High Dimensional Data Reduction in Modal Analysis with Stochastic Subspace Identification

Zhilei Luo, Boualem Merainani, Michael Döhler, Vincent Baltazart, Qinghua Zhang

High Dimensional Data Reduction in Modal Analysis with Stochastic Subspace Identification

Zhilei Luo * **Boualem Merainani** ** **Michael Döhler** *
Vincent Baltazart ** **Qinghua Zhang** *

* *Université Gustave Eiffel, Inria, COSYS-SII, I4S Team, F-35042 Rennes, France*
** *Université Gustave Eiffel, COSYS-SII, F-44344 Bouguenais, France*

Abstract: Subspace system identification methods are widely used in output-only vibration analysis of civil structures, known as operational modal analysis. With the advent of new sensor technologies, such as video camera-based full field displacement or velocity measurements, the number of measured outputs is quickly increasing. In this paper, we propose principal component analysis-based data size reduction methods for efficient application of subspace methods, while preserving the high spatial resolution of the identified mode shapes for detailed modal analysis.

Keywords: Modal Analysis; Vibrations; System identification; Subspace methods; PCA

1. INTRODUCTION

In *operational modal analysis* (OMA), the monitored civil structures, such as bridges and buildings, are excited by natural ambient disturbances (road traffic loads, wind, sea water wave, etc.). Viewing the monitored structure and the equipped vibration sensors as a system, its outputs are the vibration sensor measurements, while its inputs are natural ambient disturbances, which are unknown in most practical situations. Then the OMA based on vibration sensor measurements follows an *output-only* approach. In this approach, a linear time-invariant (LTI) state space model is built from vibration measurements made on the monitored structure (Peeters and De Roeck, 1999). The parameters of interest in modal analysis are the natural frequencies, damping ratios and vibration mode shapes. They are obtained from the state transition matrix A and the output matrix C constituting the LTI model. Stochastic subspace methods are widely used to identify both matrices (A, C) that are required to obtain the modal parameters, with typical methods being the covariance-driven subspace method and the Unweighted Principal Component algorithm (UPC, "data-driven") (Benveniste and Fuchs, 1985; Van Overschee and De Moor, 1996; Peeters and De Roeck, 1999; Döhler and Mevel, 2012). The popularity of subspace methods for operational modal analysis is, on one side, due to their favorable statistical properties, like consistency even when the noise driving system is non-stationary (Benveniste and Mevel, 2007). On the other side, they are direct methods that use straightforward numerical operations like the SVD or linear least squares to obtain the estimates, which makes them numerically efficient (Döhler and Mevel, 2012).

With recently developed measurement technologies, the number of sensors that can be used for vibration analysis are rapidly increasing, and even (near) full field measurements become possible with vision-based technologies (Feng and Feng, 2018). Using image processing methods, the displacement or velocity outputs are computed from the recorded videos at possibly thousands or more positions. This leads to a great spatial resolution of the mode shapes that can be identified from the data, which is extremely useful for applications on damage diagnosis (Avci et al., 2021) or finite element model updating (Simoen et al., 2015). On the other side, the number of modes that can be identified is still limited by the sampling frequency of the camera. The explosion of the number of outputs is a numerical challenge for the subspace methods. The possible redundancy of the outputs cannot be simply alleviated by a principal component analysis (PCA) before the identification, since the full matrix C is required to carry out the modal analysis. It should also be noted that even if full field measurements are available, system identification is still required since the measurements can only cover the visible part of the structure, so the full states cannot be measured directly.

One way to handle a large number of sensors, so far, is to select a subset of the outputs, and instead of computing output covariances between all the outputs, only the covariances of all the outputs with respect to the selected subset are computed in the subspace methods (Peeters and De Roeck, 1999; Jacobsen et al., 2008). However, since this approach depends on the selection of a subset of the outputs, the information on the other outputs is not optimally exploited. Furthermore, only the number of columns of the involved covariances is reduced.

In this paper, we propose projection strategies based on the PCA of the measurement data to efficiently perform the subspace system identification from a possibly huge number of outputs, by reducing both the rows and the columns of the involved covariances in the identification algorithm. This reduces the computational burden as well as the memory usage significantly. Despite this data size

reduction, the full size output matrix C will be recovered, so that the high spatial resolution of the mode shapes is preserved for detailed modal analysis.

The paper is organized as follows. In Section 2, the background of stochastic subspace identification and modal analysis is recalled. In Section 3, the techniques for PCA-based size reduction in the subspace algorithms are developed, and applied to numerical and experimental data in Section 4.

