CRANFIELD UNIVERSITY

A MACHINE LEARNING APPROACH FOR AUTOMATIC
OPERATIONAL MODAL ANALYSIS

VEZIO MUGNAINI

Politecnico di Torino
Dipartimento di Ingegneria Strutturale, Edile e Geotecnica

Laurea Magistrale in Ingegneria Civile
Academic Year: 2018 - 2019

Supervisor: Ceravolo Rosario
Associate Supervisor: Zanotti Fragonara Luca

October 2019

CRANFIELD UNIVERSITY

Politecnico di Torino
Dipartimento di Ingegneria Strutturale, Edile e Geotecnica

Laurea Magistrale in Ingnegneria Civile

Academic Year 2018 - 2019

Vezio Mugnaini

A MACHINE LEARNING APPROACH FOR AUTOMATIC
OPERATIONAL MODAL ANALYSIS

Supervisor: Ceravolo Rosario
Associate Supervisor: Zanotti Fragonara Luca
October 2019

ABSTRACT

Structural Health Monitoring (SHM) has been one of the main research topics in the area of civil, mechanical and aerospace engineering for the past few years. Modal parameters and their evolution in time can be used as features and indicators of the damage which a structure is subjected to and might also allow for a prediction of the residual useful life of the same. SHM needs of long time-series of data to be efficient. For this reason, output-only techniques of system identification, which can record data continuously and without the constant supervision of an operator, are particularly suitable for this aim and are typically referred to as Operational Modal Analysis (OMA).

In the output-only techniques, the knowledge of the input is replaced by the assumption that the input is a realisation of a stochastic process (white noise). The determination of a model that fits the measured data is named stochastic system identification. Despite the existence of a large number of OMA algorithms developed during the last decades, this work is exclusively focused on the Covariance-driven Stochastic Subspace Identification (SSI-COV). SSI-COV needs the definition of a model order n which is directly linked to the number of modes identified in the analysis ($n/2$). Because of the uncertainty of the SSI method, it's necessary to conduct this analysis considering a range of model orders which has, as result, the identification of several modes, generally called *poles*. The poles identified for a certain model order may have similarities or dissimilarities with the poles identified for a different model order. These are evaluated in function of the modal parameters which characterise each pole. The poles which show a low variation of the modal values with the changing of the model order are defined as stable (*physical modes*). On the other hand, the poles with high variation of the modal parameters with the changing of the model order are defined as unstable (*spurious modes*). This evidence has led to the ideation of stabilisation diagrams which show the variation of modal parameters in function of the model

order. In a *stabilisation diagram* the stability can be estimated by a visual analysis with the aim of determining a model order where the poles manifest a low variation of parameters. The adoption of a stabilisation diagram results in a manual identification of modal parameters which can lead to results affected by user experience.

In this study, an automated identification method is proposed with the aim of providing a process which is completely independent from the user experience, objective and based on the latest statistical methods of analysis. Consequently, a multistage cluster process is developed on the basis of the definition of the physical parameters characterising the modal properties.

The proposed method is tested on a numerical case considering the influence of the model order and the dimensions of the Hankel matrix on the results. Once all the parameters which control the process are validated, two experimental cases are analysed with the aim of verifying and quantifying the performance of the proposed method. In the first case a helicopter blade is used as a simplified experimental case; then, a scale reproduction of a masonry arch bridge is analysed as a complex application on a Civil Engineering structure.

Keywords:

Structural Health Monitoring, Operational Modal Analysis, Stabilisation diagram, Machine Learning, Clustering Analysis, Statistical Analysis, Masonry Arch Bridge, Helicopter Blade.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Problem Statement

Structural Health Monitoring is one of the most discussed topics of the latest years. This is because of the sensibilisation of the public opinion on the importance of the maintenance of historical and cultural heritage buildings and the safeguarding of the human life. As a matter of fact, more often than not people interact with old infrastructures, and the modernisation of these is not as quick as the increasing of transport flows and the request of new buildings.

A constant monitoring of structures can give some important information about the possible damages which may arise during their life. Moreover, an assessment of the residual life can be estimated, and maintenance strategies may be planned in time in order to avoid risks to human life and structural failure which may generate damages to other buildings.

In order to have parameters which characterise the behaviour of a structure or the changing of this with the time, different methods of investigation are conducted on the structures. The choice of these methods and the planning of study strategies depend on the typology of the object of study, the budget and other factors.

One of the most used analyses in SHM is taking information from the dynamic response of a structure. The dynamic response is evaluated on the basis of accelerations which may be generated by impulsive actions or by environmental vibrations. In case of recording of actions and reactions of the structure, the methods are defined as input-output methods. Otherwise, if the response of the structure is the only signal that is recorded, these are named as output-only methods. In the latest case, the most used method for extracting the modal parameters is the Stochastic Subspace Identification (SSI) method. This was conceived by Van Overschee and De Moor in 1996.

SSI method allows to obtain the modal parameters of a structure: frequencies, damping ratios and modal shapes. Due to the uncertainty of the results, the user is forced to conduct the SSI analysis for a range of model orders and try to identify which results are repeated for different orders. The parameters that present a high number of repetitions have a higher probability to be characterising the structure while the parameters that show a low number of repetitions are probably due to the uncertainty of the model.

For this reason, in the last few years, charts were conceived in order to show how the parameters computed by the algorithm change with the increasing of the model order. It is not always possible to understand from a chart if a parameter is stable using only a visual analysis. Therefore, some criteria were adopted in order to compare the variation of the parameters which characterise a mode identified for a certain model order with those identified for a different model order. It means the necessity to define limits a priori on the base of these parameters. Moreover, the results gotten in a chart may not be as clear as the initial prospects and more visual analysis may be needed. This means the user has to take some important decisions about the definition of thresholds and qualitative analyses. Consequently, the evaluation of the behaviour of the object of study is strongly influenced by the experience of the user.

The present dissertation treats only the SSI method and try to give a better interpretation of its results.

1.2 Aims and Objectives

In this dissertation an automatic operational modal analysis is proposed following a machine learning approach. The aim is to present a method that is completely autonomous, does not requires the supervision of the user, considers comparison criteria as numerous as possible and tries to give a statistical interpretation in all its steps.

In order to obtain these characteristics, a detailed study is carried out on different topics and the achievement of the aforementioned goals was possible keeping these objectives in mind:

- A literature review about similar cases of study which allows to identify comparison parameters, clustering algorithms and statistical tools helpful to conceive the final method;
- Implementation of an automatic operational modal analysis;
- Validation of the performance of the proposed method on a numerical case where the modal parameters of the structure are known in advance;
- Test on a simplified experimental case, a helicopter blade;
- Test on a complex experimental case, a scale reproduction of a masonry arch bridge.

1.3 Thesis Outline

This dissertation is organised in seven chapters as follows:

- **Chapter 1**: an introduction where the motivations which drive this work are explained with the goal to achieve a set of objectives.
- **Chapter 2**: a literature review. In this chapter several articles are analysed in order to define the bases of this work and to examine the methods which can be improved or developed differently. The chapter is divided in three main sections which represent the three primary topics for the definition of an AOMA:
  - Comparison parameters
  - Clustering algorithms
  - Statistical tools
- **Chapter 3**: in this chapter the methodology of the AOMA is discussed following the same steps of the algorithm with its logical pattern:
  - SSI algorithm;
  - Hard Criteria;

- Soft Criteria;

- Clusters-Modes identification;

- Removing outliers;

- Choice of representative modes;

- **Chapter 4**: a numerical case of a simple building with 9 degrees of freedom is analysed in order to validate the AOMA and to discuss important aspects of the SSI algorithm which affect the final result such as the range of the model orders and the Hankel matrix dimension. The aim of this chapter is to set the SSI parameters with the objective of obtaining the best performance of the AOMA

- **Chapter 5**: the AOMA is carried out on a simple experimental case of a helicopter blade. In this chapter the methodology mathematically explained in Chapter 3, is applied and the results of the process are showed with a step-by-step approach. The objective is to clarify the weight of each step of the proposed method on the final result. Then the results are compared with a FEM analysis and with a classical experimental modal analysis which is carried out on a stabilisation diagram with the user intervention.

- **Chapter 6**: a complex experimental case, a scale reproduction of a masonry arch bridge, is analysed showing the modal shapes identified by the AOMA and comparing them with the modal shape computed in an FE model.

- **Chapter 7**: in this chapter the conclusions are explained summarising the steps of the thesis, the objectives achieved and the further challenge with the aim of improving the quality of the results.

2 LITERATURE REVIEW

In the last few years, several studies have been carried out on the research of a method which can solve the aforementioned problems. Finding a universal technique is a challenge which has not been overcome yet.

In this chapter, a brief excursus is provided about the methods proposed by other authors. The followed strategy is of thematical type which tries to trace the logical path of the proposal algorithm.

In first, the comparison parameters are discussed. Successively, the clustering analyses are argued, considering the different typologies and the definitions of additional useful conditions. Then, the choice of the representative parameters of the identified modal clusters is examined.

Finally, a brief conclusion is exposed, trying to understand which techniques are the most efficient and which methods may influence the algorithm in a negative way.

2.1 Comparison parameters

The parameters which influence the modal identification analysis are relative to the modal variables such as:

- Frequency
- Damping ratio
- Modal shape

Frequencies and damping ratios are computed starting from the identification of the eigenvalues. Therefore, the eigenvalues can be considered as a further parameter which characterises the modes.

The stability of these parameters can be estimated by the variations between two quantities which characterise modes belonging to two different model orders. As is known, the variation of a system can be evaluated by the computing of the relative or the absolute difference

among its variables. This computing is possible both between two consecutive model orders (n and $n + 2$) and two general model orders (n and $n + 2k$). In this sense, the stability of the parameters belonging to a certain model order can be evaluated with reference to the only consecutive order or to a greater number of consecutive orders.

With regards to the eigenvectors, the Modal Assurance Criterion (MAC) is one of the most used methods to quantify the similarity between two vectors.

The evaluation of the modal shape leads to a complex eigenvector: its complexity is estimable through the phase deviation and the mean phase deviation.

2.1.1 Frequency, damping ratio and eigenvalue

The variation of frequency, damping ratio and eigenvalue can be computed as the absolute or the relative difference between two poles belonging to two different orders.

If X is the variable considered, and i and j are generical orders, both the above distances may be calculated as follows:

absolute distance	$\Delta X = |X_i - X_j|$
relative distance	$\Delta X_r = \dfrac{|X_i - X_j|}{\max{(X_i, X_j)}}$

Many authors choose the relative difference as the reference parameter to obtain all the variations in a scale of values between 0 and 1 where:

- 0 represents the perfect similarity between the two considered poles
- Vice versa, the value 1 is a sign of the complete dissimilarity.

This normalization allows to obtain a huge advantage: in fact, in this way the difference does not have a unit of measurement and all the variations may be considered with the same weight.

However, evaluating a variation with a relative distance leads to a new problem. Let us consider the values a_1, a_2, b_1, b_2 belonging to the same generic variable.

If the absolute distances between a_1, a_2 and b_1, b_2 are the same

$$\Delta a = \Delta b = |a_1 - a_2| = |b_1 - b_2|$$

And if a_1, a_2 present values much smaller than b_1, b_2, it is obvious that relative distances are different from the absolute distances and therefore:

$$\text{for} \quad a_1, a_2 \ll b_1, b_2 \quad \text{and} \quad \Delta a = \Delta b = |a_1 - a_2| = |b_1 - b_2|$$

$$\Delta a_r = \frac{|a_1 - a_2|}{\max(a_1, a_2)} \gg \Delta b_r = \frac{|b_1 - b_2|}{\max(b_1, b_2)}$$

As a consequence, if we consider a threshold for the maximum relative distance among entities of a variable, this may be stricter for small values than for bigger ones.

The problem of assigning the same scale to each variable may be overcome considering a normalisation of the difference vector trough the use of appropriate techniques such as the "standard score normalization" or the "min-max normalization".

This step is necessary in order to analyse the data with a clustering algorithm which considers different variables in the same process.

Both methods will be explained in the following chapters.

2.1.2 Modal Assurance Criterion

The similarity of two eigenvectors can be seen with the Modal Assurance Criterion. This parameter is often used by many authors and it's

considered a well-established method to evaluate the stability of the modal shapes.

2.1.3 Mean Phase and Mean Phase Deviation

The complexity of a mode shape can be measured by using the mean phase and the mean phase deviation.

Both these entities may have a big role in a stabilisation algorithm as is suggested by many authors. Reynders [1] uses the MPD for rejecting the poles where this parameter shows a high value (close to 1). Neu [2] considers the variation of MPD among poles belonging to different model orders.

In this dissertation, the values of MPD are used with the same interpretation of the latter author. The reason for this choice is given by visual results obtained during the k-means clustering. In fact, as will be showed later, the variation of MPD, indicated with the symbol dMPD, is correlated to the variations of the other parameters used in the cluster analysis.

2.2 Clustering algorithms

The clustering algorithms used in most of the works carried out on the automated identification systems are two:

- K-means clustering
- Hierarchical clustering

The former is used to separate the poles which show high stability from the poles with low stability.

The latter one is used in order to create clusters of poles with similar features.

2.2.1 K-means clustering

The authors chose this method with the aim of splitting the poles into two groups. The first one contains the poles which present low variation of the

parameters while the second one is the cluster of the poles which show high variations of the variables.

The variations considered in this work are in terms of frequency, damping ratio, eigenvalue, modal shape (MAC) and mean phase deviation. As recommended by Reynders [1], the parameters which must be considered in the computing of the similarity, should be as many as possible. The problem is finding entities that help us to characterise the poles and their stability. The addition of new parameters to the k-means clustering is not always possible.

The variations are usually computed between the most similar poles belonging to two consecutive orders. This similarity can be chosen considering the value $df_r + (1 - MAC)$, as suggested by Neu [2].

As previously discussed, in order to process more variables at the same time in a k-means analysis, the variations of the parameters must be scaled with the aim of obtaining the same weight for each one of them.

This leads many authors to consider the relative distances as reference, allowing to obtain values between 0 and 1 for the variations of each parameter. In this dissertation a different approach will be adopted in order to establish an equal weight for the variables.

A further consideration may be made on the pre-processing data before the k-means clustering. This method requires the distribution of the data object of analysis to follow a gaussian distribution, as recommended by Neu [2].

The gaussianisation process can be computed in different ways: one of the methods which allows a redistribution of the variables to a normal distribution is the Box-Cox transformation. This method is used by Neu and adopted in this work.

2.2.2 Hierarchical clustering

The hierarchical clustering allows to obtain clusters containing poles with similar features. This process needs of the definition of the following parameters:

- distance among poles
- linkage method
- choosing a threshold

For the distance among the poles, Reynders and Neu chose $df_r + (1 - MAC)$. As mentioned, in this work the distances will be considered in absolute value but the necessity to give the same weight to the variation of frequency and modal shape is a problem that persists. To cope with this, the variables df and $(1 - MAC)$ will be scaled in an appropriate way.

The linkage method is a personal choice but the average method may be considered as one of the most efficient for this kind of analysis.

The choice of a threshold is a challenge for every hierarchical clustering. The algorithm tends to collect the poles in a unique cluster. An important step analysed in this work starts from the consideration that the poles belonging to the same model orders can't be within the same cluster. This allowed to obtain a maximum number of cluster but, as evidenced by the results, this approach is necessary but insufficient to obtain the right number of clusters.

Therefore, the choose of a threshold is adopted as suggested by Neu.

2.3 Statistical tools

Statistical investigations are used with two aims:

- To give a gaussian distribution to the variables analysed during the k-means clustering;
- To remove outliers from the final identified clusters.

The Box-Cox transformation is used for the former objective, as recommended by Neu.

For the latter aim, the method adopted is different and simpler than the method used by Neu.

2.4 Conclusions

In the present dissertation, the steps followed for the identification of the modal parameters are the same used by many other authors. The changes made to the process try to give a better interpretation to the results and overcome some limits showed in the literature approaches. This is computed with few adjustments which don't overturn the well-established methods.

3 METHODOLOGY

Before showing the methodology used in the automated identification code, a brief presentation of the SSI [3] algorithm will be introduced in order to clarify the problems faced.

Subsequently, the steps of the process of identification will be explained in the same order pursued in the code.

The code is outlined as follows.

- K-means clustering
- Hierarchical clustering
- Removing outliers within the final clusters
- Choosing of the representative parameters for each final cluster

3.1 SSI algorithm

Stochastic Subspace Identification (SSI) is an algorithm which allows to identify the modal parameters of a structure starting from the accelerations recorded by sensors located on the object of study.

Usually SSI is adopted in cases where the input signal is unknown and, for this reason, it is defined as an output-only method. The typical example of application is on the structure where the excitation arises from the environmental vibrations. SSI, other than the accelerations array, needs of two parameters:

- The model order
- The dimension of the Hankel matrix

The theory at the base of the SSI algorithm is taken from a Technical Paper on the Stochastic Subspace Identification Techniques [4].

In output-only modal analysis, the two main assumptions are that the underlying physical system behaves *linearly* and it is *time-invariant*.

The model structure can be expressed with two equations:

$$x_{t+1} = A \cdot x_t + w_t \qquad\qquad \text{3-1}$$

$$y_t = C \cdot x_t + v_t \qquad\qquad \text{3-2}$$

The equation 3-1 is called the state equation and models the dynamic behaviour of the system. In other words, this equation transforms the state vector x_t (dimension $n \cdot 1$) in a new state x_{t+1} ($n \cdot 1$). The dimension n of a state vector, is called the state space dimension. The state vectors, x_t and x_{t+1}, contain the displacement and the velocity respectively at the instant t and $t + 1$. The time is discretised by the sampling frequency f_s with the relation $f_s = 1/(t - (t + 1))$. The matrix A is a squared matrix $n \cdot n$ and it models the dynamic of the physical system. The vector w_t, which represents the contribution of the input excitation to the state vector, is considered as a stochastic process (random noise).

The second equation (3-2) is called the observation equation. The vector y_t of dimension $p \cdot 1$, is the observable vector and contains the measured output. The matrix C is the observation matrix which links the state vector to the observable one. The vector v_t is considered as a stochastic process, as w_t, and it represents the contribution of the input to the output.

Another important assumption in the model system is that the system response is modelled by a stochastic process. Consequently, if E is the expectation operator, the moments of y_t are expressed by:

$$E[y_t] = 0 \qquad\qquad \text{3-3}$$

$$E[y_{t+i}y_t^T] = \Lambda_i \qquad\qquad \text{3-4}$$

Where i denotes the time lag $i\Delta t$ and Λ_i indicates the covariance function.

Since the system response of the linear state space model is a Gaussian stochastic process, it follows that x_t, w_t and v_t all are Gaussian stochastic processes as well. The input processes w_t and v_t are unknown and, for this reason, it is only possible to adopt some assumptions about their

statistical properties: they are two correlated zero-mean Gaussian white noise processes.

Defining G as the covariance matrix between system response y_t and the updated state vector x_{t+1}

$$E[y_t \, x_{t+1}{}^T] = G \tag{3-5}$$

the covariance function can be written as follows:

$$E[y_{t+i} \, y_t{}^T] = \begin{cases} \Lambda_0 = C & i = 0 \\ \Lambda_i = CA^{i-1}G & i \neq 0 \end{cases} \tag{3-6}$$

There are two additional system matrices that turn out to play an important role:

$$\Gamma_i = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-2} \\ CA^{i-1} \end{bmatrix} \tag{3-7}$$

$$\Delta_i = \begin{bmatrix} A^{i-1}G & A^{i-2}G & \cdots & AG & G \end{bmatrix} \tag{3-8}$$

Where Γ_i is the extended observability matrix and Δ_i is the reversed extended stochastic controllability matrix.

Correlations of the data with respect to a subset of reference sensor channels computed at a sequence of different time lags may be assembled into a block-Hankel matrix as follows:

$$H_i = \begin{bmatrix} \Lambda_i & \cdots & \Lambda_1 \\ \vdots & \ddots & \vdots \\ \Lambda_{2i-1} & \cdots & \Lambda_i \end{bmatrix} = \Gamma_i \Delta_i \tag{3-9}$$

The Singular Value Decomposition (SVD) technique is then used to reduce the block-Hankel matrix into suitable factors as follows:

$$H_i = USV^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1{}^T \\ V_2{}^T \end{bmatrix} = U_1 S_1 V_1{}^T \tag{3-10}$$

The observability and controllability matrices are then computed using the equations 3-9 and 3-10:

$$\Delta_i = U_1 S_1^{1/2}$$ 3-11

$$\Gamma_i = S_1^{1/2} V_1^{T}$$ 3-12

The matrix can be obtained as follows:

$$A = \Delta_i(1:p(i-1),:)^{+}\Delta_i(p+1:pi,:)$$ 3-13

Where p is the number of output channels, i is the number of block rows of the Hankel Matrix and $[*]^{+}$ indicates the pseudo-transpose.

Then the modal parameters, are computed knowing the relation of the eigenvalues and the eigenvectors with the matrix A.

$$A = \Psi M \Psi^{-1}$$ 3-14

Where Ψ is the matrix which contains the eigenvectors and M is a diagonal matrix containing the discrete-time eigenvalues μ_j. The values μ_j are directly related to the eigenvalues of the system with the following equation:

$$\lambda_j = \frac{\ln \mu_j}{\Delta t}$$ 3-15

The time lag $i\Delta t$ is a user choice as the dimension of the state vector which is related to the model order. The value of i affects the dimension of the Hankel Matrix and, consequently, the results of the modal identification. The choice of this parameter is a challenge which will be discussed in the following chapters.

The number of the model order should be chosen in function of the quantity of sensors used for the recording and the degrees of freedom of the structure. For continuous structures, the degrees of freedom are infinite and for this reason it is not possible to define the number of significant modes to consider in the analysis a priori. Consequently, the model order is not an assured quantity and therefore not definable a priori.

For this reason, the model order considered is not unique and a range of orders may be adopted and applied to the same case of study.

The modes identified in a model order are called poles. These are characterized by eigenvalues and eigenvectors which should appear as couple of conjugate complex entities. From the identification of complex eigenvalues, the determination of the frequency and the damping ratio are computed as follows:

$$\lambda = \alpha + i\beta \qquad\qquad \text{3-16}$$

$$f = \sqrt{\alpha^2 + \beta^2} \qquad\qquad \text{3-17}$$

$$dr = \frac{-\alpha}{\sqrt{\alpha^2 + \beta^2}} \qquad\qquad \text{3-18}$$

Furthermore, the SSI is affected by imprecisions and the poles identified for a certain model order may show dissimilar parameters from the poles established for a different model order. This evidence is explicated with the following meaning:

- The poles which show similar parameters by the variation of the model order are considered stable
- Vice versa, the poles which present high variation of parameters with the changing of the model order are considered unstable

The concept of similarity and dissimilarity is not always clear and for this reason the adoption of rigid limits on the variation of the parameters is not often suitable for different analyses. A statistical approach is more appropriate and allows to obtain different limits for different qualities of results.

3.2 Hard Criteria

The "*Hard Criteria*" are defined as the criteria for which it is necessary to adopt a boundary condition for the identification of modes considered as mathematics and not physics.

The poles identified for a certain model order may show parameters with a no physical meaning. In fact, for structural cases, the damping ratio that is expected in a mode should have positive values and however not higher than 10-25% (Cabboi [5], Ceravolo [6], Demarie [7]).

For this reason, a boundary condition is adopted as follows:

$$0 < dr < 0.2 \qquad\qquad 3\text{-}19$$

All the poles identified with values of damping ratio exceeding these limits are considered as mathematical modes.

Furthermore, in the SSI algorithm, the eigenvalues and the eigenvectors computed for a certain model order should present themselves as couples of conjugate complex values. If this is not the case, the poles are discarded (Reynders [1]).

3.3 Soft Criteria

The term "*Soft Criteria*" is used to refer to all the methods adopted for discarding the poles with comparison parameters among these.

In this work, the comparison parameters are computed among the two most similar poles belonging two consecutive orders.

If j is a generic pole identified for a model order n, a comparison n-dimensional vector p_j is associated to this pole. As mentioned, the components p_{j_i} are computed between the parameters of j and those of the most similar pole belonging the consecutive model order $n + 2$.

After the connection of the comparison vectors to all the poles, a k-means clustering is conducted on these, defining two clusters where:

- In the first cluster ($C1$) the poles which show low variation of variables with the increasing of the model order are distinct
- In the second cluster ($C2$), on the contrary, the poles which have variables with a large difference are selected.

Therefore, the poles which fall within $C2$ are discarded as considered "*unstable*" while the poles belonging to $C1$ are taken as stable and successively processed as explained in 3.4.

3.3.1 Comparison parameters

In this dissertation, the comparison parameters analysed are the following:

- Absolute difference of eigenvalue $d\lambda$
- Absolute difference of frequency df
- Absolute difference of damping ratio ddr
- Modal Assurance Criterion MAC
- Absolute difference of Mean Phase Deviation $dMPD$

For each pole j belonging to a model order n, the previous entities are calculated compared to all the poles k belonging the consecutive model order $n + 2$ as follows.

$$d\lambda_j{}^n = |\lambda_j{}^n - \lambda_k{}^{n+2}| \qquad \text{3-20}$$

$$df_j{}^n = |f_j{}^n - f_k{}^{n+2}| \qquad \text{3-21}$$

$$ddr_j{}^n = |dr_j{}^n - dr_k{}^{n+2}| \qquad \text{3-22}$$

$$dMAC = 1 - MAC(\phi_j{}^n, \phi_k{}^{n+2}) \qquad \text{3-23}$$

$$dMPD = |MPD(\phi_j{}^n) - MPD(\phi_k{}^{n+2})| \qquad \text{3-24}$$

where

$$MAC(\phi_j{}^n, \phi_k{}^{n+2}) = 1 - \frac{[(\phi_j{}^n)^*(\phi_k{}^{n+2})]^2}{[(\phi_j{}^n)^*(\phi_j{}^n)] \cdot [(\phi_k{}^{n+2})^*(\phi_k{}^{n+2})]} \qquad \text{3-25}$$

with $\blacksquare^*$ that denotes the conjugate transpose of the vector.

For the computation of the Mean Phase Deviation, the steps followed are the same as suggested by Phillips and Allemang [8] and they are described in the 7A.1.

In this way, a matrix of dimensions l, m is obtained for each comparison parameter, where l is the number of poles belonging to the order n and m is the number of poles belonging to the order $n + 2$.

3.3.1.1 Finding the neighbour

The next step is finding the poles of the order n which are more similar to each pole of the model order $n + 2$. To reach this goal, the only variables df and $dMAC$ are used. As known, these have two different units of measure and two different scales:

- $dMAC$ is a pure number and assumes values between 0 and 1, where 0 denotes the perfect similitude of the eigenvectors and 1 the complete difference
- df has a unit in Hz and may assume any positive value, 0 included

To give an equal scale and pure numbers to both the entities, the matrices of $dMAC$ and df are normalized using a "*min max normalization*".

$$[\widehat{dMAC}] = \frac{[dMAC] - \min([dMAC])}{\max([dMAC]) - \min([dMAC])} \qquad \text{3-26}$$

$$[\widehat{df}] = \frac{[df] - \min([df])}{\max([df]) - \min([df])} \qquad \text{3-27}$$

After this transformation, the values assumed in the generic position of both the matrices are with certainty between 0 and 1.

At this point, the two matrices may be added in order to obtain a matrix of similarity $[S]$:

$$[S] = [\widehat{dMAC}] + [\widehat{df}] \qquad \text{3-28}$$

For each row, the minimum value can be found and the logical number 1 is associated with this position while for the other elements of the row the logical 0 is given. In this way, a logical matrix $[P]$ is computed which presents only one element of the row with value equal to 1. The position $P_{i,j} = 1$ indicates that pole j of the order $n + 2$ is the *neighbour* of the pole i of the order n.

3.3.1.2 Associating the comparison vector to each pole

Multiplying the matrix $[P]$ for each comparison matrix, element by element, 5 vectors of dimension l are found, where 5 is the number of comparison methods considered in this case of study and l the number of poles within the order n.

$$\{d\lambda\} = [P].* [d\lambda] \qquad \text{3-29}$$

$$\{df\} = [P].* [df] \qquad \text{3-30}$$

$$\{ddr\} = [P].* [ddr] \qquad \text{3-31}$$

$$\{dMAC\} = [P].* [dMAC] \qquad \text{3-32}$$

$$\{dMPD\} = [P].* [dMPD] \qquad \text{3-33}$$

The new matrix of comparison $[C]$ can be assembled as follows

$$[C] = [\{d\lambda\} \ \{df\} \ \{ddr\} \ \{dMAC\} \ \{dMPD\}] \qquad \text{3-34}$$

The row i of the matrix $[C]$ contains the comparison parameters associated to the pole i.

3.3.2 Data pre-processing

As known, the k-means algorithm works well with variables which assume a distribution close to the Gaussian one [9].

Another important issue is the use of a clustering algorithm on elements characterized by a number of parameters larger than 1. In fact, it's necessary to normalise all the variables in order to obtain pure numbers belonging the same domain of existence.

3.3.2.1 Box-Cox transformation

The Box-Cox transformation [10] is a process which might allow to obtain a distribution of the variables close to the Gaussian distribution.

The Box-Cox transformation is a parametric family of transformation from y to $y(\lambda)$ trough the parameter λ

$$y(\lambda) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & if \; \lambda \neq 0 \\ \log(y) & if \; \lambda = 0 \end{cases} \qquad \text{3-35}$$

The parameter λ is obtained by a profile log-likelihood maximization.

This transformation must be conducted for each parameter considering the values assumed by all the poles.

This means there is a necessity to create a vector of the generic parameter p containing the value of each pole within all the orders considered in the analysis. This vector can be assembled as follows:

$$\{p\} = \begin{Bmatrix} \{p_{nmin}\} \\ \{p_{nmin+2}\} \\ \vdots \\ \{p_n\} \\ \vdots \\ \{p_{nmax-2}\} \\ \{p_{nmax}\} \end{Bmatrix} \qquad \text{3-36}$$

$$\{p_n\} = \begin{Bmatrix} p_{1,n} \\ p_{2,n} \\ \vdots \\ p_{i,n} \\ \vdots \\ p_{l-2,n} \\ p_{l_n,n} \end{Bmatrix} \qquad \text{3-37}$$

Where p is the generic parameter ($d\lambda$, df, ddr, $dMAC$, $dMPD$), $nmin$ and $nmax$ are the minimum and maximum order and l_n is the number of poles belonging the generic order n.

3.3.2.2 Z-score normalization

Once a better distribution of the parameters is obtained, it appears that they are not in the same scale. For this reason, a z-score normalization [11] is used in this work to obtain a mean equal to 0 and a standard deviation equal to 1 for each variable.

The steps for this normalization are the followings:

$$\mu_p = \frac{1}{\sum_{n=nmin}^{nmax} l_n} \sum_{n=nmin}^{nmax} \sum_{i=1}^{l_n} p_{i,n}$$

3-38

$$\sigma_p = \sqrt{\frac{1}{\sum_{n=nmin}^{nmax} l_n} \sum_{n=nmin}^{nmax} \sum_{i=1}^{l_n} (p_{i,n} - \mu)^2}$$

3-39

$$z_{i,n} = \frac{p_{i,n} - \mu_p}{\sigma_p}$$

3-40

The vector $\{z\}$ will be the final transformation of the vector $\{p\}$ for the generical variable and, after that, $\{z\}$ will be processed with the k-means clustering.

3.3.3 k-means clustering

The "*k-means clustering*" [12] is a method of data partitioning where the elements analysed are grouped in a number of clusters equal to *k*. For this reason, the user must define the value of *k* a priori.

This clustering is an iterative process and the main steps may be summarised as follows:

1. The staring centroid C_1, C_2, ...C_k of the *k* clusters can be defined as random or with specific values

2. For each element analysed the *distance* among the element and the centroids of the *k* clusters is computed

3. Each element is assigned to the neighbour cluster according with the point 2

4. For each cluster, a new centroid is calculated as the mean of the element's parameters belonging to the considered cluster

5. Steps 2 to 4 are repeated until the centroid of the step *i* is equal to the previous one.

The analysed elements may be characterized by a certain number *l* of parameters. This allows to process more variables in the same analysis.

