



Time series-based SHM using PCA with application to ASCE benchmark structure

Kundan Kumar¹ · Prabir Kumar Biswas² · Nirjhar Dhang³

Received: 26 January 2020 / Accepted: 16 July 2020 / Published online: 26 July 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Detecting damage at an early stage can avoid a serious catastrophic failure of structures due to inevitable cause, such as fatigue, environmental corrosion, and natural disasters. Various damage detection algorithms have been proposed based on autoregressive model using time series data, which are computationally expensive, and the selection of an optimal order of the model requires extra expertise. In this paper, computationally efficient algorithm is proposed to process the time series data using principal component analysis (PCA) in an effective way. PCA is utilized to model a feature space to compute damage sensitive features insensitive to environmental variations and measurement noise. The modeled feature space preserves the damage information along with eliminates the consequences of environmental variations and measurement noise. Furthermore, Mahalanobis squared distance is adopted to compute damage index as the severity of the damage. The proposed method is validated on analytical models of IASC–ASCE benchmark structure. The test results show that the proposed damage diagnosis method can be useful for wireless sensor network-based structural health monitoring with less computation and low data transmission rate.

Keywords Structural health monitoring · Damage detection · Eigenspace · Principal component analysis · Mahalanobis squared distance · Outlier detection

1 Introduction

Structural health monitoring (SHM) of aerospace, civil, and mechanical structures has become one of the fastest-growing multi-disciplinary research problems among the research community [1–3]. Usually, these structures operate under excessive loads (for example, large variation in temperature,

humidity, dynamic loads due to traffic and wind, etc.), and are often exposed to an open environment resulting in wear and tear due to corrosion, flood, earthquake, etc. [4, 5]. Collectively, these factors affect the predicted life span of structures or lead to catastrophic failures [6]. To avoid such catastrophic failures, continuous health monitoring of structures is very much demanded [7]. In that context, there exist numerous monitoring approaches that include a wide range of data interpretation techniques [8–10]. In SHM, data interpretation are performed mainly using two approaches: model-based and data-driven approaches. While model-based approach uses accurate finite element model [7, 11, 12], which is a time-consuming approach, the data-driven approach uses time series data analysis that allows continuous monitoring of structures in real time for damage diagnosis, and, hence, is considered as an effective method [13, 14].

In the data-driven approach, time series data sequences (vibration signals) are measured through an array of sensors and are transmitted mainly through wired [15] or wireless networks [16]. The shortcomings of wired sensor networks such as cable installation, source of power supply, maintenance and repair, etc., especially for remotely placed

✉ Kundan Kumar
erkundanec@gmail.com

Prabir Kumar Biswas
pkb@ece.iitkgp.ac.in

Nirjhar Dhang
nirjhar@civil.iitkgp.ac.in

¹ Department of Electronics and Communication Engineering, Institute of Technical Education and Research, S'O'A Deemed to be University, Bhubaneswar 751030, India

² Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721302, India

³ Department of Civil Engineering, Indian Institute of Technology, Kharagpur 721302, India

structures [17], can be mitigated using wireless sensor network-based SHM [16, 18]. Accordingly, various time series-based SHM systems using wireless sensor network have been proposed [2, 19, 20]. However, many of them use complex algorithms, which could be time-consuming. Moreover, an optimal wireless sensor network-based SHM system needs to be energy efficient with low data transmission rate [18]. Undeniably, the amount of data transmission through the network can be minimized by employing either the feature extraction or data compression at the sensor node itself [21–23].

The feature extraction is a process of identifying the underlying patterns, known as damage sensitive feature (DSF), in a measured vibration signal. In the previous studies, several DSFs have been proposed based on Fourier transform [24], wavelet transform [25], Hilbert-Huang transform [26], Auto-associative Neural Network [27], Autoregressive (AR) model [9], etc. Among them, AR model [19, 28, 29] and its alternatives, such as autoregressive with exogenous input (ARX) [30], autoregressive with moving average (ARMA) [31, 32], and autoregressive integrated moving average (ARIMA) [4, 22] models, are widely adopted models for DSFs extraction in SHM. Besides, data compression using principal component analysis (PCA) [33] has been employed on time series data to obtain orthogonal modes [also called principal components (PCs)], which could be used to compute DSFs. Moreover, Yan et al. [5] and Pozo et al. [34] have shown that DSFs can be made insensitive to different environmental conditions by discarding first few PCs from the analysis. It has been observed that the computation of orthogonal modes are computationally expensive if the data streams have significantly higher lengths [35]. Literature suggests that the corresponding computational burden can be reduced by computing orthogonal modes from the windowed data streams of time series data [8, 36]. Accordingly, Posenato et al. [37] have found promising results using moving PCA based on windowed data streams, in comparison to AR model-based and wavelet-based approaches for damage diagnosis. Furthermore, Cavadas et al. [8] have used a similar approach for the early detection of structural damage. However, these works did not consider operational and environmental variations in the damage diagnosis process. Additionally, their application in a wireless sensor network-based SHM is limited, because it processes all the sensor data in a single windowed matrix [16].

Given that, here, we propose a computationally efficient algorithm using PCA to extract the DSFs, which are insensitive to environmental variations and measurement noise. In addition, we compute the damage severity of the test structure by calculating the Mahalanobis squared distance (MSD) [38] of the test data stream from the feature distribution corresponding to undamaged data streams.

Furthermore, we validate the proposed DSFs using the ASCE benchmark structure (Johnson et al. [39]). Accordingly, the algorithm addresses the issues of (a) varying environmental conditions, (b) low energy efficiency, and (c) poor data transmission. Besides, our algorithm is computationally efficient for DSFs extraction, which can have application in wireless sensor network-based continuous health monitoring. Specifically, here, we present:

- *An efficient algorithm for DSFs extraction:* While most of the PCA-based algorithms given in the literature are computationally expensive, because they compute eigenvectors from the covariance matrix of full size [5, 8], the algorithm presented in the paper processes a comparatively smaller covariance matrix and, thereby, makes it efficient.
- *Local sensor-level analysis:* Various reports suggest the successful implementation of PCA for global sensor-level analysis [8, 37]; however, local sensor-level analysis using PCA is lacking. Accordingly, local sensor-level analysis proposed in this work could possibly pave the path for wireless sensor network-based SHM.
- *DSFs insensitive to environmental variations:* Reported approaches to make DSFs insensitive to environmental variations suggest that a number of PCs equal to the number of environmental factors present during monitoring need to be discarded [5, 34] from the analysis. Therefore, the computation by the reported algorithms is biased by the prior information of the environmental factors. Contrastingly, here, we introduce an automated approach to decide the number of PCs to be discarded that is based on cumulative percentage of variance (CPV) criterion [40], and hence, the presented methodology is independent of any prior information of the environmental factors.

Above all, the proposed feature space model can compute the DSFs corresponding to test data streams at the sensing node itself by just projecting them onto the feature space using an embedded algorithm. Accordingly, a significant data compression at the sensor node is possible that could result in a noteworthy reduction of the amount of data transmission.

