A TWO-STEP CLUSTERING METHOD FOR MODAL PARAMETERS ESTIMATION APPLIED TO SHM

Rharã de Almeida Cardoso

rhara.cardoso@gmail.com

Departamento de Engenharia Civil - Universidade Federal de Ouro Preto

Rua Diogo Vasconcelos, 122, Pilar, Ouro Preto, 35400-000, Minas Gerais, Brasil

Alexandre Abrahão Cury

Flávio de Souza Barbosa

alexandre.cury@engenharia.ufjf.br

flavio.barbosa@engenharia.ufjf.br

Departamento de Mecânica Aplicada e Computacional - Universidade Federal de Juiz de Fora

Rua José Lourenço Kelmer, s/n, Martelos, Juiz de Fora, 36036-330, Minas Gerais, Brasil

Abstract. *The subject of structural health monitoring is drawing more and more attention over the last years. Many vibration-based techniques aiming at detecting small structural changes or even damage have been developed or enhanced through successive researches. Lately, several studies have focused on the use of raw dynamic data to assess information about structural condition. Despite this trend and much skepticism, many methods still rely on the use of modal parameters as fundamental data for damage detection. Therefore, it is of utmost importance that modal identification procedures are performed with a sufficient level of precision and automation. To fulfill these requirements, this paper presents a novel automated time-domain methodology to identify modal parameters based on a two-step clustering analysis. The first step consists in clustering modes estimates from parametric models of different orders, usually presented in stabilization diagrams. In an automated manner, the first clustering analysis indicates which estimates correspond to physical modes. To circumvent the detection of spurious modes or the loss of physical ones, a second clustering step is then performed. The second step consists in the data mining of information gathered from the first step. To attest the robustness and efficiency of the proposed methodology, experimental data obtained from a simply supported beam tested in laboratory and from a railway bridge are utilized. The results appeared to be more robust and accurate comparing to those obtained from methods based on one-step clustering analysis.* ***Keywords:*** *structural health monitoring, automated modal identification, clustering, structural dynamics.*

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering
P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editores), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

1 INTRODUCTION

Over the last decade, methods using raw dynamic measurements are proving to be a viable alternative for the detection of structural damage. Santos (2014), for instance, presented a clever technique that relies on time series symbolic data objects and moving windows to detect structural novelties. Alves (2015) presented a methodology based on the use of symbolic data coupled with clustering techniques to detect damage using raw acceleration measurements. Despite all advances in this field of research, it is common to find studies in the literature that considers modal parameters (natural frequencies, damping ratios and mode shapes) as preferred inputs to assess structural behaviors or damage states. Evidently, the close relationship that exists between modal parameters and structural stiffness and mass explains this tendency. Therefore, it is not by chance that modal parameters are widely used in continuous dynamic monitoring systems of important civil structures, such as long-span bridges, stadiums, tall buildings, among others (Battista *et al.*, 2000, Gautier *et al.*, 2005).

Continuous monitoring systems are especially relevant due to their preventive nature, since they can be inserted in maintenance programs and actually support detecting unusual structural behaviors. Abnormal structural states are often assessed by observing the variation of modal parameters over time. Hence, the modal identification procedure becomes of utmost importance and deserves special attention concerning its accuracy, robustness and automation to make it viable in a context of online monitoring. Due to the necessity of an automatic feature, parametric time-domain based identification methods have been preferred amongst engineers. Mostly, the so-called SSI (Stochastic Subspace Identification) methods, which use discrete-time stochastic state-space models, are very suitable in this case due to their capacity to identify closely-spaced frequencies. Moreover, they provide useful ready-to-use data for feature extraction by means of clustering techniques or any other data mining methods.

At this point, it is worth highlighting the difference between modal parameter estimation (MPE) and modal tracking (MT). The former consists in the estimation of modal parameters from a single record of measured data and the latter corresponds to tracking the evolution of modal parameters of a structure through repeated MPE. This paper is limited to MPE, not involving what would be a complementary task to online monitoring (MT). Therefore, isolated signals of a given dynamic test in a structure are evaluated independently, with no time tracking of modal parameters.

In fact, without the automation of the MPE process, a lot of user interaction over a large amount of measured vibration data would be required, leading to an impossibility of practical applications such as the health monitoring of important infrastructures. Therefore, especially over the last decade, several methods aiming the reduction of the number of user-defined parameters have been developed (Magalhães *et al.*, 2009; Reynders *et al.*, 2012; Ubertini *et al.*, 2013, Cabboi, 2017), some of which reached the complete elimination of all manually specified parameters and, thus, are usually labelled as fully automated.

For instance, Reynders *et al.* (2012) proposed a fully automated approach for the interpretation of stabilization diagrams, which are usual outputs of SSI methods. Their approach was cleverly based on clustering techniques built over three stages: pre-cleaning by means of a classification of all modes into two categories (possibly physical or certainly spurious); a hierarchical clustering of the possible physical ones to group them together; and a final classification of the formed clusters into a physical or spurious condition. No user-defined parameter is needed in the entire process.

However, despite the observed advances, practice has been showing that even the most sophisticated automated process of MPE has to be preliminarily checked or tuned in order to fit a unique behavior of the analyzed structure. Otherwise, the method might become too much vulnerable to fail due to different identification scenarios, like different levels of frequencies and damping ratios, different complexities of modal shapes, several types of external excitation sources and different amounts of signal noise. For that reason, the present work focus on the development of a robust automated approach that depends on just few easy-to-set user-defined parameters.

Currently, the available time domain methods for operational modal analysis (OMA), explored in the work of Peeters and De Roeck (1999), are essentially based on two types of models: discrete-time stochastic state-space models and ARMA (Auto-Regressive Moving Average) or just AR (Auto-Regressive) models. The models can be identified either from correlations (or covariances) of the outputs: Covariance driven Stochastic Subspace Identification – SSI-COV; or directly from time series collected at the tested structure by the use of projections (Van Overschee and De Moor, 1996): Data driven Stochastic Subspace Identification – SSI-DATA. As reported in the work of Peeters and De Roeck (1999), these two methods are very closely related. Still, the SSI-COV has the advantage of being faster and based on simpler principles, whereas the SSI-DATA allows obtaining some further information with a convenient post-processing, as for instance, the decomposition of the measured response into modal contributions.

This work focuses on the use of the SSI-DATA method, since this method yields stabilization diagrams as outputs, which are a resourceful data for reliable structural modal identification. Moreover, when associated with clustering techniques, they are capable of providing very robust results. Thus, section 2 briefly discusses the SSI-DATA method (a more thorough description can be found in reference Van Overschee and De Moor, 1996).

To provide a comparison basis for this new methodology, a reference procedure based on the work of Magalhães *et al.* (2009) is presented. Section 3 details both methodologies. Section 4 shows the results obtained by both approaches applied to different identification scenarios: a simply supported steel beam tested in laboratory and a real test on a high-speed train (TGV) viaduct in France. Finally, a brief conclusion and comments on practical relevance of the work are drawn in section 5.

2 DATADRIVEN STOCHASTIC SUBSPACE IDENTIFICATION

Since the work published in the nineties (Van Overschee and De Moor, 1996), the application of data-driven SSI algorithms to determine the modal parameters of structures has been quite spread over the ambient vibration tests. By computing state space models from output data (structure's response) the identification problem can be enunciated as follows.

