



Thank you for downloading this document from the RMIT Research Repository.

The RMIT Research Repository is an open access database showcasing the research outputs of RMIT University researchers.

RMIT Research Repository: <http://researchbank.rmit.edu.au/>

Citation:

Neu, E, Janser, F, Afaghi Khatibi, A and Orifici, A 2017, 'Fully automated operational modal analysis using multi-stage clustering', *Mechanical Systems and Signal Processing*, vol. 84, no. A, pp. 308-323.

See this record in the RMIT Research Repository at:

<https://researchbank.rmit.edu.au/view/rmit:37839>

Version: Accepted Manuscript

Copyright Statement: © 2016 Elsevier Ltd. All rights reserved
Creative Commons Attribution-NonCommercial-NoDerivatives
4.0 International License

Link to Published Version:

<http://dx.doi.org/10.1016/j.ymssp.2016.07.031>

PLEASE DO NOT REMOVE THIS PAGE

Fully Automated Operational Modal Analysis using Multi-Stage Clustering

Eugen Neu^a, Frank Janser^a, Akbar A. Khatibi^b, Adrian C. Orifici^b

^a*FH Aachen UAS, Hohenstaufenallee 6, Aachen, Germany*

^b*RMIT University, GPO Box 2476 Melbourne, Victoria Australia, 3001*

Abstract

The interest for robust automatic modal parameter extraction techniques has increased significantly over the last years, together with the rising demand for continuous health monitoring of critical infrastructure like bridges, buildings and wind turbine blades. In this study a novel, multi-stage clustering approach for Automated Operational Modal Analysis (AOMA) is introduced. In contrast to existing approaches, the procedure works without any user-provided thresholds, is applicable within large system order ranges, can be used with very small sensor numbers and does not place any limitations on the damping ratio or the complexity of the system under investigation. The approach works with any parametric system identification algorithm that uses the system order n as sole parameter. Here a data-driven Stochastic Subspace Identification (SSI) method is used. Measurements from a wind tunnel investigation with a composite cantilever equipped with Fiber Bragg Grating Sensors (FBGSs) and piezoelectric sensors are used to assess the performance of the algorithm with a highly damped structure and low signal to noise ratio conditions. The proposed method was able to identify all physical system modes in the investigated frequency range from over 1000 individual datasets using FBGSs under challenging signal to noise ratio conditions and under better signal conditions but from only two sensors.

Keywords: automatization; operational modal analysis; stabilisation diagram; clustering; structural health monitoring

1. Introduction

Continuous Structural Health Monitoring (SHM) presupposes the automatic extraction of damage sensitive features. In the case of vibration-based SHM these features usually are the modal parameters of the system (natural frequencies, mode shapes and damping ratios). In recent years significant progress has been made in developing and refining modal parameter identification methods that use unmeasured environmental loads as the primary source of structural excitation. These methods are today known under the name of Operational Modal Analysis (OMA) [1]. OMA itself requires manual user interaction but multiple OMA-based automatization algorithms have been proposed and successfully applied to complex structure like bridges [2, 3] and wind turbines [4]. Despite some progress, the proper (and ideally fully automatic) choice of automatization parameters and thresholds as well as the identification of complex and heavily damped modes is an area of ongoing research.

The main challenge for Automated Operational Modal Analysis (AOMA) from parametric system identification algorithms is the separation between physical and mathematical modes. This challenge is commonly addressed through parameter identification at a large number of system orders n . The approach is based on the empirical observation that physical modes are identified with nearly identical properties at every system order. Their modal properties are stable. Mathematical modes on the other hand are not identified in a consistent way. Traditionally inconsistency thresholds for each modal parameter are provided by the user to separate physical from mathematical modes [5]. Furthermore, additional mode validation criteria like Mean Phase Deviation (MPD), Mode Phase Collinearity (MPC), etc. are often used to single out the physical system modes. This data are then summarized in a stability diagram, which allows the user to manually select the physical modes.

A variety of methods have been proposed to automatize the OMA process. Overviews were published in [3] and [6] and will not be repeated here. The approach to AOMA described in this work can be summarised into the following steps [2, 3]:

Email addresses: neu@fh-aachen.de (Eugen Neu), janser@fh-aachen.de (Frank Janser), akbar.khatibi@rmit.edu.au (Akbar A. Khatibi), adrian.orifici@rmit.edu.au (Adrian C. Orifici)

1. Identify mode candidates from a large number of system orders.
- 30 2. Remove as many mathematical modes as possible.
3. Use hierarchical clustering to divide the remaining modes into homogeneous sets.
4. Remove the small sets, which typically consist of mathematical modes.

The method proposed in [2] requires at least one user-defined parameter, the maximum within-cluster distance between representations of the same physical mode from different system orders. Such parameters have to be selected for every new sensor setup and system under investigation and may be sensitive to varying operational and environmental conditions. Their proper choice requires expert knowledge and, depending on the specific application, considerable manual effort may be required. Reynders et al. [3] suggested to automatically derive this parameter from the actual data. However, the proposed algorithm is limited to (nearly) real mode shapes and includes a damping ratio threshold. These are acceptable restrictions for some engineering structures, including e.g. bridges that were investigated in [2] and [3]. However, in aerospace applications damping and complexity are dominated by the fluid-structure interaction and are often substantially larger than under no-wind conditions. For such applications these constraints may be too restrictive.

A Stochastic Subspace Identification (SSI) method is commonly used to identify the mode candidates in a large range of system orders. However, the influence of the utilized system order range has not been investigated in the context of AOMA. Instead, in previous publications [2, 3] the maximum system order was chosen to be much larger than the number of expected physical modes in the investigated frequency range. Further, the insensitivity of the proposed methodologies to varying system order ranges was not proven, and no methods were discussed to detect the upper and lower bounds of the usable system order range.

In this work an innovative multi-stage clustering approach for AOMA is introduced that can be used with any parametric system identification algorithm. No user-defined thresholds are required and neither the to-be-identified damping ratios nor the mode shape complexities are limited in any way. Furthermore, the sensitivity of step one of the pro-

posed method to changes of the chosen system order range is explored and compared to existing approaches using a large number of independent datasets and two different measurement setups. One major novel contribution of this work is the consistent formulation of a clustering feature vector to separate between physical and mathematical modes in step two of the investigated method, and the subsequent application of transformation and normalisation techniques to the heavily skewed feature vector. It is at this point that the demand for small complexity as well as for a damping ratio threshold can be dropped. For hierarchical clustering we derive a statistically profound threshold value from the measured Probability Density Functions (PDFs) to separate the remaining probably physical modes into homogeneous sets. The performance of the algorithm is assessed using a large number of wind tunnel measurements with a composite cantilever that was equipped with a low number of piezoelectric sensors and a high number of Fiber Bragg Grating Sensors (FBGSs). This presents a challenging data set in terms of a highly damped system with variable noise levels, as well as more broadly representing one possible future utilisation scenario for the two sensing technologies in SHM.

2. Methodology

2.1. Experimental data

To assess the performance of the proposed AOMA methodology experimental data from a wind tunnel study are used. The experimental setup is shown in figure 1. The investigation was conducted in a closed-loop wind tunnel with an open test section. The investigated specimen was a glass fiber-reinforced polymer plate (500 mm \times 90 mm \times 4 mm), which was subjected to different flow conditions. The structural response of the specimen was measured using three sensor types: FBGSs, a unidirectional piezoelectric (PZT) accelerometer and a piezoelectric strain sensor. In addition, the dynamic properties of the inflowing wind were measured using a hot-wire anemometer. A detailed description of the experimental setup was published in [7].

The experimental setup was designed to represent two limiting cases of possible sensor setups. On one hand two piezoelectric sensors, with high dynamic range but only limited

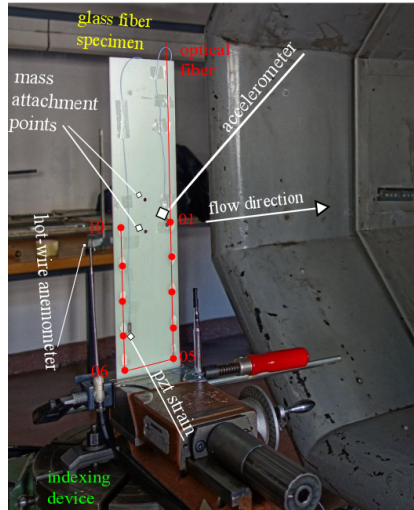


Figure 1: Wind tunnel setup.

spatial information. On the other hand ten FBGSs, with more spatial information but significantly worse dynamic range due to the investigated interrogator, which is based on Charge-Coupled Device (CCD) technology. The differences in the Signal-to-Noise Ratio (SNR) of the two cases are apparent from the Power Spectral Densities (PSDs) in figure 2. Furthermore, a preceding investigation showed the first bending mode to be strongly damped by aerodynamic forces [7], which is especially demanding for previously existing fully automated OMA procedures that rely on fixed and arbitrary chosen damping ratio thresholds [3, 6]. In summary, the first investigated use case are ten FBGSs, a sampling rate of 400 Hz, no anti-aliasing filter and a comparatively low SNR. The second use case consists of only two sensors, a unidirectional accelerometer and a piezoelectric strain sensor, measured at a sampling rate of 1600 Hz with analog anti-aliasing and a high SNR.