The methods used to compute the distance between elements may be of different nature and can be summarised in Table 3-1 - list of possible measurements of distance

.

Squared Euclidean distance	$d(x,c) = (x - c)(x - c)'$
Sum of absolute differences	$d(x,c) = \sum_{j=1}^{p} |x_j - c_j|$
Cosine	$d(x,c) = 1 - \dfrac{xc'}{\sqrt{(xx')(cc')}}$

Table 3-1 - list of possible measurements of distance [13]

In this study, the assumptions are the following:

- the variables processed are the transformations of the parameters $d\lambda$, df, ddr, $dMAC$ and $dMPD$, as described in 3.3.1 and 3.3.2

- the number of clusters is equal to 2 in order to separate the poles identified as "*stable*" to the poles classified as "*unstable*"

- the distance adopted is the Squared Euclidean distance as showed in Table 3-1, where:

- - x is the vector referred to a certain pole which contains the variables $d\lambda$, df, ddr, $dMAC$ and $dMPD$
  - c is the coordinate vector of the cluster centroid
- the starting centroids of the two clusters are placed on the two opposite sides of the 5-D space, in particular:

$$C1_0 = \begin{Bmatrix} -3 \\ -3 \\ -3 \\ -3 \\ -3 \end{Bmatrix} \qquad C2_0 = \begin{Bmatrix} +3 \\ +3 \\ +3 \\ +3 \\ +3 \end{Bmatrix}$$

- after the cluster analysis, the poles which present x within the final cluster $C2$ are considered as "*unstable*" and discarded.

3.4 Clusters-Modes identification

The poles identified as "*stable*" are processed all together with a hierarchical clustering algorithm, in order to obtain groups of similar poles.

Therefore, in this step the comparison is not exclusively done among poles of consecutive orders and, as consequence, the stability assumes a meaning extended to more model orders.

The linkage among poles needs the definition of a distance vector that is computed as established in 3.4.1.

The hierarchical clustering (3.4.2) starts considering the poles as single clusters and links the closest clusters defining new groups. In order to create new congregates, the algorithm needs of the definition of a linkage method.

The process is over once it is not possible to link new clusters anymore or, by the definition of a threshold $\tilde{d}$, when the distance among clusters overcomes $\tilde{d}$. The choice of $\tilde{d}$ is computed following a statistical approach described in 3.4.3

The final clusters may present a high number of poles or a lower one:

- If a cluster shows several poles, it means there is a high level of identification of similar poles for different model orders. This could be an evidence of the fact that cluster represents a *physical* mode.
- On the contrary, the clusters which contain fewer poles are representative of low redundancy of identification and will be considered ad *mathematical* modes.

The choice of the limit between *physical* and *mathematical* modes is computed with a k-means clustering, explained in 3.4.4

Once established which clusters are *physical* modes, a statistical analysis is conducted for each group, in order to reject potential outliers, as described in 3.5.

3.4.1 Distance vector

The distance is computed among poles belonging to different model orders and considering the frequency and modal shape.

Considering a pole j and a pole k, respectively belonging to a model order n_j and n_k, the following distances are computed

$for\ j \neq k$	$df_{jk} = \|f_j - f_k\|$	3-41
	$dMAC_{jk} = 1 - MAC(\phi_j, \phi_k)$	3-42
$for\ j \neq k$	$df_{jk} = NaN$	3-43
	$dMAC_{jk} = NaN$	3-44

The definition of these quantities entails that $df_{jk} = df_{kj}$ and $dMAC_{jk} = dMAC_{kj}$.

Each entity is recorded inside two vectors $\{df\}$ and $\{dMAC\}$ of dimensions

$$m = \frac{l \cdot (l-1)}{2}$$

Where l is the total number of poles considered in the analysis and m the number of the total possible combinations. The shape of the two vectors is showed below

$$\{df\} = \begin{Bmatrix} df_{12} \\ df_{13} \\ \vdots \\ df_{1l} \\ df_{23} \\ \vdots \\ df_{2l} \\ \vdots \\ df_{(l-1)l} \end{Bmatrix} \qquad \{dMAC\} = \begin{Bmatrix} dMAC_{12} \\ dMAC_{13} \\ \vdots \\ dMAC_{1l} \\ dMAC_{23} \\ \vdots \\ dMAC_{2l} \\ \vdots \\ dMAC_{(l-1)l} \end{Bmatrix}$$

As for point 3.3 a normalization is necessary in order to give the same weight to the two variables. Therefore, a *min-max normalization* is adopted for both the vectors.

$$\widehat{df}_{jk} = \frac{df_{jk} - \min{(df)}}{\max(df) - \min{(df)}} \qquad \text{3-45}$$

$$\widehat{dMAC}_{jk} = \frac{dMAC_{jk} - \min{(dMAC)}}{\max(dMAC) - \min{(dMAC)}} \qquad \text{3-46}$$

In this way, the NO-NaN entities in the two vectors $\{\widehat{df}\}$ and $\{\widehat{dMAC}\}$ assume values between 0 and 1.

The distance vector $\{D\}$ is computed as the sum of $\{\widehat{df}\}$ and $\{\widehat{dMAC}\}$

$$\{D\} = \{\widehat{df}\} + \{\widehat{dMAC}\} \qquad \text{3-47}$$

3.4.2 Hierarchical clustering

Hierarchical clustering is one of the most popular and easy clustering techniques. There are two types of clustering: *Bottom-up (agglomerative) methods* and *Top-down (divisive) methods* [12].

In this work, an agglomerative method is used following the algorithm structure proposed in MathWorks [14] and described below:

1. Find the similarity or dissimilarity between every pair of objects in the data set, as described in 3.4.1.

2. Group the objects into a binary, hierarchical cluster tree. In this step, pairs of objects that are in close proximity are linked using a *linkage function*. The linkage function uses the distance information generated in step 1 to determine the proximity of objects to each other. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed.

3. Determine where to cut the hierarchical tree into clusters, as described in 3.4.3.

The step 2, the *linkage method*, is described following the "*Data Mining - The Textbook*" [12] approach with a particular attention to the methods implemented on MATLAB.

The two clusters to be merged are denoted by i and j, respectively. The distance between two groups of objects is computed as a function of the $mi \cdot mj$ pairs of distances among the constituent objects. The different ways of computing distances between two groups of objects are as showed in Table 3-2

Single	The distance is equal to the minimum distance between all $mi \cdot mj$ pairs of objects. This corresponds to the closest pair of objects between the two groups.
Complete	The distance between two groups of objects is equal to the maximum distance between all $mi \cdot mj$ pairs of objects in the two groups. This corresponds to the farthest pair in the two groups.
Average	The distance between two groups of objects is equal to the average distance between all $mi \cdot mj$ pairs of objects in the groups.

Centroid	the closest centroids are merged in each iteration.

Table 3-2 - Linkage Methods [12]

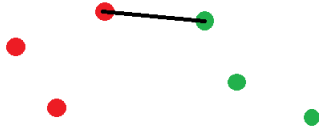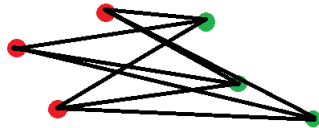These methods may be visualized as follows

Single	
Complete	
Average	
Centroid	

Table 3-3 - Linkage Methods (figures)

The method adopted in this work is the *Average Method* and the measurement of the distance may be computed with the equation 3-48

$$D(c_i, c_j) = \frac{1}{n_i \cdot n_j} \left(\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} D_{ij} \right)$$

3-48

Where n_i, n_j are the number of elements within c_i, c_j and D_{ij} is the distance between the element i belonging c_i and j belonging c_j according with 3-47.

3.4.3 Choice of a threshold

The agglomerative hierarchical clustering is a clustering algorithm which starts considering the single poles as starting clusters and, linking the neighbours, it reaches a cluster which contains all the poles.

In this study, the aim of using the hierarchical clustering is to obtain as many clusters as the identified modes in the analysis. Therefore, the choice of a threshold is fundamental in order to cut the dendrogram generated by the hierarchical clustering at certain level and, consequently, obtain a number of clusters. These clusters should represent a group of poles with similar parameters.

The choice of a threshold cannot always be done using a visual analysis, especially in the proposed method where a fully automated method is advanced with the aim of getting results not influenced by the user.

For this reason, the threshold which establishes the number of clusters is defined using a statistical approach on the base of distances among the most similar poles belonging to two consecutives model orders.

Consequently, the distribution of these distances is fitted with a Weibull distribution and it is used to find the 95th percentile nearest-distances between probable physical modes.

3.4.3.1 Distance vector

Starting from the residual poles, a pole j belonging to a model order n is selected and the distances between this and the poles k belonging to the model order $n + 1$ are computed with the same steps showed in 3.4.1.

$$\boldsymbol{df}_{jk} = |\boldsymbol{f}_j - \boldsymbol{f}_k| \qquad \text{3-49}$$

$$\boldsymbol{dMAC}_{jk} = 1 - \boldsymbol{MAC}(\boldsymbol{\phi}_j, \boldsymbol{\phi}_k) \qquad \text{3-50}$$

Consequently, the minimum distance is selected both for df_{jk} and $dMAC_{jk}$ finding the nearest distance between a pole belonging to a model order and a pole belonging to the consecutive one. If the model order $n + 1$ is empty, the distance is not computed.

Therefore, as mentioned in the previous chapter, the two vectors need of a normalization and the min-max method is adopted in this case as well.

$$\widehat{df}_{jk} = \frac{df_{jk} - \min(df)}{\max(df) - \min(df)} \qquad \text{3-51}$$

$$\widehat{dMAC}_{jk} = \frac{dMAC_{jk} - \min(dMAC)}{\max(dMAC) - \min(dMAC)} \qquad \text{3-52}$$

In this way, the NO-NaN entities in the two vectors $\{\widehat{df}\}$ and $\{\widehat{dMAC}\}$ assume values between 0 and 1.

The distance vector $\{D\}$ is computed as the sum of $\{\widehat{df}\}$ and $\{\widehat{dMAC}\}$

$$\{D\} = \{\widehat{df}\} + \{\widehat{dMAC}\} \qquad \text{3-53}$$

The vector $\{D\}$ is ready to be processed in order to have the Weibull parameters which fit the distribution of the distances.

3.4.3.2 Weibull distribution

As suggested by Neu [2], the value where to cut the dendrogram is computed from the inverse cumulative distribution function of the Weibull

distribution. The 95th percentile is used to find nearest-distances between probable physical modes.

The probability density function of a Weibull random variable is defined as follows:

$$f(x, \lambda, k) = \begin{cases} \dfrac{k}{\lambda} \left(\dfrac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}} & if\ x > 0 \\ 0 & if\ x < 0 \end{cases} \qquad \textit{3-54}$$

Where

$k > 0$ is the shape parameter

$\lambda > 0$ is the scale parameter

The maximum likelihood estimator for k is the solution for k of the following equation

$$0 = \frac{\sum_{i=1}^{n} x_i{}^{k} \ln x_i}{\sum_{i=1}^{n} x_i{}^{k}} - \frac{1}{k} - \frac{1}{n} \sum_{i=1}^{n} \ln x_i \qquad \textit{3-55}$$

The maximum likelihood estimator for the λ parameter given k is

$$\hat{\lambda}^{k} = \frac{1}{n} \sum_{i=1}^{n} x_i{}^{k} \qquad \textit{3-56}$$

The parameter estimation is implemented in MATLAB.

As mentioned, once the 95th percentile is fixed, the value of the threshold is established.

3.4.4 *Physical* and *Mathematical* modes

Once the dendrogram is cut at a certain level as described in 3.4.3, the hierarchical clustering generates a certain number of clusters. These

clusters should represent the identified modes as well as groups of poles with similar parameters in terms of frequency and modal shape.

Some clusters could show a high number of poles belonging to them while others could present fewer poles. This evidence can be interpreted as a signal of the fact that clusters with few poles are *mathematical* modes, identified by the algorithm but not actually modes of the structure. On the contrary, clusters with a high number of poles should be representative of the analysed structure and they are defined as *physical* modes.

For this reason, the clusters are split in two groups following a 2-means clustering based on the number of poles within each cluster.

3.4.4.1 Addition of empty clusters

As suggested by Reynders [1], in order to obtain a better solution for the k-means clustering and to avoid the elimination of physical clusters, empty clusters are added to this analysis with a number equal to one fifth of the number of the poles belonging to the largest cluster identified by the hierarchical clustering.

3.4.4.2 2-means clustering

Once the empty clusters are added, the clusters identified by the hierarchical clustering are processed with the 2-means clustering using as variable the number m_i of poles belonging to each cluster.

The starting centroids are placed as follows

$$C_1 = \max{(m_i)}$$

$$C_2 = 0$$

Once the cluster analysis is over, the clusters with m_i closest to the final centroid C_1 are considered as *physical* modes while, the others are considered *mathematical* modes and discarded.

3.5 Removing outliers

The process showed in 3.4 leads to a final number of clusters which represent the modes identified by the proposed method. The identification of the modal parameters from each cluster starts from the parameters which each pole present. For this reason, a method of removal of outliers within the clusters is necessary in order to avoid any parameter influencing the final result.

For each cluster i, the centroid is determined as the mean of the parameters characterizing the poles k^i belonging to i.

$$\bar{f_i} = \frac{1}{m_i} \sum_{k=1}^{m_i} f_{k^i}$$

$$\overline{dr_i} = \frac{1}{m_i} \sum_{k=1}^{m_i} dr_{k^i}$$

$$\{\bar{\psi}_i\} = \frac{1}{m_i} \left\{ \begin{array}{c} \sum_{k=1}^{m_i} \psi_{1,k^i} \\ \sum_{k=1}^{m_i} \psi_{2,k^i} \\ \vdots \\ \sum_{k=1}^{m_i} \psi_{n,k^i} \end{array} \right\}$$

Where m_i is the number of poles belonging to the i-cluster and n is the number of channels (number of components of the eigenvector).

Consequently, the difference among parameters are computed for each pole k^i in terms of df, ddr and MAC.

$$df_{k^i} = |\bar{f_i} - f_{k^i}|$$

$$ddr_{k^i} = |\bar{f_i} - f_{k^i}|$$

$$MAC_{k^i} = MAC\left(\{\bar{\psi}_i\}, \{\psi_{k^i}\} \right)$$

Using MATLAB, the outliers are removed with the MAD criterion (appendix 7A.3). Outliers are defined as the poles which present at least one of the three parameters higher than three scaled MAD from the median.

3.6 Choice of representative modes

Once the physical modes are identified as the final clusters and the outliers are removed, the problem is selecting the representative parameters for each cluster. In fact, each cluster contains a certain number of poles with similar but not equal parameters.

It is not simple choosing a unique method for extracting frequency, damping ratio and modal shape. Many authors have tried to give a personal interpretation of the results; for example, Reynders [1] selected the parameters of the pole with the closest damping ratio to the mean of *dr* that characterises the poles which belong to the selected cluster.

In this work, different methods are proposed and implemented in order to give the opportunity to select the best solution for each case by testing each method on specific case.

For each cluster, the modal parameters can be extracted with the following methods:

1. Choosing the mean of *frequency*, *damping ratio* and *eigenvector* of the poles belonging to the cluster.
2. Choosing the pole which presents the minimum distance between its *damping ratio* and the mean of the *dr* of the poles belonging to the analysed cluster.
3. Choosing the pole which presents the minimum distance between its parameters (*frequency*, *damping ratio* and *eigenvector*) and the mean of all the parameters of the poles belonging to the analysed cluster.
4. Choosing the pole which presents the minimum value of the sum of differences between the components of its *eigenvector* and the

components of the *eigenvector* computed as the mean of the eigenvectors of the poles belonging to the analysed cluster.

4 NUMERICAL CASE

4.1 Description

The proposed method is tested on a numerical case which is representative of a 3-dimensional building with three floors and it presents the following features:

- Each floor is considered as infinitely rigid; the behaviour of the building is like a shear type structure which presents three degrees of freedom for each floor: two orthogonal displacements and one rotation are allowed on the three levels.
  As a result, the building has a total of nine degrees of freedom.

- An irregular disposition of the vertical elements; for this reason, the building has an eccentricity between the centre of mass (CM) and the centre of resistance (CR).

- The acceleration sensors are three per floor and placed as showed in

$$Median(\{X\}) = X_c$$

Error! No text of specified style in document.-8

$$\{\vec{X}\} = sort_{smallest-gretest}\{X\} = X_1, X_2, \ldots X_n$$

$$X_i < X_{i+1} \quad \forall\, i \in \mathbb{N}$$

Error! No text of specified style in document.-9

$$X_c = \begin{cases} X_{(n+1)/2} & for\ n\ odd \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & for\ n\ even \end{cases}$$

Error! No text of specified style in document.-10

MAD is defined as the median of the absolute difference between a vector Y and its relative median value Y_c.

$$MAD = Median(|\{Y\} - Y_c|)$$

The three scaled MAD is obtained by multiplying the MAD by a coefficient c defined as follows:

$$c = -\frac{1}{\sqrt{2} * erfc^{-1}(3/2)} \approx 1.4826$$

Where $erfc^{-1}$ represents the Inverse Complementary Error Function which is defined as

$$erfc^{-1}\big(erfc(x)\big) = x$$

The Complementary Error Function ($erfc$) is given by the following relation

$$erfc(x) = 1 - \mathrm{erf}(x)$$

Where $\mathrm{erf}(x)$ represents the Error Function of the argument x

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt =$$

-

$$Median(\{X\}) = X_c$$

$$\left\{\vec{X}\right\} = sort_{smallest-gretest}\{X\} = X_1, X_2, \dots X_n$$

$$X_i < X_{i+1} \quad \forall\, i \in \mathbb{N}$$

$$X_c = \begin{cases} X_{(n+1)/2} & for\ n\ odd \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & for\ n\ even \end{cases}$$

MAD is defined as the median of the absolute difference between a vector Y and its relative median value Y_c.

$$MAD = Median(|\{Y\} - Y_c|)$$

The three scaled MAD is obtained by multiplying the MAD by a coefficient c defined as follows:

$$c = -\frac{1}{\sqrt{2} * erfc^{-1}(3/2)} \approx 1.4826$$

Where $erfc^{-1}$ represents the Inverse Complementary Error Function which is defined as

$$erfc^{-1}\big(erfc(x)\big) = x$$

The Complementary Error Function $(erfc)$ is given by the following relation

$$erfc(x) = 1 - \mathrm{erf}(x)$$

Where $\mathrm{erf}(x)$ represents the Error Function of the argument x

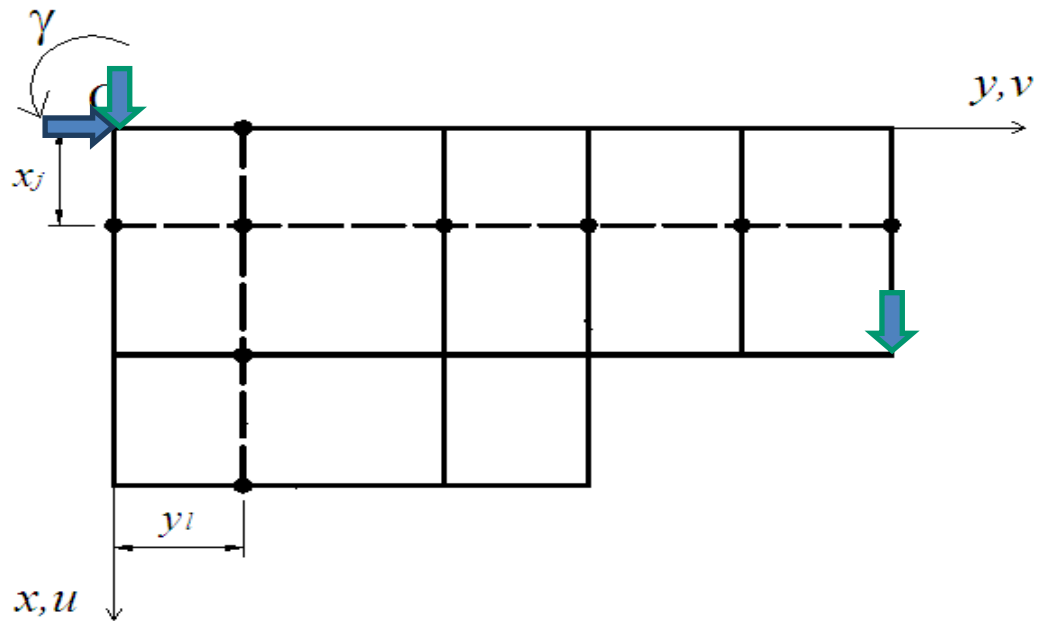$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt =$$

-

Figure 4-1 – disposition of the accelerometers

- The stiffness matrix is the following

$$\begin{bmatrix} 1.8900 & -0.8630 & -0.0576 & 0.0000 & 0.0000 & 0.0000 & -13.4000 & 6.0900 & -0.4060 \\ -0.8630 & 1.3600 & -0.5740 & 0.0000 & 0.0000 & 0.0000 & 6.0900 & -9.5600 & 4.0500 \\ -0.0576 & -0.5740 & -0.4410 & 0.0000 & 0.0000 & 0.0000 & -0.4060 & 4.0500 & -3.1100 \\ 0.0000 & 0.0000 & 0.0000 & 1.9300 & -0.7780 & 0.0022 & 7.4000 & -3.1100 & 0.0293 \\ 0.0000 & 0.0000 & 0.0000 & -0.7780 & 1.3300 & -0.5200 & -3.1100 & 5.2500 & -2.0400 \\ 0.0000 & 0.0000 & 0.0000 & 0.0022 & -0.5200 & 0.5110 & 0.0293 & -2.0400 & 2.0300 \\ -13.4000 & 6.0900 & -0.4060 & 7.4000 & -3.1100 & 0.0293 & 21.1000 & -9.6400 & 0.4700 \\ 6.0900 & -9.5600 & 4.0500 & -3.1100 & 5.2500 & -2.0400 & -9.6400 & 15.5000 & -6.3700 \\ -0.4060 & 4.0500 & -3.1100 & 0.0293 & -2.0400 & 2.0300 & 0.4700 & -6.3700 & 5.4000 \end{bmatrix} 10^9 \ [Nm^2]$$

In accordance with the reference system showed in

$$Median(\{X\}) = X_c$$

$$\{\vec{X}\} = sort_{smallest-gretest}\{X\} = X_1, X_2, \dots X_n$$

$$X_i < X_{i+1} \quad \forall \, i \in \mathbb{N}$$

$$X_c = \begin{cases} X_{(n+1)/2} & for \ n \ odd \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & for \ n \ even \end{cases}$$

MAD is defined as the median of the absolute difference between a vector Y and its relative median value Y_c.

$$MAD = Median(|\{Y\} - Y_c|)$$

The three scaled MAD is obtained by multiplying the MAD by a coefficient c defined as follows:

$$c = -\frac{1}{\sqrt{2} * erfc^{-1}(3/2)} \approx 1.4826$$

Where $erfc^{-1}$ represents the Inverse Complementary Error Function which is defined as

$$erfc^{-1}\big(erfc(x)\big) = x$$

The Complementary Error Function ($erfc$) is given by the following relation

$$erfc(x) = 1 - \mathrm{erf}(x)$$

Where $\mathrm{erf}(x)$ represents the Error Function of the argument x

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt =$$

, the previous values are referred to the following stiffness

$$\begin{bmatrix} k_{xx} & k_{xy} & k_{x\Theta} \\ k_{xy} & k_{yy} & k_{y\Theta} \\ k_{x\Theta} & k_{y\Theta} & k_{\Theta\Theta} \end{bmatrix}$$

- The mass matrix is the following

$$\begin{bmatrix}
199926 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1262530 & 0.0000 & 0.0000 \\
0.0000 & 173425 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1095179 & 0.0000 \\
0.0000 & 0.0000 & 200717 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1267529 \\
0.0000 & 0.0000 & 0.0000 & 199926 & 0.0000 & 0.0000 & 762350 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 173425 & 0.0000 & 0.0000 & 659897 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 200717 & 0.0000 & 0.0000 & 772463 \\
-1262530 & 0.0000 & 0.0000 & 762350 & 0.0000 & 0.0000 & 19742644 & 0.0000 & 0.0000 \\
0.0000 & -1095179 & 0.0000 & 0.0000 & 659897 & 0.0000 & 0.0000 & 16901864 & 0.0000 \\
0.0000 & 0.0000 & -1267529 & 0.0000 & 0.0000 & 772463 & 0.0000 & 0.0000 & 20166960
\end{bmatrix}[kg]$$

In accordance with the reference system showed in

$$Median(\{X\}) = X_c$$

Error! No text of specified style in document.-8

$$\left\{\vec{X}\right\} = sort_{smallest-gretest}\{X\} = X_1, X_2, \dots X_n$$

$$X_i < X_{i+1} \quad \forall\, i \in \mathbb{N}$$

Error! No text of specified style in document.-9

$$X_c = \begin{cases} X_{(n+1)/2} & for\ n\ odd \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & for\ n\ even \end{cases}$$

Error! No text of specified style in document.-10

MAD is defined as the median of the absolute difference between a vector Y and its relative median value Y_c.

$$MAD = Median(|\{Y\} - Y_c|)$$

Error! No text of specified style in document.-11

The three scaled MAD is obtained by multiplying the MAD by a coefficient c defined as follows:

$$c = -\frac{1}{\sqrt{2} * erfc^{-1}(3/2)} \approx 1.4826$$

Where $erfc^{-1}$ represents the Inverse Complementary Error Function which is defined as

$$erfc^{-1}\big(erfc(x)\big) = x$$

The Complementary Error Function ($erfc$) is given by the following relation

$$erfc(x) = 1 - \mathrm{erf}(x)$$

Where $\mathrm{erf}(x)$ represents the Error Function of the argument x

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt =$$

, the previous values are referred to the following mass

$$\begin{bmatrix} M_{xx} & M_{xy} & M_{x\Theta} \\ M_{xy} & M_{yy} & M_{y\Theta} \\ M_{x\Theta} & M_{y\Theta} & M_{\Theta\Theta} \end{bmatrix}$$

- The damping ratios associated to each mode are

$$dr_1 = 0.20\%$$
$$dr_2 = 0.60\%$$
$$dr_3 = 1.30\%$$
$$dr_4 = 3.50\%$$

$$dr_5 = 0.80\%$$

$$dr_6 = 1.50\%$$

$$dr_7 = 0.70\%$$

$$dr_8 = 2.90\%$$

$$dr_9 = 2.00\%$$

The geometric case is now known.

4.2 Modal parameters

Once the mass matrix, the stiffness matrix and the damping ratios are known, the modal parameters can be computed.

4.2.1 Mode 1

$Frequency$: $0.7278\ Hz$

$Damping\ ratio$: 0.20%

$$Modal\ shape: \begin{Bmatrix} 0.41250 \\ 0.6716 \\ 1 \\ -0.1016 \\ -0.1532 \\ -0.2268 \\ 0.0268 \\ 0.0393 \\ 0.0571 \end{Bmatrix}$$

4.2.2 Mode 2

Frequency: 4.1980 Hz

Damping ratio: 0.60%

$$\text{Modal shape:} \begin{Bmatrix} -0.2416 \\ -0.3263 \\ -0.4572 \\ 0.3073 \\ 0.6411 \\ 1 \\ -0.0735 \\ -0.1461 \\ -0.2256 \end{Bmatrix}$$

4.2.3 Mode 3

Frequency: 4.8239 Hz

Damping ratio: 1.30%

$$\text{Modal shape:} \begin{Bmatrix} -0.0213 \\ -0.0228 \\ -0.0254 \\ -0.2731 \\ -0.6263 \\ -1 \\ -0.0043 \\ -0.0068 \\ -0.0092 \end{Bmatrix}$$

4.2.4 Mode 4

Frequency: 11.8675 Hz

Damping ratio: 3.50%

$$\text{Modal shape:} \begin{Bmatrix} -1 \\ -0.9593 \\ 0.9555 \\ 0.1600 \\ 0.1598 \\ -0.1759 \\ -0.0362 \\ -0.0345 \\ 0.0410 \end{Bmatrix}$$

4.2.5 Mode 5

Frequency: 12.0137 *Hz*

Damping ratio: 0.80%

$$\text{Modal shape: } \begin{Bmatrix} 0.0390 \\ 0.0142 \\ -0.0318 \\ 0.9738 \\ 1 \\ -0.8089 \\ 0.0012 \\ -0.0019 \\ -0.0009 \end{Bmatrix}$$

4.2.6 Mode 6

Frequency: 12.5996 *Hz*

Damping ratio: 1.50%

$$\text{Modal shape: } \begin{Bmatrix} -1 \\ -0.8478 \\ 0.8843 \\ 0.9279 \\ 0.8468 \\ -0.7488 \\ -0.2507 \\ -0.2199 \\ 0.2000 \end{Bmatrix}$$

4.2.7 Mode 7

Frequency: 17.7984 *Hz*

Damping ratio: 0.70%

$$\text{Modal shape: } \begin{Bmatrix} 0.9751 \\ -1 \\ 0.1694 \\ -0.3543 \\ 0.3574 \\ -0.0621 \\ 0.0592 \\ -0.0650 \\ -0.0094 \end{Bmatrix}$$

4.2.8 Mode 8

Frequency: 18.3285 *Hz*

Damping ratio: 2.90%

$$
Modal\ shape:\ \begin{Bmatrix} -0.2011 \\ 0.2207 \\ -0.0403 \\ -1 \\ 0.9095 \\ -0.2261 \\ -0.0044 \\ 0.0081 \\ -0.0007 \end{Bmatrix}
$$

4.2.9 Mode 9

Frequency: 20.2295 *Hz*

Damping ratio: 2.00%

$$
Modal\ shape:\ \begin{Bmatrix} 0.4842 \\ -0.6372 \\ 0.0990 \\ -0.9432 \\ 1 \\ -0.2443 \\ 0.2837 \\ -0.3039 \\ 0.0740 \end{Bmatrix}
$$

4.3 Excitation

In order to obtain an excitation similar to the environmental one, the input signal is generated using the function *"randn"* which is implemented in MATLAB.

In this way, the signal assumes a meaning of *white noise.* White noise is a random signal with a constant power spectral density and without a periodicity with the time. The values which the white noise assumes can be generated by a standard normal distribution ($\mu = 0$, $\sigma = 1$).

In first, in order to define the input signal, a time vector t is defined as follows:

$$t = \left\{ \begin{array}{c} 0 \\ \dfrac{1}{fs} \\ \dfrac{2}{fs} \\ \vdots \\ \dfrac{n}{fs} \end{array} \right\}$$

Where fs is the sampling frequency and $n - 1$ is the number of elements which compose the time vector. In other words, the value of *n* is linked to the time of sampling Ts with the following relation:

$$Ts = \frac{n}{fs}$$

In this study the sampling frequency is considered equal to $200\ Hz$ and the time of sampling equal to $180\ s$.

The time vector assumes the following values:

$$t[s] = \left\{ \begin{array}{c} 0 \\ 0.005 \\ 0.010 \\ 0.015 \\ 0.020 \\ \vdots \\ 180 \end{array} \right\}$$

48

In order to obtain values of acceleration similar to the ones which can be recorded in a real measurement, the random values of the white noise are scaled with a coefficient equal to 0.005. This means assuming the following variance:

$$\sigma = 0.005$$

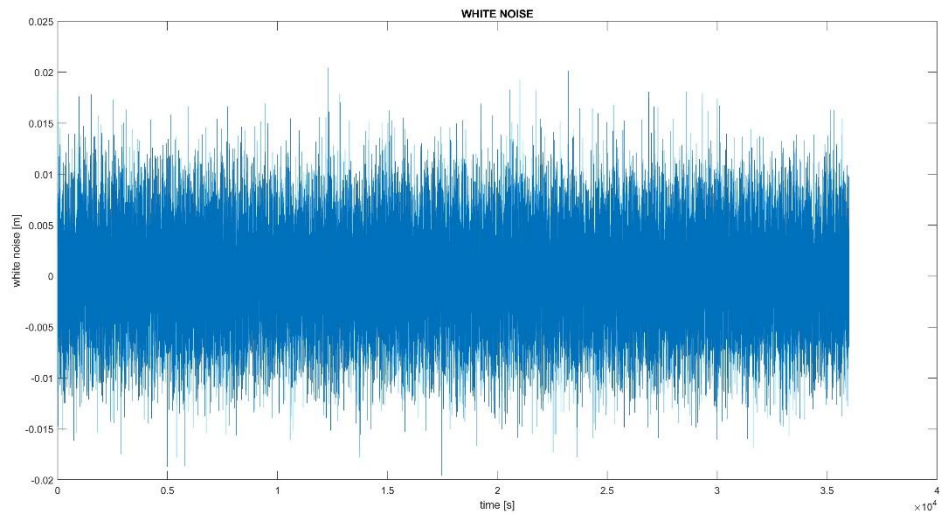Then, the white noise can be graphed as showed in Figure 4-2.

