Available online at www.eccomasproceedia.org
Eccomas Proceedia COMPDYN (2023) 3620-3635

ECCOMAS
Proceedia

COMPDYN 2023
9th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Athens, Greece, 12-14 June 2023

COMPRESSION AND RECONSTRUCTION OF AUTOCOVARIANCE FUNCTIONS FOR DAMAGE DETECTION AND LOCALIZATION UNDER VARIABLE ENVIRONMENTAL CONDITIONS

Jyrki Kullaa

Department of Automotive and Mechanical Engineering
Metropolia University of Applied Sciences
Leiritie 1, 01600 Vantaa, Finland
P.O. Box 4071, 00079 Metropolia, Finland
e-mail: jyrki.kullaa@metropolia.fi

Abstract

Vibration-based structural health monitoring using autocovariance functions (ACFs) as damage-sensitive features is studied. ACFs are spatiotemporally correlated, which can be utilized in damage detection in the time domain. For accurate ACF estimates, a long measurement period with a high sampling frequency is needed. Storing data from repeated measurements will soon become exhaustive. A huge amount of space can be saved by storing much shorter ACFs instead. Further data compression can be achieved by storing only an optimally selected set of ACFs together with a coefficient matrix, from which the discarded ACFs can be reconstructed. The selection of the stored ACFs is done individually for each measurement based on the reconstruction accuracy. This ensures that the variability between measurements due to environmental or operational conditions or damage remains in the data. The stored and reconstructed ACFs are used for damage detection and localization by applying novelty detection techniques. The proposed method was studied with a numerical model of a bridge deck having 28 accelerometers. Only about four ACFs were stored together with the coefficient matrix for reconstruction. The space saving was more than 80%. The stored and reconstructed ACFs and the full set of original ACFs performed equally well in damage detection and localization. ACFs also outperformed direct measurement data in damage detection.

Keywords: Autocovariance Function, Data Compression and Reconstruction, Damage Detection, Optimal Sensor Placement, Environmental Effects, Spatiotemporal Correlation.

ISSN:2623-3347
Peer-review under responsibility of the organizing committee of COMPDYN 2023.
doi: 10.7712/120123.10666.20149

1 INTRODUCTION

Structural health monitoring (SHM) is based on measurement data. Vibration-based SHM is a global technique, in which damage can be detected remotely from a sensor. Nevertheless, it is often necessary to equip structures with a sensor network consisting of several vibration sensors. Information about the structure's condition is extracted from the response time histories. This process typically requires training data from the undamaged structure under different environmental or operational conditions. The excitation is often unknown, but can be assumed colored white noise. In order to obtain accurate vibration characteristics from the data, a long measurement period must be used. Moreover, the sampling frequency must be sufficiently high resulting in a large amount of data. These data must be stored to train a data model of the undamaged structure and then to monitor possible damage evolution by applying unsupervised learning techniques.

Due to a large amount of data, data management soon becomes an issue. The amount of data can be considerably reduced by extracting damage-sensitive features from the time records. The most common features are probably natural frequencies, mode shapes, and modal damping of the structure. They can be obtained using output-only system identification algorithms [1]. The algorithms must run automatically, which may be an issue. Also, a high-dimensional feature vector may result in the curse of dimensionality in statistical data analysis, especially if mode shapes are used in damage detection. On the other hand, if natural frequencies are only used, they may not be sensitive enough to detect damage, and damage localization would be difficult.

In this paper, autocovariance functions (ACFs) are proposed for damage detection and localization. Each sensor results in an ACF consisting of covariances at different time lags. This is advantageous for statistical analysis, because the feature vector is low-dimensional and has several data points per measurement. Also, damage can be associated to a sensor, which can be utilized in damage localization. ACFs have the same form as a free decay of the system [1]. This form is very advantageous, because there is a strong spatial and temporal correlation between ACFs of different DOFs [2]. The estimation of ACFs can be easily automated, but the measurement period should be long for sufficient accuracy. ACFs have been found to outperform raw measurement data in time-domain damage detection [2].

Correlation functions have been proposed for damage detection also in other studies. Li and Law [3] used cross-covariance functions with one sensor acting as a reference. In many cases, only a single data point is selected from the correlation functions, either the value at lag $\tau = 0$ [4–8], or the maximum of the cross-correlation function [9, 10].

The space requirement of autocovariance functions may also be an issue. Fortunately, ACFs are much shorter signals than the measured time histories, from which they are estimated. Storing AFCs instead of the actual measurements results already in an enormous space saving. Further data compression is possible by storing only a limited set of ACFs so that the discarded ACFs can be reconstructed using the stored ACFs. This is the main objective of this study.

It may first appear that because of redundancy, the stored ACFs alone would be sufficient for damage detection and no further information would be gained using also the reconstructed ACFs. This is, however, not true as will be shown in the numerical experiment. The redundancy in a single measurement is utilized in data compression, where the environmental and operational conditions remain unchanged. However, pooling many measurements under different environmental conditions in damage detection requires a higher-dimensional feature vector. In addition, damage localization is possible also to sensors whose data are reconstructed [11], fully exploiting all sensor data.

Neither excitation nor environmental variables need be measured. Quite a few methods to eliminate environmental or operational effects from data have been proposed, in which

measurement of the underlying variables is not necessary [12]. In this study, different environmental conditions between measurements are considered using whitening transformation of the training data acquired from the undamaged structure [13].

Sensor selection for storage is performed using optimal sensor placement (OSP) algorithms. The backward sequential sensor placement (BSSP) algorithm has proven efficient and practical [14–16] and is also used in this study. Initially, a full set of ACFs is available and one ACF is removed in each round until the stopping criterion is met. The selection is based on the most accurate reconstruction of the discarded data. The cost function is thus related to the reconstruction error, which must be minimized.

Damage detection is based on changes in the dynamic characteristics of the structure. In this paper, damage detection and localization are performed in the time domain. The data are the stored and reconstructed ACFs. Statistically significant differences between the training and test data are assumed to reveal damage. The discrepancy can be projected to each sensor, and the largest value is assumed to locate damage in the vicinity of the corresponding sensor.

The paper is organized as follows. Autocovariance functions are introduced in Section 2. An algorithm to optimally select ACFs for storage and reconstruct the discarded ACFs is proposed in Section 3. Damage detection and localization using the stored and reconstructed ACFs is also outlined. In Section 4, a numerical experiment is conducted with a bridge structure under variable environmental conditions. Damage is a propagating crack in the steel girder. Concluding remarks are given in Section 5.

2 AUTOCOVARIANCE FUNCTIONS

The autocorrelation function (ACF) of a zero-mean random process $\{x(t)\}$ is [17]:

$$R_{xx}(\tau) = \mathrm{E}[x(t)x(t+\tau)] \tag{1}$$

where E[·] is the expectation operator. The ACF at lag $\tau = k\Delta t$ can be estimated from the sampled measurement record having N samples with sampling period Δt:

$$R_{xx}(k) = \frac{1}{N-k}\sum_{j=1}^{N-k} x(j)x(j+k) \tag{2}$$

where $x(k) = x(k\Delta t)$ and $R_{xx}(k) = R_{xx}(k\Delta t)$. ACFs can alternatively be estimated using FFT computations [17]. This is much faster, if the measurement period is very long.

