Fast unsupervised learning methods for structural health monitoring with large vibration data from dense sensor networks

Alireza Entezami[1,2], Hashem Shariatmadar[1], and Stefano Mariani[2]

[1] *Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Azadi Square, Mashhad, Iran*

[2] *Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy*

Abstract

Data-driven damage localization is an important step of vibration-based structural health monitoring (SHM). Statistical pattern recognition paradigm based on the prominent steps of feature extraction and statistical decision-making provides an effective and efficient SHM framework. However, these steps may become time-consuming or complex when there are large volumes of vibration measurements acquired by dense sensor networks. To deal with this issue, this study proposes fast unsupervised learning methods for feature extraction through AutoRegressive (AR) modeling and damage localization through a new distance measure called Kullback-Leibler Divergence with Empirical Probability Measure (KLDEPM). The feature extraction approach consists of an iterative algorithm for order selection and parameter estimation aiming to extract residuals in the training phase, and of another iterative process aiming to extract residuals only in the monitoring phase. The key feature of the proposed approach is the use at each iteration of correlated residual samples of the AR model as a new time series, rather than handling the measured vibration response of the structure. This is shown to highly reduce the computational burden of order selection and feature extraction; moreover, it

effectively provides low-order AR models with uncorrelated residuals. The KLDEPM method exploits a segmentation technique to subdivide random data into independent sets, and provides a distance metric based on the theory of empirical probability measure with no need to explicitly compute the actual probability distributions at the training and monitoring stages. Numerical and experimental benchmarks are then used to assess accuracy and performance of the proposed methods, and compare them with some state-of-the-art approaches. Results show that the proposed approach is successful in feature extraction and damage localization, with a reduced computational burden.

Keywords: Structural health monitoring; damage diagnosis; unsupervised learning; autoregressive model; feature extraction; Kullback-Leibler divergence.

---

1. Introduction

Structural health monitoring (SHM) has recently become a very active research area in the civil, mechanical and aerospace engineering fields. It essentially aims at assessing the health of structures, or systems in a broader sense, and at detecting any possible damage via vibration data[1]. Damage in a structure can occur because of changes in the geometrical configuration or boundary conditions, or because of material degradation due to e.g. cracks in concrete, loose bolts and broken welds in steel connections, corrosion, and fatigue. All these effects may cause non-recoverable variations of the structural stiffness and, therefore, unfavorable vibrations, local failures and even collapse. To prevent such events and decrease the high costs of maintenance and rehabilitation, if necessary, SHM is becoming a paradigm in our smart-cities age, regardless of the site-specific geographical location and economic development.

Any SHM strategy can be generally classified according to whether a damage diagnosis or prognosis is provided[2, 3]. The former class can be subdivided on its own into three main levels: early damage detection (level 1); damage localization (level 2); and damage quantification (level 3). The latter class is instead intended to predict the remaining lifetime of the structure[4]. Early damage detection is conceived as a global process that attempts to perceive if the damage is present in the whole structure; on the other hand, damage localization and quantification are local procedures, aiming to provide a clear snapshot of the damage pattern along with an assessment of its severity. Since damage is an inherently local phenomenon, rigorous and robust methods look necessary for SHM procedures targeting damage localization.

Damage diagnosis is generally carried out with model-driven or data-driven methods[5]. The model-driven approach is based on an analytical or finite element model of the structure[6]; due to the discrepancies between the model itself and the real-life structure, in which case model updating becomes a mandatory task[7-12]. Although such an approach is often successful in relation to damage diagnosis[13-15], some limitations are represented by the necessity of a detailed model, of a model updating procedure, and of data reduction from raw vibration measurements. It then results to be excessively time-consuming and makes data-driven techniques appealing as a more feasible option.

Within the data-driven methods, statistical pattern recognition stands as an efficient and powerful framework since it relies upon the direct use of raw vibration (time-domain) measurements, in terms of excitations and relevant structural responses. The relevant framework for vibration-based SHM can be then decomposed into four main steps: operational evaluation; data acquisition; feature extraction; and statistical decision-making, or statistical modeling for feature discrimination[2]. The last two steps are in need

of more research activities, due to their high importance and required effectiveness. Feature extraction refers to the process of learning meaningful information from the measurements, supposed to be pertinent to damage; such information is represented by the so-called damage-sensitive features (DSFs). Statistical decision-making is concerned with the implementation of algorithms that handle the DSFs relevant to undamaged and damaged conditions, to provide the sought damage diagnosis. As the functional relationship between the DSFs and the damaged state is difficult to define by means of model-driven procedures, statistical methods are usually adopted in the process of decision-making[2].

Dealing with vibration data in the time domain, time series modeling is adopted for feature extraction. In this regard, Fassois and Sakellariou[16] discussed parametric and non-parametric time series representations for vibration-based SHM problems. They concluded that although the use of non-parametric representations is simple, parametric time series models are more reliable. Application of parametric time series modeling to the feature extraction is based on fitting an appropriate mathematical model to the measured data and extracting DSFs such as the model residuals. Depending upon the nature of the time series (e.g. stationarity vs. non-stationary, linear vs. nonlinear, seasonal vs. non-seasonal, etc.) and the type of data acquisition (e.g. input-output vs. output-only), a broad range of representations can be adopted to extract significant DSFs. Considering linear and stationary vibration data, widely used time-invariant linear models are the AutoRegressive (AR)[17-19], AutoRegressive with eXogenous input (ARX)[20], AutoRegressive Moving Average (ARMA)[21, 22], AutoRegressive Moving Average with eXogenous input (ARMAX)[23, 24], AutoRegressive and AutoRegressive with eXogenous input (ARARX)[25] ones. On the other hand, Vector-dependent Functionally Pooled (VFP)

models[26], Gaussian Process (GP) time-series models[27], Time-Varying AutoRegressive Moving Average (TV-ARMA)[28], and Trigonometric Box-Cox ARMA Trend Seasonal (TBATS)[29] are suitable for conditions that time series data are exposed to uncertainties, operational and environmental variability, ambient vibration, and seasonal variations. Among the time-invariant linear representations, the AR model offers remarkable advantages for feature extraction: the model depends only on the response of the structure, so that it is appropriate for output-only SHM strategies; the model coefficients (parameters) are linked to the inherent structural properties, regardless of the excitation sources and their variations[30, 31]; the coefficients and residuals of the AR model are known to be sensitive to damage [30]; its implementation looks also simple.

Feature extraction via time-invariant linear representations is generally obtained with coefficient-based and residual-based algorithms[16, 32]. Having set model orders and coefficients for the undamaged state, especially with the latter ones, the time series will no longer provide a good fit to the structural response in the damaged state, and residuals will increase. Therefore, the great benefit of the residual-based approach is that it does not require any order determination and parameter estimation for the damaged, or current structural states[16].

The performance of AR models in extracting DSFs, either with coefficient-based or residual-based algorithms, is affected by the choice of the model order to enable the generation of uncorrelated residuals. From a statistical viewpoint, the so-called accuracy and sufficiency of the time series orders depend strongly on the uncorrelatedness of the residual sequences[33, 34]: an inadequate model order would not allow the AR model to track the underlying dynamics of the structure, ultimately leading to the extraction of damage-insensitive features[35]. The Akaike Information Criterion (AIC) and the Bayesian

Information Criterion (BIC) are well-known techniques for order selection [33]; several research studies have been recently conducted to propose some alternative, even more efficient and effective methods. Broersen[36] proposed the use of Combined Information Criterion (CIC) and Finite Sample Information Criterion (FSIC), to enhance the capability of determining the model order for large sampling data. In connection with operational modal analysis via time series modeling, a new order selection method for Vector AutoRegressive (VAR) representations was suggested by Vu et al.[37], regardless of the measurement noise. For SHM problems, Figueiredo et al.[35] assessed the influence of the AR orders on damage detection, using four different information criterion techniques and showing that a low order may provide insensitive features and a weak damage detectability. To robustly select the order, with an emphasis on extracting uncorrelated residuals, Entezami and Shariatmadar [38] proposed an iterative procedure via the Ljung-Box Q-test (LBQ), assuring model accuracy and adequacy. Rezaei-Pajand et al.[39] also presented a two-stage iterative algorithm, to select the model order on the basis of a residual analysis through statistical hypothesis tests and signal filtering.

Although all the above-mentioned order selection techniques have their own advantages, it may be questionable to assess whether they are indeed time-saving or time-consuming. This is especially true by considering the recent advances and improvements in sensor technology and data acquisition, overall providing a great opportunity to exploit dense sensor networks for most of the SHM applications. Handling vibration datasets with high dimensionality, an approach looks efficient if it requires a short computational time for order selection, guarantying in the end model sufficiency and accuracy by generating uncorrelated residuals. Since order selection directly affects parameter estimation and residual extraction, a positive by-product is the computational efficiency

of the coefficient-based and residual-based feature extraction algorithms. The other way around, high-order models leading to a good fit of the structural output and providing uncorrelated residuals may increase the probability of overfitting, which finally causes erroneous forecasts [33]. An effective order selection approach must be therefore time-saving, or efficient and generate uncorrelated residuals without any overfitting issue.

Besides feature extraction, the methodology has to be robust also in terms of statistical decision-making for feature discrimination. A reliable way to attain this latter goal is to adopt a well-established statistical distance metric to measure the discrepancy between different sets of samples. Successful approaches used for damage diagnosis relied upon the Mahalanobis-squared distance[40, 41], the Itakura and cepstral distances[42], the Kullback–Leibler divergence (KLD)[43, 44], the Kolmogorov-Smirnov statistical test (KSTS)[45, 46], dynamic time warping[47], the multivariate distance correlation[31], and damage indices such as the Fisher criterion[48], Q-statistic and T2-statistic[49], and the Deflection Coefficient (DC)[50].

All the aforementioned methods can be exploited via machine learning tools, in either supervised or unsupervised learning manners [38, 51, 52]. Damage diagnosis through machine learning is so decomposed into the training (baseline) and monitoring (inspection) phases [2]. During training, a statistical model or classifier is learned by accounting for the training data. In the monitoring stage, one then attempts to make a decision based on the learned model and the currently acquired data. The main difference between supervised and unsupervised learning algorithms is that the former needs the DSFs relevant to the undamaged and damaged conditions to train a classifier, whereas the latter is based on DSFs of the undamaged state only to train the classifier. The clear

benefit of the unsupervised learning algorithms is that no information (i.e. vibration data or DSFs) is required beforehand for any possible damage condition.

Going more into detail, the capability to locate single and multiple damage states characterized by different severity levels, particularly the small ones, plays a prominent role in establishing the effectiveness and robustness of the SHM approach. The decision-making process based on an unsupervised learning strategy must then be fast and overcome the main obstacle of accurate damage localization, whenever a large volume of DSFs can be extracted from data provided by dense sensor networks.

Accounting for all the above-mentioned issues and limitations, a fast unsupervised learning method is here proposed to locate structural damage within the statistical pattern recognition paradigm, by means of an iterative feature extraction approach. The approach is based on AR modeling and a novel statistical distance method, named Kullback-Leibler Divergence with Empirical Probability Measure (KLDEPM). The proposed feature extraction approach consists of: an improved iterative AR model order selection with parameter estimation and residual extraction during the training phase; iterative AR residual extraction only during the monitoring stage. The major novelty of this approach consists of using correlated residual samples of the AR model as a new time series dataset in each iteration, in place of the measured vibration response of the structure. It is shown that this innovation can considerably reduce the computational time of AR order selection and residual extraction; further to that, it leads to low-order AR models that guarantee to finally obtain uncorrelated residuals. The proposed KLDEPM method is an enhancement of the classical KLD technique, in the sense that random feature samples (i.e. the AR model residuals related to the normal and damaged conditions) are subdivided into independent segments, and information on the segmentation (in terms of numbers of

samples and segments) are accounted for in the distance calculation with the theory of empirical probability measure. All these novelties make it an effective distance approach for damage identification, and a fast tool for decision-making. Results relevant to a concrete beam[53] and the IASC-ASCE experimental benchmark structure[54] have been used to verify accuracy and performance of the proposed method. Both the iterative feature extraction and the KDLEPM method turn out to be superior to other available strategies, providing reliable DSFs and locating single and multiple damages of different severity.

The remainder of this paper is organized as follows. Section 2 briefly discusses the vibration response modeling by an AR representation. Section 3 presents the proposed iterative residual-based feature extraction approach. The novel KLDEPM method is detailed in Section 4. The results of feature extraction and damage localization for the mentioned examples are presented in Section 5. Finally, Section 6 draws the main conclusions of this work.

2. AutoRegressive modeling

Within the SHM realm, a time series is a sequence of data that actually consists of time spaced, successive measurements. Time series analysis is, therefore, a statistical tool that is meant for model identification, parameter estimation, model validation of forecasting purposes [33]. It is also known to be a powerful approach for feature extraction from vibration measurements, actually consisting of excitations and/or structural responses.

Assuming that a linear time-invariant representation can fit the structural response [55], the relevant AR model for a single-output system reads:

$$x(t) = \sum_{i=1}^{p} \theta_i x(t-i) + e(t) \quad (1)$$

where: $x(t)$ is the measured vibration response at time t; $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_p]$ denotes the vector of AR coefficients, or parameters to be estimated; p represents the model order; $e(t)$ is an uncorrelated residual sequence (an unobservable random error) that quantifies the difference between the measured vibration response and the one predicted by the model. In this work, the least-squares technique is used to estimate the coefficients $\boldsymbol{\theta}$ of the AR model.

3. Iterative residual-based feature extraction technique

The feature extraction technique here proposed is subdivided into the two stages of training and monitoring; on its own, each stage consists of an initial phase and of a subsequent iterative phase. In the mentioned initial stages, an optimal AR model order is set by an ad-hoc designed order selection method, so that model residuals are provided as DSFs to the next iterative phases. The monitoring phase exploits the information provided by training and, without dealing with order determination and parameter estimation, extracts the residuals of the AR model as the main DSFs for the current structural state. Since order selection and parameter estimation tasks are not tackled in such a monitoring stage, the proposed residual-based feature extraction method conforms to unsupervised learning. Details of the whole procedure are provided next.

3.1. Training phase: AR order selection and residual extraction

The essential step of feature extraction by time series modeling is the determination of accurate and sufficient series orders, aiming to lead to uncorrelated

residuals [33, 34]. The iterative order selection algorithm here proposed is an improvement of a former authors' approach [38]; the fundamental principle of the algorithm relies on the choice of the AR order, based on the residual analysis via the LBQ hypothesis test. This test assesses the correlation among residual sequences as follows:

$$Q_{LB} = n(n+2)\sum_{\mathrm{j}=1}^{L}\frac{\rho_{\mathrm{j}}^{2}}{n-\mathrm{j}} \tag{2}$$

where: n is the number of samples in the residual vector; ρ_{j} is the sample autocorrelation function at the j^{th} lag (j=1,2,…,L); L represents the number of lags. Test decision is based on either the null hypothesis or alternative ones linked to a certain significance level α: model residuals are thus considered uncorrelated if Q_{LB} is less than a critical value (c-value), or if the test probability value (p-value) is greater than the significance level.

Although a purely iterative order selection technique is able to robustly set the AR model order providing uncorrelated residuals, it may be time-consuming or may yield excessively high-order models. With the present approach, correlated residual sequences are used as a new time series within the iterative process, in place of the measured vibration response; in this way, the order selection turns out to be computationally more efficient than with the original (purely iterative) technique. This is because the range of correlation decreases at each iteration of the improved method. In other words, the new correlated residual samples have less correlation than the residuals of the previous iterations and the original vibration signal. This situation allows us to examine fewer sample orders and selecting a small order than the purely iterative technique due to reducing the correlation of the new time series data (the residual samples) and the sufficiency of using small sample orders. Unlike the purely iterative technique, furthermore, one does not need to use the original vibration signal at each iteration. As a

result, these reasons help to expedite the order determination and enable the improved method to be more efficient than the purely iterative technique. In the following, these two steps are described in details.

