Olav Guddal

Comparison of methods for automatic modal analysis

Master's thesis in Civil and Environmental Engineering
Supervisor: Gunnstein Thomas Frøseth
Co-supervisor: Anno Christian Dederichs
June 2022

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Structural Engineering

NTNU
Kunnskap for en bedre verden

Olav Guddal

Comparison of methods for automatic modal analysis

Master's thesis in Civil and Environmental Engineering
Supervisor: Gunnstein Thomas Frøseth
Co-supervisor: Anno Christian Dederichs
June 2022

Norwegian University of Science and Technology
Faculty of Engineering
Department of Structural Engineering

Institutt for konstruksjonsteknikk
Fakultet for ingeniørvitskap
NTNU – Noregs teknisk-naturvitskaplege universitet

TILGJENGE:

Open

MASTEROPPGÅVE 2022

FAGOMRÅDE:	DATO:	TAL SIDER:
Konstruksjonsdynamikk	11. juni 2022	85 (101 totalt)

TITTEL:

Samanlikning av metodar for automatisk modal analyse

Comparison of methods for automatic modal analysis

UTFØRT AV:

Olav Guddal

SAMANDRAG:
Denne masteroppgåva samanliknar to algoritmar føreslege av Reynders (2012) og Yang (2019) for automatisk modal analyse. Algoritmane er implementert ved bruk av objektorientert programmering i Python. Samanlikninga er gjennomført ved analyse av numerisk data generert frå ei analytisk skjerramme, og data henta frå Elgeseter bru i samband med oppgåva.

Dei to algoritmane er overordna like i framgangsmåte, og gjev for dei fleste bruksområde tilfredsstillande resultat. Reynders sin algoritme er meir optimistisk og finn på jamnen fleire systemmodar. Det er tidvis stor spreiing i klyngene med modar (clusters), som alle skal estimera den same systemmoden.
Yang sin algoritme er meir selektiv og finn på jamnen færre systemmodar. Klyngene med modar er mindre og har høgare kvalitet enn klyngene frå Reynders sin algoritme.
Mykje av forskjellen kjem av det siste steget i Yang sin algoritme. Der vert mellom 25% og 50% av klyngje-modane fjerna ved teknikken «Local Outlier Factor Detection». Hovudforskjellen kjem likevel av måten klyngjene vert til og korleis dei dårlege klyngjene vert fjerna.

Val av algoritme kjem an på problemet. Kort sagt bør Reynders sin algoritme veljast viss målet er å detektera flest mogleg modar, og Yang sin algoritme veljast viss analyse utan feil og deteksjon av modar med høg presisjon er målet. Viser elles til kapittel 5 for ein grundigare oversikt over forskjellane.

Uavhengig av metode er resultata sterkt avhengig av inputen. Til dømes var kvaliteten på måledataen frå Elgeseter bru noko dårleg, som igjen førte til middelmåtig kvalitet på analyseresultatet.

FAGLÆRAR: Gunnstein Thomas Frøseth

RETTLEIAR: Anno Christian Dederichs

UTFØRT VED: Institutt for konstruksjonsteknikk

Department of Structural Engineering
Faculty of Engineering
NTNU – Norwegian University of Science and Technology

Accessibility:

Open

MASTER'S THESIS, 2022

SUBJECT AREA:	DATE:	NO. OF PAGES:
Structural Dynamics	June 11th, 2022	85 (101 in total)

TITLE:

Comparison of methods for automatic modal analysis

Samanlikning av metodar for automatisk modal analyse

BY:

Olav Guddal

SUMMARY:
This master's thesis compares two algorithms created by Reynders (2012) and Yang (2019) for automatic modal analysis. The algorithms are implemented using object-oriented programming in Python.
The comparison was carried out by analysis of numerical data generated from an analytical shear frame, and self sampled data collected from Elgeseter bridge in connection with the thesis.

The two algorithms have similar overall procedures and give satisfactory results for most applications.
Reynders' algorithm is more optimistic and finds more system modes. There is also a greater spread in the clusters with modes, all of which must estimate the same system mode.
Yang's algorithm is more selective and finds fewer system modes. The clusters with modes are smaller in length and yield higher quality modes than the clusters created with Reynders' algorithm.
Much of the difference comes from the last step in Yang's algorithm. Between 25% and 50% of cluster modes are removed by the "Local Outlier Factor Detection" technique. The main difference comes from the clustering technique, and how they are filtered.

The choice of algorithm depends on the object of the analysis. Reynders should be chosen if detecting as many modes as possible is the main priority. Yang is the preferred choice if errors are not tolerated, and high precision is the aim. See Chapter 5 for a more thorough overview of their differences.

Measurement quality has a big affect on the results, regardless of algorithm chosen. This was the case with the lower quality measurements from Elgeseter, which led to mediocre quality of the analysis.

PROFESSOR:	Gunnstein Thomas Frøseth
SUPERVISOR:	Anno Christian Dederichs
CARRIED OUT AT:	The Department of Structural Engineering

Preface

This thesis makes up the final work of my five year long Master's Degree Programme at the Norwegian University of Science and Technology (NTNU). The thesis is written with guidance from the Structural Dynamics Group at the Department of Structural Engineering. The workload is equivalent to 30 credits, i.e one semester.

This thesis was not one of the suggested theses by the institute this year. It was, however, proposed by Gunnstein Thomas Frøseth while discussing the content of other theses and my personal interest in automating cumbersome processes. Automatic algorithms to perform modal analysis was the natural next step after writing my project thesis on manually identifying the dynamic properties of the Hardanger Bridge.

I would like to give a special thanks to my supervisors, Associate Professor Gunnstein Thomas Frøseth and PhD Candidate Anno Christian Dederichs. Gunnstein has always been ready to jump deep into debugging code or discussing the practicality of the results. Anno always had an answer when the theory was lacking necessary information and helped me get on the best possible path by giving me an overview of my work. Their interest and open door mentality have been crucial for the progression and final outcome of this thesis.

My fellow students at the "clean desk" study room have made this semester a pleasure, both on and outside of the university. So a special thanks to; Anders Kastet, Andreas Grøndahl Nourouzi, Erik Hattestad, Stian Gundersen Raniszewski, and Trygve Vangsnes.

Trondheim,
11th of June, 2022

Olav Guddal

Contents

List of Acronyms

DOF	Degree of Freedom.
EMA	Experimental Modal Analysis.
LOF	Local Outlier Factor.
MAC	Modal Assurance Criteria.
MDOF	Multi Degree of Freedom.
MOC	Modal Observability Correlation.
MP	Mean Phase.
MPC	Modal Phase Colinearity.
MPD	Modal Phase Deviation.
MPE	Modal Parameter Estimation.
MSF	Modal Scale Factor.
MTN	Modal Transfer Norm.
OMA	Operational Modal Analysis.
OMAX	Operational Modal Analysis with exogenous inputs.
PoGER	Post Global Estimation Re-scaling.
PoSER	Post Separate Estimation Re-scaling.
PreGER	Pre Global Estimation Re-scaling.
PSD	Power Spectral Density.
SDOF	Single Degree of Freedom.
SHM	Structural Health Monitoring.
SSI-COV	Covariance Driven Stochastic Subspace Identification.
SVD	Singular Value Decomposition.
ZOH	Zero Order Hold.

Chapter 1

Introduction to automatic modal analysis

1.1 Motivation

Society has a never ending demand for new and better structures that are safe, environmentally friendly, and visually appealing. These structures include high rising slender skyscrapers of more environmental materials. Mjøstårnet, the currently highest timber skyscraper towering 85.4 m above ground, is a good example of the new structures of our time, (Moelven, 2022). Another good example is the planned new highway, $E39$, along the western coast of Norway needing record breaking structures to cross deep and exposed waters and harsh fjords. The current plans include new constructions like submerged pipe-bridges never been built before. It has, however, been suggested multiple times since it was patented by Sir Edward James Reed in 1886, (Rønnquist, 2022).

With new structures, instrumentation and validation of the analysis made beforehand is crucial. The Norwegian Public Roads Administration has a requirement in their handbook for bridge design, N400, that all pipe-bridges must be instrumented for systematic recording of the structure's movements and load response, as well as surveillance of the protection systems and re-bar corrosion, (SVV, 2015).

To meet the demand to surveillance structures and validate analyses, structural engineers are continuously expanding their toolboxes. One of the latest additions are digital twins, aiming at representing all information available together with the physical properties of the structure. Digital twins can consequently represent the current health state of the structure, and also contribute to expose weaknesses in the chosen design, just as the road authorities demand. Getting information from the physical structure automatically is then key to make applications economically feasible and applicable in large scale.

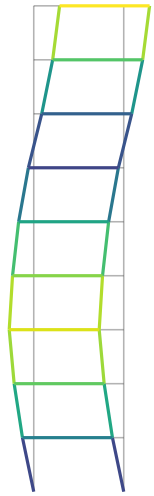


Figure 1.1: Analytical shear frame in undeformed configuration, and with a mode shape's deflection.

1.2 Structural dynamics

Nearly all structures are exposed to some sort of time varying loads. The time varying loads may have cyclic components near the eigenfrequencies of the structure that will induce resonance. Resonance yields responses multiple times greater than the static response of the loading would imply. Throughout history resonance have occurred, like with Tacoma Narrows where fluctuating winds at a frequency close to the eigenfrequency of one torsional mode resulted in the bridge collapsing, (WSDOT, 2022). Information about the eigenfrequencies and also the loading frequencies is therefore essential to ensure safety.

A common assumption within structural dynamics is that every structure has *modes* as inherent properties. Each mode is mathematically characterized by its eigenvalue and mode shape. From the eigenvalue one can derive the eigenfrequency, f, and damping ratio, ξ, as derived in Section 2.5.3. Eigenfrequency, damping ratio and mode shape are modal properties that can physically be observed during certain excitation or in controlled tests. An example of a mode's mode-shape is illustrated in Figure 1.1. The modal parameters are directly influenced by every choice made during the design phase of the structure. Examples are geometry influencing the wind flow around the structure, material choices with respect to mass and stiffness, and wanted boundary conditions chosen by the responsible engineer. Total dynamic response is the superposition of the individual responses of all modes. The excitation of each mode is highly dependent on the current loading situation.

The modes of a structure are usually not visible by the naked eye. They are, however, measurable by accelerometers, which over the last decades have been developed to measure very small vibrations. Estimation of modes from only a single data series will be referred to as Modal Parameter Estimation (MPE). Tracking of modes over multiple recordings is referred to as *modal tracking*. Modal tracking can be employed to determine the current health state of the structure, as well as early stage damage detection. Information about the location of the damage is also possible to extract. This allows for the possibility to inspect the structure at any time without requiring physical appearance, as well as notification of when and where to perform physical inspections. The mentioned techniques can be gathered into the collective term Structural Health Monitoring (SHM). Another application of modal tracking is to get enhanced knowledge about external effects, e.g. temperature, to further optimize the structural design.

As of today, the substantial amount of user interaction is limiting the the widespread of modal tracking. Many attempts on automation have been made, but none are completely autonomous. The aim of this masters thesis is to implement and test some of the current techniques available. The two investigated algorithms are further described in Reynders et al. (2012) and Yang et al. (2019).

Chapter 2

Theory

This theory chapter takes the reader through the relevant theory needed to understand the underlying assumptions and limitations of modal analysis at a level needed to perform analyses. The identification of modes and algorithms for automatic clearing of the stabilization diagram is explained in detail together with all validation criteria used in this thesis. The left hand pages are used to summarize the main findings of each right hand page where figures are not present.

2.1 Modal analysis techniques

Identification of modal parameters is crucial in the design of many structures, and the techniques have therefore evolved together with the structural design. The first identification procedures where experimental and can be dated back to the middle of the 19th century. Experimental methods are in its most basic form the most intuitive; apply loading, see what happens, and then find the relation between the applied loading and the observed response. In Experimental Modal Analysis (EMA) both applied forces and response are recorded, and their relation is solved in either the time or frequency domain. Force and response is related through the impulse-response function in the time domain, and the frequency-response function in the frequency domain. Modes are then corresponding to peaks of the frequency response function.

Operational Modal Analysis (OMA) is sometimes called the stochastic counterpart to EMA, due to the nature of the loading. The fundamental assumption in OMA is that the load spectra is gaussian white noice which implies that all modes are excited. With all modes excited, they are all possible to detect from the recorded measurement. OMA is especially advantageous on huge structures like bridges as forced excitation of such structures may be hard or even practically impossible. Combined methods where both measured and unmeasured forces are accounted for is often called Operational Modal Analysis with exogenous inputs (OMAX).

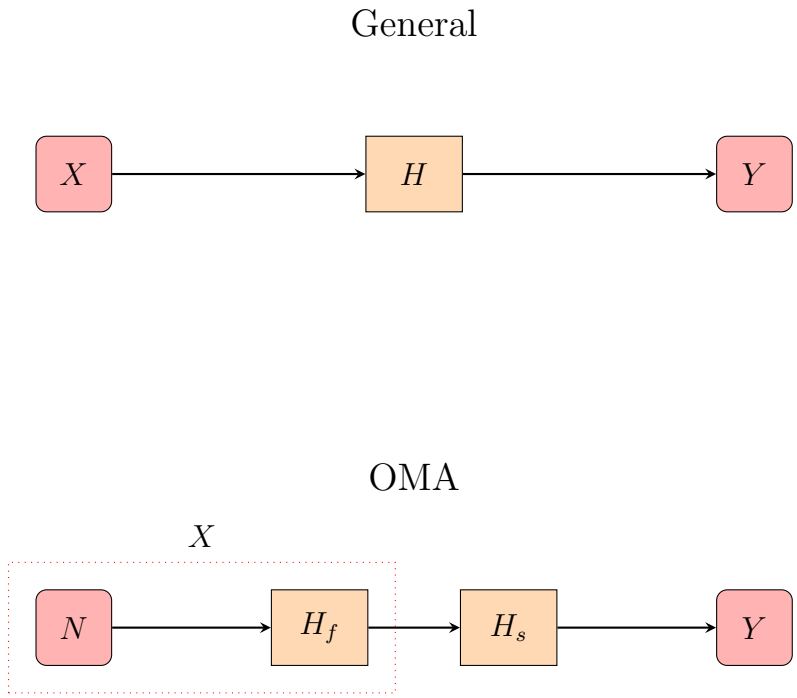


Figure 2.1: General system and OMA system.

2.2 OMA

Operational Modal Analysis (OMA) is the collective term for modal analysis methods only based on the output, Y, of the structure under the currently present loads. OMA is based on three important assumptions:

- *Linearity* - principle of superposition is valid
- *Stationarity* - dynamic properties are independent of time
- *Observability* - sensor layout is able to observe modes of interest

The input, X, that is loads acting on the structure, consists of both environmental and operational loads, which are all immeasurable. This leads to some assumptions about the loading. The first is that applying a load with the white noise characteristics, $N(\omega)$, yields a flat power spectra that excites all modes equally. The response spectrum, $Y(\omega)$, will then contain full information about the structure's dynamic properties.

No real life loads have a flat power spectra, making the modes unequally excited. The observed modes is then weighted by the spectral distribution of the loading and also the noise present. To generalize this, it is natural to assume that the structure is excited by the excitation system loaded by white noise. The excitation system is then the output of all the unknown loads on the structure combined. Graphically this is represented in Figure 2.1. The upper chart is a general system with input, X (loading), system responding, $H(\omega)$ (frequency response), and output, $Y(\omega)$ (acceleration). The lower chart is depicting the OMA-assumption with the loading X consisting of the output system H_f loaded by white noise N. H_s is then the system's frequency response characteristics.

As a result of the OMA-assumption modes are unequally excited depending on the current excitation system, which consist of windloading, trafic, etc. The response measurement Y contains information about the excitation. All modal parameters are however preserved and identifiable, without loss of accuracy. This possible due to the fact that the structure has a narrow banded response characteristic and time invariant properties, and the excitation system is characterized by a broad banded response and possibly time variant properties. The narrow banded response characteristics is depicted by the frequency response function that for most structures have high and narrow peaks. The statement of time invariant properties is true under normal conditions, that is no damage to the structure, and over the normal time span of a measurement which usually is less than hours.

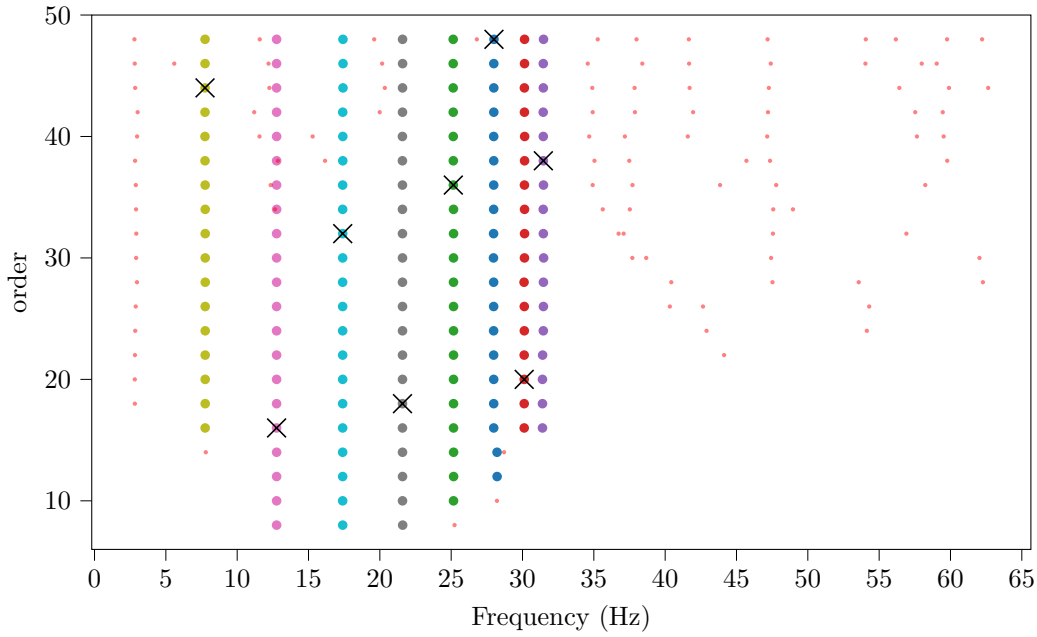


Figure 2.2: Example of an automatically cleared stabilization diagram.

2.3 Parametric system identification

Parametric identification methods have proven to be much more accurate than *non-parametric* methods, and is currently the preferred way of approach. With parametric methods one are fitting the recorded data to a model, rather than looking for the model that fits the data. The model has a order n that must be given. In theory there is two eigenvalues for each physical system mode, that is two complex conjugated eigenvalues. Much effort has been made to find the optimal n in an automated way to maximize the prediction capacity of the parametric identification methods.

Empirical tests, on the other hand, shows that the identified modal parameters appears to be nearly the same for different model orders, also when the model order is overestimated. This justifies an approach with repeated overestimation of the model order. For each model order one obtain n modes. All identified modes can then be represented in a *stabilization diagram* which is a plot of eigenfrequency vs. model order, see Figure 2.2 showing an example of an automatically cleared stabilization diagram. Modes aligning as vertical lines will typically represent the same mode, as most modes of a structure is at different frequencies. This is, however not always the case as later explained. With multiple approximations (modes) of the same physical system mode it is natural to group these, i.e. clustering. Clusters of modes estimating the same gives the possibility to perform statistical measures for assessment of the quality. The different dot colors, ● ● ● ++ , in Figure 2.2 represent the different clusters, and their picked representative is marked by a ×.

Modes that do not represent any physical mode of the system are said to be spurious. Spurious modes can be divided into two subgroups:

- noise modes: physical harmonics that are present, but do not represent a mode. For example: wind induced vibrations of suspension cables
- mathematical modes: are the result of the mathematical description. Measurement noise, computational noise and modelling inaccuracies may all create mathematical modes.

The spurious modes are marked with smaller red or grey dots, • •, as shown in Figure 2.2.

The subsequent sections will derive the models that the measurement data is fitted to, or more precisely the derivation of the COV-SSI approach leading up to the mode objects.

State equation

$$\{\dot{s}(t)\} = [A_c]\{s(t)\} + [B_c]\{u(t)\}$$

2.4 State-Space Models

2.4.1 State-Space Model in continous time

A general Multi Degree of Freedom (MDOF) system can be represented by the following second order differential equation, on matrix form:

$$[M]\{\ddot{y}(t)\} + [C]\{\dot{y}(t)\} + [K]\{y(t)\} = \{f(t)\} \tag{2.1}$$

$\{\ddot{y}\}$ $\{\dot{y}\}$ and $\{y\}$ are vectors of acceleration, velocity, and displacement respectively. $[M]$ $[C]$ and $[K]$ are mass-, damping- and stiffness-matrices of the system.

The force vector $\{f(t)\}$ can be factorized into: $\{f(t)\} = [\bar{B}]\{u(t)\}$, where $[\bar{B}]$ is a matrix defining the locations of each input and $\{u(t)\}$ describes the timevariation of the loading.

This makes the following simplification natural:

$$\{\ddot{y}(t)\} + [M]^{-1}[C]\{\dot{y}(t)\} + [M]^{-1}[K]\{y(t)\} = [M]^{-1}[\bar{B}]\{u(t)\} \tag{2.2}$$

The second order differential equation (2.2) can be rewritten as a set of first order differential equations by the introduction of the *state vector*:

$$\{s(t)\} = \left\{ \begin{array}{c} \{\dot{y}(t)\} \\ \{y(t)\} \end{array} \right\} \tag{2.3}$$

and a "dummy"-equation:

$$[M]\{\dot{y}\} = [M]\{\dot{y}\} \tag{2.4}$$

This gives the following set of first order differential equations:

$$\{\dot{s}(t)\} = \underbrace{\begin{bmatrix} -[M]^{-1}[C] & -[M]^{-1}[K] \\ [I] & [0] \end{bmatrix}}_{A_c}\{s(t)\} + \underbrace{\begin{bmatrix} -[M]^{-1}[\bar{B}] \\ [0] \end{bmatrix}}_{B_c}\{u(t)\} \tag{2.5}$$

Equation (2.5) defines the change in state $\{\dot{s}\}$ as a matrix function of the current state and current loading. The underlying assumption that mass $[M]$, damping $[C]$ and stiffness $[K]$ is constant opens for further simplifications. The *statematrix* $[A_c]$ and the *input influence matrix* $[B_c]$ makes the *state equation* compact:

$$\{\dot{s}(t)\} = [A_c]\{s(t)\} + [B_c]\{u(t)\} \tag{2.6}$$

Note the subscript, $\square_c$, denoting continuous time matrices.

Observation equation

$$\{y_l(t)\} = [C_c]\{s(t)\} + [D_c]\{u(t)\}$$

Assuming that there are l sensor locations one can define the *observation equation*:

$$\{y_l(t)\} = [C_a]\{\ddot{y}\} + [C_v]\{\dot{y}\} + [C_d]\{y\} \tag{2.7}$$

where subscripts a, v, d is acceleration, velocity and displacement.