Autocovariance functions are time-domain features. For a viscously damped SDOF system under white noise excitation with a constant power spectral density S_0, the displacement autocorrelation function at a non-negative lag τ can be expressed as [18]:

$$R_{xx}(\tau) = \frac{\pi\omega_\mathrm{n}S_0}{2k^2\zeta}\left(\cos\omega_\mathrm{d}\tau + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_\mathrm{d}\tau\right)\exp(-\omega_\mathrm{n}\zeta\tau) \tag{3}$$

where ω_n is the natural angular frequency, k is the spring constant, ζ is the damping ratio, and $\omega_\mathrm{d} = \omega_\mathrm{n}\sqrt{1-\zeta^2}$ is the damped natural angular frequency. Note that Equation (3) has the following form.

$$R_{xx}(\tau) = (A\cos\omega_\mathrm{d}\tau + B\sin\omega_\mathrm{d}\tau)\exp(-a\tau) \tag{4}$$

where A, B, a, and ω_d are constants. Equation (4) has the same form as the free vibration of a viscously damped SDOF system. This particular form was studied in [2] showing that temporal correlation applies, so that values at different lags are redundant. The same form (4) applies also to velocity and acceleration ACFs [8].

For a MDOF system, ACFs include contributions from all modes, resulting in the following formula for DOF j (see also [1, 7, 19]).

$$R_{jj}(\tau) = \sum_{s=1}^{n} e^{-\zeta_s \omega_s \tau} \left[A_{jjs} \cos(\omega_{\mathrm{d}s}\, \tau) + B_{jjs} \sin(\omega_{\mathrm{d}s}\, \tau) \right] \tag{5}$$

where n is the number of modes, and

$$A_{jjs} = \sum_{r=1}^{n} \frac{F_{rs}}{2 m_r \omega_{\mathrm{d}r} m_s \omega_{\mathrm{d}s}} \boldsymbol{\phi}_{jr} \boldsymbol{\phi}_{js}\, \boldsymbol{\phi}_r^T \mathbf{B} \mathbf{Q} \mathbf{B}^T \boldsymbol{\phi}_s \tag{6}$$

$$B_{jjs} = \sum_{r=1}^{n} \frac{G_{rs}}{2 m_r \omega_{\mathrm{d}r} m_s \omega_{\mathrm{d}s}} \boldsymbol{\phi}_{jr} \boldsymbol{\phi}_{js}\, \boldsymbol{\phi}_r^T \mathbf{B} \mathbf{Q} \mathbf{B}^T \boldsymbol{\phi}_s \tag{7}$$

$$F_{rs} = \frac{-\zeta_r \omega_r - \zeta_s \omega_s}{\left(\zeta_r \omega_r + \zeta_s \omega_s\right)^2 + (\omega_{\mathrm{d}r} + \omega_{\mathrm{d}s})^2} + \frac{\zeta_r \omega_r + \zeta_s \omega_s}{\left(\zeta_r \omega_r + \zeta_s \omega_s\right)^2 + (\omega_{\mathrm{d}r} - \omega_{\mathrm{d}s})^2} \tag{8}$$

$$G_{rs} = \frac{\omega_{\mathrm{d}r} + \omega_{\mathrm{d}s}}{\left(\zeta_r \omega_r + \zeta_s \omega_s\right)^2 + (\omega_{\mathrm{d}r} + \omega_{\mathrm{d}s})^2} + \frac{\omega_{\mathrm{d}r} - \omega_{\mathrm{d}s}}{\left(\zeta_r \omega_r + \zeta_s \omega_s\right)^2 + (\omega_{\mathrm{d}r} - \omega_{\mathrm{d}s})^2} \tag{9}$$

in which m_r, ω_r, $\omega_{\mathrm{d}r}$, ζ_r, $\boldsymbol{\phi}_r$, are, respectively, the modal mass, the undamped natural circular frequency, the damped natural circular frequency, the damping ratio, and the mode shape vector of the mode r. Matrix **B** is the load distribution matrix, and **Q** is a diagonal matrix with spectral densities of the loads. From Equation (5), it can be seen that it has the same form as a free decay of the system. The redundancy applies also in a MDOF system within a single measurement under constant environmental and operational conditions.

However, between measurements, the situation is more complicated. First, under different environmental conditions, the dynamic characteristics may differ, resulting in different ACFs. Second, even if the environmental conditions remain the same, the excitation may vary, or more specifically, matrices **B** and **Q** may be different between measurements affecting the constants A_{jjs} and B_{jjs}.

Due to reasons stated above, it is important to process each measurement individually for data compression and reconstruction. On the other hand, in damage detection, in which the training data are comprised of several measurements, the data covariance matrix includes normal variability due to different environmental or operational conditions. Therefore, damage detection is supposed to require a higher-dimensional feature vector to take these variabilities into account. This can be realized by using all ACFs, stored and reconstructed, and applying spatiotemporal correlation.

3 DAMAGE DETECTION USING COMPRESSED DATA

Data analysis for damage detection is done using a full set of ACFs. However, it is often not necessary to store all ACFs. The objective is to store only a limited set of the ACF data so that the discarded ACFs can be reconstructed with sufficient accuracy using the stored signals. It is important to process each measurement individually to ensure that the variability between measurements due to different environmental or operational conditions or damage remains in the data.

Each measurement may result in a different set of stored data, depending on the reconstruction accuracy. Because all AFCs of a measurement are initially available, data compression and reconstruction can be actually tested before making the decision. The data analysis process is illustrated in Figure 1.

