorithm [9], [15]. The residual modes after

[1]Some accelerometers that experienced failures during the long-term monitoring have been avoided. Moreover, we select a subset of accelerometers present in both phases to ensure data consistency.

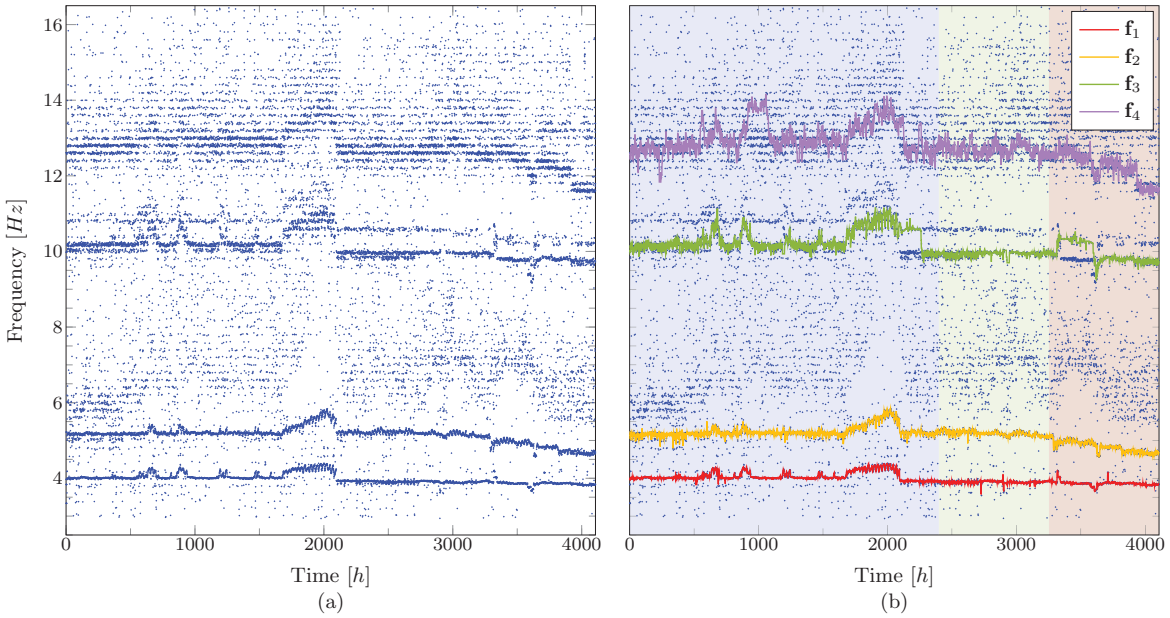


Fig. 2 (a) Fundamental frequencies extracted through SSI for each measurement; (b) First four natural frequencies estimation after the density-based tracking algorithm. Blue and green backgrounds highlight the acquisitions made during the bridge's normal condition, used respectively as training and test sets, while the red background stands for damaged condition acquisitions used in the test phase

selection are represented with $\bar{\mu}_p^{(a,n)}$, and the modes after clustering with $\bar{\mu}_p^{(a)}$. The results of this approach applied for all the acquisitions is reported in Fig. 2 (a). After that, a density-based mode tracking algorithm is proposed to track the fundamental frequencies, and it is described in the following.

A. Mode Tracking

The tracking phase is the final step in the natural frequencies extraction chain. Several algorithms can find the frequency traces starting from the estimation made through the clustering phase [18]. In this paper, a technique that needs neither the frequency starting points nor their number is proposed, contrary to several approaches that infer the number and the starting position of the fundamental frequencies through simulation of the physical structure. The proposed tracking algorithm consists of two steps: a starting phase and an online phase.

1) Starting Phase: Without any assumption about the structure, the idea is to analyze the data, $\bar{\mu}_p^{(a)}$, to find some clusters of points that could be the starting ones. To perform this task, the first $N_\mathrm{t} = 200$ acquisitions, i.e., $\bar{\mu}_p^{(a)}$ with $a = 1, 2, \ldots, N_\mathrm{t}$, are considered (see Fig. 2 (a)). From this initial data, the number of points that fall in frequency bins of bandwidth $B_\mathrm{f} = 0.4\,\mathrm{Hz}$ are counted. The histogram obtained is depicted in Fig. 2 (b). Selecting the largest values of the histogram the number of starting points and the corresponding frequencies, $f_s^{(0)}$, are estimated. In particular, the first estimated frequency is evaluated as the average values of the frequencies that fall in the respective bins. For example, according to Fig. 2 (b) the values of the starting points, $f_s^{(0)}$, in this case $s = 1, \ldots, 4$, are estimated and correspond to 4.0, 5.2, 10.1, and 12.8 Hz.