2. STOCHASTIC SUBSPACE IDENTIFICATION IN VIBRATION ANALYSIS

2.1 Stochastic Subspace Identification

In vibration analysis, the studied vibrating mechanical structure is usually described by a finite dimensional system of ordinary differential equations (ODE), capturing the essential vibration modes. Transformed into state-space form and discretized in time (Prevosto et al., 1991; Juang, 1994), the ODE model becomes

$$\begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the internal state vector, $y_k \in \mathbb{R}^r$ is the output vector corresponding to available sensor measurements (displacement, velocity, and/or acceleration); $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^r$ are random noises modeling natural excitations of the structure (e.g., traffic load, wind, or sea wave) and measurement noise, which are assumed to be white; $A \in \mathbb{R}^{n\times n}$ is the state transition matrix and $C \in \mathbb{R}^{r\times n}$ is the output matrix. While the output dimension r is the number of available sensors, the dimension n of the state vector x_k is usually unknown, but determined by some model order selection method.

With subspace identification methods, the matrices A and C are estimated from sensor measurements y_k. In this paper, the covariance-driven subspace algorithm (Benveniste and Fuchs, 1985; Peeters and De Roeck, 1999) is applied, but the developments of this paper can be similarly applied to other subspace methods. The covariance-driven algorithm is based on the output covariances $R_i = \mathbf{E}(y_{k+i}y_k^T)$, $i = 1, 2, \ldots$, and their factorization property $R_i = CA^{i-1}G$ where $G = \mathbf{E}(x_{k+1}y_k^T)$ (Van Overschee and De Moor, 1996). When arranged in a block Hankel matrix

$$\mathcal{H} = \begin{bmatrix} R_1 & R_2 & \cdots & R_q \\ R_2 & R_3 & \cdots & R_{q+1} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p+1} & R_{p+2} & \cdots & R_{q+p} \end{bmatrix} \in \mathbb{R}^{(p+1)r\times qr} \tag{2}$$

with usually $p + 1 = q$, the system matrices A and C can be retrieved from its column space, since the Hankel matrix factorizes into observability matrix $\mathcal{O} \in \mathbb{R}^{(p+1)r\times n}$ and stochastic controllability matrix $\mathcal{C} \in \mathbb{R}^{n\times qr}$,

$$\mathcal{H} = \mathcal{O}\mathcal{C}, \text{ with } \mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^p \end{bmatrix}, \quad \mathcal{C} = \begin{bmatrix} G & AG & \ldots & A^{q-1}G \end{bmatrix}. \tag{3}$$

The matrix C is the first block row of $\mathcal{O}$, and A is obtained from the shift invariance property of $\mathcal{O}$ by

$$A = \underline{\mathcal{O}}^{\dagger}\bar{\mathcal{O}} \tag{4}$$

where $\bar{\mathcal{O}}$ and $\underline{\mathcal{O}} \in \mathbb{R}^{pr\times n}$ are respectively the parts of $\mathcal{O}$ after removing the first and the last block rows, and $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse.

Using covariance estimates $\hat{R}_i = \frac{1}{N}\sum_{k=1}^{N} y_{k+i}y_k^T$ filling the Hankel matrix $\hat{\mathcal{H}}$ corresponding to (2), the observability matrix estimate is obtained from the truncated singular value decomposition (SVD) of $\hat{\mathcal{H}}$ at model order n,

$$\hat{\mathcal{H}} = [U_1\ U_0]\begin{bmatrix}\Sigma_1 & 0 \\ 0 & \Sigma_0\end{bmatrix}\begin{bmatrix}V_1^T \\ V_0^T\end{bmatrix}, \quad \hat{\mathcal{O}} = U_1\Sigma_1^{1/2}, \tag{5}$$

and estimates $\hat{A}$ and $\hat{C}$ are obtained accordingly. Because any linear transformation of the state x_k in (1) would modify the matrices A, C while keeping y_k unchanged, $\hat{A}$ and $\hat{C}$ are estimates of A and C up to some similarity transformation. Nevertheless, the vibration modal parameters are invariant under such a transformation.

2.2 Modal Analysis

In vibration analysis, it is important to obtain the modal parameters of a vibrating structure, i.e., the natural frequencies f_i, damping ratios ξ_i and mode shapes φ_i. They are obtained based on the eigenvalues λ_i and eigenvectors ϕ_i of A, and matrix C, by

$$A\phi_i = \lambda_i\phi_i, \quad \varphi_i = C\phi_i, \tag{6}$$

and recovered via the complex-valued eigenvalues μ_i of the corresponding continuous-time system (Juang, 1994; Peeters and De Roeck, 1999) through

$$\mu_i = \frac{\log(\lambda_i)}{\tau}, \ f_i = \frac{|\mu_i|}{2\pi}, \ \xi_i = \frac{-\Re(\mu_i)}{|\mu_i|}. \tag{7}$$

If $\tilde{A} = TAT^{-1}$ and $\tilde{C} = CT^{-1}$ with any invertible matrix $T \in \mathbb{R}^{n\times n}$, then $(\tilde{A}, \tilde{C})$ has the same eigenvalues λ_i and mode shapes φ_i as in (6).

