Fuzzy-driven strategy for fully automated modal analysis: Application to the SMART2013 shaking-table test campaign

P.-É. Charbonnel

Version of Record: https://www.sciencedirect.com/science/article/pii/S0888327020307743
Manuscript_ad5a368c6dc45dc3884a8a59f9451181

Fuzzy-driven strategy for fully automated modal analysis: application to the SMART2013 shaking-table test campaign

P.-É. Charbonnel[a]

[a]*DES-Service d'Études Mécaniques et Thermiques (SEMT), CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*

Abstract

A crucial step when identifying the modal signature of systems using growing order parametric methods consists in discriminating spurious modes from physical modes. In this paper, a three-stages clustering strategy is presented in a fuzzy framework for automating this selection process in the context of Input/Output and Output-Only identification. The novelty and strong point of the approach lies in the first stage where, after computation of single mode validation indicators, a modified *fuzzy c-means* clustering procedure is developed for performing a first partition. It is shown how the membership function obtained for the cluster of physical modes can be interpreted as a new synthetic modal indicator and helps with pole-splitting detection, outlier rejection and generally improves the final modal parameters estimation. The developed methodology does not involve any user-specified threshold and can be used for discriminating modes produced by any methodology consisting in fitting a growing order model to experimental data of any type. In this paper, accelerations measured during the SMART2013 shaking-table test campaign are processed using data-driven state-space identification algorithms. The automated selection process is used for tracking the modal signature of a trapezoidal shaped reinforced-concrete specimen using in turn stochastic and combined deterministic-stochastic algorithms, defining for the latter the movement of the shaking table as input. Variations in the modal signature are then correlated to the damage actually observed on the specimen and a comparison between Output-Only and Input/Output results is made in order to estimate the interaction between the specimen and the whole shaking table device.

Keywords:
Operational/Experimental Modal Analysis, Stabilization Diagram, Automation, Fuzzy Clustering, Structural Health Monitoring, Shaking table tests

1. Introduction

Analyzing the response of real structures is an essential preliminary step to modeling and control of mechanical systems. For linear structures, the *modal signature* composed of

Email address: pierreetienne.charbonnel@cea.fr (P.-É. Charbonnel)

frequencies, damping ratios and modeshapes summarizes this response. In order to characterize this key feature, measurements performed on real structures in operational context or on specimen under experimental dynamic loading are processed using identification methods described either in a time or frequency framework (see *e.g.* [1] and [2] to only mention one of the most popular references of each case and [3] for a more complete overview). Those methods lie in fitting a linear model to time or frequency data; the modal signature is then extracted after eigenvalue decomposition of the matrix describing the dynamic behavior of the system. The order n of the model to be fitted being unknown, a current practice consists in performing identification for increasing values of n. One then needs to distinguish the sought physical modes from mathematical artifacts that arise owing to model-order overspecification.

This crucial discrimination step is generally done with the help of *stabilization diagrams* where physical modes can be detected as columns in a frequency versus model-order plot. However, a careful inspection of stabilization diagrams still requires time and the trained eye of an experimented user. This paper then focuses on automating the physical modes selection process among poles identified for large model order range based on a single data set. This *automated modal analysis* must be distinguished from *modal tracking* that relies on existing knowledge on modal parameters (based on numerical model or reference data) to detect changes in the modal signature that can further be related to global or local damage [4], correlated with environment conditions [5, 6] or be used for instability or fault detection [7, 8]. At the end of this paper, the evolution the modal signature of a reinforced concrete specimen is sought on the basis of repeated analysis on different data sets; no online monitoring is performed.

The attempts for automating or at least assisting the interpretation of stabilization diagrams are not new (see *e.g.* [9, 10] and [11] for a more extensive bibliographic review). Early references relied on the computation of *modal indicators*: plotting modes with high indicator values leads to clearer stabilization diagrams. Among the most effective indicators, one can mention the Modal Phase Collinearity (MPC) [12], the Modal Transfer Norm (MTN) [13], the Modal Coherence Indicator (MCI) [14] and the Mean Phase Deviation (MPD) [9]. Other authors [15, 16, 8, 17] rely on uncertainty computation for discarding spurious modes; data sets are divided in (statistically independent) blocks and uncertainty bounds on modal parameters are obtained from perturbation theory. The poles whose modal properties have the largest standard deviations are declared spurious. Different clustering techniques have also been used for automating the selecting process usually referred to as *hierarchical clustering*, *partitioning methods* and *histogram analysis* (see ref. [9] for details). Among the most effective and followed attempts, one can cite [18] that defined a methodology based on hierarchical clustering for grouping similar modes clusters and successfully performed the modal identification of a bridge.

The main drawbacks of the aforementioned references is the systematic presence of user-specified thresholds on modal distances for achieving proper clustering or on modal indicators for discarding spurious modes (see also [19]). Those thresholds depend on the application, on the amount and quality of available data, on the level of solicitation and must sometimes be relaxed depending on (expected) modeshapes complexity. Such methodologies have proven effective in the case of long-term monitoring where single structures are instrumented with

permanent sensing devices; in that case the time spent for correctly adapting the thresholds to the problematic is of less importance compared to monitoring time. However, the seismic tests conducted in the TAMARIS/CEA facility impose a different context: the zoology of experimental systems is vast (reinforced concrete building [20], steel-frame structures [21, 22], timber-frame construction [23], tanks [24], etc.), the ground motion tests numerous, involving possibly different configurations and a wide variety of sensors. Manual user-dependent threshold calibration is a huge obstacle to automation and might be intractable in the case of specimens exhibiting important modal signature variations throughout test sequence. To overcome this obstacle, Reynders et al. in [9] proposed an efficient strategy for automating the mode selection process in three stages:

(i) Several modal indicators are first computed for each mode and a *k-means* ($k = 2$) algorithm allows to remove the modes interpreted as certainly spurious modes,

(ii) A hierarchical clustering stage is then carried out on the subsisting family containing possibly physical modes for detecting groups of similar modes identified for different values of n,

(iii) A last partition stage is conducted for retaining only the most populated groups generated at stage (ii). This last group contains the physical modes because they are expected to be identified for every n.

The maximum within-clusters distance between representations of the same physical mode for different system orders is the decisive criterion for stopping the hierarchical clustering procedure in stage (ii). This criterion, that had to be specified by the user in [18], is derived from the statistics of the group selected after stage (i) in [9].

This strategy offers a high degree of automation, however some limitations have been experienced. The *k-means* algorithm used in stage (i) is not necessarily the best option for building the first partition, as it was also noticed in [25]. For most indicators computed in stage (i), the cluster of possibly physical modes resembles an exponential distribution whereas the cluster of spurious modes tends to be normally distributed. The *k-means* algorithm is always dominated by the variables with highest variance and, when applied to normally distributed data, the algorithm tends to split datasets in approximately equally sized clusters [26]. For an optimal partitioning, one would thus have to choose the maximum order n from the number of expected observable modes (that is not known in advance) or at least specify a sufficiently large maximum n in order to guarantee that no possible physical modes will be discarded after stage (i). Moreover, when plotting the respective cardinals of the groups identified after hierarchical clustering at stage (ii), in most applications, no clear jump is noticeable, *i.e.* no clear partition between 'heavily populated groups' and 'poorly populated groups' can be made by the *k-means* algorithm that is used at stage (iii) for the final selection. Once again, *k-means* will tend to build clusters of equal size.

The present contribution proposes an enhancement of the three-stages selection strategy described in [9] with the idea of overcoming the above-mentioned limitations. The key enhancement consists in replacing the *k-means* algorithm of stage (i) by a more adaptive *fuzzy c-means* algorithm that is modified to bring a better insight in the data to be split.

The membership function, computed at stage (i), can then be interpreted as a synthetic modal indicator and will help the partitioning at stage (iii). At the end of the proposed algorithm, the means and standard deviations will be given for physical clusters in a fuzzy sense (weighted by the membership function), what will improve the overall quality of the modal estimates, empowering the most reliable candidates of each group and will help dealing with outliers and pole splitting phenomenon in the same time.

The robustness of the improved algorithm is illustrated using the measurements of the SMART2013 test campaign [20] during which a reduced scale Reinforced Concrete (RC) building was submitted to a sequence of shaking table tests. The sequence is composed of seismic tests of increasing level, that gradually activate non-linear mechanisms on the specimen, and intermediate low-level broad-band tests. The accelerations measured during intermediate tests are processed using subspace-based identification algorithms and the new pole selection strategy is carried out for an automated tracking of the modal signature along the test campaign. The evolution of the modal signature will be correlated to the damage actually observed on the specimen and a comparison between Output-Only (OO) and Input/Output (IO) results will be made in order to estimate the interaction between the specimen and the shaking table device.

The paper is organized as follows. Section 2 gives a brief recap of the SMART2013 test campaign, describes the processed dataset and details how subspace-based algorithms are tuned. The single mode validation criteria used in this paper are also defined, paying particular attention to a newly derived indicator, called *normalized Modal Transfer Norm* (nMTN) difference, that is developed for promoting the detection of weakly excited modes. Section 3 describes the fuzzy-enhanced three stages procedure that is central to this contribution. A specific focus is made on the modified fuzzy c-means procedure that has been derived at stage (i) for making a first rejection of artifacts. After quantifying its performance on clusters of known distribution, the procedure is carried out for modal analysis purpose. The way the resulting membership function is reused to enhance the procedure is illustrated using OO measurements performed at the beginning of the SMART2013 test campaign where the RC specimen is considered healthy. An thorough discussion on the comparative use of k-means vs. fuzzy-c-means algorithms at stage (i) is also present. The effect of this first partition on yielding hierarchical clustering threshold used at stage (ii) is particularly studied and this for different modal indicators-sets whose influence is also investigated. Section 4 presents the modal analysis results and the evolution of the modal signature throughout the test campaign, obtained from repeated stochastic OO and combined deterministic-stochastic IO identification. Section 5 finally gives conclusions.

2. Experimental setup and algorithms for modal identification

2.1. The SMART2013 shaking table test campaign

End 2013, within the framework of the EDF-CEA SMART project, a three-story RC specimen was tested on the six degrees of freedom AZALEE shaking table of the TAMARIS/CEA facility. The singular trapezoidal design of the RC specimen (see fig. 1) was chosen such as to emphasize torsional effects to which constructions are subjected during seismic loading. The $6 \times 6m^2$ AZALEE shaking table is equipped with eight 1000kN maximum capacity

hydraulic MTS actuators and can reproduce complex seismic loading composed of six independent components (three linear accelerations and three angular accelerations). The size of the table and its movement capacity led to the construction of a 12 tons specimen, built with respect to classical EC8 design, and with dimensions homogeneously reduced of a $s = 1/4$ scale factor. For respecting Cauchy-Froude's similitude law [27] (accelerations, strains, stresses and resistance parameters are preserved), the ground motion seismic inputs have been contracted in time of a factor $\sqrt{s}$ and a total of 33.8 tons additional masses, in a ratio $(1-s)/s$ w.r.t. reference full scale configuration, have been equally distributed on each story of the specimen (see fig. 1).

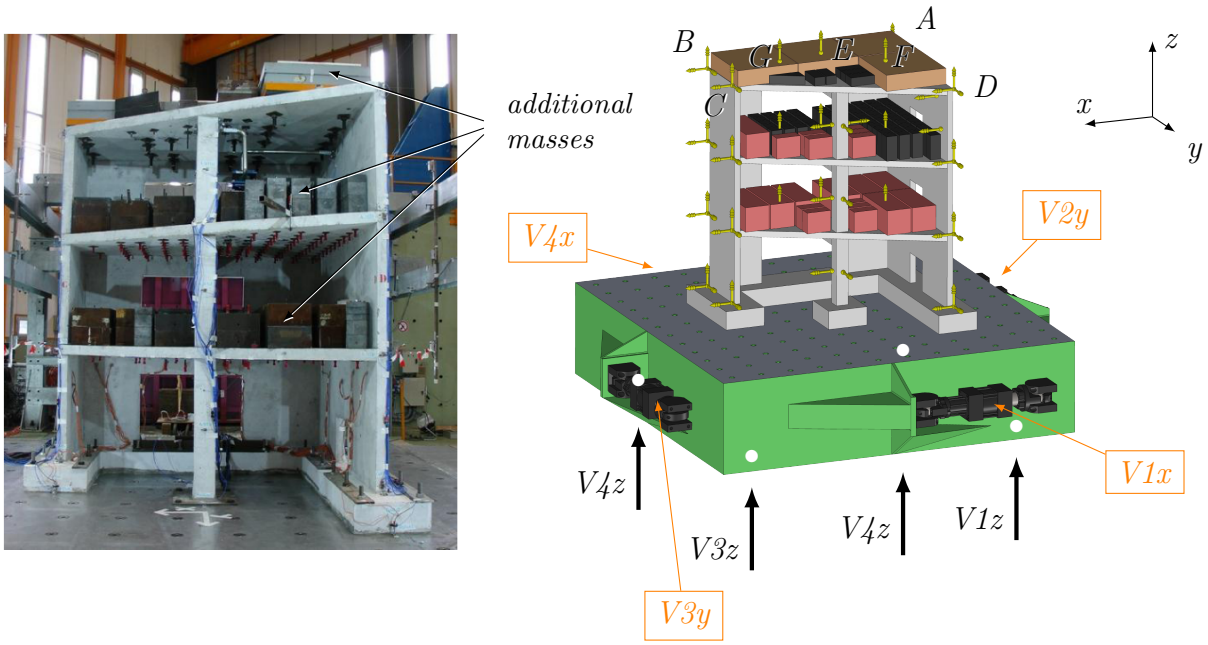


FIGURE 1: The SMART RC specimen anchored to the AZALEE shaking table with actuators and sensors position. The yellow arrows indicate the position of the accelerometers used for modal analysis. Only the horizontal actuators ($V1x$, $V2y$, $V3y$ and $V4x$) are represented, the position of the vertical actuators ($V1z$, $V2z$, $V3z$ and $V4z$) underneath the AZALEE table is labeled with white dots and their action marked with black arrows.

The specimen is instrumented with more than 200 sensors including 64 accelerometers placed on the RC specimen. The accelerometers are labeled on figure 1 with yellow arrows. Three accelerations are recorded at each corners A, B, C D of the trapeze on each story (including soleplate level), vertical accelerations are measured in-between the masses at floor level for points E, F and G, and 7 accelerometers are placed on the central column. In addition to that, the accelerations of the eight hydraulic actuators rods have been recorded during the whole test campaign. Further information regarding data acquisition strategy and experimental results can be found in [20].

Table 1 gives a brief recap of the SMART2013 test campaign. This one consists of an alternation of bi-axial seismic tests (gray lines of the table) of increasing level, which one wishes to be gradually damaging for the specimen, and of random tests (white lines) with an acceleration level chosen such that the first eigenmodes of the experimental system are

excited but without adding further damage to the RC medium. For this last type of test, the reference ground motion is defined as a bi-axial random white noise; more precisely, the reference acceleration $\ddot{u}_c$ sent to the MTS controller (see fig. 2) is defined as bi-axial white noise with decorrelated horizontal components and the level is chosen such as having a RMS acceleration level $\mathbf{E}\left(\|\ddot{u}_m\|\right) = 0.02g$ measured at the center of the shaking table.

Phase 1 : Signal SMART 2008 – PGAx = 0.2g – PGAy = 0.2g	
RUN6	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN7	Seismic signal 50%
RUN8	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN9	Seismic signal 100%
Phase 2 : Northridge signal (main shock) – PGAx = 1.78g – PGAy = 0.99g	
RUN10	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN11	Seismic signal 11%
RUN12	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN13	Seismic signal 22%
RUN14	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN15	Seismic signal 22%
RUN16	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN17	Seismic signal 44%
RUN18	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN19	Seismic signal 100%
Phase 3 : Northridge signal (after-shock) – PGAx = 0.37g – PGAy = 0.31g	
RUN20	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN21	Seismic signal 33%
RUN22	Broad-band ergodic signal (X+Y) 0.02g RMS
RUN23	Seismic signal 100%
RUN24	Broad-band ergodic signal (X+Y) 0.02g RMS

TABLE 1: Recap of the SMART2013 test campaign. The biaxial damaging seismic tests are highlighted in gray whereas the intermediate low-level biaxial random tests stay in white.

The *experimental system* is composed of every elements represented in figure 1, but also contains the MTS controller operating in closed loop as shown on figure 2. The *closed loop experimental system*, denoted $\mathfrak{Z}$ is thus defined as :

$$\mathfrak{Z} = \left\{ \begin{array}{c} \text{RC Specimen + Additional masses + AZALEE shaking table} \\ \text{Closed-loop controlled hydraulic actuators} \end{array} \right\} \tag{1}$$

Let one stress that the Three Variable Controller (TVC) provided by the manufacturer MTS is also involved in the system.

From this system, a total of 72 accelerations are acquired at a sample frequency $f_s = 1000Hz$ and filtered with $400Hz$ cut-off frequency anti-aliasing filters. This set of measurements, with typical $0.003g$ RMS noise level, will be processed, for each RUN of the test campaign, by subspace-based identification methods for the purpose of modal analysis.

2.2. Tuning of the subspace-based identification algorithms

Subspace-based identification methods form a class of algorithms dedicated to the identification of linear systems via state-space modeling. From the presumed order n of the

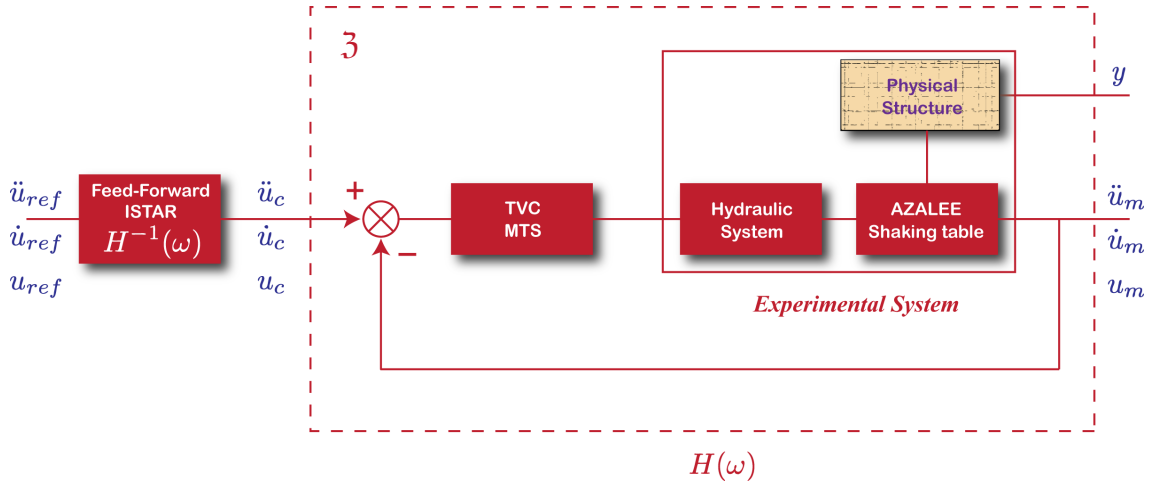


FIGURE 2: Closed-loop experimental system $\mathfrak{Z}$.

tested system and knowing a given number of output (and possibly input) measurements, the system matrices can be retrieved after robust orthogonal (resp. oblique) projections. The modal signature $(\lambda_j, \phi_j)_{j=1}^{n}$ of the system is then obtained after eigenvalue decomposition of the state-space dynamic transition matrix. No further details about subspace-based identification will be given here for the seek of conciseness. In this work the data-driven implementation described in [1] has been used specifying a CVA[1] weighting for OO analysis and combined deterministic-stochastic algorithm with unitary weighting in the IO case.