To discuss the details of the proposed algorithm and results in an orderly manner, we organize the paper as follows. The mathematical background of PCA is discussed in Sect. 2. Section 3 describes the proposed damage diagnosis algorithm. A novel DSF extraction method is described in Sect. 3.2. Section 4 describes the analytical model structure of ASCE benchmark problem for the validation of the proposed algorithm. The results are reported in Sect. 5. Finally, work is concluded in Sect. 6.

2 Principal component analysis background

The principal component analysis (PCA) is a statistical procedure that provides an efficient way to find a lower dimensional space to simplify the structure and reveals the hidden information [33]. Computationally, PCA calculates the eigenvectors and eigenvalues of the covariance of a data matrix.

Let $X = [x_1 \ x_2 \ \dots \ x_n]$ be an $m \times n$ dimensional data matrix, where x_k is a column vector with m number of data points ($m \gg n$). Here, m defines the number of observations and n defines the number variables. The covariance matrix of mean centered data matrix X can be defined as:

$$\Sigma = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})(x_k - \bar{x})^T = \frac{1}{n-1} XX^T,$$

where Σ is a square matrix with the variance of individual measurements in the main diagonal, and the covariance between the measurement types in off-diagonals. The matrix equation for eigenvalue decomposition can be written as:

$$\Sigma V = V\Lambda, \quad (1)$$

which is equivalent to write as:

$$\Sigma = V\Lambda V^T. \quad (2)$$

The columns of matrix $V = [v_1 \ v_2 \ \dots \ v_m]$ are the eigenvectors; and the diagonal elements in the matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ are the eigenvalues of the covariance matrix Σ . The eigenvectors are orthonormal vectors that satisfy the orthonormality property as in Eq. (3):

$$\forall v_j, v_k \in V, \langle v_j, v_k \rangle = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \quad (3)$$

There will be m number of eigenvalue and eigenvector pairs as the dimension of the covariance matrix Σ is $m \times m$. Among these, most of the eigenvalues are nearly equal to zero; therefore, the corresponding eigenvectors contain redundant information. Moreover, the eigenvalue decomposition of the matrix XX^T is computationally expensive as $m \gg n$. Therefore, the smaller matrix $X^T X$ of size $n \times n$ can be used instead of XX^T to compute the eigenvalue and eigenvector pairs. It can be shown that the eigenvalues of $X^T X$ are also the eigenvalues of XX^T , whereas corresponding eigenvectors are strongly related [41].

Suppose u_k and λ_k are the k th eigenvector and eigenvalue pair of $X^T X$; therefore, we can write as:

$$X^T X u_k = \lambda_k u_k. \quad (4)$$

Multiply Eq. (4) both side by X , we get:

$$XX^T (X u_k) = X \lambda_k u_k = \lambda_k (X u_k). \quad (5)$$

We can observe from Eq. (5) that as long as $X u_k$ is a non-zero vector; it will be an eigenvector of XX^T , and λ_k will be the eigenvalue of XX^T as well as $X^T X$. On this account, first, we compute the eigenvectors of the smaller matrix $X^T X$ of dimension $n \times n$, then we calculate $v_k = X u_k$ as the k th eigenvector of XX^T . The set of eigenvectors, corresponding to the covariance matrix XX^T , can be obtained as:

$$V = XU, \quad (6)$$

where $U = [u_1 \ u_2 \ \dots \ u_n]$ is the eigenvector matrix of the covariance matrix $X^T X$. It can also be verified that:

$$\lambda_{n+1} = \lambda_{n+2} = \dots = \lambda_m \approx 0. \quad (7)$$

Because the eigenvectors of the covariance matrix XX^T are the principal components (PCs) of the data matrix X , we have used eigenvectors and PCs interchangeably in the paper.

3 Proposed damage diagnosis algorithm

The detailed steps of the damage diagnosis algorithm are discussed in the following sections.

3.1 Preprocessing

Suppose $x^i(t)$ be the vibration signal measured from i th sensor placed on the structure to be monitored, where $i = 1, 2, \dots, N$. N is the number of sensors placed on the monitoring structure. The signal is partitioned into n number of data streams $x_j^i(t)$, where i and j are the indices for the sensor number and stream number, respectively. The partition of the signal into a number of data streams can be viewed as non-overlapped windowed data streams [36]. Next, each data streams are smoothed using weighted moving average filter (WMAF) [42] to remove the random noise, but at the same time, retain the sharp step responses. Binomial weighted average and exponential weighted average filters are the good choices in which filtered output retains sharp step responses. In this work, binomial WMAF is used which follow the binomial expansion of $\left[\frac{1}{2} \ \frac{1}{2}\right]^n$. The binomial filter coefficients are obtained by convolving $\left[\frac{1}{2} \ \frac{1}{2}\right]$ with itself, and the output is convolved with $\left[\frac{1}{2} \ \frac{1}{2}\right]$ for 5 times iteratively. Each data point is produced in filtered output after weighted averaging of data points from the input data stream. In literature, the WMAF is the fastest digital filter for smoothing; however, multiple passes of the moving average are comparatively slower but still faster than Gaussian

and Blackman filters [42]. A typical signal stream and its filtered output are shown in Fig. 1.

Furthermore, filtered data streams are standardized to zero mean and unit standard deviation, because PCA on non-standardized variables may lead to higher eigenvalues for the variables with higher variance, resulting in principal components (PCs) may dependence on the variables having higher variance.

3.2 Feature space model

The preprocessed data streams referring to an undamaged structure are arranged in a matrix X^i to create a feature space model for i th sensor location:

$$X^i = \begin{bmatrix} x_{11}^i & x_{12}^i & \dots & x_{1n}^i \\ x_{21}^i & x_{22}^i & \dots & x_{2n}^i \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}^i & x_{m2}^i & \dots & x_{mn}^i \end{bmatrix} \quad (8)$$

