

FEDERAL UNIVERSITY OF OURO PRETO

DEPARTMENT OF CIVIL ENGINEERING

POSTGRADUATE PROGRAM IN CIVIL ENGINEERING

APPROACHES FOR AUTOMATED DAMAGE DETECTION IN
STRUCTURAL HEALTH MONITORING

RHARÃ DE ALMEIDA CARDOSO

OURO PRETO

2018

UNIVERSIDADE FEDERAL DE OURO PRETO

DEPARTAMENTO DE ENGENHARIA CIVIL

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

TÉCNICAS PARA DETECÇÃO AUTOMÁTICA DE DANOS NO
MONITORAMENTO DA INTEGRIDADE ESTRUTURAL

RHARÃ DE ALMEIDA CARDOSO

OURO PRETO

2018

APPROACHES FOR AUTOMATED DAMAGE DETECTION IN STRUCTURAL
HEALTH MONITORING

Thesis presented to the Postgraduate Program
in Civil Engineering, Department of Civil
Engineering, Federal University of Ouro Preto,
as partial fulfillment of the requirements for
the degree of Doctor in Civil Engineering,
area: Structures and construction.

Advisor: Alexandre Abrahão Cury

Co-advisor: Flávio de Souza Barbosa

Ouro Preto
Department of Civil Engineering
2018

TÉCNICAS PARA DETECÇÃO AUTOMÁTICA DE DANOS NO MONITORAMENTO
DA INTEGRIDADE ESTRUTURAL

Tese apresentada ao programa de Pós-Graduação do Departamento de Engenharia Civil da Escola de Minas da Universidade Federal de Ouro Preto, como parte dos requisitos mínimos para obtenção do título de Doutor em Engenharia Civil, área de concentração: Estruturas e construção.

Orientador: Alexandre Abrahão Cury

Coorientador: Flávio de Souza Barbosa

Ouro Preto

Departamento de Engenharia Civil da UFOP

2018

C268t

Cardoso, Rharã de Almeida.

Técnicas para detecção automática de danos no monitoramento da integridade estrutural [manuscrito] / Rharã de Almeida Cardoso. - 2018.
143f.: il.: color; grafqs; tabs.

Orientador: Prof. Dr. Alexandre Abrahão Cury.

Coorientador: Prof. Dr. Flávio de Souza Barbosa.

Tese (Doutorado) - Universidade Federal de Ouro Preto. Escola de Minas.
Departamento de Engenharia Civil. Programa de Pós-Graduação em Engenharia Civil.

Área de Concentração: Construção Metálica.

1. Monitoramento da Integridade Estrutural. 2. Detecção de Danos. 3. Identificação Modal. I. Cury, Alexandre Abrahão. II. Barbosa, Flávio de Souza. III. Universidade Federal de Ouro Preto. IV. Título.

CDU: 624.014

TÉCNICAS PARA DETECÇÃO AUTOMÁTICA DE DANOS NO MONITORAMENTO DA INTEGRIDADE ESTRUTURAL

AUTOR: RHARÃ DE ALMEIDA CARDOSO

Esta tese foi apresentada em sessão pública e aprovada em 23 de novembro de 2018, pela Banca Examinadora composta pelos seguintes membros:



Prof. Alexandre Abrahão Cury, D. Sc. – UFJF (Presidente)



Prof. Flávio de Souza Barbosa, D. Sc. – UFJF (Coorientador)



Prof. Francisco de Assis das Neves, D. Sc. – UFOP



Prof. Carlos Cristiano Hasenclever Borges, D. Sc. – UFJF



Prof. Diogo Rodrigo Ferreira Ribeiro, D. Sc. – ISEP/FEUP



Prof. José Fernando Sousa Rodrigues, D. Sc. – Procert Engenharia

Dedicado aos meus amados pais.

AGRADECIMENTOS

Dedico meus sinceros e profundos agradecimentos ao professor Alexandre Cury, exemplo de bom mestre e amigo.

Também agradeço ao professor Flávio Barbosa pela parceria e pela abertura de oportunidades tão valiosas no âmbito acadêmico.

Com inteligência, empatia e cuidado, ambos contribuíram para tornar minha caminhada mais leve e virtuosa.

Por ter chegado até aqui, dou muitas graças por ter tido a oportunidade de passar por instituições de ensino públicas federais com tanta qualidade. De fato, o Brasil me proporcionou educação de alta qualidade e gratuita. Na minha trajetória acadêmica passei sete anos no Colégio Militar de Juiz de Fora, graduei-me em cinco anos na Universidade Federal de Juiz de Fora e realizei a pós-graduação por quase seis anos na Universidade Federal de Ouro Preto. No total, desfrutei de dezoito anos de trabalho nobre de homens e mulheres, funcionários públicos, que dão vida a estas belas instituições. Dou muitas graças a todos!

Dou graças por ter a saudável companhia dos meus colegas de trabalho da PROINFRA da UFJF. Especialmente, agradeço aos pró-reitores: engenheira Janezete Marques e professor Marcos Tanure. Eles possibilitaram, de forma harmônica, um ambiente laboral no qual eu pude desenvolver minha tese em paralelo com as demais atividades de engenharia. Agradeço aos dois e também às chefias imediatas.

Agradeço imensamente aos meus amigos da UFOP, especialmente Rafael Zaltron, Marko Rupert, Everton, Marcela e Tatiane. Agradeço também especialmente aos professores Marcílio, Célio, Arlene, Ricardo Azoubel e Assis. Todos contribuíram de sua própria forma.

Pelo fundamental suporte financeiro dou graças a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais).

Agradeço aos meus irmãos Davi e Daniel por terem dado suporte ao *projeto de vida* da família enquanto estive muito ocupado com a pesquisa. Ambos contribuíram muito para a realização desta tese, mesmo que indiretamente.

Finalmente, agradeço à minha amada esposa Iara. Sua brilhante e constante companhia me abastece com Propósito e me estimula para a Atividade. Esta tese é grandemente fruto da sua Inteligência e de seu Amor que são expressos em cada detalhe do dia-a-dia. Nossa convivência torna evidente o Pai-Mãe-Vida tão maravilhosamente se manifestando através de tudo e de todos.

*If you want to find the secrets of the Universe,
think in terms of energy, frequency and
vibration.*

Nikola Tesla

Abstract of Thesis presented to PROPEC/UFOP as partial fulfillment of the requirements for the degree of Doctor in Civil Engineering

APPROACHES FOR AUTOMATED DAMAGE DETECTION IN STRUCTURAL HEALTH MONITORING

Rharã de Almeida Cardoso

October/2018

Advisor: Alexandre Abrahão Cury

Co-advisor: Flávio de Souza Barbosa

Knowing the integrity of major structural systems in real-time and continuously during their operational service is a crucial requirement for manufacturers, owners, users and maintenance teams. This knowledge affords the stakeholders relevant structural performance information to drive design and production improvements, minimize maintenance costs, and increase the operational structure's safety for the users. Among the several activities that comprise Structural Health Monitoring (SHM), damage detection is the core task to satisfy maintenance and safety aspects. Thus, an SHM program must be aided by computational tools capable of analyzing the acquired sensorial information continuously. Then, it must promptly yield one or more indicators of damage (or novelty) occurrence in the structure. Therefore, for a damage detection technique to be compatible with the scope of SHM, it should ideally respond automatically, in an unsupervised and continuous way, based solely on ambient vibration tests when the structure is under normal operation. Hence, this thesis presents two approaches: one based on the tracking of structural modal parameters, i.e., natural frequencies, damping ratios, and mode shapes; the other based on the analysis of raw acceleration measurements. The proposed methodologies were applied to real-case structures and showed promising performances when it comes to long-term, continuous and real-time monitoring.

Keywords: Structural Health Monitoring, Damage Detection, Modal Identification.

Resumo da Tese apresentada ao PROPEC/UFOP como parte dos requisitos necessários para obtenção do grau de Doutor em Engenharia Civil

TÉCNICAS PARA DETECÇÃO AUTOMÁTICA DE DANOS NO MONITORAMENTO DA INTEGRIDADE ESTRUTURAL

Rharã de Almeida Cardoso

Outubro/2018

Orientador: Alexandre Abrahão Cury

Coorientador: Flávio de Souza Barbosa

Conhecer a integridade de sistemas estruturais de grande vulto durante o serviço, em tempo real e continuamente, é uma grande necessidade dos fabricantes, proprietários, concessionários, usuários finais e equipes de manutenção destes sistemas. Tal conhecimento provê aos gestores informações relevantes sobre o desempenho estrutural para direcionar melhorias no projeto e produção, além de minimizar custos de manutenção para o proprietário/concessionário e de aumentar a segurança de operação da estrutura para os usuários. Dentre as diversas atividades que compreendem o Monitoramento da Integridade Estrutural (MIE), a detecção de danos constitui o núcleo básico para atender aos aspectos de manutenção e segurança. Para tanto, o programa de MIE deve dispor de ferramentas computacionais capazes de analisar as informações adquiridas continuamente e em tempo real, fornecendo a cada momento um ou mais indicadores da ocorrência de dano (ou alteração) na estrutura. Portanto, para que uma técnica de detecção de danos seja compatível com o escopo do MIE ela deve, idealmente, prover respostas de forma automática, não supervisionada e contínua, baseando-se unicamente em testes de vibração ambiente com a estrutura em operação. Visando atingir estes objetivos, esta tese apresenta duas abordagens: uma baseada na evolução dos parâmetros modais, isto é, frequências naturais, taxas de amortecimento e modos de vibração; outra baseada na análise direta de medições de aceleração. As metodologias propostas foram avaliadas em estruturas reais e demonstraram desempenhos promissores quando aplicadas em monitoramentos de estruturas em longo prazo, contínuos e em tempo real.

Palavras-chave: Monitoramento da Integridade Estrutural, Detecção de Danos, Identificação Modal.

Foreword

This thesis is written as an effort to gather the last four years of research in a compact, yet comprehensible way: each chapter corresponds to a published/accepted or submitted paper. The methodologies proposed herein were assessed through numerical and experimental (laboratory and *in situ*) applications, as the techniques had their complexity increased and their accuracy improved.

The author would like to emphasize, however, that the present work is not only devoted to being a simple collection of papers. Instead, it was thought to create a smooth connection among the publications and to serve as a valuable manuscript on the subject for future reference.

Therefore, the text is structured in four parts as Fig. I shows.

Part I consists of a brief introduction, pointing out the relevance, motivation, and objectives of the developed research.

Part II and Part III constitute the core of this research. Within these parts, the reader will find one chapter for each one of the four written papers. The achieved advances and the connecting lines are reported through brief texts between the papers, at the beginning of each chapter.

Part IV concludes this thesis by highlighting the achieved milestones.

Finally, the reader will find the bibliographic references placed at the end of each chapter.

PART I INTRODUCTION	CHAPTER 1 Introduction
PART II MODAL IDENTIFICATION FOR DAMAGE DETECTION	CHAPTER 2 Paper #1: A clustering-based strategy for automated structural modal identification
	CHAPTER 3 Paper #2: A robust methodology for modal parameters estimation applied to SHM
PART III RAW DATA PROCESSING FOR DAMAGE DETECTION	CHAPTER 4 Paper #3: Unsupervised real-time SHM technique based on novelty indexes
	CHAPTER 5 Paper #4: Automated real-time damage detection strategy using raw dynamic measurements
PART IV CONCLUSION	CHAPTER 6 Achieved Milestones

Fig. I Thesis's outline.

Summary

PART I – INTRODUCTION	1
1 Introduction.....	2
1.1 Overview	2
1.2 Context.....	3
1.3 Damage and Novelty in SHM	4
1.4 Objectives.....	5
References	6
PART II – MODAL IDENTIFICATION FOR DAMAGE DETECTION.....	7
2 Paper #1 – A clustering-based strategy for automated structural modal identification ...	8
Abstract.....	9
2.1 Introduction and state of the art.....	9
2.2 SSI-DATA.....	12
2.3 Automation methods for stabilization diagram interpretation	14
2.4 Results	19
2.5 Results discussion	32
2.6 Practical relevance and potential applications	34
Funding.....	34
References	35
3 Paper #2 – A robust methodology for modal parameters estimation applied to SHM ...	36
Abstract.....	37
3.1 Introduction	37
3.2 Data-driven Stochastic Subspace Identification.....	40
3.3 Automation methods for stabilization diagram interpretation	41
3.4 Results	49
3.5 Conclusions	61

Acknowledgements	62
References	62
PART III – RAW DATA PROCESSING FOR DAMAGE DETECTION	64
4 Paper #3 – Unsupervised real-time SHM technique based on novelty indexes.....	65
Abstract.....	66
4.1 Introduction	66
4.2 Feature Extraction and Data Fusion	69
4.3 Feature Classification.....	77
4.4 Experimental applications	79
4.5 Conclusion.....	86
Acknowledgements	86
References	87
5 Paper #4 – Automated real-time damage detection strategy using raw dynamic measurements	92
Abstract.....	93
5.1 Introduction	93
5.2 Feature Extraction and Data Fusion	96
5.3 Feature Classification.....	103
5.4 Application 1: Yellow Frame.....	107
5.5 Application 2: Concrete Bridge.....	114
5.6 Conclusion.....	117
Acknowledgements	118
References	118
PART IV – CONCLUSION	122
6 Achieved Milestones.....	123

PART I

INTRODUCTION

Chapter 1

Introduction

1.1 Overview

Any mechanical structure is, ultimately, a dynamic system. Its vibrational behavior is continually *informing* about its state of integrity, or "health". This *information* can be acquired by various sensors measuring temperature, displacement, strain, acceleration, velocity, angular tilt, and so forth. Then, an appropriate software must process the acquired signals to allow the *information* to be accessible for human judgment. This is the core idea behind Structural Health Monitoring (SHM), as Fig. 1.1 summarizes.

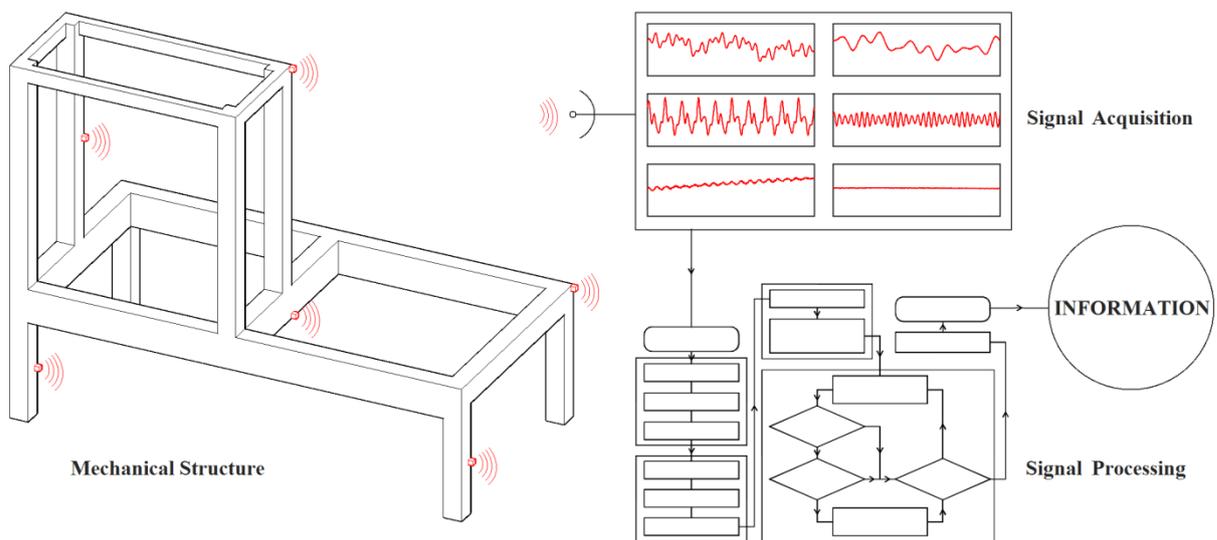


Fig. 1.1 Extraction of intelligible information for structural health monitoring.

It is worth mentioning that, even though this thesis focuses on SHM of Civil Engineering structures, the techniques herein developed can also be applied to any other dynamic system, such as electric networks, machines, aerospace structures, the human body, hydrologic basins, economic models, production lines, and so on. The proposed algorithms

are supposed to perform the very task of detecting abnormalities in the system's behavior utilizing *signal processing*. Thus, it does not matter from which system this signal (response) is collected. Naturally, any application out of the civil engineering scope would possibly require some additional calibration.

1.2 Context

There has been an increasing awareness of the importance of damage prognosis programs in civil, mechanical and aerospace structures. A robust monitoring system could advise about the structure's health state, informing about any incipient damage in real time and providing an estimate of the structure's remaining useful life. The potential benefits that would result from such technology are enormous. The maintenance procedures for structures with such prognosis systems could change from being schedule-driven to condition-based. Hence, time, cost and labor requirements would be drastically saved, thus avoiding the interdiction of such structures. This would especially benefit aging structures, which are currently represented by the vast majority of viaducts, bridges, buildings, etc., in the world. Last, but not least, such a monitoring system would increase the structure's safety for the users.

Thus, the acquisition of vibration tests and their subsequent analysis have gained great practical importance in the field of civil engineering. Studies have been published on the application of several types of instrumentation and data acquisition systems commonly used in the dynamic monitoring of large structures, such as the Rio-Niterói Bridge ^[1] in Brazil, the Z24 bridge ^[2] in Switzerland, and the Millau viaduct ^[3] in France.

Indeed, some strategic structures are monitored 24 hours a day, seven days a week – continuously or during time intervals – to provide data that can allow detecting structural problems in their initial stages, like cracks, excessive vibration, abnormal behaviors, etc. Additionally, such monitoring aims at providing a more in-depth assessment of structural reliability and vulnerability.

In this type of structures, the presence of an SHM program can be crucial for preventing disasters. Health-related information extracted from structural behaviors can save lives and avoid economic losses. Sadly, a recent example of a great catastrophe occurred in Genoa, Italy (August 14, 2018), where a stayed bridge collapsed leaving 43 people dead, as dozens of cars fell onto the riverbed, railroad tracks and streets below (Fig. 1.2). The southern

stays (from one tower) that initially went slack are the same in which a renowned structural engineer found possible damage during dynamic tests performed almost one year earlier. He warned the company that managed the bridge but said that they never followed up on his recommendation to conduct a thorough study and to instrument the bridge with permanent sensors. “Probably they underestimated the importance of the information,” said the engineer in an interview ^[4].



Fig. 1.2 Bridge collapses in Italy.

1.3 Damage and Novelty in SHM

To better understand the object of study of this work, it is imperative to distinguish the meaning of the terms structural *damage* and structural *novelty*.

The term *damage* can be defined as a *novelty* (or change) in a structural system that affects its performance, present or future, both regarding its safety and service. Such *novelties* may be caused by variations in the material properties and/or in the geometric properties of the structure, including boundary conditions. While the occurrence of *damage* implies that there has been a structural *novelty*, the opposite is not true. Identifying a *novelty* in structural behavior does not necessarily mean that performance, present or future, is compromised and,

therefore, does not imply that the structural system is damaged. For instance, a *novelty* can be caused by a structural reinforcement imposed by human intervention.

It is important to emphasize that the present work is devoted to providing tools for detecting structural *novelty*, which may or may not be characterized as *damage*. In practice, except for cases where there were structural reinforcements or intentional changes by human actions, almost every *novelty* is indeed *damage*. Therefore, the set of techniques developed within this thesis can provide useful tools to be placed into Level 1 of SHM Rytter's classification:

- Level 1: Is there any damage present in the structural system (detection)?
- Level 2: Where is the damage located (location)?
- Level 3: What is the damage magnitude (quantification)?
- Level 4: What is the residual life of the structure (prognosis)?

