Bauhaus-Universität
Weimar

FAKULTÄT BAUINGENIEURWESEN
INSTITUT FÜR STRUKTURMECHANIK
MARIENSTRASSE 15 · D-99423 WEIMAR

Assessment of Data from Dynamic Bridge Monitoring

Submitted in Partial Fulfilment of the Requirements for
the Degree of Master of Science (M.Sc.)

Supervisors:
Prof. Dr.-Ing. habil. Carsten Könke
Dr.-Ing. Volkmar Zabel

Author:
Mihai-Andrei Udrea

February, 2014

Abstract

The focus of the thesis is to process measurements acquired from a continuous monitoring system at a railway bridge. Temperature, strain and ambient vibration records are analysed and two main directions of investigation are pursued.

The first and the most demanding task is to develop processing routines able to extract modal parameters from ambient vibration measurements. For this purpose, reliable experimental models are achieved on the basis of a stochastic system identification (SSI) procedure. A fully automated algorithm based on a three-stage clustering is implemented to perform a modal parameter estimation for every single measurement. After selecting a baseline of modal parameters, the evolution of eigenfrequencies is studied and correlated to environmental and operational factors.

The second aspect deals with the structural response to passing trains. Corresponding triggered records of strain and temperature are processed and their assessment is accomplished using the average strains induced by each train as the reference parameter. Three influences due to speed, temperature and loads are distinguished and treated individually. An attempt to estimate the maximum response variation due to each factor is also carried out.

Nomenclature

Abbreviations

AGNES	Agglomerative Nesting
ARMA	Autoregressive Moving Average
CE	Complex Exponential
CMI	Consistent-Mode Indicator
DIANA	Divisive Analysis
DOF	Degree Of Freedom
EFDD	Enhanced Frequency Domain Decomposition
EGV	Eigenvector Decomposition
EMA	Experimental Modal Analysis
ERA	Eigensystem Realization Algorithm
FDD	Frequency Domain Decomposition
FE	Finite Element
FEM	Finite Element Method
FFT	Fast Fourier Transform
FRF	Frequency Response Function
GUI	Graphical User Interface
ITD	Ibrahim Time Domain
LSCE	Least Squares Complex Exponential
p-LSCFD	poly-Least Squares Complex Frequency Domain
MAC	Modal Assurance Criterion
MDOF	Multiple Degree Of Freedom
MLE	Maximum Likelihood Estimator
MP	Mean Phase
MPC	Modal Phase Collinearity
MPD	Mean Phase Deviation

MPE	Modal Parameter Estimation
MRITD	Multiple Reference Ibrahim Time Domain
OMA	Operational Modal Analysis
OS	Operating System
PP	Peak-Picking
PRCE	Polyreference Complex Exponential
SHM	Structural Health Monitoring
SM	Structural Monitoring
SSE	Sum of the Squared Error
SSI	Stochastic Subspace Identification
SSI-cov/data	Reference-based Data-driven Stochastic Subspace Identification
SSI-cov/ref	Reference-based Covariance-driven Stochastic Subspace Identification
SSI/ref	Reference-based Stochastic Subspace Identification
SVD	Singular Value Decomposition
ZOH	Zero-Order Hold

Latin Letters

a_i, b_i	"Modal a" and "modal b" coefficient
c_i	Modal damping
$d, d(i,j)$	Distance measure for hierarchical clustering
$\tilde{d}$	Cut-off distance for hierarchical clustering
f	Frequency [Hz]
$f(t)$	Load vector
f_i, f_j, f_k	Eigenfrequency [Hz]
h_k	Markov matrices for time instant k
i	Length of each of the two blocks of the output data Hankel matrix (time-lag)
j	Imaginary unit $j^2 = -1$
k_i	Modal stiffness
l	Number of outputs
m	Number of inputs
m_i	Modal mass

n	State-space model order
n_2	Number of DOF
q	Order of the controllability and observability matrices
$\ddot{q}(t), \dot{q}(t), q(t)$	Time series of accelerations, velocities and displacements
$\ddot{q}_m(t), \dot{q}_m(t), q_m(t)$	Time series of modal accelerations, velocities and displacements
r	Number of references
t	Continuous time variable
$u(t)$	Input at time t
u_k	Input at time instant k
v	Train speed
v_k	Measurement noise vector at time instant k
w_f, w_{MAC}	Weights for frequency and mode shape correlation
w_k	Process noise vector at time instant k
$x(t)$	State vector at time t
x_k	State vector at time instant k
$x_m(t)$	Modal state vector at time t
$y(t)$	Output vector at time t
y_k	Output vector at time instant k
y_k^{ref}	Reference output vector at time instant k
A, B, C, D	Discrete-time state-space matrices
A_c, B_c, C_c, D_c	Continuous-time state-space matrices
B_2	Input location matrix
C_2	Damping matrix
$C_{d(f_i,f_j)}$	Final centroid for frequencies
C_i, C_j	Centroids of clusters i and j
$C_{MAC(\{\phi_i\},\{\phi_j\})}$	Final centroid for MAC
G	"Next state-output" covariance matrix
G^{ref}	Reduced "next state-output" covariance matrix
H_1, H_2	Block Hankel matrices
H^{ref}	Output data block Hankel matrix
I	Identity matrix

K	Stiffness matrix
K	Number of clusters for K-means method
L	Selection matrix of the references from the outputs
L^T	Discrete-time modal input matrix
L_c^T	Continuous-time modal input matrix
M	Mass matrix
N	Number of samples
O_i	Extended observability matrix of order i
O_q	Observability matrix of order q
P	Forward state covariance matrix
P, Q	System matrices to derive the state-space model
Q, R, S	Process and measurement noise covariance matrices
R_i	Output covariance matrix at time lag i
R_i^{ref}	Reduced output covariance matrix at time lag i
T	Matrix of similarity transformation
$T_{1\|i}^{ref}$	Output covariances block Toeplitz matrix
U, S, V	SVD factors
V	Discrete-time modal output matrix
V_c	Continuous-time modal output matrix
$Y_f, Y_{i\|2i-1}$	Hankel matrix of all future outputs
$Y_p^{ref}, Y_{0\|i-1}^{ref}$	Hankel matrix of past reference outputs

Greek Letters

α, β	Rayleigh damping constants
$\delta(t)$	Dirac delta
δ_k	Kronecker delta
λ_w	Wavelength of excitation
$\lambda_i, \lambda_j, \lambda_k$	Continuous-time eigenvalue
μ_i	Discrete-time eigenvalue
ξ_i, ξ_k	Modal damping ratios
τ	Time differential
φ_i	Real eigenvector
ϕ_i, ϕ_j	Complex eigenvectors

ω	Circular frequency [rad/s]
ω_i	Circular eigenfrequency [rad/s]
Γ	Diagonal matrix assembled from $2\xi_i\omega_i$
Γ_i^{ref}	Extended controllability matrix of order i
Γ_q	Controllability matrix of order q
Δt	Time interval (sampling period)
Θ	Matrix of complex eigenvectors
Λ	Diagonal matrix assembled from λ_i
Λ_c	Diagonal matrix assembled from the continuous-time λ_i, λ_i^*
Λ_d	Diagonal matrix assembled from the discrete-time λ_i, λ_i^*
Σ	State covariance matrix
Φ	Matrix of real eigenvectors
Ψ	Matrix of complex eigenvectors of the state-space model
Ω	Diagonal matrix assembled from ω_i

Mathematical Notations

$\{\bullet\}$	Vector
$[\bullet]$	Matrix
$\begin{bmatrix} \ddots & & \\ & \bullet & \\ & & \ddots \end{bmatrix}$	Diagonal matrix
$[\bullet]_{i,j}$	Row i, column j of a matrix
$[\bullet]_{:,j}$	Column j of a matrix
$\{\bullet\}^T$	Transpose of a vector
$[\bullet]^T$	Transpose of a matrix
$[\bullet]^{-1}$	Inverse of a matrix
$[\bullet]^\dagger$	Moore-Penrose pseudo-inverse of a matrix
$(\bullet)^*$	Complex conjugate
$\lvert\bullet\rvert$	Absolute value
$\lVert\bullet\rVert_2$	Euclidean norm
$\mathbf{E}[\bullet]$	Expected value operator
$\mathrm{Im}\{\bullet\}$	Imaginary part of a complex vector
$\mathrm{Re}\{\bullet\}$	Real part of a complex vector

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the last decades, the necessary maintenance of structures together with their inherent ageing and degradation have faced the structural engineers and the scientific community to constant challenges. The numerical models used in the design phase are based on various simplifications which approximate the real behaviour of a structure. Therefore, experimental data collected from different types of sensors plays an important role in validating the theoretical assumptions and in predicting the operational response of a great variety of structures from civil engineering to automotive, aerospace or industrial engineering. For this purpose, next to the previous mentioned fields, disciplines like electrical engineering (e.g. signal processing, control, telecommunications, computer engineering) and mechanical engineering are also interconnected and benefit from each other and their individual progress.

1.1 Study Context

Since the natural tendency in structural engineering has been to design more and more complex and slender structures, also the experimental tools had to keep pace in order to provide reliable data to support the calibration, updating and eventual optimization of numerical models. One major concern if experimentally investigating a structure is to identify the key parameters that define its dynamic behaviour: *natural frequencies, mode shapes* and *damping ratios*. These are referred to as modal parameters and depend on the geometry, material properties and boundary conditions of the system. They are used to describe the modal model as a set of linearly combined, time-invariant particular modes of vibration, also called resonant modes. An in-depth knowledge and understanding of the vibration problems caused by the excitation of a structure at one of its natural frequencies can avoid unpleasant issues, such as discomfort, malfunctioning, damage or even total collapse.

Of particular interest is the structural response of railway bridges to dynamic loads since they are an important part of the surface transportation system and their operational failure can cause severe economic, environmental and social consequences. The focus on their dynamic behaviour has increased in recent years, due to the introduction of high speed trains which include trains capable of reaching speeds of over 200 km/h.

A study conducted by the European Commission in 2009 [4] indicates an increasing trend in the length of high speed rails in Europe from about 700 km in 1985 to more than 6000 km in 2010 (Fig. 1.1). In order to meet the scheduled 15 000 km by 2020 and 21 000 km after 2020, approximately 800 to 1000 km of high speed lines are expected to be constructed annually.

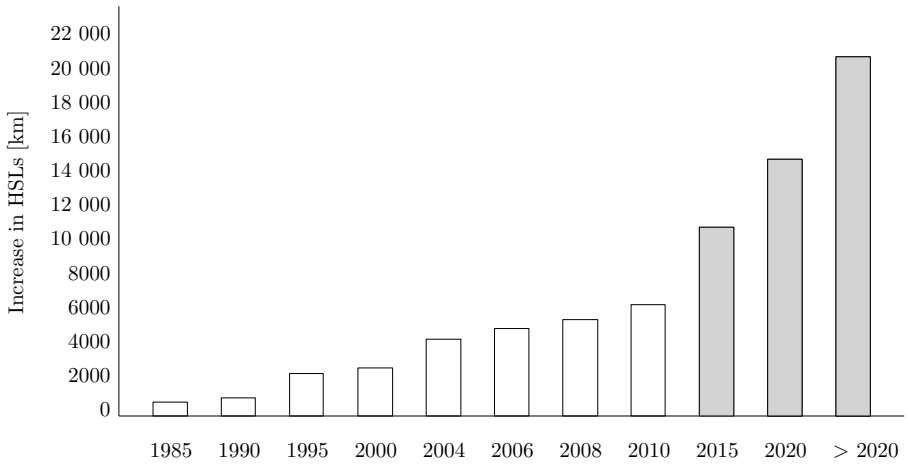

Figure 1.1: Length of high speed rails in Europe since 1985 and predicted length (grey filled bars) according to [4]

From a structural point of view, this generates additional problems caused by large vibrations in the bridge superstructure at train passages. Next to the train-superstructure interaction, the most important parameters which influence the maximum deflection, such as the train speed, number of trains and their travel path, span length, axle loads, mass, stiffness and damping of the system, underline the complex phenomenon of the railway bridges dynamic response. As a consequence, the classical consideration of the dynamic effects by magnifying the static loads using dynamic amplification factors based mainly on the natural frequency and the span length, may yield unreliable results.

Short-span bridges raise even new challenges since their natural frequency may lie in the range of the excitation frequency. For example, after introducing the TGV on the Paris-Lyon line, short-span bridges showed cracks and crumbles of concrete, high ballast attrition due to high accelerations and great track irregularities [27]. The excessive bridge deck vibration can also cause the loss of the wheel-rail contact, destabilisation of the ballast and exceedance of the stress limits. In the context of all the uncertainties which can not be fully grasped by numerical models, the assessment of the data from dynamic bridge monitoring is vital.

Structural monitoring (SM) has proved to be an essential tool to deliver accurate information about the parameters which characterize a real system. This enabled not only an estimation of the integrity of a structure (the field of structural health monitoring and damage detection), but also a valuable insight into the physical model which ultimately led to an increase in safety and maintainability. The addressed issues include the excitation methods (forced, ambient or local excitation), the data transmission, the sensors placement and the sensing structural response or measured quantities (strain, displacement, velocity, acceleration, temperature, etc.) Due to intrinsic sensitivities, a SM system is adapted to a particular structure and accordingly, there is no universal technique to gather all the possible information and solve every

problem which may occur. However, together with the development of statistical methods which take into account experimental errors, incompleteness, environmental and operational conditions, great progress has been achieved concerning the tools used to process the data.

The modal identification based on the data acquired from monitoring systems can be mainly classified into two groups:

- *Input-output* or *experimental modal analysis* (EMA): is based on the estimation of a set of frequency response functions (FRF) which relate the response of the structure to the applied forces. Since the generated loads from the excitation source (impact hammer, electrodynamic shakers or eccentric mass vibrators) of the system have to be completely known, this procedure requires controlled tests or recreated load scenarios. The extraction of the modal parameters from the FRF can be performed using a wide variety of mathematical curve-fitting algorithms between the measured and the theoretical functions. The main drawback of this approach is that it requires either a laboratory environment which can be significantly different than its real operating conditions, or on-site forced vibration tests that are costly and involve heavy equipment.
- *Output-only* or *operational modal analysis* (OMA): considers the ambient excitations as a zero mean Gaussian white noise and estimates the modal model only on the basis of the recorded response. Its recent development has been facilitated by the advancement of concepts and algorithms from numerical linear algebra, such as singular value decomposition (SVD), orthogonal decomposition (QR) and eigenvector decomposition (EGV). An important issue comes from the white noise assumption and the difficulty of estimating the correct system order. This leads in most cases to the occurrence of additional computational modes, also called mathematical or spurious modes, without physical meaning, which need to be removed.

Cunha et al. [5] gives a brief overview and a short description of most of the EMA and OMA techniques.

1.2 Outline of the Thesis

The thesis presents the theoretical background of the topic and the main challenges, as well as their application for a study case. It is organized as follows:

Chapter 1 introduces the study context and the relevance of the research with a brief overview of the two main types of modal identification, input-output and output-only.

Chapter 2 presents the most important models of vibrating structures and connects the theoretical background of deterministic to stochastic system identification. Special emphasis is put on the stochastic subspace identification, which is applied later in the study case.

Chapter 3 deals with the extraction of modal parameters from stabilization diagrams and includes the main clustering algorithms used for an automatic extraction of the modal parameters. Several approaches existent in the literature are also briefly mentioned.

Chapter 4 represents the research study conducted for the measurements acquired by a continuous dynamic monitoring system at a railway bridge. The first part involves modal parameter estimation and modal tracking. The second part analyses the structural response due to passing trains, where the influences of speed, temperature and loads are discussed.

Chapter 5 summarizes the main results and concludes with possibilities for future research.

Additionally, appendices are included to document the developed routines and highlight the influence of several stability criteria and signal decimation on the identification results. They also contain an example showing the performance of the automatic procedure for a set of measurements from one day.

Chapter 2

System Identification

The goal of system identification procedures in structural engineering applications is to extract modal parameters estimates from recorded experimental data. Its theoretical foundation lies on the correlation between the properties of different dynamical models which are extensively explained in some reference books like Ewins [6], Juang [11], Natke [13] or Gawronski [7].

There are various possibilities to derive the analytical models. The main classification is done using the representation domain criterion and groups them into time domain (differential equations) or frequency domain (transfer functions) models. In the following sections, only the time domain models will be shortly described and one path to achieve the system identification will be presented.

The time domain models are represented by linear differential equations that can be either second-order, which are usually used in structural engineering, or first-order, also known as state-space formulation, preferred in control engineering. As pointed out in [7], the state-space standardization of structural models allows for the extension of known control system properties into structural dynamics. From this point of view, the FE model merely represents the starting point to develop the basis appropriate for experimental modelling.

As a side note, the notations and workflow of this chapter comply to those from Peeters [15], Magalhães [12] and Zabel [25]. For clarity reasons, the matrices and vectors brackets syntax as in [6] and [25] is chosen over the bold format.

2.1 Time-domain Models for Dynamic Systems

In order to derive the stochastic model used in the operational modal analysis employed in this thesis, it should be understood how to relate the classical finite element (FE) model to the state-space model. Afterwards, an important intermediate step is to transform the continuous-time relations, the case of analytical models, to discrete-time relations, the case of experimental models.

2.1.1 Finite Element Models

The well-known linear time invariant modelling of the dynamic behaviour of a discrete mechanical system with multiple degree of freedom (MDOF) is the matrix form of Newton's second law of motion:

$$[M]\{\ddot{q}(t)\} + [C_2]\{\dot{q}(t)\} + [K]\{q(t)\} = \{f(t)\} = [B_2]\{u(t)\} \tag{2.1}$$

where $[M], [C_2], [K] \in \mathbb{R}^{n_2 \times n_2}$ denote the mass, damping and stiffness matrices; $\{q(t)\}, \{\dot{q}(t)\}, \{\ddot{q}(t)\} \in \mathbb{R}^{n_2}$ are the displacement, velocity and acceleration column vectors at continuous time t. The load vector $\{f(t)\} \in \mathbb{R}^{n_2}$ contains the forces applied to the system and can be replaced by a vector of $m < n_2$ dimensions, $\{u(t)\} \in \mathbb{R}^{m}$, multiplied by a matrix that maps the m inputs to the n_2 degrees of freedom (DOF), $[B_2] \in \mathbb{R}^{n_2 \times m}$. $[B_2]$ is composed of ones and zeros and specifies the input location.

Second-order structural models can be described in either *nodal* or *modal* coordinates. The classical FEM approach is to compute nodal models whereas in structural dynamics a modal description is preferred to reduce the order of the system. In the context of system identification, the modal representation is a natural outcome of an experiment [7] and it is not possible, nor necessary to measure all DOF of a FE model in order to obtain an accurate experimental model [15].

The damping matrix $[C_2]$ cannot be assembled in the same way as $[M]$ and $[K]$ due to the lack of identifiable or measurable material constants that govern the global damping behaviour of a structure [15]. For this reason, the cases of proportional and general viscous damping are treated separately.

Proportional Viscous Damping Model

After solving the classical eigenvalue problem and inserting the eigenvectors $\{\varphi_i\} \in \mathbb{R}^{n_2}$ $(i = 1, \ldots, n_2)$ into the matrix $[\Phi] \in \mathbb{R}^{n_2 \times n_2}$, Eq. (2.1) can be pre-multiplied by $[\Phi]^T$ and a coordinate transformation $\{q(t)\} = [\Phi]\{q_m(t)\}$ can by introduced, resulting in:

$$\begin{aligned}[\Phi]^T[M][\Phi]\{\ddot{q}_m(t)\} + [\Phi]^T[C_2][\Phi]\{\dot{q}_m(t)\} + [\Phi]^T[K][\Phi]\{q_m(t)\} \\ = [\Phi]^T[B_2]\{u(t)\}\end{aligned} \tag{2.2}$$

where the vector $\{q_m(t)\} \in \mathbb{R}^{n_2}$ contains the so-called modal displacements. At this stage, the orthogonality properties displayed in Eq. (2.3) apply also for the particular case of proportional damping as shown in Eq. (2.4).

$$[\Phi]^T[M][\Phi] = \begin{bmatrix} \ddots & \\ & m_i \\ & & \ddots \end{bmatrix} \qquad [\Phi]^T[K][\Phi] = \begin{bmatrix} \ddots & \\ & k_i \\ & & \ddots \end{bmatrix} \tag{2.3}$$

$$[\Phi]^T[C_2][\Phi] = \begin{bmatrix} \ddots & \\ & c_i \\ & & \ddots \end{bmatrix} = \begin{bmatrix} \ddots & \\ & 2\xi_i\omega_i m_i \\ & & \ddots \end{bmatrix} = [\Gamma]\begin{bmatrix} \ddots & \\ & m_i \\ & & \ddots \end{bmatrix} \tag{2.4}$$

where $\xi_i = c_i/2m_i\omega_i$ is the classical definition of modal damping ratios, ω_i [rad/s] denotes the angular eigenfrequencies, m_i and k_i are the modal masses and stiffnesses,

respectively. Introducing Eq. (2.3) and Eq. (2.4) into Eq. (2.2), it yields:

$$[I]\{\ddot{q}_m(t)\} + [\Gamma]\{\dot{q}_m(t)\} + \left[\Omega^2\right]\{q_m(t)\} = \begin{bmatrix} \ddots & & \\ & \frac{1}{m_i} & \\ & & \ddots \end{bmatrix} [\Phi]^T [B_2]\{u(t)\} \tag{2.5}$$

where the identity matrix $[I]_{n_2 \times n_2}$ and $\left[\Omega^2\right] = \begin{bmatrix} \ddots & & \\ & \omega_i^2 & \\ & & \ddots \end{bmatrix} \in \mathbb{R}^{n_2 \times n_2}$ result from the common practice of mass normalization. In this case, the equation is decoupled, since all the left hand side matrices are diagonal.

The solution of the homogeneous equation has the form $\{q(t)\} = \{\varphi_i\} e^{\lambda_i t}$and leads to the complex eigenvalues λ_i in Eq. (2.6) and the same eigenvectors as if the system was undamped, Eq. (2.7):

$$\lambda_i^2 + 2\xi_i\omega_i\lambda_i + \omega_i^2 = 0$$

$$\lambda_i,\ \lambda_i^* = -\xi_i\omega_i \pm j\sqrt{1-\xi_i^2}\,\omega_i \tag{2.6}$$

$$[K]\{\varphi_i\} = [M]\{\varphi_i\}\omega_i^2 \tag{2.7}$$

The superscript * is the convention for complex conjugate. The damping proportionality is often achieved by assuming a linear combination of the stiffness and mass matrices, well-known as Rayleigh damping.

$$[C_2] = \alpha[M] + \beta[K]$$

where α and β are non-negative scalar constants. This constraint on the damping behaviour implies that the damping is distributed in a structure in the same way as the mass and stiffness are. However, this is not always the case and a general approach should be considered as well.

General Viscous Damping Model

In the case of non-proportional damping, the modal parameters are extracted by reducing the second order of the equation of motion (2.1) to first order. This is achieved by inserting the equality $[M]\{\dot{q}(t)\} = [M]\{\dot{q}(t)\}$ into Eq. (2.1) and introducing additional notations:

$$\underbrace{\begin{bmatrix} [C_2] & [M] \\ [M] & [0] \end{bmatrix}}_{[P]} \underbrace{\begin{Bmatrix} \{\dot{q}(t)\} \\ \{\ddot{q}(t)\} \end{Bmatrix}}_{\{\dot{x}(t)\}} + \underbrace{\begin{bmatrix} [K] & [0] \\ [0] & [-M] \end{bmatrix}}_{[Q]} \underbrace{\begin{Bmatrix} \{q(t)\} \\ \{\dot{q}(t)\} \end{Bmatrix}}_{\{x(t)\}} = \begin{bmatrix} [B_2] \\ [0] \end{bmatrix} \{u(t)\} \tag{2.8}$$

The vector $\{x(t)\} \in \mathbb{R}^n$ is called the *state vector*. The model order n is equal to $2 \times n_2$. Using a solution formulation similar to the previous model, $\{x(t)\} = [\Psi] e^{\Lambda_c t}$, the corresponding eigenvalue problem is:

$$[P][\Psi][\Lambda_c] + [Q][\Psi] = [0] \tag{2.9}$$

where $\Psi \in \mathbb{C}^{n \times n}$ is assembled by n complex column vectors and the diagonal matrix $[\Lambda_c] = \begin{bmatrix} \ddots & & \\ & \lambda_i & \\ & & \ddots \end{bmatrix} \in \mathbb{C}^{n \times n}$ contains n complex eigenvalues λ_i [rad/s]. The structure of

Λ_c and Ψ is:

$$[\Lambda_c] = \begin{bmatrix} [\Lambda] & [0] \\ [0] & [\Lambda^*] \end{bmatrix} \qquad [\Psi] = \begin{bmatrix} [\Theta] & [\Theta^*] \\ [\Theta]\,[\Lambda] & [\Theta^*][\Lambda^*] \end{bmatrix} \tag{2.10}$$

where $[\Lambda], [\Theta] \in \mathbb{R}^{n_2 \times n_2}$ contain the eigenvalues and eigenvectors of the original second order differential equations. Since the eigenvectors of the general damping model are not the same as in the proportional damping case, they are symbolized with $[\Theta]$ instead of $[\Phi]$ and are not diagonalizing any of the matrices $[M], [C_2], [K]$. The complex eigenvalues are given here also by Eq. (2.6). If Eq. (2.10) is introduced in Eq. (2.9), it results:

$$\begin{bmatrix} [M]\,[\Theta]\left[\Lambda^2\right] + [C_2]\,[\Theta]\,[\Lambda] + [K]\,[\Theta] & [M]\,[\Theta^*]\,[\Lambda^*]\,[\Lambda] + [C_2]\,[\Theta^*]\,[\Lambda] + [K]\,[\Theta^*] \\ [M]\,[\Theta]\,[\Lambda] - [M]\,[\Theta]\,[\Lambda] & [M]\,[\Theta^*]\,[\Lambda^*] - [M]\,[\Theta^*]\,[\Lambda^*] \end{bmatrix}$$
$$= \begin{bmatrix} [M]\,[\Theta]\left[\Lambda^2\right] + [C_2]\,[\Theta]\,[\Lambda] + [K]\,[\Theta] & [M]\,[\Theta^*]\left[\Lambda^2\right] + [C_2]\,[\Theta^*]\,[\Lambda] + [K]\,[\Theta^*] \\ [0] & [0] \end{bmatrix}$$
$$= [0] \tag{2.11}$$

One could notice that the terms in Eq. (2.11) which are nonzero replicate the general equation of motion (2.1) and prove the system order reduction approach.

The system matrices $[P]$ and $[Q]$ can be diagonalized using the orthogonality properties of $[\Psi]$:

$$[\Psi]^T [P]\,[\Psi] = \begin{bmatrix} \diagdown & & \\ & a_i & \\ & & \diagdown \end{bmatrix} \qquad [\Psi]^T [Q]\,[\Psi] = \begin{bmatrix} \diagdown & & \\ & b_i & \\ & & \diagdown \end{bmatrix} \tag{2.12}$$

The two matrices in Eq. (2.12) are called *modal a matrix* and *modal b matrix*. Pre-multiplying Eq. (2.12) by the matrix $[\Psi]^{-T}$ and substituting the obtained expressions into Eq. (2.9) yields:

$$\lambda_i = -\frac{b_i}{a_i} \tag{2.13}$$

2.1.2 State-space Models

The state-space models were developed in control theory and describe mathematically the physical systems as sets of inputs, outputs and state variables. They are generally defined by:

$$\begin{aligned} \{\dot{x}\} &= [A]\,\{x\} + [B]\,\{u\} \\ \{y\} &= [C]\,\{x\} \end{aligned} \tag{2.14}$$

with the initial state $\{x\,(0)\} = \{x_0\}$. The vector $\{x\}_{n\times 1}$ is the previously mentioned state vector and shows the relation between the system state and the input, $\{u\}_{m\times 1}$ is the system input and $\{y\}_{l\times 1}$ is the system output. The system state-space representation is given by the matrices $[A]_{n\times n}, [B]_{n\times m}, [C]_{l\times n}$. The l dimension in the second equation is less than the number of DOF considered, since the structural response cannot be measured at every single point, as already mentioned in Section 2.1.1.

For a dynamic system, Eq. (2.14) can be derived either from the system model given in time domain by differential equations or from its transfer function representation in frequency domain. Only the first case will be presented here.

Continuous-time Model

The starting point is to pre-multiply Eq. (2.8) by:

$$[P]^{-1} = \begin{bmatrix} [0] & [M]^{-1} \\ [M]^{-1} & -[M]^{-1}[C_2][M]^{-1} \end{bmatrix}$$

This step normalizes the term in $\{\dot{x}(t)\}$ and gives:

$$\begin{aligned} \{\dot{x}(t)\} &= -[P]^{-1}[Q]\{x(t)\} + [P]^{-1}\begin{bmatrix} [B_2] \\ [0] \end{bmatrix}\{u(t)\} \\ &= \underbrace{\begin{bmatrix} [0] & [I] \\ -[M]^{-1}[K] & -[M]^{-1}[C_2] \end{bmatrix}}_{[A_c]\in\mathbb{R}^{n\times n}}\{x(t)\} + \underbrace{\begin{bmatrix} [0] \\ [M]^{-1}[B_2] \end{bmatrix}}_{[B_c]\in\mathbb{R}^{n\times m}}\{u(t)\} \end{aligned} \tag{2.15}$$

which is similar to the first equation of Eq. (2.14) . The *state matrix* A_c and *input matrix* B_c bear the subscript $_c$ which means continuous in time. Using the modal decomposition of $[P]$ and $[Q]$, $[A_c]$ can be rewritten as a product of the eigenvectors and eigenvalues from Eq. (2.10):

$$\begin{aligned} [A_c] &= -[P]^{-1}[Q] \\ &= -[\Psi]\begin{bmatrix} \diagdown & & \\ & \frac{1}{a_i} & \\ & & \diagdown \end{bmatrix}[\Psi]^T[\Psi]^{-T}\begin{bmatrix} \diagdown & b_i & \diagdown \end{bmatrix}[\Psi]^{-1} \\ &= [\Psi][\Lambda_c][\Psi]^{-1} \end{aligned} \tag{2.16}$$

which corresponds to the standard eigenvalue problem $[A_c][\Psi] = [\Psi][\Lambda_c]$. Consequently, $[\Psi]$ and $[\Lambda_c]$ are assembled with the eigenvectors and eigenvalues of $[A_c]$.

The second equation format from Eq. (2.14), also known as observation equation, can be applied to the dynamic model and extended by introducing $[C_a]$, $[C_v]$, $[C_d] \in \mathbb{R}^{l\times n_2}$ which are the acceleration, velocity and displacement measurements location matrices:

$$\{y(t)\} = [C_a]\{\ddot{q}(t)\} + [C_v]\{\dot{q}(t)\} + [C_d]\{q(t)\} \tag{2.17}$$

The accelerations $\{\ddot{q}(t)\}$ from Eq. (2.17) can be replaced using Eq. (2.1) and obtain:

$$\begin{aligned} \{y(t)\} = &\underbrace{\begin{bmatrix} \Big([C_d] - [C_a][M]^{-1}[K]\Big) & \Big([C_v] - [C_a][M]^{-1}[C_2]\Big) \end{bmatrix}}_{[C_c]\in\mathbb{R}^{l\times n}}\begin{Bmatrix} \{q(t)\} \\ \{\dot{q}(t)\} \end{Bmatrix} \\ &+ \underbrace{[C_a][M]^{-1}[B_2]}_{[D_c]\in\mathbb{R}^{l\times m}}\{u(t)\} \end{aligned} \tag{2.18}$$

Next to the *output/observation matrix* $[C_c]$, the additional *direct transmission/feed-through matrix* $[D_c]$ is introduced when compared to the general formulation from Eq. (2.14). This is equal to zero if there are no accelerometers used ($[C_a] = [0]$).

The complete state-space model is given by Eqs. (2.15) and (2.18):

$$\begin{aligned} \{\dot{x}(t)\} &= [A_c]\{x(t)\} + [B_c]\{u(t)\} \\ \{y(t)\} &= [C_c]\{x(t)\} + [D_c]\{u(t)\} \end{aligned} \tag{2.19}$$

In this form, the matrices $[A_c], [B_c], [C_c]$ and the state vector $\{\dot{x}(t)\}$ are not unique yet. To achieve this, an alternative state model can be developed using a so-called *similarity transformation*:

$$\{x(t)\} = [T]\{z(t)\} \tag{2.20}$$

where $[T] \in \mathbb{C}^{n \times n}$ is a quadratic non-singular complex matrix. If applied to Eq. (2.19), this coordinate transformation yields:

$$\begin{aligned} \{\dot{z}(t)\} &= [T]^{-1}[A_c][T]\{z(t)\} + [T]^{-1}[B_c]\{u(t)\} \\ \{y(t)\} &= [C_c][T]\{z(t)\} + [D_c]\{u(t)\} \end{aligned} \tag{2.21}$$

The matrices pre-multiplying the vectors of the system describe the same input-output relations as the original matrices, but the new state vector $\{z(t)\}$ loses the meaning of physical displacements and velocities.

A special case of similarity transformation leads to a modal form of the state-space model. A similar equation to Eq. (2.17) can be obtained if $[T]$ is substituted by $[\Psi]$ and $\{z(t)\}$ by the modal state vector $\{x_m(t)\} \in \mathbb{R}^{n \times n}$:

$$\{x(t)\} = [\Psi]\{x_m(t)\}. \tag{2.22}$$

Using the modal decomposition $[A]_c = [\Psi][\Lambda_c][\Psi^{-}1]$ and inserting Eq. (2.22) in Eqs.(2.21), it yields:

$$\begin{aligned} \{\dot{x}_m(t)\} &= [\Psi]^{-1}[\Psi][\Lambda_c][\Psi]^{-1}[\Psi]\{x_m(t)\} + [\Psi]^{-1}[B_c]\{u(t)\} \\ &= [\Lambda_c]\{x_m(t)\} + \underbrace{[\Psi]^{-1}[B_c]}_{[L_c]^T}\{u(t)\} \\ \{y(t)\} &= \underbrace{[C_c][\Psi]}_{[V_c]}\{x_m(t)\} + [D_c]\{u(t)\} \end{aligned} \tag{2.23}$$

Thus, the so-called *modal state-space model* is obtained:

$$\begin{aligned} \{\dot{x}_m(t)\} &= [\Lambda_c]\{x_m(t)\} + [L_c]^T\{u(t)\} \\ \{y(t)\} &= [V_c]\{x_m(t)\} + [D_c]\{u(t)\} \end{aligned} \tag{2.24}$$

where $[L_c]^T$ is the modal input matrix and contains the modal participation factors and $[V_c]$ is the modal output matrix and indicates which part of the mode shapes can be observed from data (displacements, velocities or accelerations measurements) [15]. Together with $[\Lambda_c]$, they give the modal parameters of the structure.

The theory behind to model described above is very similar to the classical modal analysis. Consequently, the model reduction by considering only a limited number of modes which contribute to the global structural response, can be applied as well.

