Engineering Structures 178 (2019) 245–257

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Interpretation of stabilization diagrams using density-based clustering algorithm

Ruben L. Boroschek*, Joaquin A. Bilbao

Dept. of Civil Engineering, University of Chile, Chile

ARTICLE INFO

Keywords:
System identification
Stability diagrams
Clustering
OPTICS

ABSTRACT

The estimation of modal parameters is a critical requirement in structural health monitoring, damage detection, design validation, among other topics. The most prevalent methodology for manual identification is via an interpretation of a stabilization diagram. A density-based algorithm for automatically interpreting this type of diagram is proposed. The method employs three stages of interpretation. First, hard criteria are used to discard distinct spurious modes. Second, a density-based algorithm, Ordering Points to Identify the Clustering Structure (OPTICS), is used to cluster data. Finally, the modal parameters are selected taking into account the density distribution of the clustered values. Automation on the procedure is proposed, tested and applied to the vibration measurements of a building structure that has been continuously monitored since 2009. The results indicate a satisfactory interpretation, despite the low signal-to-noise ratios, the effect of induced electric noise, the low density of the sensors, different ambient conditions, and the occurrence of earthquake events.

1. Introduction

The estimation of the modal parameters of civil infrastructure is critical in structural health monitoring, damage detection, design validation, among other topics. Large quantities of data and the identification of basic modal parameters need to be processed in short periods of time. The process of modal parameter estimation can be divided into the following three steps: data acquisition, cleansing and preprocessing, model parameter identification and the selection of a representative set of structural modal parameters. The first two processes usually need to be tuned once and can be automatically executed, whereas the last process requires a considerable amount of user interaction, which is the main focus of research concerning automation for identification techniques.

This paper focuses on parametric system identification for civil engineering structures. The technique is dependent on the parameter estimation of a viscoelastic model with an unknown number of participating modes. The process is usually iterative, in which several models of a different order or size are identified. The model order is an integer value that is equivalent to the number of eigenvalues in the model. Specifying a low model order can produce an inaccurate data fit, whereas a higher model order can yield spurious modal parameters. In general, the order is overestimated [1]; thus, a procedure is required to discriminate between real structural modes and spurious modes. The most common procedure is the construction of a diagram in which the real modes can be identified. This diagram is referred to as a stabilization diagram.

A stabilization diagram can be defined as a model order vs. a frequency plot (and generally includes a mode shape index and damping) for an extensive range of model orders. The main hypothesis, which is obtained from the empirical observation of numerous modal identification problems, is that the physical modes constantly appear with similar frequencies, along with damping and mode shapes as the model order sequentially increases. Thus, the objective of the stabilization diagram is to identify the stable parameters that represent the physical modes. Note that other situations exist in which stable parameters appear in the stabilization diagram, such as machine vibrations or electrical signals that are not associated with the study structure, and this has to be identified as spurious information.

Depending on the quality of the data, the number of sensors and the efficiency of the identification algorithm, the stable physical parameters form a column in the stability diagram. In the manual case, the definition or identification of a stable column is dependent on the judgment of the analyst.

The majority of the strategies in the literature aim to automate the same process that would be followed by a manual analysis of the stabilization diagram. In these methods, the process can be divided into three stages:

(1) Stage 1: Detect and remove as many spurious modes as possible from the stabilization diagram. The removal is generally based on the basic physical characteristics of viscoelastic and proportionally damped structures.
(2) Stage 2: Define sets of similar possible physical modes to define the stable columns and delete possible sets of spurious modes.
(3) Stage 3: Select a representative mode for each set of physical modes (stable columns).

* Corresponding author.
E-mail addresses: rborosch@ing.uchile.cl (R.L. Boroschek), jbilbao@ing.uchile.cl (J.A. Bilbao).

https://doi.org/10.1016/j.engstruct.2018.09.091
Received 14 December 2017; Received in revised form 25 April 2018; Accepted 30 September 2018
Available online 22 October 2018

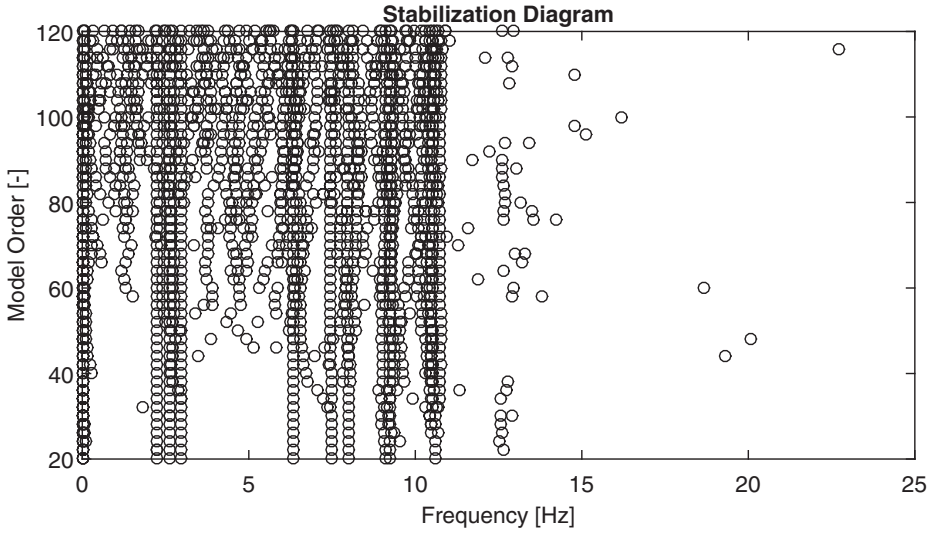

Fig. 2.1. Stabilization diagram, TC record.

Several methodologies for the automatic or semi-automatic identification of stability diagrams have been proposed Bakir et al. [2], Cabboi et al. [3], Magalhaes et al [1], Reynders et al. [4]. For example, in stage one, Magalhaes et al [1] manually fixes thresholds to define variations for the modal properties that belong to a single mode. In stage two, a hierarchical clustering approach is applied. Bakir [2] also fixes thresholds in stage one. However, instead of applying a hierarchical clustering approach in the second stage, he employs an overlapping histogram for the selection of the identified modal properties. Cabboi et al. [3] uses an alternate method of linkage for stage two, which is based on the distance between the parameters and a distance limit that requires adjustment. Stage three is often a direct result of the different methodologies that are proposed by each author and often determines the final parameters from the average or median values. The procedure proposed by Reynders et al. [4] presents a four-stage process. The first stage establishes a series of soft and hard criteria. Based on the results of the application of these criteria, spurious modes are initially selected and deleted from the diagram using a k-means clustering algorithm. In the second stage, sets of similar (close) modes are determined using a hierarchical clustering algorithm, and then spurious modes are removed using k-means clustering, considering the number of elements of each set. In stage three, one mode is selected from each remaining set of physical modes. The final selection is based on the accuracy of the modal parameters, which are estimated and measured in terms of the damping ratio, Modal Phase Collinearity (MPC) or Modal Phase Deviation (MPD).

In this paper, a density-based clustering algorithm based upon an existing algorithm called Ordering Points to Identify the Clustering Structure (OPTICS [5]) is proposed and analyzed. This procedure provides more tools to efficiently analyze the stabilization diagram, and also automation on this algorithm is presented and analyzed. The analysis includes the results yielded by the application of this methodology on a live monitored building. The results provide guidelines and considerations that are aimed to improve the interpretation of stabilization diagrams.

The organization of this paper is as follows. In Section 2, the case study is presented from which the raw stabilization diagram for a single record is exposed to be used for the analysis. Section 3 exposes the proposed methodology and analyzes it using the results of the single record provided by the case study. Section 4 presents the results obtained by applying the algorithm on a given month (2282 records). Finally, Section 5 concludes the paper.

2. Introduction and raw stabilization diagram of a single record

To illustrate the proposed methodology, an automatic interpretation method is applied to the “Torre Central (TC)” Building, which is located in the Faculty of Physical and Mathematical Sciences of the University of Chile in Santiago, Chile. The structure has nine floors and two basements, with an approximate height of 30.2 [m] above ground level. The thickness of each of the shear walls is 35 [cm], and the thickness of each of the slabs is 25 [cm]. Eight accelerometers are installed on the structure, two accelerometers are installed on the basement, and three accelerometers are installed on the third and eighth floors. Additionally, three humidity sensors have been installed at 20, 10, and 5 m below the surface, connected to the accelerometer data acquisition system. The monitoring system also gathers information from a meteorological station 40 m distant from the central tower, which collects data from temperature, precipitation, and wind speed, among others [6]. We have tested our proposal in an analytical model and it yielded good results (Bilbao and Boroschek [7]), which are not described here.

