

Damage detection of bridges based on spectral sub-band features and hybrid modeling of PCA and KPCA methods

Hossein Babajanian Bisheh^{1a} and Gholamreza Ghodrati Amiri^{*2}

¹School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran

²Natural Disasters Prevention Research Center, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran

(Received December 28, 2021, Revised April 7, 2022, Accepted April 24, 2022)

Abstract. This paper proposes a data-driven methodology for online early damage identification under changing environmental conditions. The proposed method relies on two data analysis methods: feature-based method and hybrid principal component analysis (PCA) and kernel PCA to separate damage from environmental influences. First, spectral sub-band features, namely, spectral sub-band centroids (SSCs) and log spectral sub-band energies (LSSEs), are proposed as damage-sensitive features to extract damage information from measured structural responses. Second, hybrid modeling by integrating PCA and kernel PCA is performed on the spectral sub-band feature matrix for data normalization to extract both linear and nonlinear features for nonlinear procedure monitoring. After feature normalization, suppressing environmental effects, the control charts (Hotelling T^2 and SPE statistics) is implemented to novelty detection and distinguish damage in structures. The hybrid PCA-KPCA technique is compared to KPCA by applying support vector machine (SVM) to evaluate the effectiveness of its performance in detecting damage. The proposed method is verified through numerical and full-scale studies (a Bridge Health Monitoring (BHM) Benchmark Problem and a cable-stayed bridge in China). The results demonstrate that the proposed method can detect the structural damage accurately and reduce false alarms by suppressing the effects and interference of environmental variations.

Keywords: cable-stayed bridge; environmental effects; hybrid PCA-KPCA; spectral sub-band features; structural damage detection

1. Introduction

The structural health monitoring (SHM) process provides timely and exact information regarding the performance of structures under gradual, and sometimes sudden, changes in their condition to ensure durability, safety, and serviceability of the structures and prevent economic and human losses.

In the field of SHM, vibrationbased damage identification approaches have been extensively developed for the past few decades to diagnosis damage of civil infrastructure systems, especially bridges. The underlying idea of these methods is that damage-derived alterations in the physical properties will result detectable changes in the structural characteristic properties, which in turn will affect the vibration responses (Doebbling *et al.* 1996). Thus, the damage assessment process can be

*Corresponding author, Professor, E-mail: ghodrati@iust.ac.ir

^aPostdoctoral Researcher, E-mail: h.babajanian@yahoo.com

implemented by using damage-sensitive features reflecting the irregularity in the dynamic information or the changes of vibration data that occur in the damaged structure. Data interpretation techniques can be categorized into two classes in SHM: inverse or model-based techniques and data-driven techniques (Zhang *et al.* 2019, Razavi and Hadidi 2020, Sajedi and Liang 2020, Pan *et al.* 2018). Model-based methods rely on finite element (FE) analyses to connect the structural model to measurements by updating the FE-model. In contrast, data-driven methods employ only the information extracted from processed, or direction, dynamic measurements to predict damage.

Data-driven monitoring techniques in which damage can be identified by employing only the measured vibration responses have been widely accepted in online SHM under operational and environmental conditions. These approaches aim to provide damage information by extracting sensitive features from the recorded dynamic response of the healthy structure, monitoring the structure by any alteration in the features. Many studies have been conducted in data-driven structural damage assessment based on changes in natural frequencies (Gillich *et al.* 2017), modal strain energy (Moughty and Casas 2017), or modal curvature (Shokrani *et al.* 2018). Moreover, artificial neural networks (Nguyen *et al.* 2019), Bayesian networks (Yin and Zhu 2018), genetic algorithms (Ramezani and Bahar 2021), support vector machines (Shyamala *et al.* 2018), and signal processing techniques are used, among others, with the aim of structural damage identification.

Signal processing based approaches can be defined as a tool to extract sensitive features from the dynamic response of structures and to interpret this information into a recognition of damage, which can then be employed for decision making regarding maintenance management strategies. Feature extraction based on signal processing is considered as one of the main components and challenging aspect of vibration-based SHM (Goyal and Pabla 2016). Various features extracted using signal processing techniques can be used to obtain recognition information, including time-domain features (Cao *et al.* 2017, Azim *et al.* 2020) (like standard deviation, mean, root mean squares, kurtosis, skewness, crest factor, etc.), frequency domain features (Pedram *et al.* 2018, Oliver *et al.* 2016, Zhang and Aoki 2019) (such as Fourier coefficient, frequency bands, energy in different frequency bands, and others), and time-frequency domain features (Pan *et al.* 2018) (such as amplitude levels in time-frequency bands, energy concentration, time-frequency distribution, etc.). Time-frequency analysis such as wavelet transform, Hilbert-Huang transform, short-time Fourier transform, and empirical wavelet transform have been extensively used to extract damage sensitive features from the measured dynamic response (Balafas *et al.* 2018, Li *et al.* 2019, Xin *et al.* 2019, Hamidian *et al.* 2018, Ahmadi *et al.* 2021).

Data-driven damage detection algorithms are often significantly affected by environmental and operational variations that may lead to false alarms in detecting damage. In the SHM field, the procedure of separating the environmental effects from the structural response is usually called data normalization, which can be complex depending on the type of structure.

In recent years, output-only approaches, which characterize the effects produced by the environmental variations without measuring them, have been applied to structural damage recognition. Examples of output-only techniques are the auto-associative neural network (Sohn *et al.* 2002), time series models (Achilli *et al.* 2021), cointegration (Liang *et al.* 2018), and principal component analysis (PCA). Sousa Tomé *et al.* (2019) presented a data-based methodology for identifying the existence and location of damage in a cable-stayed bridge using multivariate data analysis methods. They proposed the combined application of multiple linear regression and principal component analysis methods to suppress the effects of environmental and operational variations. The changes in principal components of frequency response function (FRF) were also used by Esfandiari *et al.* (2020) for structural model updating and structural damage estimation.

Kesavan and Kiremidjian (2012) developed a wavelet-based damage detection technique using the principal component analysis. They proposed the damage-sensitive feature vector obtained as a function of the wavelet energies at the fifth, sixth, and seventh dyadic scales. Shokrani *et al.* (2018) introduced a PCA-based method to detect and localize damage under the effects of environmental variations by performing PCA on mode shape curvatures. Nie *et al.* (2019) proposed a feature extraction method based on wavelet packet transform (WPT) combining with PCA employed for damage identification in vibration-based SHM.

Although the linear PCA has been widely utilized for feature extraction approaches to process high-dimensional, highly correlated, and noisy data, it can not reveal nonlinear relationships between variables for nonlinear systems. More recently, nonlinear PCA methods, known as kernel PCA (KPCA), have attracted significant attention from researchers in the damage detection field. Ghoulem *et al.* (2020) implemented kernel PCA to investigate damage identification of nonlinear cable structure and compared it with PCA. Reynders *et al.* (2014) developed an output-only method for structural health monitoring using Gaussian kernel PCA to estimate and eliminate nonlinear environmental and operational effects on the monitored features. Santos *et al.* (2015) compared the performance of linear and nonlinear PCA for detecting damage in a three-story frame structure under the influence of operational and environmental variations. The comparative study concluded that using the PCA-based algorithm with RBF kernel leads to the best classification performance and reduces the number of false alarms.

The present paper aims to detect structural damage of bridges under changing environmental conditions using the normalized feature-based strategy. A data-driven methodology for online early structural damage identification is proposed and implemented in a numerical model of a bridge benchmark problem and a real cable-stayed bridge (Yonghe bridge). The feature extraction methods based on spectral sub-band features and a hybrid model of PCA and KPCA for data normalization are combined to present damage index. In the first stage, we propose using two spectral sub-band features, namely SSCs and LSSEs, as damage-sensitive features for extracting damage information from the measured dynamic responses of the structures. These features provide frequency information by applying a filter-bank to the power spectrum of the vibration signals. In the second stage, a hybrid PCA-KPCA of the extracted features is proposed by closely integrating linear PCA and nonlinear KPCA of the features to suppress the effects of environmental variations and reduce false alarms. The Hotelling T^2 and SPE control charts are employed to identify the abnormal procedure operation condition and distinguish damage in the structures. The hybrid PCA-KPCA method is compared to KPCA by the SVM classifier to evaluate the effectiveness of its performance in detecting damage. Numerical and full-scale studies are carried out to validate the feasibility and effectiveness of the proposed method. The results demonstrate the ability of the proposed normalized features in detecting structural damage successfully and reducing false alarms by suppressing the interference of environmental effects.