The task to solve the state-space equations (2.24) is relatively difficult, since it involves a vectorial differential equation. An important step is to understand how to approach the scalar first order differential equation:

$$\{\dot{x}(t)\} = a\{x(t)\} + b\{u(t)\}$$

Pre-multiplying this equation with e^{-at} and denoting $\{x(t_0)\} = \{x_0\}$, it results:

$$\begin{aligned}
e^{-at}\{\dot{x}(t)\} &= e^{-at}a\{x(t)\} + e^{-at}b\{u(t)\} \\
e^{-at}b\{u(t)\} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-at}\{x(t)\}\right) \\
\int_0^t e^{-a\tau}b\{u(\tau)\}\,d\tau &= e^{-a\tau}\{x(\tau)\}\Big|_0^t \\
\int_0^t e^{a(t-\tau)}b\{u(\tau)\}\,d\tau &= \{x(t)\} - e^{at}\{x_0\} \\
\{x(t)\} &= e^{at}\{x_0\} + \int_0^t e^{a(t-\tau)}b\{u(\tau)\}\,d\tau
\end{aligned} \tag{2.25}$$

Now, if Eq. (2.25) is applied to the vectorial representation, the solution is achieved as:

$$\{x(t)\} = e^{[A_c](t-t_0)}\{x(t_0)\} + \int_{t_0}^{t} e^{[A_c](t-\tau)}[B_c]\{u(\tau)\}\,d\tau \tag{2.26}$$

Discrete-time Model

In practice, a signal is always achieved at a defined sampling rate Δt [s]. For this reason, the solution given by Eq. (2.26) needs to be adapted to a discrete-time series with $k \in \mathbb{N}$ time instants ($t = k\Delta t$):

$$\{x((k+1)\Delta t)\} = e^{[A_c]\Delta t}\{x(k\Delta t)\} + \int_{k\Delta t}^{(k+1)\Delta t} e^{[A_c]((k+1)\Delta t-\tau)}[B_c]\{u(\tau)\}\,\mathrm{d}\tau \tag{2.27}$$

Assuming that the excitation $\{u(\tau)\}$is constant between two sampling instants, which is also referred to as a zero-order hold (ZOH) assumption, it can be replaced by $\{u(\tau)\} = \{u(k\Delta t)\}$. Introducing the notation $\tau' = (k+1)\,\delta t - \tau$, it follows:

$$\{x((k+1)\Delta t)\} = e^{[A_c]\Delta t}\{x(k\Delta t)\} + \left[\int_0^{\Delta t} e^{[A_c]\tau'}\mathrm{d}\tau'\,[B_c]\right]\{u(k\Delta t)\} \tag{2.28}$$

If the discrete-time state vector is $\{x_k\} = \{x(k\Delta t)\}$, $\{u_k\}$ and $\{y_k\}$ the sampled input and output, then, using the additional notations:

$$\begin{aligned}
[A] &= e^{[A_c]\Delta t} & [B] &= \int_0^{\Delta t} e^{[A_c]\tau'}\mathrm{d}\tau'\,[B_c] = ([A]-[I])\,[A_c]^{-1}\,[B_c] \\
[C] &= [C_c] & [D] &= [D_c]
\end{aligned} \tag{2.29}$$

leads to the *discrete-time state-space model*:

$$\begin{aligned}
\{x_{k+1}\} &= [A]\{x_k\} + [B]\{u_k\} \\
\{y_k\} &= [C]\{x_k\} + [D]\{u_k\}
\end{aligned} \tag{2.30}$$

The discrete-time state-space model can be also transferred to the modal domain using the the eigenvalue decomposition of the continuous state matrix $[A_c]$, as shown in

Eq. (2.31), and employing the McLaurin series expansion of the exponential function, presented in Eq. (2.32):

$$[A_c] = [\Psi]\,[\Lambda_c]\,[\Psi]^{-1} \tag{2.31}$$

$$[A] = e^{[A_c]\Delta t} = [I] + [A_c]\,\Delta t + \frac{1}{2!}\left([A_c]\,\Delta t\right)^2 + \frac{1}{3!}\left([A_c]\,\Delta t\right)^3 + \ldots \tag{2.32}$$

The final format is thus obtained as:

$$[A] = e^{[A_c]\Delta t} = e^{[\Psi][\Lambda_c][\Psi]^{-1}\Delta t} = [\Psi]\,e^{[\Lambda_c]\Delta t}\,[\Psi]^{-1} = [\Psi]\,[\Lambda_d]\,[\Psi]^{-1} \tag{2.33}$$

where

$$[\Lambda_d] = \begin{bmatrix} \ddots & & \\ & \mu_i & \\ & & \ddots \end{bmatrix} \tag{2.34}$$

The eigenvectors of $[A]$ are the same as for $[A_c]$ and the eigenvalues of the discrete-time state-space model μ_i contained in the diagonal matrix $[\Lambda_d]$ are related to the continuous ones by:

$$\mu_i = e^{\lambda_i \Delta t} \Leftrightarrow \lambda_i = \frac{\ln(\mu_i)}{\Delta t} \tag{2.35}$$

It is straightforward now to introduce a similar definition of the discrete modal participation matrix and the observed mode shapes as in Eq. (2.24):

$$[L]^T = [\Psi]^{-1}\,[B] \qquad [V] = [C]\,[\Psi] \tag{2.36}$$

and define a *modal discrete-time state-space model*:

$$\begin{aligned} \{\dot{x}_{m,k+1}(t)\} &= [\Lambda_c]\,\{x_{m,k}(t)\} + [L_c]^T\,\{u(t)\} \\ \{y(t)\} &= [V_c]\,\{x_{m,k}(t)\} + [D_c]\,\{u(t)\} \end{aligned} \tag{2.37}$$

Stochastic Model

The previously described models are based on the deterministic assertion that the input time functions are completely known. But the fundamental concept behind operational modal analysis is that the signal is recorded without any knowledge of the source excitation. For this reason, it is necessary to introduce the concepts of stochastic processes with their statistical properties.

A stochastic process is a collection of random time dependent variables $\{x(t)\}$, defined by continuous or discrete realizations. The practical and most common assumptions are *stationarity, ergodicity* and *zero-mean*. A process is called stationary if the statistical properties are constant over time and ergodic if the statistical properties can be calculated either considering average values over many realizations at a certain time instant, or using the average values of just one realization over time. The zero-mean hypothesis implies detrending the measured signals [12].

According to these considerations, the deterministic model described by Eq. (2.30) can be adapted by including the process noise $\{w_k\} \in \mathbb{R}^{n\times 1}$ to take into account random disturbances and modelling inaccuracies and the measurement noise $\{v_k\} \in \mathbb{R}^{l\times 1}$ due to the inherent sensor inaccuracy.

This leads to the following *discrete-time combined deterministic-stochastic state-space model*:

$$\begin{aligned} \{x_{k+1}\} &= [A]\{x_k\} + [B]\{u_k\} + \{w_k\} \\ \{y_k\} &= [C]\{x_k\} + [D]\{u_k\} + \{v_k\} \end{aligned} \tag{2.38}$$

Under the zero-mean and white noise (constant power spectral density) assumptions, the covariance matrices for two arbitrary time instants p and q can be expressed as:

$$\left[\mathbf{E}\left[\left\{\begin{array}{c}\{w_p\}\\ \{v_p\}\end{array}\right\}\left[\begin{array}{cc}\{w_q\}^T & \{v_q\}^T\end{array}\right]\right]\right] = \left[\begin{array}{cc}[Q] & [S]\\ [S]^T & [R]\end{array}\right]\delta_{pq} \tag{2.39}$$

where $\mathbf{E}[\bullet]$ denotes the expected value and δ_{pq} is the Kronecker delta. The matrices $[Q] \in \mathbb{R}^{n\times n}, [S] \in \mathbb{R}^{n\times l}, [R] \in \mathbb{R}^{l\times l}$ are the covariance matrices of the noise sequences $\{w_k\}$ and $\{v_k\}$.

As mentioned in Van Overschee et al. [23], the matrix pair $\{[A], [C]\}$ is assumed to be *observable*, which suggests that all the modes in the system can be observed in the output $\{y_k\}$ and thus can be identified. The matrix pair $\left\{[A], \left[\begin{array}{cc}[B] & \left[Q^{1/2}\right]\end{array}\right]\right\}$ is considered to be *controllable*, which involve that all the modes of the system are excited by either the deterministic input $\{u_k\}$ and/or the stochastic input $\{w_k\}$.

However, this is just an intermediate step to model a pure stochastic system. To achieve this, the excitation $\{u_k\}$ and the noise signals $\{w_k\}$ and $\{v_k\}$ are considered undistinguishable and merged into the state-space model. The properties of the noise components are passed to the stochastic excitation as well. Then, the *discrete-time stochastic state-space model* is defined by:

$$\begin{aligned} \{x_{k+1}\} &= [A]\{x_k\} + \{w_k\} \\ \{y_k\} &= [C]\{x_k\} + \{v_k\} \end{aligned} \tag{2.40}$$

The main properties of linear time invariant stochastic processes, including the non-uniqueness of the state space description is summarized in Van Overschee et al. [23] and will be briefly mentioned in the next paragraphs. The stationarity of the stochastic process implies:

$$\left[\mathbf{E}\left[\{x_k\}\{x_k\}^T\right]\right] = [\Sigma] \qquad \{\mathbf{E}[\{x_k\}]\} = \{0\} \tag{2.41}$$

where the state covariance matrix $[\Sigma]$ is independent of the time instant k. This means that $[A]$ is a stable matrix. Since $\{w_k\}$ and $\{v_k\}$ are zero-mean white noise vector sequences, independent of $\{x_k\}$, it follows that:

$$\left[\mathbf{E}\left[\{x_k\}\{w_k\}^T\right]\right] = [0] \qquad \left[\mathbf{E}\left[\{x_k\}\{v_k\}^T\right]\right] = [0] \tag{2.42}$$

The output covariance matrices $[R_i] \in \mathbb{R}^{l\times l}$ with i as an arbitrary time-lag and the "next-state-output" covariance matrix $[G] \in \mathbb{R}^{n\times l}$ are defined as:

$$[R_i] = \left[\mathbf{E}\left[\{y_{k+i}\}\{y_k\}^T\right]\right] \qquad [G] = \left[\mathbf{E}\left[\{x_{k+1}\}\{y_k\}^T\right]\right] \tag{2.43}$$

From all the previously indicated properties, the following relations can be deduced, as proved in [23] and [25]:

$$[\Sigma] = [A][\Sigma][A]^T + [Q] \qquad [R_0] = [C][\Sigma][C]^T + [R] \qquad [G] = [A][\Sigma][C]^T + [S] \tag{2.44}$$

For $i = 1, 2, \ldots$:

$$[R_i] = [C]\,[A]^{i-1}\,[G] \qquad [R_{-i}] = [G]^T \left([A]^{i-1}\right)^T [C]^T \tag{2.45}$$

In the next sections, it will be shown how particularly Eq. (2.45) plays a key role to estimate the output covariance sequence from the measured data and how the problem can be related to the impulse response deterministic identification.

2.2 Deterministic System Identification

2.2.1 Overview

The deterministic system identification algorithms compute the state-space models (see Section 2.1.2) from known input-output data. There is a wide range of input-output modal identification methods which are briefly reviewed in Cunha et al. [5]. It has been concluded that the time domain methods tend to overcome the limitations of spectral estimates concerning leakage errors and frequency resolution. The most known procedures are grouped into:

- *Indirect methods*: Complex Exponential (CE), Least-Squares Complex Exponential (LSCE), Polyreference Complex Exponential (PRCE), Ibrahim Time Domain (ITD), Eigensystem Realization Algorithm (ERA)
- *Direct methods*: Autoregressive Moving Average (ARMA)

The purpose of this chapter is to introduce one of these algorithms and to point out its connection with the stochastic subspace identification applied later in the context of operational modal analysis. The essential concept is that stochastic realization algorithms are actually extensions of the classical algorithms fed with covariances instead of impulse responses.

The Eigensystem Realization Algorithm (ERA) is considered to offer a high level of robustness in engineering applications since it does not require parametrization. It is based on the realization method of Ho and Kalman [10] and developed by Juang [11] into the ERA method. The description below is based on the main derivations from [11] as presented in [7].

2.2.2 Eigensystem Realization Algorithm

The discrete-time state-space model defined by Eq. (2.30) can be developed setting the initial condition $\{x_0\} = \{0\}$ as:

$$\begin{aligned}
k = 0 \qquad & \{x_0\} = \{0\} \\
& \{y_0\} = [D]\{u_0\} \\
k = 1 \qquad & \{x_1\} = [B]\{u_0\} \\
& \{y_1\} = [C][B]\{u_0\} + [D]\{u_1\} \\
k = 2 \qquad & \{x_2\} = [A][B]\{u_0\} + [B]\{u_1\} \\
& \{y_2\} = [C][A][B]\{u_0\} + [C][B]\{u_1\} + [D]\{u_2\} \\
k = 3 \qquad & \{x_3\} = [A]^2[B]\{u_0\} + [A][B]\{u_1\} + [B]\{u_2\} \\
& \{y_2\} = [C][A]^2[B]\{u_0\} + [C][A][B]\{u_1\} + [C][B]\{u_2\} + [D]\{u_3\} \\
\vdots & \\
k = k \qquad & \{x_k\} = \sum_{i=1}^{k} [A]^{i-1}[B]\{u_{k-1}\} \\
& \{y_k\} = \sum_{i=1}^{k} [C][A]^{i-1}[B]\{u_{k-1}\} + [D]\{u_k\}
\end{aligned} \tag{2.46}$$

The impulse response function (IRF) implies $u_i(0) = 1$ at $k = 0$ and $u_i(k) = 0$ otherwise, and inserted in Eq. (2.46) gives the solution:

$$\{x_k\} = \begin{cases} \{0\}, & k = 0 \\ [A]^{k-1}[B]_{:,i}, & k \geq 0 \end{cases} \qquad \{y_k\} = \begin{cases} [D]_{:,i}, & k = 0 \\ [C][A]^{k-1}[B]_{:,i}, & k \geq 0 \end{cases} \tag{2.47}$$

where the impulse is applied at the i^{th} input. For the system with m inputs and l outputs, two important definitions can be introduced. The so-called *controllability matrix* $[\Gamma_q] \in \mathbb{R}^{n\times(m\times q)}$ and *observability matrix* $[O_q] \in \mathbb{R}^{n\times(l\times q)}$ of order q are determined by:

$$[\Gamma_q] = \begin{bmatrix} [B] & [A][B] & \dots & [A]^{q-1}[B] \end{bmatrix} \qquad [O_q] = \begin{bmatrix} [C] \\ [C][A] \\ \vdots \\ [C][A]^{q-1} \end{bmatrix} \tag{2.48}$$

where $q \geq \max(m, l)$. Matrix $[\Gamma_q]$ is nothing else but the matrix format of the first expression in Eq. (2.47) and matrix $[O_q]$ results if the excitation part is completely dismissed in Eq. (2.46) ($\{u_k\} = \{0\}$). The two properties have been already mentioned in Section 2.1.2 where the stochastic model was introduced. Their benefit can be practically achieved if the dimensions $m \times q$ and $l \times q$ are greater than the system order n to such a degree that the identification errors can be minimized.

The second expression in Eq. (2.47) gives the so-called Markov parameters or matrices of a system. They play a key role in the ERA algorithm since for a time instant k they

contain the impulse response (the sequence $[D], [C][B], [C][A][B], [C][A]^2[B], \ldots$) at a certain position on its corresponding column, so it can usually be directly measured or obtained from input-output records. For $k \geq 0$ the Markov matrices $[h_k] \in \mathbb{R}^{m \times l}$ are defined as:

$$[h_k] = [C][A]^k[B] \tag{2.49}$$

and will be used to assemble Hankel matrices. Taking two Hankel matrices, $[H_1]$ and a shifted $[H_2] \in \mathbb{R}^{(m\times q)\times(l\times q)}$:

$$[H_1] = \begin{bmatrix} [h_1] & [h_2] & \dots & [h_q] \\ [h_2] & [h_3] & \dots & [h_{q+1}] \\ \vdots & \vdots & \ddots & \vdots \\ [h_q] & [h_{q+1}] & \dots & [h_{2q-1}] \end{bmatrix} \qquad [H_2] = \begin{bmatrix} [h_2] & [h_3] & \dots & [h_{q+1}] \\ [h_3] & [h_4] & \dots & [h_{q+2}] \\ \vdots & \vdots & \ddots & \vdots \\ [h_{q+1}] & [h_{q+2}] & \dots & [h_{2q}] \end{bmatrix}$$

they can be assembled using the controllability and observability matrices:

$$[H_1] = [O_q][\Gamma_q] \qquad [H_2] = [O_q][A][\Gamma_q] \tag{2.50}$$

These relations can be easily proven by doing the multiplication. However, only the Hankel matrices are known from the measured input responses, and taking into account that the decomposition is not unique, it is possible to apply another factorization, for example Cholesky, LU or QR. The singular value decomposition (SVD) leads to:

$$[H_1] = [U][S][V]^T \tag{2.51}$$

where

$$[S] = \begin{bmatrix} \begin{bmatrix} \diagdown & & \\ & \sigma_n & \\ & & \diagdown \end{bmatrix} & [0] \\ [0] & [0] \end{bmatrix} \qquad [U][U]^T = [I] \qquad [V]^T[V] = [I] \tag{2.52}$$

The SVD directly estimates the rank of a matrix as the number of non-zero singular values. This information also serves to determine the unknown matrices from the first decomposition in Eq. (2.50) as follows:

$$[O_q] = [U][S]^{1/2} \qquad [\Gamma_q] = [S]^{1/2}[V]^T \tag{2.53}$$

Extracting $[A]$ from $[H_2]$ in Eq. (2.50) and introducing Eq. (2.53) allows to find the state matrix as:

$$[A] = [O_q]^\dagger [H_2][\Gamma_q]^\dagger = [S]^{-1/2}[U]^T[H_2][V][S]^{-1/2} \tag{2.54}$$

where $[\bullet]^\dagger$ represents the pseudo-inverse of a matrix, also known as Moore-Penrose pseudo-inverse. The input matrix $[B]$ complies to the form of the controllability matrix $[\Gamma_q]$ and is obtained from its first m columns:

$$[B] = [\Gamma_q][E_m] = [S]^{1/2}[V]^T[E_m] \tag{2.55}$$

and similarly, $[C]$ from the first l rows of the observability matrix $[O_q]$:

$$[C] = [E_l]^T[P] = [E_l][U][S]^{1/2} \tag{2.56}$$

where $[E_m] = \begin{bmatrix} [I_m] & [0] & \dots & [0] \end{bmatrix}^T$ and $[E_l] = \begin{bmatrix} [I_l] & [0] & \dots & [0] \end{bmatrix}^T$.

At this moment, the natural frequencies and mode shapes of the system can be identified using the already introduced eigenvalue problem $[A] = [\Psi][\Lambda_d][\Psi]^{-1}$ (2.33) and the eigenvector problem $[V] = [C][\Psi]$ (2.36).

Moreover, the similarity between the covariances in Eq. (2.45) and the measured impulse response in Eq. (2.47) is obvious and it can be concluded that the matrices $([A], [B], [C], [D])$ are for deterministic models equivalent to the matrices $([A], [G], [C], [R_0])$ for stochastic models. Therefore, the factorization procedures described in this section can be applied also for the output-only covariance-driven methods.

2.3 Stochastic System Identification

2.3.1 Overview

The stochastic system identification algorithm solves the stochastic state-space models (see Section 2.1.2) from the recorded output data. The main assumption is that the input is a realization of a stochastic process (zero-mean Gaussian white noise) which replaces the deterministic knowledge of the excitation. The techniques can be grouped according to their domain of application [5] into:

- *Frequency domain (non-parametric) methods*: Peak-Picking (PP), Frequency Domain Decomposition (FDD), Enhanced Frequency Domain Decomposition (EFDD)
- *Time domain (parametric) methods*: Reference-based Stochastic Subspace Identificaion (SSI/ref), Multiple Reference Ibrahim Time Domain (MRITD), Least-Squares Complex Exponential (LSCE), Ibrahim Time Domain (ITD)

It is noteworthy to remark that most of the output-only methods stem from the classical input-output methods and adapt the frequency or impulse response functions to stochastic inputs. It is also possible to divide the methods according to the type of data that they require: raw time data, covariances or spectra as presented in [15].

The SSI algorithms have proved to be powerful, robust and reliable methods. The novel approach called reference-based stochastic subspace identification (SSI/ref) introduced by Peeters et al. [16] reformulated the classical algorithm so that it uses the covariances between the outputs and only a limited set of reference outputs instead of the covariances between all outputs. Since it reduces the dimension of the problem, this added to the efficiency with beneficial effects on the computational costs. Currently, there are two implementations of SSI/ref with similar accuracy: covariance-driven (SSI-cov/ref) and data-driven (SSI-data/ref). The SSI-data/ref utilizes a QR decomposition of the data Hankel matrix and allows a direct application to the response time series without passing through the estimates of their covariance functions. Although SSI-data/ref presents some advantages over the SSI-cov/ref, it is considered to be slower [15] and its application is more demanding. Therefore, only the SSI-cov/ref was selected for the study case of this thesis and the algorithm will be explained in the next section.

2.3.2 Covariance-driven Stochastic Subspace Identification

The procedure starts with the discrete-time stochastic state-space model defined by Eq. (2.40). The properties of the noise process were already discussed in Section 2.1.2 for the stochastic excitation. The main interest here is to manipulate the output vector $\{y_k\} \in \mathbb{R}^{l\times 1}, k = 0, \ldots, N-1$ in order to extract the relevant modal information. The solution presented here is the modified version of the SSI algorithm as proposed in [16] and [15]. The advanced concepts of Kalman filter states and projections will not be discussed here.

It is typical for ambient measurement campaigns to divide the locations of interest into several setups with overlapping sensors due to physical limitations. The reference outputs are considered those sensors common for every setup and well-placed on the structure, so that they carry information about all the existing modes of vibration. However, when only one setup is available, the use of reference sensors increases the speed of the algorithms by decreasing the redundancy.

The first step is to arrange the response vector such that the first r rows contain the reference signals:

$$\{y_k\} = \begin{bmatrix} \left\{y_k^{ref}\right\} \\ \left\{y_k^{-ref}\right\} \end{bmatrix} \tag{2.57}$$

where $\left\{y_k^{ref}\right\} \in \mathbb{R}^{r\times 1}$ are the reference outputs and $\left\{y_k^{-ref}\right\} \in \mathbb{R}^{(l-r)\times 1}$ are the others. In comparison to the impulse response case, the Hankel matrix $\left[H^{ref}\right] \in \mathbb{R}^{(r+l)i\times N}$ is composed from the white noise input and can be divided into a past reference and a future part. If two equal blocks of length i are taken:

$$\left[H^{ref}\right] = \frac{1}{\sqrt{N}} \left[\begin{array}{cccc} \left\{y_0^{ref}\right\} & \left\{y_1^{ref}\right\} & \cdots & \left\{y_{N-1}^{ref}\right\} \\ \left\{y_1^{ref}\right\} & \left\{y_2^{ref}\right\} & \cdots & \left\{y_N^{ref}\right\} \\ \vdots & \vdots & \ddots & \vdots \\ \left\{y_{i-1}^{ref}\right\} & \left\{y_i^{ref}\right\} & \cdots & \left\{y_{i+N-2}^{ref}\right\} \\ \hline \left\{y_i^{ref}\right\} & \left\{y_{i+1}^{ref}\right\} & \cdots & \left\{y_{i+N-1}^{ref}\right\} \\ \left\{y_{i+1}^{ref}\right\} & \left\{y_{i+2}^{ref}\right\} & \cdots & \left\{y_{i+N}^{ref}\right\} \\ \vdots & \vdots & \ddots & \vdots \\ \left\{y_{2i-1}^{ref}\right\} & \left\{y_{2i}^{ref}\right\} & \cdots & \left\{y_{2i+N-2}^{ref}\right\} \end{array}\right] = \left[\begin{array}{c} \left[Y_{0|i-1}^{ref}\right] \\ \hline \left[Y_{i|2i-1}\right] \end{array}\right] = \left[\begin{array}{c} \left[Y_p^{ref}\right] \\ \hline \left[Y_f\right] \end{array}\right] \tag{2.58}$$

The data is scaled by a factor $1/\sqrt{N}$. The indices p and f stand for past and future, respectively. The subscripts of the matrices $\left[Y_{0|i-1}\right]$ and $\left[Y_{i|2i-1}\right]$ indicate the first and the last element in the first column of a block Hankel matrix.

The next step is to calculate the reduced covariance matrices between all the outputs and the references which are exactly the first r columns of the full covariance matrices:

$$\left[R_i^{ref}\right] = \left[\mathbf{E}\left[\{y_{k+i}\}\left\{y_k^{ref}\right\}^T\right]\right] \in \mathbb{R}^{l\times r} \tag{2.59}$$

Similarly to Eq. (2.43), the *reduced "next state-output" covariance matrix* G^{ref} results as:

$$\left[G^{ref}\right] = \left[\mathbf{E}\left[\{x_{k+1}\}\left\{y_k^{ref}\right\}^T\right]\right] \in \mathbb{R}^{n\times r} \tag{2.60}$$

Developing the relations, the equivalent factorization as for the full covariances case introduced in Eq. (2.45) leads to:

$$\left[R_i^{ref}\right] = [C]\,[A]^{i-1}\,[G] \qquad \left[R_i^{ref}\right]^T = \left[G^{ref}\right]^T\left([A]^{i-1}\right)^T[C]^T \qquad i = 1,2,\ldots \tag{2.61}$$

The main element for further computations is the block Toeplitz matrix $\left[T_{1|i}^{ref}\right] \in \mathbb{R}^{li\times ri}$ which is created from the obtained covariance matrices:

$$\left[T_{1|i}^{ref}\right] = \left[Y_f\right]\left[Y_p^{ref}\right]^T = \begin{bmatrix} \left[R_i^{ref}\right] & \left[R_{i-1}^{ref}\right] & \ldots & \left[R_1^{ref}\right] \\ \left[R_{i+1}^{ref}\right] & \left[R_i^{ref}\right] & \ldots & \left[R_2^{ref}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \left[R_{2i-1}^{ref}\right] & \left[R_{2i-2}^{ref}\right] & \ldots & \left[R_i^{ref}\right] \end{bmatrix} \tag{2.62}$$

Using Eq. (2.61) this can be expressed as:

$$\left[T_{1|i}^{ref}\right] = \underbrace{\begin{bmatrix} [C] \\ [C]\,[A] \\ \vdots \\ [C]\,[A]^{i-1} \end{bmatrix}}_{[O_i]\in\mathbb{R}^{li\times n}} \underbrace{\begin{bmatrix} [A]^{i-1}\left[G^{ref}\right] & \ldots & [A]\left[G^{ref}\right] & \left[G^{ref}\right] \end{bmatrix}}_{\left[\Gamma_i^{ref}\right]\in\mathbb{R}^{n\times ri}} \tag{2.63}$$

The obtained result $\left[T_{1|i}^{ref}\right] = \left[O_i\right]\left[\Gamma_i^{ref}\right]$ resembles what has been achieved in Eq. (2.50). The matrix $[O_i]$ is called the *extended observability matrix* and $\left[\Gamma_i^{ref}\right]$ is the *extended controllability matrix*. From now on, it is straightforward to carry out the same procedure as described for the impulse response in Section 2.2.2. After applying the SVD it follows:

$$\left[T_{1|i}^{ref}\right] = [U]\,[S]\,[V]^T = \begin{bmatrix} [U_1] & [U_2] \end{bmatrix}\begin{bmatrix} [S_1] & [0] \\ [0] & [0] \end{bmatrix}\begin{bmatrix} [V_1]^T \\ [V_2]^T \end{bmatrix} = [U_1]\,[S_1]\,[V_1]^T \tag{2.64}$$

The number of the non-zero singular values which are to be found on the diagonal of $[S_1]$ gives again the rank of the decomposed matrix. The matrices $[U]$ and $[V]$ are explicitly

expressed here by their components. The extended observability and controllability matrices can also be identified by splitting Eq. (2.64):

$$[O_i] = [U_1]\,[S_1]^{1/2} \qquad \left[\Gamma_i^{ref}\right] = [S_1]^{1/2}\,[V_1]^T \tag{2.65}$$

Analogous to Section 2.2.2, considering a shifted block Toeplitz matrix $\left[T_{2|i+1}^{ref}\right]$ with a time span from 2 to $2i$:

$$\left[T_{2|i+1}^{ref}\right] = [O_i]\,[A]\,\left[\Gamma_i^{ref}\right] \tag{2.66}$$

allows to compute the state matrix:

$$[A] = [S_1]^{-1/2}\,[U_1]^T\,\left[T_{2|i+1}^{ref}\right]\,[V_1]\,[S_1]^{-1/2} \tag{2.67}$$

The matrix $[C]$ is obtained after extracting the first l rows from $[O_i]$ and the matrix $\left[G^{ref}\right]$ is achieved from the last r columns of $[\Gamma]_i^{ref}$ (see Eq. (2.55) and Eq. (2.56) in Section 2.2.2). Knowing $[A]$ and $[C]$, the modal decomposition described by Eq. (2.33) and Eq. (2.36) offers the solution of the identification problem. To sum up the results, the eigenvalues come up as:

$$\lambda_k = \frac{\mu_k}{\Delta t} \qquad f_k = \frac{|\lambda_k|}{2\pi} \qquad \xi_k = -\frac{\mathrm{Re}\,(\lambda_k)}{|\lambda_k|} \tag{2.68}$$

and the mode shapes which stem from matrix $[C]$ occur in conjugate pairs. Therefore, a state-space model of order n provides modal parameters for $n/2$ modes and from this perspective, the reason why the equality $n = 2 \times n_2$ was introduced for the state-vector in Section 2.1.1 can be better understood.

Now, that the system identification theory for the SSI-cov/ref has been clarified, it still remains to discuss how to identify the system order n and the time-lag i. If an automated modal analysis is sought, this topic becomes even more complex and deserves a stand-alone examination which will be carried out in the next chapter.

Chapter 3

Modal Parameter Estimation

The key aspects which will be discussed in this chapter refer to the *modal parameter estimation* (MPE) which involves the methodology to identify and select proper dynamic characteristics of a system after the identification procedure has been chosen (in this case, SSI-cov/ref). In the first section, the well-known approach which involves a lot of user interaction is described and the main criteria to extract the modal parameters are highlighted. The second section considers the challenging task of automated operational modal analysis and includes a brief description of the most popular clustering algorithms which originate from data mining theory as well as some approaches applied to experimental measurements as proposed in literature.

3.1 Manual Interpretation

The main issue in system identification techniques is to determine the model order n. This corresponds to the expected number of eigenvalues, twice the number of eigenfrequencies, achieved by a physical insight of the system. Although there are several procedures to find a precise n, the main concern in modal analysis is not a good model per se, but rather the physical relevance of each individual mode.

The experimental nature of the problem implies the existence of noise in the acquired data, a limited number of samples or possible non-linear behaviour of the tested structure [12]. Consequently, the derived state-space model matrices and the resulted modal parameters should be viewed as estimates. According to the theory (see Section 2.3.2), the number of non-zero singular values gives the system order and can be visually inspected as in Fig. 3.1. It can be seen that the singular values can drop significantly at some point, which can be interpreted to correspond to the order n. However, this is not always the case and it is also possible that the singular values can never reach exactly zero and sometimes not even an obvious gap can be distinguished.

The results of the SSI-cov/ref is a set of so-called *poles* which include information about the modal parameters. The common practice for an appropriate selection, is to over-specify the system order n, based on the empirical observation that the physical poles appear nearly at the same natural frequency, while the mathematical or spurious poles are randomly distributed. Despite the occurrence of these spurious modes, an increased order allows to identify also weakly excited modes.

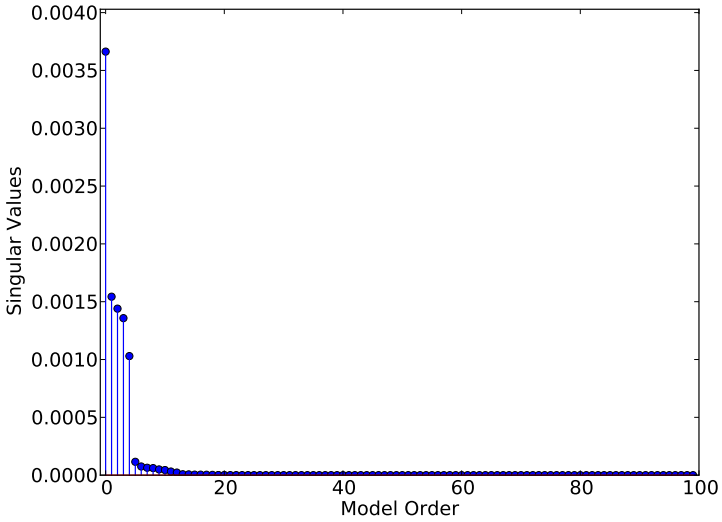

Figure 3.1: Singular values from SSI-cov/ref

The fundamental outcome is a frequency versus model order chart, where the physical modes tend to stabilize and appear in a column-like shape. Hence, the resulting plot is known as *stabilization diagram* and plays a vital role in modal parameter estimation. It is common to introduce certain stability requirements which are meant to discriminate between physical and spurious poles. If a certain mode between two consecutive model orders satisfy some pre-defined limits for its frequencies, damping ratios and mode shapes distance measures, it is considered to be a stable pole. The conditions are defined as:

$$\begin{aligned} d\left(f_n, f_{n+1}\right) &= \frac{\left|f_n - f_{n+1}\right|}{\max\left(\left|f_n\right|, \left|f_{n+1}\right|\right)} \times 100\% < lim_f\% \\ d\left(\xi_n, \xi_{n+1}\right) &= \frac{\left|\xi_n - \xi_{n+1}\right|}{\max\left(\left|\xi_n\right|, \left|\xi_{n+1}\right|\right)} \times 100\% < lim_\xi\% \\ MAC\left(\{\phi_n\}, \{\phi_{n-1}\}\right) &\times 100\% > lim_{MAC}\% \end{aligned} \tag{3.1}$$

where the *modal assurance criterion* (MAC) is introduced and shows the level of correlation between two mode shapes:

$$MAC\left(\{\phi_n\}, \{\phi_{n-1}\}\right) = \frac{\left(\{\phi_n\}^T \{\phi_{n-1}\}\right)^2}{\left(\{\phi_n\}^T \{\phi_n\}\right)\left(\{\phi_{n-1}\}^T \{\phi_{n-1}\}\right)} \tag{3.2}$$

The MAC values lie between 0 and 1, with 1 indicating fully consistent mode shapes. Some popular limits for the indicators are:

$$lim_f = 1\% \qquad lim_\xi = 5\% \qquad lim_{MAC} = 98\% \tag{3.3}$$

Fig. 3.2 and Fig. 3.3 show typical stabilization diagrams for an ambient vibration record using different time-lags i, where poles stable in frequency, damping or mode shape were also considered. The block rows/columns of the Toeplitz matrix are taken as 50 and 100, respectively, and they show the sensitivity of the results to this parameter. This has an influence on the matrix dimension and affects the quality of the model.

As reported by Cabboi et al. [2], an adequate choice of i has an impact on a stable damping identification and can also lead to the occurrence of the *modal splitting*

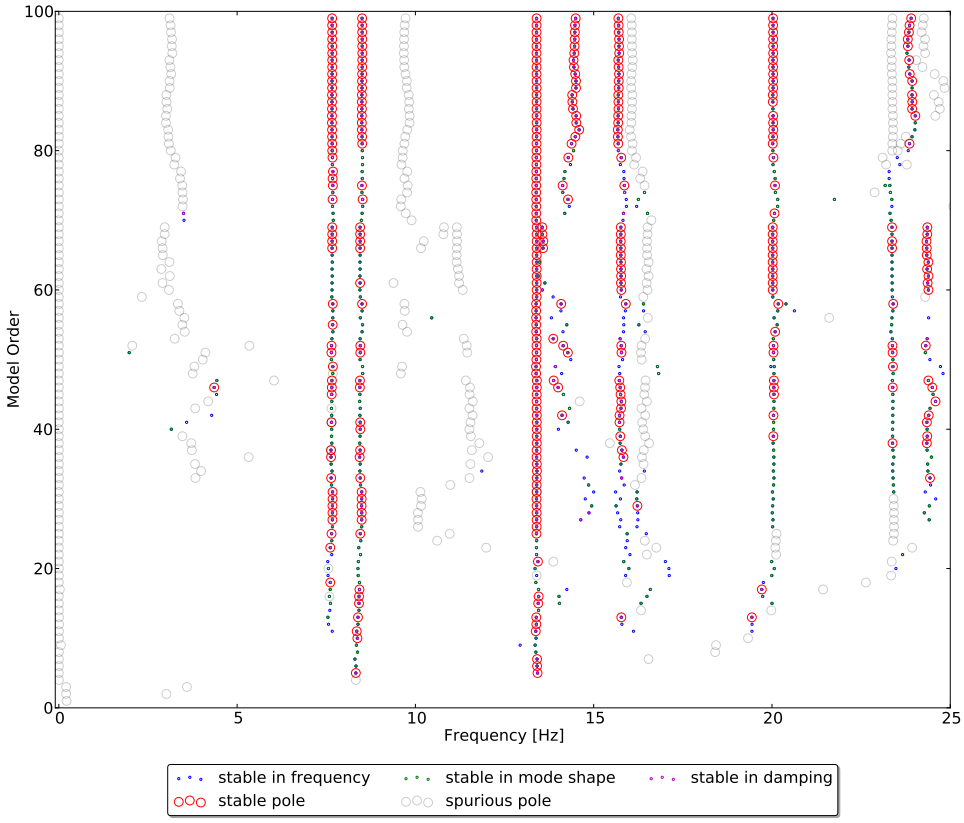

Figure 3.2: Stabilization diagram, time-lag $i = 50$

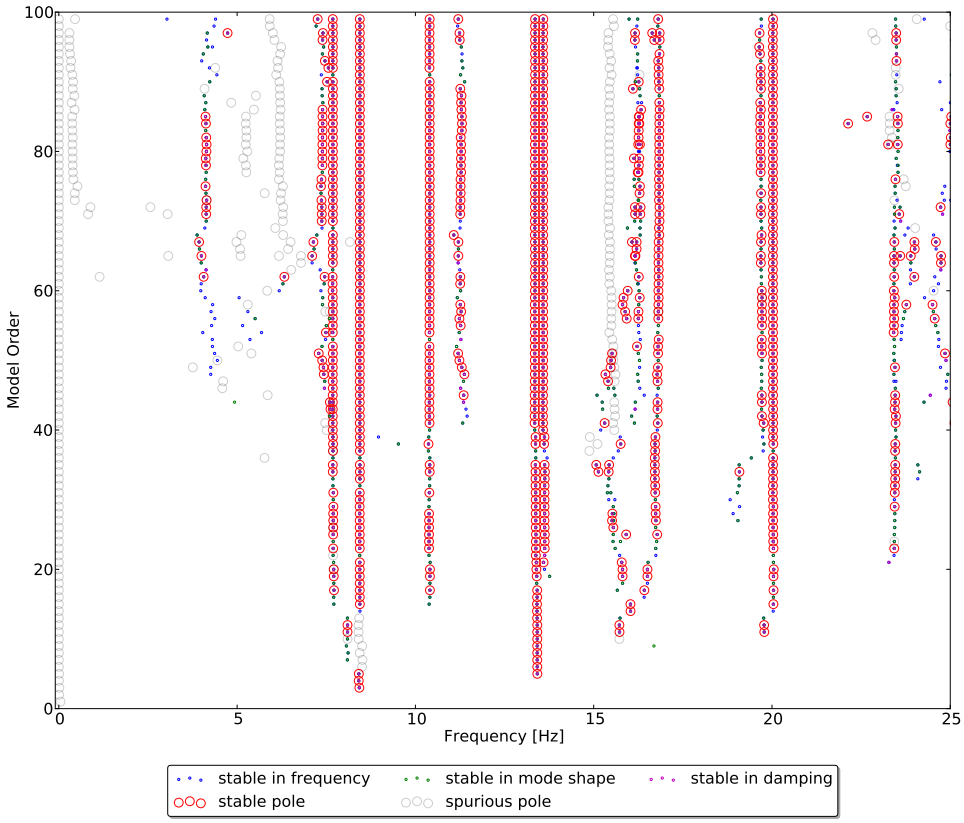

Figure 3.3: Stabilization diagram, time-lag $i = 100$

phenomenon, a subject which is still under research and deserves a lot of attention. An evidence of this can be seen in Fig. 3.3 around 14 Hz. Thus, a higher value of i can better identify close-spaced modes, but can also introduce additional spurious poles due to the noise inserted in the output covariance function with little structural information.