Given s measurements of the output $\boldsymbol{y}_k \in \mathbb{R}^l$ generated by the unknown stochastic system of order *n*:

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering
P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editores), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \mathbf{A}\boldsymbol{x}_k + \boldsymbol{w}_k \\
\boldsymbol{y}_k &= \mathbf{C}\boldsymbol{x}_k + \boldsymbol{v}_k
\end{aligned}
\tag{1}
$$

where $\boldsymbol{x}_k \in \mathbb{R}^n$ is the state vector at a discrete time instant k, $\boldsymbol{w}_k$ and $\boldsymbol{v}_k \in \mathbb{R}^n$ are zero mean, white vector sequences and l is the number of outputs of the process (number of instrumented channels).

Knowing matrices **A** and **C**, one can calculate the modal frequencies ω_i, the modal damping ratios ξ_i and the corresponding modal shapes $\boldsymbol{\phi}_i$. This calculation is performed through the eigenvalue decomposition of matrix **A** as the following statements show:

$$
\lambda_i = \frac{ln(\mu_i)}{\Delta t} \quad \rightarrow \quad
\begin{cases}
\omega_i = |\lambda_i| \;\; (rad/s) \;\rightarrow\; f_i = \dfrac{|\lambda_i|}{2\pi} \;\; (Hz) \\
\xi_i = \dfrac{-Re(\lambda_i)}{|\lambda_i|} \times 100 \;\; (\%) \\
\boldsymbol{\phi}_i = \mathbf{C}\boldsymbol{\psi}_i
\end{cases}
\tag{2}
$$

where μ_i is the i^{th} discrete-time eigenvalue of **A** and $\boldsymbol{\psi}_i$ is its corresponding eigenvector. The eigenvalues of **A** are also called *poles*. They are complex numbers and appear mostly in conjugated pairs. Only those poles which have a conjugated pair and have positive imaginary component are taken into account for the equation (2). Therefore, up to $n/2$ modes are expected to be estimated. The $Re(\bullet)$ is the symbol for the real part of $\bullet$ and λ_i denotes the continuous-time eigenvalue of **A**.

When one works with real structure response data, it is very hard to determine the system's order. Thus, one of the tools to overcome this problem is the well-known stabilization diagrams, where the modes estimates are plotted for different system orders.

3 AUTOMATION METHODS FOR STABILIZATION DIAGRAM INTERPRETATION

The modes estimates, which are the outputs of SSI-DATA algorithm, are usually presented in stabilization diagrams. The analysis of such diagrams is supposed to separate physical modes from spurious (numerical) ones. The latter inevitably exist among the data, since they are a natural consequence of the parametric model attempt to fit the response data.

The process of detecting physical modes within a stabilization diagram can be understood as its interpretation, which, ideally, must be carried out automatically. To this end, this section presents two techniques, both based on hierarchical clustering analysis. The first one, explained in section 3.1, was proposed in the work of (Magalhães *et al*, 2009) and will serve as a baseline. The second one presents a richer modal identification method by proposing a second clustering analysis, shown in section 3.2. The proposed approach is an enhancement from the first and contains some elements found in Reynders *et al.* (2012). In the remainder of the text, these two techniques will be referenced as "reference methodology" and "proposed methodology", respectively.

Although this work uses SSI-DATA routine to provide stabilization diagrams, it is relevant to point out that both interpreters are suitable for the analysis of the outputs generated by any other parametric identification technique that yields modes estimates (frequencies, damping ratios and modal shapes) for several model orders.

3.1 Reference Methodology

This approach is based on the routines presented by Magalhães *et al.* (2009) and does not apply a pre-filter to the modes estimates data. Thus, stabilization diagrams would still be full of certainly spurious modes, like those with extremely high or negative damping ratios.

This method is based on a hierarchical clustering algorithm. Hence, in order to stablish a concept of similarity/dissimilarity between objects, a dissimilarity (distance) measure between all pairs of modes estimates must be calculated. The adopted metric for the evaluation of such a distance is:

$$d_{i,j} = \left| \frac{f_i - f_j}{f_j} \right| + \left(1 - \mathrm{MAC}_{i,j}\right) \tag{3}$$

where f_i and f_j are, respectively, the natural frequencies of the modes estimates i and j. The mode shape similarity between these two modes estimates is evaluated by the well-known Modal Assurance Criterion (MAC, Ewins, 2010). Moreover, one can note that for this metric it is possible to have $d_{i,j} \neq d_{j,i}$, which leads to an oddly asymmetric dissimilarity matrix.

Hierarchical clustering algorithms differ in the way they compute the distance between two already formed clusters. In this methodology, the single-linkage criterion is used, which means that the distance between two clusters will be considered as being equal to the smallest distance, also computed with equation (3), between objects inside these two clusters. This information allows constructing the hierarchical tree.

In the following step, one must define the tree's cut-level. This is usually dependent on the expected number of clusters. However, this number is not trivial to predict due to the unknown quantity of clusters grouping spurious modes. Thus, the alternative strategy is the use of a distance limit (d_{lim}). Such threshold is the criterion to prone the branches of the hierarchical tree, generating clusters that are distant from each other more than that value. Hence, the lower is the limit, the higher is the number of resulting clusters. In this methodology, such threshold is a user-defined parameter.

At this point, one has each cluster representing a single mode. Then, the clusters are ranked according to their number of elements to help deciding whether the mode is spurious or not. Thus, the top n clusters with more elements are selected. Since physical modes are very consistent for models with different orders, their clusters present a much higher number of elements than the clusters that contain numerical estimates (which present a higher scatter between models of different orders). Therefore, it is expected to find that the n selected modes are indeed physical. In practice, it is common to realize that the number of physical modes n can be guessed, for instance, by a simple preliminary frequency domain analysis. In this methodology, this number of modes is also a user-defined parameter.

The next step is to take into account the damping ratios by means of an outlier analysis within each selected cluster. This internal filtering eliminates the modes estimates with

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering
P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editores), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

extreme damping ratio values, which are out of the range defined by ±2.698σ (standard deviation) that embraces 99.3% of the samples in a Gaussian distribution.

Finally, each selected cluster yields mean values of the modal parameters (natural frequency, modal damping ratio and mode shape), which are the outputs of the methodology.

3.2 Proposed Methodology

The main idea of the proposed methodology is to enclose three fundamental steps:

- use a time-domain method for fitting parametric models (such as SSI-DATA) to get modes estimates (stabilization diagram);
- carry out the interpretation of such diagrams using enhanced statistical techniques;
- perform a novel cluster analysis over the various interpretations results (*realizations*) obtained in the previous step.

Thus, the proposed approach contains three main subroutines, namely: stabilization diagram generator (SSI-DATA), stabilization diagram interpreter (first clustering analysis) and the second clustering analysis routine. The flowchart of Fig. 1 shows each step of the proposed methodology.

During steps 1 and 2 (Fig. 1), the user must feed the routines with some input parameters, namely:

- n – number of desired modes to be identified as final output
- f_s – signals sampling frequency
- f_{min} and f_{max} – minimum and maximum frequencies to delimit a frequency band to be considered in the analysis
- m_{lim} – minimum MPC value of a mode estimate to be considered as physical

In step 3, the limits for the number of desired modes (n_{min} and n_{max}) are computed depending on the value of n. Those values are set as 50% less or 50% more than n, rounded to the nearest integer. For instance, for the case when n is equal to 4, n_{min} and n_{max} would be equal to 2 and 6, respectively.

In step 4, the routine loads a certain number n_{mat} of signal matrices ($\mathbf{M}_1$, $\mathbf{M}_2$, …, $\mathbf{M}_{nmat}$). After that, a loop over each one of these matrices is initiated with index k=1 (step 5). Thus, the k^{th} signal matrix is kept loaded for step 6. Based on it, three values for maximum model orders are set internally (ord_1, ord_2 and ord_3). The bigger is the size of signal matrix $\mathbf{M}_k$, the smaller are these values (step 7). This is due to the fact that high-order state-space models imply on huge block Hankel matrices, which may lead to an exorbitant computational effort for solving the pseudo-inverse operation of equation (2).