1.4 Objectives

At first sight, an algorithm capable of detecting damage within an SHM context might be seen as a product of a simple strategy. However, the extensive literature review presented throughout the four papers showed that there are some challenges not yet transposed within this subject. In general, three obstacles pose an unpractical aspect to the existing detection algorithms (Level 1) for civil structures:

1. The lack of *unsupervised* techniques capable of detecting structural damage without relying on any prior knowledge about the structure's condition;
2. The lack of strategies capable of robustly detecting structural damage *automatically*;
3. The lack of algorithms suitable for performing *real-time* detection.

Thus, the present work aims at presenting damage detection strategies that are *unsupervised*, *automated* and suitable for *real-time* SHM. **These strategies are separated into two approaches: one based on the automatic identification of structural modal parameters, i.e., natural frequencies, damping ratios and mode shapes (Part II); the other based on the analysis of raw acceleration measurements (Part III).**

References

- [1] R.C. Batista and M.S. Pfeil. Reduction of vortex-induced oscillations of Rio-Niterói Bridge by dynamic control devices, *Wind Engineering and Industrial Aerodynamics*, **2000**, 84(3), pp. 273-288.
- [2] J. Maeck and G. De Roeck. Damage assessment using vibration analysis on the Z24-Bridge, *Mechanical Systems and Signal Processing*, **2003**, 17(1), pp. 133-142.
- [3] Y. Gautier, C. Cremona, O. Moretti. Experimental modal analysis of the Millau Bridge, Proceedings of the EVACES (Experimental Vibration Analysis for Civil Engineering Structures), **2005**.
- [4] J. Glanz *et al.* Genoa bridge collapse: the road to tragedy. *New York Times*, September 6, **2018**.
<https://www.nytimes.com/interactive/2018/09/06/world/europe/genoa-italy-bridge.html>

PART II

MODAL IDENTIFICATION FOR DAMAGE DETECTION

Chapter 2

Paper #1 – A clustering-based strategy for automated structural modal identification

One of the most common strategies to detect damage is based on the *identification* and *tracking* of structural modal parameters. Since the latter cannot exist without the former, the development and validation of tools for the automatic modal identification of structures under normal operation are the first main objective of this research. Therefore, this first research proposes a methodology for modal identification automatically, making it suitable for unsupervised real-time modal parameter based strategies for damage detection in SHM. In 2017, a paper was published in the journal “*Structural Health Monitoring*”.

Paper Information

Status Published in 2017

DOI <https://doi.org/10.1177/1475921716689239>

Publisher Structural Health Monitoring
ISSN 1475-9217; Qualis N/A; Impact Factor (2018): 3.798

Citation Cardoso RA, Cury A, Barbosa F (2017). A clustering-based strategy for automated structural modal identification. *Structural Health Monitoring*, 17(2): 201-217.

Cardoso RA, Cury A, Barbosa F (2017). A clustering-based strategy for automated structural modal identification. *Structural Health Monitoring*, 17(2): 201-217.

Abstract

Structural health monitoring of civil infrastructures has great practical importance for engineers, owners and stakeholders. Numerous researches have been carried out using long-term monitoring, such as the Rio–Niterói Bridge in Brazil, the former Z24 Bridge in Switzerland and the Millau Bridge in France. In fact, some structures are continuously monitored to supply dynamic measurements that can be used for the identification of structural problems such as the presence of cracks, excessive vibration or even to perform a quite extensive structural evaluation concerning its reliability and life cycle. The outputs of such an analysis, commonly entitled modal identification, are the so-called modal parameters, that is, natural frequencies, damping ratios and mode shapes. Therefore, the development and validation of tools for the automatic modal identification during normal operation is fundamental, as the success of subsequent damage detection algorithms depends on the accuracy of the modal parameters' estimates. This work proposes a novel methodology to perform, automatically, the modal identification based on the modes' estimates data generated by any parametric system identification method. To assess the proposed methodology, several tests are conducted using numerically generated signals, as well as experimental data obtained from a simply supported beam and from a motorway bridge.

2.1 Introduction and state of the art

The dynamic parameters of a structure are deeply related to its mechanical properties. In fact, the deviation of natural frequencies and mode shapes often indicates structural changes. Hence, if these deviations are properly assessed, it is possible to have an idea about the structure's condition, that is, to infer about the presence of damage or of any other drastic change that could set up a danger condition. Hence, the identification of natural frequencies, damping ratios and modal shapes is of utmost importance to the structural health monitoring of large civil structures such as tall buildings and long-span bridges [1]. The permanent online monitoring of such parameters increases the chance of detecting unusual behaviour of the structure in a preventive way.

First, for the sake of clarity, one should clearly distinguish *modal parameter estimation* (MPE) from *modal tracking*. The former consists of the estimation of modal parameters from a single record of measured data and the latter corresponds to tracking the evolution of the modal parameters of a structure through repeated MPE. This article focuses on the MPE process, which means that structural responses are evaluated independently. Therefore, what would naturally be the second step, related to modal tracking, will not be considered here.

Thus, concerning the parametric system identification techniques, one can enumerate the following two steps for the MPE:

1. Computation of the modes' estimates for several orders of the parametric model, by fitting them to the response series;
2. Interpretation (in case of manual interpretation, the stabilization diagram would be one of the useful graphical tools used) of the modes' estimates data by detecting spurious modes and selecting the physical ones.

In fact, without the automation of the MPE process, the user would be required to set several input parameters, making the modal identification procedure impractical to long-term applications such as the online health monitoring of structures. Therefore, especially over the last decade, several methods aiming at the reduction in the number of user-defined parameters are being developed [2–5], some of which reached the complete elimination of all manually specified parameters and, hence, are usually labelled as fully automated.

For instance, Reynders et al. [3] proposed a fully automated approach for the interpretation of a stabilization diagram. It was cleverly based on clustering techniques through three stages: a diagram pre-cleaning by means of a classification of all modes into two categories (possible physical or certainly spurious), a hierarchical clustering of the possible physical ones to group them together (automatic detection of vertical lines on diagram) and a final classification of the formed clusters into a physical or spurious condition. No user-defined parameter is needed whatsoever in the entire process.

On the other hand, the structural dynamic monitoring practice has been showing that even the most well-crafted automated process of MPE has to be, at a first use, manually checked or tuned to fit a possible specific behaviour of the analysed structure to obtain satisfactory results. As a consequence, the development of new methods does not need to

focus on the elimination of all user-defined parameters (fully automation), but on the construction of a robust automated approach with some few easy configurable user settings, which are defined once, before starting the repetitive automated process itself. Thereby, this article presents an automated method for interpreting the stabilization diagram with just two easily user-defined parameters. Nowadays, the available time domain methods for operational modal analysis (OMA), explored in the work of Peeters and De Roeck [6], are essentially based on two types of models: discrete-time stochastic states-space models and auto-regressive moving average (ARMA) or just auto-regressive (AR) models.

The formulations that use state-space models, designated stochastic subspace identification (SSI) methods, constitute the parametric approach that is more commonly adopted for civil engineering applications. The model can be identified either from correlations (or covariances) of the outputs: covariance-driven stochastic subspace identification (SSI-COV), or directly from time series collected at the tested structure by the use of projections [7]: data-driven stochastic subspace identification (SSI-DATA). As reported in the work of Peeters and De Roeck [6], these two methods are very closely related. Still, the SSI-COV has the advantage of being faster and based on simpler principles. The SSI-DATA, however, yields further outputs such as the decomposition of measured responses into modal contributions.

Considering the existence of a large number of alternative algorithms for OMA developed during the last decades, which sometimes are based on similar principles, this work focuses on the description of the SSI-DATA method. Since this method yields stabilization diagrams as outputs, the authors understand that these diagrams are a resourceful data for reliable structural modal identification. Moreover, when associated with clustering techniques, they are capable of providing very robust results. Thus, section 2.2 briefly discusses this method.

Section 2.3 presents a reference methodology and the novel approach developed in this article. Besides some peculiarities, the proposed method has an innovative way of measuring distances between modes' estimates during the clustering process. Section 2.4 shows the results for numerical and practical applications. Considerations and discussions over the results are drawn in section 2.5. Finally, in section 2.6, a short paragraph points out the practical relevance of the work developed.

2.2 SSI-DATA

Since the work published in the 1990s [7], the application of SSI-DATA algorithms to determine the modal parameters of structures has been quite spread over the ambient vibration tests. State-space models are computed from output data (structure response) and can have their problem stated as following.

Given s measurements of the output $\mathbf{y}_k \in \mathbb{R}^l$ generated by the unknown stochastic system of order n

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{v}_k\end{aligned}\quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector at a discrete-time instant k , \mathbf{w}_k and $\mathbf{v}_k \in \mathbb{R}^n$ are zero mean, white vector sequences with covariance matrix (at instants p and q with δ meaning the Kronecker delta)

$$\mathbb{E}\left[\begin{pmatrix} \mathbf{w}_p \\ \mathbf{v}_p \end{pmatrix} \begin{pmatrix} \mathbf{w}_q^T & \mathbf{v}_q^T \end{pmatrix}\right] = \begin{pmatrix} \mathbf{Q} & \mathbf{S}_v \\ \mathbf{S}_v^T & \mathbf{R} \end{pmatrix} \delta_{pq} \geq 0, \quad (2)$$

determine the below:

- The order n of the unknown system;
- The system matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{C} \in \mathbb{R}^{l \times n}$.

For this purpose, the SSI-DATA algorithm must consist of the following eight steps:

1. Organize the response series \mathbf{y} (the only measured quantity) in a Hankel matrix \mathbf{Y} which can be partitioned into a past measurement's matrix \mathbf{Y}_p and a future measurement's matrix \mathbf{Y}_f .
2. Calculate the orthogonal projection \mathbf{P} of the row space of \mathbf{Y}_f on the row space of \mathbf{Y}_p

$$\mathbf{P} = \mathbf{Y}_f \mathbf{Y}_p^T \left(\mathbf{Y}_p \mathbf{Y}_p^T \right)^\dagger \mathbf{Y}_p \quad (3)$$

where \bullet^\dagger denotes the Moore-Penrose pseudo-inverse of the matrix \bullet .

3. Compute the singular value decomposition (SVD) of the projection weighted by matrices \mathbf{W}_1 and \mathbf{W}_2

$$\mathbf{W}_1 \mathbf{P} \mathbf{W}_2 = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (4)$$

4. Determine the system order n by inspecting the rank of \mathbf{S} (the number of non-zero singular values) and partition the SVD accordingly to obtain \mathbf{U}_1 and \mathbf{S}_1

$$\mathbf{U} \mathbf{S} \mathbf{V}^T = (\mathbf{U}_1 \quad \mathbf{U}_2) \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{pmatrix} \quad (5)$$

5. Compute the extended observability matrix $\mathbf{\Gamma}$

$$\mathbf{\Gamma} = \mathbf{W}_1^{-1} \mathbf{U}_1 \mathbf{S}_1^{1/2} \quad (6)$$

6. Determine, in a least square sense, the system (dynamic) matrix \mathbf{A} from $\mathbf{\Gamma}$ as

$$\mathbf{A} = \underline{\mathbf{\Gamma}}^\dagger \bar{\mathbf{\Gamma}} \quad (7)$$

where $\underline{\mathbf{\Gamma}}$ denotes $\mathbf{\Gamma}$ without the last l (number of outputs) rows and $\bar{\mathbf{\Gamma}}$ denotes $\mathbf{\Gamma}$ without the first l rows.

7. Determine \mathbf{C} as the first l rows of $\mathbf{\Gamma}$.
8. Finally, knowing \mathbf{A} and \mathbf{C} , one can calculate the modal frequencies ω_i , the modal damping ratios ξ_i and the corresponding modal shapes ϕ_i . This calculation is made through the eigenvalue decomposition of matrix \mathbf{A} as the following statements show

$$\lambda_i = \frac{\ln(\mu_i)}{\Delta t} \rightarrow \begin{cases} \omega_i = |\lambda_i| \text{ (rad / s)} \rightarrow f_i = \frac{|\lambda_i|}{2\pi} \text{ (Hz)} \\ \xi_i = \frac{-\text{Re}(\lambda_i)}{|\lambda_i|} \times 100 \text{ (\%)} \\ \phi_i = \mathbf{C} \psi_i \end{cases} \quad (8)$$

where μ_i is the i^{th} discrete-time eigenvalue of \mathbf{A} and ψ_i is its corresponding eigenvector. The eigenvalues of \mathbf{A} are also called *poles*. They are complex numbers and appear mostly in conjugated pairs. Only those poles which have a conjugated pair and have positive imaginary component are taken into account for the equation (8). Therefore, up to $n/2$ modes are expected to be estimated. The $\text{Re}(\bullet)$ is the symbol for the real part of \bullet and λ_i denotes the continuous-time eigenvalue of \mathbf{A} .

These eight steps summarize the SSI-DATA routine which is, essentially, a fit of a state-space model to the temporal output data by means of the geometric projection of the row space of the future measurements onto the past measurements.

There are several valid sets of weighting matrices \mathbf{W}_1 and \mathbf{W}_2 to be used in equations (4) and (6). They determine the state-space basis in which the model will be identified. Special choices of these matrices correspond to algorithms described in the literature: principal component (PC), unweighted principal component (UPC) and canonical variate analysis (CVA). The PC algorithm is the one used in the applications shown in section ‘Results’.

It is necessary to point out that when one works with real structure response data, it is very hard to determine the system’s order through inspection of singular values like stated in the fourth step. This happens because the matrix \mathbf{S} is actually full rank. Thus, a threshold would be necessary to consider a singular value as being null. Therefore, instead of creating ways for defining this threshold, the system order n is readily given as an input to the algorithm, which solves the problem, however, it creates the need for a further check whether this arbitrary order is able or not to accurately represent the actual data. One of the tools for this check is exactly the well-known stabilization diagram, where the modes’ estimates are plotted for various different system orders.

2.3 Automation methods for stabilization diagram interpretation

As previously mentioned, the mode estimates (usually presented in a stabilization diagram) are supposed to be analysed to get physical modes apart from the spurious (numerical) ones. Spurious modes inevitably exist among the data, since they are a natural consequence of the parametric model’s attempt to better fit the response data.

Naturally, due to the large amount of collected data from online structural monitoring, a manual interpretation of modes estimates data is unfeasible. Thus, the obvious solution is the automation of such a process. For this purpose, in this article, one proposes a novel approach, which was mainly inspired by the automation methodology proposed in a PhD thesis [2] as well as by some elements found in a recently published fully automated method [3]. The main new features found in the proposed method rely on the innovative way of

measuring distances between modes' estimates during the clustering process. Also, some additional modifications regarding filtering of spurious modes are included. Besides, a novel cluster regrouping technique is proposed.

For comparison purposes, section 2.3.1 presents a reference methodology, which corresponds to the basis of the algorithms proposed by Magalhães [2]. In section 2.3.2, the novel approach proposed by the authors is explained and all its contributions are highlighted. It is relevant to point out that both methods are suitable for the analysis of the outputs generated by any parametric identification technique that produces modes' estimates (frequencies, damping ratios and modal shapes) for several model orders.

2.3.1 Reference methodology

In this approach, no pre-filter is applied to the modes' estimates data, which means that the stabilization diagram would be still full of certainly spurious modes, that is, like those with extremely high or negative damping ratios. This method is based on a hierarchical clustering algorithm.

First, the dissimilarity (distance) measure between all pairs of estimated modes is calculated. The adopted 'metric' for calculation of such a distance is

$$d_{i,j} = \left| \frac{f_i - f_j}{f_j} \right| + (1 - \text{MAC}_{i,j}) \quad (9)$$

where f_i and f_j are the natural frequencies of the modes' estimates i and j , respectively. The mode shape similarity between these two modes' estimates is evaluated by the well-known modal assurance criterion (MAC). Moreover, one can note that for this metric, it is possible that $d_{i,j} \neq d_{j,i}$, which leads to an oddly asymmetric dissimilarity matrix.

Hierarchical clustering algorithms differ in the way they compute the distance between two already formed clusters. In this methodology, the single-linkage criterion is used, which means that the distance between two clusters will be considered as being equal to the smallest distance, also computed with equation (9), between objects inside these two clusters. This information allows constructing the hierarchical tree.

In the following step, one must define the tree's cut-level. This is usually dependent on the expected number of clusters. However, this number is not trivial to predict due to the

unknown quantity of clusters grouping spurious modes. Thus, the alternative strategy is the use of a distance limit (d_{lim}). Such a distance is the criterion to prune the branches of the hierarchical tree, generating clusters that are sufficiently distant from each other. Hence, the lower the limit, the higher the number of resulting clusters.

At this point, one has each cluster representing a single mode. To decide whether the mode is spurious or not, the clusters are ranked according to their number of elements. Thus, the top n_m clusters with more elements are selected. Once physical modes are very consistent for models with different orders, their clusters present a much higher number of elements than the clusters that contain numerical estimates (which present a higher scatter between models of different orders). Therefore, it is expected to find that the n_m selected modes are indeed physical. In practice, it is common to realize that the number of physical modes n_m can be guessed, for instance, by a simple preliminary frequency domain analysis.

Then, the damping ratios are taken into account by means of an outlier analysis within each selected cluster. This internal filtering eliminates the modes' estimates with extreme damping ratio values, which are out of the range defined by $\pm 2.698\sigma$ (standard deviation) that embraces 99.3% of the samples in a Gaussian distribution. Finally, each selected cluster produces average values of the modal parameters (natural frequency, modal damping ratio and mode shape), which are the outputs of the methodology.

2.3.2 Proposed methodology

This method is also based on a hierarchical clustering algorithm. However, differently from the previous approach, one intends avoiding unnecessary computational efforts and aims to obtain better results from the clustering process. Thus, a pre-filter is applied on the modes' estimates data, which removes all modes whose damping ratios are not between 0% and 15% (recommended for civil engineering structures). In addition, the stabilization diagram will have eliminated all modes with modal phase collinearity (MPC) indicator lower than a limit m_{lim} in the $[0, 1]$ interval of this criterion. The MPC index of a vector ϕ was stated by Juang and Pappa [8], revised by Pappa *et al.* [9] and specifically studied, together with MAC, by Vacher *et al.* [10]. It was used by a number of authors, such as Reynders *et al.* [3] and Yun *et al.*, [11] to help distinguishing physical modes from spurious ones. MPC values closer to 1 indicates that the mode shape vector is 'monophase' and is more likely to be physical, while

values closer to 0 show that the mode is either a spurious mode or the mode is significantly complex. Such an indicator is computed as

$$\text{MPC}(\phi) = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 \quad (10)$$

where λ_1 and λ_2 are the eigenvalues of the matrix

$$\begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \quad (11)$$

composed, as shown in equations (12), by the variances and covariance of the vectors ϕ_r and ϕ_i which are the real and imaginary parts of ϕ , respectively

$$\begin{aligned} S_{xx} &= \phi_r^T \phi_r \\ S_{yy} &= \phi_i^T \phi_i \\ S_{xy} &= \phi_r^T \phi_i \end{aligned} \quad (12)$$

Finally, the user may also have removed all modes' estimates out of a range of frequencies specified (much like a 'band-pass' filter on the modes' estimates data).

After having the modes' estimates data processed by the pre-filter, a plot of the stabilization diagram would reveal that a significant amount of modes was removed, yielding a clearer aspect. Depending on the situation, and this is not rare, the removed modes may even represent more than 50% of the initial data. These removed modes' estimates are considered as certainly spurious (or undesired) and will not be processed by the subsequent clustering algorithm.