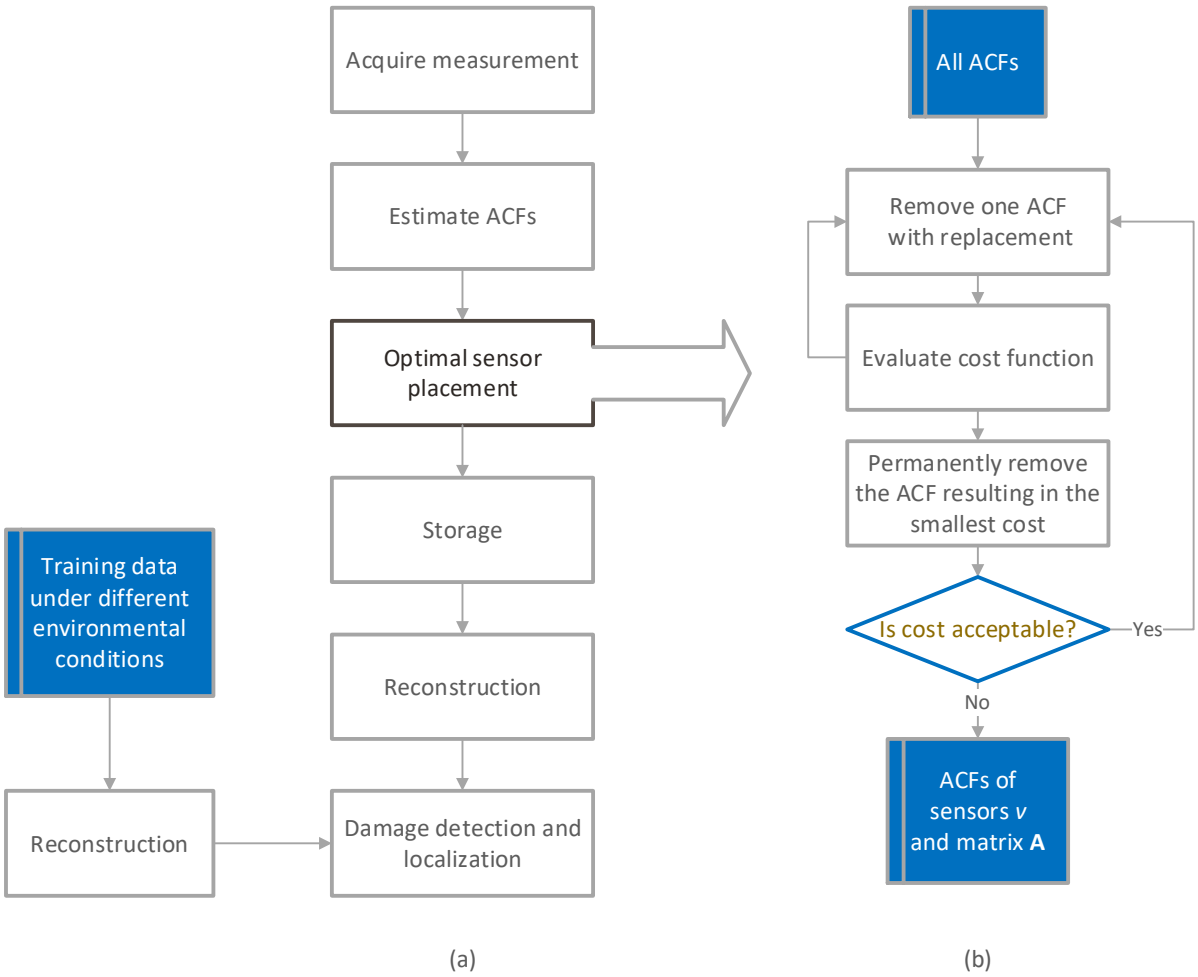


Figure 1: a) Flowchart for data compression and reconstruction for damage detection and localization. b) Flowchart for optimal sensor placement (OSP) to select the ACFs for storage.

3.1 Data compression and reconstruction

The flowchart to optimally select sensors for storage is depicted in Figure 1b. Consider a sensor network with p sensors resulting in p ACF estimates $\hat{R}_{ii}(\tau),\ i = 1,2,\ldots,p$. For simplicity, denote the ACF estimates with

$$\mathbf{r}(k) = \begin{Bmatrix} \hat{R}_{11}(k\Delta t) \\ \hat{R}_{22}(k\Delta t) \\ \vdots \\ \hat{R}_{pp}(k\Delta t) \end{Bmatrix} \tag{10}$$

After a single measurement, the ACFs from all sensors are available. If only a subset of the ACFs is stored, a lot of space can be saved. Assume that the ACFs of channels v are only stored, while the discarded ACFs of the remaining channels u must be reconstructed.

Divide $\mathbf{r}$ into stored data $\mathbf{r}_v$ and discarded data $\mathbf{r}_u$. The discarded data can be reconstructed using spatiotemporal correlation, as follows. Assuming model order m, the following data vectors are defined.

$$\mathbf{y}_u(k) = \boldsymbol{r}_u(k), \qquad \mathbf{y}_v(k) = \begin{Bmatrix} \mathbf{r}_v(k) \\ \mathbf{r}_v(k+1) \\ \mathbf{r}_v(k+2) \\ \vdots \\ \mathbf{r}_v(k+m) \end{Bmatrix} \tag{11}$$

$$\mathbf{y}(k) = \begin{Bmatrix} \mathbf{y}_u(k) \\ \mathbf{y}_v(k) \end{Bmatrix} \tag{12}$$

The data covariance matrix is

$$\boldsymbol{\Sigma} = \mathrm{E}[\mathbf{y}\mathbf{y}^T] = \begin{bmatrix} \mathrm{E}(\mathbf{y}_u\mathbf{y}_u^T) & \mathrm{E}(\mathbf{y}_u\mathbf{y}_v^T) \\ \mathrm{E}(\mathbf{y}_v\mathbf{y}_u^T) & \mathrm{E}(\mathbf{y}_v\mathbf{y}_v^T) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{y}_u\mathbf{y}_u} & \boldsymbol{\Sigma}_{\mathbf{y}_u\mathbf{y}_v} \\ \boldsymbol{\Sigma}_{\mathbf{y}_v\mathbf{y}_u} & \boldsymbol{\Sigma}_{\mathbf{y}_v\mathbf{y}_v} \end{bmatrix} \tag{13}$$

Note that for model order equal to $m = 0$, $\mathbf{y}_v(k)$ is simply equal to $\mathbf{r}_v(k)$, representing spatial correlation only.

The discarded data can be reconstructed by applying linear minimum mean-square error (MMSE) estimation. The mean square-error is minimized by the conditional mean [20]:

$$\mathrm{E}(\mathbf{y}_u|\mathbf{y}_v) = \boldsymbol{\Sigma}_{\mathbf{y}_u\mathbf{y}_v}\boldsymbol{\Sigma}_{\mathbf{y}_v\mathbf{y}_v}^{-1}\mathbf{y}_v = \mathbf{A}\mathbf{y}_v \tag{14}$$

where $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{y}_u\mathbf{y}_v}\boldsymbol{\Sigma}_{\mathbf{y}_v\mathbf{y}_v}^{-1}$ is the coefficient matrix that has to be stored together with the stored signals $\mathbf{r}_v$. The covariance matrix of the conditional distribution $p(\mathbf{y}_u|\mathbf{y}_v)$ is [20]:

$$\mathrm{cov}(\mathbf{y}_u|\mathbf{y}_v) = \boldsymbol{\Sigma}_{\mathbf{y}_u\mathbf{y}_u} - \boldsymbol{\Sigma}_{\mathbf{y}_u\mathbf{y}_v}\boldsymbol{\Sigma}_{\mathbf{y}_v\mathbf{y}_v}^{-1}\boldsymbol{\Sigma}_{\mathbf{y}_v\mathbf{y}_u} \tag{15}$$

which can be used to estimate the reconstruction error for optimal sensor selection.

3.2 Optimal sensor placement

The selection of ACFs for storage utilizes optimal sensor placement (OSP) methods. Starting with a full set of ACFs, one ACF is removed in each round until the stopping criterion is met. This is referred as backward sequential sensor placement (BSSP) algorithm [16].