The records cover 15 min segments of continuous measurements from June 2009 to July 2014 with a sampling frequency rate that ranges from 100 to 200 [Hz]. The channels located in the basement were omitted from the analysis given that they would most likely represent the behaviors of the soil and foundation compared with the structural vibration and its modes. Consequently, only six channels were considered.

The identification of the structure (construction of the stabilization diagrams) is made using SSICOV [8] with the implementation of a fast multi-order computation algorithm of the system matrices presented by Döhler et al. [9].

Fig. 2.1 shows the stabilization diagram for a single record (09-26-2009 01:00–01:15). Several spurious modes exist.

3. Proposed methodology

3.1. Stage one

The first stage of the algorithm is aimed at detecting and removing as many spurious modes as possible from the stabilization diagram. To discriminate between spurious and physical modes, different validation criteria are used. These criteria can be categorized as hard validation criteria, which yield binary answers, and soft validation criteria, which yield values within a certain range.

Only hard validation criteria with relatively loose limits are proposed to be employed as a first stage. In this manner, a more intuitive approach to the cleaning of the stabilization diagram, which is easily adjusted to the context of the identification process, can be developed. Note that soft validation criteria can be transformed into hard validation criteria by imposing thresholds on their values.

The hard validation criteria considered on the methodology are

Table 3.1
Hard validation criteria summary.

Hard Validation Criteria
$d(f) \leq 5[\%]$
$d(\beta) \leq 20[\%]$
$1-MAC_d \leq 5[\%]$
$1-MPC \leq 50[\%]$
$MPD \leq 50[\%]$
$MCF \leq 50[\%]$
$\min(f) = 0[Hz]$
$\max(f) = 25[Hz]$
$\min(\beta) = 0[\%]$
$\max(\beta) = 20[\%]$
$presenceofconj(\lambda)$
$presenceofconj(\phi)$

presented in Table 3.1. It is noted that the criteria that can be used is not limited to those exposed in Table 3.1. Moreover, the limits should be evaluated beforehand, following the judgement by the analyst in order to provide optimal results.

3.1.1. Stage one validation criteria

3.1.1.1. Distance criteria. Distance validation criteria are aimed at determining the degree of "stability" by measuring the minimum normalized relative distance between the mode in the analysis and the modes in the nearest model order. This distance can be determined at a model order k and a mode i, by either the eigenfrequencies f_i^k, the damping ratios β_i^k, the continuous-time eigenvalue λ_{ci}^k or the mode shape $\boldsymbol{\phi}_i^k$.

Note that (theoretically) the continuous-time eigenvalue is a function of the eigenfrequency and damping ratio [10]:

$$\lambda_{ci} = -\beta(2\pi f_i) \pm j(2\pi f_i)\sqrt{1-\beta_i^2} \tag{3.1}$$

The distance criteria in terms of eigenfrequencies $d(f_i^k)$, damping ratios $d(\beta_i^k)$, and the continuous-time eigenvalue $d(\lambda_{ci}^k)$ can be generally defined as:

$$d(\chi_i^k) = min\left(\frac{|\chi_i^k - \chi_j^{k+1}|}{max(|\chi_i^k|, |\chi_j^{k+1}|)}\right) \tag{3.2}$$

where χ_i^k represents either f_i^k, β_i^k or λ_{ci}^k. This parameter corresponds to the minimum of the (normalized) distances between the analyzed mode χ_i^k and every mode found in the next model order χ_j^{k+1}. A distance value near 0 would yield a larger probability that the analyzed mode is part of a stable column, whereas a high distance value may indicate a spurious unstable mode or another mode. Note that distance criteria on the eigenvalue was not considered due to its redundancy on frequency and damping.

The distance criteria determined by the mode shape $\boldsymbol{\phi}_i^k$ are obtained using the modal assurance criteria (MAC, [11]) value:

$$MAC(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j) = \frac{|\boldsymbol{\phi}_i^* \boldsymbol{\phi}_j|^2}{||\boldsymbol{\phi}_i||_2^2 \, ||\boldsymbol{\phi}_j||_2^2} \tag{3.3}$$

where $[\ ]^*$ denotes the conjugate transpose of a vector. The distance criteria as a function of the MAC value of the mode shape $\boldsymbol{\phi}_i^k$ are defined as:

$$MAC_d(\boldsymbol{\phi}_i^k) = max(MAC(\boldsymbol{\phi}_i^k, \boldsymbol{\phi}_j^{k+1})) \tag{3.4}$$

This definition yields the maximum MAC value that can be obtained between the analyzed mode shape $\boldsymbol{\phi}_i^k$ and all of the mode shapes found in the next model order $\boldsymbol{\phi}_j^{k+1}$. A high MAC value presents a larger degree of correlation between the mode shapes.

3.1.1.2. Mode shape complexity criteria. In civil engineering, structures can be considered to be linear and proportionally damped at low amplitude responses, with the exception of structures that are supported by rubber isolators (high-damping rubber or natural rubber) and structures with added energy dissipators. Linear and proportionally damped systems always present real-valued mode shape vectors; however, their evaluation is performed via a general modal analysis (i.e., the eigenvectors of the state matrix), which will present complex vectors. Each element of the vector will have the same phase angle, therefore, they can be normalized by the mean phase to obtain real-valued mode shape vectors.

Considering that most structures can be assumed to be proportionally damped, with the exclusion of some exceptional cases, measuring the complexity of the mode shape, in terms of the monophasic behavior of the mode shape vector, helps to determine if a mode is to be considered as a physical or spurious mode. Two methods for determining this property have been proposed: the modal phase collinearity (MPC) and the mean phase deviation (MPD).

MPC is defined as [12]:

$$MPC(\boldsymbol{\phi}_j) = \left[\frac{2\lambda_1}{\lambda_1 + \lambda_2} - 1\right]^2 \tag{3.5}$$

where λ_1 and λ_2 are the maximum eigenvalue and minimum eigenvalue (respectively) of the covariance matrix $\boldsymbol{S}_{cov}$ between the real part and imaginary part of the mode shape:

$$\boldsymbol{S}_{cov} = \begin{bmatrix} Re(\boldsymbol{\phi}_i)^T Re(\boldsymbol{\phi}_i) & Re(\boldsymbol{\phi}_i)^T Im(\boldsymbol{\phi}_i) \\ Re(\boldsymbol{\phi}_i)^T Im(\boldsymbol{\phi}_i) & Im(\boldsymbol{\phi}_i)^T Im(\boldsymbol{\phi}_i) \end{bmatrix} \tag{3.6}$$

Perfect collinearity in the complex plane (monophasic vector) is achieved if only one eigenvalue differs from zero, in which case MPC would assume a value equal to 1. If the two eigenvalues are equal ($\lambda_1 = \lambda_2$), no collinearity (extreme case) exists; thus, the MPC would assume a value equal to 0.

The MPD is defined as [4]:

$$MPD(\boldsymbol{\phi}_j) = \frac{\sum_{o=1}^{n} w_o \arccos\left|\frac{Re(\phi_{j_o})V_{22} - Im(\phi_{j_o})V_{12}}{\sqrt{V_{12}^2 + V_{22}^2}\ |\phi_{j_o}|}\right|}{\sum_{o=1}^{n} w_o}, \ \sum_{o=1}^{n} w_o \neq 0 \tag{3.7}$$

The subscript o is used to define each element of the mode shape vector, w_o represents weighting factors that are equal to $|\phi_{j_o}|$ to provide a higher weight to higher amplitudes. V_{22}, V_{12} are the (2,2) and (1,2) elements of the V matrix, which are determined from the singular value decomposition of the matrix that is formed by the real and imaginary parts of the mode shape vector:

$$\boldsymbol{USV}^t = [Re(\boldsymbol{\phi}_j) \ \ Im(\boldsymbol{\phi}_j)] \tag{3.8}$$

As its name implies, the MPD measures the weighted average of the deviation of each element from the mean phase. Thus, a low value (limited to 0) would represent no deviation from the mean phase (monophasic behavior), whereas a higher value (limited to 45°) represents the degree of variation of the phase for each element. This parameter is normalized by 45° to obtain values in the range [0, 1].

The average of both of these criteria can also be used (known as the modal complexity factor (MCF) [3]):

$$MCF(\boldsymbol{\phi}_i) = \frac{1 - MPC(\boldsymbol{\phi}_i) + MPD(\boldsymbol{\phi}_i)}{2} \tag{3.9}$$

Note that this expression considers MPD to be normalized.