At first glance, the diagrams suggests that the vertical alignments which appear for some of the frequencies can be interpreted as possible physical modes. Nevertheless, a more detailed inspection is mandatory and the subjective judgement plays here a decisive role.

The restrictions given by Eq. (3.1) are usually adapted according to the quality of the data and the level of conservatism. As already mentioned, a stable pole satisfies all three criteria, but this does not mean that it necessarily offers the best level of confidence. Sometimes, poles that show stability only in one or a couple of these requirements can slightly exceed the bounds, but prove to be more appropriate.

To help even an experienced user, it is common to include more information to reflect the quality of the results. For example, solutions with negative or excessively high damping values can be eliminated. Among the available definitions existing in literature, three deterministic validation criteria were considered in the thesis to be representative for the mode shape complexity: *modal phase collinearity* (MPC), *mean phase* (MP) and *mean phase deviation* (MPD) and were derived by Heylen et. al [9]. These values were also adopted in the powerful MATLAB toolbox developed at Katholieke Universiteit Leuven, MACEC 3.2 [19].

The MPC is a correlation index which evaluates whether the real and imaginary parts of a mode shape lie on a straight line in the complex plane and is assigned as:

$$MPC\left(\{\phi_i\}\right) = \frac{\left\|\mathrm{Re}\left(\{\tilde{\phi}_i\}\right)\right\|_2^2}{\left\|\mathrm{Re}\left(\{\tilde{\phi}_n\}\right)\right\|_2^2 + \left\|\mathrm{Im}\left(\{\tilde{\phi}_i\}\right)\right\|_2^2} + \frac{\frac{1}{\epsilon_{MPC}}\mathrm{Re}\left(\{\tilde{\phi}_i\}\right)^T \mathrm{Im}\left(\{\tilde{\phi}_i\}\right)\left(2\left(\epsilon_{MPC}^2+1\right)\sin^2\left(\Theta_{MPC}\right)-1\right)}{\left\|\mathrm{Re}\left(\{\tilde{\phi}_i\}\right)\right\|_2^2 + \left\|\mathrm{Im}\left(\{\tilde{\phi}_i\}\right)\right\|_2^2} \tag{3.4}$$

where $\{\phi_i\} \in \mathbb{C}^{n\times 1}$ represents a single mode shape and

$$\{\tilde{\phi}_i\} = \{\phi_{io}\} - \frac{\sum_{o=1}^{n}\{\phi_{io}\}}{n} \qquad \epsilon_{MPC} = \frac{\left\|\mathrm{Re}\left(\{\tilde{\phi}_i\}\right)\right\|_2^2 - \left\|\mathrm{Im}\left(\{\tilde{\phi}_i\}\right)\right\|_2^2}{2\,\mathrm{Re}\left(\{\tilde{\phi}_i\}\right)^T \mathrm{Im}\left(\{\tilde{\phi}_i\}\right)}$$

$$\Theta_{MPC} = \arctan\left(|\epsilon_{MPC}| + \mathrm{sign}\left(\epsilon_{MPC}\right)\sqrt{1+\epsilon_{MPC}^2}\right)$$

The values range from 0 (not collinear at all) to 1 (perfect collinearity or mono-phase behaviour). For proportionally damped structures, real modes with high MPC index are expected. A mode with a low index is rather complex, indicating a computational or noisy mode. However, this is common for highly damped structures and should be cautiously interpreted. As explained by Vacher et. al [22], MPC is nothing more than the MAC computed between a vector and its conjugate so it can be alternatively calculated as:

$$MPC\left(\{\phi_i\}\right) = MAC\left(\{\phi_i\}, \{\phi_i\}^*\right) \tag{3.5}$$

The MP is the angle obtained after a straight line fitting through the mode shape components in the complex plane, such as the orthogonal distance is minimized:

$$MP\left(\{\phi_i\}\right) = \arctan\left(\frac{-\left[V\right]_{1,2}}{\left[V\right]_{2,2}}\right) \tag{3.6}$$

where $[U] \in \mathbb{R}^{n\times 2}$, $[S] \in \mathbb{R}^{2\times 2}$ and $[V] \in \mathbb{R}^{2\times 2}$ emerge from the SVD:

$$[U]\,[S]\,[V]^T = \left[\begin{array}{cc} \mathrm{Re}\left\{\phi_i\right\} & \mathrm{Im}\left\{\phi_i\right\} \end{array}\right]$$

The MPD is a statistical indicator that measures the phase scatter of an eigenvector and can be seen as the weighted mean deviation of the vector components from the MP. It is defined as:

$$MPD\left(\{\phi_i\}\right) = \frac{\sum_{o=1}^{n} w_o \arccos\left|\dfrac{\mathrm{Re}\left\{\phi_{io}\right\}\left[V\right]_{2,2} - \mathrm{Im}\left\{\phi_{io}\right\}\left[V\right]_{1,2}}{\left|\{\phi_{io}\}\right|\sqrt{\left[V\right]_{1,2}^2 + \left[V\right]_{2,2}^2}}\right|}{\sum_{o=1}^{n} w_o} \tag{3.7}$$

where $\sum_{o=1}^{n} w_o \neq 0$.

Next to their useful information for a manual interpretation of the diagrams, these validation criteria are indispensable also when developing an automated selection algorithm. This aspect will be highlighted in the following section, where the indices will be used to automatically select a validated set of modal parameters.

3.2 Automatic Interpretation

Given the amount of user interaction necessary to interpret the stabilization diagrams, a lot of effort has been lately invested to automatize the MPE. The parametric system identification techniques (e.g. SSI-cov/ref) are considered to be more accurate than non-parametric estimates [18], but they also involve a lot of time spent for fine-tuning the input parameters and post-process the data. Therefore, especially for vibration-based structural health monitoring (SHM), it is not feasible to inspect every single record from a huge database and an automated or semi-automated procedure is crucial.

Since many of the proposed approaches rely heavily on clustering algorithms, this section will firstly introduce the most popular clustering methods and then it will briefly review some of their applications to dynamic monitoring systems. The chapter will conclude with the selected routine employed for the case study of an existing monitoring system investigated in Chapter 4.

3.2.1 Introduction to Cluster Analysis

Clustering denotes the process of dividing data into groups called *clusters* and assigning each individual object to these groups in such a way that the homogeneity within a group and the heterogeneity between groups are maximized. In other words, the data objects

within a cluster are very similar to each other, while they are strongly dissimilar to the objects in other clusters. The clustering concept emerges from various fields like data mining, statistics, machine learning and biology and has been widely implemented in numerous applications, including image processing, market research, data analysis and pattern recognition.

The presentation of the clustering approaches in this chapter are based on the descriptions given by Han and Kamber [8] and Tan et al. [21].

The categories which subdivide clustering methods can be organized by means of different criteria. For example, one possible classification is based on whether the clusters contain sub-clusters or not. A clustering where the data objects belong to exactly one subset is termed as *partitional/unnested*, whereas if the clusters include sub-clusters, they are called *hierarchical/nested*.

Another distinction can be made concerning the number of clusters assigned to each object. This discriminates between *exclusive* clusters where each object belongs to a single cluster, *overlapping/non-exclusive* clusters where an object can belong to more than one group and *fuzzy* clusters where every object is located in every cluster with a membership weight between 0 and 1. Fuzzy clustering stems from the mathematical theory of fuzzy sets and is closely related to probabilistic clustering techniques which compute the probability that each point is contained in each class, where the probabilities sum to 1.

Moreover, from the perspective of outliers (observation points distant from other observations), one can distinguish between *complete* clustering where every object is assessed to a cluster and *partial* clustering where some objects might not belong to a well-defined group.

Before choosing a specific approach and understanding the differences between the available methods, one should keep in mind that there are several ways to cluster the same data set as it is shown in Fig. 3.4.

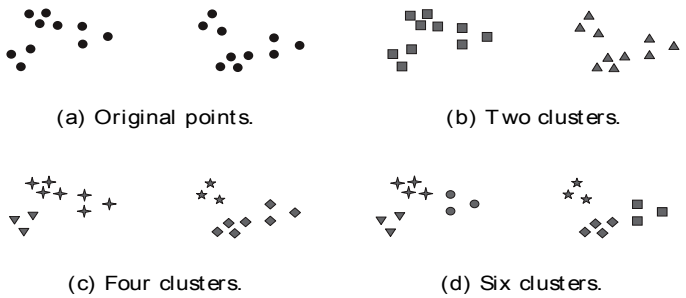

Figure 3.4: Different possibilities to cluster the same set of points [21]

The major clustering algorithms are divided into *partitioning methods, hierarchical methods, density-based methods, grid-based methods, model-based methods* and others. Only two of the oldest algorithms were selected to be applied for the dynamic data analysis and will be shortly explained in the next paragraphs.

K-Means

K-means is a prototype-based partitioning method which aims at assembling K partitions or clusters from a data set. It starts with a user defined K number of initial clusters represented by their *centroid* or *center of gravity*. Using an iterative relocation technique, it tries to regroup the members to the closest centroids that update their position accordingly, until there is not any point left to change the clusters and the centroids stop moving. This condition is expressed by an objective function that minimizes the proximity measure of each point x to the nearest cluster centroid c_i. The typical proximity measure is the Euclidean distance and the criterion used is the *sum of the squared error* (SSE):

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} \|c_i, x\|_2^2 \tag{3.8}$$

The procedure attempts to make the obtained clusters as compact and distinct as possible. Fig. 3.5 illustrates the clustering steps of the K-means algorithm.

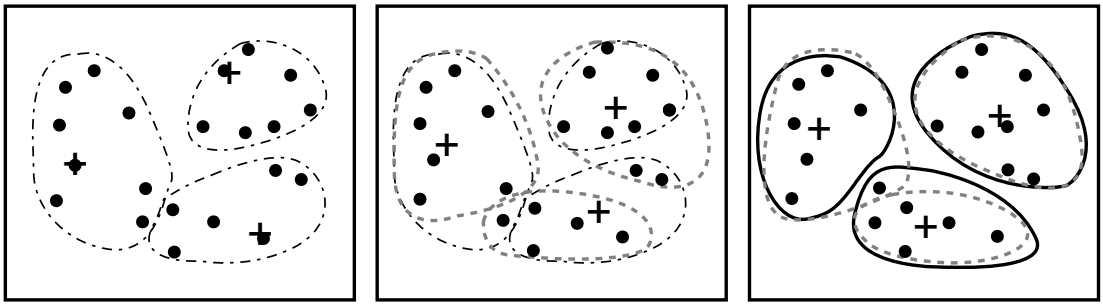

Figure 3.5: Clustering steps using the K-means algorithm ("+" stands for each centroid) [8]

It can be seen that after the initialization of the three cluster centroids marked by a "+" symbol in Fig. 3.5(a), the objects are grouped according to the nearest center. Afterwards, the centroids are recalculated and new clusters drawn by dashed curves are obtained in Fig. 3.5(b). The centroids move until the final iteration when the members are again reassigned leading to Fig. 3.5(c).

The strength of K-means resides in its simplicity and easy implementation as well as in its intuitive objective function that can be seen as an optimization of the intracluster similarity. It is also relatively efficient in processing large data sets due to its low complexity $O(Ktn)$, where n is the total number of points, t the number of iterations and K the number of clusters [8]. Normally, K and t are much lower than n.

However, the main drawback of the method is the requirement to specify in advance a certain K number of clusters and to define the initial position of the centroids. A proper initialization is actually a key step of the basic K-means procedure and a poor selection can lead to a local minimum, thus a suboptimal clustering. There are several approaches described in [21] to treat this problem. Another aspect is the impossibility to deal with categorical data because the mean values should be definable. Moreover, as mentioned in [8], nonconvex shapes detection creates difficulties and the method is sensitive to outliers since they have a strong impact on the mean value.

Agglomerative and Divisive Hierarchical

Agglomerative hierarchical is a widespread classical hierarchical clustering method which assigns each object to a cluster and then merges each group step by step until all the objects belong to a single cluster. The reversed procedure is called *divisive hierarchical* and starts with all objects placed into one cluster which is subsequently split until each object is contained in its own cluster. The first approach is a so-called bottom-up strategy, whereas the second one is known as a top-down strategy. Both of them accept a termination condition which interrupts the algorithm earlier.

Although the agglomerative and divisive hierarchical methods are two sides of the same coin, the first method is by far the most common. The procedure applied to a set of five data points is depicted in Fig. 3.6. The acronyms AGNES and DIANA stand for *agglomerative nesting* and *divisive analysis*, respectively.

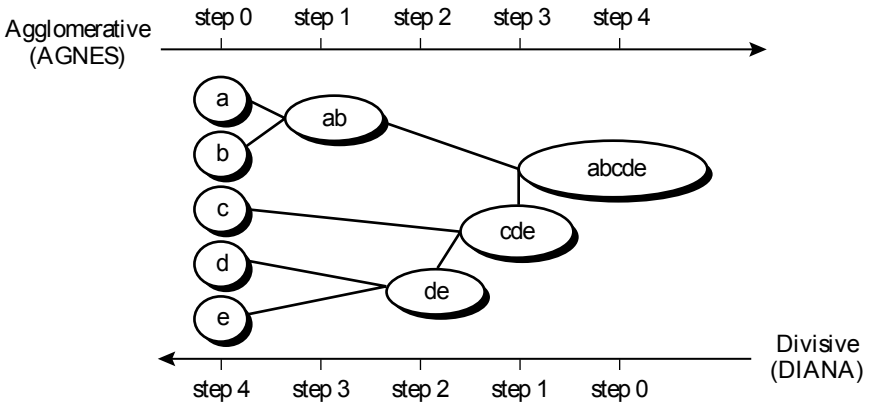

Figure 3.6: Agglomerative and divisive hierarchical clustering on the set $\{a, b, c, d, e\}$ [8]

The core of the algorithm is the proximity or (dis)similarity matrix which stores pairwise distances between all the points in a data set. Due to its symmetrical structure, only the upper or lower triangular parts are usually kept, thus the matrix has the following format:

$$\begin{bmatrix} 0 & & & & \\ d\,(2,1) & 0 & & & \\ d\,(3,1) & d\,(3,2) & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ d\,(n,1) & d\,(n,2) & \cdots & \cdots & 0 \end{bmatrix} \tag{3.9}$$

where $d\,(i,j)$ is the distance measure between the members i and j and the most important types are:

$$\begin{aligned} \textit{Minimum distance}: \quad & d_{min} = \min_{x \in C_i, x' \in C_j} d_{i,j} \\ \textit{Maximum distance}: \quad & d_{min} = \max_{x \in C_i, x' \in C_j} d_{i,j} \\ \textit{Average distance}: \quad & d_{min} = \frac{1}{n_i n_j} \sum_{x \in C_i} \sum_{x' \in C_j} d_{i,j} \end{aligned}$$

with x and x' points belonging to two different clusters and n_i and n_j the number of elements in the clusters C_i and C_j.

If the inter-cluster similarity uses the minimum distance, it is called *single linkage* and computes the similarity of the closest pair. For the maximum distance, it is referred to as *complete linkage* which is the similarity between the furthest pair. And finally, for the average distance, the concept *average linkage* denotes the average similarity between groups. Each of these approaches has its own pros and cons and they should be chosen according to the data quality.

The clustering process has a specific representation known as *dendrogram* and exemplified in Fig. 3.7 for the same set of data as in Fig. 3.6. It is typical only for this type of algorithm and provides a visual interpretation of the procedure which is also a reason for its popularity.

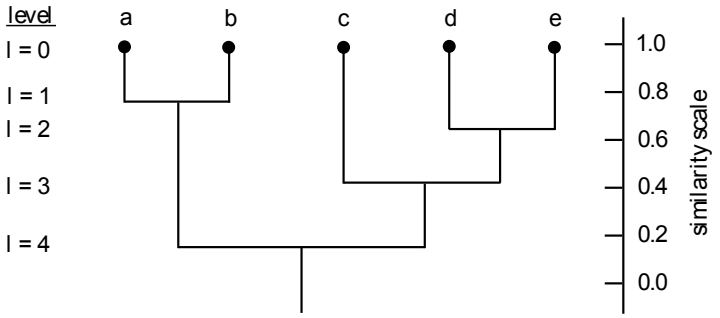

Figure 3.7: Dendrogram represenation for the set $\{a, b, c, d, e\}$ [8]

The dendrogram consists of U-shaped lines connecting the members in a hierarchical tree. The tree analogy is also suggested by referring to the initial single object clusters as *leaves*. Each node is merged for a specific distance value displayed on the vertical axis. A cut-off threshold value can be defined at which the merging process (in case of agglomerative hierarchical clustering) can be stopped.

Although the hierarchical clustering method is fairly simple, it does not involve the optimization of an objective function. The decision to merge or split two clusters is irreversible and if it turns out later to have been a poor selection, it is impossible to be undone. This can also be advantageous because it prevents the drawback of the local optimization criterion from which the K-means method suffers. In addition to this, this method does not rely on any initialization of the starting points.

The time complexity $O\left(n^2 \log n\right)$ and the space complexity $O\left(n^2\right)$ make the algorithm computationally expensive and not appropriate for very large data sets.

To overcome the difficulties of both K-means and (agglomerative) hierarchical clustering and take full advantage of their strengths, it is common to apply the two methods successively. The algorithms can also be enhanced or more complex methods can be employed, which will not be discussed here.

3.2.2 Overview of Existing Approaches

The problematic of automated modal testing includes two issues which are not always treated separately or even differentiated, namely the MPE and the *modal tracking*. MPE is usually concerned with the automatic extraction of the modal parameters from single measurements, while the modal tracking inspects the change of modal parameters in time for continuous monitoring systems. It is common to use a baseline of modal parameters to initialize a modal tracking method, but this is not always the case and sometimes the whole data set is considered, without a preliminary selection through a MPE from an individual record.

Modal Parameter Estimation

There are many recent research publications which investigated various routines starting either from nonparametric system identification methods in frequency domain or from parametric methods in time domain.

The main issues which are addressed when performing an automated MPE correspond to the three steps involved in a manual interpretation: the application of several criteria to obtain a clear stabilization diagram, the quality assessment of the vertical columns using several indicators and the final selection of the physical modes estimates. For this purpose, various methodologies were developed on the basis of the main clustering approaches described in the previous section.

One of the earliest applications of an algorithm similar to hierarchical clustering was developed by Pappa et al. [14] for the ERA method used to analyse the data provided by a hammer test. The autonomous approach adopted distance measures based on natural frequency differences and MAC values. The consistent-mode indicator (CMI), specific for the ERA algorithm, served to evaluate the accuracy of each estimate and was included in the threshold values for the removal of infeasible modal identification.

An example of a successful pure hierarchical clustering approach is reported by Magalhães [12] where he resorted to an automatic selection of the modal parameters without constructing the stabilization diagram. The similarity measure introduced was also calculated taking the eigenfrequency differences and MAC values into consideration. All the mode estimates, stable or unstable, were included and the only user-defined parameters were the maximum limit for the intercluster distance and the number of expected modes.

Other approaches employ partitioning methods and can be exemplified by the work of Verboven et al. [24]. They used the Fuzzy C-means clustering method, which is an enhancement of the K-means algorithm described in Section 3.2.1, that assigns for each object a membership weight to every cluster. The system identification technique, frequency-domain maximum likelihood estimator (MLE), was optimized to handle large data sets from forced vibration tests on a slat track. The validation criteria included both classical mode shape indicators, as well as novel stochastic indices like pole-zero pairs or pole-zero correlation.

Another application of Fuzzy C-means clustering on stabilization diagrams was performed by Carden and Brownjohn [3] who used the covariance-driven SSI for the

data recorded from Z24 Bridge. The starting point was the procedure developed for this study case by Peeters and De Roeck [17] which automatically selected poles that were n_s times stable and evaluated the average of a column of stable poles to decide upon the representative result. In [3], the sensor grid was considered too sparse to provide meaningful mode shape estimates and hyperspherical clusters were formed by taking only the frequency and damping information in the complex plane.

Clustering based on histograms is another validated procedure that provided promising results. Scionti et al. [20] tried to replicate as much as possible the decision path followed by an experienced user. For this purpose, they divided the frequency axis of the stabilization diagram by overlapping bins for which weights to account for the occurrence of stable poles and alignment were assigned. The best pole was obtained after an additional clustering based on damping ratios was carried out for each bin. However, some user defined parameters were still needed (bin-width, stability criteria). A similar approach was reported also by Bakir [1], who proposed a new cluster validity index to handle the problems caused by the scale of measurements. This threshold value was also decided by human judgement.

A fully automated novel technique was developed by Reynders et al. [18] and serves as reference for the assessment of the dynamic monitoring data in Chapter 4. The three-stage clustering allows to obtain the modal parameters from stabilization diagrams without any user defined parameters. In the first stage, the spurious poles represented by vectors of so-called single-mode validation criteria are removed using a K-means partitioning method. The remained possible physical poles are subsequently grouped employing a hierarchical clustering with an automatically calculated cut-off distance based on the results from the previous step. In the last stage, the final set of physical modes is selected after an additional K-means clustering based on the number of elements in each group. The method was successfully validated for the Z24 bridge (Fig. 3.8) and the Wetteren footbridge.

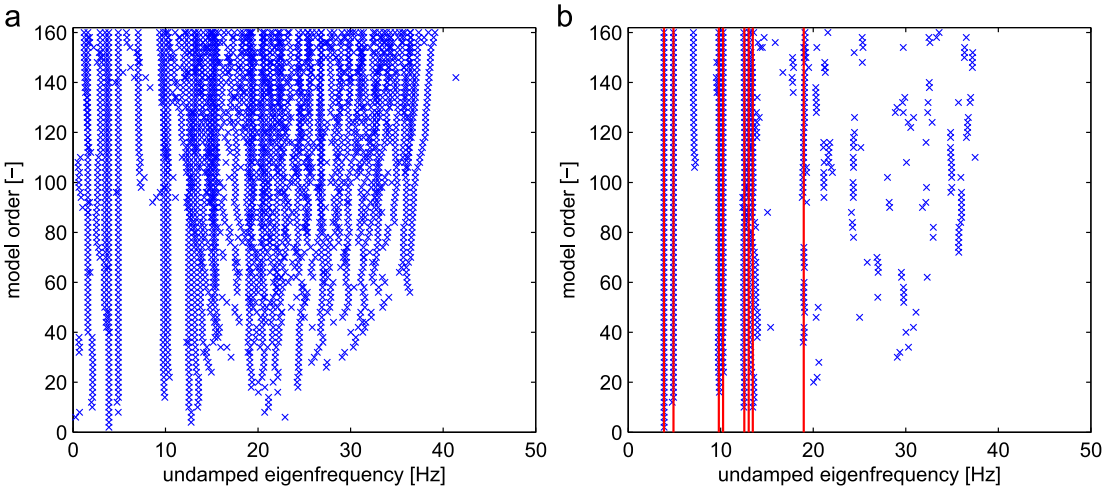

Figure 3.8: Full stabilization diagram (a) and automatically cleared diagram (b) for the Z24 bridge [18]

Modal Tracking

As already stated, the modal tracking issue is not always included in every study which deals with automated MPE.

From the aforementioned papers, the methodology presented in [3] performs the clustering on a couple of records altogether and is concerned with the tracking of the cluster centres within two pre-defined frequency bands, in order to detect any structural changes and eventual damage.

There are, however, several publications where the modal tracking follows the automatic MPE, also for SHM purposes.

The procedure described in [12] implies the comparison between each set of estimates obtained from three methods (FDD, SSI-cov/ref and p-LSCF) to a set of reference values identified from all the datasets acquired in the first day of continuous measurement. Only those modes with a MAC ratio higher than 0.8 and a frequency shift lower than 15% were kept. After the identification, the modal values were fed to an algorithm which eliminated the environmental and operational effects. It has been concluded that the SSI-cov/ref results (Fig. 3.9) are more sensitive to small daily variations, but the scatter of the mode estimates inside the same cluster is higher than for the p-LSCF method.

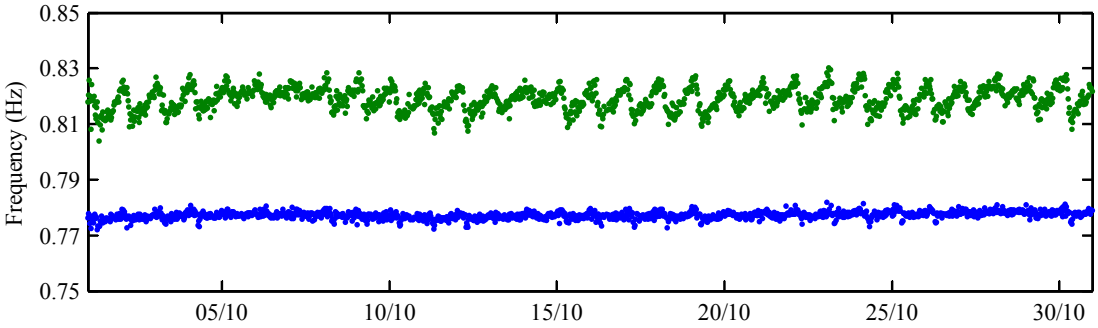

Figure 3.9: The first two eigenfrequencies for Infante D. Henrique Bridge [12]

The similar approach applied in [17] extracted the modes of interest from preliminary modal surveys. The modal tracking procedure was able to capture up to 18% frequency differences which were explained by normal environmental changes.

Another research where the baseline modal parameters were identified from the data of the first day of measurement, was conducted on the historic San Michele bridge by Cabboi et al. [2]. After setting appropriate values for the key input parameters of the SSI-cov/ref algorithm (model order and time-lag), the time evolution of the tracked natural frequencies could underline once again a clear dependence on environmental and operational conditions and the essential need to remove these influences for an accurate damage detection.

An attempt to highlight these aspects for a dynamic monitoring system, prefaced by the development of necessary tools, will be accomplished in the next chapter.

Chapter 4

Case Study: Spanische Allee Bridge

Several steel or composite railway bridges were experimentally investigated in the frame of the research project DYNABRIDGE carried out by the Institute for Structural Mechanics of Bauhaus-Universität Weimar and the Steel Engineering department of RWTH Aachen. The goal of the project was to identify and quantify the influences on their dynamic behaviour through measurements campaigns as well as to permit a better prediction of their expected response by means of numerical modelling.

One of the inspected bridges was the S-Bahn (city train) short-span bridge at Spanische Allee, located in the south-western part of Berlin, in the vicinity of Nikolassee station. (Fig. 4.1).

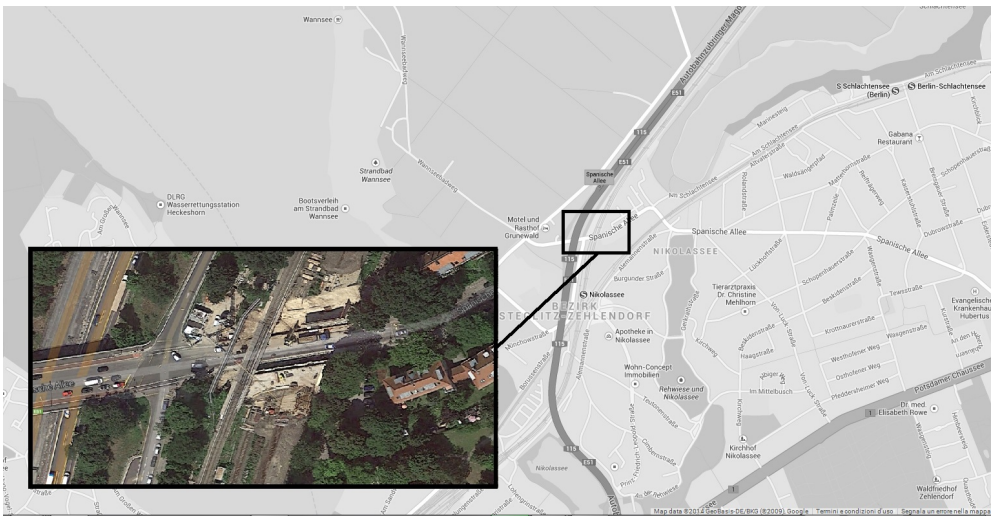

Figure 4.1: Overview of the bridge over Spanische Allee (www.maps.google.com)

Four new superstructures with spans from 17.84 m to 18.30 m have been installed. The analysed S355 steel through bridge has a width of 4.48 m and is delimited by inclined longitudinal beams. Each beam is stiffened, has an I cross-section and a depth of 1.33 m. (Fig. 4.2). A first assessment of the modal parameters of the superstructure was performed on the basis of a numerical model created in SLang 5.05, as well as experimental tests performed in the PST storage yard [26]. The outcome will be discussed later, in Section 4.2.2. However, these results were just preliminary information that served as a starting point for a detailed analysis of the

dynamic behaviour of the bridge under in-operating conditions and at different stages of construction. For this purpose, a monitoring system was set up and will be described in the following section.

Figure 4.2: The bridge superstructure [26]

4.1 The Monitoring System

4.1.1 Description of the Monitoring System

The installation of the monitoring system in the construction phase can provide a helpful insight on how the dynamic characteristics change during the deployment process and on the long-term behaviour of the bridge under service loads. Therefore, both dynamic and static data was acquired using velocity geophones and strain gauges. The geophones record velocities on two channels, horizontally (GP1) and vertically (GP2), and are appropriate for long-term measurements due to their robustness, while the linear strain gauges account for the deformations due to live loads in the longitudinal (DMS1) and transversal (DMS2) directions. The location of the monitoring sensors is depicted in Fig. 4.3. Additionally, two temperature sensors were connected in order to allow the assessment of the environmental conditions.

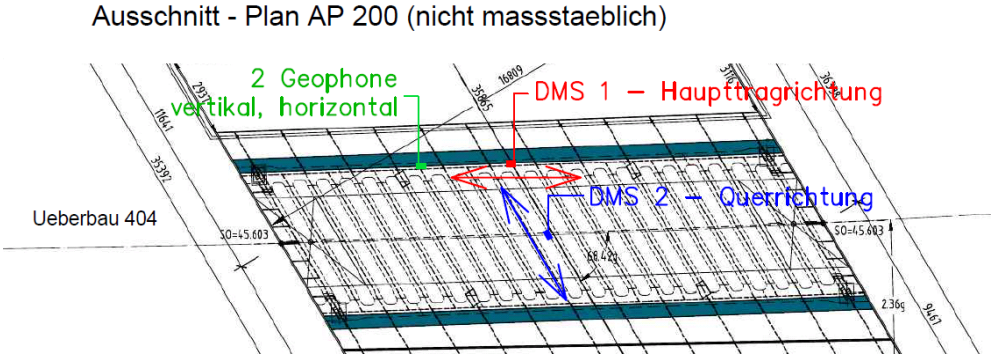

Figure 4.3: Position of the sensors [26]

4.1.2 Monalyse Database

The Monalyse database contains the recorded data which was collected and stored in a specific Python 3 binary file format, called *shelve*. The shelve module is a dictionary-like object, implemented on top of the *anydbm* module that picks one of the available interfaces when creating a database. The conversion of the files was done in a 64-bit operating system (OS), Linux distribution and constrains their accessibility to this particular configuration.

The files were organized according to their Windows OS timestamps which were also unique keys for the *shelve* files. Regular measurements of strain and temperature were acquired at a sampling rate of 100 Hz and recorded for about 30 s every 30 min. Ambient excitation data was sampled at 500 Hz for about 10 min every three hours. Moreover, each passing train triggered measurements with all channels at a sampling rate of 100 Hz and collected for around 80 s. Next to the raw values, every file was updated with its corresponding time and data statistics to help further post-processing. Also boolean information about eventual sensor faults and overload cases was included.

4.2 Assessment of Modal Parameters

One of the most challenging tasks was the automatic determination of the modal properties of the structure. Especially for the bridge monitoring, the dynamic characteristics are obtained efficiently by performing an in-operation modal analysis. It is not only expensive, but also not practical to use artificial excitation (shakers, drop weights). Therefore, the system identification has to be carried out from output-only measurements.

Since the main concern was to evaluate the data from a continuous monitoring system, an autonomous procedure had to be integrally developed in the same programming language in which the measurement files were stored, namely Python 3. Therefore, both the modal parameter estimation and the modal tracking had to further process the results from a system identification method. Due to their algebraic nature and accuracy in the identification of closely-spaced modes [2], the SSI techniques are considered appropriate for automated procedures. Therefore, the SSI-cov/ref (see Chapter 3) was the core algorithm to be implemented to deliver the estimates for the unsupervised assessment.

There are three main steps which can summarize the extraction of the modal parameters for vibration-based SHM: a) pre-processing of the measurements; b) system identification; c) automated interpretation of the results. It is important to regard carefully the quality of data before proceeding with any further analysis.

The code written for this purpose takes full advantage of the object-oriented nature of Python and allows to call different methods separately. For the clustering part, several packages in the *scipy* library are available. A manual interpretation as well as an automated processing of the files is included and can be accessed independently. However, since the program is not a stand-alone software based on a graphical user interface (GUI), the accessibility is not straightforward or very intuitive and needs to

be explained. For this purpose, the components and complete procedure are described in Appendix A. The settings for the input files are also detailed.

4.2.1 Modal Parameter Estimation

It is recommended to carry out the modal parameter estimation (MPE) after the signal has been processed. For example, it can be convenient to reduce the amount of data by decimation or reduce the noise by applying filtering techniques or offset removal. Decimation is a two-steps process that implies firstly, to eliminate the high-frequency components with a digital lowpass filter and secondly, to downsample the output sequence by an integer factor N. This means to keep every multiple of N output sample and neglect the others. It is very important to filter the signal before removing data points in order to avoid the problems caused by aliasing. The effects of the decimation on the stabilization diagrams are shown in Appendix B.1.

Manual Interpretation

For the manual interpretation part, the stabilization diagram which requires user interaction was constructed after editing, overwriting and customizing some widgets available in the *matplotlib* library and is not a native GUI application. Nevertheless, it includes useful features like radio and check buttons which can be seen in Fig. 4.4.