Inserting Equation (2.2):

$$\begin{aligned}\{y_l(t)\} = &\left([C_v] - [C_a][M]^{-1}[C]\right)\{\dot{y}(t)\} \\ &+ \left([C_d] - [C_a][M]^{-1}[K]\right)\{y(t)\} \\ &+ \left([C_a][M]^{-1}[\bar{B}]\right)\{u(t)\}\end{aligned} \tag{2.8}$$

This may also be simplified by introducing new matrices. *The output influence matrix* $[C_c]$ and *direct transmission matrix* $[D_c]$ is defined as follows:

$$[C_c] = \begin{bmatrix}[C_v] - [C_a][M]^{-1}[C] & [C_d] - [C_a][M]^{-1}[K]\end{bmatrix} \tag{2.9}$$

$$[D_c] = \begin{bmatrix}[C_a][M]^{-1}[\bar{B}]\end{bmatrix} \tag{2.10}$$

The *observation equation* simplifies to:

$$\{y_l(t)\} = [C_c]\{s(t)\} + [D_c]\{u(t)\} \tag{2.11}$$

The state equation (2.6) and observation equation (2.11) constitutes the *coninuous-time state space model*.

2.4.2 State-space model in disctrete time

Real life measurements are not continuous and a connection between discrete time and continous time is needed. To convert the continuous time model one needs assumptions about the input. The simplest is Zero Order Hold (ZOH) which state that the input is piece wise constant within every time sample. This leads to the following relations between continuous-time matrices and corresponding discrete-time matrices:

$$[A] = e^{[A_c]\Delta t} \tag{2.12}$$

$$[B] = ([A] - [I])[A_c]^{-1}[B_c] \tag{2.13}$$

$$[C] = [C_c] \tag{2.14}$$

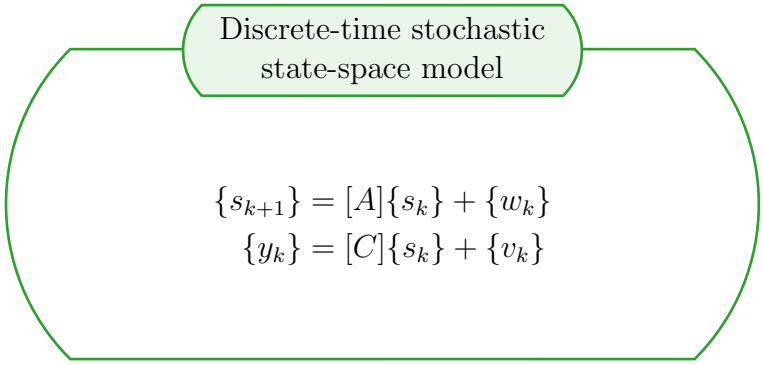
$$[D] = [D_c] \tag{2.15}$$

The *discrete time state-space model* is then:

$$\{s_{k+1}\} = [A]\{s_k\} + [B]\{u_k\} \tag{2.16}$$

$$\{y_k\} = [C]\{s_k\} + [D]\{u_k\} \tag{2.17}$$

Note that the lack of subscript, $\square_c$, is referring to discrete time matrices


Discrete-time stochastic
state-space model
$\{s_{k+1}\} = [A]\{s_k\} + \{w_k\}$
$\{y_k\} = [C]\{s_k\} + \{v_k\}$

2.4.3 Discrete-time stochastic state-space model

The *discrete time state-space model* is deterministic because the driving forces, i.e. the input, is deterministic. To accommodate stochastic loading one can introduce the two stochastic components; *process noise* $\{w_k\}$ and *measurment noise* $\{v_k\}$ at the discrete time instants $t_k = k\Delta t$:

$$\{s_{k+1}\} = [A]\{s_k\} + [B]\{u_k\} + \{w_k\} \tag{2.18}$$

$$\{y_k\} = [C]\{s_k\} + [D]\{u_k\} + \{v_k\} \tag{2.19}$$

The process noise will then contain the external disturbances and model inaccuracies and the measurement noise contain the inaccuracy of the sensors.

Performing OMA the excitation of the system is immeasurable. This makes the loading input $\{u_k\}$ unknown, and it may be removed. When the process noise, $\{w_k\}$ and the measurement noise, $\{v_k\}$, is the only driving forces, the *discrete-time stochastic state-space model* is obtained:

$$\{s_{k+1}\} = [A]\{s_k\} + \{w_k\} \tag{2.20}$$

$$\{y_k\} = [C]\{s_k\} + \{v_k\} \tag{2.21}$$

The discrete-time stochastic state-space model consists of two equations. The first one relates the current state, $\{s_k\}$ with the next state, $\{s_{k+1}\}$ through the state matrix, $[A]$, plus the process noise, $\{w_k\}$. The second equation states that the observed system response, $\{y_k\}$, is the product of the output influence matrix, $[C]$, and the current state, $\{s_k\}$, plus the measurement noise, $\{v_k\}$.

Having related the previous state with the next state, and the observed output with the current state, the next step is determining the state matrix and the observation matrix. Discrete measurements will in this context work as estimates of the relations, but the amount of noise in each step is unknown. Therefore the relation cannot explicitly be solved, but rather estimated through covariance calculations.

Fundamental relations

$$
\begin{aligned}
[\Sigma] &= [A][\Sigma][A]^T + [Q] \\
[R_0] &= [C][\Sigma][C]^T + [R] \\
[G] &= [A][\Sigma][C]^T + [S] \\
[R_i] &= [C][A]^{i-1}[G]
\end{aligned}
$$

2.4.4 Fundamental relations

The covariance between the two driving loads at discrete time instant q and p can be then be expressed as:

$$E\left[\begin{Bmatrix} w_p \\ v_p \end{Bmatrix}\begin{Bmatrix} \{w_q\}^T & \{v_y\}^T \end{Bmatrix}\right] = \begin{bmatrix} [Q] & [S] \\ [S]^T & [R] \end{bmatrix}\delta_{pq} \tag{2.22}$$

Where δ_{pq} is the Kronecker delta function which is one for $q = p$ and zero for $q \neq p$, i.e. no correlation between the loading at different discrete time instants.

The observed output y_k inherits the properties of the loading $\{w_k\}$ and $\{v_k\}$, that is, a zero mean Gaussian process:

$$[R_i] = E[\{y_{k+1}\}\{y_k\}^T] \tag{2.23}$$

$[R_i]$ is containing all the available information from the recorded measurements.

The state s_k is also a zero mean Gaussian process:

$$[\Sigma] = E[\{s_k\}\{s_k\}^T] \tag{2.24}$$

The state $\{s_k\}$ is uncorrelated with both the process noise:

$$[\{s_k\}\{w_k\}^T] = 0 \tag{2.25}$$

and the measurement noise:

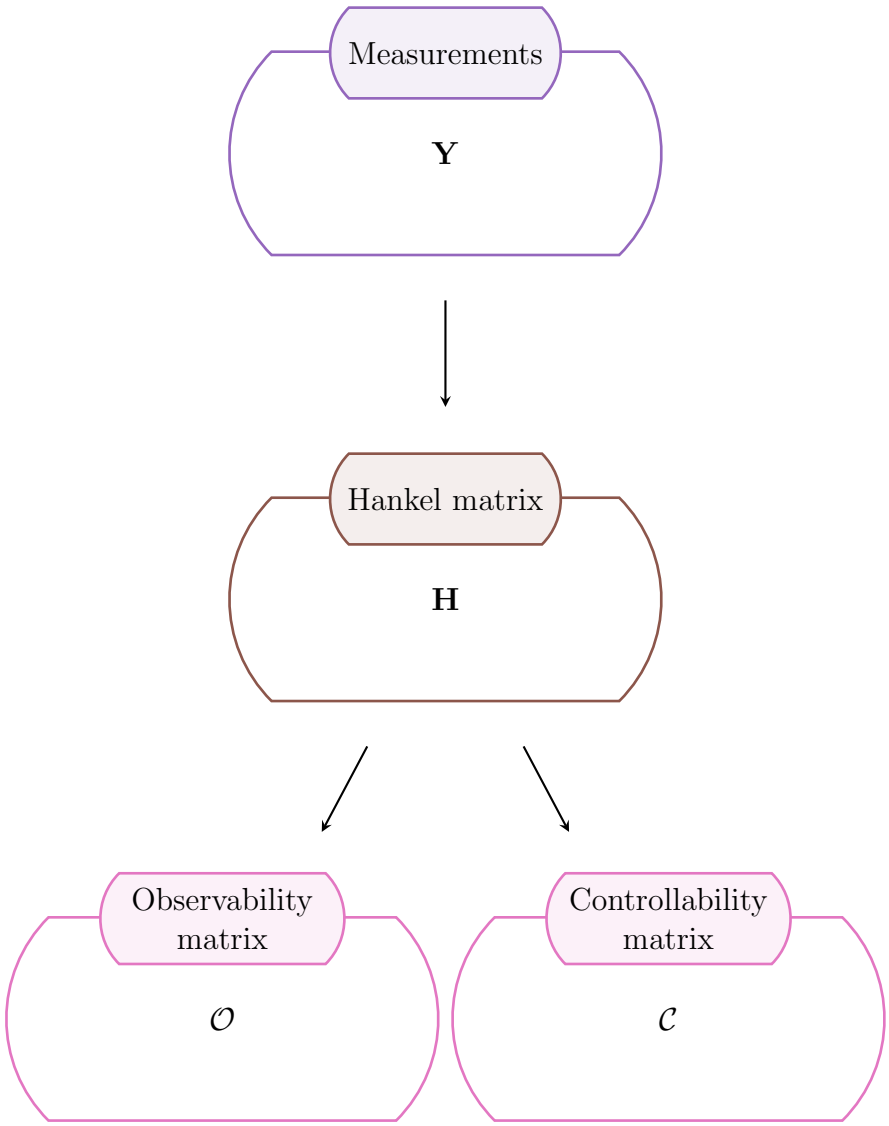
$$E[\{s_k\}\{v_k\}^T] = 0 \tag{2.26}$$

These properties leads to the following relations:

$$[\Sigma] = [A][\Sigma][A]^T + [Q] \tag{2.27}$$
$$[R_0] = [C][\Sigma][C]^T + [R] \tag{2.28}$$
$$[G] = [A][\Sigma][C]^T + [S] \tag{2.29}$$
$$[R_i] = [C][A]^{i-1}[G] \tag{2.30}$$
$$\text{where: } [G] = E[\{s_{k+1}\}\{y_k\}^T] \tag{2.31}$$

[G], the *next state-output covariance matrix*, is then measuring the covariance between the currently observed output, $\{y_k\}$, and the the next state $\{s_{k+1}\}$.

Having defined the stochastic subspace formulation it is possible to derive modes with multiple methods. This thesis is using the Covariance Driven Stochastic Subspace Identification (SSI-COV) approach that will be discussed in the next section.


Measurements
Y
Hankel matrix
H
Observability matrix
$\mathcal{O}$
Controllability matrix
$\mathcal{C}$

2.5 Identifying modes

2.5.1 Covariance-Driven Stochastic Subspace Identification

Assuming that measured data is organized in a matrix as follows:

$$\underset{l\times N}{[Y]} = \begin{Bmatrix} \{y_0\}^T \\ \{y_1\}^T \\ \vdots \\ \{y_l\}^T \end{Bmatrix} \tag{2.32}$$

where $\{y_l\}$ is the response vector of sensor l with N discrete measurements.

The estimated output correlation at time lag i is then:

$$\underset{l\times l}{[\hat{R}_i]} = \frac{1}{N-1}\underset{l\times N}{[Y_{1:N-i}]}\underset{N\times l}{[Y_{i:N}]}^T \tag{2.33}$$

The estimate at different time lags i can be gathered into a block-Hankel matrix, with i block rows:

$$\underset{il\times il}{[H_i]} = \begin{bmatrix} [R_1] & [R_2] & \dots & [R_i \\ [R_2] & [R_3] & \dots & [R_{i+1}] \\ \vdots & \vdots & \ddots & \vdots \\ [R_i] & [R_{i+1}] & \dots & [R_{2i+1}] \end{bmatrix} \tag{2.34}$$

Utilizing the property in Equation (2.30) one can create a general decomposition of the block-Hanckel into the observability $\mathcal{O}_i$ and controlablity matrix $\mathcal{C}_i$:

$$\underset{il\times il}{[H_i]} = \underset{il\times l}{[\mathcal{O}_i]}\underset{l\times il}{[\mathcal{C}_i]} \tag{2.35}$$

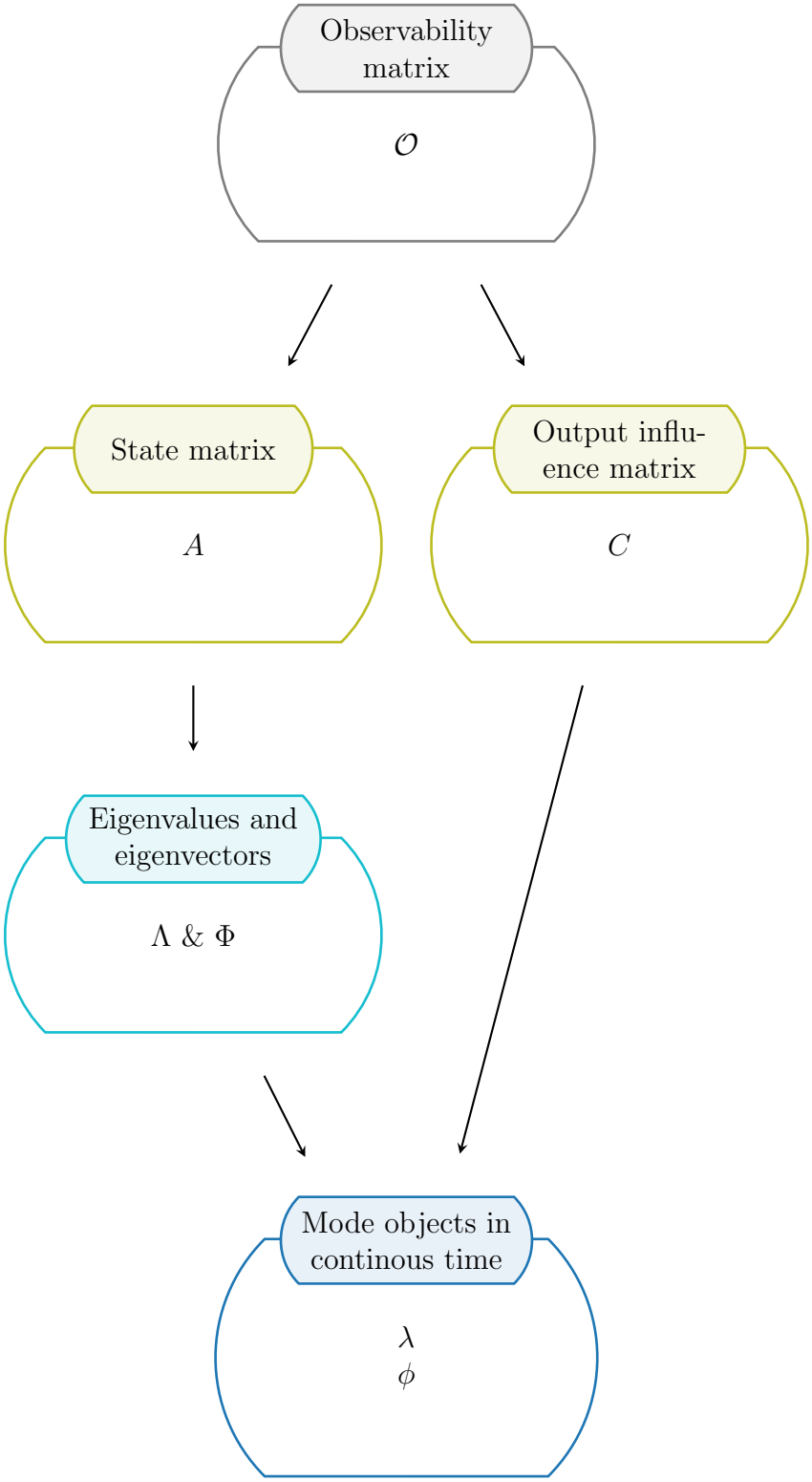
where:

$$\underset{il\times l}{[\mathcal{O}_i]} = \begin{bmatrix} [C] \\ [C][A] \\ [C][A]^2 \\ \vdots \\ [C][A]^{i-1} \end{bmatrix} \tag{2.36}$$

$$\underset{l\times il}{[\mathcal{C}_i]} = \begin{bmatrix} [G] & [A][G] & [A]^2[G] & \dots & [A]^{i-1}[G] \end{bmatrix} \tag{2.37}$$

With i output correlation matrices, R_i, whose size is $l \times l$, the maximum number of block rows, n, is limited to:

$$i \times l >= n \tag{2.38}$$


Observability matrix
$\mathcal{O}$
State matrix
A
Output influence matrix
C
Eigenvalues and eigenvectors
Λ & Φ
Mode objects in continous time
λ
ϕ

Through Singular Value Decomposition (SVD) of the block hankel matrix with truncation at order, o, one obtains:

$$\text{SVD}(H_i) = \begin{bmatrix}[U_o] & [U_d]\end{bmatrix} \begin{bmatrix}[\Sigma_o] & [0] \\ [0] & [\Sigma_d]\end{bmatrix} \begin{bmatrix}[V_o]^T \\ [V_d]^T\end{bmatrix} \tag{2.39}$$

$$\approx [U_o][\Sigma_o][V_o]^T \tag{2.40}$$

Combining Equations (2.35) and (2.40) one can get an expression for the observability matrix, $\mathcal{O}_i$:

$$\mathcal{O}_i = [U_o][\Sigma_o]^{\frac{1}{2}} \tag{2.41}$$

From Equation (2.36) one can observe that the output influence matrix, $[C]$, can be approximated from the first block row. And by removing the first block row of $[\mathcal{O}_i]$, taking the pseudo inverse, and multiplying with all but the last block row of $[\mathcal{O}_i]$ one can approximate $[A]$ as:

$$\underset{l\times l}{[A]} = [\mathcal{O}_i^{\downarrow}]^{+}[\mathcal{O}_i^{\uparrow}] \tag{2.42}$$

2.5.2 Decomposing A and C matrices

After having found the A and C matrix at one model order, n, one can find the eigenvalues through eigenvalue decomposition of A:

$$A = \Psi\Lambda\Psi^{-1} \tag{2.43}$$

where Ψ is a matrix with eigenvectors and Λ a diagonal matrix with eigenvalues.

By combining the relation between continuous time and discrete time matrices given in Equation (2.12) and Equation (2.43) the relation between continuous time and discrete time eigenvectors are:

$$A = e^{[A_c]\Delta t} = e^{\Psi_c\Lambda_c\Psi_c^{-1}\Delta t} = \Psi_c e^{\Lambda_c\Delta t}\Psi_c^{-1} \tag{2.44}$$

$$\vdots$$

$$\lambda = e^{\lambda_c\Delta t} \tag{2.45}$$

where Ψ_c and Λ_c is the continuous time counterpart of Ψ and Λ.

The modeshapes Φ are obtained as:

$$\Phi = C\Psi \tag{2.46}$$

Having determined Λ and Φ at a given model order n one can create the n mode-objects stored inside the matrices Λ and Φ. Each mode can be represented as an object with eigenvalue λ and mode shape ϕ as main attributes. The eigenvalue λ can be used to derive eigenfrequency ω, damping ξ and damped eigenfrequency ω_d as shown in the subsequent Section 2.5.3.

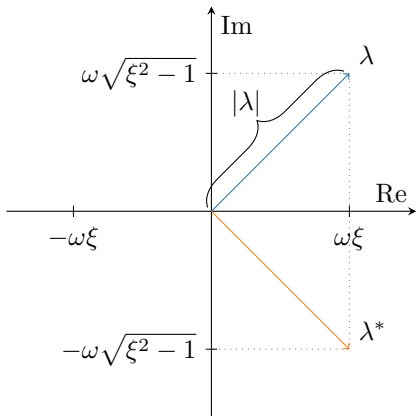


Figure 2.3: Eigenvalue pair related in the Argand diagram.

2.5.3 Deriving dynamic properties

A general underdamped system may be represented by the general solution:

$$y = Ae^{\lambda t} \tag{2.47}$$

where λ is a complex number. Inserting this into the general equation for a Single Degree of Freedom (SDOF) yields:

$$Ae^{\lambda t}\left(\lambda^2 m + \lambda c + k\right) = 0 \tag{2.48}$$

Applying the quadratic formula gives:

$$\lambda = \frac{-c \pm \sqrt{c^2 - 4mk}}{2m} \tag{2.49}$$

This equation may have multiple solutions as the square root may be positive, negative or zero, depending on the values of m, c, k. Critical damping is defined as the damping value c giving the square root of zero:

$$c_{cr}^2 - 4mk = 0 \longrightarrow c_{cr} = \sqrt{4mk} = 2m\omega \tag{2.50}$$

The damping ratio is defined as:

$$\xi = \frac{c}{c_{cr}} = \frac{c}{2m\omega} \tag{2.51}$$

Substitution of c in the quadratic formula gives:

$$\begin{aligned} \lambda &= \frac{-2m\omega\xi \pm \sqrt{(2m\omega\xi)^2 - 4mk}}{2m} \\ &= -\omega\xi \pm \sqrt{\omega^2\xi^2 - \frac{k}{m}} \\ &= -\omega\xi \pm \omega\sqrt{\xi^2 - 1} \end{aligned} \tag{2.52}$$

Most civil engineering structures are underdamped, i.e. $\xi^2 - 1 < 0$, leading to:

$$\lambda = -\omega\xi \pm i\omega\sqrt{1 - \xi^2} \tag{2.53}$$

Expressions for eigenfrequency, ω, damping, ξ and damped eigenfrequency, ω_d, is then obtained:

$$\omega = |\lambda| = \sqrt{\lambda\lambda^*} \tag{2.54}$$

$$\xi = -\frac{Re(\lambda)}{\omega} = -\frac{Re(\lambda)}{|\lambda|} \tag{2.55}$$

$$\omega_d = \omega\sqrt{1 - \xi^2} = Im(\lambda) \tag{2.56}$$

And their relation can be depicted as shown in Figure 2.3.