3. DIMENSION REDUCTION FOR SUBSPACE IDENTIFICATION

With the development of new sensor technologies, in some applications the number of available sensors becomes so large that the Hankel matrix in (2) exceeds the processing capability of typical computers. In this case, the Hankel matrix $\mathcal{H}$, or the sensor data size, should be somehow reduced before applying subspace identification algorithms. In (Peeters and De Roeck, 1999; Jacobsen et al., 2008), $\mathcal{H}$ in (2) has been reduced by replacing $R_i \in \mathbb{R}^{r\times r}$ with the covariances between the full output vector $y_{k+i} \in \mathbb{R}^r$ and a subset of $s < r$ outputs $y_k^{\text{sub}} \in \mathbb{R}^s$, i.e., by $R_i^{\text{sub}} = \mathbf{E}(y_{k+i}(y_k^{\text{sub}})^T) \in \mathbb{R}^{r\times s}$, which corresponds to selecting the respective s columns of R_i. Then the factorization (3) of the Hankel matrix still holds where G is replaced by G^{sub} containing the respective columns of G. The subset of outputs is chosen such that the correlation between the chosen outputs is minimal and the correlation of the chosen outputs with the remaining outputs is maximal (Jacobsen et al., 2008). In this paper, instead of using a subset of outputs for the size reduction, we propose a PCA-based reduction to profit from the information in all outputs. In addition, both the rows and columns of $\mathcal{H}$ will be reduced in order to more substantially decrease the requirement on computation resources. Moreover, despite the reduction,

the full size C matrix and the mode shapes φ_i corresponding to all the available sensors before reduction will be retrieved for detailed vibration analysis.

3.1 Covariances between full data and reduced data

The reduced size output vector y_k^{red} is formed by the principal components of the full size output vector, obtained by PCA as follows. Let Y denote the matrix of full size output data,

$$Y = [y_1 \ y_2 \ \dots \ y_N] \in \mathbb{R}^{r\times N}, \tag{8}$$

where N is the number of samples, and let $\tilde{Y} = S^{-1}Y$ be the data matrix normalized by the standard deviation of each output with $S = \text{diag}(\sigma_1, \dots, \sigma_r)$ and σ_j being the standard deviation of the j-th output. The full size correlation matrix is computed as

$$\hat{R}_0 = \tfrac{1}{N}\tilde{Y}\tilde{Y}^T \in \mathbb{R}^{r\times r}. \tag{9}$$

The singular value decomposition (SVD) of (9) yields

$$\hat{R}_0 = \begin{bmatrix} U^{\text{red}} & U_0 \end{bmatrix} \begin{bmatrix} \Sigma^{\text{red}} & 0 \\ 0 & \Sigma_0 \end{bmatrix} \begin{bmatrix} (U^{\text{red}})^T \\ U_0^T \end{bmatrix}. \tag{10}$$

In this decomposition, $U^{\text{red}} \in \mathbb{R}^{r\times s}$ corresponds to the first s principal components, and $\Sigma^{\text{red}} \in \mathbb{R}^{s\times s}$ is the corresponding singular value matrix, s is the chosen reduced data size ($s < r$) with $\min\{ps, qs\} \geq n$. The data reduction is achieved by projecting the outputs to the reduced basis

$$y_k^{\text{red}} = (U^{\text{red}})^T y_k. \tag{11}$$

Then, the right side reduced output covariance is

$$R_i^{\text{rr}} = \mathbf{E}(y_{k+i}(y_k^{\text{red}})^T) \in \mathbb{R}^{r\times s} \tag{12}$$

and the right side reduced block Hankel matrix is

$$\mathcal{H}^{\text{rr}} = \begin{bmatrix} R_1^{\text{rr}} & R_2^{\text{rr}} & \cdots & R_q^{\text{rr}} \\ R_2^{\text{rr}} & R_3^{\text{rr}} & \cdots & R_{q+1}^{\text{rr}} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p+1}^{\text{rr}} & R_{p+2}^{\text{rr}} & \cdots & R_{q+p}^{\text{rr}} \end{bmatrix} \in \mathbb{R}^{(p+1)r\times qs}. \tag{13}$$

It can be easily seen that the size of the Hankel matrix is reduced from $(p+1)r \times qr$ to $(p+1)r \times qs$. Since $R_i = CA^{i-1}G$ and $R_i^{\text{rr}} = R_i U^{\text{red}}$, the right side reduced output covariance can be factorized as

$$R_i^{\text{rr}} = CA^{i-1}(GU^{\text{red}}), \tag{14}$$

leading to the factorization

$$\mathcal{H}^{\text{rr}} = \mathcal{O}\mathcal{C}^{\text{red}}, \tag{15}$$

where $\mathcal{O}$ is defined in (3) and

$$\mathcal{C}^{\text{red}} = \begin{bmatrix} (GU^{\text{red}}) & A(GU^{\text{red}}) & \cdots & A^{q-1}(GU^{\text{red}}) \end{bmatrix}, \tag{16}$$

which is assumed to maintain full row rank. Hence, the original matrices A and C, and the corresponding modal parameters, can be consistently identified from an estimate of the right side reduced Hankel matrix $\mathcal{H}^{\text{rr}}$.