In the sequence, a second loop begins (step 8). This loop is carried over each previously computed value of maximum order of state-space model. Such an information is an input to the first main subroutine (step 9): the stabilization diagram generator (SSI-DATA).

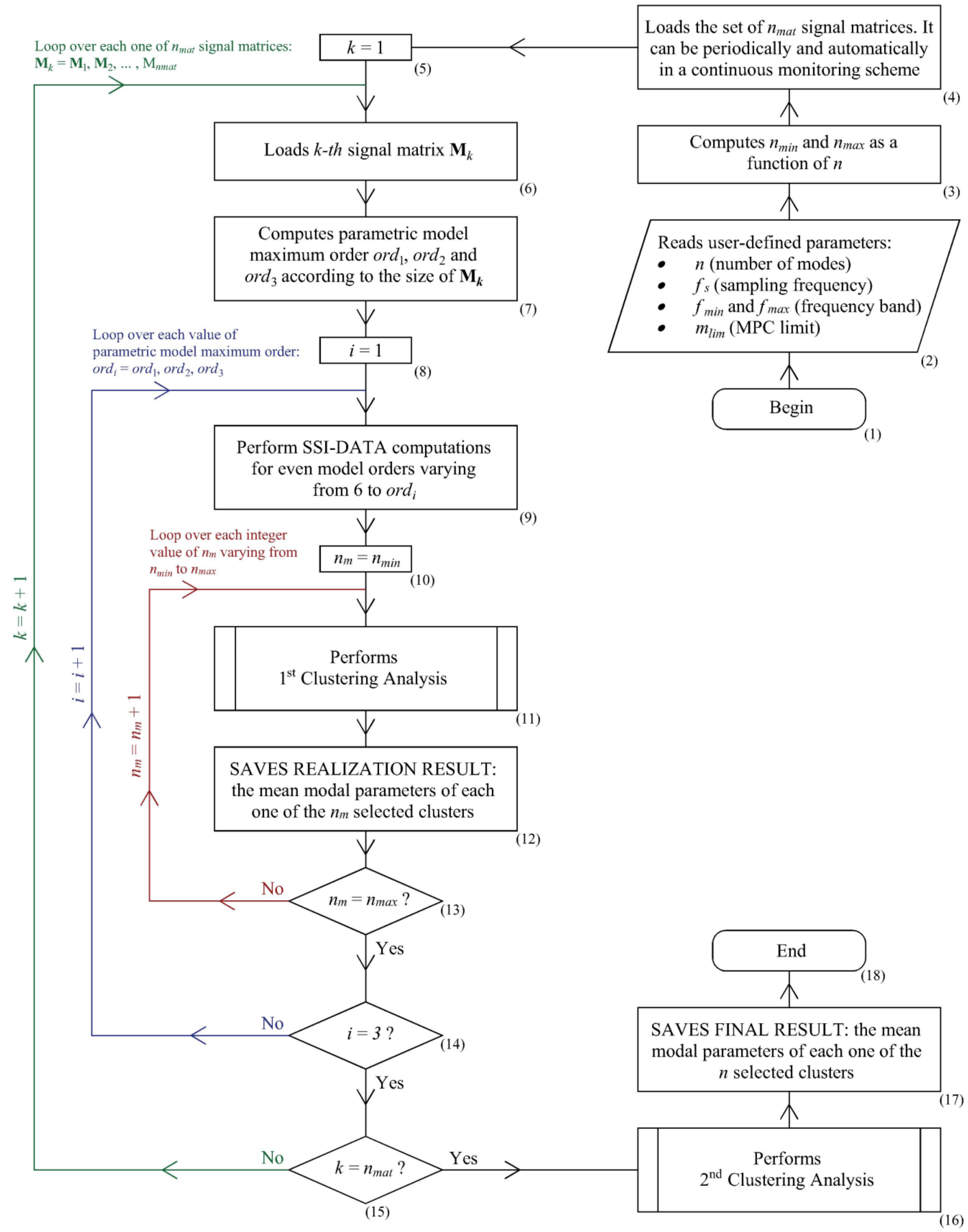


Figure 1. Flowchart of the proposed method

At this point, the third loop (step 10) begins, which is over each value of n_m (number of modes of this iteration), varying from n_{min} and n_{max}. Once SSI-DATA yields the stabilization diagram in the previous step, an enhanced interpreter is called. This consists in step 11, which is the second main subroutine (first clustering analysis). Its flowchart is depicted in Fig. 2.

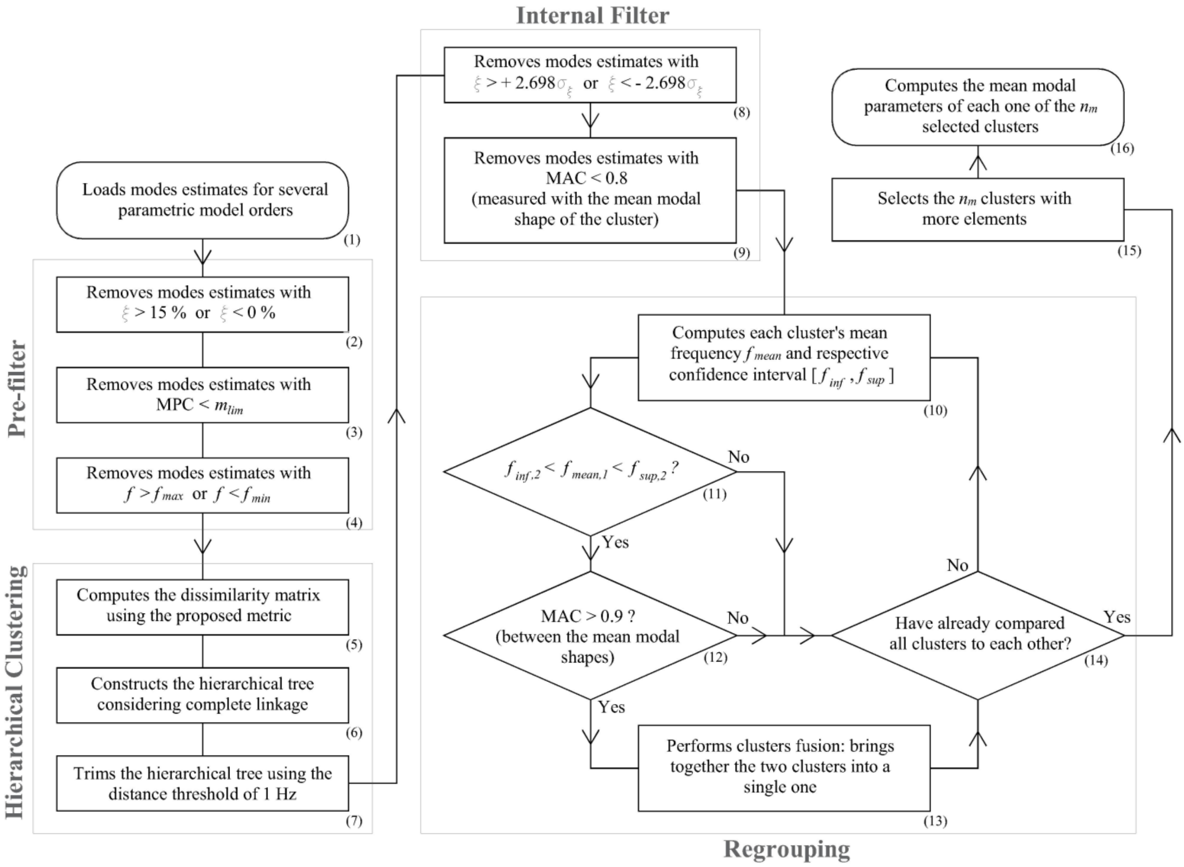


Figure 2. Flowchart of the enhanced interpreter (1st clustering analysis) – step 11 of flowchart of Fig. 1

Compared to the reference interpreter, the main new features found in the flowchart (Fig. 2) rely on the innovative way of measuring distances between modes estimates during the clustering process. Moreover, modifications regarding filtering of spurious modes and a novel technique for regrouping clusters, which were originally separated incorrectly, are also included at this step.