3.1.1.3. General hard criteria. Hard validation criteria are binary values that represent the compliance of a given test. The most commonly used criteria include the following, which are used in the proposed methodology:

(1) *Presence of negative or high damping ratios*: In general, structures are always stable; therefore, negative damping values should rarely be encountered in practice. A hard criterion is defined by eliminating all modes with negative damping. Given that high damping ratios

are unlikely, defining an upper limit for the damping ratio (usually between 15[%] and 20 [%] if it is not an isolated or artificially damped structure) is also applied to remove spurious modes.

(2) *Presence of conjugate pairs:* In theory, for every continuous-time eigenvalue λ_{c_i} and mode shape $\boldsymbol{\phi}_i$, a second mode with conjugate properties should also be present. Thus, the lack of complex conjugate pairs for λ_{c_i} and $\boldsymbol{\phi}_i$ indicates that the mode is spurious.

(3) *Frequency limit:* When determining the modal properties of a structure, high-frequency modes may be unnecessary in the analysis of the structure. Therefore, a limit in the frequency can be applied.

The results of the use of the criteria presented are shown in Fig. 3.1. Note that 74% of the "spurious" modes were deleted using these broad hard limits.

3.2. Stage two

Stage two of the interpretation process seeks to define sets of similar (physical) modes (stable columns). This stage can be considered as the core of the identification process. Bakir [2] uses a frequency-based overlapping histogram to define these stable columns, which is the most common tool in manual analyses. Magalhaes et al. [1] and Reynders et al. [4] uses a hierarchical clustering approach to group modes to a cut-off distance, noting that Reynders et al. [4] automatically defines this cut-off distance. Cabboi et al. [3] proposes a grouping algorithm that is based on the distance between elements by sequentially considering only one mode per model order for the definition of each set.

A tool for efficiently analyzing and automating the detection of stable columns based on a density-clustering algorithm is proposed. The selected clustering methodology is OPTICS [5], which is an extension of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [14]. The use of DBSCAN was initially presented by the authors in [7]. The main concept of OPTICS is a reachability plot.

The procedure starts by calculating the weighted distance between all objects in a set. In our case, each object is constructed using the modal frequency and shape. Then the distance that cluster a minimum set of objects is identified for each component, this is called the Core Distance. A procedure is developed to find all the objects that are close together and that can be assigned to physical modes and those that are outliers. This procedure is based on the concept of Reachability Distance and the ordering of the objects. The results are visualized on what is call the Reachability Plot.

A detailed presentation of the various definitions is provided in [13,14]. Note that the following notation has been modified in the context of the current application.

3.2.1. Reachability plot construction

In order to explain the construction of reachability plots, a few definitions must be presented.

Let D be a set of possibly physical modes and let $p \in D$ be an object that represents the mode properties. Each object p will be defined in our case by its frequency and mode shape:

$$p_i = \{f_i, \boldsymbol{\phi}_i\}$$

Let the weighted distance between two objectsp_i, $p_j$be a function of $\{f, \boldsymbol{\phi}\}$ defined as follows:

$$dist(p_i, p_j) = \frac{|f_i - f_j|}{\max(f_i, f_j)} + w(1 - MAC(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)) \tag{3.10}$$

The damping ratio distance is not considered; it will often be misleading due to identification inaccuracies and the high chance of two or more (different) modes with similar damping ratios. The Eigenvalue distance is also discarded given that it is a combination on frequency and damping. A weighted distance is used to be able to consider a different importance of the difference for the frequency or the mode shape. Typically w equals one. Nevertheless, in some cases the effect of the mode shape difference could be reduced by assigning a small value.

Let $|N_\epsilon(p)|$ be the number of objects at a distance ϵ from$p$:

$$|N_\epsilon(p)| = card\{p_i |\ dist(p, p_i) \leq \epsilon\} \tag{3.11}$$

where *card* indicates the number of elements in a set (cardinal). Fig. 3.2 illustrates this definition.

We define the variable *MinObj* as the minimum number of objects in a set to consider the set as a cluster. We can define the variable "Core Distance" (CD) as the distance at which we have enclosed a minimum number of objects *MinObj* inside the set. Note that if *MinObj* is larger than the number of elements in the set D, the CD value will be undefined, i.e., the number of elements is insufficient for creating a cluster with the given minimum number of elements. Therefore, CD is defined as:

$$CD_{MinObj}(p) = \begin{Bmatrix} Undefined \quad if \quad |N_{\epsilon \to \infty}(p)| < MinObj \\ \min_{\epsilon}(|N_\epsilon(p)| \leq MinObj), \quad otherwise. \end{Bmatrix} \tag{3.12}$$

Fig. 3.3 illustrates this definition considering $MinObj = 3$ (case a) and $MinObj = 7$ (case b), where the minimum distance that is required to enclose *MinObj* is shown (i.e., the Core Distance).

Let the reachability distance (RD) of the object p_i with respect to the object p_j be the maximum value between the CD of p_j and the distance between them. Note that if *MinObj* is larger than the number of elements in the set D, this value will be undefined.

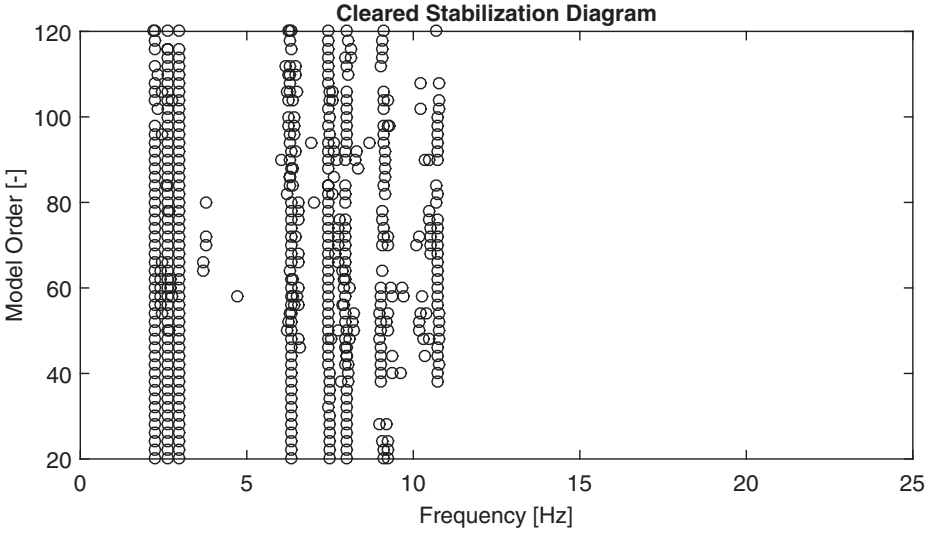

Fig. 3.1. Cleared stabilization diagram (deleted: 74% of the total number of modes), TC record.

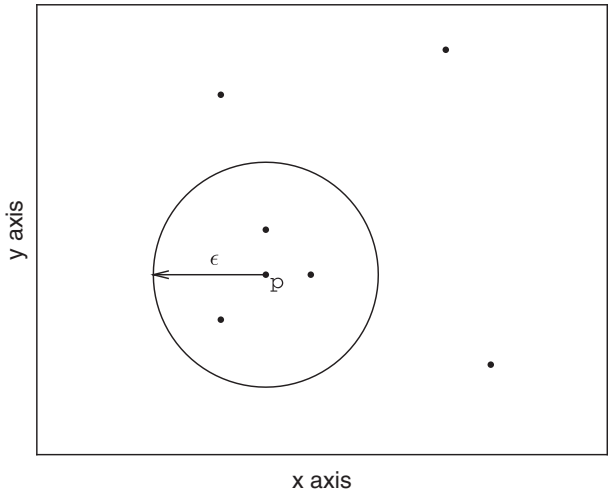

Fig. 3.2. Number of objects at a distance from reference object $\boldsymbol{p}$: $|\boldsymbol{N}(\boldsymbol{p})| = 3$.

$$RD_{MinObj}(p_i, p_j) = \begin{Bmatrix} Undefined \;\; if \; |N_{\epsilon \to \infty}(p)| < MinObj \\ \max(CD_{MinObj}(p_j), dist(p_i, p_j)), \quad otherwise. \end{Bmatrix} \tag{3.13}$$

Fig. 3.3 illustrates this definition, considering the following two cases of interest: (a) where the RD is controlled by the distance between the objects, and (b) where the RD is controlled by the CD.

The reachability plot is defined as a graphical representation of the order obtained for the objects using the reachability distance metric, from which clusters can be identified.