To calculate the distance between the modes' estimates i and j , a novel metric is used

$$d_{i,j} = |f_i - f_j| + (1 - \text{MAC}_{i,j})c \quad (13)$$

where c is a constant arbitrarily set to 5 Hz (during initial studies, the authors varied this constant from 1 to 10 Hz without obtaining relevant differences in the outputs, which points out the low sensitiveness of the proposed methodology to such a parameter. As the authors proved in this article, the weighting value set to 5 Hz was appropriate to all presented applications. So far, there was no need to set this constant to another value to ensure the

success of the method and, therefore, no deeper study about this topic was performed), which after some tests was established by the authors for being a satisfactory weight value. One should note that differently from equation (9), this metric leads to a perfectly symmetric dissimilarity matrix, that is, $d_{i,j}=d_{j,i}$.

Then, the hierarchical clustering is applied. However, differently from the reference methodology, the distance between two already formed clusters is measured according to the complete linkage criterion, which means that such a distance is equivalent to the largest possible distance between two of their elements. Then, the hierarchical tree will be trimmed exactly like in the reference methodology by using a distance threshold d_{lim} .

Differently from what the previous methodology does, the internal filter is here applied within the clusters before they are ranked according to their number of elements. This is done to penalize the clusters with high scatter (likely to be representing a numerical mode) before they could be ranked among the top n_m modes that will be selected as physical. Thereby, the internal filter removes outlier modes within the cluster considering not only damping ratios but also mode shapes. Mode estimates that have damping ratios out of the $\pm 2.698\sigma$ range and/or have a lower than 0.8 MAC related to the cluster average mode shape are eliminated.

After this grouping process, depending on the d_{lim} value, it is possible to find that different clusters represent the very same physical mode. In other words, a single physical mode could have been split into two or more different clusters. Evidently, this phenomenon is to be avoided. Therefore, the authors propose a regrouping procedure to minimize this problem.

The regrouping routine is based on the confidence interval of the frequencies' estimates. Each single cluster has its frequency mean value and its lower and upper limits delimiting a confidence interval. All clusters are compared two by two. If the first cluster has its average frequency estimate between the confidence limits of the second cluster, they are possibly representing the same physical mode, and therefore, they are selected to go through a further analysis: the evaluation of the MAC coefficient among their mode estimates. If the MAC is higher than 0.9, the clusters are grouped into one single cluster.

Finally, the top n_m clusters (modes) with more elements are selected as being physical. Then, each selected cluster has its average modal parameters (natural frequency, damping ratio and mode shape) calculated.

A flowchart of the developed methodology can be found in Fig. 2.1. The differences to the reference method, that is, the contribution of this work, is highlighted in red over the diagram shown in this figure.

The complete description and study of the proposed methodology can be found, in detail, in the work of Cardoso [12].

2.4 Results

2.4.1 Numerical experiment

Introduction

To initially evaluate the efficiency of the proposed method a numerical experiment was performed. The signals correspond to a hypothetical structure. Five degrees of freedom (DOFs) were considered.

Equation (14) describes a free damped vibration of a single-DOF dynamic system. In this equation, ω_D represents the damped natural frequency that is approximately equal to, for small damping ratios ζ , the undamped natural frequency $\omega = 2\pi f$. Besides, $y(t)$ is the signal amplitude at a time instant t , ρ is an amplitude factor and θ is a phase constant.

$$y(t) = \sum_{i=1}^2 \rho_i \cos(\omega_{D_i} t + \theta_i) e^{-\zeta_i \omega_i t} \quad (14)$$

The variables in the equation above were chosen in a manner that two modes with close natural frequencies were obtained. Table 2.1 shows the values of the dynamic parameters from this hypothetical system.

Table 2.1: Dynamic parameters of the hypothetical system.

	f (Hz)	ζ (%)
Mode 1	20.80	1.00
Mode 2	20.90	2.00

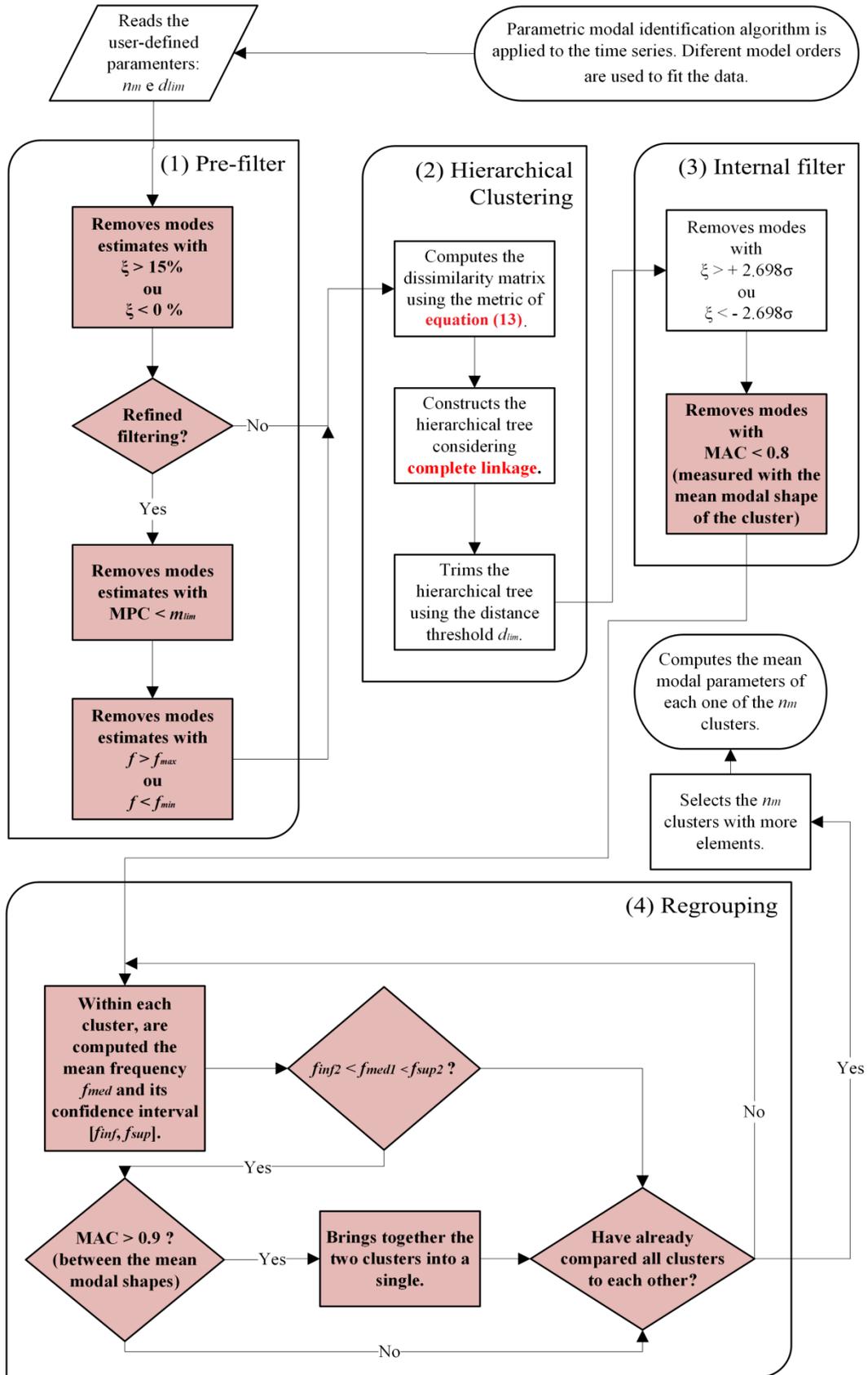


Fig. 2.1 Flowchart of the proposed methodology.

The modal shapes will not be analysed in this application. The objective here is to verify the capability of the proposed method for identifying correctly these closely spaced modes. First, a signal without noise will be processed. Then, 20% of white noise is artificially added. Such a noise was added to the signal according to equation (15). In this expression, y_{noise} is the signal with noise, y_{max} is the maximum absolute value of y , β is a factor containing the noise level (for instance, $0.2 = 20\%$) and χ is a random signal (white noise) generated by the computer, with amplitude varying from -1 to $+1$.

$$y_{noise}(t) = y(t) + y_{max} \cdot \beta \cdot \chi(t) \quad (15)$$

A 1000 Hz sampling rate was adopted, corresponding to 10 seconds of signal's length. In this numerical experiment, the proposed methodology parameters were set equal to $d_{lim} = 1$ Hz e $n_m = 2$.

Signal without noise

The pure signal (without noise), is shown in Fig. 2.2.

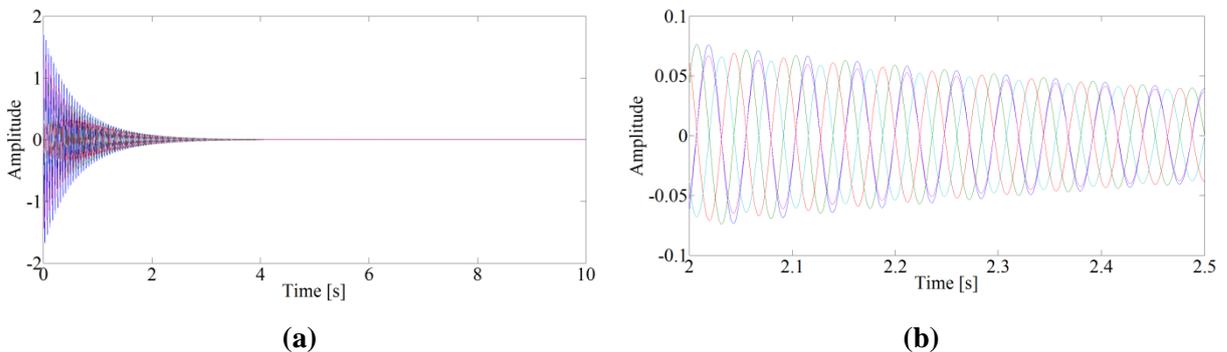


Fig. 2.2 Time history of the generated signal (five channels) without noise: (a) complete signal and (b) zoom.

The SSI-DATA (PC) method was used to identify models with orders varying from 10 to 80. Once the dynamic system only has two physical modes, a large number of spurious modes appeared. This high-order modelling was intentional and aimed to turn the stabilization diagram more ‘polluted’, to test the capability of the developed method in distinguishing spurious modes from physical ones. In fact, in this study a model with order 4 would be enough to represent accurately the structural system.

As noted in Fig. 2.3, the proposed filters cleaned up dramatically the stabilization diagram: 1004 of 1325 (75.8%) modes’ estimates were considered as certainly spurious and, therefore, were readily removed from the stabilization diagram. The red dots represent which

modes were considered physical by the algorithm (top two clusters with more elements). Observing the processed diagram, one can say that, apparently, just one stable column was detected. However, when a zoom is given (Fig. 2.4), one can clearly see, the two closely spaced modes detected correctly.

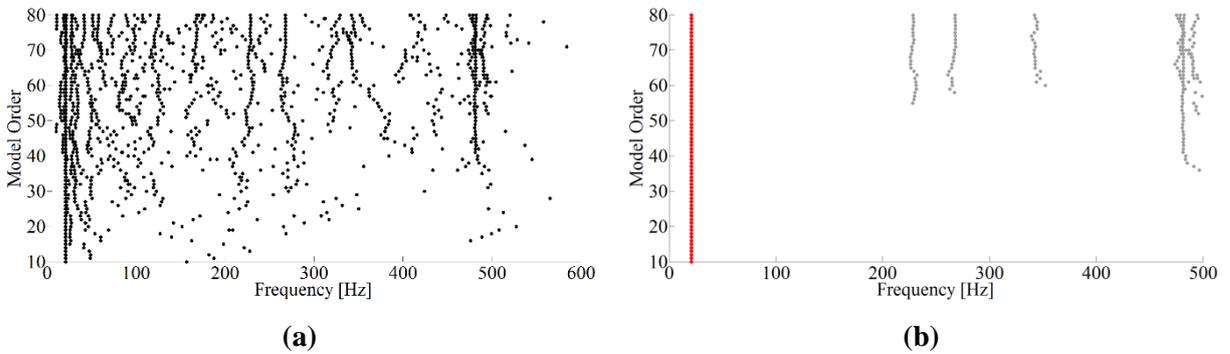


Fig. 2.3 Stabilization diagram: (a) before and (b) after the automated identification by the proposed method.

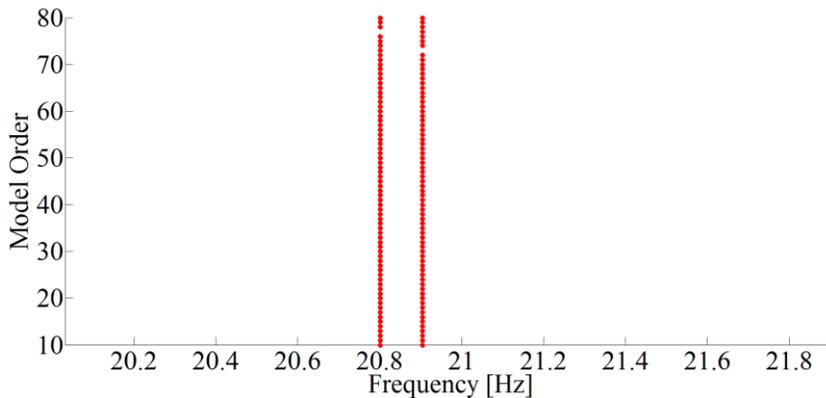


Fig. 2.4 Zoom of the stabilization diagram after algorithm’s application.

A frequency versus damping ratio plot of both identified modes can be found in Fig. 2.5. One can note, by such a figure, that there was no dispersion at all between the modes’ estimates within each one of the two clusters. Of course, this was already expected, once the signal is numerically generated and is free of noise.

Finally, Table 2.2 summarizes the results of the automated identification. It can be seen that the relative errors and the intra cluster standard deviations (σ) were technically null.

Signal with noise

Intending to create a scenario where the proposed methodology could be tested harder, the experiment in this section includes the influence of noise in the signal. Thus, according to

equation (15), 20% ($\beta = 0.2$) of noise was added to the original signal. The high level of embedded noise, depicted in Fig. 2.6, can be seen.

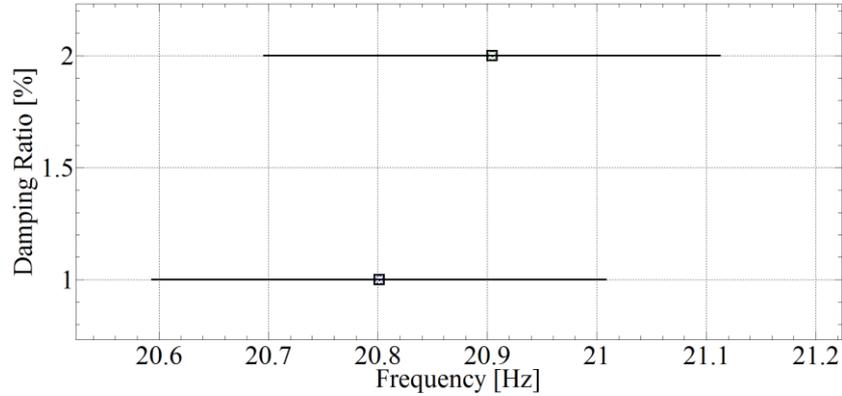


Fig. 2.5 Frequency versus damping ratio plot of the two physical modes (clusters).

Table 2.2: Results of the automated modal identification.

	f (Hz)	σ_f (Hz) / error (%)	ζ (%)	σ_ζ (%) / error (%)
Mode 1	20.79	0.003 / 0.05	0.96	0.00 / 4.00
Mode 2	20.94	0.017 / 0.19	2.06	0.01 / 2.91

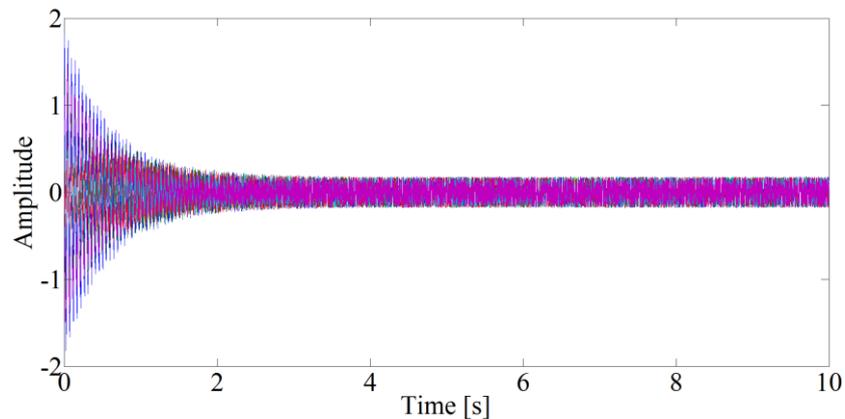


Fig. 2.6 Complete time series of the five channels with a 20% noise.

To allow a better insight about the effect of the noise level, Fig. 2.7 shows a part of the signal before and after the noise addition. It is clear that from a certain time instant, the amplitude of the noise is higher than the pure signal itself.

As in the previous application, intending to obtain the modes' estimates, the SSI-DATA algorithm was applied to models with orders varying from 10 to 80. The obtained stabilization diagram is depicted in Fig. 2.8.

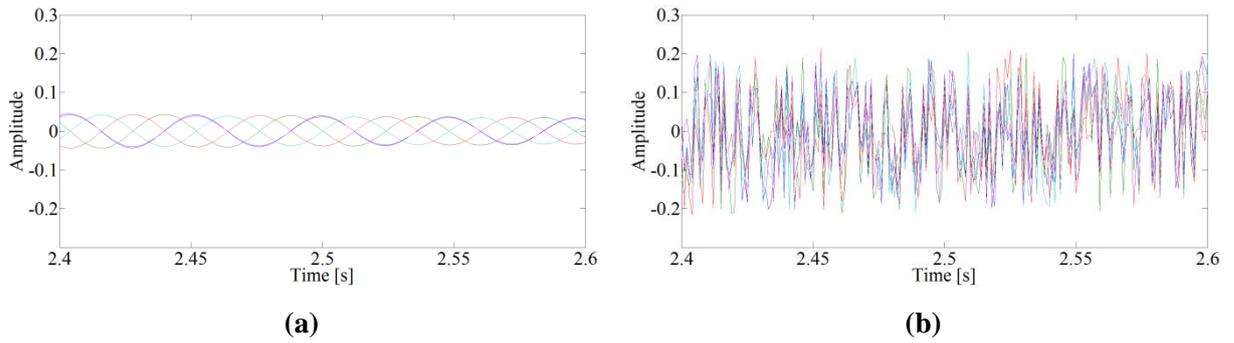


Fig. 2.7 Part of the five channels' signal: (a) before and (b) after noise addition.

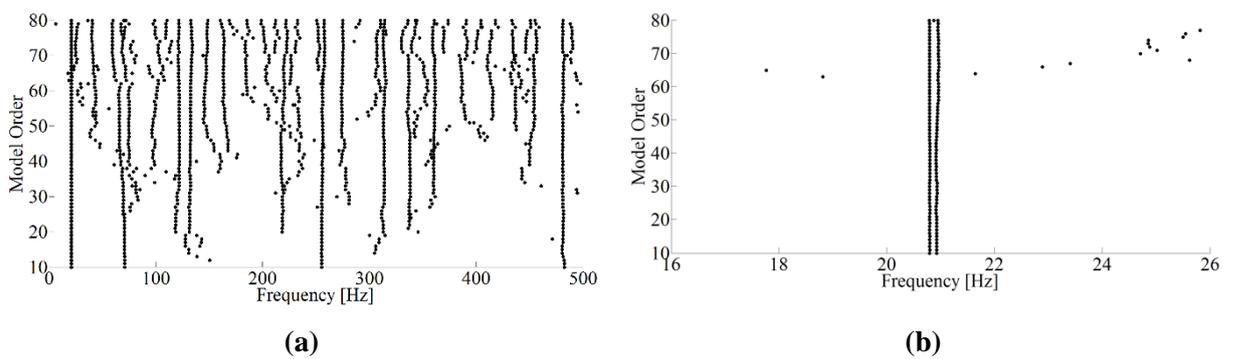


Fig. 2.8 Stabilization diagram generated by SSI-DATA (PC) algorithm: (a) complete diagram and (b) zoom.

