# Cluster Analysis for Automated Operational Modal Analysis: A Review

M. Danial A. Hasan<sup>1,2\*</sup>, Z. A. B. Ahmad<sup>2</sup>, M. Salman Leong<sup>1</sup>, L. M. Hee<sup>1</sup> and M. Haffizzi Md. Idris<sup>1</sup>

<sup>1</sup>Institute of Noise and Vibration, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

<sup>2</sup>School of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Abstract. Recent developments in the field of modal-based damage detection and vibration-based monitoring have led to a renewed interest in automated procedures for the operational modal analysis (OMA). The development of automated operational modal analysis (OMA) procedures marked a fundamental step towards the elimination of any user intervention since traditional modal identification requires a lot of interaction by an expert user. A key for effective automation of OMA is depended on welldefined modal indicators for a clear indication about which modes are to be selected as the physical modes. In all modal analysis, the construction of stabilization diagrams is necessary in order to illustrate, and decide, if a mode is physical or not for predefined range of the model order. On the other hand, the use of stabilization diagram tools involves a large amount of user interaction, costly, time-consuming process and certainly unsuited for online applications. Therefore, the development of automatic procedures for the analysis of stabilization diagrams by resembling decision-making process of a human has been carried out in recent years. For the sake of clearness, the automation of the interpretation of stabilization diagrams can generally be divided into two steps in order to speed up the process: a) elimination of noise modes and b) clustering of physical modes in order to obtain the most representative values of the estimated parameters of each clustered mode. In recent years, several alternative procedures have been proposed for clustering techniques. Therefore, this review aims to provide relevant essential information on the recent developments of cluster analysis in automated OMA. A literature review of existing clustering algorithm has been carried out to find best practice criteria for automated modal parameter identification which involving the general concepts of these techniques as well as the pro and cons of applying these clustering techniques are also discussed and summarised.

### **1** Introduction

Recent developments in the field of modal-based damage detection and vibration-based monitoring have led to a renewed interest in automated procedures for the operational modal analysis (OMA) or output-only identification of dynamic parameters. The development of automated operational modal analysis (OMA) procedures marked a fundamental step towards the elimination of any user intervention since traditional modal identification requires a lot of interaction by an expert user. It frequently used for repetitive test or numerous data sets for the same OMA test. This is crucial for the application of structural health monitoring (SHM) where the input data need to be processed or analyzed automatically so that the variations of modal parameter identification can be straightforwardly identified [1]. A key for effective automation of OMA is depended on welldefined modal indicators for a clear indication about which modes are to be selected as the physical modes.

In all modal analysis, the construction of stabilization diagrams is necessary in order to illustrate, and decide, if a mode is physical or not for predefined range of the model order [1,2]. Since the model of the system is often oversized, thus, the plot will contain noise modes and mathematical modes. The noise modes are caused by physical reasons, while the mathematical modes are generated to ensure the mathematical description of the measured data. Theoretically, the physical poles should be stabilized and can be easily identifiable along the vertical alignment of stable poles, whereas the computational or mathematical poles are scattered, showing the criterion of the unstable poles in the stabilization diagram. This is based on the comparison of the poles associated with a given model order with those attained from a one-order lower model [3].

<sup>\*</sup> Corresponding author: <u>muhd\_danial200@yahoo.com</u>

On the other hand, the use of stabilization diagram tools involves a large amount of user interaction, costly, time-consuming process and certainly unsuited for online applications. Therefore, the development of automatic procedures for the analysis of stabilization diagrams by resembling decision-making process of a human has been carried out in recent years. For the sake of clearness, the automation of the interpretation of stabilization diagrams can generally be divided into two steps in order to speed up the process: a) elimination of noise modes and b) clustering of physical modes in order to obtain the most representative values of the estimated parameters of each clustered mode [4]. In recent years, several alternative procedures have been proposed for clustering techniques. In the following section will be discussed about the main procedures of cluster analysis.

Thus, the automated OMA consists of the following steps and clearly illustrated in Figure 1:

- (1) Measure the responses of the structure and estimate the modal parameters using a high model order, n modes.
- (2) Construct Stabilization diagram by estimating poles with an increasing model order and illustrate, and decide, either mode is physical or computational modes
- (3) Classify the n modes in physical and computational modes using a clustering algorithm.