For IO analysis, two configurations are considered:

IO-1 : Firstly, defining as input the biaxial acceleration movement extracted from the hydraulic actuators sensors. In this case, the modal signature identified will be that of a subsystem $\bar{\mathfrak{Z}}_1$ defined as:

$$\bar{\mathfrak{Z}}_1 \;=\; \Big\{\text{RC Specimen} + \text{Additional masses} + \text{AZALEE shaking table}\Big\} \tag{2}$$

IO-2 : Secondly, the biaxial movement of the soleplate (floor 0) extracted from the bottom accelerometers is defined as input. The identified modes this time will be those of the smaller subsystem $\bar{\mathfrak{Z}}_2$ composed of:

$$\bar{\mathfrak{Z}}_2 \;=\; \Big\{\text{RC Specimen (cantilivered at soleplate)} + \text{Additional masses}\Big\} \tag{3}$$

The number i of block Hankel matrices has been chosen, as advised by [13], such as to verify:

$$i \geq \frac{f_s}{2f_0} \tag{4}$$

where $f_s = 1/\Delta t$ is the sample frequency and $f_0 \approx 5Hz$ is the smallest expected eigenfrequency. The maximum order n_{max} of the state-space model to be interpolated to the

[1]Canonical Variate Analysis

dataset is generally chosen as large as possible (at least twice the number n_{ph} of observable modes expected on the frequency range of interest). In this study, the value $n_{max} = 300$ is chosen. One ends up with a maximum number of $J \leq n_{max}(n_{max}+1)$ mode candidates $(\lambda_j, \phi_j)_{j=1}^{J}$ coexisting in stabilization diagrams, from which one would like to extract the different variations of the n_{ph} physical modes (columns of the diagram).

One can note that, in this paper, the order n is the only tuning parameter that is investigated; no variations of i (as done for example in [19]) is taken into account, but the presented methodology could perfectly enable the partition of the broader mode set $(\lambda_j, \phi_j)_j$ that would have been generated if multiple values of i were to be considered. Regarding the SMART2013 dataset, the too low values of i that did not meet the criterion (4) led to irrelevant results (even failing to catch low frequency modes) and too high values led to CPU limitations. Several values of i within $[100; 200]$ have been tested and produced extremely similar results, at least for initial configuration (RUN6) in the OO case; the value $i = 200$ is chosen for each even RUN of table 1 to account for expected frequency drop ($f_0 \approx 2.5Hz$ at the end of the test campaign).

2.3. Single-mode validation criteria

Following the lines of [9], as preliminary step to automated selection process, several numerical indicators are computed for each mode of the set $(\lambda_j, \phi_j)_{j=1}^{J}$ for quantifying their ability to be 'good mode candidates'. Table 2 synthesizes the different indicators tested in this work.

Indicator	☐	Definition	Ideal phys. mode (C_1)	Typical artifact (C_2)
X_1	✓	$\Delta(f_j, f_{\hat{j}})$	0	1
X_2	✓	$\Delta(\xi_j, \xi_{\hat{j}})$	0	1
X_3	✓	$\Delta(\lambda_j, \lambda_{\hat{j}})$	0	1
X_4	✓	$1 - \mathrm{MAC}(j, \hat{j})$	0	1
X_5	✓	$\Delta(\mathrm{MTN}_j, \mathrm{MTN}_{\hat{j}})$	0	1
X_6	✓	$\Delta(\mathrm{MPD}_j, \mathrm{MPD}_{\hat{j}})$	0	1
X_7	✓	$n\mathrm{MTN}_j$	0	1
X_8	✓	MPD_j	0	1
X_9		MTN_j	Large	0
X_{10}		MPC_j	0	1
X_{11}		MCI_j	1	0

TABLE 2: Single mode validation criteria tested in this work. The checkmarks indicate which indicators are retained for defining the automated modal selection strategy proposed in this paper.

One first defines a distance between two modes (λ_j, ϕ_j) and (λ_k, ϕ_k) (possibly produced for different values of n) as:

$$d(j,k) \;=\; \frac{|\lambda_j - \lambda_k|}{\max\left(|\lambda_j|, |\lambda_k|\right)} \;+\; 1 - \mathrm{MAC}\left(\phi_j, \phi_k\right) \tag{5}$$

where the classical *Modal Assurance Criterion* (MAC) is involved (see e.g. [28]) for measuring the colinearity between two modeshapes.

Continuity criteria are then used for establishing if some properties of the same mode (such as frequency, damping ratio, etc.) are found for other values of n, in particular $n+2$. Let $I_{n+2} = \{1, 2, \ldots, n+2\}$ be the set of modes indices identified at order $n+2$. For each mode j identified at order n, one seeks the closest mode $\hat{\jmath}$ in the sense of the modal distance (5):

$$\hat{\jmath} \;=\; \arg \min_{l \in I_{n+2}} d(j,l) \tag{6}$$

For every modal feature $\mathcal{X}$, whose continuity is to be verified from an order n to $n+2$, one defines a measure of this continuity under the form:

$$\Delta\left(\mathcal{X}_j, \mathcal{X}_{\hat{\jmath}}\right) \;=\; \frac{\left|\mathcal{X}_j - \mathcal{X}_{\hat{\jmath}}\right|}{\max\left(\left|\mathcal{X}_j\right|, \left|\mathcal{X}_{\hat{\jmath}}\right|\right)} \tag{7}$$

This indicator is used on undamped frequencies f_j, damping ratios ξ_j, poles λ_j, Modal Transfer Norm MTN_j and Mean Phase deviation MPD_j.

The MPD gives a measure of the modeshape complexity and characterizes the non-alignment of the modeshape in the complex plane. When a mode j is physical and when the associated damping ξ_j is small, the components of the modeshape ϕ_j are organized along a straight line passing through the origin in the complex plane. This property no longer stands in the case of having complex modes with local dissipative mechanisms or closely spaced modes coupled due to damping. This indicator described therefore be used sparingly. The interested reader is referred to [9] for mathematical derivations and to table 4 for graphical illustrations.

A last indicator used in this work is the MTN. Introduced initially in [16], the authors highlighted that physical modes have a high MTN, whereas spurious (noise or mathematical) modes have a much lower MTN. For modal selection, this indicator is somehow too discriminating; what has been observed on numerous applications involving complex structures is that this criterion tends to reinforce clear dominant modes, sometimes relegating less participative physical modes to the rank of artefacts. In order to favor the selection of weakly excited modes, a normalized version of the MTN is proposed and derived as follows. In the OO case, the Power Spectral Density (PSD) matrix $\mathcal{S}_{yy}(\omega)$ of observable outputs y at pulsation ω writes by definition as the Fourier transform of the auto-covariance sequence $\left(\mathcal{R}_{yy}(k)\right)_k$:

$$\mathcal{S}_{yy}(\omega) \;=\; \sum_{k=-\infty}^{-1} \mathcal{R}_{yy}(k)\; z^{-k} \;+\; \mathcal{R}_{yy}(0) \;+\; \sum_{k=1}^{\infty} \mathcal{R}_{yy}(k)\; z^{-k} \tag{8}$$

with $z = e^{\imath\omega\Delta t}$. Now using the fact that $\mathcal{R}_{yy}(-k) = \mathcal{R}_{yy}(k)^T$ and $\mathcal{R}_{yy}(k) = CA^{k-1}G$ introducing the classical state-space matrices (A, C, G), one can easily show that the PSD reads[2]:

$$\mathcal{S}_{yy}(\omega) \;=\; C\left(zI - A\right)^{-1} G \;+\; \mathcal{R}_0 \;+\; G^T\left(z^{-1}I - A^T\right)^{-1} C^T \tag{9}$$

Now expressing the system matrices in the modal basis Ψ, one introduces:

$$\Psi \Lambda \Psi^{-1} \;=\; A \quad , \quad g \;=\; \Psi^{-1} G \quad , \quad \Phi = C\Psi \tag{10}$$

[2]where $\mathcal{R}_{yy}(0)$ has been replaced by $\mathcal{R}_0$ as long as there is no ambiguity

to write the contribution of mode j to the power spectral density as:

$$\mathcal{S}_{yy,\,j}(\omega) \;=\; \phi_j \left(z-\lambda_j\right)^{-1} g_j + \mathcal{R}_{0,j} + g_j^T \left(z^{-1}-\lambda_j\right)^{-1} \phi_j^T \qquad \text{with} \qquad \mathcal{R}_{0,j} \;=\; \frac{\phi_j g_j}{\lambda_j} \tag{11}$$

The modal transfer norm is then defined as:

$$\mathrm{MTN}_j \;=\; \left\| \mathcal{S}_{yy,\,j}(\omega) \right\|_\infty \;=\; \max_\omega \left| \boldsymbol{\lambda}_1 \left(\mathcal{S}_{yy,\,j}(\omega) \right) \right| \tag{12}$$

where the $\boldsymbol{\lambda}_1\,(\bullet)$ operator extracts the highest singular value of argument matrix.

One shall now remark that, in the neighborhood of resonant pulsations ω_j, the PSD should be dominated by the response of corresponding mode j, what writes:

$$\mathcal{S}_{yy}(\omega=\omega_j) \;\approx\; \mathcal{S}_{yy,\,j}(\omega_j) \;=\; \mathrm{MTN}_j \tag{13}$$

Using the Δ operator introduced in (7), one defines a *normalized* MTN distance as:

$$n\mathrm{MTN}_j \;=\; \Delta(\mathrm{MTN}_j, \overline{\mathrm{MTN}}_j) \qquad \text{with} \qquad \overline{\mathrm{MTN}}_j \;=\; \mathcal{S}_{yy}(\omega=\omega_j) \tag{14}$$

where $\mathcal{S}_{yy}(\omega)$ is an estimation of PSD matrix obtained thanks to formula (9) with system matrices and thus involving all modes identified at order n. $\overline{\mathrm{MTN}}_j$ is the targeted value that an ideal physical mode contribution should reach at resonance. A physical mode will then have a nMTN close to 0 whereas artefacts with no local dominant influence on the PSD will exhibit closer-to-unit values.

In the IO case, the MTN can be adapted as explained in [13]. After, splitting the state and output vectors of the state-space model into a *deterministic* and *stochastic* contribution, the MTN is defined as the highest singular value of the concatenated matrices $\left[H(\omega), (\mathcal{S}_{y^s y^s})^+(\omega)\right]$ where $H(\omega)$ is the transfer matrix of the deterministic system verifying:

$$H(\omega) \;=\; C\left(zI-A\right)^{-1} B + D \tag{15}$$

using the classical system matrices (A, B, C, D) and where $(\mathcal{S}_{y^s y^s})^+(\omega)$ is the positive power spectral density of the stochastic part of the measured outputs y^s (see [13] for details). From this redefinition of the MTN, an IO-adapted nMTN can then be defined analogously to OO case using (14) with $\overline{\mathrm{MTN}}_j = [H(\omega_j), (\mathcal{S}_{y^s y^s})^+(\omega_j)]$.

Other indicators such as MPC [12], MCI [14] and MTN in its classical definition [13] have also been tested. In this work, only the first eight criteria of table 2 are retained for carrying out the automated modal selection. A complete discussion on the choice of indicators and its impact on the modal selection performance of several algorithms is available at section 3.5 processing the SMART2013/RUN6 data in the OO case.

3. Three-stages automated selection strategy

The strategy, initially proposed by [9] and used and improved in this paper, aims at separating the *physical modes* from the *spurious modes* produced by a growing model-order identification algorithm. Somehow, this strategy tries to mimic the classification that would be done manually by an expert user and is composed of three stages:

stage (i) : Based on the preliminiary calculation of modal indicators detailed in section 2.3 this stage consists in partitioning the modes of the diagram into two families using *clustering* techniques. The first family $\mathcal{E}^{sp}$ contains the modes that can be classified as 'certainly spurious': the members of this first family, less stable regarding modal indicators, will be discarded. The second family $\mathcal{E}^{ph}$ contains physical modes, or at least modes with good stability properties that make them potential physical modes;

stage (ii) : The modes of the family $\mathcal{E}^{ph}$ are then divided into sub-groups $\left(\mathcal{E}_i^{ph}\right)_i$ such that $\mathcal{E}^{ph} = \oplus_i \mathcal{E}_i^{ph}$. In each group $\mathcal{E}_i^{ph}$, one expects to find only variations of the same mode i obtained for different values of n. *Hierarchical clustering* techniques are implemented for this stage using a threshold, specified at section 3.2, that directly depends on the quality of the first partition done at stage (i);

stage (iii) : The sub-groups of $\mathcal{E}^{ph}$ are then split into two groups: the *physical modes* and the *spurious modes*. A single modeshape is chosen for the representation of mode i (associated with median damping ratio) and the mean and standard deviation for frequency and damping ratio is computed based on the members populating group $\mathcal{E}_i^{ph}$.

This methodology, initially proposed by [9] and also followed by [25], has been adapted here and will be detailed in the next subsections; emphasis will be put when the implemented methodology differs from the above-mentioned references.

3.1. Stage (i) : fuzzy clustering

3.1.1. Preliminary observations

The objective of this first stage is to make a first partition of the modes into two groups, that of possible physical modes $\mathcal{E}^{ph}$ and that of certainly spurious modes $\mathcal{E}^{sp}$. Each of the $J \leq n_{max}(n_{max}+1)$ modes present in the diagram (modes with negative damping are removed) is associated to a vector in $\mathbb{R}^d$ containing the $d = 8$ retained modal indicators of table 2. Indicators are then gathered in column for each mode j in a matrix X of size $8 \times J$ and having coefficients within $[0;1]$. Of course, one has: $\operatorname{card}\mathcal{E}^{sp} + \operatorname{card}\mathcal{E}^{ph} = J$. What is important is to avoid the situation of having a too small $\mathcal{E}^{ph}$, meaning that physical modes would be wrongly classified in $\mathcal{E}^{sp}$. On the contrary having a too large $\mathcal{E}^{ph}$ would have a negative impact on stage (ii). At stage (ii), $\mathcal{E}^{ph}$ is subdivided using hierarchical clustering with a threshold computed from the statistical properties of the family $\mathcal{E}^{ph}$; having too many artefacts in this group would lead to irrelevant threshold (see discussion 3.5).

Hard clustering *k-means* algorithm is used in [9, 25] for stage (i). The centers C_1 (of $\mathcal{E}^{ph}$) and C_2 (of $\mathcal{E}^{sp}$) are sought as:

$$\left\{C_1, C_2\right\} \;=\; \arg\min_{C} \sum_{k=1}^{2} \sum_{j=1}^{J} \|X_j - C_k\|_2^2 \tag{16}$$

However, this algorithm is known for having two strong limitations [26]:

(a) Due to its euclidean metrics, the *k-means* algorithm is suitable for normally distributed clusters of similar variance which is not the case for modal indicators;

(b) Even when applied to normally distributed data, *k-means* tends to split the datasets in approximately equally-sized clusters.

Figure 3 shows a scatter plot of the modal indicators nMTN and MPD. One can observe two not normally distributed clusters of in-homogeneous size that furthermore do not exhibit good convex properties. On figure 3, the marginal probability density functions (pdf) f have

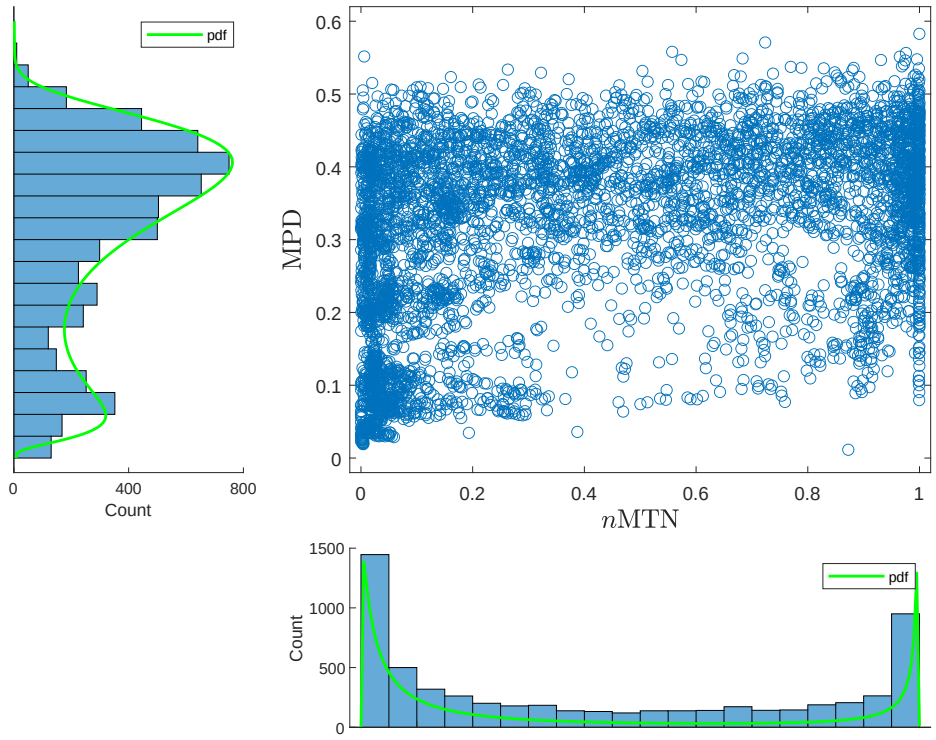


FIGURE 3: Scatter plot of the nMTN vs. MPD indicators. The estimated marginal probability density functions (in green) have been superimposed to the histograms.

been estimated in the maximum likelyhood sense using a combination of two direct- and reversed- log-normal distributions:

$$f(t;\alpha,\tau^{(1)},\mu^{(1)},\sigma^{(1)},\tau^{(2)},\mu^{(2)},\sigma^{(2)}) \;=\; \alpha\log\mathcal{N}(t^{+};\mu^{(1)},\sigma^{(1)}) \;+\; (1-\alpha)\log\mathcal{N}(t^{-};\mu^{(2)},\sigma^{(2)}) \tag{17}$$

with $t^{+}=t-\tau^{(1)}$ and $t^{-}=-(t-\tau^{(2)})$ and:

$$\log\mathcal{N}(t;\mu,\sigma) \;=\; \frac{1}{\sigma\sqrt{2\pi}t}\,\exp\left(-\frac{(\log t-\mu)^2}{2\sigma^2}\right) \tag{18}$$

The limitations of the *k-means* algorithm and the disadvantages of using it at stage (i) have also been noticed in [25]. In this last reference, the indicators $X_{[1:4,6]}$ of table 2 are

used for discrimination and the observation is made that the marginals can be reasonably well approximated by exponential-like distributions. Then, a change of variable $X \to \mathcal{M}(X)$ involving logarithmic-like functions (Box-Cox transform) is proposed for transforming the feature vector into a shape that resembles a normal distribution. Thanks to this transformation, the partition enabled by *k-means* is improved, provided that the maximum model order n_{max} is chosen such that $\operatorname{card}\mathcal{E}^{sp} \approx \operatorname{card}\mathcal{E}^{ph}$. This implies that the number of expected physical modes is known in advance, which is not the case in most practical applications.

As claimed in the introduction, in this paper a different way is followed. Somehow, a change of variable involving a mapping $\mathcal{M}$ is also used, but this time, the transformation is identified from the structure of the data itself, following a *spectral clustering* approach (see e.g. [29] for a detailed tutorial). A spectral transformation will allow to circumvent limitation (a) and a modified *fuzzy c-means* clustering algorithm will enable to overcome limitation (b) as explained in the following lines. More robustness w.r.t. non-convexity and inhomogeneous clusters is then expected. Before carrying out the newly developed algorithm on modal indicators for partitioning $\mathcal{E}^{ph}$ and $\mathcal{E}^{sp}$, it will first be tested on two academic test-cases (representative to a certain extent of modal indicators distributions) for enabling comparison with classical *k-means* and *fuzzy c-means* algorithms on known distributions.