Initial step: Let $\mathbf{x}=[x(1)\ \ldots\ x(n)]\in\Re^n$ be the vector of measured vibration responses of the structure associated with the undamaged or normal condition. AR(1) (i.e. an AR model of order one) is fitted to **x**, in order to extract the model residuals $\mathbf{E_x}=[e(2)\ldots e(n)]\in\Re^{n-1}$. Before extracting such residuals, model coefficients $\boldsymbol{\theta}_{\mathbf{x}_1}$ are estimated by the least-squares technique. Since the residual sequence $e(1)$ for this first-order model is approximately null, it can be neglected [33]. In case of using large volume of vibration measurements, featuring n relatively large, the residuals of AR(1) turn out to be correlated; hence, $\mathbf{E_x}$ is handled as a new series dataset for the subsequent iterative step. This is done since, if the residuals are correlated, some information present in the original dataset is not captured by the model. It is worth mentioning that, if the analysis of AR(1) via the LBQ test allows ascertaining that residuals are uncorrelated, which is indeed a rare event, one can respectively adopt $p=1$ and $\mathbf{E_x}$ as model order and main DSF for the normal condition.

Iterative step: By increasing the model order with $k = 2,3,\ldots$, AR(k) is fitted to the correlated residual sequence furnished by the previous step if $k=2$, or by the previous iteration if $k>2$, in order to extract the new residual vector $\mathbf{E}_{\mathbf{x}_k}=[e(\frac{k(k+1)}{2}+1)\ldots e(n)]\in\ \Re^{n-k(k+1)/2}$. Once again, before residual extraction the least-squares technique is adopted to estimate the coefficients $\boldsymbol{\theta}_{\mathbf{x}_k}$ of the new AR model. Alike in the initial step, the first k samples of the residual vector of AR(k) are null and are therefore dropped from $\mathbf{E}_{\mathbf{x}_k}$ at the current iteration. If the sequences of $\mathbf{E}_{\mathbf{x}_k}$ satisfy the null hypothesis by the LBQ test ($Q_{LB}\leq c$-value), the iterative process is terminated; otherwise

(Q_{LB}>c-value), the model order k is increased and the correlated residual vector is exploited as a new dataset ($\hat{\mathbf{x}} = \mathbf{E}_{\mathbf{x}_k}$). When the termination condition is attained, the number of iterations provides an optimal and robust AR order, and the relevant uncorrelated residual vector $\mathbf{E}_{\mathbf{x}_k}$ can be used as the main DSF for the training phase. Model coefficients at each iteration are stored, to be later used in the residual extraction during the monitoring phase. For the sake of convenience, Algorithm 1 presents the pseudo-code for the initial and iterative steps of the proposed feature extraction method in the training phase.

Algorithm 1. Initial and iterative steps of the proposed feature extraction method for the training phase

Initialization

nm=Number of test measurements;

ns=Number of sensors;

p0=Minimum AR order (default p0=1);

pm=Maximum AR order;

Initial Step

for $i \leftarrow 1$ to nm **do**

  for $j \leftarrow 1$ to ns **do**

    - Read the vector of vibration time-domain samples regarding the normal condition for the j^{th} sensor and i^{th} test measurement
    - Estimate the coefficient of AR(p0), θ_{p0}
    - Extract the model residuals $e(1) \ldots e(n)$
    - Implement the LBQ test and obtain the test statistics Q_{KL} and its c-value

    if Q_{LB}<c-value **then**

      - Store p0, θ_{p0}, $e(2) \ldots e(n)$ as the optimal AR order, the model coefficient, and the main features at the j^{th} sensor and i^{th} test measurement
      - Terminate the iterative loop

    else

      Iterative Step

      for $k \leftarrow$ p0+1 to pm **do**

        - Set $e\left(\frac{k(k+1)}{2}+1\right) \ldots e(n)$ as the new time series data
        - Estimate the coefficients of AR(k), $\theta_1 \ldots \theta_k$
        - Store the model coefficients $\theta_1 \ldots \theta_k$ for the k^{th} iteration, j^{th} sensor, and i^{th} test measurement
        - Extract the model residuals $e\left(\frac{k(k+1)}{2}+1\right) \ldots e(n)$
        - Implement the LBQ test and obtain the test statistics Q_{KL} and its c-value

        if Q_{LB}>c-value **then**

          - Continue the iterative loop

        else

          - Store k, $\theta_1 \ldots \theta_k$, and $e\left(\frac{k(k+1)}{2}+1\right) \ldots e(n)$ as the optimal AR order, the model coefficients, and the final features for the k^{th} iteration, j^{th} sensor, and i^{th} test measurement

        end

      end

    end

  end

end

3.2. Monitoring phase: residual extraction

In order to obtain the DSFs relevant to the current structural state, the initial and iterative steps previously detailed need to be repeated again to extract the residual sequences.

Initial step: Let $\mathbf{y}=[y(1)\ \dots\ y(n)]\in\Re^n$ be now the vector of measured vibration responses associated with the current structural state. By fitting AR(1) to these data, the residual vector $\mathbf{E_y}=[\varepsilon(2)\dots\varepsilon(n)]\in\Re^{n-1}$ is obtained. If the residual vector $\mathbf{E_x}$ of the initial step of the training phase is used as the main DSF for the normal condition, one can do the same with $\mathbf{E_y}$ for the current state; otherwise, $\mathbf{E_y}$ is used as a new time series in the subsequent iterative step.

Iterative step: The iterative procedure in this step is governed by the variable order $r=2,3,\dots,k$. By using the model coefficients $\boldsymbol{\theta}_{\mathbf{x}_r}$ estimated in the training phase, at the r^{th} iteration the residuals of AR(r) are extracted. If $r = k$, the residual vector $\mathbf{E}_{\mathbf{y}_r}=[\varepsilon(\frac{r(r+1)}{2}+1)\dots\varepsilon(n)]\in\Re^{n-r(r+1)/2}$ is used as the DSF for the current structural condition; otherwise, the iterative process continues by choosing $\mathbf{E}_{\mathbf{y}_r}$ as the new time series data ($\hat{\mathbf{y}}$). For simplicity, Algorithm 2 provides the pseudo-code for the initial and iterative steps of the proposed feature extraction method in the monitoring stage.

Algorithm 2. Initial and iterative steps of the proposed feature extraction method for the monitoring phase

`Initialization`

`nm`=Number of test measurements;

`ns`=Number of sensors;

`Load`

Read the optimal AR orders, the model coefficients for all sensors and all test measurements from **Algorithm 1**

if the optimal AR order=`p0` **or** k=1 **then**

  `Initial Step`

  for i←1 to `nm` **do**

    for j←1 to `ns` **do**

      - Read the vector of vibration time-domain samples regarding the damaged state for the j^{th} sensor and i^{th} test measurement
      - Fit AR(`p0`) to the vibration time-domain samples
      - Extract the model residuals ε(`p0`+1)…$\varepsilon(n)$ as the main features for the j^{th} sensor and i^{th} test measurement

    end

  end

else

  `Iterative Step`

  for i←1 to `nm` **do**

    for j←1 to `ns` **do**

      - Set k as the optimal AR order at the j^{th} sensor and i^{th} test measurement

      for r← `p0`+1 to k **do**

        - Fit AR(r) using the known model coefficients $\theta_1 \ldots \theta_r$ regarding the r^{th} iteration obtained from **Algorithm 1**
        - Extract the model residuals $\varepsilon\left(\frac{r(r+1)}{2}+1\right) \ldots \varepsilon(n)$

        if r=k **then**

          - Store $\varepsilon\left(\frac{r(r+1)}{2}+1\right) \ldots \varepsilon(n)$ as the main features of the damaged state at the j^{th} sensor and i^{th} test measurement
          - Terminate the iterative loop at the iteration r

        else

          - Set $\varepsilon\left(\frac{r(r+1)}{2}+1\right) \ldots \varepsilon(n)$ as the new time series data
          - Continue the iterative loop by a new value for r

        end

      end

    end

  end

end

4. A novel statistical distance measure for damage localization

4.1. Kullback-Leibler Divergence (KLD)

The KLD technique proposed by Kullback and Leibler in [56] is a well-known statistical distance metric. It is a non-symmetric measure of the difference between two probability distributions, or better between two probability density functions. If $\mathbf{P_x}$ and $\mathbf{P_y}$ are the probability distributions of the random samples $\mathbf{E}_{\mathbf{x}_k}$ and $\mathbf{E}_{\mathbf{y}_r}$, the general formulation for the KLD reads:

$$D_{KL}\left(\mathbf{P_x} \,\|\, \mathbf{P_y}\right) = \sum_{z} \mathbf{P_x}(z) \log \frac{\mathbf{P_x}(z)}{\mathbf{P_y}(z)} \tag{3}$$

Since $\mathbf{P_x}$ and $\mathbf{P_y}$ are both positive, Eq. (3) always provides a positive distance or divergence measure. In practical terms, $D_{KL}=0$ means that the two probability distributions have the same behavior, and any deviation from the zero value is representative of dissimilarity between them.

4.2. Kullback-Leibler Divergence with Empirical Probability Measure (KLDEPM)

Unlike the conventional KLD technique, the main advantage of the KLDEPM method is that one does not need to compute the probability distributions of the data. The key features of this method are the segmentation of random samples and the adoption of an empirical probability measure [57]. The calculation of the empirical probability measure directly depends on the way data samples are segmented.

The strategy used in the proposed KLDEPM method is based on the maximum entropy approach. The residual vector $\mathbf{E}_{\mathbf{y}_r}$ is arranged with entries in ascending order, in such a way that it begins with the minimum residual value and ends with the maximum

one. Subsequently, the rearranged vector $\mathbf{E}_{\mathbf{y}_r}$ is divided into m sub-vectors in the following way:

$$\begin{aligned} &\varepsilon_{\min} \le \mathbf{C}_1 \le \varepsilon_{N_y} \\ &\varepsilon_{(\mathrm{i}-1)N_y} < \mathbf{C}_{\mathrm{i}} \le \varepsilon_{\mathrm{i}N_y} \\ &\varepsilon_{N_y(m-1)} < \mathbf{C}_m \le \varepsilon_{\max} \end{aligned} \tag{4}$$

where i=2,3,…,m-1. In Eq. (4), N_y denotes the number of samples within each segment of $\mathbf{E}_{\mathbf{y}_r}$, empirically given as the square root of the total number of samples in each residual vector, that is:

$$N_y = \sqrt{n - \frac{k(k+1)}{2}} \tag{5}$$

Accordingly, the number of segments turns out to be $m = \frac{n-\frac{k(k+1)}{2}}{N_y}$. Since both N_y and m have to be positive integers, one needs to round off the outcomes of Eq. (5) to the nearest integer. Once vector $\mathbf{E}_{\mathbf{y}_r}$ has been segmented, the segmentation procedure is repeated almost the same for the vector $\mathbf{E}_{\mathbf{x}_k}$, since it is not necessary to arrange it in ascending order. Next, the sequences of $\mathbf{E}_{\mathbf{x}_k}$ that fall into the domains of $\mathbf{C}_1,\mathbf{C}_2,\ldots,\mathbf{C}_m$ are determined, given that N_x represents now the number of samples in each segment.

Let $\hat{P}_x$ and $\hat{P}_y$ be the empirical probability measures of the residual vectors $\mathbf{E}_{\mathbf{x}_k}$ and $\mathbf{E}_{\mathbf{y}_r}$, respectively. On the basis of the information on the segmentation provided by N_y, N_x, and m, such measures for the i^{th} segment are given by:

$$\hat{P}_x\left(\mathbf{C}_{\mathrm{i}}\right) = \frac{N_{x_{\mathrm{i}}}}{n - \frac{k\left(k+1\right)}{2}} = \frac{2N_{x_{\mathrm{i}}}}{2n - k\left(k+1\right)} \tag{6}$$

$$\hat{P}_y\left(\mathbf{C}_{\mathrm{i},\mathrm{i}\neq m}\right)=\frac{N_y}{n-\frac{k\left(k+1\right)}{2}}=\frac{2N_y}{2n-k\left(k+1\right)} \tag{7}$$

where $N_{x\mathrm{i}}$ denotes the number of samples of $\mathbf{E}_{\mathbf{x}_k}$ belonging to the segment $\mathbf{C}_\mathrm{i}$. It is worth remarking that only the first m-1 segments of the residual vector $\mathbf{E}_{\mathbf{y}_r}$ have the same number of samples; for the last segment, the empirical probability measure is written as follows:

$$\hat{P}_y\left(\mathbf{C}_m\right)=\frac{n-\frac{k\left(k+1\right)}{2}-mN_y}{n-\frac{k\left(k+1\right)}{2}}=\frac{2n-k\left(k+1\right)-2mN_y}{2n-k\left(k+1\right)} \tag{8}$$

Having obtained the empirical probability measures, the KLDEPM is formulated as:

$$\hat{D}_{KL}\left(\mathbf{E}_{\mathbf{x}_k}\,\|\,\mathbf{E}_{\mathbf{y}_r}\right)=\sum_{\mathrm{i}=1}^{m}\hat{P}_x\left(\mathbf{C}_\mathrm{i}\right)\log\frac{\hat{P}_x\left(\mathbf{C}_\mathrm{i}\right)}{\hat{P}_y\left(\mathbf{C}_\mathrm{i}\right)} \tag{9}$$

The great advantage of the proposed KLDEPM method, as detailed here above, is that it computes the distance between the two random datasets via the information given by segmentation and empirical probability measures, without directly handling the random samples as with the classical KLD technique. Since all the values obtained with the segmentation are positive, $\widehat{D}_{\mathrm{KL}}$ is positive as well. Ideally, a zero distance would denote a similarity between the residual vectors relevant to the normal and current structural states, which practically means that damage has not occurred.

4.3. Unsupervised learning strategy

In order to locate damage by means of a univariate distance method such as the KLDEPM, the distance must be computed between the residual vectors for the

undamaged and damaged conditions at each sensor location. Considering N_s sensors deployed over the structure, one can build a damage localization vector by assembling the $\widehat{D}_{\mathrm{KL}}$ values associated with all the sensors as follows:

$$\mathbf{D}=\left[\hat{D}_{KL}\left(\mathbf{E}_{\mathbf{x}_k}\,\|\,\mathbf{E}_{\mathbf{y}_r}\right)_1 \quad \hat{D}_{KL}\left(\mathbf{E}_{\mathbf{x}_k}\,\|\,\mathbf{E}_{\mathbf{y}_r}\right)_2 \quad \cdots \quad \hat{D}_{KL}\left(\mathbf{E}_{\mathbf{x}_k}\,\|\,\mathbf{E}_{\mathbf{y}_r}\right)_{N_s}\right]\in\Re^{N_s} \tag{10}$$

An efficient and rigorous way for the localization of damage is to define a threshold limit for the distance values obtained from all the sensors. Let the training phase consist of N_c normal conditions, where N_c represents e.g. the number of measurements during the training period: the distance between the residual vectors related to each normal condition and all the other ones is computed according to Eq. (10) at all sensor locations. The obtained distance matrices $\mathbf{T}_l\in\Re^{N_s\times N_c}$ for each undamaged state, where l=1,2,…,N_c, are then gathered in a global distance matrix $\mathbf{T}\in\mathrm{R}^{N_s\times N_c^2}$. A threshold limit is next defined, based on the upper bound of the 95% confidence interval for each column of **T**, according to:

$$\rho_{\mathrm{t}}=\mu_T+Z_{\alpha/2}\frac{\sigma_T}{\sqrt{N_s}},\quad \mathrm{t}=1,2,...,N_c^2 \tag{11}$$

where μ_T and σ_T are the mean and the standard deviation of each column of **T**; $Z_{\alpha/2}$ is the confidence coefficient, equal to 1.96 for the 95% confidence interval under an α=0.05 significance level. The mean value of the entries of $\boldsymbol{\rho}=[\rho_1\ldots\rho_{N_c^2}]$ finally represents the threshold limit for damage localization. In this regard, the sensor locations with the distance values more than the threshold limit are identified as the damaged areas of the structure. This means that the sensors should be arranged close to the possible damaged areas to ensure the damage sensitivity. Under such circumstances, an optimal and accurate sensor placement is of paramount importance to locate damage. On the other hand, it

should be mentioned that the strategy of finding the sensor location with distance value larger than the threshold limit is often valid for non-complex structural systems. This is because the damage localization procedure cannot simply be boiled down to tracking the location of a larger discrepancy in sensed signals or their damage indicators. Damage in structural systems may manifest in complex forms depending on the system's complexity, boundary conditions, and localized nonlinear elements. Therefore, this strategy needs possible extensions and combinations with further SHM localization criteria.