Stabilization diagram	Clusters Analysis	Selected modal estimates
20 15 0 10 0.5 1 1.5 2.5 Frequency (Hz)	st 10 te s te s	(%) a) a) b) b) c) c) c) c) c) c) c) c) c) c

**Fig. 1.** Steps of the proposed methodology for automated OMA.

This review aims to provide relevant essential information on the recent developments of cluster analysis in automated OMA. A literature review of existing clustering algorithm has been carried out to find best practice criteria for automated modal parameter identification which involving the general concepts of these techniques as well as the pro and cons of applying these clustering techniques are also discussed and summarised.

## 2 Cluster analysis

Cluster analysis is a technique to classify or group objects regarding their characteristics. The classified or

clustered objects should then reveal high internal (within-cluster) homogeneity and high external (between cluster) heterogeneity [5]. In the case of parametric identification techniques that apply the model of several orders, the aims are to cluster the mode estimates that possess a similar physical mode. For instance, all the modes estimate the corresponding natural frequency and modal damping ratio based on the results provided by five model orders as shown in Figure 2 below. The easiest way to do cluster analysis is by making the figure simpler by showing only the physical modes. In a real application, some further points randomly scattered would be present. Typically, the cluster analysis is to group the points that are near to each other (circles of the figure). The concepts graphically illustrated for a case where only two variables are considered. The maximum number of variables used for graphical interpretation is three, more than that is considered impossible. The variables or modal quantity can be natural frequency, modal damping ratio, mode shape (projected onto a fixed vector), modal participation and mode shape scaling [1].



**Fig. 2.** Scheme to illustrate the application of clustering algorithms.

Most commonly used clustering algorithms can be classified into three general categories: hierarchical, partitioning methods and histogram analysis that will be discussed in the next section.

#### 2.1. Hierarchical clustering

Hierarchical clustering algorithms are constructed by a hierarchy of a treelike structure. Theoretically at initial, each object is considered as a cluster. Then, the two nearest clusters (or individuals) are joint together to become a new aggregate cluster which can reduce the number of clusters by one in each step until the distance between all remaining clusters is larger than a user-defined threshold value. Finally, all individuals are grouped into one large cluster as shown in Figure 3 below [6].



Fig. 3. Hierarchical clustering.

Thus, the execution of the of the hierarchical algorithms consist of the following key steps:

- calculation of the similarity between every pair of objects in the dataset,
- (2) connecting of the objects in a hierarchical tree and,
- (3) lastly, the definition of a rule to cut the hierarchical tree at a certain level, assigning all the objects of each branch to a single cluster.

There is a different application of hierarchical algorithms such as single linkage, complete linkage, average linkage, Ward's method, and the centroid method depends on the distance between clusters [5].

In a literature, Verboven P. et al. used hierarchical clustering for analysis stabilization diagrams by the application of the LSCF method (Least Squares Complex Frequency Domain) to datasets collected in an experimental modal analysis [7]. In the proposed method, only the group of the estimated pole is considered, while mode shapes are neglected and only applied in a second phase to assess the quality of the formed groups. Meanwhile, Pappa et al. were possibly the primary applied such an approach, using the eigenfrequency difference and the MAC value as distance measures [8]. Even though, such an approach was not clearly defined as 'hierarchical clustering' but it worked effectively for automating the Eigensystem Realization Algorithm (ERA) for an experimental modal analysis (EMA) of the Space Shuttle tail rudder [9]. The following research was expanded by applying genetic algorithms, to find the 'optimal' ERA parameter values (besides n) [10]. Moreover, Chauhan and Tcherniak [11] did a bit changes from the original approach of Pappa et al. [8]. Goethals et al. introduced another way of distance measure, integrating the eigenfrequency and damping ratio difference [12]. Thus, such an approach capable to detect closely spaced modes via the presence of modes with the same