3.1.2. Two reference 2D test cases

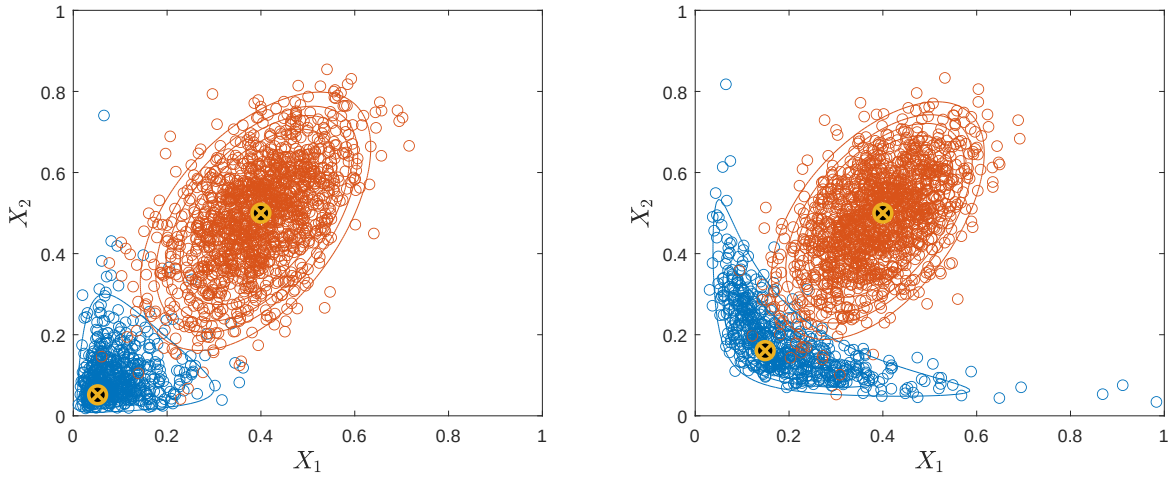
For illustrating the comparative performance of the different clustering algorithms developed in this section, two test-cases involving clusters in $[0;1]^2$ are considered. For each one of them, the clusters are defined using a bivariate expression of the pdf (17) that was fitted to the marginals distribution of nMTN and MPD indicators. The identified parameters $(\tau^{(k)}, \mu^{(k)}, \sigma^{(k)})_{k=1,2}$ however have been changed for producing simpler distributions that will facilitate the discussion on the performance of the different clustering algorithms. (For reproducibility, the parameters used for the test-cases are given in Appendix A). So, one considers:

- A first case (see fig. 4a) where the two clusters exhibit good convex properties,
- A second case (see fig. 4b) where the blue cluster has a non-convex shape produced by a bivariate lognormal distribution with negative off-diagonal term in the covariance matrix.

In both cases, the parameters for generating the red cluster are chosen such that its scatter plot resembles this of a bivariate normal distribution. Identifying for convenience $\mathcal{E}^{ph}$ to the blue cluster and $\mathcal{E}^{sp}$ to the red cluster of convex spread, the following ratio $a = \operatorname{card}\mathcal{E}^{ph}/J$ is considered. This cardinality ratio is fixed to $a = 0.3$; in the modal analysis context, this situation would correspond to the case of having greater number of artefacts than physical modes in the diagram. The in-between clusters distance (simple euclidean distance between the theoretical positions of the two cluster centers, see fig. 4) is $\bar{d} = 0.60$ in the convex case, $\bar{d} = 0.42$ in the non-convex case.

3.1.3. Clustering algorithms

Spectral clustering is generally of valuable help when dealing with non convex clusters (see e.g. [29, 30] and references therein). As previously explained, a mapping $\mathcal{M}$ is derived



(a) Convex case

(b) Non-convex case

FIGURE 4: Test case – Two overlapping reference clusters of known pdf. The solid lines represent the isovalues of the bivariate pdf. The theoretical positions of the two cluster centers are marked with yellow circled crosses.

from the structure of the data for transforming non-convex possibly overlapping clusters X into hopefully better separated clusters of more convex shape Y, through a change of variable writing $Y = \mathcal{M}(X)$. The mapping $\mathcal{M}$ (more precisely transformed Y data) is constructed from the smallest eigenvalues (0 in the ideal case) of the so-called *graph Laplacian* L following the steps described hereafter.

First, a similarity matrix S is assembled. The quantity S_{jl} measures to what extend a given point X_j is closed to its neighbors X_l. Almost unit values are associated to close points whereas smaller values concern more distant points. There are several ways to construct the similarity matrix. In this work, a Gaussian similarity function is considered: $S_{jl} = \exp\left(-\frac{1}{2}\frac{\|X_j - X_l\|^2}{\sigma^2}\right)$ where the parameter σ controls the width of the neighborhood where points l are to be considered close to point j.

The test-case results have proved to be extremely sensitive to the choice of this control parameter. A small change in a manually found appropriate σ leads to poor quality results, to the point that no generally suitable value has been found. The same observation is made in [31] and [32] and in this work, following the lines of [31], an adaptive value σ^2 is chosen according to the local neighborhood of each point j. The distance from any point l to point j is defined as $\|X_j - X_l\| / \sigma_j$ using the classical euclidean norm and where the local scaling parameter σ_j is associated to the density of the neighborhood of point j; the converse is $\|X_l - X_j\| / \sigma_l$. In practice, σ_j is identified to the euclidean distance from point j to its p-th closest neighbor i.e. $\sigma_j = \|X_j - X_p\|$; in this work as in [31], the choice $p = 7$ is made, however, values within $[5; 10]$ were also tested and proved good adaptability to different cluster densities, spreads and sizes. Similarity matrix thus writes:

$$S_{jl} = \exp\left(-\frac{1}{2}\frac{\|X_j - X_l\|^2}{\sigma_j \sigma_l}\right) \tag{19}$$

The variant proposed by [32], consisting in building the similarity matrix from the results of a preliminary fuzzy clustering, has also been tested and has led to results of slightly improved quality for the 2D test cases. However, a certain loss of robustness has been experienced partitioning the modal indicators when dimensionality increases. The adaptive σ_j similarity construction [31] has shown the best overall performance.

The *graph Laplacian* is then defined as $L = D - S$, where the so-called degree matrix D only contains diagonal elements $d_j = \sum_{l=1}^{J} S_{jl}$. According to classical graph theory (see e.g. [29]) the multiplicity of the 0 eigenvalue of the graph Laplacian L corresponds to the number of (well) separated clusters. Two equivalent weightings are classically used for improving the condition number of L: the symmetric $L_{sym} = D^{-1/2} L D^{-1/2}$ and random walk $L_{rw} = D^{-1} L$ weighting. The random walk weighting is used in this work and the transformed data Y are obtained from the null (or close to zero) singular vectors of L_{rw} after normalization of each row. Algorithm 1 synthesizes the construction of such mapping $\mathcal{M}$.

Input : Data $X \in \mathbb{R}^{J \times d}$
Output : Data $Y \in \mathbb{R}^{J \times c}$
Parameter : c (=2, number of clusters)

1 Compute the similarity matrix S using (19) and adaptative σ_j based on [31]
2 Compute the weighted graph Laplacian $L_{rw} = I - D^{-1} S$
3 Solve the eigenvalue problem on L_{rw} and retain the eigenvectors $U_{k=1\ldots c}$ corresponding to the c smallest eigenvalues
4 Compute $Y_{k=1\ldots c}$ as : $Y_{jk} = U_{jk} / (\sum_{k=1}^{c} U_{jk}^2)^{1/2}$

Algorithm 1: Computation of $Y = \mathcal{M}(X)$.

The scatter plot of the sought clusters in the Y_1 vs. Y_2 plane now shows convex shaped clusters (see e.g. fig. 5a and 5b on test-cases). Traditional hard or soft clustering methods can then be used. In this work, a *fuzzy c-means* (FCM) algorithm is used for accounting the not-necessary spherical spread of the clusters and a modification is made to handle the potential asymmetry in terms of cardinality. The modified FCM algorithm is described in the following lines.

Compared to hard clustering methods, soft clustering methods somehow relax the notion of belonging to a group by introducing *membership functions* $\mu_{k=1\ldots c}$ with values in $[0; 1]$ and verifying:

$$\forall j, \quad \sum_{k=1}^{c} \mu_{jk} \;=\; 1 \tag{20}$$

Thus, a point j will be more likely to belong to a cluster k if the value μ_{jk} is close to one. Using a non-necessary euclidean norm $\|\bullet\|_A$, the centers and membership functions are the solutions of the constrained minimization problem:

$$\{C, \mu\} \;=\; \arg\min_{C,\mu} \underbrace{\sum_{k=1}^{c} \sum_{j=1}^{J} \mu_{jk}^m \, \|X_j - C_k\|_A^2}_{J(C,\mu;X)} \tag{21}$$

with μ verifying eq. (20). The exponent m, chosen superior to 1, enables one to tune the degree of "fuzziness" of the partition: $m = 1$ corresponds to hard clustering and $m = 2$ is the classically used value that is also chosen in this work. The hermitian norm $\|\bullet\|_A$ can be defined using a positive definite matrix A, identically chosen for all clusters, as initially defined in [33]. The euclidean (identity matrix) and mahalanobis distance (inverse of the covariance matrix) are special cases for the choice of a constant A. Constant A matrices will have less robustness with respect to asymmetrical clusters.

Alternatively, following the lines of [34], the metric $\|\bullet\|_A$ can be defined using a c-tuple $A = (A_1, A_2, \ldots, A_c)$ for naturally accounting for the potential ellipticity of each cluster. The objective function in (21) becomes: $J(C, \mu, A; X) = \sum_k \sum_j \mu_{jk}^m \|X_j - C_k\|_{A_k}^2$. The matrices A_k must verify positivity and definiteness and are introduced in the minimization problem as optimization variables. Ideally, one would like the metric to handle different scalings along each direction in the feature space. The matrices A_k will then be computed iteratively by the minimization algorithm thus allowing each cluster k to adapt its distance norm to the local topological structure of the data.

The objective functional to minimize will now include the c-tuple A : $J(C, \mu, A; X)$. Note that the cost function J, linear in A, can be made as small as desired by simply making the A_k less positive. The A_k has to be constrained in some sort and what the authors advocate in [34] is to constraint their determinants:

$$\forall k, \quad |A_k| \;=\; \rho_k \;=\; 1 \tag{22}$$

This involves a volume constraint; for what has been noted in the simulations, the local control volume for clustering is roughly inversely proportional to $|A_k|$. An alternative way tested in this work is to set the following constraint:

$$\sum_{k=1}^{c} \rho_k^{\alpha} \;=\; 1 \tag{23}$$

with $\alpha = -1$. The idea behind this constraint is to let the metric adapt to the local density of the feature space and hopefully enable a better separation of inhomogeneous clusters.

The augmented cost function $\bar{J}$ of the constrained minimization problem now writes:

$$\bar{J}\left(C, \mu, \lambda, A, \beta, \rho, \gamma; X\right) \;=\; J\left(C, \mu; X\right) + \sum_{j=1}^{J} \lambda_j \left(\sum_{k=1}^{c} \mu_{jk} - 1 \right) + \sum_{k=1}^{c} \beta_k \left(|A_k| - \rho_k \right) + \gamma \left(\sum_{k=1}^{c} \rho_k^{\alpha} - 1 \right) \tag{24}$$

introducing the following Lagrange multipliers $(\lambda_j)_{j=1}^{J}$, $(\beta_k)_{k=1}^{c}$ and γ. Now, writing the stationarity of $\bar{J}$ with respect to every solution field in $(C, \mu, \lambda, A, \beta, \rho, \gamma)$ gives rise to a set of equations, that, after elimination of the Lagrange multipliers (λ, β, γ), leads to:

$$\forall k, \qquad C_k \;=\; \frac{\displaystyle\sum_{j=1}^{J} \mu_{jk}^m X_j}{\displaystyle\sum_{j=1}^{N} \mu_{jk}^m} \tag{25}$$

$$\forall j,k, \qquad \mu_{jk} = \frac{1}{\sum_{p=1}^{c} \left(\frac{\|X_j - C_k\|^2_{A_k}}{\|X_j - C_p\|^2_{A_p}} \right)^{\frac{1}{m-1}}} \tag{26}$$

$$\forall k, \qquad \rho_k = \frac{c\,|D_k|^{\frac{1}{d-1}}}{\sum_{p=1}^{c} |D_p|^{\frac{1}{d-1}}} \tag{27}$$

$$\forall k, \qquad A_k^{-1} = \left(\frac{1}{\rho_k\,|\Sigma_k|} \right)^{\frac{1}{d}} \Sigma_k \tag{28}$$

with $\Sigma_k = \left(\sum_{j=1}^{J} \mu_{jk}^m (X_j - C_k)(X_j - C_k)^T\right) / \left(\sum_{j=1}^{J} \mu_{jk}^m\right)$ and $D_k = \left(\sum_{j=1}^{J} \mu_{jk}^m\right)^d / |\Sigma_k|$. This set of equations is solved using a classical fixed-point algorithm with alternate search directions (see alg. 2) that proved quick and effective convergence for all tested datasets. (Further clues indicating why such algorithms should converge, esp. their link to maximum likelyhood estimation can be found in [34]). If the position of one or several centers is known (as for the case of identifying the center of an exponentially distributed cluster), the update step [5] of algorithm 2 can be skipped. Finally, a new algorithm combining spectral and fuzzy clustering can be derived (see alg. 3).

Input : Data $X \in \mathbb{R}^{J \times d}$
Output : Centers $C \in \mathbb{R}^d$, Membership functions $\mu \in \mathbb{R}^{J \times c}$
Parameters : c, α, Initial centers C_0 (if known), ϵ_0, update parameter up

---

Initializations $C = C_0$, $A = (\mathbf{1}_1, \dots \mathbf{1}_c)$
while $J(C, \mu; X) > \epsilon_0$ **do**
  Compute membership functions μ_k using (26)
  if $up(k)$ **then**
    Update centers C_k using (25)
  end
  Compute cost function $J(C, \mu; X)$
  Compute Σ_k and D_k
  Compute determinant ρ_k using (27)
  Update matrix A_k using (28)
end

Algorithm 2: Modified fuzzy c-means algorithm.

3.1.4. Results on the 2D test-cases

Figure 5 shows the clustering results obtained by running algorithm 3 on the 2D test-cases of figure 4. The total number of points is $J = 2000$. In both convex and non-convex cases, the clusters are represented in the Y- and X-plane where the color-scale of the markers is chosen according to the value of the membership function μ_1 associated to the first cluster $\mathcal{E}^{ph}$. The members j belonging to the first cluster (chosen such as $\mu_{j1} \geq 0.5$) are then circled in blue

Input : Data $X \in \mathbb{R}^{J\times d}$
Output : Centers $C \in \mathbb{R}^{d}$, Membership functions $\mu \in \mathbb{R}^{J\times c}$
Parameters : c, α, Initial centers C_0 (if known), ϵ_0, update parameter up

---

1 Stack initial centers if needed $X = [X; C_0]$
2 Compute mapping $Y = \mathcal{M}(X; c)$ using alg. 1
3 Dissociate transformed data Y from transformed centers C_0^Y : $[Y; C_0^Y] = Y$
4 Compute membership functions $[C^Y, \mu] = \text{FCM}(Y; c, \alpha, C_0^Y, \epsilon_0, up)$ using alg. 2
5 Recompute final centers C from μ using (25)

Algorithm 3: New spectral fuzzy c-means algorithm.

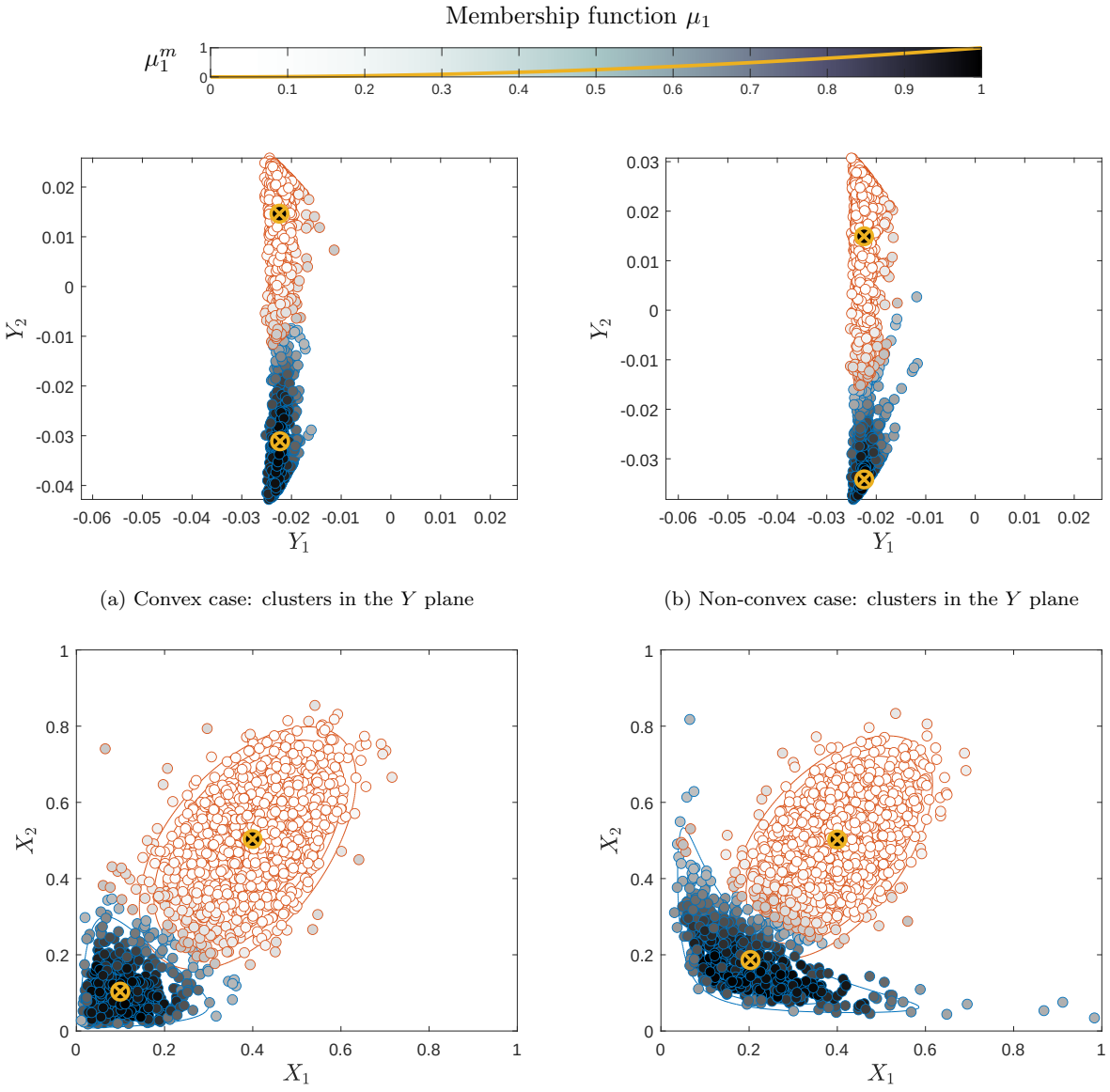
whereas the points of the second cluster (verifying $\mu_{j1} < 0.5$) are circled in red. The final clustering results (see fig. 5) exhibit a very satisfying partition in each case. Unsurprisingly, the membership functions have close-to-1 values near centers and close-to-0.5 values in the overlapping gray area.