After the application of the proposed algorithms, the stabilization diagram was the one shown in Fig. 2.9. One can note that the filters had less impact in this case: only 220 of the 1543 (14.3%) modes' estimates were removed. Despite that, the red dots indicate that the developed method has identified, correctly, the physical modes.

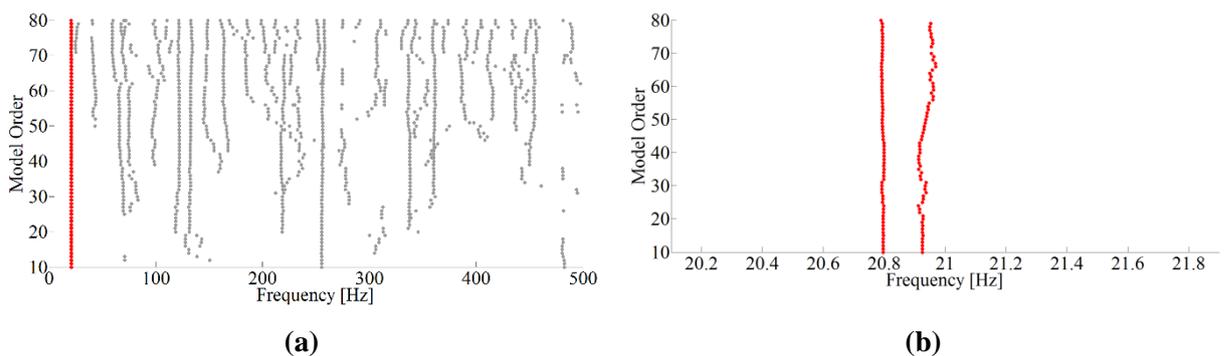


Fig. 2.9 Stabilization diagram after the proposed algorithm's application: (a) complete diagram and (b) zoom.

One can note that despite the large number of spurious modes, the proposed method clustered the modes' estimates in such a way that, clearly, only two groups gathered a

significant amount of elements (corresponding to the two physical modes). This fact can be verified by means of Fig. 2.10. In such a figure, the red-headed stems represent the two selected clusters with more objects.

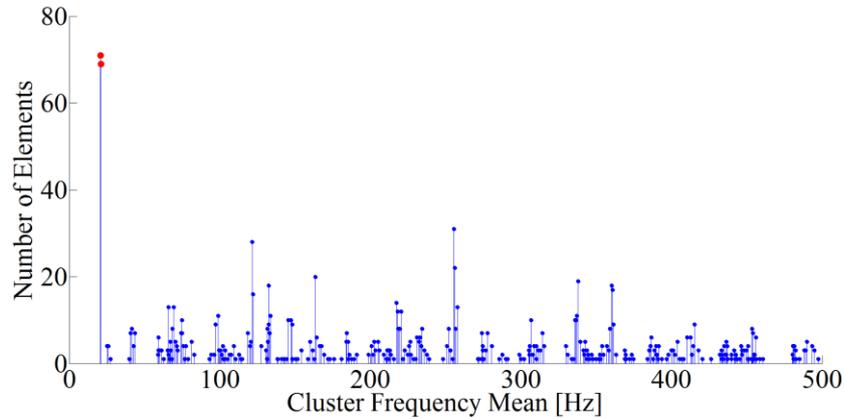


Fig. 2.10 Number of elements per cluster.

The frequency versus damping ratio plot of the selected modes estimates can be seen in Fig. 2.11. It is possible to note that a little dispersion in terms of frequency occurs within the cluster of higher frequency (green circles).

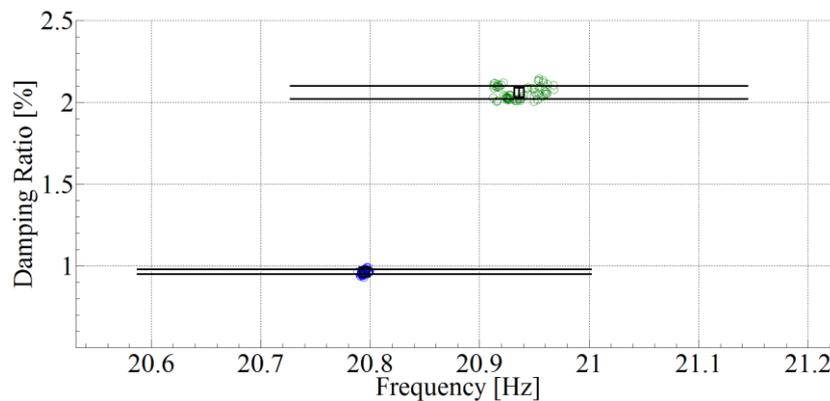


Fig. 2.11 Frequency versus damping ratio plot of the two physical modes (clusters).

The final results and comparisons are gathered in Table 2.3. One can see that in spite of the noise, the method was able to identify, with fair precision, the two modes with, respectively, 20.8 and 20.9 Hz. Considering this artificial scenario, it is possible to say that the proposed methodology, even handling a signal with a significant level of noise, is good enough to identify, correctly, modes with very similar frequencies. In other words, the developed algorithms were able to create a single cluster for each one of the physical modes, treating them as really being different.

Table 2.3: Results of the automated modal identification.

	f (Hz)	σ_f (Hz) / error (%)	ζ (%)	σ_ζ (%) / error (%)
Mode 1	20.79	0.003 / 0.05	0.96	0.00 / 4.00
Mode 2	20.94	0.017 / 0.19	2.06	0.01 / 2.91

2.4.2 Laboratory experiment

This section presents the experimental tests conducted at COPPE/UFRJ laboratory on a simply supported steel beam depicted in Fig. 2.12. This beam is 1.46 m long with rectangular cross section (76.2×8.0 mm) and was instrumented with six piezoelectric accelerometers (PCB, 336C31). Data acquisition was carried out using Lynx ADS2002 equipment, which essentially is a conditioning/amplifier regulating system. This study considered random vibration tests (using a shaker). The random excitation was applied throughout the duration of the tests.

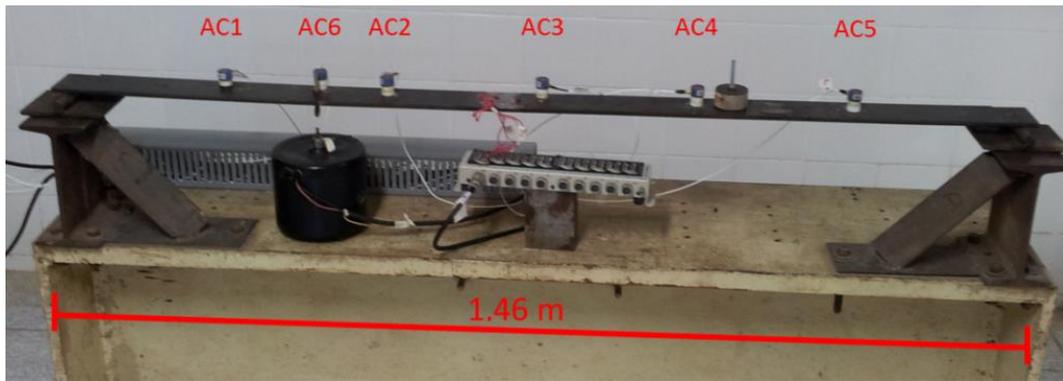


Fig. 2.12 Instrumented steel beam.

Again, the PC variant of SSI-DATA method was applied to the response series (six channels), leading to modes' estimates of models with even orders from 10 to 120. Intending to create a basis for comparisons, the reference and the proposed methodologies are applied to the very same data set.

After a first quick manual judgement, the user-defined parameters were set up: $d_{lim} = 0.01$ and $n_m = 8$ (reference methodology); $d_{lim} = 1$ Hz and $n_m = 5$ (proposed methodology). The m_{lim} filter parameter was set to 0.9. Then, the different methodologies performed the automatic interpretation of such a data. The results of the processes, for both methods, can be checked in Figs. 2.13 to 2.15.

From Fig. 2.13, it is possible to see that the proposed filters had a great impact on the initial data set (Fig. 2.13b). Because of such a feature, 1015 (57.83%) of the 1755 modes' estimates were readily removed. The stabilization diagrams depicted in this figure shows red dots for the automatically identified physical modes.

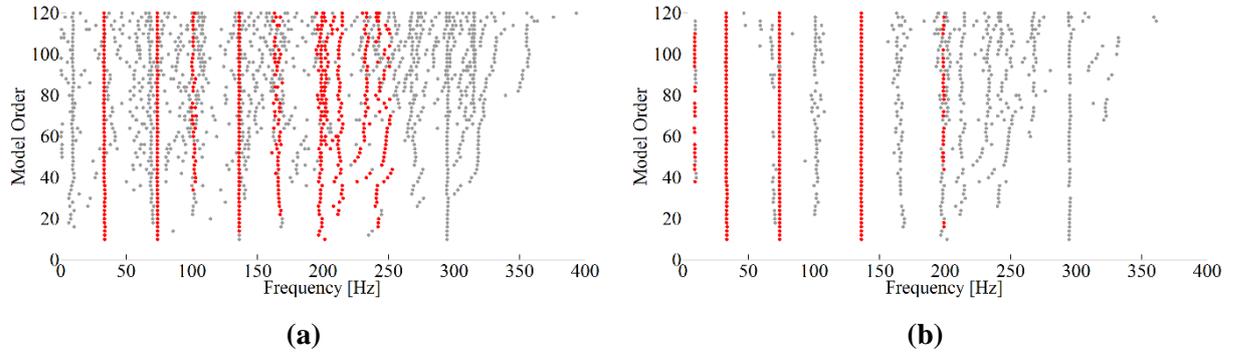


Fig. 2.13 Stabilization diagrams after processing by (a) the reference methodology and (b) the proposed methodology.

The number of elements within each formed cluster can be checked graphically in Fig. 2.14. Stems' plot shown in Fig. 2.14a was generated by the reference methodology, while the stems' plot shown in Fig. 2.14b was the result of the proposed methodology. The red head stems indicate the top nm clusters with more elements, which will be considered as being physical by the algorithms.

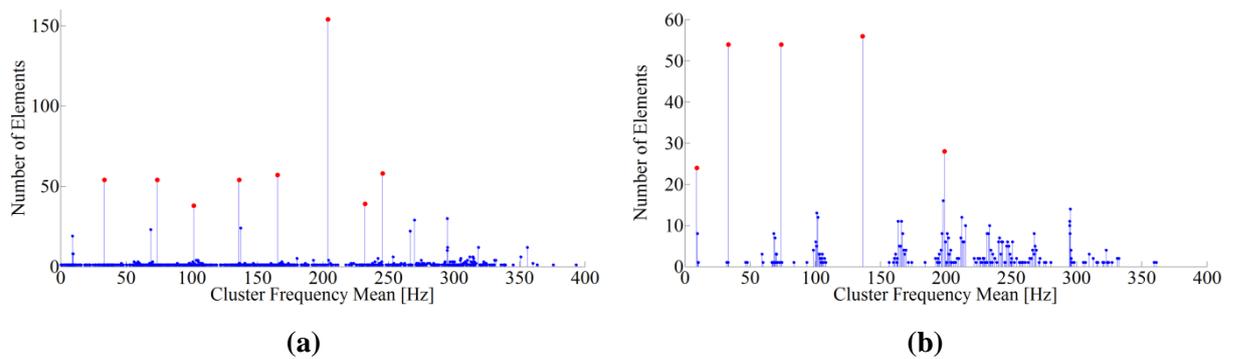
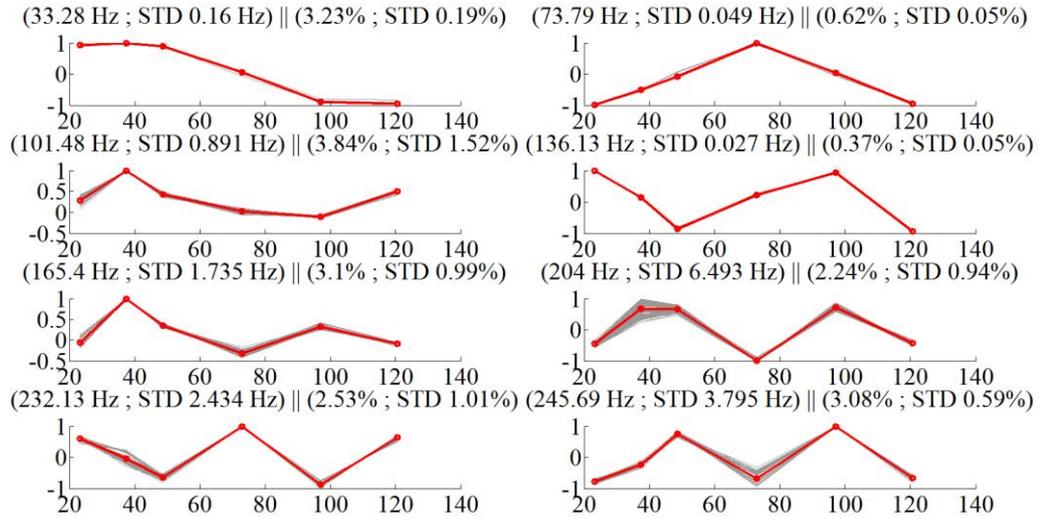


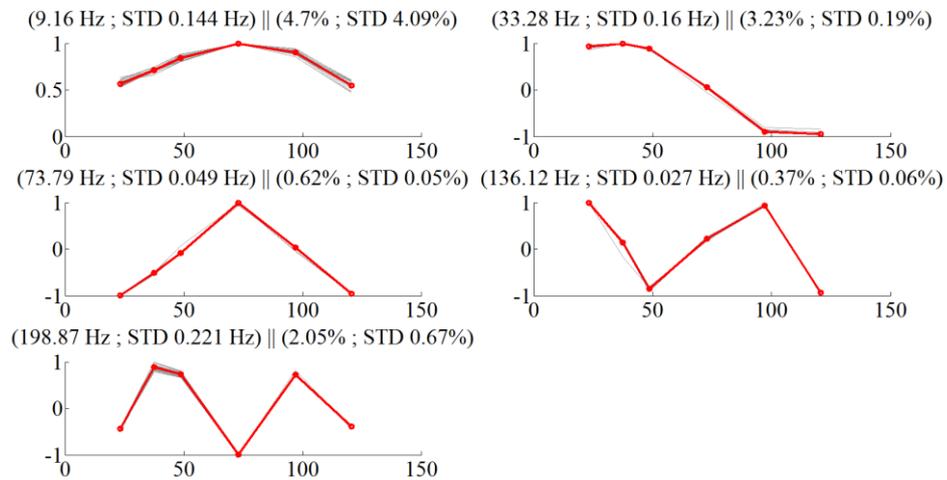
Fig. 2.14 Number of elements grouped into each cluster by (a) the reference methodology and (b) the proposed methodology.

Fig. 2.15 summarizes the results obtained by both methodologies. In such a figure, the first two columns of graphics are the results from the reference methodology, while the last two columns are from the proposed one. It can be noted, in this figure, the modal shapes of each mode estimate within the clusters (grey lines), as well as the mean modal shape of them

(red lines). The small font texts show the values (with standard deviations) of natural frequencies and damping ratios for each mode.



(a)



(b)

Fig. 2.15 Modal parameters obtained by (a) the reference method and (b) the proposed method.

By simple inspection of Fig. 2.15, it is possible to see that the reference method was not able to identify the first mode with frequency close to 9 Hz because its cluster gathered just a few elements (less than 25 as shown in Fig. 2.14). On the other hand, the proposed methodology did identify the first mode. It is remarkable that this experiment had a peculiar happening: the first mode was excited with a low energy when compared to the second, third and fourth, as would reveal a fast Fourier transform (FFT) of the signal.

By analysing Table 2.4, one can conclude that the reference method not only ‘missed’ the first mode but also ‘captured’ the fifth mode with an unacceptable standard deviation (emphasized in bold) in terms of frequency. Differently, the proposed method identified the fifth mode with quite low standard deviation of frequencies. As can be seen, although the proposed method identified the first mode with high standard deviation of damping ratio (emphasized in bold), it was able to ‘capture’ the first mode of 9.16 Hz with just 0.14 Hz of standard deviation.

Table 2.4: Results of the automatic modal identification for the five vertical modes.

Mode	Reference Methodology		Proposed Methodology	
	f (Hz) / σ (Hz)	ζ (%) / σ (%)	f (Hz) / σ (Hz)	ζ (%) / σ (%)
1	-	-	9.16 / 0.144	4.70 / 4.09
2	33.28 / 0.160	3.23 / 0.19	33.28 / 0.160	3.23 / 0.19
3	73.79 / 0.049	0.62 / 0.05	73.79 / 0.049	0.62 / 0.05
4	136.13 / 0.027	0.37 / 0.05	136.12 / 0.027	0.37 / 0.06
5	204.00 / 6.493	2.24 / 0.94	198.87 / 0.221	2.05 / 0.67

2.4.3 PI-57 Oise Bridge

The PI-57 Bridge is a double-deck bridge located near the town of Senlis in France, crossing the Oise River and carrying the A1 motorway, which connects Paris to Lille (Fig. 2.16). The bridge, a 116.50-m-long, cast-in-place, post-tensioned segmental structure built in 1965, consists of three continuous spans of 18.00, 80.50 and 18.00 m. The two lateral spans play the role of counterweights.



Fig. 2.16 PI-57 Bridge.

Dynamic tests were performed under ambient excitation: the traffic was used as a source of excitation. Sixteen piezoelectric accelerometers (Bruel & Kjaer 4507B-005 with sensitivity of 1 V/g, frequency range from 0.4 to 6000 Hz, maximum operational level of 75 g, temperature range from -54 to $+100$ °C) have been installed on the bridge deck (Lille/Paris – Fig. 2.17). For the acceleration recording, a data programmable controller Gantner E-PAC DL was used and connected to an 8 GB USB flash drive. Data were transferred by a TCP/IP modem. Accelerations were filtered within the 0–30 Hz frequency range and sampling was set to 0.004 s (250 Hz) for 5 min. To make the data processing amenable, structural data were only recorded every 3 h over a 24-h time period and stored on a buffer hard disc.

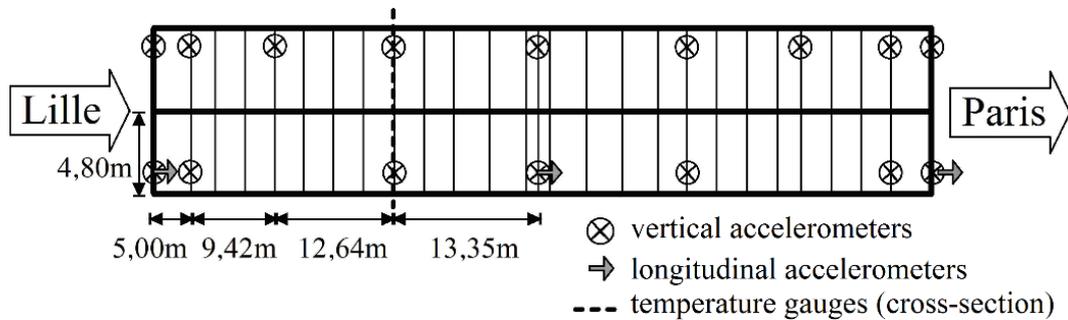


Fig. 2.17 Plane view of the bridge with monitoring system.