Note that unlike in damage detection, each measurement under constant environmental and operational conditions is analyzed separately. Therefore, each measurement may result in different ACFs selected with a unique coefficient matrix.

Once the model order m is determined, the covariance matrix is estimated using Equation (13). One sensor in turn is removed with replacement, and all the removed ACFs are estimated using the retained ACFs. Once all ACFs have been gone through and the cost function evaluated, one ACF is permanently removed corresponding to the minimum value of the cost function. The process is repeated until the stopping criterion is met. Then the retained ACFs together with the coefficient matrix **A** in Equation (14) are stored.

The cost function and the stopping criterion are related to the error of reconstruction, Equation (15), which is minimized in each round.

3.3 Damage detection and localization

Damage detection starts by defining the training data, which are the selected ACFs (either directly estimated or stored and reconstructed) from the undamaged structure. Once the model order m is determined, another covariance matrix is estimated for all training data including influences of different environmental or operational conditions, but not damage.

The time-shifted covariance matrix estimate, with a time shift k, is computed by

$$\mathbf{R}_k = \mathrm{E}[\mathbf{x}(j)\mathbf{x}^T(j+k)] \approx \frac{1}{N-k}\sum_{j=1}^{N-k}\mathbf{x}(j)\mathbf{x}^T(j+k) \tag{16}$$

where $\mathbf{x}(k) = \mathbf{x}(k\Delta t)$ are the ACFs of the training data at a time instant $k\Delta t$, Δt is the sampling period, and N is the total number of samples. Equation (16) is similar to Equation (2), but in Equation (16) cross-covariances are also included. In Equation (2), N denotes the length of a single measurement, while in Equation (16) N denotes the total length of the training data. If m is the model order, the covariance matrix is:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_m \\ \mathbf{R}_1^T & \mathbf{R}_0 & \cdots & \mathbf{R}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_m^T & \mathbf{R}_{m-1}^T & \cdots & \mathbf{R}_0 \end{bmatrix} \tag{17}$$

For a model order equal to $m = 0$, $\mathbf{R} = \mathbf{R}_0$ corresponding to spatial correlation only.

Whitening transformation is applied to the training data using the covariance matrix in Equation (17). Whitening acts as data normalization considering also the effects of environmental or operational variability [13]. After whitening transformation, the training data points are located inside a unit hypersphere. The test data are then subjected to the same transformation. If damage (or any other novelty) is present, the corresponding transformed data points are presumably located outside the unit hypersphere. This can be detected e.g. by principal component analysis. Retaining only the first principal component reveals the direction of the largest variance. This direction projected on each sensor DOF and searching for the maximum is assumed to identify the sensor closest to damage.

A generalized extreme value statistics (EVS) probability distribution [21] is identified for the first principal component scores of the training data. An EVS control chart is then designed for the scores with appropriate control limits and subgroup size [22]. In this paper, the subgroup size equal to 100 and the probability of a false alarm equal to 0.001 have been used.

4 NUMERICAL EXPERIMENT

4.1 Structure and monitoring data

Sensor data were generated numerically using a finite element model of a bridge deck with a concrete slab and steel stiffeners (Figure 2). The structure was 30 m long and 11 m wide. More details can be found in [23]. The structure was equipped with 28 accelerometers. Vertical random excitation was applied at two points shown in green in Figure 2. The unmeasured excitations were mutually independent and different in each measurement in the frequency range between 0 and 20 Hz with random amplitudes and phases. The first 30 modes were used in the simulation. Modal damping was assumed with damping ratios of $\zeta_{1\text{–}2} = 0.01$, $\zeta_{3\text{–}4} = 0.02$, and $\zeta_{5\text{–}30} = 0.03$.

Steady state analysis was performed in the frequency domain using modal superposition. The sampling frequency was 100 Hz and the measurement period was more than 43 min. Each sensor produced 2^{18} samples per measurement. Gaussian random noise with equal standard deviations was added to each sensor. The average signal-to-noise ratio was 30 dB.

The bridge ends were at random temperatures between –20°C and +40°C. The relationship between the Young's modulus E of the concrete slab and temperature T (°C) was piecewise linear (see Figure 3a):

$$E = \begin{cases} E_0 + k_1 T, & T < 0 \\ E_0 + k_2 T, & T \geq 0 \end{cases} \tag{18}$$

where $E_0 = 40$ GPa, $k_1 = -1.0$ GPa/°C, and $k_2 = -0.25$ GPa/°C. The temperature distribution between the bridge ends was linear, except for Gaussian noise $N(0, 0.2°C)$, which was added to

the temperature values in each element row. Note that the temperature or the Young's modulus were not measured, but they were considered latent variables.

Damage was an open crack in a steel girder with an increasing severity (Figure 2). Due to damage and the temperature effect, the natural frequencies varied between measurements. Figure 3b shows the seven lowest natural frequencies of the structure in all measurements. The data points on the right-hand side of the vertical line are from the damaged structure. It is difficult to detect damage visually from the frequency changes due to the strong environmental effect. Natural frequencies were not used in damage detection in this paper.

The first 100 measurements were acquired from the undamaged structure and each damage level was monitored with six measurements under different and unknown environmental conditions. Training data were the first 70 measurements. Extreme value statistics (EVS) control charts were designed using the same training data.