Table 2.1: Hard and soft criteria used in this work

Hard criteria

nr.	Criterion
VH 1	$\xi > 0$
VH 2	$\xi < 20\%$
VH 3	Complex conjugate exists: that is one $\overline{\lambda_j}$ for every λ_j

Note:

Soft criteria are normalized to values from 0 to 1 where:

0 is ideal physical

1 is ideal spurious

Soft criteria

nr.	Criterion
VS 1	$d(\lambda_j, \lambda_l)$
VS 2	$d(f_j, f_l)$
VS 3	$d(\xi_j, \xi_l)$
VS 4	$n\mathrm{MAC}(\phi_j, \phi_l)$
VS 5	$n\mathrm{MTN}^S_{\infty j}$
VS 6	$d(\mathrm{MTN}^S_{\infty j}, MTN^S_{\infty l})$
VS 7	$n\mathrm{MPC}(\phi_j)$
VS 8	$\mathrm{MPD}(\phi_j)/90°$
VS 9	$n\mathrm{MTN}^d_{\infty j}$
VS10	$d(\mathrm{MTN}^d_{\infty j}, \mathrm{MTN}^d_{\infty l})$
VS11	$d(q_j, q_l)$
VS12	$n\mathrm{MOC}(\tilde{o}_j, \tilde{o}_l)$

2.6 Filter criteria

The stabilization diagram may at first be full of spurious modes, depending on the degree of overestimation and the amount of noise in the signal that the modes originate from. To ensure precise filtering of the spurious modes many filter criteria have been derived. The criteria used in this thesis is summarized in Table 2.1. The observant reader may notice great similarity with criteria used in Reynders et al. (2012). That is true, with the addition of *MOC* and adjustment to make all soft criteria, VS □, ideally spurious at the value of 1 and ideally physical at 0. For the criteria comparing modes the value of 0 can be interpreted identical modes, i.e. no difference, and 1 as totally different.

Which modes to compare is a natural question to ask. Modes approximating the same physical system mode should appear as vertical lines in the stabilization diagram. The most common approach is therefore to compare each mode with the closest mode in the closest model order below. Attempts using the closest mode in the the whole stabilization diagram have been made, but is not common practice. Quantification of which mode is closest in a model order can be done with all the mentioned filter criteria. Some are better than other and drawing lines between a mode and the mode that it finds the closest can then be used as a control.

In a manual analysis all these criteria can be used as filters by setting threshold values. The thresholds are subjects of trial and error, combined with the experience of skilled analysists. An understanding of the distribution of modes within each criteria is therefore needed. The following section is summarizing the different criteria, and their characteristics.

2.6.1 Hard criteria

The hard criteria is different from the soft criteria in the sense that they yield binary numbers determining if a mode is possibly physical or not. Damping ratios, ξ, less than zero does not damp the system but rather drive the system. Real life structures are always damped and this is hard criteria VH 1. Damping ratios above 20% could arguably occur, but most applications in civil engineering structures are way below 20%. Hard criteria, VH 2, may be adjusted to the structure considered. The presence of a complex conjugate pair was shown in section 2.5.3 and is the basis of hard criteria VH 3.

2.6.2 Soft criteria

The eigenvalue λ contains information about the eigenfrequency ω, damping ξ and then naturally the damped eigenfrequency w_d which is dependent on both ω and ξ. This was shown in Section 2.5.3. Possible criteria are then many, starting with the eigenvalue:

$$d(\lambda_j, \lambda_l) = \frac{|\lambda_j - \lambda_l|}{\max(|\lambda_j|, |\lambda_l|)} \tag{VS 1}$$

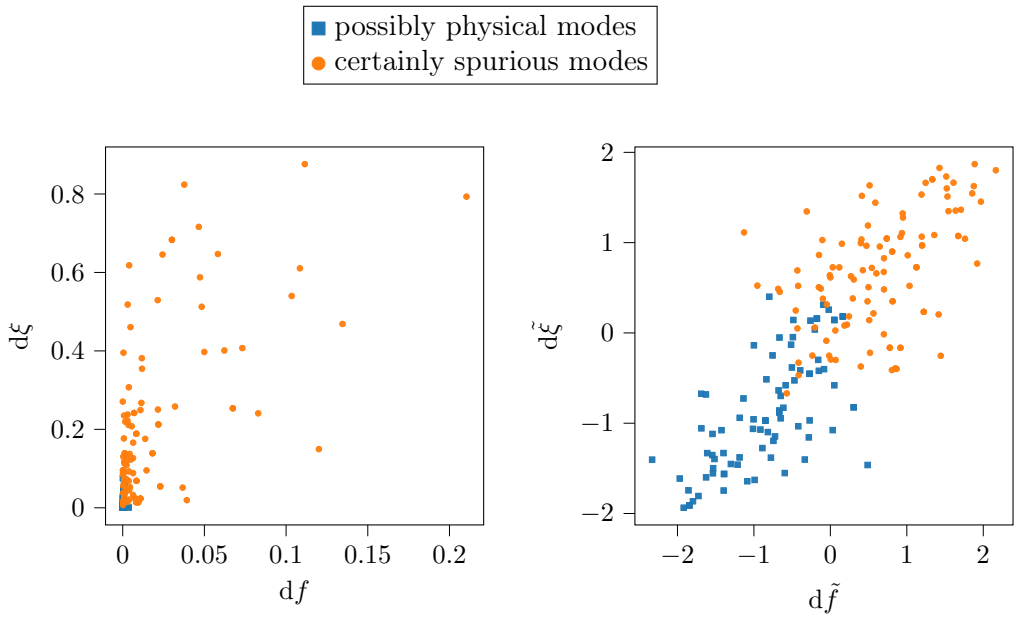


Figure 2.4: Change in frequency, df, vs change in damping, dξ of all modes before,□, and after,□̃ box-cox transformation and normalization.

Eigenfrequency and damping may also be separate criteria. Frequencies in Hz are more commonly used because frequency, and the period $T = 1/f$, is physically intuitive. With frequency f in Hz:

$$d(f_j, f_l) = \frac{|f_j - f_l|}{\max(f_j, f_l)} \tag{VS 2}$$

$$d(\xi_j, \xi_l) = \frac{|(\xi_j - \xi_l)|}{\max(\xi_j, \xi_l)} \tag{VS 3}$$

The change in frequency, df, is usually very small, and especially for the modes that are good estimates. Change damping ratio, ξ, is usually higher and more scattered. Rainieri et al. (2014) suggested that modes with $df < 0.01$ and $d\xi < 0.05$ should be considered stable. The spread of modes is visually depicted in the leftmost plot in Figure 2.4, plotting change in frequency df vs. change in damping $d\xi$. The modes shown is from analysing the shear frame in the numerical study, Section 3.1. It is very clear that the spurios modes are more widespread, and that frequency deviations are smaller. The rightmost plot shows the spread of modes after box-cox transformation and normalization. This a step of the algorithm suggested by Yang to ensure that the clustering is a precise as possible, but this is further discussed in Section 2.8.

Most modes are seperated in terms of f, as earlier mentioned, but this not always the case. Extra criteria is therefore needed to seperate closely spaced modes. Modal Assurance Criteria (MAC) compares the unscaled mode shapes ϕ_j, ϕ_l of mode j and l. A high similiarity results in MAC values near 1, and the adjustment to nMAC may be defined:

$$\mathrm{MAC}(\phi_j, \phi_l) = \frac{|\phi_j^* \phi_l|^2}{||\phi_j||_2^2 ||\phi_l||_2^2} \tag{2.57}$$

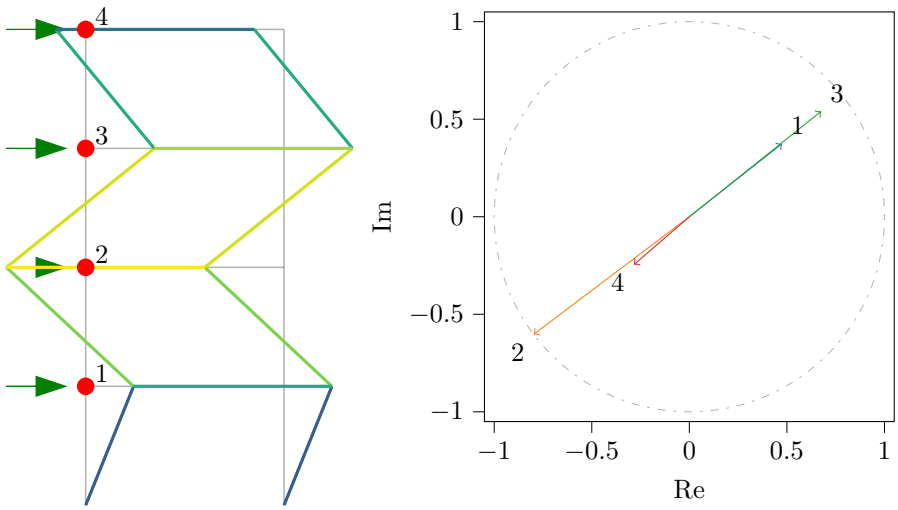
$$\mathrm{nMAC}(\phi_j, \phi_l) = 1 - \mathrm{MAC}(\phi_j, \phi_l) \tag{VS 4}$$

Modal Transfer Norm (MTN) was introduced by Reynders et al. (2008). MTN measures the contribution from a single mode to the total response. The superscripts $\square^d$ and $\square^s$ is to differentiate between deterministic MTN and stochastic MTN. Subscript $\square_{\infty j}$ is to declare that MTN equals the peak gain of a transfer function containing only mode j.

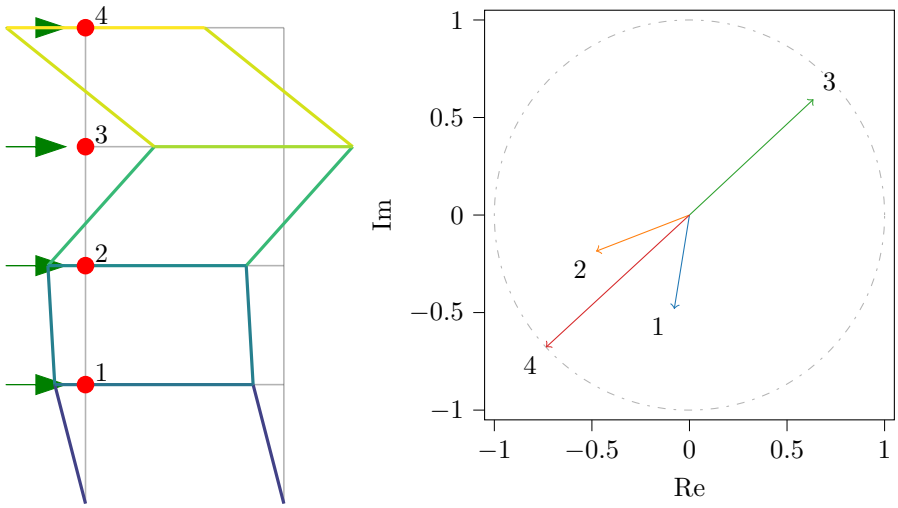
$$\mathrm{MTN}^s_{\infty j} = \max \quad \sigma \left(\frac{\phi_j g_{dj}^T}{e^{i\omega_j T} - \lambda_{dj}} + \frac{\phi_j g_{dj}^T}{2\lambda_{dj}} \right) \tag{2.58}$$

$$\mathrm{nMTN}^s_{\infty j} = 1 - \mathrm{MTN}^s_{\infty j} \tag{VS 5}$$

where $\sigma(\square)$ is the set of singular values, g_{dj} is the discrete stochastic participation vector of mode j, λ_{dj} the discrete time eigenvalues which can be obtained by Equation (2.45).



(a) Rayleigh damped shearframe



(b) Shearframe with non-diagonal damping matrix.

Figure 2.5: Fourth mode shape of two four story shear frames plotted on geometry and in the complexity plot. Modes are identified with Covariance Driven Stochastic Subspace Identification (SSI-COV) and Reynders algorithm. Numbers shown indicate sensor layout/degree of freedom.

Stabilization diagrams showing only the modes with the highest MTN, i.e. most contributing modes, are typically very clear. The change in MTN may also be defined in order to compare two modes:

$$d(\mathrm{MTN}^s_{\infty j}, \mathrm{MTN}^s_{\infty l}) = \frac{|\mathrm{MTN}^s_{\infty j} - \mathrm{MTN}^s_{\infty l}|}{\max(|\mathrm{MTN}^s_{\infty j}|, |\mathrm{MTN}^s_{\infty l}|)} \tag{VS 6}$$

The deterministic counterparts to the given equations are:

$$\mathrm{MTN}^d_{\infty j} = \max \quad \sigma \left(\frac{\phi_j g^T_{dj}}{e^{i\omega_j T} - \lambda_{dj}} 2\lambda_{dj} \right) \tag{2.59}$$

$$\mathrm{nMTN}^d_{\infty j} = 1 - \mathrm{MTN}^s_{\infty j} \tag{VS 9}$$

$$d(\mathrm{MTN}^d_{\infty j}, \mathrm{MTN}^d_{\infty l}) = \frac{|\mathrm{MTN}^d_{\infty j} - \mathrm{MTN}^d_{\infty l}|}{\max(|\mathrm{MTN}^d_{\infty j}|, |\mathrm{MTN}^d_{\infty l}|)} \tag{VS10}$$

Mode shape complexity measures

Mode shapes can visually be plotted on the geometry, or in the *complex plot*, as shown in Figure 2.5. Each vector in the complex plot is corresponding to one DOF, and the real part of the vector is the same as the deflection of the frame. If animated in time, the vectors in the complex plane would be rotating anticlockwise and the real part of the vectors represent the oscillating deflection of the DOFs.

Structures without damping, or proportional damping, have modes that are *normal*. Normal modes are modes without complexity, which practically means that all DOFs are synchronized to be at their max at the same time. In the complex plot this means that all DOF point along the same line, but not necessarily the same direction. An example of a normal mode is shown in on the Rayleigh damped shear frame in Figure 2.5b. The vectors are not perfectly straight lined, but that is due to the fact that the modes are found through OMA with applied measurement noise.

Introduction of non-proportional damping leads to modes that are complex, i.e. out of phase, and every DOF reach its max at different time instants. Large damping variation along the diagonal of the damping matrix gives the largest complexity which is observed as vectors pointing in all directions. A complex mode is shown on the shear frame in Figure 2.5a. The only difference between the two shear frames in Figure 2.5 is the damping matrix C.

To quantify the complexity between the l DOFs multiple measures are possible. Modal Phase Colinearity (MPC) measures how straight aligned all the DOFs are in one number, where 0 is no co-linearity and 1 is perfect co-linearity. One should, however, use MPC with caution as not all structural modes yield straight-line characteristics in the complex plot. Examples of such structures are double symmetric structures and axisymmetric structures. Real life mode shapes with vectors pointing in all directions, that is making up circle in the complex plot, have also been identified for nearly axisymmetric structures (Dooms et al., 2006).

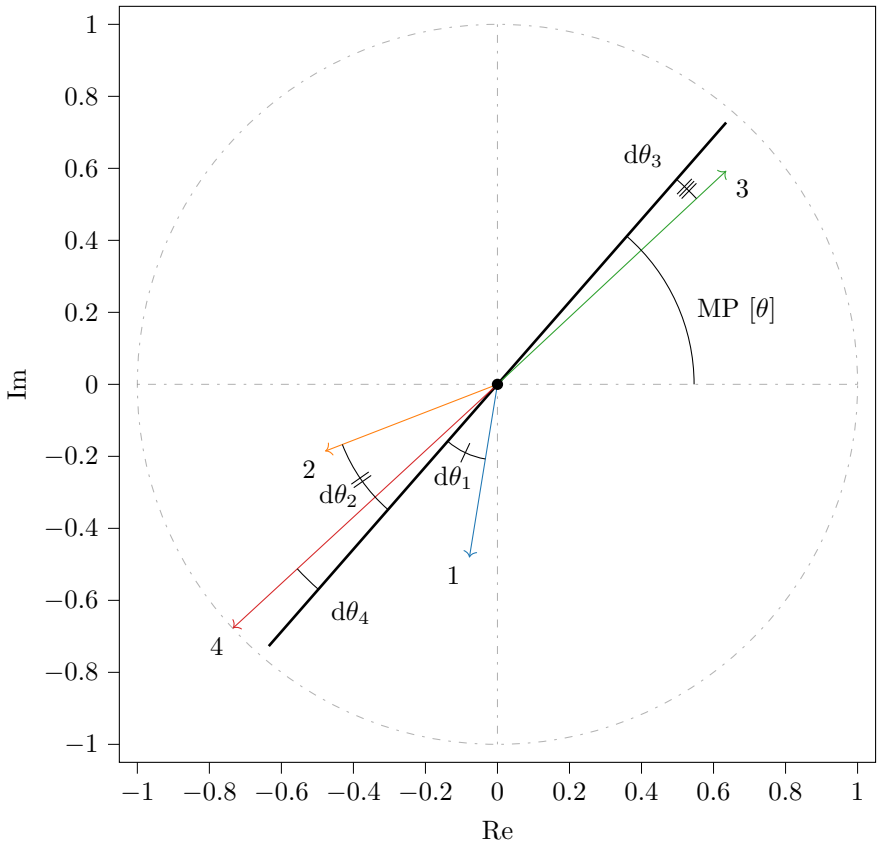


Figure 2.6: Argand diagram of a mode shape with mean phase, and individual contribution to the MPD of each DOF.

MPC and the normalized version used in this thesis is defined as:

$$\mathrm{MPC}(\phi_j) = \frac{\|\mathrm{Re}(\tilde{\phi})\|^2 + \frac{2(\epsilon_{MPC}^2+1)\sin^2(\theta_{MPC})}{\epsilon_{MPC}}\|\mathrm{Re}(\tilde{\phi}^T)\|^2\|\mathrm{Im}(\tilde{\phi})\|^2}{\|\mathrm{Re}(\tilde{\phi})\|^2 + \|\mathrm{Im}(\tilde{\phi})\|^2} \tag{2.60}$$

$$n\mathrm{MPC} = 1 - \mathrm{MPC}(\phi_j) \tag{VS 7}$$

where:

$$\tilde{\phi_j} = \phi_j - \bar{\phi_j} = \phi_j - \sum_{r=1}^{l}\frac{\phi_{j.r}}{l} \tag{2.61}$$

$$\epsilon_{MPC} = \frac{\|\mathrm{Im}(\tilde{\phi_j})\|^2 - \|\mathrm{Re}(\tilde{\phi_j})\|^2}{2\|\mathrm{Re}(\tilde{\phi_j}^T)\|^2\|\mathrm{Im}(\tilde{\phi_j})\|^2} \tag{2.62}$$

$$\theta_{MPC} = \arctan\left(|\epsilon_{MPC}| + \mathrm{sgn}(\epsilon_{MPC})\sqrt{1+\epsilon_{MPC}^2}\right) \tag{2.63}$$

$\|\square\|$ denotes the l_2 norm and sgn is the signum function, extracting the sign of a real number, in the given functions.

As an alternative Modal Phase Deviation (MPD) can be used. Then the Mean Phase (MP) must be defined. Visually MP is identified as the angle from the positive real part to the best fit line pointing the same direction as the DOFs, see Figure 2.6:

$$\mathrm{MP}_j = \bar{\psi_j} = \sum_{r=1}^{l}\frac{\psi_{j.r}}{l} \tag{2.64}$$

$$\tag{2.65}$$

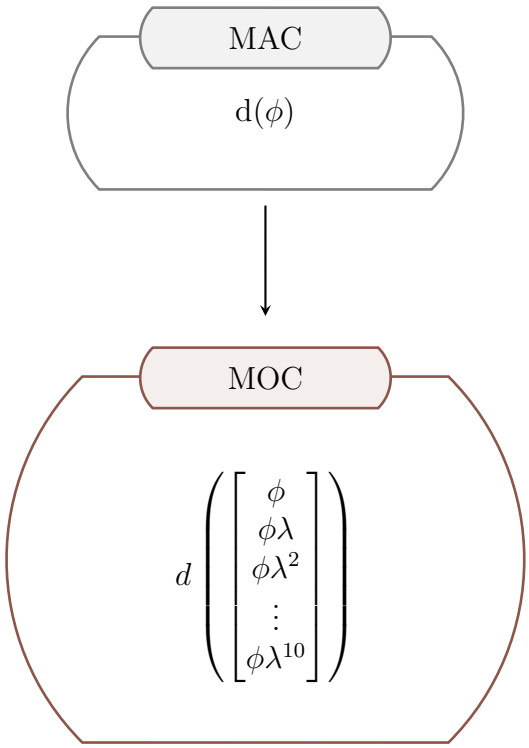
where r is the considered DOF and l the total number of sensor positions/DOFs:

$$\psi_{j.r} = \begin{cases} \arctan\left(\dfrac{\mathrm{Re}(\phi_{j.r})}{\mathrm{Im}(\phi_{j.r})}\right) & if \quad \left(\dfrac{\mathrm{Re}(\phi_{j.r})}{\mathrm{Im}(\phi_{j.r})}\right) >= 0 \\ \arctan\left(\dfrac{\mathrm{Re}(\phi_{j.r})}{\mathrm{Im}(\phi_{j.r})}\right) + \pi & otherwise \end{cases} \tag{2.66}$$

The Modal Phase Deviation (MPD) is then simply:

$$\mathrm{MPD}(\phi_j) = \sum_{r=1}^{l}\sqrt{\frac{(\psi_{j.r} - \mathrm{MP}_j)^2}{l}} \tag{VS 8}$$

The last criteria used in this thesis to asses modal complexity is the newest one. It was first introduced by Yaghoubi et al. (2018). Its derivation starts by transformation of the dynamic system to modal form by the use of the following transformation: $\tilde{x}(t) = \psi^{-1}x(t)$, where $[\Psi]$ is the matrix with eigenvectors $\{\psi\}$ of $[A]$.