As can be seen on figures 5c and 5d, the clusters in the Y plane can have elliptical spreads, that sometimes can be degenerated into lines. This stresses the importance of having a metric that adapts to the shape of the clusters to detect.

A study of the influence of the total point number J has also been performed. 10 different data sets following the reference distributions of the two test cases have been generated for increasing values of J ranging from 100 to 4000. For each data set, three different algorithms have been run: the classical *k-means* algorithm used in [9, 25], the modified *fuzzy c-means* algorithm 2 and the new *spectral fuzzy c-means* algorithm 3. In each configuration, the identified cardinality-ratio $a = \text{card}\,\mathcal{E}^{ph}/J$ has been computed and compared to its 0.3 reference value. Figure 6 condenses the results of this study for both convex (6a) and non-convex (6b) cases. The colored areas contain the identified values of a for the 10 generated data sets for each J. The solid lines of corresponding color are associated to the mean values of a w.r.t. to J.

What can be seen in the convex case (fig. 6a) is that the k-means algorithm gives a slightly biased value $a \approx 0.345$, which in our modal analysis context would lead to include artefacts in the group of possibly physical modes. The new modified- and spectral- FCM algorithms in this case exhibit a similar behavior and offer a better partition with values of a closer to 0.3. In the non-convex case, as expected, the spectral-FCM algorithm enables a clear improvement of the partition with close-to-0.3 values for a, whereas modified-FCM algorithm gives more biased values $a \approx 0.35$ and k-means shows a more unstable biased behavior ($a \approx 0.4$) with larger dispersion. One can also notice from those two plots that none of those algorithms shows good performances when the data sets are too small ($J < 500$). This is another reason for preferring large n_{max} values in the modal analysis context when tuning growing model-order identification algorithms.

One may now question the performances of the modified spectral-FCM algorithm when the in-between clusters distance decreases. A last study has been done, generating again 10 data sets for different values of $(\bar{d}, a)$ and containing this time a fixed value of $J = 2000$



(a) Convex case: clusters in the Y plane

(b) Non-convex case: clusters in the Y plane

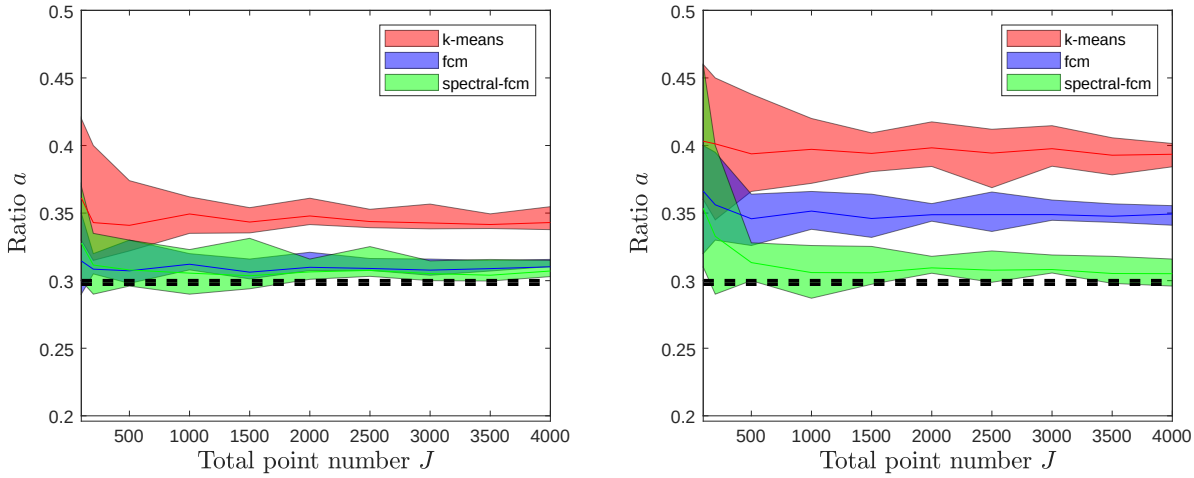
(c) Convex case: clusters in the X plane

(d) Non-convex case: clusters in the X plane

FIGURE 5: Test-case – Final clustering results using alg. 3. In the four plots, the color of each point j is chosen according to the value of the membership function μ_{j1} (see color scale at the top). Points of the first ($\mu_{j1} \geq 0.5$) and second cluster ($\mu_{j1} < 0.5$) are respectively circled in blue and red. Yellow circled crosses mark the identified centers.

members. A colored plot of the identified $a = \text{card}\,\mathcal{E}^{ph}/J$ is shown figure 7 for varying values of $\bar{d}$. One can observe for the two test-cases a clear improvement of the partition when $\bar{d} > 0.4$. However, a general trend still persists: the results are better when trying to dissociate relatively homogeneous clusters (values $a \in [0.3; 0.7]$ in fig. 7).

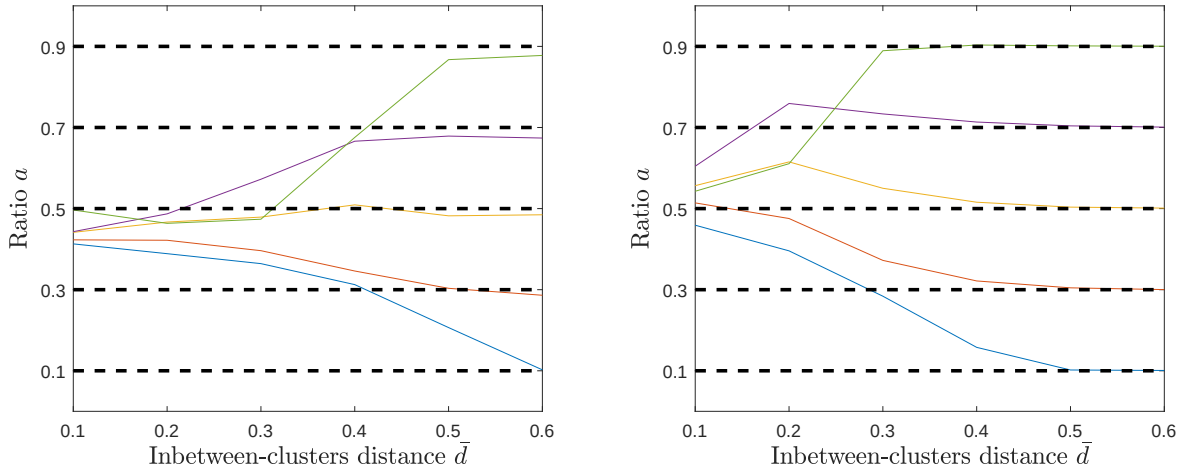
In conclusion to this test-case study, the newly derived spectral-FCM algorithm (see alg. 3) has shown good robustness with respect to both non-convexity and inhomogeneity of the clusters to be separated. This algorithm is now used in lieu and place of the classical k-means algorithm for separating the group $\mathcal{E}^{ph}$ of possibly physical modes from the group



(a) Convex case

(b) Non-convex case

FIGURE 6: Influence of the total point number J on the clustering results. For each value of J, the identified value of a for each of the 10 generated data sets are contained in the colored areas. The solid lines of corresponding colors are associated to the mean values of a. The thick dashed line marks the targeted reference value.



(a) Convex case.

(b) Non-convex case.

FIGURE 7: Performance of the new spectral-FCM algorithm for varying values of $(a, \bar{d})$. Each color corresponds to a fixed value of a: (blue) $a = 0.1$, (red) $a = 0.3$, (yellow) $a = 0.5$, (purple) $a = 0.7$, (green) $a = 0.9$. The thick dashed lines mark the targeted reference values.

$\mathcal{E}^{sp}$ of spurious modes at stage (i). Also note that modified FCM alg. 2 also exhibited good similar performance in the non-convex case.

3.1.5. Modal analysis results

Algorithm 3 is carried out for separating $\mathcal{E}^{ph}$ from $\mathcal{E}^{sp}$ in $\mathbb{R}^8$ after computation of the $X_{[1:8]}$ modal indicators of table 2 for the SMART2013 dataset. In the rest of current section 3, derivations will be illustrated using test results recorded for RUN6 in the OO context.

Figure 8 shows a scatter plot of four out of the eight validation criteria used for dissociating the two families. The theoretical center C_1 of $\mathcal{E}^{ph}$ is not updated during the FCM algorithm (diamond and square are superposed) so that the membership function μ_{j1} can be interpreted as a performance indicator for each mode j. Indeed, modes with the higher membership function value are closer to the theoretical ideal center C_1 and have thus more chances to be good physical mode candidates. This membership function μ_1 is preciously kept as performance indicator and will be reused at stage (iii) (see section 3.3).

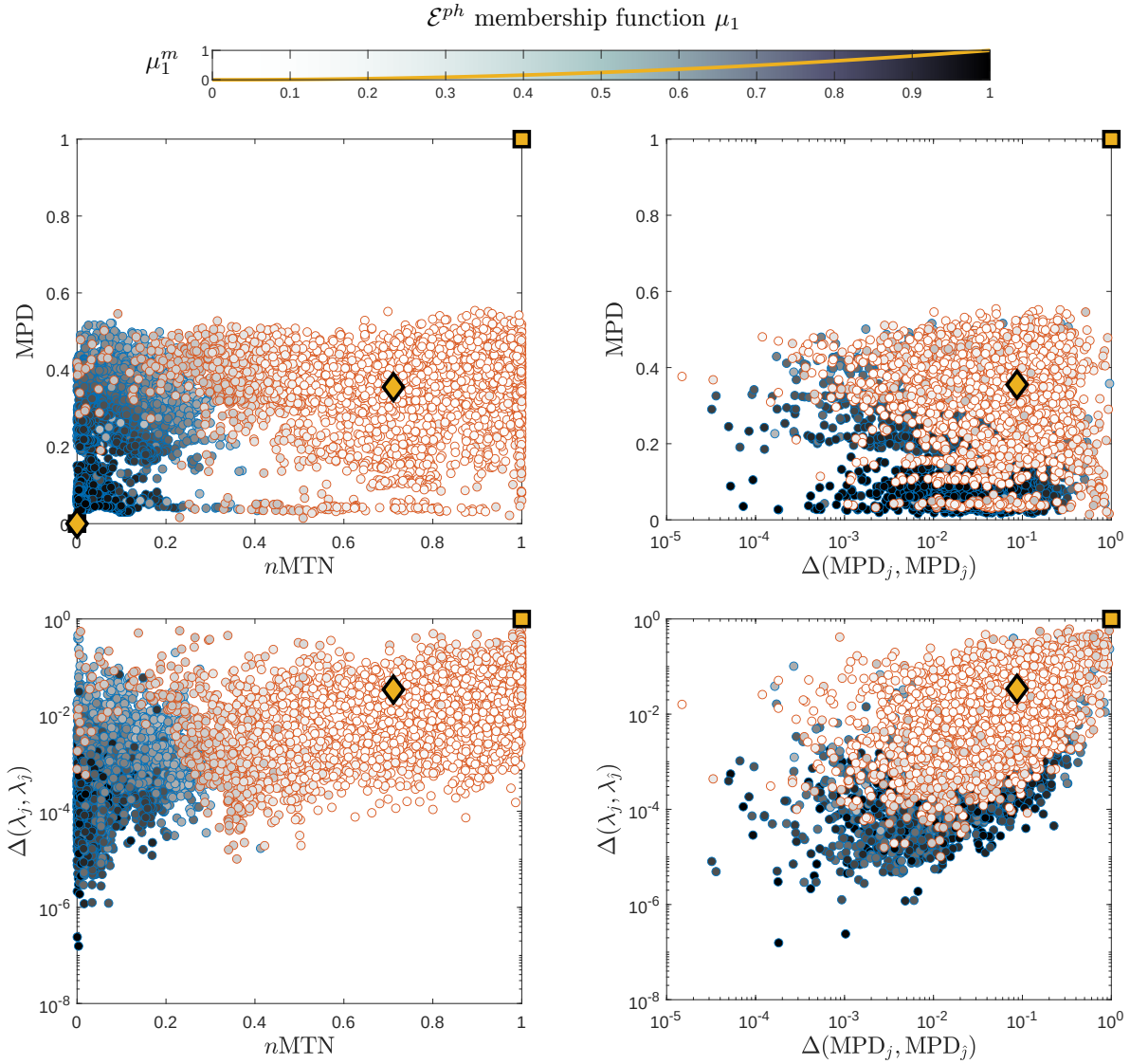


FIGURE 8: Stage (i) – Clustering results on single mode validation criteria $X \in \mathbb{R}^8$ obtained by alg. 3 – OO identification RUN6 – Here only for 4 out of 8 criteria (see tab. 2) are represented. The color of each point is defined according to the values of the membership function μ_1 of group $\mathcal{E}^{ph}$ (see gray scale at the top). Possibly physical modes in $\mathcal{E}^{ph}$ are circled in blue, spurious modes in $\mathcal{E}^{sp}$ are circled in red. The initial guesses for centers C_i are marked with squares and their updated positions are marked with diamonds.

3.2. Stage (ii) : hierarchical clustering

This stage is at the center of the automated selection process. The idea is to form sub-groups $\left(\mathcal{E}_i^{ph}\right)_{i=1}^{I}$ of $\mathcal{E}^{ph}$, each one containing variations of the same mode obtained for different model order n. This is done using *hierarchical clustering* and this stage is not treated differently than in the reference [9]. Hierarchical clustering algorithms start from a situation where each mode belongs to its own cluster and progressively group the modes using a distance criterion for building a unique cluster containing all modes. This iterative clustering process can be visually summarized using the so-called *dendrogram* shown figure 9.

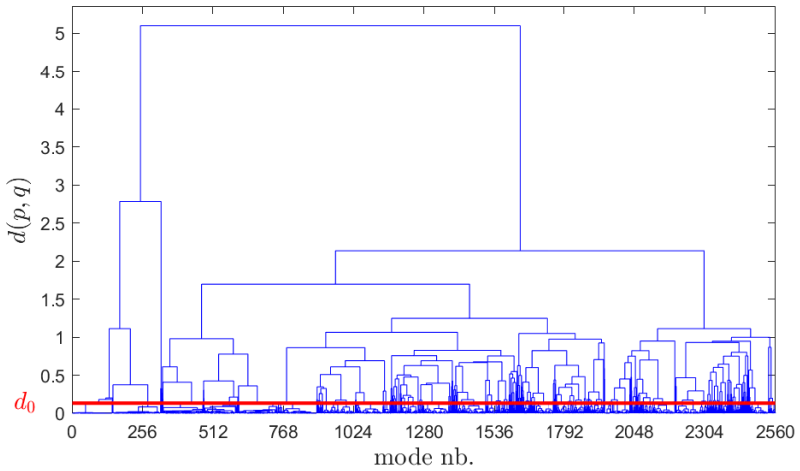


FIGURE 9: Stage (*ii*) – Dendrogram obtained from hierarchical clustering – OO identification RUN6 – The cut-off distance d_0 is represented in red.

The main steps for hierarchical clustering can be summarized as follows:

1. Every mode in the stabilization diagram is placed in its own group and the distance $d(p,q)$ between two groups p and q is computed using the inter-mode distance (5);

2. The two closest groups are merged and the mutual distance between groups is calculated as the average distance between their different elements;

3. Previous stage is then repeated until the mutual distance between groups is greater than a threshold value d_0. This threshold is derived from the statistical properties of the family $\mathcal{E}^{ph}$ identified at stage (i) as:

$$d_0 \;=\; \mathbf{E}\left(d^{ph}\right) \;+\; 2\boldsymbol{\sigma}(d^{ph}) \tag{29}$$

where $\mathbf{E}\left(d^{ph}\right)$ and $\boldsymbol{\sigma}(d^{ph})$ are respectively the mean and standard deviation values of the inter-modes distance (5) denoted d^{ph} and computed for the possibly physical modes in $\mathcal{E}^{ph}$.

In the application (see fig. 9), the cut-off distance d_0 has led to the creation of $I = 239$ sub-groups of $\mathcal{E}^{ph}$ containing similar modes according to the modal distance (5).

3.3. Stage (iii) : final selection

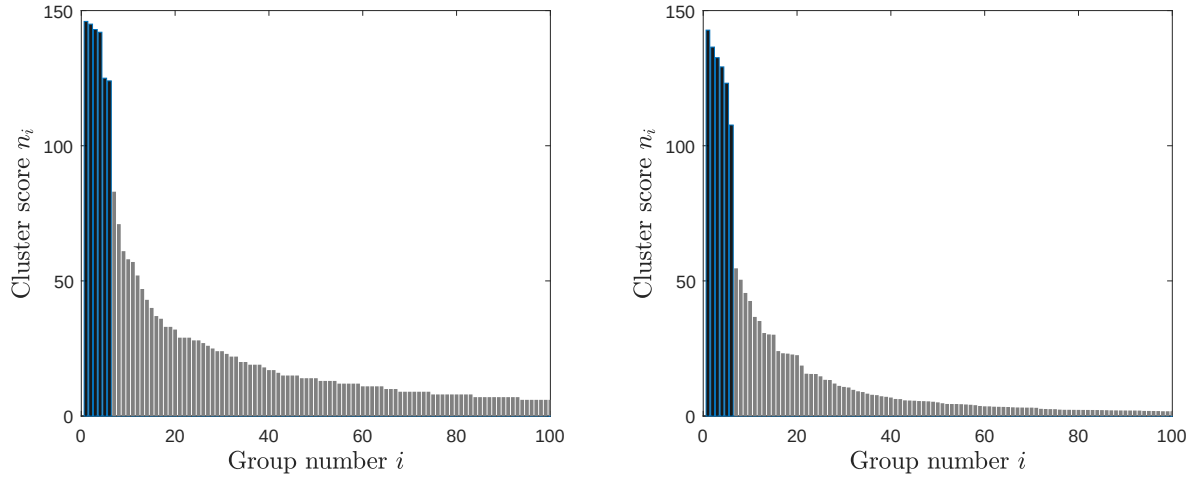
One last stage is now necessary to discard the groups $\mathcal{E}_i^{ph}$ that contain a too small number of mode occurrences. The goal here is to mimic the detection of columns in the stabilization diagrams when visual inspection is carried out. In order to proceed to the last selection stage without introducing a new user-specified threshold, a k-means ($k = 2$) algorithm is used in [9] to separate the set of potential physical modes $\mathcal{E}^{ph} = \oplus_{i=1}^{I} \mathcal{E}_i^{ph}$ into two subgroups according to the number of modes contained in $\mathcal{E}_i^{ph}$. Thus, defining $n_i = \operatorname{card} \mathcal{E}_i^{ph}$, the "centers" $n_{c,1}$ and $n_{c,2}$ of the physical and non-physical modes are sought as:

$$(n_{c,1}, n_{c,2}) \;=\; \arg \min_{n_c} \sum_{k=1}^{2} \sum_{i=1}^{I} (n_i - n_{c,k})^2 \tag{30}$$

In this work a last improvement is made to benefit from the results of stage (i). In order to favor modes with good stability properties, the $\mathcal{E}^{ph}$ membership function is used for redefining the n_i values as:

$$n_i \;=\; \sum_{k=1}^{\operatorname{card} \mathcal{E}_i^{ph}} \mu_{k1}^m \tag{31}$$

Figure 10 shows the more clear partition that is obtained using the new definition of n_i. The jump in the n_i distribution appears more clearly on figure 10b and is more likely to be identified by the k-means algorithm.