3.2 Covariances between reduced data

The above right side reduction has the advantage of preserving the original A and C matrices, despite the data size reduction. However, it reduces only the columns of the Hankel matrix to be built and analyzed, leaving the number of its rows unchanged. To further reduce the size of the Hankel matrix, both its columns and rows will be reduced. Such a both sides reduction would not preserve the original modal parameters, because it will lead to a reduced size C matrix with s rows instead of r rows. In order to overcome this drawback, an additional computation step will be proposed to retrieve the original full size C matrix and the corresponding modal parameters.

The both side reduced output covariances are

$$R_i^{\text{br}} = \mathbf{E}(y_{k+i}^{\text{red}}(y_k^{\text{red}})^T) \in \mathbb{R}^{s\times s}, \tag{17}$$

where the exponent "br" refers to "Both side Reduced", and the corresponding both sides reduced Hankel matrix

$$\mathcal{H}^{\text{br}} = \begin{bmatrix} R_1^{\text{br}} & R_2^{\text{br}} & \cdots & R_q^{\text{br}} \\ R_2^{\text{br}} & R_3^{\text{br}} & \cdots & R_{q+1}^{\text{br}} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p+1}^{\text{br}} & R_{p+2}^{\text{br}} & \cdots & R_{q+p}^{\text{br}} \end{bmatrix} \in \mathbb{R}^{(p+1)s\times qs}. \tag{18}$$

Compared with the right side reduced $(p+1)r \times qs$ Hankel matrix, the size of the both sides reduced Hankel matrix is further reduced to $(p+1)s \times qs$. The factorization of this reduced covariance becomes

$$R_i^{\text{br}} = (U^{\text{red}})^T CA^{i-1}(GU^{\text{red}}), \tag{19}$$

leading to the factorization

$$\mathcal{H}^{\text{br}} = \mathcal{O}^{\text{red}}\mathcal{C}^{\text{red}}, \text{ where } \mathcal{O}^{\text{red}} = \begin{bmatrix} (U^{\text{red}})^T C \\ (U^{\text{red}})^T CA \\ \vdots \\ (U^{\text{red}})^T CA^p \end{bmatrix} \tag{20}$$

and $\mathcal{C}^{\text{red}}$ is as defined in (16). To estimate A, $((U^{\text{red}})^T C, A)$ needs to be observable. Then, A is consistently retrieved

$$A = (\underline{\mathcal{O}}^{\text{red}})^{\dagger}\bar{\mathcal{O}}^{\text{red}} \tag{21}$$

where $\bar{\mathcal{O}}^{\text{red}}$ and $\underline{\mathcal{O}}^{\text{red}} \in \mathbb{R}^{ps\times n}$ are respectively the parts of $\mathcal{O}^{\text{red}}$ after removing the first and the last block rows.

In contrast to the factorization of $\mathcal{H}^{\text{rr}}$ in (15) yielding the full size observability matrix $\mathcal{O}$, the factorization of $\mathcal{H}^{\text{br}}$ in (20) gives $\mathcal{O}^{\text{red}}$, whose first block row is $(U^{\text{red}})^T C$ instead of C in the full size $\mathcal{O}$ as in (3). This $(U^{\text{red}})^T C$ is in fact a reduced size C matrix. If it was used in modal analysis, the resulting mode shapes φ_i^{red} would be vectors of size $s \times 1$, instead of the full size $r \times 1$ vectors φ_i. In order to avoid this loss of details in modal analysis, the full size C matrix will be computed in an extra step.