MAC
d(ϕ)
MOC
$d\left(\begin{bmatrix} \phi \\ \phi\lambda \\ \phi\lambda^2 \\ \vdots \\ \phi\lambda^{10} \end{bmatrix}\right)$

The observability matrix can then be written on modal form:

$$
\begin{aligned}
\underset{il\times l}{[\tilde{\mathcal{O}}_i]} &= \begin{bmatrix} [\tilde{C}] \\ [\tilde{C}][\tilde{A}] \\ [\tilde{C}][\tilde{A}]^2 \\ \vdots \\ [\tilde{C}][\tilde{A}]^{i-1} \end{bmatrix} = \begin{bmatrix} [C][\Psi] \\ [C][\Psi][\Lambda] \\ [C][\Psi][\Lambda]^2 \\ \vdots \\ [C][\Psi][\Lambda]^{i-1} \end{bmatrix} = \begin{bmatrix} [\Phi] \\ [\Phi][\Lambda] \\ [\Phi][\Lambda]^2 \\ \vdots \\ [\Phi][\Lambda]^{i-1} \end{bmatrix} \\
&= \begin{bmatrix} \{\phi_1\} & \dots & \{\phi_i\} \\ \{\phi_1\}\lambda_1 & & \{\phi_i\}\lambda_i \\ \{\phi_1\}\lambda_1^2 & \ddots & \{\phi_i\}\lambda_i^2 \\ \vdots & & \vdots \\ \{\phi_1\}\lambda_1^{i-1} & \dots & \{\phi_i\}\lambda_i^{i-1} \end{bmatrix} = \begin{Bmatrix} \{\tilde{o}_1\} & \{\tilde{o}_2\} & \dots & \{\tilde{o}_i\} \end{Bmatrix}
\end{aligned}
\tag{2.67}
$$

The defined observability vector, $\tilde{o}_i$, is a vertical stack of block rows where each block contains the mode shape, ϕ_i, multiplied with the corresponding eigenvalue, λ, to the power of the block row minus 1. The number of block rows are in this thesis chosen to be 10. The definition of MOC is similar to the definition of MAC:

$$
\mathrm{MOC}(\tilde{o}_j, \tilde{o}_l) = \frac{|\tilde{o}_j^* \tilde{o}_l|^2}{||\tilde{o}_j||_2^2 ||\tilde{o}_l||_2^2} \tag{2.68}
$$

$$
\mathrm{nMOC}(\tilde{o}_j, \tilde{o}_l) = 1 - \mathrm{MOC}(\tilde{o}_j, \tilde{o}_l) \tag{VS12}
$$

By setting the number of block rows to 1, MOC yields the exact same results as MAC. The great advantage of MOC is that modes with similar mode shapes, but different frequencies are separated. That can be the case if the number of sensors used in the measurement scheme is insufficient. This phenomenon is called spatial aliasing, and will always be a consideration when working with a limited number of sensors.

When measured forces are available, that is in an EMA or OMAX context, it is natural to use Modal Scale Factor (MSF). MSF is defined as:

$$
\mathrm{MSF}(\hat{\phi}_j, \hat{\phi}_l) = \frac{\hat{\phi}_l^* \hat{\phi}_j}{\|\hat{\phi}_l\|^2} \tag{2.69}
$$

where $\hat{\phi}$ is the unit modal mass normalized mode shape ϕ. When MSF yields values close to ± 1 the two mode shapes, ϕ_j and ϕ_l have close amplitude and phase. Stabilization criteria VS11 can then be defined:

$$
d(q_j, q_l) = \left| \ln \left(\frac{\mathrm{MSF}(\hat{\phi}_j, \hat{\phi}_l)}{\mathrm{sgn}(\max(\mathrm{Re}(\mathrm{MSF}(\hat{\phi}_j, \hat{\phi}_l)), \mathrm{Im}(\mathrm{MSF}(\hat{\phi}_j, \hat{\phi}_l))} \right) \right| \tag{VS11}
$$

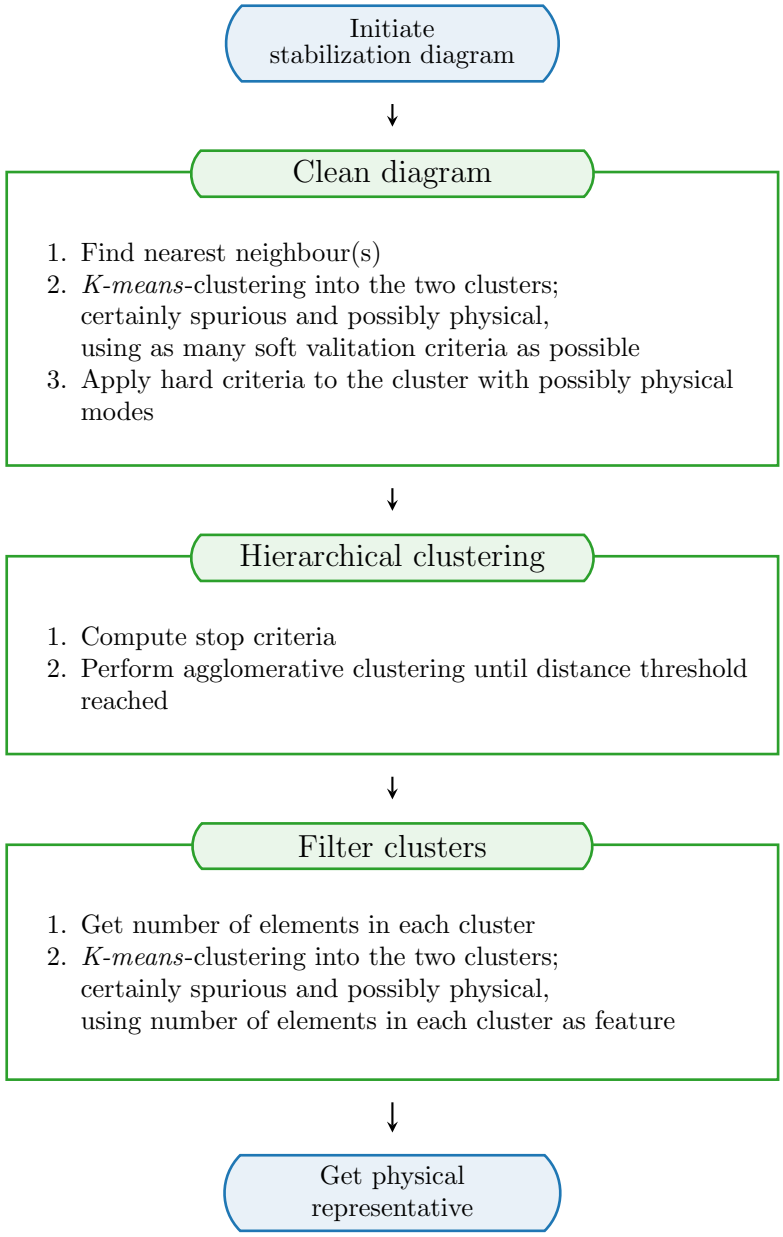


Figure 2.7: Flowchart of Reynders' algorithm.

2.7 Reynders' algorithm

Reynders et al. (2012) suggested an algorithm that automates the interpretation of the stabilization diagram. The algorithm is hereby referred to as Reynders' algotithm. The approach is based on the following five target criteria, (Reynders et al., 2012):

1. not rely on more than one data record or on prior estimates for any of the modal parameters
2. be as physically intuitive as possible and follow the course of a manual analysis
3. produce similar results as in a manual analysis
4. work in an EMA, OMA and OMAX framework and with any parametric system identification algorithm
5. not contain parameters that need to be specified or tuned by the user

The algorithm follows three main steps, as shown in Figure 2.7, and all steps will in the following subsections be explained. Worth noting is that the original paper defining the algorithm may not contain the exact sub steps in its own description, but the process is similar except when modifications are stated.

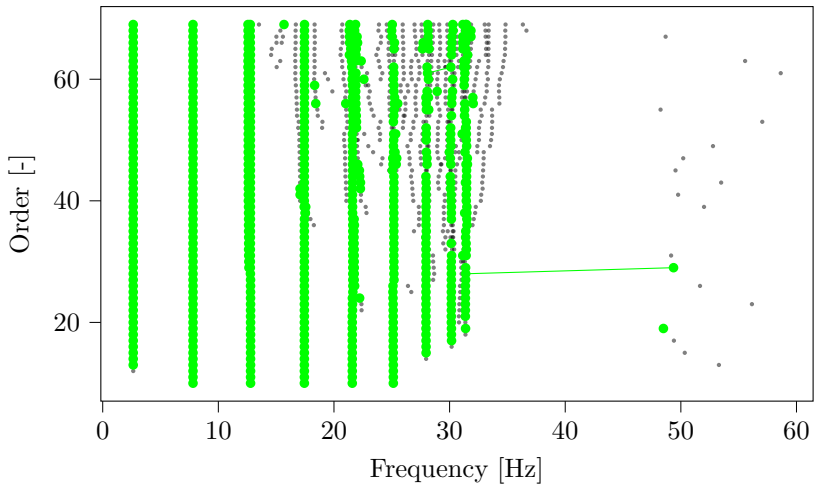
Step 1. Automated clearing of a stabilization

Step 1 intents to remove the majority of the spurious modes. In a manual analysis this is done by setting threshold values on properties like damping ξ, or change in properties like frequency, df, between a mode and the closest mode in the order above or below. Reynders' algorithm is not using threshold values but rather clustering based on as many criteria as possible in order to separate the modes into the two categories; certainly spurious modes, and possibly physical modes. Reynders is using the closest mode in the closest model order below for all criteria needing another mode for comparison as seen in Table 2.1. The closest mode in the model order above or below is hereby referred to as *nearest neighbours* to simplify reading. Reynders does not state how, or on which criteria, the neighbours are determined. In this thesis different criteria combinations was tried and the euclidean norm of VS 2 and VS 3 proved good performance:

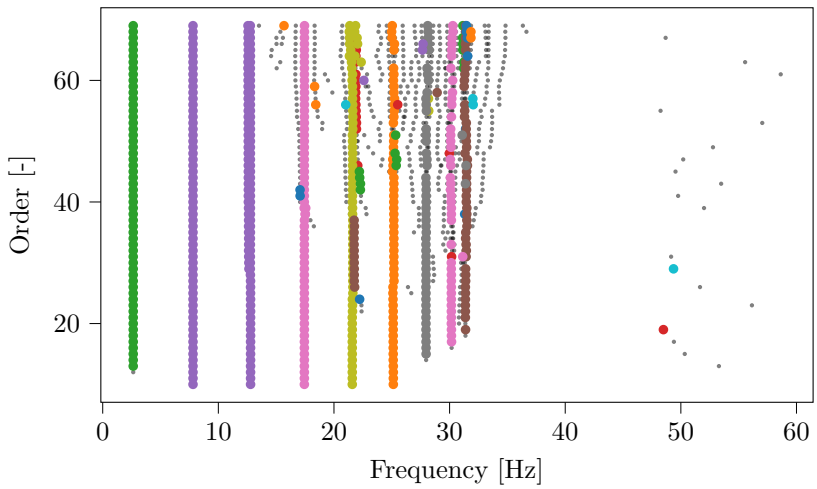
$$d = ||d(f_j, f_l), d(\xi_j, \xi_l)|| \tag{2.70}$$

Sophisticated algorithms like *Nearest Neighbours* in *scipy* (Python-library) could possibly be used for determination, but as of today they do not have the possibility to use custom distance metrics, (Pedregosa et al., 2011). Brute force was therefore chosen to find the nearest neighbours. If the implementation in this thesis were to be optimized a feasible way of approach is to only calculate the distance metric to the modes within a given change in frequency, df.

To avoid loss of modes, by comparison with modes that are certainly spurious, it is in this thesis chosen to use the closest neighbour above or below. This is neither described in Reynders nor in Yang, but it is an easy way to overcome gaps where there is a missing approximation. Another benefit is that all modes will have a neighbour,



(a) Filtered



(b) Clustered

Figure 2.8: Filtered and clustered stabilization diagram created by shear frame data.

also at the highest and lowest model order depending on direction of comparison, which makes the implementation simpler.

Having found a neighbour for every mode the next step is to construct the validation vector of each mode with length equal the number of soft validation criteria, n_{vs}. The clustering is then performed in the n_{vs} dimensional space $\mathbb{R}^{n_{vs}}$. Algorithm employed to perform the clustering is in this thesis k-means from *scipy* with two clusters, $k = 2$. K-means create k cluster centers and iteratively works through all objects to minimize the euclidean distance between every object and its cluster center. The objective function of K-means is:

$$O_m = \sum_{k=1}^{2} \sum_{j=1}^{n_m} \mathbf{B}_{k.j} \|\mu_k - \mathrm{vs}_{m.j}\|^2 \tag{2.71}$$

where m_j is mode object number j with its soft validation vector $\mathrm{vs}_{m.j}$, and $\mathbf{B}_{k.j}$ is a Boolean determining if mode m_j is belonging to cluster center μ_k or not. With k-means from scipy it is both possible to specify the initial clusters or let the algorithm choose it by it self. This thesis lets the k-means algorithm determine the centers its self.

With two clusters the possibly physical one is the one with cluster center closest to the origin, i.e. soft validation criteria, VS $\square$, closest to zero. This cluster is then filtered by the hard criteria shown in Table 2.1. After step 1 is performed the stabilization diagram should appear as if threshold values was applied, just like in an manual analysis. However, all modes not aligning as vertical lines will not necessarily be removed as these are discarded in later steps. This is shown in Figure 2.8. The connection between every possibly physical mode and it's nearest neighbour is plotted, but only a few lines are visible. Note that no lines are passing the vertical gaps, which means that there is less modes lost using the closest neighbour above or below.

Step 2. Grouping similar modes

The grouping of similar modes is performed by the use of Agglomerative Clustering which is a bottom up Hierarchical Clustering technique. At first, all modes are placed into separate clusters. Iteratively the two clusters that are closest are merged based on the distance between them. Usually this distance is between the cluster centers, but it may also be between the outermost elements.

In the first step of Reynders' algorithm the distance between any mode, M_j, and mode, M_l, is computed as:

$$\begin{aligned} d(M_j, M_l) &= \text{VS 1} + \text{VS 4} \\ &= d(\lambda_j, \lambda_l) + n\mathrm{MAC}(\phi_j, \phi_l) \end{aligned} \tag{2.72}$$

The process of clustering modes is performed until the distance between any two clusters reach the stop distance threshold. The distance stop threshold used is the mean plus two standard deviations $d = \mu + 2\sigma$, of the distance between all the possibly physical modes and their respective closest neighbour. This threshold can be extracted from the soft validation vectors used in Step 1.

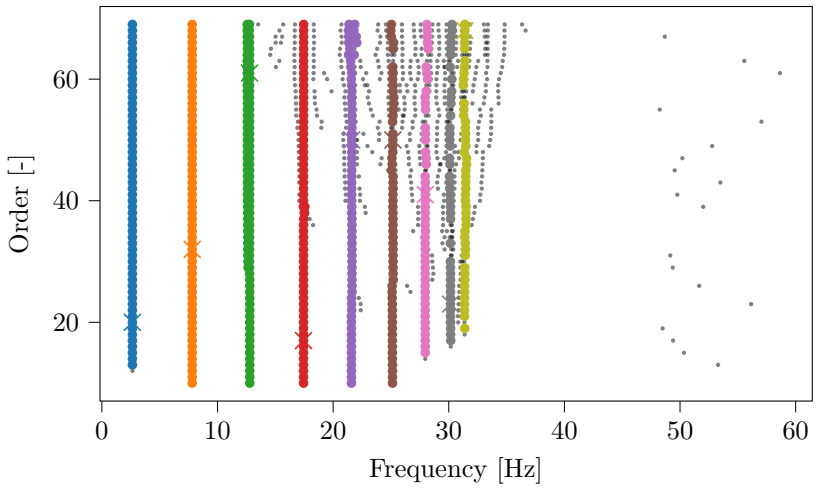


Figure 2.9: Automatically interpreted stabilization diagram with Reynders' method.

Step 3. Filter clusters

The output of the hierarchical clustering yields clusters of very different lengths, as shown in Figure 2.8. The best possible cluster contains one estimate, mode, of the physical system at every model order in the analysis. In reality this is nearly never the case, but one can in general state that clusters with many modes are typically representing the physical system modes, and clusters with only a few modes are spurious. Therefore Reynders' algorithm employs K-means with two clusters, $k = 2$, based on the length of the clusters. This results in the removal of the smallest clusters.

Having only a limited set of clusters a physical representative is chosen from each cluster. The physical representative is chosen by the median damping to avoid the affect of outliers.

Wort noting is if the model order is to low one might get biased modes, that is modes containing a combination of different system modes. Typically this occurs when to two system modes have close frequencies. The problem can be solved by setting the highest model order high enough to observe the split point and only use points above the split. In Reynders' algorithm there is no implemented fix to the biasing problem, and the user should therefore plot all, also spurious, modes in the stabilization diagram to visually check if biasing and splitting occurs.

The stabilization diagram in Figure 2.9 shows the outcome of Reynders' algorithm; a cleared stabilization diagram with picked representatives, $\times$. There was in this case perfect match between the number of identified clusters and the number of system modes. There are gaps where modes are "missing", which is a result of the fitting process. Note also that the green cluster has two parallel columns of dots. This is not due to biasing, or aliasing, but the fitted model; modes have to be placed somewhere. This results in two approximations of the same system mode, where one or both is spurious in the sense that they are bad measurements. In this case there is not external affects, making it a good example of mathematical modes. The picked representative is as seen in the longest column, which most likely is best.

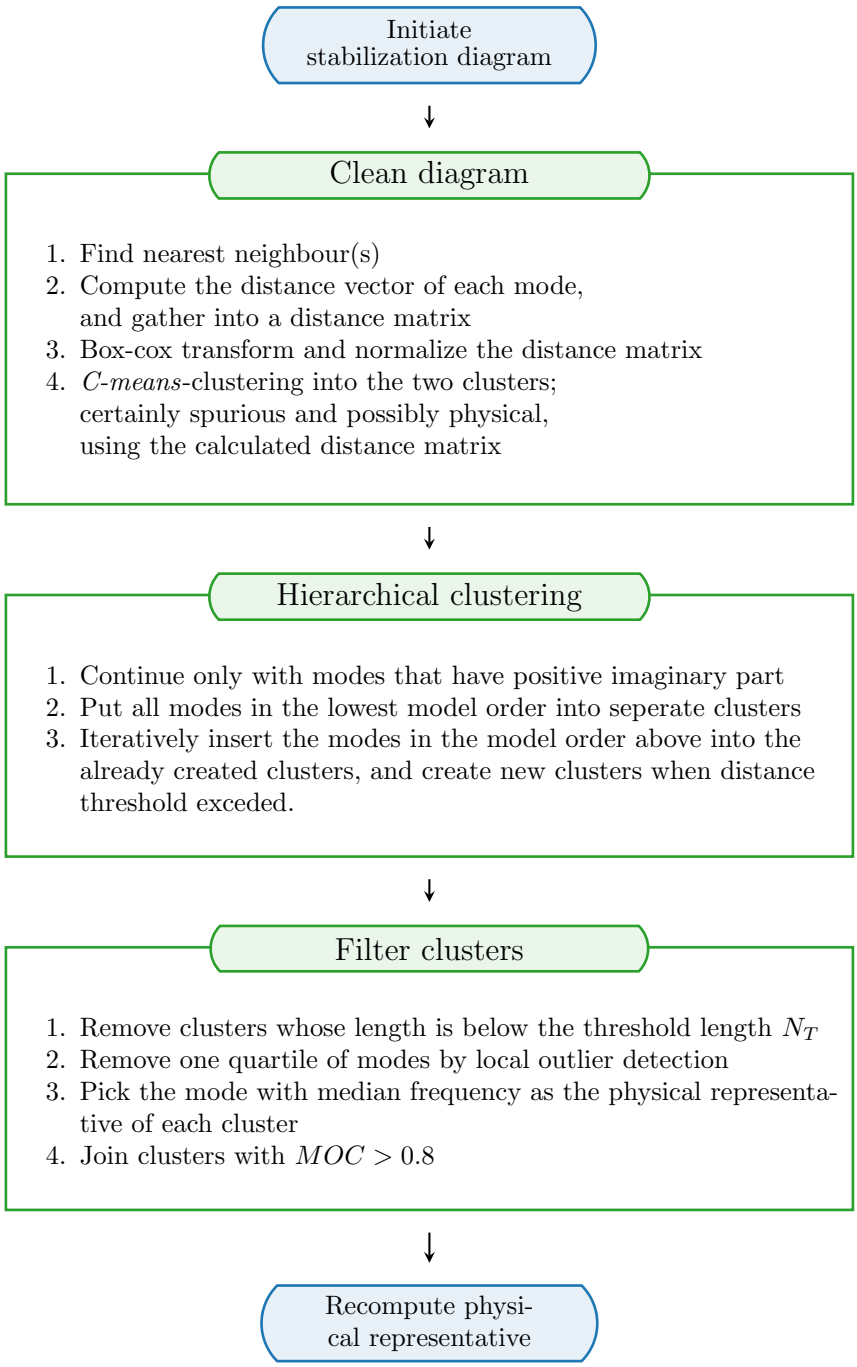


Figure 2.10: Flowchart of Yang's algorithm.

2.8 Yang's algorithm

Yang's algorithm have great similarity with Reynders' algorithm. Step ways they are not divided into the same steps in their respective papers, but should have been in the opinion of the writer of this thesis. The flowchart in Figure 2.10 is therefore divided into the same three main steps as Reynders' algorithm was.

One main difference between Yang's and Reynders' is the shift in focus from using as many validation criteria as possible to using fewer and more strict criteria. Modal Observability Correlation (MOC) is widely used throughout the algorithm and replaces the use of MAC completely.

Step 1. Clean diagram

The first step when cleaning the stabilization diagram is calculating the nearest neighbour in the order below. In contrast to Reynders the distance measure between any mode, M_j, and mode, M_l, is specified when determining neighbours:

$$\begin{aligned} d(M_j, M_l) &= \text{VS 2} + \text{VS 12} \\ &= d(f_j, f_l) + n\text{MOC}(\tilde{o}_j, \tilde{o}_l) \end{aligned} \tag{2.73}$$

In order to avoid loosing modes and problems with endpoints the implementation in this thesis is using the closest neighbour in the order above or below, just as described in Section 2.7. Having determined a neighbour for each mode the next step is constructing the feature vector of each mode, and gather all vectors into a distance matrix, D. The distance vector between mode, M_j, and it's neighbour mode, M_l, is given as:

$$\begin{aligned} d(M_j, M_j) &= \begin{bmatrix} \text{VS 2} & \text{VS 3} & \text{VS 12} & \text{VS 7} \end{bmatrix} \\ &= \begin{bmatrix} d(f_j, f_l) & d(\xi_j, \xi_l) & n\text{MOC}(\tilde{o}_j, \tilde{o}_l) & n\text{MPC}(\phi_j, \phi_l) \end{bmatrix} \end{aligned} \tag{2.74}$$

Fuzzy C-means clustering is performing better on normal distributed data. The Box-Cox transformation by Box et al. (1964) is therefore used on each feature, y, of the the distance matrix, D. y_i is then a column of D and $y_{i,T}$ denotes the Box-Cox transformed feature y_i.

$$y_T = \begin{cases} \dfrac{(y^\lambda - 1)}{\lambda} & \text{for } \lambda ! = 0 \\ \ln(y) & \text{for } \lambda = 0 \end{cases} \tag{2.75}$$

where λ is a solution of the profile likelihood function.