The reachability plot is obtained by evaluating the RD between each object and the rest of the set in a sequential manner. In each iteration, the RD between the selected object and the rest of the set is computed. The closest object (in terms of RD) is identified. A sequential label is assigned to it (thus creating the order of the objects) and it is used as the next object for the computation of RD. The set is reduced in each iteration by deleting the previously analyzed object. The resulting reachability plot is presented in Fig. 3.4.

A few observations on the procedure are indicated below:

(1) The order of the objects is part of the result of the algorithm.
(2) The first object can be arbitrarily selected without hindering the ability to detect the cluster structure.
(3) The RD value is always assigned to the object that had the minimum RD with respect to the analyzed object. Therefore, the first object analyzed does not have an RD value. The solution proposed is to assign 1.1 times the value of the largest RD value obtained for the rest of the objects. Note that this does not alter the results provided by the reachability plot.
(4) The RD values are computed at each iteration for every object left in the set, and it will rewrite the previous RD value as long as it is smaller than the one that the object already had. This means that if at a certain point in the iteration no objects had their RD value rewritten, or the minimum RD value was not defined in the current iteration, then the next object in the analysis is defined by a previous iteration and not by the object in analysis.
(5) The RD assigned to each object, measures how connected it is to the reference object (or a previous object, taking into account the special case explained earlier). This parameter is controlled by the maximum value between the distance between the two objects being analyzed or by the Core Distance (see Eq. (3.12)). From a physical point of view, the distance function measures only the distance between the objects, in terms of frequency and mode shape. Meanwhile the CD measures how far one should go to find a set of *MinObj* in the neighborhood of the object.

Four generic cases can be discussed:

Case (a) If the distance between the two objects is large, but CD is small, this means that the objects are not in the same cluster (see Fig. 3.3(a), with p_j being the current object and p_i the previous object). This information is saved assigning the distance between the objects to the RD value.
Case (b) If the distance between the two objects is small, but CD is large, this means that both objects are probably outliers, (see Fig. 3.3(b)). This information is saved assigning the CD to the RD value. When CD is large. this means there are not too many objects close together.
Case (c) If both the distance between the two objects and the CD are large, this means that the object is not part of the cluster of the previous object. This information is saved assigning the larger value to RD.
Case (d) Finally, if both the distance and CD are small, this probably means that both objects are part of the same cluster. This information is saved assigning the largest of these relatively small values to the RD.

Note however, especially for the last two cases, that the “large” and “small” values are relative to the rest of the objects.

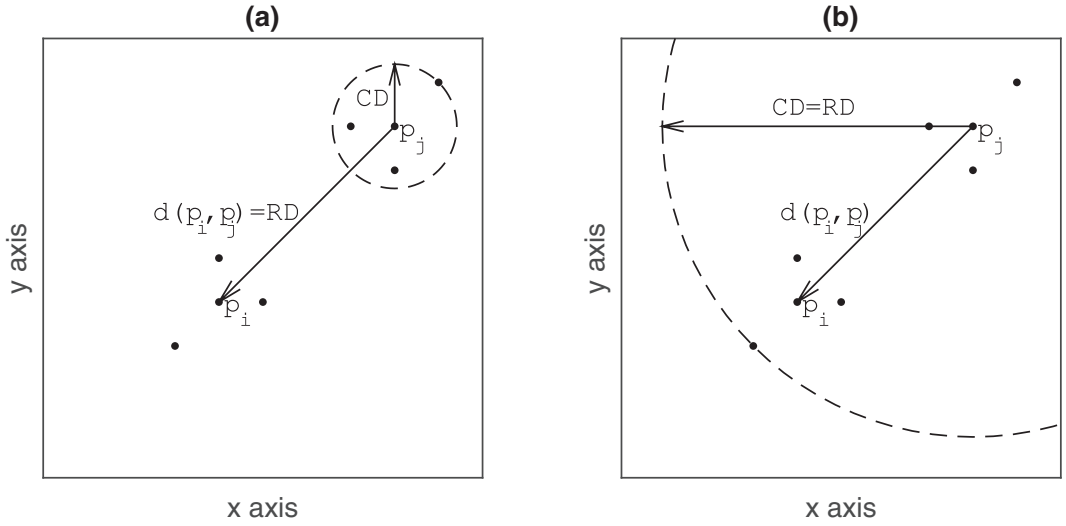

Fig. 3.3. Reachability distance, considering (a) Case for distance between objects larger than CD, and (b) Case for Core Distance larger than distance between all objects considered (***MinObj*** = 7).

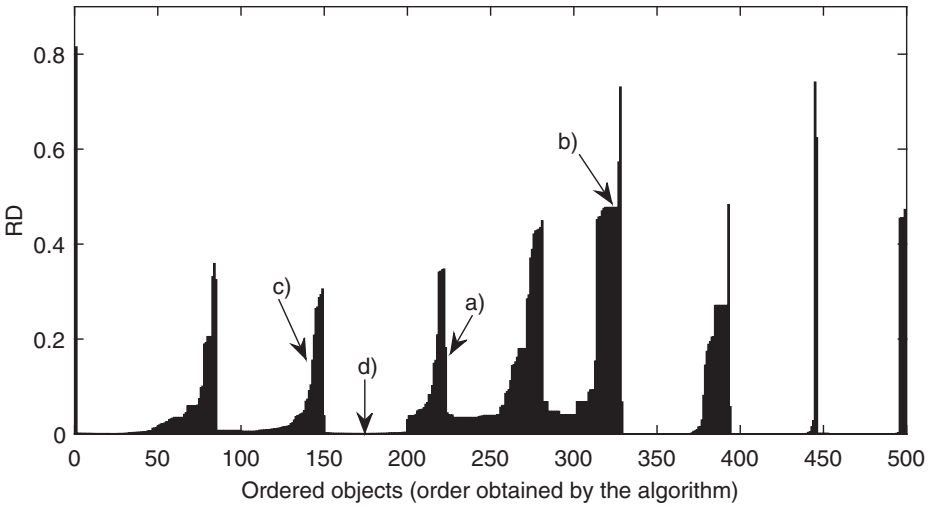

Fig. 3.4. Reachability plot, TC record.

3.2.2. Reachability plot analysis

Now that the reachability plot has been developed (see Fig. 3.4), the clustering structure can be observed and derived from this plot where the low RD regions represent objects that are part of the same cluster, and the high RD regions most likely represent outliers.

In particular, each case discussed in the previous section can be found in the reachability plot: Case a) can be found at the objects with a high RD value just before the start of a low RD value region, Case b) can be found in objects with a high RD value at the middle of a high RD value region, Case c) can be found in objects with a high RD value at the start of a high RD value region, and Case d) can be found in objects with a low RD value at a low RD value region. These cases are illustrated in Fig. 3.4.

As noted in the previous section, “large” and “small” values are relative to the whole set, and by inspection of the reachability plot, these relative values can be observed from the plot.

3.2.3. Reachability plot analysis automation

The final step is the automatic detection of clusters from the reachability plot. To attain automation, we propose the following methodology, motivated by the recommendations in [5].

Considering the observations found in the previous sections, a cluster will be defined as the set of objects starting at the last object of a high RD value region and ending at the last object of a low RD value region. Every object that is not considered as part of a cluster will be considered as spurious.

To identify the clusters from the reachability plot, two concepts will be used: contiguous objects that clearly display a different RD value (positive or negative difference, we will define this difference as slopes) and segments of contiguous objects with similar RD values (valley)

3.2.3.1. Slope between close objects. In order to measure the difference or slope between contiguous objects, the following normalized distance between subsequent RD values is employed:

$$s_{RD}(p_k) = \frac{RD_{k+1} - RD_k}{\max(RD)} \tag{3.14}$$

A high (positive) distance indicates an upward slope of the RD plot, and a high (negative) distance indicates a downward slope, and values near zero indicate nearly constant RD values.

To differentiate valley objects from those in high RD regions with positive and negative slopes, the following definitions are constructed based on a reference value of slope, ξ:

$$\begin{aligned} &UPobject_{\xi}(p_k)\text{: } if\,(s_{RD}(p_k) > \xi) \\ &DOWNobject_{\xi}(p_k)\text{: } if\,(s_{RD}(p_k) < -\xi) \end{aligned} \tag{3.15}$$

In order to maintain automation we propose to define ξ as $\xi = mean(|s_{RD}|)/2$. Fig. 3.5 shows the slope $s_{RD}(p_k)$ values for each object, where the ξ and $-\xi$ thresholds are shown (in red). Given that the thresholds are hard to distinguish, Fig. 3.6 shows the reachability plot, where the UPobjects and DOWNobjects are identified by red and green, respectively. Note that the baseline is change to -0.1 for illustration purposes.