5. Applications

5.1. A numerical concrete beam

To verify the accuracy and performance of the proposed methods, a numerical benchmark model is first considered as shown in Fig. 1. This benchmark was constructed by Kullaa et al.[53] as a realistic simulation of a simply supported concrete beam, cracked at the mid-span cross-section and featuring a length of 5m and a height of 0.5m. The beam was modeled in ABAQUS[58], space discretized with 4-node two-dimensional finite elements. At the beam ends, the constraints were provided at the neutral axis. For excitation purposes, a uniform transverse random load was applied to the top surface.

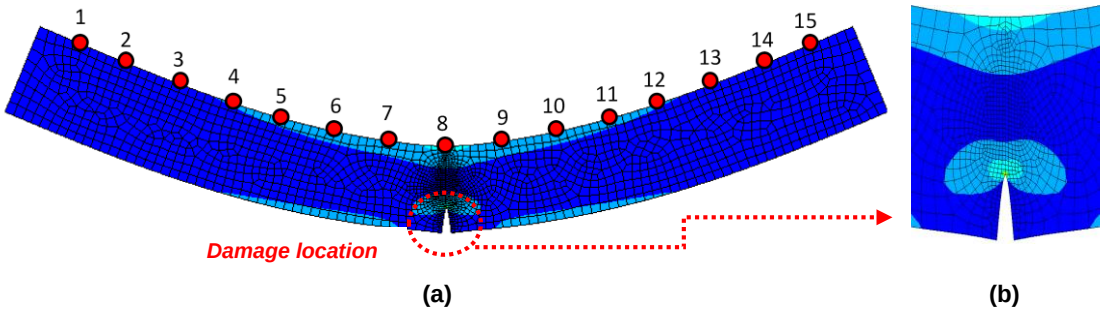

Fig. 1. The numerical benchmark model of a concrete beam: (a) sensor locations and damaged area, (b) element mesh in the vicinity of the crack [53]

Fifteen sensors were supposed to be mounted on the top surface of the beam, to measure the time histories of the acceleration in the transverse (vertical) direction. The measurement period was assumed to last two seconds, to include 4001 samples. A noise featuring a 30 dB signal-to-noise ratio was added to the digital acceleration records, as obtained with the finite element analyses. The histories were low-pass filtered at 1000Hz, in order to keep five vibration modes of the structure possibly excited in the test results.

The crack at the bottom of the beam mid-span was used to simulate a single damage state: crack lengths of 10, 30, 50, and 100mm are used in this work to varying the amount of damage. For each damage scenario, twelve test measurements are considered, so that the first ten ones are relevant to the undamaged condition, while the information provided by the last two varies depending on the crack length. To have the same amount of data in the undamaged and damaged conditions, the structural states and, therefore, the corresponding measurements are arranged as shown in Table 1; accordingly, the acceleration response at each sensor location always contains 8002 data samples. The healthy states are thus assumed to be five, whereas the damage cases, at varying damage level, are the remaining four ones. As a sample, Fig. 2 shows the acceleration responses at the sensor 8 in Cases 1 and 5.

Table 1. Structural conditions relevant to the numerical concrete beam

Case no.	Label	Measurement no.	Damage level/crack length (mm)
1	Healthy	1-2	-
	Healthy	3-4	-
	Healthy	5-6	-
	Healthy	7-8	-
	Healthy	9-10	-

2	Damaged	11-12	10
3	Damaged	11-12	30
4	Damaged	11-12	50
5	Damaged	11-12	100

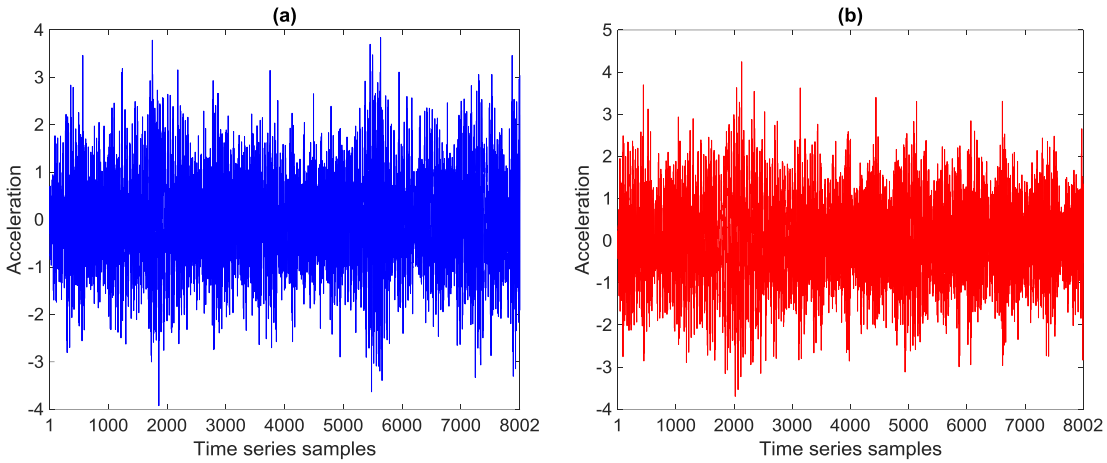

Fig. 2. The acceleration responses of Sensor 8: (a) Case 1, (b) Case 5

Before moving to the results related to damage detection, the reasons for choosing the AR representation are discussed. An effective way to identify the appropriateness of this model is through the Leybourne-McCabe (LMC) hypothesis test [59]: in essence, this test assesses the null hypothesis that a univariate time series conforms to an AR process, against the alternative hypothesis relevant to a non-stationary AutoRegressive Integrated Moving Average (ARIMA) process. In other words, if the test statistics Q_{LM} is smaller than a *c*-value, which refers to as before to the null hypothesis, one can argue that the time series is stationary and the AR model is suitable to represent it.

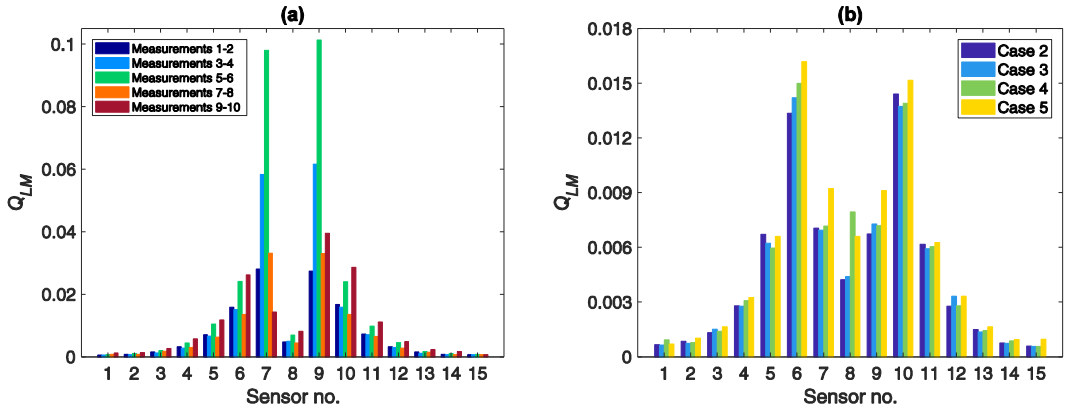

Fig. 3. Stationarity assessment and AR model identification by the LMC hypothesis test: (a) Case 1, (b) Cases 2-5

Fig. 3 shows the LMC test statistics for the healthy and damaged conditions. It can be seen that all the test statistics are smaller than the *c*-value, which amounts to 0.1460 under a 5% significance limit; all the acceleration time histories in all the cases can then be considered stationary processes conforming to AR ones. Accordingly, the AR model for feature extraction is assumed accurate for SHM purposes. The other reason that may help to verify the choice of AR modeling is to use the Box-Jenkins methodology[33], which presents a graphical model identification approach. It utilizes the autocorrelation function (ACF) and partial autocorrelation function (PACF) of time series data and identifies a suitable time series model among the AR, MA, and ARMA representations. In this regard, if the samples of ACF have an exponential decay form or behave as a sine wave without any trend in being zero and the samples of PACF gradually cut off after a specific lag, these indicate that time series data conform to the AR process. This means that the most proper time series model is AR. In contrast, if the samples of both the ACF and PACF have exponential decay forms or treat as sine waves, one can conclude that the ARMA model is suitable for the time series data.

With these descriptions, Fig. 4 illustrates the ACF and PACF plots of some vibration signals of the concrete beam for the first two tests (i.e. the measurements 1-2) of Case 1. As can be seen, all samples of the ACF behave as sine waves, while the samples of the PACF gradually cut off approximately after 50 lags. These observations confirm that the AR representation is suitable for modeling the acceleration responses. It needs to mention that the same conclusion can be reached for the other vibration signals in other cases.

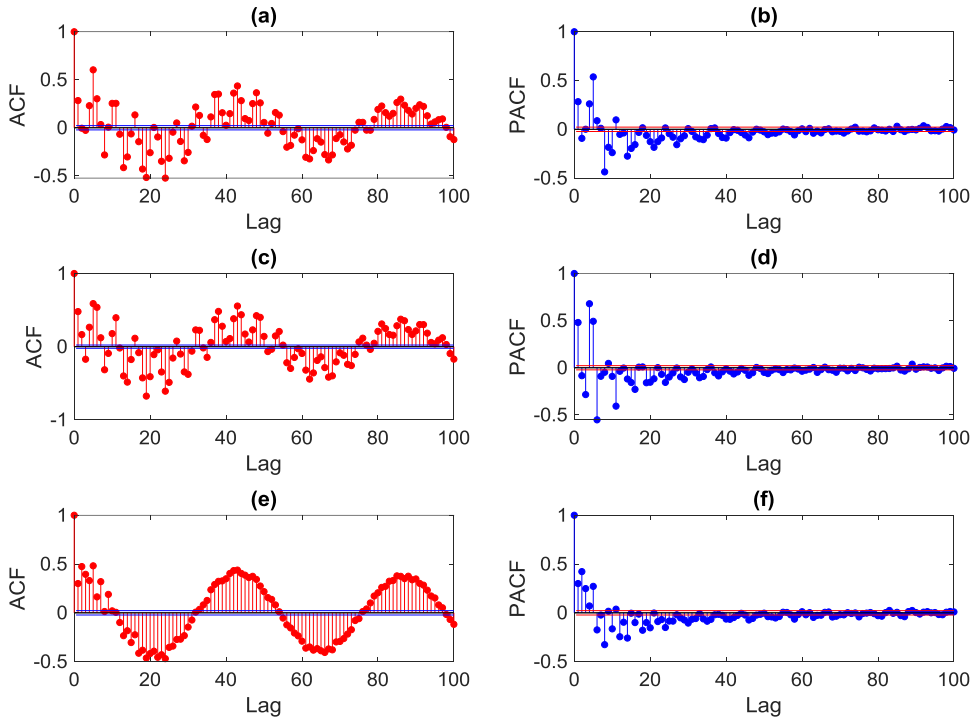

Fig. 4. Graphical model identification between the AR and ARMA representations by the Box-Jenkins methodology in Case 1: (a) ACF at Sensor 4, (b) PACF at Sensor 4, (c) ACF at Sensor 8, (d) PACF at Sensor 8, (e) ACF at Sensor 11, (f) PACF at Sensor 11

Using the test measurements for Case 1, Fig. 5(a) shows the identified AR orders at all sensor locations, as obtained with the proposed two-stage order selection method. To demonstrate the superiority of this method, the orders obtained with the original

iterative approach and the BIC technique are shown in Fig. 5(b) and 5(c). It can be seen that the range of the AR orders obtained with the improved method is upper bounded by 26, whereas the original approach and the BIC technique provide results with orders up to 52 or 65, respectively.

To further show how the improved order selection method yields low-order AR models, still preserving their adequacy and accuracy by generating uncorrelated residuals, Fig. 6 illustrates the variations of the LBQ test statistics with the number of iterations for the exemplary results related to sensors 5 and 14 and the measurements 9-10 of Case 1, see Table 1. Using a 5% significance level, in this case, the *c*-value turns out to be 31.4104: graphs show that the obtained orders, namely k=20 for sensor 5 and k=26 for sensor 14, enable the AR models to generate uncorrelated residuals, with outcomes of the LBQ test smaller than the *c*-value. For more evaluations, Fig. 7 indicates the statistics for analyzing the uncorrelatedness and Gaussianity (normality) of the residual samples in Cases 1-5. In this regard, the LBQ test is applied to investigate whether the residual samples are uncorrelated. For the problem of Gaussianity of the residuals, the Anderson-Darling (AD) hypothesis test[60] is applicable. The outputs of this test are similar to the LBQ test, in which case if the AD test statistics to be smaller than the *c*-value, one can infer that the residual samples are Gaussian.