(a) Initial version, $n_i = \operatorname{card} \mathcal{E}_i^{ph}$

(b) Improved version, n_i defined wrt (31)

FIGURE 10: Stage (iii) – Final clustering stage – OO identification RUN6 – Cluster score n_i for each group $\mathcal{E}_i^{ph}$.

Remark 3.1 (Pole splitting)**.** When large model orders n_{max} are specified, the so-called pole splitting phenomenon can occur, *i.e.* for high model orders, a single physical mode tends to be modeled by two (or more) similar modes in the stabilization diagram for several values of n (see fig. 11). The hierarchical clustering algorithm of stage (ii) will produce a group $\mathcal{E}_i^{ph}$

containing too many representations of mode i and this can deteriorate the quality of the last partitioning stage (iii). To ensure that only one representation of mode i for any model order n is present in $\mathcal{E}_i^{ph}$, the membership function μ_1 is used once again. If a group $\mathcal{E}_i^{ph}$ contains two occurrences p and q of the same mode for the same value of n, the membership function enables one to automatically select the more 'reliable' candidate k by simply taking: $k = \arg\max_{p,q}(\mu_{p1}, \mu_{q1})$. □

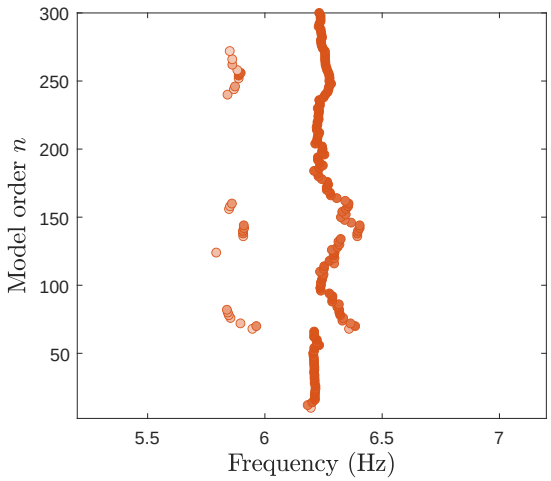


FIGURE 11: Pole splitting and duplicated pole removal – OO identification RUN6 – Group $\mathcal{E}_1^{ph}$ associated to first eigenfrequency. The modes k present in $\mathcal{E}_1^{ph}$ are represented in red with an opacity defined according to μ_{k1}. The right column representations are more reliable than the left column representations regarding the performance criterion μ_{k1}.

3.4. Final results and statistical properties

Figure 12 shows the stabilization diagram obtained in the OO context for RUN6. Keeping the same convention, the face color of each mode k is chosen according to the value of the membership function μ_{k1} and the artefacts are circled in red. $\hat{I} = 6$ modes have been automatically selected (circled in blue and highlighted with blue lines in fig. 12) by the three stages algorithm on the frequency range $[0; 100Hz]$.

Figure 13a shows the cleared stabilization diagram obtained after stage (iii) and duplicated pole removal. One color is chosen per selected family $\mathcal{E}_i^{ph}$ and the opacity of the colors marking the poles is defined according to the membership function μ_1. Thus, the most reliable modes appear in brighter colors while the less reliable ones are nearly transparent.

Table 3 collects the values of the eigenfrequencies and damping ratios of the $\hat{I} = 6$ automatically selected modes using the new spectral FCM algorithm for partitioning indicators $[X_1, \ldots, X_8]$ at stage (i). The modeshapes of the 6 selected modes are represented in table 4 and a comparison with other selection algorithms involving different modal indicators sets is made in section 3.5 and collected in table 3. Please note that the mean and standard deviation values are given in a fuzzy sense for each group i according to the membership function values μ_{k1} with $k \in \mathcal{E}_i^{ph}$. More precisely, for any modal feature $\mathcal{X}$, the *fuzzy mean*

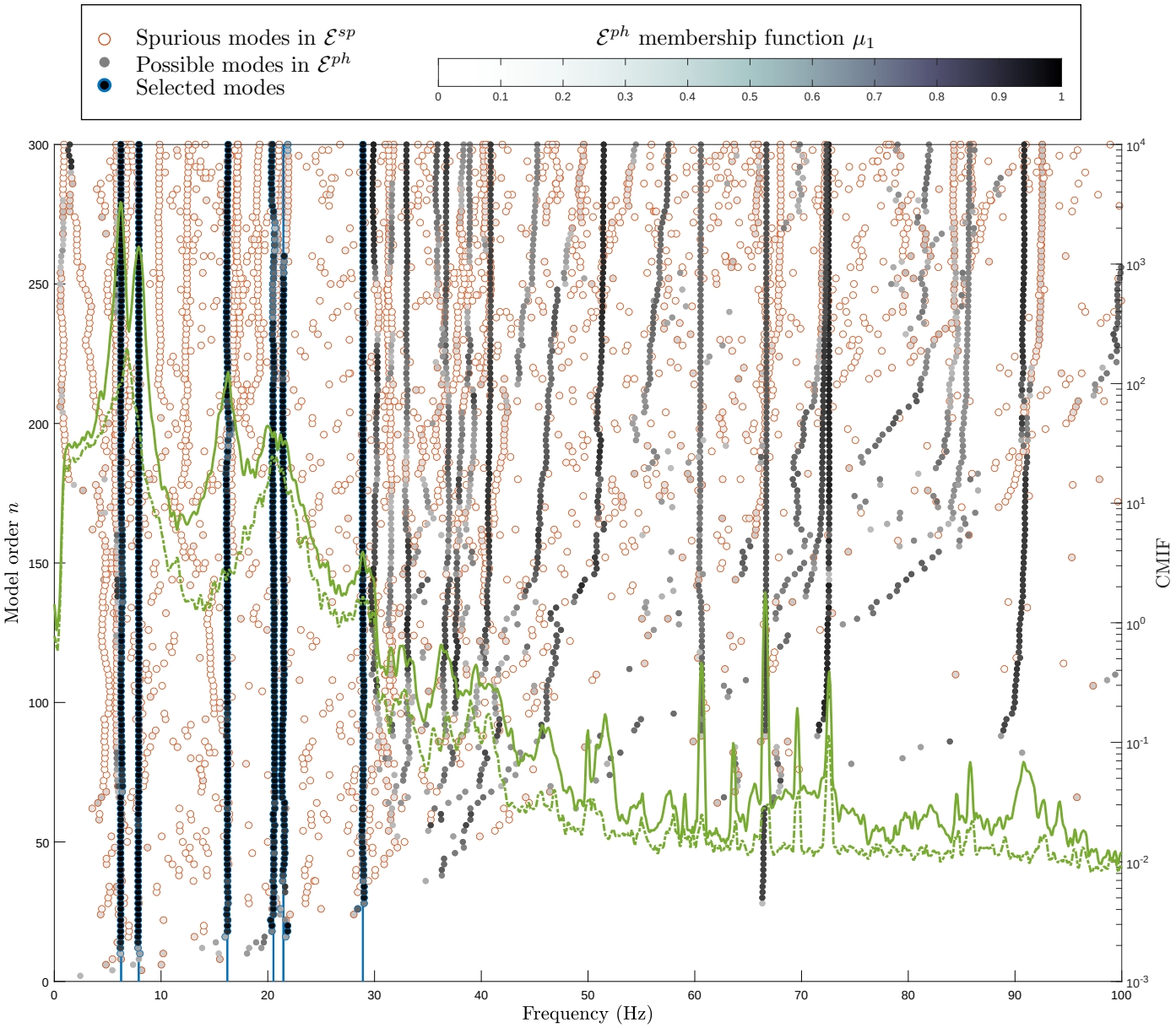
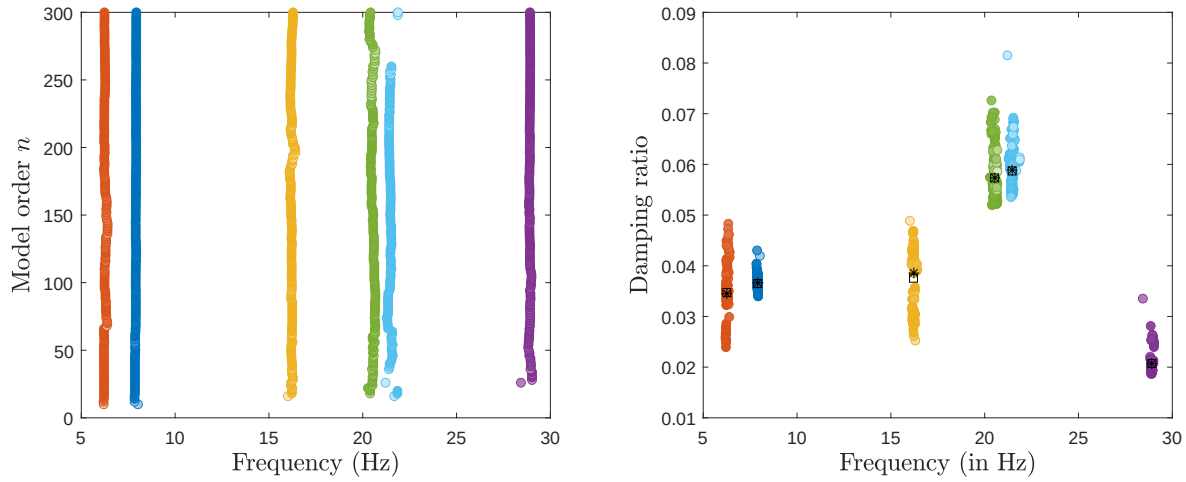


FIGURE 12: Stabilization diagram and modes selected by proc. FCM-$X_{[1:8]}$ – OO identification RUN6 – The continuous and dotted green lines correspond to the first and second eigenvalues of $\mathcal{S}_{yy}(\omega)$ plotted against frequency (log-scale on the right).

and *fuzzy standard deviation* is computed using:

$$\tilde{\mathbf{E}}\left(\mathcal{X}_i\right) = \frac{\sum\limits_{k\in\mathcal{E}_i^{ph}} \mu_{k1}^m \mathcal{X}_k}{\sum\limits_{k\in\mathcal{E}_i^{ph}} \mu_{k1}^m} \quad , \quad \tilde{\boldsymbol{\sigma}}\left(\mathcal{X}_i\right) = \sqrt{\tilde{\mathbf{E}}\left(\left[\mathcal{X}_i - \tilde{\mathbf{E}}\left(\mathcal{X}_i\right)\right]^2\right)} \tag{32}$$



(a) Final stabilization diagram

(b) Damping ratios vs. eigenfrequencies

FIGURE 13: Final modal analysis results – OO identification RUN6 – One color per mode group $\mathcal{E}_i^{ph}$ and opacity of colors defined according to μ_1. The squares $\square$ mark the mean damping values while the stars $*$ point the modes with median damping values chosen for modeshape representation.

The idea behind this is still to give more importance to the modes with the greatest confidence index. No outlier rejection strategy, as used e.g. in [25], is implemented in this work. One here relies on the fuzzy framework (more precisely on the performance index μ_1) to minimize the impact of the outliers on the statistical properties of the modes. What have been empirically found is that outliers generally have low μ_{k1} values as one can see on figure 13b (outliers are nearly transparent).

The results can finally be summarized in a last figure using the *Complex Mode Indicator Function* (CMIF). The CMIF [35] is defined as the ∞-norm of the power spectral density matrix $\mathcal{S}_{yy}(\omega)$ in the OO context, of the transfer function $H(\omega)$ in an IO context. This synthetic frequency indicator has the property of having clear local maxima in the neighborhood of natural frequencies ω_i and is clearly related to the previously introduced MTN reading $\text{MTN}_i = \text{CMIF}(\omega = \omega_i)$. An indication on how well the $\mathcal{S}_{yy}(\omega)$ (resp. $H(\omega)$) matrix is reconstructed from the $\hat{I}$ automatically selected modes can be obtained by comparing two CMIF curves:

- The first version is defined by the ∞-norm of the PSD matrix computed using the Welch's periodogram method (see e.g. [36] for details). In this work, regularly spaced 60% overlapping Blackman windows of $5s$ length are used.
- The second version is obtained after reconstruction of the PSD matrix using the $\hat{I}$ modes selected after SSI:

$$\mathcal{S}_{yy}(\omega) \;=\; \sum_{i=1}^{\hat{I}} \mathcal{S}_{yy,\,i}(\omega) \tag{33}$$

  where each single-mode contribution $\mathcal{S}_{yy,\,i}(\omega)$ is computed using (11).

Figure 14 compares these two versions. In addition to the CMIF reconstructed from (33)

using the mean values $\tilde{\mathbf{E}}(f_i)$ and $\tilde{\mathbf{E}}(\xi_i)$ (black solid line), an envelope curve is also plotted using undamped frequency values in $[\tilde{\mathbf{E}}(f_i) - \tilde{\boldsymbol{\sigma}}(f_i); \tilde{\mathbf{E}}(f_i) + \tilde{\boldsymbol{\sigma}}(f_i)]$ and damping ratios in $[\tilde{\mathbf{E}}(\xi_i) - \tilde{\boldsymbol{\sigma}}(\xi_i); \tilde{\mathbf{E}}(\xi_i) + \tilde{\boldsymbol{\sigma}}(\xi_i)]$. Reconstructed CMIF and Welch CMIF are very close over the entire frequency range $[0; 30Hz]$. The very good match between the two curves demonstrates the good modal reconstruction of the PSD matrix $\mathcal{S}_{yy}(\omega)$ and validates the automatically identified modal signature in terms of eigenfrequencies, damping ratios, modeshapes and participation factors.

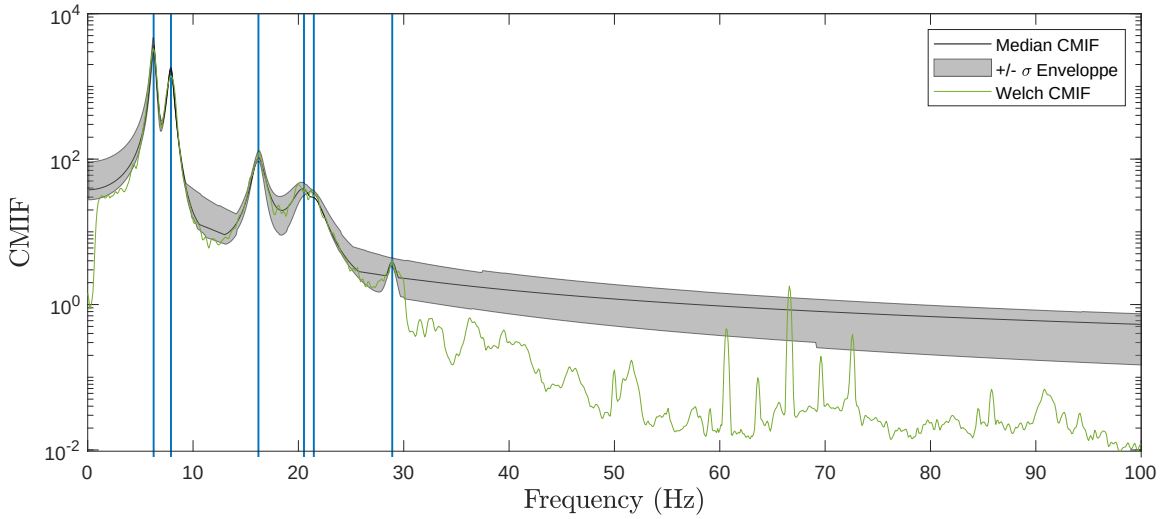


FIGURE 14: Comparison of CMIF curves – OO identification RUN6 – CMIF reconstructed from the modal signature (in black) vs. CMIF obtained from Welch's periodogram (in green). The mean values of the selected eigenfrequencies appear as blue vertical lines.

3.5. Discussion: choice of modal indicators and comparison with other algorithms

The 6 automatically selected modes are clear structural modes of the RC specimen coupled with hydraulics (see additional physical interpretation section 4.1) with low MPD and higher MTN values. However, inspecting the diagram fig. 12, one can notice that several modes on the frequency range $[30Hz; 100Hz]$ have been discarded by the algorithm although they visually exhibit nice stable display with μ_1 values above 0.7. Those higher frequency modes are associated to higher MPD levels (as reported tab. 3) and the question if whether or not they should have been retained by the automated selection algorithm is legitimate.

Further investigation is made in this sense and is explained in the following lines. The current procedure (referred to as FCM-$X_{[1:8]}$ in table 3) has been run a second time, removing MPD from the indicators-set on which fuzzy clustering is performed at stage (i). What can be observed in table 3 (see FCM-$X_{[1:7]}$) is that modes with stable behavior and higher MPD level are now selected by the algorithm. For this second procedure, a larger number of mode candidates was selected in $\mathcal{E}^{ph}$ leading to higher threshold value d_0 (see two last rows of table 3). The adaptability of the threshold to the nature of the user-specified indicators set is a desirable feature.

Proc.	FCM-$X_{[1:8]}$		FCM-$X_{[1:7]}$		KM-$X_{[1:8]}$		KM-$X_{[1:7]}$		KM-log$X_{[1:4,6]}$		KM-$X_{[1:5,8:10]}$		Manual		
Mode	$\tilde{\mathbf{E}}(f)$	$\tilde{\mathbf{E}}(\xi)$	$\tilde{\mathbf{E}}(f)$	$\tilde{\mathbf{E}}(\xi)$	$\mathbf{E}(f)$	$\mathbf{E}(\xi)$	$\mathbf{E}(f)$	$\mathbf{E}(\xi)$	$\mathbf{E}(f)$	$\mathbf{E}(\xi)$	$\mathbf{E}(f)$	$\mathbf{E}(\xi)$	$\mathbf{E}(f)$	$\mathbf{E}(\xi)$	MPD
1	6.26	3.5	6.26	3.5	6.26	3.5	6.26	3.5	6.24	3.4	6.26*	3.5	6.25	3.5	0.10
2	7.91	3.6	7.91	3.7	7.91	3.7	7.91	3.7	7.91	3.6	7.90	3.7	7.89	3.7	0.05
3	16.21	3.8	16.21	3.8	16.21	3.8	16.21	3.8	16.20*	4.2	16.21	3.7	16.21	3.8	0.12
4	20.53	5.7	20.53	5.8	20.52	5.8	20.52	5.8	20.55	5.4	20.54	5.7	20.52	5.8	0.11
5	21.46	5.9	21.47	5.9	21.51	6.0	21.52	6.1	21.47	5.6	21.45	5.8	21.48	6.1	0.15
6	28.91	2.1	28.91	2.1	28.90	2.1	28.90	2.1			28.91	2.0	28.92	2.1	0.14
7			30.04	3.7	30.06	3.8	30.05	3.7					30.05	3.7	0.32
8			33.01	2.1			33.12	2.1					33.32	2.1	0.43
9			36.79	2.5	36.81	2.4	36.81	2.4					36.89	2.4	0.46
10			40.76	1.9	40.75	1.9	40.75	1.9					40.77	1.9	0.28
11			51.23	2.7	51.22	2.7	51.21	2.5					51.20	2.5	0.44
12													60.61	0.6	0.66
13					66.92	3.1	66.92	2.9					66.88	3.1	0.48
14			72.13	0.7	72.16	0.7	72.15	0.7					72.21	0.7	0.41
15			85.43	0.9	85.23	0.9	85.24	0.9					85.63	0.9	0.46
16			90.44	1.7									90.63	1.8	0.44
d_0	0.0805		0.1077		0.1153		0.1155		0.0162*		0.0613		0.1		
a	0.42		0.48		0.52		0.52		0.42		0.30		-		

TABLE 3: Modal selection results for different algorithms – OO identification RUN6 – Mean values of undamped frequencies $\mathbf{E}(f)$ (in Hz) and damping ratios $\mathbf{E}(\xi)$ (in %) for different automated selection procedures and manually selected modes. The names of the tested procedures indicate the modal indicators X used referring to table 2 and the clustering algorithm carried out for partitioning stage (i) (FCM: fuzzy c-means, KM: k-means). Please note that mean values $\tilde{\mathbf{E}}(\bullet)$ are given in a fuzzy sense when FCM algorithm is used. The two last rows indicate the (automatically-)chosen threshold d_0 employed for each procedure and the ratio $a = \text{card}\,\mathcal{E}^{ph}/J$ achieved by the clustering algorithm. The MPD values in the last column corresponds to the MPD value of manually selected modeshapes. The markers * indicate when singularity in the results is observed (see text for details).