The reduced data y_k^{red} does not contain sufficient information about the full size C. For this reason more information than the both side reduced $\mathcal{H}^{\text{br}}$ is required. Consider

$$\gamma^{\text{rr}} = \begin{bmatrix} R_1^{\text{rr}} & R_2^{\text{rr}} & \cdots & R_q^{\text{rr}} \end{bmatrix} \in \mathbb{R}^{r\times qs}. \tag{22}$$

Due to the factorizations of R_i^{rr} in (14), γ^{rr} factorizes as

$$\gamma^{\text{rr}} = C\mathcal{C}^{\text{red}} \tag{23}$$

where $\mathcal{C}^{\text{red}}$ is defined in (16) and has already been computed when A was computed from (20) and (21). Then the full size matrix C is consistently recovered through

$$C = \gamma^{\text{rr}}(\mathcal{C}^{\text{red}})^{\dagger}. \tag{24}$$

This leads to the following reduced size subspace identification and full size modal parameters:

1. Compute correlation of output data in (9).
2. Compute SVD in (10) to obtain the first s principal components U^{red}, where s is chosen for the reduced data size.
3. Compute the covariances of the projected output data $(U^{\text{red}})^T y_k$ and the Hankel matrix estimate $\hat{\mathcal{H}}^{\text{br}}$ corresponding to (17)–(18).

4. Compute the SVD of $\hat{\mathcal{H}}^{\text{br}}$ and truncate it at the chosen model order

$$\hat{\mathcal{H}}^{\text{br}} = [U_1 \; U_0] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_0^T \end{bmatrix}, \tag{25}$$

$$\hat{\mathcal{O}}^{\text{red}} = U_1 \Sigma_1^{1/2}, \quad \hat{\mathcal{C}}^{\text{red}} = \Sigma_1^{1/2} V_1^T. \tag{26}$$

5. Obtain $\hat{A}$ from $\hat{\mathcal{O}}^{\text{red}}$ as in (21).
6. Compute output covariances between full and reduced data $\hat{\gamma}^{\text{rr}}$ as in (22), and estimate $\hat{C}$ as in (24).
7. Compute modal parameters through (6) and (7).

4. APPLICATION TO A CANTILEVER BEAM

In this section, the proposed data size reduction schemes are applied for modal analysis of simulated data from a finite element beam model as well as on experimental data from its counterpart in the lab, comparing:

- **FULL.** Full data size identification method (Section 2.1)
- **PCACOR-rr.** Proposed right side reduction scheme on Hankel matrix (Section 3.1)
- **PCACOR-br.** Proposed both sides (left and right) reduction scheme on Hankel matrix (Section 3.2)

The numerical application is used to validate the developed methods in a Monte Carlo simulation, before applying the methods to an experimental data set.

4.1 Description of test structure

The test specimen is a cantilever beam of size 30 mm × 2000 mm × 5 mm made out of steel, and mounted on a shake table in the Structures and Integrated Instrumentation Laboratory at Gustave Eiffel University, see Fig. 1 (left). A numerical counterpart, namely a 3D finite element (FE) model has been made for simulation and comparison. The Timoshenko beam model, adopted from (Przemieniecki, 1985) and illustrated in Fig. 1 (right), has Young's modulus of $2.1 \cdot 10^{11}$ Pa, mass density of 7850 kg/m^3 and Poisson's ratio of 0.21. It is composed of 160 beam elements. Each node has three translational and three rotational degrees of freedom (DOFs), i.e. d_x, d_y, d_z and r_{xy}, r_{yz}, r_{xz}. The DOFs at node 1 are constrained.

4.2 Numerical Results

In a Monte Carlo simulation, $N_{MC} = 100$ displacement vibration data sets of the beam are generated from Gaussian white noise excitation at all 160 DOFs of the beam in the (horizontal) x-direction, with time step $\tau = 1/500$ for a data length of $N = 10^5$. Gaussian white measurement noise is added on each output with 5% of the standard deviation of the respective output signal. Only the generated output data are used in the following, but not the inputs.

The datasets are processed with the proposed methods, using parameters $p + 1 = q = 61$ for the block Hankel matrices. For the size reductions, the first three principal components of the output correlations are used, as the sum of their respective singular values represent over 99% of the total variance in the PCA. They are selected for the projection of the output signals for the data reduction.