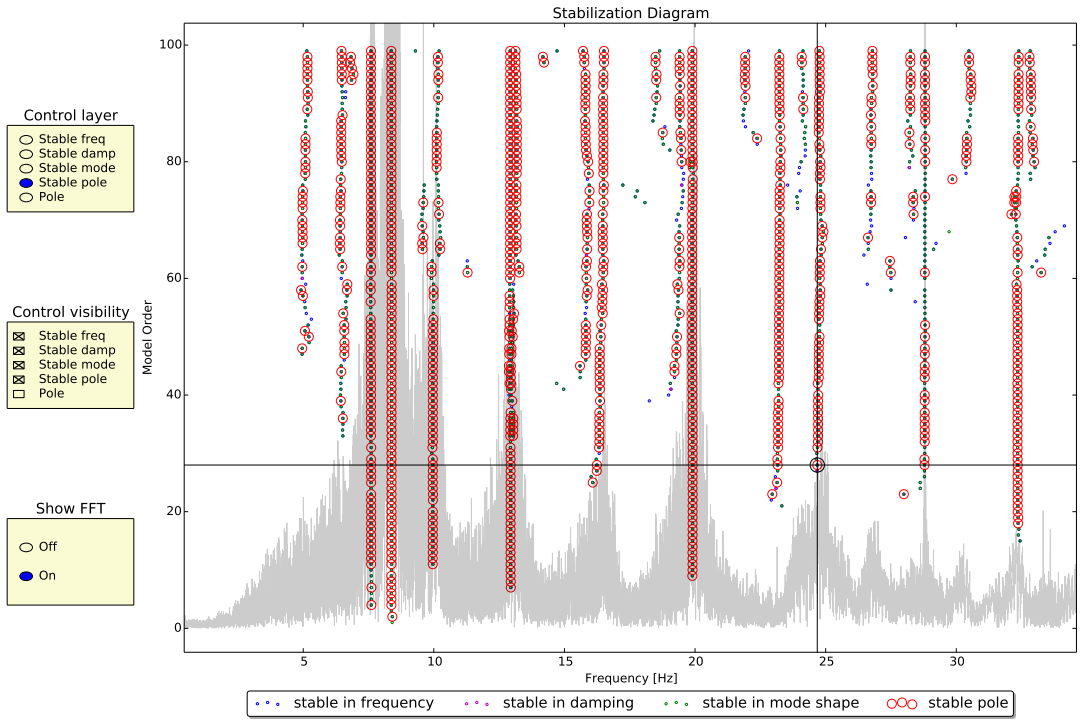

Figure 4.4: Stabilization diagram for a measurement on 13-09-2012

The widget forces the cursor to snap to the closest point only on a specific layer of poles, according to the selected option. If the image is too dense, one can deselect some

layers. Additionally, the spectrum resulting from the Fast Fourier Transform (FFT) can be superimposed on the diagram in order to correlate the poles with the peaks of the graph. The de-selection of an already picked pole is also allowed and shades the previous selection on the diagram. Every data point is evaluated according to the indicators presented in Section 3.1 which are displayed on the toolbar as follows:

Frequency, Model order, Damping ratio
Mean phase collinearity (MPC), Mean phase (MP), Mean phase deviation (MPD)

As already mentioned, the stabilization criteria influence the number of stable poles. How the results change for different predefined values included in the input file, is discussed in Appendix B. The input and the output formats are also presented in Appendix A.1.

The interactive inspection is important to evaluate one single record, but is not feasible for a database of hundreds of records. It is also extremely helpful to navigate through the results for a couple of measurements and create the baseline for the modal tracking procedure. For the evaluation of the database however, an unsupervised algorithm was introduced.

Automatic Interpretation

The automatic interpretation is achieved on the basis of the methodology described in [18] and introduced in Section 3.2.2. This solution particularly stands out due to its fully automated approach that eliminates completely any user intervention. The next paragraphs will detail and exemplify the background of the obtained results according to the description in [18] and applied to the data at the Spanische Allee bridge. A detailed example was chosen from the records on 13-09-2012 and each step of the algorithm will be illustrated using this data.

All stages which are involved in a manual analysis have a corresponding autonomous analysis based on the main clustering techniques. Firstly, the stabilization diagram needs to be cleared from spurious modes. Secondly, only certain well-shaped columns indicate physical modes and should be inspected. And finally, the representative value from several modes of interest are selected.

The algorithm related to the first step separates all the poles into spurious and possibly physical by means of a set of so-called soft and hard validation criteria. It consists of the following procedural chain:

1. calculate single-mode validation criteria
2. separate all the poles into two groups using the K-means clustering algorithm considering only the soft validation criteria (Table 4.1)
3. apply the hard validation criteria to the resulting possibly physical modes (Table 4.2)

To realize this, each mode is represented by a vector that contains n_{vs} soft validation criteria and n_{vh} hard validation criteria. The K-means partitioning method with $K = 2$

clusters identifies the certainly spurious modes and the possibly physical modes on the basis of soft validation criteria by minimizing the objective function introduced in Section 3.2.1, Eq. (3.8). It is important to respect this sequence because the hard validation criteria may remove all spurious modes and the clustering algorithm will significantly reduce the number of physical modes.

Next to the indicators defined in Section 3.1, a similar eigenfrequency distance measure between two modes i and j can be written in terms of continous-time eigenvalues

$$d\left(\lambda_i, \lambda_j\right) = \frac{|\lambda_i - \lambda_j|}{\max\left(|\lambda_i|, |\lambda_j|\right)} \tag{4.1}$$

which were introduced for the first time in Eq. (2.6) and contain information about damping and natural frequency at the same time.

Six soft validation criteria that take values between 0 and 1 are summarized in Table 4.1. It is possible that the clustering algorithm does not always reach convergence and several adjustments of the ideal values for the initialization of the centroids were permitted.

Table 4.1: Soft validation criteria and their ideal values

Criterion	Value	Ideal pyhsical	Ideal spurious
VS1	$d\left(f_i, f_j\right)$	0	1
VS2	$d\left(\xi_i, \xi_j\right)$	0	1
VS3	$d\left(\lambda_i, \lambda_j\right)$	0	1
VS4	$\mathrm{MAC}(\{\phi_i\}, \{\phi_j\})$	1	0
VS5	$\mathrm{MPC}(\{\phi_i\})$	1	0
VS6	$\mathrm{MPD}(\{\phi_i\})/90°$	0	1

Three hard validation criteria yield binary measures and are listed in Table 4.2.

Table 4.2: Hard validation criteria and their boolean values

Criterion	Check	Possibly physical	Certainly spurious
VH1	$\xi_i > 0$	1	0
VH2	$\xi_i < 15\%$	1	0
VH3	Conjugate pair?	1	0

Considering the six-dimensional space of the vectors of soft validation criteria in which the partitioning took place, different representations can be used to visualize the results. One possible approach is to depict the poles in 2D projections, but 3D projections can be employed as well. Figs. 4.5 to 4.10 illustrate several sections through the six-dimensional space. They also point out how the initial centroids move at the end of the iteration and in which range the two groups of poles lie. At this stage, the physical poles are referred to as "possibly" because they will be subjected later to further discrimination. It should be noted that the MPD values which appear in the graphs are normalized.

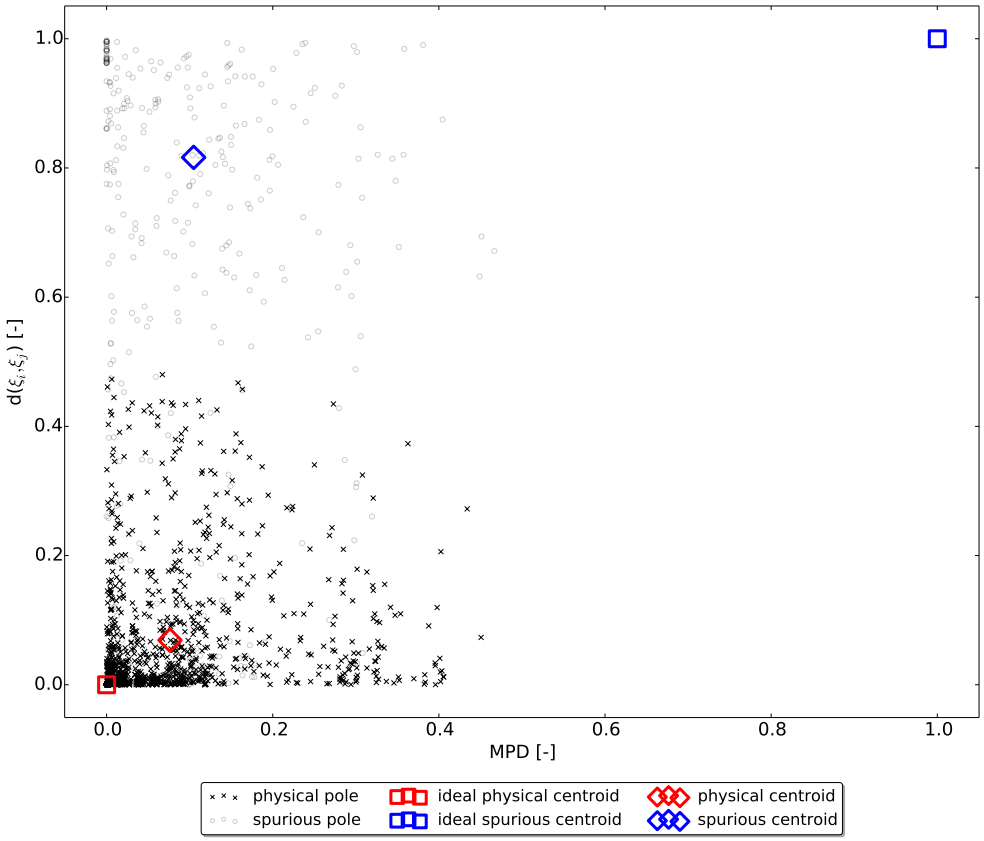

Figure 4.5: Possibly physical and certainly spurious poles (MPD vs. damping ratio)

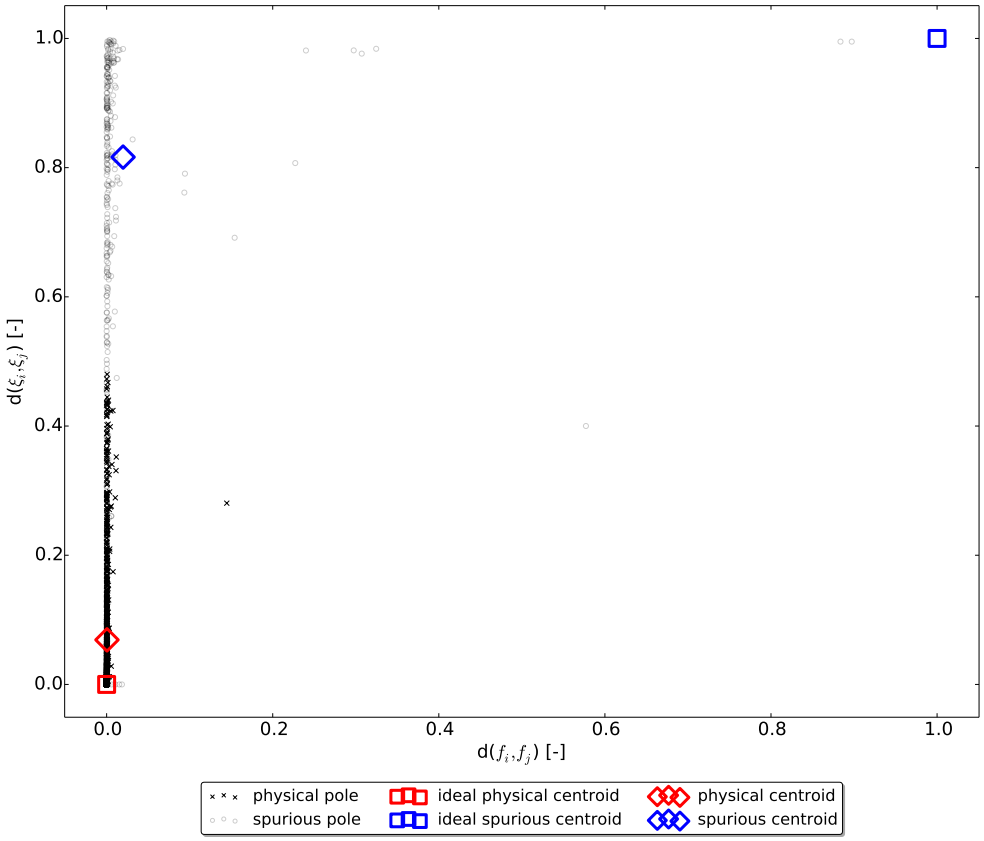

Figure 4.6: Possibly physical and certainly spurious poles (frequency vs. damping ratio)

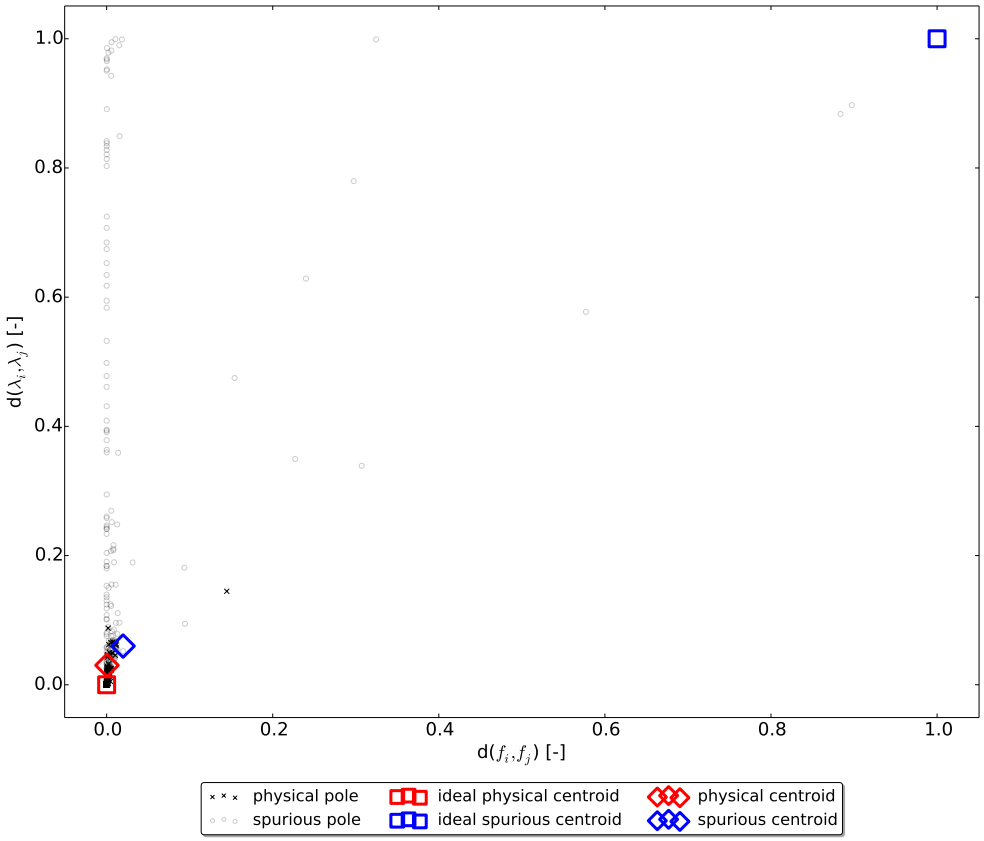

Figure 4.7: Possibly physical and certainly spurious poles (frequency vs. eigenvalue)

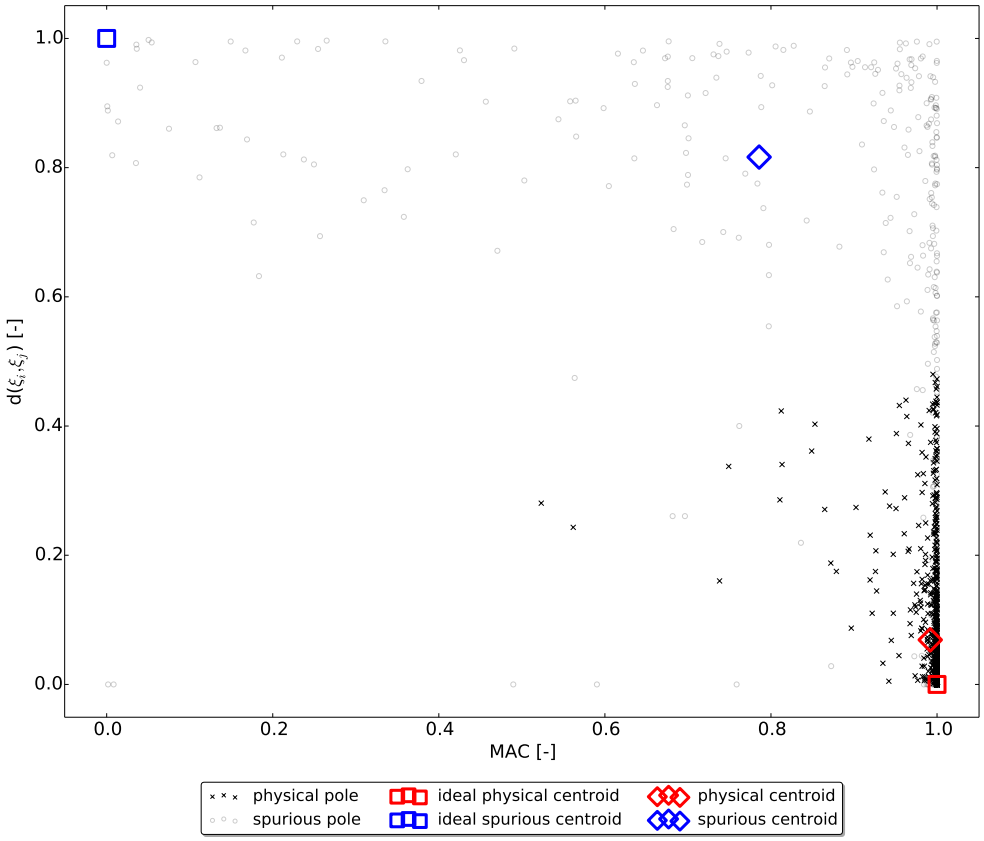

Figure 4.8: Possibly physical and certainly spurious poles (MAC vs. damping ratio)

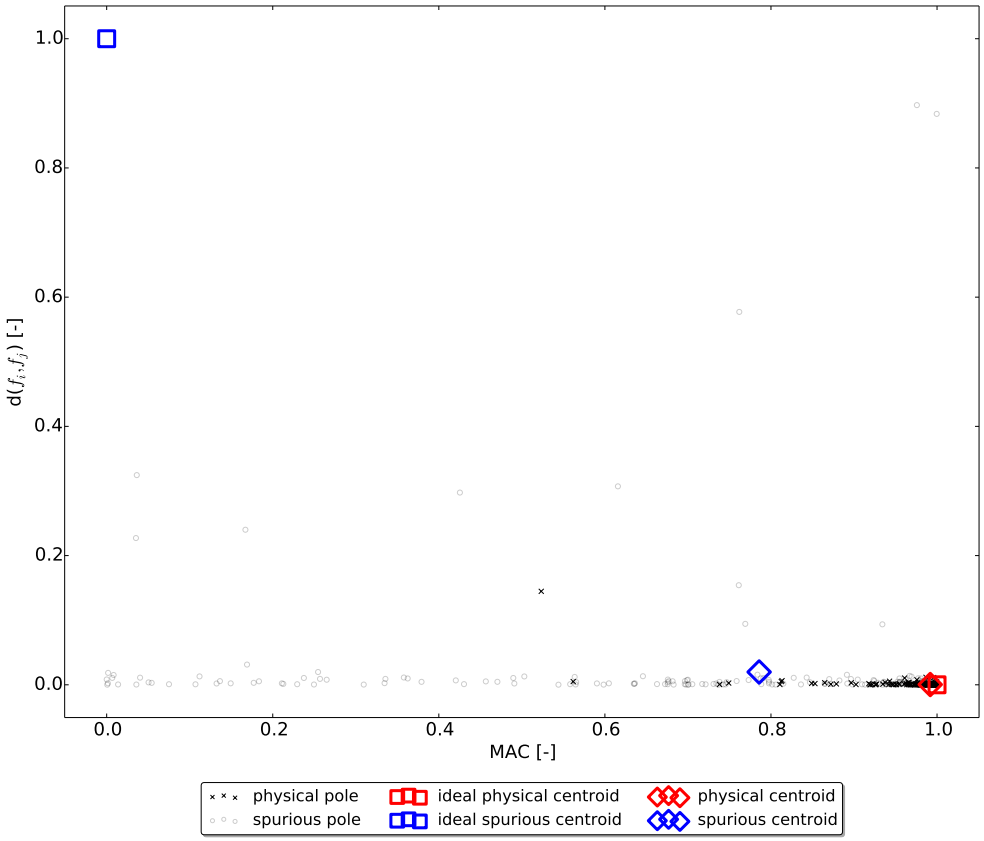

Figure 4.9: Possibly physical and certainly spurious poles (MAC vs. frequency)

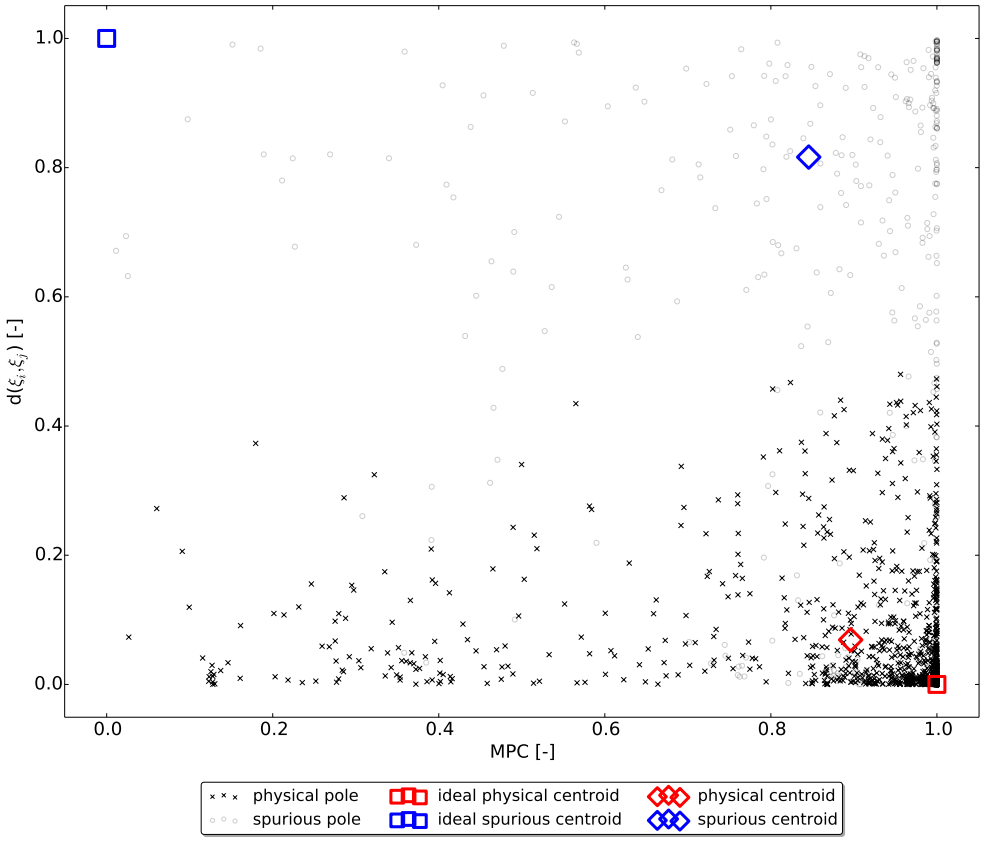

Figure 4.10: Possibly physical and certainly spurious poles (MPC vs. damping ratio)

After performing the automated clearing by filtering the obtained data through the hard validation criteria, the number of poles in the stabilization diagram reduce considerably and the computational effort for the next clustering stage is diminished. This can be easily perceived inspecting Fig. 4.11 and one can notice that this approach succeeds in the same way as the stability criteria applied by the user would do.

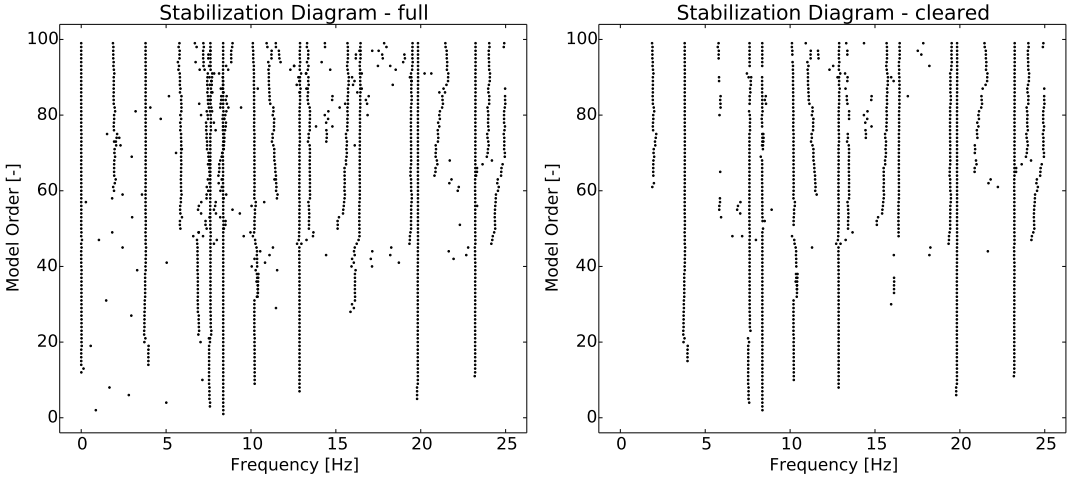

Figure 4.11: Full and automatically cleared stabilization diagrams

Now that most of the remaining poles have a potential physical interpretation, the related modes should be grouped based on a similarity measure. The agglomerative hierarchical clustering approach requires to pre-define a threshold value, but in this case the parameter will be determined automatically from the outcome of the previous step. The three main sequences can be summarized as follows:

1. compute the mutual distance between any two poles i and j and assemble the proximity matrix based on the following measure:

$$d(i,j) = d(f_i, f_j) \times w_f + (1 - MAC(\{\phi_i\}, \{\phi_j\})) \times w_{MAC}$$

2. merge the closest two clusters and update the proximity matrix using the single linkage measure between their elements

3. repeat the previous procedure until the distance between the two closest clusters i and j exceeds the allowed threshold computed according to the formula:

$$\tilde{d} = C_{d(f_i,f_j)} \times w_f + \left(1 - C_{MAC(\{\phi_i\},\{\phi_j\})}\right) \times w_{MAC}$$

where w_f and w_{MAC} denote the weights introduced for the frequency and mode shape correlation, while $C_{d(f_i,f_j)}$ and $C_{MAC(\{\phi_i\},\{\phi_j\})}$ represent their corresponding centroids after the first stage of clustering. The defined distance between two modes has also been used in [12, 2]. In comparison to the measure applied in [18], which includes also the damping information, this version has been preferred due to the high scatter of the damping ratios. Especially for this application, it is also appropriate to assess more importance to the relative differences in natural frequencies than to the mode shape differences as a result of the limited number of sensors that yield only two values for

the mode shapes. A new way to define automatically the threshold values is proposed, and is a reliable alternative to the one in [18].

The essential part of building the similarity matrix is computationally demanding and it makes sense to initially retain only those solutions in the range of interest. Even several minutes of computation time saved for an individual calculation become relevant if they accumulate for hundreds of records. For this reason, the analysis for Spanische Allee bridge was carried out in the frequency range 0 Hz to 25 Hz.

After the clustering is realized, the history of the merging process can be tracked on the dendrogram depicted in Fig. 4.12. The cut-off level is represented by a red dashed line that indicates the final set of the physical modes intersected. Every clustering process completed above the limit is ignored.

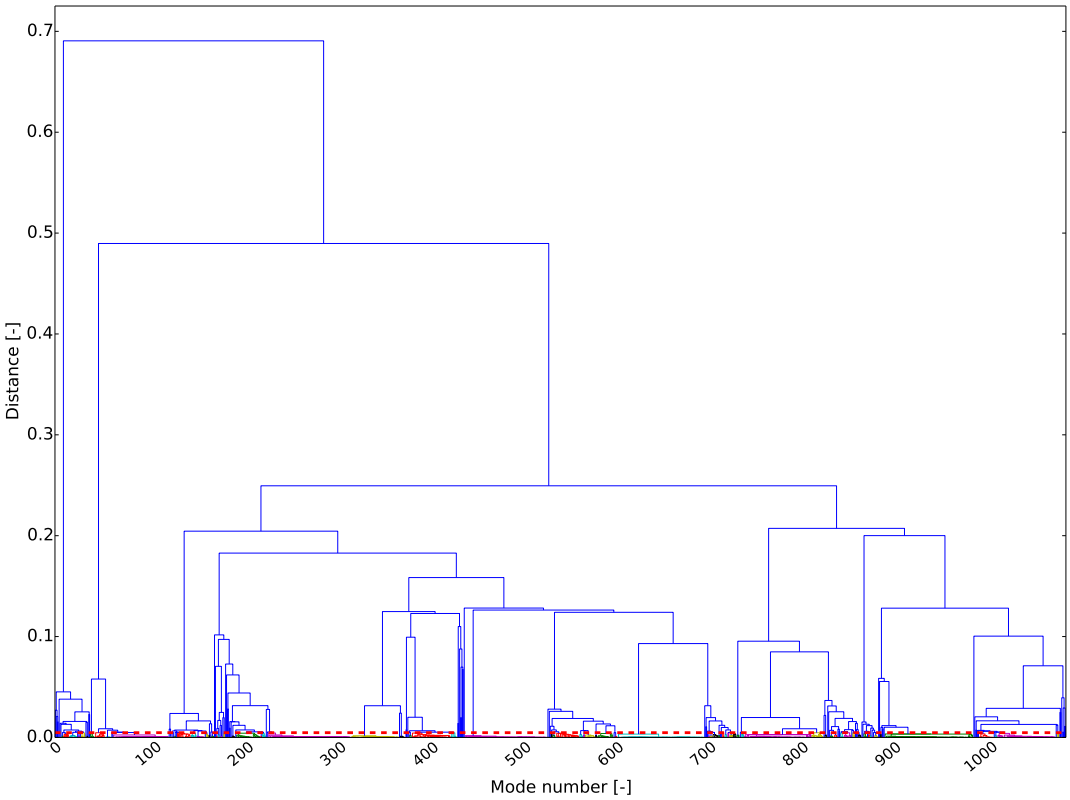

Figure 4.12: Dendrogram representing the hierarchical clustering algorithm

For the analysed measurement, the low threshold of 0.0047 suggests the necessity to discriminate between closely spaced modes. Multiple trials for different records prove that this is a sensitive parameter. The lower the limit is, the higher is the number of the resulting clusters and the estimates of the same physical mode might split into several clusters. On the contrary, the higher is the value, the lower is the number of the clusters that could consequently merge different modes into the same group. How the value adapts for different recordings is shown in Appendix C and highlights once again why a fully automated algorithm is imperative.

Each resulting cluster contains a specific number of elements that vary for different thresholds. In Fig. 4.13 the obtained clusters are centred around the representative frequency values and displayed up to 7 Hz.

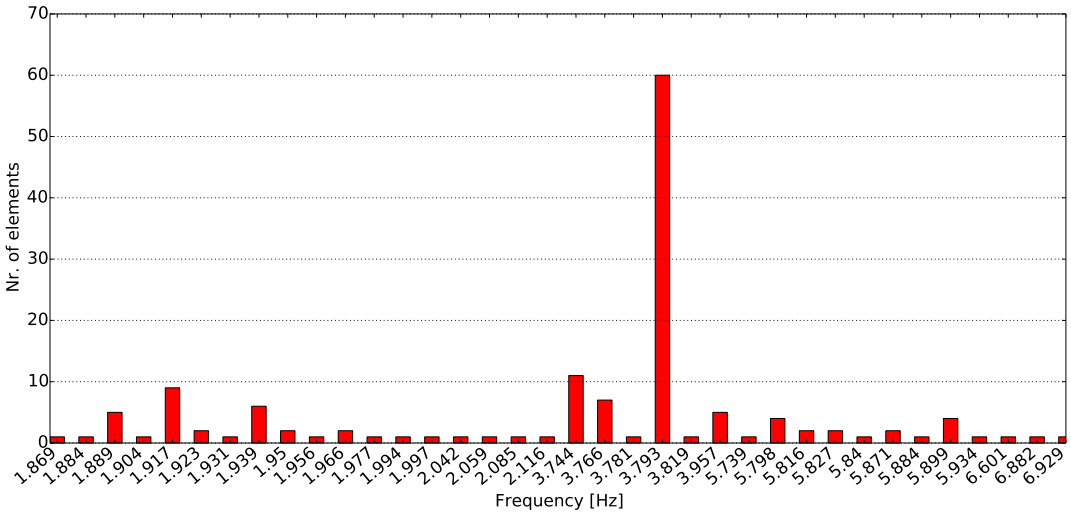

Figure 4.13: The clusters resulted in the frequency range 0 Hz to 7 Hz

Up to now, it is uncertain how many reliable modes are contained in the clustering results. The decisive parameter for an accurate evaluation is the number of members in each cluster that reflects the rate of appearance for every possibly physical mode. A low rate is an indicator of a random identification, in contrast to a high rate typical for stable poles that ideally start to emerge from a certain model order on. Usually, only a specified number of modes sorted according to their cluster dimension is kept (see [12]). But this approach requires again human intervention. Alternatively, it is intuitive to resort to another K-means partitioning method with $K = 2$. The objective function in Eq. (3.8) minimizes the distances between the centroids of the physical and spurious modes and operates over the number of elements in each set. The outcome of this last stage of clustering is illustrated in Fig. 4.14.

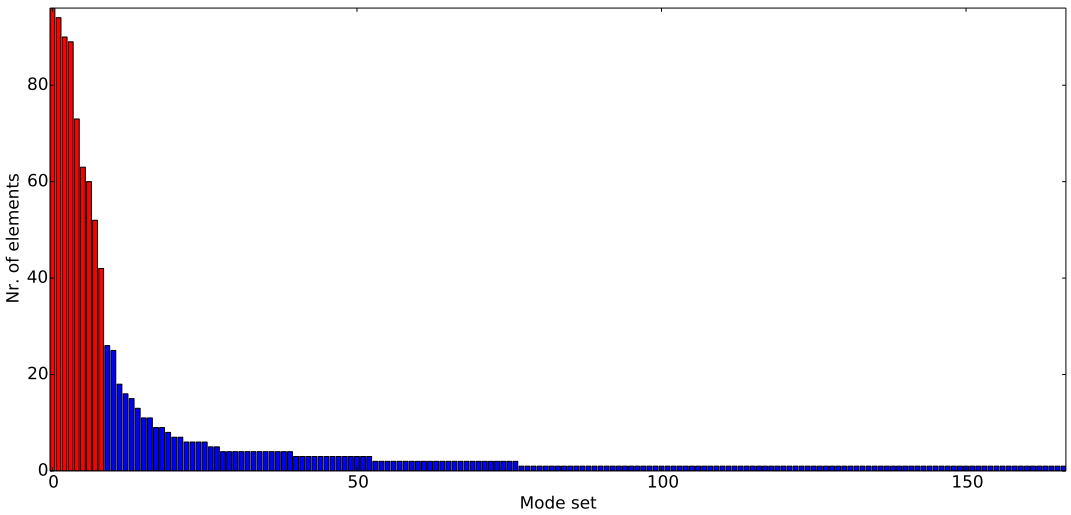

Figure 4.14: Unsupervised classification of the mode sets based on their number of elements
Red bars: physical sets; Blue bars: spurious sets

Another important issue is the way to select a representative value from the final mode sets. One possible approach is to compute the average values in terms of frequency,

damping ratio and mode shape, as in [12]. This is however an approximation and might not correspond exactly to any of the obtained solutions. Another approach is to exclude the damping ratios when computing the mean values, and select the median damping ratio next to the average frequency, mode shape, MPC and MPD [2]. This accounts for the observation that the damping ratios are normally scattered and the outliers weigh less. Eventually, one might naturally select the complete solution corresponding to the median damping ratio value, as in [18]. This path is also followed in the current application.

Fig. 4.15 is a suggestive representation of the final sets of modes emphasized in the first picture using blue vertical lines and superimposed on the classical stabilization diagram. The red dashed line in the clusters vs. frequency plot indicates the level at which the last partitioning algorithm separates the two groups of poles. Only those clusters with the number of elements above the automatically detected limit are accepted.