Figure 4-2 – White Noise example ($\mu = 0, \sigma = 0.005$)

Therefore, in order to excite each degree of freedom with a different intensity, a track array is considered as follows:

$$track = \begin{Bmatrix} 0.6761 \\ 0.1690 \\ 0.1690 \\ -0.5070 \\ 0.3380 \\ 0.3380 \\ -0.0101 \\ -0.0101 \\ 0.0034 \end{Bmatrix}$$

The signal is applied to the origin O of the axes x, y (

$$Median(\{X\}) = X_c$$

$$\{\vec{X}\} = sort_{smallest-gretest}\{X\} = X_1, X_2, \dots X_n$$

$$X_i < X_{i+1} \quad \forall\, i \in \mathbb{N}$$

$$X_c = \begin{cases} X_{(n+1)/2} & for\ n\ odd \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & for\ n\ even \end{cases}$$

MAD is defined as the median of the absolute difference between a vector Y and its relative median value Y_c.

$$MAD = Median(|\{Y\} - Y_c|)$$

The three scaled MAD is obtained by multiplying the MAD by a coefficient c defined as follows:

$$c = -\frac{1}{\sqrt{2} * erfc^{-1}(3/2)} \approx 1.4826$$

Where $erfc^{-1}$ represents the Inverse Complementary Error Function which is defined as

$$erfc^{-1}\big(erfc(x)\big) = x$$

The Complementary Error Function ($erfc$) is given by the following relation

$$erfc(x) = 1 - \text{erf}(x)$$

Where $\text{erf}(x)$ represents the Error Function of the argument x

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt =$$

) considering the previous track array.

4.4 Displacements, Velocities and Accelerations

Once the geometry of the problem and the excitation are established, the computing of the displacements of O is possible for each floor thanks to the following integral

$$u_i = -\left(\frac{\Psi_i}{\sqrt{\Psi_i' M \Psi_i}}\right)' M\, T_r \frac{1}{\omega_i \sqrt{1 - (dr_i)^2}} \int_0^T wn_i(t) \cdot (e^{-dr_i\, \omega_i\, t} \cdot \sin\left(\omega_i \sqrt{1 - (dr_i)^2}\, t\right)) dt$$

4-1

51

Where

- Ψ_i, ω_i, dr_i are the eigenvector, the angular frequency and the damping ratio associated to the i-mode
- M is the mass matrix
- T_r is the track vector
- $wn_i(t)$ is the white noise, the acceleration which the structure is subjected to
- t is the time
- T is the duration of the excitation

In this case the time is discrete and for this reason the white noise is defined for each time interval $1/fs$ where fs is the sampling frequency.

As a consequence, the integral is transformed in a summation.

The displacements of each accelerometer are computed considering the floor as a rigid body.

Velocities and accelerations are obtained deriving the displacements with the time.

At this point, the output signal is known for each accelerometer.

4.5 Noise

The accelerations computed for each sensor are free of any noise. In order to create a model close to what is expected of reality, different levels

$$signal\ disturbed = signal + pp * (standard\ deviation(signal)) \qquad \textit{4-2}$$

of noise are added to the signals with the following approach.

Where pp indicates the percentage of noise level.

The percentages of noise which are adopted in this chapter are: 1%, 5%, 10%, 20%, 30%, 50%, 75%, 100%, 150%, 200%. These are used with different values in function of the aim of study.

4.6 Influence of Model Order and Hankel Matrix dimension

In this chapter different values of model order and dimensions of the Hankel Matrix are considered in order to understand the influence of these parameters on the mode identification. Furthermore, percentages of level noise are adopted to give more credibility to the signals which are processed by SSI.

In order to quantify the performance of the Automatic Operational Modal Analysis, two entities are used as significant parameters: *Precision* and *Recall.*

4.6.1 Precision and Recall

Precision and recall are the measures used in the information retrieval domain to quantify how well an information retrieval system retrieves the relevant documents requested by a user [15].

Processing the signals with SSI and, consequently, applying the proposed method to the results of SSI, a certain number of modes are identified. These modes may or may not be representative of the modes of the numerical case. Furthermore, the number of identified modes can be different from the number of modes which represent the structure - in this case nine. This evidence may be summarised with the following definitions:

- *True Positive* (TP): mode identified by the algorithm which matches a real mode of the structure.
- *False Positive* (FP): mode identified by the algorithm which does not match any mode of the structure.
- *Precision*: ratio between the number of modes identified as TP and the total number of modes identified by the algorithm (TP+FP).
- *Recall*: ratio between the number of modes identified as TP and the number of modes which are representatives of the structure.

Precision and recall can be two useful indices to estimate the performance of the proposed method and SSI. In fact, the AOMA is a process which elaborates the results of SSI and for this reason the quantity and the quality of the identified modes are strongly influenced by the efficiency of SSI.

Recall and precision are evaluated in function of frequency and eigenvector in two separated charts.

A mode is considered true positive in terms of frequency if the difference of f between the real mode and the identified mode is lower than $0.1\ Hz$. On the other hand, a mode is considered TP in terms of eigenvector if the *MAC* computed on the identified mode and the real mode is higher than 0.9. These hypotheses are summarised in the following table.

	TRUE POSITIVE	*FALSE POSITIVE*
Frequency	$\|f_r - f_i\| < 0.1\ Hz$	$\|f_r - f_i\| \geq 0.1\ Hz$
Eigenvector	$MAC(\Psi_i, \Psi_r) > 0.9$	$MAC(\Psi_i, \Psi_r) \leq 0.9$

Table 4-1 – True Positive and False Positive

With the previous consideration, precision and recall can be computed as follows:

$$Precision = \frac{nTP}{nTP + nFP}$$	*4-3*
$$Recall = \frac{nTP}{9}$$	*4-4*

Table 4-2 – Precision and Recall

Where nTP is the number of modes identified as *True Positive* and nFP is the number of modes identified as *False Positive.*

4.6.2 Model order, Hankel Matrix and Noise Level

At first, in this analysis, three different levels of noise are considered in order to understand how much of an influence this parameter can have on precision and recall of the AOMA. The three levels of noise are: 10%, 30% and 50%.

For each analysis a range of model orders is considered as follows:

- The minimum order is established equal to 10, which is higher than the number of degrees of freedom of the numerical case.
- The maximum order is variable. It starts from 40 and ends at 140 with an increasing value of 4.

The dimensions of the Hankel Matrix for each order are considered between 20 and 170 with intervals of 6.

In this way, Precision and Recall can be computed for the nodes of a grid of dimension 26-by-26, where 26 are both the maximum orders and the dimensions of the Hankel Matrix which are examined.

4.6.3 Results

The results of the analyses described in chapter 4.6 are summarised in the following figures.

Each chart is characterised by these features:

- The *x label* represents the maximum order (minimum order fixed equal to 10);
- The *y label* represents the dimension of the Hankel Matrix;
- A *colour legend* is showed in order to assign the values of precision or recall to each colour;
- The values of Precision and Recall are computed for each node of the grid;
- The values of precision and recall which are not in correspondence of a node are interpolated.

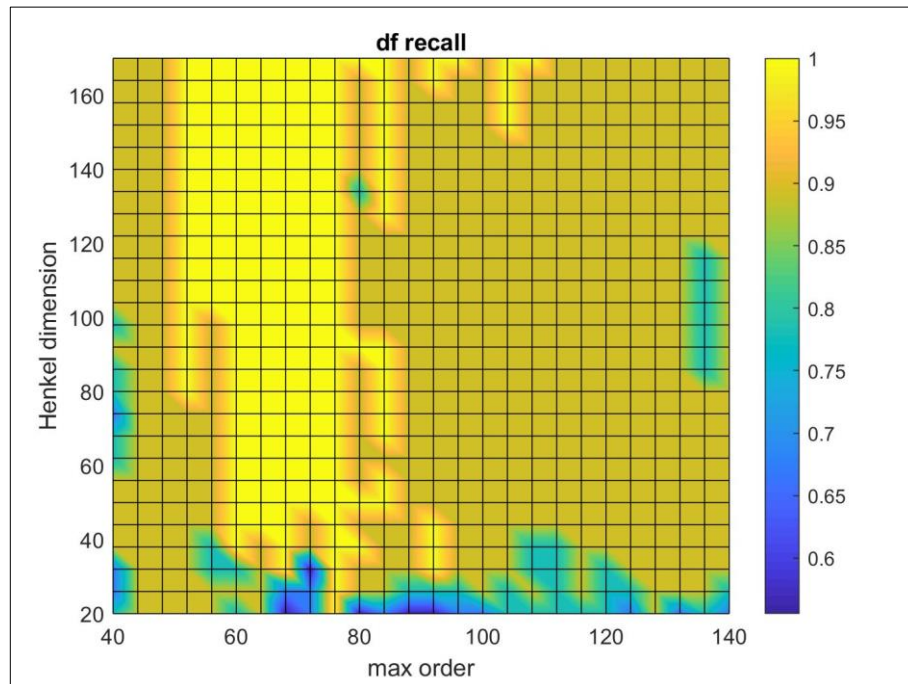### 4.6.3.1 Noise level 10%

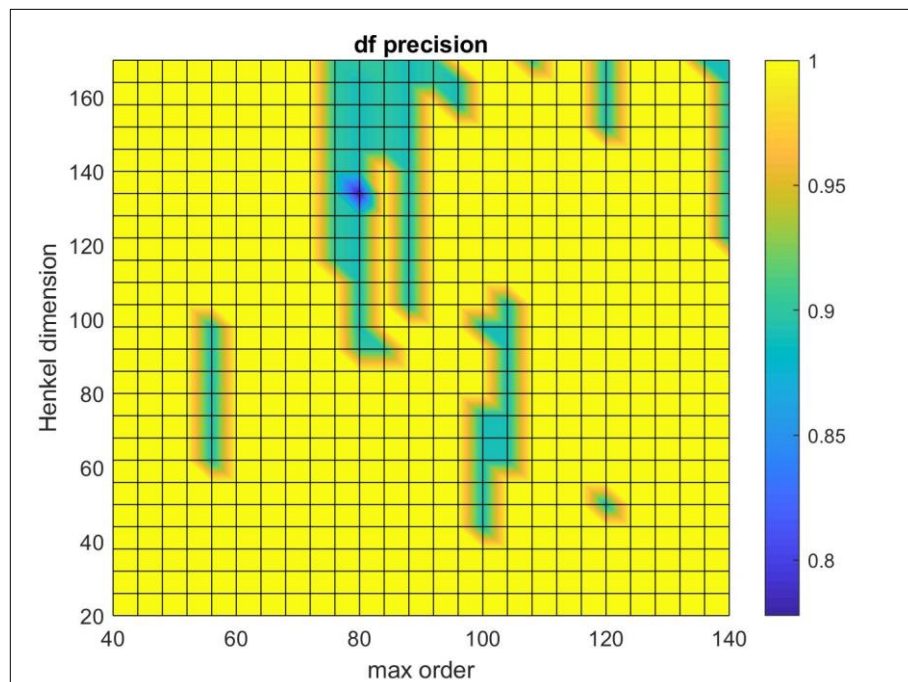*Figure 4-3 – recall in terms of difference of frequency (noise level 10%)*

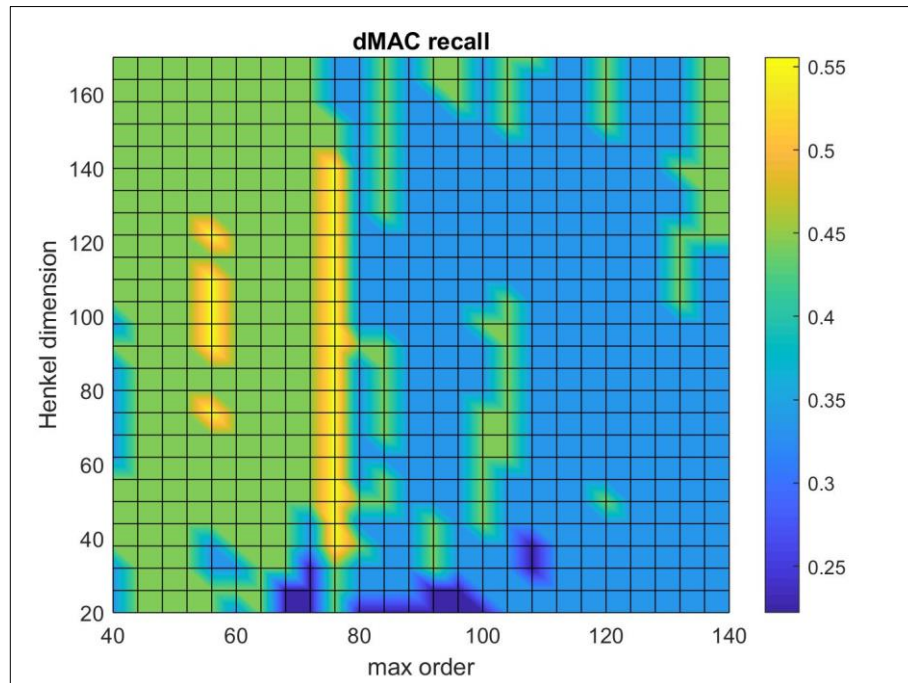*Figure 4-4 – precision in terms of difference of frequency (noise level 10%)*

56

Figure 4-5 – recall in terms of MAC (noise level 10%)

Figure 4-6 – precision in terms of MAC (noise level 10%)

57

4.6.3.2 Noise level 30%

Figure 4-7 - recall in terms of difference of frequency (noise level 30%)

Figure 4-8 – precision in terms of difference of frequency (noise level 30%)

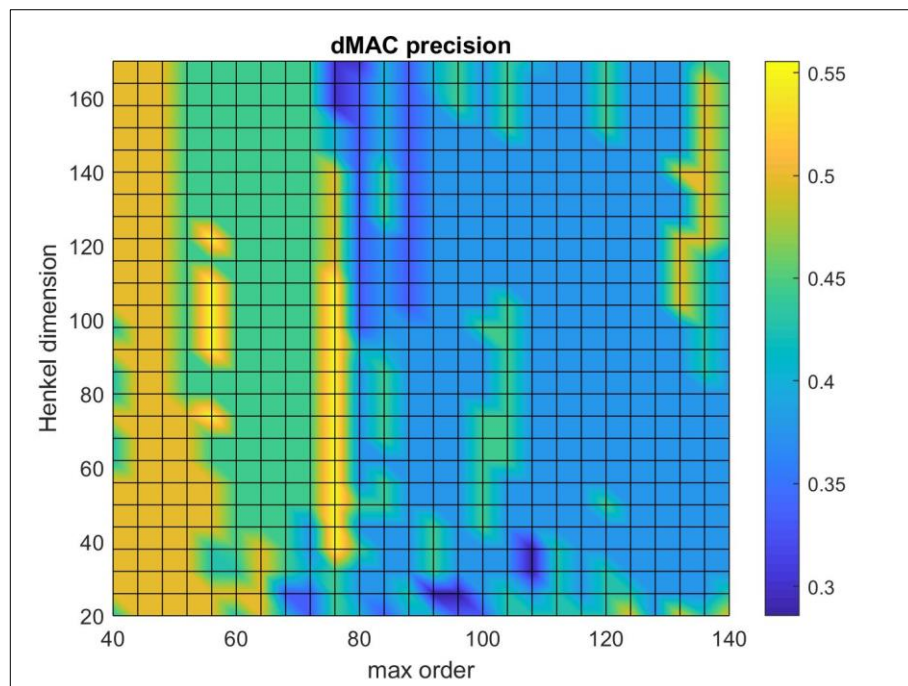*Figure 4-9 – recall in terms of MAC (noise level 30%)*

Figure 4-10 – precision in terms of MAC (noise level 30%)
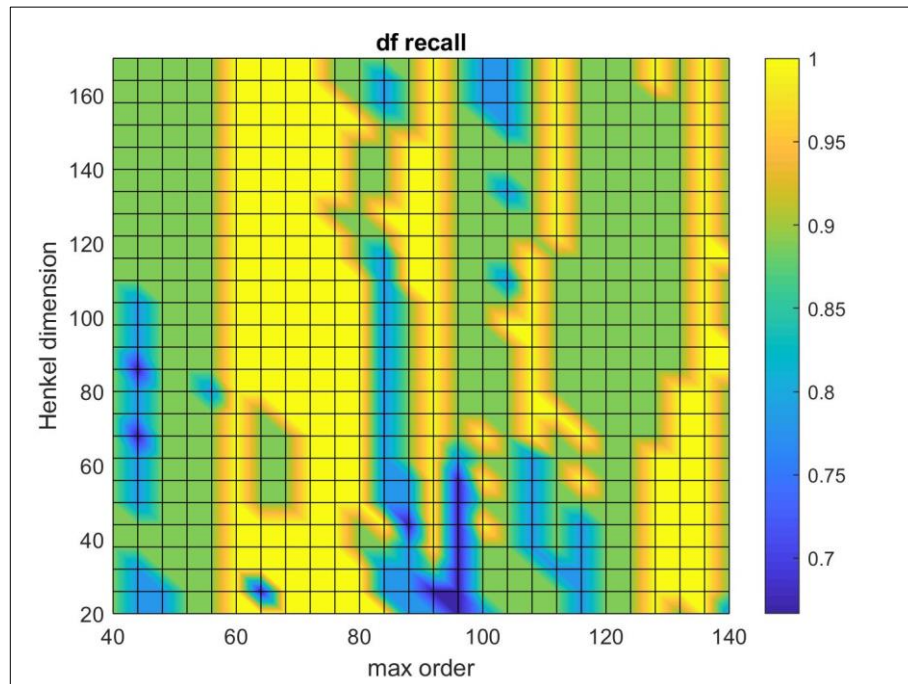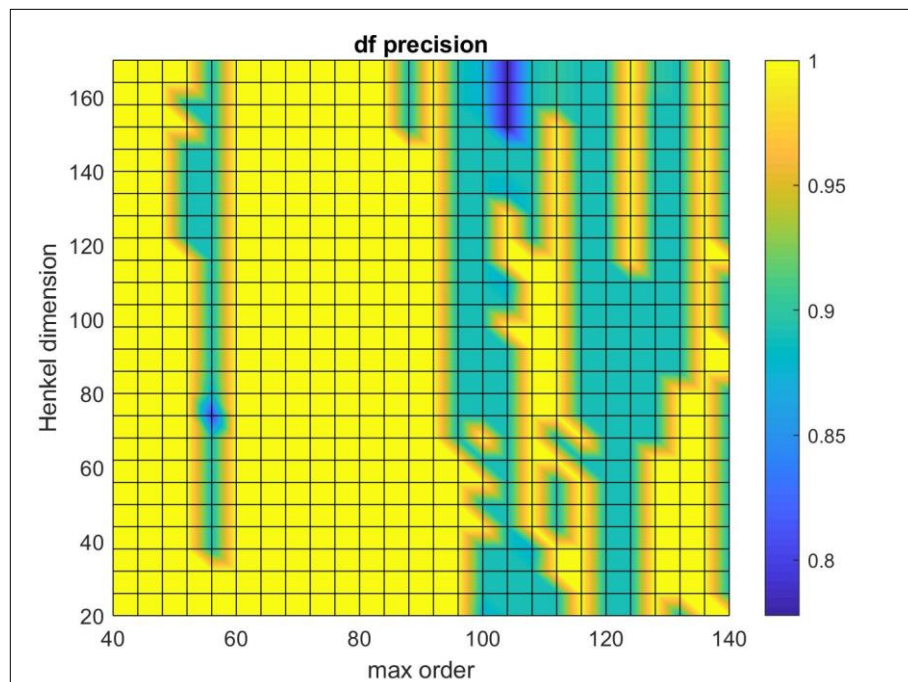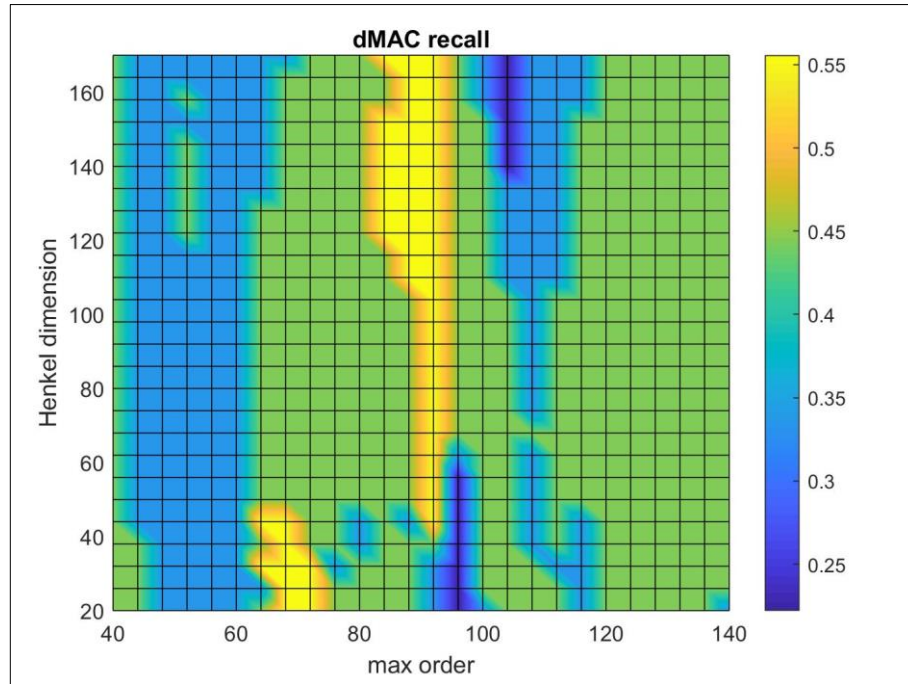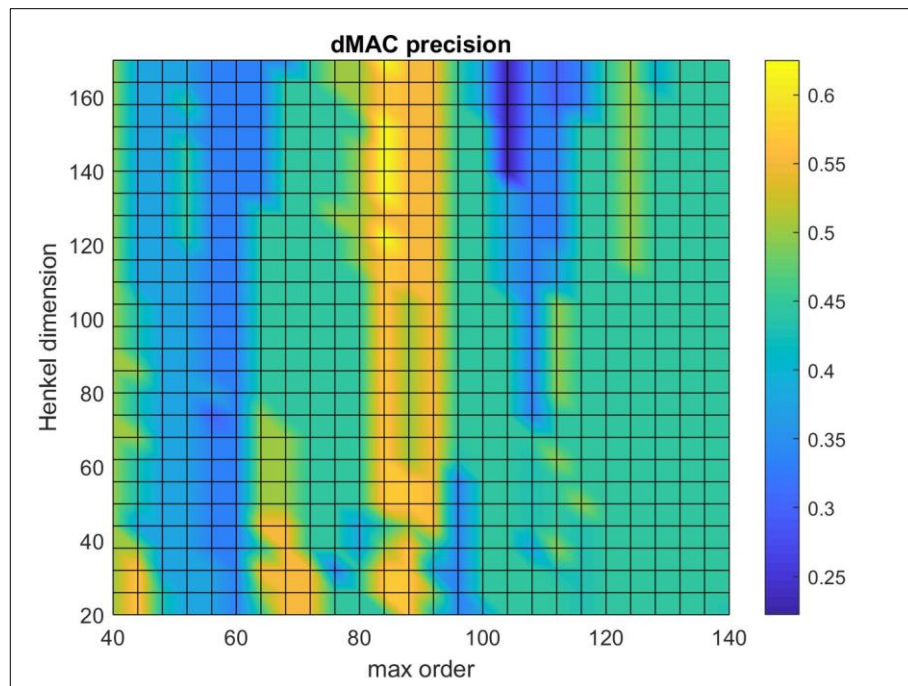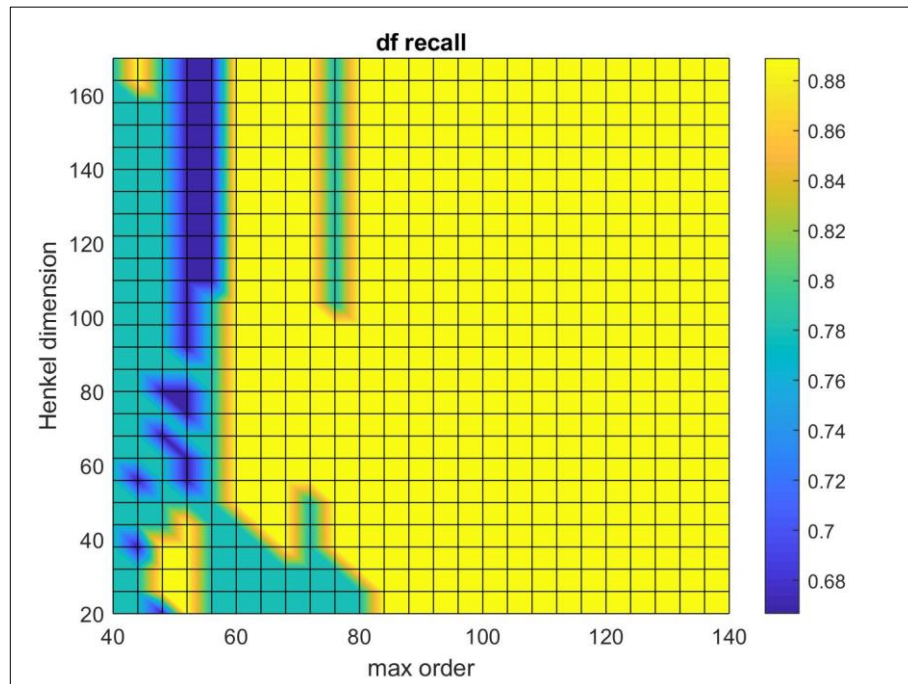
4.6.3.3 Noise level 50%

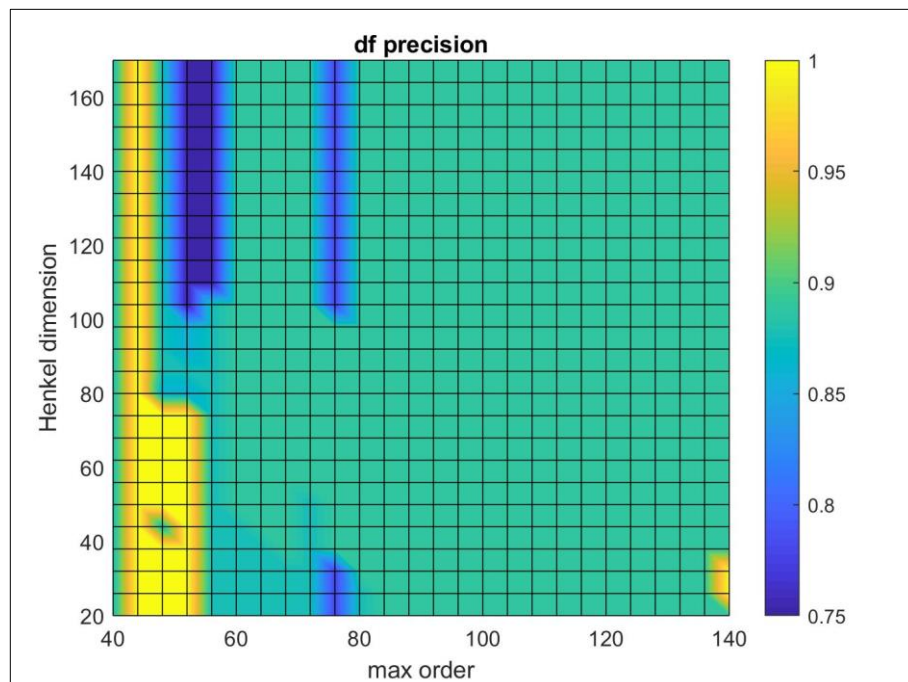*Figure 4-11 – recall in terms of difference of frequency (noise level 50%)*

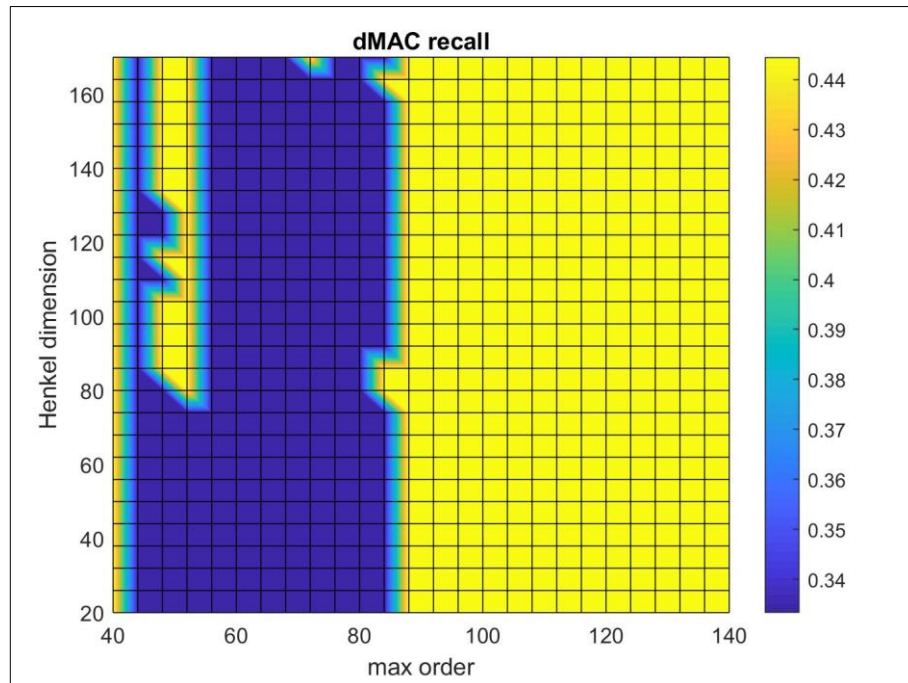*Figure 4-12 – precision in terms of difference of frequency (noise level 50%)*
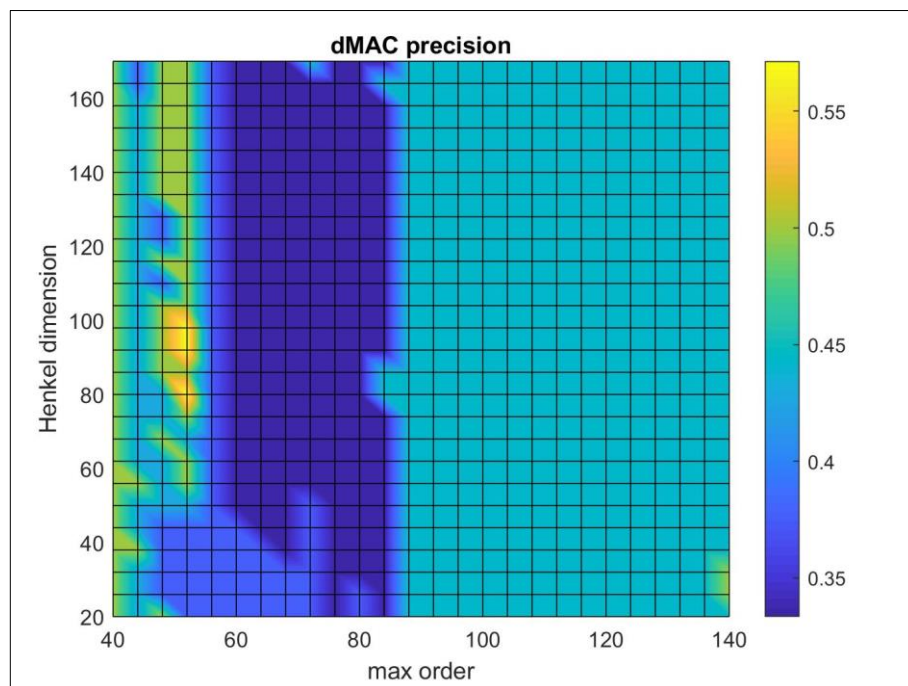
Figure 4-13 – recall in terms of MAC (noise level 50%)

Figure 4-14 – precision in terms of MAC (noise level 50%)

61

4.6.4 Conclusions

The proposed method has a high value of Recall for the identification of the frequencies. In fact, for a level noise lower than 50%, dimensions of the Hankel Matrix between 20 and 170, and considering different ranges of model order, the modes identified by the algorithm which are representative of the case of study are at least six.