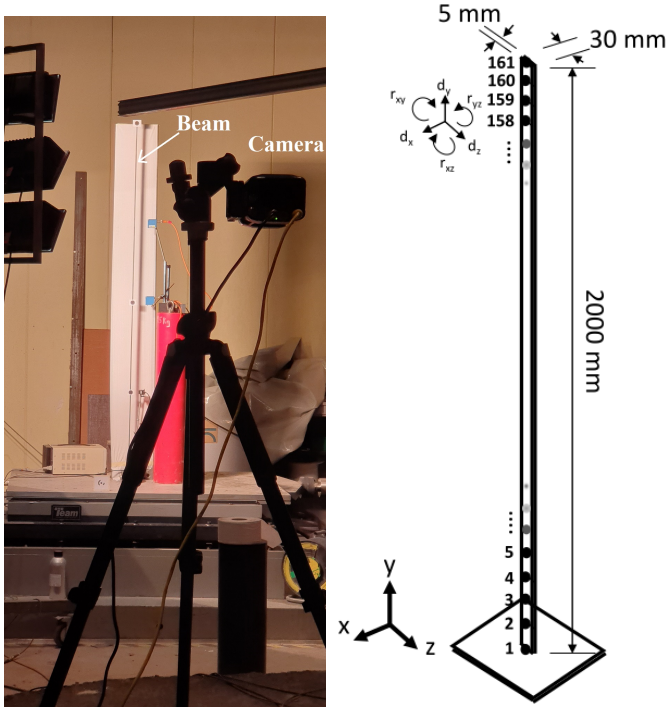

Fig. 1. Experimental setup and finite element (FE) model of cantilever beam

The computational times for the entire processing of one dataset are compared for both reduction techniques and the full data size identification in Table 1. The results show that in our setting the computation times are reduced by factor 40 from the full data to PCACOR-rr, and by another factor of 15 when using the PCACOR-br scheme. These results show a good computational efficiency of the data size reduction schemes.

Table 1. Computational time of identification algorithms for one dataset.

Data size reduction schemes	Computational time (seconds)
FULL	170.4
PCACOR-rr	4.3
PCACOR-br	0.3

To evaluate the precision and the accuracy of the approach, the bias and standard deviation of the identified frequencies is evaluated from the Monte Carlo simulation. Eight structural modes are identified in all datasets. In Fig. 2, the empirical bias and standard deviations of all the identified frequencies are shown, normalized by the frequencies, as

$$E_i = \frac{|\bar{f}_i - f_i|}{f_i}, \quad D_i = \frac{\sigma_{f_i}}{f_i}, \tag{27}$$

respectively, where $\bar{f}_i$ and σ_{f_i} are the empirical mean and standard deviation from the Monte Carlo simulation. The results exhibit that the bias is very low and remains in the same range for all the methods. For some of the modes, the bias with the size reductions is even smaller than with the full size data. Similarly, the standard deviations with the size reductions are getting mostly smaller when using the size reduction techniques. These results show that for most of the identified frequencies there is no loss in accuracy or precision when using the size reduction techniques.

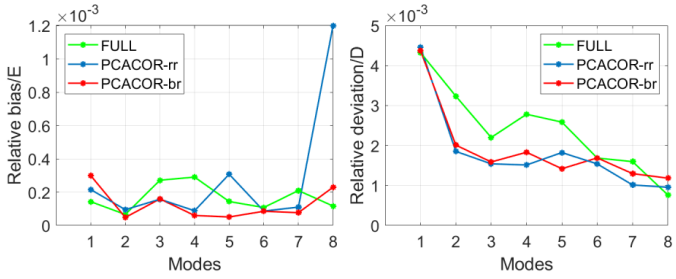

Fig. 2. Relative bias (left) and standard deviation (right) of identified natural frequencies of the eight modes

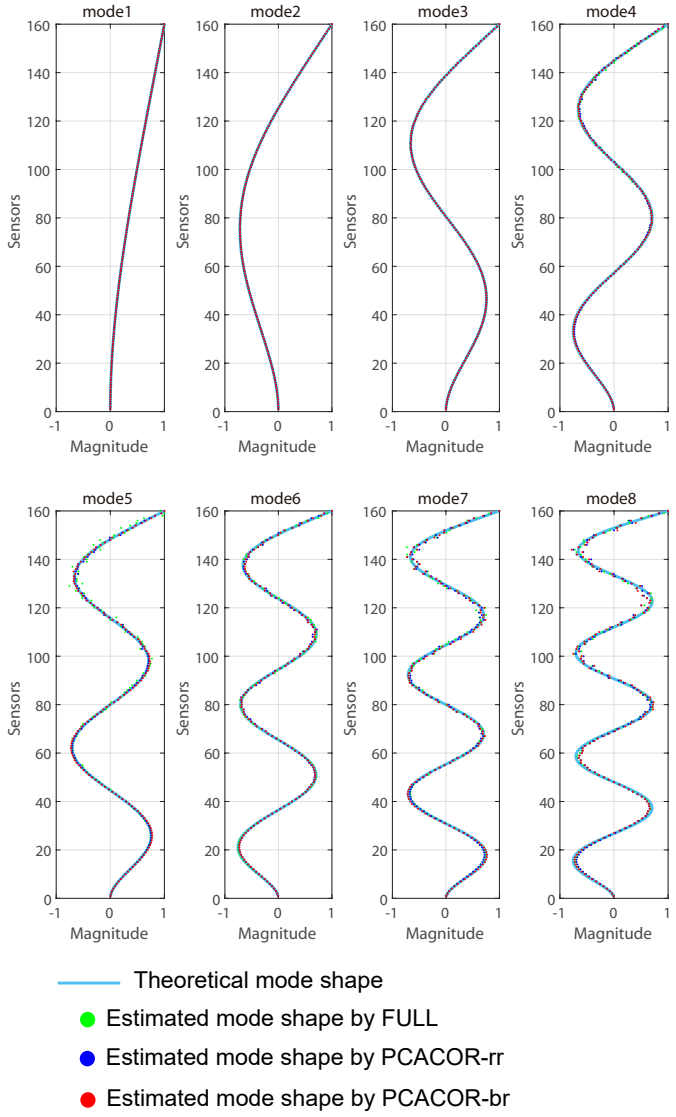

Fig. 3. Identified mode shapes of the beam from a simulated dataset with and without the reduction techniques, and numerical mode shapes.