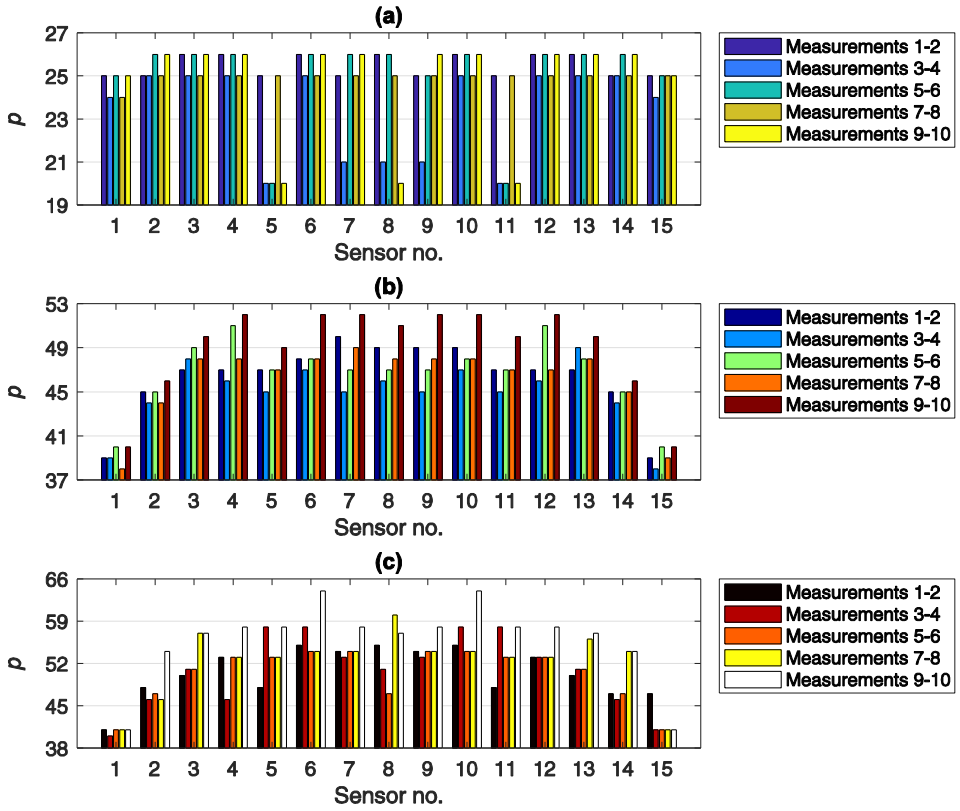

Fig. 5. AR model orders for all test measurements relevant to Case 1, as obtained with (a) the proposed method, (b) the original iterative method, and (c) the BIC method

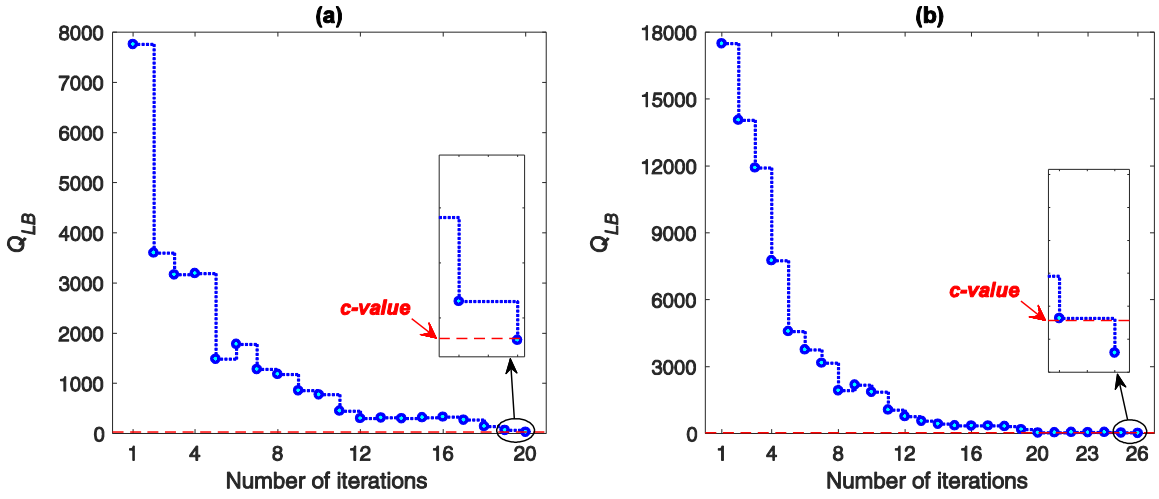

Fig. 6. LBQ test statistics for the test measurements 9-10 of Case 1: (a) Sensor 5, (b) Sensor 14

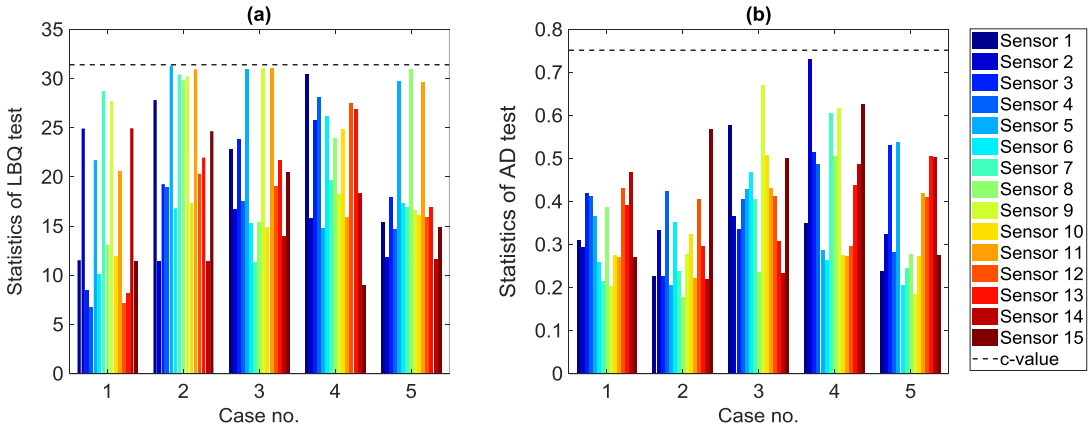

Fig. 7. Analysis of the residual samples of the AR models in Cases 1-5 of the numerical beam: (a) assessment of the residual uncorrelatedness by the LBQ test, (b) assessment of the residual Gaussianity by the AD test

As can be seen in Fig. 7(a) and 7(b), the statistics of both the LBQ and AD tests are smaller than the *c*-values under the 5% significance level. These observations prove the uncorrelatedness and Gaussianity of the residual samples extracted from the AR models. Regarding the computational efficiency for the AR order selection, Fig. 8(a) shows the computational time of the improved and original iterative algorithms, and of the BIC approach, obtained by using the acceleration time histories from all the sensors. Fig. 8(b) provides instead a comparison among the times needed for the complete process of residual-based feature extraction, by allowing for the datasets of measurements 1-2 of Case 1 for the training period and of measurements 11-12 of Case 2 for the monitoring phase. In the plot, the data relevant to the main steps of AR order selection, parameter estimation and residual extraction during training, and residual extraction during monitoring are detailed. These results have been obtained by running the analyses with Matlab R2017a on a PC featuring an Intel™ Core i7-3770, 3.40–3.90 GHz CPU and 16G RAM.

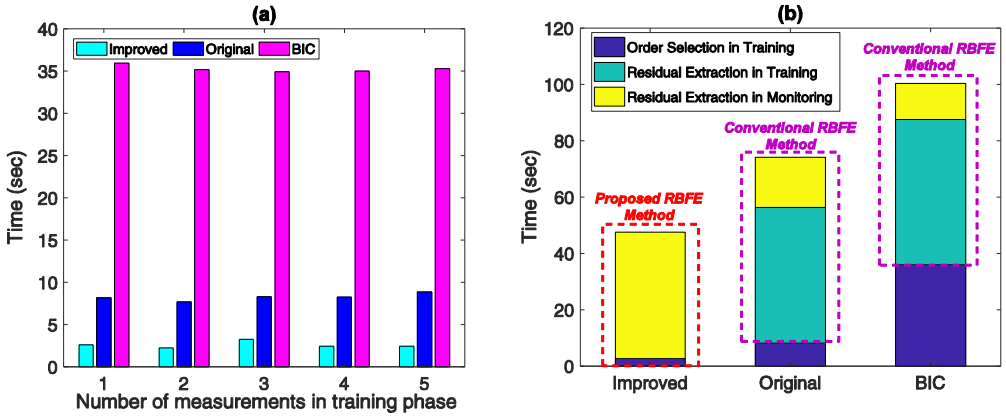

Fig. 8. Comparison among the method-dependent computational times relevant to (a) AR order selection, and (b) complete feature extraction process (RBFE: Residual-Based Feature Extraction)

Fig. 8 clearly points out that the proposed method is more efficient than the other techniques, as far as the AR order selection is concerned; the other way around, it is less efficient than the others as for the residual extraction in the monitoring phase. Anyway, the proposed method generally provides a faster feature extraction since it does not feature a further parameter estimation step in the training phase, which is shown to be relatively time-consuming for the conventional approach. As the process of order selection has the highest influence on the overall computational time of feature extraction, the novel iterative algorithm thus represents a time-saving unsupervised learning strategy for feature extraction.

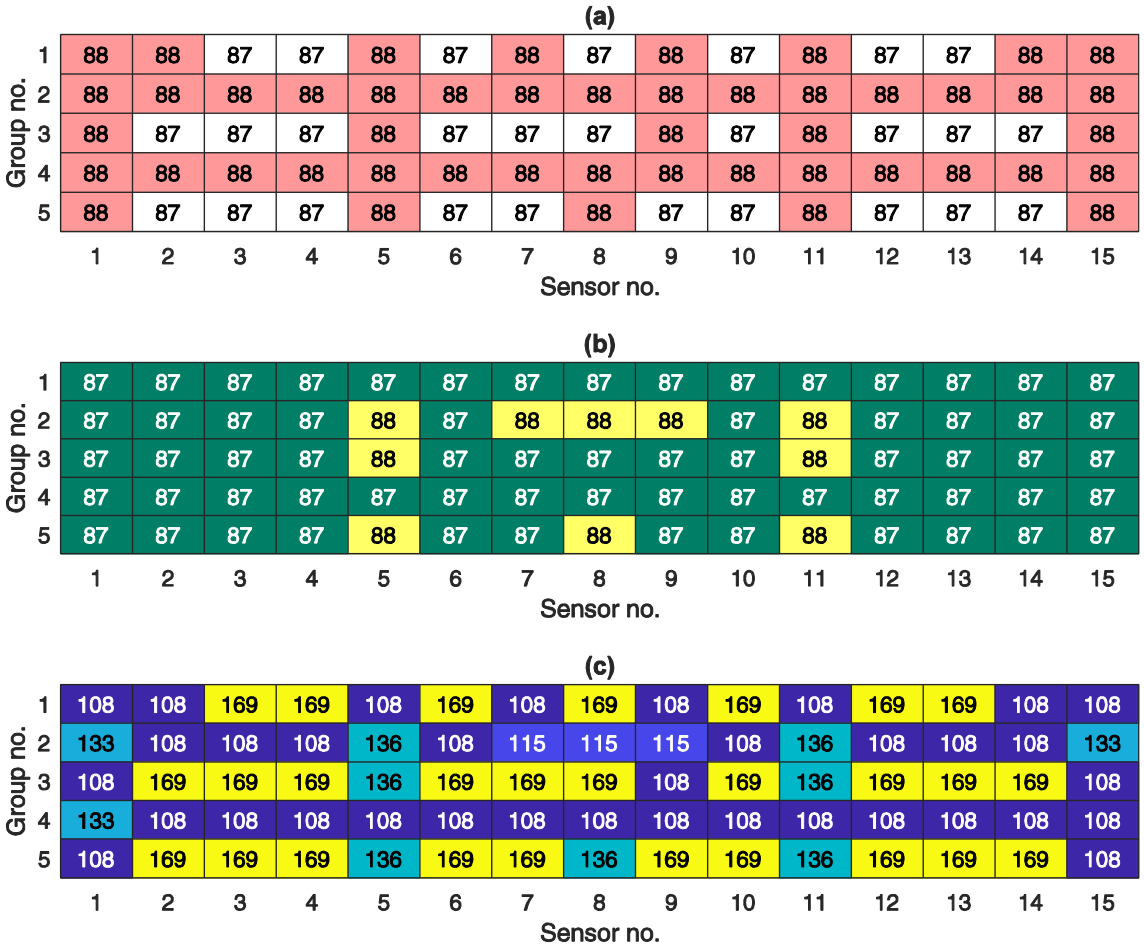

Fig. 9. Case-dependent segmentation of the residual vectors at all sensor locations: (a) number of segments; (b) number of samples of $\mathbf{E}_{\mathbf{y}_r}$ in the first m-1 segments; (c) number of samples of $\mathbf{E}_{\mathbf{y}_r}$ in the last segment

By adopting the model order and coefficients obtained in the training phase, the residual vectors $\mathbf{E}_{\mathbf{x}_k}$ and $\mathbf{E}_{\mathbf{y}_r}$ at each sensor are extracted and respectively used as the main DSFs for the healthy and damaged conditions. The setting of N_y and m on the basis of segmentation depends on the number of the residual sequences, which is equal to $\sqrt{n-k(k+1)/2}$; referring to the example in Fig. 6(a), the order k of the AR model at sensor 5 results to be 20, so that the number of samples within each segment and the number of segments are rounded to N_y=m=88. Fig. 9 collects the results in terms of N_y and m values at all sensor locations; here the group numbering along the vertical axes

refers to the healthy cases reported in Table 1. Finally, also the number N_x of samples related to the residual vectors in the normal conditions have to be set; as an example, Fig. 10 shows exemplary values for the residual vectors relevant to sensors 1-4 and to the first measurement of Case 1.

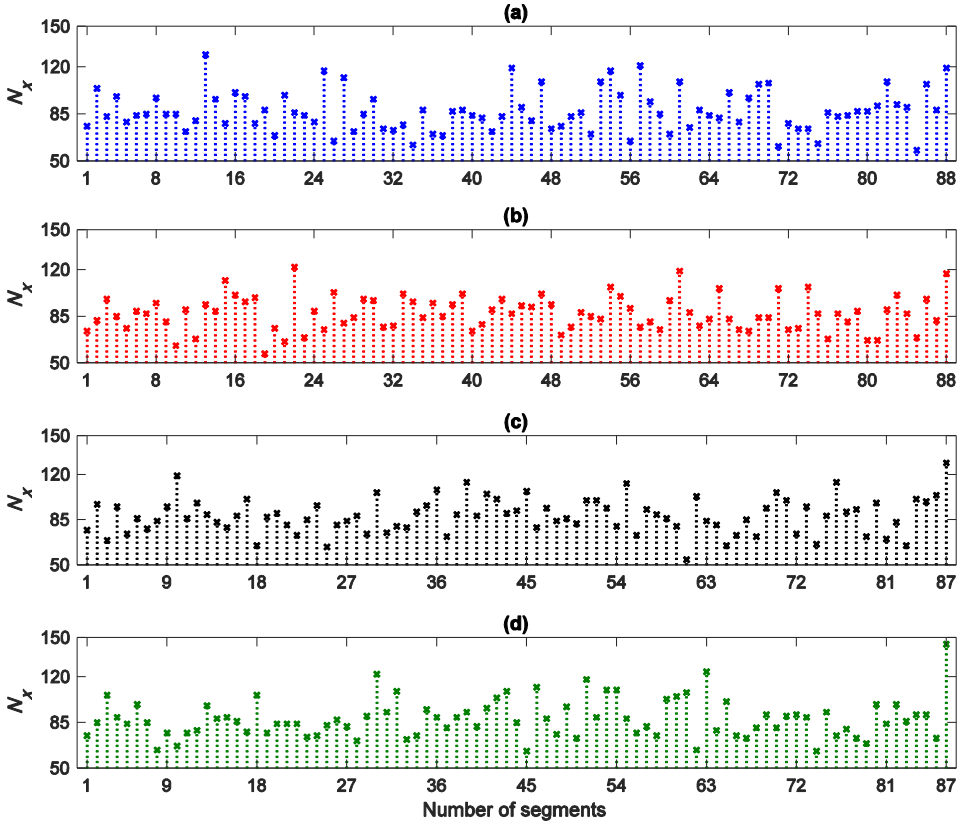

Fig. 10. Number of samples in $\mathbf{E}_{\mathbf{x}_k}$ based on the segmentation of $\mathbf{E}_{\mathbf{y}_r}$ in the first measurement of Case 1 relevant to (a) sensor 1, (b) sensor 2, (c) sensor 3, and (d) sensor 4

This setting is then used in Eqs. (7)-(11), to compute the empirical probability measures for all the segments, and the distance quantities via KLDEPM at all sensors for damage localization. In this example, the single vertical crack shown in Fig. 1 is located near sensor 8, hence the attempt is to assess whether the proposed KLDEPM method is able to identify such location as the damaged area. For comparison purposes, damage localization is performed also by the classical KLD method and by the state-of-the-art

KSTS technique. On the basis of the methodology for threshold estimation described in Section 4.3, the threshold values for the KLDEPM, KLD, and KSTS methods are respectively set as 0.1007, 0.2171, and 0.0361. The sensor location associated with a distance value greater than these threshold levels is accordingly identified as the damaged area: Fig. 11 illustrates the results of damage localization obtained with the three methods, wherein the mean value of the distances between the residual vectors referring to the damaged and to all the normal conditions is used to identify it.