After the transform each feature of the distance matrix is normalized, that is subtracting the mean and dividing by the standard deviation:

$$y_{T.N} = \frac{y - \mu_y}{\sigma_y} \tag{2.76}$$

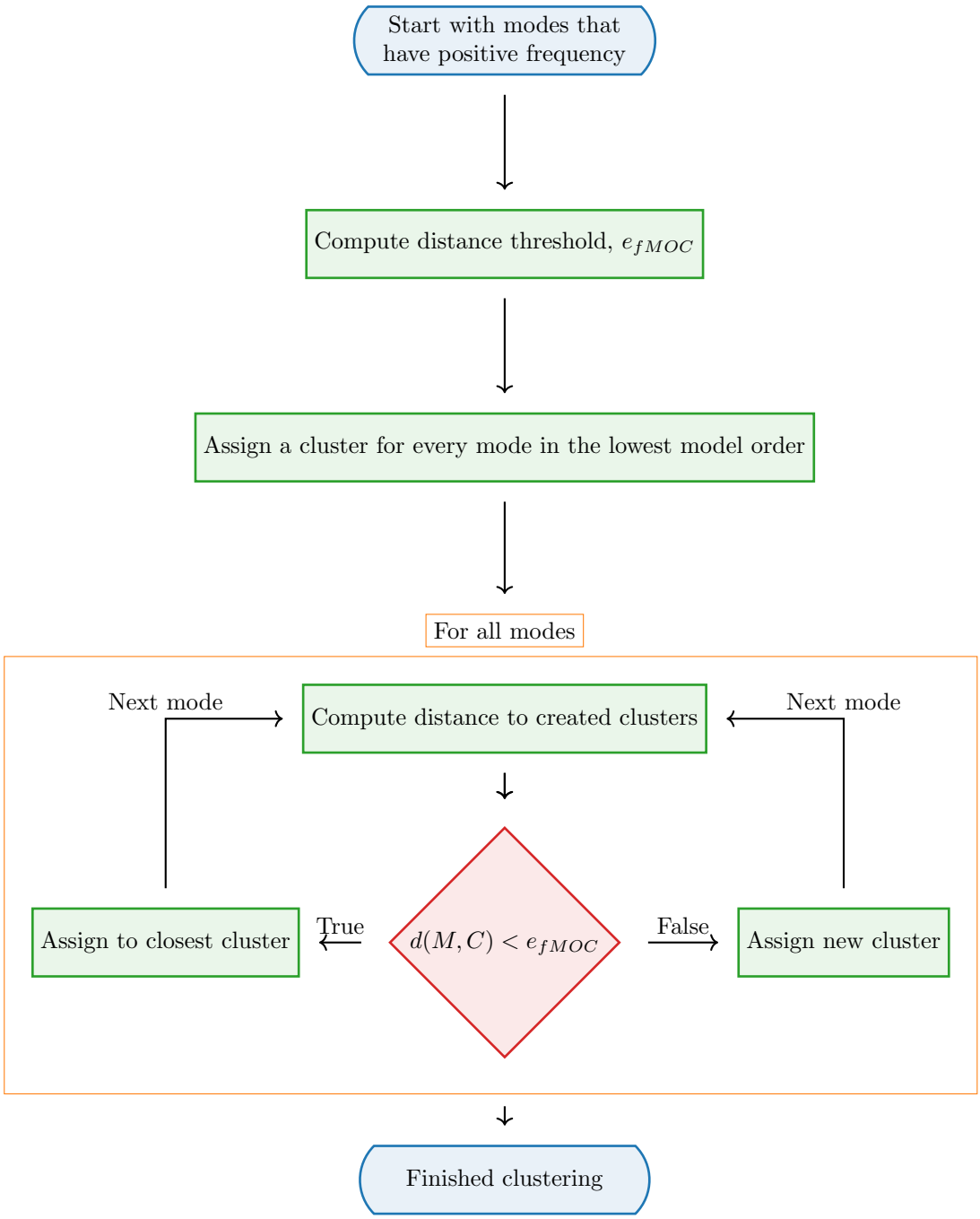


Figure 2.11: Flowchart of clustering method suggested by Yang.

The input and output of this process of transformation and normalization was depicted in Figure 2.4.

C-means clustering is then performed with the distance matrix, D, given. Yang suggests specifying initial cluster centers. The implementation in this thesis has skipped this step as the final result is the same in all normal cases and computational gain may be small and possibly negative.

The output of step 1. is similar to step 1 of Reynders'; a filtered stabiliztion diagram

Step 2. Hierarchical clustering

Yang suggest a new way to cluster modes, that possibly is more computationally efficient depending on implementation. The clustering process is depicted in Figure 2.11.

In order to end up with modes that have positive frequency, and reduce computational effort, only modes with positive imaginary part of the eigenvalue is kept.

The distance threshold, e_{fMOC}, is given by the mean plus two standard deviations of all possibly physical modes and their respective neighbours:

$$d = [d(M_1, M_{1.N}), d(M_2, M_{2.N}), \ldots, d(M_{nModes}, M_{nModes.N})] \tag{2.77}$$

$$e_{fMOC} = \mu(d) + 2 \times \sigma(d) \tag{2.78}$$

At first all modes at the lowest model order is put in separate clusters. Then all other modes are appended to the existing clusters, or put in new clusters if the distance between a mode and the clusters is greater than the calculated threshold, e_{fMOC}. The distance between a mode and a cluster is calculated as:

$$d(M, C) = \frac{1}{N_C} \sum_{i=1}^{N_C} d(M, M_i) \tag{2.79}$$

where N_C is the number of modes in cluster C, M_i is mode number i in the cluster, and $d(M, Mi)$ is given by Equation (2.73).

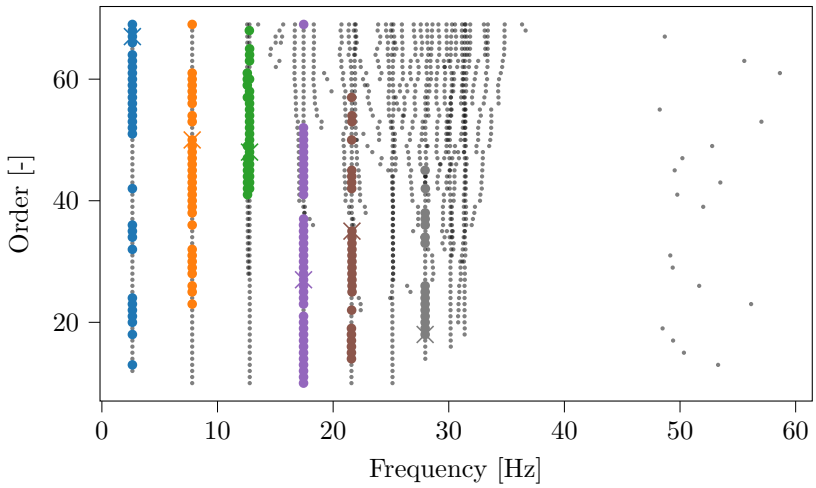


Figure 2.12: Automaticly interpreted stabilization diagram with Reynders method.

Step 3. Filter clusters

As with Reynders there are clusters with spurious modes, that are characterized by short length. These short clusters are removed by setting a cluster length threshold, $N_T = \alpha \times (N_{max} - N_{min})/\beta$. N_{max}, N_{min}, and β is then the max model order, min model order, and step between model orders respectively. α is a factor that may be chosen in range $0.3 - 0.5$ without much difference in the final outcome. The implementation in this thesis uses 0.3 as this yields the highest number of detections.

With many modes in each cluster some are outliers. Yang suggest removing one quartile of the modes by Local Outlier Factor (LOF) of frequency and damping. LOF is a relatively modern method that was first described in Breunig et al. (2000). It is not stated in Yang et al. (2019) whether to remove one quartile of one property first and then one more quartile of the remaining mordes, or one quartile of both combined or a separate calculations and merging afterward. In this implementation LOF is performed for damping and frequency separately. Modes that are outliers in either are removed, making the number of removed modes span between one and two quartiles. The number of modes is most likely closest to one quartile as the spurious modes typically are outliers in both frequency and damping.

Next, modes with median frequency is chosen as the physical representative. Median is a earlier mentioned better than average as outliers have a big affect on the average. The last operation is joining all clusters with MOC higher than 0.8 between their physical representatives, and recalculation of the physical representative in the merged clusters.

The outcome of Yang's algorithm is shown in Figure 2.12. The input data is exactly the same as the data used when describing the steps of Reynders' in Figures 2.8 and 2.9. Notice the major difference between Reynders' and Yang's in number of modes detected from this numerical measurement.

Chapter 3

Numerical studies

This chapter contains two fully digital studies. The scope of the studies were to compare the performance of the two automatic algorithms introduced in 2.7 and Section 2.8. To ensure that the performance is validated without flaws the testing is performed on numerical data created by loading an analytical shear frame.

3.1 Numerical study of shear frame

Determining the sources of error and their respective contributions to the total error is a big and important part of all research work. A analytical shear frame have therefore been implemented at the Department of Structural Engineering at NTNU. The shear frame is implemented as an object in *Python* and can be accessed through the freely available package *strid* on GitHub (Frøseth et al., 2022). With this shear frame one can vary the number of floors, the mass of each floor, the stiffness of each column-pair, the damping of the system, and the applied loading. The output acceleration response of the chosen shear frame can then be simulated and recorded exactly. This gives the possibility to check the performance of system identification algorithms compared to exact solutions. All errors apart from the computation error is introduced though the user, eliminating common problems in real life analyses like sudden loads, present harmonics, effect of temperature change, and so on.

To ensure that the algorithms have a challenge, that the model is not to complex, and that reproduction can be done, a simple shear frame with 9 floors have been chosen in this work, as seen in Figure 3.1. The mass of each floor is $m = 10^3$ and the stiffness of each story, that is pair of columns, $k = 10^7$. Damping is chosen to be classical Rayleigh damping, with damping ratio $\xi = 0.05$ of the first and last mode. Mass and stiffness proportional damping matrices ensures that the system is invertible and has unique solutions. The number of unique physical system modes L are always the same as the number of floor for the planar shear frame.

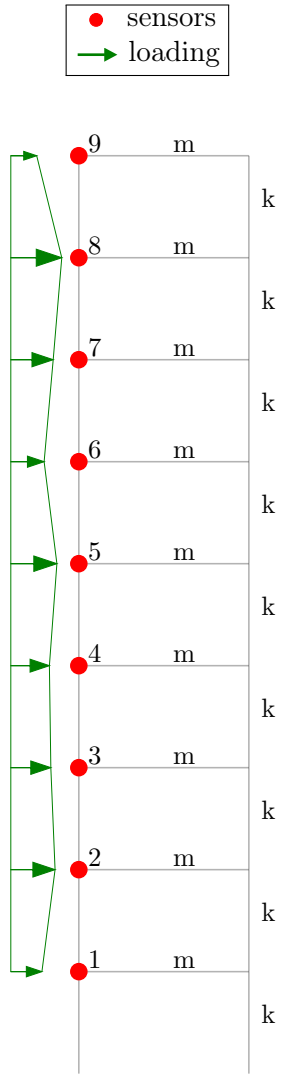


Figure 3.1: Visualization of shear frame.

3.1.1 OMA parameters

When performing OMA there are many considerations before making the measurements and parameters to determine when analysing the recorded measurements. The first consideration is choosing the right setup for the structure under consideration. Sensor layout must be tight enough to detect the system modes under investigation, and avoid zero points. With the analytical shear frame, sensor layout is given. The fundamental assumption with the planar shear frame is that each floor is moving only horizontally. This results in totally 9 modes that are unique and one must therefore have 9 sensors to capture all modes, just as shown in Figure 3.1. Sampling period, T, must be long enough to capture all information of the modes under investigation. What sampling period that is sufficient is not easy to quantify. Common practice is measurements in the range 10 to 30 minutes. $T = 600s$ has been chosen in the analyses of the shear frame. Sampling frequency, f_s must be sufficient to detect all modes of interest, but low enough to avoid impact of high frequency noise. The sampling frequency, f_s, is a parameter in the first study and will be further discussed in Section 3.2.1.

When analysing the data the first step is retrieving the modes. In this work it has been chosen to use covariance driven subspace identification, COV-SSI, as described in the Section 2.5. The number of block rows, n_{br} and the wanted model orders, N, are the user specified parameters of COV-SSI. The number of block rows n_{br} is determining the number of time shifts in the covariance matrix. A big number of n_{br} is needed to capture the modes with highest periods, i.e. lowest frequency. The number of block rows, n_{br}, is chosen according to the estimate given in Yang et al. (2019):

$$n_{br} >= \frac{f_s}{f_{min.sf}} \tag{3.1}$$

The number of block rows are for simplicity kept constant at the highest number suggested by Yang for all analyses. Given that the highest sampling frequency in this work is five times the highest eigenfrequency of the shear frame, $f_s = 5 \times f_{max.sf}$:

$$n_{br} = \frac{5 \times f_{max.sf}}{f_{min.sf}} = \frac{5 \times 31.40}{2.63} \approx 60 \tag{3.2}$$

The number of modes at a given model order is by definition the same as the model order, N. At first sight a model order of twice the number of modes is sufficient, as half of the modes are complex conjugates. Empirically this has proven to be insufficient and the most common approach is heavily overestimation and then application of filtering afterwards. Depending on whether the filtering is manual, through threshold values, or automatic through algorithms, the desired max model order will be different. Anyhow, trial and error of the maximum model order is the most common way of approach. Neu et al. (2017) suggests an approximate relationship between the number of possibly physical modes N_{pp}, certainly spurious

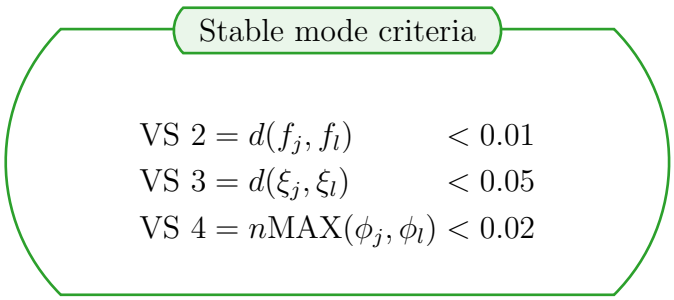


Figure 3.2: Criteria determining if a mode is stable.

modes N_{cs} and unique physical system modes L. Given:

$$N_{pp} \propto L n_{max} \tag{3.3}$$

$$N_{cs} = \frac{1}{2} n_{max}(n_{max} - 1) - N_{pp} \tag{3.4}$$

The number of unique physical system modes L is then 9. Given that most clustering algorithms prefer near equal size of the clusters, a ratio of $\frac{N_{cs}}{N_{pp}} = 1$ is a good start. Applying this assumption gives the maximum model order N_{max} as:

$$n_{max} \approx 4L - 1 \tag{3.5}$$

$$= 4 \times 9 - 1 = 35 \tag{3.6}$$

In this work the minimum model order is set to be $N_{min} = 20$ for all analyses of the shear frame. The maximum model order is further discussed in Section 3.2.1. The incremental step from the minimum model order, N_{min}, to the maximum model order, N_{max}, is kept constant at $\beta = 1$.

3.1.2 Stability of modes

Before going into the details some vocabulary must be clarified. A mode can be used as a reference to every mode, i.e. dot, in the stabilization diagram. It is also used as a reference to the clusters that yet have not been categorized as stable or not. By this it is implicitly assumed that the algorithms only have good quality clusters. A system mode, a true mode, or combined to a true system mode, is all referring to the real life modes of the structure under investigation.

When determining if a system mode is identified or not, the accuracy limit is a question to debate. Another source of debate is to which extent one picked representative from the cluster can represent the cluster as one. For simplicity the picked representative of each cluster is used in this thesis. Noteworthy is that Reynders' and Yang' uses the mode with median damping and frequency respectively as the physical representative of the cluster. One should however use all modes when looking for the true properties, as the mean of many approximations will be better than one single observation. The accuracy limits determining if a cluster is stable or not is in this plot chosen to be the limits defined by Rainieri et al. (2014):

$$\text{VS } 2 = d(f_j, f_l) \qquad < 0.01 \tag{3.7}$$

$$\text{VS } 3 = d(\xi_j, \xi_l) \qquad < 0.05 \tag{3.8}$$

$$\text{VS } 4 = n\text{MAX}(\phi_j, \phi_l) < 0.02 \tag{3.9}$$

These criteria could arguably be to strict, but trends of the algorithm's performance should still be possible to detect. Another argument to use strict criteria is that to loose limits increases the risk of mixing closely spaced modes.

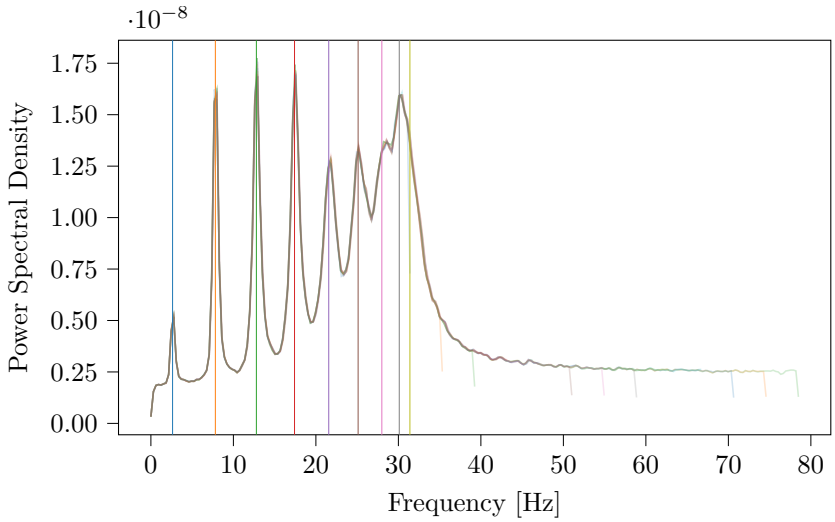


Figure 3.3: Power spectral density and true eigenfrequencies

3.2 Study 1 | Sensitivity to sampling frequency and model order

A comparison of Reynders and Yang will in this study be performed by variation of the user specified parameters; sampling frequency, f_s, and the highest model order, N_{max}. Setting these parameters is a usually a process of trial and error combined with experience. It is therefore valuable to have an algorithm that is less dependent on these parameters. The aim is to identify trends in the performance of the algorithms.

3.2.1 Setup

The Nyquist frequency is by definition half of the sampling frequency, and it is the theoretical upper limit of what frequency content that can be captured without distortion. Frequency content above half the Nyqiust frequency may be indistinguishable from each other and will appear to be folded around the Nyquist frequency which also is known as the folding frequency. The sampling rate should therefore be twice the highest frequency content looked for in the signal. This is also known as the Nyquist rate.

The shear frame is analysed for step wise changes in in sampling frequency starting at the Nyquist rate, $f_s = f_{Ny.rate} = 2 \times f_{max.sf}$, and step wise increments of factor 0.25 up to five times the maximum eigenfrequency of the shear frame $f_s = 5 \times f_{max.sf}$. In Figure 3.3 one can see the Power Spectral Density (PSD) of all the 13 different OMA signals. The loading signal was created with a sampling frequency 10 times the highest eigenfrequency of the shear frame. This way, all runs are down sampled copies of the same loading. The original signal is the response of white noise loading, with the addition of white noise characteristic measurement noise of intensity equal one standard deviation of the exact response. Also shown is the 9 eigenfrequencies of the shear frame. From the peaks one can see that the modes are, as expected, unequally excited and that the last three modes are quite close and hard to identify, thus creating a challenge for the algorithms.

The minimum model order, $N_{min} = 5$, and the step, $\beta = 1$, was described in Section 3.1.1. The highest model order is then the only property of variation. The time consumption of the algorithms is increasing exponentially with the highest model order, N_{max}. It therefore very beneficial to limit the number of modes. This study uses max model order from 10, and increasing by 10 up to 140. In total this gives $13 \times 13 = 169$ analyses.

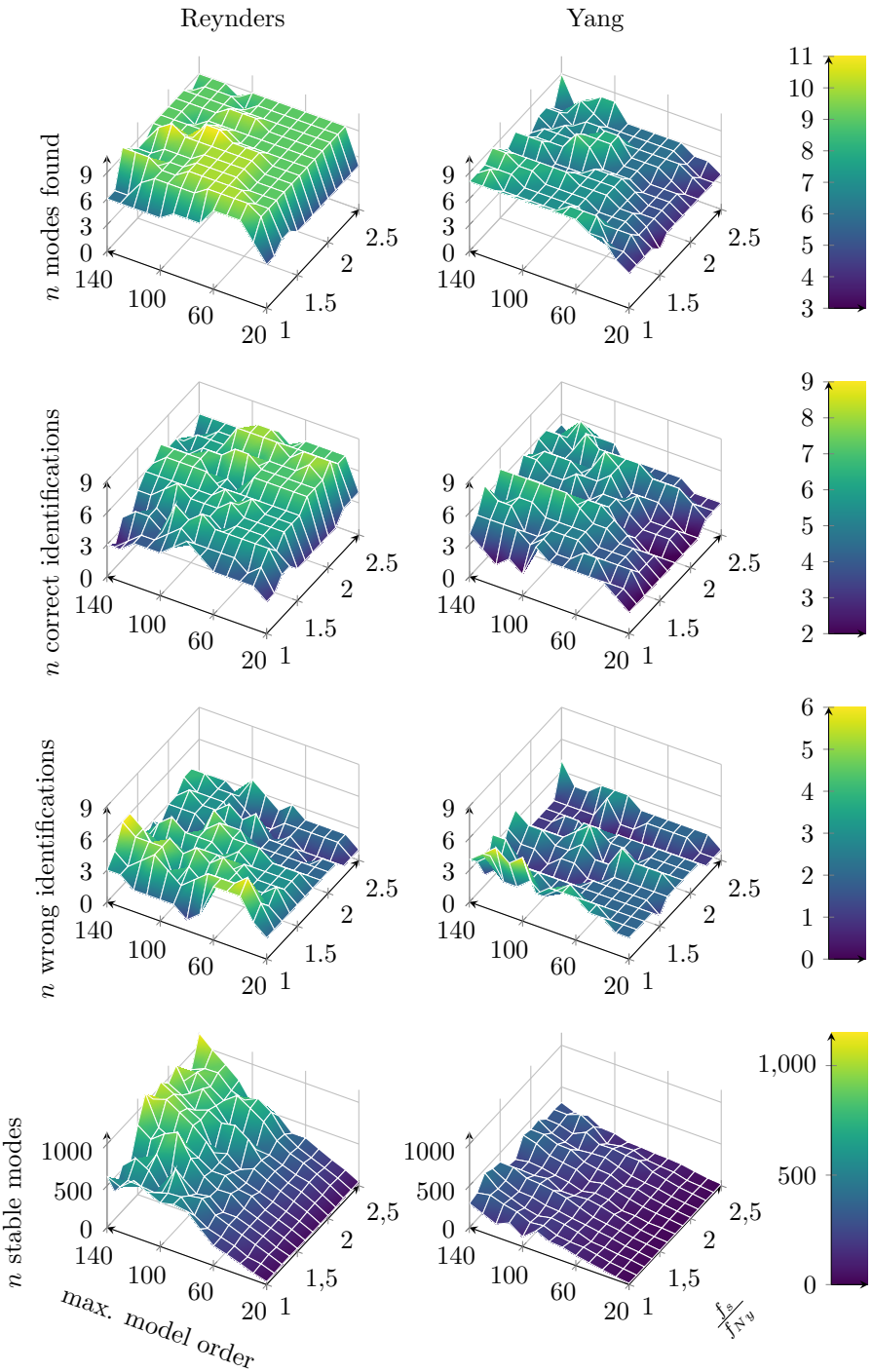


Figure 3.4: Comparison of Yang's and Reynders'.