The values of Precision for the frequency identification are generally higher than the values of Recall. This means that in each combination of noise level, Hankel Matrix, and model order, the algorithm has found a low number of *false positive* elements. In other words, the process works well to discard the poles identified by SSI which are spurious modes.

The values of Recall and Precision associated with the identification of the modal shapes (eigenvectors) are lower than the values associated with the frequency identification. This is evidence of the fact that the computing of the modal shapes is a process affected by a higher level of uncertainty. This problem is related to the identification of the eigenvectors by SSI-COV and it is not attributable to the AOMA. However, the modal shapes which are identified for each analysis are 4-5.

Having said that, it is necessary to make some considerations about the influences of the Hankel Matrix and the maximum model order on the AOMA. Considering the Figure 4-3, which represents the Recall of the frequency identification for a noise level equal to 10%, the higher values of Recall are showed in a range of model orders with minimum value of 10 and maximum value between 60 and 80. In those analyses, a number of 25-35 model orders are processed by the AOMA. Furthermore, as shown by the Figure 4-4 which is related to the Precision for the same analyses, for these values of model order the Precision is equal to 1. The dimension of the Hankel Matrix which guarantees a Recall of value 1 should be higher than 44. Van Overschee suggested the following dimension for the Hankel Matrix:

$$n(Hankel\ Matrix) = ent(\frac{2 * maximum\ order}{number\ of\ accelerometers})$$

In the case of study, the dimension of the Hankel Matrix computed with the previous equation, should be between 13 and 18. These values are far below the dimension of 44 which is found with the analyses explicated in this chapter. It is naturally important to analyse other numerical cases in order to reach clear conclusions which demonstrate that the dimension of the Hankel Matrix suggested by Van Overschee is underestimated. However, in the analyses which will be done in the following chapters, the dimension of the Hankel Matrix will be assumed in accordance with the following equation:

$$n(Hankel\ Matrix) = ent(\frac{\mathbf{6} * maximum\ order}{number\ of\ accelerometers}) \qquad 4\text{-}5$$

As a conclusion, the algorithm works well with the following parameters:

- dimension of the Hankel Matrix computed with the equation 4-5
- a range of model orders with minimum value equal to the number of degrees of freedom of the structure and maximum value higher than four times the minimum order, guaranteeing at least 25 model orders in each analysis.

4.7 Influence of noise

After the conclusions about the model order and the dimension of the Hankel matrix discussed in 4.6.4, in this chapter different noise levels are applied to the signal with the aim of understanding the influences on the results.

Precision and Recall are computed with the same meaning as explained in 4.6.1. The analysis is conducted using the following variables:

- Ranges of model order with:

    o   Minimum value equal to 10;

    o   Maximum value from 60 to 120 with intervals of 4.

- 10 different noise levels: 1%, 5%, 10%, 20%, 30%, 50%, 75%, 100%, 150% and 200%.

As explained in chapter 4.6, Precision and Recall are computed for frequency and modal shape.

4.7.1 Results

The results are summarised in the following charts.

Each chart is characterised by these features:

- The *x label* represents the maximum order (minimum order fixed equal to 10);
- The *y label* represents the noise level;
- A *colour legend* which assigns the values of precision or recall to each colour;
- The values of Precision and Recall are computed for each node of the grid;
- The values of Precision and Recall which are not in correspondence with a node are interpolated.
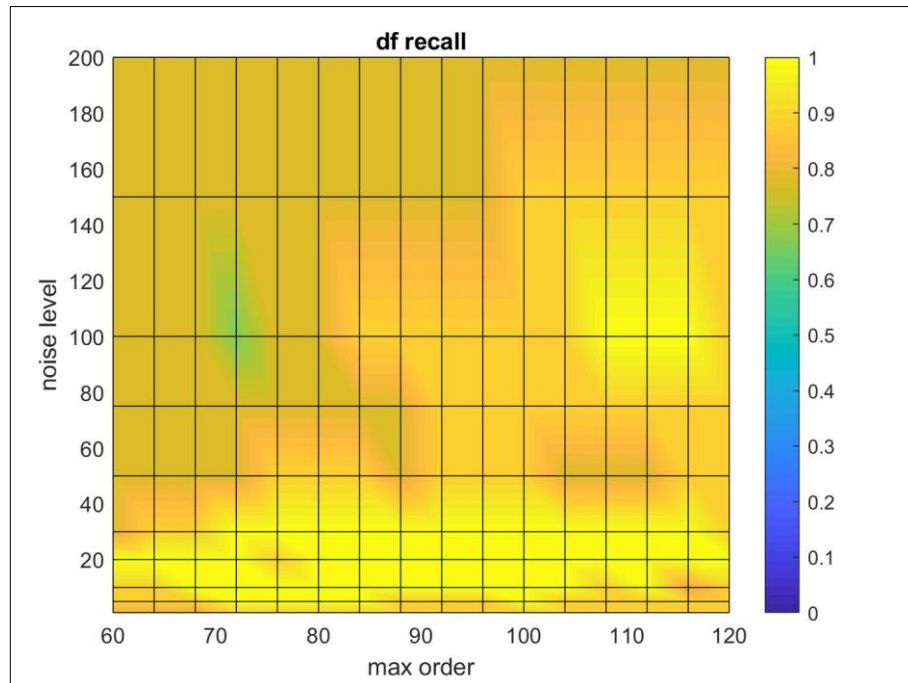
Figure 4-15 – recall in terms of difference of frequency (variable noise levels)
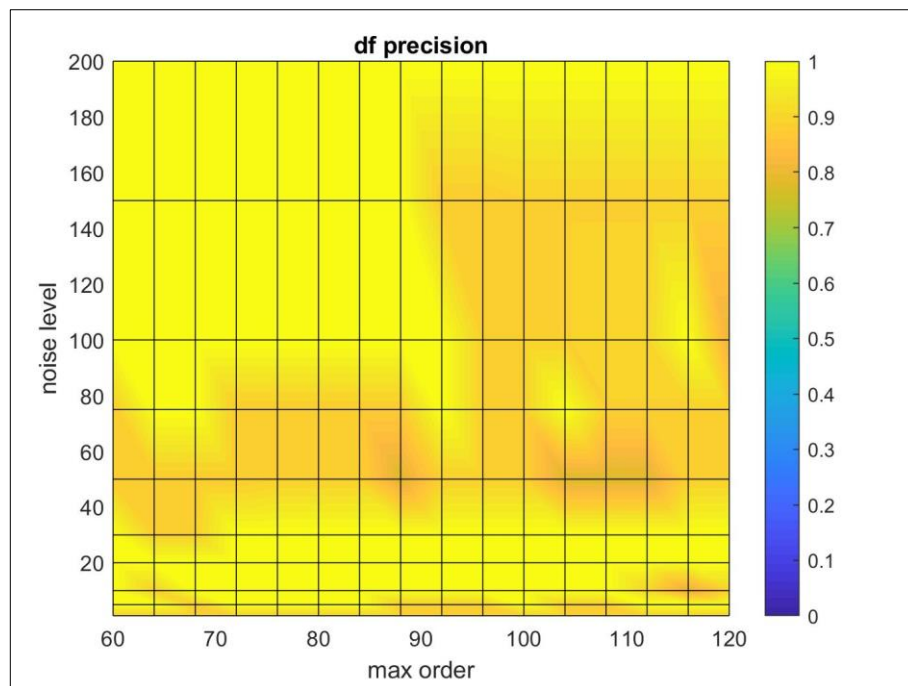

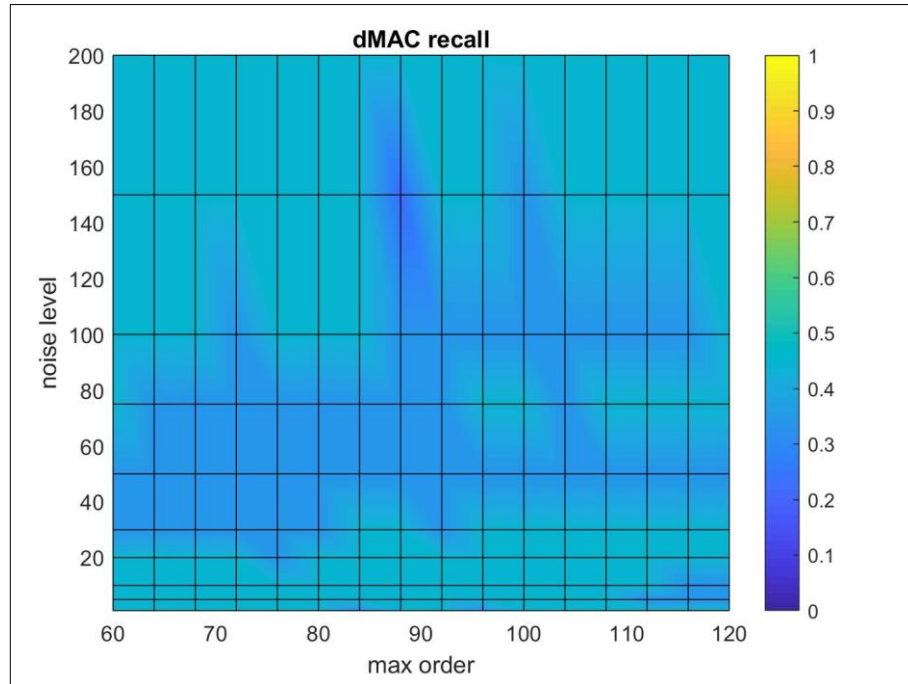*Figure 4-16 – precision in terms of difference of frequency (variable noise levels)*

Figure 4-17 – recall in terms of MAC (variable noise levels)
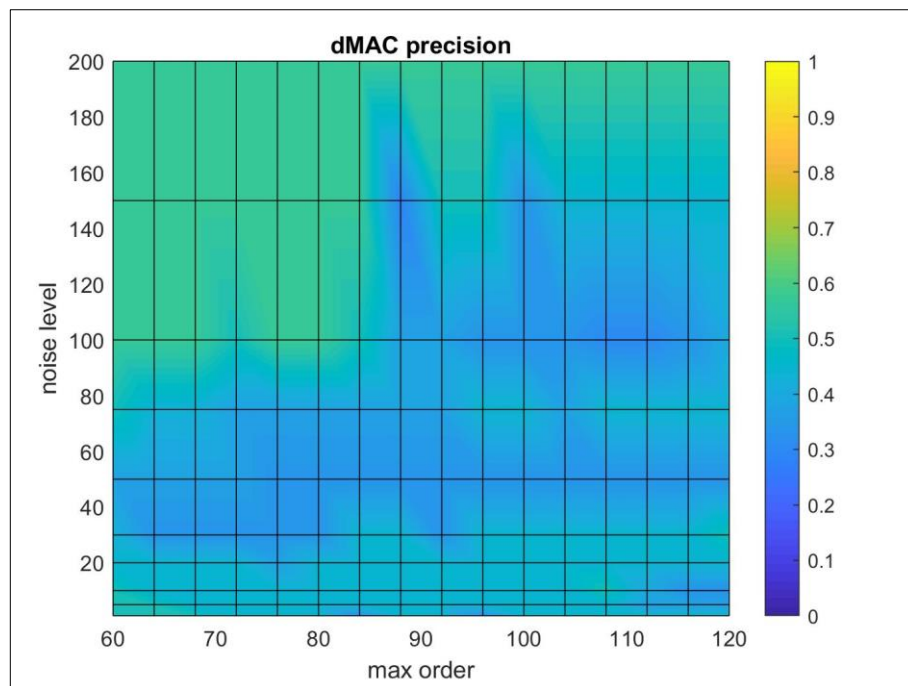

Figure 4-18 – precision in terms of MAC (variable noise levels)

4.7.2 Conclusions

As in the analysis explained in chapter 4.6, the values of Recall and Precision computed for the difference of frequency are higher than the values calculated for the MAC. This confirms that SSI is more powerful on the identification of frequencies than on the identification of the modal shapes.

The values of Recall for the difference of frequency are between 6/9 and 9/9. As a result, in each analysis at least six modes belonging to the numerical case are identified. The influence of the noise level is not relevant as it was conceivable. As shown in Figure 4-15, in order to obtain high values of Recall, the maximum model order should increase with the increase of the noise level. At the same time, high values of the maximum order are a probable reason of a lower precision (Figure 4-16) – this means the identification of false positive modes.

The true positive modes identified with the MAC analysis, are between 2 and 4.

In conclusion, the influence of the noise level is not relevant. The identification of the modal frequencies is exhaustive in most of the case.

4.8 Comparison between two different noise levels

In this chapter, two noise levels are applied to the signal in order to compare each step of the AOMA and to identify the main differences on the results.

The analyses are conducted considering the following features:

- Noise levels equal to 5% and 100%;
- Range of model orders between 10 and 72;
- Dimension of the Hankel matrix computed with the equation 4-5.

4.8.1 Application of Hard Criteria

The poles which present negative damping ratio or with values higher than 20% are discarded, as explained in 3.2.

Noise level	5%	100%
Number of mathematical poles	112	84

Table 4-3 – number of mathematical poles identified by Hard Criteria (Noise levels 5% vs. 100%)

4.8.2 Application of Soft Criteria

Following the steps explained in 3.3, the poles which are discarded by the soft criteria are showed with a black cross while the poles considered as stable are indicated with a red circle in the following figures.
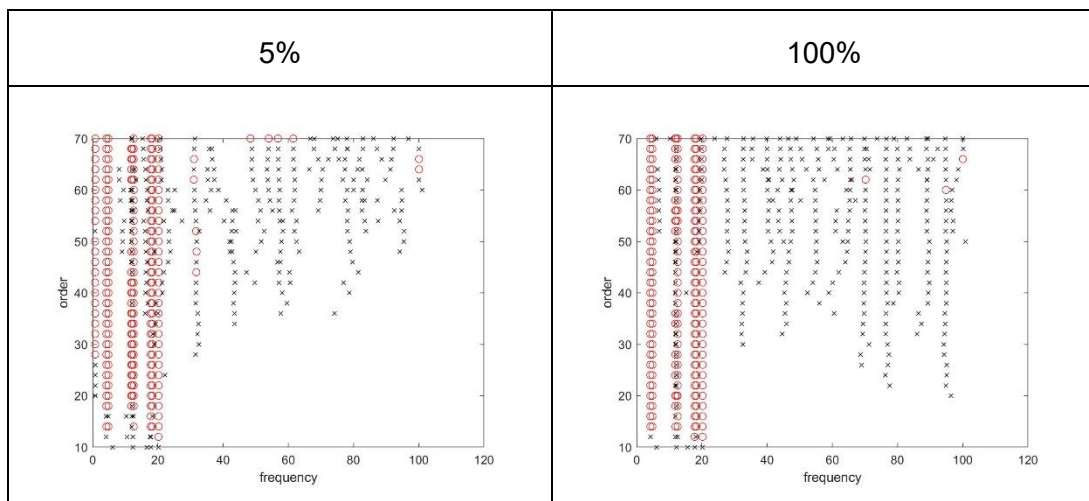
5%	100%

Table 4-4 – Stabilisation diagram for frequency (noise levels 5% vs. 100%)

Noise level	5%	100%
Number of mathematical poles	270	341

Table 4-5 – number of mathematical poles identified by Soft Criteria (Noise levels 5% vs. 100%)

The poles identified by SSI for a noise level equal to 5% show a higher stability than the poles identified for a noise level of 100%. This an evidence of the fact that the poles which are discarded in case of noise level 100% are 341 while, for a noise level equal to 5%, the poles which are removed are 270.

4.8.3 Clusters-Modes identification

The Clusters-Modes identification is conducted on the two different noise levels following the steps of chapter 3.4.

The results are showed in the following pictures where a colour is assigned to the poles belonging the same cluster.
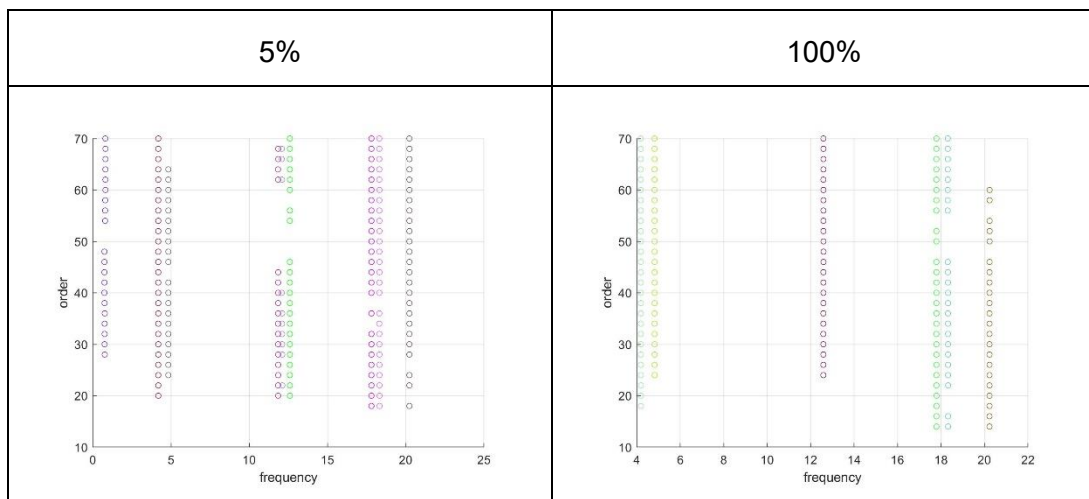
Table 4-6 – Clusters-Modes identification (Noise levels 5% vs. 100%)

The modes identified by the algorithm are 9 for a 5% of noise level and 6 for a noise level equal to 100%. The levels of precision and recall can be obtained from the Figure 4-15Figure 4-16Figure 4-17Figure 4-18.

4.8.4 Modal parameters

The frequencies identified in the two analyses are showed in Table 4-7 – frequencies identified for noise levels 5% and 100% Table 4-7.

Noise level	5%	100%	Target frequency
f_1	0.7836	-	0.7277
f_2	4.1919	4.1915	4.1979
f_3	4.8338	4.8250	4.8239
f_4	11.8482	-	11.8675
f_5	12.0964	-	12.0137
f_6	12.5982	12.5895	12.5996
f_7	17.8093	17.7895	17.7983
f_8	18.3176	18.3171	18.3285
f_9	20.2347	20.2305	20.2294

Table 4-7 – frequencies identified for noise levels 5% and 100%

We can summarise the results in the following table.

Noise level	5%	100%
Modes	9	6
True positive	9	6
False positive	0	0

Table 4-8 – true positive and false positive modes for frequency (noise levels 5% vs. 100%)

5 HELICOPTER BLADE

In this chapter, a helicopter blade is analysed as a simple experimental case in order to test the performance of the AOMA.

5.1 Description

The analysis is carried out on an Airbus Helicopters H135 bearingless main rotor blade.

The Blade is characterised by a length of 5.1 m and a variable cross section. The modal parameters were computed using a Finite Element Model.

9 triaxial acellerometers were placed on the blade but only the flapwise and edgewise dicrections and a channel in the blade axis direction were recorded with a total of 19 channels. The output signal was recorded with a sampling frequency of $2560\ Hz$. The measuerements were carried out in a controlled laboratory environment. The blade was excited for 600 seconds using a Random-on-Random (RoR) excitation waveform over a frequency range 0-100 Hz.

Figure 5-1 – Airbus Helicopters H135

5.2 AOMA application

SSI-Cov is set with these parameters:

- Range of model orders with minimum value equal to 20 and maximum value of 100 with intervals of 2.
- Dimension of the Hankel matrix computed with the equation 4-5

The steps which allow the identification of modal parameters are shown in the following chapters.

5.2.1 Hard Criteria

The number of poles identified for each model order are equal to half the dimension of the model order. Then, considering a range of model order between 20 and 100, the number of poles are:

$$\sum_{i=10}^{50} i = \frac{10 + 50}{2} \cdot 40 = 1200$$

The modes which show a dr negative or higher than 20% are **499**.

The residual poles (701) are then processed with Soft Criteria.

5.2.2 Soft Criteria

Following the steps explained in the chapter 3.3, the number of poles identified as stable or unstable are shown in Table 5-1.

Furthermore, the charts which represent the results of the 2-means clustering process are shown in Figure 5-2 to Figure 5-6 in order to clarify the influences of each comparison parameter on the identification of stable/unstable poles. Cluster 1 is representative of the stable poles while Cluster 2 groups the unstable poles.

Soft Criteria	Stable Poles	315
	Unstable Poles	386

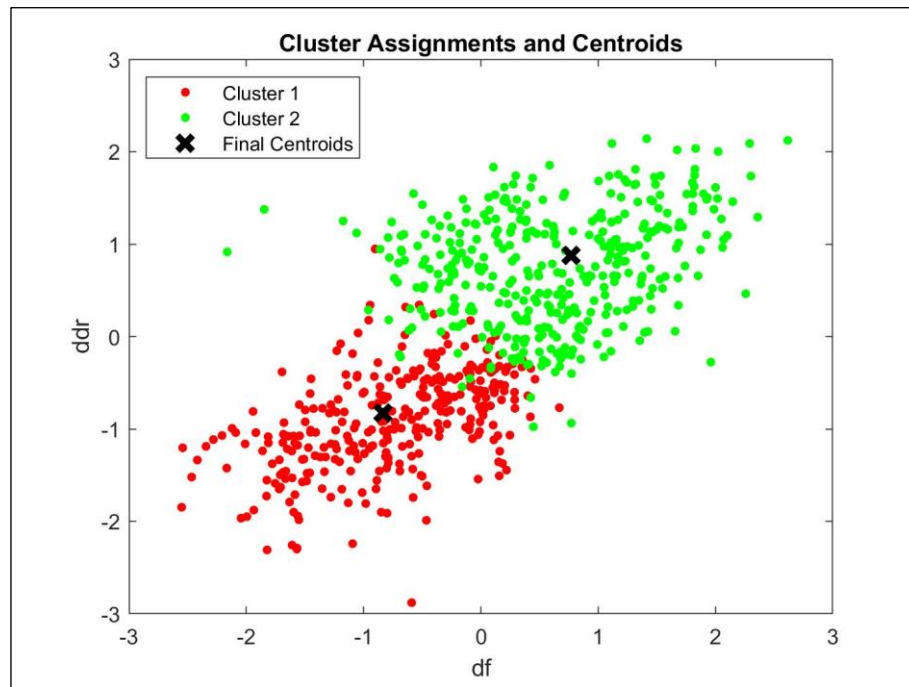*Table 5-1 – Poles identified as stable or unstable by Soft Criteria (Helicopter Blade)*

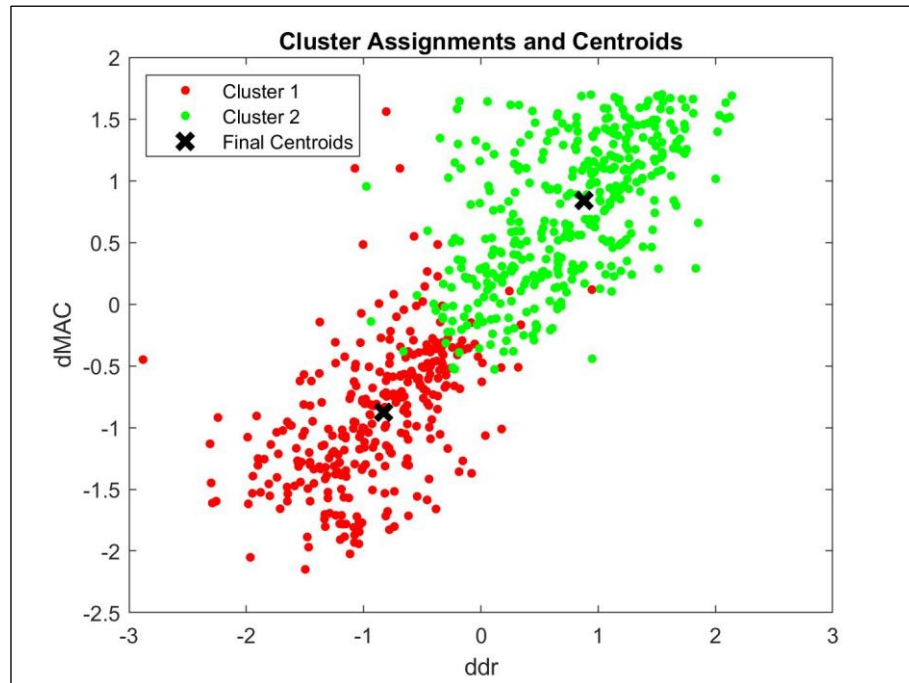*Figure 5-2 – stable/unstable poles (df-ddr)*

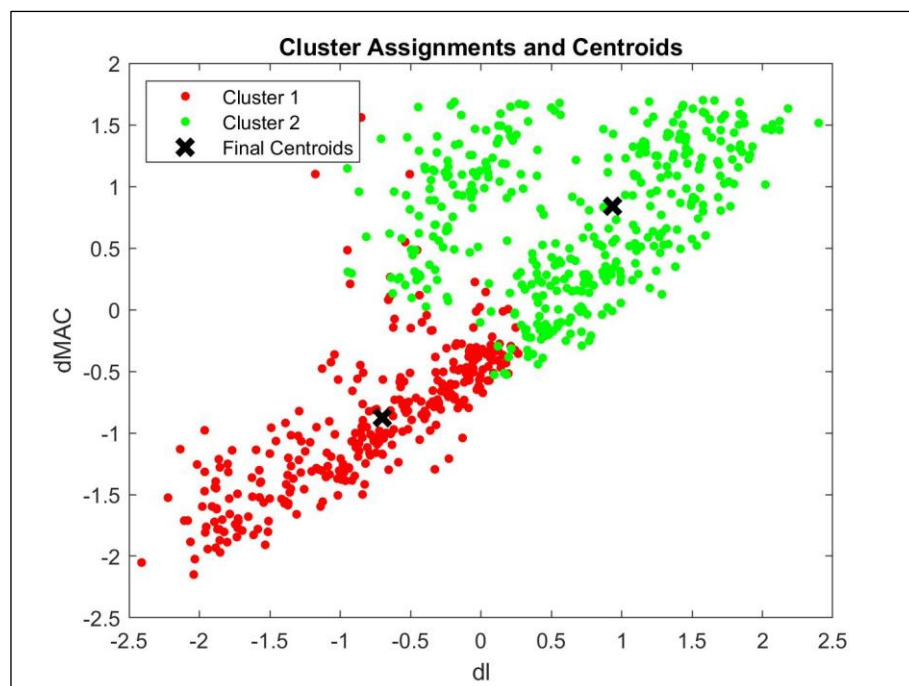*Figure 5-3 – stable/unstable poles (ddr-dMAC)*
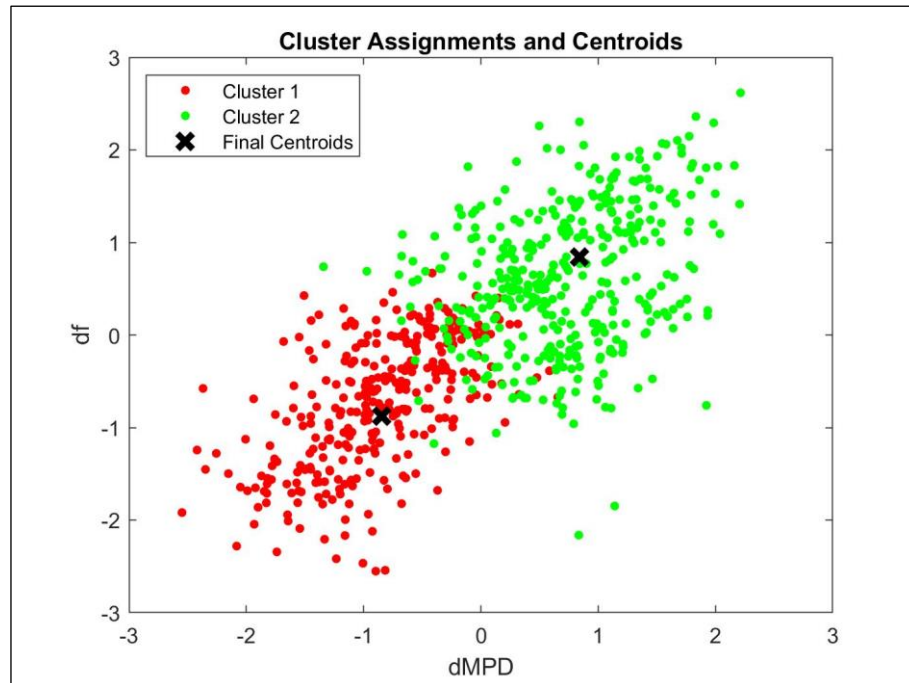

Figure 5-4 – stable/unstable poles (dλ-dMAC)

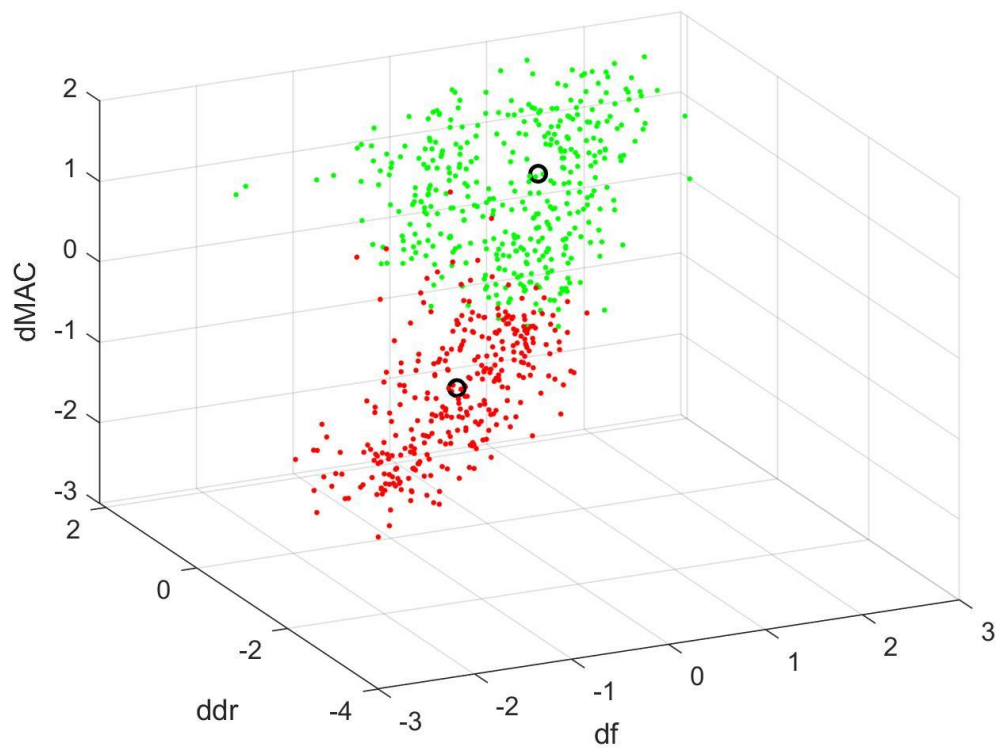*Figure 5-5 – stable/unstable poles (dMPD-df)*

Figure 5-6 – stable/unstable poles 3D (df-ddr-dMAC)

As it is shown by the previous charts, thanks to the transformations of comparison parameters explained in Data pre-processing (3.3.2), the contributes of df, ddr, $dMAC$, $dMPD$ are similar to each other. This evidence is highlighted by the shape of the two clusters which is almost unchanged despite the different points of view.