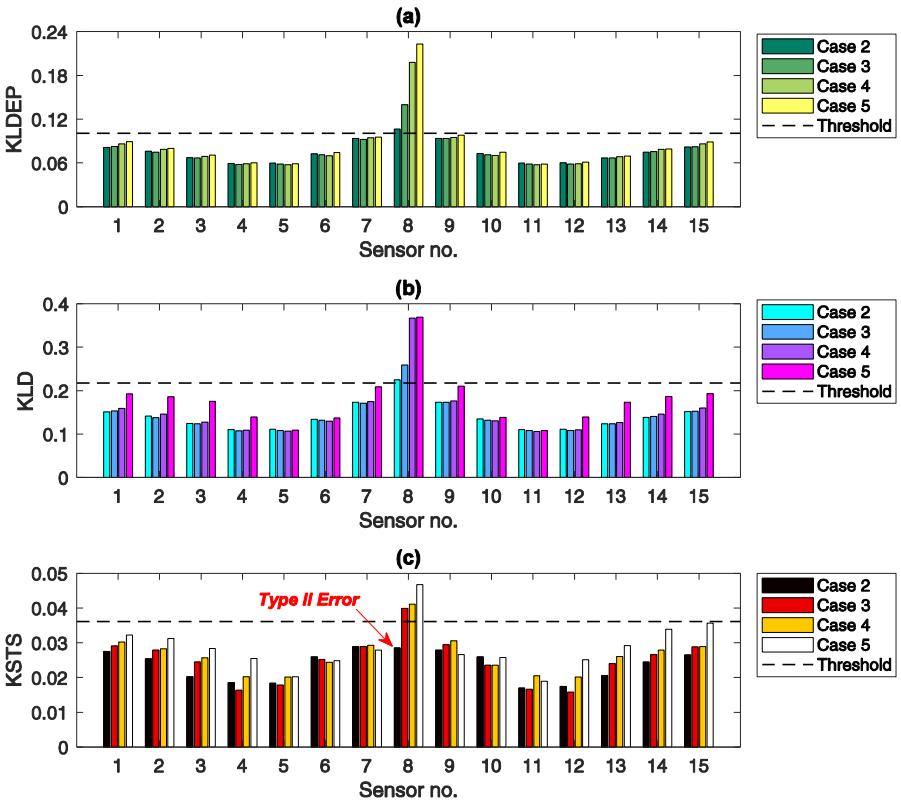

Fig. 11. Damage localization in the numerical beam, as obtained with the (a) KLDEPM, (b) KLD, and (c) KSTS methods

As reported in Fig. 11, all the distance quantities furnished by KLDEPM and KLD exceed the threshold limits at sensor 8, so these methods prove successful in locating the damage at a varying severity level. Although the same outcome is reported by KSTS for

Cases 3-5, it can be seen that this technique fails in identifying the location in case of small damages (as Case 2 corresponds to a 10mm crack length), giving rise to a false-negative damage indication or type II error. Fig. 11 also conveys the important conclusion that one of the main reasons for achieving the accurate damage localization with any distance method (except Case 2 with the KSTS technique) is due to the high sensitivity to damage of the AR model residuals obtained with the proposed feature extraction method.

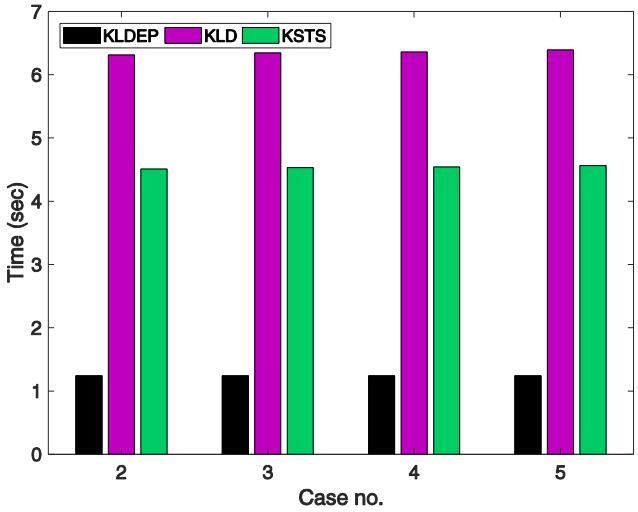

Fig. 12. Comparison among KLDEPM, KLD, and KSTS methods, in terms of the computational time for damage localization in the numerical beam

Besides the issue of damage identification, results related to the computational times of damage localization are compared in Fig. 12. The proposed KLDEPM method is shown to be much faster than the others, and it thus results to provide the best performance also in this regard. The classical KLD technique requires instead the largest computing time, while the KSTS technique is a bit less computational demanding, though it must be borne in mind that it also leads to the erroneous damage localization in Case 2.

Overall, Figs. 11 and 12 lead to the conclusion that the proposed KLDEPM method is able to locate damage by means of a fast unsupervised learning strategy.

5.2. A four-story steel structure

The experimental datasets of the second phase of the IASC-ASCE benchmark [54] are now exploited, to verify the robustness of the proposed method in the presence of multiple damage cases. The benchmark consists of a scaled four-story steel structure, made of a 2-bay-by-2-bay steel frame with a size of 2.5×2.5×3.6m, as shown in Fig. 13(a). The members were hot-rolled grade 300W steel, with a nominal yield strength of 300 MPa. The columns and floor beams respectively featured B100×9 and S75×11 cross-sections. For each bay, the bracing system consisted of two diagonal threaded steel rods with a diameter of 12.7mm. To make the mass distribution realistic, slabs were placed in each bay with a mass of 1000kg at the first, second and third floors, and of 750kg at the fourth floor.

The structure was excited in different ways; in this work, account is taken of the experimental responses obtained with an electro-dynamic shaker mounted on the top level of the structure. Acceleration time histories caused by this excitation were measured by the 15 accelerometers shown in Fig. 13(b), at a frequency of 200Hz; each file of acceleration response thus consists of 24,000 samples. The accelerometers located on the east and west frames measured the accelerations along the strong axis, or north-south (N-S) direction: sensors 1, 4, 7, 10 and 13 were placed on the west side, while sensors 3, 6, 9, 12, and 15 were placed on the east side, with numbering moving from the base and going up to the fourth floor. Accelerometers 2, 5, 8, 11, and 14 were instead placed in the

vicinity of the central columns, again with numbering from the base to the fourth floor, to measure the accelerations along the weak axis or east-west (E-W) direction.

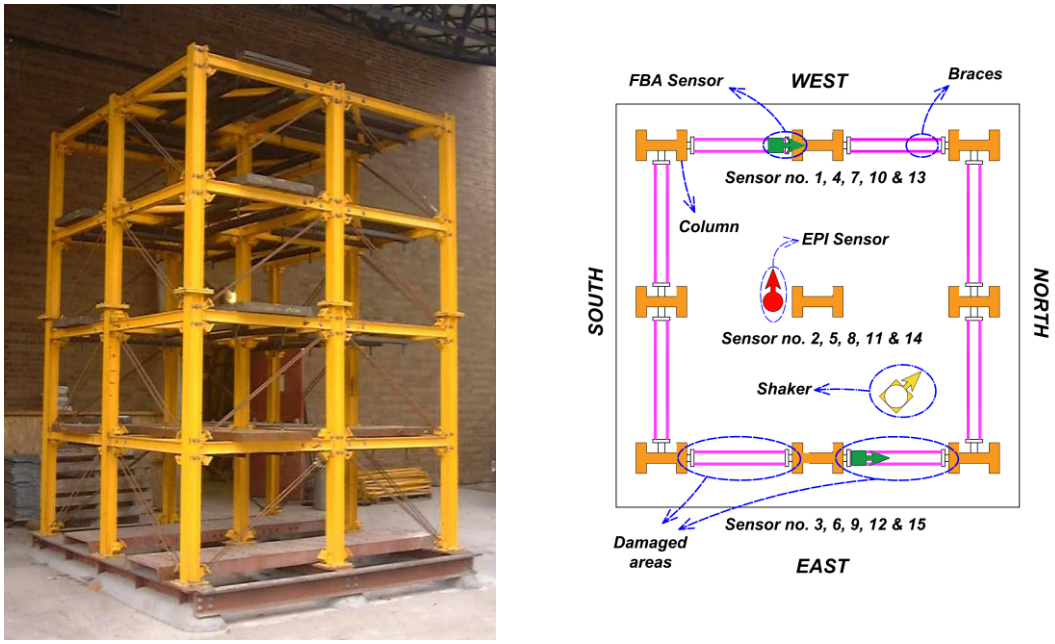

Fig. 13. (a) The four-story steel structure of the IASC-ASCE SHM benchmark problem[54]; (b) details of the structure plan for the base, first, second, third, and fourth floors

Acceleration time histories acquired with the sensors 1-3 mounted on the base are not incorporated in this analysis since they do not provide relevant information concerning the dynamic behavior of the structure. Nine damage scenarios were simulated for this structure, by removing some braces from the east side and southeast corner (first damage pattern) or by loosening bolts at the beam-column connections of the east side (second damage pattern). The five damage scenarios belonging to the aforementioned first damage pattern are considered next, as detailed in Table 2. For example, Fig. 14 shows the excitation force and acceleration responses at the sensor 9 in Cases 1 and 2.

Table 2. Healthy and damaged conditions of the IASC-ASCE structure (Phase II)

Case no.	Label	Description
1	Healthy	Fully braced configuration – No damage

2	Damaged	All braces removed from floors 1-4 at the east side
3	Damaged	All braces removed from floors 1-4 at the southeast corner
4	Damaged	Braces removed from floors 1 and 4 at the southeast corner
5	Damaged	Braces removed from floor 1 at the southeast corner

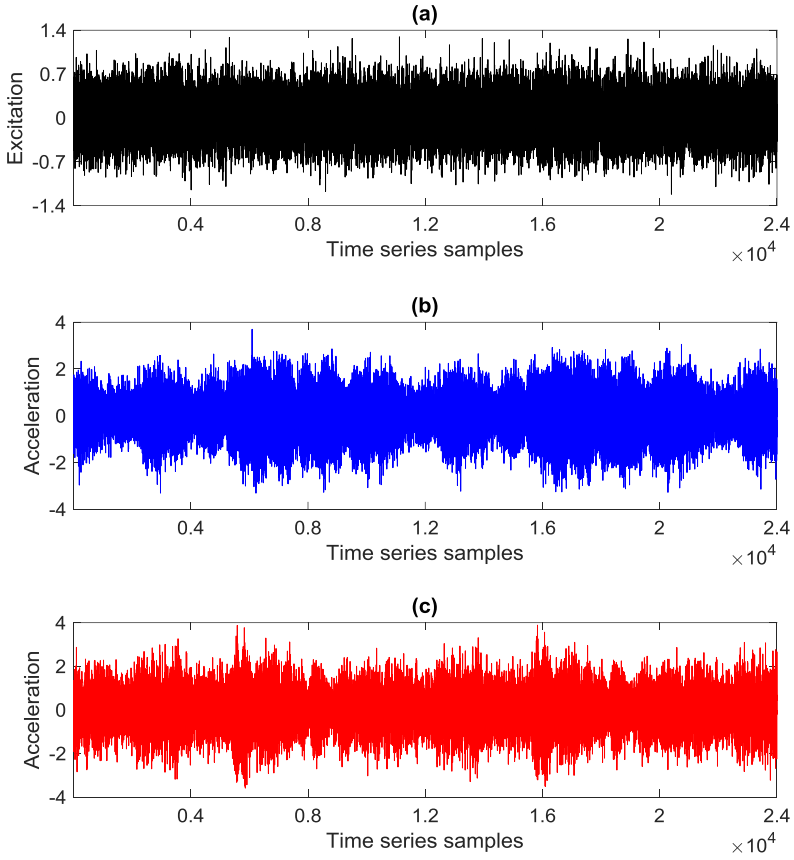

Fig. 14. (a) The excitation force, (b) the acceleration response at the sensor 9 in Case 1, (c) the acceleration response at the sensor 9 in Case 2

First of all, the LMC hypothesis test is employed to assure the accuracy of the adopted AR model for feature extraction. Fig. 15 collects the LMC test statistics at all sensors for Cases 1-5. Using a 5% significance limit, the *c*-value is now equal to 0.1460; the plot shows that all the values of Q_{LM} are smaller than this test *c*-value, so that once again all the measured acceleration responses can be assumed stationary and compatible with an AR process. Furthermore, Fig. 16 shows the plots of the ACF and PACF for the

graphical model identification via the Box-Jenkins methodology for some vibration signals of Case 1. It is seen that the samples of the ACF (the left plots) are similar to sine waves without any trend in being zero. By contrast, one can discern that the samples of the PACF (the right plots) gradually cut off approximately after more than 50 lags. These observations demonstrate that the most suitable time series model between AR and ARMA for the vibration signals is AR. It is worth remarking that the same conclusion can be achieved for the other responses.

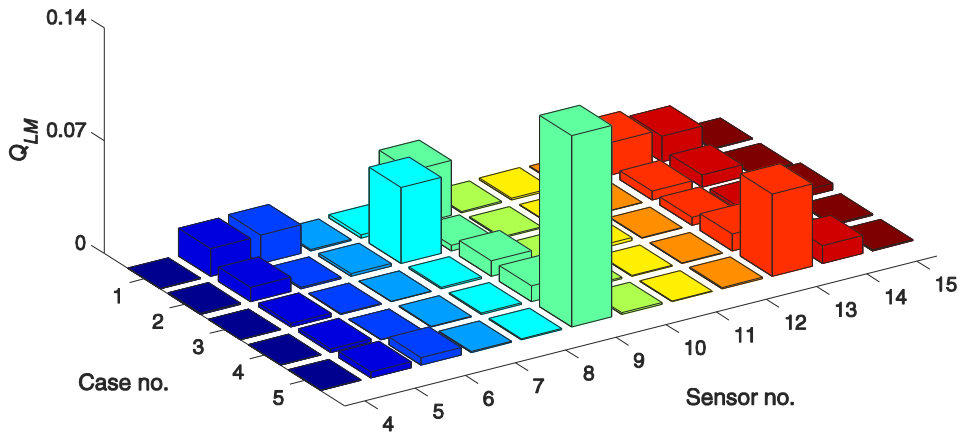

Fig. 15. Stationarity assessment of the AR model identification by the LMC hypothesis test

To handle different normal conditions in the training phase, which look essential for the threshold estimation, the acceleration responses associated with Case 1 are randomly contaminated by five different noise levels. In this way, five different acceleration datasets for the healthy state of the IASC-ASCE structure are built, and they can be considered equivalent to five normal conditions or five test measurements for Case 1 in the training stage (N_c=5). Fig. 17 gathers the orders of all the AR models furnished by the improved and original iterative methods as well as by the BIC technique, handling the mentioned five measurements for Case 1. The results show that the improved method

gives rise to orders smaller than the other techniques also in this analysis. Overall, the BIC technique turns out to be always the worst one in terms of performance.