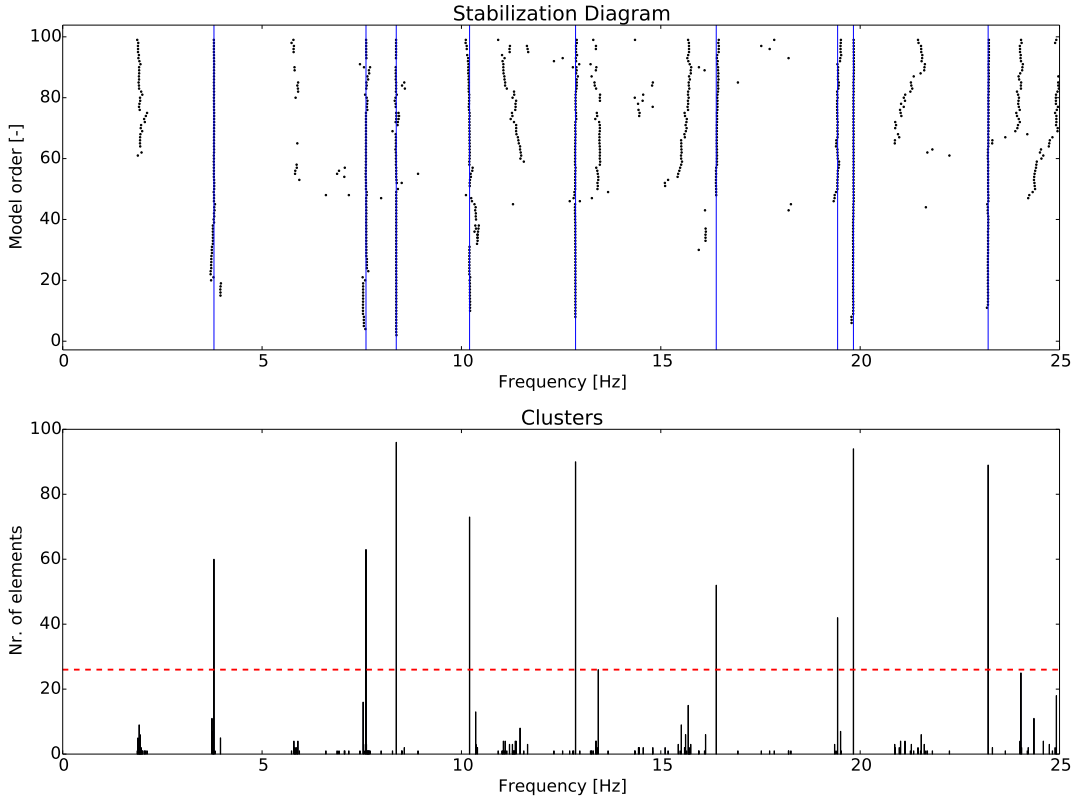

Figure 4.15: The final set of physical modes (above) and the result of the K-means clustering algorithm (below)

The last plot demonstrates how a physically intuitive and completely autonomous procedure can be applied successfully to output-only measurements. Once all the records have been processed, the next step of the modal parameters assessment is to track every possible change in the dynamic behaviour of the structure and to correlate it with external influences, like environmental factors or construction stages.

4.2.2 Modal Tracking

In the context of continuous dynamic monitoring, the modal parameters identified at each time instance must be linked to a set of reference values. It can not be expected that the same set of modes will be identified for every single record, so it is also not straightforward which modes should be tracked. A starting baseline list of modal parameters can be defined after a manual inspection of several results. This can be achieved after reviewing the outcome of the measurements from a one day period. In the present study, the results were not filtered for each record, thus all data was plotted as in Fig. 4.16 and only the modes with an appearance rate higher than 80% were selected.

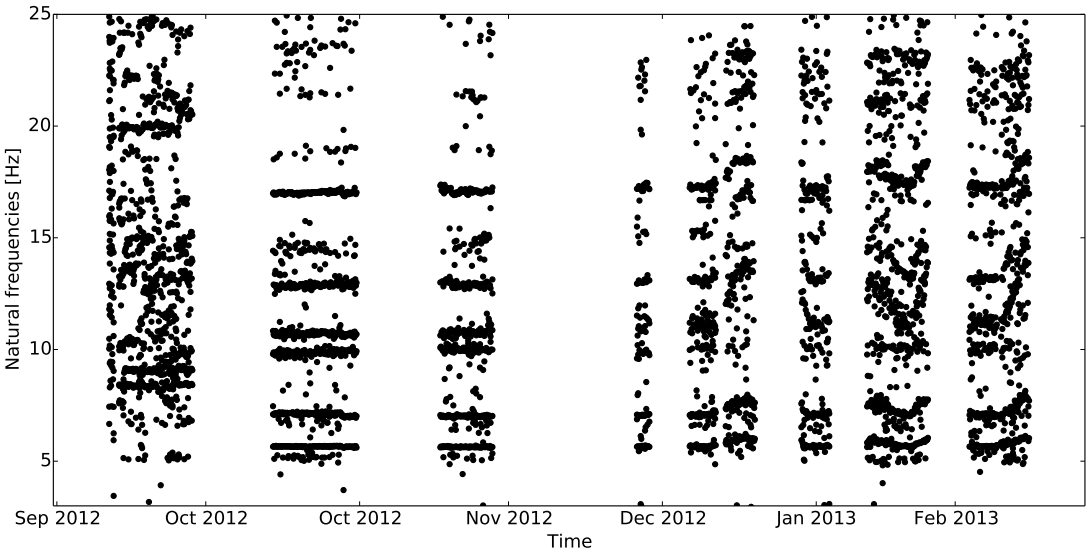

Figure 4.16: All the frequencies resulting from the automatic SSI

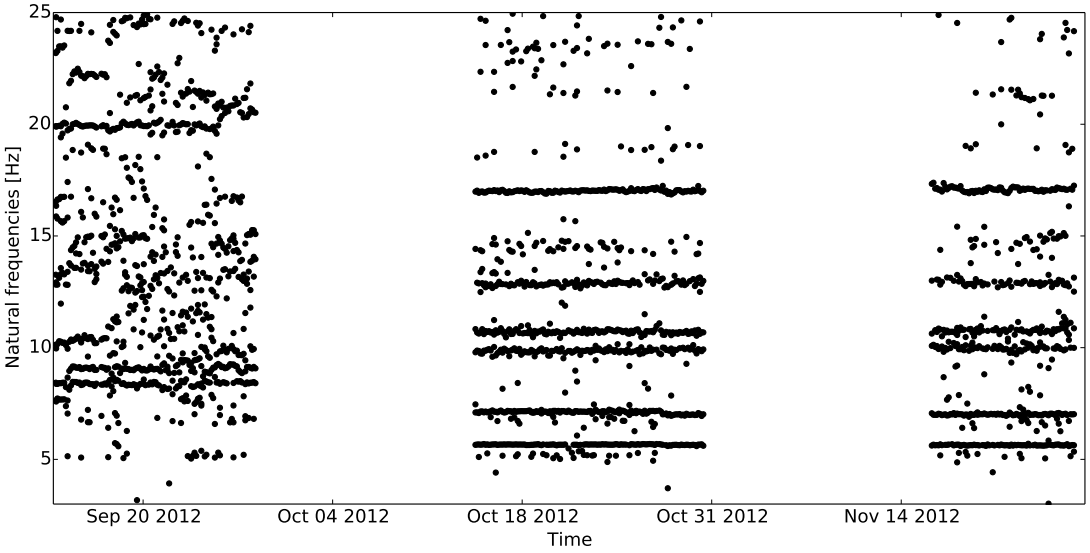

Figure 4.17: All the frequencies resulting from the automatic SSI during September 2012, October 2012 and November 2012

At this point, several aspects should be emphasized. From the zoomed-in plot (Fig. 4.17)

is apparent that in September 2012, the results are very scattered and only 2 frequencies could be effectively tracked. In October 2012 and November 2012, the evolution of 6 eigenfrequencies is visible, but 3 of them, between circa 9 Hz and 14 Hz, are not properly excited later on (after December 2012 in Fig. 4.16) and their strong variation makes the tracking procedure almost impossible.

However, the effect of different construction stages is revealed and it is in agreement with the dates in the construction log. On the 14^{th} of September, the superstructure was put into place, shortly after the monitoring system was installed. Fig. 4.18 reveals a sudden increase of the two eigenfrequencies, which starts at 7:00 am and stabilizes around 16:00 pm, when the hydraulic jacks were completely removed.

The period between the 28^{th} of September and the 14^{th} of October corresponds to the time frame when the ballast was added. Unfortunately, there is no data available but the frequencies observed around 9.80 Hz and 10.7 Hz during the next two weeks in October seem to correspond to those identifiable in September. As expected, after adding the ballast, the structure is stiffer and the eigenfrequencies higher.

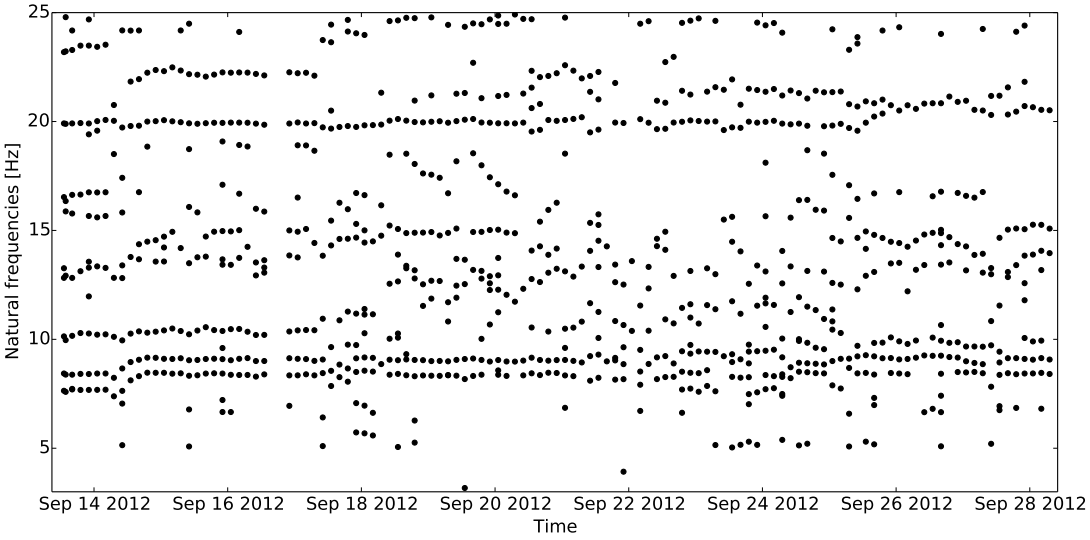

Figure 4.18: All the frequencies resulting from the automatic SSI during September 2012

Two short-term measurements (Table 4.3) were already analysed, one taken in the storage yard for a similar structure (Measurement 1) and another one for the same structure monitored over a longer term and taken after completion (Measurement 2).

Table 4.3: Results from measurements at the plant (Measurement 1) and for the completed structure (Measurement 2) [26]

Mode nr.	Measurement 1 (12.07.2012)		Measurement 2 (21.11.2012)	
	Frequency [Hz]	Damping ratio [%]	Frequency [Hz]	Damping ratio [%]
1	5.06	3.27	5.66	0.8
2	7.93	2.89	7.07	2.0
3	8.82	1.07	14.33	3.8
4	9.96	0.62	17.06	1.0

These results serve as a comparison basis with the solutions from the automatic modal analysis. The 2 frequencies indicated in Fig. 4.18 for the 13^{th} and 14^{th} of September correspond to the second and third frequency already assigned in Measurement 1. Theoretically, the values from Measurement 2 should be identified also in the long-term results.

In order to carry on the modal tracking, the procedure involves the following steps:

1. the calculation of the MAC index and the frequency distance between each reference mode and the detected modes in data set i
2. the selection of the closest spaced mode, which does not exceed a frequency variation of 10 % and exhibit a MAC value greater than 95 %
3. for the next data set $i + 1$, the repeat of previous steps by keeping the same reference list

This algorithm accounts for the environmental and operational changes, but due to the required high threshold values, it can fail whenever two closely spaced modes appear and their variation is difficult to track.

After successfully processing 570 files, a first attempt was to compute the modal parameters taking the whole record into consideration. It has been observed that most of the measurements captured abnormal load conditions caused by train passages. This phenomenon contradicts the white noise assumption on which the SSI algorithm is based.

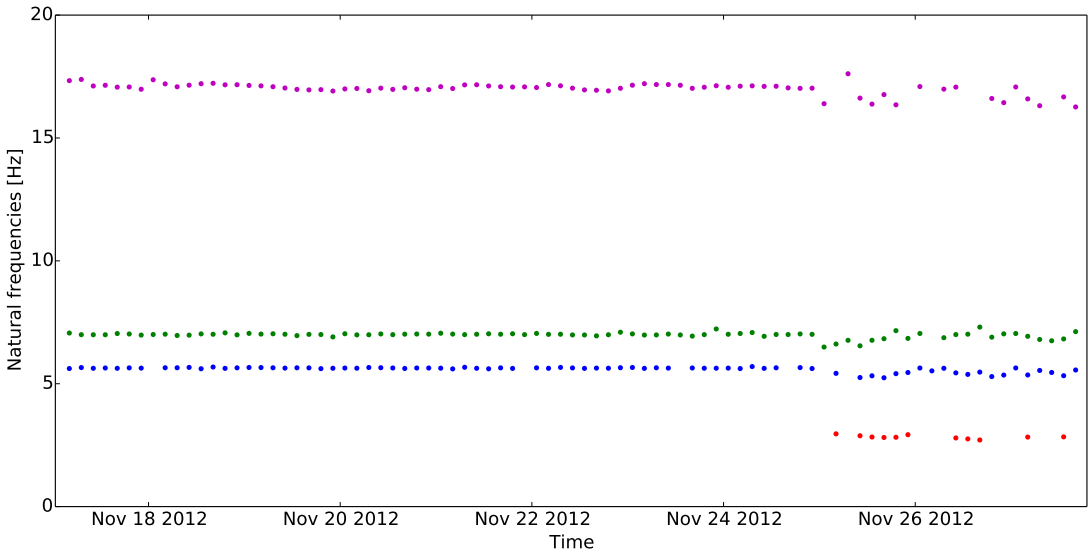

Figure 4.19: Evolution of the natural frequencies during November 2012 and the evidence of effects from train passages

Fig. 4.19 shows the appearance of an additional frequency at around 3 Hz, exactly after the bridge was put into operation (the 25^{th} of November, 2012). The figure also highlights disturbances in all the other modes, starting with the same day.

As a second attempt, the train passages were detected and the corresponding parts of the signals were eliminated. A noteworthy effect was the removal of the frequency at 3

Hz and the qualitative improvement of the results, as shown in Fig. 4.20 and Fig. 4.21. Most likely, the first mode identified in the first case was caused by the forced vibration induced by passing trains and was not an inherent property of the structure.

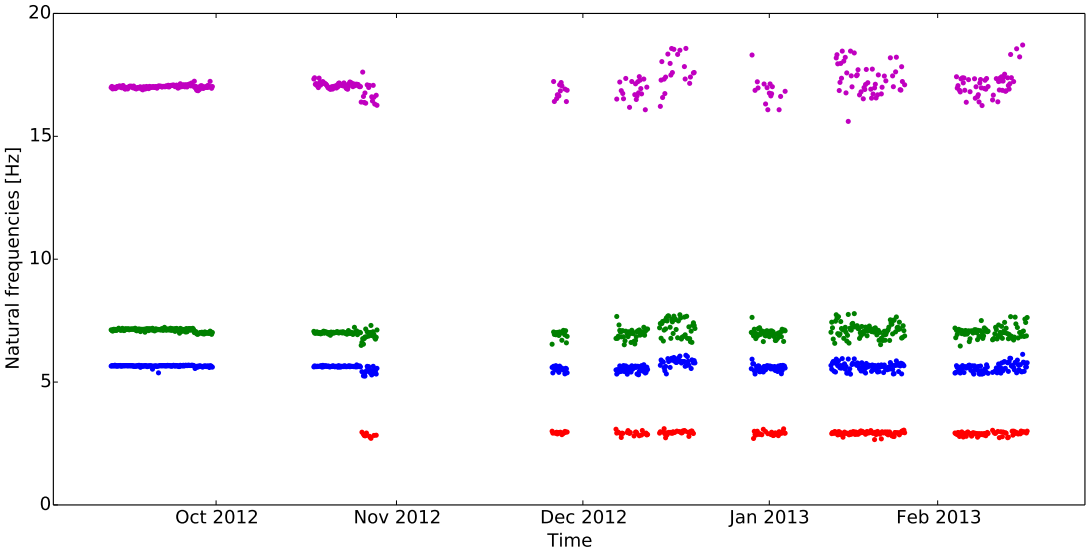

Figure 4.20: Evolution of the natural frequencies between October 2012 and March 2013, without removing the train passages from the signal

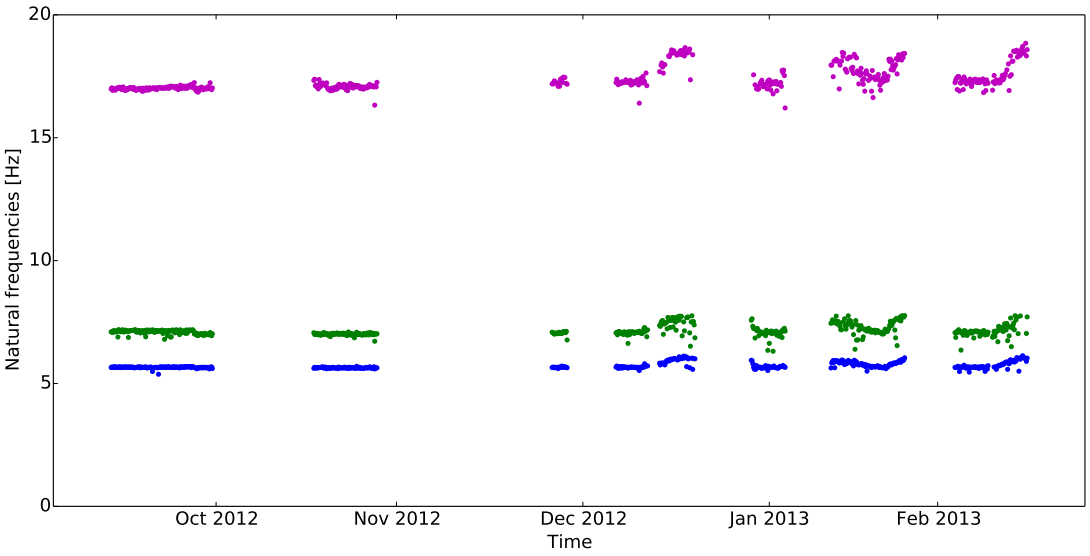

Figure 4.21: Evolution of the natural frequencies between October 2012 and March 2013, after removing the train passages from the signal

Moreover, the temporary forced vibration decreased the sensitivity of the computed modes to environmental changes. For this reason, the scatter for each mode in Fig. 4.21 is clearly lower and the variation of the frequencies can be better observed. Unfortunately, several disturbances and connectivity problems caused various gaps visible on the graphs and restrict a continuous identification of the modal parameters for every single day. However, they do not affect the global trend of the outcome.

Table 4.4 summarizes the results of the automated procedure and contains the most relevant statistical indicators. The high success rate suggests that the algorithm performs very well for strong excited modes and can track modal parameters that exhibit fluctuations of up to 22 %. The maximum difference is computed between the lowest and the highest values in a data set and includes the eventual outliers as well. When comparing the results with those for short-term measurements (Measurement 2) in Table 4.3, there is a clear correspondence between the first, second and fourth mode of vibration.

Table 4.4: Results of the automatic modal analysis

Mode nr.	Success rate [%]	Eigenfrequency			
		Min. [Hz]	Mean [Hz]	Max. [Hz]	Max. diff. [%]
1	92	5.38	5.72	6.13	14
2	96	6.31	7.14	7.76	22
3	90	16.21	17.35	18.84	16

Concerning the damping ratios, the box plot in Fig. 4.22 displays the median values and the variability of the values outside the upper and lower quartiles together with the outliers symbolized with blue crosses.

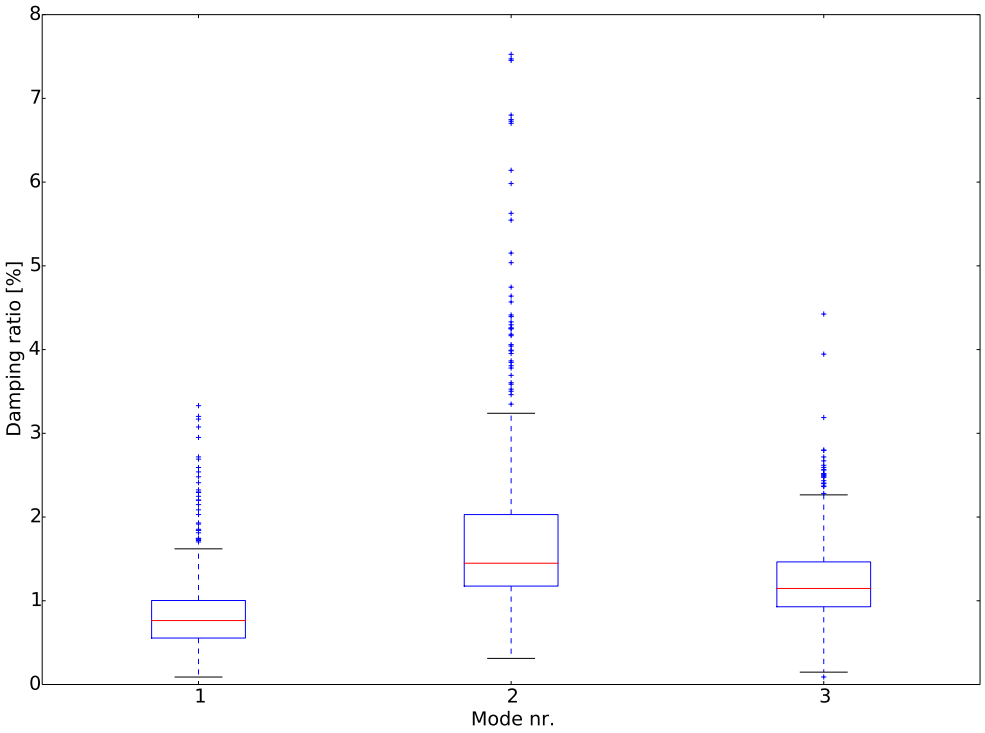

Figure 4.22: Damping ratios resulted from the automatic analysis

The median damping ratios are equal to 0.764 %, 1.449 % and 1.147 % for the first, second and third mode, respectively. These values are close to those manually assessed in Table 4.3 for Measurement 2.

In order to assess the data in the context of environmental changes, two temperature sensors were placed on the bridge, one of which was enclosed in a metal box. Samples were acquired every 30 min to ensure a very good temporal resolution appropriate to the unsteady weather at the location. Fig. 4.23 emphasizes a strong variation during 6 months of acquisition and indicates higher temperatures for the sensor placed inside the box. Variations of about 25 °C were recorded for several months.

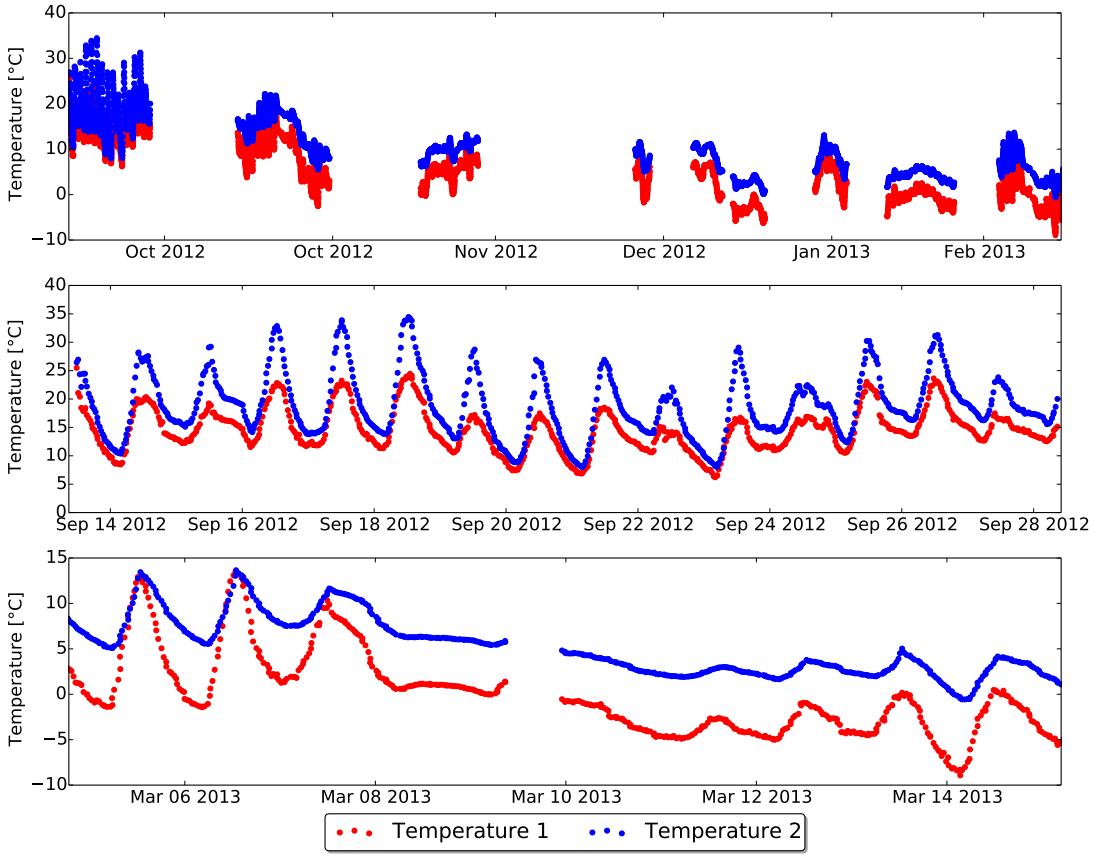

Figure 4.23: Evolution of the temperature for the two sensors during the time period Oct 2012-Mar 2013 (top) and close-up around Oct 2012 (middle) and Mar 2013 (bottom)

Given the significant temperature changes, which span between 25 °C and -9 °C, their influence on the modal parameters should not be neglected. Therefore, the considerable variations of the identified eigenfrequencies should be explained by this observation. By adding the temperature graph to Fig. 4.21, it becomes possible to correlate the natural frequencies and temperature changes, as in Fig. 4.24.

Several conclusions can easily be derived by inspecting the whole picture. Firstly, the natural frequencies are very sensitive to temperatures below 0 °C. All three modes show a strong tendency to increase when negative temperatures occur, which suggests that the structure looses its flexibility due to freezing. The stiffness of the supports is strongly dependent on the stiffness of the rails, thus whenever a change of this parameter appears, it is reflected in the structural behaviour as well. Secondly, there is no significant change of the frequencies if the temperature fluctuates but remains positive. This aspect can be seen by zooming-in on the evolution of these parameters during October 2012 (Fig. 4.25).

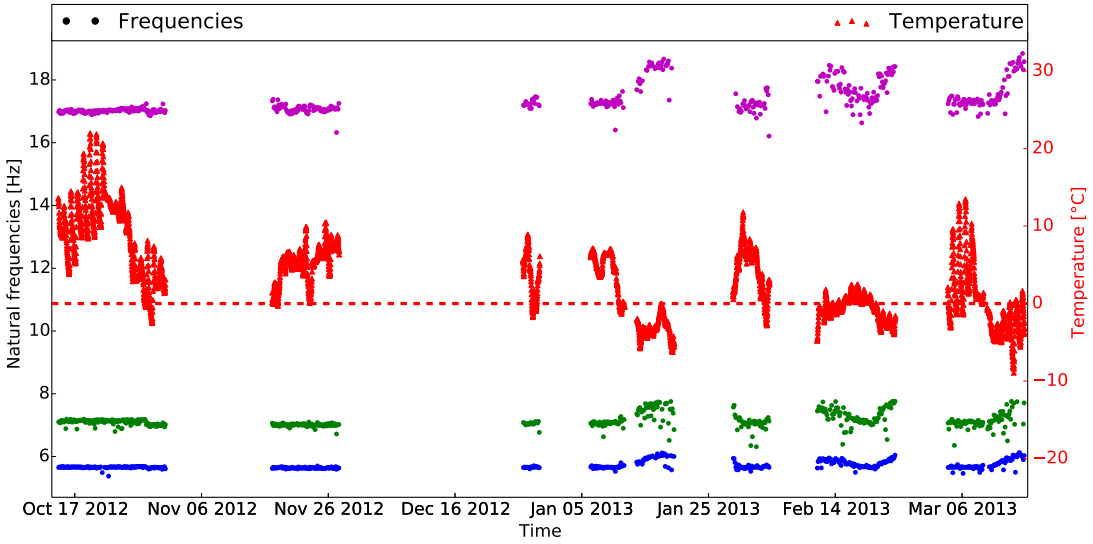

Figure 4.24: Evolution of temperature and natural frequencies during the time period Oct 2012-Mar 2013

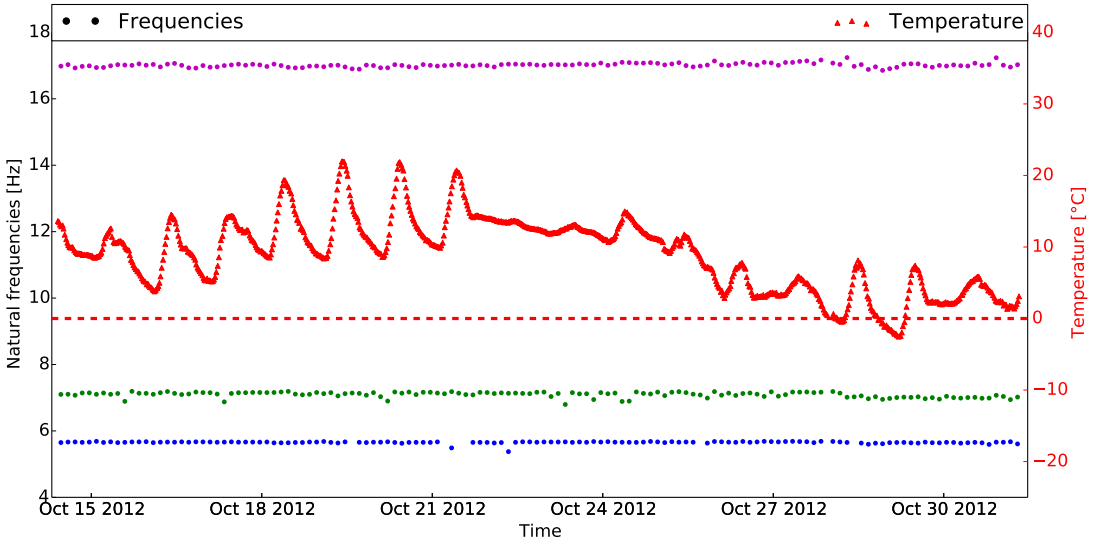

Figure 4.25: Evolution of temperature and natural frequencies during Oct 2012

However, this is not the case for the time span December to March, when negative temperatures were constantly recorded.

For this period, Figs. 4.26 to 4.28 reveal to which extent the frequency variations are correlated to the temperature which is represented using a reversed scale. The first natural frequencies show a variation of about 0.5 Hz, whereas the second and third ones of approximately 1 Hz and 2 Hz, respectively. Due to the low spatial resolution, it is not possible to relate specific mode shapes to temperature variations.

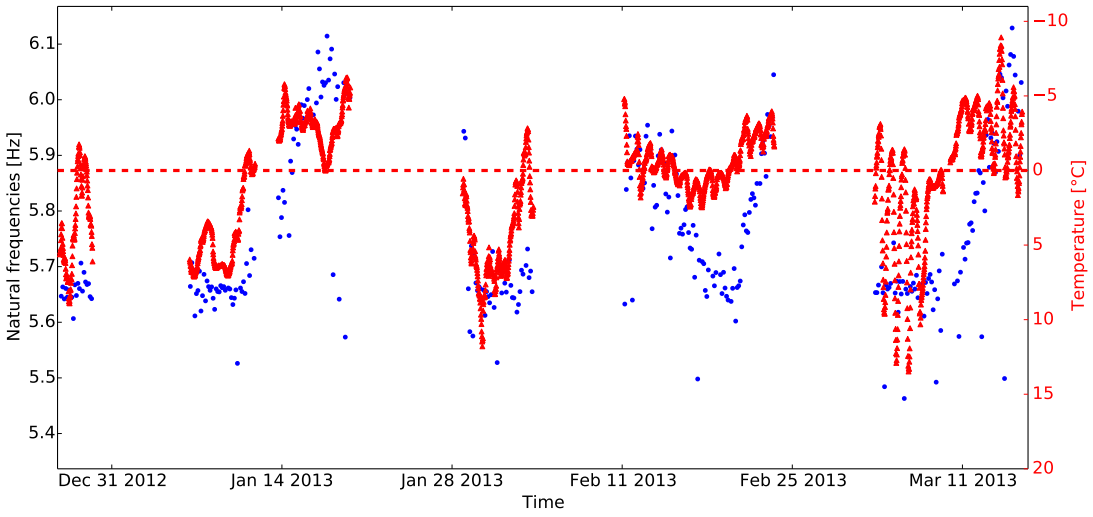

Figure 4.26: Temperature *vs* first eigenfrequency during Dec 2012 to Mar 2013

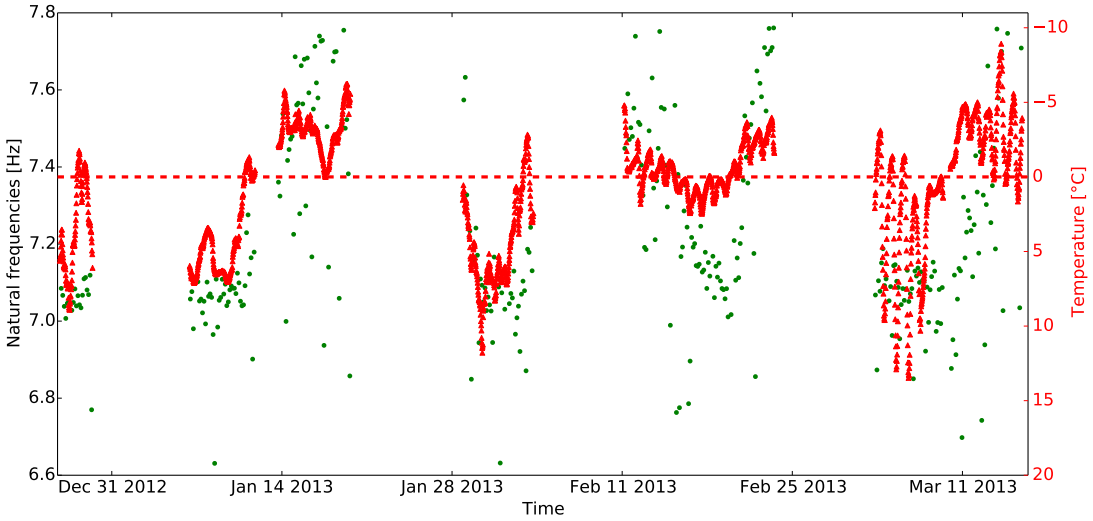

Figure 4.27: Temperature *vs* second eigenfrequency during Dec 2012 to Mar 2013

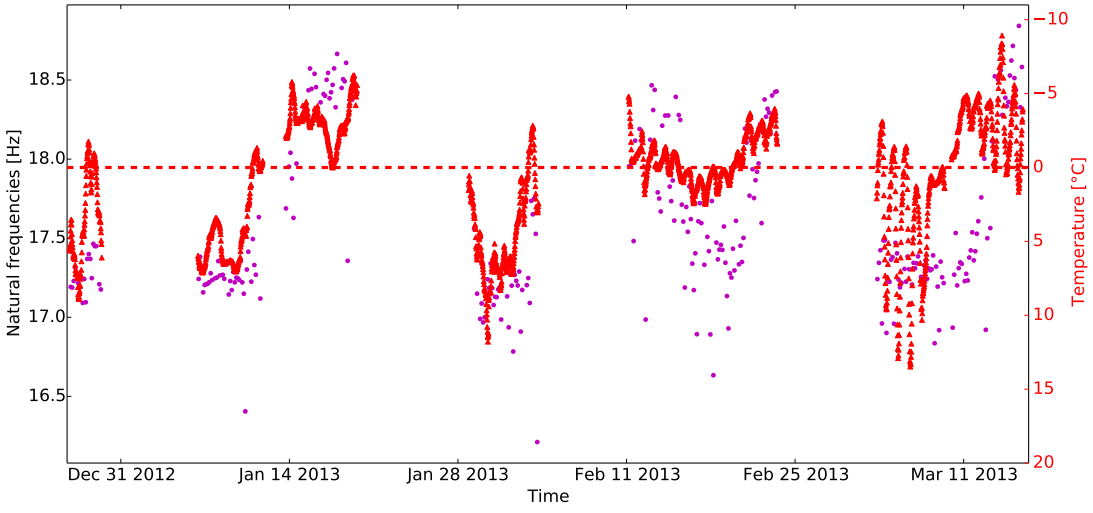

Figure 4.28: Temperature *vs* third eigenfrequency during Dec 2012 to Mar 2013

4.3 Structural Response to Passing Trains

Besides the modal analysis, the data acquired by the installed monitoring system can be used to derive several conclusions regarding the response of the structure to passing trains. The information provided includes the strain variation that is directly related to the stiffness and deflection of the bridge. From the triggered measurements, one can assess the variation of statical deformation and temperature, but also to determine the speed of moving loads. The passing S-Bahn belongs to DB-Baureihe 481/482 class of electrical railcars, whose geometry is shown in Fig. 4.29.