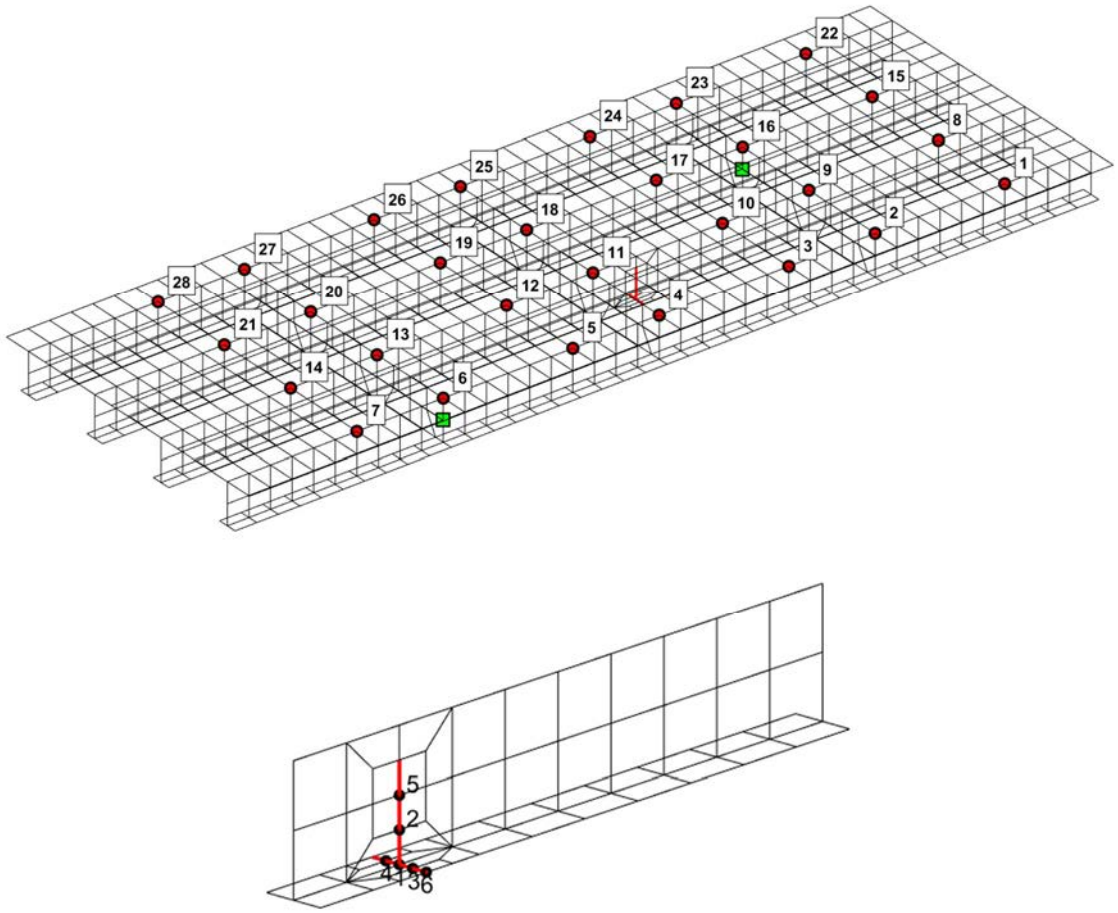
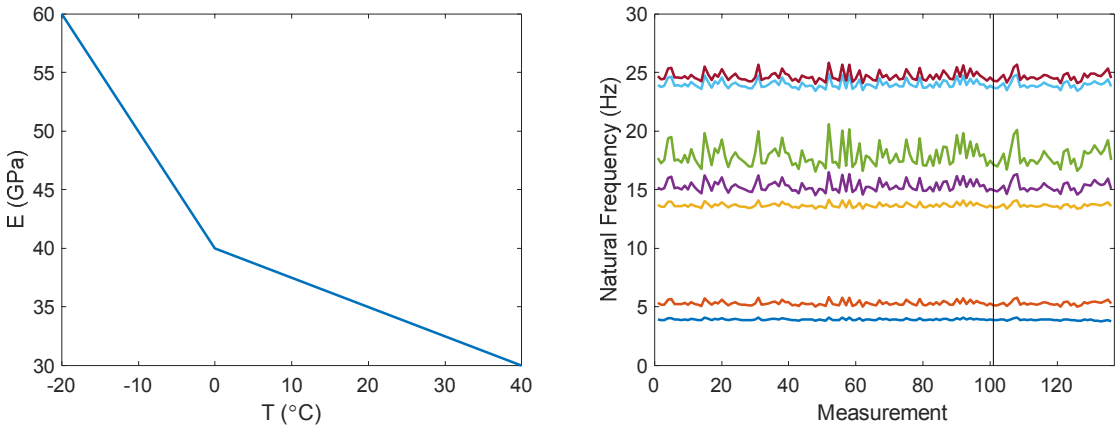


Figure 2: Top: Bridge deck with 28 accelerometers, two excitation points (green), and damage in a steel girder. Bottom: Detail of the crack damage with numbers indicating the order of damage evolution.



(a) (b)

Figure 3: (a) Young's modulus of concrete versus temperature. (b) Variation of the seven lowest natural frequencies due to temperature and damage. Frequencies on the right of the vertical line are from the damaged structure.

4.2 Data compression and reconstruction

Autocovariance functions were estimated from each sensor signal at time lags 0–3 s, resulting in 300 samples in each ACF ($n = 300$). As an example, the ACF of sensor 1 in measurement 4 is plotted in Figure 4. Compared with the actual sensor data, ACFs resulted in a huge space saving.

Having all 28 ACFs of a single measurement ($p = 28$), the next step was to select a limited number of ACFs for storage so that the discarded ACFs could be reconstructed with sufficient accuracy. The algorithm described in Section 3 was applied. The stopping criterion was that the estimation error, or more precisely, the square root of the conditional variances in Equation (15) had to be less than the estimated random errors of the ACF estimates at lag $\tau = 0$. The normalized rms error is [17]:

$$\varepsilon\left[\hat{R}_{xx}(0)\right] \approx (BT)^{-1/2} \tag{19}$$

where B is the bandwidth and T is the record length. With $B = 20$ Hz and $T = 2621$ s resulted in the normalized rms error $\varepsilon = 0.0044$. Multiplying this with $\hat{R}_{xx}(0)$ gave the absolute error. This value was computed for all data records and used in sensor selection.

Starting with all 28 ACFs, the backward sequential sensor placement (BSSP) algorithm was applied by removing a single sensor (ACF) in turn so that the maximum difference between the current and allowed reconstruction errors in any sensor in the network was minimized. Sensor removal continued until the accuracy requirement was violated. Model order $m = 4$ was used in data compression and reconstruction. Figure 4 shows the direct ACF estimate of sensor 1 in measurement 4 together with the reconstruction estimated using only three stored ACFs and the coefficient matrix of size 25×15. The difference is hardly visible. Only at higher lags is the discrepancy more pronounced.

The required number of stored ACFs varied between 3 and 6. Even with the relatively large stored coefficient matrices for each measurement, the total space saving was more than 81%.

With a higher model order, the accuracy of reconstruction would have been higher and the number of stored sensors would have been smaller. However, the size of the coefficient matrix would have increased. The size of the data matrix is $(n - m) \times (m + 1)p$. With a large model order m, the covariance matrix becomes singular. In the present case, the maximum allowed

model order would have been $m_{\max} = 9$. Therefore, a relatively small model order had to be used. In damage detection, the model order can be selected independently as can be seen in the following.

Once a single ACF was permanently removed, the mean error of the reconstructed ACFs was evaluated. The mean error as a function of the number of stored ACFs is plotted in Figure 5 for measurement 4. It can be seen that storing 11 ACFs instead of all 28 ACFs did not significantly increase the average noise level. When the number of stored ACFs was further decreased below 11, the reconstruction error considerably increased. In this case, only 3 ACFs had to be stored before the accuracy criterion was violated.

The reconstruction errors in each measurement, or more specifically, the square roots of the diagonal terms of $\mathrm{cov}(\mathbf{y}_u|\mathbf{y}_v)$ in Equation (15), are plotted in Figure 6 for all 28 sensors. The heavy line on top is the error limit. The criterion was not applied to the stored ACFs, and for those sensors the plotted value was zero.

A histogram of the selected sensors for storage in all measurements is shown in Figure 7a. The most often selected sensors can be clearly seen. However, both the number and location of the stored sensors varied between measurements. The placement of the stored three sensors in measurement 4 is plotted in Figure 7b.