Exactly like in the previous experiments, the SSI-DATA (PC) was applied to estimate structural modes considering both the reference methodology and the proposed methodology. However, differently from the previous test, in this application, the first method had d_{lim} set to 0.1. On the other hand, the proposed algorithm had its distance threshold parameter set to the very same value of the last two experiments ($d_{lim} = 1$ Hz), with one difference: here the proposed method also eliminated (through its pre-filter) all the modes' estimates with frequencies out of the 0–50 Hz range. Both methods had the parameter n_m set to 8.

The stabilization diagrams after processing can be checked in Fig. 2.18. Again, as can be seen in the diagram generated by the proposed methodology (Fig. 2.18b), the filters had a significant impact on the initial data: 81.39% (1347) of the initial 1655 modes' estimates were removed.

The number of elements per cluster can be seen in Fig. 2.19. Stems with red heads represent the n_m selected clusters with more elements.

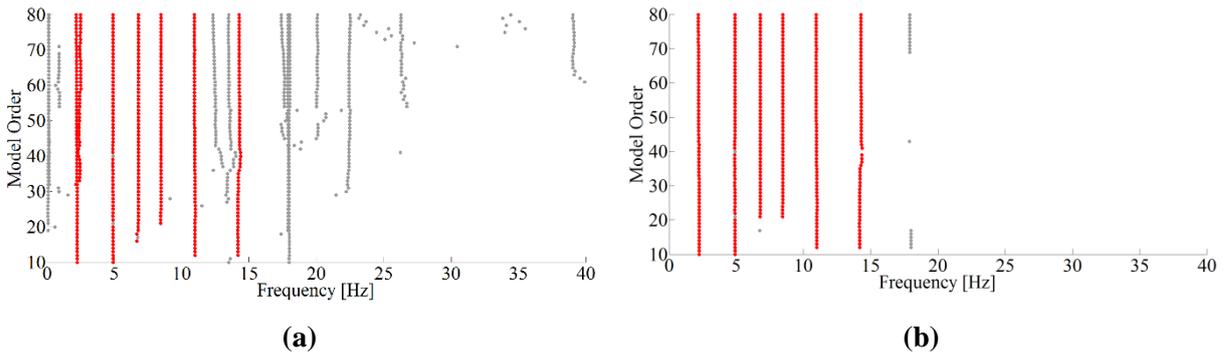


Fig. 2.18 Stabilization diagrams after processing by (a) the reference methodology and (b) the proposed methodology.

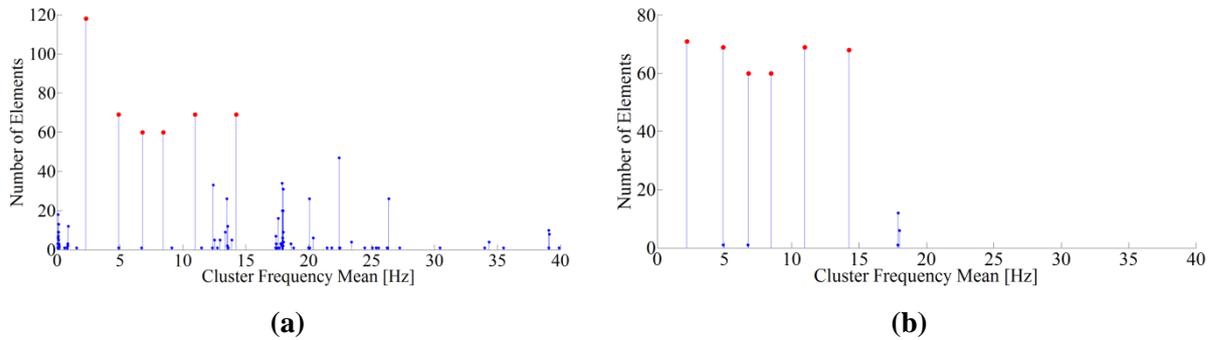


Fig. 2.19 Number of elements grouped into each cluster by (a) the reference methodology and (b) the proposed methodology.

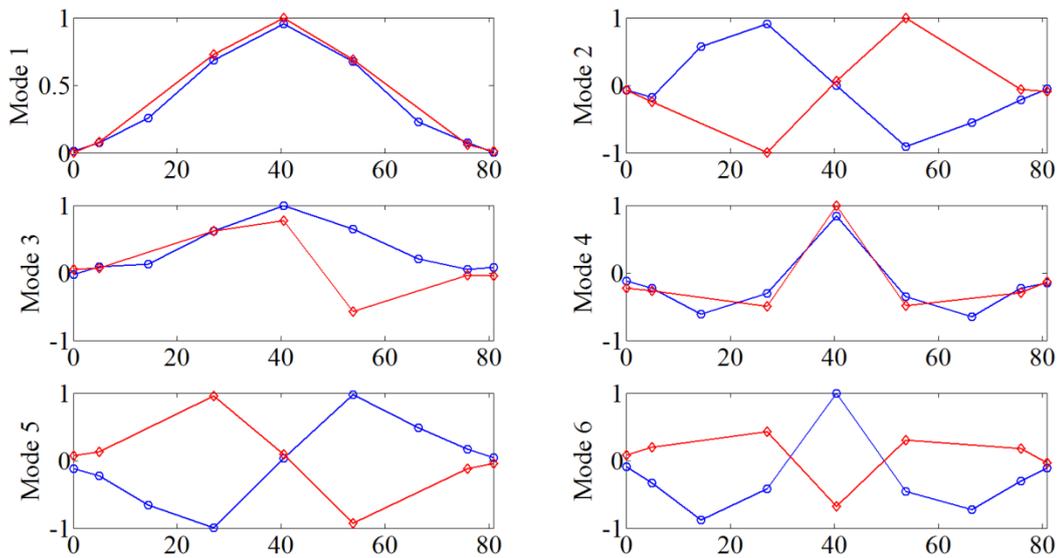


Fig. 2.20 Mode shapes obtained by both methodologies.

Regarding the mode shapes, both methodologies technically found the very same results, which are shown in Fig. 2.20. In this figure, the blue line corresponds to the upper

nine vertical accelerometers in the plane view of Fig. 2.17, while the red corresponds to the lower line (in the plane view) of seven vertical accelerometers, as shown in the same figure.

As shown in the results summarized in Table 2.5, for this application, the methodologies achieved very similar results, except for the higher standard deviation of the first mode identified by the reference methodology (emphasized in bold).

Table 2.5: Results of the automatic modal identification for both methodologies.

Mode	Reference Methodology		Proposed Methodology	
	f (Hz) / σ (Hz)	ζ (%) / σ (%)	f (Hz) / σ (Hz)	ζ (%) / σ (%)
1	2.30 / 0.118	9.02 / 8.58	2.21 / 0.027	2.43 / 0.64
2	4.89 / 0.003	0.93 / 0.03	4.89 / 0.003	0.93 / 0.03
3	6.78 / 0.004	1.92 / 0.02	6.78 / 0.004	1.92 / 0.02
4	8.46 / 0.006	2.87 / 0.15	8.46 / 0.006	2.87 / 0.15
5	10.96 / 0.022	3.60 / 0.11	10.96 / 0.022	3.60 / 0.11
6	14.24 / 0.054	2.62 / 0.18	14.24 / 0.052	2.62 / 0.18

2.5 Results discussion

First, concerning the numerical experiment, one can conclude that the proposed method is suitable for identifying modes with very close natural frequencies, even in the presence of high levels of noise. However, it should be noted that the correct closely spaced modes' identification is mainly due to the parametric modal identification method that was used (SSI-DATA).

Considering the laboratory and the bridge applications presented, it is possible to note that compared to the reference approach, the proposed methodology works with a much lower number of mode estimates due to the pre-filter application. As a result, a faster and more accurate clustering process is achieved. In the laboratory test, for instance, 1015 (57.83%) of the 1755 initial modes' estimates were automatically considered as certainly spurious and, therefore, were immediately removed. This fact can be noted by comparing the full stabilization in Fig. 2.13a and the filtered one in Fig. 2.13b. In a similar way, concerning the bridge experiment, as can be seen in Fig. 2.18, the pre-filter also deleted a large number of undesired noisy modes. More precisely, this quantity reached 81.39% (1347) of the initial 1655 modes' estimates. These expressive numbers of removed modes are mainly due to the MPC limit, which was set to 0.9. However, at this point, it is important to say that such a parameter must be used with caution [3].

The hierarchical clustering process over the modes' estimates, or over the few remaining ones in case of the proposed method, deserves special attention. For the laboratory experiment, (Figs. 2.13 to 2.15), one may observe that the reference method was not able to identify the first vertical bending mode. However, in spite of the low-energy excitation of this mode during the tests, the proposed method did find this mode. Several values for the distance threshold (d_{lim}) were tested the reference method to try to induce the algorithm to 'catch' the first mode. However, the results were conclusive: as one increases the value of d_{lim} , before the low frequencies' clusters could gain enough modes to be ranked among the top n_m , the higher frequencies' clusters start to gather, in a too permissive way, a lot of modes (resulting in a fast growth of their stems –Fig. 2.14a). Consequently, a high scatter between modes belonging to a high-frequency cluster can be observed. Evidently, this excessive scattering is undesirable. A possible way of forcing the reference algorithm to catch the first mode is by keeping the d_{lim} value fixed and increasing the n_m parameter in an extreme conservative way. However, one can say that this solution is not elegant, since a lot of spurious clusters (modes) would be selected along with the physical ones.

The high-frequency scattering phenomenon happens, mainly, for two reasons. First, the distance metric, which is written in equation (9), presents a dimensionless ratio of frequencies and, therefore, does not measure a distance between low-frequency modes with the same rigour as it measures a distance between high-frequency modes. Second, the single-linkage criterion for measuring the distance between two already formed clusters promotes a decrease in the heights of the hierarchical U-shaped branches, that is, it makes easier to 'gather' two clusters when compared to the complete linkage criterion.

On the other hand, the novel methodology proposed in this article is immune to such a phenomenon. That is due to the fact that the distance metric, stated in equation (13), does not contain any frequency ratio and, therefore, ensures that the distances between modes (with either high or low frequency) will be expressed by an absolute value in Hertz.

Another remarkable point is that the success of the algorithm proposed in this work is not as sensitive to the distance threshold as the reference algorithm is. For instance, in all tests, the user-defined limit for the proposed algorithm did not need to be adjusted ($d_{lim} = 1$ Hz). This was achieved because of the novel metric and the proposed filters.

The experiments showed that the peculiarities of the two examined automation methods produce significant differences at the final process output. However, the proposed

methodology, when compared to the reference in the studied applications, generated better results and, therefore, may represent an advance towards a more robust automatic modal identification method.

2.6 Practical relevance and potential applications

As the results show, the features included in the proposed methodology allowed accurate results for the automatic modal identification of civil engineering structures. With this algorithm, the large amount of data generated by the online dynamic monitoring can be treated automatically without significant loss of quality in the outputs. For the future, the suggested routine can be embedded in a standalone program with a graphical user interface to turn the application even more friendly and useful.

Although this article has used the SSI-DATA method to process the structure response, the proposed algorithm is also able to treat the outputs of any other parametric identification method that generates a stabilization diagram.

The proposed method is, therefore, a useful tool to modal identification of civil engineering structures subjected to ambient vibrations (OMA). Since this kind of test is by far the most practical one, it is of utmost importance that the algorithms dedicated to automatically interpret the data have a reliable efficiency, which was achieved in this work.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was financially supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). The authors also thank IFSTTAR (former LCPC – Laboratoire Central des Ponts et Chaussées) and SANEF (Société des Autoroutes du Nord et de l’Est de la France) for the data used in this paper (*project 0560V407*).

References

- [1] Peeters B and De Roeck G. One year monitoring of the Z24-bridge: environmental effects versus damage effects. *Earthq Eng Struct*, **2001**; 30: 149–171.
- [2] Magalhães F. *Operational modal analysis for testing and monitoring of bridges and special structures*. PhD Thesis, University of Porto, Portugal, **2010**.
- [3] Reynders E, Houbrechts J and De Roeck G. Fully automated (operational) modal analysis. *Mech Syst Signal Pr*, **2012**; 29: 228–250.
- [4] Ubertini F, Gentile C and Materazzi AL. Automated modal identification in operational conditions and its application to bridges. *Eng Struct*, **2013**; 46: 264–278.
- [5] Cabboi A. *Automatic operational modal analysis: challenges and applications to historic structures and infrastructures*. PhD Thesis, Università degli Studi di Cagliari, Italy, **2013**.
- [6] Peeters B and De Roeck G. Reference-based stochastic subspace identification for output-only modal analysis. *Mech Syst Signal Pr* **1999**; 13(6): 855–878.
- [7] Van Overschee P and De Moor B. *Subspace identification for linear systems: theory, implementation and applications*. Dordrecht: Kluwer Academic Publishers, **1996**.
- [8] Juang JN and Pappa RS. An eigensystem realization algorithm for modal parameter identification and model reduction. *J Guid Control Dynam*, **1985**; 8(5): 620–627.
- [9] Pappa RS, Elliott KB and Schenk A. Consistent-mode indicator for the eigensystem realization algorithm. *J Guid Control Dynam*, **1993**; 16(5): 852–858.
- [10] Vacher P, Jacqier B and Bucharles A. Extensions of the MAC criterion to complex modes. In: *Proceedings of the international conference on noise and vibration engineering (ISMA 2010)*, Leuven, 20–22 September **2010**.
- [11] Yun GJ, Lee SG and Shang S. An improved mode accuracy indicator for eigensystem realization analysis (ERA) techniques. *KSCE J Civ Eng*, **2012**; 16(3): 377–387.
- [12] Cardoso RA. *Development of methodology for automatic modal identification of structures*. MSc dissertation, Federal University of Ouro Preto, Brazil, **2015**.

Chapter 3

Paper #2 – A robust methodology for modal parameters estimation applied to SHM

The previous chapter showed a method able to perform modal identification with little human judgment. Briefly, a stabilization diagram is interpreted by unsupervised machine learning algorithms. Although a high level of automation and accuracy has been achieved, the choice of the parametric model order and the choice of the number of desired modes consisted in a particular limitation for real-time SHM applications. Hence, complementary studies were carried out. This research culminated with the implementation of a new strategy: by automatically varying the model order and the number of desired modes, several stabilization diagrams are generated. Each one of them is interpreted by using the clustering procedures already presented. Afterward, the results of the various interpretations are subjected to another clustering step.

Consequently, despite the increase in computational effort, more automation and robustness were achieved. Thus, any abnormal deviations can be promptly detected, since natural frequencies, damping ratios and mode shape amplitudes are tracked over time. **With this methodology, the objective related to the first approach of this thesis, i.e., to detect damage based on the tracking of structural modal parameters, is fulfilled.** The resulting paper was published in the journal “*Mechanical Systems and Signal Processing*”.

Paper Information

Status	Published in 2017
DOI	https://doi.org/10.1016/j.ymssp.2017.03.021
Publisher	Mechanical Systems and Signal Processing ISSN 0888-3270; Qualis A1; Impact Factor (2018): 4.370
Citation	Cardoso R, Cury A, Barbosa F (2017). A robust methodology for modal parameters estimation applied to SHM. <i>Mechanical Systems and Signal Processing</i> , 95(2017): 24-41.

Cardoso R, Cury A, Barbosa F (2017). A robust methodology for modal parameters estimation applied to SHM. *Mechanical Systems and Signal Processing*, 95(2017): 24-41.

Abstract

The subject of structural health monitoring is drawing more and more attention over the last years. Many vibration-based techniques aiming at detecting small structural changes or even damage have been developed or enhanced through successive researches. Lately, several studies have focused on the use of raw dynamic data to assess information about structural condition. Despite this trend and much skepticism, many methods still rely on the use of modal parameters as fundamental data for damage detection. Therefore, it is of utmost importance that modal identification procedures are performed with a sufficient level of precision and automation. To fulfill these requirements, this paper presents a novel automated time-domain methodology to identify modal parameters based on a two-step clustering analysis. The first step consists in clustering modes estimates from parametric models of different orders, usually presented in stabilization diagrams. In an automated manner, the first clustering analysis indicates which estimates correspond to physical modes. To circumvent the detection of spurious modes or the loss of physical ones, a second clustering step is then performed. The second step consists in the data mining of information gathered from the first step. To attest the robustness and efficiency of the proposed methodology, numerically generated signals as well as experimental data obtained from a simply supported beam tested in laboratory and from a railway bridge are utilized. The results appeared to be more robust and accurate comparing to those obtained from methods based on one-step clustering analysis.

3.1 Introduction

Over the last decade, methods using raw dynamic measurements are proving to be a viable alternative for the detection of structural damage. Santos [1], for instance, presented a clever technique that relies on time series symbolic data objects and moving windows to detect structural novelties. Alves [2] presented a methodology based on the use of symbolic data coupled with clustering techniques to detect damage using raw acceleration measurements. Despite all advances in this field of research, it is common to find studies in the literature that considers modal parameters (natural frequencies, damping ratios and mode shapes) as preferred inputs to assess structural behaviors or damage states. Evidently, the close

relationship that exists between modal parameters and structural stiffness and mass explains this tendency. Therefore, it is not by chance that modal parameters are widely used in continuous dynamic monitoring systems of important civil structures, such as long-span bridges, stadiums, tall buildings, among others [3,4].

Continuous monitoring systems are especially relevant due to their preventive nature, since they can be inserted in maintenance programs and actually support detecting unusual structural behaviors. Abnormal structural states are often assessed by observing the variation of modal parameters over time. Hence, the modal identification procedure becomes of utmost importance and deserves special attention concerning its accuracy, robustness and automation to make it viable in a context of online monitoring. Due to the necessity of an automatic feature, parametric time-domain based identification methods have been preferred among engineers. Mostly, the so-called SSI (Stochastic Subspace Identification) methods, which use discrete-time stochastic state-space models, are very suitable in this case due to their capacity to identify closely-spaced frequencies. Moreover, they provide useful ready-to-use data for feature extraction by means of clustering techniques or any other data mining methods.

At this point, it is worth highlighting the difference between *modal parameter estimation* (MPE) and *modal tracking* (MT). The former consists in the estimation of modal parameters from a single record of measured data and the latter corresponds to tracking the evolution of modal parameters of a structure through repeated MPE. This paper is limited to MPE, not involving what would be a complementary task to online monitoring (MT). Therefore, isolated signals of a given dynamic test in a structure are evaluated independently, with no time tracking of modal parameters.

In fact, without the automation of the MPE process, a lot of user interaction over a large amount of measured vibration data would be required, leading to an impossibility of practical applications such as the health monitoring of important infrastructures. Therefore, especially over the last decade, several methods aiming the reduction of the number of user-defined parameters have been developed [5–8], some of which reached the complete elimination of all manually specified parameters and, thus, are usually labelled as fully automated.

For instance, Reynders [6] proposed a fully automated approach for the interpretation of stabilization diagrams, which are usual outputs of SSI methods. Their approach was cleverly based on clustering techniques built over three stages: pre-cleaning by means of a

classification of all modes into two categories (possibly physical or certainly spurious); a hierarchical clustering of the possible physical ones to group them together; and a final classification of the formed clusters into a physical or spurious condition. No user-defined parameter is needed in the entire process.