2) Online Phase: In this phase, a rectangular window that considers three acquisitions jointly (the one selected and the previous two) of width $w_s = 0.7\,\mathrm{Hz}$ is used to track each frequency evolution. This value is selected because lower values of w_s make the tracking algorithm less sensitive to prompt frequency variation; on the contrary, greater values make the tracking less stable and prone to drift effects. For each acquisition a the window is centered in $f_s^{(a-1)}$ and updated evaluating the mean value of the modes that belong to that acquisition a and the previous two ($a-1$ and $a-2$), and fall in the interval defined by $f_s^{(a-1)} \pm w_s/2$ (i.e., for the first acquisition the window is centered in the starting points $f_s^{(0)} \pm w_s/2$). More in detail, we can define the set of modes $\boldsymbol{\mu}_w^{(a)}$ that fall in the specified interval as:

$$\widehat{\boldsymbol{\mu}}^{(a)} = \left\{ \bar{\mu}_p^{(a)} : \bar{\mu}_p^{(a)} \in \left[f_s^{(a-1)} \pm w_s/2 \right] \right\}$$
$$\boldsymbol{\mu}_w^{(a)} = \left[\widehat{\boldsymbol{\mu}}^{(a)}, \widehat{\boldsymbol{\mu}}^{(a-1)}, \widehat{\boldsymbol{\mu}}^{(a-2)} \right]$$

where the concatenated vector $\boldsymbol{\mu}_w^{(a)}$ has cardinality N_w which varies with the acquisitions and tracks. Then, the updating rule can be expressed as:

$$f_s^{(a)} = \frac{1}{N_w} \sum_p \mu_{p,w}^{(a)}.$$

Repeating this procedure iteratively for each acquisition a, and each track s, the first four fundamental frequency tracks $\mathbf{f}_s = \left\{ f_s^{(a)} \right\}_{a=1}^{N_\mathrm{a}}$ with $s \in \{1, \ldots, 4\}$ are extracted and stored

in the following matrix (see also Fig. 2 (b)):

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} f_1^{(1)} & f_1^{(2)} & \dots & f_1^{(N_\mathrm{a})} \\ f_2^{(1)} & f_2^{(2)} & \dots & f_2^{(N_\mathrm{a})} \\ f_3^{(1)} & f_3^{(2)} & \dots & f_3^{(N_\mathrm{a})} \\ f_4^{(1)} & f_4^{(2)} & \dots & f_4^{(N_\mathrm{a})} \end{bmatrix}^{\mathrm{T}}.$$

IV. Dimensionality Reduction

In this section, we introduce two possible strategies to reduce the dimensionality of the damage-sensitive features extracted, intending to reduce the amount of data stored and transmitted through the sensor network and increase the anomaly detection performance.

As described in [20], the damage is introduced at the acquisition $a = N_\mathrm{d} = 3253$, corresponding to the installation of a lowering system. Therefore, from now on, the matrix $\bar{\mathbf{X}} = \mathbf{F}_{1:2N_\mathrm{d}-N_\mathrm{a}-1,:}$ contains the training points (blue background in Fig. 2b), $\bar{\mathbf{Y}} = \mathbf{F}_{2N_\mathrm{d}-N_\mathrm{a}:N_\mathrm{d}-1,:}$ contains the test points in standard condition (green background in Fig. 2 (b)), and $\bar{\mathbf{U}} = \mathbf{F}_{N_\mathrm{d}:N_\mathrm{a},:}$ contains the test points in damaged condition (red background in Fig. 2 (b)). The three subsets of acquisitions that correspond to training, test, and damaged points are, respectively, $\mathcal{I}_\mathrm{x} = \{1, ..., 2N_\mathrm{d} - N_\mathrm{a} - 1\}$, $\mathcal{I}_\mathrm{y} = \{2N_\mathrm{d} - N_\mathrm{a}, ..., N_\mathrm{d} - 1\}$, and $\mathcal{I}_\mathrm{u} = \{N_\mathrm{d}, ..., N_\mathrm{a}\}$.

Let us define the offset $\hat{\mathbf{x}}$ as the column vector containing the row-wise mean of the matrix $\bar{\mathbf{X}}$, and the rescaling factor $x_\mathrm{m} = \max_{a,s} |\bar{x}_{a,s} - \hat{x}_a|$. Before proceeding with the anomaly detection, the matrices $\bar{\mathbf{X}}$, $\bar{\mathbf{Y}}$ and $\bar{\mathbf{U}}$ are centered and normalized subtracting the offset $\hat{\mathbf{x}}$ row-wise and dividing each entry by the rescaling factor x_m. The resulting data matrices are $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{U}$, of size $N_\mathrm{x} \times D$, $N_\mathrm{y} \times D$, and $N_\mathrm{u} \times D$, respectively, with $D = 4$ features.