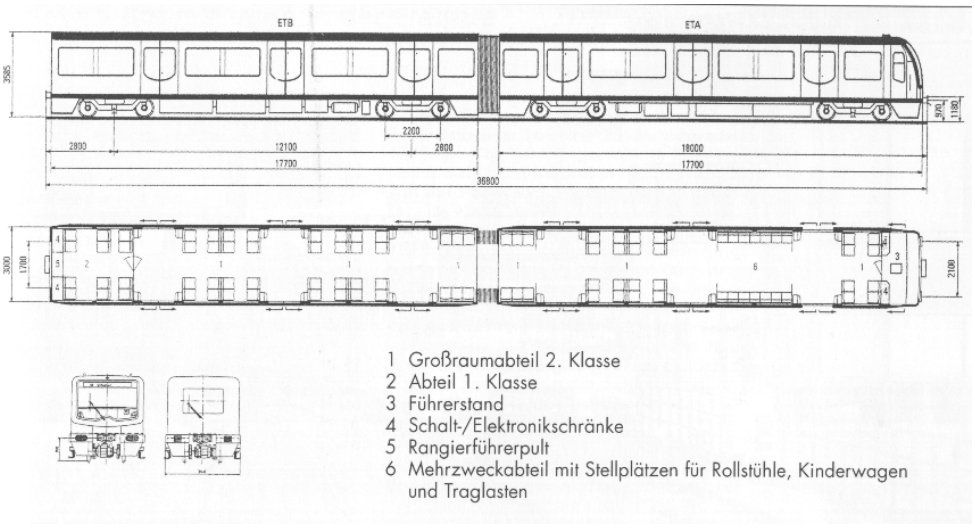

Figure 4.29: Scheme of the S-Bahn passing Spanische Allee bridge (www.berliner-verkehr.de)

For the strain measurements, it is helpful to identify the peaks not only to calculate the speed, but also to assign the average deflection induced by passing trains. For this purpose, the signal needs to be pre-processed by offset correction and by smoothing the hourglass-like shape, as exemplified in Fig. 4.30 for the record on 08-01-2012.

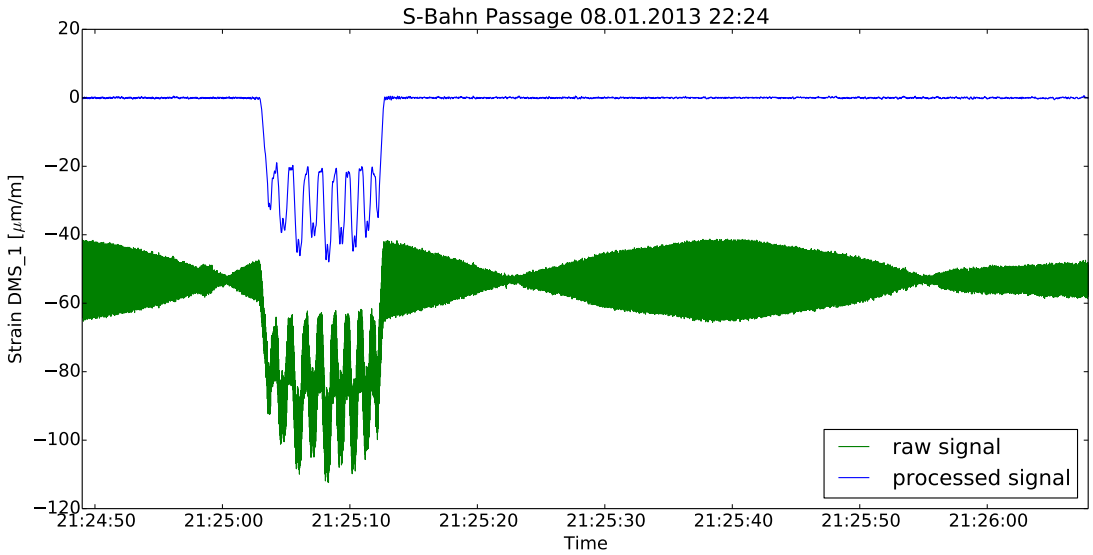

Figure 4.30: Example of a triggered strain measurement on longitudinal direction (DMS1)

Given the distance between bogies, an algorithm to detect the corresponding peaks in the time histories could be implemented. This is possible despite the limited number of strain gauges, as long as the type of passing trains and their configurations are known, which is the case in particular for Spanische Allee bridge.

After removing the noise from the signal, the peaks can be better identified and are marked with red circles in Fig. 4.31 and Fig. 4.32. Both the longitudinal and transversal strain gauges can theoretically be used. In practice, the most reliable computations proved to be those using the DMS1 strain gauge.

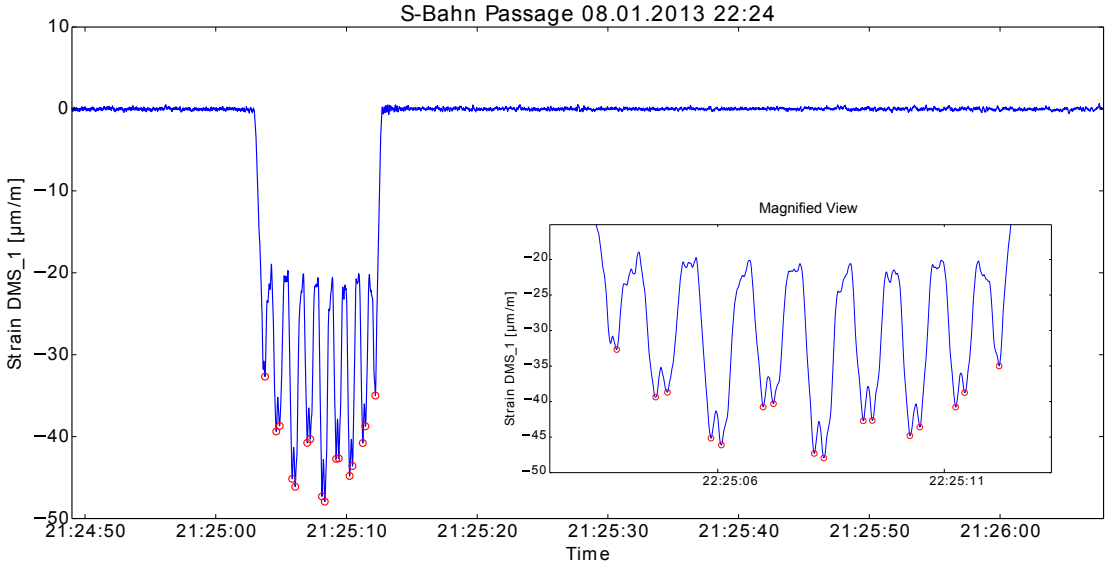

Figure 4.31: Example of peak identification - longitudinal direction

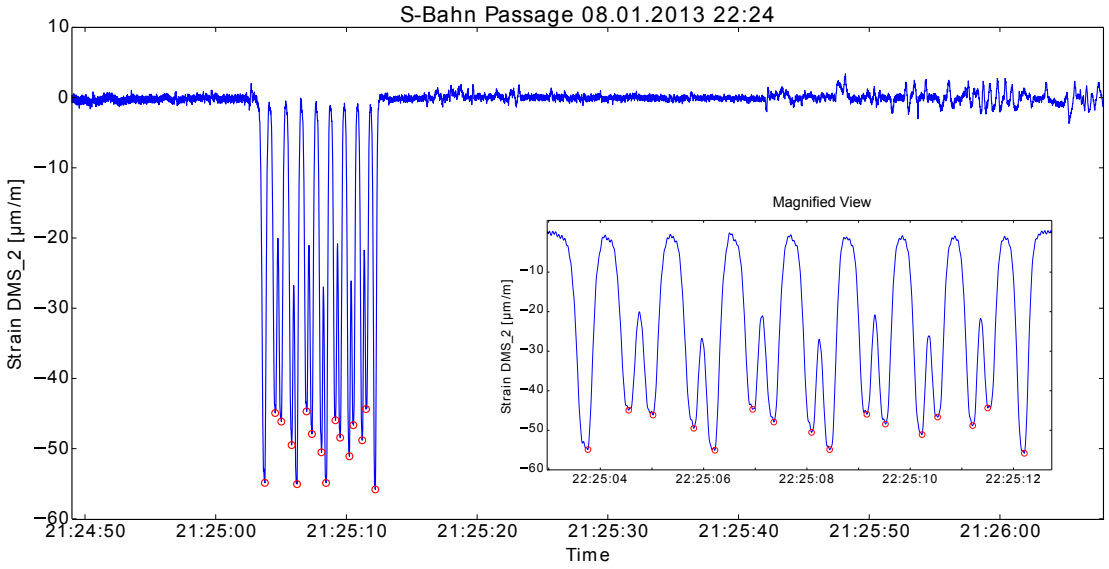

Figure 4.32: Example of peak identification - transversal direction

Although the algorithm performs very well most of the time, it is possible that sometimes the peak identification fails or yields biased estimations due to the strong dependency on the data quality and the small shifts which the filtering process may introduce.

Therefore, the speed results should be seen as rough estimations and considered from a statistical perspective.

The investigations were carried out after processing 5183 events recorded during 54 days, between 21-11-2012 and 15-03-2013. For the assessment of deformations, the mean of the peaks detected in the part of the signal corresponding to passing trains was used.

4.3.1 Influence of the Speed of Passing Trains

Since the bridge is relatively close to Nikolassee S-Bahn station (Fig. 4.1), the range of the speed results should lie below 100 km/h. Moreover, it is known that most of the passing trains are accelerating to Berlin, so an increasing trend in the velocity estimations are expected to be observed. These assertions are confirmed by the typical outcome depicted in Fig. 4.33 for several time instances on 20-02-2013.

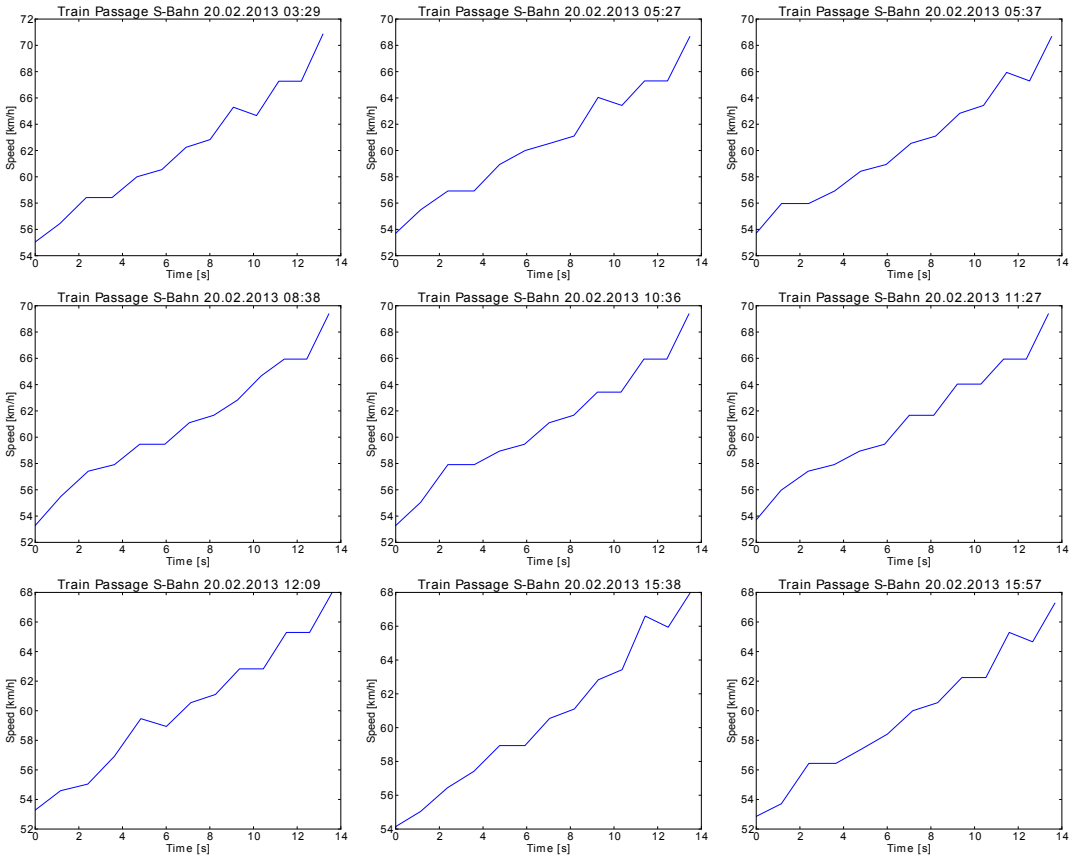

Figure 4.33: Speed profiles of passing trains for several events during 20-02-2013

The horizontal axis represents the relative time that starts with the detection of the first peak and the graphs are built using the speed assigned to every passing bogie. It can be noticed that the values are located between 52 and 70 km/h and the slope is nearly constant, which indicates an uniform acceleration.

The *shelve* files were updated with the new information concerning the speed, acceleration and strains induced by each passing train. Fig. 4.34 displays the distribution of the acceleration and speed of passing trains after removing the outliers. The mean speed has been considered representative for each passing train.

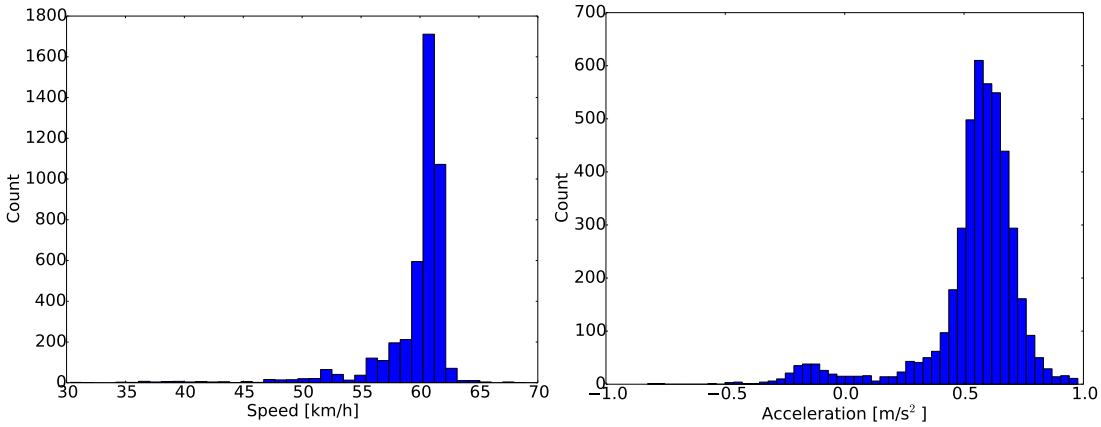

Figure 4.34: Histograms for the speed (left) and acceleration (right) results

It can be observed that the acceleration results exhibit a higher dispersion with the mean of nearly 0.6 m/s^2 and that the speed results are concentrated around 60 km/h. Consequently, the coefficient of variation in the considered range takes the value 0.384 for acceleration and 0.057 for speed. Another important remark is that sometimes the train decelerates, which means that it approaches the station from the other direction, i.e. from Berlin to Potsdam. The number of negative accelerations is small enough to indicate that these events occurred temporarily when the other lane was closed due to construction.

Especially for high speed railway lines, the dynamic problem of passing trains is related to the resonance effect induced at certain speeds. The length of the axle separation of the trains and the main eigenfrequencies of the bridges represent the key parameters. These are related by the so-called *wavelength* of excitation:

$$\lambda_w = \frac{v}{f_0}$$

where f_0 is the first natural frequency of the superstructure and v is the train speed. Resonance occurs whenever the characteristic distance between axles is a multiple of the wavelength of excitation, or in other words, the load frequency matches a multiple of the natural frequency of the structure. The worst-case scenario is perfect equality, when the vibration amplitudes strongly build up and may occur for high speed train passages in case of short-span bridges with relatively high eigenfrequencies.

Considering the results in Section 4.2.2, the critical speed for the Spanische Allee bridge was estimated for the first four harmonics of the first natural frequency and included in Table 4.5.

Table 4.5: Critical speeds for the first four harmonics of the first natural frequency

Harmonic nr.	First natural frequency [Hz]		Critical speed [km/h]	
	Min.	Max.	Min.	Max.
1	5.38	6.13	234	267
2	10.76	12.26	117	133
3	16.14	18.39	78	89
4	21.52	24.52	58	66

Even though the first and two harmonics are responsible for a significant vibration amplification, the higher order harmonics can also explain the observed structural response, which is the case for Spanische Allee bridge. As Fig. 4.35 indicates, the maximum values for strains appear at speeds around 58 km/h which correspond to the excitation of the fourth harmonic of the first natural frequency (Table 4.5).

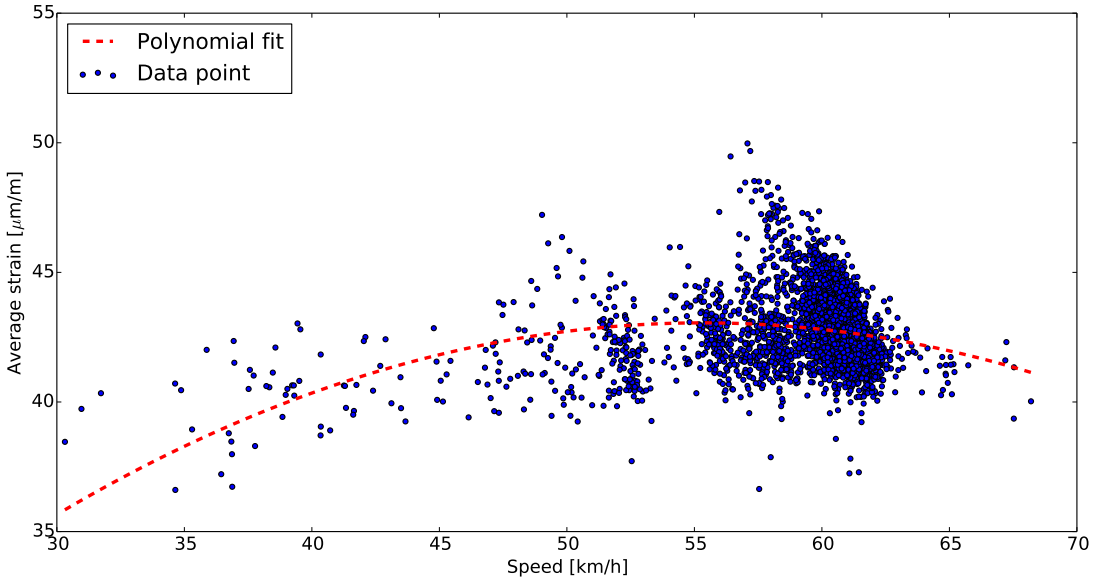

Figure 4.35: Variation of average strains with speed of passing trains

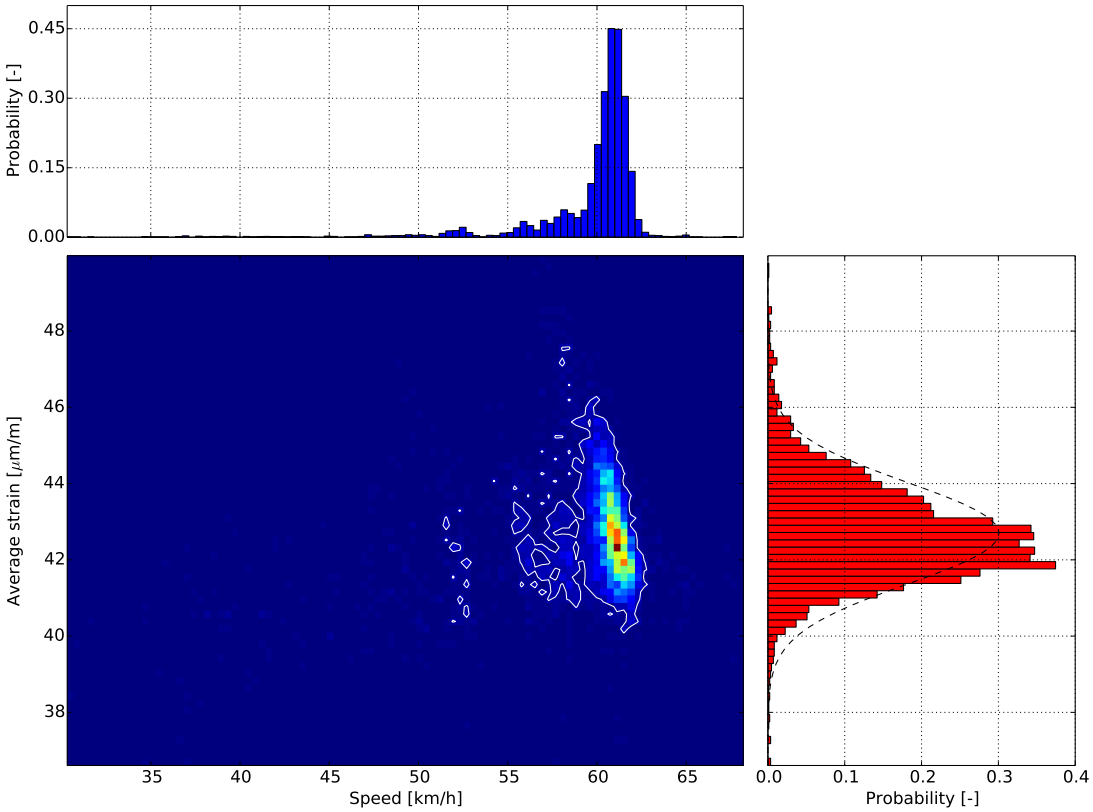

Figure 4.36: 2D histogram of average strains and speeds of passing trains white contour in the colour map - 1 standard deviation of average strains

Different loading and temperature conditions are responsible for the scatter of the strains at the same train speed. Nevertheless, the highest density can be observed at circa 60 km/h. At higher speeds, the values decrease and can be fitted by a second order polynomial regression as in Fig. 4.35, which indicates the trend of the data.

It is meaningful to describe the outcome in a 2D histogram plot, where a colour map is obtained from the density of data, as in Fig. 4.36. The maximum number of points in each bin is represented with red, and the minimum with dark blue. Both of the two histograms reflecting the speed and average strain distributions are displayed on the axes. The strain distribution complies to a normal/Gaussian distribution illustrated with a black dashed line on the corresponding histogram. As it can be observed also in Fig. 4.35, the highest density of data is reached at around 60 km/h with an average strain of about 43 μm/m.

4.3.2 Influence of Temperature

A similar representation as in Section 4.3.1 can be obtained for the average strains and temperatures, as shown in Fig. 4.37 and Fig. 4.38. Despite the noticeable dispersion explained by the other parameters which influence the relative deformation, one can remark a linear proportionality between the two variables. The maximum density is located at 0 °C for average strains of about 42 μm/m.

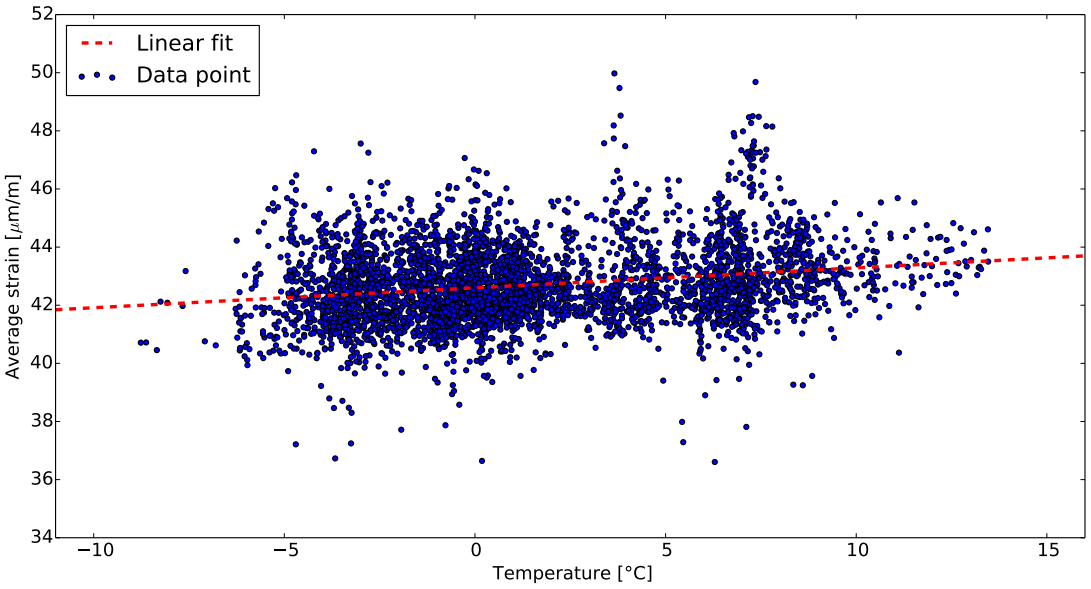

Figure 4.37: Variation of average strains with temperature

The slight increase of average strains with temperature agrees with the conclusions derived in Section 4.2.2, namely that the bridge becomes stiffer at lower temperatures, especially below the freezing point. Therefore, the same explanation regarding the supports applies for the static deflections as well.

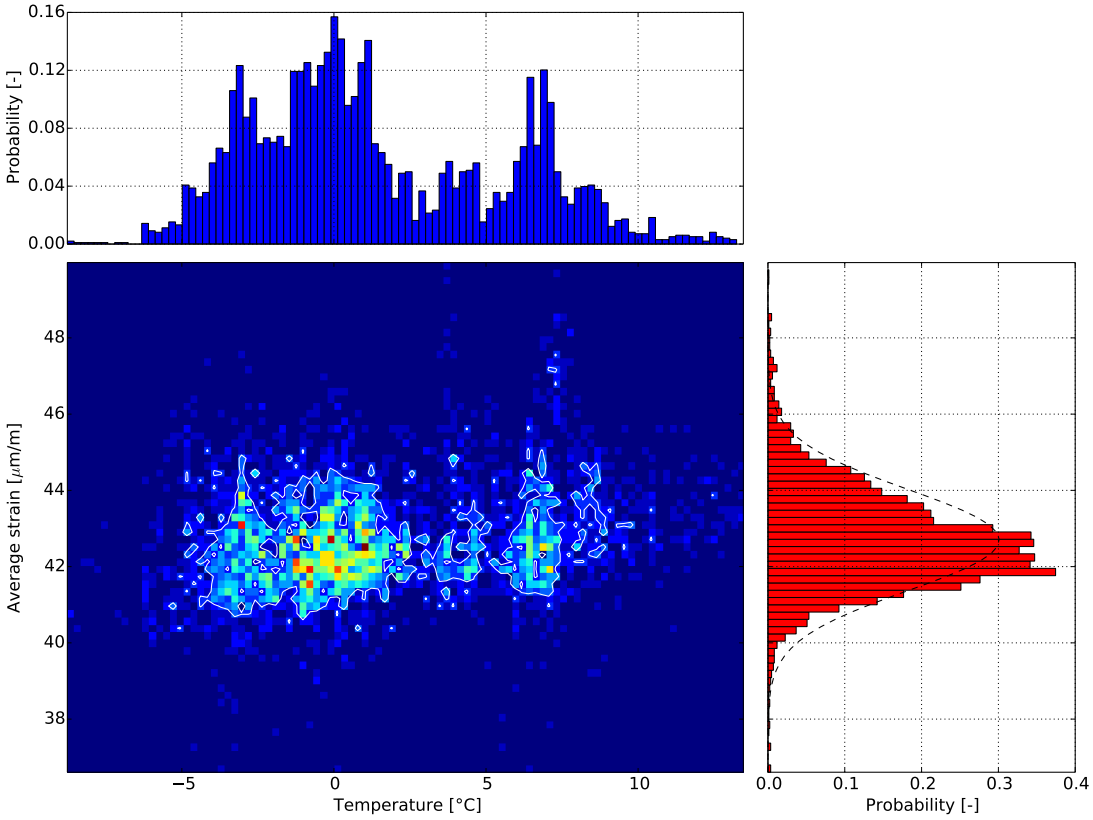

Figure 4.38: 2D histogram of average strains and temperatures
white contour in the colour map - 1 standard deviation of average strains

In order to disregard the influence of other factors, one can analyse the variation of strains induced by passing trains with known loads. This happens at those time instances when the lowest deflection is recorded and can be assumed that the number of passengers is small enough that only the dead load matters. For example, this is the case on a typical work day, at 5:00 am in the morning.

4.3.3 Influence of Loads of Passing Trains

Besides the speed and temperature, another contribution to the structural response is given by the loads of passing trains. These can not be calculated from data, but can be related directly to the strains calculated for each train either as an average of the strains produced by every bogie, or by treating every bogie separately. Due to its clarity and simplicity, the first approach was also preferred in the next few paragraphs.

The global picture of all deformations recorded for 54 days is depicted in Fig. 4.39. A closer look into data indicates a certain periodicity in the average strains evolution. Fig. 4.40 and Fig. 4.41 demonstrate this aspect for two months, February and March, respectively. The highest peaks occur during the week, while the lowest ones, which appear on the 16^{th} and 17^{th} of February and the 9^{th} and 10^{th} of March, correspond to weekends.

Since the impact of temperature is relatively small for short-term spans, these variations should be attributed solely to the influence of the loads. The pattern repeats both for

February and March, even though the temperature fluctuation in March is higher.

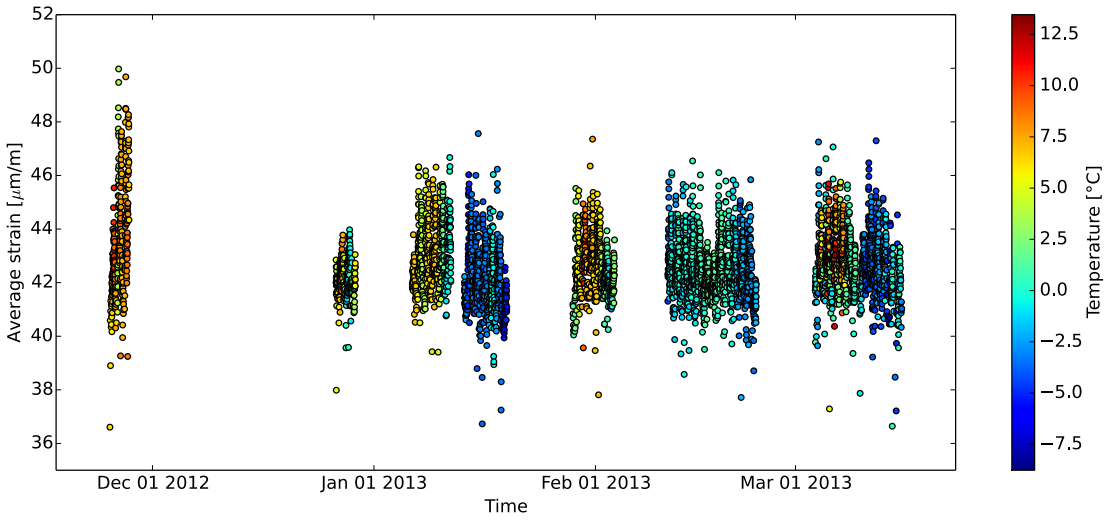

Figure 4.39: Average strains recorded during November 2012 to March 2013

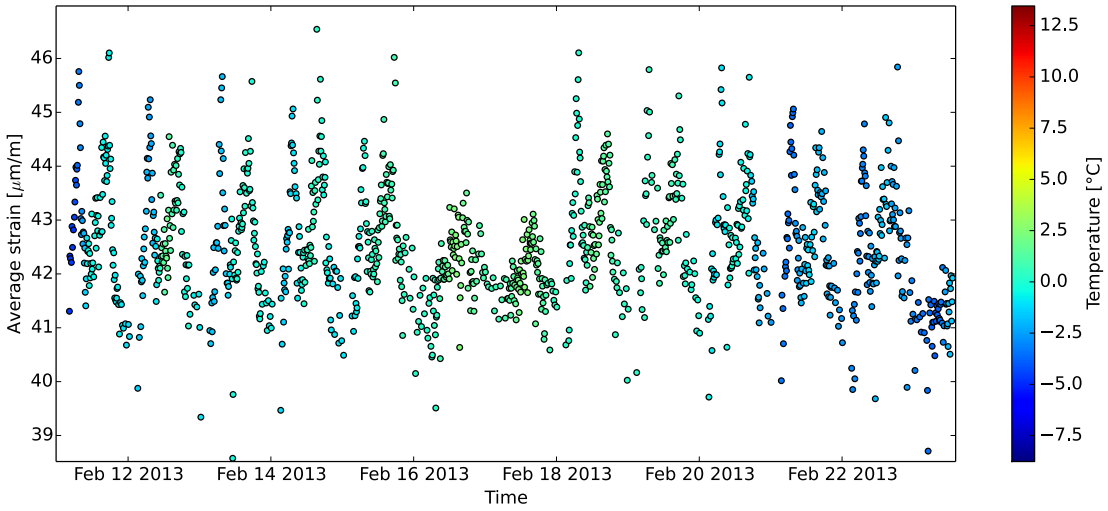

Figure 4.40: Average strains recorded during 11-02-2013 to 23-02-2013

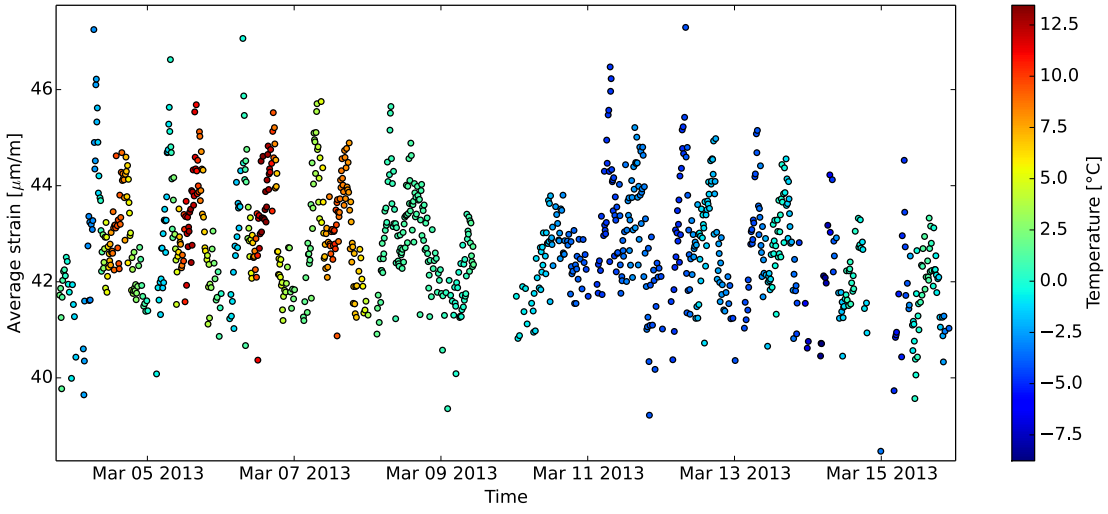

Figure 4.41: Average strains recorded during 04-03-2013 to 16-03-2013

The cycles re-appear during each day and are shown in Fig. 4.42 and Fig. 4.43. They reveal an interesting tendency which characterizes the passenger density during a typical day. Two peaks at around 7:00 am and 16:00 pm mark the rush hours.