The data compression ratio was computed as follows. If all ACF data were stored, the number of floating-point numbers in each measurement was 28·300 = 8400 numbers. Storing three ACFs (3·300 = 900 numbers) and the coefficient matrix **A** (Equation 14) of size 25×15 resulted in 1275 numbers. Consequently, only 15.2% of the total data had to be stored. Note also that the space saving was huge compared to the case if all raw acceleration data had been stored.

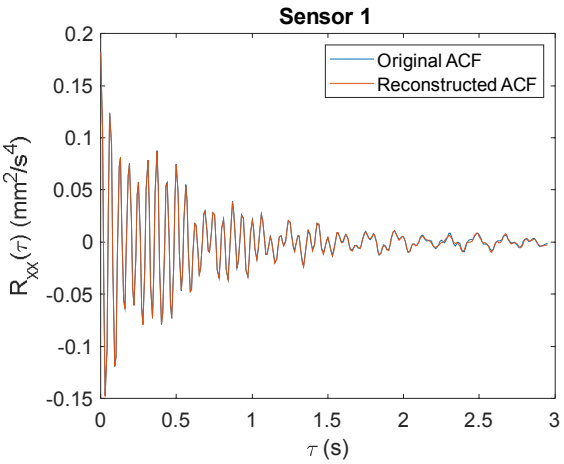


Figure 4: Directly estimated and reconstructed ACF of accelerometer 1 in measurement 4.

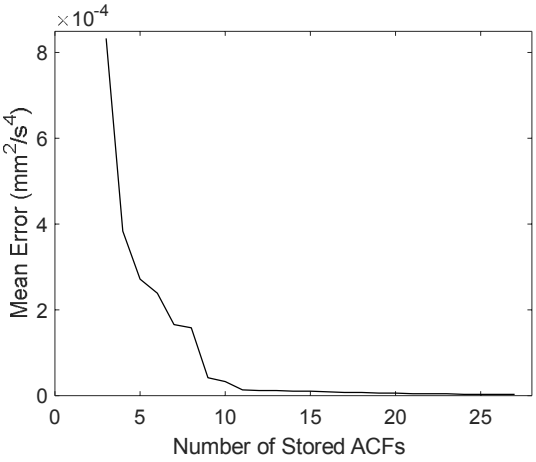


Figure 5: Mean reconstruction error as a function of the number of stored ACFs in measurement 4.

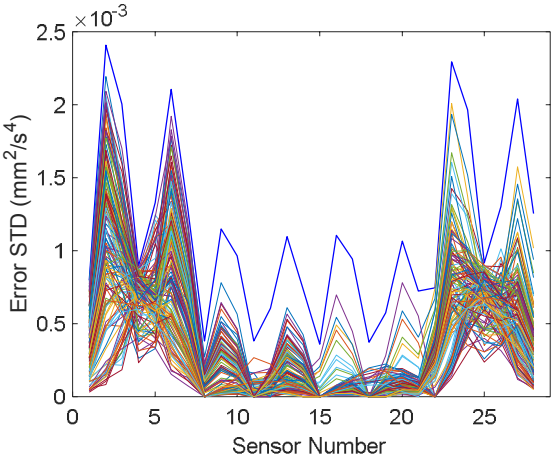
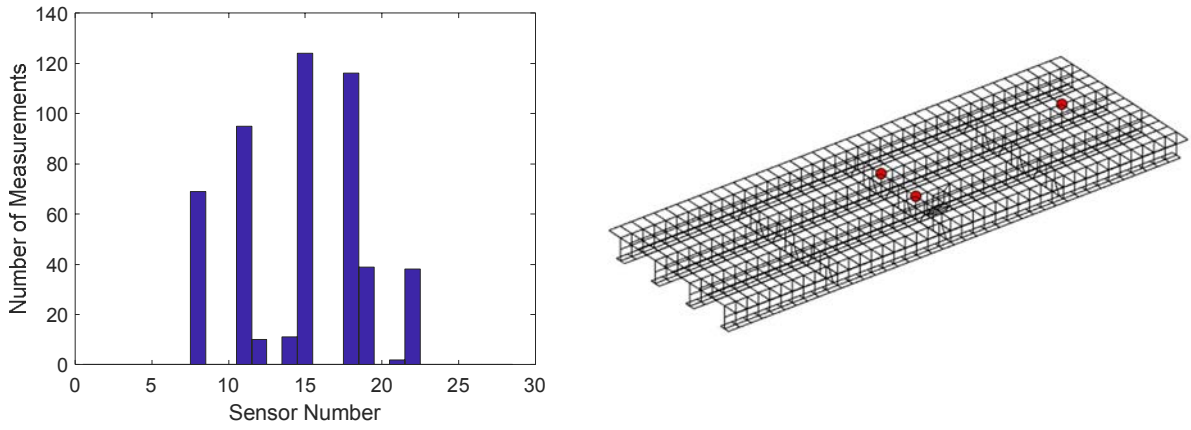


Figure 6: Accuracy criterion (blue heavy line on top) and the reconstruction errors of each sensor in all measurements.



(a) (b)

Figure 7: a) Histogram of selected sensors for storage in all measurements. b) Selected sensors for storage in measurement 4: sensors 11, 15, and 18.

4.3 Damage detection and localization

Damage detection was studied using four different data: (1) ACFs estimated from the measurement data, (2) stored and reconstructed ACFs, (3) raw acceleration measurements, and (4) the stored ACFs only. Model order equal to 10 was used in all cases.

EVS control charts designed with a subgroup size of 100 are plotted in Figure 8. The data points to the left of the blue vertical line correspond to the training data, while the black vertical lines indicate the onset of each six damage levels. Logarithmic scaling was applied to the vertical axis for clarity.

When using all directly estimated ACFs, all damage levels could be detected with only three false positives (Figure 8a). Detection of damage level 1 was however not clear. As a comparison, using the stored and reconstructed ACFs resulted in a slightly better detection performance with two false alarms (Figure 8b). This was surprising, because the accuracy probably decreased in data compression. Notice also that in one measurement, the maxima exceeded the lower control limit, which could be ignored.

If the measurement data were directly used [23], only the largest damage level could be detected (Figure 8c). Also, if the four most stored ACFs of sensors 8, 11, 15, and 18 were only used, damage levels 5 and 6 were only detected (Figure 8d). Due to different environmental conditions between measurements, a higher-dimensional feature vector would have been needed to remove the environmental influences. This could be achieved by using all 28 ACFs instead of the four stored ACFs alone.

Damage localization was done by plotting the squared projection of the first principal component on each sensor (Figure 9). Using the direct ACF estimates, damage was localized to sensor 11, and using the stored and reconstructed ACFs, damage was localized to sensor 10. The correct position was between sensors 10 and 11, slightly closer to sensor 11.

Applying spatial correlation only with the model order equal to $m = 0$, the control charts in Figure 10 resulted. Compared to the spatiotemporal correlation, damage detection was less reliable. Direct ACF estimates outperformed the reconstructed ones. The number of false positives remained at the same level as with spatiotemporal correlation.