2. Spectral sub-band features

2.1 Spectral sub-band centroids

Spectral Sub-band Centroids (SSCs) related to spectral peak locations of the signal illustrate at what frequency in the filter-bank of the power spectrum the center of mass is located. For each sub-band, the first centroid is computed to extract the frequency information of the power spectrum and

to find SSCs for a frame. Because these features are basically calculated from the power spectrum in which locations of spectral peaks are approximately unchanged under noisy conditions, SSC can be computed reliably. We assume the FFT amplitude spectrum of a segment by $S[k]$, where $k=1, \dots, N$ is called the discrete frequency indicator. The indicator $k=N$ relates to half the sampling frequency $f_s/2$ and the SSC for the m -th sub-band is defined as (Nicolson *et al.* 2018)

$$c_m = \frac{\sum_{k=q_1(m)}^{q_h(m)} kW_m[k]S^\gamma[k]}{\sum_{k=q_1(m)}^{q_h(m)} W_m[k]S^\gamma[k]} \quad (1)$$

where $W_m[k]$ denotes frequency response of the m -th bandpass filter coefficients. $q_1(m)$; $q_h(m)$ are lower and higher edges of m -th sub-band, respectively, and γ is called a constant applied for controlling the dynamic range parameter of the power spectrum. The key component of this method is the filter-bank, dividing the power spectrum into several frequency bands. In this work, the SSCs are calculated by dividing the frequency band uniformly on mel scale, and a value of $\gamma=1$ is used.

2.2 Log spectral sub-band energies

The Spectral Sub-band Energy (SSE) coefficients are calculated from the Power Spectral Density (PSD) using band filters uniformly on the mel scale. For a frame, the SSE that forms the final spectro-temporal representation of the signal is defined as (Chatterjee and Paliwal 2016)

$$X_i(b) = \sum_k h_b(k)\hat{P}_i(k), \quad 0 \leq b \leq B - 1 \quad (2)$$

$$\hat{P}_i(k) = \frac{1}{N} \left| \sum_{n=0}^{K-1} x_i(n)w(n)e^{-j2\pi kn/K} \right|^2, \quad 0 \leq k \leq K - 1 \quad (3)$$

where h_b refers to the b -th filter bank, $P_i(k)$ is power spectral density, K is the DFT length, N the frame length, and $w(n)$ is a windowing function. Log spectral subband energies (LSSEs) are local in both time and frequency. Spectral sub-band energy coefficients are scaled by the natural logarithm to form LSSEs

$$LSSE_i(b) = \log \sum_k h_b(k)\hat{P}_i(k), \quad 0 \leq b \leq B - 1 \quad (4)$$

3. PCA and kernel PCA

The PCA is a multivariate statistical method in which the original variables are transformed into a lower dimension space of uncorrelated principal components properly by their variances without missing much information. It extracts a new set of uncorrelated variables with high variances, called principal components (PCs), which illustrate changes of the dominant data. The variables with low variances are the residual matrix, which can be due to the noise information or environmental effects. Soo Lon Wah *et al.* (2018) The coefficients of the PCs, which are linear combinations of the

variables, can be calculated using eigenvectors of the correlation (or covariance) matrix of data. This technique with an orthogonal rotate retains only the PCs, also called the number of factors, by selecting the first k eigenvectors. Consider a data matrix $X \in \mathbb{R}^{n \times m}$, where each row represents a sample, and each column represents a feature. By applying PCA, the data matrix X is decomposed as

$$X = \hat{X} + \bar{X} = \sum_{i=1}^k t_i p_i^T + \bar{X} \tag{5}$$

where k denotes the number of retained principal components, $t_i \in \mathbb{R}^n$ represents the i -th score vector or principle component, and $p_i \in \mathbb{R}^m$ is the corresponding loading vector. Also, $\hat{X} \in \mathbb{R}^{n \times m}$ denotes the matrix reconstructed using the principal components, and $\bar{X} \in \mathbb{R}^{n \times m}$ is called the residual matrix resulting from the k factors. The PCs in the matrix X are formed in descending order according to their respective eigenvalues. The first few PCs are related to the factors with large variances, which can reect most of the information of original variables, while the last PC represents the factor(s) generating the minimum variances. In kernel principal component analysis, a nonlinear mapping $\Phi(\cdot)$ projects the original data X onto a higher dimensional feature space \mathcal{F} to correlate the data linearly. Then, linear PCA is performed on the new feature space \mathcal{F} . To avoid defining nonlinear mapping, a kernel function is applied to complete the nonlinear transformation. By applying the function $\ker(\cdot, \cdot)$, the dot product of two feature vectors $\Phi(x_i)$ and $\Phi(x_j)$ can be computed according to

$$\ker(x_i, x_j) = \Phi^T(x_i)\Phi(x_j) \tag{6}$$

for any $x_i, x_j \in \mathbb{R}^m$, no need to do the nonlinear mapping. Typical kernel functions are the Gaussian kernel $\ker(x_i, x_j) = \exp(-\|x_i - x_j\|^2/c)$, where $c > 0$ is called the kernel width, the radial bases function (RBF) kernel $\ker(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$, where $\gamma > 0$ is a kernel parameter that must be decided, and the polynomial kernel $\ker(x_i, x_j) = (x_i^T + d_0)^{d_1}$, where d_0 and d_1 are the polynomial kernel parameters (Santos *et al.* 2015). Although PCA represents a robust and effective monitoring technique, it cannot withstand nonlinear processes. In contrast, kernel PCA is able to extract nonlinear properties. Because of the complexity of the procedure monitoring and the performance of KPCA, depending on selected kernel parameters, it is most difficult to warranty that kernel PCA can accurately capture the procedure characteristics. In addition, many identification procedures depend on both linear and nonlinear features. Integrating linear PCA and nonlinear KPCA can be represented as a new procedure monitoring by a hybrid model structure. When KPCA is employed in damage identification, two multivariate control charts, including the principal components score T^2 and squared prediction error (SPE) plots, can be used for detecting abnormal conditions. One of the multivariate statistical techniques as a procedure control chart used for detecting abnormal behavior of a monitoring system is the Hotelling T^2 statistic. The presence of abnormal values in the control limit can indicate the existence of damage. Investigating the variability of data by analyzing the score matrix in the space of PCs, The T^2 statistic supposes that the underlying procedure pursues a multivariate normal distribution. For the i -th experiment, Hotelling's T^2 statistic that is the sum of the normalized squared scores can be defined as follows

$$T^2 = t\Lambda^{-1}t^T \tag{7}$$

where Λ denotes the diagonal matrix of the inverse of the eigenvalues associated with the retained

principal components, and $t=[t_1 t_2 \dots t_l]$ includes the linear and nonlinear principal components. The confidence threshold (or limits) for T^2 -statistic can be computed using statistical F -distribution in the for

$$T_{l,n,\alpha}^2 = \frac{l(n-1)}{n-l} F_{l,n-l,\alpha} \quad (8)$$

where n denotes the number of examples in the training data set applied in the computation of the model, F is called the Fisher-Snedecor's F -distribution, l is the number of the linearnonlinear principal components retained, and α parameter, percentage of the F -distribution, is the standard deviation. The squared prediction error (SPE), also called the Q -statistic, is the measure of fit of an example to the PCA model and is defined as

$$SPE = \sum_{j=1}^n t_j^2 - \sum_{j=1}^l t_j^2 \quad (9)$$

For the SPE, the control limit obtained by fitting a weighted x^2 -distribution to the reference distribution can be calculated from its approximate distribution

$$SPE_\alpha \sim g x_h^2 \quad (10)$$

Where g denotes a weighting parameter included to compute for the SPE magnitude, and h calculates for the freedom degrees. Detailed information on these control charts is given in Lee *et al.* (2004).