A. Feature Extraction

This technique consists of mapping a set of data in a low-dimensional feature space, trying to reduce an error function that represents the distance between the original data and the remapped data obtained from the mapping subspace. In this work, we decide to use the ANN as auto-encoders to accomplish this task [9], [11]. In this case, the normalized feature matrix $\mathbf{X}$ is fed to an ANN with the classic bottleneck structure that provides:

- Mapping layers, which consist of one or more hidden layers, with the number of neurons in each layer decreasing progressively till the last one named bottleneck. In the bottleneck, the number of neurons is usually lower than the number of input features;
- Demapping layers, composed of one or more hidden layers where the number of neurons increases progressively.

The input and output layers have the same dimension of the feature space, and the labels during the training phase must be set equals to the input data point. With this structure, the data are mapped in lower-dimensional feature space (with dimension equal to the number of neurons present in the bottleneck layer) and then reconstructed through the demapping layers minimizing the error with respect to the input data.

B. Feature Selection

This approach proposes selecting the most reliable features among all the available with some metrics to eliminate noisy components that can deteriorate the damage detection capability of the one-class classification (OCC) algorithms. In this work, we decided to evaluate some statistical features (i.e., variance, kurtosis, and entropy) to identify the most reliable fundamental frequencies among the four extracted, which are widely discussed in Section VI.

V. Survey of Anomaly Detection Techniques

In this section we briefly review PCA, KPCA, and ANN, which are often adopted for OCC [22]–[26].

A. Principal Component Analysis

This technique remaps the training data from the feature space $\mathbb{R}^D$ in a subspace $\mathbb{R}^P$ (where $P < D$ is the number of components selected) that minimizes the error (defined as Euclidean distance) between the data in the feature space and their projection in the chosen subspace [27]. More in detail, to find the best subspace to project the training data, we need to evaluate the $D \times D$ sample covariance matrix:

$$\mathbf{\Sigma}_\mathrm{x} = \frac{\mathbf{X}^\mathrm{T}\mathbf{X}}{N_\mathrm{x} - 1}. \tag{1}$$

By eigenvalue decomposition $\mathbf{\Sigma}_\mathrm{x}$ can be factorized as $\mathbf{\Sigma}_\mathrm{x} = \mathbf{V}_\mathrm{x}\mathbf{\Lambda}_\mathrm{x}{\mathbf{V}_\mathrm{x}}^\mathrm{T}$, where $\mathbf{V}_\mathrm{x}$ is an orthonormal matrix whose columns are the eigenvectors, while $\mathbf{\Lambda}_\mathrm{x}$ is a diagonal matrix that contains the D eigenvalues. The eigenvalues magnitude represents the importance of the direction pointed by the relative eigenvector. In our setting we select the largest component, hence $P = 1$, therefore the best linear subspace of dimension one is $\mathbf{v}_\mathrm{P}$, which coincides with the eigenvector related to the largest eigenvalue of $\mathbf{\Sigma}_\mathrm{x}$. The projection into the subspace is obtained multiplying the data by $\mathbf{v}_\mathrm{P}$, i.e., $\mathbf{x}_\mathrm{P} = \mathbf{X}\mathbf{v}_\mathrm{P}$, $\mathbf{y}_\mathrm{P} = \mathbf{Y}\mathbf{v}_\mathrm{P}$, and $\mathbf{u}_\mathrm{P} = \mathbf{U}\mathbf{v}_\mathrm{P}$.

To evaluate the error between the projected points and the starting ones, it is necessary to reconstruct the data in the original feature space, i.e., $\widetilde{\mathbf{X}} = \mathbf{x}_\mathrm{P}\mathbf{v}_\mathrm{P}^\mathrm{T}$, $\widetilde{\mathbf{Y}} = \mathbf{y}_\mathrm{P}\mathbf{v}_\mathrm{P}^\mathrm{T}$, and $\widetilde{\mathbf{U}} = \mathbf{u}_\mathrm{P}\mathbf{v}_\mathrm{P}^\mathrm{T}$. After the reconstruction, it is possible to evaluate the error as the Euclidean distance between the original and reconstructed data.