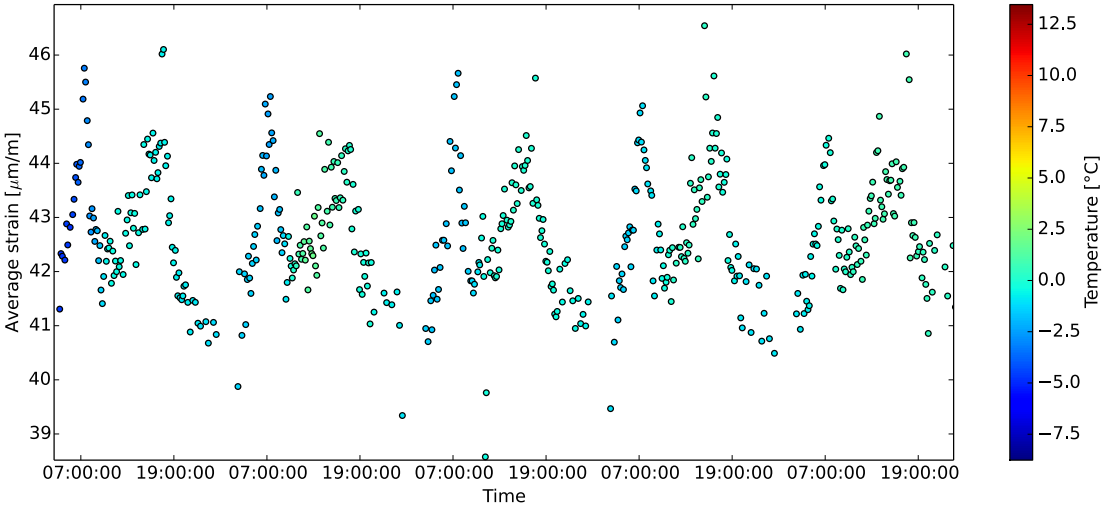

Figure 4.42: Average strains recorded during one week of February 2013

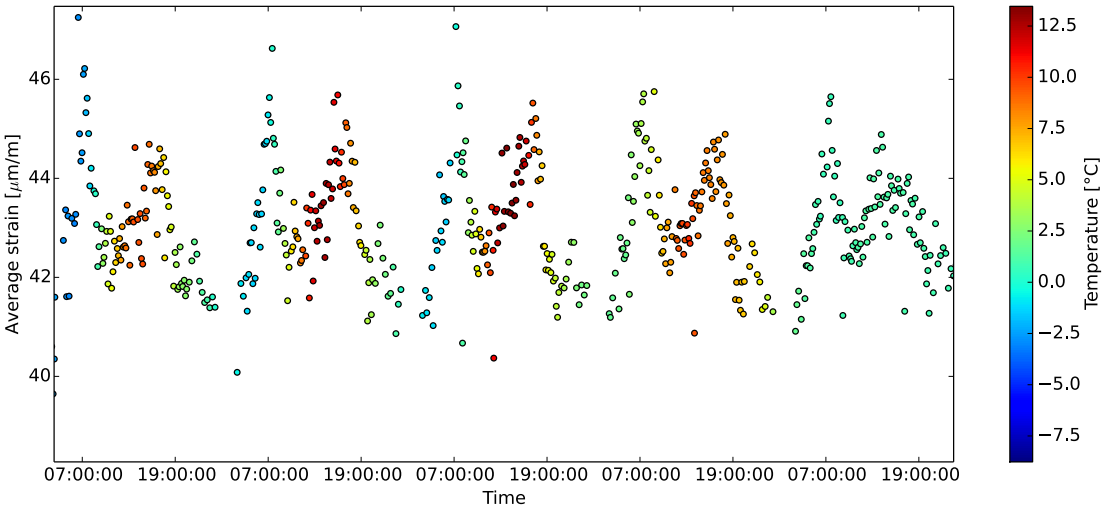

Figure 4.43: Average strains recorded during one week of March 2013

Information about daily variations can be extracted after calculating the maximum average strain difference for every single day and taking the minimum value as the reference parameter. Using this data, the range of the maximum strain variation for the whole period of monitoring can be assessed.

Fig. 4.44 illustrates the maximum differences assigned for every day of triggered measurements (blue dots). It can be seen that all values lie below 30 % of maximum variation. So far, there is no discrimination carried out between the various parameters which influence the results.

In order to refine the computation, all data was considered but the range of values was restricted to a fixed speed and temperature. Given the histograms reflecting the speed and temperature distributions in the previous sections, only the values corresponding to a speed of 60±1 km/h were retained. Table 4.6 includes the results for the relevant

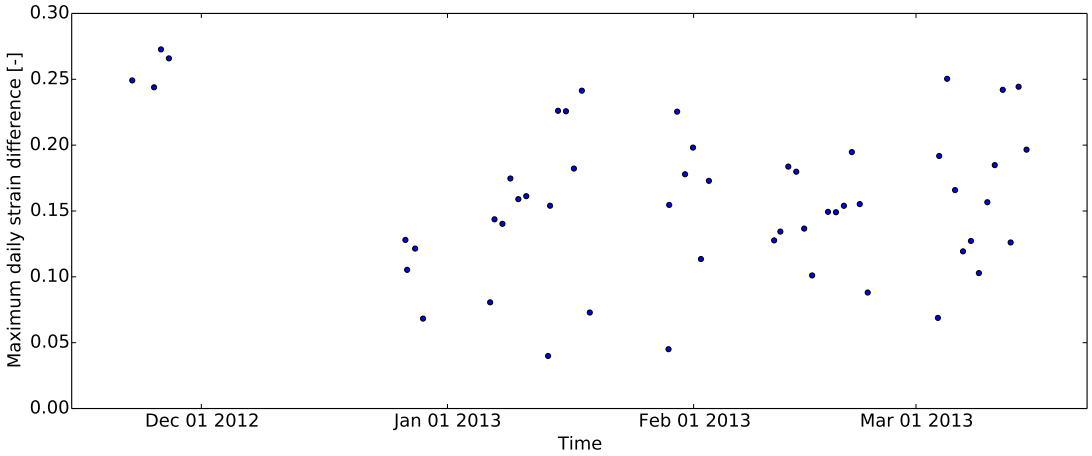

Figure 4.44: Daily maximum strain differences between November 2012 and March 2013

range of temperatures between -6 °C and +9 °C. Every entry in the table was considered within 1 °C deviation that is small enough to not affect the results and large enough to capture a sufficient number of events (1819).

Table 4.6: Maximum average strain differences for trains passing at 60±1 km/h

Temperature ±0.5 [°C]	Nr. of events	Max. average strain diff. [%]
-5.5	25	13.7
-4.5	64	17.4
-3.5	169	13.0
-2.5	155	17.4
-1.5	194	15.1
-0.5	255	22.0
+0.5	223	12.9
+1.5	138	11.3
+2.5	64	9.2
+3.5	57	10.1
+4.5	83	12.9
+5.5	58	13.8
+6.5	154	14.4
+7.5	87	15.5
+8.5	93	15.2
	Weighted mean:	15.0

The weighted mean of the results yields to a maximum strain difference due to loads of 15.0 %. This valuable information helps define the loads in a numerical model more

accurately by taking their uncertainty into account. Moreover, the maximum strain difference induced by trains passing at 60±1 km/h is 22.8% for the whole range of temperatures, which includes both the influences of temperature and loads. After estimating the load variation at 15.0 % and subtracting it from the global value, it leads to a strain variation due to temperature of about 7.8 %. This is smaller than the fluctuation of the eigenfrequencies, 14 % (Table 4.4). It can be concluded that the natural frequencies are more sensitive to temperature than deflections due to passing trains.

Chapter 5

Conclusions and Future Research

The scope of the thesis is to develop and apply various tools to process and extract information from long-term measurements acquired by a monitoring system at a railway bridge. For this purpose, the focus has been split into two parts: (automatic) operational modal analysis and assessment of the structural response to passing trains.

The core algorithm for system identification is the reference-based covariance-driven stochastic subspace identification (SSI-cov/ref) that allowed to estimate the modal parameters from ambient vibration measurements. In order to manage all data and track the changes of eigenvalues in time, a fully automatic procedure based on a three-stage clustering algorithm has been implemented. This was achieved after presenting the theoretical background for both the system identification method and the main clustering techniques. The results were finally assessed in the context of environmental changes.

The main observations are enumerated as follows:

- Although the SSI-cov/ref algorithm is one of the most robust and efficient system identification methods, the assessment of meaningful results is not always straightforward. The main issue is the interpretation of the stabilization diagrams which prove to be sensitive to several input parameters. A proper choice of the size of block matrices and the system order has a strong influence on the distribution of poles. Even for an automatic approach, there is a need for fine-tuning these decisive parameters.

- How the signal is processed has also an impact on the results quality. Decimating the signal helps to better identify poles in lower frequency ranges, but also introduces additional non-physical (spurious) poles that need to be manually inspected. Additionally, a poor signal quality leads to highly scattered poles which makes a reliable selection impossible.

- When dealing with large amounts of data, an automatic procedure is crucial. The implemented scheme relies on a three-stage combination of K-means and agglomerative hierarchical clustering algorithms, which are independent from any user interaction. Its robustness and accuracy was proven after inspecting the results for different measurements.

- All the records for a time span between September 2012 and March 2013 were processed for tracking the change of modal parameters due to temperature or construction stages. Although the quality of the outcome in September is poor, a subsequent increase of 2 of the eigenfrequencies when the structure was put on the bearings and after the ballast was added could be observed.
- In order to assess the sensitivity of modal parameters to climatic changes, the mode tracking was carried out further for the 3 modes which exhibited an identification rate higher than 80 %. It was demonstrated that the passing trains which were sometimes recorded affected the quality of results and the corresponding part of the signal needed to be removed. A better identification was achieved and it allowed us to correlate the temperature with the eigenfrequencies fluctuations. All 3 eigenfrequencies tend to increase for negative temperatures and decrease for positive ones with a variation that ranges from 14 % to 22 %.

The second part of the analysis has involved the various influences that had an impact on the structural response to passing trains. These were investigated separately for speed, temperature and loads. All the conclusions were derived using the average strain computed after identifying the peaks induced by each passing train. They can be summarized as follows:

- Most of the speed results are concentrated around 60 km/h and the corresponding strains are higher than for lower speeds. A possible explanation is given by the observation that the speed matches the estimated critical speed for the fourth harmonic of the first natural frequency.
- A linear proportionality between temperature and average strains was noticed. The tendency is that the strains increase with temperature and this agrees with the phenomenon observed for eigenfrequencies, namely that the flexibility of the supports is dependent on the rails elongation and/or contraction.
- The loads of the passing trains could not be directly assigned. However, the variations of strains reflected a certain periodicity, with two peaks occurring regularly during the rush hours. The maximum strain differences computed for each day by taking the minimum values as reference parameters, lie below 30 %. Further discrimination was possible after fixing the temperature and speed and revealed an average variation due to loads of 15 %. This result is extremely useful for estimating the load uncertainties in a numerical model.

Since only the data between September 2012 and March 2013 were analysed, the same procedures should be extended for all the remaining data recorded by the monitoring system which has not yet been converted.

Several improvements for future development can be considered. The code created for the automatic modal parameter estimation, is not a standalone software with a proper graphical user interface (GUI). Consequently, the application of the Python 3 routines is not very straightforward and it is solely based on input and output files. Visualization panels for the signals and results should be included as well. Additionally, plotting procedures for the mode shapes estimations should be further implemented.

Another issue is related to the computational speed of the agglomerative hierarchical clustering algorithm. Assembling the proximity matrix proved to be computationally

expensive and different methods to perform the clustering of the poles can be applied. For example, the clustering can be carried out by comparing the results between two consequent system orders, instead of considering all data. Also, more sophisticated algorithms could be introduced.

The problematic of dealing with the results provided by different setups was not treated in this thesis. The difficult task of merging mode shapes obtained from several setups, without a prior knowledge of their correspondence, is another important and actual research topic.

The damping ratio evaluation possesses a high degree of uncertainty, as it proved to be dependent on the input parameters. Together with a sensitivity analysis, an attempt to increase the accuracy of estimation should be carried out.

For damage identification purposes, the effects of environmental and operational factors on natural frequencies should be minimized. Various static and dynamic multiple linear regression models are useful statistical tools proposed in the literature and their application would be of high interest. The cases of modal splitting, harsh environmental conditions and/or strong damage in the context of modal tracking, are still open issues which need to be addressed.

References

[1] P. G. Bakir. Automation of the stabilization diagrams for subspace based system identification. *Expert Systems with Applications*, 38(12):14390 – 14397, 2011.

[2] A. Cabboi, F. Magalhães, C. Gentile, and Cunha. Á. Automatic Operational Modal Analysis: Challenges and Practical Application to a Historical Bridge. In *6th ECCOMAS Conference on Smart Structures and Materials*, 2013.

[3] E. P. Carden and J. M. W. Brownjohn. Fuzzy clustering of stability diagrams for vibration-based structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering*, 23(5):360–372, 2008.

[4] European Commission. *European high-speed rail - an easy way to connect*, 2009.

[5] Á. Cunha, E. Caetano, F. Magalhães, and C. Moutinho. *From Input-Output to Output-Only Modal Identification of Civil Engineering Structures*. Final Report F11, SAMCO Association, Europe, 2006.

[6] D.J. Ewins. *Modal Testing, 2nd edition*. Research Studies Press, Baldock, England, 2000.

[7] W. Gawronski. *Advanced Structural Dynamics and Active Control of Structures*. Springer-Verlag, New York, USA, 2004.

[8] J. Han and M. Kamber. *Data Mining: Concepts and Techniques, 2nd edition*. Morgan Kaufmann Publishers, San Francisco, USA, 2001.

[9] W. Heylen, S. Lammens, and P. Sas. *Modal Analysis Theory and Testing*. KULeuven, Belgium, 1998.

[10] B.L. Ho and R.E. Kalman. Effective construction of linear state-variable models from input–output data. *Regelungstechnik*, 14:545–548, 1966.

[11] J.-N. Juang. *Applied System Identification*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1994.

[12] F. Magalhães. *Operational Modal Analysis for Testing and Monitoring of Bridges and Special Structures*. PhD thesis, Universidade do Porto, Faculdade de Engenharia, Portugal, 2010.

[13] H.G. Natke. *Einführung in Theorie und Praxis der Zeitreihen- und Modalanalyse*. Vieweg, Braunschweig, Germany, 1992.

[14] R. S. Pappa, G. H. James III, and D. C. Zimmerman. *Autonomous Modal Identification of the Space Shuttle Tail Rudder*. Report NASA TM-112866, National Aeronautics and Space Administration, June 1997.

[15] B. Peeters. *System Identification and Damage Detection in Civil Engineering*. PhD thesis, Katholieke Universiteit Leuven, Belgium, 2000.

[16] B. Peeters and G. De Roeck. Reference based stochastic subspace identification for output-only modal analysis. *Mechanical Systems and Signal Processing*, 13(6):855–878, 1999.

[17] B. Peeters and G. De Roeck. One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering & Structural Dynamics*, 30(2):149–171, 2001.

[18] E. Reynders, J. Houbrechts, and G. De Roeck. Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing*, 29(0):228 – 250, 2012.

[19] E. Reynders, M. Schevenels, and G. De Roeck. *MACEC 3.2: A MATLAB Toolbox for Experimental and Operational Modal Analysis*. Report BWM-2011-01, Katholieke Universiteit Leuven, February 2011.

[20] M. Scionti, J. Lanslots, I. Goethals, A. Vecchio, H. Van Der Auweraer, B. Peeters, and B. De Moor. Tools to Improve Detection of Structural Changes from In-Flight Flutter Data. In *International Conference on Recent Advances in Structural Dynamics*, Southampton, UK, 2003.

[21] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Pearson Education Inc., Boston, USA, 2006.

[22] P. Vacher, B. Jacquier, and A. Bucharles. Extensions of the MAC Criterion to Complex Modes. In *ISMA 2010 International Conference on Noise and Vibration Engineering*, 2010.

[23] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory - Implementation - Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[24] P. Verboven, E. Parloo, P. Guillaume, and M. Van Overmeire. Autonomous Structural Health Monitoring. Part I: Modal Parameter Estimation and Tracking. *Mechanical Systems and Signal Processing*, 16(4):637 – 657, 2002.

[25] V. Zabel. *Experimental Structural Dynamics*. Lecture notes, Bauhaus-Universität Weimar, Fakultät Bauingenieurwesen, Weimar, Germany, 2011.

[26] V. Zabel, S. Höll, and H. Bigelow. *Brücke EÜ Spanische Allee, Auswertung der numerischen und experimentellen Analyse V1*. FOSTA Project DYNABRIDGE P 941, Bauhaus-Universität Weimar: Institut für Strukturmechanik, RWTH Aachen: Institut für Stahlbau, August 2012.

[27] M. Zacher. Dynamic of Railway Bridges. In *5th ADAMS/Rail Users' Conference*, Haarlem, Netherlands, 2000.

Appendices

Appendix A

Documentation

The purpose of this appendix is to clarify the structure of the Python code written for modal parameter estimation (MPE) and help any eventual user who wants to apply or further develop the available routines.

A.1 Input and Output Files

All the input parameters are stored in a *json* file which is read as a dictionary whenever an instance of the object is created. Every information can be accessed by its keys and is enclosed in curly brackets. How the file is organized is exemplified in Fig. A.1 for the monitoring system at the Spanische Allee bridge. These settings can be appended for any other system under a different main key (location name).

Config SSI.txt

```
"Spanische_Allee":
   "Data":
        "Channels":["GP_1","GP_2"],
        "Samples": "Start": [], "End": [],
        "Sampling rate": 500,
        "Process": "Decimate": [], "Filter": "no", "Smooth": "Window": "", "Length": "",
   "System identification":
        "Channels assignment": "Reference": [1, 2], "Roving": [1, 2],
        "Model parameters": "Model order": 100, "Block rows": 150, "Block columns": 150,
        "Stabilization criteria": "Frequency": 0.01, "Damping": 0.05, "MAC": 0.99,
        "Clustering parameters":
                "Kmeans  iteration": 2000,
                "Hierarchical": "Weight frequency": 1.5, "Weight MAC": 0.5,
                                "Intercluster distance": []
```

Figure A.1: Example of an input file

Firstly, the data and the signal processing tools are included under the "Data" key. Secondly, everything related to the SSI-cov/ref algorithm is mentioned under the "System identification" entry.

All settings will be listed in the following paragraphs.

DATA

- **Channels**: the name of the channels which stem from the *shelve* files introduced exactly as they appear in the database.
- **Samples**: the length of the signal; takes the whole signal if nothing is specified
  - *Start*: the number of the samples that can be skipped; for *None* leave empty square brackets [ ]
  - *End*: the number of the samples that are considered; for all until the end leave empty square brackets [ ]
- **Sampling rate**: the sampling rate of the measurement [Hz]
- **Process**: several signal processing tools
  - *Decimate*: the decimation factor used when applying an eighth-order low-pass Chebyshev Type I filter in both forward and reverse directions to remove all phase distortion; for *None* leave empty square brackets [ ]
  - *Filter*: smooths the data convoluting the signal with a scaled window; accepts "yes" or "no" as arguments
    - Window: the type of the window applied (flat, hanning, hamming, bartlett, blackman); default is hamming
    - Length: an odd integer which is the length of the smoothing window; default is 11

SYSTEM IDENTIFICATION

- **Channels assignment**: defines the type of channels for computing the correlation matrices
  - *Reference*: a list with the indexes of the reference sensors
  - *Roving*: a list with the indexes of the roving sensors
- **Model parameters**: reads the key parameters used to compute the modal estimations
  - *Model order*: the maximum system order
  - *Block rows*: the number of block rows for assembling the Toeplitz matrix
  - *Block columns*: the number of block columns for assembling the Toeplitz matrix; recommended the same as the number of block rows
- **Stabilization criteria**: the allowable distances between two consecutive model order solutions; relevant for the manual interpretation of the stabilization diagram
  - *Frequency*: the maximum normalized difference in terms of frequencies
  - *Damping*: the maximum normalized difference in terms of damping
  - *MAC*: the maximum difference in terms of MAC values

- **Clustering parameters**: various parameters to be used in the clustering process
  - *Kmeans iteration*: the number of iterations of the K-means algorithm to run
  - *Hierarchical*: several options taken by the agglomerative hierarchical clustering algorithm
    * Weight frequency: assign the relative importance of the frequency distance when calculating the distance measure; for equal consideration insert 1
    * Weight MAC: assign the relative importance of the MAC distance when calculating the distance measure; for equal consideration insert 1
    * Intercluster distance: user defined cut-off distance; for the autonomous computation leave empty square brackets [ ]

After the identification of the modal parameters, different files will be generated. The solutions of the SSI-cov/ref method will be stored in a *shelve* file that can be easily opened in Python for further processing. Additionally, the manual or automatic selection will create a *text* file with all information and can be accessed on any platform. One possible output file after a user interpretation is illustrated in Fig. A.2.

Modal info manual.txt

```
MANUAL MODAL ANALYSIS
=====================
Time:                       2012-09-13 13:12:47+02:00
Frequencies [Hz]:           [3.79423839616017, 7.6167604741677497, 8.3653170344719019]
Damping [%]:                [3.7537850422509984, 1.0683505483712121, 0.45176254088839207]
Mode shapes:
[[ -9.02952423e-05-0.00032941j   7.30642708e-04+0.00281197j]
 [ -3.49389159e-03-0.00041739j   1.85531627e-03+0.0003115j ]
 [  3.90129030e-04+0.00791054j  -4.37530742e-03-0.04250885j]]
Model order:            [66, 85, 45]
df  [-]:                [3.1586190129701217e-06, 7.3336965616551837e-05, 1.3589258527872597e-06]
dxi [-]:                [0.00030151427756427702, 0.0081024102675919633, 0.0056533335233676148]
MAC [-]:                [0.99999762127468628, 0.99793933922057088, 0.99999997270281404]
MPC [-]:                [0.99999044269351534, 0.99844461202972135, 0.99963566131417414]
MP  [°]:                [75.424312836581862, 7.4163076242344754, 84.224634379077386]
MPD [-]:                [0.0009840703889336601, 0.012557025351947559, 0.0060773441254039808]

SSI parameters
======================
Maximum order :             100
Block rows :                150
Block columns :             150
Decimation :                1
Filtering :                 None
```

Figure A.2: Example of an output file after manually selecting the modes of interest

It should be noted that the components of the mode shape vectors are displayed horizontally and every line corresponds to a specific mode. The time information is converted from the timestamp of the measurements.

Similar to the previous file, all necessary information is stored for the automatic detection, as in Fig. A.3. Due to paper size constrains, although the algorithm identified nine modes, only the first four values are included in the presented output file. For the frequency, damping and mode shape indicators, the standard deviation within the selected clusters exhibits low values.

Modal info autom.txt

```
AUTOMATIC MODAL ANALYSIS
========================

Time:                        2012-09-13 13:12:47+02:00
Frequencies [Hz]:            [3.7934697200079119, 7.6062795240104268, 8.3661253781046376, 10.20429713483674]
Damping [%]:                 [3.7503975842025952, 0.61039135698273683, 0.44908546886644646, 0.72204609378862539]
Mode shapes:
[[  8.89491218e-05+0.00030103j  -7.31644054e-04-0.00280565j]
 [ -4.89423901e-03-0.00073693j   4.24816190e-03-0.00024839j]
 [ -3.55773158e-04-0.00791008j   4.30478403e-03+0.04249422j]
 [ -1.13369342e-04+0.00047152j  -3.90918773e-03+0.00022896j]]
Number of poles:         [60, 63, 96, 73]
df  [-]:                 [2.4042640659915734e-05, 3.0950850832260129e-05, 8.1725883490092474e-06, 4.0714471427393213e-05]
dxi [-]:                 [0.019205021212255962, 0.013320217482355253, 0.0059258388635684434, 0.041819090381895276]
MAC [-]:                 [0.99999275081883054, 0.99735467373352449, 0.99999975984819467, 0.99999480482479164]
MPC [-]:                 [0.9999524776062908, 0.95837418764706628, 0.99959732290578374, 0.94550219985660944]
MP  [°]:                 [75.362821744341133, 3.5138129777640961, 84.321937038511521, -3.5992481249799071]
MPD [-]:                 [0.0021945685127455769, 0.065402139174779328, 0.0063892706326382775, 0.091681179034126717]

SSI parameters
======================
Maximum order :          100
Block rows :             150
Block columns :          150
Decimation :             2
Filtering :              None
Cut-off distance :       0.00470916071883
Weight freq dist :       1.5
Weight MAC :             0.5

Standard deviation
======================
Frequencies [-]:         [0.01475809058872527, 0.0099913873411479406, 0.016669946287197226, 0.0090236460830511109]
Damping [-]:             [0.16477633268829076, 0.011319141462420006, 0.20805774967816201, 0.050164951198957128]
MAC [-]:                 [1.1046043506415703e-05, 0.0001042736944537756, 0.00073240945213146904, 0.050225455730115248]
MPC [-]:                 [0.00020889715438584606, 0.0082682164399577283, 0.0056039542065790746, 0.0015317241954485267]
MPD [-]:                 [0.0012654115679763037, 0.0088611038528181897, 0.011899749410097603, 0.0056339607983704265]
```

Figure A.3: Example of an output file after automatically selecting the modes

A.2 Software Components

The concept of the code is to allow a flexible use of different functionalities. In the context of object-oriented programming, this can be easily achieved once a class with different methods has been created. The diagram in Fig. A.4 displays the available functions that can be called after an instance of the object is created.

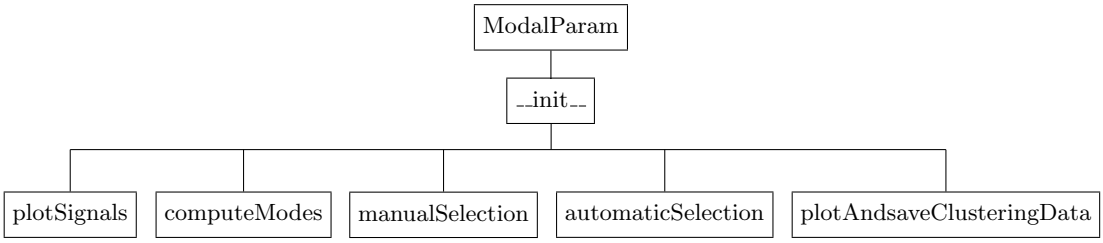

Figure A.4: Flowchart of the main class with its methods

As shown, any instance of the **ModalParam** object passes through the *"__init__"* method, using the information given in the required arguments:

measurement, config, timestamp, write_SSI, write_m, write_a,
read_SSI (optional), read_shelve_a (optional), format_plot (optional)

where:

- *measurement* can be either a Monalyse dictionary or a simple *numpy* array of $n \times m$ dimensions with n the number of samples and m the number of channels.
- *config* is the *json* file previously introduced which stores all the necessary configuration
- *timestamp* is the Windows timestamp of the measurement
- *write_SSI* is the name of the *shelve* where to store all the results from the SSI-cov/ref calculation
- *write_m* is a two-dimensional list containing two names: the *text* and *shelve* files where to write the modal information from the manual interpretation of the stabilization diagram
- *write_a* is a two-dimensional list containing two names: the *text* and *shelve* files where to write the modal information from the automatic analysis
- *read_SSI* has the same entries as *write_SSI* and is used for reading the results, for example to perform the automatic analysis.
- *read_shelve_a* is the name of the file used to read the information stored from a previously automatic computation.
- *format_plot* is the extension of the saved figure (supported formats: emf, eps, pdf, png, ps, raw, rgba, svg, svgz)

If the last three entries are not explicitly mentioned, the program will try to read the same *shelves* which were created under *write_SSI* and *write_a* and the default extension for saving the plots will be *pdf*. In this way, the user has the flexibility to control what the code reads for computation and the methods are decoupled as much as possible. All the functions are explained below and allow to be called separately whenever a specific result is sought.

- *"plotSignals"* creates a plot with two graphs, one for the original signal and the other for the eventually processed signal
- *"computeModes"* calculates all the solutions given by the SSI-cov/ref algorithm
- *"manualSelection"* starts the interactive widget using the stability criteria defined in the input file
- *"automaticSelection"* begins the clustering process; it requires the results obtained after calling the "computeModes" method
- *"plotAndsaveClusteringData"* shows all relevant plots for the autonomous MPE

Moreover, once an analysis has been completed, the results can be accessed and/or visualized at any time, without the need to re-calculate everything. All definitions are contained in

SSIcovRef.py

which is imported by another file where several routines can be included, like for example indicating the project path where to generate all the folders, reading the configuration files or running a loop over all measurements. The name of this Python file is suggestively called

Main_SSIcovRef.py

Appendix B

Influence of Input Parameters

The SSI-cov/ref algorithm requires several pre-defined parameters which are normally assigned after multiple runs. The purpose of this appendix is to picture how the decimation can affect the identified poles and what are the effects when applying different stabilization criteria in the context of a manual interpretation of the results.

B.1 Decimation

Different decimation factors were applied for the record on *13-09-2012, 16:09:59+02:00*.

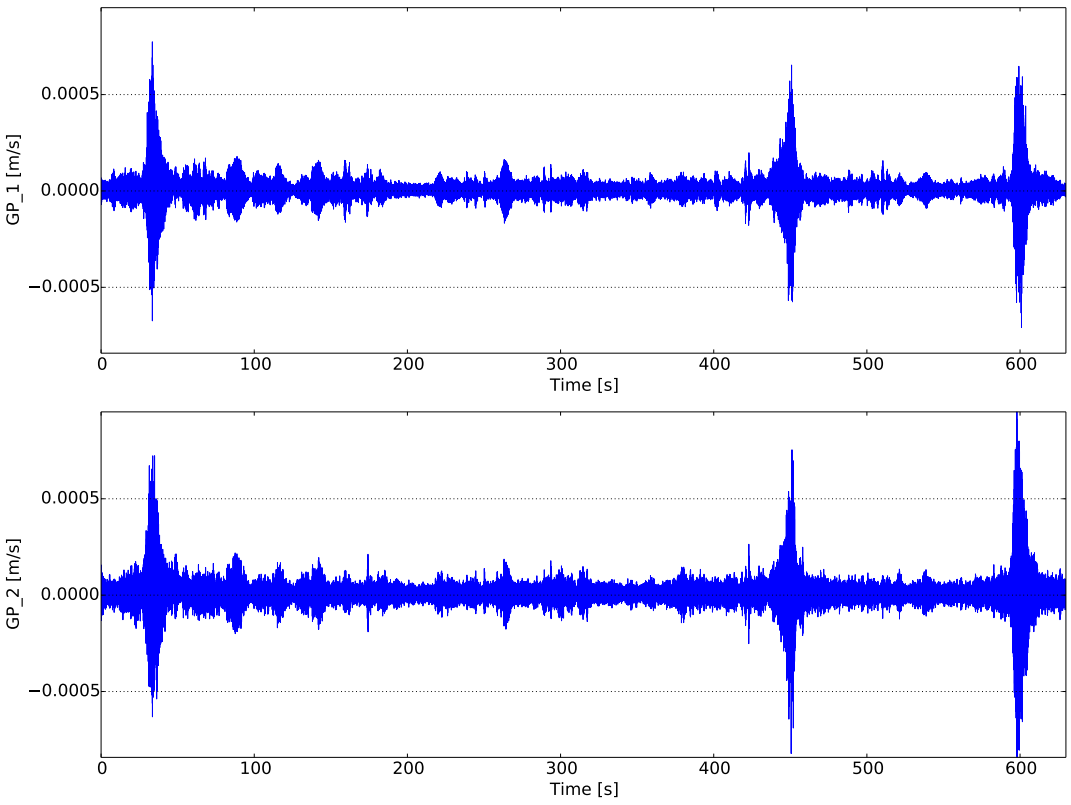

Figure B.1: The records for GP1 and GP2 without using decimation

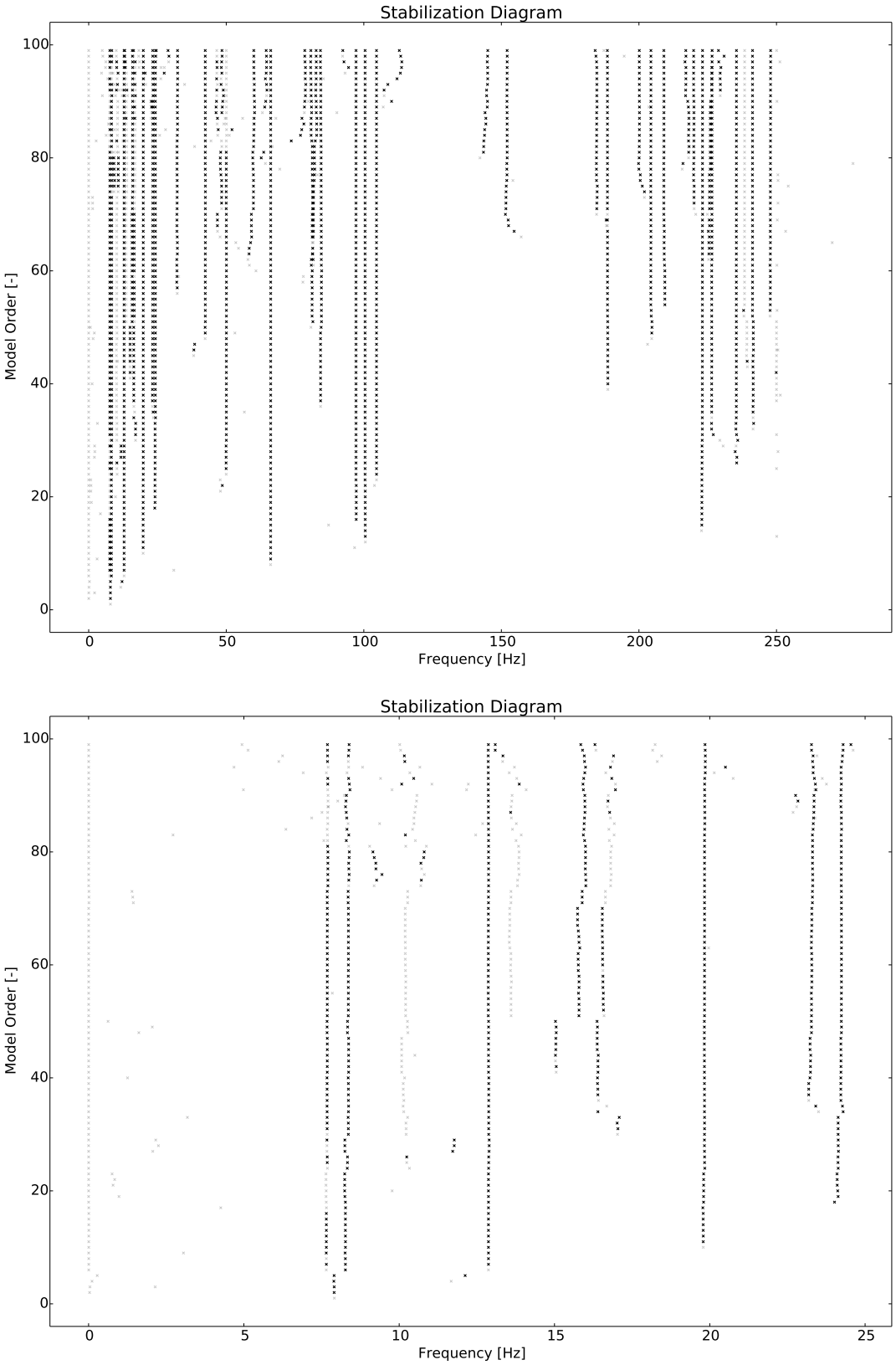

Figure B.2: The stabilization plot for the whole frequency range (above) and for the frequency range 0 Hz to 25 Hz (below);
black-possibly physical poles, grey-certainly spurious poles, no decimation