4. Hybrid model based on PCA and KPCA for damage detection

The proposed hybrid model technique combines linear PCA and kernel PCA using a serial model structure to consider linear and nonlinear properties for monitoring procedures. As shown in Fig. 1, this approach consists of four-step: (1) linear PCA is applied to obtain linear features, (2) nonlinear KPCA is used to obtain nonlinear features, (3) after fusing the linear and nonlinear principal components achieved from the previous steps, the KPCA is performed to the new feature matrix to compute KPCA decomposition, and (4) the squared prediction error (SPE) plot and principal components score plot T^2 are applied for detecting abnormal conditions. Given the training data matrix $X \in \mathbb{R}^{n \times m}$, the PCA decomposition is used as (Deng *et al.* 2018)

$$X = \sum_{i=1}^{k_L} t_{L_i} p_{L_i}^T + \bar{X} \quad (11)$$

where t_{L_i} and p_{L_i} are i -th linear score vector and corresponding loading vector. k_L is the number of retained principal components, and $\bar{X} = [\bar{x}_1 \bar{x}_2 \bar{x}_3 \dots, \bar{x}_N]^T$ is the data matrix in residual space of the PCA. The loading vector can be computed from the eigenvectors of the covariance matrix of data X as

$$\frac{1}{n-1} X X^T p_{L_i} = \lambda_{L_i} p_{L_i} \quad (12)$$

where λ_{L_i} is the i -th eigenvalue. For a testing vector x_t , $t_{L_i} = x_t^T p_{L_i}$. By applying the KPCA

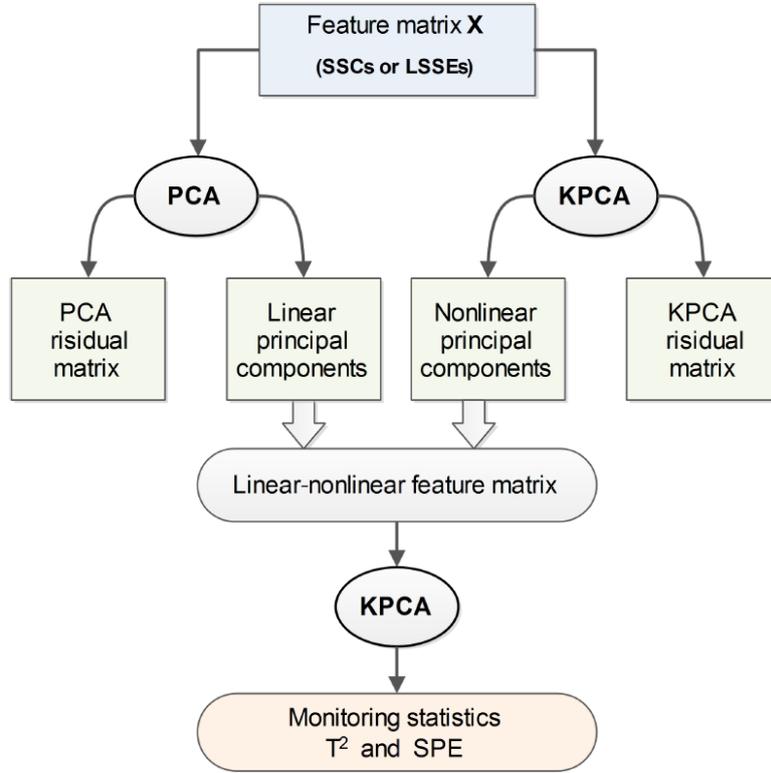


Fig. 1 Schematic for hybrid modeling based on PCA and KPCA

analysis of a feature matrix SSCs or LSSEs in a higher dimensional space \mathcal{F} , KPCA decomposition is obtained as

$$\Phi(X) = \sum_{i=1}^{k_N} t_{N_i} (p_{N_i})^T + E \tag{13}$$

where k_N denotes the number of retained kernel principal components, $t_{N_i} \in \mathbb{R}^n$ is the i -th nonlinear score vector, $p_{N_i} \in \mathcal{F}$ is related to the row of the loading vector, and $E \in \mathbb{R}^n \times \mathcal{F}$ is called the residual matrix of the KPCA. The nonlinear score vectors t_{N_i} and loading vectors p_{N_i} of the KPCA can be computed by the decomposition of the covariance matrix as

$$\frac{1}{n-1} \Phi^T(X) \Phi(X) p_{N_i} = \lambda_{N_i} p_{N_i} \tag{14}$$

where λ_{N_i} is the i -th eigenvalue of matrix and there exist projection vectors $\alpha_i = [\alpha_{i,1} \ \alpha_{i,2} \ \dots \ \alpha_{i,n}]^T$.

$$p_{N_i} = \sum_{j=1}^n \alpha_{i,j} \Phi(x_j) = \Phi^T(X) \alpha_i \tag{15}$$

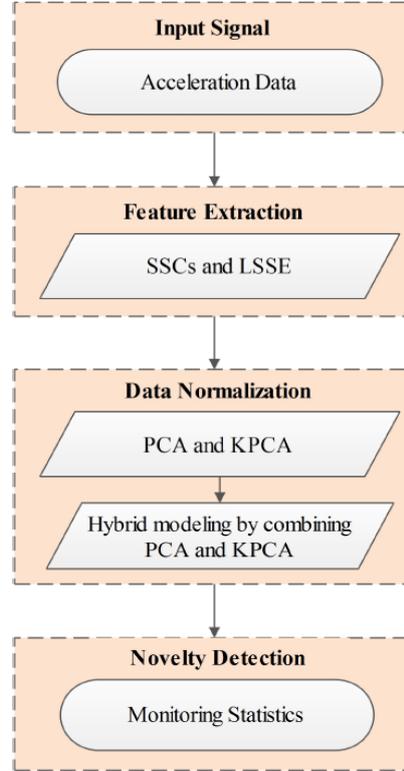


Fig. 2 Flowchart showing the methodology for damage identification

$$t_{N_i} = \Phi^T(x_t) p_{N_i} = \sum_{j=1}^n \alpha_{i,j} \Phi^T(x_j) \Phi(x_t) \quad (16)$$

A new feature space is constructed by combining the linear PCA and nonlinear KPCA modeling of the feature matrix as

$$\tilde{X} = [X - \bar{X} \quad \Phi(X) - E] \quad (17)$$

The KPCA analysis is performed on the linear-nonlinear feature matrix \tilde{X} to the decomposition process. After the feature normalization procedure using the above hybrid modeling, two monitoring statistics (T^2 and SPE) are constructed for novelty detection. In order to identify the damage condition for the SPE and T^2 monitoring charts, the threshold value is defined based on the confidence limit over the entire training features in the undamaged states. For each monitoring statistic, the confidence limit is required to examine if a damage occurs. The confidence limits for T^2 and Q statistics are computed using F distribution and weighted χ^2 -distribution.

5. Process of damage detection

According to the above analysis, it can be stated that the damage features normalized by a hybrid

model of PCA and KPCA will have an obvious jump after the damage occurred. The proposed method can remove the environmental factors from the measured response of a structure and reduce false alarms, based on which structural damage can be detected accurately. The detail process is described as follows (Fig. 2):

1. A dataset containing the acceleration time-series of the measured structural response is split into specified sections to the feature extraction procedure. The dataset related to the normal operating status (healthy) is considered as the baseline data.

2. Spectral subband features for a frame of raw data are calculated as damage features. These indicators include SSCs and LSSEs as potential candidates for extracting damage information. A comparison is also made between two feature matrices to investigate their ability and sensitivity in showing structural damage.

3. A data normalization procedure is performed on each of the feature matrices by hybrid modeling of linear PCA and nonlinear KPCA. Combining PCA and KPCA modeling presents a new and viable alternative to procedure monitoring. Using the model, which considers both linear and nonlinear features, can remove the effect caused by the environmental change except the structural damage, also leading to improvement in the identification process performances. In order to determine the optimal number of principal features in this paper, the few principal components (PCs) with the largest eigenvalues are selected as the retained features, yielding high classification accuracy in detecting damage by the SVM classifier.

4. After the damage feature matrix is normalized, tow monitoring statistics or control charts are applied for novelty detection. The Multivariate control charts include Hotelling's T^2 statistics and the Squared Prediction Error (SPE) for online monitoring. Once damage occurs, the dynamic balance is disturbed, which leads to an obvious jump in the control chart. It should be noted that all calculations were carried out using the MATLAB software.

6. Numerical benchmark and full-scale studies

6.1 Numerical model simulations

As the first phase to demonstrate the feasibility of the proposed method for damage identification, a numerical model of a Bridge Health Monitoring Benchmark Problem is used for the simulations. The numerical benchmark problem was conducted by University of Central Florida (UCF), and the vibration data of the bridge model were accessible. The physical structure that is a steel grid model includes two spans in the longitudinal direction with continuous girders across the middle columns, and seven transverse beams connecting tow the girders, as shown in Fig. 3. The bridge has tow 5.49 m longitudinal girders, 1.83 m wide, and 1.07 m columns. Details about the benchmark problem can be found at the benchmark bridge website (<http://cee.ucf.edu/people/catbas/benchmark.htm>) and Gul and Catbas (2011). A finite element model of the bridge was generated using 181 elements and 176 nodes that consist of 1056 degrees of freedom. Three damage patterns were simulated by various levels, such as reduced stiffness at connections and boundary condition changes. Several accelerometers were placed on the model to record dynamic data under random loading. Four accelerometers located on nodes of the model ($N1$, $N2$, $N4$, and $N5$ in Fig. 3) are selected to collect the vertical accelerations. Three damage cases (i.e., cases A , B , and C) with varying levels of damage scenarios are considered to illustrate the effectiveness of the proposed methodology in detecting damage.