model order in the same cluster; and then distinguished using the MAC value. Allemang et al. applied another distance measure, namely the MAC value between extended, pole-weighted mode shape vectors that are obtained from each mode instead of the mode shape [13]. Verboven et al. proposed an alternative approach, where it is presumed that the number of modes in one cluster is a previously identified but this is rarely to occur [14]. An effective application of hierarchical clustering was stated by Magalhães F. et al., who analyzed more than 2500 high-quality data sets collected on a 280 m-span concrete arch bridge [15]. Besides that, the similar researchers are then proposed alternative hierarchical algorithms to calculate the distance between already formed clusters by using the single linkage [6]. By using this approach, the distance between two clusters is equal to the smallest distance between objects inside the two clusters. In the selection of the tree cut level for the hierarchical tree is based on the maximum limit for the distance between any point and its closest point of the same cluster. It has the great advantage of requiring only two user-defined parameters, which are the maximum limit for the distance between any point and its closest point of the same cluster and the number of expected modes as well as does not require the previous construction of a stabilization diagram, because all mode estimates, stable or unstable, are considered. However, it has the drawback of demanding much more user-defined parameters particularly in the selection of stable poles.

The hierarchical algorithms have the benefit of being deterministic and allowing a good selection of the final number of clusters, based on the previously constructed hierarchical tree. However, they have the drawback of being computationally demanding in the presence of many individuals because the similarity of each pair needs to be computed. In addition, it very sensitive to outliers.

#### 2.2. Partitioning methods

Partitioning methods are from non-hierarchical clustering procedures and often referred to as K-means clustering. Generally, this method works by assigning objects into a predefined number of clusters (K) using the following procedure.

- (1) Specify the number of random seeds (kernals) or provide seeds as the initial cluster centers.
- (2) Assign samples to 'nearest' seed by previously specified threshold distance.
- (3) Iteratively reassign samples to groups in order to minimize within-group variability (i.e., assigned to a group with 'closest' centroid).

The result is a set of clusters that are as compact and well-separated as possible. The following procedure was illustrated in Figure 4 below.



Fig. 4. Partitioning clustering.

In references, non-hierarchical algorithms were applied for autonomously analysis data in stabilization diagrams [12,16,17]. In the early phase, the estimated modes are displayed like in the scheme presented in figure 2 and a clustering technique is joint with selflearning algorithms which allow a better selection of the algorithm parameters and the evaluation of physical or computational of the attained clusters. An enhancement of the K-means algorithm, named as Fuzzy C-mean clustering algorithm was then introduced which operated by giving a membership grade instead of relating an object to a certain group when distinguishing physical poles from computational modes resulting from a frequencydomain maximum likelihood estimator (MLE) for a single model order that applied to experimental modal analysis data [16,18]. The difference is based on a total of six characteristics that include, for instance, the standard deviation of the pole estimate, which is an output of the implemented identification algorithm, and indices that assess the complexity of the mode shape estimates. Otherwise, Scionti and Lanslots applied fuzzy C-means clustering to group the modes, present in a stabilization diagram, directly into a userdefined number of clusters and represented in damping vs frequency diagram [19]. However, this approach had drawbacks of predefined the number clusters, several non-intuitive improvements to the basic C-means clustering algorithm and did not provide a reliable outcome regarding a combination with genetic algorithms. The following approach, a Fuzzy C- means clustering algorithm with using the representation of the poles in the z-plane was then

introduced to cluster all the mode estimates from several datasets instead of using classical representation on a damping vs frequency diagram due to the coefficient of variation of the damping estimates is significantly larger than frequency estimates [17]. However, the shape of clusters leads to more spherical nature.

The partitioning methods have the advantage of being fast processing algorithm than hierarchical clustering for many variables. Besides that, it possibly will produce tighter clusters than hierarchical clustering. However, these clustering procedures have the drawback of the need to predefine the number of clusters and the necessity to choose the clusters seeds. Besides that, they have the limitation of not being deterministic nature of the solution, as leads to produce inconsistent results due to the frequent use of a random selection of the seeds. In addition, most of these clustering techniques are prone to finding elliptical and spherical clusters.

#### 2.3. Histogram analysis

Histogram analysis is based on the counted number of (stable) modes in a narrow bin of the frequency axis in the stabilization diagram [20].

In references, Scionti et al. [21] use this as the basis for an automated modal parameter estimation procedure that can reduce the user-defined parameters, including the bin width. Its performance was assessed based on manually selected modes for dispersed data. A great combination was shown for the PolyMAX identification method [22] that can provide a clear stabilization diagram but biased for modal damping ratio estimates [23]. On the other hand, histogram analysis was brought poor performance for the least-squares complex exponential (LSCE) identification method [24].