The modal parameters were automatically extracted from the FBGSs and the PZT sensors individually. In both cases the measurement data were first passed through a fourth-order Butterworth high-pass filter with 0.5 Hz cutoff frequency to remove the strain offset from the mean wind load and the slow temperature drift of the FBGSs. The time-domain data were then slightly cropped to remove the transient filter response. This resulted

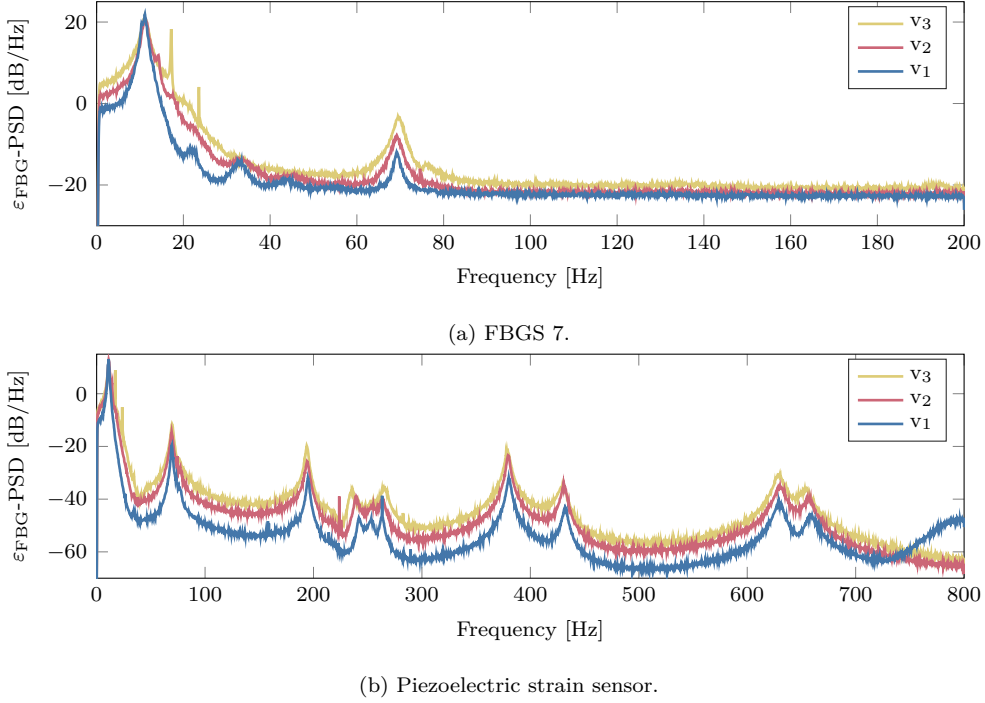


Figure 2: PSDs at three velocities and otherwise constant operational conditions.

in datasets of approximately 40 seconds in length with 450 repetitions of the period corresponding to the first natural frequency in each dataset.

105 2.2. Definitions

The relative difference between scalar (real or complex) values X_i and X_j is calculated using the formulation in equation (1) throughout this text.

$$dX_{i,j} = \frac{|X_i - X_j|}{\max(|X_i|, |X_j|)} \quad (1)$$

Equation (1) is used to measure the relative natural frequency distance df_u , the relative damping distance $d\xi$, the relative pole distance $d\lambda$ and the relative mean phase deviation $dMPD$. The Modal Assurance Criterion (MAC), which defines a relative correlation
 110 between two modes, is defined according to Eq. (2).

$$MAC_{i,j} = \frac{|\phi_i^T \cdot \phi_j^*|^2}{(\phi_i^T \cdot \phi_i^*) (\phi_j^T \cdot \phi_j^*)} \quad (2)$$

where ϕ_i and ϕ_j are mode shapes, which can be either real or complex. The Mean Phase Deviation (MPD) is a measure of mode shape complexity. It describes the mean phase angle deviation of the individual mode shape components from a straight line in the complex plane. It is calculated using a total least squares fit of the mode shape in the complex plane (Eq. (3)) and a weighted sum of phase angle deviations (Eq. (4)) [3].

$$\mathbf{USV}^T = [\text{Re}(\phi_i) \quad \text{Im}(\phi_i)] \quad (3)$$

$$MPD_i = \frac{\sum_{n=1}^{N_\phi} w_n \arccos \left| \frac{\text{Re}(\phi_{jn})V_{22} - \text{Im}(\phi_{in})V_{12}}{\sqrt{V_{12}^2 + V_{22}^2} |\phi_{in}|} \right|}{\sum_{n=1}^{N_\phi} w_n} \quad (4)$$

where V_{12} and V_{22} are the individual components of the right singular matrix of the singular value decomposition \mathbf{USV}^T . N_ϕ is the number of mode shape components, w_n are weighting factors that are chosen as $|\phi_{in}|$ in this work.

3. Automated Operational Modal Analysis

The approach to AOMA described in this work follows the four-step procedure described in section 1, where some steps involve multiple procedures:

1. Identify mode candidates from a large number of system orders.
2. Remove as many mathematical modes as possible.
 - (a) Remove certainly mathematical modes using hard validation criteria.
 - (b) Split modes into consistent and non-consistent sets using k-means clustering.
3. Divide the remaining modes into homogeneous sets using hierarchical clustering.
 - (a) Derive cutoff distance from the probability distribution of the consistent modes.
 - (b) Cluster the mode candidates based on a complex distance measure.

- 130 (c) Remove all but one mode from a single system order in one cluster.
4. Remove the small sets, which typically consist of mathematical modes.
- (a) Reject sets that are smaller than a threshold derived from the largest set size.
- (b) Use outlier rejection to remove natural frequency and damping outliers.
- (c) Select a single mode representative from the remaining modes in each cluster.

135 The steps are described in the subsequent sections and demonstrated using the experimental data from the wind tunnel investigation.

3.1. System Identification

The parametric system identification method used in this work is the data-driven Stochastic Subspace Identification - Canonical Variate Analysis. The method is based on the procedure described in [8]. The fundamentals of the SSI methods have often been described and will not be repeated here. The important concept and the common base for all parametric models, with the model order as the only parameter, is that the algorithm expects a single input parameter (the model order n), and responses with n sets of modal properties (in the OMA case n poles $\lambda_1 \dots \lambda_n$ and n unscaled mode shapes $\phi_1 \dots \phi_n$).

145 The number of block rows in the Hankel matrices was chosen to be $i = 2 \cdot n_{max}/N$ in accordance with the suggestion in [8], where n_{max} is the maximum investigated system order and N is the number of sensors in the investigated setup. The manual investigation of multiple randomly chosen datasets from the two investigated sensor setups and the variation of the parameter in a range surrounding the chosen value confirmed the response to be nearly invariant to i in the investigated number of block rows range.

150

3.2. Hard validation criteria for certainly mathematical modes

Whether a mode represents a physical mode or a mathematical mode can usually not be deduced from its isolated modal properties. However, there are certain indicators for mathematical modes. Stable systems do not have negative damping. Poles without imaginary part do not represent a system capable of oscillation. Physical poles always occur in complex conjugate pairs at a single model order n . These three criteria, which

155

test whether a mode is certainly mathematical, are sometimes called Hard Validation Criteria (HVC) [3] and can be expressed using the following formulas:

$$\text{Re}(\lambda_i) \geq 0 \quad \text{Im}(\lambda_i) = 0 \quad \lambda_i \stackrel{n\lambda_i = n\lambda_j}{\neq} \lambda_j^* \quad (5)$$

Poles that meet this criteria are removed immediately. The application of the HVC
 160 (and the a priori removal of the negative frequency range) as a first step reduces the computational effort of the algorithm. Beside the improved performance, the application of the HVC before or after k-means clustering did not have a significant influence on the final choice of physical modes in our tests.