B. Kernel Principal Component Analysis

In our feature space, the linear boundaries found by principal component analysis (PCA) represent a severe limitation [28]. kernel principal component analysis (KPCA) firstly maps the data with a non-linear function, named kernel, then applies the standard PCA to find a linear boundary in the new feature space. Such boundary becomes non-linear in the original feature space. A crucial point in KPCA is the selection of the kernel that leads to linearly separable data in the new

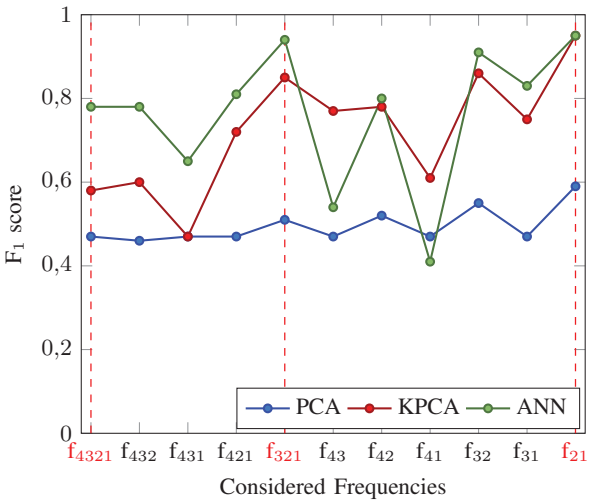


Fig. 3 F_1 score varying the considered fundamental frequencies, vertical dashed red lines indicate the best configuration with 4, 3, and 2 frequencies

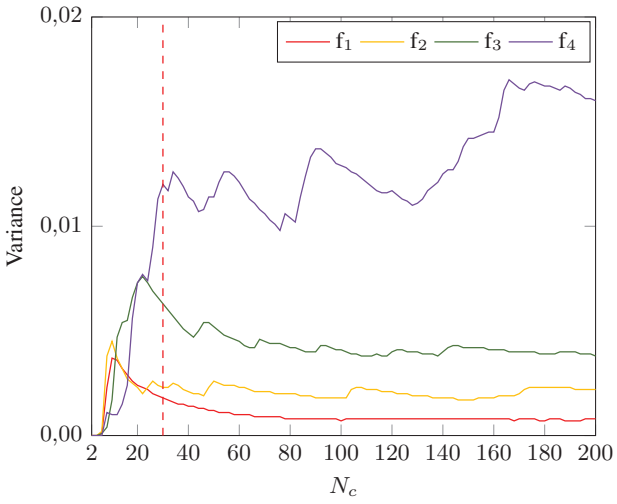


Fig. 4 Modal frequencies variance for different points; vertical dashed red lines indicate the minimum number of points for the correct frequency sorting

feature space. In [29], when the data distribution is unknown, the radial basis function (RBF) kernel is proposed as a good candidate to accomplish this task. Given a generic point $\mathbf{z}$ that correspond to a $1 \times D$ vector, we can apply the RBF as:

$$K_n^{(\mathbf{z})} = e^{-\gamma||\mathbf{z}-\mathbf{x}_n||^2}, \quad \text{with} \quad n = 1, 2, \ldots, N_\text{x} \qquad (2)$$

where γ is a kernel parameter (which controls the width of the Gaussian function) that must be set properly, $\mathbf{x}_n$ is the nth row of $\mathbf{X}$, and $K_n^{(\mathbf{z})}$ is the nth component of the point $\mathbf{z}$ in the kernel space. Overall, the vector $\mathbf{z}$ is mapped in the vector $\mathbf{k}^{(\mathbf{z})} = [K_1^{(\mathbf{z})}, K_2^{(\mathbf{z})}, \ldots, K_{N_\text{x}}^{(\mathbf{z})}]$. Remapping all the data in the kernel space, we obtain the subsequent matrices $\mathbf{K}_\text{x}$ of size $N_\text{x} \times N_\text{x}$ for training, $\mathbf{K}_\text{y}$ of size $N_\text{y} \times N_\text{x}$ for validation, and $\mathbf{K}_\text{u}$ of size $N_\text{u} \times N_\text{x}$ for test, respectively.

Applying now the PCA to the new data sets, it is possible to find non-linear boundaries in the original feature space.

C. Autoassociative Neural Network

The same structure described in IV-A can be used to perform anomaly detection. After the training phase, the network is fed with the new test points $\mathbf{Y}$ and $\mathbf{U}$; if a new data point belongs to the normal class, the reconstruction error is expected to be low; otherwise, if the data refer to an anomaly, the output of the demapping layers is likely to be very different from the input.