3.2.3.2. Finding regions of changing RD values. Using the previously defined slope of objects, regions of increasing or decreasing RD values are identified as intervals of UP or DOWN objects, such that they maintain the sign of the slope and avoid the inclusion of possible clusters in these areas.

There are two possible situations for a constant positive slope or constant negative slope. A constant positive slope region is defined as follows:

An interval I = [s,e] is to be considered as a constant positive slope region if:

a. s is an $UPobject_{\xi}(s)$.
b. e is an $UPobject_{\xi}(e)$.
c. Each point between $p_k \in [s, e]$ is at least as high as its predecessor:
d. $\forall\, p_k, s < p_k \leq e$: $p_k \geq p_{k-1}$
e. I does not contain more than $MinObj$ consecutive points, which are relatively constant (i.e., are not $UPobject_{\xi}$); otherwise, it can be a cluster
f. I is maximal: $\forall\, J$: $(I \subseteq J, UParea_{\xi}(J) \Rightarrow I = J)$, which states that an constant positive slope region cannot exist inside a similar region.

The constant negative slope region is defined similarly. The UP slope region are displayed in red and the DOWN slope region are displayed in green in Fig. 3.7; note that the baseline is changed to -0.1 for illustration purposes.

3.2.3.3. Cluster detection. After the identification of slope regions, clusters are defined as the set of objects inside a region of low constant RD values limited by regions where the normalized difference between two consecutive RD values is higher than a predefined threshold. The resulting algorithm for cluster detection is defined as follows:

(1) Start from the first object.
(2) If the object is the end of a high negative slope region, start the construction of a new cluster. If not, select the next object.
(3) If the construction has started (step 2), three cases exist for the next object:
   a. If the object is not part of a high slope region (DOWN or UP),

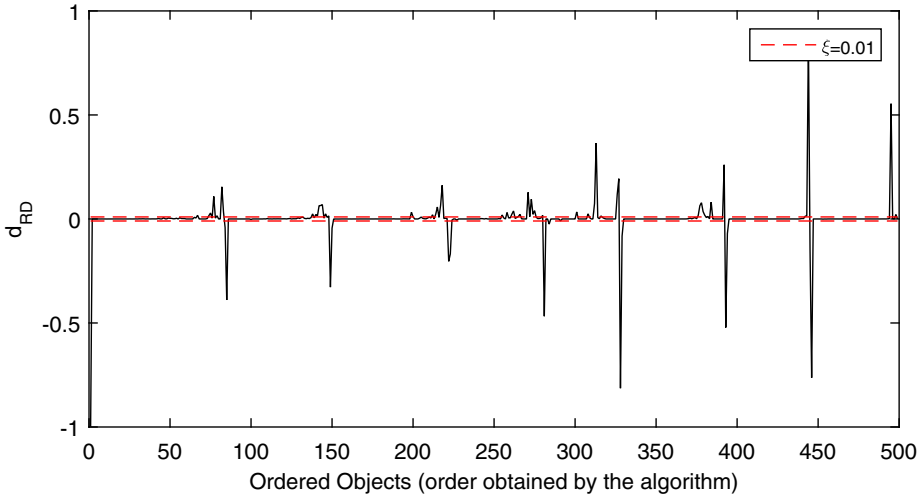

Fig. 3.5. Normalized distance between subsequent RD values, and ξ threshold value, TC record.

then assign it to the cluster in construction.

b. If the object belongs to a high negative slope region, delete the construction and start the formation of a new cluster (i.e., return to step 2).

c. If the object is part of a high positive slope region, end the cluster at this object.

(4) Repeat steps 2 and 3 until the last object is analyzed.

(5) If any cluster has less than *MinObj* objects, the cluster is disregarded.

Note that the last step is introduced for consistency in the algorithm: *MinObj* should be understood as the minimum number of elements in a cluster.

Fig. 3.8 presents the results of the application of the algorithm in terms of the reachability plot, where each cluster is identified by color; note that the baseline is changed to -0.1 for illustration purposes. In Fig. 3.9, the resulting clusters are identified from the stabilization diagram (using the same colors to distinguish the clusters).

3.2.3.4. Parameter automatization. Given the construction of the proposed methodology, the only parameter that needs to be defined is *MinObj*, i.e., the minimum number of elements that define a cluster. To analyze the effect of this value on the reachability plot, we have evaluated the effect of *MinObj* in the RD Plot. We have obtained the reachability plot for different values of *MinObj*, which vary from 1 to 100. Each plot has been grouped in a single 3D plot, as shown in Fig. 3.10. In this figure, the color intensity scale represents the RD value as a function of the ordered objects and *MinObj*. For example, if we consider 1/3 of the total number of model orders, we obtain the red line in Fig. 3.10, which corresponds to the values observed in Fig. 3.4. 2/3 of the total number of model orders corresponds to the purple dashed line in Fig. 3.10 and 1/6 of the total number of model orders corresponds to the green line.

Note that a lower *MinObj* value would cause the RD values to be controlled by the maximum distance between the modes, which loses the meaning of defining a *MinObj* parameter. Whereas a higher *MinObj* value would cause the RD values to be controlled by a higher CD value, which would cause more distant modes to be grouped (which is a direct consequence of requiring a large *MinObj* value).

Given the results obtained from this analysis, it becomes apparent that the reachability plot is not too sensitive to a reasonable value for *MinObj*. Thus, we propose to set *MinObj* equal to 1/3 of the total number of model orders. This means that, in order for a cluster to be considered as such, it should form a stable column covering at least a third of the stabilization diagram. However, this parameter should be defined at the same time as the selection of hard validation criteria to be used, and should be treated as an added tool to the analysis: setting *MinObj* as a low percentage of the total number of model orders will allow the consideration of modes that have low stability (i.e. it will include modes that are relatively less stable), and in the other hand,

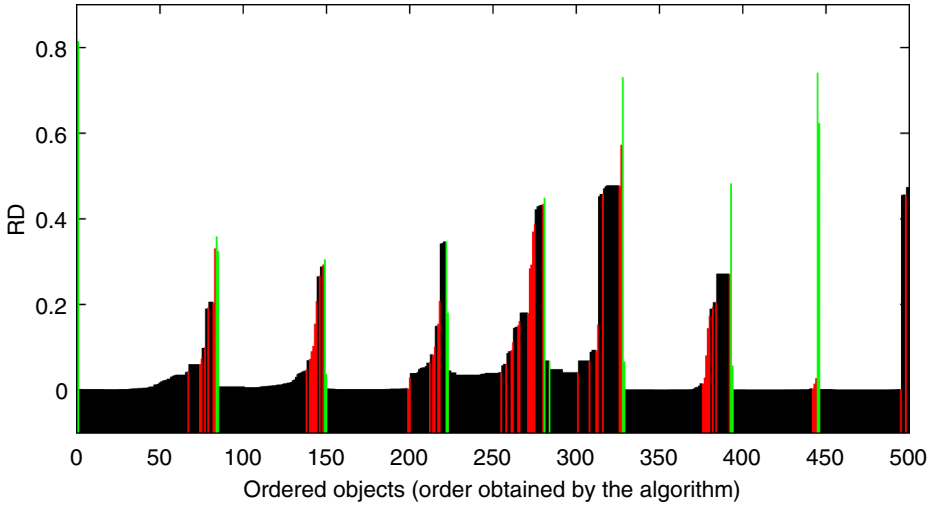

Fig. 3.6. UPobjects and DOWNobjects (red and green, respectively), TC record. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

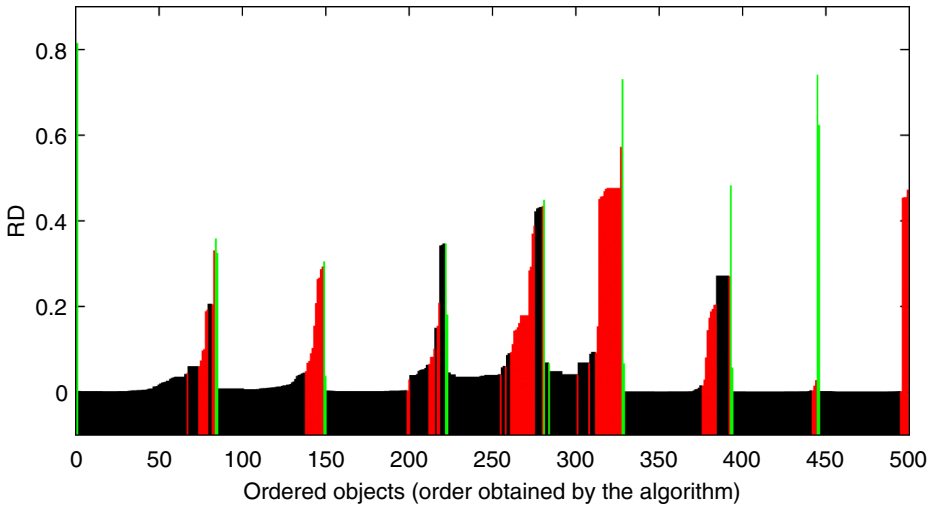

Fig. 3.7. UP and DOWN slope region (red and green, respectively), TC record. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

setting *MinObj* as a high percentage of the total number of model orders will limit the selection of less stablemodes, thus only allowing well defined modes to be considered as a result.