The flow diagram of the feature extraction process is shown in Fig. 2. The covariance of the data matrix X^i is calculated using Eq. (9):

$$\Sigma^i = \frac{1}{m-1} X^{iT} X^i \quad (9)$$

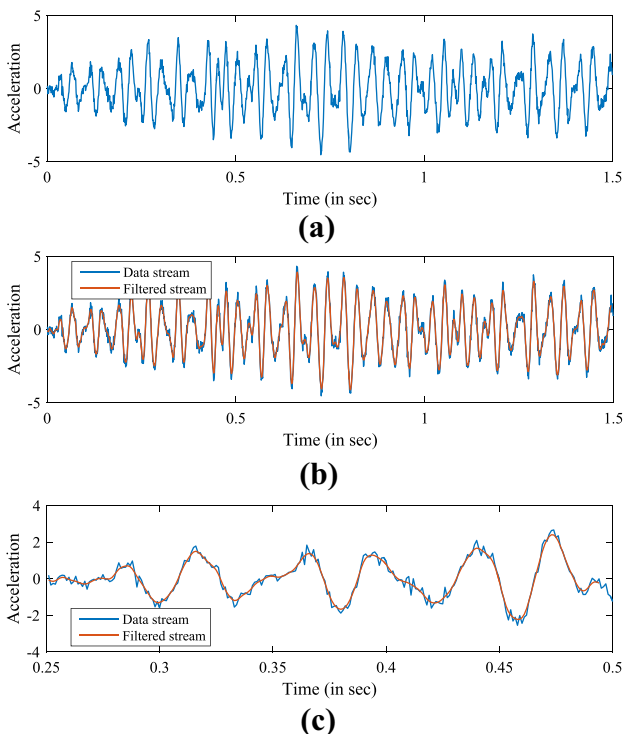


Fig. 1 A typical time-series data stream: **a** undamaged data stream, **b** filtered data stream, and **c** part of the filtered data stream

The eigenvalue and eigenvector pairs of the covariance matrix Σ^i are obtained using PCA which satisfy the principal decomposition as:

$$\Sigma^i = U^i \Lambda^i U^{iT} \quad (10)$$

where $U^i = [u_1 \ u_2 \ \dots \ u_n]$ is the eigenvector matrix having each column as eigenvector, and the diagonal elements of Λ^i are eigenvalues arranged in descending order. The principal vectors for the feature space are computed as:

$$V^i = X^i U^i \quad (11)$$

where $V^i = [v_1 \ v_2 \ \dots \ v_n]$. Here, all the principal vectors do not contain the useful information in the feature space. Yan et al. [5] have suggested that various environmental factors have a strong influence on the vibration features mainly along the direction of the principal components (PCs) associated with the higher eigenvalues. They have computed residual error by re-mapping the projected data back to the original space by discarding operational and environmental factor characterized space. However, Abolhassani et al. [43] discarded PCs corresponding to smaller eigenvalues to improve the signal-to-noise ratio (SNR) in the speech processing domain. These two notions are integrated to make the DSF more robust to change in environmental conditions and measurement noise. Therefore, we, here, assume that the eigenvectors corresponding to higher eigenvalues are affected due to change in varying environmental loads. However, eigenvectors corresponding to lower eigenvalues are affected due to redundant information and measurement noise. Accordingly, mid-range eigenvectors are selected to construct the feature space model. The reduced size feature space model is determined as:

$$V_{red}^i = [v_{a+1}, v_{a+2}, \dots, v_b] \quad (12)$$

where the parameters a and b define the range of eigenvalues that need to be chosen carefully. The first a number of eigenvectors corresponding to higher eigenvalues represents the perturbation in the data samples; however, eigenvector corresponding to b smallest eigenvalues contains redundant information and measurement noise.

To choose the appropriate value of the parameter a and b , cumulative percentage of variance (CPV) criterion is introduced. The CPV is the statistical property of the PCA which is the cumulative sum of the percentage of total variance (PTV). The PTV accounted by the k th eigenvector is obtained as $\frac{\lambda_k}{\sum_{j=1}^n \lambda_j} \times 100$. However, CPV accumulated by the first ℓ PCs is defined as:

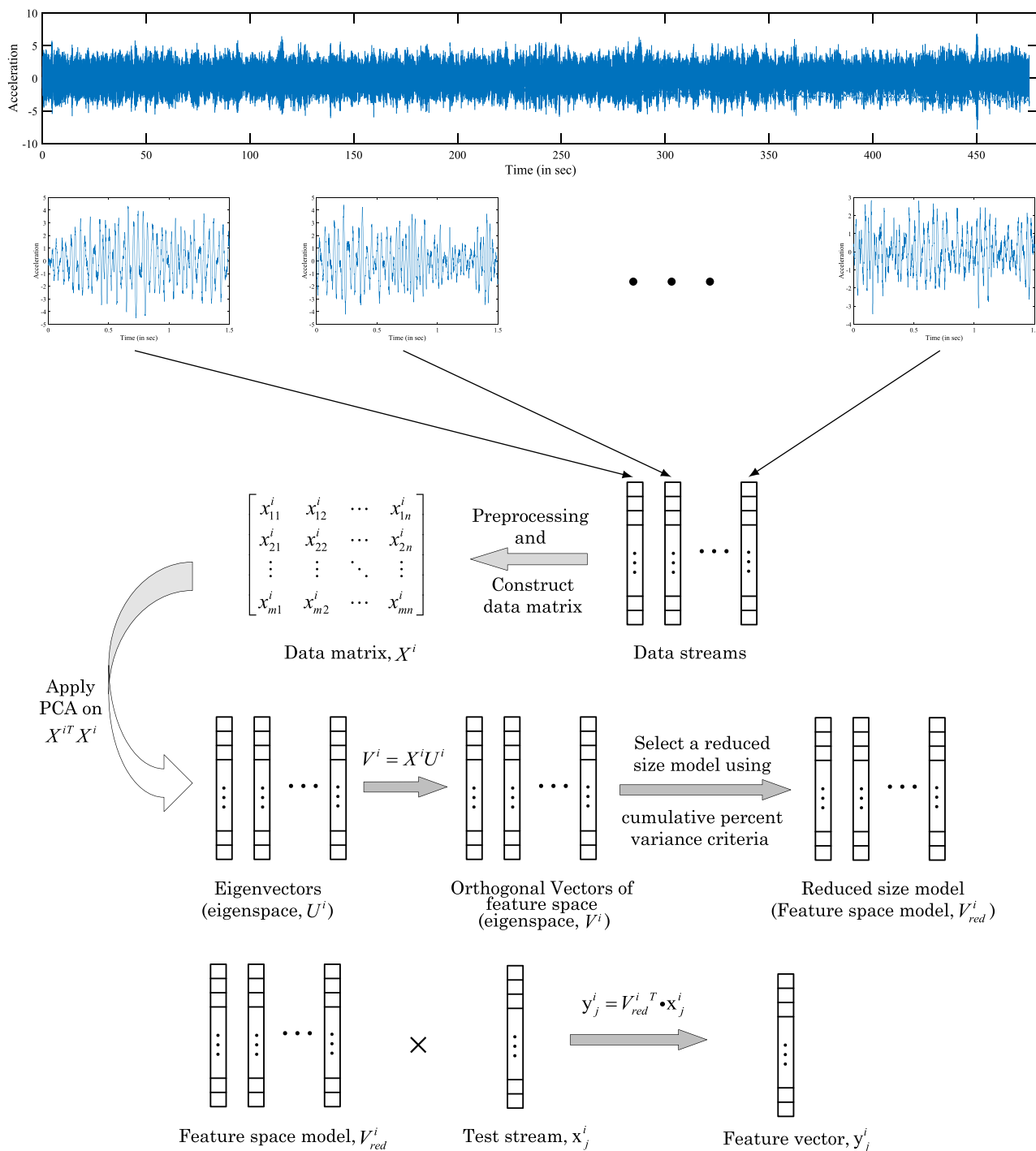


Fig. 2 Feature extraction process

$$p_\ell = \left[\frac{\sum_{j=1}^{\ell} \lambda_j}{\sum_{j=1}^n \lambda_j} \right] \times 100. \tag{13}$$