3.2.2 Results

The results from the numerical study on the shear frame is visualized as surfaces in the three dimensional plots in Figure 3.4. The two horizontal axes are the maximum model order, N_{max} and the sampling frequency, f_s as a factor of the Nyquist rate. The vertical span of the surfaces is shown in their respective colorbar. Ex. The number of modes found is ranging from 3 to 11.

The first row of surfaces shows that Reynders' consistently finds more modes than Yang, which at first sight is good. By a closer look at the colorbar it is clear that Reynders' finds more modes than the system has in a great portion of the runs. The main reasons why extra modes are found can be summarized into four cases:

Table 3.1: Causes of clusters not getting approved

Case 1	modes are multiply defined
Case 2	spurious modes by mistake clustered
Case 3	cluster found is of to low quality to be stable
(Case 4)	cluster representative is not representative for the cluster

Case 4 is in parenthesis, as it was assumed that the mode with median damping and frequency is good enough to represent the cluster.

The second row of surfaces shows the number of correct identifications, that is how many true system modes that can be matched with the identified modes within the stability criteria defined in Equations (3.7) to (3.9). Two trends are visible; Reynders' performs best on moderate to low model orders, and with high sampling frequency. Yang's performs best on high model orders, and sampling frequency is of less importance as long as it is high enough.

The third row of surfaces is really just the first subtracted the second row. Is it, however included to increase readability. Looking at the surfaces it is clear that Reynders consistently has more wrong identifications, except when the maximum model order is very low and the sampling frequency is high.

The Fourth and last row of surfaces shows how the total number of stable modes at the different runs. Total number of stable modes is the sum of the modes that are stable. Worth noticing from these surfaces are that the number of stable modes in the clusters of Yang's is approximately half of Reynders'. It is given in the setup of the test that the input of the two algorithms are the exact same modes. Therefore one can conclude that Yang is more restrictive. Worth noting in Reynders' is that the number of stable modes keep increasing even though the number of identified modes is decreasing at high maximum model order and sampling frequency.

In this analysis there were none multiply defined system modes, making all the modes appearing in the plot of wrong identifications either spurious modes or low quality clusters. To determine if the modes clustered are spurious or low quality approximations is hard to quantify. The content of the clusters is therefore analysed.

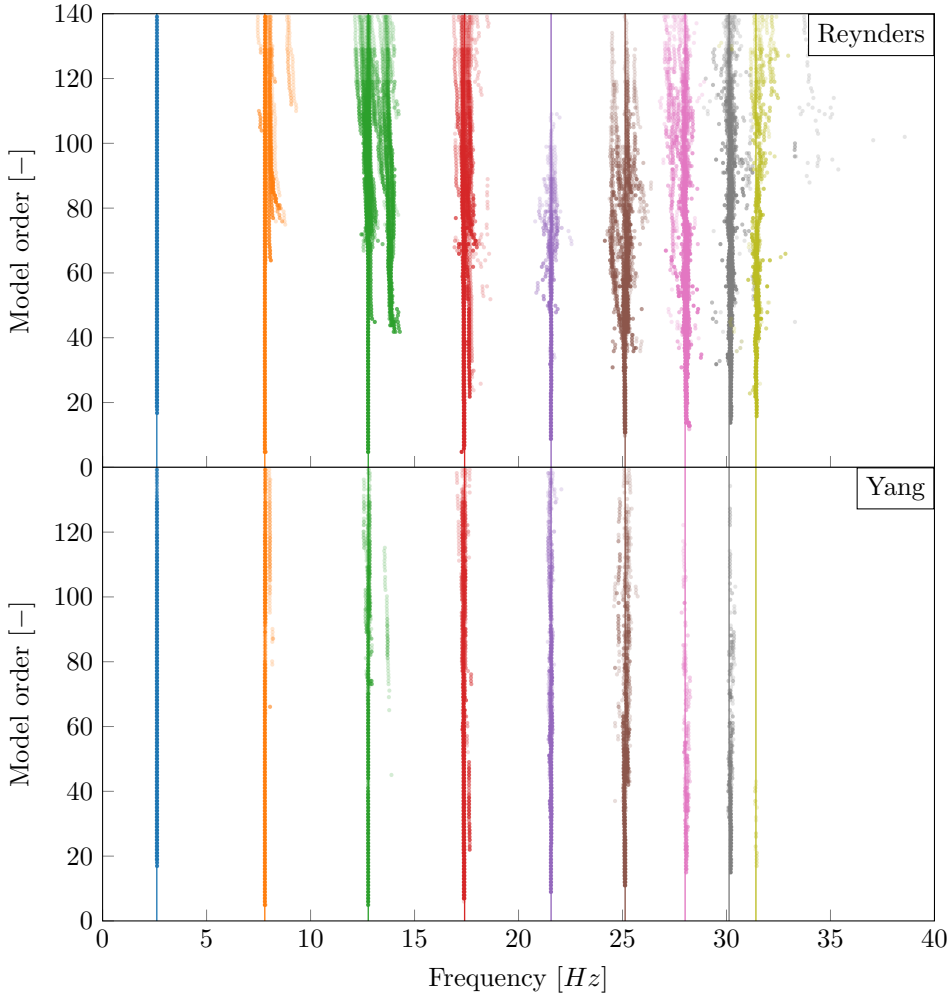


Figure 3.5: Stabilization diagrams with the modes of all stable clusters.

Table 3.3: Number of modes in stable clusters

Mode	1	2	3	4	5	6	7	8	9
reynders	7944	14026	11796	12598	2030	7076	9168	8432	3849
yang	4492	7407	2076	6870	2303	3218	608	1147	29

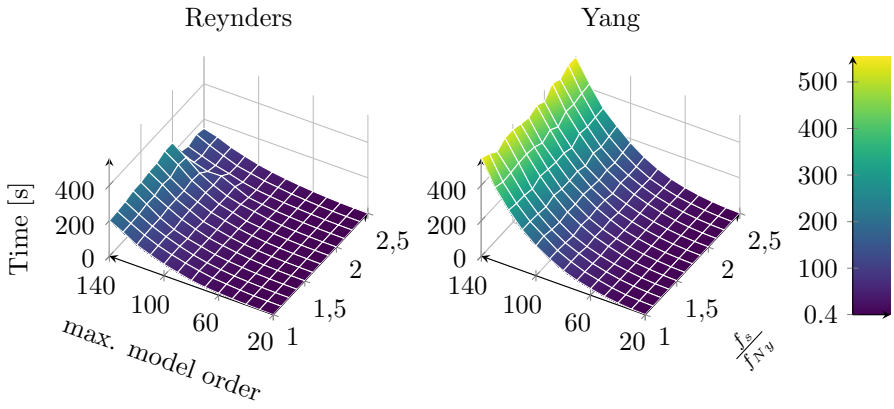
Table 3.2: Percentage modes approved

Mode	ok f		ok ξ		ok ϕ		ok all	
	Reyn.	Yang	Reyn.	Yang	Reyn.	Yang	Reyn.	Yang
1	100%	100%	100%	100%	100%	100%	100%	100%
2	90%	98%	88%	97%	90%	98%	87%	97%
3	61%	94%	27%	51%	66%	97%	24%	50%
4	87%	97%	77%	89%	83%	97%	76%	89%
5	91%	100%	60%	65%	85%	100%	59%	65%
6	82%	96%	64%	80%	80%	97%	63%	80%
7	90%	100%	65%	83%	76%	100%	60%	83%
8	96%	100%	75%	89%	86%	100%	71%	89%
9	93%	100%	55%	55%	68%	100%	45%	55%
Mean	88%	98%	68%	79%	82%	99%	65%	79%
Missed	73%	99%	26%	21%	65%	98%	25%	21%

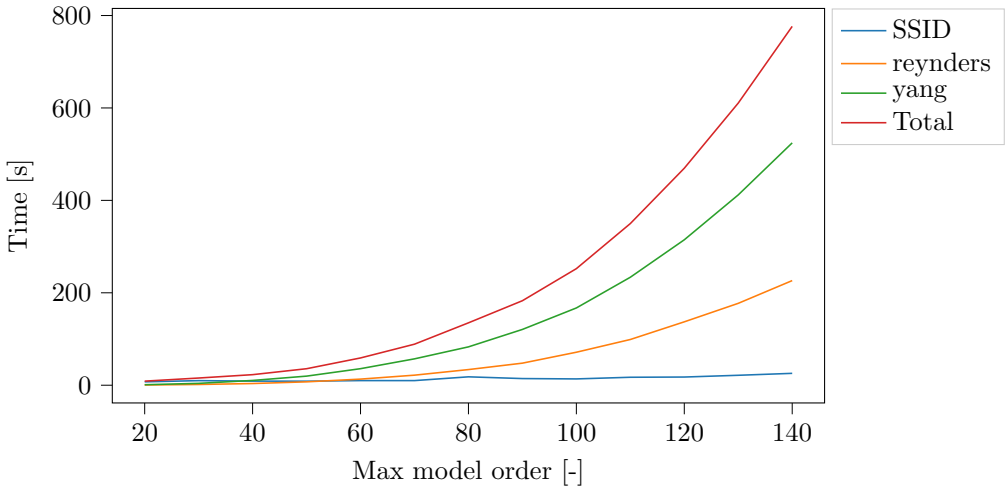
Table 3.2 shows the percentage of the modes inside the clusters that fulfill the criteria in Equations (3.7) to (3.9) determining if a mode is stable or not. Each mode is compared with the true properties of the shear frame. Row 1 to 9 is referring to all the approved clusters representing system mode 1 to 9. Row 10 is the Mean of all system modes. Last row is checking how many of the modes in the non stable clusters that would have been approved they could be merged with a true mode.

There are two main observations to pull out of Table 3.2. The first is that Yang consistently has a higher percentage of modes that are stable, visually observed as whiter columns. Comparing the three properties, frequency, damping and mode shape, it is clear that the damping criteria is the most strict criteria. Alternatively one can argue that damping is not a very well suited criteria to determine if a mode is stable because damping is not very well estimated. The damping scatter is large as earlier visualized, ex. Figure 2.4, and observing that 100% of the modes clustered by Yang fulfils the frequency and mode shape criteria, damping criteria in Equation (3.8) should be revised or replaced.

The stabilization diagrams in Figure 3.5 shows that the spread in the clusters is greater for Reynders than for Yang. Especially mode 3, green, is having a lot of mathematical modes in it's clusters, and Reynders has the greatest spread.



(a) Surface showing run times.



(b) Plot showing run times at $f_s = 1.5 \times f_{Nyquist}$.

Figure 3.6: Run time of algorithms.

The run time of the different algorithms in this study is depicted in Figure 3.6. The surfaces in Figure 3.6a shows the run time of Reynders and Yang at every run in this study. The plot in Figure 3.6b is obtained by cutting the surfaces at $f_s = 1.5 \times f_{Nyquist}$, as well as adding the time used for the system identification with SSI-COV. The total time is for both algorithms and the system identification combined.

The surfaces clearly shows that increasing the number of modes is the main factor increasing the run time of the algorithms. The different sub steps of the algorithms have different time complexities, but they are all exponential with the number of modes. The cut frequency, f_s, to create the plot should therefore yield the same plots. An exception is, however, noticeable at high model orders or Reynders'. The reason why run time is varying with the sampling frequency is unidentified. One possible reasons is a major change in number of modes removed at the first step of the algorithm, leading to less modes in the hierarchical clustering of step two.

Worth noticing about the run times is the fact that none of the algorithms are optimized and the run times is only from one computer with other processes running at the time of the analysis. Reynders' algorithm is also more optimized as all clustering steps is performed using external packages from libraries like *scipy*, that are optimized. Another difference is that Reynders is using K-means, where Yang is using C-means, in the the first step of the algorithms. The hierarchical clustering step of Yang could most likely be optimized, considering that no effort has been made in the implementation in this thesis. Small computations like the distance calculation to all clusters, instead of only the near ones, add up when the number of modes and clusters is high. This should therefore be considered if implementation is revised.

The identification of the modes through SSI-COV is insignificant compared to the run time of the two algorithms in this case. In other cases with higher number of block rows, and longer time series SSI-COV will be much longer. From the graphs one can conclude that Reynders is approximately half the run time of Yang. The difference in absolute time is therefore not very significant before high number of modes is chosen, here by setting a high max model.

The importance of the run time is naturally a question of application as further discussed after the next study, Section 3.3. From this section one can conclude that the difference between Reynders' and Yang's algorithm is significant, especially with high number of modes.

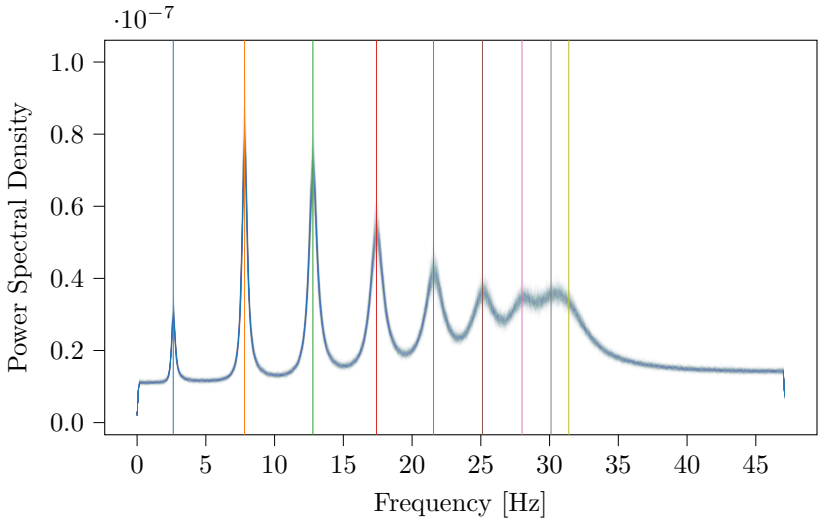


Figure 3.7: Power spectral density of all 200 runs.

3.3 Study 2 | Repeated runs on different input

In this study a comparison of Reynders' and Yang's will be performed by repeated runs on different input signals, but with the same user specified parameters. The aim is comparing the performance in terms of the most important properties; accuracy and detection rate.

3.3.1 Setup

Specifying the best parameters for both Reynders' and Yang's is not straightforward. This study bases it's chosen parameters on the results from the first study, Section 3.2, and the currently achieved user experience.

The user specified parameters are chosen to be similar to the first study, where applicable. This permits comparison and validation of the findings in the first study. Sampling period is still $T = 600s$, i.e. 10 minutes. The number of block rows is $n_{br} = 60$. From the results of the first study, shown as surfaces in Figure 3.4, a sampling frequency of $f_s = 1.5 \times f_{Nyquist}$, and a max model order $N_{max} = 60$, is appropriate for both algorithms.

The loading is in this study the new element of every run. It does however yield the same properties as the loading used in the first study. That is, response measurements of white noise loading loaded with white noise, at the magnitude of one standard deviation of the response measurements. All 200 runs has a different random loading, and the trace of the PSD-matrix of every load is plotted in Figure 3.7 together with the true eigenfrequencies. The last three modes are less excited than the other modes, and closely spaced in frequency, thus making a challenge for the algorithms to detect.

A general loading has two main properties determining if a mode is excited or not. That is, frequency content and it's spatial distribution. To excite a mode, frequency content in the loading near the eigenfrequency of the mode must be present in multiple DOFs. If there are DOFs without this frequency content, or at worst in counter phase, these DOFs are reducing the excitation. Physically this can be imagined as all loading positions pushing the frame at a synchronized frequency, but pushing at different time instances depending on the modeshape.

Given that a single loading may excite only some modes the performance of the algorithms can't be determined on just one run. This is the reason why this study runs 200 analyses.

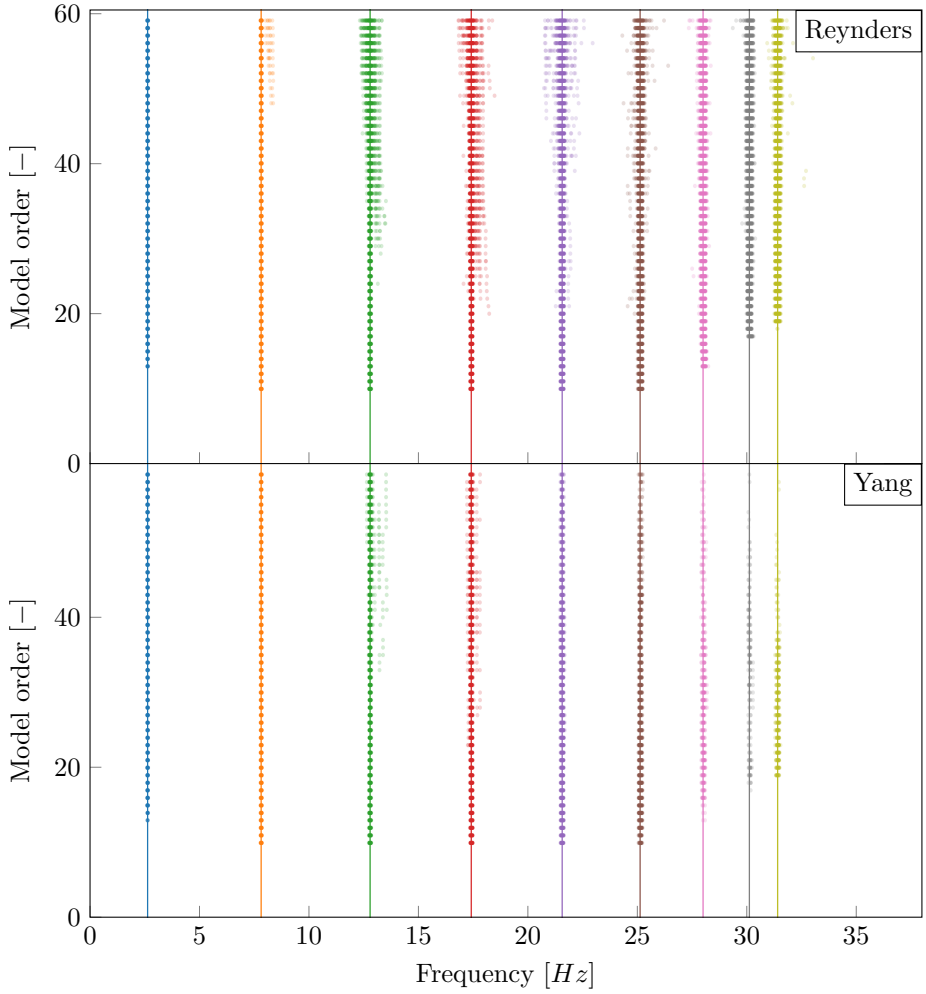


Figure 3.8: Stabilization diagram with all modes in the stable clusters.

Table 3.4: Number of modes in stable clusters

Mode	1	2	3	4	5	6	7	8	9
Reynders	4500	5897	5986	6371	6269	6518	5055	3493	4044
Yang	2555	3363	3015	2562	2105	1467	477	164	347

3.3.2 Results

The results of this study is presented using some of the same Figures and Tables as seen in the first study, Section 3.2, to compare and validate the first study. Explanation of the details in these Figures and Tables are therefore left out.

Looking at Figure 3.8, and the associated Table 3.5, the same trends can be identified. In both studies one can see that Reynders consistently has more modes in the stable clusters than Yang has. Without looking at the numbers, this is observed as a darker second row of Tables 3.2 and 3.5. Alternatively visually observing the number of points, which is the same as the tables, in the scatter in Figures 3.5 and 3.8.

Table 3.5: Percentage modes approved

Mode	ok f		ok ξ		ok ϕ		ok all	
	Reyn.	Yang	Reyn.	Yang	Reyn.	Yang	Reyn.	Yang
1	100%	100%	99%	98%	100%	100%	99%	98%
2	99%	100%	95%	99%	99%	100%	95%	99%
3	91%	98%	71%	83%	91%	98%	70%	83%
4	93%	98%	70%	84%	91%	98%	70%	84%
5	97%	100%	70%	87%	94%	100%	70%	87%
6	99%	100%	74%	91%	98%	100%	74%	91%
7	100%	100%	73%	91%	98%	100%	72%	91%
8	100%	100%	77%	89%	91%	99%	71%	89%
9	100%	100%	80%	94%	92%	100%	76%	94%
Mean	98%	100%	79%	91%	95%	99%	77%	91%
Misssed	91%	100%	10%	10%	91%	100%	10%	10%

By inspection of the modes inside the clusters one obtain Tables 3.2 and 3.5. The same trends are observed in this study; damping is the most strict criteria and Yang has better content in it's clusters. Worth noticing in this study is that the trends are more clear; the missed modes, i.e. unstable clusters, have lower percentage of approved damping. Also the difference between Reynders is greater.

Table 3.6: Frequency and Damping statistics

(a) Frequency statistics

Mode [#]	True f [Hz]	$\bar{df}$ [Hz]		$\sigma(f)$ [Hz]	
		Reynders	Yang	Reynders	Yang
1	2.63	0.000	0.000	0.011	0.010
2	7.81	0.004	0.001	0.039	0.013
3	12.79	0.009	0.010	0.090	0.074
4	17.41	0.015	0.002	0.104	0.054
5	21.56	0.004	0.001	0.103	0.038
6	25.12	0.006	0.009	0.081	0.043
7	27.99	0.002	0.011	0.071	0.051
8	30.11	0.006	0.013	0.075	0.050
9	31.40	0.011	0.006	0.090	0.049
Mean all modes		0.006	0.006	0.074	0.043

(b) Damping statistics

Mode [#]	True xi [%]	$\bar{d\xi}$ [%]		$\sigma(\xi)$ [%]	
		Reynders	Yang	Reynders	Yang
1	5.000	0.006	0.002	0.146	0.147
2	2.700	0.017	0.005	0.323	0.089
3	2.827	0.141	0.042	0.600	0.241
4	3.255	0.183	0.056	0.648	0.320
5	3.731	0.154	0.015	0.578	0.141
6	4.174	0.103	0.037	0.397	0.131
7	4.547	0.081	0.061	0.238	0.127
8	4.827	0.106	0.119	0.213	0.145
9	5.000	0.080	0.079	0.280	0.115
Mean all modes		0.097	0.046	0.380	0.162

The average frequency deviation, $\bar{df}$, and the frequency standard deviation, $\sigma(f)$, is shown for all system modes in Table 3.6a. The modes making the statistical properties of every system mode is the modes inside the stable clusters, and total number of stable modes can be read in Table 3.4. The average frequency deviation, $\bar{df}$, between Reynders and Yang is by meaning all modes equal at the magnitude of 0.006 Hz. Looking at the numbers for every mode Yang has more of the best and worst estimations. This may be due to the lower number of stable modes, making the influence of every element greater. The frequency standard deviation is consistently better for yang, and on average 58% of Reynders.