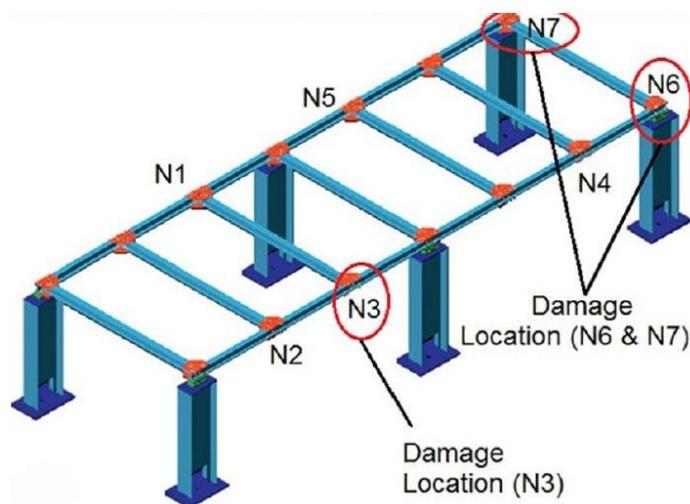


Fig. 3 Node numbers for steel grid structure used for the numerical benchmark study (Gul and Catbas 2011)

6.1.1 Case A: Release moment at node N3

For this case, the moment of the transverse beam at node $N3$ was released by eliminating the bolts connecting the longitudinal girder to the transverse beam. The vertical acceleration response data at nodes $N1$ and $N2$ were recorded from the model, and 10% white noise was artificially added. The response data is framed, and 40 filter banks are employed to calculate the spectral sub-band features for damage identification. Using spectral sub-band centroids (SSCs) and log spectral sub-band energies (LSSEs) are proposed as potential candidates for extracting damage-sensitive feature matrices. Since not all vectors of the feature space contain damage information, the most efficient feature vectors are found by a PCA-based method to create a new feature space that carries the most significant information about the structural damage. For this purpose, after the feature extraction process, the proposed hybrid modeling of linear PCA and nonlinear KPCA is performed on spectral sub-band features (SSCs and LSSEs) to normalize data and reduce the dimension of the feature space. A comparison of two spectral sub-band features is carried out to investigate their ability for structural damage detection. In the proposed hybrid modeling of PCA-KPCA, $k_L=1$ PCs for the PCA and $k_N=3$ nonlinear PCs for the KPCA with the Gaussian kernel are considered to normalize the SSCs feature matrix. According to Figs. 4 and 5, the control charts of hybrid modeling, both T^2 and SPE statistics, for this case at node $N1$ can successfully detect damage while the KPCA T^2 statistic cannot identify the damage. Fig. 6 illustrates the performance of the SSCs features normalized for this case at node $N2$.

Results show that the damage index efficiently discriminates between healthy and damaged states. However, sensor $N1$ shows less accuracy than $N2$ since it is away from the damage location. As another alternative, the above normalization procedure is performed on the log spectral subband energies to obtain the damage index. Figs. 7 and 8 illustrate that the LSSEs features normalized by hybrid modeling of PCA-KPCA lead to satisfying accuracy for classification or prediction of the bridge state for this case at node $N1$ and $N2$. It is obvious that both T^2 and SPE statistics obtained from LSSEs features through the normalization process can detect damage with good precision. Moreover, the LSSEs perform better than the SSCs in showing damage information and changes in structural conditions.

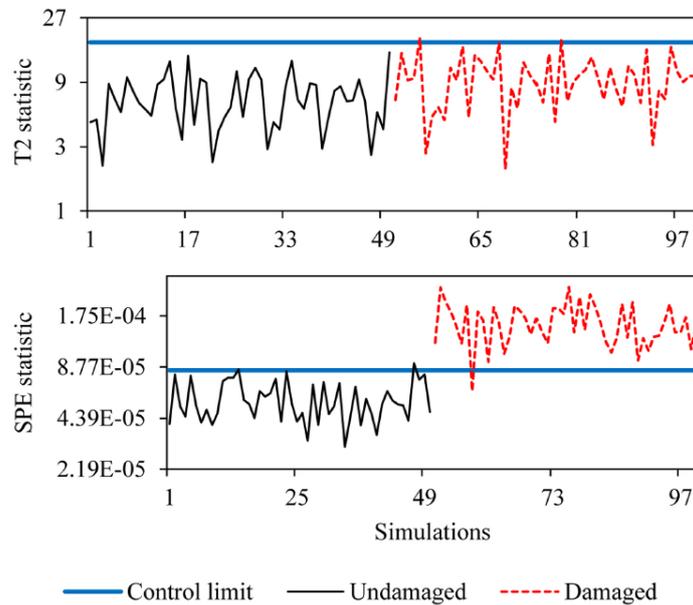


Fig. 4 Control charts of KPCA obtained using SSCs features for Case A at node N1

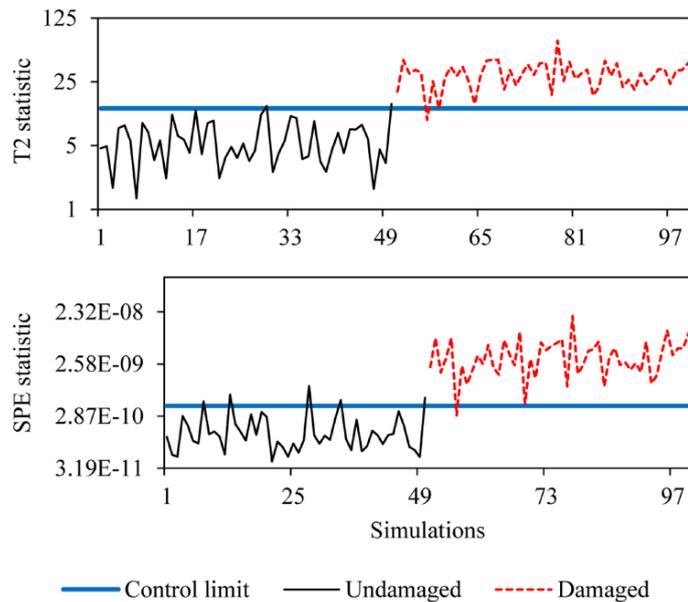


Fig. 5 Control charts of hybrid modeling (PCA-KPCA) obtained using SSCs features for Case A at node N1

6.1.2 Case B: Releasing moment and removing plate at N3

In this case, the gusset plates at node N3 and the bolts connecting the transverse beam to longitudinal girders were removed. The time series of the accelerations were recorded from the bridge at nodes N1 and N2, and 10% white noise was added. The spectral sub-band features, SSCs and LSSEs, are extracted from the response data to obtain damage sensitive feature matrices. Next,

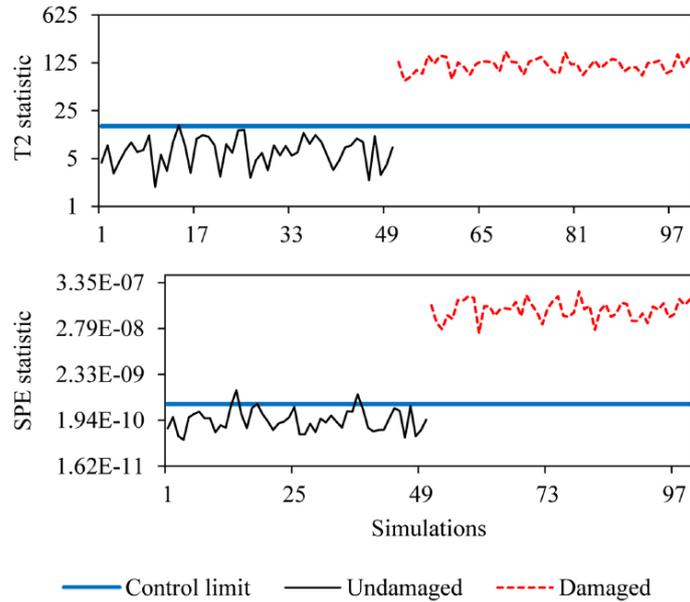


Fig. 6 Control charts of hybrid modeling (PCA-KPCA) obtained using SSCs features for Case A at node $N2$