The PTV and CPV for a typical undamaged data matrix are plotted in Fig. 3, where p_a and p_b are CPV corresponding to a th and b th PCs. In the reduced size feature space model, DSF corresponding to data stream x_j^i is obtained using Eq. (14):

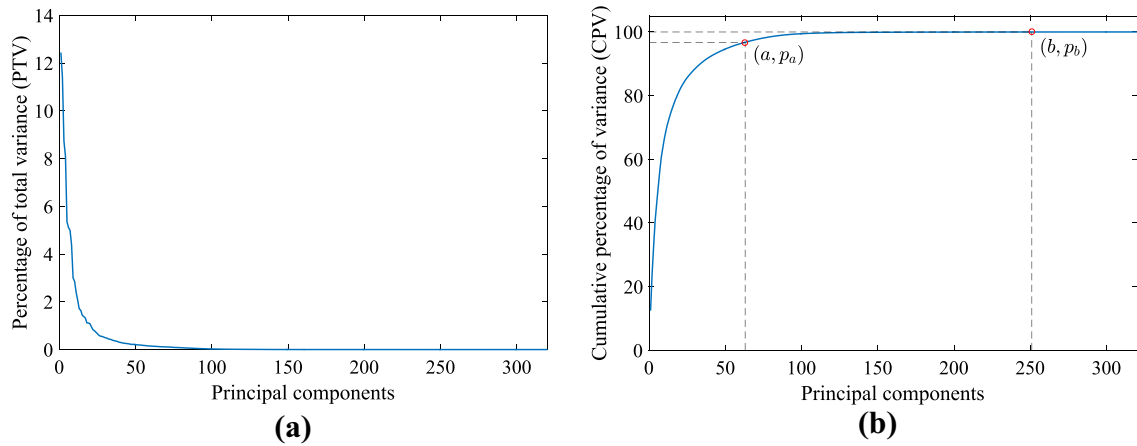


Fig. 3 Percentage of total variance and their cumulative sum for a typical data matrix corresponding to undamaged structure

$$y_j^i = V_{red}^{iT} X_j^i, \tag{14}$$

where y_j^i is the DSF vector insensitive to varying environmental conditions and measurement noise. In the same manner, we compute the DSFs for all the undamaged data streams and store them to derive a model to compute damage severity.

3.3 Damage indices

The extracted DSFs are modeled using Mahalanobis squared distance (MSD) based approach to determine the severity of damage in the structure. Mahalanobis squared distance (MSD) has been used as a distance measure for multivariate statistics outlier detection [38]. This distance measure is widely adopted for outlier detection, because it computes the distance from the sample population of similar characteristics, and it also needs less computational effort [9]. Mahalanobis distance of a test feature vector y_{test}^i from a distribution of baseline feature vectors at i th sensor location is defined as in Eq. (15):

$$d_{test}^i = (y_{test}^i - \bar{y}_i)^T \Sigma_y^{i-1} (y_{test}^i - \bar{y}_i), \tag{15}$$

where Σ_y^i is the covariance matrix of baseline features (serve as sample population of similar characteristics) and \bar{y}_i is mean feature vector corresponding to i th sensor location. The Mahalanobis distance d_{test}^i gives the damage index (DI) corresponding to i th sensor location. However, global DI, DI_{test} , is obtained by accumulating all DIs corresponding to all sensor as:

$$DI_{test} = \sum_{i=1}^N \log_{10}(d_{test}^i). \tag{16}$$

3.4 Decision-making

To make the decision, whether the structure is damaged or undamaged, statistical process control (SPC) technique is applied on DIs of the healthy data streams. The SPC technique provides a framework to monitor the change in mean and variance of DIs correspond to a new measurement. We, here, adopted the X-bar control chart-based SPC to decide the threshold and make the decision system automated. The preliminary criteria to apply the control chart technique is that DIs are to be normally distributed [44]. Yan et al. [5] have defined the three-sigma limits, upper control limit (UCL), center limit (CL), and lower control limit (LCL), as given in Eq. (17), in terms of mean \bar{DI} and standard deviation σ of the DIs corresponding to undamaged structure:

$$\begin{aligned} CL &= \bar{DI}, \\ UCL &= \bar{DI} + 3\sigma, \text{ and} \\ LCL &= \bar{DI} - 3\sigma. \end{aligned} \tag{17}$$

To classify the undamaged and damaged structural conditions, UCL is regarded as the threshold.

4 Application to ASCE benchmark structure

Various damage detection methodologies have been proposed for different structures. Wherever, the side-by-side comparison is very difficult. Johnson et al. [39] have provided a common platform for consistent evaluation of different methods using IASC–ASCE benchmark problem in SHM. For this reason, we have chosen ASCE benchmark structure to validate our proposed algorithm.

The potential of the proposed algorithm is validated on the analytical model of the benchmark structure which is a four-story 2-bay by 2-bay steel braced frame, as shown in Fig. 4. The structure is 3.6 m tall with a 2.5 m × 2.5 m base located at the Earthquake Engineering Research Laboratory of the University of British Columbia (UBC). IASC–ASCE task group [45] has provided a MATLAB computer code to simulate various analytical models. Vibration responses in the form of acceleration signal are simulated for the following analytical models:

- 12-DOF (symmetric), load on each floor (Ambient vibration)
- 120-DOF (symmetric), load on each floor (Ambient vibration)
- 120-DOF (unsymmetric), roof is excited diagonally (Shaker vibration).

The 12-DOF and 120-DOF models are used to consider the model error; however, symmetric and unsymmetric mass

distribution on each floor is taken into account to simulate uneven mass distribution of real structure. Ambient loads on all stories are used to excite the structure which is considered as the wind load to simulate the ambient vibrations. However, the diagonal load on the roof is used to excite the structure to simulate the shaker vibration. The 12-DOF model is a shear-building model that constrains all motion except two horizontal translations and one rotation per floor. On the other hand, all the floor nodes in the 120-DOF model have the same horizontal translation and in-plane rotation. The acceleration signals were measured from 16 different sensor locations ($N = 16$) (see Fig. 4) for structural health diagnosis. Four sensors were placed on each floor, where two sensors measured acceleration in the x -direction (strong) and other two measured in the y -direction (weak). Additionally, 10% of noise is added to all simulated acceleration signals to introduce randomness in the data. Damage was introduced in the structure by means of removing braces from the different stories of the frame. Table 1 presents the different damage patterns, used to validate the proposed algorithm. For more details about the