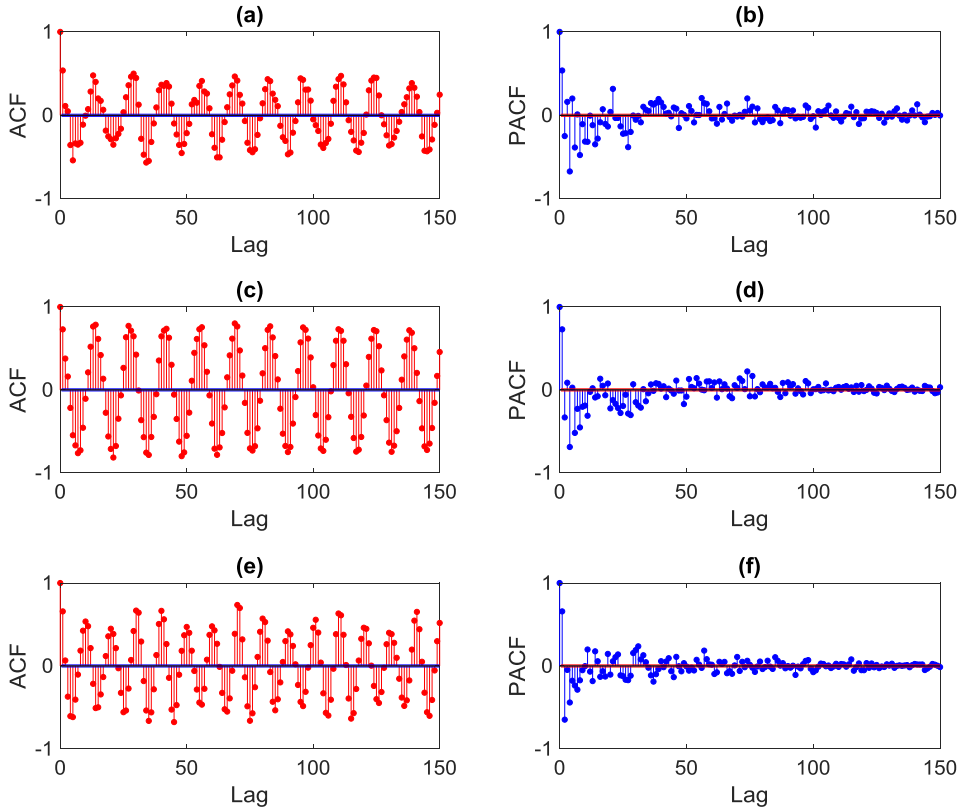

Fig. 16. Graphical model identification between the AR and ARMA representations by the Box-Jenkins methodology in Case 1: (a) ACF at Sensor 6, (b) PACF at Sensor 6, (c) ACF at Sensor 12, (d) PACF at Sensor 12, (e) ACF at Sensor 14, (f) PACF at Sensor 14

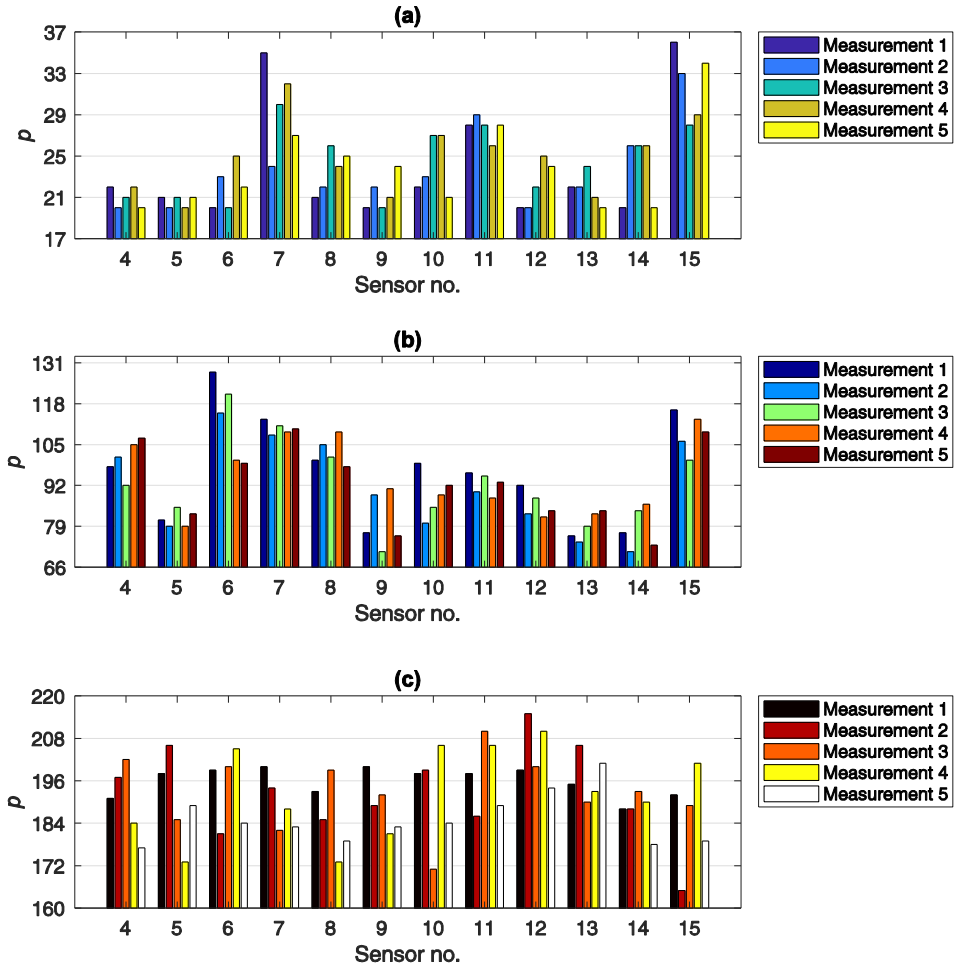

Fig. 17. AR model orders for all the five measurements relevant to Case 1: outcomes (a) of the proposed method, (b) of the original method, (c) and of the BIC method

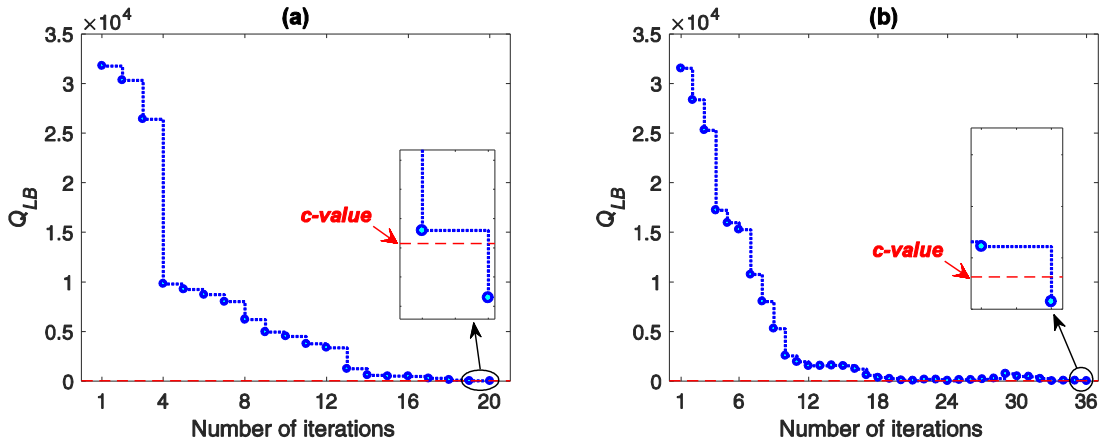

Fig. 18. LBQ test statistics for the first measurement of Case 1: (a) Sensor 9, (b) Sensor 15

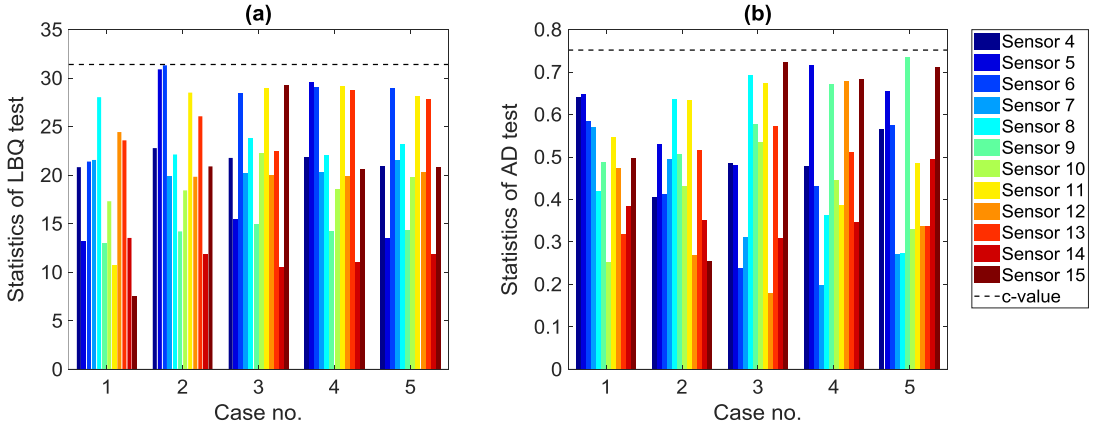

Fig. 19. Analysis of the residual samples of the AR models in Cases 1-5 of the IASC-ASCE structure: (a) assessment of the residual uncorrelatedness by the LBQ test, (b) assessment of the residual Gaussianity by the AD test

The proposed order selection method is able to guarantee model accuracy with uncorrelated residuals: Fig. 18 depicts the LBQ test statistics for the two lowest and highest AR orders, respectively obtained with the first measurement at sensors 9 and 15. In the two cases, the Q_{LB} value at iterations 20 and 36 becomes smaller than the *c*-value threshold, which amounts to 31.4104 for a 5% significance limit; similar conclusions are arrived at for all the other locations and measurements. Moreover, Fig. 19 shows the statistics of the LBQ and AD tests for analyzing the uncorrelatedness and Gaussianity of the residual samples regarding all sensors in Cases 1-5. Once again, it is seen that all statistics are smaller than the tests' *c*-values, which approve that the extracted residual samples are uncorrelated and Gaussian.

Figs. 20 and 21 illustrate the computational time relevant to the AR order selection, and to the complete process of residual-based feature extraction. In Fig. 20, the vibration responses from all the sensors in the five measurements of Case 1 are adopted to compute the time needed for the determination of the AR orders. As for Fig. 21, the

acceleration datasets of Case 2 and the first measurement of Case 1 only are handled to set the computational time of residual extraction in the training and monitoring phases.

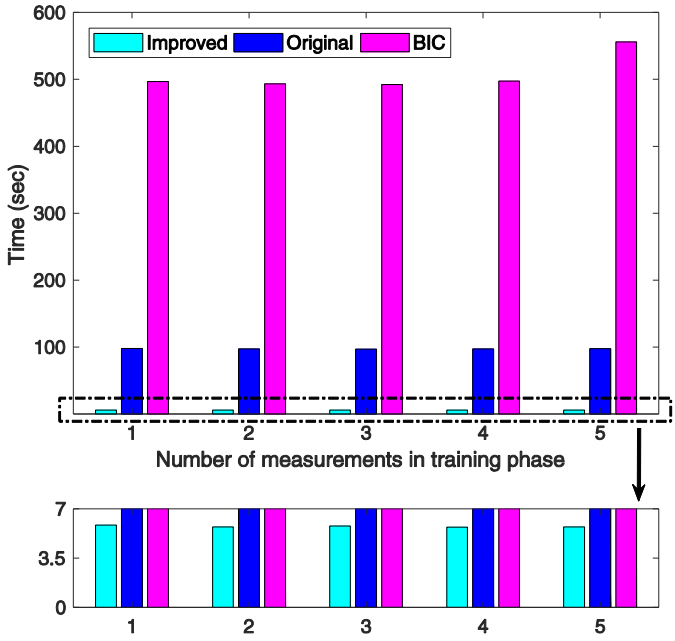

Fig. 20. Computational times for AR order selection in the IASC-ASCE structure, as obtained with the improved, original and BIC methods

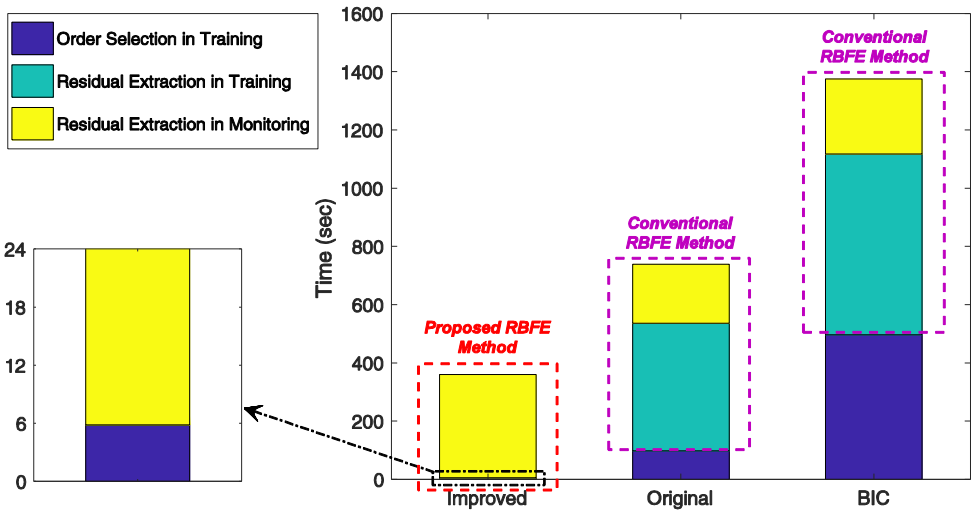

Fig. 21. Computational times for the complete process of feature extraction in the IASC-ASCE structure (RBFE, Residual-Based Feature Extraction): comparison among the performances of the improved, original and BIC methods

Results show that the improved order selection method needs very little computational effort if compared to the other approaches. The new residual-based feature extraction method provides also a better performance against the conventional approach. As already shown for the cracked beam, the step of residual extraction in the monitoring stage requires more time than with the conventional technique, due to the repeated iterative process for residual extraction. This may be a problem for SHM applications because the feature extraction in the training phase can be practically performed only for one time in advance, while during the monitoring phase, the feature extraction requires to be fast implemented for more than one time as long as the damage detection is required. Nevertheless, it should be mentioned that the improved method possesses several advantages such as determining a small and adequate order for the AR model, simultaneously obtaining the model order, coefficients, and residuals in the training phase, and providing a total short time for a complete feature extraction, all of which

enable the improved method to be a time-saving unsupervised learning strategy for feature extraction.

Once the residuals of the AR model at each sensor have been extracted for the undamaged and damaged conditions, their sensitivity to damage is assessed. To this aim, the L_2-norms of the residual vectors $\mathbf{E}_{\mathbf{x}_k}$ and $\mathbf{E}_{\mathbf{y}_r}$ at all sensors are computed, and the results are collected in Fig. 22; here, the samples in the range 1-60 belong to the measurements during training, while the subsequent samples 61-108 are related to monitoring. As the L_2-norm values relevant to the damaged conditions highly increase in comparison with the values referring to the normal conditions, a confirmation is provided of the sensitivity to damage of the residuals extracted with the proposed feature extraction method.

Figs. 23 and 24 show details of the segmentation of the residual sequences for Cases 2-5, in terms of the number of segments, and the number of samples within all the m segments. In the following, the number of samples (N_x) of $\mathbf{E}_{\mathbf{x}_k}$ is determined as shown in Fig. 24 for some exemplary sensor locations. It is to note that the black areas in Fig. 23 are related to the information provided by sensors 1-3 mounted on the base of the structure, which as mentioned earlier has not been considered in the procedures for feature extraction and damage localization.