It's possible to show the results of the Soft Criteria in a *stabilisation diagram* with frequencies on the x label and model order on the y label (Figure 5-7). The stable poles are represented with red circles while the unstable poles with black crosses. The same chart can be seen by plotting the poles in function of damping ratio and frequency (Figure 5-8).
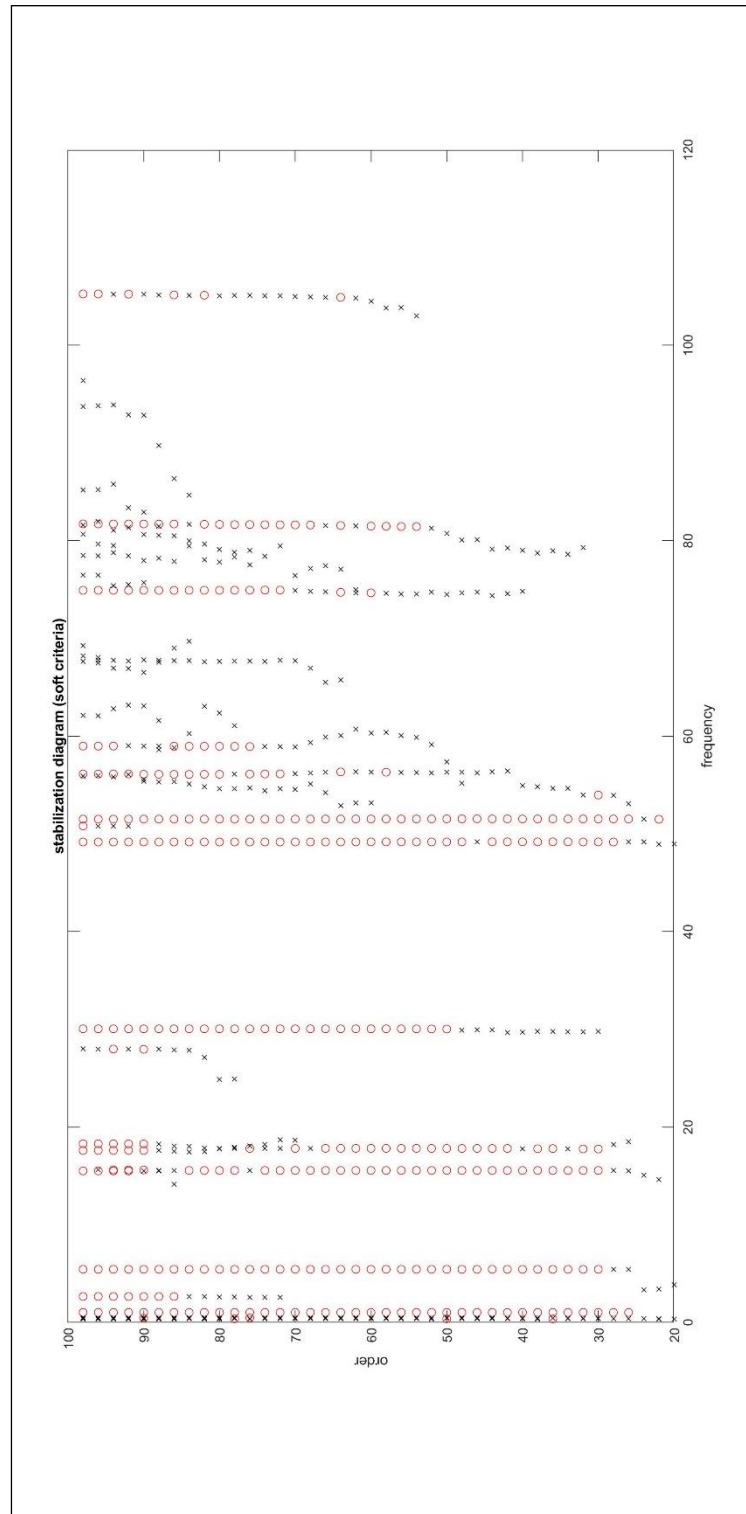
Figure 5-7 – stabilisation diagram (automatic detection of stable/unstable poles)
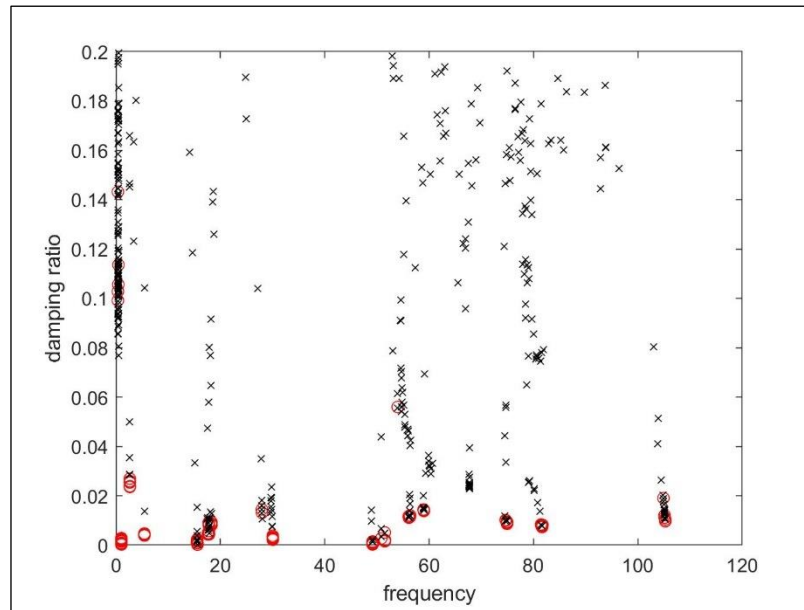
Figure 5-8 – stable/unstable poles (frequency vs. damping ratio)

Figure 5-7 and Figure 5-8 can be combined in order to obtain a 3D chart which shows the stable and unstable poles in function of frequency, damping ratio and model order (Figure 5-9).
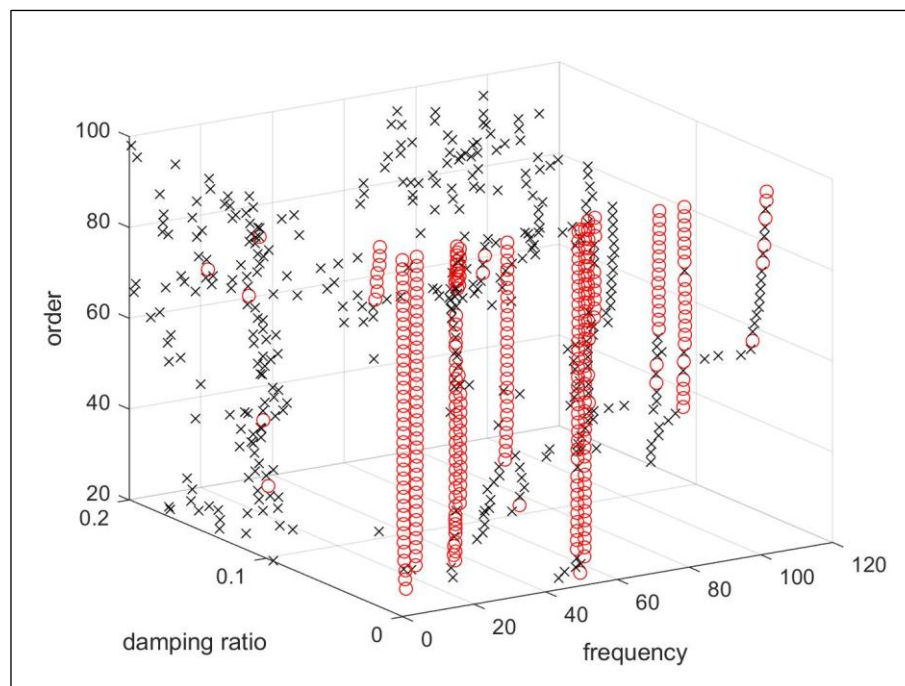
Figure 5-9 – 3D stabilisation diagram (f-dr)

78

It's not possible to clearly identify a model order which establishes frequency and damping ratio of the modes computed in this analysis. For this reason, further analyses are necessary in order to avoid the influence of the user experience on the final result.

5.2.3 Clusters-modes identification

As explained in 3.4, the aim of this process is to merge the poles with similar parameters in order to obtain groups which represent the number of identified modes.

The linkage of similar poles can be graphed using a *dendrogram* which shows the distances among the merged clusters. The distance is computed as described in 3.4.3.1.

Figure 5-10 – Dendrogram for hierarchical clustering

Figure 5-11 – zoom of a specific part of the three

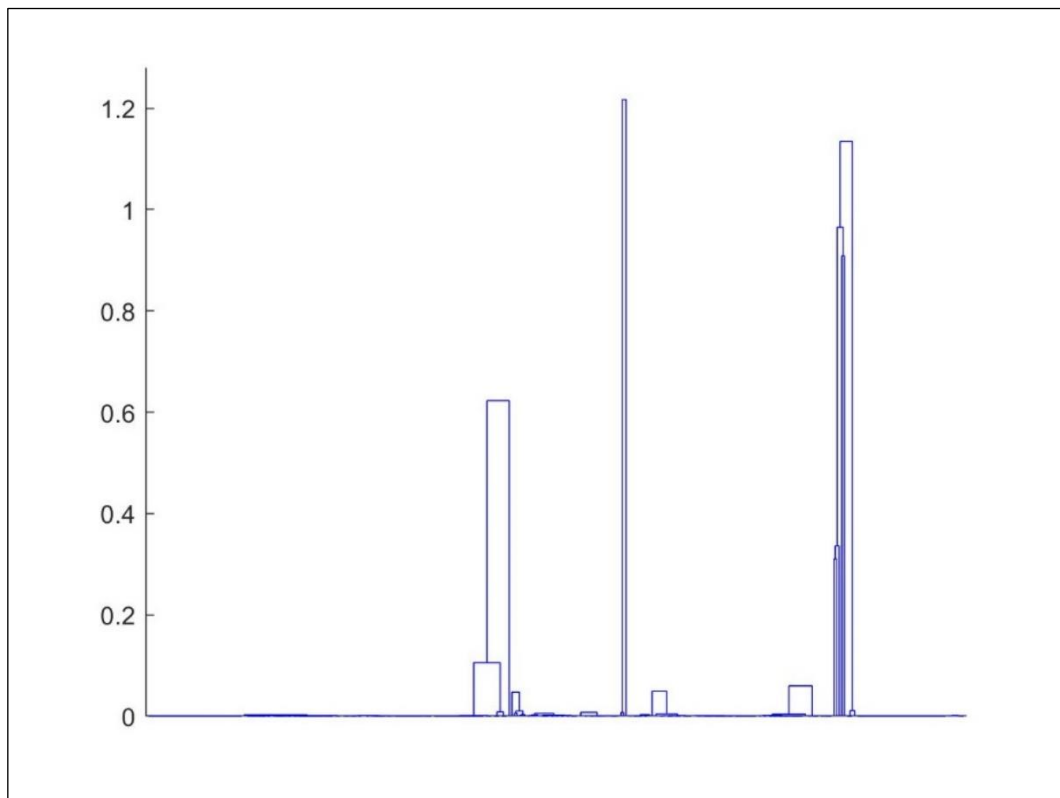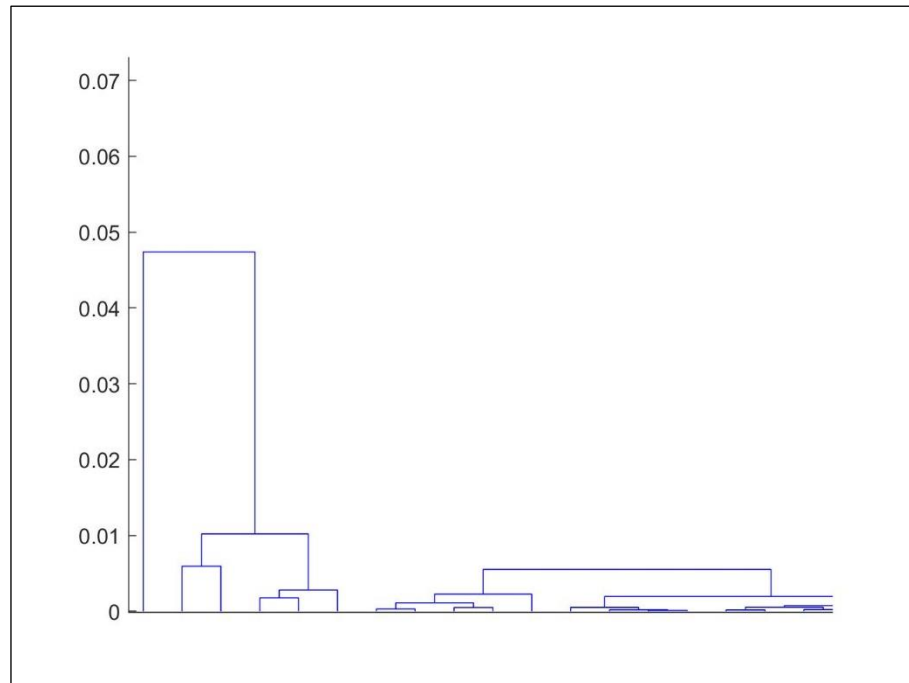The distances among the neighbour poles belonging to two consecutive model orders are analysed with a Weibull distribution in order to obtain the threshold which defines the number of clusters (Figure 5-12).
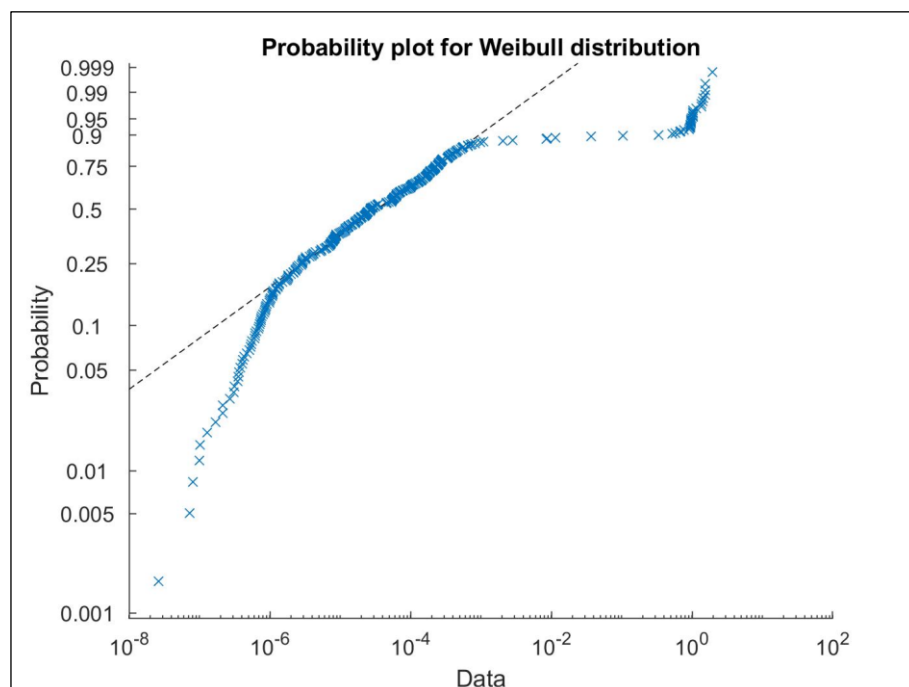
Figure 5-12 – Probability plot for Weibull distribution

The 95th percentile of the Weibull distribution defines the maximum distance between two clusters:

$$threshold = 0.1056$$

Then, it's possible to cut the dendrogram (Figure 5-13).
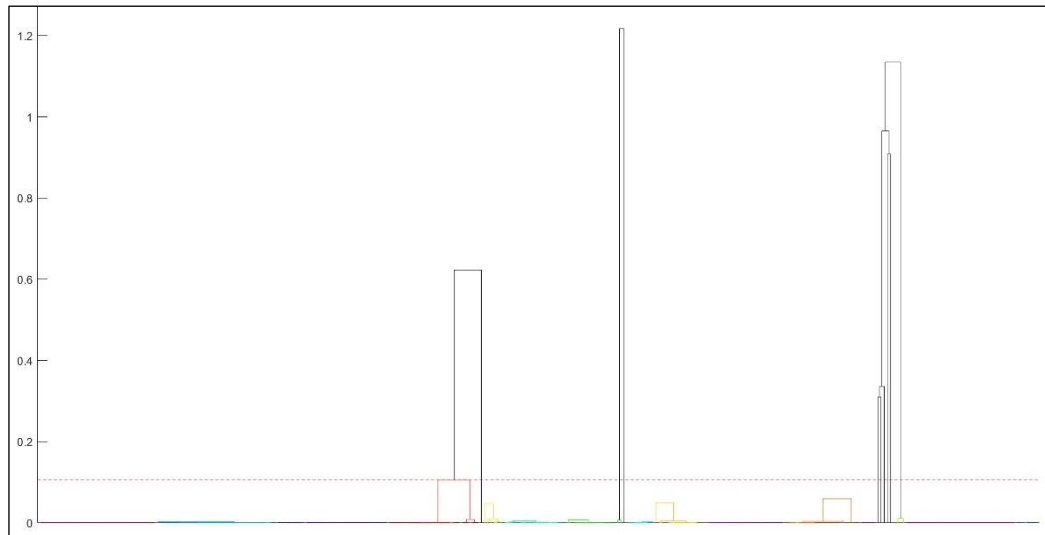
Figure 5-13 – Dendrogram (cut-off)

The starting 315 poles which were considered as stable are merged in 23 clusters.

The Silhouette (7A.2) is computed for each pole belonging to the 23 different clusters in order to obtain a visual analysis of the quality of hierarchical clustering and cut-off processes.

As shown in Figure 5-14, the silhouette assumes positive values for each pole and in the majority of the clusters, the silhouette is between 0.8 and 1. This is an evidence of the fact that the choice of the threshold allows to obtain well separated clusters characterised by different modal parameters.
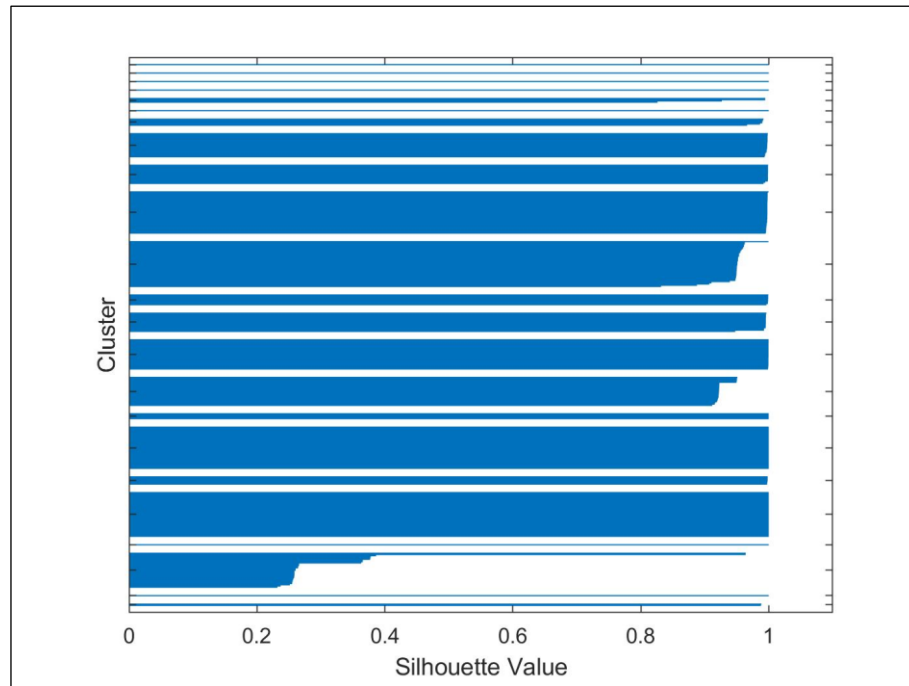
Figure 5-14 – Silhouette values for each pole belonging to the identified clusters
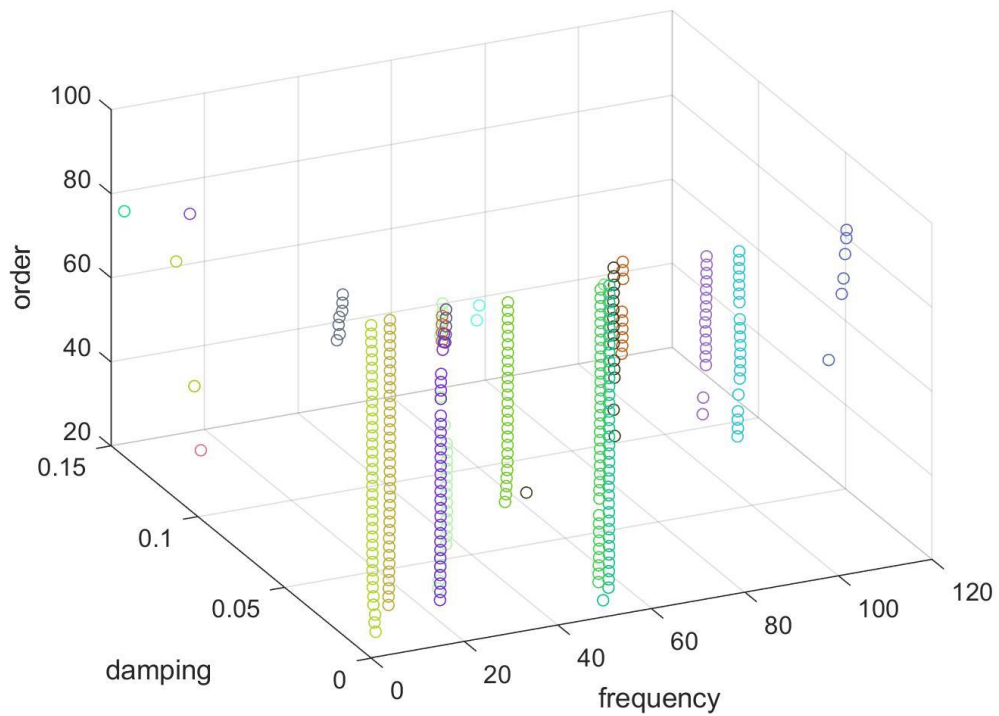
Figure 5-15 – frequency, damping ratio and model order of the identified modes

82

Then, the dimension of each cluster is analysed in order to avoid modes which are represented by fewer poles. In fact, the 23 clusters are characterised by a number of internal poles between 1 and 38. The dimension of each group is shown in Figure 5-16.
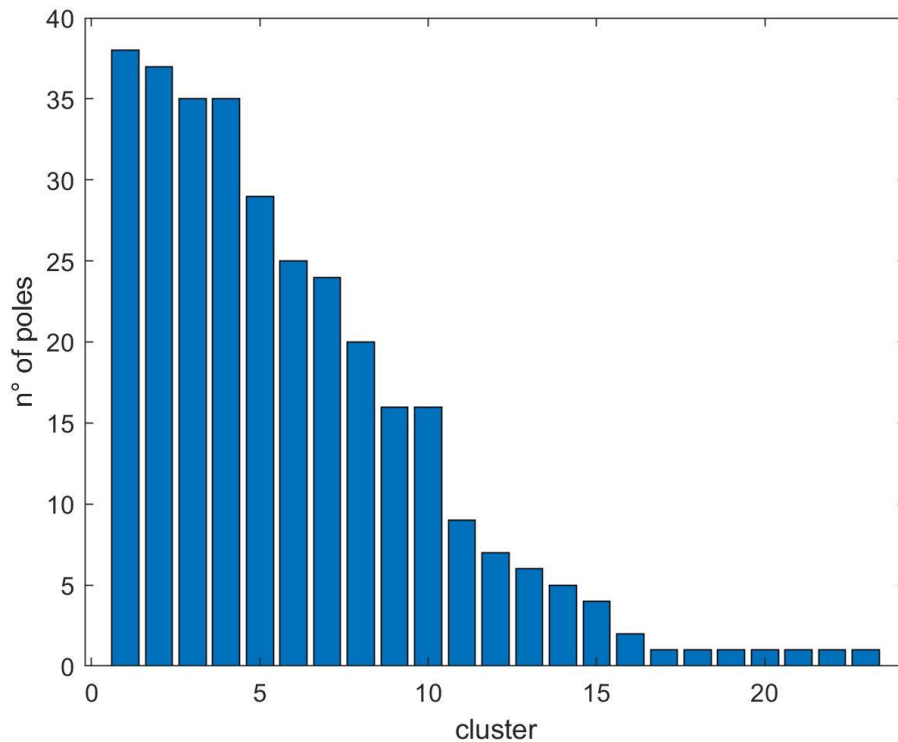
Figure 5-16 – number of poles within each cluster

The process described in 3.4.4 which allows to separate mathematical modes into possible physical modes, leads to a threshold of 14.14. In other words:

- clusters with at least 15 poles are considered as physical modes
- clusters with a number of poles lower than 15 are defined as mathematical modes and, consequently, discarded.

The analysis has identified 10 physical modes which are graphed in the following charts.
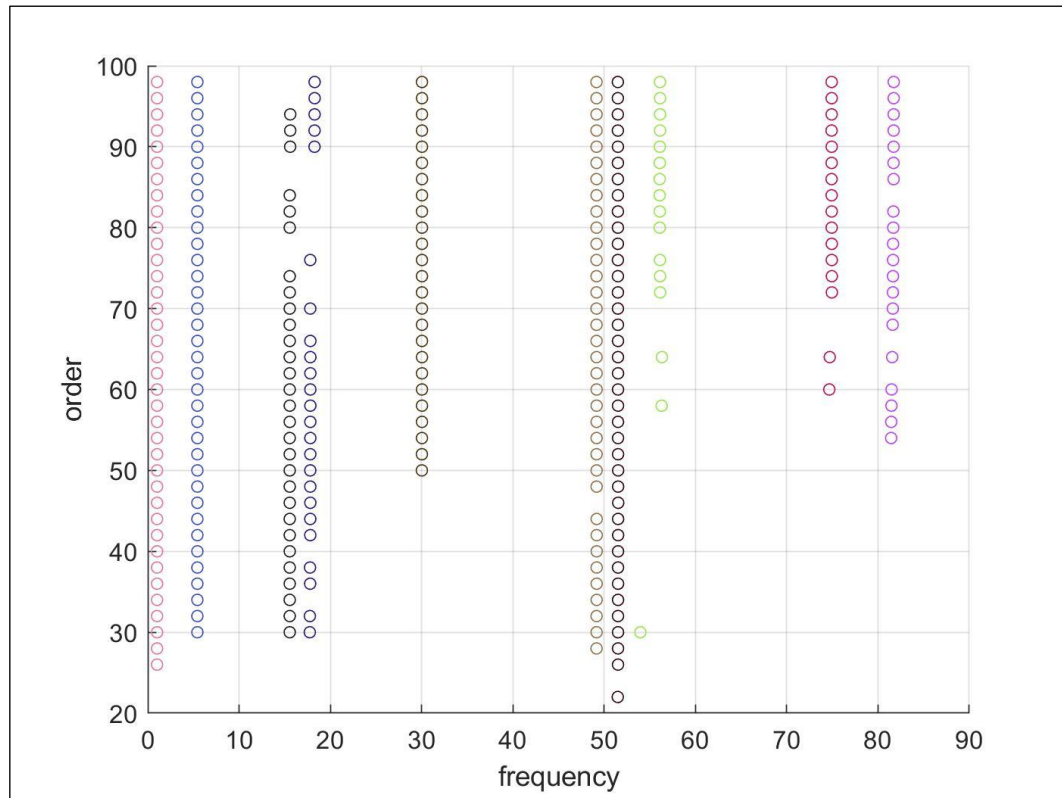
Figure 5-17 – physical modes (frequency – model order)
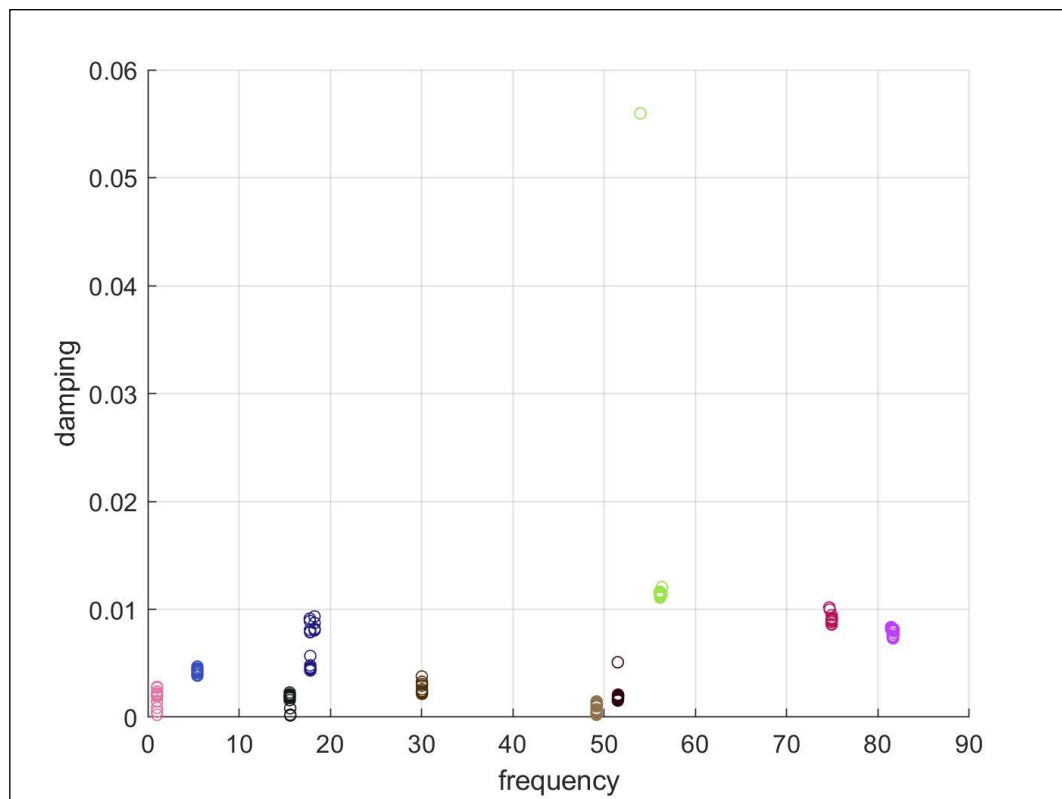
Figure 5-18 – physical modes (frequency damping ratio)

Then, for each cluster, an outlier-analysis is computed in order to discard poles which can affect the extraction of the modal parameters.
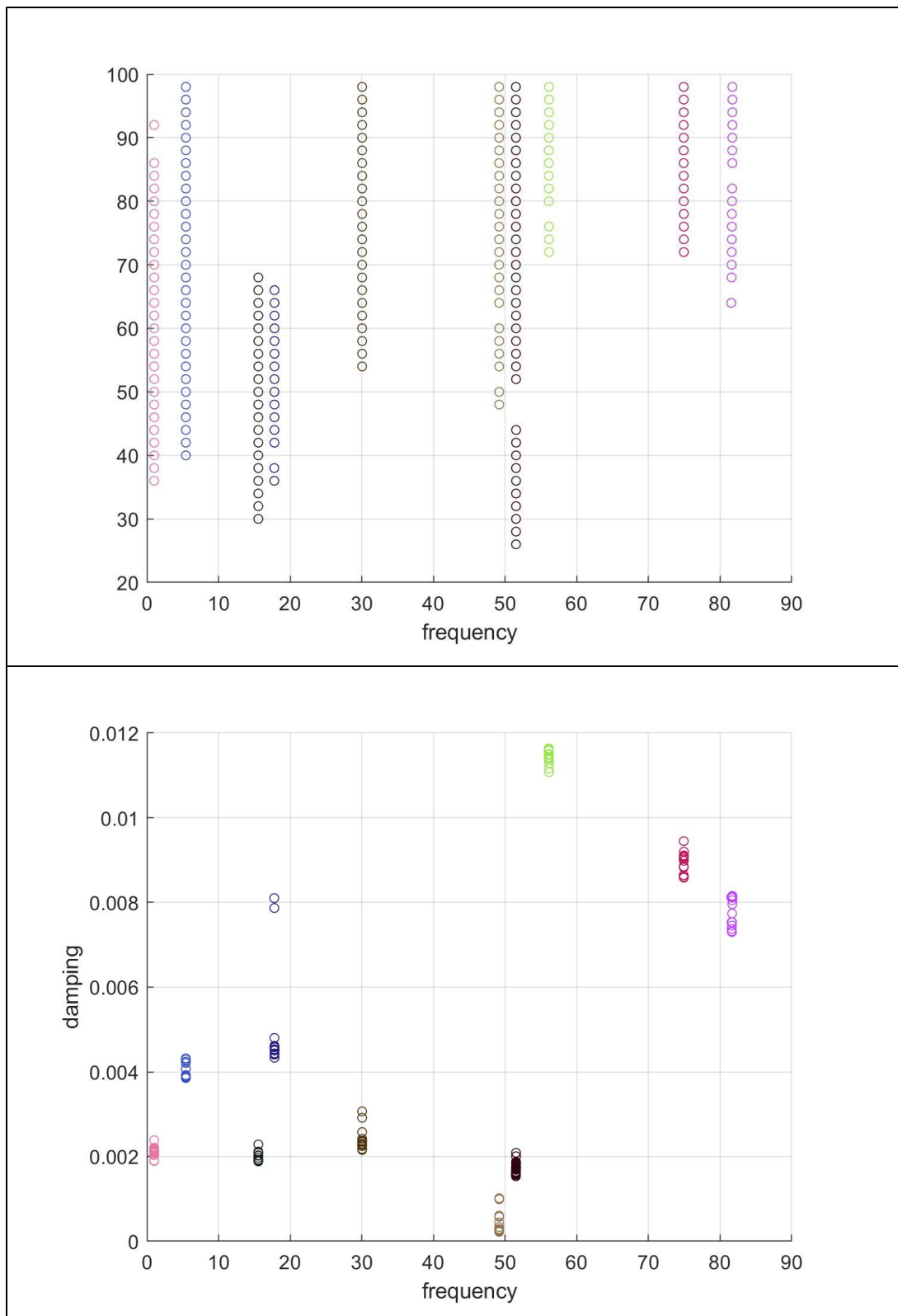
Figure 5-19 – Final clusters (removing outliers)

5.2.4 Modal parameters

The modal parameters identified by the analysis are computed as described in 3.6 and the results are summarised in the following tables.

method	1	2	3	4	FEM
f_1	1.0175	1.0175	1.0174	1.0174	0.88
f_2					3.34
f_3	5.4320	5.4320	5.4319	5.4325	4.88
f_4	15.5487	15.5489	15.5491	15.5481	14.98
f_5	17.8024	17.7995	17.8086	17.8086	22.27
f_6					28.03
f_7	30.0395	30.0392	30.0394	30.0394	29.39
	49.1636	49.1605	49.1616	49.1616	
f_8	51.5098	51.5166	51.5163	51.5163	51.15
	56.1060	56.0894	56.0959	56.0959	
f_9	74.9254	74.9343	74.9135	74.9135	75.12
f_{10}	81.6667	81.6828	81.6486	81.6486	81.80

Figure 5-20 – frequency identification for different methods

The choice of a method of extraction of the modal parameters does not affect the frequency identification (maximum difference of frequency among the methods $\sim 0.03\ Hz$).