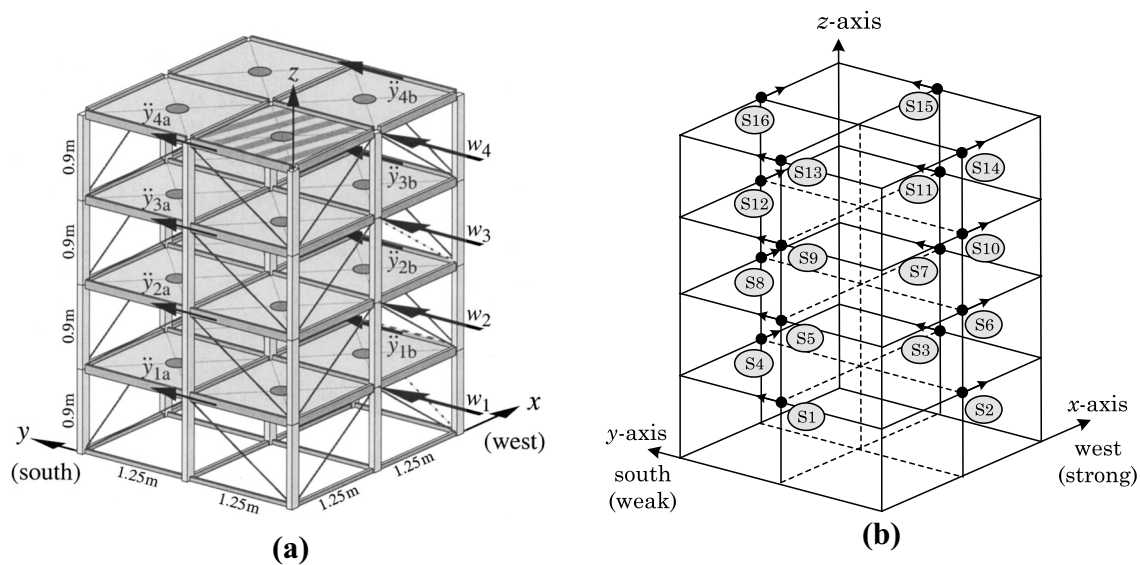


Fig. 4 ASCE Benchmark structure: **a** dimensions (Johnson et al. [39]), and **b** sensor locations (S1–S16) and the direction of the acceleration measurements for analytical model

Table 1 Numerically simulated test cases [39]

	Configuration
Case 1	No damage: fully braced configuration
Case 2	Pattern 1: removal of all braces on the 1st floor
Case 3	Pattern 2: removal of all braces on the 1st and 3rd floors
Case 4	Pattern 3: removal of one brace on the 1st floor
Case 5	Pattern 4: removal of one brace on the 1st and 3rd floors
Case 6	Pattern 5: Pattern 4 + beam-column connection weakened at 1st floor
Case 7	Pattern 6: area of one brace is reduced to 2/3 on the 1st floor

damage patterns, one can refer to the research paper of Johnson et al. [39].

5 Results and discussion

The proposed algorithm is validated using time series data collected from analytical model of IASC–ASCE benchmark structure. Both ambient and shaker vibrations are used to analyze the capability of the proposed DSF in different damage scenarios and excitation.

5.1 Selection of model parameters m, n

Before validating the proposed algorithm, in this section, selection of the model parameters m and n is discussed that how to choose them for SHM applications. In the context of this work, choosing model parameters m is similar to choosing window size in case of moving PCA proposed by Cavadas et al. [8]. Here, m should be sufficiently large enough, so that it should not be influenced by measurement noise and small enough to be stationary in nature. Stationary test of the undamaged data streams with different length is presented in Table 2.

Another parameter n is the number of operational and environmental variability corresponding to similar structural condition (i.e., undamaged). This may also depend on the number of data points collected for time-history signals in some applications as adopted in [8]. Our observation is that n should be sufficiently large, so that it can accommodate various possible operational and environmental conditions.

5.2 Numerical validation

The test cases tabulated in Table 1 are simulated for 480 s at the sampling rate of 1000 Hz, resulting in 480,000 data points in each time history signals. Collectively, we have simulated total seven time history signals (i.e., one time series with 480 s of duration for each test case). Among all the test cases, Case 1 (Undamaged pattern) is used to construct the feature space model. To split the time-history signals in a number of data streams, a stationary test is performed (see Sect. 5.1). From Table 2, it was observed that data streams are stationary for minimum length of 1000 data points. In this study, the acceleration signal of Case 1 is divided into a number of streams with each data stream having 1500 data points ($480,000/1500 = 320$ undamaged data streams), i.e., $m = 1500$ and $n = 320$. Each data streams are preprocessed

and standardized to remove the trends present in the signal. Similarly, the acceleration signals corresponding to other test cases, Case 2 – Case 7 (6 cases), are split into a number of data streams followed by preprocessing and standardization. Resulting in total 2240 ($320 + 6 * 320$) data streams are generated for numerical validation of the proposed algorithm.

Furthermore, entire 320 undamaged data streams corresponding to i th sensor are arranged in a data matrix X^i as in Eq. (8). Next, principal component analysis (PCA) is applied on the matrix X^i to obtain the reduced size feature space V_{red}^i . The principal components (PCs) of the matrix X^i are computed by the proposed methodology. Consequently, a total of 320 PCs are obtained. The percentage of total variance (PTV) and the cumulative percentage of variance (CPV) for the Case 1 are plotted in Fig. 3. Based on our assumption, the PCs corresponding to higher and relatively lower eigenvalues are discarded to obtain the feature space model for DSF extraction. Discarding the portion of data stream spanned by the PCs corresponding to lower eigenvalues is equivalent to eliminate higher frequency contents from the data stream. On the other hand, discarding the PCs corresponding to higher eigenvalues removes the perturbations from the data streams. Selecting parameter a for a desired CPV, e.g., 90%, 95%, is very subjective that how much perturbations are allowed in the data streams [40]. However, setting CPV for b in the range of 97.0–99.7% is a good choice to improve the SNR for wide range of SHM applications. In this work, a set of PCs are selected in between a and b , where a and b are the principal component indices corresponding to p_a and p_b . Percentage band selection is not much difficult as it seems. As we know that all the training data streams came from the same healthy structure, and hence, perturbation in the data is due to environmental variations and measurement noise. Therefore, a percentage band between $p_a = 95%$ and $p_b = 99.7%$ has been selected for damage diagnosis. However, a sensitivity analysis of these parameters on the performance is discussed in Sect. 5.4.

The DSFs are extracted from all the data streams corresponding to various damage patterns at each sensor location. It is worth noting that size of the feature vector may not be the same for different sensor location, because the dynamics of the structure in x -direction and y -direction may not be same due to the geometry of the structure. For the ASCE benchmark structure, the structural response strength in the x -direction (weak) and y -direction (strong) is not same due to its geometrical behavior. Total seven damage patterns were investigated for the damage severity quantification.