VI. Numerical Results

In this section, the proposed algorithms are applied to the *Z-24* bridge data set to detect anomaly based on the fundamental frequencies estimation [7], [30], [31], and a reduced number of features. The performance is evaluated through the following metrics considering only the test set:

$$\text{Accuracy} = \frac{T_\text{P} + T_\text{N}}{T_\text{P} + T_\text{N} + F_\text{P} + F_\text{N}}$$
$$\text{Precision} = \frac{T_\text{P}}{T_\text{P} + F_\text{P}}$$
$$\text{Recall} = \frac{T_\text{P}}{T_\text{P} + F_\text{N}}$$
$$\text{F}_1 \text{ score} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}$$

where T_P, T_N, F_P, and F_N, represent respectively true positive, true negative, false positive, and false negative predictions. Such indicators are obtained comparing the actual labels $[\zeta^{(1)}, \ldots, \zeta^{(N_\text{a})}]$, with those predicted by the OCC $[\widehat{\zeta}^{(1)}, \ldots, \widehat{\zeta}^{(N_\text{a})}]$. In this application, labels are 0 for normal condition and 1 for anomaly condition, respectively. Therefore,

$$T_\text{P} = \sum_{a \in \mathcal{I}_\text{u}} \mathbb{1}\left\{\zeta^{(a)}, \widehat{\zeta}^{(a)}\right\} \quad \text{and} \quad T_\text{N} = \sum_{a \in \mathcal{I}_\text{y}} \mathbb{1}\left\{\zeta^{(a)}, \widehat{\zeta}^{(a)}\right\}$$

with $F_\text{N} = N_\text{u} - T_\text{N}$, and $F_\text{P} = N_\text{y} - T_\text{P}$. In the case of unbalanced classes in the test set, the F_1 score represents a more reliable metric to evaluate the performance with respect to accuracy.

The feature space has dimension $D = 4$, and unless otherwise specified the three data sets used for training, test in normal condition, and damaged condition, have cardinality $N_\text{x} = 2399$, $N_\text{y} = 854$, and $N_\text{u} = 854$, respectively. For PCA, the number of components selected is $P = 1$. For KPCA, after several tests the values of P and γ that ensure the minimum reconstruction error are $P = 3$ and $\gamma = 8$. Regarding the autoassociative neural network (ANN) we adopted a fully connected network with 7 layers of, respectively, 50, 20, 10, k, 10, 20 and 50 neurons with k number of features extracted, with ReLU activation functions for the feature extraction task, and a fully connected network with 5 layers of, respectively, 100, 50, 1, 50 and 100 for the anomaly detection. All the

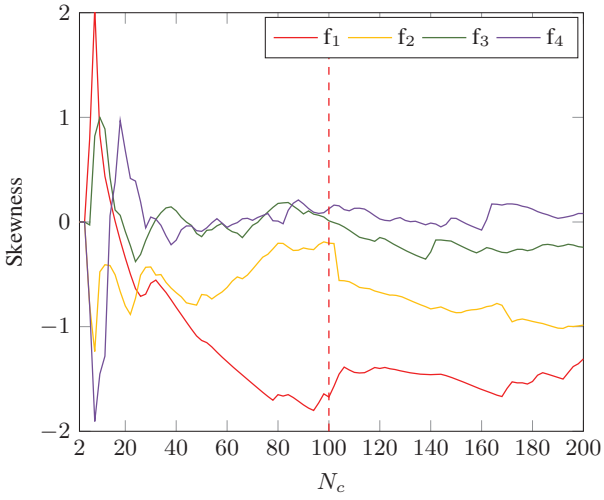


Fig. 5 Modal frequencies skewness varying the number of points; vertical dashed red lines indicate the minimum number of points for the correct frequency sorting

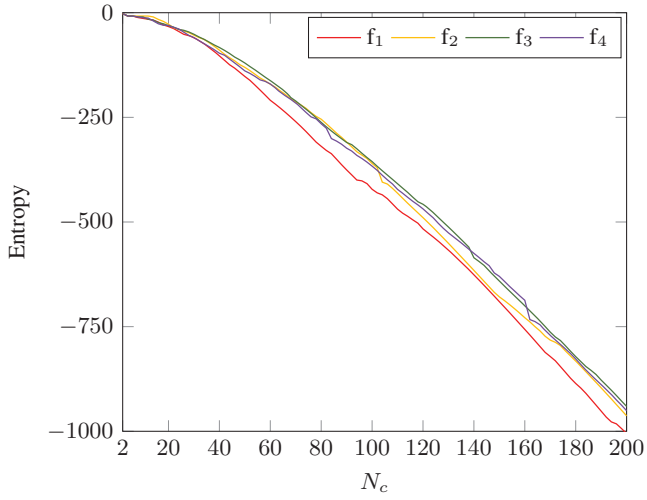


Fig. 6 Modal frequencies entropy varying the number of points

neural networks (NNs) are trained for a number of epochs $N_e = 5000$ with a learning rate $\rho = 0.05$. The error function adopted for a training set $\mathbf{X}$ is:

$$\mathcal{E}_{\mathbf{X}} = -\sum_{n=1}^{N_x}\sum_{c=1}^{C} t_{n,c} \ln \widetilde{t}_{n,c}, \tag{3}$$

where N_x is the number of points in the training set, C is the number of classes ($C = 2$), $t_{n,c} = 1$ if the nth acquisition belongs to the cth class and zero otherwise, and $\widetilde{t}_{n,c}$ is the activation function value for point n of the cth output neuron [10].

A. Frequencies Selection

First of all, a brute force approach is implemented to evaluate the fundamental frequencies that provide the best performance of the anomaly detection algorithms in terms of F_1 score with all the possible combinations of features. As reported in Fig. 3, the best performance is achieved by the algorithms with the same feature configuration, highlighted with the red vertical dashed lines, that allows sorting the features with increasing importance as f_4, f_3, f_2, and f_1. Hence a good metric to describe the fundamental frequencies reliability must sort the frequencies in the same order. With this aim, three statistical metrics are reported as a good candidate to accomplish this task.

The first one is the sample variance of the frequencies extracted defined as:

$$\text{Variance} = \frac{\sum_{i=1}^{N_c}(f_s^{(i)} - \bar{f}_s)^2}{N_c - 1}$$

where N_c is the number of acquisition considered, $f_s^{(i)}$ is the ith acquisition of the sth fundamental frequency, and $\bar{f}_s$ stands for the mean value of the sth frequency evaluated in the interval $\{1, \ldots, N_c\}$. As reported in Fig. 4, this feature works well after $N_c = 30$ observations; as depicted, the variance of reliable features is lower with respect to the noisy ones; hence this method can be successfully used to sort the frequencies in the correct order. The second metric proposed is the skewness that measure the asymmetry of the probability distribution, i.e.,

$$\text{Skewness} = \frac{\frac{1}{N_c}\sum_{i=1}^{N_c}(f_s^{(i)} - \bar{f}_s)^3}{\left(\sqrt{\frac{1}{N_c}\sum_{i=1}^{N_c}(f_s^{(i)} - \bar{f}_s)^2}\right)^3}.$$

As shown in Fig. 5, also this metric can be used to sort the frequencies correctly; the method becomes reliable after around $N_c = 100$ observations. Finally, in Fig. 6 the entropy is evaluated as a further metric:

$$\text{Entropy} = -\sum_{i=1}^{N_c} P(f_s^{(i)}) \log_{10} P(f_s^{(i)})$$

where the probability density function $P(f_s^{(i)})$ is evaluated numerically implementing data binning. In this case, the trend is descendent because the information introduced by new measurements decreases by increasing the number of observations. This metric can also be used to sort the frequencies, but in this case, f_3 and f_4 are inverted for some values of N_c and f_2 is very close to the previous two and can be missorted.

B. Performance Comparison

This paragraph provides a performance comparison of the anomaly detection algorithms and dimensionality reduction techniques. In Fig. 7 the performance of PCA, KPCA, and ANN is evaluated in term of F_1 score, varying the dimensionality of the features considered. Dashed curves are referred to features extracted with auto-encoder implemented with ANN. As we can observe from the plot, the dimensionality reduction slightly decreases the algorithms' performance; hence, for this application, feature extraction does not improve the system's anomaly detection capability. On the contrary, continuous lines are referred to feature

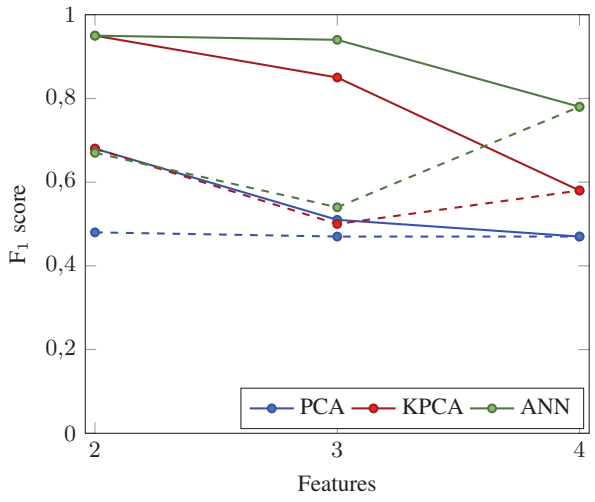


Fig. 7 F_1 score varying the dimensionality of the feature space, dashed lines correspond to frequency extraction technique, and continuous lines are referred to as the frequency selection approach