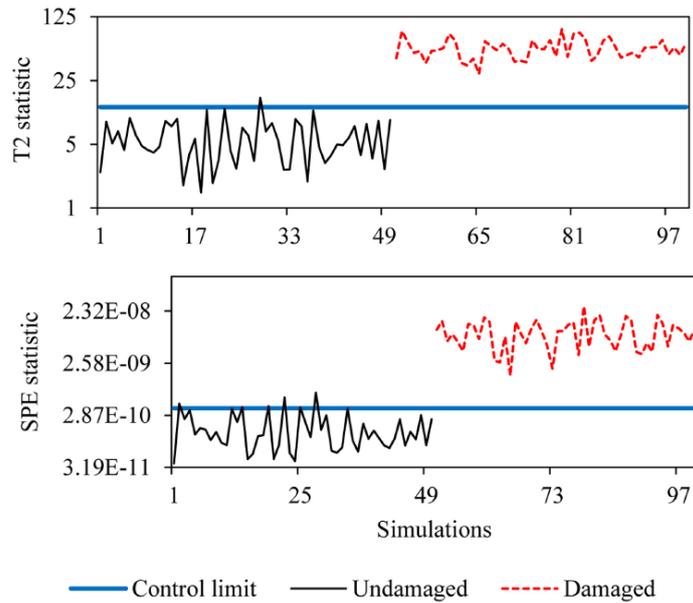


Fig. 7 Control charts of hybrid modeling (PCA-KPCA) obtained using LSSEs features for Case A at node $N1$

the feature normalization process is performed by hybrid modeling of linear PCA and nonlinear KPCA, and two control charts (T^2 and SPE) are applied to anomaly detection. In order to evaluate the effectiveness of the proposed methodology when LSSEs features extracted from data at node $N1$, the monitoring charts are presented in Fig. 9. It can be seen that the normalized features are very sensitive to damage and efficient in identifying damage to this case. In Tables 1 and 2, the

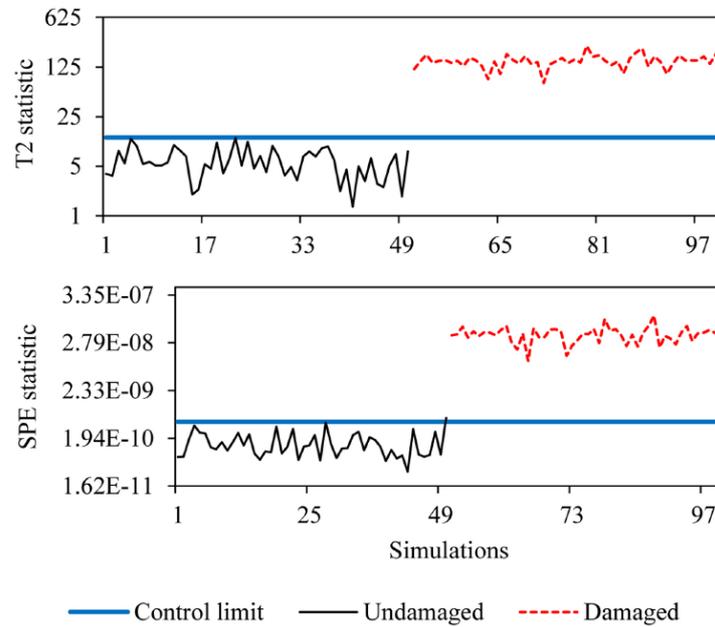


Fig. 8 Control charts of hybrid modeling (PCA-KPCA) obtained using LSSEs features for Case A at node N_2

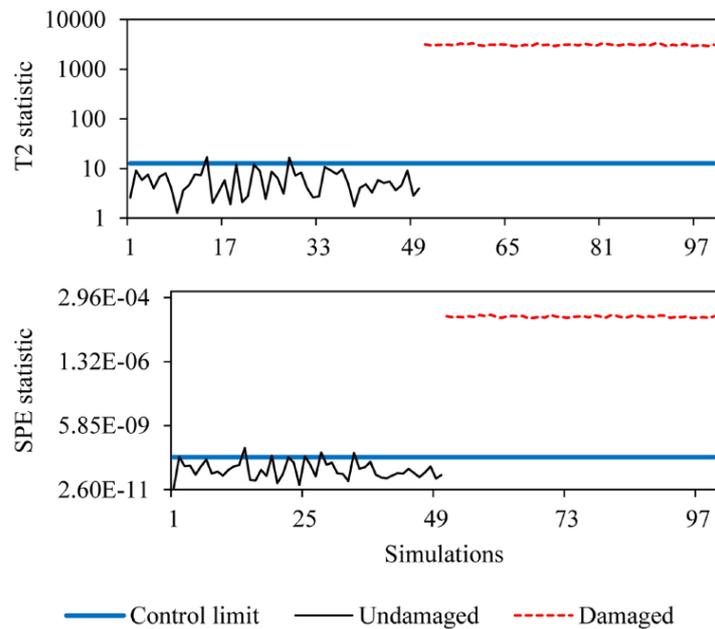


Fig. 9 Control charts of hybrid modeling (PCA-KPCA) obtained using LSSEs features for Case B at node N_1

classification precisions of spectral sub-band features normalized is studied using SVM as a classifier to show discriminative between damaged and healthy conditions. Results illustrate that the proposed methodology is successful in detecting damage with 100% classification accuracy.

Table 1 Mean of the monitoring statistic values obtained by extracting the SSCs features normalized using hybrid modeling of PCA and KPCA

Case	Situation	T^2		SPE		Accuracy (%)
		Undamaged	Damaged	Undamaged	Damaged	
Case-B	N1	6	444	1.8E-10	1.4E-6	100
	N2	7	1727	2.3E-10	1.1E-5	100
Case-C	N4	7	2978	2.4E-10	4.4E-5	100
	N5	2.2	498	9.4E-6	1.4E-3	100

Table 2 Mean of the monitoring statistic values obtained by extracting the LSSEs features normalized using hybrid modeling of PCA and KPCA

Case	Situation	T^2		SPE		Accuracy (%)
		Undamaged	Damaged	Undamaged	Damaged	
Case-B	N1	6	3109	1.8E-10	5.9E-5	100
	N2	5	378	2.3E-10	8.2E-7	100
Case-C	N4	5	1305	4.8E-8	-5.4E-6	100
	N5	5	1139	1.4E-10	1.3E-5	100

6.1.3 Case C: Boundary support restraint at N6 and N7

For this case, changes in the boundary conditions occurred and moment releases of the column supporting at nodes N6 and N7 were eliminated. The response data was collected from the deck of the bridge at nodes N4 and N5, and 10% white noise was added artificially. Feature extraction and normalization processes were performed for datasets recorded at this location. In order to evaluate the performance of the proposed method in detecting damage, SPE and T^2 statistic values are presented in Tables 1 and 2. Results clearly demonstrated that spectral sub-band features normalized by hybrid modeling of PCA-KPCA effectively discriminate between healthy and damaged conditions, which gives average accuracy of 100%, for this case at node N4 and N5.

6.2 Application to Tianjin Yonghe bridge

To validate the effectiveness of the proposed method in suppressing the effect of environmental change in practical application, the data recorded from a full-scale bridge, Yonghe cable-stayed bridge, is applied and processed. All the monitoring data coming from a real bridge established and issued as a benchmark problem by the Center of Structural Monitoring and Control at Harbin Institute. The vibration response of the full-scale cable-stayed bridge employed in this problem was recorded by a structural health monitoring system before and after the bridge was damaged. This benchmark problem can use to validate the damage detection methods (Li *et al.* 2014).

6.2.1 Description of the cable-stayed bridge

Tianjin Yonghe Bridge, shown in Fig. 10, is one of the earliest cable-stayed bridges in Mainland China and connects Hangu and Tianjin cities. It comprises two towers, 260 m main span, two side spans of 25.15 m and 99.85 m each, and 11 m width with a continuous prestressed box girder. The bridge was built in 1987, and after 19 working years, cracks were found in the bottom of the mid-span girder. Rehabilitation works were carried out, and a structural health monitoring system was



Fig. 10 General view of the Tianjin Yonghe Bridge (Liang *et al.* 2018)

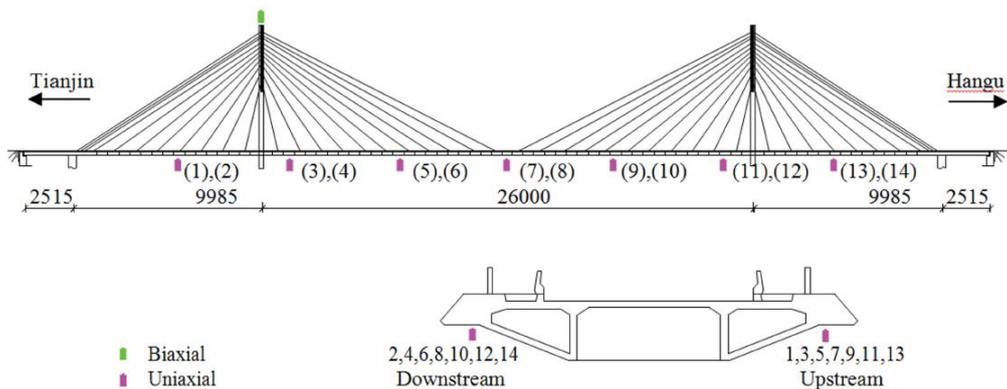


Fig. 11 Elevation of Tianjin Yonghe Bridge with the main dimensions and accelerations health monitoring system (Kaloop and Hu 2015)

designed and implemented for the bridge during the repair process. The acceleration monitoring system includes 14 uniaxial accelerometers that were permanently installed on the bridge deck to monitor and collect vibration responses, as shown in Fig. 11. After the rehabilitation and maintenance procedure, the cable-stayed bridge was reopened for operation at the end of 2007. In August 2008, two kinds of damage, including a crack occurred in the closure segment at both spans and damage in the piers of the bridge, were found through the bridge inspection.