3.3. Stage three

The third and final stage of the proposed methodology is aimed at defining the representative properties of each set of similar modes.

The methodologies found by the different authors can be categorized as follows: computation of the mean frequency, mean mode shape and median damping for each set (Cabboi et al. [3], Magalhaes et al. [1]), and the selection of a single object from the set by detecting the set nearest the median value or the lowest MCF value (Reynders et al. [4]). Note that Bakir et al. [2] does not mention an approach regarding this subject.

Accurate values can be obtained by any of the methodologies that are proposed by the different authors. The only problem that could arise is the selection of an "outlier" object or the inclusion of this object into the computed mean or median values.

To avoid the consideration of an "outlier" object, the proposed methodology aims to assign the densest object in each set as the representative element of the set. The proposed methodology iteratively uses CD to delete objects from the set. In each iteration, the CD is computed using *MinObj* as half of the number of elements of the set, and the objects that have a CD larger than the average CD are removed from the set. This process is repeated until only one object remains, or the iteration stops removing objects, in which case one randomly chosen object is selected. Note that this last consideration ensures the convergence of the algorithm.

Therefore, the proposed methodology for each set is defined as follows:

(1) Use *MinObj* = round (number of elements/2).
(2) Compute the core distance (CD) with respect to *MinObj* and the distance function defined in equation (3.10) for each object.
(3) Every element that has a CD larger than the mean CD value is deleted from the iteration.
(4) Repeat steps 1 through 3 using the remaining elements (thus, decreasing the number of elements for the computation of *MinObj*). If the objects from the previous iteration are the same objects for the current iteration, remove one (any) element, which ensures the end of the iteration.
(5) The last remaining element is the selected element of the cluster.

Fig. 3.11 displays the cluster number that is associated with the mode number and the frequency for all elements in each cluster in our example set. The red crosses in this figure represent the selected frequency values. Fig. 3.12 displays the damping vs. cluster number. As expected, a larger dispersion is observed and the density-based selected value is represented by the red crosses. Fig. 3.13 shows every mode shape (with real and

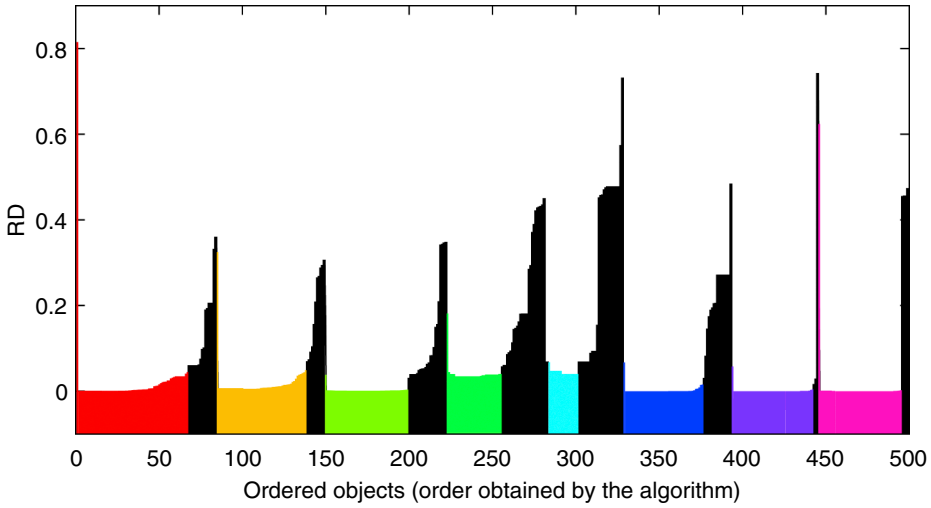

Fig. 3.8. Clusters from the reachability Plot (distinguished by color), TC record. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

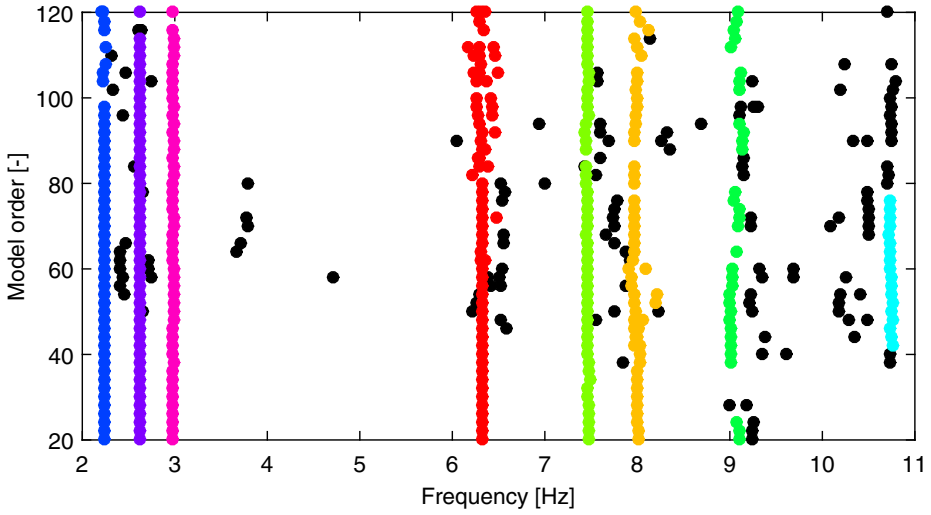

Fig. 3.9. Clusters identified (distinguished by color), TC record. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

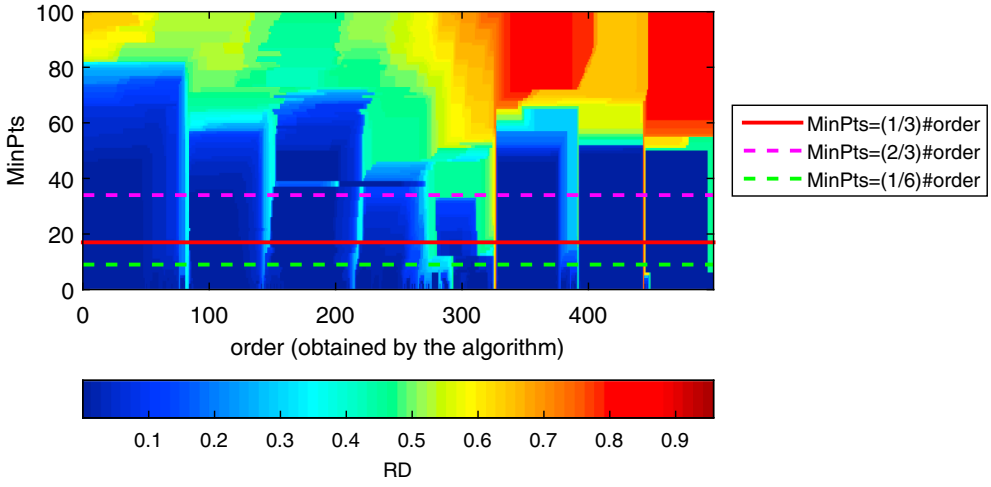

Fig. 3.10. Reachability plots for an extensive range of MinObj, TC record.

imaginary parts) for each cluster (mode number), and the density-based selected mode shape is indicated in red (solid red line represents the real part and a dashed red line represents the imaginary part).

4. Real case application

Fig. 4.1 shows the frequencies identified for August 2009 TC records (a total of 2610 results, each one obtained every 15 min during the entire month) using the proposed methodology. Only ambient records are considered with a PGA limit of less than 0.002 [g] to avoid the incorporation of earthquake events that occurred during the same time period. In this figure, eight frequencies can be observed. Note that a frequency limit of 10 Hz was considered given the low quality of the modes higher than 10 Hz.