3.3. K-means based mathematical pole removal

165 The second step of the presented AOMA algorithm is to separate the modes into two sets, probably physical modes and modes that are marked as certainly mathematical. It is important to note that it is not necessary to remove all mathematical modes at this stage of the algorithm. This will be done in subsequent clustering stages. The primary characteristic of physical modes, which distinguishes them from mathematical ones, is
 170 their similarity to their siblings at other system orders. Hence, for each mode $\lambda_{n,i}, \phi_{n,i}$ at the current model order the nearest neighbour $\lambda_{n+1,j}, \phi_{n+1,j}$ from the next higher order is found. If a similar mode is found at the next higher order, chances are high that the mode at hand is a physical mode. Otherwise the mode is probably mathematical. The distance measure used by Reynders et al. [3] and in this work is

$$dpMAC_{i,j} = d\lambda_{i,j} + (1 - MAC_{i,j}) \quad (6)$$

175 where $d\lambda$ is the pole distance according to Eq. (1) and the modes i and j are from two consecutive model orders n and $n + 1$ respectively.

The next step of the proposed algorithm is to use the information about the nearest-neighbour to create a Soft Validation Criteria (SVC) vector. Reynders et al. [3] suggested

to build such a vector from “as many relevant single-mode validation criteria as possible”. However, our investigation shows that more care must be taken to properly select, transform and normalise the variables for the feature vector. According to our definition physical modes can be separated from mathematical ones based on their similarity to modes at other orders. This is exactly what relative difference measures (Eq. (1) and (2)) describe. Single mode criteria like Modal Transfer Norm (MTN) [3] or MPD (Eq. (4)) on the other hand are strength and complexity measures of individual modes. Clustering based on these properties will divide the modes into weak and powerful modes and into real and more complex modes. However, the system under investigation may have weakly excited and/or complex modes, which would then be incorrectly flagged as mathematical.

A second argument against the use of combined feature vectors from variables with different informative value is their sometimes very dissimilar probability distribution. Figure 3a shows the MPD and relative MPD difference (dMPD) distributions from a large number of orders ($n = \{2, 4, 6, \dots, 300\}$). Certainly mathematical modes according to the HVC in Eq. (5) were removed beforehand. The MPD shape clearly resembles a multimodal distribution. The mathematical modes seem to be normally distributed in the middle of the possible MPD range, whereas the physical modes, which are nearly real in this case, are squeezed to the left side of the diagram. The dMPD shape, on the other hand, resembles some type of exponential distribution. When k-means clustering is applied to a feature vector consisting of only these two variables the result is dominated by the MPD distribution (figure 3b). The datasets are predominantly separated into real and complex but not into consistent and inconsistent. The reason for this behavior is that variables with larger variances always dominate k-means clustering. Further, when applied to normally distributed data, k-means tends to split the datasets in approximately equally sized clusters [9]. This is not true for exponentially distributed data. To allow for the occurrence of weakly excited and complex physical modes and to give all variables equal weight we therefore suggest to formulate the feature vector in the following way:

$$\mathbf{p}_i = \left[d\lambda_{i,j} \quad df_{u,i,j} \quad d\xi_{i,j} \quad (1 - MAC_{i,j}) \quad dMPD_{i,j} \right]^T \quad (7)$$

where $d\lambda$, df_u and $d\xi$ are the normalised pole distance, natural frequency distance and damping ratio distance. In this form the feature vector only has one informative value, namely the proximity to the nearest neighbour, and the shape of all variables approximately resembles the same probability distribution.

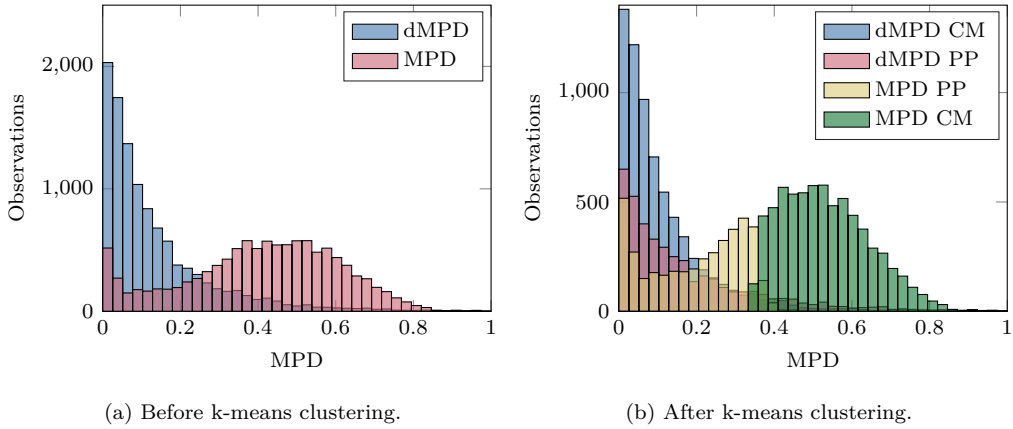


Figure 3: Result of k-means clustering when variables with strongly deviating distributions are used. PP stands for probably physical, CM for certainly mathematical.

The nearest-neighbor detection process results in heavily skewed feature distributions. A comparison of different exponential family distributions showed that the Weibull distribution seems to be the best fit for the variables in Eq. (7) as well as in Eq. (6). Figure 4a shows the best fits of the Weibull, the exponential and the half-normal distribution to the combined distance measure according to Eq. (6). When k-means clustering is applied to data that is skewed to such an extent the resulting clusters will not be of approximately equal size. Instead, a very large and a very small cluster will be created. Figure 4b shows the results of such a clustering. The smaller cluster is barely visible in the diagram and only begins near the 3σ boundary.

For the investigated problem this means that the vast majority of modes will still be flagged as possibly physical after k-means clustering and nearly no mode candidates will be removed. Whether this behavior is desired depends on the expected ratio of physical

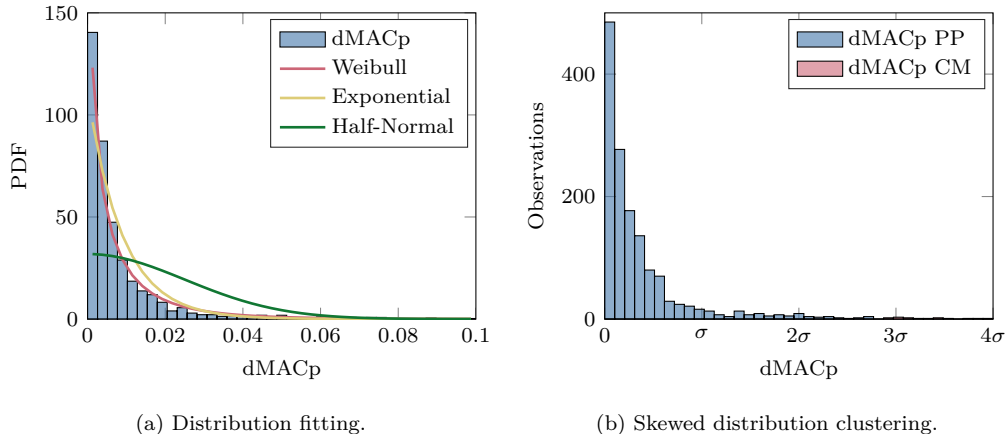


Figure 4: Result of k-means clustering when variables with different distributions are used.

to mathematical modes. The total number of physical modes N_p and mathematical modes N_m can be expressed by the following (approximate) relationships:

$$N_p \propto n_{max} \cdot L \quad (8)$$

$$N_m = \frac{1}{2} n_{max} \cdot (n_{max} + 1) - N_p \quad (9)$$

where n_{max} is the maximum investigated order and L the number of unique physical system modes in the investigated frequency range. Hence, the low orders are dominated by physical poles and the higher orders are dominated by mathematical ones. If the system order is much larger than $2L$, which is the basic requirement for the stabilisation-based mode separation, the number of mathematical modes will be larger than or in the same order as the number of physical modes. From this it follows that separation based on the skewed distribution will not result in the desired detection of a significant number of mathematical modes. Therefore, we suggest to transform the feature vector (7) into a shape that more resembles a normal distribution. The power transformation is done according to Eq. (10) using the approach described by Box and Cox [10].

$$h_{T,i}(m) = \begin{cases} (p_i^{\gamma_m}(m) - 1) \cdot \gamma_m^{-1} & , \gamma_m \neq 0 \\ \ln(p_i(m)) & , \gamma_m = 0 \end{cases} \quad (10)$$