Like the reference method, this diagram interpreter is based on a hierarchical clustering algorithm. However, differently from that approach, one intends avoiding unnecessary computational effort and aims to obtain better results from the clustering process. Thus, a pre-filter is applied on the modes estimates data, which removes all modes whose damping ratios are not between 0 and 15% (common values for civil engineering structures). In addition, all modes with Modal Phase Collinearity (MPC) indicator lower than the limit m_{lim} in the [0, 1] interval of this criterion are eliminated. The MPC index of a vector $\boldsymbol{\phi}$ was stated by Juang and Pappa (1985), revised by Pappa *et al.* (1993) and specifically studied, together with MAC (Modal Assurance Criterion), by Vacher *et al.* (2010). A number of authors, such as Reynders et al. (2012) and Yun et al. (2012), used this index to help distinguishing physical modes from spurious ones. An MPC value closer to 1 indicates that the mode shape vector is "monophasic" and, hence, is more likely to be physical, while values closer to 0 show that the mode is either a noise mode or the mode is significantly complex.

Furthermore, the user may also remove all modes estimates out of a predefined range of frequencies (much like a “band-pass” filter on the modes estimates data). All operations explained until this point correspond to steps 2, 3 and 4 of the flowchart in Fig. 2.

After having the modes estimates data processed by this pre-filter, a plot of the stabilization diagram reveals that a significant amount of modes were removed, yielding a clearer aspect. Depending on the situation, and this is not rare, the removed modes may even represent more than 50% of the initial data. These removed modes estimates are considered as certainly spurious (or undesired) and will not be processed by the subsequent clustering algorithm.

In step 5 (Fig. 2), for the distance calculation between the modes estimates *i* and *j*, the following metric is proposed:

$$d_{i,j} = \left| f_i - f_j \right| + \left(1 - \mathrm{MAC}_{i,j} \right) c \tag{4}$$

where c is a constant arbitrarily[1] set to 5 Hz, which after some tests was stablished by the authors for being a satisfactory weight value. One should note that, differently from Equation (3), this metric leads to a perfectly symmetric dissimilarity matrix, i.e. $d_{i,j}=d_{j,i;}$

At this point, the hierarchical clustering is applied. However, differently from the reference method, the distance between two already formed clusters is measured according to the complete linkage criterion, which means that such a distance is equivalent to the largest possible distance between two of their elements. With this information, one can construct the hierarchical tree that will be trimmed by a constant distance threshold of 1 Hz (steps 6 and 7 in Fig. 2). So far, there is no need to set this threshold to another value in order to ensure the success of the method and, therefore, it remains fixed as internal constant of the routine (Cardoso, 2015).

Contrary to what the reference technique does, the internal filter is applied within the clusters before they are ranked according to their number of elements. This is done to penalize the clusters with high scatter (likely to be representing a numerical mode) before they could be ranked among the top n_m modes that will be selected as physical ones. Thus, the internal filter removes outlier modes within the cluster considering not only damping ratios, but also mode shapes. Mode estimates having damping ratios out of the $\pm 2.698\sigma$ range and/or having MAC values lower than 0.8 related to the cluster average mode shape are eliminated (steps 8 and 9 in Fig. 2).

After this grouping process, depending on the d_{lim} value, it is possible to find that different clusters are representing the very same physical mode. In other words, a single physical mode could have been split into two or more different clusters. Evidently, this phenomenon needs to be avoided. Therefore, the authors proposed a regrouping procedure in order to minimize this problem.

[1] During initial studies, the authors varied this constant from 1 to 10 Hz without obtaining relevant differences in the outputs, which points out the low sensitiveness of the enhanced interpreter to such a parameter. As the authors proved in Cardoso (2015) and in this paper, the weighting value set to 5 Hz was appropriate to all presented applications. So far, there was no need to set this constant to another value in order to ensure the success of the method and, therefore, no deeper study about this topic was performed.

The regrouping routine (steps 10 to 14 in Fig. 2) is based on the confidence interval of the frequencies. Each single cluster has its frequency mean and its lower and upper limits of frequencies delimiting a certain confidence interval. All clusters are compared two by two. If the first cluster has its average frequency within the confidence limits of the second cluster, they are possibly representing the same physical mode. Therefore, they are selected to go through a further analysis: the MAC between among their mode estimates. If the MAC is higher than 0.9, the clusters are grouped into one single cluster.

Finally, the top n_m clusters (modes) with more elements are selected as being physical (step 15 in Fig. 2). Then, for the methodology output, each selected cluster has its mean modal parameters (natural frequency, damping ratio and mode shape) calculated (step 16 in Fig. 2). Such results correspond to one single realization and are saved to incorporate the dataset to be analyzed by the second clustering analysis (step 12 of Fig. 1).

After the algorithm flows out of all three loops (steps 13, 14 and 15 in Fig. 1), the third and last main routine is fed with several realizations (stabilization diagrams interpretations) to perform a second clustering analysis. This corresponds to step 16 in Fig, 1. Fig. 3 presents the detailed flowchart of the second clustering analysis.

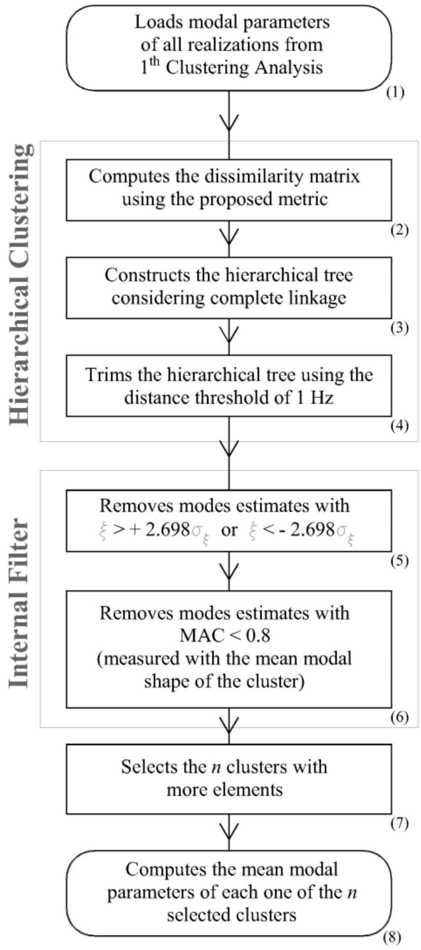


Figure 3. Flowchart of the 2nd cluster analysis – step 16 of the proposed method shown in Fig. 1

By comparing flowcharts of Fig. 2 and Fig. 3, it is clear that the second cluster analysis algorithm is very simple and similar to the first one. The only difference is that the former has no pre-filter or regrouping procedures.

The output from the second clustering analysis (step 17 and 18 in Fig. 1) is the final output of the proposed methodology, i.e. *n* identified modes (steps 16, 17 and 18 in Fig. 1).