Note that in order to identify a mode in a single 15 min record, the mode should be succesfully excited, i.e. it should be present in the stabilization diagram as a stable column. Given that not every mode is necessarily present in every single record, the history of frequencies yields more modes than the single analysis. This explains why only seven modes below 10 Hz were found in the single record analysis, compared with the results obtained from the history of frequencies.

The effect of daily temperature variation is clearly visible in all frequency series, but the two frequencies between 7.5 and 8.0 Hz are more pronounced. Many of the identified spurious frequencies are due to low signal-to-noise ratios, the effect of induced electric noise, the low densities of the sensors and small earthquake events. These non-persistent frequencies can be easily removed using a modal tracking algorithm.

As stated in Section 3.2.1, the MinObj value should be considered as the number of objects that a stable column should have in order to be considered as such. Therefore, MinObj could be set up higher in order to only consider stable columns that cover a higher number of objects in the stabilization diagram. Fig. 4.2 shows the same analysis but using MinObj as 3/4 of the total number of model orders. From this result, and as expected, less outliers are found in the frequency series at the expense of having less objects found near 9 Hz frequency. The reason being that these modes are difficult to detect due to noisy signals and other variables as previously discussed. As a consequence of this problem, the single record analysis only presented one frequency near 9 Hz, given that the other was not clearly present on the stabilization diagram analyzed. If it is desired that this frequency be included, the MinObj selected in figure Fig. 4.1 could be used.

5. Conclusions

A methodology to analyze and automatically interpret stabilization diagrams using a density-based clustering algorithm has been proposed. This new methodology allows an improved manual analysis to be performed based on the use of reachability plots to easily detect clustering structures on the stabilization diagram.

The proposed methodology was presented in detail through its

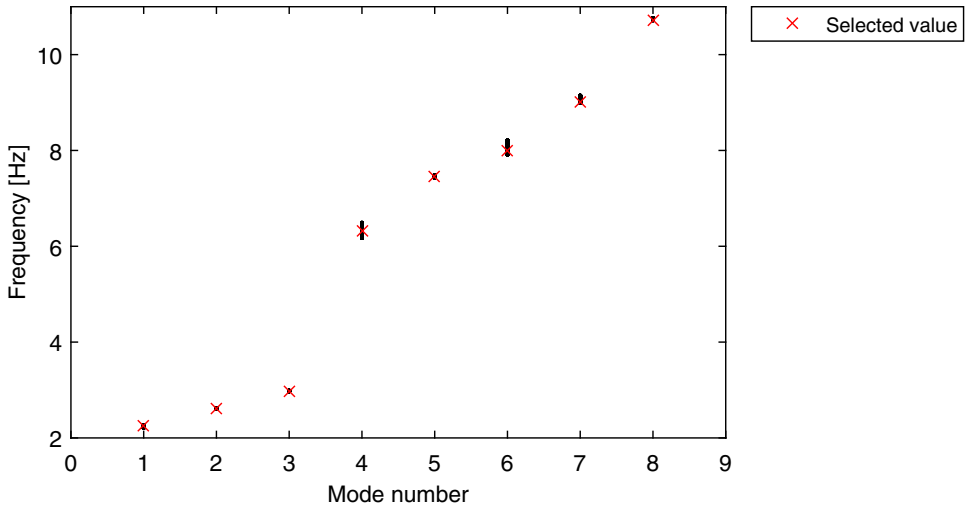

Fig. 3.11. Selected frequencies, TC record.

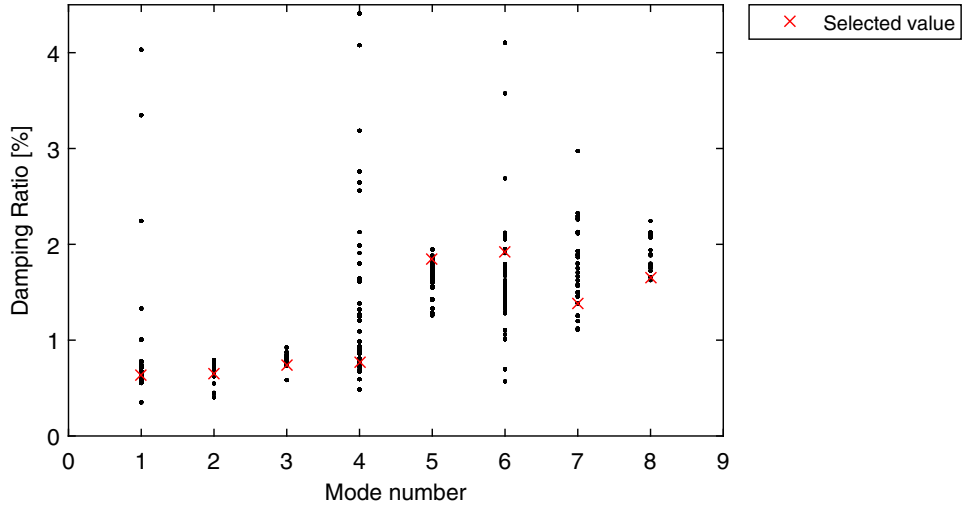

Fig. 3.12. Selected damping ratios, TC record.

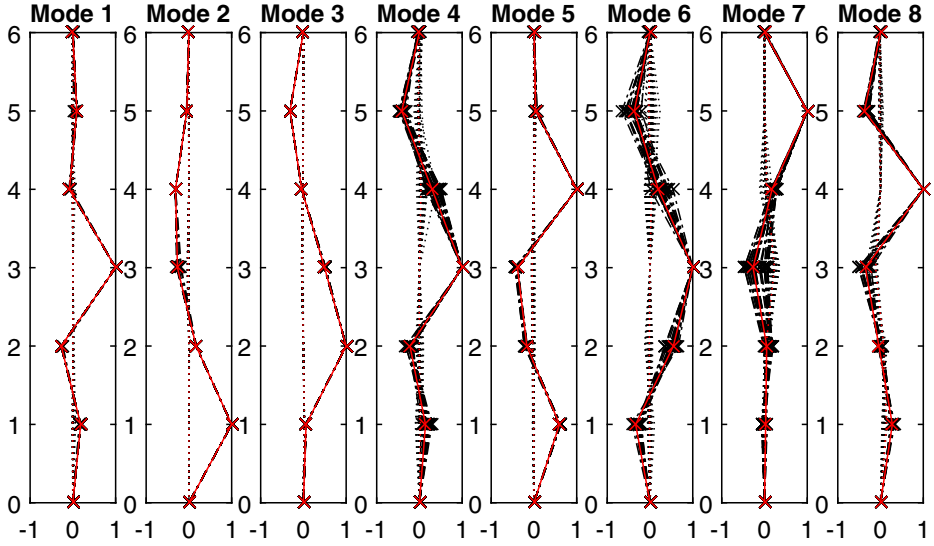

Fig. 3.13. Selected mode shapes, TC record.

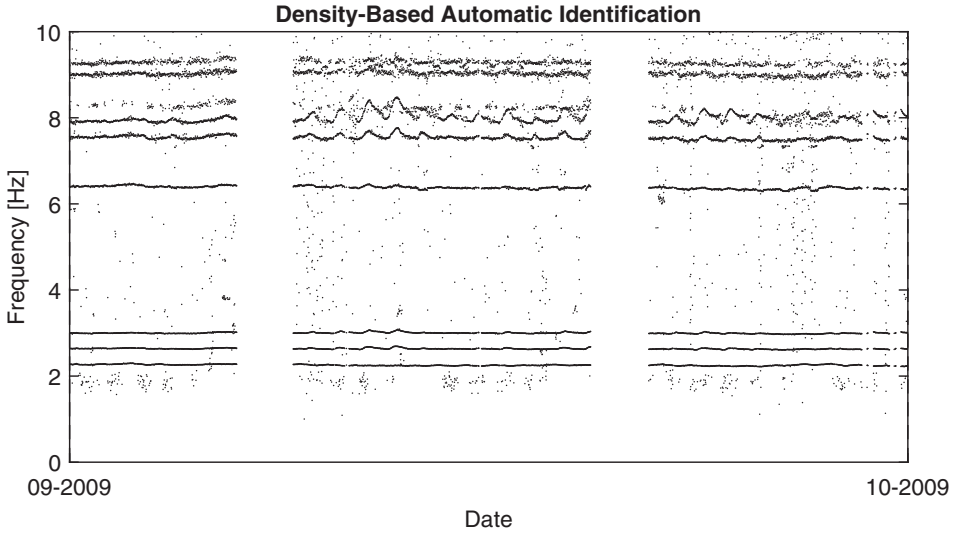

Fig. 4.1. Proposed methodology results, TC records using MinOj = (1/3)#order.