The target frequencies are computed with the FEM model. The frequencies 3.34 Hz and 28.03 Hz are not identified by the AOMA. This

problem should arise because of the model order. In fact, in Figure 5-7 the two aforementioned frequencies are identified for model orders higher than 70 and, for this reason, a maximum of 15 poles may be established for each frequency. This number is not enough. As mentioned in the numerical case (Conclusions 4.6.4), the proposed method works well with a minimum of 25-35 model orders.

The frequencies 49.2 Hz and 56.1 Hz are established as physical modes but the FEM model cannot confirm these two results.

The damping ratios identified for each mode are shown in Figure 5-21.

method	1	2	3	4
dr_1	0.2125	0.2128	0.2083	0.2083
dr_2				
dr_3	0.3957	0.3923	0.3909	0.3862
dr_4	0.1991	0.1971	0.2021	0.1905
dr_5	0.4971	0.4801	0.4577	0.4577
dr_6				
dr_7	0.2330	0.2344	0.2274	0.2274
	0.0411	0.0446	0.0256	0.0256
dr_8	0.1722	0.1725	0.1709	0.1709
	1.1410	1.1397	1.1638	1.1638
dr_9	0.8921	0.8982	0.8845	0.8845
dr_{10}	0.7773	0.7739	0.7302	0.7302

Figure 5-21 – damping ratio identification for each method

Even then, the choice of the method does not lead to considerable difference in terms of results.

5.2.5 Comparison with a manual identification

A manual identification has been carried out on the SSI results considering:

- A range of model orders between 20 and 100
- Dimension of the Hankel matrix in accordance with the equation 4-5
- Comparison parameters equal to
  - $df < 0.05\%$
  - $0\% < dr < 10\%$
  - $(1 - MAC) < 5\%$

These impositions lead to the following stabilisation diagram.
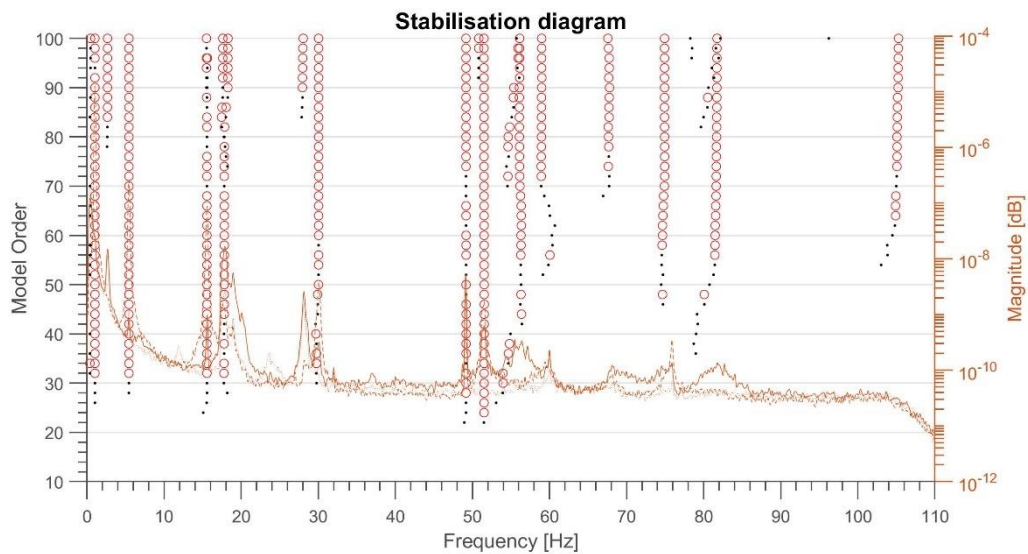
Figure 5-22 – stabilisation diagram (manual identification)

As shown in Figure 5-22, it's not simple to identify a model order through a visual analysis.

The modes identified as representative of the blade by a manual identification, which is carried out by an experienced user, present the following frequencies:

Manual identification		AOMA method 1	
f	dr	f	dr
1.02	0.20	1.02	0.21
2.63	2.25		
5.43	0.44	5.43	0.40
15.55	0.17	15.55	0.20
17.80	1.18	17.80	0.50
28.17	1.00		
30.04	0.26	30.04	0.23
51.49	0.31	51.51	0.17
56.24	1.86	56.11	1.14
59.21	2.10		
68.15	1.68		
75.12	1.22	74.93	0.89
81.80	1.09	81.67	0.78
102.94	2.35		
105.60	0.71		

Table 5-2 – comparison Manual and Automatic identification

The parameters show similar values for the identified modes but the number of modes is lower than the modes identified with a manual identification.

5.2.6 Number of modes in function of the maximum model order

In this chapter, different ranges of model orders are considered in order to analyse the influence on the number of modes identified by the proposed method.

The minimum model order is fixed equal to 20 and the maximum model order is considered between 70 and 180 with an interval of 10. The results are summarised in the following chart.
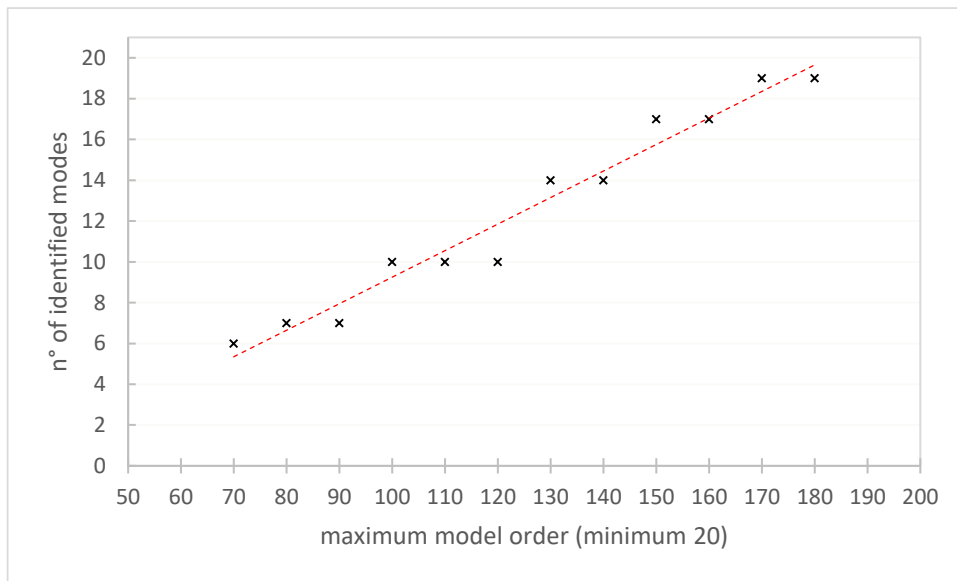

Figure 5-23 – number of identified modes in function of the maximum order

The number of modes identified by the AOMA increases with the growing of the maximum model order. So, we can assert that the maximum model order is one of the most important parameters which affect the process of identification.

6 MASONRY ARCH BRIDGE

A scale reproduction of a masonry arch bridge is analysed as a complex experimental case of a Civil Engineering structure.

This chapter is based on the following PhD Thesis: "Application of the SHM Methodologies to the Protection of Masonry Arch Bridges from Scour", Gianluca Ruocci, Politecnico di Torino, 2010. [16]

The features of the scaled model are taken from the Thesis of G. Ruocci [16] and from its relative paper [17]. The data are provided by the Politecnico di Torino.

6.1 Description

The scaled model of the masonry arch bridge shown in Figure 6-1 was built in the laboratory of the Structural and Geotechnical Engineering Department at the Politecnico di Torino. This model is not a reproduction of a real existing bridge but was designed taking masonry arch bridges common features, geometric proportions and historical design codes into account.

Figure 6-1 – The masonry arch bridge built in laboratory

The scale ratio was thought equal to 1:2.

The model is a twin-arch bridge with a length of 5.90 m, a width of 1.60 m and it is 1.75 m high. The two arches have a radius of 2.00 m, an angular opening of 30° a thickness of 0.20 m. The distance among the supports is equal to 2.00 m. The model was built with handmade clay scaled bricks (130x65x30 mm) in order to respect the adopted modelling scale law. Low compressive strength elements were chosen and a mortar with poor mechanical properties was used to bound them with the aim of reproducing the typical materials of a historical construction. The geometric details of the model are showed in Figure 6-2.
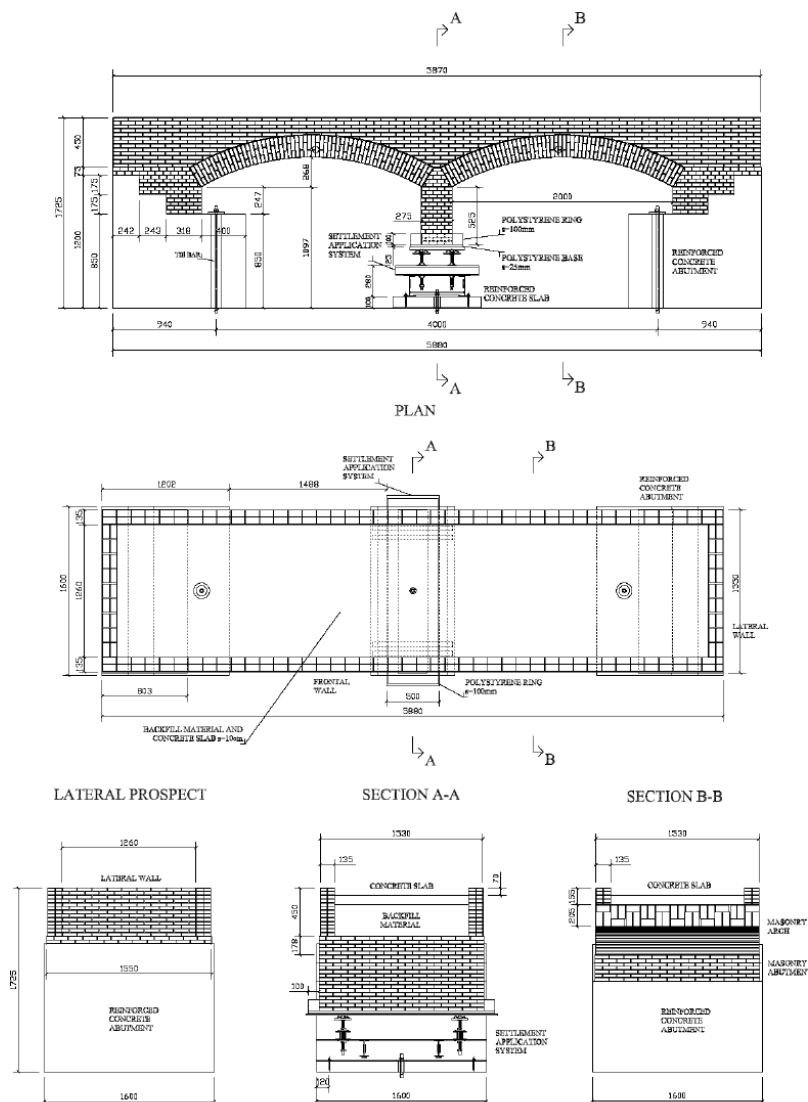

Figure 6-2 – geometric details of the masonry arch bridge

The aim of the whole experimental test was to reproduce the effects of scour at the foundation of the central pier of the twin-arch bridge model.

In order to create the effects of scour, a settlement application system was installed under the foundation of the central pier and supported by a reinforced concrete slab previously created and fixed to the laboratory floor.
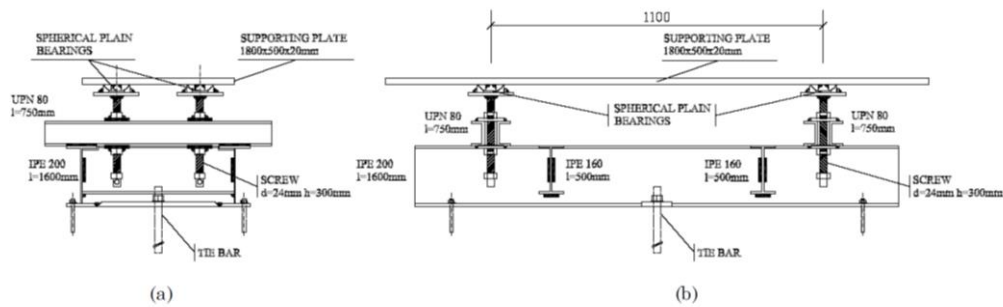

Figure 6-3 – the settlement application system: front view (a) and lateral view (b)

In order to simulate the streambed material surrounding the foundation of the central pier, a polystyrene mould was introduced. In this way a polystyrene layer interfaces the pier and the settlement application device and a polystyrene ring surrounds the pier.

6.2 Numerical estimation of the modal parameters

A 3D numerical model of the arch bridge was realized in the ADINA Finite Element package [18] to estimate its modal parameters.

The model consists of about 10000 nodes and 8800 elements, which can be distinguished among solid block, spring and shell elements.

A linear elastic behaviour was selected for all the elements groups. The elastic properties of the masonry elements groups were assigned according to the results of characterisation tests. The steel settlement application system was not modelled but three sets of translational springs working along the coordinated axes were used to simulate its flexible support. The materials parameters and the springs stiffness coefficients used in the FE model are listed in Table 6-1.

Material	Parameter	Value
Reinforced concrete	E [N/m²]	$3.0 \cdot 10^{10}$
	ν [-]	0.15
	ρ [kg/m³]	2400
Masonry	E [N/m²]	$1.5 \cdot 10^{9}$
	ν [-]	0.20
	ρ [kg/m³]	1900
Backfill material	E [N/m²]	$5.0 \cdot 10^{7}$
	ν [-]	0.1
	ρ [kg/m³]	2000
Concrete	E [N/m²]	$0.5 \cdot 10^{10}$
	ν [-]	0.15
	ρ [kg/m³]	2200
Translational springs	K_x [N/m]	$3.0 \cdot 10^{6}$
	K_y [N/m]	$3.0 \cdot 10^{6}$
	K_z [N/m]	$3.0 \cdot 10^{6}$

Table 6-1 – list of materials used in the FE model

The results of the numerical estimation are summarised in the Figure 6-4.

Mode I: 33.7 Hz - 1st vertical bending

Mode II: 33.8 Hz - 1st lateral bending

Mode III: 36.8 Hz - 1st longitudinal bending

Mode IV: 50.9 Hz - 1st torsional bending

Mode V: 60.3 Hz - 2nd longitudinal bending

Mode VI: 63.2 Hz - 2nd lateral bending

Mode VII: 69.7 Hz - 2nd vertical bending

Mode VIII: 83.3 Hz - 2nd torsional bending

Mode IX: 88.8 Hz – 3rd longitudinal bending

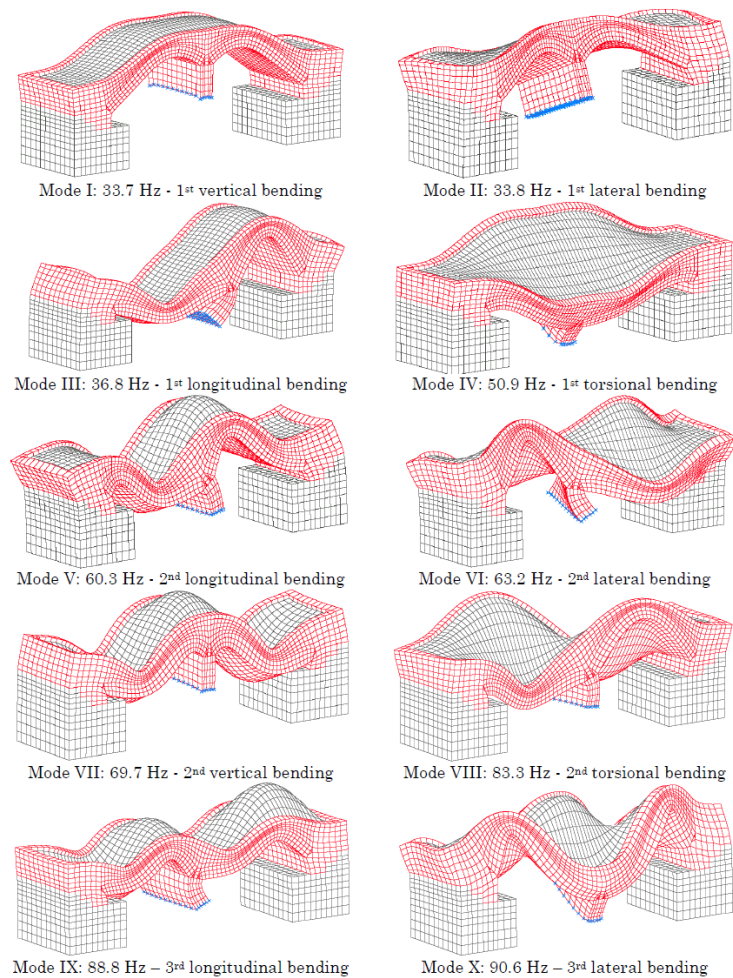Mode X: 90.6 Hz – 3rd lateral bending

Figure 6-4 – estimated frequencies and modal shapes (FE model)

94

6.3 Experimental Test Programme

Several damage steps have been applied to the structure in accordance with hydraulic flume tests as shown in Table 6-2.Table 6-2 - Damage steps, middle pier settlement, pier rotation, polystyrene removed

Experimental campaign	Damage steps	Settlement [mm]	Rotation [rad]	Polystyrene [%]
1st campaign	HS (healthy state)	0	0	0
2nd campaign	DS1	0	0	18%
	DS2	0.25	0	25%
	DS3	1	$4.21 \cdot 10^{-4}$	37.5%
	DS4	2.25	$1.01 \cdot 10^{-3}$	47%
3rd campaign	DS5	2.25	$1.23 \cdot 10^{-3}$	56%
	DS6	2.8	$1.23 \cdot 10^{-3}$	72%
	DS7	3.6	$1.27 \cdot 10^{-3}$	81%
	DS8	4.7	$1.30 \cdot 10^{-3}$	91%
	DS9	7.6	$1.28 \cdot 10^{-3}$	100%

Table 6-2 - Damage steps, middle pier settlement, pier rotation, polystyrene removed

Table 6-3 shows the timeline of the experimental tests. Different excitation sources were applied to the bridge model but the ambient vibrations (AV) source is the only which will be analysed in this work.

The experimental test involved three different experimental campaigns. The first campaign regarded the undamaged structure (October 2008 to March 2009).

The second campaign (April 2009) started after the application of additional masses on the central pier, in order to take into account the weight of the missing part of the pier. In the same campaign the first four settlement steps were applied on the upstream side of the pier. In addition, parts of the polystyrene ring were removed at each step to simulate the erosion of streambed around the foundation according to the hydraulic flume tests.

During the latter campaign (September 2010 to October 2010) five further settlement steps were applied. In this phase the removal of the polystyrene ring continued until all the polystyrene was removed.

	Time	Step	Excitation	Measurements
1st campaign	Oct. 2008	HS	AV, IH	ACC, SG, T, OPT
	Nov. 2008	HS	AV, IH	ACC, SG, T, OPT
	Jan. 2009	HS	AV, IH	ACC, SG, T, OPT
	Feb. 2009	HS	AV, IH	ACC, SG, T, OPT
	Mar. 2009	HS	AV, IH	ACC, SG, T, OPT
2nd campaign	Apr. 2009	HS	AV, IH	ACC, SG, T, OPT
		DS1	AV, IH, S	ACC, SG, T, OPT
		DS2	AV, IH, S	ACC, SG, T, OPT
		DS3	AV, IH, S	ACC, SG, T, OPT
		DS4	AV, IH, S	ACC, SG, T, OPT
3rd campaign	Sep. 2010	DS5	AV, IH, S	ACC, SG, T, OPT
		DS6	AV, IH, S	ACC, SG, T, OPT
	Oct. 2010	DS7	AV, IH, S	ACC, SG, T, OPT
		DS8	AV, IH, S	ACC, SG, T, OPT
		DS9	AV, IH, S	ACC, SG, T, OPT

Table 6-3 – Experimental test timeline

6.4 Experimental Setups

The sampling frequency was fixed to a value of 400 Hz to acquire the signals produced by ambient noise and a 180 seconds time lap was adopted. The random excitation is fixed in a 10 to 100 Hz band. Two setups were used for each vibration test in order to capture the higher number of natural modes. Each setup consisted of 18 channels leading to 36 instrumented positions.

The selected sensors for the dynamic tests performed on the structure were capacitive accelerometers. The employed dynamic acquisition system was composed of a set of 18 monoaxial PCB Piezotronics accelerometers with a sensitivity of 1 V/g, a measurement range of ±3 g, a broadband resolution of 30 µg and a weight of 17.5 g. The accelerometers

were connected through coaxial cables to the LMS Difa-Scadas data acquisition system which also provided the signal amplification. The acquired signals were recorded on the hard drive of a laptop computer interfaced with the data acquisition system and running a specific signal acquisition software.
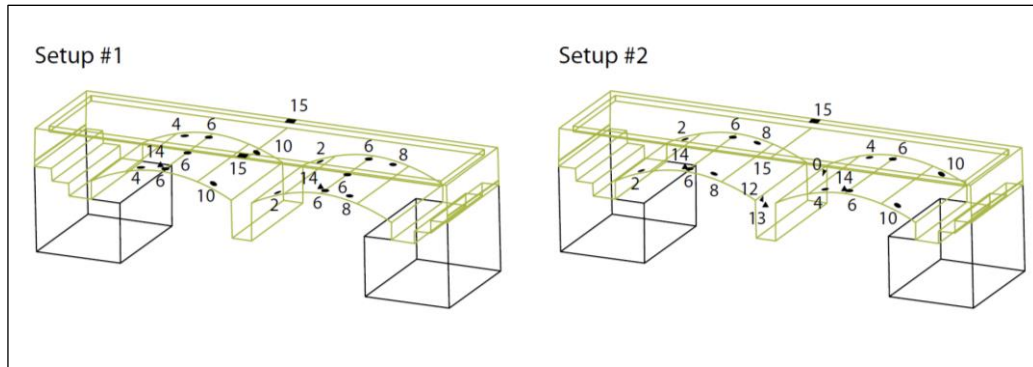

Figure 6-5 – Experimental setups

6.5 AOMA application

The analysis is carried out on the measurements which were recorded the 2nd April 2009 in the 2nd campaign of acquisition (DS1).

SSI is set with the following characteristics:

- Minimum model: 40
- Maximum model: 90
- Number of analysed model orders: 25
- Hankel matrix computed in accordance with the equation 4-5

6.5.1 Hard criteria

The number of identified poles is 793.

The poles which present a negative damping ratio or higher than 20% are **154**. The remaining 639 poles are processed with the Soft Criteria

6.5.2 Soft Criteria

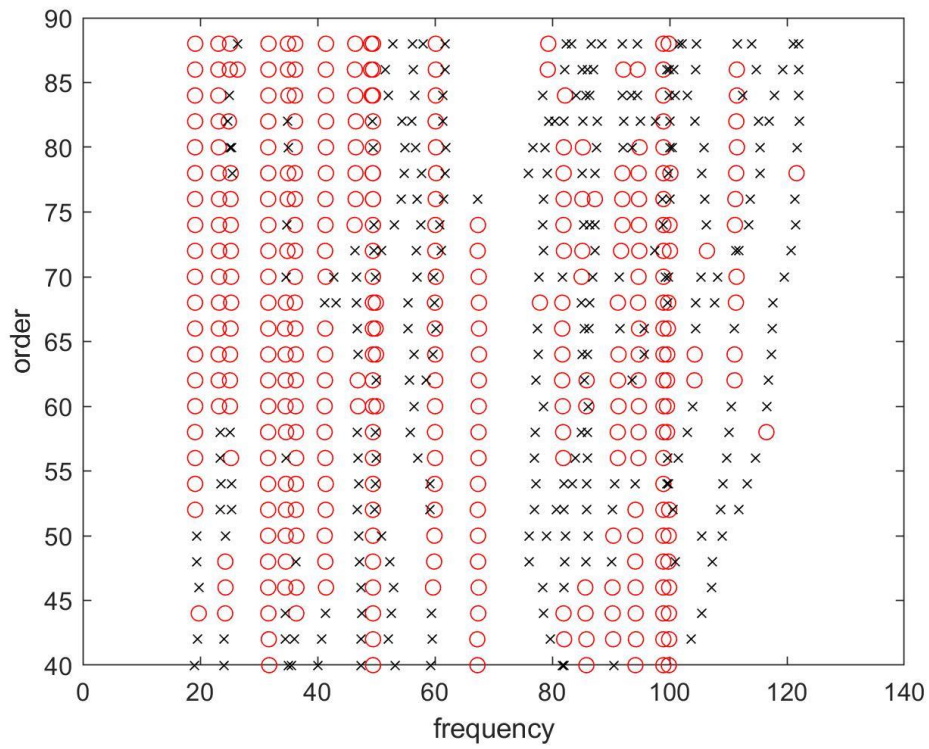The Soft Criteria analysis leads to the following stabilisation diagram.

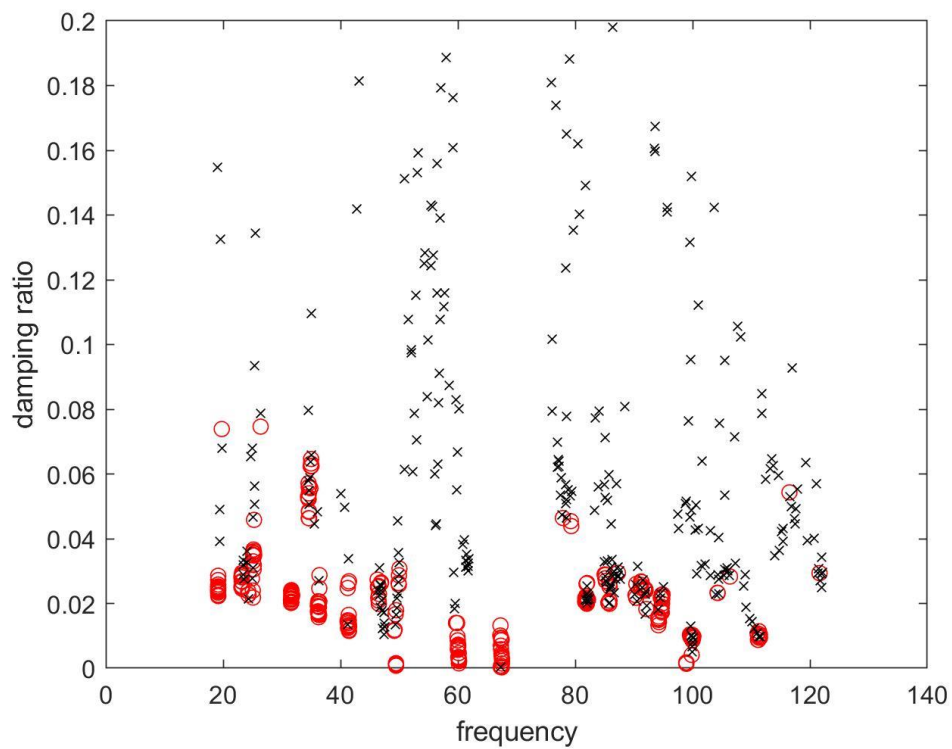*Figure 6-6 – stabilisation diagram (automatic detection)*

Figure 6-7 – stable/unstable poles (frequency-damping)

The number of poles which are classified as stable is 345, while the unstable poles are 294.

The maximum values of the comparison parameters which characterise the elements belonging to the cluster of the stable poles are showed in Table 6-4.

Comparison parameter	Maximum value allowed to stable poles
df	0.2616 Hz
ddr	0.83%
$dMAC$	0.4467
$dMPD$	0.1002

Table 6-4 – maximum values of comparison parameters for stable poles

The distribution of comparison parameters which characterise the poles belonging to the stable cluster might be of particular interest in order to understand how many poles reach the upper boundaries of the comparison parameters (Figure 6-8).

In a classical stabilisation diagram, usually the definition of stable poles starts from the introduction of rigid limits for the comparison parameters. It is possible to assume as example these thresholds:

Comparison parameter	Stable poles	Unstable poles
df	$< 0.1 \, Hz$	$> 0.1 \, Hz$
$dMAC$	< 0.05	> 0.05

Although the maximum value of $(1 - MAC)$ which represent the stable poles is very high, the elements which show a value of $(1 - MAC)$ lower than 0.05 are more than the 97%.