6.2.2 Results of the proposed methodology

In the present section, the results of the damage identification methodology employed to the Yonghe bridge as a benchmark problem are described. This analysis is carried out using the available data set that included time histories of accelerations of the bridge deck before and after the damage occurred. Time series data for health state include 24 data sets of one hour recorded with the sampling frequency of 100 HZ on 17 January 2008. For the damage state, the data sets consist of acceleration data of one hour repeated for 24 hours and recorded by deck sensors at the same locations on 31 July 2008. The dynamic monitoring data sets recorded from 12 different days, from

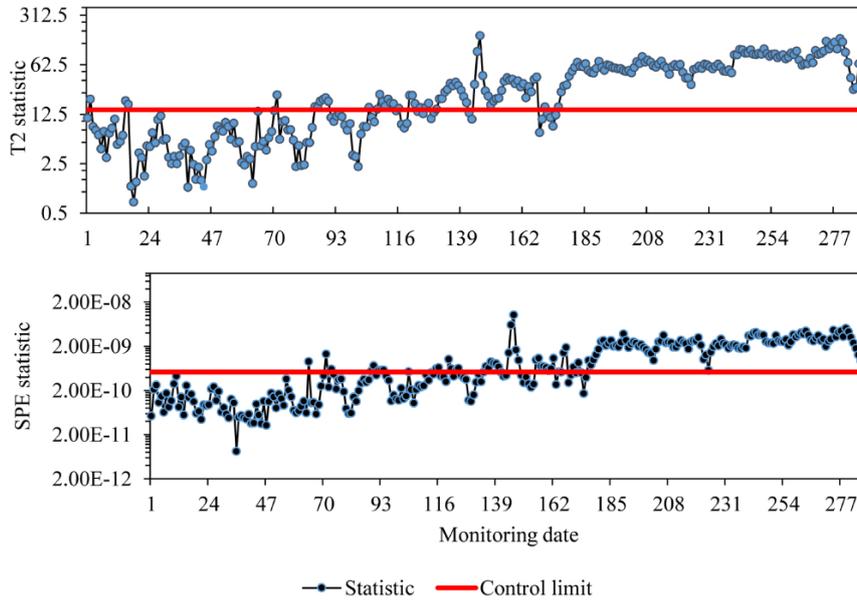


Fig. 12 Control charts of hybrid modeling (PCA-KPCA) obtained using SSCs features for Yonghe bridge at Sensor-1

January to July in 2008, namely January 1, January 17, February 3, March 19, March 30, April 9, May 5, May 18, May 31, June 7, June 16 and July 31 are available to analyze change procedure of the bridge health state. Therefore, selecting the time-series monitoring data from January to July 2008 can envelop the deterioration procedure of the bridge from health to damage. In order to determine damage-sensitive features, the SSCs are extracted from the segmented raw data to highlight the most relevant damage information. Twenty filter banks (passband filters) were employed to calculate the spectral sub-band features, leading to proper segregation of the vibration data in the presence of noises. Since not all sub-band features carry damage information, the significant features are determined through the data normalization process to discard irrelevant or redundant information and suppress the environmental effects. Hybrid modeling of linear PCA and nonlinear KPCA is proposed to normalize data and transform the SSCs feature space to new feature space. In the proposed hybrid modeling of PCA-KPCA, $k_L=2$ PCs for the PCA and $k_N=5$ nonlinear PCs for the KPCA with the Gaussian kernel are considered for reducing the dimensionality of the SSCs feature matrix. After the features are normalized, the SPE and T^2 monitoring charts are applied for anomaly detection. In this study, 95% confidence limit is computed as the damage identification threshold. For the monitoring point AC1 of the bridge, the damage identification result obtained by combining SSCs features and the PCA-based method is illustrated in Fig. 12. The result shows that normalized SSC features, as damage index, effectively discriminate between healthy and damaged conditions for both SPE and T^2 monitoring charts. It is obvious that hybrid modeling of PCA and KPCA exhibits good fitting performance for the spectral subband centroids before April 9. It means that an obvious structural alteration compared with the confidence limit did not occur from January 1 to April 9. In the feature normalization step, a comparative study between KPCA and hybrid PCA-KPCA methods is carried out using support vector machine (SVM) as a classifier to evaluate their performance in detecting damage. In the classification procedure, repeated random sub-sampling

Table 3 Classification accuracy of SSCs features normalized by two methods for the cable-stayed bridge (%)

Situation	KPCA	Hybrid modeling of PCA and KPCA	Performance improvement
Sensor-1	92.1	99.6	7.5
Sensor-2	93.2	98.4	5.2
Sensor-4	94.1	98.3	4.2
Sensor-6	92.1	97.3	5.2

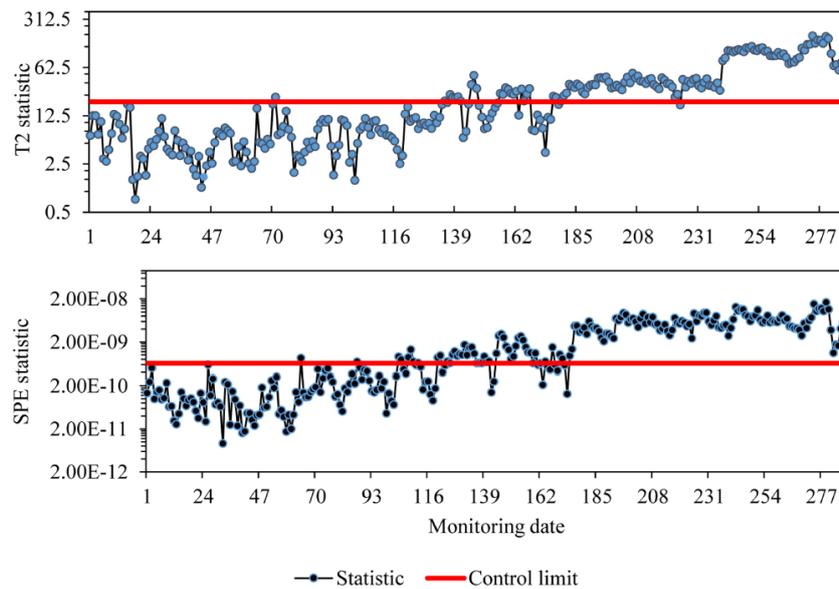


Fig. 13 Control charts of hybrid modeling (PCA-KPCA) obtained using SSCs features for Yonghe bridge at Sensor-2

validation technique is employed to cross-validation, and 50% of data is randomly selected as the train set and the other 50% as the test set. Time histories of the accelerations for the health condition (data recorded in January 2008) and the damaged condition (data recorded in July 2008) are considered as input data. The classification accuracies of SSCs features normalized by KPCA, and hybrid PCA-KPCA methods are computed by the SVM classifier for the monitoring points (AC1, AC2, AC4, and AC6), as presented in Table 3. Compared with the KPCA, results indicate that the use of the hybrid method to feature normalization improves the performance of the damage detection process so that the classification accuracy of 99.6% is achieved for the monitoring point AC1. Moreover, the SSCs extraction and feature normalization procedures are performed using the measured data of the acceleration recorded in the selected location of the sensor AC2. Fig. 13 refers to the bridge behavior from healthy to damaged conditions obtained by monitoring charts for the monitoring dates. According to the result, both SPE and T2 control charts can identify damage successfully. As potential candidates for showing damage features, the LSSEs are extracted from the acceleration data recorded in the selected locations of the sensors AC2 and AC4. Twenty filter banks (passband filters) are used to obtain the spectral subband features as input for the normalization process. The hybrid PCA-KPCA algorithm is performed to create effective features in the new feature space, of which the dimension is reduced by selecting the number of principal