4 RESULTS

4.1 Laboratory experiment

This section presents an experimental test conducted in laboratory on a simply supported steel beam. This beam is 1.46 m long with rectangular cross-section (76.2 x 8.0 mm) and is instrumented with six piezoelectric accelerometers (PCB, 336C31). Data acquisition is carried out using Lynx ADS2002 equipment, which essentially is a conditioning/amplifier regulating system. This study considered random vibration tests (using a shaker). The random excitation is applied throughout the duration of the tests. A 25 s signal, sampled at 4000 Hz, is shown in Fig. 4.

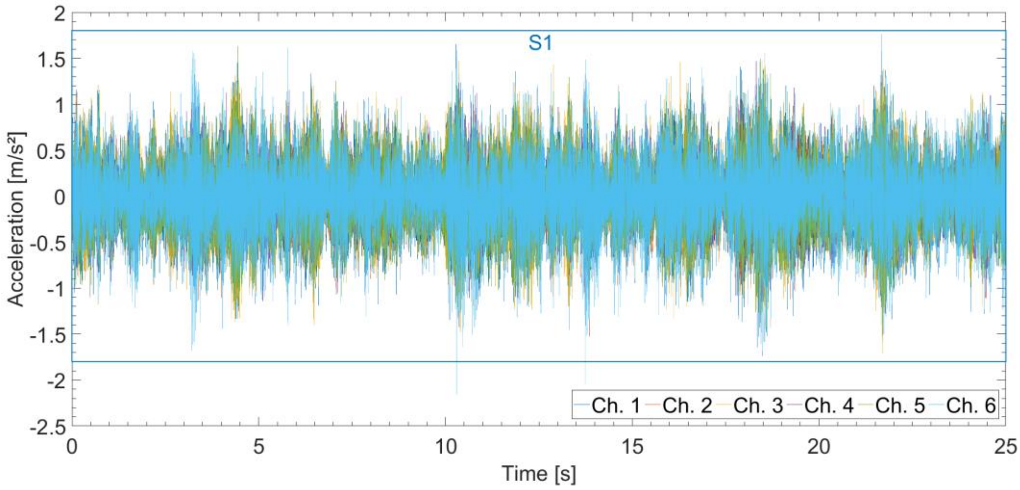


Figure 4. Signal from instrumented steel beam

In this study, the reference methodology has the d_{lim} value set to 0.01, while the proposed approach has the m_{lim} set to 0.9. It was asked both to perform the identification of 5 modes (n.

Figure 5 highlights the different performances reached by both techniques. The stabilization diagram provided by the former (Fig. 5a) is full of spurious modes estimates, while the latter shows that its pre-filters worked adequately (Fig. 5b). In those stabilization diagrams, the red colored points correspond to the physical modes. Additionally, the number of elements in each cluster, presented by the stem plots depicted in Figs. 10c and 10d show, respectively, that the reference interpreter not only missed the first mode, but also gathered several spurious modes in its fifth cluster (frequency close to 200 Hz). Conversely, due to its different metric, filters and regrouping technique, the enhanced interpreter was indeed able to correctly identify all five modes (the red headed stems stand for the five selected clusters related to physical modes).

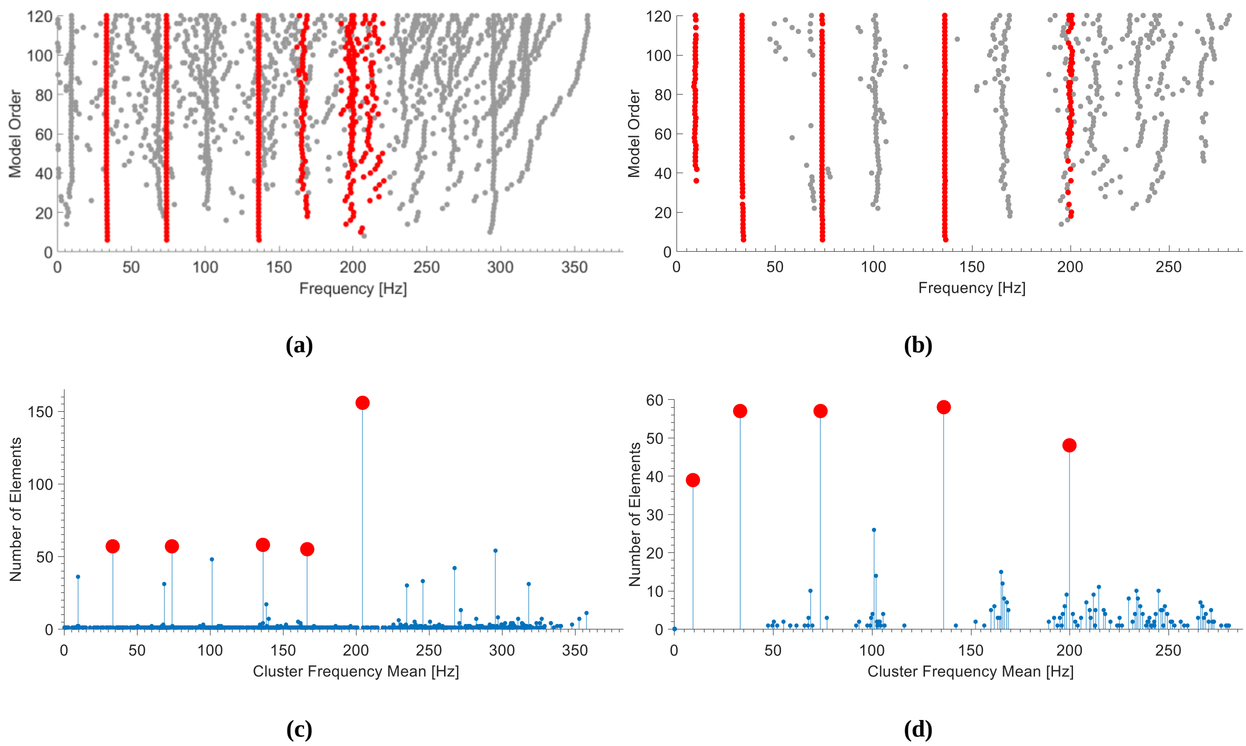


Figure 5. Modes estimates interpretation by reference interpreter (a, c) and enhanced interpreter (b, d) – realization S1M5R120

The results obtained by the enhanced interpreter shown in Fig. 6 correspond to a single realization of the proposed method (highlighted by the green text in the CSD displayed in Fig. 6).

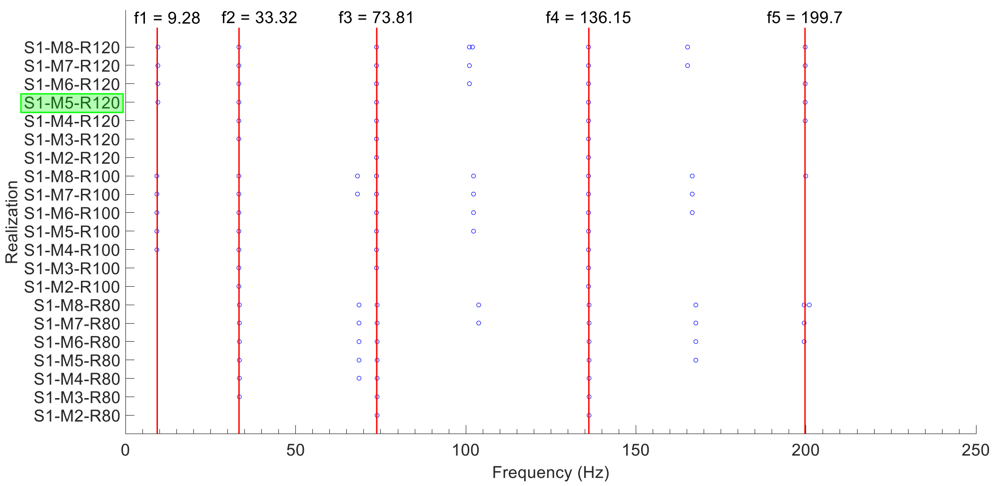


Figure 6. Clustered Stabilization Diagram of the analyzed signal

The mode shapes identified by the proposed approach are presented in Fig. 7.