The results of a third and a forth procedure are presented in table 3 where k-means is used at stage (i), referring this time to KM-$X_{[1:8]}$ and KM-$X_{[1:7]}$. Modal selection results in this case are very similar: the presence of the MPD in the indicators-set did not sensitively change the partition realized by k-means that split the data into equally-sized $\mathcal{E}^{ph}$ and $\mathcal{E}^{sp}$, and led to nearly identical values for threshold d_0. As was also noticed on the test-cases of section 3.1, the quality of the k-means partition greatly depends on the total number of modes $J = 1/2\ n_{max}(n_{max}+1)$ to be split and implies an optimal choice for n_{max}. The dependence of the automatically chosen threshold d_0 defined by (29) to n_{max} can be observed on figure 15 for both hard- and fuzzy-clustering cases. The two curves were obtained repeating procedures FCM-$X_{[1:8]}$ and KM-$X_{[1:8]}$ for n_{max} ranging from 200 to 500. What is noticeable is the greater independence of d_0 to the user-specified n_{max} using the new FCM algorithm: this is an other desirable characteristic for preferring fuzzy clustering to k-means.

The procedures described in [25] and [9] were also tested for comparison. According to the lines of [25], the indicators-set was restrained to $[X_{1\ldots4}, X_6]$ and a log-transform was applied to each X_i before performing k-means. The centers were initialized as advised in the reference and the threshold d_0 was computed fitting a Weibull distribution to d^{ph}, the inter-modes distance (5) restrained to $\mathcal{E}^{ph}$, and defining it as the 95th percentile threshold:

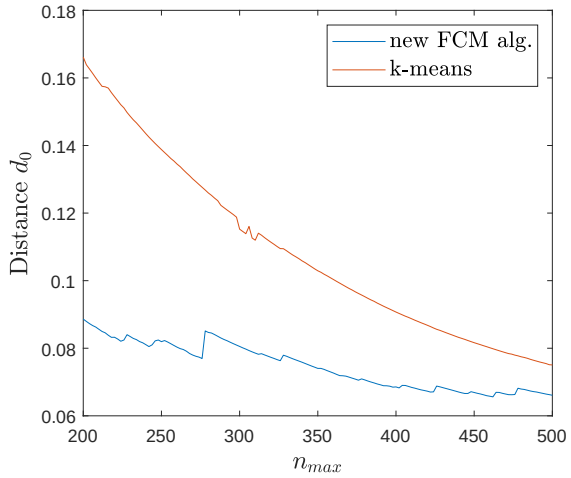
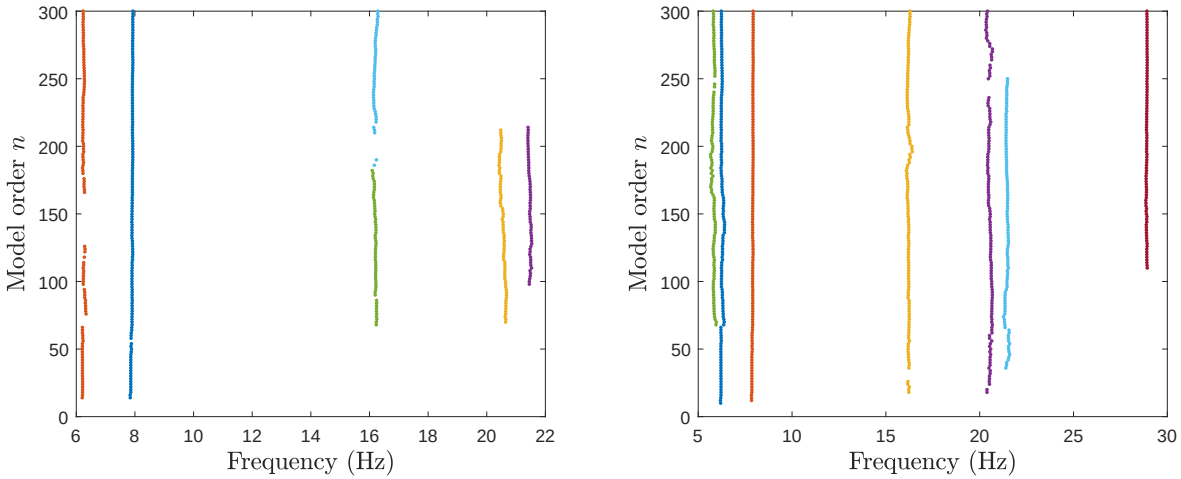


FIGURE 15: Evolution of threshold distance d_0 with maximal user-specified model order n_{max} – OO identification RUN6 – Partition is realized on the indicators set $[X_1, \ldots, X_8]$ of table 2.

$\mathcal{P}(d^{ph} < d_0) = 0.95$. This definition has led to $d_0 = 0.0145$ and only one mode ($7.91Hz$) on the frequency range $[0; 100Hz]$ was selected after stage (iii). The definition has then been replaced by (29) yielding the more appropriate selection of 6 modes. The results associated to this version are collected in table 3 referring to KM-log$X_{[1:4,6]}$ and final stabilization diagram is presented in figure 16a. One has to notice that only 5 modes are reported in table 3: due to the low threshold value $d_0 = 0.0162$, two different groups $\mathcal{E}_{3a}^{ph}$ and $\mathcal{E}_{3b}^{ph}$ were formed for the third mode after stage (ii) and both of them were selected after stage (iii). The group with the larger variance has been manually discarded while the other one was reported in table 3 and marked with a star. Processing the accelerometers of the SMART2013 test campaign, a clear trend of the procedure KM-log$X_{[1:4,6]}$ to produce too low, sometimes unusable, threshold values d_0 has been noted for every run, although it showed good overall performance in [25]. In the reference, the quasi-analytical nature of the tested specimen (cantilevered glass fiber-reinforced polymer plate) enables the discrimination of artefacts based solely on stabilization criteria; physical modes are stable and nicely decoupled while artefacts are more randomly distributed. In the SMART2013 dataset however, the modal density is much higher, especially above $30Hz$. Those higher frequency modes intend to model more complex mechanisms occurring at the contact of additional masses with floors and exhibit good overall stabilization properties to the extend that they are massively included in $\mathcal{E}^{ph}$ when performing k-means on $X_{[1:4,6]}$. Their presence somehow disturbs the definition of the threshold d_0 that becomes too low and inappropriate, finally missing the selection of the 6 linear dominant modes. This trend is here accentuated by the fact that accelerometers are processed: their high frequency range favors the emergence of numerous modes of that kind in the diagrams.

Following this time the methodology proposed in [9], a k-means partition is realized on the indicators-set $[X_{1\ldots5}, X_{8\ldots10}]$. As advised in the reference, indicators were normalized before partitioning. Final modal selection results are collected in table 3 referring to KM-$X_{[1:5,8:10]}$ and the selected modes are placed in the diagram of figure 16b. For this dataset, final selection results are very similar to those obtained with FCM-$X_{[1:8]}$. One has to notice

that, contrary to what could be expected using k-means, a large number of artefacts were removed at stage (i). This is explained by the presence in the indicators-set of the Modal Phase Collinearity (MPC); this indicator is strongly correlated with the MPD, empowering modes with a line distribution in the complex plane. The clustering stage is thus clearly dominated by the MPD/MPC distribution and many modes, too weak in the sense of those two indicators, were removed from $\mathcal{E}^{ph}$. Furthermore, the procedure identified the spurious pole-splitting mode around $6Hz$ as physical. Those spurious modes paradoxically have all the attributes of good physical mode candidates (high MTN, reasonably low MPD and MPC as well as good stabilizing properties) and the low threshold d_0 resulting from the tight partition realized by k-means enabled the hierarchical clustering stage to build a heavily populated cluster of modes one wishes to discard. The pole-splitting group was manually removed from table 3 and first mode is marked with a star.



(a) Proc. KM-log$X_{[1:4,6]}$ described in ref. [25]

(b) Proc. KM-$X_{[1:5,8:10]}$ described in ref. [9]

FIGURE 16: Final modal selection for refs. [25] and [9] – OO analysis RUN6.

The results of the different automated selection procedures were also confronted to manual selection where stabilization charts are cleared out by means of user-defined threshold d_0 and final selection is performed after careful inspection of the modeshapes and indicators values corresponding to the columns with best stabilizing properties. A group of four expert users was consulted for interpreting the diagram of figure 12; all of them benefited from the additional knowledge of (a) the Welch's CMIF and (b) the modeshapes of the numerical FE model developed in the framework of the SMART2013 benchmark [20]. It should be noted that manual interpretation of stabilization diagram greatly depends on the experience and engineering judgment of the analyst; provided results are user dependent. The four experts selected the six first modes but reported having encountered difficulties in the selection process on the interval $[30; 100Hz]$. They all reported that the simultaneous plot of the CMIF in the stabilization diagram and the knowledge of the numerical modeshapes were of great help for resolving ambiguous cases. Only one manual selection result is collected in table 3. Compared to the proposed automated selection procedure FCM-$X_{[1:7]}$, two additional modes near 60 and $66Hz$ were retained. Those modes, associated to simultaneous 2nd and 3rd floor

movements, were discarded by the automated procedure because of their high MPD level. Two out of the four experts discarded them as well for the same reason.

In conclusion to this OO analysis, the new spectral-fcm algorithm proposed for stage (i) enabled a better definition of the cluster $\mathcal{E}^{ph}$ of potential physical modes. Yielding thresholds d_0 using (29) have proved to be:

- More adapted to the physical meaning of the user-specified set of single-mode validation criteria. The analyst now has more flexibility to include weakly excited modes or more complex modes in the final selection;
- Less sensitive to the delicate choice of n_{max}, that, without prior knowledge on the studied structure and dealing with complex systems is generally chosen as high as possible.

Let one note that the newly developed nMTN indicator has played a important role for dealing with weakly excited modes and polesplitting detection. When removing the nMTN from the indicators-set, the FCM-$X_{[1:6,8]}$ procedure only selected the seven following modes $\{1, \ldots, 6, 10\}$ referring to the numbering of table 3 and a group corresponding to the pole-splitting column for mode 1 was included in the selection. This stresses the importance of the nMTN in the set of considered single mode validation criteria. Let one finally remark that in some rare cases (found while processing other experimental data-sets), the spectral-fcm algorithm 3 did not achieve proper partition, gathering the family $\mathcal{E}^{ph}$ around a unique mode corresponding to anti-aliasing filter frequency. This kind of issue can be easily bypassed specifying a reduced frequency range on which modes are to be selected or using fuzzy alg. 2 instead of alg. 3. In this last case, the threshold d_0 has more chances to be sub-optimal, but the rest of the methodology remains unchanged and still gives satisfactory results benefiting from the advantages of the fuzzy framework for pole-splitting and final mode selection as described in section 3.3.

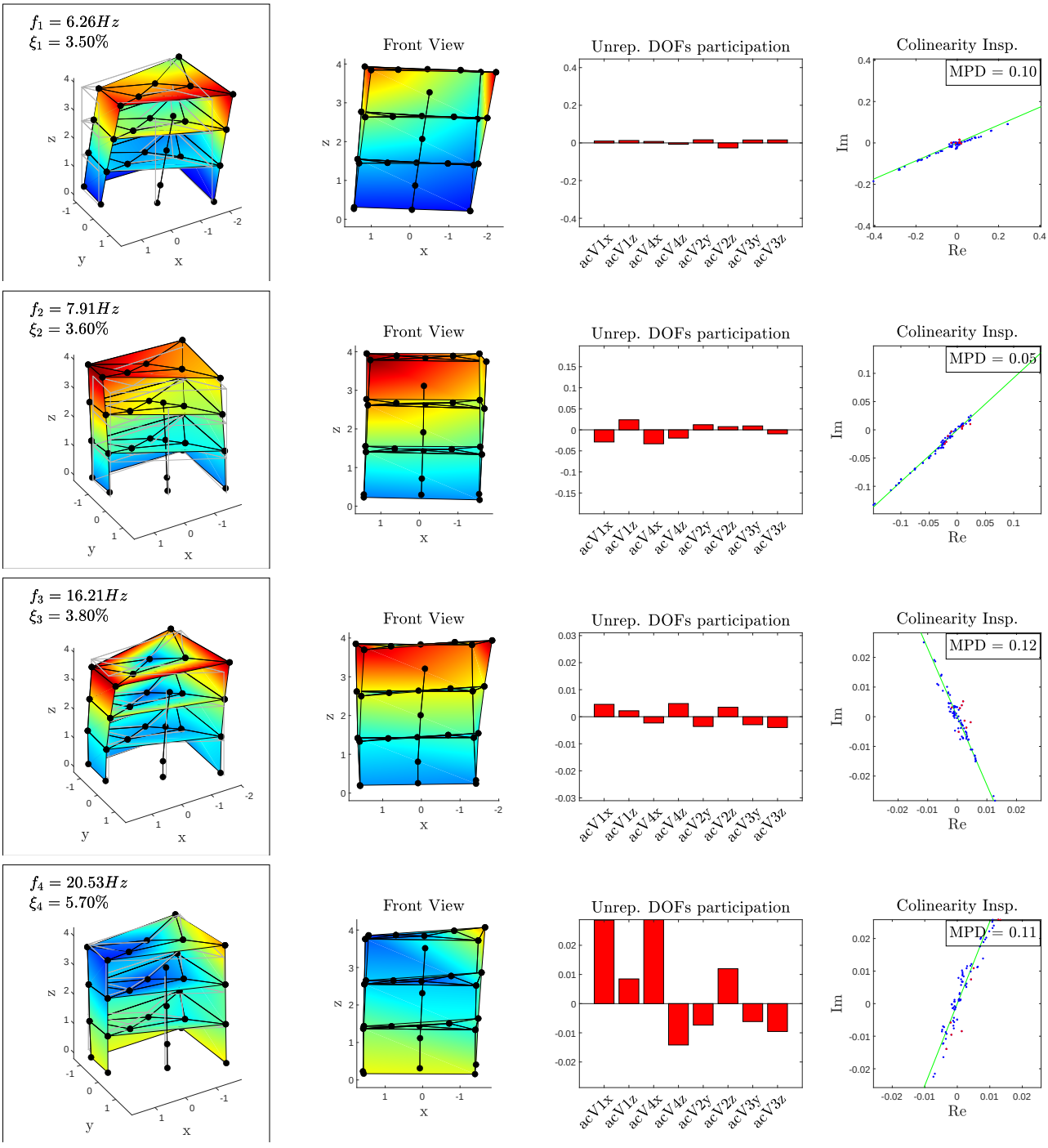
4. SMART2013 shaking-table test campaign : evolution of the modal signature

4.1. Modal signature of the initial configuration

4.1.1. Output-Only framework

Table 4 represents the modeshapes associated to the six modes identified by the FCM-$X_{[1:8]}$ procedure and collected in table 3. Those six modes dominate the dynamic behavior of the system $\mathfrak{Z}$ on the low frequency range (see also fig. 14). On each line of the table, spatial representations of the modeshapes are given for the sensors on the RC specimen and the participation of the accelerometers on the hydraulic actuators (not represented) are plotted as histograms with full scale harmonized with the spatial plots. An inspection of the modeshape colinearity is finally proposed by representing the modeshapes $\phi_i = C\psi_i$ in the complex plane: each dot k stands for the component $\phi_{i,k}$ of modeshape i at measurement point k. The blue dots are associated to the sensors placed on the RC specimen, the red dots to the sensors on the hydraulics.

TABLE 4: Modeshapes identified for RUN6 – OO analysis.



Modeshapes identified for RUN6 – OO analysis.

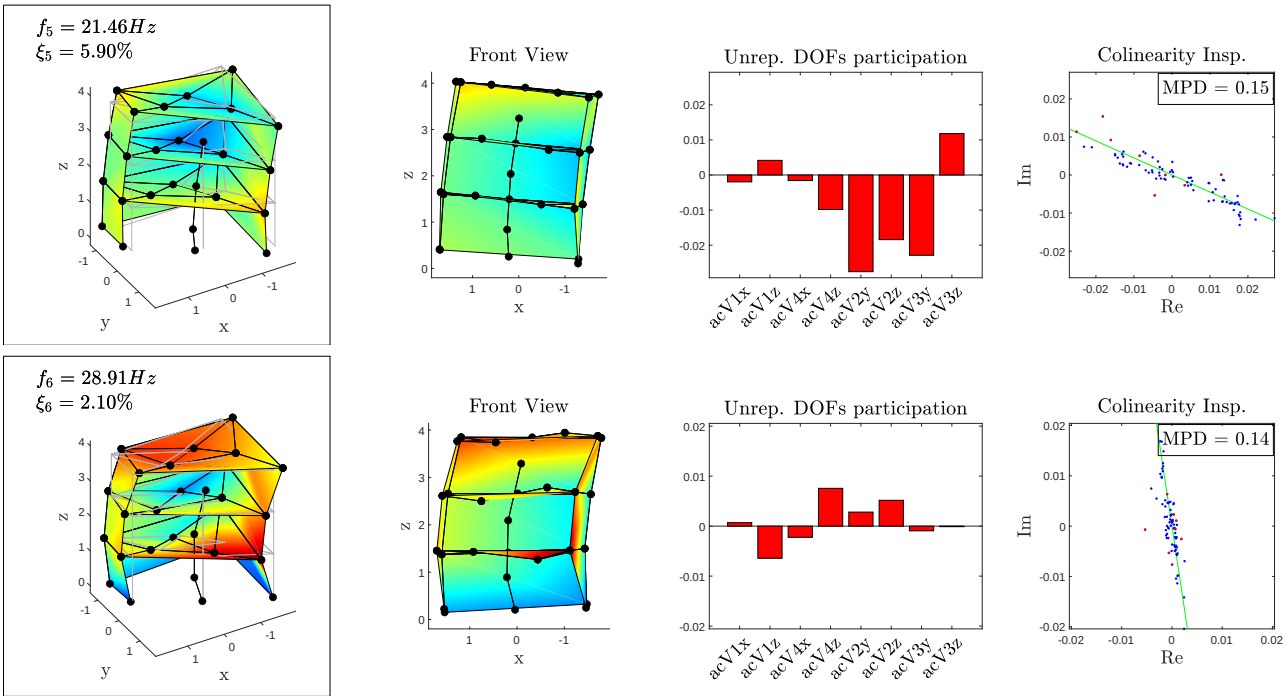


TABLE 4 – Modeshapes identified for RUN6 – OO analysis.
The two left spatial plots represent the modeshapes projected at RC specimen's sensors location. Then ensues the modal participation of the sensors placed on the eight hydraulic actuators (accelerometers $acV*$ named "unrepresented degrees of freedom") taking the form of a bar plot with full scale harmonized with the first two plots. Finally the full modeshapes ϕ_i are represented in the complex plane with blue dots for the sensors on the RC specimen and red dots for hydraulics.