The optimal transformation parameter γ_m for each individual feature variable $p_i(m)$ is found by a profile log-likelihood maximisation. For the dataset that is used as an example

throughout this chapter the following γ -vector is found:

$$\gamma = [0.06 \quad 0.05 \quad 0.19 \quad 0.10 \quad 0.23]$$

A simpler approach, where all features are transformed into the logarithmic scale (Eq. (11)) showed satisfactory results as well.

$$\mathbf{h}_{\mathbf{T},i} = \ln(\mathbf{p}_i) \quad (11)$$

Features with large dispersion will dominate the clustering process [9]. A feature like $d\xi$ which is known to have a significantly larger variance than df_u would have a larger influence on the clustering process. This is the exact opposite behavior to what is common practice in manual analysis or when $d\xi$ and df_u thresholds are manually chosen for AOMA [2]. Therefore in this work a final normalisation to standard scores is applied to the feature vector, to give every variable equal weight, using the standard deviation $\sigma(h_{T,i}(m))$ and the mean $\bar{h}_{T,i}(m)$ of the individual (transformed) features:

$$h_{N,i}(m) = (h_{T,i}(m) - \bar{h}_{T,i}(m)) / \sigma(h_{T,i}(m)) \quad (12)$$

Equation (12) shows the final form of the proposed feature vector $\mathbf{h}_{\mathbf{N},i}$. The goal of k-means clustering is to minimise the within-cluster sum of squares (Eq. (13)).

$$\{S_1, S_2\} = \arg \min_S \sum_{k=1}^2 \sum_{\mathbf{h}_{\mathbf{N},i} \in S_k} \|\mathbf{h}_{\mathbf{N},i} - \boldsymbol{\mu}_k\|^2 \quad (13)$$

Equation (13) returns two sets, S_1 and S_2 , which contain the probably physical and the certainly mathematical modes. $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the centroids of the sets S_1 and S_2 and are initialized with $+\sigma(\mathbf{h}_{\mathbf{N},i})$ and $-\sigma(\mathbf{h}_{\mathbf{N},i})$ respectively.

The results of the k-means clustering process based on this feature vector are shown in the scatterplot matrix in figure 5. The upper right triangle matrix of the scattermatrix (figure 5) shows the correlation coefficients between the individual features in equation

(12). Features with high correlation will create more “weight” in the clustering process [9]. Hence, the $d\lambda$ and df_u features have a strong influence on the clustering process, which in this case is a desired effect, since the frequency distance is an excellent indicator of stability.

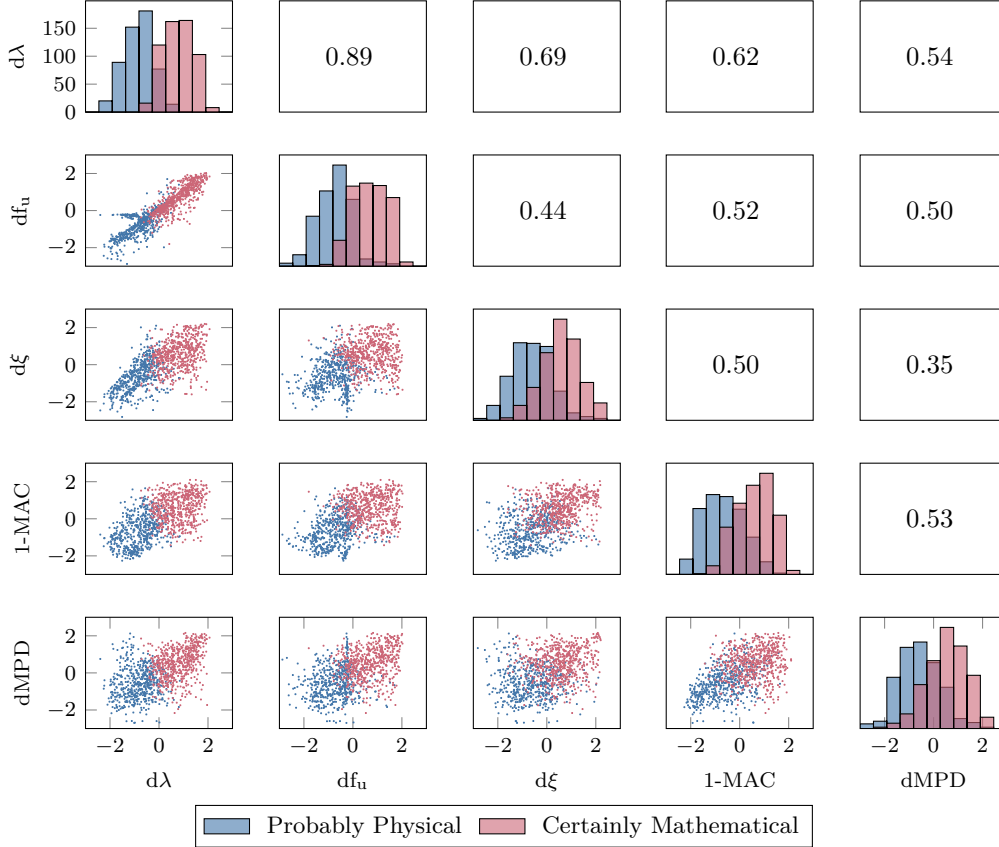


Figure 5: Scatterplot matrix of soft validation criteria. The data have been transformed into log-scale and normalised by the standard deviation. The upper right triangle numbers are correlation coefficients between the individual feature variables.

There is no distinct boundary region between the two sets marked as Probably Physical and Certainly Mathematical in figure 5. Instead, both sets merge seamlessly. Hence, it is likely that some mathematical modes will be marked as probably physical or/and that valid physical modes will be marked as mathematical, depending on the ratio of physical to mathematical modes in the initial joint set. Since the feature vectors were

transformed to resemble a normal distribution (Eq. (10)), the resulting sets S_1 and S_2 will be of approximately equal size [9]. Therefore the maximum order n_{max} should be chosen from a range where the number of mathematical mode representatives N_m exceeds the number of physical mode representatives N_p (see Eq. (8) and (9)). The sensitivity of the algorithm to system order changes and the proper choice of system orders is investigated in section 3.8.

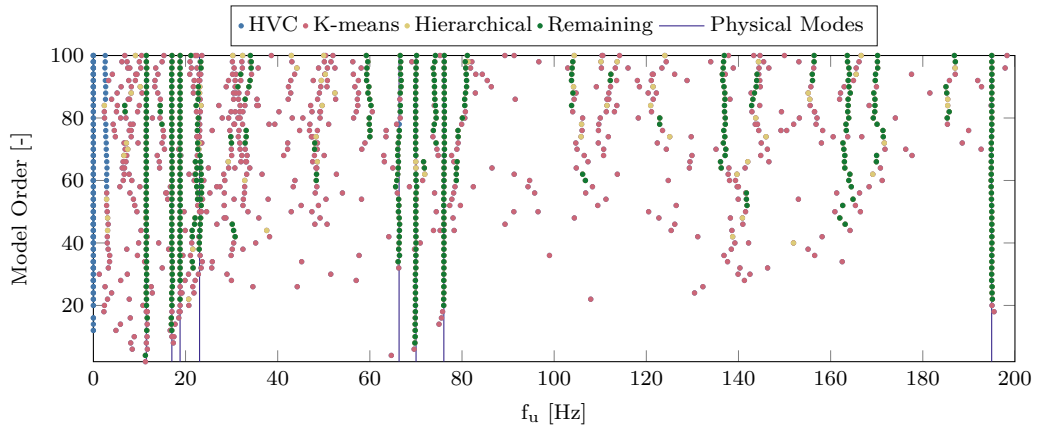
The stabilisation diagram in figure 6 shows that the algorithm successfully marked the majority of spurious modes as mathematical. Nearly no modes that appear to be stable in the diagram were not marked as such. The clustering process was tested for large ranges of model orders and always proved to be able to remove the majority of mathematical modes.