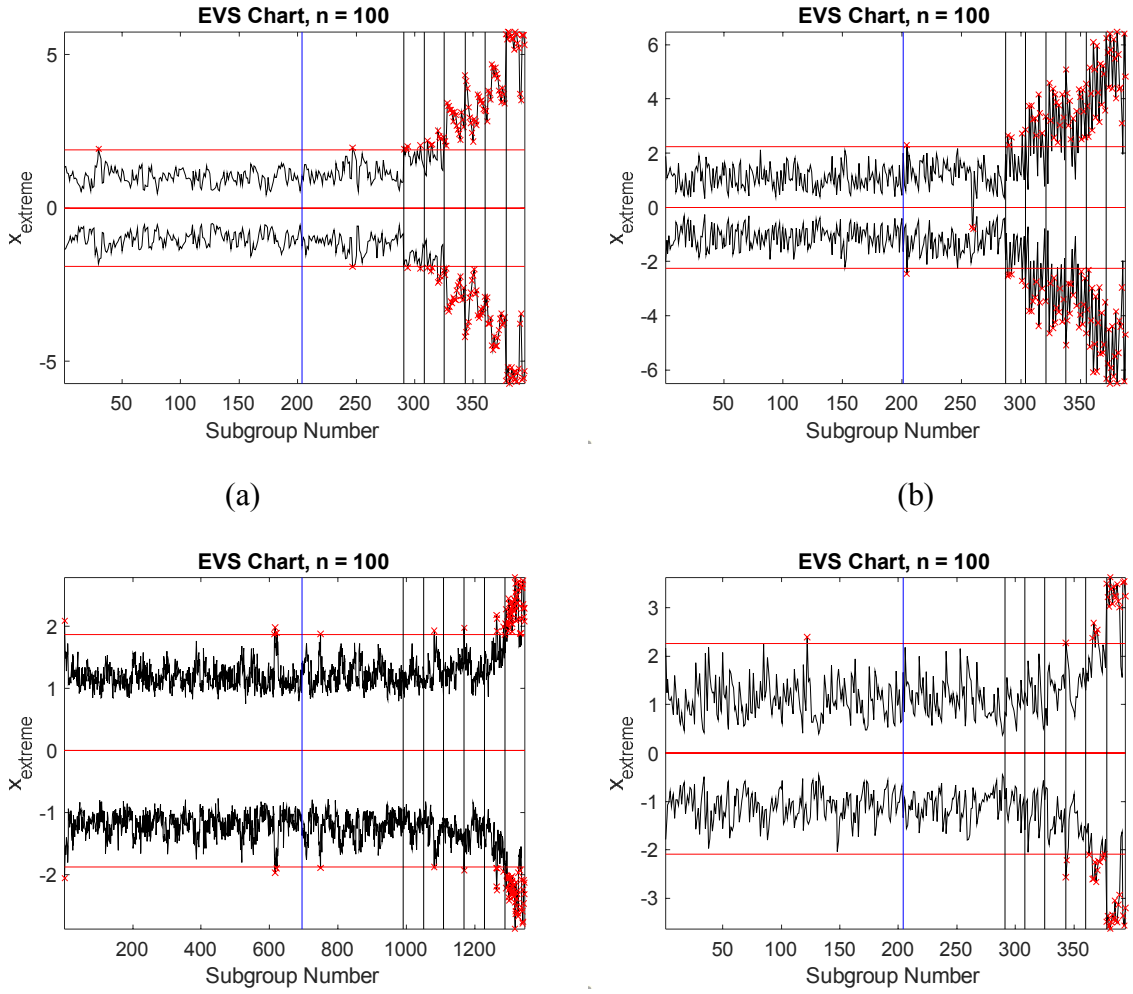


Figure 8: Damage detection using model order $m = 10$. (a) All ACFs, (b) stored and reconstructed ACFs, (c) raw measurement, and (d) ACFs of sensors 8, 11, 15, and 18. The vertical lines correspond to the end of training data (blue) and the six damage levels (black).

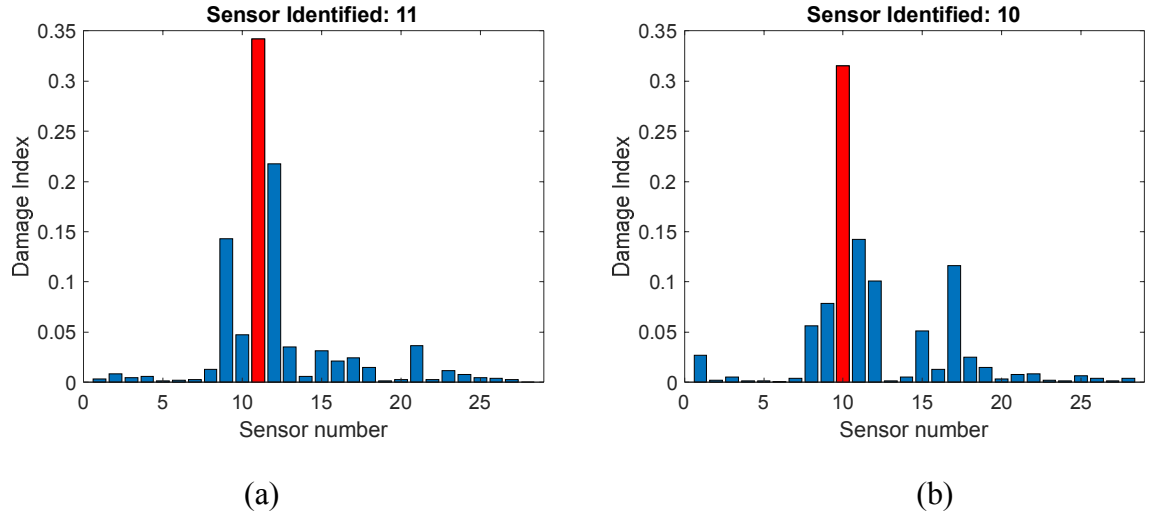
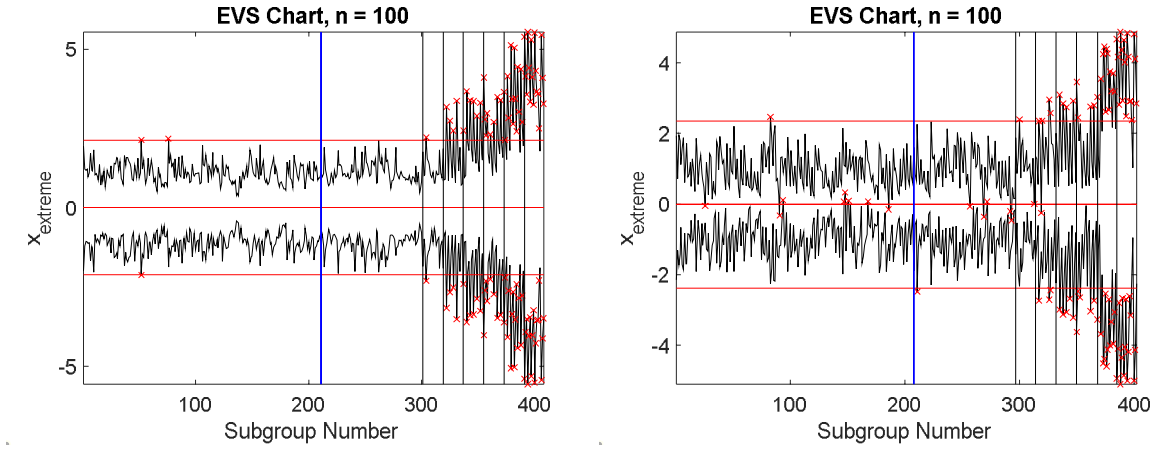


Figure 9: Damage localization. (a) All ACFs, (b) stored and reconstructed ACFs. The correct damage position was between sensors 10 and 11.