Looking at Table 3.6b the damping estimates is consistently better for Yang than Reynders. The damping ratio, ξ, is represented as a percentage, and the deviations are in percentage points. On average of all modes the mean damping deviation, $\bar{d\xi}$, of Yang's' is 47% of Reynders', and the damping standard deviation, $\sigma(\xi)$ is 43% of Reynders'.

Table 3.7: Detections

Mode	Right		Multi		Missed	
	Reynders	Yang	Reynders	Yang	Reynders	Yang
1	50%	50%	0%	0%	50%	50%
2	58%	59%	0%	0%	42%	41%
3	52%	56%	0%	0%	48%	44%
4	57%	57%	0%	0%	42%	43%
5	60%	58%	0%	0%	40%	42%
6	66%	50%	0%	0%	34%	50%
7	58%	20%	0%	0%	42%	80%
8	44%	8%	0%	0%	56%	92%
9	46%	16%	0%	0%	54%	84%
Mean	55%	41%	0%	0%	45%	59%

Table 3.7 shows that Reynders consistently detects more modes than Yang. The detection percentage is especially bad of the last three modes for Yang.

All the results in this study is in accordance with the trends found in the first study. Yang is more selective, yield higher quality clusters, at the cost of less detections.

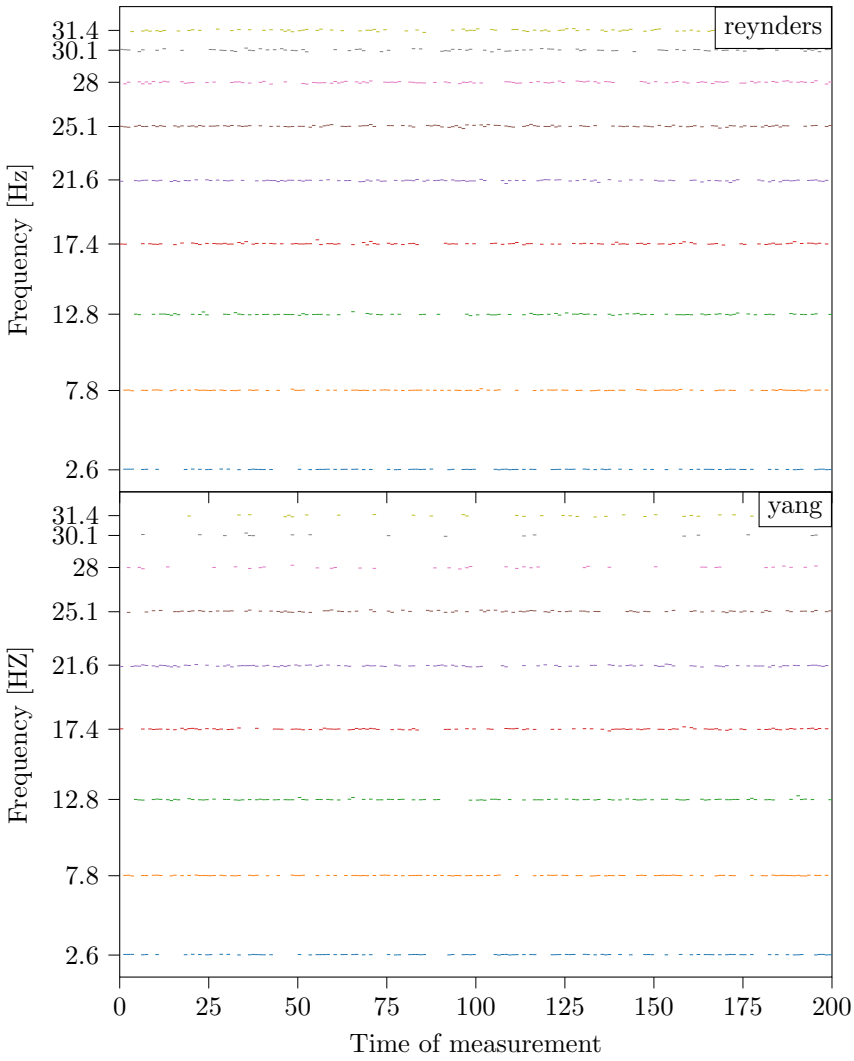


Figure 3.9: Identified frequencies at every run.

3.4 Discussion of numerical study

All the results in the second study is in accordance with the trends found in the first study. Yang is more selective, yield higher quality clusters, at the cost of lower detection rate. Reynders on the other hand is more optimistic, yielding wider clusters, at the cost of more unstable modes in the clusters. Which algorithm is then the preferred one? This is as, we will discuss, a question of application chosen and wanted result.

Automatic algorithms are specially beneficial in an Structural Health Monitoring (SHM) context, as briefly mentioned in the introduction. Given that acceleration measurements are performed automatically at a given interval, or by triggered events, there is a great cost benefit in having an automated algorithm analysing every measurement. Most likely automated algorithms is necessary to make SHM applicable in large scale. With an automated approach the expert user is only involved when there are changes; like the finding of a new mode, or modes are missing.

Typical events used to trigger an acquisition can be an abnormal acceleration, which is possible to detect with passive sensors, high winds, or a nearby earthquake. In the case of such an event one can argue that it is more important that the mode is found every time analysing than finding it very precisely. Given that a mode is found every time, it is also possible to quantify how bad a measurement is compared to the normal situation. On the other hand, if a mode is only found when measurements are of high enough quality, there will be time periods without inspections. The consequence of periods without measurements can be fatal if critical damage is not exposed.

By imagining that each of the 200 runs in the second study, Section 3.3, represent one measurement of a SHM measurement scheme one can represent the tracking as shown in Figure 3.9. The x-axis is then representing the time instant of the measurement, e.x. every day at 16:00 o clock, unless special trigger events occur. The y-axis shows the identified frequencies. As shown will Reynders algorithm give more detections, especially for the three highest frequency modes. Assuming that the measurements are daily, the maximum number of days without measurements is 7 and 41 for Reynders and Yang respectively, i.e a significant difference.

Another use of automated algorithms is simplifying the work done in a manual analysis. Say that an engineer wants to identify the modes of a new bridge to validate an analysis. In this case the benefit of using an automated algorithm is that one can get an estimate within minutes or seconds, depending on the length of the time series. By using Reynders' algorithm one will most likely identify more modes without to many trial end error runs determining a sufficient model order etc. Using Yang's algorithm one will most likely be fiddling with more trial and error, if modes are detected at all.

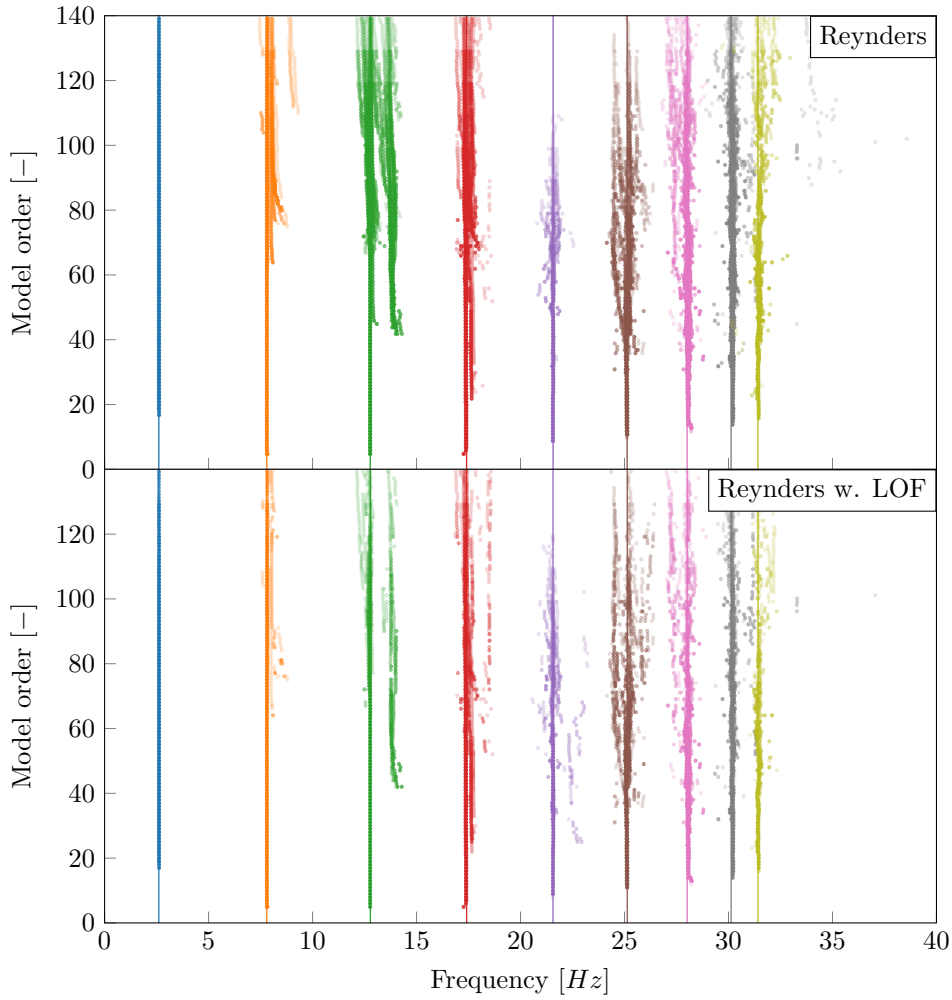


Figure 3.10: Stabiliation diagram equivalent to Figure 3.5 after filtering with LOF.

Table 3.8: Number of modes in stable clusters

Mode	1	2	3	4	5	6	7	8	9
Yang	4492	7407	2076	6870	2303	3218	608	1147	29
Reynders	7944	14026	11796	12598	2030	7076	9168	8432	3849
Reyn. w.LOF	4502	8206	3736	8273	2404	4885	4838	4463	1996
Diff.	3442	5820	8060	4325	−377	2190	4330	3969	1853

3.4.1 Affect of filtering with LOF

Comparing the two algorithms it was clear that Yang had less, and higher quality modes in its clusters. An additional mini-study has therefore been performed to check if this is due to more precise clustering or just the removal of modes with LOF. Because all runs were saved, and the implementation used in this thesis is object oriented, LOF-filtering was possible to add afterwards on the results from the first study, Section 3.2. This way, everything is the same and comparison valid.

The resulting stabilization diagram shows both improvement and deterioration in terms of visual spread, Figure 3.10. The first three modes appear to be improved, where the next three modes are deteriorated. The total number of modes in stable clusters is greatly reduced for all but mode 5 that has a slight increase. This increase is due to more clusters getting stable according to Equations (3.7) to (3.9). Especially mode three, green, is greatly reduces. Looking at the stabilization diagram it is clear that many wrong detections of mode three is removed.

Table 3.9: Percentage modes approved after filtering with LOF.

Mode	ok f		ok ξ		ok ϕ		ok all	
	Reyn.	Yang	Reyn.	Yang	Reyn.	Yang	Reyn.	Yang
1	100%	100%	100%	100%	100%	100%	100%	100%
2	97%	98%	95%	97%	97%	98%	95%	97%
3	74%	94%	43%	51%	75%	97%	40%	50%
4	89%	97%	83%	89%	86%	97%	82%	89%
5	89%	100%	64%	65%	79%	100%	63%	65%
6	80%	96%	67%	80%	77%	97%	67%	80%
7	90%	100%	65%	83%	76%	100%	60%	83%
8	97%	100%	78%	89%	86%	100%	74%	89%
9	95%	100%	61%	55%	73%	100%	52%	55%
Mean	90%	98%	73%	79%	83%	99%	70%	79%
Missed	65%	99%	21%	21%	59%	98%	20%	21%

Inspecting the content inside the clusters it is clear that a higher percentage of the modes are approved. The mean of all modes getting approved by all criteria raises from 65% in Table 3.2 to 70% in Table 3.9. The increase is smaller than expected given the drastic reduction in the number of modes in the stable cluster. It can be concluded that the outlier detection is not enough to improve the quality of Reynders' clusters up to the level of Yang's alone. The clustering process of Yang is therefore the step that makes a notable difference.

Chapter 4

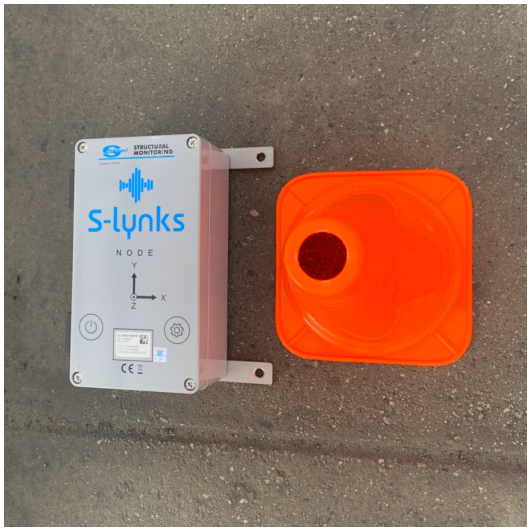
Measurements on Elgeseter bridge

This study is testing the same algorithms, but now on real life data. The data is self-sampled by portable sensors on Elgeseter bridge near the Norwegian University of Science and Technology. The aim with this study is to continue the discussion on advantages and disadvantages discovered in the numerical study by utilizing the algorithm on self recorded data.

4.1 About Elgeseter bridge

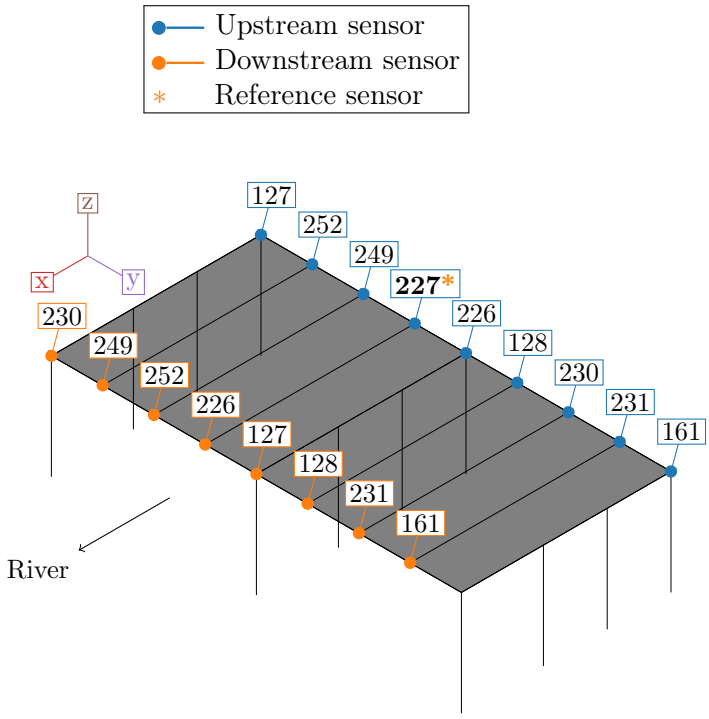
Elgeseter bridge is a monolithic concrete beam bridge resting on groups of slender columns dividing it into nine spans. The bridge did originally have six driving lanes, but only four as of today to make space for pedestrians and bikers, making the total width 23.4 m. The span length is equal at 22.5 m, except for the first and last span that are 21.5 m, giving a total length just above 200 m. Columns are arranged in groups of four, where each column is about 80 cm in diameter with length up to about 16 meter mid river. The south end is fixed to the abutment, and the north end on a bearing support giving freedom to expand with temperature. All columns are resting on foundations connected to friction piles. The friction piles on land and in the river is made of concrete and wood respectively.

The bridge has shown to be growing at a more or less stable rate of 3.5 mm per year due to alkali reactions in the concrete. All columns were top cut, displaced, and reattached summer 2003, (Thorenfeldt, 2015). This to ensure that increasing crack widths, leading to reduced shear capacity and risk of corrosion, did not reduce safety. The bridge-column interaction may have changed as a result of the reattachment.



(a) Sensor next to cone.

(b) Sensors and main unit on upstream measurement.



(c) Measurement setup of span 2 and 3 on Elgeseter bridge.

Figure 4.1: Sensors and measurement setup.

4.2 Setup

Elgeseter is not a monitored bridge, and modal analysis based on response measurement data has not earlier been performed as of the writers knowledge. The measurements presented in this work have been performed by wireless sensors from *Sercel*, pictured in Figure 4.1a. The sensors measures triaxial response, that is x-, y-, and z-direction. Communication with the sensors is done through a gateway by either radio or WiFi-signals. The range on WiFi is limited, making measurements using WiFi limited to only a few spans. This is, however, not a huge problem as the number of available sensors is only 9.

The width is massive compared to the height of the bridge, creating a very large sideways stiffness. Therefore one can assume that the horizontal modes are very little excited in an OMA-setup. Sensor layout must be chosen to capture the expected modes, which is vertical, torsional, and possibly pile modes. To differentiate the vertical modes a sufficient number of sensors must be placed in each span. Torsional modes are captured with sensors on both upstream and downstream side. Pile modes are best detected by placing sensors on the piles. Sensors were only temporarily placed without drilling. The best placement was therefore on deck, close to column tops.

With only 9 sensors, and the limited WiFi range, multiple measurements must be performed and merged to get a sensor layout tight enough. Given that the expected modes should be detectable in only two spans it was chosen to measure on span 2 and 3 as shown in Figure 4.1c. There is as shown two setups consisting of the upstream sensors, marked in blue, and the downstream, marked in orange. The numbers; [127, 128, 161, 226, 227, 230, 231, 249, 252], denotes sensor numbers, and their respective location is the connected dot. Worth noting is the reference sensor, number 227, that is kept at the upstream side during both recordings. This leads to only eight sensors on the downstream side. Note also that the x-axis is perpendicular to the length axis, due to the conventions of the sensors as seen in Figure 4.1a.

In this OMA setup, merging of measurements is achieved by merging each non-simultaneous recording with the common reference sensor, i.e. 227. The reference sensor is necessary because the mode shapes can not be normalized without in an OMA context. The process of merging the data may follow the classic Post Separate Estimation Re-scaling (PoSER) approach, the more refined Post Global Estimation Re-scaling (PoGER) approach or the advanced Pre Global Estimation Re-scaling (PreGER) approach. The details of these processes are more thoroughly discussed in Reynders et al. (2010).

All in all, the difference between the three mentioned methods is when to normalize and what to normalize. The PoSER is the most intuitive approach in the sense that the MPE is performed on each measurement setup separately and normalization and merging of the partial mode shapes is done as the final step. With PoGER the block hankel matrices from each setup are stacked and MPE is performed on the stacked matrix. Normalization is then performed just as with

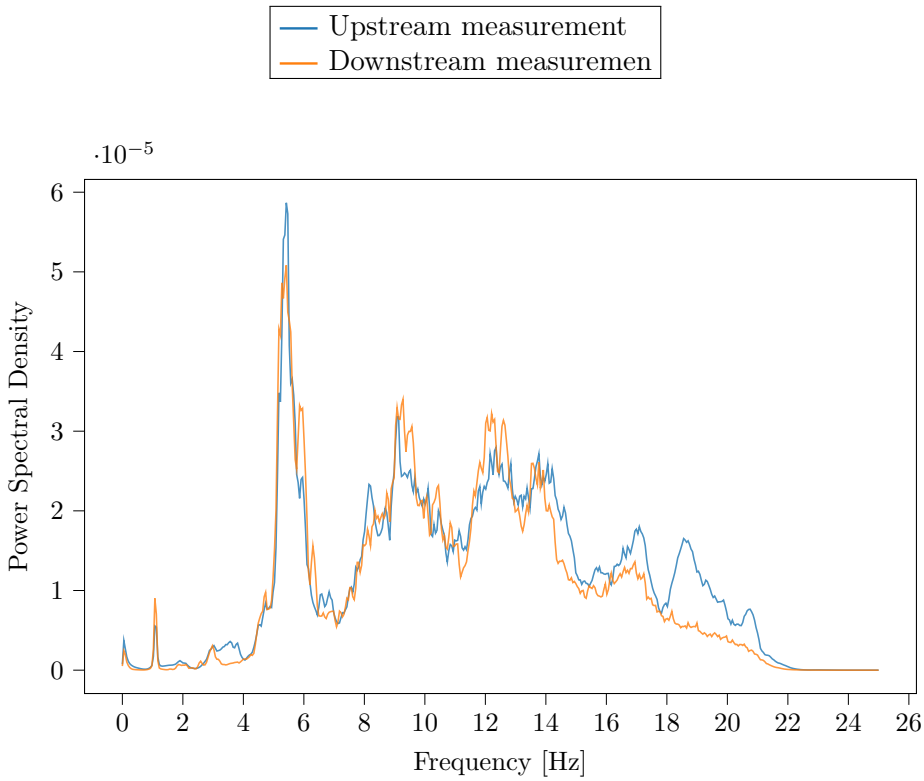


Figure 4.2: Power spectral density of down sampled signal.

PoSER on the identified modes. PreGER normalizes the hankel matrices and stacks them to one matrix. MPE is then performed on a combined system with no further need of normalization.

Chosen in this work is PoSER, as it can be used with the SSI-COV-algorithm of *strid*. This is beneficial as a manual analysis is easy to perform using the user interface of . The difference between the algorithms may also be clearer when the output of the different setups are to be merged. It is, however, a lot of bookkeeping and special care must be taken when the data is normalized and merged.

The normalization process of PoSER, and PoGER, is really just scaling and aligning the reference sensors so that the remaining DOFs in the mode shape of each setup are in phase. A mode shape can be represented in a complex plot, as shown with the shear frame in Figure 2.5. The scaling part is to normalize the length of the reference sensors. Physically this means that the excitation is scaled to be equal, as response is just a consequence of the applied loading. The alignment part is rotating the reference sensors to point in the same direction. Physically this means that the two setups are aligning in phase, i.e. reach max displacement at the same time.

One of the major problems with the PoSER-approach is how to merge modes. Most of the previously discussed criteria is not applicable anymore. The criteria comparing mode shapes can not be used as it makes no sense comparing the mode shape of the upstream and downstream side, unless the mode shape is symmetric, but that information is not known beforehand. The damping estimation is as discussed in the first study not very well estimated. The only good criteria left is then frequency, and a limit of $df < 1\%$ is chosen in accordance with previously used criteria. It should be noted that the limit is loose when not combined with other measures and closely spaced modes may accidental be mixed.