3.4. Hierarchical clustering based mode separation

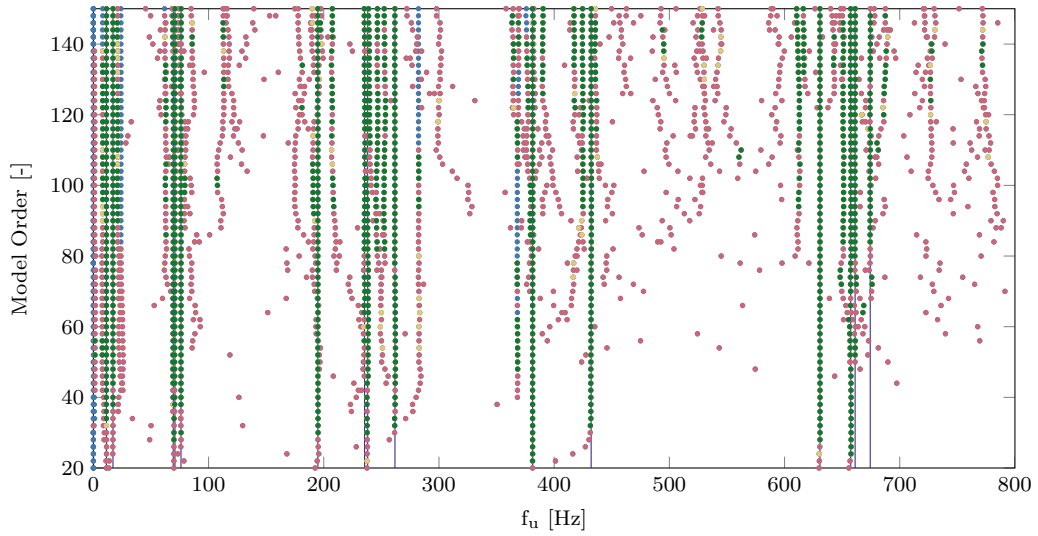
In section 3.3 a clustering algorithm was applied to separate the mode candidates into probably physical and certainly mathematical modes. The number of clusters (two) was therefore known in advance. The goal in this section is to separate the remaining modes into clusters which represent the individual physical modes of the system. Their number is not known beforehand in the vast majority of cases. The classic clustering approach for cases where the number of clusters is not known in advance is agglomerative hierarchical clustering [9]. All agglomerative hierarchical clustering procedures can be described in three steps:

1. Each observation starts in its own cluster.
2. The two nearest clusters are combined into a new cluster.
3. The procedure is repeated until all observations are contained in a single cluster.

The basic idea behind the application of hierarchical clustering to AOMA is to stop the clustering process when the distance between the nearest two clusters is larger than a certain threshold. This threshold can be understood as the distance up to which modes from different orders are considered to belong to the same physical mode. Often such thresholds are manually fit to the specific example under analysis [2]. Reynders et al.



(a) Stabilisation diagram from FBGSs



(b) Stabilisation diagram from the accelerometer and the PZT strain sensor.

Figure 6: The colors indicate at which point the individual modes have been marked as mathematical. The clusters which were automatically classified as physical are marked with vertical lines.

275 [3] calculated the threshold from the sum of the mean and two standard deviations of the probably physical mode distances. In this work the threshold is derived from the distribution of probably physical modes, since these were found to be not normally distributed. The inverse cumulative distribution function of the Weibull distribution is used to find the 95th percentile nearest-distances between probably physical modes

$$P(dpMAC_{PP,i,j} \leq \tilde{d}_{dpMAC}) = 0.95 \quad (14)$$

280 where \tilde{d}_{dpMAC} is the 95th percentile threshold and $dpMAC_{PP,i,j}$ is the distance between two neighbours from different model orders according to Eq. (6). Figure 4a shows exactly this distribution. Applied to the data shown in figure 4a the threshold value calculated using Eq. (14) is 0.046.

Figure 7 shows the sensitivity of the cutoff distance (Eq. (14)) to changes in the considered model order range. For this investigation the cutoff distance was derived from 285 different model order intervals ($n = \{2, 4, 6, \dots, n_{max}\}$), where n_{max} was varied from 50 to 300. The cutoff distance (14) was determined from 64 independent Fiber Bragg Grating (FBG) and PZT sensor measurements to assess the variance of the process. The large magnitude difference between the two curves is the result of the different number of 290 sensors, different number of physical poles in the investigated frequency range, different noise levels, etc. A user who wants to determine the threshold value for a new system or sensor setup would have to manually investigate an interval that is larger than the one spread between the FBG and PZT curves. The FBG threshold is nearly constant between a maximum order of 100 and 250 and starts to rise afterwards. The threshold 295 value drops to smaller values below $n_{max} = 100$. This behavior can be deduced from the discussion in section 3.3. At very low maximum orders the majority of modes are physical, whereas at very high maximum orders the majority of modes are mathematical. The clustering process discussed in section 3.3 only works well within these boundaries. The PZT curve is constant throughout the investigated order range. The investigated 300 upper limit for n_{max} is quite excessive. More typical n_{max} values are 100 [2] and 160 to 200 [3]. In this range both curves are basically independent of the maximum order. This underlines that the proposed method is insensitive to the investigated model order in a

wide model order range. How this compares to existing methods is discussed in section 3.8.

In addition to the definition of similarity or distance, the distance between multi-member clusters also needs to be calculated. Magalhães et al. [2] use the single-linkage method (Eq. (15)), where the smallest distance between two clusters defines their overall distance. Reynders et al. [3] on the other hand use average linkage (Eq. (16)), which defines the distance between two clusters as the average distance between all members of one cluster with all members of the other cluster.

$$d_{r,s} = \min(\text{dist}(x_{ri}, x_{sj})) \quad , i \in \{1, \dots, n_r\}, j \in \{1, \dots, n_s\} \quad (15)$$

$$d_{r,s} = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \text{dist}(x_{ri}, x_{sj}) \quad (16)$$

305 where n_r and n_s are the total number of individual members x_r and x_s in the clusters r and s respectively. The single-linkage approach is known to create “chains” through the data, when the observations are not clearly separated [9]. Our investigation showed that, especially when higher model orders were considered, physical modes in near proximity to each other were grouped into a single cluster when the linkage procedure could “jump”
 310 over mathematical modes to build a single large cluster. The average linkage procedure is computationally more expensive but was much better able to create compact clusters of individual physical modes. The complete-linkage procedure, where the largest distance between two clusters defines their overall distance, was discarded because of its sensitivity to outliers [9].

315 The process of hierarchical clustering is often visualised using a dendrogram, where the node height represents distance at which two clusters are joined. A dendrogram, with the corresponding cutoff distance according to Eq. (14) is shown in figure 8. To create the dendrogram the model order range was chosen to be very small ($n = \{20, 22, 24, \dots, 50\}$) for visualisation purposes and the derived threshold distance is lower than normal.

320 The result of the hierarchical clustering process will be either large clusters that will almost exclusively consist of modes that represent physical system modes, or small clusters consisting of mathematical modes. If the maximum order n_{max} is not chosen high enough some smaller clusters may actually be representations of very weakly excited

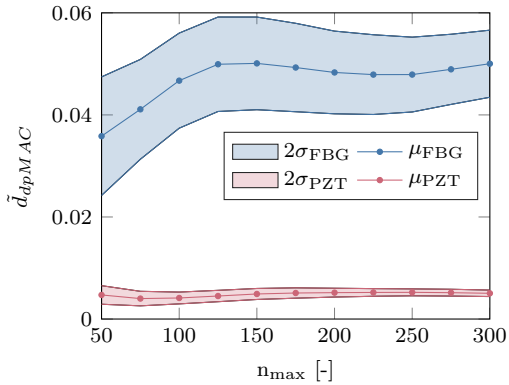


Figure 7: Order influence on cutoff threshold.

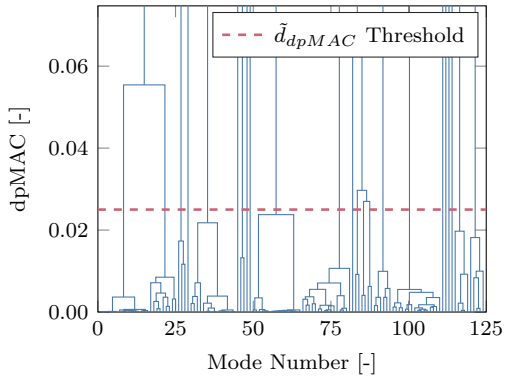


Figure 8: Hierarchical clustering dendrogram.

physical modes, which only occur at higher system orders. However, when large maximum orders are used, the so-called pole splitting phenomenon can occur. In this case a single physical system mode will be represented by two physical modes at higher system orders. Pole splitting is shown in figure 9. Furthermore, in rare cases, a mathematical mode could by chance have properties that are very similar to a physical system mode. Under this circumstance physical and mathematical modes of the same system order n would be joined into a single cluster. To make sure that only one representation of a pole is present in each cluster, repeated poles at single system orders are sought out and all but the one with the highest proximity to the cluster centroid according to Eq. (6) are removed from the cluster (figure 9).