However, despite the observed advances, practice has been showing that even the most sophisticated automated process of MPE has to be preliminarily checked or tuned in order to fit a unique behavior of the analyzed structure. Otherwise, the method might become too much vulnerable to fail due to different identification scenarios, like different levels of frequencies and damping ratios, different complexities of modal shapes, several types of external excitation sources and different amounts of signal noise. For that reason, the present work focus on the development of a robust automated approach that depends on just few easy-to-set user-defined parameters.

Currently, the available time domain methods for operational modal analysis (OMA), explored in the work of Peeters and De Roeck [9], are essentially based on two types of models: discrete-time stochastic state space models and ARMA (Auto- Regressive Moving Average) or just AR (Auto-Regressive) models. The models can be identified either from correlations (or covariances) of the outputs: Covariance driven Stochastic Subspace Identification – SSI COV; or directly from time series collected at the tested structure by the use of projections [10]: Data driven Stochastic Subspace Identification – SSI-DATA. As reported in the work of [9], these two methods are very closely related. Still, the SSI-COV has the advantage of being faster and based on simpler principles, whereas the SSI-DATA allows obtaining some further information with a convenient post-processing, as for instance, the decomposition of the measured response into modal contributions.

This work focuses on the use of the SSI-DATA method, since this method yields stabilization diagrams as outputs, which are a resourceful data for reliable structural modal identification. Moreover, when associated with clustering techniques, they are capable of providing very robust results. Thus, Section 3.2 briefly discusses the SSI DATA method (a more thorough description can be found in reference [10]).

To provide a comparison basis for this new methodology, a reference procedure based on the work of [5] is presented. Section 3.3 details both methodologies. Section 3.4 shows the results obtained by both approaches applied to different identification scenarios: numerically generated signals, a simply supported steel beam tested in laboratory and a real test on a high-

speed train (TGV) viaduct in France. Finally, a brief conclusion and comments on practical relevance of the work are drawn in Section 3.5.

3.2 Data-driven Stochastic Subspace Identification

Since the work published in the nineties [10], the application of data-driven SSI algorithms to determine the modal parameters of structures has been quite spread over the ambient vibration tests. By computing state space models from output data (structure's response) the identification problem can be enunciated as follows.

Given s measurements of the output $\mathbf{y}_k \in \mathbb{R}^l$ generated by the unknown stochastic system of order n :

$$\begin{aligned}\mathbf{x}_{k+l} &= \mathbf{A}\mathbf{x}_k + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{v}_k\end{aligned}\quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector at a discrete time instant k , \mathbf{w}_k and $\mathbf{v}_k \in \mathbb{R}^n$ are zero mean, white vector sequences and l is the number of outputs of the process (number of instrumented channels).

Knowing matrices \mathbf{A} and \mathbf{C} , one can calculate the modal frequencies ω_i , the modal damping ratios ξ_i and the corresponding modal shapes ϕ_i . This calculation is performed through the eigenvalue decomposition of matrix \mathbf{A} as the following statements show:

$$\lambda_i = \frac{\ln(\mu_i)}{\Delta t} \rightarrow \begin{cases} \omega_i = |\lambda_i| \text{ (rad / s)} \rightarrow f_i = \frac{|\lambda_i|}{2\pi} \text{ (Hz)} \\ \xi_i = \frac{-\text{Re}(\lambda_i)}{|\lambda_i|} \times 100 \text{ (\%)} \\ \phi_i = \mathbf{C}\psi_i \end{cases} \quad (2)$$

where μ_i is the i^{th} discrete-time eigenvalue of \mathbf{A} and ψ_i is its corresponding eigenvector. The eigenvalues of \mathbf{A} are also called *poles*. They are complex numbers and appear mostly in conjugated pairs. Only poles which have a conjugated pair and have positive imaginary component are taken into account for the Eq. (2). Therefore, up to $n/2$ modes are expected to be estimated. The $\text{Re}(\bullet)$ is the symbol for the real part of \bullet and λ_i denotes the continuous-time eigenvalue of \mathbf{A} .

When one works with real structure response data, it is very hard to determine the system's order. Thus, one of the tools to overcome this problem is the well-known stabilization diagrams, where the modes estimates are plotted for different system orders.

3.3 Automation methods for stabilization diagram interpretation

The modes estimates, which are the outputs of SSI-DATA algorithm, are usually presented in stabilization diagrams. The analysis of such diagrams is supposed to separate physical modes from spurious (numerical) ones. The latter inevitably exist among the data, since they are a natural consequence of the parametric model attempt to fit the response data.

The process of detecting physical modes within a stabilization diagram can be understood as its interpretation, which, ideally, must be carried out automatically. To this end, this section presents two techniques, both based on hierarchical clustering analysis. The first one, explained in Section 3.1, was proposed in the work of [5] and will serve as a baseline. The second one presents a richer modal identification method by proposing a second clustering analysis, shown in Section 3.2. The proposed approach is an enhancement from the first and contains some elements found in [6]. In the remainder of the text, these two techniques will be referenced as “reference methodology” and “proposed methodology”, respectively.

Although this work uses SSI-DATA routine to provide stabilization diagrams, it is relevant to point out that both interpreters are suitable for the analysis of the outputs generated by any other parametric identification technique that yields modes estimates (frequencies, damping ratios and modal shapes) for several model orders.

3.3.1 Reference methodology

This approach is based on the routines presented by [5] and does not apply a pre-filter to the modes estimates data. Thus, stabilization diagrams would still be full of certainly spurious modes, like those with extremely high or negative damping ratios.

This method is based on a hierarchical clustering algorithm. Hence, in order to establish a concept of similarity/dissimilarity between objects, a dissimilarity (distance) measure

between all pairs of modes estimates must be calculated. The adopted metric for the evaluation of such a distance is:

$$d_{i,j} = \left| \frac{f_i - f_j}{f_j} \right| + (1 - \text{MAC}_{i,j}) \quad (3)$$

where f_i and f_j are, respectively, the natural frequencies of the modes estimates i and j . The mode shape similarity between these two modes estimates is evaluated by the well-known Modal Assurance Criterion [11]. Moreover, one can note that for this metric it is possible to have $d_{i,j} \neq d_{j,i}$, which leads to an oddly asymmetric dissimilarity matrix.

Hierarchical clustering algorithms differ in the way they compute the distance between two already formed clusters. In this methodology, the single-linkage criterion is used, which means that the distance between two clusters will be considered as being equal to the smallest distance, also computed with Eq. (3), between objects inside these two clusters. This information allows constructing the hierarchical tree.

In the following step, one must define the tree's cut-level. This is usually dependent on the expected number of clusters. However, this number is not trivial to predict due to the unknown quantity of clusters grouping spurious modes. Thus, the alternative strategy is the use of a distance limit (d_{lim}). Such threshold is the criterion to prone the branches of the hierarchical tree, generating clusters that are distant from each other more than that value. Hence, the lower is the limit, the higher is the number of resulting clusters. In this methodology, such threshold is a user-defined parameter.

At this point, one has each cluster representing a single mode. Then, the clusters are ranked according to their number of elements to help deciding whether the mode is spurious or not. Thus, the top n clusters with more elements are selected. Since physical modes are very consistent for models with different orders, their clusters present a much higher number of elements than the clusters that contain numerical estimates (which present a higher scatter between models of different orders). Therefore, it is expected to find that the n selected modes are indeed physical. In practice, it is common to realize that the number of physical modes n can be guessed, for instance, by a simple preliminary frequency domain analysis. In this methodology, this number of modes is also a user-defined parameter.

The next step is to take into account the damping ratios by means of an outlier analysis within each selected cluster. This internal filtering eliminates the modes estimates with extreme damping ratio values, which are out of the range defined by $\pm 2.698\sigma$ (standard deviation) that embraces 99.3% of the samples in a Gaussian distribution.

Finally, each selected cluster yields mean values of the modal parameters (natural frequency, modal damping ratio and mode shape), which are the outputs of the methodology.

3.3.2 Proposed methodology

The main idea of the proposed methodology is to enclose three fundamental steps:

- use a time-domain method for fitting parametric models (such as SSI-DATA) to get modes estimates (stabilization diagram);
- carry out the interpretation of such diagrams using enhanced statistical techniques;
- perform a novel cluster analysis over the various interpretations results (*realizations*) obtained in the previous step.

Thus, the proposed approach contains three main subroutines, namely: stabilization diagram generator (SSI-DATA), stabilization diagram interpreter (first clustering analysis) and the second clustering analysis routine. The flowchart of Fig. 3.1 shows each step of the proposed methodology.

During steps 1 and 2 (Fig. 3.1), the user must feed the routines with some input parameters, namely:

- n – number of desired modes to be identified as final output;
- f_s – signals sampling frequency;
- f_{min} and f_{max} – minimum and maximum frequencies to delimit a frequency band to be considered in the analysis;
- m_{lim} – minimum MPC value of a mode estimate to be considered as physical.

In step 3, the limits for the number of desired modes (n_{min} and n_{max}) are computed depending on the value of n . Those values are set as 50% less or 50% more than n , rounded to the nearest integer. For instance, for the case when n is equal to 4, n_{min} and n_{max} would be equal to 2 and 6, respectively.

\mathbf{M}_k , the smaller are these values (step 7). This is due to the fact that high-order state-space models imply on huge block Hankel matrices, which may lead to an exorbitant computational effort for solving the pseudo-inverse operation of Eq. (2).

At this point, the third loop (step 10) begins, which is over each value of n_m (number of modes of this iteration), varying from n_{min} and n_{max} . Once SSI-DATA yields the stabilization diagram in the previous step, an enhanced interpreter is called. This consists in step 11, which is the second main subroutine (first clustering analysis). Its flowchart is depicted in Fig. 3.2.

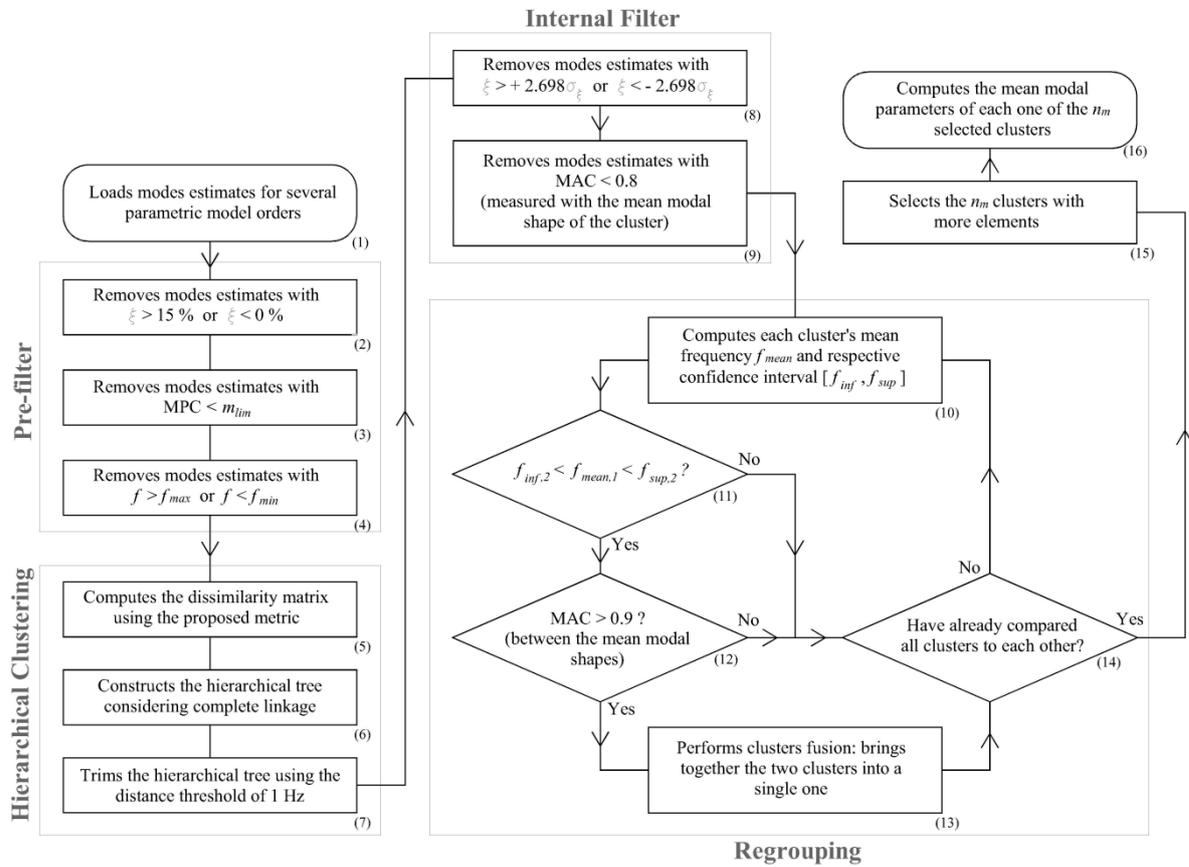


Fig. 3.2 Flowchart of the enhanced interpreter (1st clustering analysis) – step 11 of flowchart of Fig. 3.1.

Compared to the reference interpreter, the main new features found in the flowchart (Fig. 3.2) rely on the innovative way of measuring distances between modes estimates during the clustering process. Moreover, modifications regarding filtering of spurious modes and a novel technique for regrouping clusters, which were originally separated incorrectly, are also included at this step.

Like the reference method, this diagram interpreter is based on a hierarchical clustering algorithm. However, differently from that approach, one intends avoiding unnecessary computational effort and aims to obtain better results from the clustering process. Thus, a pre-filter is applied on the modes estimates data, which removes all modes whose damping ratios are not between 0 and 15% (common values for civil engineering structures). In addition, all modes with Modal Phase Collinearity (MPC) indicator lower than the limit m_{lim} in the [0, 1] interval of this criterion are eliminated. The MPC index of a vector ϕ was stated by [12], revised by [13] and specifically studied, together with MAC (Modal Assurance Criterion), by [14]. A number of authors, such as [6,15], used this index to help distinguishing physical modes from spurious ones. An MPC value closer to 1 indicates that the mode shape vector is “monophasic” and, hence, is more likely to be physical, while values closer to 0 show that the mode is either a noise mode or the mode is significantly complex.

Furthermore, the user may also remove all modes estimates out of a predefined range of frequencies (much like a “bandpass” filter on the modes estimates data). All operations explained until this point correspond to steps 2, 3 and 4 of the flowchart in Fig. 3.2.

After having the modes estimates data processed by this pre-filter, a plot of the stabilization diagram reveals that a significant amount of modes were removed, yielding a clearer aspect. Depending on the situation, and this is not rare, the removed modes may even represent more than 50% of the initial data. These removed modes estimates are considered as certainly spurious (or undesired) and will not be processed by the subsequent clustering algorithm.

In step 5 (Fig. 3.2), for the distance calculation between the modes estimates i and j , the following metric is proposed:

$$d_{i,j} = |f_i - f_j| + (1 - MAC_{i,j})c \quad (4)$$

where c is a constant arbitrarily¹ set to 5 Hz, which after some tests was established by the authors for being a satisfactory weight value. One should note that, differently from Eq. (3), this metric leads to a perfectly symmetric dissimilarity matrix, i.e. $d_{i,j} = d_{j,i}$.

¹ During initial studies, the authors varied this constant from 1 to 10 Hz without obtaining relevant differences in the outputs, which points out the low sensitiveness of the enhanced interpreter to such a parameter. As the authors proved in Cardoso [16] and in this paper, the weighting value set to 5 Hz was appropriate to all presented applications. So far, there was no need to set this constant to another value in order to ensure the success of the method and, therefore, no deeper study about this topic was performed.

At this point, the hierarchical clustering is applied. However, differently from the reference method, the distance between two already formed clusters is measured according to the complete linkage criterion, which means that such a distance is equivalent to the largest possible distance between two of their elements. With this information, one can construct the hierarchical tree that will be trimmed by a constant distance threshold of 1 Hz (steps 6 and 7 in Fig. 3.2). So far, there is no need to set this threshold to another value in order to ensure the success of the method and, therefore, it remains fixed as internal constant of the routine [16].

Contrary to what the reference technique does, the internal filter is applied within the clusters before they are ranked according to their number of elements. This is done to penalize the clusters with high scatter (likely to be representing a numerical mode) before they could be ranked among the top n_m modes that will be selected as physical ones. Thus, the internal filter removes outlier modes within the cluster considering not only damping ratios, but also mode shapes. Mode estimates having damping ratios out of the $\pm 2.698\sigma$ range and/or having MAC values lower than 0.8 related to the cluster average mode shape are eliminated (steps 8 and 9 in Fig. 3.2).

After this grouping process, depending on the d_{lim} value, it is possible to find that different clusters are representing the very same physical mode. In other words, a single physical mode could have been split into two or more different clusters. Evidently, this phenomenon needs to be avoided. Therefore, the authors proposed a regrouping procedure in order to minimize this problem.

The regrouping routine (steps 10 to 14 in Fig. 3.2) is based on the confidence interval of the frequencies. Each single cluster has its frequency mean and its lower and upper limits of frequencies delimiting a certain confidence interval. All clusters are compared two by two. If the first cluster has its average frequency within the confidence limits of the second cluster, they are possibly representing the same physical mode. Therefore, they are selected to go through a further analysis: the MAC between among their mode estimates. If the MAC is higher than 0.9, the clusters are grouped into one single cluster.

Finally, the top n_m clusters (modes) with more elements are selected as being physical (step 15 in Fig. 3.2). Then, for the methodology output, each selected cluster has its mean modal parameters (natural frequency, damping ratio and mode shape) calculated (step 16 in Fig. 3.2). Such results correspond to one single realization and are saved to incorporate the dataset to be analyzed by the second clustering analysis (step 12 of Fig. 3.1).

After the algorithm flows out of all three loops (steps 13, 14 and 15 in Fig. 3.1), the third and last main routine is fed with several realizations (stabilization diagrams interpretations) to perform a second clustering analysis. This corresponds to step 16 in Fig. 3.1. Figure 3.3 presents the detailed flowchart of the second clustering analysis.

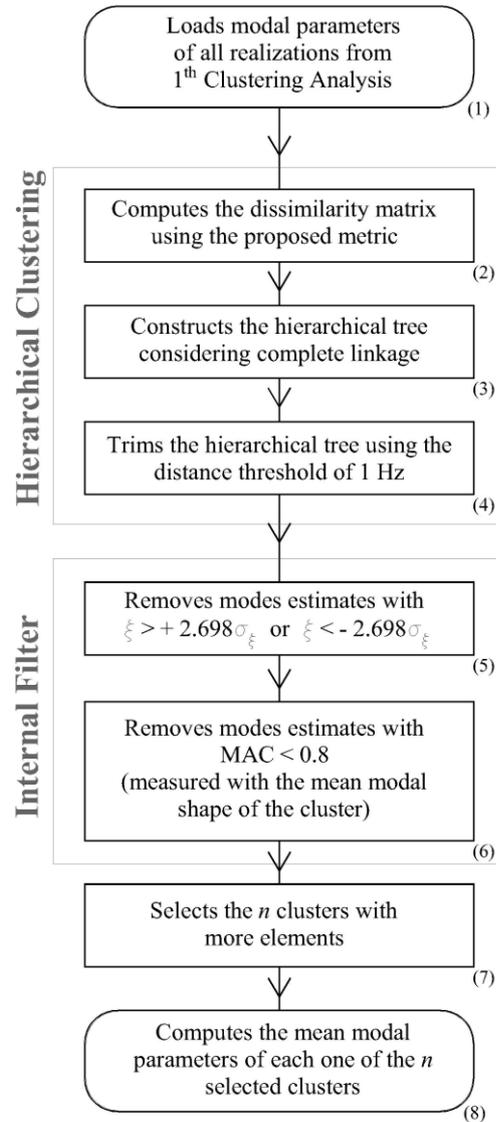


Fig. 3.3 Flowchart of the 2nd cluster analysis – step 16 of the proposed method shown in Fig. 3.1.