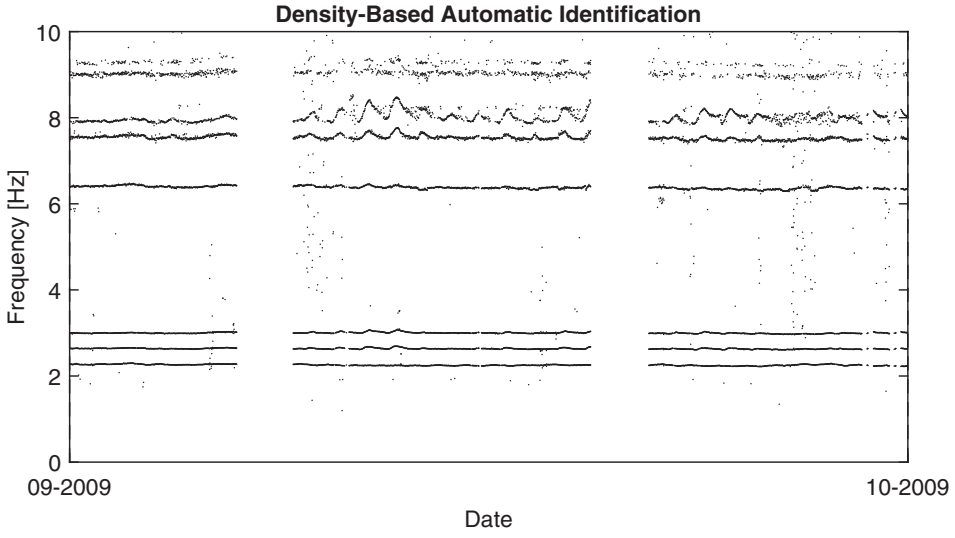

Fig. 4.2. Proposed methodology results, TC records using MinObj = (3/4)#order.

application on a real case study (TC building) for a single 15 min record and was subsequently applied to a month of records. Both cases yielded good results in spite of the presence of spurious modes due to low signal-to-noise ratios, the effect of induced electrical noise, and the low densities of the sensors and small earthquake events.

The density-based methodology consists of three stages. The first stage is a classification that is based on basic physical parameters, of which the most effective parameters are as follows: limits on damping values (greater than 0 and typically lower than 10% for typical structures and 25% for artificially damped structures, nonlinear structures and structures with friction contacts or connections), MPD and 1-MPC lower than 0.50 and the existence of a conjugate pair for frequency and mode. Additional limits on sequential model orders on a stabilization diagram for MAC and frequency substantially contribute to the identification of a stable modal parameter, and a rather low difference value has a strong effect. Thus, we confirm the standard recommended values for mode shape difference $1 - \mathrm{MAC} < 5\%$ and frequency difference of $d(f) < 5\%$. For the second stage, we identify stable columns in the stability diagram using OPTICS as a density-based classification algorithm. In the third and final stage, the selection of a single group of parameters for each mode is selected using a density-based approach again.

The only relevant parameter that needs to be adjusted to attain automation is the minimum number of modal objects in a given cluster, *MinObj*. An analysis was carried out demonstrating that the methodology is not too sensitive on this parameter and can be defined in an intuitive manner by setting it as the desired number of elements that a stable column should have to be considered as a valid mode. In the proposed procedure we generalize its implementation setting the minimum number of objects to 1/3 of the total number of model orders.

The proposed automation of the methodology allows for a simple and reliable procedure that can be set-up (once) by following the judgement of the analyst through the definition of hard validation criteria and the definition of what should be considered as a valid cluster or stable column through the definition of the *MinObj* parameter.

Furthermore, the reachability distance, RD, values obtained from the construction of the reachability plot can provide relevant information regarding the confidence of a given identified mode: a high mean RD value of a cluster would mean that the stable column was not as dense or consistent when compared to a low mean RD value of a cluster.

Acknowledgment

We acknowledge Pedro Soto and the Department of Civil Engineering at the University of Chile for providing the records of the Torre Central Building.

Appendix A

```
function [ O,CD,RD ] = Optics_Algorithm( DM,MinObj)
%   Optics_Algorithm [ O,CD,RD ] = Optics_Algorithm( dM,MinObj)
%   Input:
%       DM: Distance Matrix between points: dM(i,j)=d(pi,pj). (symmetric)
%       MinObj: Minimum number of objects needed to consider a set as a
%       cluster.
%   Output:
%       O: order vector.
%       CD: Core Distance vector.
%       RD: Reachability Distance vector.

% Written by: Joaquín A. Bilbao
% Reviewed by: Rubén L. Boroschek
% June, 2016

%% Initialization:

% empty core distance vector:
CD=zeros(1,size(DM,1));
% large initial values for reachability distance vector:
RD=ones(1,size(DM,1))*10^10;
% empty order "O" vector:
O=zeros(1,size(DM,1));
% label "S" vector, identifying each element in DM:
S=1:size(DM,1);

%% Core Distance:
for i=1:size(DM,1)
    % sorting of the distances for the i-th object:
    D=sort(DM(i,:));
    % Core Distance for the i-th object:
    CD(i)=D(MinObj+1);
end

%% Reachability Distance:

% selection index. Starting with the first object:
ind=1;
% counter employed to assign to order "O" vector:
j=0;
% iteration, ends when there are no more objects to analyze:
while ~isempty(S)
    % add counter:
    j=j+1;
    % get the label associated to selection index (ind):
    label_ind=S(ind);
    % remove the object from vector "S":
    S(ind)=[];
    % assign label to next element of order "O" vector:
    O(j)=label_ind;
    % Reachability distance of the current object with respect to every
    % object left in "S" vector:
    RD_S=max([ones(1,length(S))*CD(label_ind);DM(label_ind,S)]);
    % identify the objects that have a larger RD value assigned:
```

```
        id_larger=(RD(S))>RD_S;
        % replace the value of the identified objects with the new value
        % obtained.
        RD(S(id_larger))=RD_S(id_larger);
        % detect the minimum RD-valued object to be selected for the next
        % iteration. Save index (ind):
        [~,ind]=min(RD(S));
    end
    % assign the first object with an amplified value of the maximum RD value
    % obtained in the process:
    RD(1)=max(RD(2:end))+.1*max(RD(2:end));

    end
```

References

[1] Magalhaes F, Cunha A, Caetano E. Online automatic identification of the modal parameters of a long span bridge. Mech Syst Sig Process 2009:316–29.

[2] Bakir PG. Automation of the stabilization diagrams for subspace based system identification. Expert Syst Appl 2011:14390–7.

[3] Cabboi A, Magalhaes F, Gentile C, Cunha A. Automatic operational modal analysis: challenges and practical application to a historical bridge. 6th ECCOMAS conference on smart structures and materials, Torino. 2013.

[4] Reynders E, Houbrechts J, De Roeck G. Fully automated (operational) modal analysis. Mech Syst Sig Process 2012;29:228–50.

[5] Ankerst M, Markus B, Kriegel MH-P, Sander J. OPTICS: ordering points to identify the clustering structure. Proc. ACM SIGMOD'99 Int. Conf. on Management of Data, Philadelphia PA. 1999.

[6] Boroschek R. Structural health monitoring performance during the 2010 Gigantic Chile earthquake. Earthquakes and health monitoring of civil structures. Springer; 2010. p. 197–216.

[7] Boroschek R, Bilbao J. Evaluation of an automatic selection methodology of model parameters from stability diagrams on a damaged building. 6th International Operational Modal Analysis Conferenceq, Gijón, Spain. 2015.

[8] Peeters B, De Roeck G. Reference-base stochastic subspace identfication for output-only modal analysis. Mech Syst Signal Process 1999;13:855–78.

[9] Döhler M, Mevel L. Fast multi-order computation of system matrices in subspace-based system identification. Control Eng Pract 2012;20:882–94.

[10] Veletsos AS, Ventura CE. Modal analysis of non-classically damped linear systems. Earthq Eng Struct Dynam 1986;14:217–43.

[11] Allemang R, Brown D. A correlation coefficient for modal vector analysis. Proceedings of the 1st International Modal Analysis Conference. 1982. p. 110–6.

[12] Pappa R, Elliott K, Schenk A. A consistent-mode indicator for the eigensystem realization algorithm Report NASA TM-107067 National Aeronautics and Space Administration; 1992

[13] Ankerst M, Breunig MM, Kriegel H-P, Sander J. OPTICS: ordering points to identify the clustering structure. CM SIGMOD international conference on Management of data, NY, USA. 1999.

[14] Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd international conference on knowledge discovery and data mining. 1996. p. 226–331.