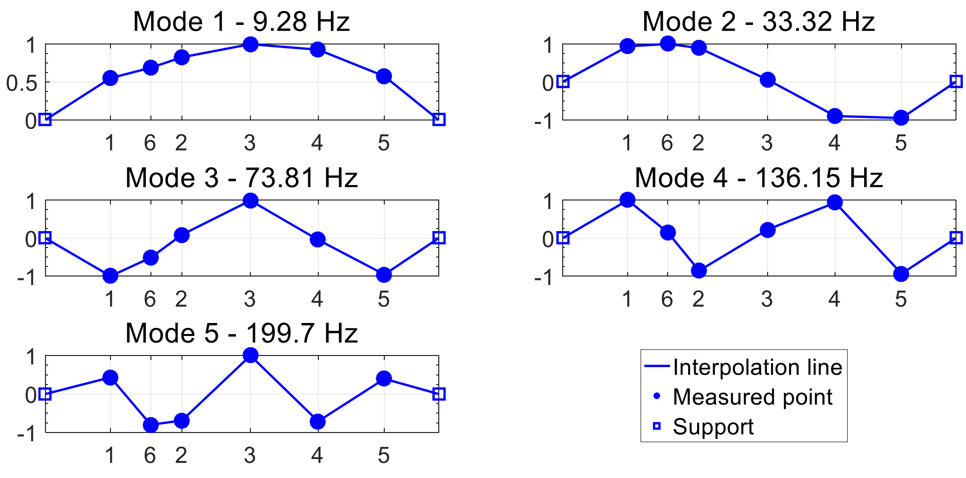


Figure 7. Modal shapes obtained by the proposed method

Finally, Table 1 summarizes the results obtained by the proposed methodology and by three attempts for the reference approach (varying d_{lim}). In all cases, the latter was not able to identify the first mode. Besides, it presented high values of intra-cluster standard deviations for the fifth mode. This fact is mainly due to the metric used in the reference interpreter. As noted in equation (3), the metric measures distances between modes estimates in a relative manner in terms of frequency, yielding a biased behavior with low-frequency modes distances and with high-frequency modes distances. Hence, clusters gathering high-frequency modes estimates tend to gather too much elements ('gathering phenomenon') erroneously. Conversely, clusters with low-frequency modes estimates tend to split into other clusters ('splitting phenomenon'). As for the proposed methodology, all standard deviations were very low due to the enhanced diagram interpreter, which considered an innovative metric, additional filters and a regrouping technique.

Table 1. Results of the automatic modal identification for the five vertical modes

Mode	**Parameter**	**Reference Interpreter (R120)**			**Proposed**
		d_{lim} = 0,01	d_{lim} = 0,005	d_{lim} = 0,001	
01	*f*(Hz)/σ(Hz)	missing	missing	missing	9.28/0.120
	ξ (%)/σ (%)				7.13 1.05
02	*f*(Hz)/σ(Hz)	33,28/0,160	33,26/0,139	33,20/0,035	33.32/0.047
	ξ (%)/σ (%)	3,23/0,19	3,20/0,14	3,17/0,12	3.34/0.06
03	*f*(Hz)/σ(Hz)	73,79/0,049	73,78/0,042	73,78/0,042	73.81/0.025
	ξ (%)/σ (%)	0,62/0,05	0,62/0,05	0,62/0,05	0.62/0.01
04	*f*(Hz)/σ(Hz)	136,13/0,027	136,13/0,027	136,13/0,027	136.15/0.015

	ξ (%)/σ (%)	0,37/0,05	0,37/0,05	0,37/0,05	0.31/0.03
05	*f*(Hz)/σ(Hz)	204,00/6,493	199,38/1,882	198,75/0,137	199.70/0.183
	ξ (%)/σ (%)	2,24/0,94	2,27/0,86	1,71/0,70	2.42/0.15

4.2 Railway viaduct

Located on the southeast high speed track in France, at the kilometric point 075+317, the tested viaduct (Fig. 8) crosses the secondary D939 road, between the towns of Sens and Soucy. Since it was built in the eighties, the increases in the operational speed of high-speed trains (TGV) have moved the excitation frequency closer to the first natural frequency of the structure. Hence, in order to prevent resonance-induced damages, the viaduct passed by a structural intervention, which consisted in tightening a system of rods near the bearings, by a torque wrench. Thus, additional stiffness and, therefore, an increase in the natural frequencies would be provided. Fig. 9 shows the instrumentation plan of the structure.

Figure 8. PK 075+317 Viaduct on the Paris-Lyon high speed train track

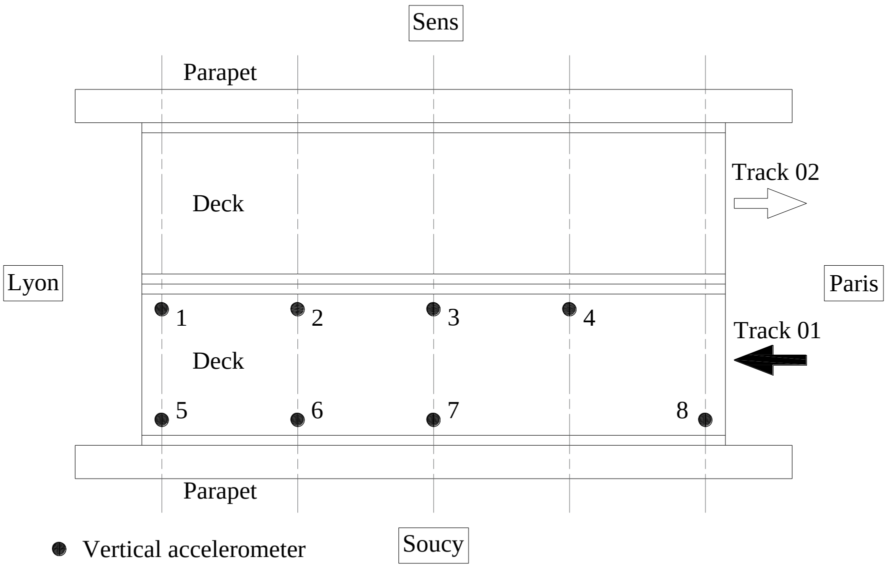


Figure 9. Plane view of the viaduct pointing sensors location

During approximately 1.6 seconds, the eight installed accelerometers registered, at 4096 Hz, the signal matrix S1 shown in Fig. 10. It corresponds to a forced vibration period while the train was passing over the structure.