3.5. The choice of physical clusters

Two types of clusters will be created by the hierarchical clustering process: Large clusters that represent physical system modes and small clusters consisting of mathematical modes. Figure 10 shows the number of modes in each cluster after hierarchical clustering for a FBGS dataset. Here the dividing line between physical and mathematical clusters is derived from the number of observations in the largest cluster. A 50% threshold is shown in figure 10, which was used to separate the clusters into physical and mathematical ones. Of course, in this case two sets are barely above the threshold value and could have been marked as mathematical if they would have been only slightly smaller. The distance

between physical and mathematical clusters can be increased when the minimum model order n_{min} is not set to 2 but to a higher value, e. g. 20. This way the size differences
 345 between the physical clusters will get smaller (compare figure 6a and 6b). With these conditions considered, further investigation showed that the gap between physical and mathematical sets spans a region from approximately 75% to 25% of the largest set size in the majority of the investigated FBG and PZT datasets. Hence, the number of physical mode sets returned by the algorithm is nearly invariant to the threshold, as long as
 350 the threshold percentage lies between 25% and 75%.

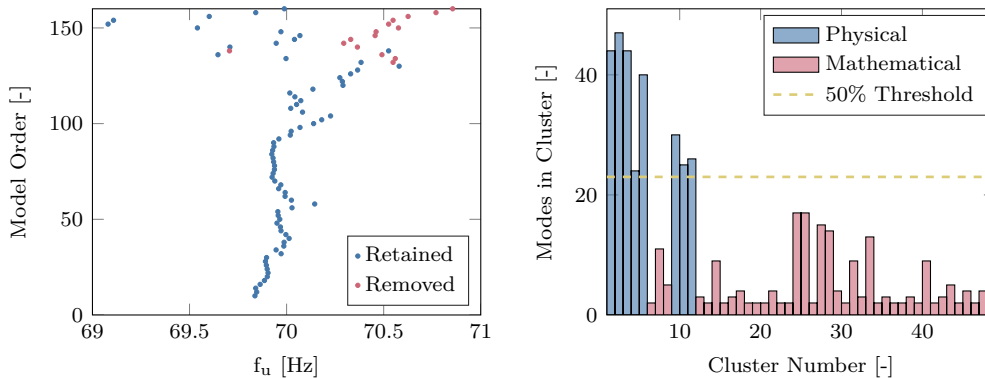


Figure 9: Pole splitting and duplicated pole removal. Figure 10: Separation into physical and mathematical clusters.

3.6. Outlier rejection

Outlier deletion is controversial, especially when the underlying probability distribution is unknown or small sets are investigated. To our best knowledge it has never been studied whether modal properties from different model orders, which are associated with the
 355 same physical system mode, tend to be normally distributed or not. Physical poles often follow trends with increasing system order, which sometimes are suddenly disrupted (see e. g. figure 9). Our examination shows no dominating probability distribution. However, empirical evidence also shows that sometimes “obvious” outliers are present in a cluster (e. g. seventh cluster from left in figure 11) and that the identification variance from
 360 large numbers of measurements can be improved when outlier rejection is applied to the identifications from the individual measurements. Hence, we apply the modified

Thompson Tau technique [11] to remove frequency and damping ratio outliers from the physical clusters.

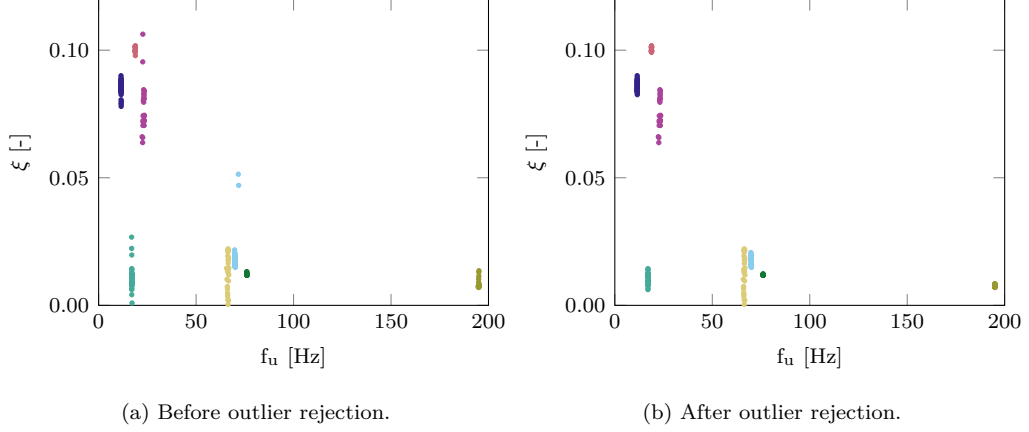


Figure 11: Outlier rejection using modified Thompson Tau technique.

The modified Thompson Tau technique considers only one outlier at a time and is repeated until no more outliers are found. The algorithm first looks for the observation with the largest absolute value deviation from the mean:

$$\delta = \max (|X_i - \bar{X}|) \quad (17)$$

In our case the dummy variable X is either the natural frequency f_u or the damping ratio ξ . In the next step the modified Thompson τ is calculated from the student's t PDF.

$$\tau = \frac{t_{\alpha/2} \cdot (n - 1)}{\sqrt{n} \cdot \sqrt{n - 2 + t_{\alpha/2}^2}} \quad (18)$$

Here n is the number of observations and $t_{\alpha/2}$ is the critical student's t value, which is a function of the number of observations n and the significance level α . $t_{\alpha/2}$ can be calculated from the inverse of student's t cumulative distribution function. α is set to 0.01 to limit the removal to strong outliers. The final step of the algorithm is to test whether

the absolute value deviation is larger than τ multiplied by the standard deviation of X ,
 375 in which case the data point is rejected (Eq. (19)). The algorithm is repeated, starting
 from Eq. (17), until no more outliers are found.

$$\delta > \tau \cdot \sigma(X) \tag{19}$$

3.7. Choosing the final modal representation

Each physical cluster obtained from the hierarchical clustering step in section 3.4 contains
 a large number of modes. Hence, the questions arises how to chose a single representa-
 380 tion of the individual modal properties. Magalhães et al. [2] used the average natural
 frequency, damping ratio and mode shape calculated from all observations in each phys-
 ical cluster. Reynders et al. [3] chose the mode with the median damping value as single
 epitome of the physical cluster. Finally, Schwochow and Jelicic [12], who proposed a
 stabilisation diagram based AOMA methodology, suggested to use the modes from the
 385 lowest possible model order, which still has an observation in each physical cluster. That
 way all chosen modes are from a single system model order and build a “consistent”
 representation of the system. All the aforementioned methods have their advantages
 and disadvantages, which depend on the planned application of the algorithm. Each of
 them can be used with the proposed AOMA methodology. In this work, the approach
 390 described in [2] was used.

3.8. Model order sensitivity and comparison to existing algorithms

The influence of the utilized system order range was never addressed in [2] or [3]. Instead,
 the maximum system order was chosen to be much larger than the number of expected
 physical modes in the investigated frequency range. The figures 12a and 12b show a
 395 comparison between the probably physical cluster S_1 and the certainly mathematical
 cluster S_2 for the feature vector proposed in [3] and the feature vector proposed in this
 work (Eq. (12)). For this investigation the ratios were derived from different model order
 intervals ($n = \{2, 4, 6, \dots, n_{max}\}$), where n_{max} was varied from 50 to 300. Furthermore,
 the identification was determined from 64 individual datasets measured under constant

operational conditions to assess the variance of the process. The results for the newly
 400 derived feature vector are in good agreement with the discussion in section 3.3. The ratio
 of probably physical modes N_{PP} to certainly mathematical modes N_{CM} is approximately
 one and nearly invariant to the maximum model order. The ratios derived from the
 clustering process with the feature vector proposed in [3] show a different behavior. The
 405 ratio obtained from piezoelectric sensor data (figure 12b) is constant and below 0.5,
 whereas the FBGS ratio increases with n_{max} and moreover shows a strong variance. The
 reason for the large variance is unclear. The smaller ratio in the PZT data (figure 12b)
 can be explained with the reduced influence of the MPC, which only returns a measure
 of complexity for setups with three or more sensors. The results show that the newly
 410 developed feature vector shows a more consistent behavior for different sensor setups and
 throughout the maximum order range. However, other measures have to be applied to
 examine the consistency of the two resulting sets.