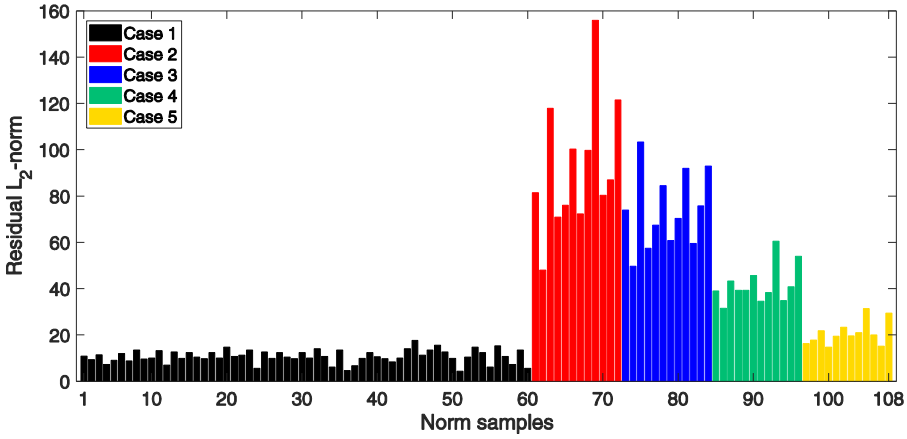

Fig. 22. L_2-norm of the residuals in the training (1-60) and monitoring (61-108) phases

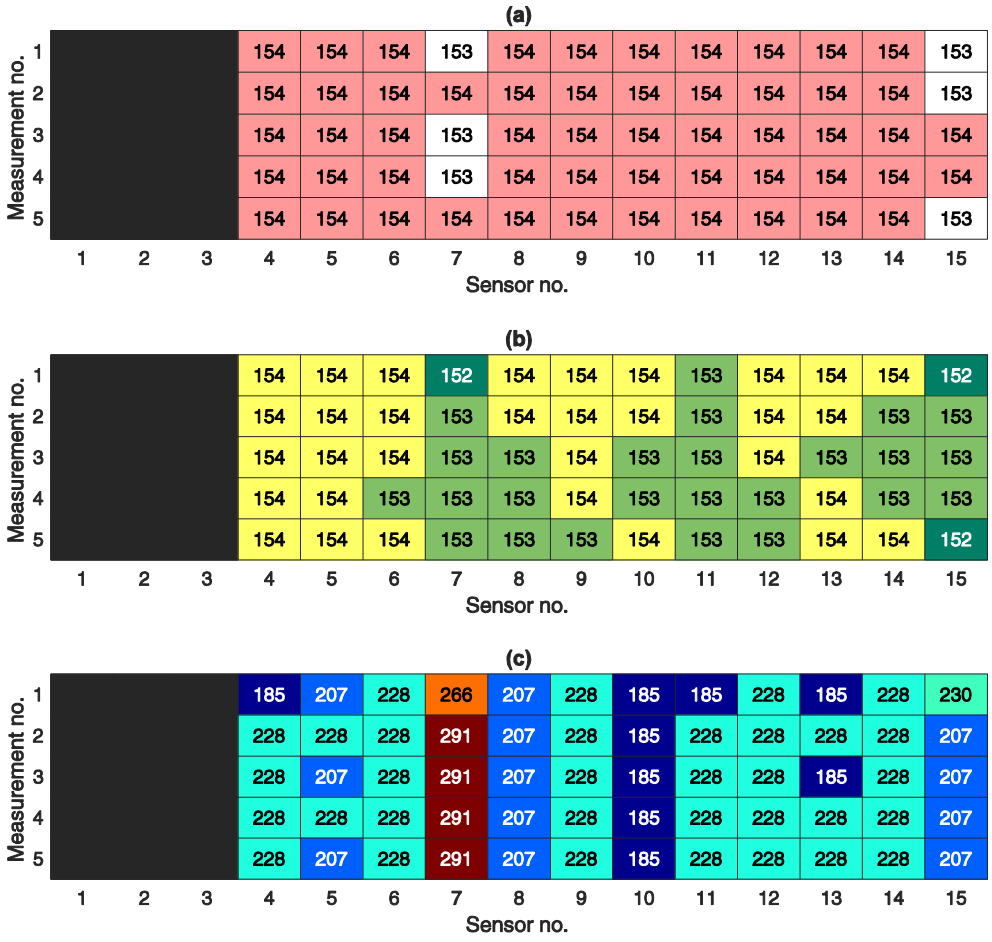

Fig. 23. Measurement-dependent segmentation of the residual vectors at all sensor locations: (a) number of segments, (b) number of samples of $\mathbf{E}_{\mathbf{y}_r}$ in the first m-1 segments, (c) number of samples of $\mathbf{E}_{\mathbf{y}_r}$ in the m-th segment

The KLDEPM method is adopted to locate damage, providing the results reported in Fig. 25; for comparison, Figs. 26 and 27 collect the same results, as obtained with the KLD and KSTS techniques. In these graphs, thresholds respectively amount to 0.0234, 0.3577, and 0.0461 for the KLDEPM, KLD, and KSTS methods. As before, the mean distance values between the residual vectors relevant to each damaged state and to the normal conditions are computed, to provide the results in terms of damage identification.

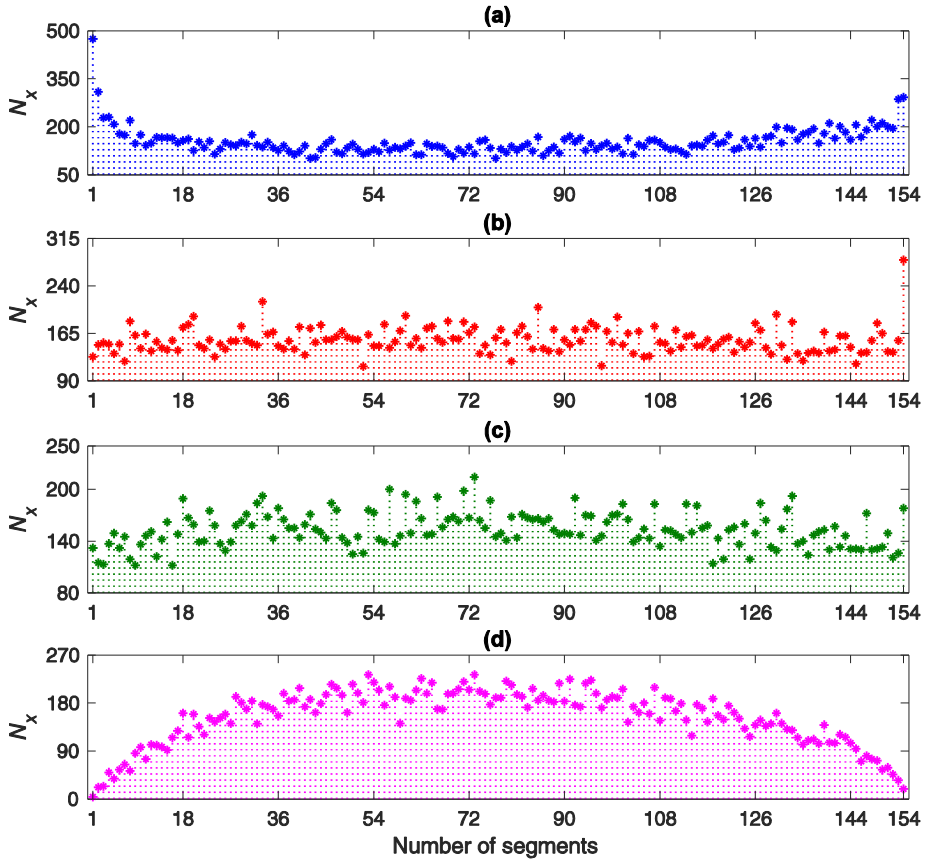

Fig. 24. Number of samples in $\mathbf{E}_{\mathbf{x}_k}$ based on the segmentation of $\mathbf{E}_{\mathbf{y}_r}$, for the first measurement of Case 1 at (a) sensor 4, (b) sensor 8, (c) sensor 11, and (d) sensor 14

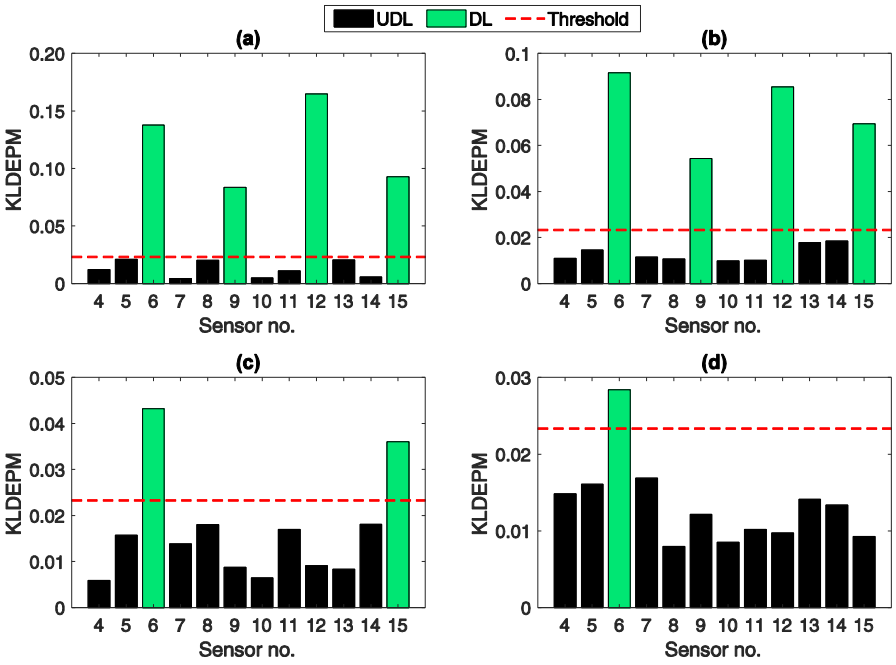

Fig. 25. Damage localization in the IASC-ASCE structure by the proposed KLDEPM method: (a) Case 2, (b) Case 3, (c) Case 4, and (d) Case 5

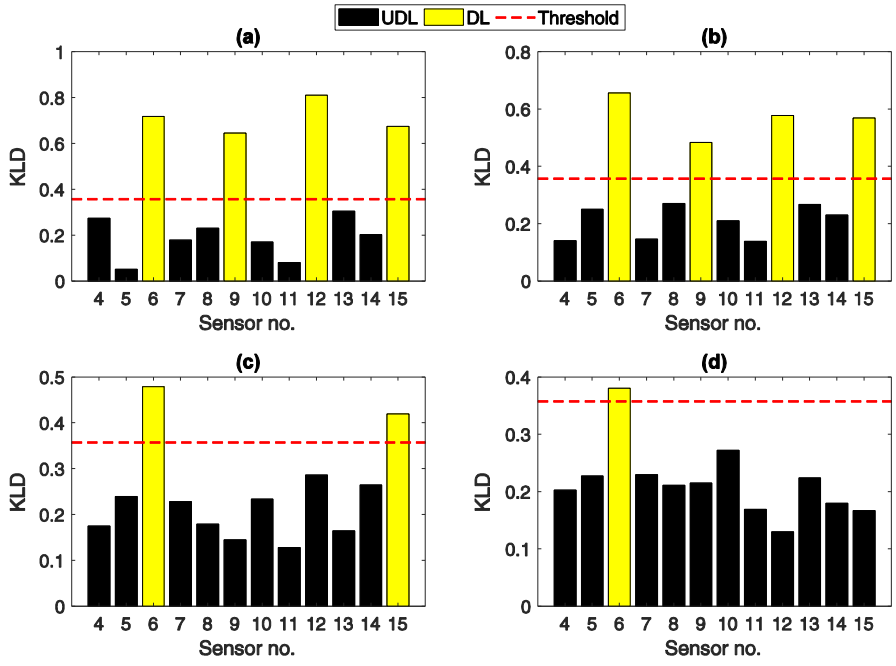

Fig. 26. Damage localization in the IASC-ASCE structure by the classical KLD technique: (a) Case 2, (b) Case 3, (c) Case 4, and (d) Case 5

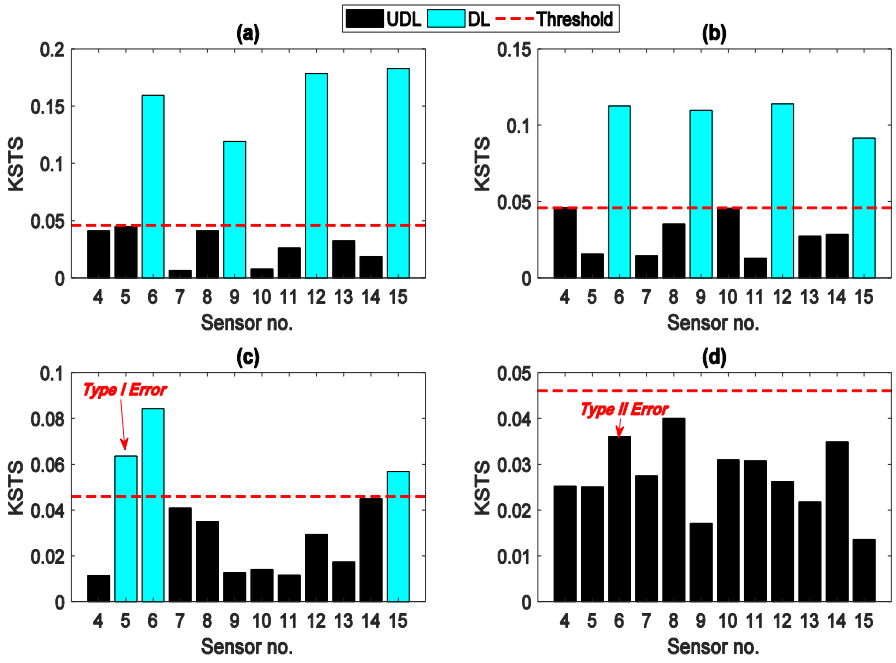

Fig. 27. Damage localization in the IASC-ASCE structure by the KSTS technique: (a) Case 2, (b) Case 3, (c) Case 4, and (d) Case 5

Regarding Cases 2 and 3, Figs. 25(a)-(b), 26(a)-(b) and 27(a)-(b) show that the damage locations (DLs) are correctly identified by the three methods at the east side of floors 1-4, since the distance values at sensors 6, 9, 12, and 15 exceed the threshold. For Cases 4 and 5, in Figs. 25(c)-(d) and 26(c)-(d) it is seen that both the KLDEPM and KLD methods succeed in locating the damaged areas at the southeast corner of either floor 1 or 4. These outcomes are related to distance values exceeding the threshold at the sensors 6 and 15 for Case 4, and at the sensor 6 for Case 5. Fig. 27(c) shows for Case 4 that KSTS can also identify the damage locations at the sensors 6 and 15, but a false-positive damage indication (Type I error) is observable at sensor 5, which is instead an undamaged location (UDL). This technique is also incapable of identifying the damaged area in Case 5, due to the reported false-negative indication (Type II error) at sensor 6 shown in Fig. 27(d). Furthermore, Fig. 28 compares the time required for damage identification via the three

distance approaches: the best performance is provided again by the KLDEPM method. Thus, also in this case it emerges as a fast decision-making process within an unsupervised learning strategy for damage localization.