# 3 Conclusions

This review will serve as a base for future studies in enhancing the automation of OMA method as modal information engine in structural health monitoring (SHM) systems by reducing some common drawbacks of available automated OMA methods as stated below.

- Identification of actual modes was based on several statically set parameters;
- A time-consuming calibration process for each monitored structure was required at startup;
- The static identification of thresholds and parameters was often inadequate to follow natural changes in modal properties of structures due to damage or environmental effects.

Thus, an alternative approach was required to avoid the tuning of analysis parameters at startup and these have recently been recognized and accepted also by other authors on how important neglecting predefined parameters [20].

The authors would like to extend their greatest gratitude to the Institute of Noise and Vibration UTM for funding the current study under the Higher Institution Centre of Excellence (HICoE) Grant Scheme (R.K130000.7843.4J227). Additional funding for this research came from the UTM Research University Grant (Q.K130000.2543.11H36) and the Fundamental Research Grant Scheme (R.K130000.7840.4F653) from The Ministry of Higher Education, Malaysia.

## References

- 1. R. Bricker, C. Venture, *Introduction to operational modal analysis*, (2015)
- 2. B. Peeters, *System Identification and Damage* Detection in Civil Engineering, Thesis (Ph.D.), (2000)
- 3. C. Rainieri, G. Fabbrocino, *Operational Modal Analysis of Civil Engineering Structures*, (2014)
- 4. A. Cabboi, *Automatic Operational Modal Analysis: Challenges And Applications To Historic Structure and Infrastructure*, Thesis (Ph.D.), (2012)
- 5. JF. Hair, RE. Anderson, RL. Tatham, WC Black, *Multivariate data analysis*, (1998)
- 6. F. Magalhães, *Operational modal analysis for testing and monitoring of bridges and special structures*, Thesis (Ph.D.), (2010).
- P. Verboven, P. Guillaume, B. Cauberghe, E. Parloo, S. Vanlanduit, *Proceedings of the IMAC-XXI Conference and Exposition on Structural Dynamics*, 3–6 (2003)
- 8. RS. Pappa, GH. James, DC. Zimmerman, J. Spacecr Rockets, **35**, 163–169 (1998)
- J-N. Juang, RS. Pappa, J. Guid. Control Dyn., 8, 620–627 (1985)
- 10. KS. Chhipwadia, DC Zimmerman, GH. James Iii, 17th IMAC proceedings, **3727**, 819 (1999)
- 11. S. Chauhan, D. Tcherniak, *IMAC proceedings*, (2008)
- 12. I. Goethals, B. Vanluyten, B. De Moor, *ISMA* proceedings, 991–1003 (2004)
- 13. RJ. Allemang, DL. Brown, AW. Phillips, Survey of Modal Techniques Applicable to Autonomous/Semi-Autonomous Parameter Identification, (2010)
- P. Verboven, B. Cauberghe, E. Parloo, S. Vanlanduit, P. Guillaume, Mech. Syst. Signal Process., 18, 759–780 (2004)
- F. Magalhaes, A. Cunha, E. Caetano, Mech. Syst. Signal Process., 23, 316–329 (2009)
- P. Verboven, E. Parloo, P. Guillaume, M. Van Overmeire, Mech. Syst. Signal Process., 16, 637–

657 (2002)

- 17. EP. Carden, JMW. Brownjoh, J. COMPUT-AIDED CIV. INF., **23**, 360–372 (2008)
- S. Vanlanduit, P. Verboven, P. Guillaume, J. Schoukens, J. Sound Vib., 265, 647–661 (2003)
- M. Scionti, JP. Lanslots, Adv. Eng. Softw., 36, 768–779 (2005)
- 20. E. Reynders, J. Houbrechts, G. De. Roeck, Mech. Syst. Signal Process., **29**, 228–250 (2012)
- M. Scionti, J. Lanslots, I. Goethals, A. Vecchio, H. Van der Auweraer, B. Peeters, et al., 8th RASD proceedings, (2003)
- 22. B. Peeters, H. Van der Auweraer, P. Guillaume, J. Leuridan, J. SHOCK VIB., **11**, 395–409 (2004)
- 23. T. De Troyer, *Frequency-domain modal analysis* with aeroelastic applications, Ph.D. Thesis, (2009)
- 24. J. Lanslots, B. Rodiers, B. Peeters, *ISMA proceedings*, (2004.)