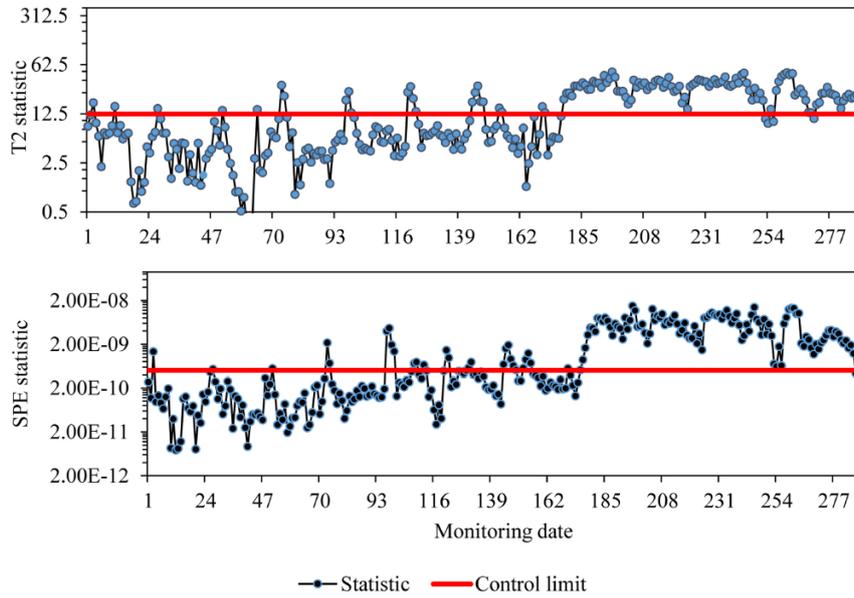


Fig. 14 Control charts of hybrid modeling (PCA-KPCA) obtained using LSSEs features for Yonghe bridge at Sensor-2

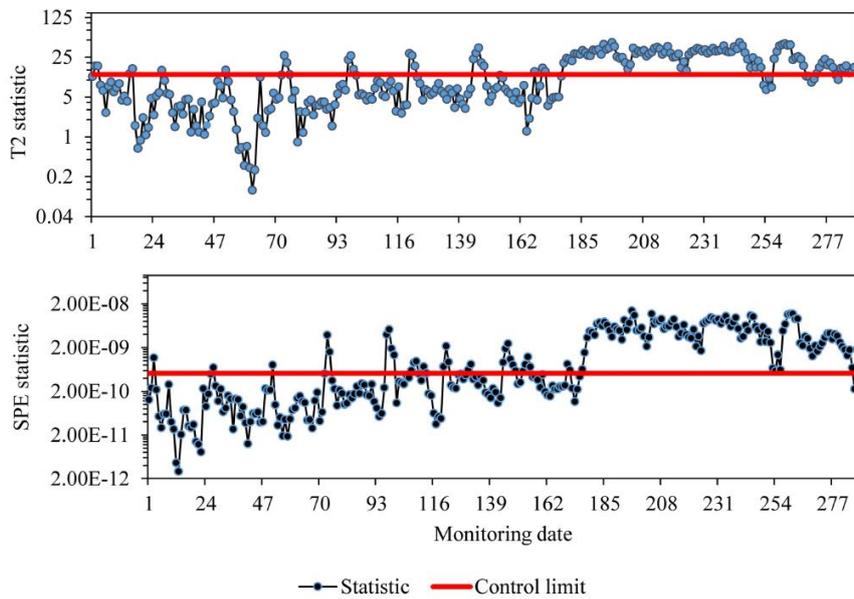


Fig. 15 Control charts of hybrid modeling (PCA-KPCA) obtained using LSSEs features for Yonghe bridge at Sensor-4

components. In the hybrid PCA-KPCA, $k_L=2$ PCs for the PCA and $k_N=4$ nonlinear PCs for the KPCA with the Gaussian kernel are considered to remove irrelevant features. After the feature normalization process, the SPE and T^2 control charts are used for abnormal detection. Figs. 14 and 15 show the bridge behavior from healthy to damaged conditions for the monitoring points AC2

Table 4 Classification accuracy of LSSEs features normalized by tow methods for the cable-stayed bridge (%)

Situation	KPCA	Hybrid modeling of PCA and KPCA	Performance improvement
Sensor-1	89.6	98.9	8.3
Sensor-2	90.2	97.3	7.1
Sensor-4	90.1	96.2	6.1
Sensor-6	90.4	96.1	5.7

and AC4, respectively. According to both SPE and T^2 monitoring charts, the results show that normalized LSSE features, as damage index, can detect the damage successfully and present a good fitting performance of the bridge behavior. Table 4 presents the classification precisions obtained by the SVM classifier for LSSE features normalized by KPCA and hybrid PCA-KPCA techniques for the monitoring points (AC1, AC2, AC4, and AC6). It is found that the hybrid method has better accuracy than the KPCA method for feature normalization since they give a classification accuracy of about 99% for the monitoring point AC1. The elimination of the effects of environmental and Fig. 13 Control charts of hybrid modeling (PCA-KPCA) obtained using SSCs features for Yonghe bridge at Sensor-2 operational changes was performed by reducing the damage-sensitive feature matrix without significant loss of damage information. The proposed feature matrix is transformed using the hybrid modeling of PCA and KPCA, which fuses the linear and nonlinear features, into new uncorrelated features arranged by their variances while preserving damage information to remove the irrelevant or redundant information. The new features with high variances indicate the dominating data variations, while other features with low variances are often thought to be the noise information or the influences of environmental and operational changes.

7. Conclusions

A novel data-driven methodology based on normalized spectral sub-band features for early damage identification is proposed in this paper to suppress, or at least minimize, the interference of environmental influences from the structural dynamic responses. As potential candidates in revealing damage information from data, the SSCs and LSSEs are proposed and utilized to separate vibration data in the presence of noises and is considered as the damage identification features. Since not all sub-band features extracted by SSCs or LSSEs processing carry damage information, the feature normalization algorithm based on hybrid modeling of linear PCA and nonlinear KPCA is proposed. The data normalization process aims to determine effective sub-band features and discard irrelevant or redundant information, leading to suppress the environmental effects. After the damage features are normalized, the SPE and T^2 monitoring charts are employed for anomaly detection. Numerical simulation of bridge benchmark problem and then a full-scale example of a cable-stayed bridge are used to validate the feasibility of the proposed method for damage identification. The numerical simulation study has illustrated that the performance of spectral sub-band features normalized by hybrid PCA-KPCA is highly promising as a damage index for the task of damage detection, despite the existence of numerical noise. For the LSSE features, both SPE and T^2 monitoring charts can detect damage with good accuracy to varying levels of damage cases of the bridge. Although normalized SSCs can identify damage successfully, the LSSE features are very sensitive to structural damage. Besides, the real cable-stayed bridge study under varying environmental effects powerfully

proves that the proposed method can successfully detect the health condition and suppress the influences of environmental variations. Finally, a comparative study between KPCA and hybrid PCA-KPCA methods is carried out using the SVM classifier to evaluate their performance in detecting damage. Compared with the KPCA, the hybrid method for feature normalization improves the classification accuracy between around 5% to 8%.

References

- Achilli, A., Bernagozzi, G., Betti, R., Diotallevi, P.P., Landi, L., Quqa, S. and Tronci, E.M. (2021), "On the use of multivariate autoregressive models for vibration-based damage detection and localization", *Smart Struct. Syst.*, **27**(2), 335-350. <https://doi.org/10.12989/sss.2021.27.2.335>.
- Ahmadi, H.R., Mahdavi, N. and Bayat, M. (2021), "A new index based on short time fourier transform for damage detection in bridge piers", *Comput. Concrete*, **27**(5), 447-455. <https://doi.org/10.12989/cac.2021.27.5.447>.
- Azim, M.R., Zhang, H. and Gül, M. (2020), "Damage detection of railway bridges using operational vibration data: Theory and experimental verifications", *Struct. Monit. Mainten.*, **7**(2), 149-166. <https://doi.org/10.12989/smm.2020.7.2.149>.
- Balafas, K., Kiremidjian, A.S. and Rajagopal, R. (2018), "The wavelet transform as a Gaussian process for damage detection", *Struct. Control Hlth. Monit.*, **25**(2), e2087. <https://doi.org/10.1002/stc.2087>.
- Cao, M., Sha, G., Gao, Y. and Ostachowicz, W. (2017), "Structural damage identification using damping: a compendium of uses and features", *Smart Mater. Struct.*, **26**(4), 043001.
- Chatterjee, A. and Paliwal, K. (2016), "Spectral subband centroids for tone vocoder simulations of cochlear implants", *Int. J. Signal Pr. Syst.*, **4**(4), 289-294.
- Deng, X., Tian, X., Chen, S. and Harris, C.J. (2018), "Deep principal component analysis based on layerwise feature extraction and its application to nonlinear process monitoring", *IEEE Trans. Control Syst. Technol.*, **27**(6), 2526-2540. <https://doi.org/10.1109/TCST.2018.2865413>.
- Doebling, S.W., Farrar, C.R., Prime, M.B. and Shevitz, D.W. (1996), "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review", Technical Report, Los Alamos National Lab., NM, United States.
- Esfandiari, A., Nabiyani, M.S. and Rofooei, F.R. (2020), "Structural damage detection using principal component analysis of frequency response function data", *Struct. Control Hlth. Monit.*, **27**(7), e2550. <https://doi.org/10.1002/stc.2550>.
- Ghoulem, K., Kormi, T. and Bel Hadj Ali, N. (2020), "Damage detection in nonlinear civil structures using kernel principal component analysis", *Adv. Struct. Eng.*, **23**(11), 2414-2430. <https://doi.org/10.1177/1369433220913207>.
- Gillich, G., Ntakpe, J., Wahab, M.A., Praisach, Z. and Mimis, M. (2017), "Damage detection in multi-span beams based on the analysis of frequency changes", *J. Phys.: Conf. Ser.*, **842**, 012033.
- Goyal, D. and Pabla, B. (2016), "The vibration monitoring methods and signal processing techniques for structural health monitoring: a review", *Arch. Comput. Meth. Eng.*, **23**(4), 585-594. <https://doi.org/10.1007/s11831-015-9145-0>.
- Gul, M. and Catbas, F.N. (2011), "Structural health monitoring and damage assessment using a novel time series analysis methodology with sensor clustering", *J. Sound Vib.*, **330**(6), 1196-1210. <https://doi.org/10.1016/j.jsv.2010.09.024>.
- Hamidian, D., Salajegheh, E. and Salajegheh, J. (2018), "Damage detection technique for irregular continuum structures using wavelet transform and fuzzy inference system optimized by particle swarm optimization", *Struct. Eng. Mech.*, **67**(5), 457-464. <https://doi.org/10.12989/sem.2018.67.5.457>.
- Kalooop, M.R. and Hu, J.W. (2015), "Stayed-cable bridge damage detection and localization based on accelerometer health monitoring measurements", *Shock Vib.*, **2015**, Article ID 102680. <https://doi.org/10.1155/2015/102680>.