Table 2 Stationary test to the number of data points for Case 1

	Number of data points in data stream (m)							
	250	500	750	1000	1250	1500	1750	2000
Mean value	0.132	0.031	0.027	0.006	0.004	0.006	0.006	0.006

Mahalanobis squared distance (MSD) of test feature vectors from the distribution of baseline (undamaged) features are computed at each sensor location. The DIs for test data streams are obtained using Eq. (16) and shown in Fig. 5 for three different analytical models of the ASCE benchmark structure.

Figure 5 shows that proposed method assigned a meaningful damage severity to different levels of damage. Interestingly, it can be noticed that proposed feature is sensitive to small damages (Case 7) in the structure. As we closely analyzed the damage patterns, we found that Case 4 has damage in weak direction; however, Case 5 and Case 6 have one damage in weak direction and another in the strong direction. In the ambient vibration case, the structure was excited along weak direction, and the damage in strong direction has very less impact on the structure. Therefore, DI for Case 4, Case 5, and Case 6 for ambient vibration are approximately same (Fig. 5a and b). However, in case of shaker vibration, shaker load was applied diagonally at 45° of *x*-axis which excited the structure equally in both *x*- and *y*-directions. Therefore, a sharp difference between the DIs for Case 4 and Case 5 has been observed (Fig. 5c). In both the scenarios (under ambient and shaker excitation), Case 7

(small damage) is clearly identified. We have also analyzed the average DI of individual damage pattern, where average DI has been computed by averaging 320 damage indices corresponding to the damage pattern. Table 3 presents the average DI of the entire test cases for different analytical models. Note that average DI for Case 5 and Case 6 are exactly the same in the case of 12-DOF model, because the floor was perfectly rigid, and bending of the floor beams was not allowed, unlike 120-DOF model as mentioned by Johnson

Table 3 Average damage severity for each test case corresponding to analytical models

Scenarios	12-DOF sym-metric	120-DOF sym-metric	120-DOF unsym-metric
Case 1	3.356	3.355	3.468
Case 2	4.505	4.521	4.920
Case 3	4.750	4.520	5.050
Case 4	3.804	3.846	3.952
Case 5	3.805	3.850	4.241
Case 6	3.805	3.857	4.244
Case 7	3.515	3.563	3.540

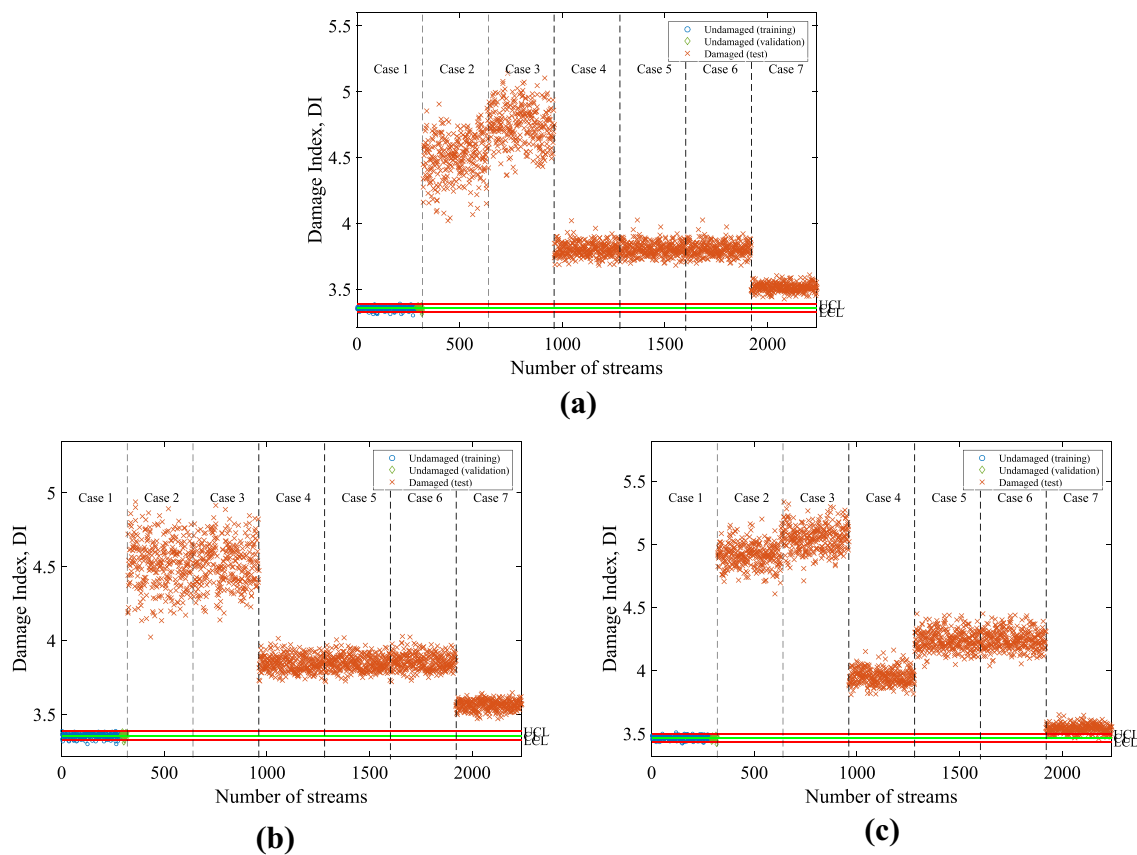


Fig. 5 Damage index for analytical models: **a** 12-DOF symmetrical ambient, **b** 120-DOF symmetrical ambient, and **c** 120-DOF unsymmetrical shaker

et al. [39]. However, for other models, a slight difference between Case 5 and Case 6 has been observed due to the weak beam-column connection. We have also verified that the proposed damage sensitive feature (DSF) is insensitive to a new source of excitation producing much higher acceleration. For that, we have simulated a few more time series data streams for the same damaged cases considering much higher force intensity in the MATLAB computer code. The average DIs are computed for these new time series data streams, and we found that average DIs are insensitive to different force intensities applied on the structure.

For the quantitative analysis of the proposed algorithm, a threshold is decided using undamaged data streams of Case 1. 90% data streams of Case 1 are used for training and rest 10% data streams for validation. X-bar control chart is used to obtain UCL, CL, and LCL. For the classification, UCL is considered as the threshold. If the DI exceeds the threshold, structural conditions are classified as damaged, otherwise undamaged. Classification accuracy for all three analytical models are presented in Table 4. Results show that the proposed DSF is very efficient to detect all types of damages in the structure with high classification accuracy.

For real-time applications, in the proposed framework, the feature space models can be feed to individual sensors placed on the monitoring structure using embedded algorithms. Moreover, as the sensor node will record the vibration signals during continuous monitoring, the test data streams can be trimmed and projected onto the modeled feature space to compute the corresponding DSFs. The computed DSFs from all the synchronized sensor nodes can be transmitted through wireless channels to the receiver end for further processing, where the damage index can be computed as severity of damage. The change in the damage index can be monitored for continuous health monitoring of the structure.