(a) (b)

Figure 10: Damage detection using model order = 0. (a) All ACFs, (b) stored and reconstructed ACFs. The vertical lines correspond to the end of training data (blue) and the six damage levels (black).

5 CONCLUSION

Data management in structural health monitoring applications is important, because the system is operating for a long time and each measurement produces lots of data. In order to save disk space without compromising damage detection performance, the following conclusions were drawn.

- Instead of using raw data, autocovariance functions were suggested as damage-sensitive features. They are much shorter signals to store, they are easy to estimate, and spatiotemporal correlation applies. Their accuracy can also be controlled with the measurement period.
- Storing only a limited set of ACFs, a further space saving can be obtained. In the given example, the additional space saving was more than 80%.
- The stored and reconstructed data were used to detect and localize damage.
- Spatiotemporal correlation was utilized both in data compression and reconstruction as well as in damage detection. The model orders need not be equal. Increasing the model order also increases the size of the covariance matrix. In data compression, the model order should be kept small making sure that the data covariance matrix is not singular and at the same time taking care that the number of stored ACFs is minimal.
- In damage detection, a full set of ACFs was used. The variability between measurements required a higher-dimensional feature vector than only the stored signals in order to consider also the environmental or operational variability.
- The original and reconstructed ACFs performed equally well in damage detection. In the given numerical experiment, it was seen that spatiotemporal correlation outperformed spatial-only correlation in damage detection.
- ACFs outperformed raw data in damage detection.
- Several functions were used in damage detection and localization: Whitening was applied to the training data to take different environmental or operational conditions into account. Principal component analysis was used to find the direction of the largest change, which

was also used in damage localization. Finally, extreme value statistics control chart was plotted to detect damage.

- Damage localization to a reconstructed DOF was possible.
- Experimental results are needed to validate the proposed technique.

ACKNOWLEDGEMENTS

This research was supported by Metropolia University of Applied Sciences.

REFERENCES

[1] R. Brincker, C. Ventura, *Introduction to operational modal analysis.* Wiley, 2015.

[2] J. Kullaa, Damage detection and localization using autocorrelation functions with spatiotemporal correlation. Z. Wu, T. Nagayama, J. Dang, R. Astroza, eds. *Experimental Vibration Analysis for Civil Engineering Structures. Lecture Notes in Civil Engineering, vol 224*. Springer, Cham. 83–95, 2023.

[3] X.Y. Li, S.S. Law, Matrix of the covariance of covariance of acceleration responses for damage detection from ambient vibration measurements. *Mechanical Systems and Signal Processing*, **24**, 945–956, 2010.

[4] M. Zhang, R. Schmidt, Sensitivity analysis of an auto-correlation-function-based damage index and its application in structural damage detection. *Journal of Sound and Vibration*, **333**, 7352–7363, 2014.

[5] M. Zhang, R. Schmidt, Study on an auto-correlation-function-based damage index: Sensitivity analysis and structural damage detection. *Journal of Sound and Vibration*, **359**, 195–214, 2015.

[6] W. Li, Y. Huang, A method for damage detection of a jacket platform under random wave excitations using cross correlation analysis and PCA-based method, *Ocean Engineering*, **214**, 107734, 2020.

[7] Z. Yang, L. Wang, H. Wang, Y. Ding, X. Dang, Damage detection in composite structures using vibration response under stochastic excitation. *Journal of Sound and Vibration*, **325**, 755–768, 2009.

[8] L. Wang, Z. Yang, T.P. Waters, Structural damage detection using cross correlation functions of vibration response. *Journal of Sound and Vibration*, **329**, 5070–5086, 2010.

[9] I. Trendafilova, A method for vibration-based structural interrogation and health monitoring based on signal cross-correlation. *Journal of Physics Conference Series*, **305**, 012005, 2011.

[10] L.-S. Huo, X. Li, Y.-B. Yang, H.-N. Li, Damage Detection of Structures for Ambient Loading Based on Cross Correlation Function Amplitude and SVM. *Shock and Vibration* **2016**, 3989743, 2016.

[11] J. Kullaa, Damage detection and localization under variable environmental conditions using compressed and reconstructed Bayesian virtual sensor data. *Sensors*, **22**, 306, 2022.

[12] H. Sohn, Effects of environmental and operational variability on structural health monitoring, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365**, 539–560, 2007.

[13] J. Kullaa, Whitening transformation in damage detection. A.E. Del Grosso, P. Basso eds. *Smart structures: Proceedings of the 5th European Conference on Structural Control — EACS 2012*, Genoa, Italy, June 18–20, 2012.

[14] D.C. Kammer, Sensor placement for on-orbit modal identification and correlation of large space structures, *Journal of Guidance, Control, and Dynamics*, **14**, 251–259, 1991.

[15] D.C. Kammer, Effects of noise on sensor placement for on-orbit modal identification of large space structures, *Journal of dynamic systems, measurements and control—Transactions of the ASCE*, **114**, 436–443, 1992.

[16] C. Papadimitriou, G. Lombaert, The effect of prediction error correlation on optimal sensor placement in structural dynamics, *Mechanical Systems and Signal Processing*, **28**, 105–127, 2012.

[17] J.S. Bendat, A.G. Piersol, *Random data: Analysis and measurement procedures. 4th Edition.* Wiley, 2010.

[18] R.W. Clough, J. Penzien, *Dynamics of structures. 2nd Edition*. McGraw-Hill, 1993.

[19] G.H. James III, T.G. Carne, J.P. Lauffer, The natural excitation technique (NExT) for modal parameter extraction from operating structures. *Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis* **10**, 260–277, 1995.

[20] H.W. Sorenson, *Parameter Estimation: Principles and Problems*. Marcel Dekker, 1980.

[21] S. Coles, *An introduction to statistical modeling of extreme values*. Springer, 2001.

[22] D.C. Montgomery, *Introduction to statistical quality control, 3rd edition*, Wiley, 1997.

[23] J. Kullaa, Robust damage detection in the time domain using Bayesian virtual sensing with noise reduction and environmental effect elimination capabilities, *Journal of Sound and Vibration*, **473**, 115232, 2020.