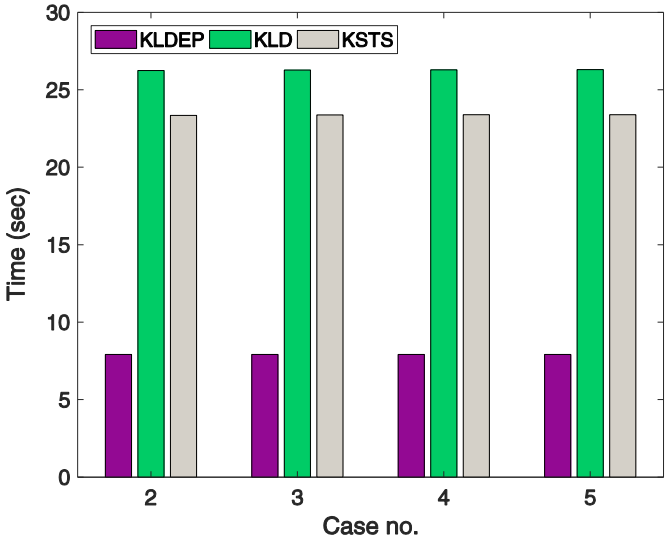

Fig. 28. Comparison among the KLDEPM, KLD, and KSTS methods, in terms of the computational time for damage localization in the IASC-ASCE structure

In most cases of SHM problems, the environmental and operational variability conditions are inevitable[50]. To indicate the effects of these conditions on the procedures of feature extraction and damage localization, it is assumed that Case 5 is an undamaged state with the environmental variability attributable to the reduction in stiffness by removing the only braces on one bay at the first story of the IASC-ASCE structure. Note that this assumption was already used by Entezami and Shariatmadar[37]. In fact, the basis of this assumption comes from the simulation of the environmental variations by reducing the structural stiffness, which was simulated on the laboratory frame of the Los Alamos National Laboratory in the USA[31]. Under such circumstances, the normal conditions are Cases 1 and 5 (N_c=2), which are applied to estimate a threshold limit for damage localization. All the steps of feature extraction (i.e. AR order determination, parameter

estimation, and residual extraction in the training and monitoring phases) are implemented to locate damage in Cases 2-4 as shown in Fig. 29, where the dashed lines are the threshold limit equal to 0.0544. As can be discerned, the same accurate results as Fig. 25 are observable in Fig. 29. This means that the proposed methods succeed in extracting reliable features and accurately locating damage in Cases 2-4 even under the simulated environmental variability.

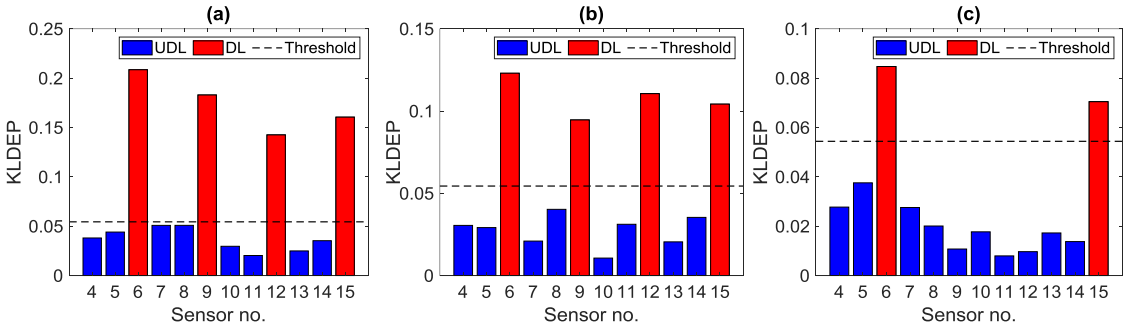

Fig. 29. Damage localization in the IASC-ASCE structure using the proposed KLDEPM method by considering the simulated environmental variability: (a) Case 2, (b) Case 3, (c) Case 4

6. Conclusions

An innovative feature extraction approach, through AR modeling and a statistical distance method termed KLDEPM, has been proposed to locate structural damage based on an unsupervised learning strategy. The proposed method consists of an improved AR order selection, an iterative residual extraction algorithm for the training stage and another iterative residual extraction process for the monitoring phase. The KLDEPM method is based on the theory of empirical probability measure and rests on a segmentation of the residual sequences into independent sets. A numerical concrete beam and the IASC-ASCE benchmark structure have been considered, to assess the accuracy and performance

of the proposed method. Comparative analyses with alternative state-of-the-art procedures have been conducted, to testify the superior performance of the offered method.

The results of feature extraction have shown that the improved order selection method guarantees the accuracy and sufficiency of the AR model by generating uncorrelated residuals, and also provides model orders smaller than those obtained with the alternative techniques. Accordingly, this method results to be less computationally demanding, also because the process of order selection has the highest influence on the total time required for feature extraction. By means of the L_2-norms in the undamaged and damaged conditions, the residuals of the AR model extracted from the proposed feature extraction approach have been confirmed to be highly sensitive to damage.

In the statistical decision-making process for damage localization, results have proven that the KLDEPM method is capable of locating damage of varying severity, particularly for scenarios featuring small damage levels. Since the same conclusion has been obtained with the (less computationally efficient) classical KLD technique, the statistical distance measuring the discrepancy between two datasets by a logarithmic function has shown to be a reliable tool for damage identification. The comparative analysis has also confirmed that KLDEPM is superior in locating small damages, with Type I and Type II errors never observed.

One can then conclude that the proposed feature extraction and KLDEPM methods can be efficient and successful for SHM purposes, particularly regarding damage localization, with the great advantage of working within the frame of an unsupervised learning strategy.

Acknowledgment

The authors are indebted to Prof. Jyrki Kullaa at Aalto University, and to the IASC-ASCE Structural Health Monitoring Task Group for providing the benchmark datasets. This research was supported by Grant No. 96007230 from the Iran National Science Foundation (INSF).

References

1. Cawley P. Structural health monitoring: Closing the gap between research and industrial deployment. *Struct Health Monit*. 2018; 17: 1225-44.

2. Farrar CR and Worden K. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons Ltd, 2013.

3. Kopsaftopoulos F and Fassois S. Vibration based health monitoring for a lightweight truss structure: experimental assessment of several statistical time series methods. *Mech Syst Sig Process*. 2010; 24: 1977-97.

4. Kopsaftopoulos FP and Fassois SD. A vibration model residual-based sequential probability ratio test framework for structural health monitoring. *Struct Health Monit*. 2015; 14: 359-81.

5. Barthorpe RJ. On model-and data-based approaches to structural health monitoring. University of Sheffield, 2010.

6. Sakaris CS, Sakellariou JS and Fassois SD. Vibration-based damage precise localization in three-dimensional structures: Single versus multiple response measurements. *Struct Health Monit*. 2015; 14: 300-14.

7. Mottershead JE, Link M and Friswell MI. The sensitivity method in finite element model updating: A tutorial. *Mech Syst Sig Process*. 2011; 25: 2275-96.

8. Capellari G, Eftekhar Azam S and Mariani S. Damage detection in flexible plates through reduced-order modeling and hybrid particle-Kalman filtering. *Sensors*. 2016; 16: 2.

9. Azam SE, Mariani S and Attari N. Online damage detection via a synergy of proper orthogonal decomposition and recursive Bayesian filters. *Nonlinear Dyn*. 2017; 89: 1489-511.

10. Eftekhar Azam S and Mariani S. Online damage detection in structural systems via dynamic inverse analysis: A recursive Bayesian approach. *Eng Struct*. 2018; 159: 28-45.

11. Capellari G, Chatzi E and Mariani S. Structural Health Monitoring Sensor Network Optimization through Bayesian Experimental Design. *ASCE ASME J Risk Uncertain Eng Syst A Civ Eng*. 2018; 4: 04018016.

12. Sarmadi H, Karamodin A and Entezami A. A new iterative model updating technique based on least squares minimal residual method using measured modal data. *Appl Math Modell*. 2016; 40: 10323-41.

13. Fan W and Qiao P. Vibration-based damage identification methods: a review and comparative study. *Struct Health Monit*. 2011; 10: 83-111.

14. Meruane V and Heylen W. Damage detection with parallel genetic algorithms and operational modes. *Struct Health Monit*. 2010; 9: 481-96.

15. Entezami A, Shariatmadar H and Sarmadi H. Structural damage detection by a new iterative regularization method and an improved sensitivity function. *J Sound Vib*. 2017; 399: 285-307.

16. Fassois SD and Sakellariou JS. Statistical time series methods for structural health monitoring. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, Ltd, 2009, p. 443-72.

17. Hoell S and Omenzetter P. Optimal selection of autoregressive model coefficients for early damage detectability with an application to wind turbine blades. *Mech Syst Sig Process*. 2016; 70-71: 557-77.

18. Sohn H, Czarnecki JA and Farrar CR. Structural Health Monitoring Using Statistical Process Control. *Journal of Structural Engineering*. 2000; 126: 1356-63.

19. Da Silva S, Gianini Gonsalez C and Lopes Jr V. Adaptive filter feature identification for structural health monitoring in an aeronautical panel. *Struct Health Monit*. 2010; 10: 481-9.

20. Adams DE and Farrar CR. Classifying linear and nonlinear structural damage using frequency domain ARX models. *Structural Health Monitoring*. 2002; 1: 185-201.

21. Entezami A and Shariatmadar H. Damage localization under ambient excitations and non-stationary vibration signals by a new hybrid algorithm for feature extraction and multivariate distance correlation methods. *Struct Health Monit*. 2019; 18: 347-75.

22. Nair KK, Kiremidjian AS and Law KH. Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure. *J Sound Vib*. 2006; 291: 349-68.

23. Xie L and Mita A. An Innovative Substructure Damage Identification Approach for Shear Structures Based on ARMAX Models. *Procedia Eng*. 2017; 188: 119-24.

24. Ay AM and Wang Y. Structural damage identification based on self-fitting ARMAX model and multi-sensor data fusion. *Struct Health Monit*. 2014; 13: 445-60.

25. Sohn H, Allen DW, Worden K and Farrar CR. Statistical damage classification using sequential probability ratio tests. *Struct Health Monit*. 2003; 2: 57-74.

26. Kopsaftopoulos F, Nardari R, Li Y-H and Chang F-K. A stochastic global identification framework for aerospace structures operating under varying flight states. *Mech Syst Sig Process*. 2018; 98: 425-47.

27. Avendaño-Valencia LD, Chatzi EN, Koo KY and Brownjohn JM. Gaussian process time-series models for structures under operational variability. *Frontiers in Built Environment*. 2017; 3: 69.

28. Bogoevska S, Spiridonakos M, Chatzi E, Dumova-Jovanoska E and Höffer R. A data-driven diagnostic framework for wind turbine structures: A holistic approach. *Sensors*. 2017; 17: 720.

29. Shi H, Worden K and Cross EJ. A cointegration approach for heteroscedastic data based on a time series decomposition: An application to structural health monitoring. *Mech Syst Sig Process*. 2019; 120: 16-31.

30. Yao R and Pakzad SN. Autoregressive statistical pattern recognition algorithms for damage detection in civil structures. *Mech Syst Sig Process*. 2012; 31: 355-68.

31. Entezami A and Shariatmadar H. Structural health monitoring by a new hybrid feature extraction and dynamic time warping methods under ambient vibration and non-stationary signals. *Meas*. 2019; 134: 548-68.

32. Figueiredo E, Figueiras J, Park G, Farrar CR and Worden K. Influence of the Autoregressive Model Order on Damage Detection. *Comput-Aided Civ Infrastruct Eng*. 2011; 26: 225-38.

33. Box GE, Jenkins GM, Reinsel GC and Ljung GM. *Time Series Analysis: Forecasting and Control*. 5th ed.: John Wiley & Sons, 2015.

34. Hyndman RJ and Athanasopoulos G. *Forecasting: principles and practice*. OTexts, 2014.

35. Figueiredo E, Figueiras J, Park G, Farrar CR and Worden K. Influence of the Autoregressive Model Order on Damage Detection. *Comput-Aided Civ In Eng*. 2011; 26: 225-38.

36. Broersen PM. Finite sample criteria for autoregressive order selection. *IEEE Trans Signal Process*. 2000; 48: 3550-8.

37. Vu VH, Thomas M, Lakis A and Marcouiller L. Operational modal analysis by updating autoregressive model. *Mech Syst Sig Process*. 2011; 25: 1028-44.

38. Entezami A and Shariatmadar H. An unsupervised learning approach by novel damage indices in structural health monitoring for damage localization and quantification. *Struct Health Monit*. 2017; 17: 325-45.

39. Rezaiee-Pajand M, Entezami A and Shariatmadar H. An iterative order determination method for time-series modeling in structural health monitoring. *Adv Struct Eng*. 2017; 21: 300-14

40. George RC, Mishra SK and Dwivedi M. Mahalanobis distance among the phase portraits as damage feature. *Struct Health Monit*. 2018; 17: 869-87.

41. Yeager M, Gregory B, Key C and Todd M. On using robust Mahalanobis distance estimations for feature discrimination in a damage detection scenario. *Struct Health Monit*. 2019; 18: 245-53.

42. Zheng H and Mita A. Localized damage detection of structures subject to multiple ambient excitations using two distance measures for autoregressive models. *Struct Health Monit*. 2009; 8: 207-22.

43. Eybpoosh M, Berges M and Noh HY. An energy-based sparse representation of ultrasonic guided-waves for online damage detection of pipelines under varying environmental and operational conditions. *Mech Syst Sig Process*. 2017; 82: 260-78.

44. Delpha C, Diallo D and Youssef A. Kullback-Leibler Divergence for fault estimation and isolation : Application to Gamma distributed data. *Mech Syst Sig Process*. 2017; 93: 118-35.

45. Roy K, Bhattacharya B and Ray-Chaudhuri S. ARX model-based damage sensitive features for structural damage localization using output-only measurements. *J Sound Vib*. 2015; 349: 99-122.

46. Zhan Y and Mechefske CK. Robust detection of gearbox deterioration using compromised autoregressive modeling and Kolmogorov–Smirnov test statistic—Part I: Compromised autoregressive modeling with the aid of hypothesis tests and simulation analysis. *Mech Syst Sig Process*. 2007; 21: 1953-82.

47. Zhen D, Wang T, Gu F and Ball AD. Fault diagnosis of motor drives using stator current signal analysis based on dynamic time warping. *Mech Syst Sig Process*. 2013; 34: 191-202.

48. Mosavi AA, Dickey D, Seracino R and Rizkalla S. Identifying damage locations under ambient vibrations utilizing vector autoregressive models and Mahalanobis distances. *Mech Syst Sig Process*. 2012; 26: 254-67.

49. Mujica L, Rodellar J, Fernandez A and Güemes A. Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. *Struct Health Monit*. 2011; 10: 539-53.

50. Liu G, Mao Z, Todd M and Huang Z. Damage assessment with state–space embedding strategy and singular value decomposition under stochastic excitation. *Struct Health Monit*. 2014; 13: 131-42.

51. Entezami A, Shariatmadar H and Karamodin A. Data-driven damage diagnosis under environmental and operational variability by novel statistical pattern recognition methods. *Struct Health Monit*. 2018; 0.

52. Figueiredo E, Park G, Farrar CR, Worden K and Figueiras J. Machine learning algorithms for damage detection under operational and environmental variability. *Struct Health Monit*. 2011; 10: 559-72.

53. Kullaa J, Santaoja K and Eymery A. Vibration-based structural health monitoring of a simulated beam with a breathing crack. *Key Engineering Materials*. Trans Tech Publ, 2013, p. 1093-100.

54. Dyke SJ, Bernal D, Beck J and Ventura C. Experimental phase II of the structural health monitoring benchmark problem. *Proceedings of the 16th ASCE engineering mechanics conference*. 2003.

55. Sohn H, Czarnecki JA and Farrar CR. Structural Health Monitoring Using Statistical Process Control. *J Struct Eng*. 2000; 126: 1356-63.

56. Kullback S and Leibler RA. On information and sufficiency. *Ann Math Stat*. 1951; 22: 79-86.

57. Wang Q, Kulkarni SR and Verdú S. Divergence estimation of continuous distributions based on data-dependent partitions. *IEEE Trans Inf Theory*. 2005; 51: 3064-74.

58. Dassault Systemes Simulia Crop. ABAQUS/CAE V6.12. 2012.

59. Leybourne SJ and McCabe BP. A consistent test for a unit root. *J Bus Econ Stat*. 1994; 12: 157-66.

60. Anderson TW and Darling DA. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *The annals of mathematical statistics*. 1952; 23: 193-212.