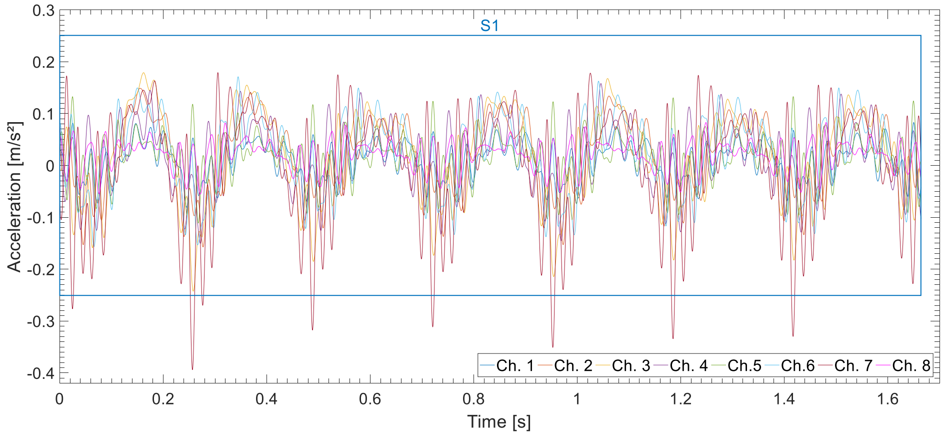


Figure 10. Signal acquired during the transit of the train

Again, the reference interpreter and the proposed methodology were applied to the signal. In this study, the number of desired modes is set to four. However, in a conservative way, both methods were asked to provide six modes. Additionally, a maximum frequency of 50 Hz was set, due to prior knowledge of the structure's dynamic behavior.

Figure 11 depicts the results obtained by the reference interpreter. It is clear that this method was not able to identify the second mode (8 Hz). Three attempts were made varying d_{lim}: 0.01, 0.05 and 0.10. One can note the "gathering" phenomenon happening again as the distance threshold increases, from left to right.

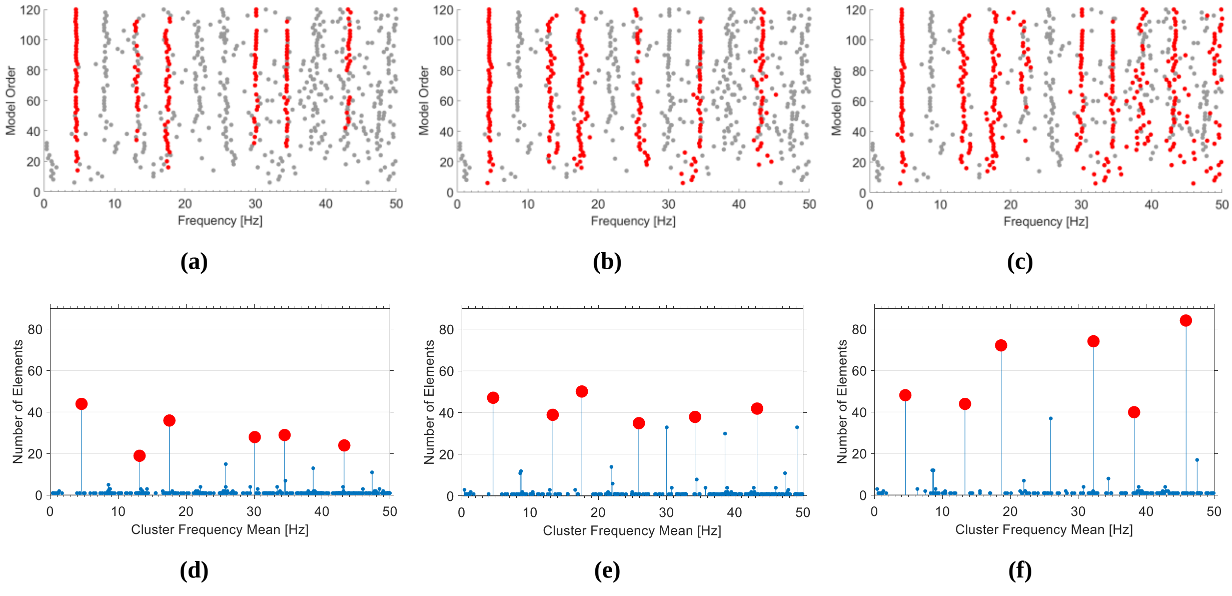
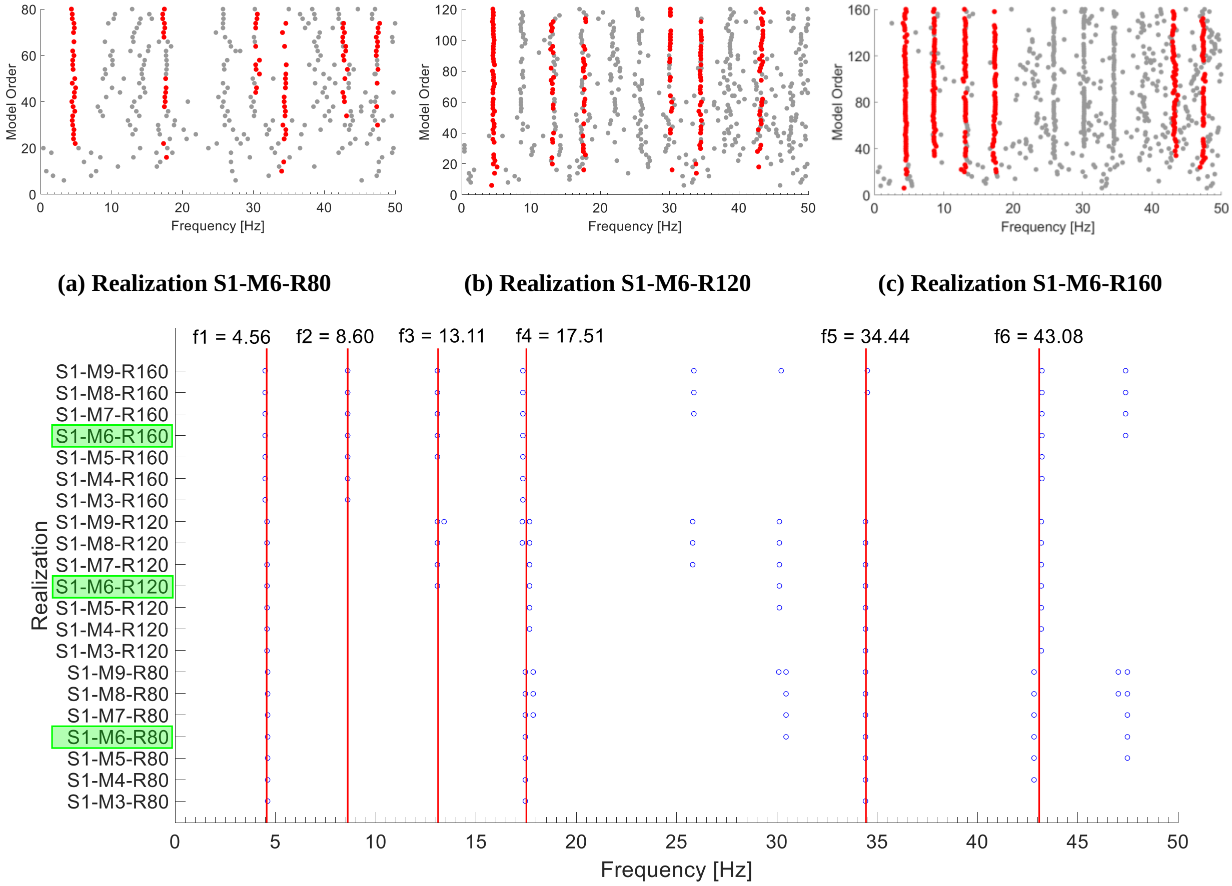


Figure 11. Modes estimates interpretation by reference interpreter, with d_{lim} varying: 0.01 (a, d) , 0.05 (b, e) and 0.10 (c, f)

For the proposed method, however, all four modes were identified correctly. In Figs. 17a, 17b and 17c, three realizations are shown in detail. All 21 realizations and the final result can be checked in the CSD presented in Fig. 12d.