The figures 12c and 12d show the 95th percentile \tilde{d}_{dpMAC} according to Eq. (14) derived
 from the probably physical set S_1 using the distance measure introduced in Eq. (6).
 415 In other words, the two figures 12c and 12d show a comparison of nearest-neighbour
 distances in the probably physical set when the new and the feature vector proposed by
 [3] are used to separate probably physical from certainly mathematical modes. Two things
 are striking: Reynders et al.'s feature vector shows a strong n_{max} sensitivity, whereas the
 feature vector proposed in this work is nearly n_{max} invariant and the nearest neighbours
 420 have much larger distances in the sets derived with Reynders et al.'s feature vector
 than in the sets derived with the consistent, transformed and normalised one. \tilde{d}_{dpMAC}
 is used as the stopping criterion for the hierarchical clustering procedure (section 3.4)
 and can therefore be directly compared to the manual cutoff distance used in [2] (0.02)
 and the automatically derived OMA threshold shown in [3] (0.24). Manual thresholds,
 425 modified to be comparable to Eq. (6), which are used in free or commercial tools to
 create stabilisation diagrams [13, 14] are all below 0.06. The distance derived from the
 newly developed feature vector is much closer to the one chosen in the manual analysis,
 whereas the distance calculated with the feature vector proposed by Reynders et al. [3]
 is in the same range as the one published in [3] but much larger than what would be
 430 used in a manual analysis.

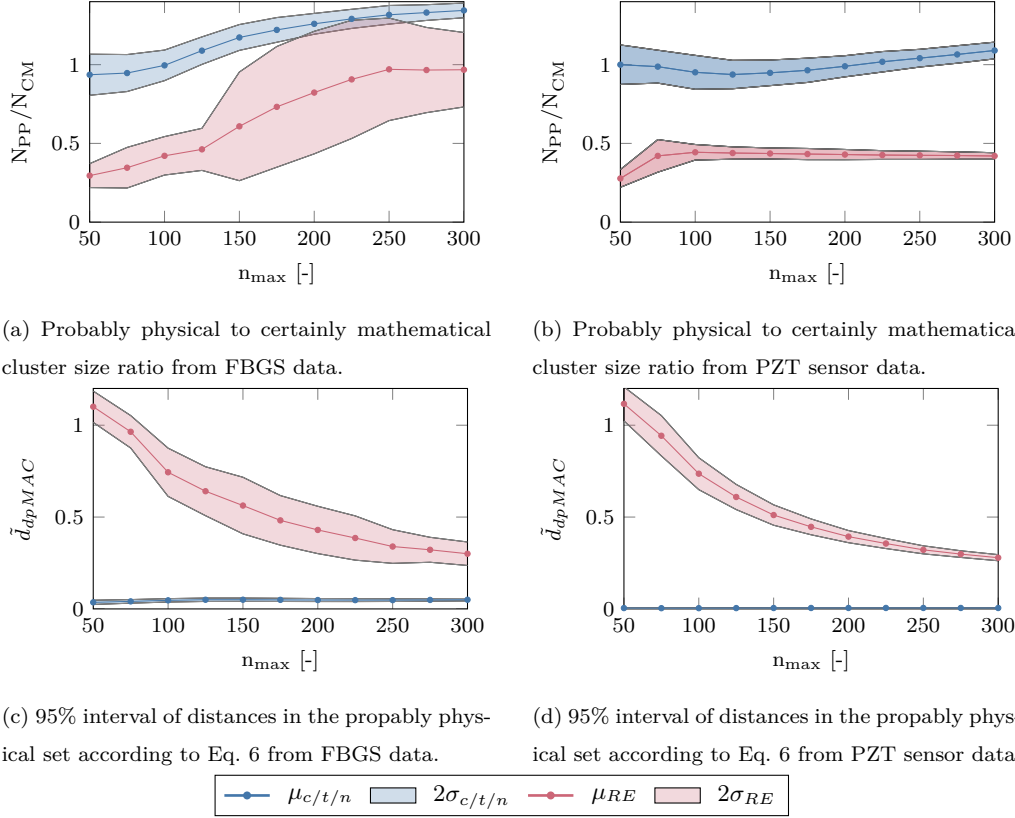


Figure 12: Model order range sensitivity comparison between the consistent, transformed and normalised feature vector ($c/t/n$) and the feature vector (RE) proposed in [3].

4. Modal Analysis Results

The AOMA methodology described in section 3 was applied to a total of 1152 datasets, 576 measured with FBSs and 576 measured with piezoelectric sensors, to assess the performance of the proposed algorithm. The investigated datasets are from measurements at a constant velocity (v_3), three Angles Of Attack (AOAs) and with two additional masses, resulting in 9 different operation points, each measured 64 times. The detected natural frequencies are shown in figure 13 and can be compared to Experimental Modal Analysis (EMA) and Finite Element Analysis (FEA) results shown in table 1. The EMA results were obtained from a hammer impact test, which was carried out in the wind tunnel but without any wind excitation. The data were collected and processed according to the

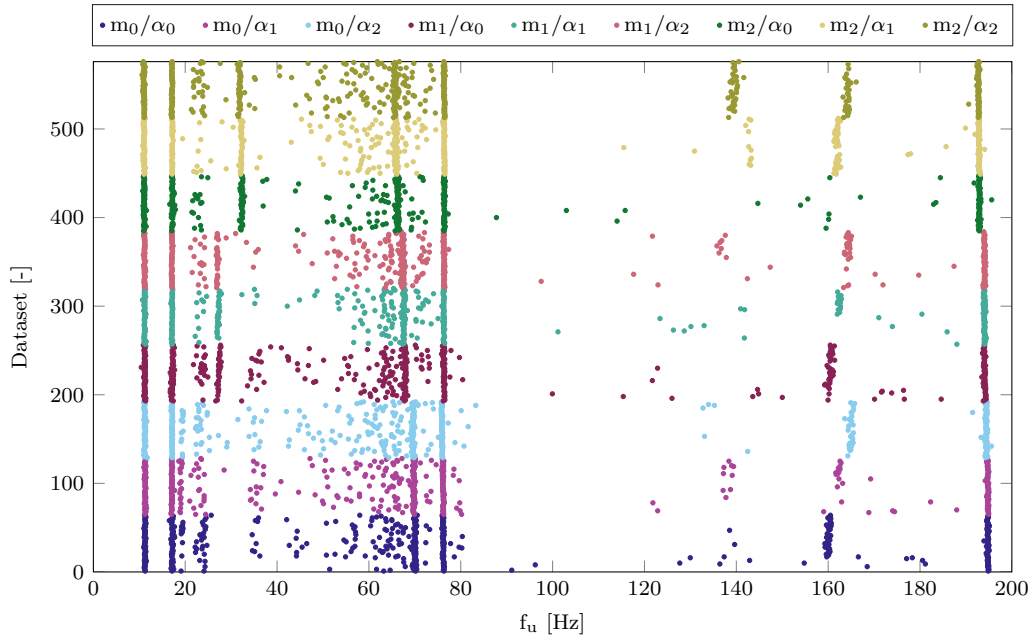
method described in [15]. The FEA results were obtained from numerical modal analysis using the commercial software ANSYS Mechanical. The system was modelled using a full 3D solid representation of each fiber layer, orthotropic material and fixed clamping conditions. The material properties were tuned to fit the EMA results.

Table 1: Natural frequency results from EMA and FEA. The deformation type of each mode is given in the first row, where B stands for bending, T for torsion and F for a for-and-aft in-plane bending mode.

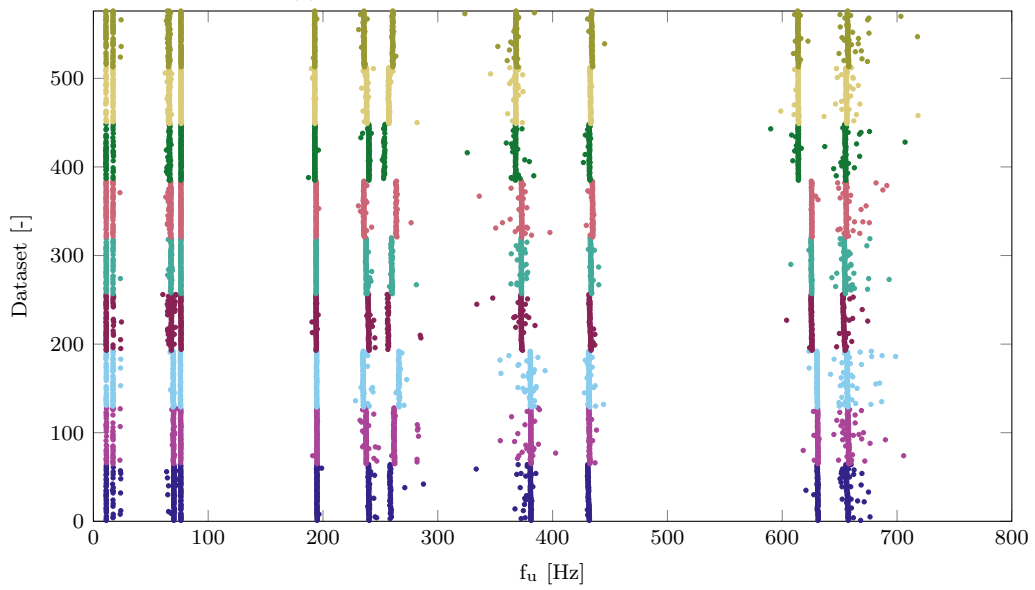
Mode		B1	B2	T1	B3	T2	F1	T3	B4	T4	B5
EMA	f_u [Hz]	11.1	69.0	78.9	193.7	244.4	-	379.9	435.0	628.5	661.6
FEA	f_u [Hz]	11.1	69.4	79.0	194.2	245.3	261.1	380.1	435.2	627.5	660.8