For the difference of frequency, the 87% of stable poles presents a value of df lower than 0.1 Hz.
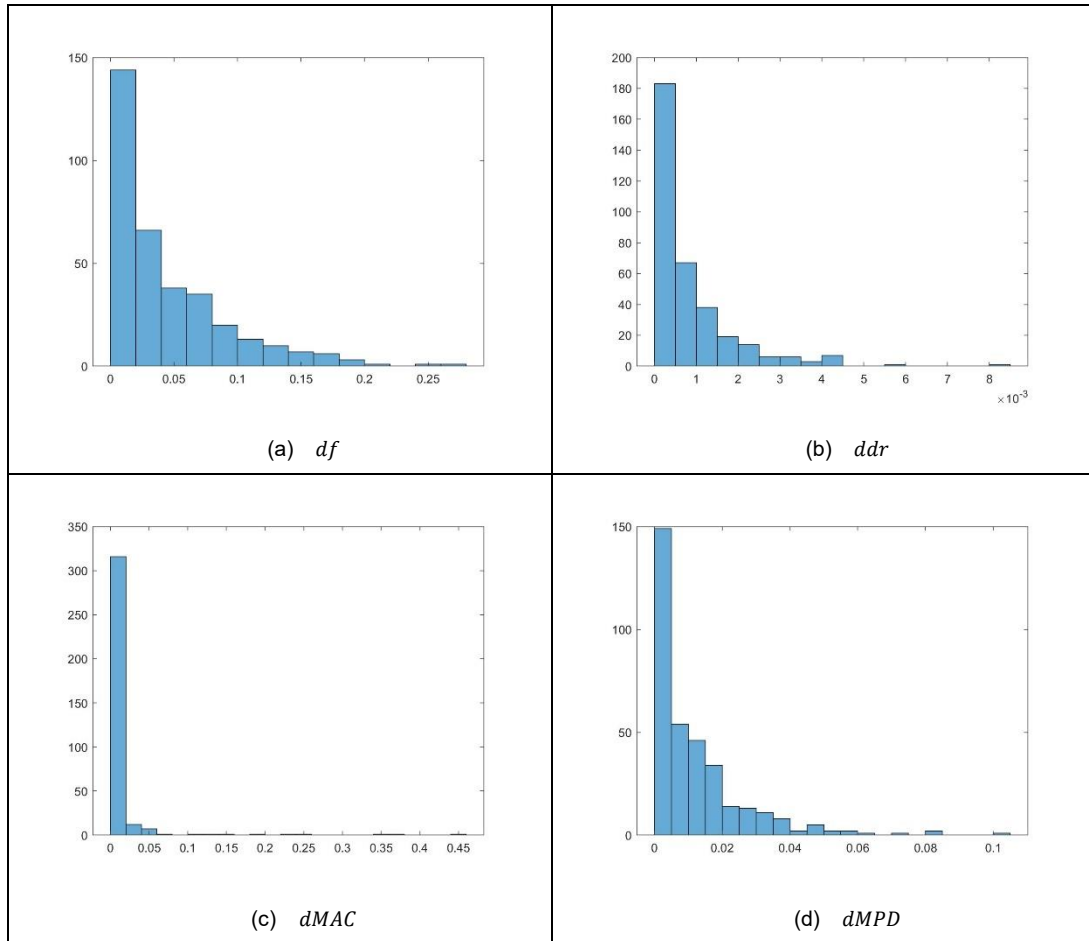

Figure 6-8 – comparison parameters distribution for stable poles: difference of frequency (a), difference of damping ratio (b), 1-MAC (c), difference of MPD (d)

6.5.3 Clusters-modes identification

The hierarchical clustering leads to the dendrogram shown in Figure 6-9.
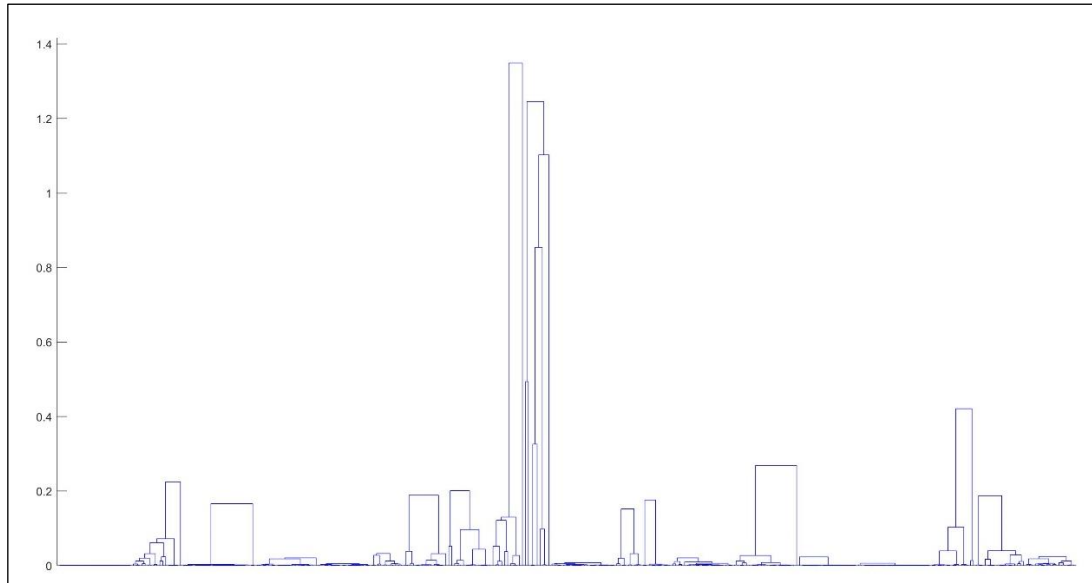
Figure 6-9 – dendrogram for hierarchical clustering (masonry arch bridge)

The computing of the threshold which results in the identification of a certain number of clusters leads to a high value. This is consequence of the fact that there is more distance among poles of this experimental case respect to the helicopter blade analysed in chapter 5.
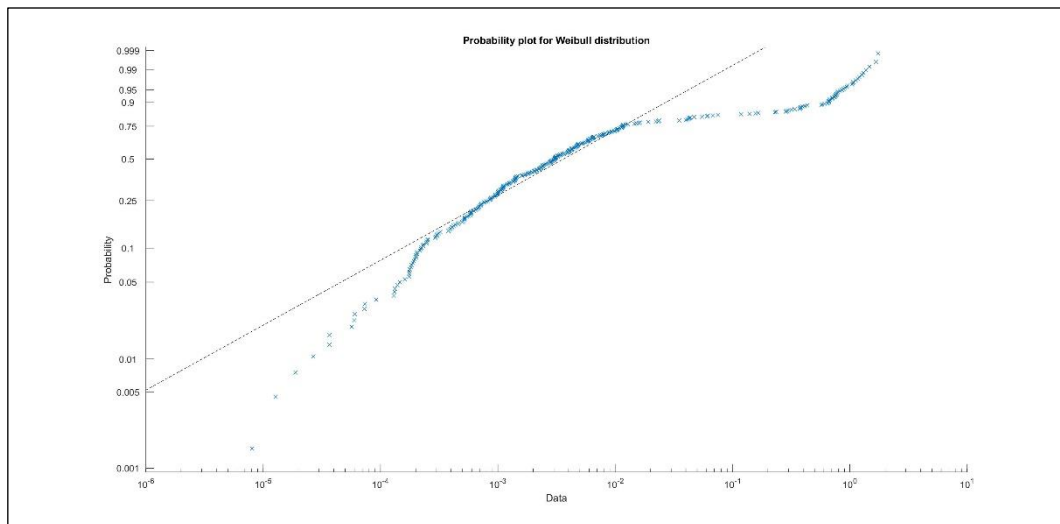

Figure 6-10 – probability distribution (Weibull) – threshold identification

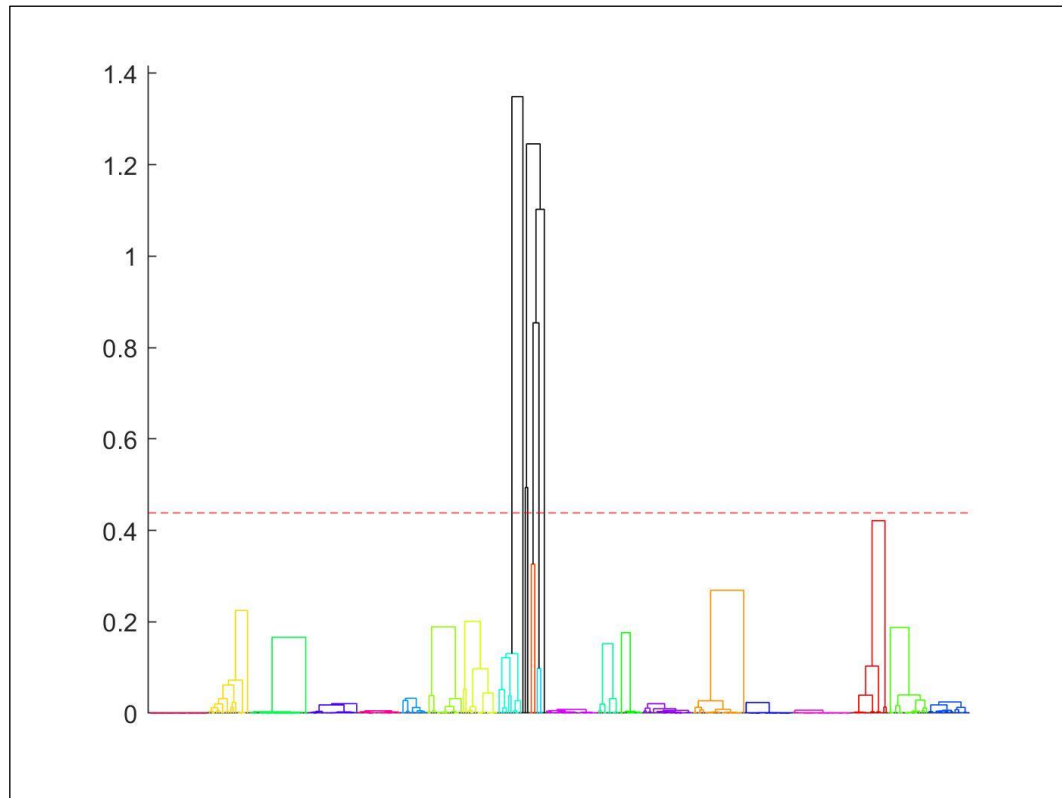$$threshold(p_{95\%}) = 0.4377$$

Figure 6-11 – cut-off – masonry arch bridge dendrogram

The number of identified clusters is equal to 25.

In order to quantify the quality of the hierarchical clustering process and the relative choice of the threshold, the silhouette factor (A.2) is computed for each pole belonging to the different 25 clusters (Figure 6-12).
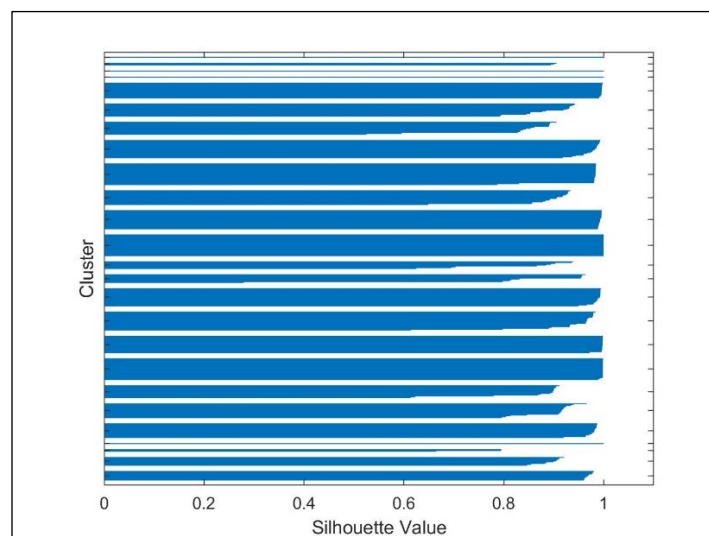

Figure 6-12 – Silhouette value (clustering quality)

The clusters have a number of poles between 1 and 25. The distribution of the poles within each cluster is shown in Figure 6-13.
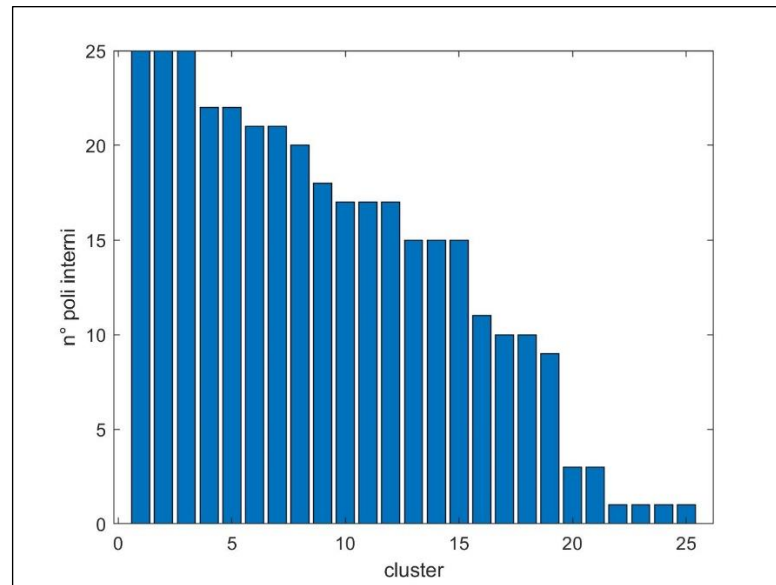

Figure 6-13 – poles within each cluster

The 2-means clustering is carried out on the number of poles which characterise each cluster in order to split mathematical modes from physical modes (3.4.4). The result is summarised in the following table.

Clusters	Number of poles within the cluster
Physical modes	> 10
Mathematical modes	≤ 10

The clusters which are identified as physical modes are 16 while the mathematical modes are 9.

Then, the outliers are removed from each cluster using the approach explained in 3.5.

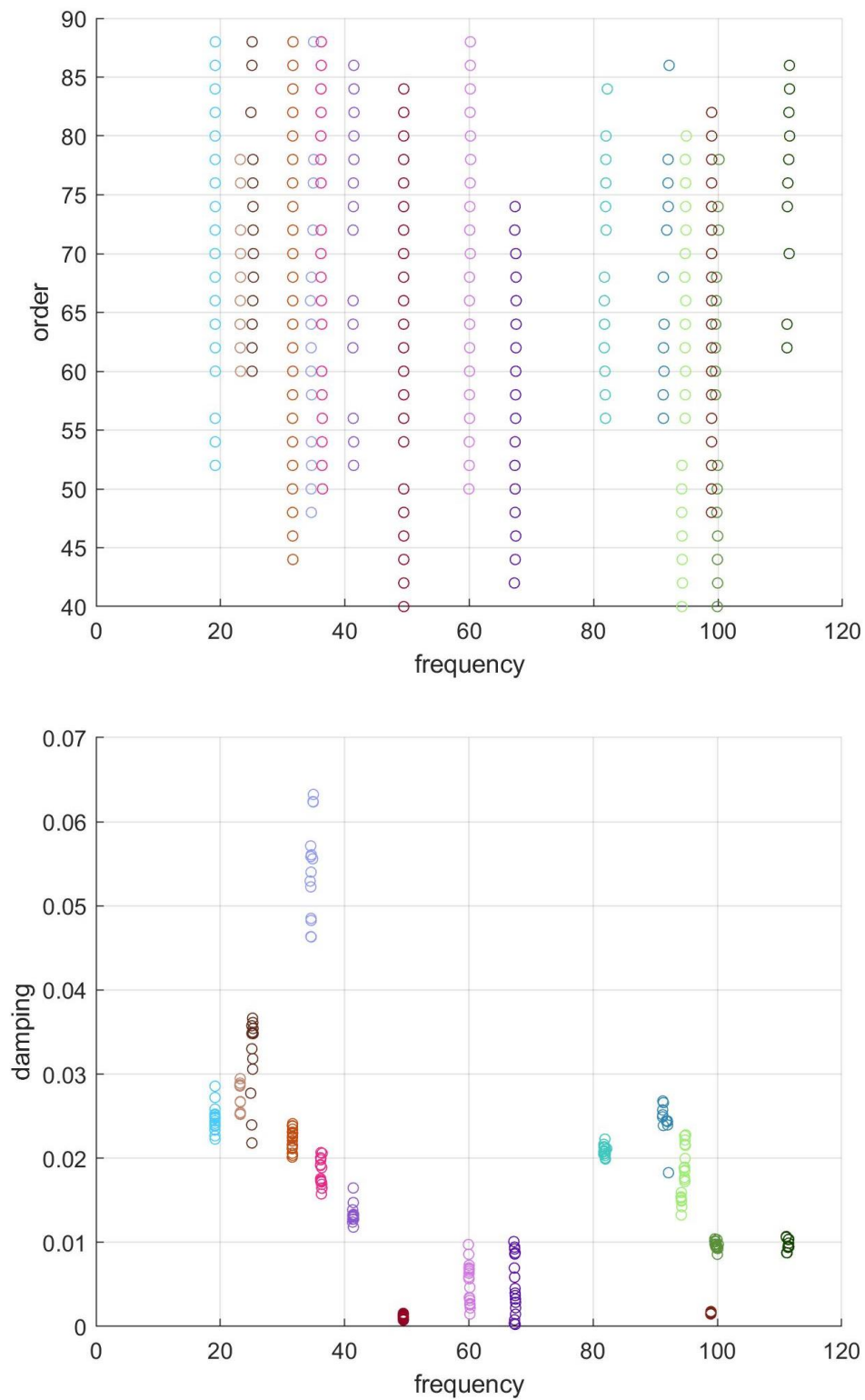The final clusters are shown in a frequency – model order diagram and in a frequency – damping ratio diagram.


Figure 6-14 – frequency vs model order and frequency damping ratio (physical modes)

The distribution of the values of the eigenvectors might be of particular interest in order to have a further diagram which represents the stability of the poles.

In the following figures, the eigenvectors are plotted for each cluster where the clusters-modes are sorted from the smallest to the largest frequency (Mode 1, Mode 2... Mode 16).
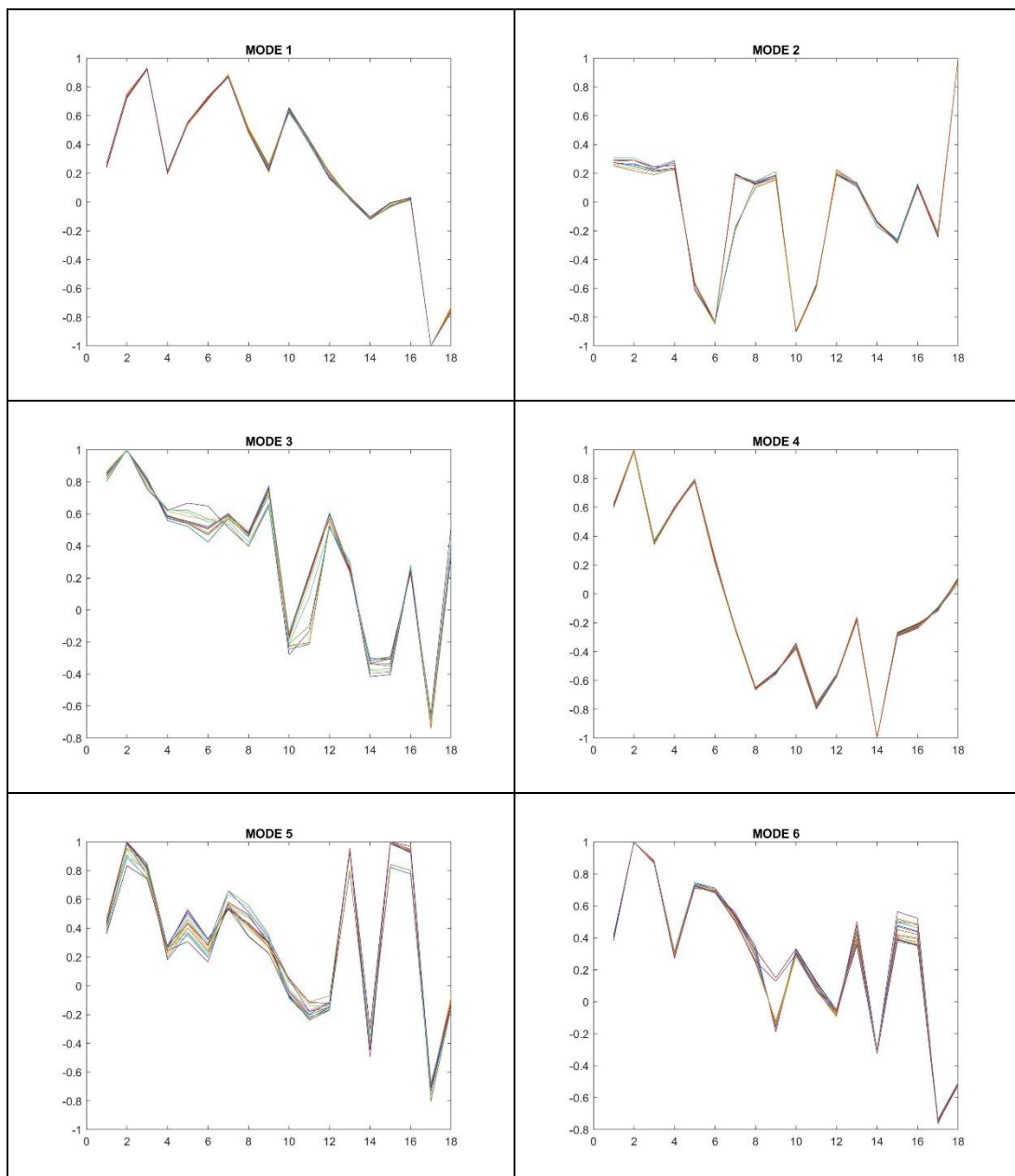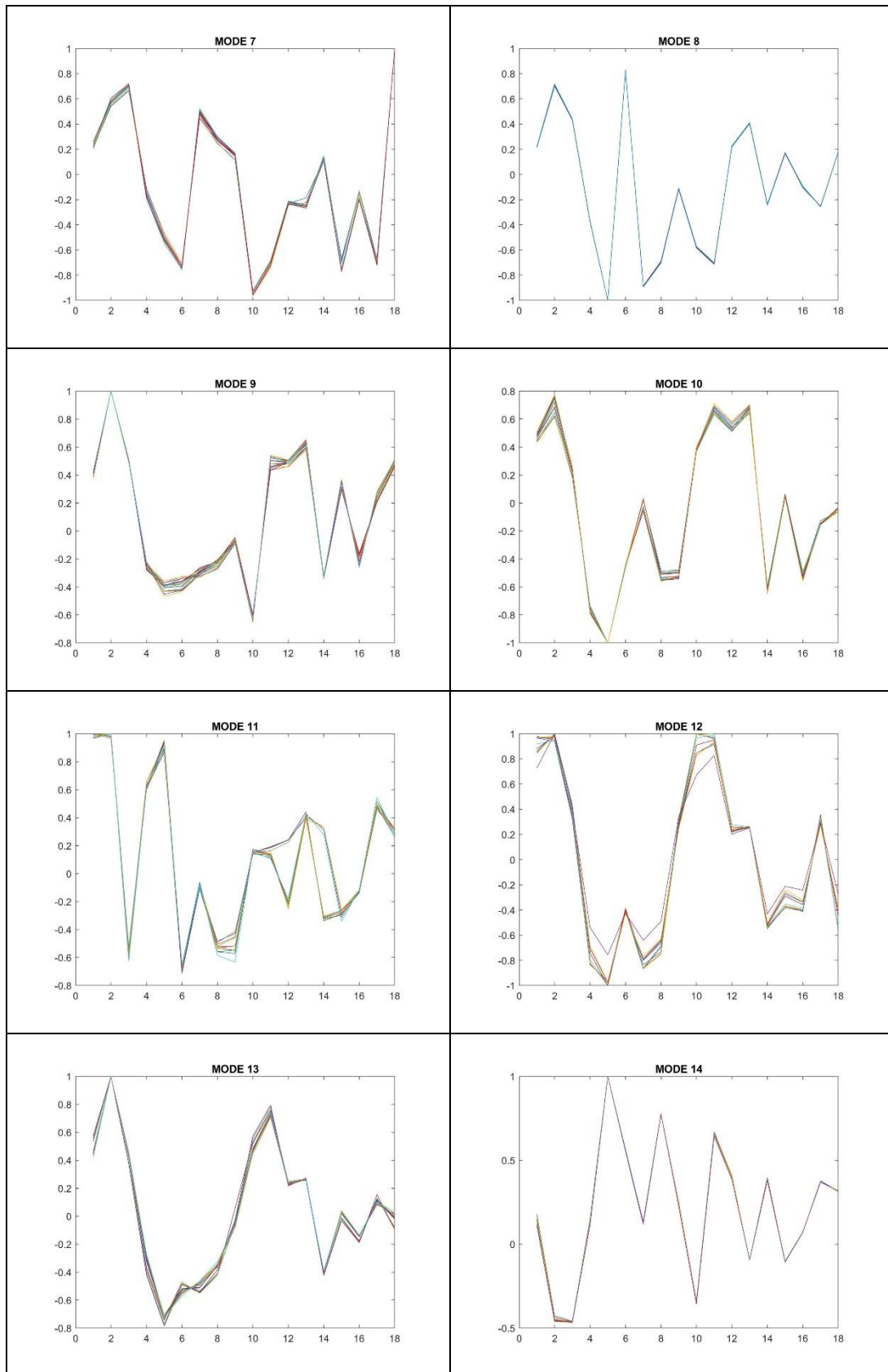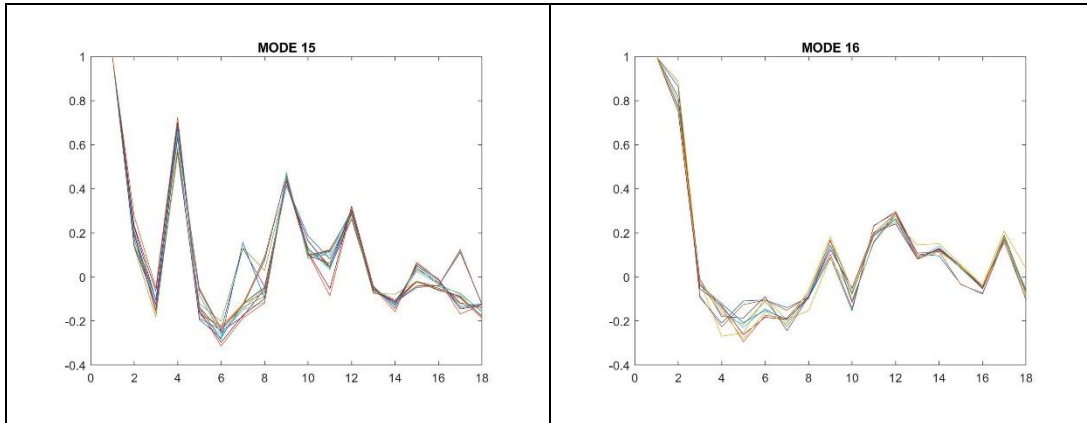
Figure 6-15 – eigenvectors within each cluster

Some modes show uncertainty for part of the elements which compose the eigenvector. On the other hand, Modes 1, 4, 8 and 14 present a low dispersion of the values which will constitute the modal shapes.

6.5.4 Modal parameters

The choice of the modal parameters is carried out with the method 1. In this way, each value which characterises the mode is computed with the mean of the parameters relative to the poles belonging to the respective cluster.
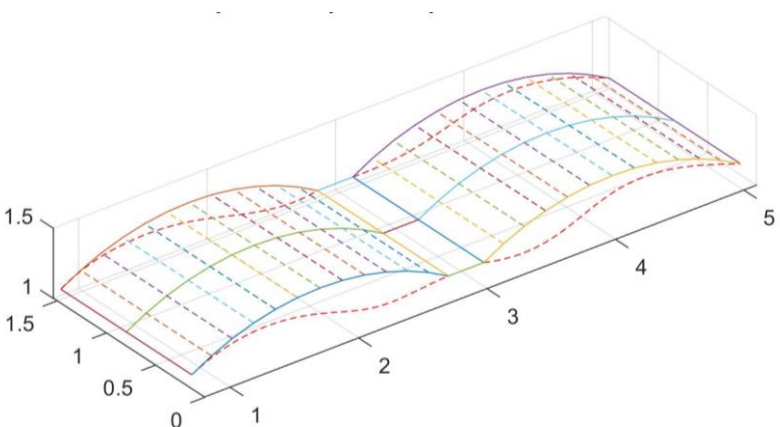
The frequencies and the damping ratios which are identified in this computing are showed in the following table.
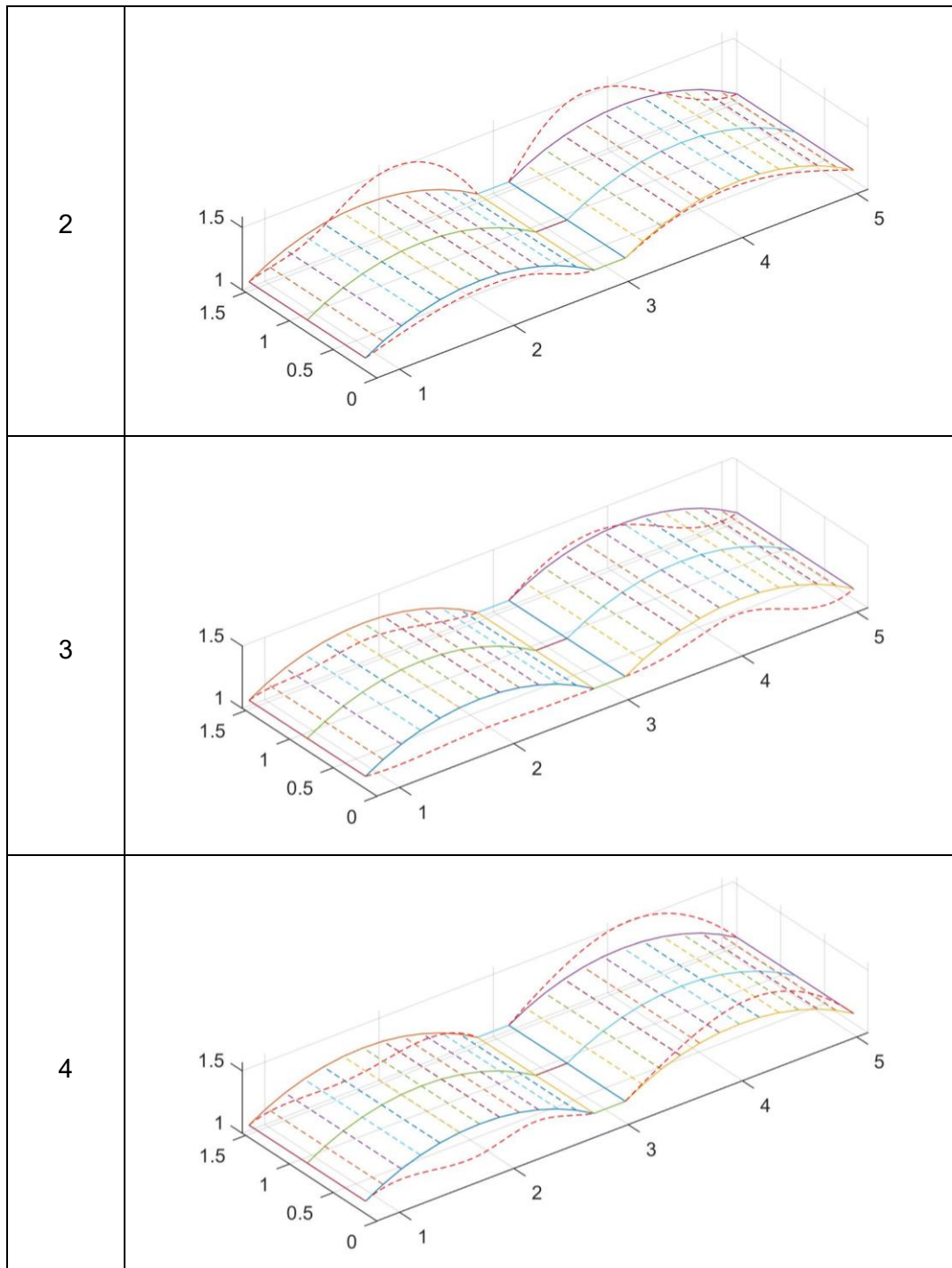
MODE	frequency	Damping ratio
1	19.14	2.46%
2	23.20	2.72%
3	25.14	3.21%
4	31.61	2.22%
5	34.68	5.44%

6	36.23	1.82%
7	41.38	1.33%
8	49.45	0.12%
9	60.08	0.50%
10	67.41	0.48%
11	81.87	2.09%
12	91.55	2.44%
13	94.55	1.81%
14	98.95	0.16%
15	99.82	0.97%
16	111.31	0.98%

Table 6-5 – frequencies and damping ratios (automatic identification)

The modal shapes are plotted in the following table.