Finally, the accuracy of the mode shape estimation is evaluated on one dataset. The identified mode shapes are presented in Fig. 3, where they are compared to the theoretical mode shapes from the numerical model. It can be observed that the mode shapes are identified very well with high spatial resolution, even with the data reduction using merely three principal components. Few differences are only visible for the higher modes. This

Table 2. MAC of mode shapes in simulation

mode	FULL	PCACOR-rr	PCACOR-br
1	1.000	1.000	1.000
2	1.000	1.000	1.000
3	1.000	1.000	1.000
4	0.998	0.999	1.000
5	0.998	0.998	0.999
6	0.997	0.996	0.998
7	0.989	0.993	0.992
8	0.990	0.988	0.984

can also be validated quantitatively in Table 2, where the modal assurance criterion (Allemang, 2003) between the identified and the theoretical modes from the numerical model is shown. The value of modal assurance criterion (MAC) between two mode shapes is defined as

$$MAC(\varphi, \psi) = \frac{|\varphi^H \psi|^2}{(\varphi^H \varphi)(\psi^H \psi)}. \tag{28}$$

It is bounded between 0 and 1, where 1 indicates a perfect match. The MAC values are all very close to 1, and there seems to be no loss in accuracy when using the data size reduction. This shows a high accuracy in mode shape identification with the proposed data reduction scheme.

4.3 Application on Experimental Data

Finally, the algorithms have been applied to experimental data from the beam in the laboratory. A PHANTOM MIRO 320S camera with resolution 1920×1200 was used to record video flows of the beam while the shake table was driven in random mode horizontally, perpendicular to the camera axis. The frame rate was 512 frames/s, ensuring identification of the first eight modes. The excitation noise was band limited between 1 Hz and 200 Hz, and datasets of 33 s length were recorded. Displacements were extracted in 160 regions of interests (ROIs), coinciding with the nodes of the numerical model, through a two-step process: estimation of the integer pixel shift using cross correlation (Sutton et al., 2009) and subpixel shift by quadratic surface fitting (Pan et al., 2005) for refinement.

From the obtained dataset, the modal parameters were estimated with the presented methods, see Table 3, which are all in very good agreement. In Fig. 4, the identified mode shapes are plotted, which are of good accuracy and show no degradation with the reductions. This is also quantified in Table 4, where the MAC values between the mode shape estimates and the numerical FE mode shapes are shown. No significant difference with or without the size reductions is visible, indicating good accuracy of the proposed methods also on experimental data.

Table 3. Estimated frequencies f (Hz) and damping ratios ξ (%) from experimental data

mode	FULL		PCACOR-rr		PCACOR-br	
	f	ξ	f	ξ	f	ξ
1	0.999	0.16	0.996	0.18	1.009	0.52
2	6.741	0.18	6.742	0.19	6.741	0.19
3	19.01	0.09	19.02	0.10	19.01	0.13
4	37.35	0.15	37.34	0.16	37.35	0.16
5	62.03	0.14	62.03	0.14	62.04	0.17
6	92.57	0.05	92.57	0.05	92.57	0.05
7	129.7	0.08	129.7	0.02	129.7	0.19
8	172.1	0.06	172.1	0.06	172.1	0.06