By comparing flowcharts of Figs. 3.2 and 3.3, it is clear that the second cluster analysis algorithm is very simple and similar to the first one. The only difference is that the former has no pre-filter or regrouping procedures.

The output from the second clustering analysis (step 17 and 18 in Fig. 3.1) is the final output of the proposed methodology, i.e. n identified modes (steps 16, 17 and 18 in Fig. 3.1).

3.4 Results

3.4.1 Numerically generated signals

To evaluate the efficiency of the proposed methodology handling with close frequency modes, a numerical experiment is conducted by simulating signals generated from a hypothetical structure. To this end, Eq. (5) describes a free damped vibration of a single degree-of-freedom dynamic system. In this equation, ω_D represents the damped natural frequency, which is approximately equal to the undamped natural frequency $\omega = 2\pi f$ (for small damping ratios n). Besides, $y(t)$ is the signal amplitude at a time instant t , ρ is an amplitude factor and θ is a phase constant.

$$y(t) = \sum_{i=1}^2 \rho_i \cos(\omega_{Di}t + \theta_i) e^{-\xi_i \omega_i t} \quad (5)$$

The aforementioned variables are chosen in a way so that two modes with close natural frequencies are obtained. Table 3.1 shows the values of the dynamic parameters of this hypothetical structure. The mode shapes are not considered in this application.

Table 3.1: Dynamic parameters of the hypothetical structure.

	f (Hz)	ζ (%)
Mode 1	15.10	2.00
Mode 2	15.20	2.50

Moreover, 10% of white noise is added to the simulated signals, according to Eq. (6). In this expression, y_{noise} is the signal with noise, y_{max} is the maximum absolute value of y , β is a factor containing the noise level (for instance, $0.1 = 10\%$):

$$y_{noise}(t) = y(t) + y_{max} \cdot \beta \cdot \chi(t) \quad (6)$$

where χ is a random signal (Gaussian white noise) generated by the computer with an amplitude varying from -1 to $+1$.

A 1000 Hz sampling rate is adopted, yielding a 10 s signal. Three signal windows (S1, S2 and S3) are extracted from the original signal in order to generate more realizations for the modal identification process. Both methods are requested to identify two modes ($n = 2$), while the other inputs are kept in default values (no pass-band filter and $m_{lim} = 0$).

Signal without noise

Figure 4 shows the time history of the generated signal (five channels) without noise as well as the three windows created.

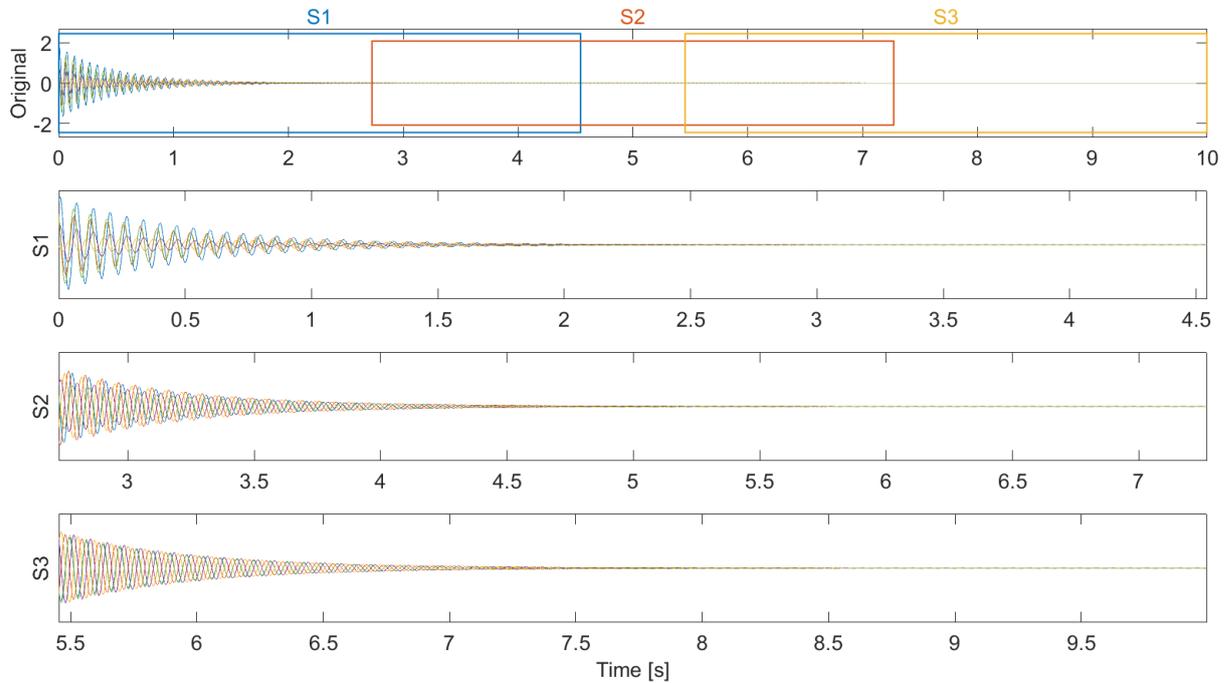


Fig. 3.4 Time history of the generated signal (five channels) without noise.

For each one window, SSI-DATA yielded three stabilization diagrams of maximum orders of 100, 140 and 180. For each stabilization diagram, the interpreter provided three different results varying the number of desired modes (1, 2 and 3). Consequently, the number of realizations was 27.

The results are presented by the scattered blue points in the Clustered Stabilization Diagram (CSD) seen in Fig. 3.5. The word “Clustered” stands for the reason that the scattered points are not in terms of mode estimates frequency, but in terms of frequency means of clustered modes estimates considered as being physical by that specific realization (interpretation of a stabilization diagram). In such diagrams, the red lines highlight the result (in terms of frequency) of the second clustering analysis, i.e. the final output of the proposed method.

Vertical axis corresponds to each realization, which receives a label according to its signal matrix (S), the number of desired modes (M) and the maximum order of stabilization diagram (R). For instance, S2-M2-R140 stands for: second signal window (S2), number of modes equal to two and a stabilization diagram with a maximum order of 140.

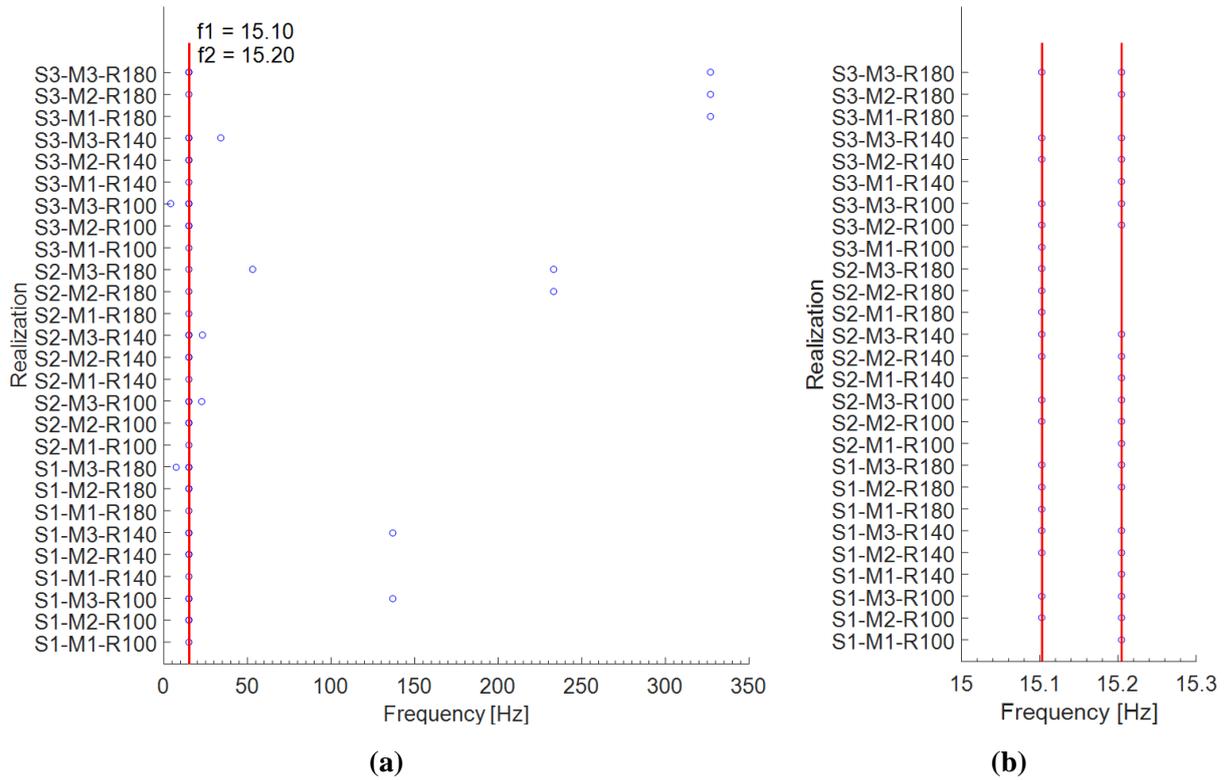


Fig. 3.5 Clustered stabilization diagram (a) and its zoom (b).

Table 3.2 summarizes the results for both reference (varying d_{lim}) and proposed methodology. One observes that the relative errors and the intra-cluster standard deviations (σ) are zero for both approaches. However, the reference methodology is affected by the d_{lim} parameter. For the cases when $d_{lim} = 0.05$ and $d_{lim} = 0.10$, the modal identification was unsuccessful.

Table 3.2: Results from the reference interpreter. All values were truncated.

Mode	Parameter	Reference Interpreter – R140				Proposed
		$d_{lim} = 0.001$	$d_{lim} = 0.01$	$d_{lim} = 0.05$	$d_{lim} = 0.10$	
01	$f(\text{Hz}) / \sigma(\text{Hz})$	15.10 / 0.000	15.10 / 0.000	15.10 / 0.000	missing	15.10 / 0.000
	$\xi(\%) / \sigma(\%)$	2.00 / 0.00	2.00 / 0.00	2.00 / 0.00	missing	2.00 / 0.00
02	$f(\text{Hz}) / \sigma(\text{Hz})$	15.20 / 0.000	15.20 / 0.000	missing	missing	15.20 / 0.000
	$\xi(\%) / \sigma(\%)$	2.50 / 0.00	2.50 / 0.00	missing	missing	2.50 / 0.00

Signal with noise

To further assess the robustness of the proposed approach, a new simulation considering 10% ($b = 0.1$) of noise added to the original signal is presented. Figure 3.6 depicts the noisy signal.

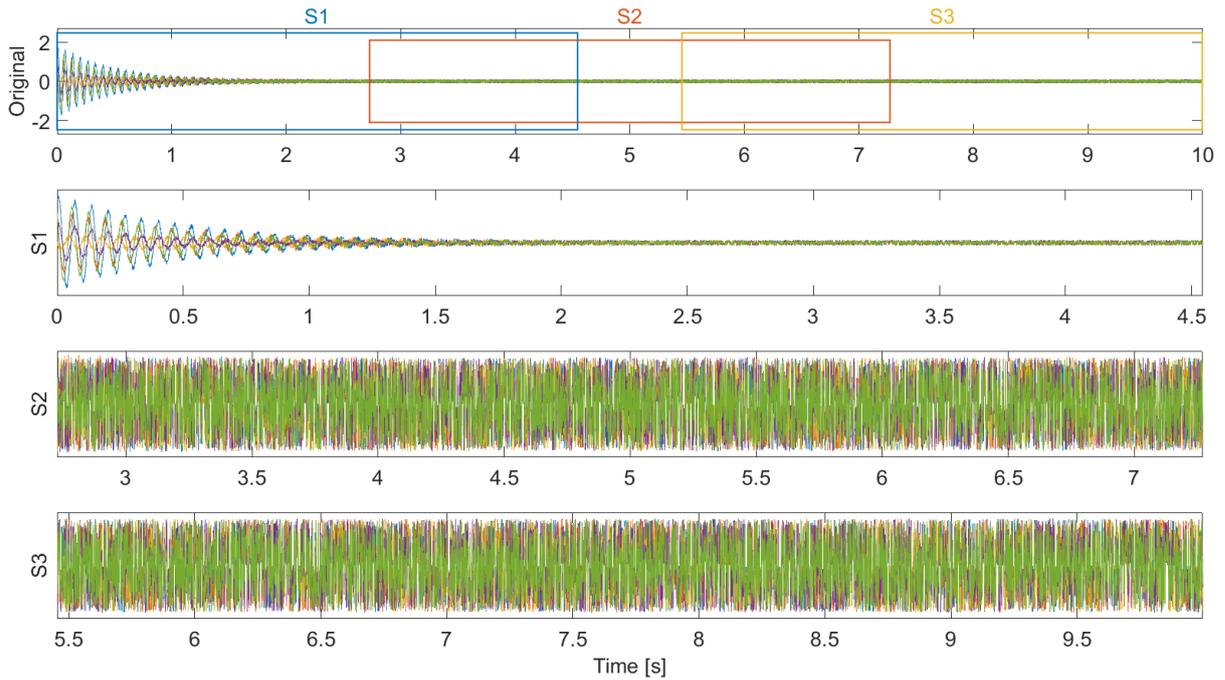


Fig. 3.6 Time history of the generated signal (five channels) with noise.

One observes that both second and third windows (S2 and S3) are completely affected by the presence of noise. This can be confirmed by computing noise-to-signal ratios, as shown in Table 3.3.

Table 3.3: Noise/signal ratio considering RMS values.

	Ch. 1	Ch. 2	Ch. 3	Ch. 4	Ch. 5
Original	0.2433	0.3499	0.7339	0.5915	0.3148
S1	0.1668	0.2441	0.5880	0.4431	0.2182
S2	0.9955	1.0059	0.9934	0.9994	0.9996
S3	0.9913	1.0011	0.9993	1.0005	1.0047

As the ratio gets close to one, the more corrupted with noise the signal is. Moreover, it can be noticed that the noise levels in the first window (S1) are also high, with 35% and 73% of noise-to-signal ratio for channels 2 and 3, respectively.

To highlight the difficulty of such an analysis, Fig. 3.7 depicts an FFT (Fast Fourier Transform) plot of the signal using 2^{16} points to get a fine frequency resolution (0.015 Hz). Clearly, it is impossible to identify these two simulated close-spaced modes.

Despite this challenging scenario, the proposed method performed the modal identification with high precision. Fig. 3.8 shows the CSD. It is interesting to observe that

windows S2 and S3 did not provide useful information for the identification of both two modes. By inspecting Fig. 3.8, one observes high-scattered blue points corresponding to realizations of S2 and S3. Nevertheless, the novel second clustering analysis proposed in this paper was able remove those spurious frequencies.

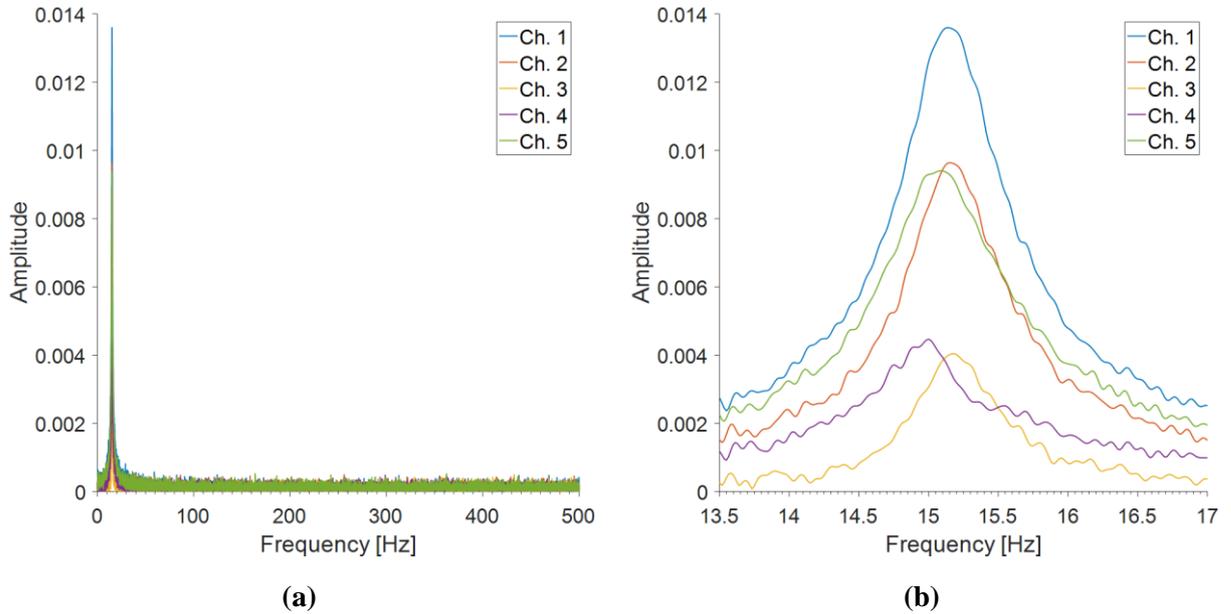


Fig. 3.7 FFT of each channel (a) and its zoom (b).

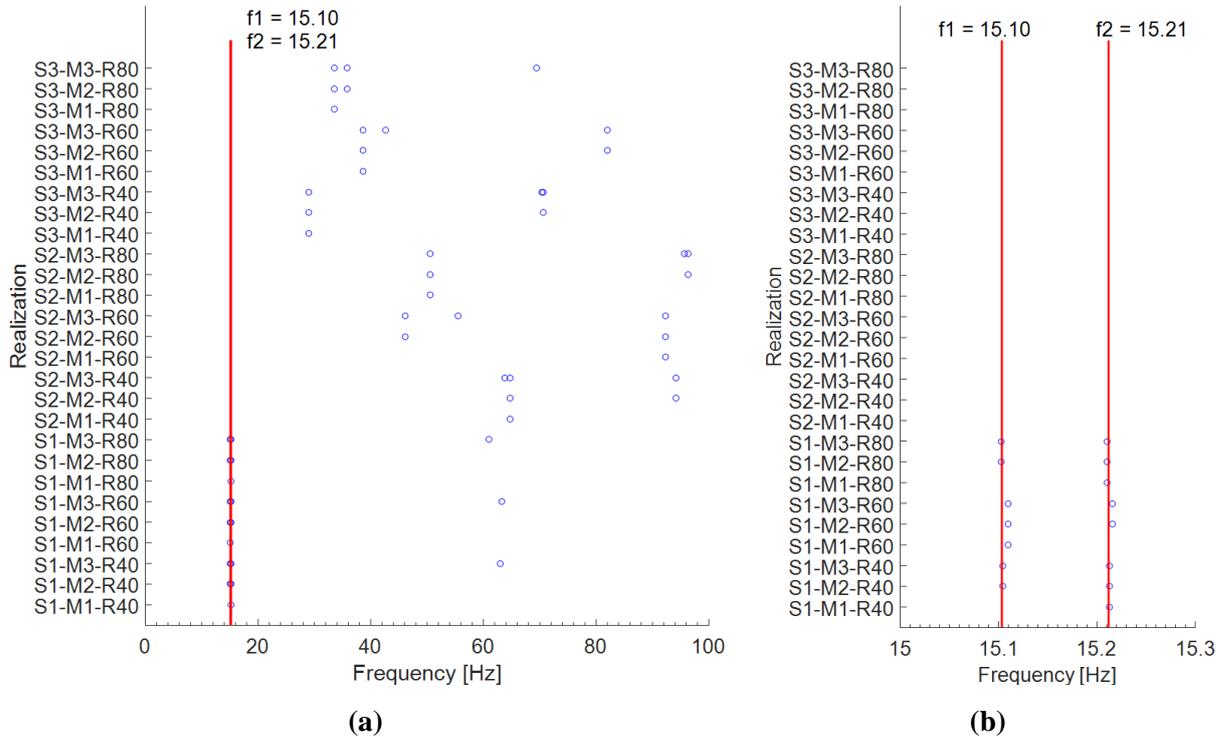


Fig. 3.8 Clustered stabilization diagram (a) and its zoom (b).