The sampling frequencies of the sensors is 250 Hz by default. This is way higher than the modes under investigation. From experience it has proven beneficial to down sample the signal to remove high frequency content that are of little interest. To much down sample is however not beneficial, as the excess modes when overestimating the model order must be place somewhere. A down sampling factor $q = 5$ was therefore chosen to give the sampling frequency of $f_s = 50$Hz. The power spectral density is plotted in Figure 4.2. The shape of the upstream and downstream measurement is in general having the same shape, meaning that the excitation is near equal. To have the same excitation is not necessary, but may improve the results as the scaling will be smaller and finding the same modes will be easier. With different excitation one may not find the same mode, making merging impossible.

To reduce the workload for the algorithms, all modes above 12Hz are removed. This will remove some system modes, but most likely none that the algorithms are able to detect well due to the noise at high frequencies. This is substantiated by the power spectral density plot shown in Figure 4.2.

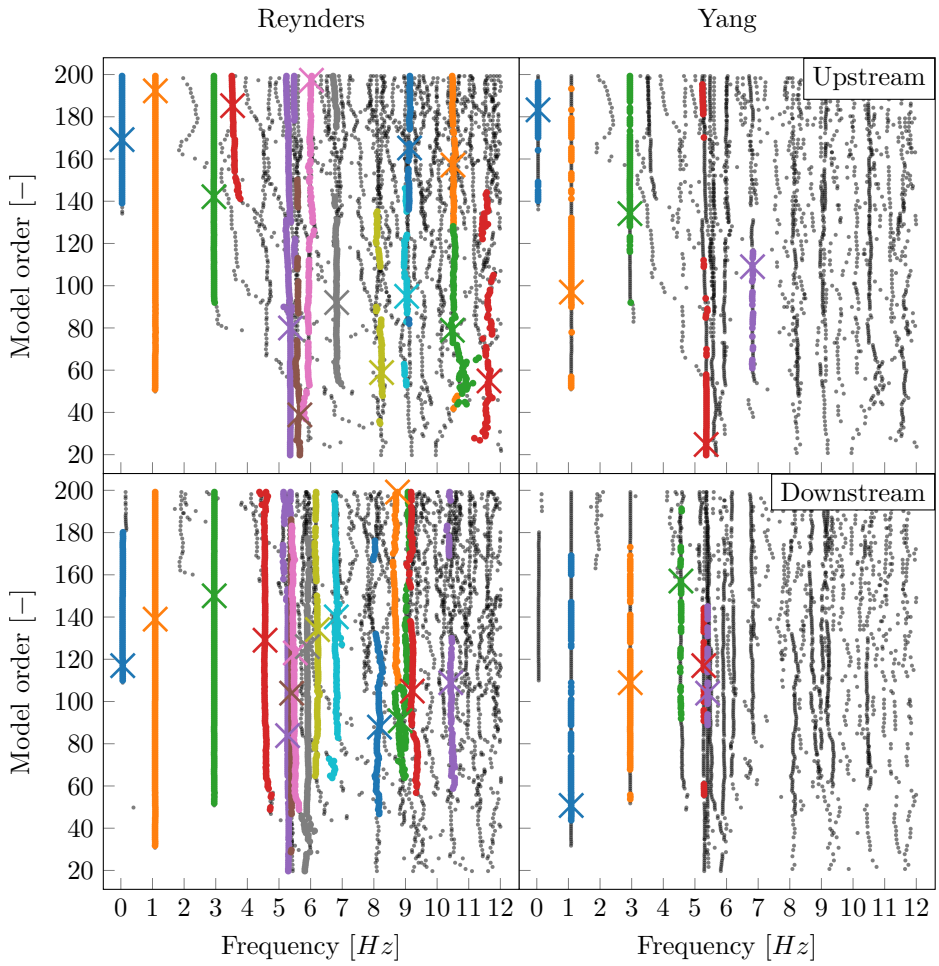


Figure 4.3: Stabilization diagram of upstream and downstream side for both Reynders' and Yang's.

4.3 Results

The resulting stabilization diagrams after automatic clearing is shown in Figure 4.3. The the rows and columns are denoting the measurement side and used algorithm respectively. One can clearly see that Reynders' identify more modes, visually by more clusters of different colors. It is also clear that the upstream and downstream side recording leads to different modes in the subspace identification process, SSI-COV. There is consequently difference in which modes that is identified modes. E.g. the mode at f 4.5 Hz is detected by both algorithms on the downstream side, and appearing as one vertical line in stabilization diagram. On the upstream side it is not detected, and appears only at some model orders. Determining the source of the difference in detected modes can not be done by one measurement. One can, however, state that minor changes to the excitation, noise or any of the user specified parameters makes a notable difference in SSI-COV.

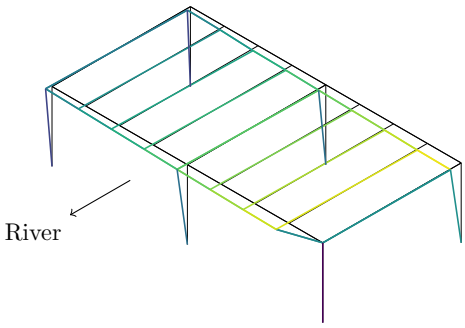
Both algorithms identifies a mode with frequency of 0.044 Hz at the upstream side. This mode is not apparent on the downstream side, but there is one close at f 0.062 Hz. These modes are discarded due to the low frequency. Finding a mode with period of $1/0.05 = 20$ s is not expected for this kind of bride.

By looking at the scatter of the modes in the frequency range between 7 and 12 hertz one can conclude that the noise levels compared to the excitation is to low to get good results. This substantiate the simplification done by removing all modes above 12 Hz. Note also that the majority of modes not clustered is at higher frequencies, just as assumed when determining the sampling frequency. A rule of thumb can therefore be to have a sampling frequency of approximately 4 times the highest frequency of interest.

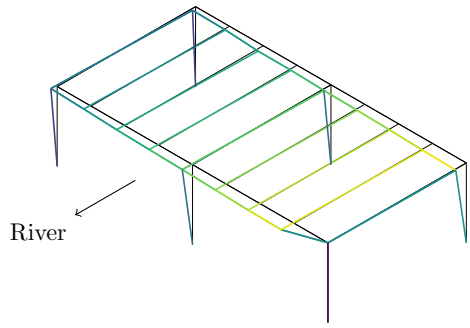
The characteristics observed in the numerical studies, Section 3.2, is still present and easily observable. The number of clusters is less for Yang than Reynders, and the clusters are with less and more precise modes. This is observed by less clusters, and missing modes in the vertically aligning clusters. E.g. The mode at 1.1 Hz.

Looking at the identified frequencies it is clear that some match, and others not.

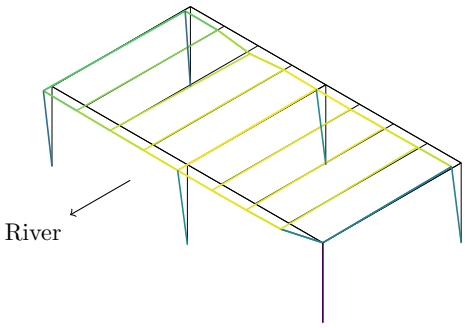
A manual analysis is also performed after looking at the stabilization diagrams shown in Figure 4.3. This gives a perception of which modes to look for. Combined with user experience, and the possibility to compare the upstream and downstream on split screen, this gives less modes detected on only one side. With manual picking all but 2 modes are merged as shown in Table 4.1.



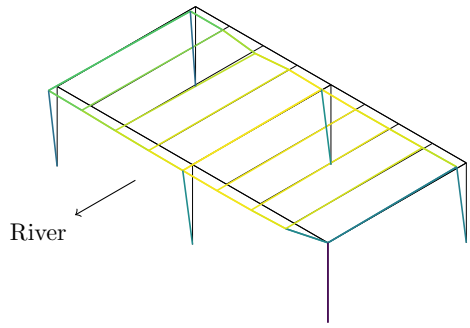
(a) Mode 1 found with Reynders'.



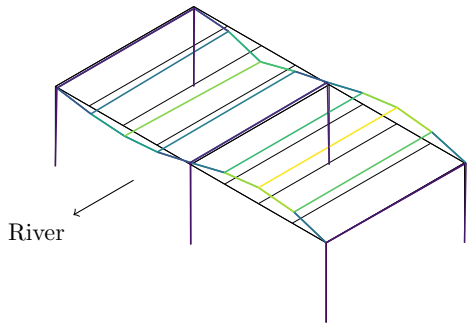
(b) Mode 1 found with Yang's.



(c) Mode 2 found with Reynders'.



(d) Mode 2 found with Yang's.



(e) Mode 3 found manually.

Figure 4.4: Mode shape plot of first and second mode found with the algorithms and mode three found with manual analysis.

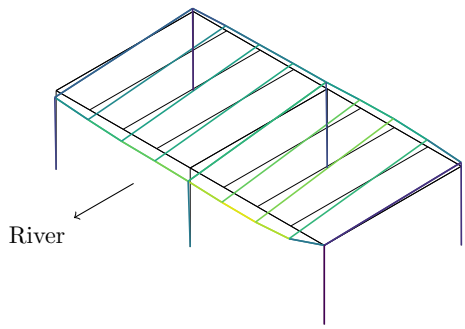
Table 4.1: Identified frequencies

	Reynders			Yang			Manual		
	US	DS	Merg.	US	DS	Merg.	US	DS	Merg.
1	0.043	-	-	0.044	-	-	1.094	1.088	1.091
2	-	0.062	-	1.094	1.087	1.090	2.942	2.955	2.949
3	1.094	1.088	1.091	2.941	2.952	2.947	4.516	4.550	4.533
4	2.939	2.953	2.946	-	4.548	-	5.293	5.304	5.299
5	3.532	-	-	-	5.272	-	5.478	5.468	5.472
6	-	4.553	-	5.349	5.391	5.369	5.562	5.566	5.563
7	-	5.266	-	6.807	-	-	5.979	6.002	5.990
8	5.356	5.391	5.373	-	-	-	6.839	6.829	6.834
9	-	5.510	-	-	-	-	8.042	7.994	8.015
10	5.652	-	-	-	-	-	8.996	8.998	8.992
11	-	5.892	-	-	-	-	9.555	-	-
12	6.016	-	-	-	-	-	-	9.657	-
13	-	6.208	-	-	-	-	-	-	-
14	6.811	6.805	6.808	-	-	-	-	-	-
15	8.230	8.161	8.195	-	-	-	-	-	-
16	-	8.732	-	-	-	-	-	-	-
17	-	8.818	-	-	-	-	-	-	-
18	9.019	-	-	-	-	-	-	-	-
19	9.112	9.199	9.155	-	-	-	-	-	-
20	10.462	10.411	10.436	-	-	-	-	-	-
21	10.484	10.411	10.447	-	-	-	-	-	-
22	11.619	-	-	-	-	-	-	-	-

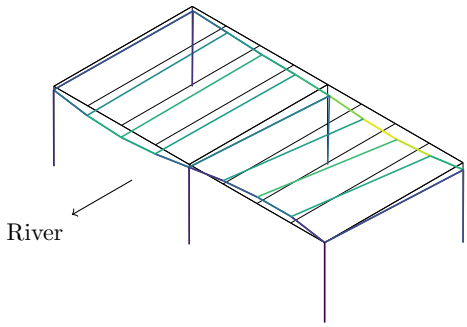
Table 4.1 confirms the differences from the stabilization diagrams. Reynders is in total finding 14, and 15, modes for the upstream and downstream side respectively. Yang identifies five for both sides. After merging, highlighted by grey rows, Reynders identifies 8 different modes, Yang identifies 3 modes, and the manual analysis identifies 10 modes. The blue cells denote cells with equal frequencies in each row. This is the case in row 20 and 21 where the same downstream mode has been merged twice. The frequency of row 20 is closest, and may therefore by used. It must however be pointed out that this mixing of closely spaced modes is the drawback of PoSER.

The manual analysis was performed by manually picking modes in the stabilization diagram on split screen. This gives the possibility to look for the best match of each diagram, and is the reason why all but two modes on the upstream and downstream side in Table 4.1 has a match.

Figure 4.4 shows the first and second mode of Elgeseter bridge. These mode shapes



(a) Mode 4 found by manual analysis.



(b) Mode 5 found by manual analysis.

Figure 4.5: Mode shape plot of third and fifth mode found manually.

are clearly appearing as rigid body motions of the bridge deck, making this pile modes. The piles are neither completely fixed nor totally free to rotate at the endpoints. By analysing the expansion of the bridge it was calculated that the moments in the foundation end is approximately 90% of the moments in the top end, according to Thorenfeldt (2015). Comparing with the ratio between the theoretical mode shape frequencies of the shear frame in the numerical study one obtains:

$$\frac{f_2}{f_1} = \frac{2.946}{1.091} = 2.700 \qquad \frac{f_{T.2}}{f_{T.1}} = \frac{7.81}{2.63} = 2.97 \tag{4.1}$$

The ratio between the first and second mode has a deviation that may be due to the difference in stiffness between the top and bottom.

The first vertical mode of multi span bridges is expected to look like the mode shape in Figure 4.4e, and one can conclude that this mode is most likely the first vertical mode of Elgeseter bridge. This mode was not detected by the automatic algorithms, as previously discussed.

The fourth and fifth manual mode found is shown in Figures 4.5a and 4.5b. It is clear that these modes are torsional movement involving vertical motion of the bridge over the piles, which in most bridge analyses is assumed to be rigid. This may be due to flex in the slender columns or in the foundation of the piles. The piles are most likely in loose sediments or clay-material. Note also from Table 4.1 that there are possibly multiple modes in this frequency range, possibly different torsional modes dependant on that are slightly different depending on the stiffness of each column? Determining the actual mode shapes could therefore be an interesting analysis.

To summarize this study one can conclude that automatic algorithms are not best suited for merging of measurements with PoSER. This is due to the high precision needed to get any modes merged. PreGER is likely better suited since merging is done before the MPE, making the algorithms unaffected.

One can, however, state from this study that the differences observed in the numerical study maintains for real life data, making the numerical study great for testing future algorithms or methods without the need to do measurements outside.

Chapter 5

Concluding remarks and further work

The automatic interpretation of the stabilization diagram is for most purposes well performed by the algorithms suggested in Reynders et al. (2012) and Yang et al. (2019). This is substantiated by analyses on real life data and artificial data. Choosing which algorithm to use is a consideration depending on the desired results.

The steps of both algorithms are in general the same. With the implementation done in this thesis, it is possible to combine every sub step of the algorithms. In this way one can choose which filtering step, clustering step, and cluster filtering to use. Observed in this study is that local outlier detection is increasing the quality of the clusters, but not as much as the clustering method suggested by Yang, see flowchart in Figure 2.11. The problem with the clustering method of Yang is that it is to strict for some applications. The quality of the measurements must be higher because new clusters are created if the deviation to previously created clusters are to big. Combined with a fixed length threshold for the clusters this makes it unsuited for applications with few good modes and where finding all modes is crucial, like SHM. One possible quick fix it to apply the cluster filtering step of Reynders where the clusters are partitioned into two by their length. This because some measurements have less clear modes due to noise and little excitation making the fixed threshold unsuited.

The general trend is that Reynders' identifies more modes than Yang at the cost of lower quality clusters. In an context where identifying modes is the most important, Reynders' is best suited. If only precise estimates is wanted, Yang's is best suited.

Having performed numerous analyses on both real life and artificial data it is clear that there are some main challenges before modal parameter estimation is fully automatic. The interpretation of the stabilization diagram is fully automatic. The outcome is, however, highly dependant on the quality of the input. This was the case with the data from Elgeseter bridge where the manual analysis was able to find more modes than the algorithms. The main challenge is therefore not the clustering of modes, but rather getting high quality precise modes to cluster.

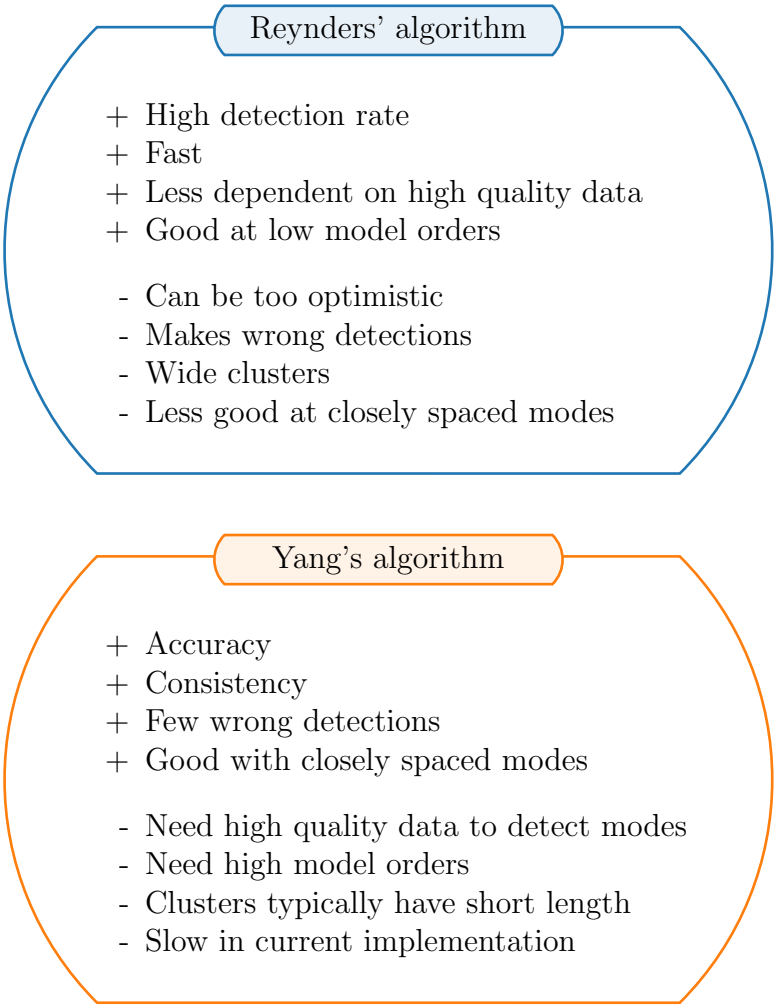


Figure 5.1: Most important findings.

These challenges can be summarized as:

- Quantify the noise present, or better add anti-noise filters
- Determine the sampling frequency
- Determine the number of block rows
- Determine the model orders, especially the maximum model order

If the noise present were quantified, Yang could have adapted the length threshold, i.e. much noise → few good modes → lower cluster length threshold, and opposite. Measures about the assumed quality could also be possible with noise levels determined. Even better would of course have been efficient noise filters.

Determination of sampling frequency is another tricky parameter. Throughout the work of this thesis a rule of thumb with 4 times the highest expected eigenfrequency seemed well suited. This should however be studied further. Automation is one step further if the best sampling frequency could be determined by a relation with the highest expected eigenfrequency. even if the expected highest frequency must be given as an input. It may also be possible to determine the highest frequency from the power spectral density of the input signal.

The problem with determining the number of block rows can be automated by using the approximation given by Yang. Quantification of the performance with this rule of thumb must then be performed. The same goes for determining the model orders with the approximations given by Neu, Section 3.1.1. With increased computational power this problem may vanish as massive over estimation is no problem. However, heavily overestimating the model order leads to more spurious modes to remove, and greater challenge for the algorithms. With the current implementation, and under the tested applications, Yang is best suited when the model order is heavily over estimated. This is substantiated by the surface plots in Figure 3.4, and also the clustering technique; if modes are not fitting the current clusters, new clusters are created. Reynders' on the other hand, is grouping based on distance between the elements, and with many elements there will certainly be spurious modes near the cluster that accidentally are merged.

The most important findings about the two algorithms are summarized in Figure 5.1.

Bibliography

Articles

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/J.2517-6161.1964.TB00553.X

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers.

Dooms, D., Degrande, G., De Roeck, G., & Reynders, E. (2006). Finite element modelling of a silo based on experimental modal analysis. *Engineering Structures*, *28*(4), 532–542. https://doi.org/10.1016/J.ENGSTRUCT.2005.09.008

Neu, E., Janser, F., Khatibi, A. A., & Orifici, A. C. (2017). Fully Automated Operational Modal Analysis using multi-stage clustering. *Mechanical Systems and Signal Processing*, *84*, 308–323. https://doi.org/10.1016/j.ymssp.2016.07.031

Reynders, E., & Roeck, G. D. (2008). Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis. *Mechanical Systems and Signal Processing*, *22*(3), 617–637. https://doi.org/10.1016/J.YMSSP.2007.09.004

Reynders, E., Houbrechts, J., & De Roeck, G. (2012). Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing*, *29*. https://doi.org/10.1016/j.ymssp.2012.01.007

Thorenfeldt, E. (2015). Alkalireaksjoner-Karbon-fiberforsøk Elgeseter bru.

Yaghoubi, V., Vakilzadeh, M. K., & Abrahamsson, T. J. (2018). Automated modal parameter estimation using correlation analysis and bootstrap sampling. *Mechanical Systems and Signal Processing*, *100*, 289–310. https://doi.org/10.1016/J.YMSSP.2017.07.004

Yang, X.-M., Yi, T.-H., Qu, C.-X., Li, H.-N., & Liu, H. (2019). Automated Eigensystem Realization Algorithm for Operational Modal Identification of Bridge Structures. *Journal of Aerospace Engineering*, *32*(2), 04018148. https://doi.org/10.1061/(asce)as.1943-5525.0000984

Books, web pages and software

Frøseth, G. T., & olavguddal. (2022). *Gunnstein/strid: Strid - v0.4.3* (Version v0.4.3). Zenodo. https://doi.org/10.5281/zenodo.6540518

Moelven. (2022). Mjøstårnet. https://www.moelven.com/mjostarnet/.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). *Scikit-learn: Machine learning in Python*.

Rainieri, C., & Fabbrocino, G. (2014). *Operational Modal Analysis of Civil Engineering Structures An Introduction and Guide for Applications*.

Reynders, E., MagalhãesMagalh, F., Mevel, L., De Roeck, G., & Cunha, A. (2010). *Pre- and Post-identification Merging for Multi-setup OMA with Covariance-driven SSI* (tech. rep.).

Rønnquist, A. (2022). Submerged floating tube bridges. https://www.ntnu.edu/kt/research/dynamics/research/long-span/submerged-floating-tube.

SVV. (2015). *Håndbok N400 - Bruprosjektering*.

WSDOT. (2022). Tacoma Narrows Bridge history. https://wsdot.wa.gov/TNBhistory/.

Appendices

The appendix to this thesis is delivered as a separate file in the submission.

The appendix contains:

- *autostrid.py* - implementation of the algorithms
- *process - fs and order.py* - Study 1 setup
- *process - force.py* - Study 2 setup
- Raw-data from sampling at Elgeseter bridge

For others interested in code or sampled data, make contact.

Olav Guddal

Comparison of methods for automatic modal analysis

NTNU
Kunnskap for en bedre verden