(a) Realization S1-M6-R80 **(b) Realization S1-M6-R120** **(c) Realization S1-M6-R160**

(d) Clustered Stabilization Diagram of analyzed signal

Figure 12. Different realizations results (a, b, c) and CSD of analyzed signal (d)

Fig. 13 shows the modal shapes evaluated by the proposed methodology. The measured modal amplitudes are marked as dots, whereas interpolated points or supports are represented as hollow squares. The blue line represents modal amplitudes measured by sensors 1, 2, 3 and 4 whereas the red line stands for accelerometers 5, 6, 7 and 8 (see Fig. 9).

Finally, Table 2 gathers the results for the modal identification. One can note that the reference interpreter was unable to identify the second mode regardless of the value of d_{lim}. Once again, the proposed method has proved to be more robust and accurate when compared to the reference approach.

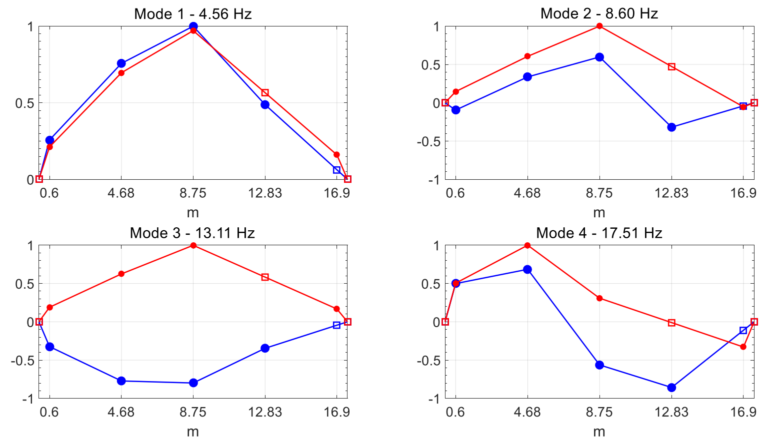


Figure 13. Four identified modal shapes: flexional, flexional-torsional, torsional, and flexional-torsional, respectively

Table 2. Results of the automatic modal identification

Mode	**Parameter**	**Reference Interpreter R120**			**Proposed**
		$d_{lim} = 0.01$	$d_{lim} = 0.05$	$d_{lim} = 0.10$	
01	f(Hz)/σ(Hz)	4.579/0.085	4.586/0.119	4.572/0.151	4.563/0.056
	ξ (%)/σ (%)	2.23/1.07	2.34/1.20	2.36/1.19	3.59/1.59
	MAC_{min}	0.9981	0.9937	0.9937	0.9999
02	ξ (%)/σ (%)				8.603/0.000
	f(Hz)/σ(Hz)	missing	missing	missing	1.19/0.00
	MAC_{min}				1.0000
03	f(Hz)/σ(Hz)	13.156/0.168	13.344/0.384	13.311/0.392	13.108/0.108
	ξ (%)/σ (%)	4.25/0.99	3.58/1.82	3.64/1.84	2.78/1.15
	MAC_{min}	0.9770	0.8787	0.5808	0.9937
04	ξ (%)/σ (%)	17.520/0.208	17.595/0.387	18.653/1.948	17.511/0.178
	f(Hz)/σ(Hz)	1.98/0.99	2.41/1.78	3.30/2.72	1.72/0.81
	MAC_{min}	0.9131	0.7323	0.5990	0.9778

5 CONCLUSIONS

This paper presented a novel methodology that uses a parametric time-domain modal identification algorithm and hierarchical clustering techniques to perform automated modal identification. The developed routines used SSI-DATA algorithm to generate modes estimates for several model orders, followed by two steps of clustering analysis. The first one is the clustering analysis of modes estimates coming from a single stabilization diagram, outputting a given number of modes with a high chance of being physical. After many different stabilization diagrams are interpreted, the second step of clustering analysis is carried out to finally select the modes definitely considered physical.

Experimental applications presented in sections 4.1 and 4.2 showed that the enhanced diagram interpreter provided a more robust tool to eliminate spurious modes. In both applications, all physical modes were correctly identified. The laboratory experiment provided a structure with natural frequencies ranging from 9 to 200 Hz. In contrast to what happened with the reference method, it was shown that the proposed method is perfectly suitable to handle either low or high values of natural frequencies. Besides, the railway viaduct test allowed concluding that the developed routines are reliable to be applied to real dynamic tests.

The proposed methodology represents, therefore, a useful tool to modal identification of civil engineering structures subjected to ambient vibrations (Operational Modal Analysis). In this context, the present work offers a suitable method capable of performing, satisfactorily, automated interpretation of modal estimates, which can be obtained by any parametric time-domain identification algorithm.

ACKNOWLEDGMENTS

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) and UFJF (Universidade Federal de Juiz de Fora) for the financial support, COPPE/UFRJ and IFSTTAR (former LCPC – Laboratoire Central des Ponts et Chaussées) and SNCF (Société Nationale des Chemins de fer Français) for the data used in this paper (project 01V0527 RGCU “Evaluation dynamique des ponts”).

REFERENCES

Alves, V., Cury, A., Roitman, N., Magluta, C. & Cremona, C., 2015. *Novelty detection for SHM using raw acceleration measurements*. Structural Control & Health Monitoring, v. 22, p. 1193-1207.

Battista, R., & Pfeil, M., 2000. Reduction of vortex-induced oscillations of Rio-Niterói Bridge by dynamic control devices. Journal of Wind Engineering and Industrial Aerodynamics, vol. 84, n. 3, pp. 273–288.

Cabboi, A., Magalhães, F., Gentile, C. & Cunha, A.,2017. *Automated modal identification and tracking: Application to an iron arch bridge.* Structural Control & Health Monitoring, v. 24 (1).

Cardoso, R., 2015. *Development of methodology for automatic modal identification of structures*. MSc dissertation by the Federal University of Ouro Preto, Brazil.

Ewins, D., 2000. *Modal Testing: Theory, Practice and Application.* Research Studies Press Ltd., USA.

Gautier, Y., Moretti, O., & Cremona, C., 2005. Experimental modal analysis of the Millau Bridge. Proceedings of Experimental Vibration Analysis for Civil Engineering Structures.

Juang, J. & Pappa, R., 1985. *An eigensystem realization algorithm for modal parameter-identification and model reduction.*. Journal of Guidance Control and Dynamics, 8(5): 620-627.

Magalhães, F., Cunha, A. & Caetano E., 2009. Online automatic identification of the modal parameters of a long span arch bridge. Mech Syst Signal Pr 23(2): 316-329.

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

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

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

Santos, J., 2014. *Smart structural health monitoring techniques for novelty identification in civil engineering structures*. PhD thesis by the University of Lisbon, Superior Technical Institute, Portugal.

Ubertini, F., Gentile, C. & Materazzi, A., 2013. *Automated modal identification in operational conditions and its application to bridges.* Engineering Structures 46: 264-278

Vacher, P., Jacqier, B. & Bucharles, A., 2010. Extensions of the MAC criterion to complex modes. International Conference on Noise and Vibration Engineering, ISMA 2010.

Van Overschee, P. & De Moor, B., 1996. Subspace identification for linear systems: theory, implementation and applications. Kluwer Academic Publishers.

Yun, G., Lee, S. & Shang, S., 2012. An improved mode accuracy indicator for eigensystem realization analysis (ERA) techniques. *KSCE Journal of Civil Engineering*, 16(3): 377-387.