MODE	MODAL SHAPE
1	

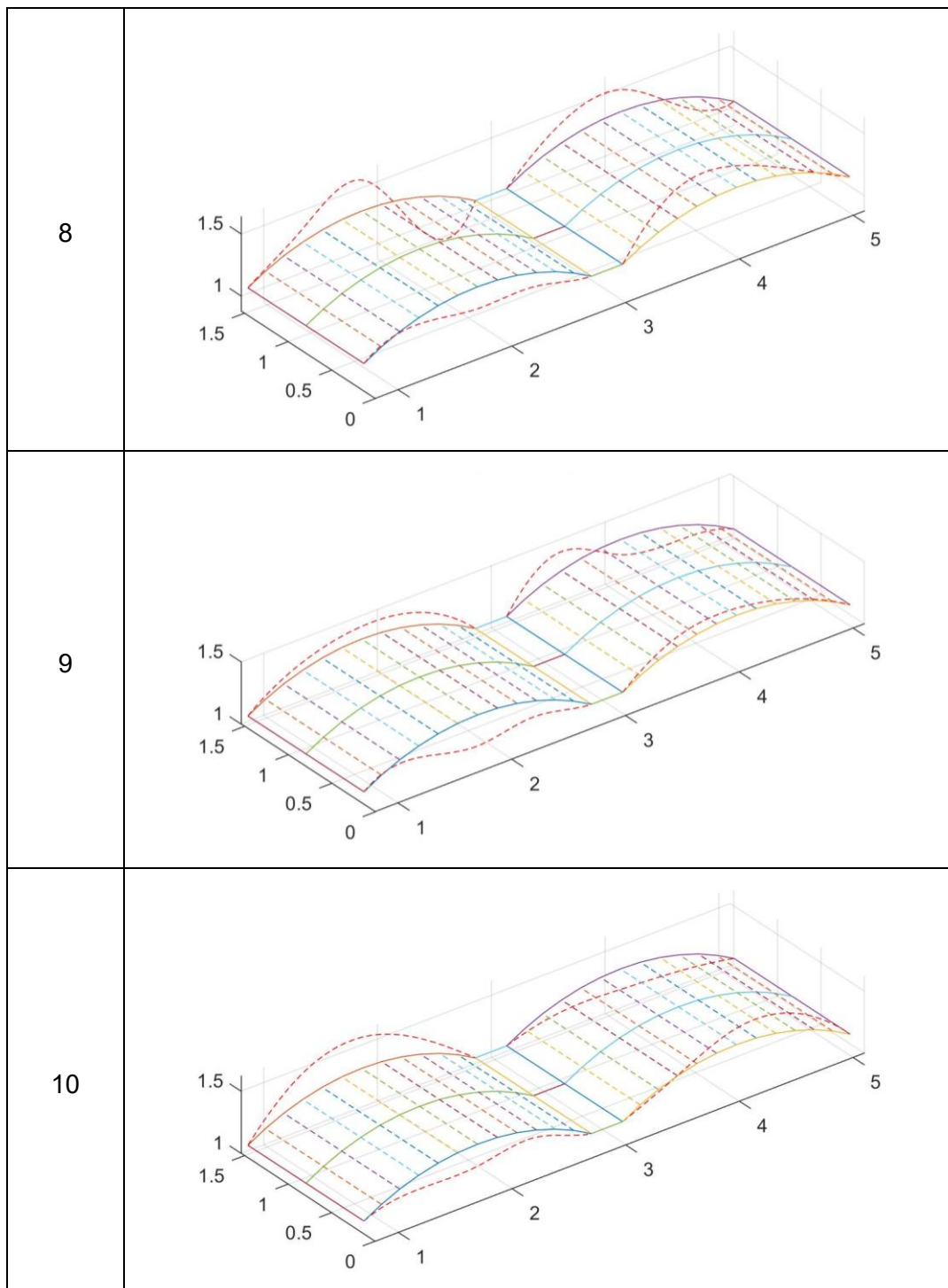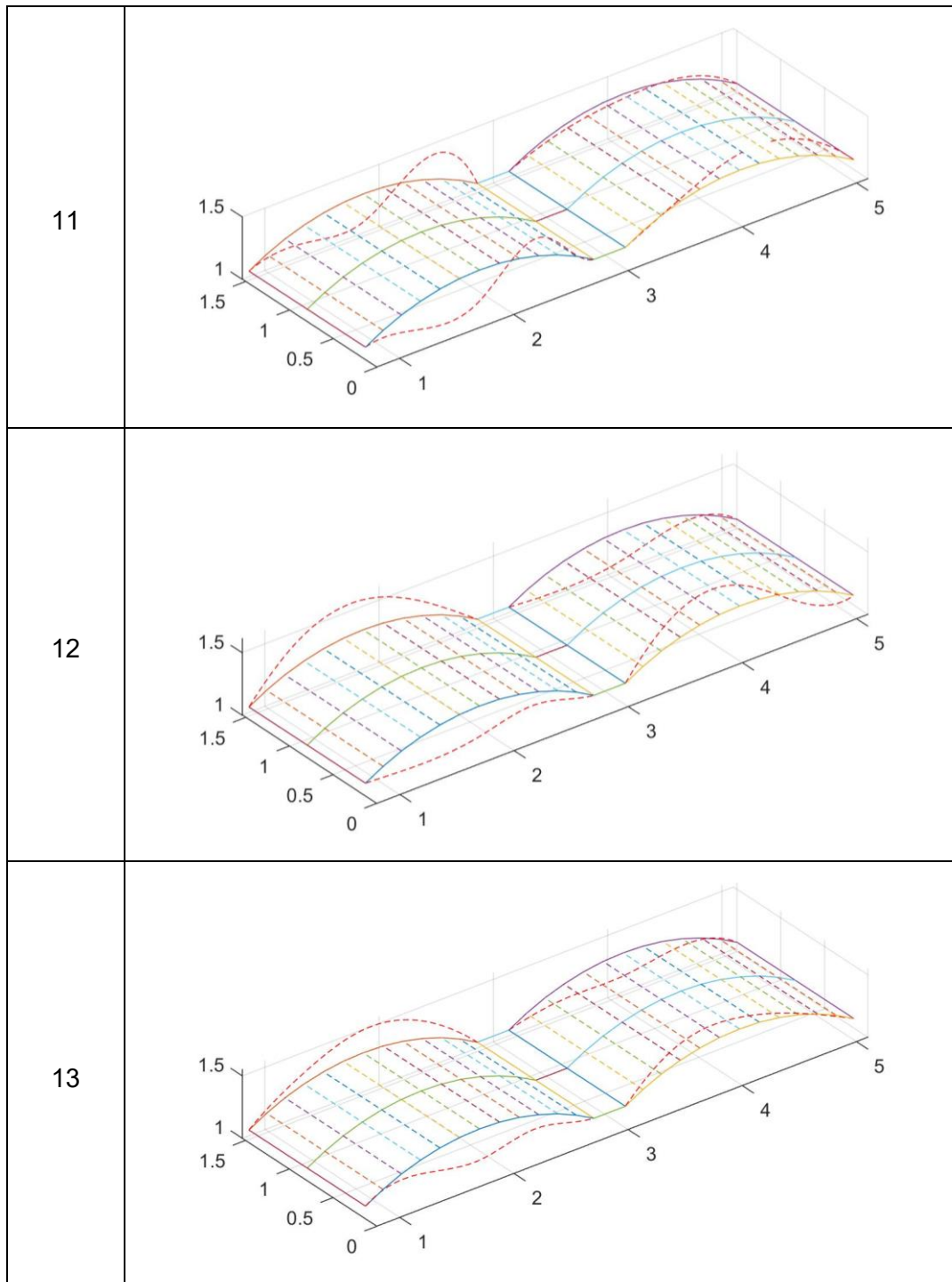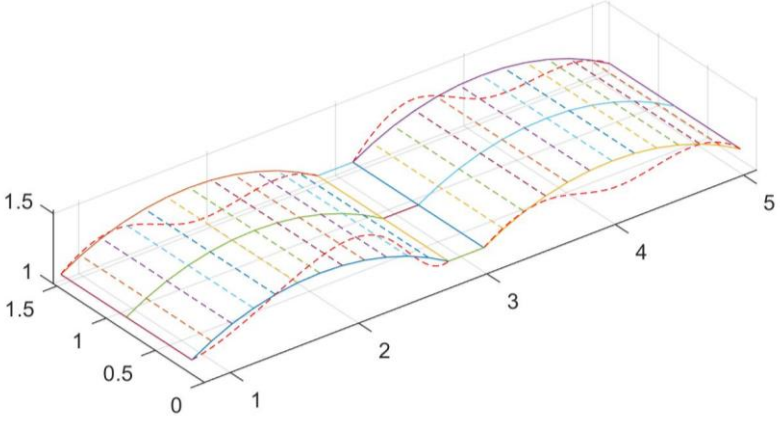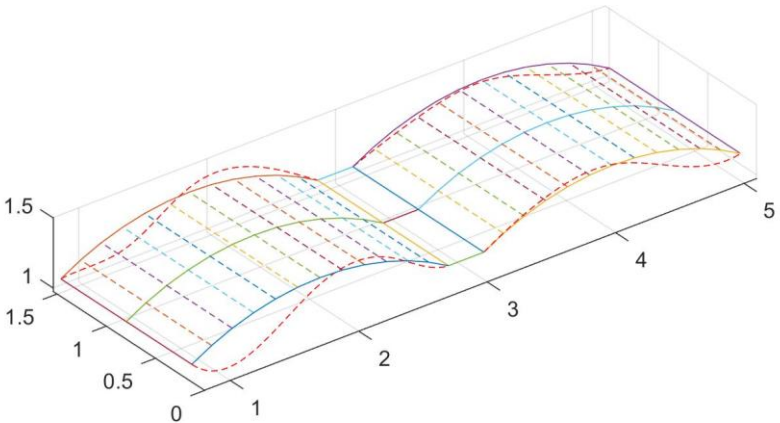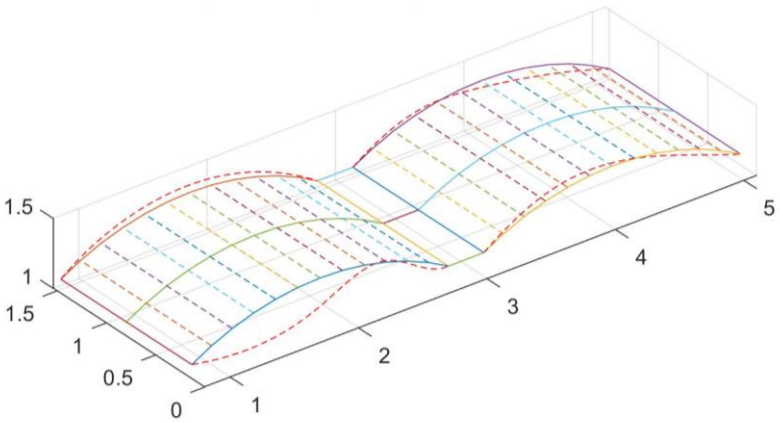| 2	
3	
4	

5	
6	
7	

8	
9	
10	

11	
12	
13	

14	
15	
16	

Table 6-6 – Modal shapes (1-16)

113

6.5.5 Influence of model order

The range of the model orders which was adopted in the previous analysis, was established in function of the threshold level and the silhouette values.

In other words, the minimum and the maximum value of the model orders which are analysed in the previous analysis are selected in accordance with the following objectives:

- Considering a number of model orders between 25 and 35, as suggested in 4.6;
- Obtaining positive values of Silhouette;
- Obtaining the minimum value of cut-off;

In the following chart, the value of the threshold is showed in function of minimum and the maximum model order considered in each analysis. The minimum model order is considered between 20 and 40 with intervals of 2. The maximum model order is between 70 and 90 with intervals of 2. In this way the analysis which are computed are 121.
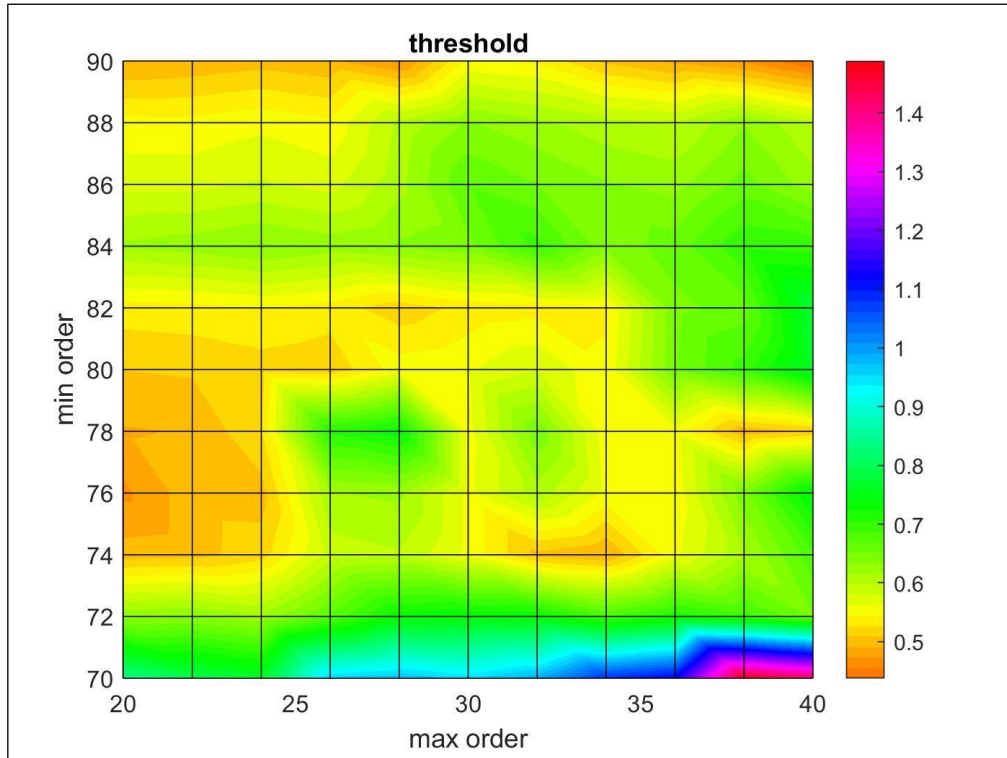
Figure 6-16 – influence of the model order on the threshold

The Figure 6-16 shows the minimum values of the threshold for the model orders 20-76, 28-90 and 40-90. This last range was adopted for the identification of the modal parameters which are showed in 6.5.4.

For the 121 combinations of min-max model orders, the modal parameters are estimated and analysed with a cluster analysis. The maximum number of clusters-modes which are identified in a single analysis is 19. So, a k-means analysis is carried out considering:

- A number of clusters k=19;
- A distance among the modes computed considering the frequency as the only comparison parameter.

The final clusters are showed in the following figure.
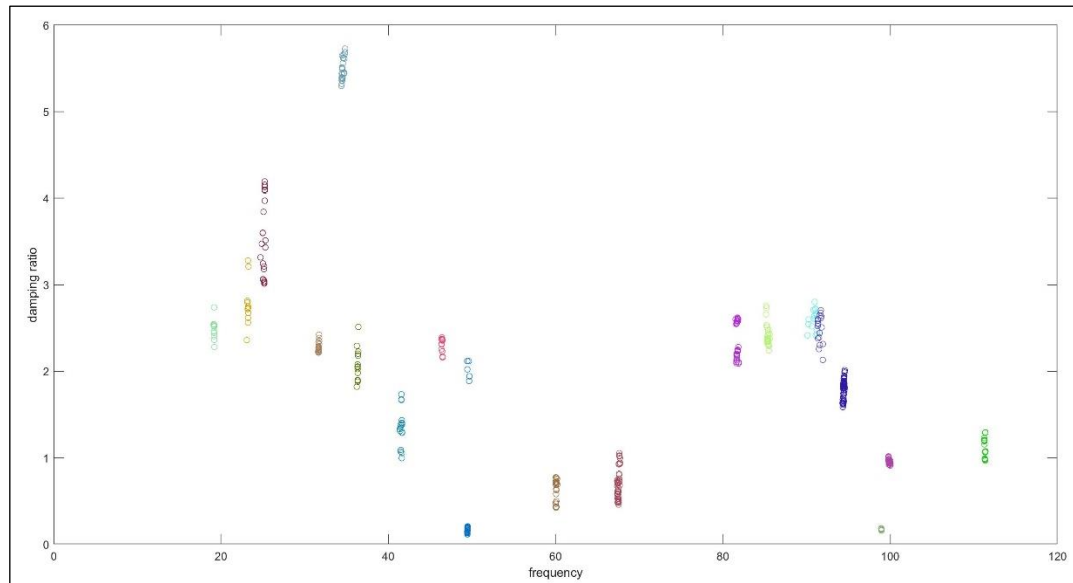
Figure 6-17 – clusters of the identified modes in the 121 combinations of min-max order

The number of modes within each cluster represents the level of identification. The clusters with a high number of internal modes have a higher probability to be representative of the structure.
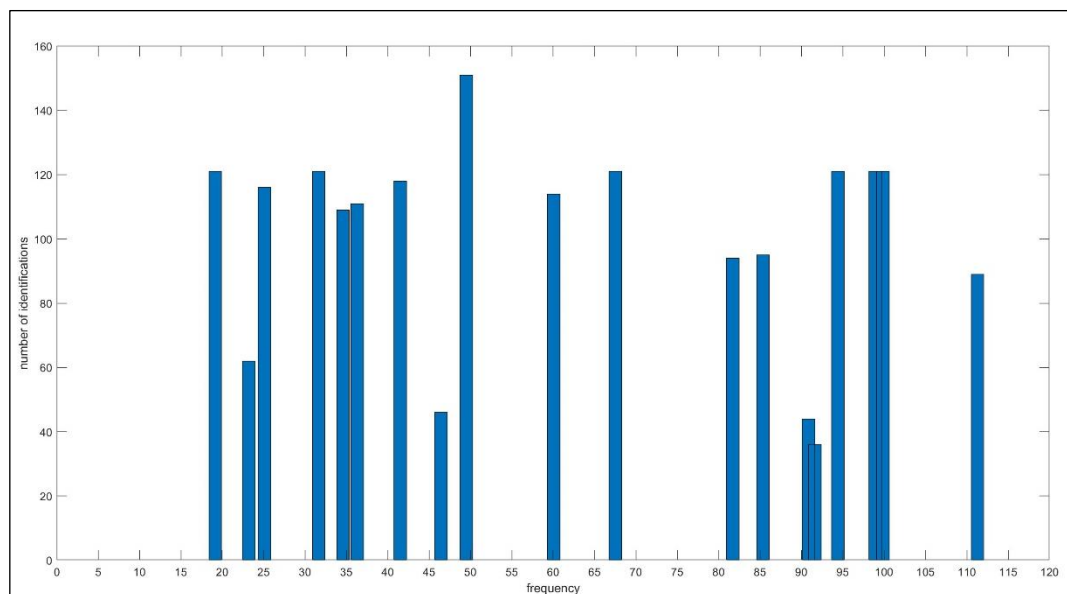
Figure 6-18 – number of identifications for frequency

116

As shown by Figure 6-18, most of the frequencies are identified in the majority of the analyses. The mode close to 49.5 Hz is identified for a number of time higher than 121. This is an evidence of the fact that in some analyses the mode is split in two clusters.

The modes with frequencies around 23.2 Hz, 46.4 Hz, 90.9 Hz and 91.6 Hz show a low repetition and it is possible that the aforementioned modes are mathematical modes.

7 CONCLUSIONS

This Thesis provides a method of Automatic Operational Modal Analysis which is able to identify the modal parameters of a structure which is analysed by a Stochastic Subspace Identification algorithm. SSI was considered as an output-only method where the examined signals were generated by an Ambient Vibration Test.

Starting from the definition of the SSI-Cov algorithm, a Literature Review was carried out in order to provide foundation of knowledge on the topic, identify the methods which can be adopted in the AOMA and try to give a different interpretation to the processes which could show limitations in some applications.

Then, the steps of the proposed method were discussed in order to clarify the logical path which starts from the results of the SSI-Cov algorithm and culminates with the identification of a finite number of modes and their representative modal parameters.

The numerical case of a simple building with 9 degrees of freedom was analysed with the aims of:

- Verifying the performance of the method
- Identifying the boundary conditions which help the method to achieve the identification of all the modes which characterise the structure with modal parameters as closest as possible to the numerical values. In this step, it was possible to assert that the method works well with a number of model orders between 25 and 35, and a dimension of the Hankel Matrix equal to 6 times the ratio between the maximum model order and the number of channels.
- Analysing the influence of the noise level which may affect the signal quality and influence the final result.

The simple experimental case of a helicopter blade was considered in order to show the performance of each step of the algorithm, clarify the

weight of each step on the final result, compare the results with an FE model. Furthermore, it showed the difference between the performance of a classical experimental identification, carried out manually on the basis of a stabilisation diagram, and the AOMA. In the same case, the influence of the maximum model order on the number of identified modes was analysed in order to start reflecting on the importance of a good choice of the minimum and maximum model orders.

Finally, the AOMA was carried out on the scale reproduction of a masonry arch bridge. The modal shapes identified by the proposed method were shown establishing a substantial difference with the results of an FE model. An important analysis was performed on the influence of the minimum and maximum model orders on the threshold level which establishes the number of clusters deriving from hierarchical clustering. Furthermore, the results of the AOMA for different combinations of min-max model orders are processed with a cluster analysis in order to show the frequency of identification of each mode.

The implementation of the AOMA results in a process which is completely independent from the user experience and it allows to identify the modal parameters without the choice of limits for the comparison parameters a priori.

The main achievements can be summarised as follows:

- The only threshold which is prefixed is the range of the damping ratio (0-20%). This range is widely fixed and the modes which are identified in the two experimental cases show values of damping ratio largely below the upper boundary (20%). This evidence leads to considerate that the aforementioned range does not influence in a strong way the process of estimation of the modal parameters.
- The comparison parameters which split the poles into stable and unstable poles affect the process with comparable weights. This objective is achieved thanks to:
  - A good normalisation of the comparison parameters;

- A good transformation of the variables (Box-Cox) that allows to redistribute the parameters in order to fit a Standard Normal Distribution.

- The stable and unstable poles are defined with mobile thresholds which adapt their values in function of the quality and the quantity of analysed poles. This allows to not define limits a priori and, consequently, it does not require user intervention.

- If on one hand the comparison parameters allow to identify the stable and unstable poles, on the other hand this analysis is carried out among poles belonging to two consecutive model orders and it assumes the only meaning of local stability. The hierarchical clustering takes part of the process in order to obtain a method which overcomes the above mentioned issue. In fact, the HC allows to analyse all the poles belonging to the different model orders in a single analysis. In other words, the agglomerative clustering and its relative and subsequently processes, assume a meaning of global stability, merging the poles with similar modal parameters.

- The removal of outliers allows to obtain a method which gives a meaning of internal stability to each cluster.

- As shown by the helicopter blade analysis, the choice of a method which identifies the modal parameters representing each mode-cluster does not affect the final results strongly. This is an evidence of the fact that the final poles within each cluster show a high similarity.

The proposed method works well with a range of model orders between 25 and 35 but it's not simple to identify the minimum and maximum model order to adopt in an analysis a priori. This challenge may be overcome considering:

- The quality of the Silhouette values which characterise the hierarchical clustering process and its relative cut-off.

- The value of the threshold which identifies the height of the cut-off - low values may mean a good identification of the poles and good performances of the analysis with the Soft Criteria.

The above parameters are difficult to automate and this could be a challenge to achieve with further implementations of the proposed method.

Furthermore, the analysis can be carried out with different ranges of model orders, as proposed in the final chapter relative to the masonry arch bridge (6.5.5). In this way, the modes which are identified in each analysis are then processed with a k-means clustering which groups the modes and gives an important information: the number of times that a mode is identified considering different ranges of model orders. It might be pleonastic to assert that the modes with a high repetition of identification are strongly representative of the structure.

REFERENCES

[1] J. H. G. D. R. Edwin Reynders, "Fully automated (operational) modal analysis," *Mechanical Systems and Signal Processing,* pp. 228-250, 31 January 2012.

[2] F. J. A. A. K. A. C. O. Eugen Neu, "Fully Automated Operational Modal Analysis using multi-stage clustering," *Mechanical Systems and Signal Processing,* pp. 308-323, 26 July 2016.

[3] B. D. M. P. Van Overschee, Subspace Identification for linear systems, Katholieke Universiteit Leuven (Belgium): Kluwer Academic Publishers, 1996.

[4] P. Andersen, "Technical Paper on the Stochastic Subspace Identification Techniques," Structural Vibration Solutions, [Online]. Available: http://www.svibs.com/resources/ARTeMIS_Modal_Help_v3/SSI_ssi.htm.

[5] F. M. C. G. a. Á. C. Alessandro Cabboi, "Automated modal identification and tracking: Application to an iron arch bridge," 28 February 2016. [Online]. Available: http://www.wileyonlinelibrary.com.

[6] G. P. L. Z. F. S. M. &. G. A. Rosario Ceravolo, "Vibration-Based Monitoring and Diagnosis of Cultural Heritage: A Methodological Discussion in Three Examples," *International Journal of Architectural Heritage,* pp. 375-395, 2016.

[7] G. V. D. a. D. Sabia, "A machine learning approach for the automatic long-term structural health monitoring," 2018. [Online]. Available: journals.sagepub.com/home/shm.

[8] Allyn W. Phillips and Randall J. Allemang, "Application of Modal Scaling to the Pole Selection Phase of Parameter Estimation," Structural

Dynamics Research Laboratory - Department of Mechanical Engineering - University of Cincinnati, Cincinnati, Ohio, U.S.A.

[9] C. F. A. M. A. E. R. W. L. R. Ka Yee Yeung, "Model-Based Clustering and Data Transformations for Gene Expression Data," Department of Computer Science & Engineering, University of Washington, Seattle, WA 98195, 2001.

[10] G. E. P. B. a. D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society,* Vols. Series B (Methodological), Vol. 26, no. 2, pp. 211-252, 1964.

[11] S. M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, University of California, Berkeley: Elsevier Academic Press, 2004.

[12] C. C. Aggarwal, Data Mining, Yorktown Heights, New York, U.S.A.: Springer, 2015.

[13] "MathWorks," Introduced before R2006a. [Online]. Available: https://uk.mathworks.com/help/stats/kmeans.html#buefs04-Distance.

[14] I. The MathWorks, "MathWorks," [Online]. Available: https://uk.mathworks.com/help/stats/hierarchical-clustering.html.

[15] K. M. Ting, "Encyclopedia of Machine Learning," Springer Link, 2010. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_652.

[16] G. Ruocci, Application of the SHM Methodologies to the Protection of Masonry Arch Bridges from Scour, Torino: Politecnico di Torino, 2010.

[17] A. Q. L. Z. F. R. C. A. D. S. Gianluca Ruocci, "Experimental Testing of a Masonry Arch Bridge Model Subject to Increasing Level of Damage," in *4th International Conference on Advances in Experimental Structural Engineering*, Ispra, 21027 Varese, Italy, 2011.

[18] I. ADINA R & D, "Theory and Modeling Guide," ADINA R & D, Inc., Watertown, MA 02472 USA, 2012.

APPENDICES

A.1 Mean Phase and Mean Phase Deviation

The characteristics of a complex vector can be estimated using the mean phase and the mean phase deviation. The description of these two parameters follows the same explication proposed by Allyn W. Phillips and Randall J. Allemang [8].

The mean phase (MP) is the average angle between the complex and the imaginary part of a vector Ψ and can be computed by the following steps.

In first, an eigenvalue problem is considered

$$[Im(\Psi) \quad Re(\Psi)]^T [Im(\Psi) \quad Re(\Psi)] \{v\} = \bar{\lambda} \{v\}$$

Error! No text of specified style in document.-1

Then, the eigenvector $\{v\}_{\lambda_{min}}$ associated to the smallest eigenvalue λ_{min} is selected and the two components used for the computing of MP

$$\{v\}_{\lambda_{min}} = \begin{Bmatrix} v_1 \\ v_2 \end{Bmatrix}$$

Error! No text of specified style in document.-2

$$MP = \theta_{MP} = tan^{-1}(\frac{v_2}{-v_1})$$

Error! No text of specified style in document.-3

The above relation returns the principal angle for the vector, bounded by ± 90°. If it is desired that the angle indicates the direction of largest response, it is necessary to pay attention to the sign of the numerator and denominator of the inverse tangent function. In general, the mean phase is

126

used to determine whether the modal vector is dominantly real-valued or imaginary-valued. A mean phase of 0° means that the vector is oriented around the real axis while a mean phase of ± 90° means that the vector is oriented around the imaginary axis.

Mean phase deviation (MPD) represents the scatter of the (modal) vector about the mean phase angle on a fraction or percentage basis. MPD can be computed as follows:

$$MPD = \frac{\left\| Im\left\{ e^{-i\theta_{MP}}\{\Psi\} \right\} \right\|}{\left\| Re\left\{ e^{-i\theta_{MP}}\{\Psi\} \right\} \right\|}$$

Error! No text of specified style in document.-4

A mean phase deviation equal to 0 means that the vector is a normal mode oriented about the mean phase angle while a mean phase larger than 0 means that the vector is a complex mode oriented about the mean phase angle.

A.2 Silhouette

The Silhouette coefficient is a criterium used to validate the quality of the clustering process. The method provides a graphical representation of how well an element is classified in a certain cluster. The definition of the Silhouette value starts from the computing of two distances:

- The average distance among the analysed element i within a cluster C_ι and the other elements j which belong to C_ι.

$$D_i^{in} = \frac{1}{n_\iota - 1} \sum_{j=1}^{n_\iota} D(i,j)$$

Error! No text of specified style in document.-5

Where n_ι is the number of poles within C_ι and $D()$ is the distance function adopted in the clustering algorithm.

D_i^{in} represents the similarity of i with the other elements which belong to the same cluster. Low values of D_i^{out} establish a good closeness among the elements

- The average distance among the element i and the elements k belonging to the closest cluster C_κ.

$$D_i^{out} = \frac{1}{n_\kappa} \sum_{k=1}^{n_\kappa} D(i,k)$$

$$= min_{\chi \neq \iota} \left(\frac{1}{n_\chi} \sum_{x=1}^{n_\chi} D(i,x) \right) \quad for \; \chi = 1,2,\ldots N_C$$

Where n_χ is the number of elements within a cluster $C_\chi \neq C_\iota$ and N_C is the total number of clusters identified by the analysis.

D_i^{out} represents the dissimilarity of i with the other elements which belong to the closest and different cluster. High values of D_i^{out} mean a good separation among i and the other clusters.

The Silhouette coefficient related to the element i can be computed with the following equation:

$$S_i = \frac{D_i^{out} - D_i^{in}}{\max(D_i^{out}, D_i^{in})}$$

This formulation leads to Silhouette coefficients between -1 and 1. Large positive values indicate highly separated clustering, and negative values are indicative of some level of "mixing" of data points from different clusters [12].

A.3 MAD criterion

The Mean Absolute Deviation (MAD) criterion is one of the most used method to detect outliers.

The definition of MAD starts from the computing of the $Median$. The Median of a set of numbers can be found arranging the numbers from the smallest to the greatest value and identifying the central value. If the number of analysed values is even, the Median is defined as the mean of the two central elements.

$$Median(\{X\}) = X_c$$

$$\{\vec{X}\} = sort_{smallest-gretest}\{X\} = X_1, X_2, \dots X_n$$

$$X_i < X_{i+1} \quad \forall\, i \in \mathbb{N}$$

$$X_c = \begin{cases} X_{(n+1)/2} & for\ n\ odd \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & for\ n\ even \end{cases}$$

MAD is defined as the median of the absolute difference between a vector Y and its relative median value Y_c.

$$MAD = Median(|\{Y\} - Y_c|)$$

The three scaled MAD is obtained by multiplying the MAD by a coefficient c defined as follows:

$$c = -\frac{1}{\sqrt{2} * erfc^{-1}(3/2)} \approx 1.4826$$

Where $erfc^{-1}$ represents the Inverse Complementary Error Function which is defined as

$$erfc^{-1}\big(erfc(x)\big) = x$$

The Complementary Error Function ($erfc$) is given by the following relation

$$erfc(x) = 1 - \mathrm{erf}(x)$$

Where $\mathrm{erf}(x)$ represents the Error Function of the argument x

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt =$$

Error Function