445 Figure 13a shows the results of the FBG identification. A number of modes are identified consistently from nearly every dataset, whereas others are only identified sporadically. According to the EMA and FEA analysis four physical modes should be present in the frequency range from 0 Hz to 200 Hz (table 1). These four modes are successfully identified from nearly every dataset. However, a significant number of additional modes
450 are detected as well. The consistently detected mode at 17 Hz was identified as a narrow-banded excitation caused by the rotating wind tunnel blades [7]. Since OMA identifies a joined system, consisting of the excitation and the structural system response, the wind tunnel excitation is classified as a system mode. Other identified modes can be explained with the lack of an anti-aliasing filter in the utilized FBG interrogator. Hence, all the
455 high frequency modes are folded into the low-frequency range and those that are excited above the noise floor are detected by the algorithm and identified as physical modes. The proposed AOMA methodology, in combination with the data-driven SSI method used, show excellent mode detection capabilities under challenging SNR conditions. For example, the bending mode B3 is consistently identified from FBG data, even if the mode
460 is barely excited above the noise floor (see figure 2a).

Figure 13b shows the result of the identification from the two piezoelectric sensors. The investigated structure was equipped with only a single accelerometer and a single piezoelectric strain sensor. In order to obtain relevant consistency indicators from the MAC and the MPD the modal properties were identified from the joined strain and accelerometer
465 measurements. At the k th natural energy equilibrium state the modal parameters



(a) Identified modes from 576 FBG datasets.



(b) Identified modes from 576 piezoelectric sensor datasets.

Figure 13: The colors indicate different operational conditions. Data from three different mass setups and three AOAs are shown. All other operational conditions were kept constant.

measured with both sensors are entirely the same [16]. Thus, the MAC and MPD in Eq. (6) and Eq. (7) are applied to a pattern of motion at the k th equilibrium state where one component of the resulting mode shape represents an acceleration and the other represents a strain. Eleven vertical lines are visible in the diagram, the ten expected physical modes according to table 1 and the narrow-banded excitation from the wind tunnel at 17 Hz. Hence, the algorithm successfully detected every physical mode in the investigated frequency range. It is noticeable that considerably less spurious or noise modes are visible in figure 13a when compared to figure 13b, especially in the region between 20 Hz and 70 Hz. A comparison to the PSDs in figure 2 reveals that the majority of these noise modes can be attributed to the missing anti-aliasing filter.

Figure 13 shows clearly visible stepwise changes for some of the identified natural frequencies that correspond to varying mass configurations and AOA. Further evaluation of the data in figure 13b shows that many of the natural frequencies and damping ratios identified under different operational conditions build distinguishable clusters. In this work an aeroelastic application of the proposed automation algorithm is investigated. Under certain operational conditions strong aerodynamic damping ($\xi > 0.1$) can and in fact did occur (see figure 11). Still, the automation technique was able to reliably detect these modes. The general applicability of OMA for in-flight modal parameter extraction of wings was already demonstrated in a number of studies [17, 18, 19]. These also confirmed the detectability of velocity and AOA-induced variability. Robust automation techniques, like the procedure described in this work, are another important building block for future applications of AOMA in passive flutter testing or SHM.

5. Conclusions

In this work a multi-stage clustering approach for automated operational modal analysis is presented, which improves existing approaches in multiple aspects. The algorithm is fully automatic. No parameters or thresholds have to be provided by the user. Neither the damping ratios nor the complexities of the to-be-identified modes are limited in any way. In contrast to existing methods, the procedure is shown to be insensitive to the chosen system order ranges. The methodology was applied to a large number of challenging wind

495 tunnel measurements with, in part, poor SNR conditions, highly damped modes and/or
identification from only two sensors. Nevertheless, the method was able to consistently
identify all physical modes in the investigated frequency range.

References

- [1] R. Brincker, P. H. Kirkegaard, Special Issue on Operational Modal Analysis, *Mechanical Systems and Signal Processing* 24 (5) (2010) 1209–1212.
- 500 [2] F. Magalhães, Á. Cunha, E. Caetano, Online Automatic Identification of the Modal Parameters of a Long Span Arch Bridge, *Mechanical Systems and Signal Processing* 23 (2) (2009) 316–329.
- [3] E. Reynders, J. Houbrechts, G. de Roeck, Fully Automated (Operational) Modal Analysis, *Mechanical Systems and Signal Processing* 29 (2012) 228–250.
- 505 [4] C. Devriendt, F. Magalhaes, W. Weijtjens, G. de Sitter, A. Cunha, P. Guillaume, Structural Health Monitoring of Offshore Wind Turbines using Automated Operational Modal Analysis, *Structural Health Monitoring* 13 (6) (2014) 644–659.
- [5] A. G. Piersol, T. L. Paez, C. M. Harris, *Harris’ Shock and Vibration Handbook*, McGraw-Hill, New York, USA, 6th edn., 2010.
- 510 [6] C. Rainieri, G. Fabbrocino, Development and Validation of an Automated Operational Modal Analysis Algorithm for Vibration-based Monitoring and Tensile Load Estimation, *Mechanical Systems and Signal Processing* 60-61 (2015) 512–534.
- [7] E. Neu, F. Janser, A. A. Khatibi, A. C. Orifici, Operational Modal Analysis of a Cantilever in a Wind Tunnel using Optical Fiber Bragg Grating Sensors, in: *Proceedings of the 6th International Operational Modal Analysis Conference*, Gijón, Spain, 2015.
- 515 [8] P. v. Overschee, B. L. R. d. Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*, Kluwer Academic, Boston, USA, 1996.
- [9] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, *Multivariate Data Analysis*, Pearson Education Limited, Harlow, United Kingdom, 7th edn., 2014.
- 520 [10] G. E. P. Box, D. R. Cox, An Analysis of Transformations, *Journal of the Royal Statistical Society. Series B (Methodological)* 26 (2) (1964) p 211–252.
- [11] J. M. Cimbala, *Outliers: Lecture on Instrumentation, Measurements, and Statistics*, URL <http://www.mne.psu.edu/cimbala/me345/>, accessed: 12 Juli, 2016, 2011.
- 525 [12] J. Schwochow, G. Jelicic, Automatic Operational Modal Analysis for Aeroelastic Applications, in: *Proceedings of the 6th International Operational Modal Analysis Conference*, Gijón, Spain, 2015.
- [13] A. Brandt, ABRAVIBE–A Toolbox for Teaching and Learning Vibration Analysis, *Sound and Vibration* 1 (11) (2013) 12–17.
- [14] *Structural Vibration Solutions A/S, ARTeMIS Modal Documentation*, version 3.0, 2013.
- 530 [15] A. Brandt, R. Brincker, Impact Excitation Processing for Improved Frequency Response Quality, in: *Proceedings of the 28th International Modal Analysis Conference*, Jacksonville, Florida, USA, 2010.

- [16] L. Y. Yam, T. P. Leung, D. B. Li, K. Z. Xue, Theoretical and Experimental Study of Modal Strain Analysis, *Journal of Sound and Vibration* 191 (2) (1996) 251–260.
- [17] J. Debille, B. Peeters, The Benefits of Operational Modal Analysis of Aircraft and Spacecraft Structures, in: *Proceedings of European Test & Telemetry Conference*, Toulouse, France, 2005.
- [18] B. Peeters, T. d. Troyer, P. Guillaume, In-flight Modal Analysis - a Comparison between Sweep and Turbulence Excitation, in: *Proceedings of the 22th International Conference on Noise and Vibration Engineering*, Leuven, Belgium, 2006.
- [19] E. Neu, F. Janser, A. A. Khatibi, C. Braun, A. C. Orifici, Operational Modal Analysis of a Wing Excited by Transonic Flow, *Aerospace Science and Technology* 49 (2016) 73–79.