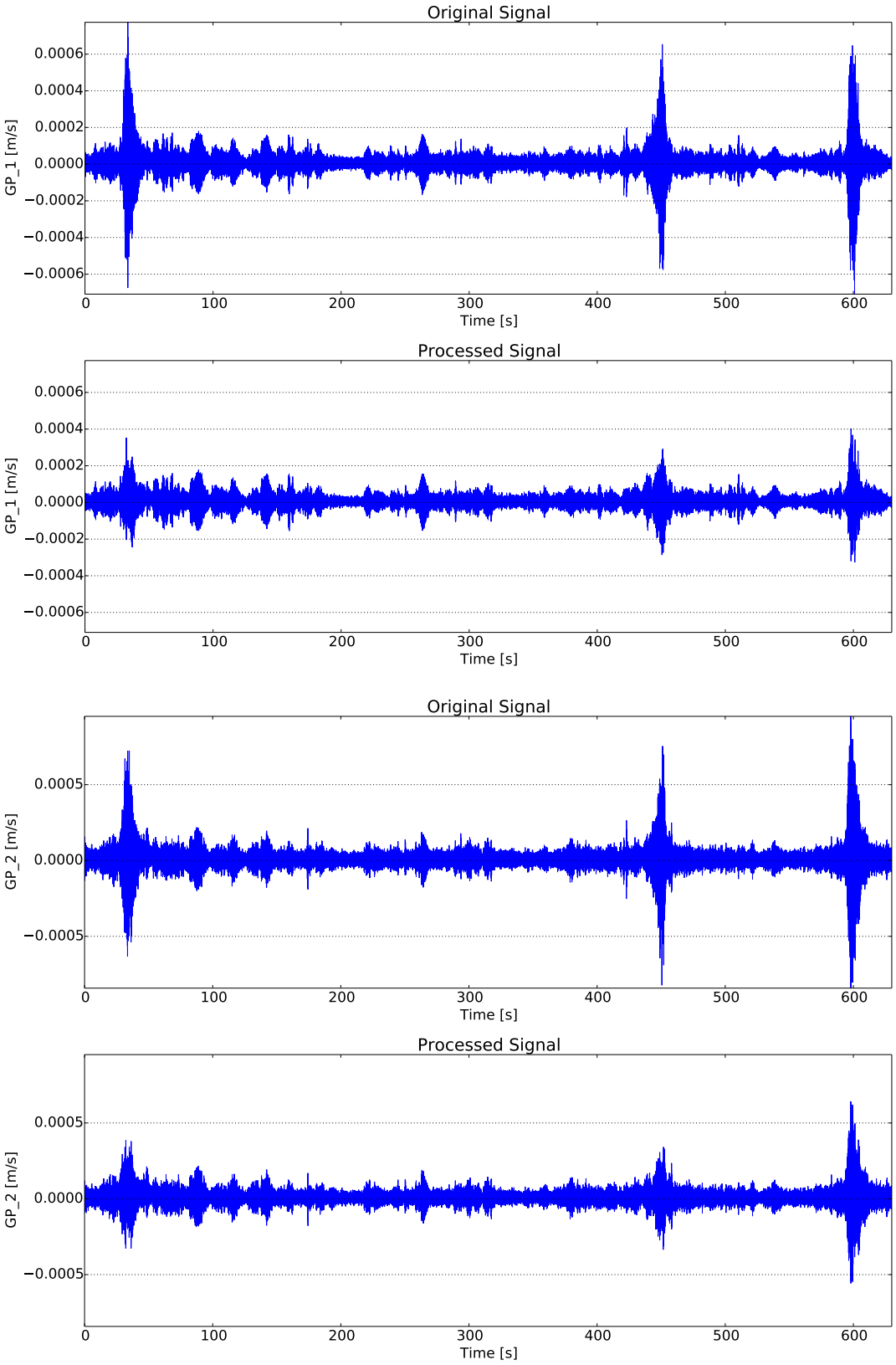

Figure B.3: The records for GP1 and GP2 using a decimation factor of 4

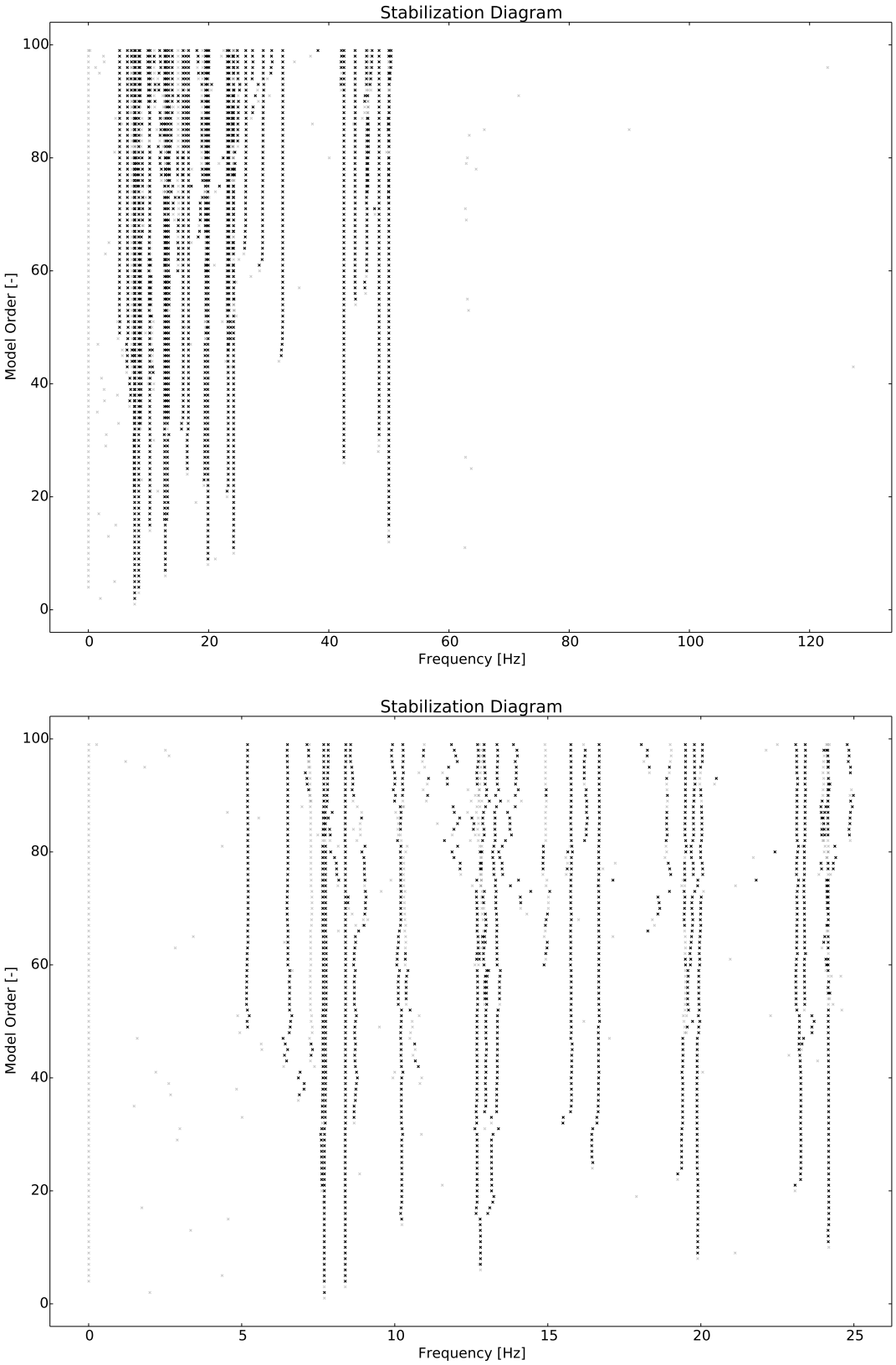

Figure B.4: The stabilization plot for the whole frequency range (above) and for the frequency range 0 Hz to 25 Hz (below);
black-possibly physical poles, grey-certainly spurious poles, decimation factor 4

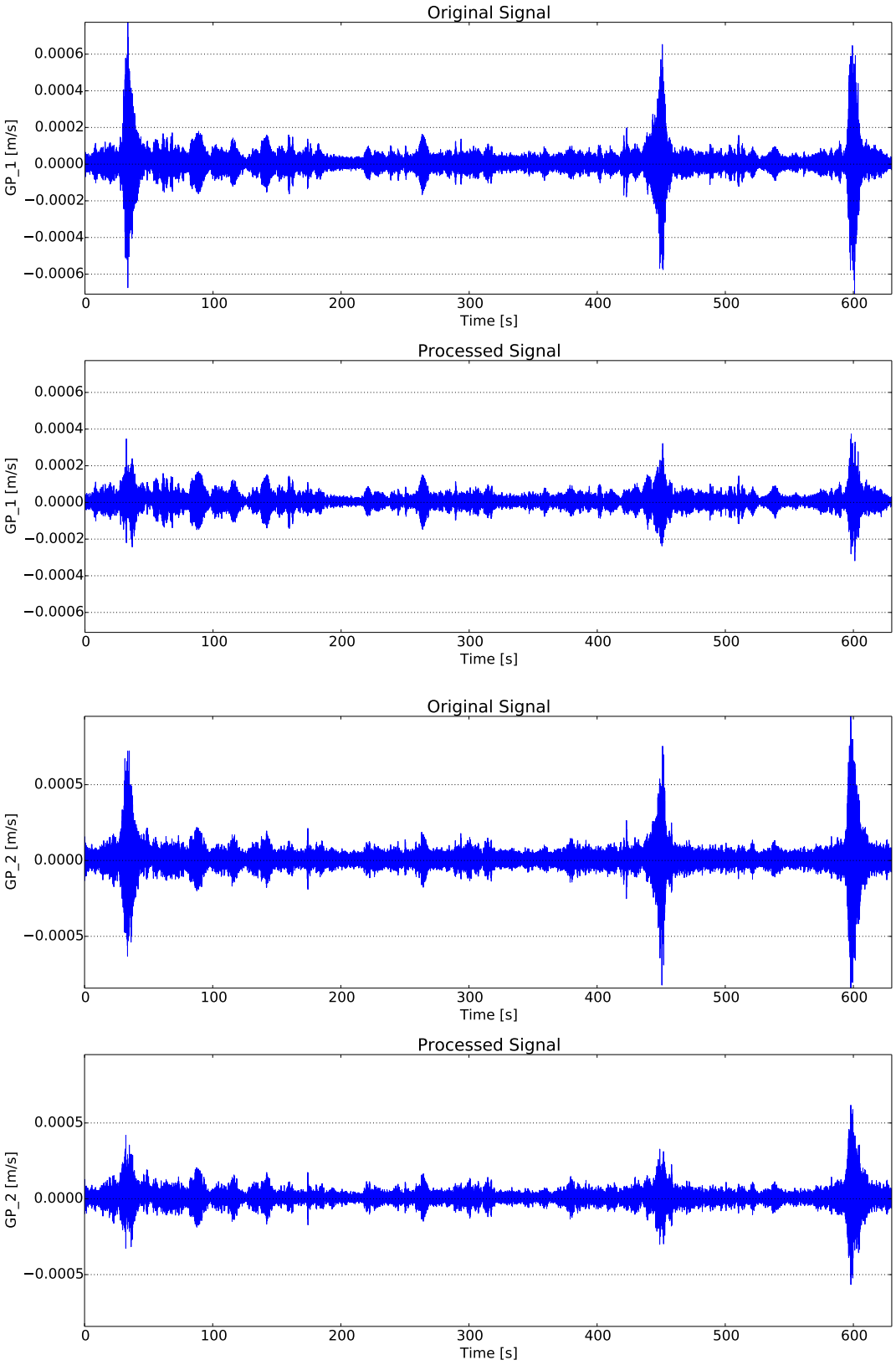

Figure B.5: The records for GP1 and GP2 using a decimation factor of 8

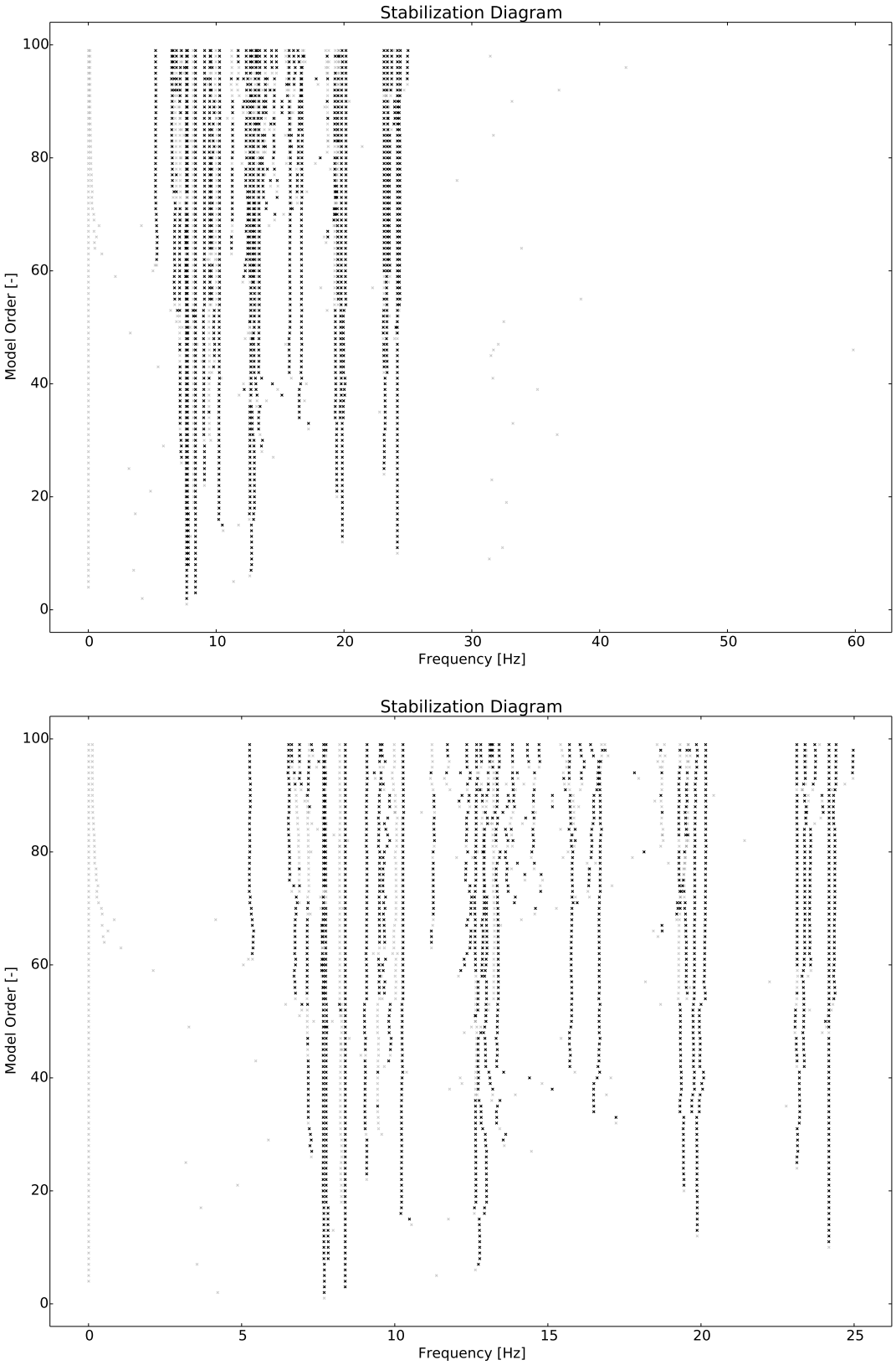

Figure B.6: The stabilization plot for the whole frequency range (above) and for the frequency range 0 Hz to 25 Hz (below);
black-possibly physical poles, grey-certainly spurious poles, decimation factor 8

B.2 Stabilization Criteria

The same record at *13-09-2012, 16:09:59+02:00* was chosen to illustrate the effect of different stabilization criteria on the SSI-cov/ref results. The results are shown for the frequencies up to 25 Hz.

As expected, the higher the threshold values, the less stable poles can be selected. Due to the high MAC values in the case of measurements taken with a limited number of sensors, the stabilization criterion for the mode shapes is not a very sensitive parameter. On the other hand, given its scatter, the parameter related to the damping ratios has a strong influence and is usually allowed to take higher values than the similar measure for frequencies.

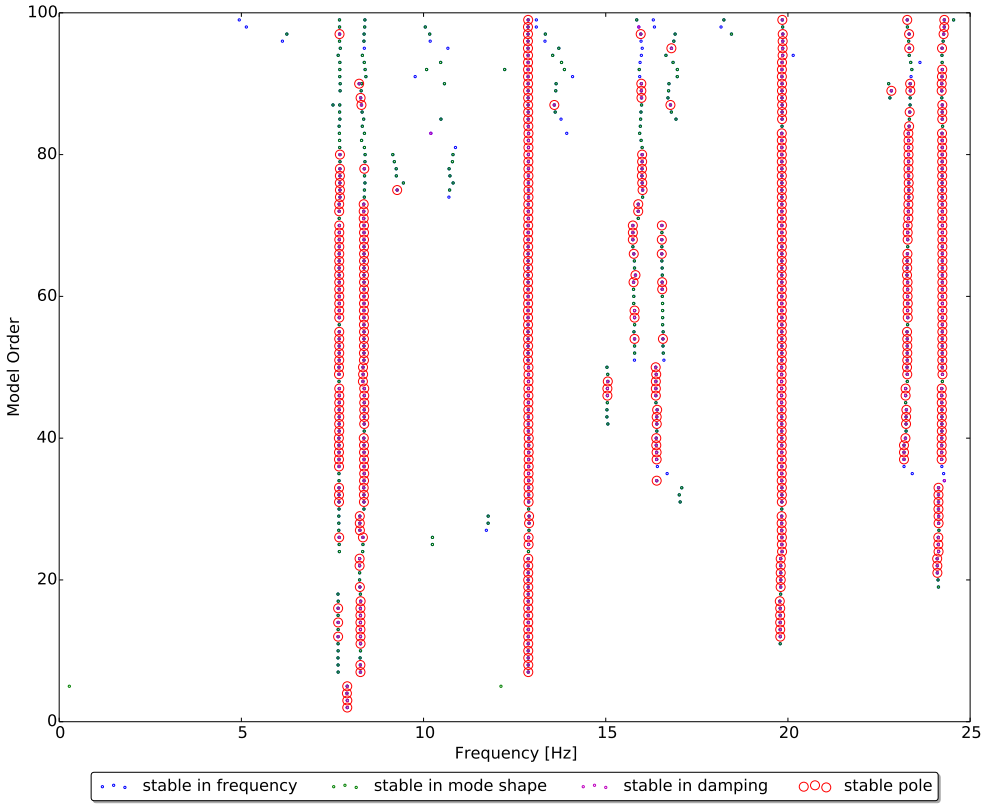

Figure B.7: Stabilization criteria: 1% frequency, 5% damping ratio, 98% MAC

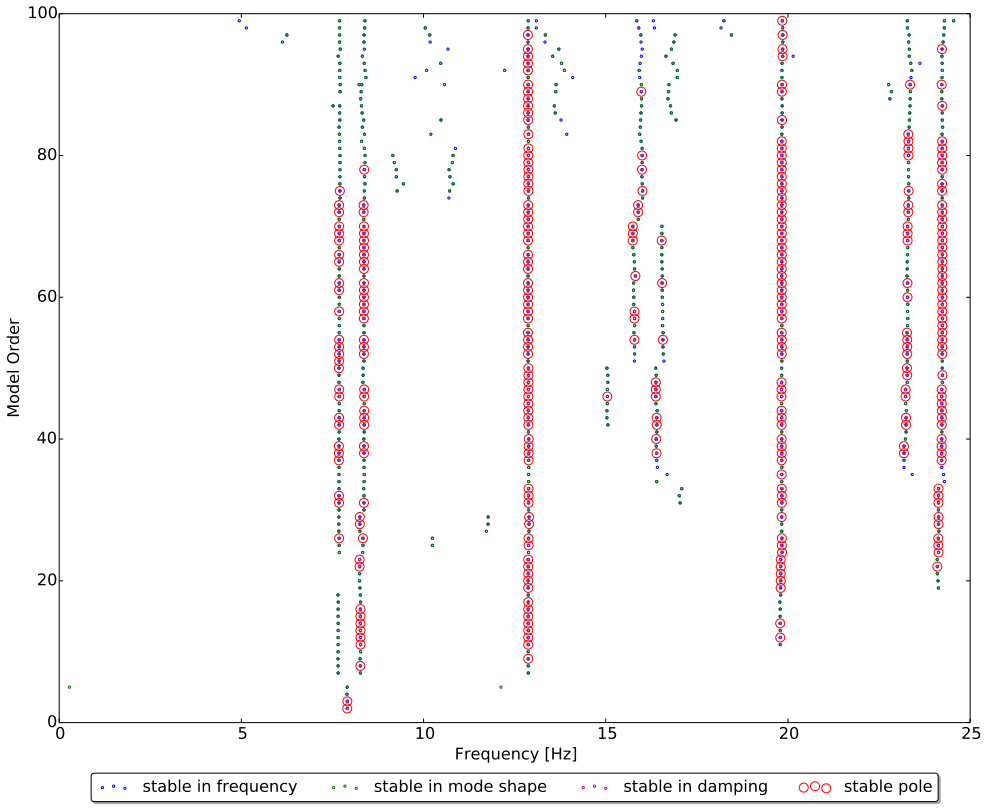

Figure B.8: Stabilization criteria: 1% frequency, 1% damping ratio, 98% MAC

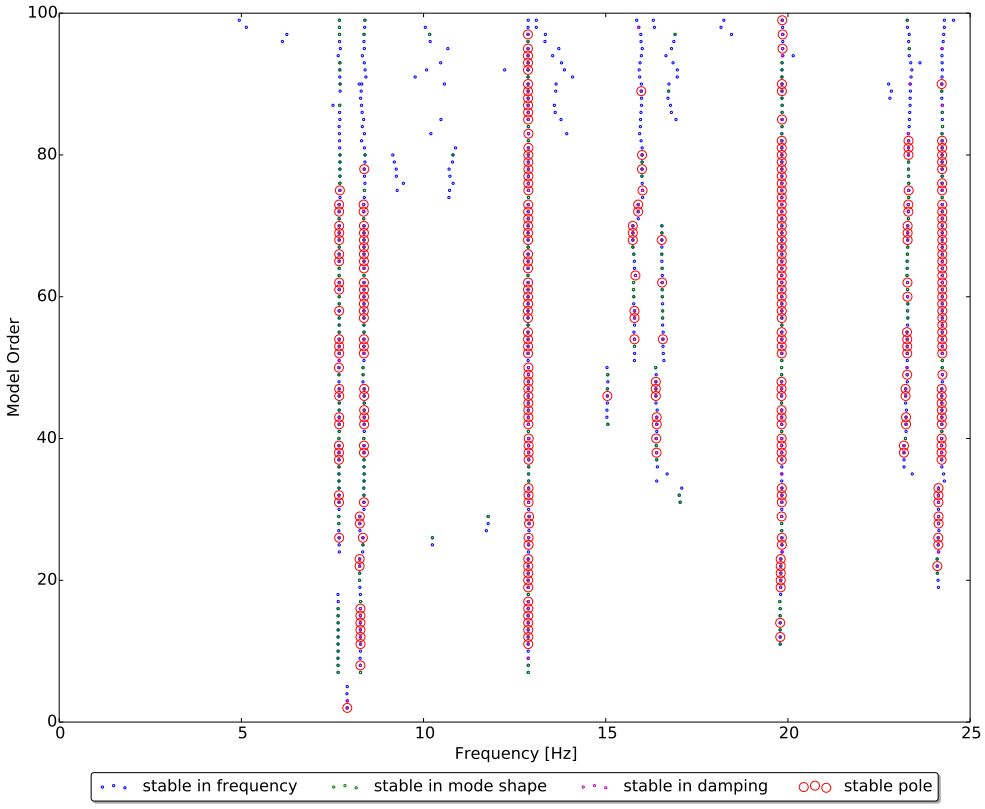

Figure B.9: Stabilization criteria: 1% frequency, 1% damping ratio, 99.99% MAC

Appendix C

Automatic Identification for One Day of Measurements

13-09-2012 13:12:47+02:00

Decimation: 1; Weight frequency distance: 1.5; Weight MAC distance: 0.5

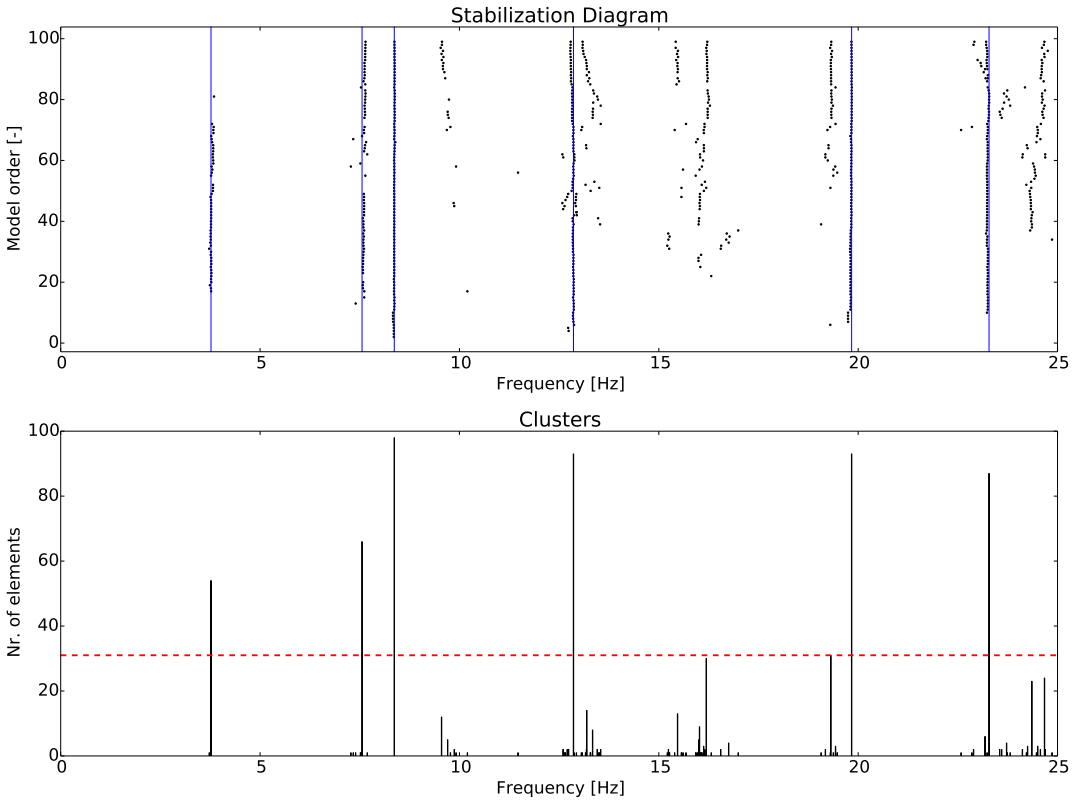

Figure C.1: The final set of physical modes (above) with their corresponding clusters (below) for the record at 13:12:47

Cut-off intercluster distance: 0.0062

Table C.1: Several results from the automated identification for the record at 13:12:47

Poles	Frequency		Damping ratio		MAC		MPC	
	Repr.[Hz]	Std.[-]	Repr.[%]	Std.[-]	Repr.[-]	Std.[-]	Repr.[-]	Std.[-]
54	3.769	0.00921	0.917	0.07065	0.999	0.00006	0.998	0.00048
66	7.557	0.01945	1.740	0.03684	0.964	0.00036	0.912	0.01152
98	8.366	0.03391	0.510	0.30030	0.999	0.02357	0.999	0.02900
93	12.86	0.01642	2.652	0.14868	0.997	0.04375	0.957	0.01326
93	19.83	0.02220	0.764	0.58602	0.999	0.00489	0.989	0.02830
87	23.28	0.02553	0.523	0.31478	0.999	0.00060	0.994	0.00775

Repr. - the representative value within a cluster Std. - standard deviation

13-09-2012 13:49:47+02:00

Decimation: 1; Weight frequency distance: 1.5; Weight MAC distance: 0.5

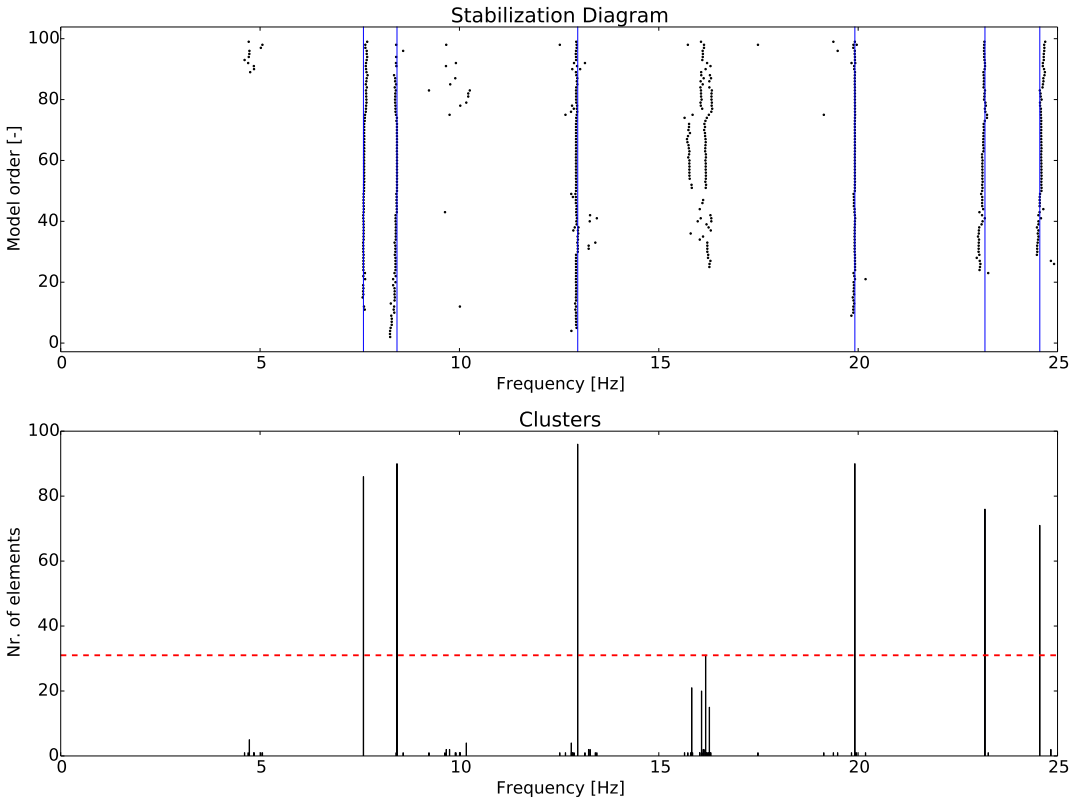

Figure C.2: The final set of physical modes (above) with their corresponding clusters (below) for the record at 13:49:47

Cut-off intercluster distance: 0.0089

Table C.2: Several results from the automated identification for the record at 13:49:47

Poles	Frequency		Damping ratio		MAC		MPC	
	Repr.[Hz]	Std.[-]	Repr.[%]	Std.[-]	Repr.[-]	Std.[-]	Repr.[-]	Std.[-]
86	7.592	0.02007	1.590	0.23486	0.999	0.02467	0.996	0.00859
90	8.433	0.04445	1.201	0.38994	0.999	0.00137	0.998	0.00488
96	12.96	0.01558	2.073	0.05895	0.997	0.00550	0.999	0.00262
90	19.91	0.03060	0.884	0.17914	0.999	0.00196	0.996	0.00515
76	23.17	0.05521	1.193	0.16692	0.999	0.00783	0.919	0.03757
71	24.55	0.05105	1.250	0.12958	0.999	0.00179	0.897	0.05283

Repr. - the representative value within a cluster Std. - standard deviation

13-09-2012 16:09:59+02:00

Decimation: 1; Weight frequency distance: 1.5; Weight MAC distance: 0.5

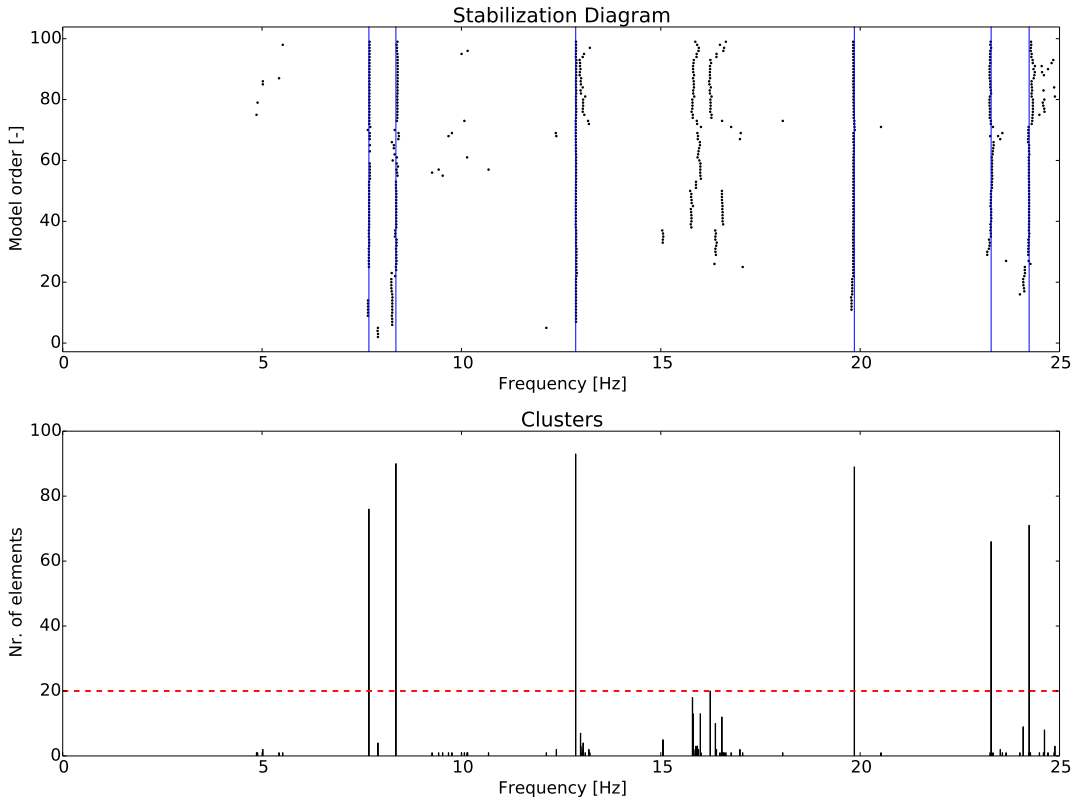

Figure C.3: The final set of physical modes (above) with their corresponding clusters (below) for the record at 16:09:59

Cut-off intercluster distance: 0.0056

Table C.3: Several results from the automated identification for the record at 16:09:59

Poles	Frequency		Damping ratio		MAC		MPC	
	Repr.[Hz]	Std.[-]	Repr.[%]	Std.[-]	Repr.[-]	Std.[-]	Repr.[-]	Std.[-]
76	7.679	0.00511	0.472	0.06937	0.999	0.00069	0.997	0.00922
90	8.356	0.05292	1.708	0.24887	0.999	0.01158	0.994	0.00234
93	12.86	0.01501	1.886	0.08010	0.997	0.00037	0.993	0.00770
89	19.85	0.01249	1.010	0.12803	0.999	0.00069	0.975	0.00471
66	23.28	0.04734	0.807	0.14371	0.999	0.00127	0.970	0.02890
71	24.23	0.03109	0.729	0.10515	0.999	0.00343	0.699	0.01596

Repr. - the representative value within a cluster Std. - standard deviation

13-09-2012 19:09:59+02:00

Decimation: 1; Weight frequency distance: 1.5; Weight MAC distance: 0.5

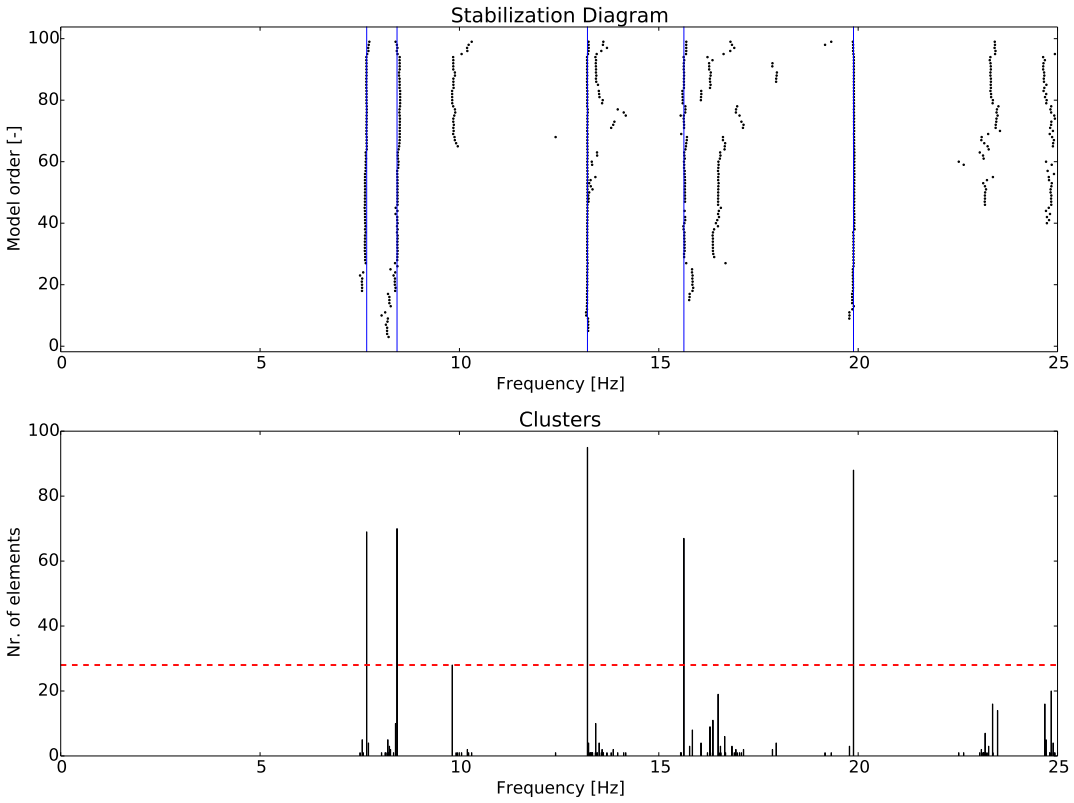

Figure C.4: The final set of physical modes (above) with their corresponding clusters (below) 19:09:59

Cut-off intercluster distance: 0.0043

Table C.4: Several results from the automated identification for the record at 19:09:59

Poles	Frequency		Damping ratio		MAC		MPC	
	Repr.[Hz]	Std.[-]	Repr.[%]	Std.[-]	Repr.[-]	Std.[-]	Repr.[-]	Std.[-]
69	7.674	0.00968	1.858	0.13441	0.999	0.00023	0.962	0.00618
70	8.434	0.01427	1.652	0.03558	0.999	0.00019	0.994	0.00155
93	13.21	0.02806	2.229	0.33865	0.997	0.00095	0.989	0.00273
67	15.62	0.01801	1.267	0.18091	0.999	0.00117	0.999	0.01016
88	19.88	0.02511	0.970	0.33859	0.999	0.00578	0.998	0.01828

Repr. - the representative value within a cluster Std. - standard deviation

Statement of Authorship

I hereby affirm that this thesis has been written by myself and is based on my own work. No illegitimate appliances have been used. All sources have been referenced and are listed in the bibliography. Where illustrations from foreign sources have been used, they have been marked accordingly.

Mihai-Andrei Udrea

Weimar, February 2014