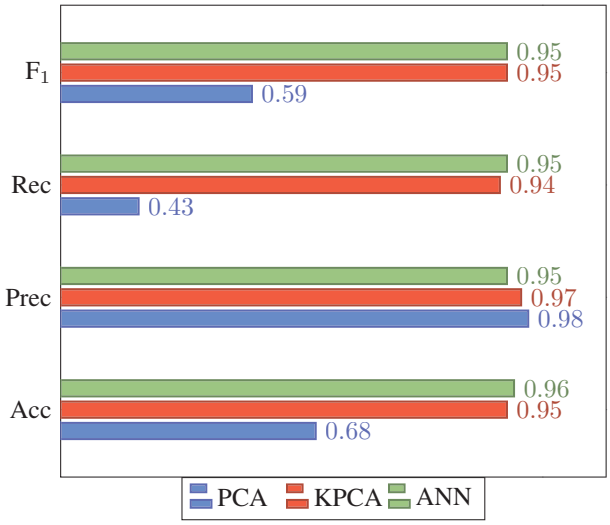


Fig. 8 Comparison of the classification algorithms in terms of F_1 score, recall, precision, and accuracy with the best feature configuration

selected through the metrics described in the previous paragraph; in this case, the selection of the correct fundamental frequencies tend to increase the detection performance, hence it is strongly suggested for this application for the fact that both reduce the dimensionality of the problem and increase the anomaly detection capability. In Fig. 8 the performance of the algorithms in terms of F_1 score, recall, precision, and accuracy is reported in the best configuration, hence considering only the first two fundamental frequencies selected by the described metrics (f_1 and f_2). As we can see, the F_1 score of both KPCA and ANN are around 95%, and the accuracy for ANN is greater than 96% that represents a remarkable result in this anomaly detection application.

VII. Conclusion

In this paper, we present a SHM system that aims to extract the most reliable damage sensitive features to implement anomaly detection with high accuracy. An overview of some widely used anomaly detection algorithms is provided. Two different paradigms to implement dimensionality reduction are presented and compared in terms of damage detection capability. In this sense, a best practice is to select the most reliable fundamental frequencies among all the extracted ones, which provide low dimensionality of the problem and better performance with respect to the feature extraction implemented through auto-encoders. A set of metrics with which select the most reliable frequencies is presented and widely discussed. Finally, a detailed comparison of algorithm performance is reported in the best system configuration, and it is shown that in these conditions, it is possible to detect damage with accuracy greater than 96%.

References

[1] R. Ferrari, D. Froio, E. Chatzi, F. Pioldi, and E. Rizzi, "Experimental and numerical investigations for the structural characterization of a historic RC arch bridge," in *Proc. Int. Conf. on Comp. Methods in Structural Dynamics and Earthquake Eng.*, vol. 1, Athens, Greece, May 2015, pp. 2337–2353.

[2] A. Benedetti, M. Tarozzi, G. Pignagnoli, and C. Martinelli, "Dynamic investigation and short-monitoring of an historic multi-span masonry arch bridge," in *Proc. Int. Conf. on Arch Bridges*, vol. 11, Porto, Portugal, Oct. 2019, pp. 831–839.

[3] A. Benedetti, G. Pignagnoli, and M. Tarozzi, "Damage identification of cracked reinforced concrete beams through frequency shift," *Materials and Structures*, vol. 51, pp. 1–15, Oct. 2018.

[4] A. Benedetti, C. Colla, G. Pignagnoli, and M. Tarozzi, "Static and dynamic investigation of the taro masonry bridge in parma, italy," in *Proc. Int. Conf. on Struct. Analysis of Hist. Const.*, vol. 18, Cusco, Perù, Sep. 2019, pp. 2264–2272.

[5] G. D. Roeck, "The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience," *J. of Structural Control*, vol. 10, no. 2, pp. 127–134, May 2003.

[6] K. Worden, C. Farrar, J. Haywood, and M. Todd, "A review of nonlinear dynamics applications to structural health monitoring," *Structural Control and Health Monitoring*, vol. 15, no. 4, pp. 540–567, Jul. 2008.

[7] G. Fabbrocino and C. Rainieri, *Operational modal analysis of civil engineering structures*. New York: Springer-verlag, May 2014.

[8] S. Taylor, K. Farinholt, E. Flynn, E. Figueiredo, D. L. Mascarenas, E. A. Moro, G. Park, M. Todd, and C. Farrar, "A mobile-agent-based wireless sensing network for structural monitoring applications," *Measurement Science and Technology*, vol. 20, no. 4, pp. 1–14, Jan. 2009.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Verlag, Aug. 2006.