The three first modes are structural modes of the RC specimen with a low participation of the hydraulics: flexion mode along x at $6.26Hz$, flexion mode along y at $7.91Hz$ and torsion mode around z at $16.2Hz$. The two following modes around $20Hz$ are hydraulic modes that mainly involve the mass of the set {Table + RC Specimen} and the stiffness of the oil column contained in hydraulic jacks. The oil column resonances of the actuators in the x and y direction are clearly decoupled. A slight response in flexion of the RC specimen is to observe for the second hydraulic mode along y. The sixth mode around $29Hz$ is a second order flexion mode along x of the RC specimen with a slight implication of two first floors.

Other modes have been identified above $30Hz$ using the FCM-$X_{[1:7]}$ algorithm (see table 3). They mainly involve floor movements with expected local dissipative phenomena of hysteretic nature (sliding friction between additional masses and concrete surface at anchoring): the modeshapes exhibit a higher complexity that is translated into MPD values above 0.3. Those modes with lower participation factor are not plotted in this paper for conciseness.

4.1.2. Input/Output framework

In the IO context, the biaxial ground motion of the AZALEE shaking table is defined as input of the experimental system. Two different analysis are performed specifying in the input vector u the two independent components contained in the row space of the data associated to accelerometers, located on the hydraulic actuators rods in a first study IO-1, on the the soleplate (floor 0) in a second study IO-2. Comparing the results obtained in the OO and IO frameworks, the objective is to quantify the the coupling of the tested structure with the hydraulics and AZALEE shaking table. Table 5 collects the results obtained after carrying out procedure FCM-$X_{[1:7]}$ for each case and recalls the OO results for commodity. Figure 17 presents the stabilization diagram obtained for the IO-1 analysis where the modes selected by the FCM-$X_{[1:7]}$ procedure are highlighted with blue lines. The figure is completed by the plot of the two singular values of the transfer function $H(\omega)$ computed from the Welch's periodogram and verifying:

$$\mathcal{S}_{yy}(\omega) \;=\; H(\omega)\;\mathcal{S}_{uy}(\omega) \tag{34}$$

where $\mathcal{S}_{uy}(\omega)$ is the cross-PSD matrix between inputs and outputs. $H(\omega)$ is then computed after classical Moore-Penrose pseudo-inversion. Unsurprisingly, one visually observes that the automated selection algorithm correctly prioritizes the selection of frequencies where the H-CMIF indicator is peaking. Figure 18 compares the H-CMIF obtained from Welch's periodograms (see eq. 34) with the modal decomposition of expression (15) involving the modehshapes, participation factor together with undamped frequency and damping ratio living respectively in $[\tilde{\mathbf{E}}\,(f_i) - \tilde{\boldsymbol{\sigma}}\,(f_i)\,;\tilde{\mathbf{E}}\,(f_i) + \tilde{\boldsymbol{\sigma}}\,(f_i)]$ and $[\tilde{\mathbf{E}}\,(\xi_i) - \tilde{\boldsymbol{\sigma}}\,(\xi_i)\,;\tilde{\mathbf{E}}\,(\xi_i) + \tilde{\boldsymbol{\sigma}}\,(\xi_i)]$.

The OO and IO results of table 5 exhibit good correlation and the H-CMIF reconstructed after modal selection in figure 18 is in good accordance with the Welch's estimation. As could be expected, the two modes 4 and 5 associated to important hydraulic participation near $20Hz$ (oil column resonance) are removed by the oblique projection in the IO context. Compared to the OO case, a greater proportion of artefacts is present on figure 17 and removed by the new FCM algorithm (see ratio a) and the higher order modes above $30Hz$ now appear more clearly in the stabilization diagram. For those modes, associated to local floor movements and higher order flexion of the RC specimen in the x and y direction, no significant difference between OO and IO analysis is to be noted in terms of undamped frequencies and modeshapes (see MAC indicator tab. 5). Regarding the three first modes however, a significant 6 to 17% increase in mean undamped frequency values is observed in the IO case. The closed-loop controlled hydraulic devices included in the definition of system $\mathfrak{Z}$ and excluded from $\mathfrak{Z}_1$ and $\mathfrak{Z}_2$ (see definitions (2) and (3)) in the IO case obviously adds a non-negligible dynamic impedance to the experimental system, what mainly affects global modes at low frequency.

Furthermore, one can notice that the two IO analysis gave very similar results: the undamped frequencies identified for $\bar{\mathfrak{Z}}_1$ and $\bar{\mathfrak{Z}}_2$ (with and without the AZALEE shaking table) differ from less than 1.5% on the three first modes and less than 0.5% for higher order modes. The coupling between AZALEE shaking table and the 45.8 tons {Specimen + Additional masses} can be qualified as negligible. When trying to reproduce experimental results with numerical models (see e.g. SMART2013 benchmark [37]) more efforts should be made for including the effect of the closed-loop controlled hydraulic actuators rather than

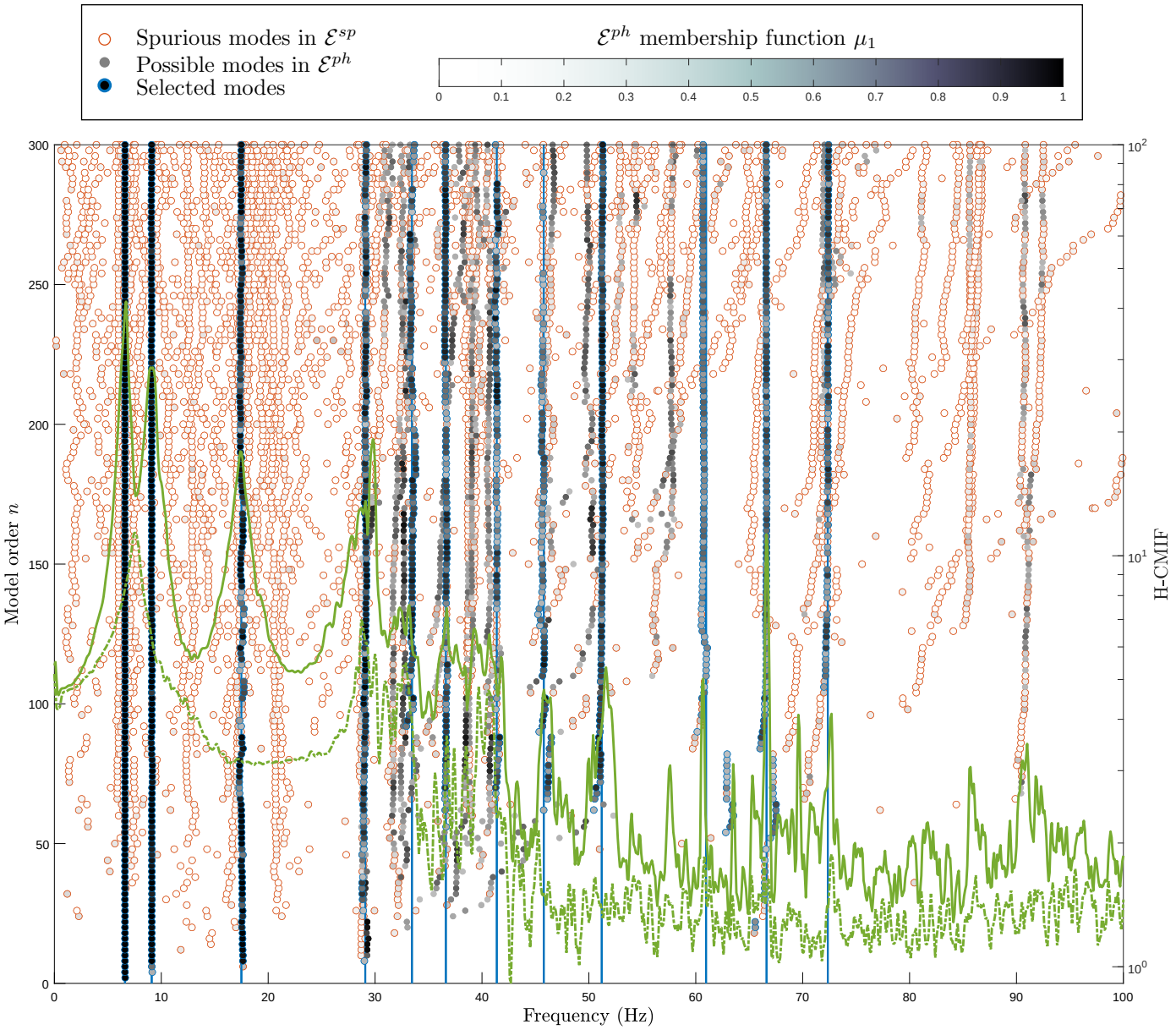


FIGURE 17: Stabilization diagram and modes selected by proc. FCM-$X_{[1:7]}$ – IO-1 identification RUN6 – The continuous and dotted green lines correspond to the first and second singular values of $H(\omega)$ plotted against frequency (log-scale on the right).

this of the AZALEE shaking table itself.

Inspecting fig. 17, one can wonder about the presence stable modes around $13.3Hz$ and $22.6Hz$. There are indeed modes at those frequencies associated to a piping system with embedded mass that is connected to the RC wall and central column at the third floor level. The device is visible on figure 1 and was an attempt for representing a pumping equipment present on typical nuclear electrical secondary buildings [20]. The coupling between this de-

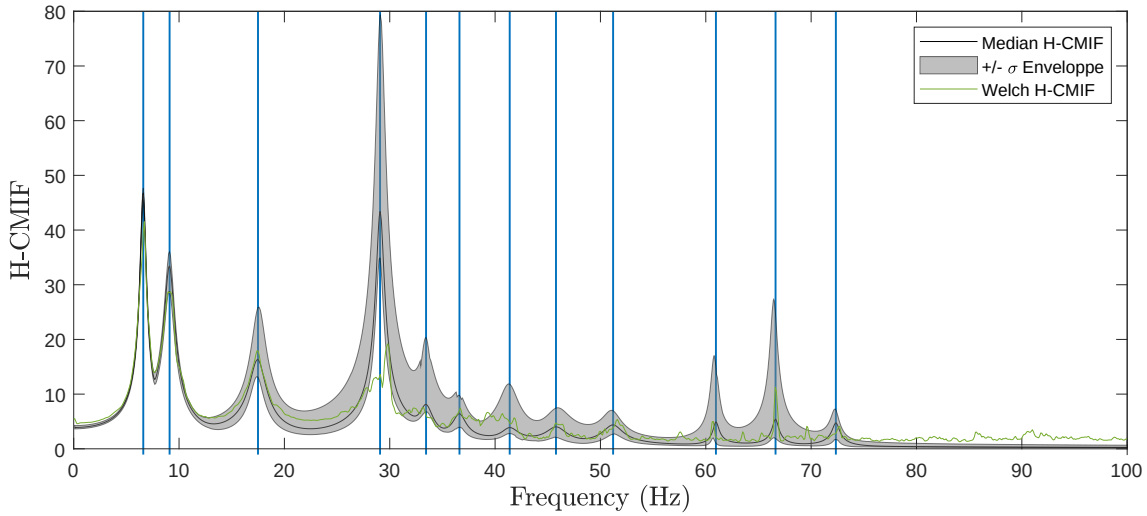


FIGURE 18: First singular value of the transfer function (H-CMIF) – IO-1 identification RUN6 – H-CMIF reconstructed from the modal signature (in black) vs. H-CMIF obtained from Welch's periodogram (in green). The mean values of the selected eigenfrequencies appear as blue vertical lines.

	OO – System $\mathfrak{Z}$				IO-1 – System $\mathfrak{Z}_1$					IO-2 – System $\mathfrak{Z}_2$				
Mode	$\tilde{\mathbf{E}}(f)$	$\tilde{\boldsymbol{\sigma}}(f)$	$\tilde{\mathbf{E}}(\xi)$	$\tilde{\boldsymbol{\sigma}}(\xi)$	$\tilde{\mathbf{E}}(f)$	$\tilde{\boldsymbol{\sigma}}(f)$	$\tilde{\mathbf{E}}(\xi)$	$\tilde{\boldsymbol{\sigma}}(\xi)$	MAC	$\tilde{\mathbf{E}}(f)$	$\tilde{\boldsymbol{\sigma}}(f)$	$\tilde{\mathbf{E}}(\xi)$	$\tilde{\boldsymbol{\sigma}}(\xi)$	MAC
1	6.26	0.048	3.5	0.55	6.60	0.004	4.3	0.11	1.00	6.68	0.003	4.4	0.11	0.99
2	7.91	0.023	3.7	0.12	9.10	0.022	5.4	0.19	0.98	9.23	0.018	5.3	0.22	0.98
3	16.21	0.048	3.8	0.58	17.49	0.097	4.3	0.68	0.95	17.60	0.102	4.4	0.88	0.94
4	20.53	0.106	5.8	0.49										
5	21.46	0.089	5.9	0.50										
6	28.91	0.049	2.1	0.20	29.08	0.093	1.4	0.21	0.92	28.82	0.102	2.4	0.44	0.95
7	30.02	0.179	3.7	0.61						29.95	0.079	1.7	0.41	0.97
8	33.01	0.034	2.1	0.20	33.45	0.155	2	0.28	0.80	33.60	0.055	2.1	0.25	0.78
9	36.79	0.124	2.5	0.10	36.62	0.057	1.7	0.18	0.97	36.65	0.050	1.7	0.14	0.97
10										38.54	0.083	1.8	0.23	-
11	40.76	0.122	1.9	0.32	41.39	0.084	2.1	0.24	0.90					
12					45.78	0.229	2.2	0.39	-	45.62	0.088	1.8	0.15	-
13	51.23	0.205	2.7	0.62	51.20	0.126	1.8	0.23	0.97	50.51	0.152	2.2	0.25	0.93
14					60.96	0.754	0.4	0.11	-	60.83	0.685	0.4	0.28	-
15					66.62	0.262	0.4	0.12	-	66.64	0.254	0.5	0.13	-
16	72.14	0.162	0.7	0.31	72.35	0.075	0.5	0.09	0.42					
17	85.44	0.357	0.9	0.40										
18	90.45	0.543	1.7	0.40						90.54	0.117	1.0	0.20	0.93
d_0			0.1077				0.1260					0.1283		
a			0.48				0.36					0.44		

TABLE 5: Modal selection results obtained for RUN6 in the OO and IO context – Mean values and standard deviations of undamped frequencies $\tilde{\mathbf{E}}(f), \tilde{\boldsymbol{\sigma}}(f)$ (in Hz) and damping ratios $\tilde{\mathbf{E}}(\xi), \tilde{\boldsymbol{\sigma}}(\xi)$ (in %) obtained after carrying out selection procedure FCM-$X_{[1:7]}$. The two last rows indicate the (automatically-)chosen threshold d_0 employed for each procedure and the ratio $a = \operatorname{card} \mathcal{E}^{ph}/J$ achieved by the clustering algorithm. The MAC values appearing in the table measure the correlation between modeshapes computed in the IO case and those obtained in the OO case (extracted at compatible sensors position).

vice and the RC structure is very low, such that the modes are barely observable (extremely low MTN and close-to-0 nMTN) and has been classified as spurious. The proper characterization of such modes can be done including in the analysis the triaxial accelerometer that was placed on the piping system. For conciseness and because of the low interest of these modes for our study, no further detail is given.

Damping ratios, in turn, are coherently identified for both OO and IO analysis and give results within $[1; 5.5\%]$ for the modes involving the reinforced-concrete structure. The uncertainties being much higher regarding damping, no real general trend is to observe.

4.2. Evolution of the modal signature – Correlation with observed damage

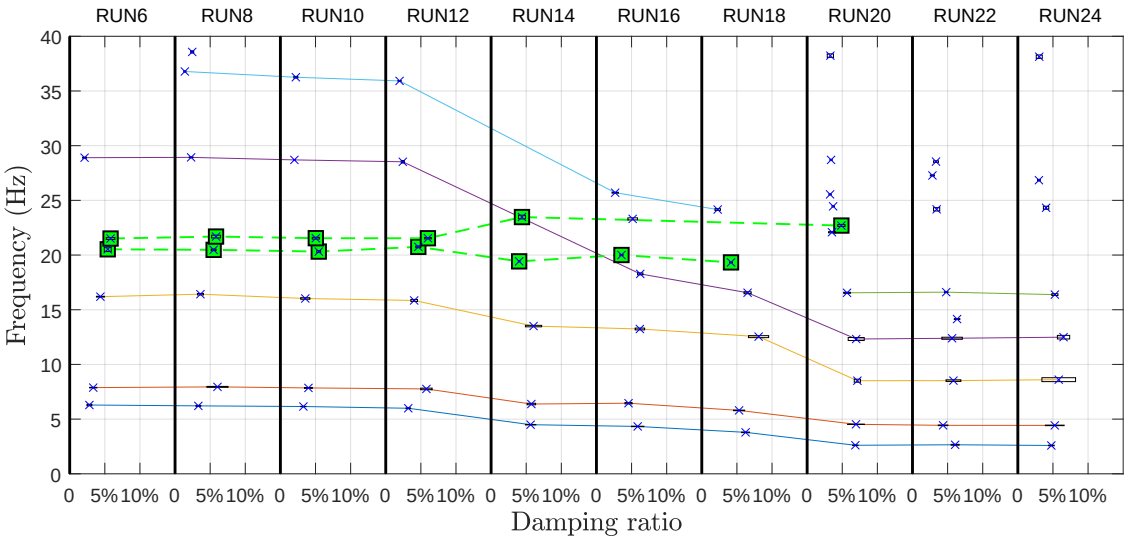
Figures 19a and 19b show the evolution of the modal signature in terms of frequency (vertical) axis and damping ratio (local horizontal axis) along the SMART2013 test campaign respectively in the OO and IO context. Groups of similar modes are made from the results of independent analysis and modes are joined together with solid lines of different colors for the six first modes of table 5. For building those groups, a last post-processing is carried out using hierachical clustering on $1 - \text{MAC}(\psi_i, \psi_j)$ distance and with a stop criterion on the maximum cardinal number of each group (equal to ten, the number of experimental runs processed in this work). On figure 19a, the hydraulic modes are pointed out with green squares.

One can note for the RC-involved modes a progressive reduction of the six main natural frequencies with two notable drops after RUN13 and 19 for which a sudden rupture of the steel reinforcements was observed at the base of the connection between the small wall V04 and the soleplate, together with a strong damage level of concrete medium (see fig. 20). A sensitive damping ratio increase is also to noticeable. One of the contributions of this work thus concerns the experimental characterization of damping ratios for a RC structure having different damage rates. Corresponding values are gathered in table 6 where damping ratios are corrected by a factor $\sqrt{s}$ to account for Cauchy-Froude's similitude ($s = 1/4$ scale factor) and obtain typical damping ratios to be expected on full scale reinforced-concrete structures.