- Kesavan, K.N. and Kiremidjian, A.S. (2012), "A wavelet-based damage diagnosis algorithm using principal component analysis", *Struct. Control Hlth. Monit.*, **19**(8), 672-685. <https://doi.org/10.1002/stc.462>.
- Lee, J.M., Yoo, C., Choi, S.W., Vanrolleghem, P.A. and Lee, I.B. (2004), "Nonlinear process monitoring using kernel principal component analysis", *Chem. Eng. Sci.*, **59**(1), 223-234. <https://doi.org/10.1016/j.ces.2003.09.012>.
- Li, R., Gu, H., Hu, B. and She, Z. (2019), "Multi-feature fusion and damage identification of large generator stator insulation based on lamb wave detection and SVM method", *Sensor.*, **19**(17), 3733. <https://doi.org/10.3390/s19173733>.
- Li, S., Li, H., Liu, Y., Lan, C., Zhou, W. and Ou, J. (2014), "SMC structural health monitoring benchmark problem using monitored data from an actual cable-stayed bridge", *Struct. Control Hlth. Monit.*, **21**(2), 156-172. <https://doi.org/10.1002/stc.1559>.
- Liang, Y., Li, D., Song, G. and Feng, Q. (2018), "Frequency Co-integration-based damage detection for bridges under the influence of environmental temperature variation", *Measure.*, **125**, 163-175. <https://doi.org/10.1016/j.measurement.2018.04.034>.
- Moughty, J.J. and Casas, J.R. (2017), "A state of the art review of modal-based damage detection in bridges: development, challenges, and solutions", *Appl. Sci.*, **7**(5), 510. <https://doi.org/10.3390/app7050510>.
- Nguyen, D.H., Bui, T.T., De Roeck, G. and Wahab, M.A. (2019), "Damage detection in Ca-Non Bridge using transmissibility and artificial neural networks", *Struct. Eng. Mech.*, **71**(2), 175-183. <https://doi.org/10.12989/sem.2019.71.2.175>.
- Nicolson, A., Hanson, J., Lyons, J. and Paliwal, K. (2018), "Spectral subband centroids for robust speaker identification using marginalization-based missing feature theory", *Int. J. Signal Pr. Syst.*, **6**(1), 12-16.
- Nie, Z., Guo, E. and Ma, H. (2019), "Structural damage detection using wavelet packet transform combining with principal component analysis", *Int. J. Lifecy. Perform. Eng.*, **3**(2), 149-170.
- Oliver, J.A., Kosmatka, J.B., Farrar, C.R. and Conte, J.P. (2016), "Frequency domain statistical damage identification applied to an experimental composite plate", *57th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, San Diego, California, January*.
- Pan, H., Azimi, M., Yan, F. and Lin, Z. (2018), "Time-frequency-based data-driven structural diagnosis and damage detection for cable-stayed bridges", *J. Bridge Eng.*, **23**(6), 04018033. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001199](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001199).
- Pedram, M., Esfandiari, A. and Khedmati, M.R. (2018), "Frequency domain damage detection of plate and shell structures by finite element model updating", *Invers. Prob. Sci. Eng.*, **26**(1), 100-132. <https://doi.org/10.1080/17415977.2017.1309398>.
- Ramezani, M. and Bahar, O. (2021), "Structural damage identification for elements and connections using an improved genetic algorithm", *Smart Struct. Syst.*, **28**(5), 643-660. <https://doi.org/10.12989/sss.2021.27.5.643>.
- Razavi, M. and Hadidi, A. (2020), "Assessment of sensitivity-based FE model updating technique for damage detection in large space structures", *Struct. Monit. Mainten.*, **7**(3), 261-281. <https://doi.org/10.12989/smm.2020.7.3.261>.
- Reynders, E., Wursten, G. and De Roeck, G. (2014), "Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification", *Struct. Hlth. Monit.*, **13**(1), 82-93. <https://doi.org/10.1177/1475921713502836>.
- Sajedi, S.O. and Liang, X. (2020), "A data-driven framework for near real-time and robust damage diagnosis of building structures", *Struct. Control Hlth. Monit.*, **27**(3), e2488. <https://doi.org/10.1002/stc.2488>.
- Santos, A., Silva, M., Sales, C., Costa, J. and Figueiredo, E. (2015), "Applicability of linear and nonlinear principal component analysis for damage detection", *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, May*.
- Shokrani, Y., Dertimanis, V.K., Chatzi, E.N. and N. Savoia, M. (2018), "On the use of mode shape curvatures for damage localization under varying environmental conditions", *Struct. Control Hlth. Monit.*, **25**(4), e2132. <https://doi.org/10.1002/stc.2132>.
- Shyamala, P., Mondal, S. and Chakraborty, S. (2018), "Numerical and experimental investigation for damage detection in FRP composite plates using support vector machine algorithm", *Struct. Monit. Mainten.*, **5**(2),

- 243-260. <https://doi.org/10.12989/smm.2018.5.2.243>.
- Sohn, H., Worden, K. and Farrar, C.R. (2002), "Statistical damage classification under changing environmental and operational conditions", *J. Intel. Mater. Syst. Struct.*, **13**(9), 561-574. <https://doi.org/10.1106/104538902030904>.
- Soo Lon Wah, W., Chen, Y.T., Roberts, G.W. and Elamin, A. (2018), "Separating damage from environmental effects affecting civil structures for near real-time damage detection", *Struct. Hlth. Monit.*, **17**(4), 850-868. <https://doi.org/10.1177/1475921717722060>.
- Sousa Tomé, E., Pimentel, M. and Figueiras, J. (2019), "Online early damage detection and localisation using multivariate data analysis: Application to a cable-stayed bridge", *Struct. Control Hlth. Monit.*, **26**(11), e2434. <https://doi.org/10.1002/stc.2434>.
- Xin, Y., Hao, H. and Li, J. (2019), "Operational modal identification of structures based on improved empirical wavelet transform", *Struct. Control Hlth. Monit.*, **26**(3), e2323. <https://doi.org/10.1002/stc.2323>.
- Yin, T. and Zhu, H.P. (2018), "Probabilistic damage detection of a steel truss bridge model by optimally designed Bayesian neural network", *Sensor.*, **18**(10), 3371. <https://doi.org/10.3390/s18103371>.
- Zhang, J. and Aoki, T. (2019), "A frequency-domain noniterative algorithm for structural parameter identification of shear buildings subjected to frequent earthquakes", *Comput.-Aid. Civil Infrastr. Eng.*, **35**(6), 615-627. <https://doi.org/10.1111/mice.12502>.
- Zhang, Z., Sun, C., Li, C. and Sun, M. (2019), "Vibration based bridge scour evaluation: A data-driven method using support vector machines", *Struct. Monit. Mainten.*, **6**(2), 125-145. <https://doi.org/10.12989/smm.2019.6.2.125>.