[10] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine Learning Refined*. Cambridge University Press, 2016.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[12] E. Favarelli and A. Giorgetti, "Machine learning for automatic processing of modal analysis in damage detection of bridges," *IEEE Trans. on Instrumentation and Measurement*, 2020.

[13] L. Pucci, E. Testi, E. Favarelli, and A. Giorgetti, "Human activities classification using biaxial seismic sensors," *IEEE Sensors Letters*, vol. 4, no. 10, pp. 1 – 4, Oct. 2020.

[14] Z-24 bridge data. [Online]. Available: https://bwk.kuleuven.be/bwm/z24

[15] E. P. Carden and J. M. W. Brownjohn, "Fuzzy clustering of stability diagrams for vibration-based structural health monitoring," *Computer-Aided Civil and Infrastructure Eng.*, vol. 23, no. 5, pp. 360–372, May 2008.

[16] C. Wu, H. Liu, X. Qin, and J. Wang, "Stabilization diagrams to distinguish physical modes and spurious modes for structural parameter identification," *Journal of Vibroeng.*, vol. 19, no. 4, pp. 2777–2794, Jun. 2017.

[17] D. Brigante, C. Rainieri, and G. Fabbrocino, "The role of the modal assurance criterion in the interpretation and validation of models for seismic analysis of architectural complexes," in *Proc. Int. Conf. on Structural Dynamics (Eurodin)*, vol. 199, Rome, Italy, Sep. 2017, pp. 3404–3409.

[18] A. Cabboi, F. Magalhães, C. Gentile, and Á. Cunha, "Automated modal identification and tracking: Application to an iron arch bridge," *Structural Control and Health Monitoring*, vol. 24, no. 1, p. e1854, Feb. 2017.

[19] M. Pastor, M. Binda, and T. Harčarik, "Modal assurance criterion," *Procedia Engineering*, vol. 48, pp. 543–548, 2012.

[20] E. Reynders and G. D. Roeck, "Continuous vibration monitoring and progressive damage testing on the z 24 bridge," *Encyclopedia of structural health monitoring*, vol. 10, pp. 127–134, May 2009.

[21] E. Reynders, J. Houbrechts, and G. D. Roeck", "Fully automated (operational) modal analysis," *Mechanical Systems and Signal Processing*, vol. 29, pp. 228–250, May 2012.

[22] A. Santos, M. Silva, C. Sales, J. Costa, and E. Figueiredo, "Applicability of linear and nonlinear principal component analysis for damage detection," in *Proc. IEEE Int. Instr. and Meas. Tech. Conf. (I2MTC)*, Pisa, Italy, May 2015, pp. 869–874.

[23] E. Favarelli, E. Testi, L. Pucci, and A. Giorgetti, "Anomaly detection using wifi signal of opportunity," in *Proc. IEEE Int. Conf. on Signal Proc. and Comm. Sys. (ICSPCS)*, Surfers Paradise, Gold Coast, Australia, Dec. 2019, pp. 1–7.

[24] E. Favarelli, E. Testi, and A. Giorgetti, "One class classifier neural network for anomaly detection in low dimensional feature spaces," in *Proc. IEEE Int. Conf. on Signal Proc. and Comm. Sys. (ICSPCS)*, Surfers Paradise, Gold Coast, Australia, Dec. 2019, pp. 1–7.

[25] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Trans. Image Process*, vol. 28, no. 11, pp. 5450–5463, Nov. 2019.

[26] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," *CoRR*, vol. abs/1802.06360, Aug. 2018.

[27] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interd. Reviews: Comp. Stat.*, vol. 2, no. 4, pp. 433–459, 2010.

[28] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. Int. conf. on artificial neural networks*, vol. 1327, no. 6. Lausanne, Switzerland: Springer, Oct. 1997, pp. 583–588.

[29] B. Schölkopf, A. Smola, E. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comp.*, vol. 10, pp. 1299–1319, Jul. 1998.

[30] M. Silva, A. Santos, E. Figueiredo, R. Santos, C. Sales, and J. Costa, "A novel unsupervised approach based on a genetic algorithm for structural damage detection in bridges," *Eng. Applications of Artificial Intelligence*, vol. 52, pp. 168–180, Jun. 2016.

[31] M. Silva, A. Santos, R. Santos, E. Figueiredo, C. Sales, and J. C. Costa, "Agglomerative concentric hypersphere clustering applied to structural damage detection," *Mechanical Systems and Signal Processing*, vol. 92, pp. 196–212, Feb. 2017.