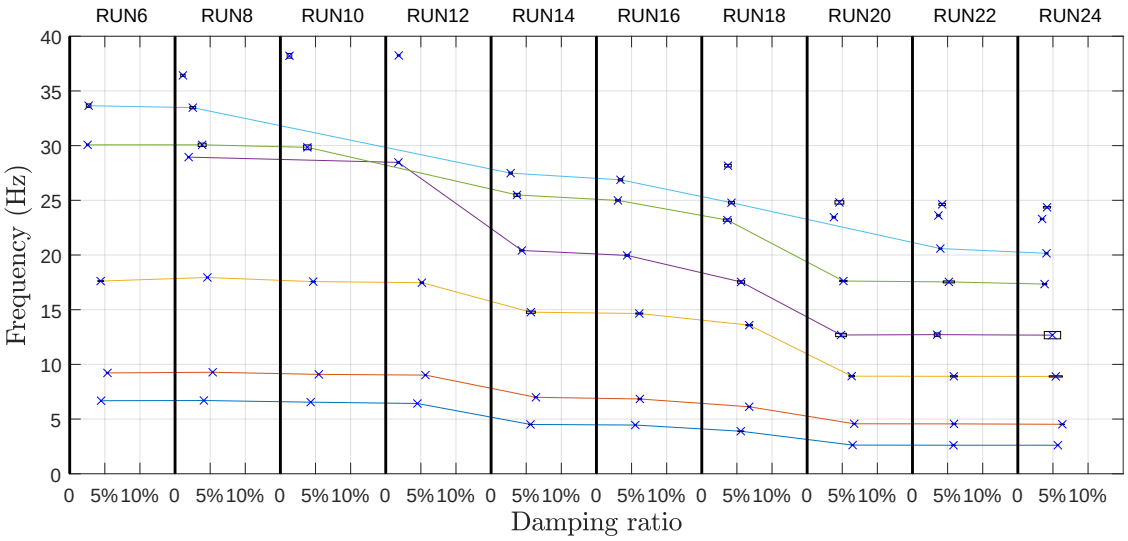
Damage level (RC structure)	Healthy	Light	Extended
Damping ratio (experimental specimen)	2-5%	5-7%	5-8%
Damping ratio (full scale structure)	1-2.5%	2.5-3.5%	2.5-4%

TABLE 6: Damping coefficients obtained imposing low $0.02g$ RMS broad-band ergodic ground motion to RC structure for different damage states. Light damage: cracks in the concrete matrix, possible rupture of reinforcements in most loaded area. Extended damage: rupture of all reinforcements in the most loaded area, complete loss of stiffness of the RC structure locally observed.

The stability of the modeshapes characteristics along the test campaign has also been investigated, computing for each RUNk the MPD_i^k and $\text{MAC}(\phi_i^6, \phi_i^k)$ values for modes $i \in [1; 6]$ and considering the modeshapes identified for RUN6 as reference. No clear trend could be observed except for a slight increase in the MPD level for the first mode (flexion $/x$) and third mode (torsion $/z$) after RUN19 rising from typical 0.1 to 0.2 values. The second mode (flexion $/y$) however, has shown stable close-to-0.05 MPD and close-to-1 MAC values. A slight decrease of $\text{MAC}(\phi_3^6, \phi_3^k)$ was also noted. The slight MPD_1 increase can be explained by the presence of local dissipative phenomena of hysteretic nature (friction between concrete's



(a) OO analysis.

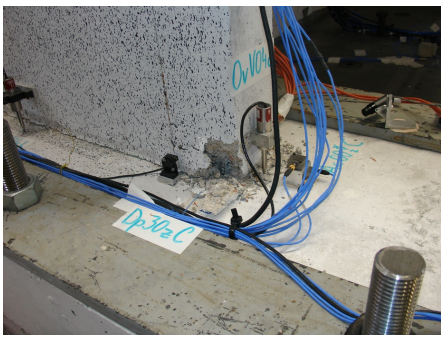


(b) IO-1 analysis.

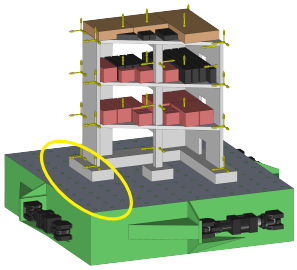
FIGURE 19: Evolution of the modal signature. One color per mode, the mean values $(\tilde{\mathbf{E}}(f_i), \tilde{\mathbf{E}}(\xi_i))$ are marked with crosses with while centered rectangles of dimensions $\tilde{\boldsymbol{\sigma}}(f_i) \times \tilde{\boldsymbol{\sigma}}(\xi_i)$ materialize their corresponding standard deviations. Green dashed lines point modes with important hydraulic participation. See fig. 20 for a visual description of associated damage on the RC specimen.

crack lips at the base of wall V04) and MAC variations indicate the effect of the local loss of stiffness (rupture of reinforcements and cracked concrete medium) on the modal signature. Modes 4 to 6 in turn exhibited generally higher MPD levels within $[0.3; 0.4]$ and MAC values associated to such dispersion that no real conclusion could be drawn. These less stable modal properties reflect above all the difficulty of the subspace algorithms to identify less observable modes. Let one finally note that the instrumentation considered in this work is maybe too sparse (especially in the damaged area) to enable further accurate tracking of the modeshapes characteristics.

After RUN13



After RUN19



Position of wall V04

FIGURE 20: Main damage area of the SMART2013 specimen: base of the wall V04. (After RUN13) first crack appearing near the junction between wall and soleplate – sudden rupture of one or more vertical reinforcements. (After RUN19) rupture of the rest of the vertcal reinforcements at the base of wall V04 – complete loss of rigidity between wall and soleplate (breathable crack) – high damage level to concrete.

5. Conclusion

This paper describes a fuzzy-driven multistage clustering strategy for automating the time consuming modal selection process when performing growing model-order system identification. Just like earlier references, in the continuity of which this work is part, the presented automated selection strategy does not contain any case-dependent manually-chosen threshold nor parameter and the fuzzy framework has proven to bring more robustness to the methodology in general and particularly with respect to maximum model order specification, outliers and polesplitting phenomena. Compared to other k-means based methodologies, a better adequacy of the final modal selection with the set of modal indicators specified for primary partition has also been demonstrated. The clarifying role of the newly derived nMTN indicator was also highlighted.

The potential of the approach has been illustrated using data from the SMART2013 test campaign during which a 1/4 reduced scale reinforced concrete specimen was submitted to a sequence of seismic ground motions of increasing level. The new automated modal selection strategy was applied to low-level broadband intermediate tests and enabled the characterization of the modal signature of the specimen for different damage states. Both Output-Only and Input/Output system identification have been carried out for estimating the coupling between the shaking table device and the experimental specimen. One important conclusion to be seen on the first natural frequencies is that the impedance associated to the AZALEE shaking table itself seem negligible compared to that of the hydraulic actuators. Modeling the shaking-table-to-specimen interaction should focus on the interaction with the controlled hydraulic device.

Progresses still need to be made to deal with the case of high modal densities which may still be problematic. Some of the single mode validation criteria used in this work such as the MPD or nMTN may loose relevance in the case of multimodal behavior and could lead coupled physical modes to be wrongly classified as artifacts. One has to think in terms of modal subspace, considering the combination of several modes at very close frequencies.

Acknowledgments

The tandem EDF-CEA is deepfully acknowledged for funding this research activity and giving access the SMART2013 test campaign database. The work reported in this paper has been also supported by the SEISM Institute. F. Voldoire (EDF) is more personally thanked for the fruitful discussions while following up this research activity. C. Feau (CEA) and A. Le Maoult (CEA) are also thanked for being such valuable beta-testers: their permanent feedback has been an asset for the development of many of the automated selection tools proposed in this paper. The MSSP reviewers are finally acknowledged for their numerous comments and advice, that were of great help to shape this contribution in its current form.

Appendix A. Parameters used for the 2D test-cases (sec. 3.1)

The bivariate expression of the lognormal pdf used for defining the 2D reference clusters of fig. 4 reads:

$$\log \mathcal{N}(x;\mu,\Sigma) \;=\; \frac{1}{2\pi\,|x_1|\,|x_2|\,\sqrt{|\Sigma|}}\,\exp\,-\frac{1}{2}(\log x-\mu)\;\Sigma^{-1}\;(\log x-\mu)^T \tag{A.1}$$

The following parameters were used for cluster 1 (in blue):

a. convex case

$$\mu^{(1)}=\begin{bmatrix}-2.5\\-2.5\end{bmatrix}\quad,\quad \Sigma^{(1)}=\begin{bmatrix}0.40 & 0.06\\0.06 & 0.40\end{bmatrix} \tag{A.2}$$

b. non-convex case

$$\mu^{(1)}=\begin{bmatrix}-1.8\\-1.8\end{bmatrix}\quad,\quad \Sigma^{(1)}=\begin{bmatrix}0.325 & -0.225\\-0.225 & 0.250\end{bmatrix} \tag{A.3}$$

A reversed lognormal pdf $\log \mathcal{N}(x^-;\mu,\Sigma)$ was used for generating cluster 2 (in red) defining $x^- = -(x-\tau)$ and choosing:

a. convex case

$$\tau^{(2)}=\begin{bmatrix}3\\3\end{bmatrix}\quad,\quad \mu^{(2)}=\begin{bmatrix}0.96\\0.92\end{bmatrix}\quad,\quad \Sigma^{(2)}=\begin{bmatrix}0.0015 & 0.0010\\0.0010 & 0.0027\end{bmatrix} \tag{A.4}$$

b. non-convex case

$$\tau^{(2)}=\begin{bmatrix}2.8\\2.8\end{bmatrix}\quad,\quad \mu^{(2)}=\begin{bmatrix}0.88\\0.84\end{bmatrix}\quad,\quad \Sigma^{(2)}=\begin{bmatrix}0.0015 & 0.0010\\0.0010 & 0.0027\end{bmatrix} \tag{A.5}$$

The choice of those parameters is not crucial, however, some care has been taken to tune them for generating convex and non-convex clusters with equivalent overlapping areas. Other pdf choices leading to similar clusters in terms of spread, convexity and overlapping could be used and would lead to analogous results.

References

[1] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems : Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.

[2] P. Guillaume, P. Verboven, S. Vanlanduit, H. Van der Auweraer, and B. Peeters. A poly-reference implementation of the least-squares complex frequency-domain estimator. In *Proc of IMAC XXI*, Kissimee, Florida, February 2003.

[3] R. J. Allemang, D. L. Brown, and A. W. Phillips. Survey of modal techniques applicable to autonomous/semi-autonomous parameter identification. In *International Conference on Noise and Vibration Engineering (ISMA)*, 2010.

[4] Arnaud Deraemaeker and Keith Worden, editors. *New Trends in Vibration Based Structural Health Monitoring*. Springer, 2010.

[5] F. Ubertini, N. Cavalagli, A. Kita, and G. Comanducci. Assessment of a monumental masonry bell-tower after 2016 central italy seismic sequence by long-term shm. *Bulletin of Earthquake Engineering*, 16:775–801, 2018.

[6] C. Gentile, A. Ruccolo, and F. Canali. Long-term monitoring for the condition-based structural maintenance of the milan cathedral. *Construction and Building Materials*, 228:117101, 2019.

[7] L. Mevel, M. Basseville, and A. Benveniste. Fast in-flight detection of flutter onset: A statistical approach. *Journal of Guidance, Control, and Dynamics*, 28(3):431–438, 2005.

[8] M. Döhler and L. Mevel. Subspace-based fault detection robust to changes in the noise covariances. *Automatica*, 49:2734–2743, 2013.

[9] E. Reynders, J. Houbrechts, and G. De Roeck. Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing*, 29:228–250, 2012.

[10] C. Devriendt, F. Magalhães, W. Weijtjens, G. De Sitter, A. Cunha, and P. Guillaume. Structural health monitoring of offshore wind turbines using automated operational modal analysis. *Structural Heath Monitoring*, 13(6):644–659, 2014.

[11] C. Rainieri and G. Fabbrocino. Development and validation of an automated operational modal analysis algorithm for vibration-based monitoring and tensile load estimation. *Mechanical Systems and Signal Processing*, 60-61:512–534, 2015.

[12] R. S. Pappa, K. B. Elliott, and A. Schenk. Consistent-mode indicator for the eigensystem realization algorithm. *Journal of Guidance, Control, and Dynamics*, 16(5):852–858, 1993.

[13] E. Reynders and G. De Roeck. Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis. *Mechanical Systems and Signal Processing*, 22:617–637, 2008.

[14] J. Lardies and M.-N. Ta. Modal identification of stay calces from outout-only measurements. *Mechanical Systems and Signal Processing*, 25:133–150, 2011.

[15] R. Pintelon, P. Guillaume, and J. Schoukens. Uncertainty calculation in (operational) modal analysis. *Mechanical Systems and Signal Processing*, 21:2359–2373, 2007.

[16] E. Reynders, R. Pintelon, and G. De Roeck. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical Systems and Signal Processing*, 22(4):948–969, 2008.

[17] P. Mellinger, M. Döhler, and L. Mevel. Variance estimation of modal parameters from output-only and input/output subspace-basedsystemidentification. *Journal of Sound and Vibration*, 2016.

[18] F. Magalhães, , C. Álvaro, , and E. Caetano. Online automatic identification of the modal parameters of a long span arch bridge. *Mechanical Systems and Signal Processing*, 23:316–329, 2009.

[19] F. Ubertini, C. Gentile, and A. L. Materazzi. Automated modal identification in operational conditions and its application to bridges. *Engineering Structures*, 46:264–278, 2013.

[20] B. Richard, S. Cherubini, F. Voldoire, P.-E. Charbonnel, T. Chaudat, S. Abouri, and N. Bonfils. Smart 2013: Experimental and numerical assessment of the dynamic behavior by shaking table tests of an asymmetrical reinforced concrete structure subjected to high intensity ground motions. *Engineering Structures*, 109:99–116, 2016.

[21] Brian M. Broderick, Jamie Goggins, Darko Beg, Ahmet Y. Elghazouli, Philippe Mongabure, Alain Le Maoult, Alan Hunt, Suhaib Salawdeh, Primoz Moze, Gerard O'Reilly, and Franc Sinur. *Assessment of the Seismic Response of Concentrically-Braced Steel Frames*, pages 327–344. Springer International Publishing, Cham, 2015.

[22] Vincent Crozet, Ioannis Politopoulos, and Thierry Chaudat. Shake table tests of structures subject to pounding. *Earthquake Engineering & Structural Dynamics*, 48(10):1156–1173, 2019.

[23] C. Boudaud, J. Baroth, and L. Daudeville. Influence of joint strength variability in timber-frame structures: propagation of uncertainty through shear wall finite element models under seismic loading. *Can. J. Civ. Eng.*, 43:609–618, 2016.

[24] Antonio C. Caputo, Fabrizio Paolacci, Oreste S. Bursi, and Renato Giannini. Problems and Perspectives in Seismic Quantitative Risk Analysis of Chemical Process Plants. *Journal of Pressure Vessel Technology*, 141(1), 12 2018. 010901.

[25] E. Neu, F. Janser, A. A. Khatibi, and A. C. Orifici. Fully automated operational modal analysis using multi-stage clustering. *Mechanical Systems and Signal Processing*, 84:308–323, 2017.

[26] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate Data Analysis*. Pearson Education Limited, 7th edition edition, 2014.

[27] H. G. Harris and G. Sabnis. *Structural Modeling and Experimental Techniques*. CRC Press, 2nd edition edition, mar 1999.

[28] D. J. Ewins. *Modal Testing : Theory, Practice and Application*. Research Studies Press, Baldock, Hertfordshire, UK, 2000.

[29] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[30] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *Ann. Statist.*, 36(2):555–586, 04 2008.

[31] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems*. MIT Press, 2004.

[32] Feng Zhao, Hanqiang Liu, and Licheng Jiao. Spectral clustering with fuzzy similarity measure. *Digital Signal Processing*, 21:701–709, 2011.

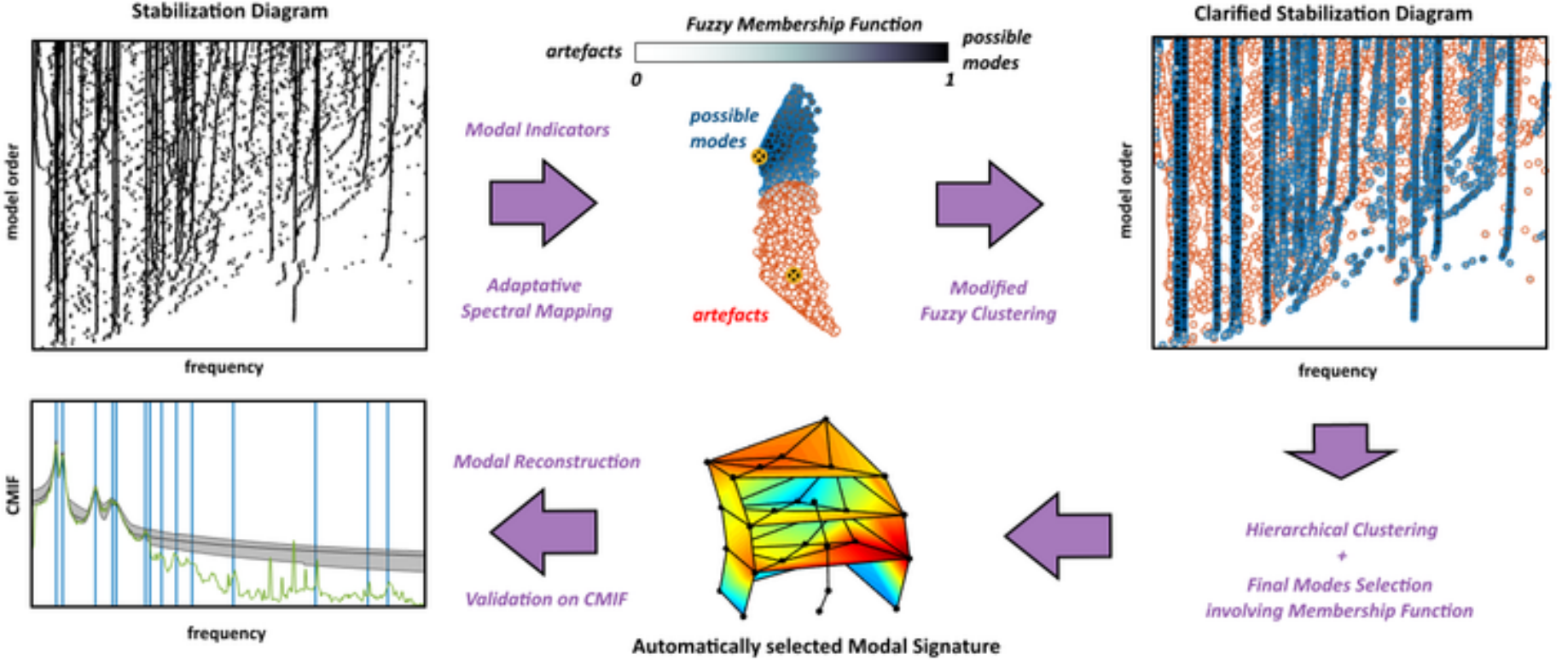
[33] J. C. Bezdek and J. C. Dunn. Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. *IEEE Transactions on Computers*, 24(8):835–838, 1975.

[34] Donald E. Gustafson and William C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Decision and Control including the 17th Symposium on Adaptive Processes*, 1978.

[35] C. Y. Shih, Y. G. Tsuei, R. J. Allemang, and D. L. Brown. Complex mode indication function and its applications to spatial domain parameter estimation. *Mechanical Systems and Signal Processing*, 2(4):367–377, October 1988.

[36] P. Stoica and R. Moses. *Spectral analysis of signals*. Prentice Hall, 2004.

[37] B. Richard, F. Voldoire, M. Fontan, J. Mazars, T. Chaudat, S. Abouri, and N. Bonfils. Smart 2013: Lessons learned from the international benchmark about the seismic margin assessment of nuclear rc buildings. *Engineering Structures*, 116:207–222, 2018.


Stabilization Diagram
model order
frequency
Fuzzy Membership Function
artefacts
possible modes
0
1
Modal Indicators
Adaptative Spectral Mapping
possible modes
artefacts
Modified Fuzzy Clustering
Clarified Stabilization Diagram
model order
frequency
Hierarchical Clustering
+
Final Modes Selection involving Membership Function
Automatically selected Modal Signature
Modal Reconstruction
Validation on CMIF
CMIF
frequency