5.3 Influence of parameter m and n

In this section, an experiment is performed on the various length of data streams to study the influence of parameter m and n to the classification accuracy (shown in Table 5). It is worth mentioning that for $m < n$, the feature space models are computed directly from the covariance matrix XX^T . It can be observed that the length of data stream should not be less than 1000 data points for a better classification accuracy

Table 5 Influence of parameter m and n to the classification accuracy

m	n	12-DOF Analytical Model, Ambient	120-DOF Analytical Model, Ambient	120-DOF Analytical Model, Shaker
250	1920	66.05	65.34	69.91
500	960	66.91	66.78	78.38
750	640	70.54	70.83	89.46
1000	480	99.82	99.87	92.81
1250	384	99.91	99.93	96.44
1500	320	99.94	99.89	98.98
1750	274	99.33	99.87	98.63
2000	240	96.70	98.05	96.90

which validates our stationary test performed in the previous section (see Table 2).

5.4 Analysis of performance sensitivity to p_a and p_b

In this section, the sensitivity of the classification performance to the parameters p_a and p_b is analyzed. First, we fixed the parameter value of p_b to 99.7% according to 3-sigma rule [46] and varied the parameter p_a in the range of 15 – 95% to study the sensitivity to p_a . From Table 6, we infer that an increase in the number of PCs (after a certain number of PCs) to be discarded in the DSF extraction process improves the classification accuracy. We also find that the proposed algorithm achieves maximum accuracy at $p_a = 95%$. Therefore, for the sensitivity analysis corresponding to p_b , we fixed the value of p_a equal to 95% and varied the parameter p_b in the range of 97–99.7% (see Table 7). It should be mentioned here that most of the misclassified results shown in both the tables actually correspond to Case 7.

5.5 Computational time

In the present section, the computational time of the proposed DSF extraction process is evaluated. We have performed an experiment to find an optimal number of AR model coefficients, which defines the dynamics of the ASCE benchmark structure, using Akaike Information Criterion (AIC) [47]. Our results indicate that most of the variation in the data due to change in dynamics of the structure can be

Table 4 Classification accuracy for each test case corresponding to different analytical models

Excitation	Classification accuracy in %						
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
12DOF Ambient	99.37	100	100	100	100	100	100
120DOF Ambient	98.75	100	100	100	100	100	100
120DOF Shaker	99.06	100	100	100	100	100	94.06

Table 6 Analysis of performance sensitivity to p_a for fixed value of $p_b = 99.7\%$

p_a	Avg. no. of PCs discarded corresponding to higher eigenvalues (a)	No. of misclassified data streams	% of misclassified data streams	Accuracy
15	2	79	4.11	95.89
35	5	82	4.27	95.73
55	9	83	4.32	95.67
75	19	78	4.06	95.94
85	32	65	3.38	96.62
90	46	40	2.08	97.92
95	71	20	1.04	98.96

Table 7 Analysis of performance sensitivity to p_b for fixed value of $p_a = 95\%$

p_b	Avg. no. of PCs discarded corresponding to lower eigenvalues ($320 - b$)	No. of misclassified data streams	% of misclassified data streams	Accuracy
97.0	231	555	24.78	75.22
97.9	218	417	18.62	81.38
98.8	198	341	15.22	84.78
99.7	50	20	1.04	98.96

captured by AR model of order in the range of 5–8. These results corroborate well with the reported values in the context of AR model-based damage diagnosis. The reports suggest that AR model of order in the range of 5–8 is sufficient to detect all damaged cases of ASCE benchmark structure [32]. Therefore, in the current work, we have chosen the eighth-order AR model for the comparison with our proposed DSF in terms of computational time. Accordingly, the average processing time for the proposed DSF extraction is compared with the average processing time to estimate eighth-order AR model coefficients. This experiment was performed in MATLAB R2015b on a computer with a third-generation core i5 64-bit processor running Windows 7 and 4 GB of RAM. To calculate the average time, 320 data streams of case 1 corresponding to analytical model were used. AR model coefficients of 8th order models are estimated from 320 data streams and averaged to compute the processing time per data stream. The average processing time for AR coefficient estimation is obtained as 1.76 s/stream; however, the proposed DSF takes 155 millisecc/stream. This shows that the proposed DSF extraction is much faster than the estimation of AR coefficients. Less processing time saves power consumption at wireless sensor node. Thus, the proposed DSF would be very efficient for the wireless sensor network-based SHM.

6 Conclusions

In this paper, a time series-based damage diagnosis algorithm has been proposed by introducing a new computationally efficient damage sensitive feature (DSF). A

reduced size feature space model has been proposed to extract DSFs using multivariate analysis of time series data using PCA. The proposed feature extraction process eliminates the consequence of environmental variations and measurement noise during the feature extraction itself by selecting a number of principal components (PCs) corresponding to mid-range eigenvalues. The range of eigenvalues has been selected successfully using cumulative percentage of variance (CPV) criterion. The major advantage of the proposed DSF is that it is simple and computationally efficient. The applicability of the proposed DSF has been validated on ASCE benchmark structure under different excitations and structural conditions. The results show that the levels of damage are well detected, and the proposed DSF is also sensitive to small damages in the structure.