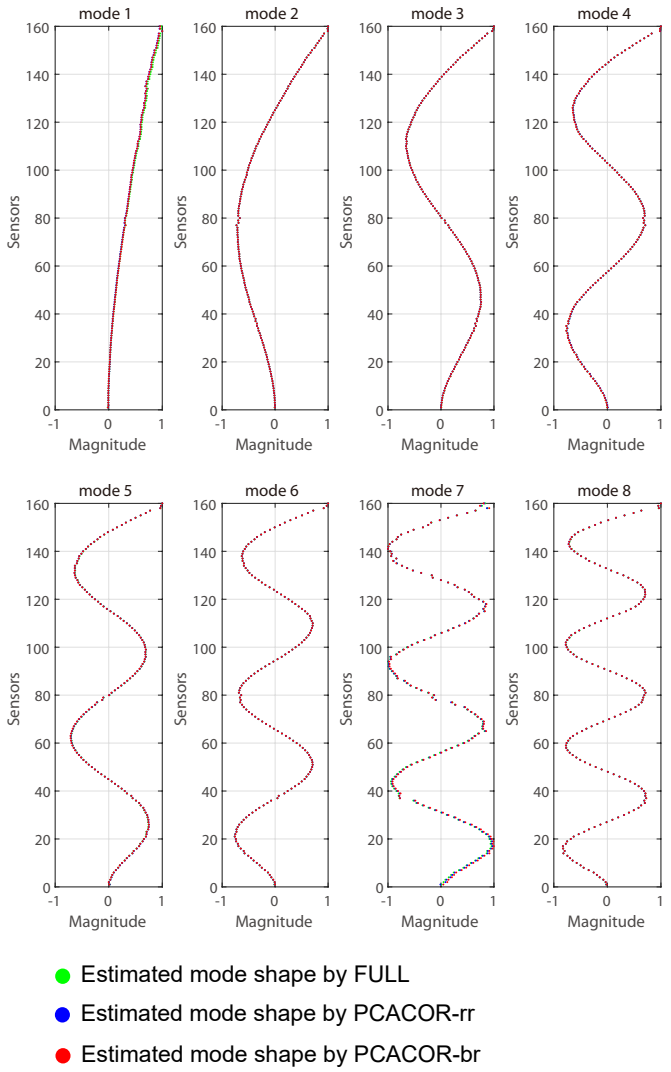

Fig. 4. Mode shapes of beam from experimental data.

Table 4. MAC of experimental beam mode shapes with respect to numerical mode shapes

mode	FULL	PCACOR-rr	PCACOR-br
1	0.999	0.999	0.999
2	1.000	1.000	1.000
3	0.999	0.999	0.999
4	0.999	0.999	0.999
5	0.998	0.998	0.998
6	0.998	0.998	0.998
7	0.962	0.964	0.964
8	0.995	0.995	0.995

5. CONCLUSION

In this paper, a novel data reduction scheme based on principal component analysis is proposed for subspace-based modal analysis with high dimensional output data. The proposed method is first validated on numerical motion data from a finite element model of a vibrating cantilever beam. The computational time is reduced by factor 600 compared to the full data size computation. No loss in accuracy nor precision has been observed in a Monte Carlo simulation, while the high spatial resolution and accurate estimates of mode shapes are preserved. The proposed schemes have also been applied successfully on experimental vibration data from vision-based motion estimation, which is promising for applicability to full field measurements. Future research includes the optimal choice of the number of principal components.

ACKNOWLEDGEMENTS

The support from the ANR “France Relance” program is gratefully acknowledged.

REFERENCES

Allemang, R.J. (2003). The modal assurance criterion (MAC): Twenty years of use and abuse. *Sound and Vibration*, 37(8), 14–21.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., and Inman, D.J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing*, 147, 107077.

Benveniste, A. and Fuchs, J.J. (1985). Single sample modal identification of a nonstationary stochastic process. *IEEE Transactions on Automatic Control*, AC-30(1), 66–74.

Benveniste, A. and Mevel, L. (2007). Nonstationary consistency of subspace methods. *IEEE Transactions on Automatic Control*, AC-52(6), 974–984.

Döhler, M. and Mevel, L. (2012). Fast multi-order computation of system matrices in subspace-based system identification. *Control Engineering Practice*, 20(9), 882–894.

Feng, D. and Feng, M.Q. (2018). Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection–a review. *Engineering Structures*, 156, 105–117.

Jacobsen, N.J., Andersen, P., and Brincker, R. (2008). Applications of frequency domain curve-fitting in the EFDD technique. In *Proc. 26th International Modal Analysis Conference (IMAC)*.

Juang, J.N. (1994). *Applied system identification*. Prentice Hall, Englewood Cliffs, NJ, USA.

Pan, B., Xu, B., Chen, D., and Feng, J. (2005). Sub-pixel registration using quadratic surface fitting in digital image correlation. *ACTA Metro. sinica*, 26(2), 128–134.

Peeters, B. and De Roeck, G. (1999). Reference-based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing*, 13(6), 855–878.

Prevosto, M., Olagnon, M., Benveniste, A., Basseville, M., and Le Vey, G. (1991). State space formulation: a solution to modal parameter estimation. *Journal of Sound and Vibration*, 148(2), 329–342.

Przemieniecki, J.S. (1985). *Theory of matrix structural analysis*. Courier Corporation.

Simoen, E., De Roeck, G., and Lombaert, G. (2015). Dealing with uncertainty in model updating for damage assessment: A review. *Mechanical Systems and Signal Processing*, 56, 123–149.

Sutton, M.A., Orteu, J.J., and Schreier, H. (2009). *Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications*. Springer.

Van Overschee, P. and De Moor, B. (1996). *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer, Dordrecht, The Netherlands.