The extraction of the proposed DSF is much faster than estimation of AR coefficients for the same length of time series signal. Therefore, the proposed approach can work efficiently for the wireless sensor network, where measured signals can be processed at the sensor node itself through embedded algorithms. The damage detection results are promising. Although we have discussed the algorithm in the context of wireless sensor network-based SHM, it can be utilized for cable-based SHM. Furthermore, field testing is required to validate the proposed algorithm in an extensive change in environmental conditions along with various damage patterns like loosening of bolts or cracking at joints.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Chang PC, Flatau A, Liu SC (2003) Review paper: health monitoring of civil infrastructure. *Struct Health Monit* 2(3):257–267
- Das S, Saha P, Patro SK (2016) Vibration-based damage detection techniques used for health monitoring of structures: a review. *J Civ Struct Health Monit* 6(3):477–507
- Neves A, González I, Leander J, Karoumi R (2017) Structural health monitoring of bridges: a model-free ANN-based approach to damage detection. *J Civ Struct Health Monit* 7(5):689–702
- Le HV, Nishio M (2015) Time-series analysis of GPS monitoring data from a long-span bridge considering the global deformation due to air temperature changes. *J Civ Struct Health Monit* 5(4):415–425
- Yan AM, Kerschen G, De Boe P, Golinval JC (2005) Structural damage diagnosis under varying environmental conditions-Part I: A linear analysis. *Mech Syst Signal Process* 19(4):847–864
- Imam B, Chryssanthopoulos M (2010) A review of metallic bridge failure statistics. In: *Proceedings of the Fifth International IABMAS Conference on Bridge Maintenance, Safety and Management*, pp 3275–3282
- Farrar CR, Worden K (2007) An introduction to structural health monitoring. *Philos Trans R Soc A: Math Phys Eng Sci* 365(1851):303–315
- Cavadas F, Smith IF, Figueiras J (2013) Damage detection using data-driven methods applied to moving-load responses. *Mech Syst Signal Process* 39(1):409–425
- Figueiredo E, Park G, Farrar CR, Worden K, Figueiras J (2011) Machine learning algorithms for damage detection under operational and environmental variability. *Struct Health Monit* 10(6):559–572
- Erazo K, Sen D, Nagarajaiah S, Sun L (2019) Vibration-based structural health monitoring under changing environmental conditions using kalman filtering. *Mech Syst Signal Process* 117:1–15
- Salawu OS (1997) Detection of structural damage through changes in frequency: A review. *Eng Struct* 19(9):718–723
- Doebbling SW, Farrar CR, Prime MB (1998) A summary review of vibration-based damage identification methods. *Shock Vib Digest* 30(2):91–105
- Farrar CR, Worden K (2012) *Structural health monitoring: a machine learning perspective*. Wiley, Hoboken
- Deraemaeker A, Worden K (2012) *New trends in vibration based structural health monitoring*, vol 520. Springer Science & Business Media, Berlin
- Walia SK, Patel RK, Vinayak HK, Parti R (2015) Time-frequency and wavelet-based study of an old steel truss bridge before and after retrofitting. *J Civ Struct Health Monit* 5(4):397–414
- Nair K (2007) *Damage diagnosis of algorithms for wireless structural health monitoring*. PhD thesis
- Lynch JP (2007) An overview of wireless structural health monitoring for civil structures. *Philos Trans R Soc A* 365(1851):345–372
- Chae MJ, Yoo HS, Kim JY, Cho MY (2012) Development of a wireless sensor network system for suspension bridge health monitoring. *Autom Constr* 21:237–252
- Gul M, Catbas FN (2009) Statistical pattern recognition for structural health monitoring using time series modeling: theory and experimental verifications. *Mech Syst Signal Process* 23(7):2192–2204
- Whelan M, Zamudio NS, Kernicky T (2018) Structural identification of a tied arch bridge using parallel genetic algorithms and ambient vibration monitoring with a wireless sensor network. *J Civ Struct Health Monit* 8(2):315–330
- Kumar K, Biswas PK, Dhang N (2019) Damage diagnosis of steel truss bridges under varying environmental and loading conditions. *Int J Acoust Vibr* 24(1):56–67
- Le HV, Nishio M (2019) Structural change monitoring of a cable-stayed bridge by time-series modeling of the global thermal deformation acquired by gps. *J Civ Struct Health Monit* 9(5):689–701
- Jayawardhana M, Zhu X, Liyanapathirana R, Gunawardana U (2017) Compressive sensing for efficient health monitoring and effective damage detection of structures. *Mech Syst Signal Process* 84:414–430
- Liu X, Lieven NAJ, Escamilla-Ambrosio PJ (2009) Frequency response function shape-based methods for structural damage localisation. *Mech Syst Signal Process* 23(4):1243–1259
- Mahato S, Teja MV, Chakraborty A (2017) Combined wavelet-Hilbert transform-based modal identification of road bridge using vehicular excitation. *J Civ Struct Health Monit* 7(1):29–44
- Roveri N, Carcaterra A (2012) Damage detection in structures under traveling loads by Hilbert-Huang transform. *Mech Syst Signal Process* 28:128–144
- Kramer MA (1991) Nonlinear principal component analysis using autoassociative neural networks. *AIChE J* 37(2):233–243
- Kumar K, Biswas PK, Dhang N (2016) Statistical damage detection approach in SHM based on error prediction model. *Int J Struct Civ Eng Res* 5(5):300–307
- Goi Y, Kim CW (2017) Damage detection of a truss bridge utilizing a damage indicator from multivariate autoregressive model. *J Civ Struct Health Monit* 7(2):153–162
- Yao R, Pakzad SN (2012) Autoregressive statistical pattern recognition algorithms for damage detection in civil structures. *Mech Syst Signal Process* 31:355–368
- Nair KK, Kiremidjian AS, Law KH (2006) Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure. *J Sound Vib* 291(1):349–368
- Nair KK, Kiremidjian AS (2007) Time series based structural damage detection algorithm using Gaussian mixtures modeling. *J Dyn Syst Meas Contr* 129(3):285–293
- Shlens J (2014) A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*
- Pozo F, Arruga I, Mujica EL, Ruiz M, Podivilova E (2016) Detection of structural changes through principal component analysis and multivariate statistical inference. *Struct Health Monit* 15(2):127–142
- Shokrani Y, Dertimanis VK, Chatzi EN, Savoia M (2016) Structural damage localization under varying environmental conditions. In: *11th HSTAM international congress on mechanics*, Athens, Greece
- Posenato D, Lanata F, Inaudi D, Smith IF (2008) Model-free data interpretation for continuous monitoring of complex structures. *Adv Eng Inform* 22(1):135–144
- Posenato D, Kripakaran P, Inaudi D, Smith IF (2010) Methodologies for model-free data interpretation of civil engineering structures. *Comput Struct* 88(7):467–482
- Kim CW, Kitauchi S, Chang KC, McGetrick PJ, Sugiura K, Kawatani M (2014) Structural damage diagnosis of steel truss bridges by outlier detection. In: *Safety, reliability, risk and life-cycle performance of structures and infrastructures*. CRC Press, pp 4631–4638
- Johnson EA, Lam HF, Katfygiotis LS, Beck JL (2004) Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data. *J Eng Mech* 130(1):3–15

40. Valle S, Li W, Qin SJ (1999) Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods. *Ind Eng Chem Res* 38(11):4389–4401
41. Leskovec J, Rajaraman A, Ullman JD (2014) *Mining of massive datasets*. Cambridge University Press, Cambridge
42. Smith SW (1997) *The scientist and engineer's guide to digital signal processing*, 2nd edn. California Technical Publishing, San Diego
43. Abolhassani AH, Selouani S, O'Shaughnessy D (2007) Speech enhancement using PCA and variance of the reconstruction error in distributed speech recognition. In: *IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*, IEEE, pp 19–23
44. Fugate ML, Sohn H, Farrar CR (2001) Vibration-based damage detection using statistical process control. *Mech Syst Signal Process* 15(4):707–721
45. Dyke SJ, Agrawal AK, Caicedo JM, Christenson R, Gavin HP, Johnson EA, Nagarajaiah S, Narasimhan S, Spencer BF (2010) *Database for structural control and monitoring benchmark problems*. In: *Dataset, Network for Earthquake Engineering Simulation (database)*
46. Montgomery DC, Runger GC (2010) *Applied statistics and probability for engineers*. Wiley, Hoboken
47. Ljung L (1999) *System identification: theory for the user*, 2nd edn. PTR Prentice Hall, Upper Saddle River

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.