Table 3.4: Results from the reference interpreter. All values were truncated.

Mode	Parameter	Reference Interpreter – R140				Proposed
		$d_{lim} = 0.001$	$d_{lim} = 0.01$	$d_{lim} = 0.05$	$d_{lim} = 0.10$	
01	$f(\text{Hz}) / \sigma(\text{Hz})$	15.12 / 0.002				15.103 / 0.000
	$\xi(\%) / \sigma(\%)$	1.91 / 0.01	missing	missing	missing	1.96 / 0.06
02	$f(\text{Hz}) / \sigma(\text{Hz})$	15.17 / 0.005			15.17 / 0.009	15.212 / 0.001
	$\xi(\%) / \sigma(\%)$	2.90 / 0.02	missing	missing	2.90 / 0.03	2.52 / 0.19

Table 3.4 summarizes the results for both reference (varying d_{lim}) and proposed methodology. The relative errors and the intra-cluster standard deviations (σ) are close to zero for both approaches. However, the reference methodology is more severely affected by the d_{lim} parameter. For the cases when $d_{lim} = 0.01$, $d_{lim} = 0.05$ and $d_{lim} = 0.10$, the modal identification did not yield any results. Moreover, the proposed methodology achieved better results in terms of frequency and damping ratio estimates.

3.4.2 Laboratory experiment

This section presents an experimental test conducted in laboratory on a simply supported steel beam. This beam is 1.46 m long with rectangular cross-section (76.2×8.0 mm) and is instrumented with six piezoelectric accelerometers (PCB, 336C31). Data acquisition is carried out using Lynx ADS2002 equipment, which essentially is a conditioning/amplifier regulating system. This study considered random vibration tests (using a shaker). The random excitation is applied throughout the duration of the tests. A 25 s signal, sampled at 4000 Hz, is shown in Fig. 3.9.

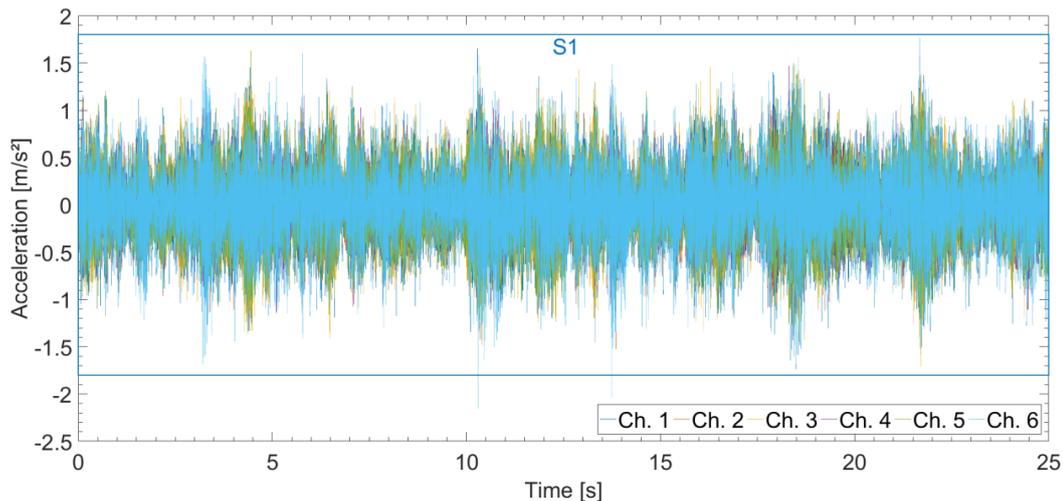


Fig. 3.9 Signal from instrumented steel beam.

In this study, the reference methodology has the d_{lim} value set to 0.01, while the proposed approach has the m_{lim} set to 0.9. For both methods, the number of requested modes was set to five.

Figure 3.10 highlights the different performances reached by both techniques. The stabilization diagram provided by the former (Fig. 3.10a) is full of spurious modes estimates, while the latter shows that its pre-filters worked adequately (Fig. 3.10b). In those stabilization diagrams, the red colored points correspond to the physical modes. Additionally, the number of elements in each cluster, presented by the stem plots depicted in Fig. 3.10c and d show, respectively, that the reference interpreter not only missed the first mode, but also gathered several spurious modes in its fifth cluster (frequency close to 200 Hz). Conversely, due to its different metric, filters and regrouping technique, the enhanced interpreter was indeed able to correctly identify all five modes (the red headed stems stand for the five selected clusters related to physical modes).

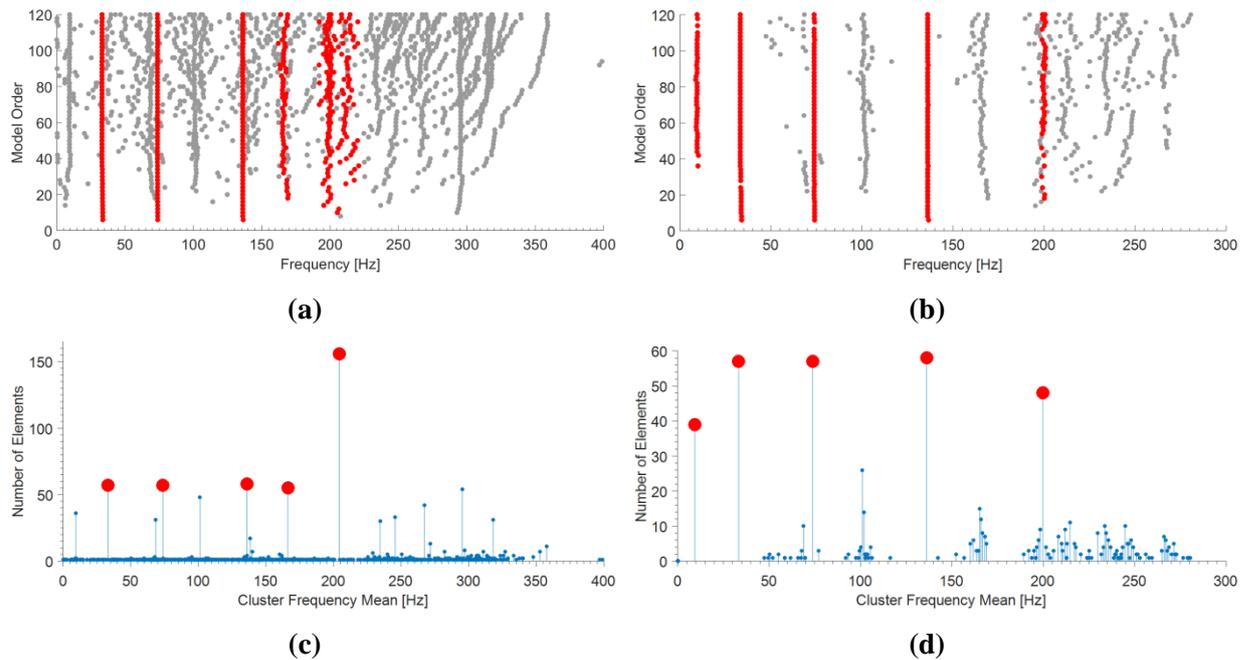


Fig. 3.10 Modes estimates interpretation by reference interpreter (a, c) and enhanced interpreter (b, d) – realization S1M5R120.

The results obtained by the enhanced interpreter shown in Fig. 3.10 correspond to a single realization of the proposed method (highlighted by the green text in the CSD displayed in Fig. 3.11).

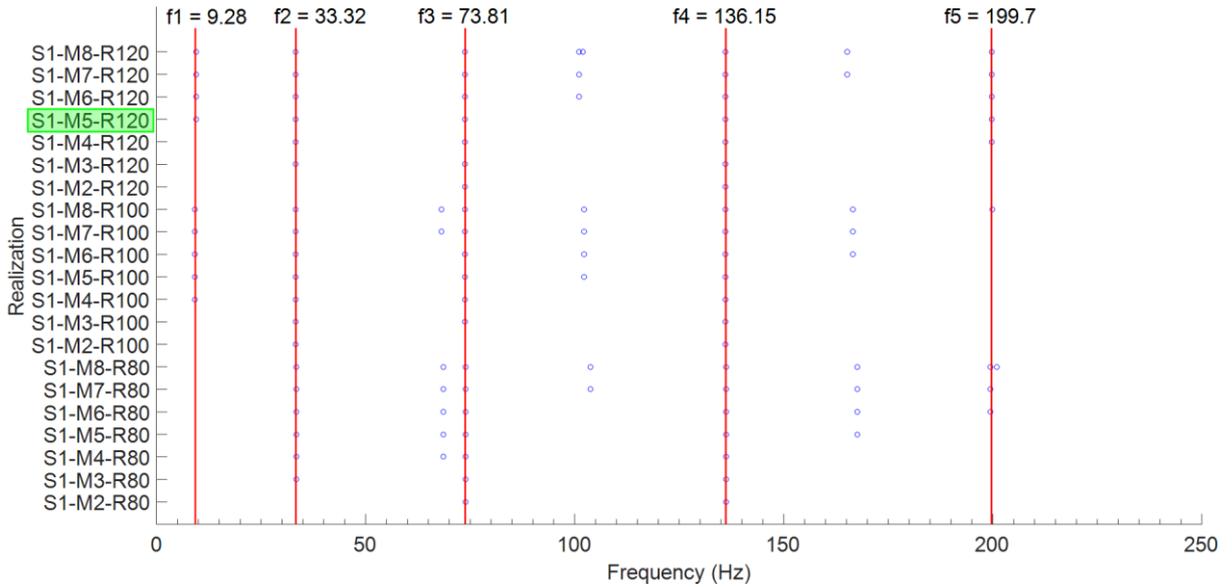


Fig. 3.11 Clustered stabilization diagram of the analyzed signal.

The mode shapes identified by the proposed approach are presented in Fig. 3.12.

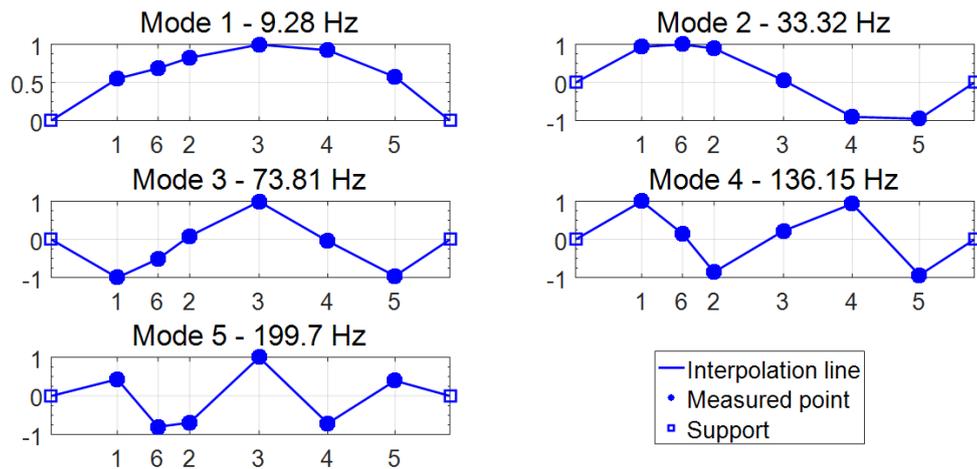


Fig. 3.12 Modal shapes obtained by the proposed method.

Finally, Table 3.5 summarizes the results obtained by the proposed methodology and by three attempts for the reference approach (varying d_{lim}). In all cases, the latter was not able to identify the first mode. Besides, it presented high values of intra-cluster standard deviations for the fifth mode. This fact is mainly due to the metric used in the reference interpreter. As noted in Eq. (3), the metric measures distances between modes estimates in a relative manner in terms of frequency, yielding a biased behavior with low-frequency modes distances and with high-frequency modes distances. Hence, clusters gathering high-frequency modes estimates tend to gather too much elements ('gathering phenomenon') erroneously. Conversely, clusters with low-frequency modes estimates tend to split into other clusters

(‘splitting phenomenon’). As for the proposed methodology, all standard deviations were very low due to the enhanced diagram interpreter, which considered an innovative metric, additional filters and a regrouping technique.

Table 3.5: Results of the automatic modal identification for the five vertical modes.

Mode	Parameter	Reference Interpreter (R120)			Proposed
		$d_{lim} = 0,01$	$d_{lim} = 0,005$	$d_{lim} = 0,001$	
01	f (Hz) / σ (Hz)	missing	missing	missing	9.28 / 0.120
	ξ (%) / σ (%)				7.13 / 1.05
02	f (Hz) / σ (Hz)	33,28/0,160	33,26/0,139	33,20/0,035	33.32 / 0.047
	ξ (%) / σ (%)	3,23/0,19	3,20/0,14	3,17/0,12	3.34 / 0.06
03	f (Hz) / σ (Hz)	73,79/0,049	73,78/0,042	73,78/0,042	73.81 / 0.025
	ξ (%) / σ (%)	0,62/0,05	0,62/0,05	0,62/0,05	0.62 / 0.01
04	f (Hz) / σ (Hz)	136,13/0,027	136,13/0,027	136,13/0,027	136.15 / 0.015
	ξ (%) / σ (%)	0,37/0,05	0,37/0,05	0,37/0,05	0.31 / 0.03
05	f (Hz) / σ (Hz)	204,00/6,493	199,38/1,882	198,75/0,137	199.70 / 0.183
	ξ (%) / σ (%)	2,24/0,94	2,27/0,86	1,71/0,70	2.42 / 0.15

3.4.3 Railway viaduct

Located on the southeast high speed track in France, at the kilometeric point 075 + 317, the tested viaduct (Fig. 3.13) crosses the secondary D939 road, between the towns of Sens and Soucy. Since it was built in the eighties, the increases in the operational speed of high-speed trains (TGV) have moved the excitation frequency closer to the first natural frequency of the structure. Hence, in order to prevent resonance-induced damages, the viaduct passed by a structural intervention, which consisted in tightening a system of rods near the bearings, by a torque wrench. Thus, additional stiffness and, therefore, an increase in the natural frequencies would be provided. Figure 3.14 shows the instrumentation plan of the structure.

During approximately 1.6 s, the eight installed accelerometers registered, at 4096 Hz, the signal matrix S1 shown in Fig. 3.15. It corresponds to a forced vibration period while the train was passing over the structure.

Again, the reference interpreter and the proposed methodology were applied to the signal. In this study, the number of desired modes is set to four. However, in a conservative

way, both methods were asked to provide six modes. Additionally, a maximum frequency of 50 Hz was set, due to prior knowledge of the structure’s dynamic behavior.



Fig. 3.13 PK 075 + 317 Viaduct on the Paris-Lyon high speed train track.

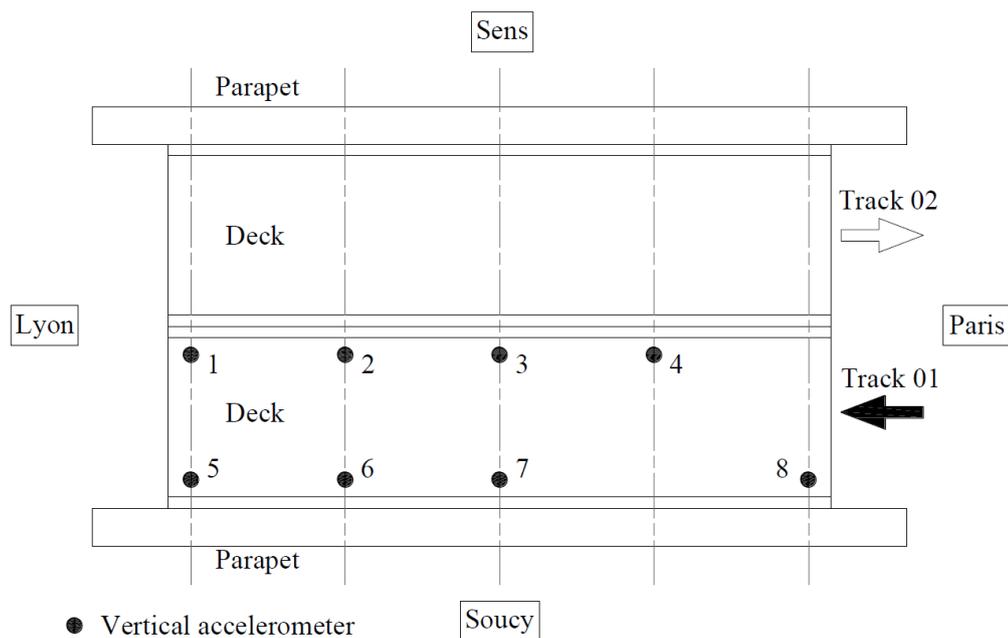


Fig. 3.14 Plane view of the viaduct pointing sensors location.

Figure 3.16 depicts the results obtained by the reference interpreter. It is clear that this method was not able to identify the second mode (8 Hz). Three attempts were made varying d_{lim} : 0.01, 0.05 and 0.10. One can note the “gathering” phenomenon happening again as the distance threshold increases, from left to right.

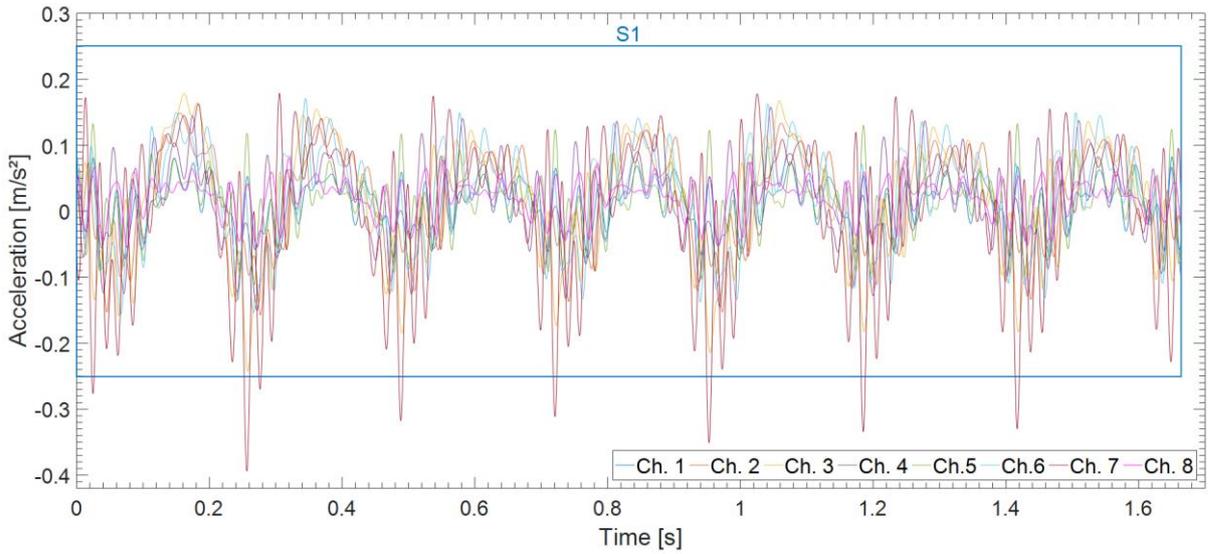


Fig. 3.15 Signal acquired during the transit of the train.

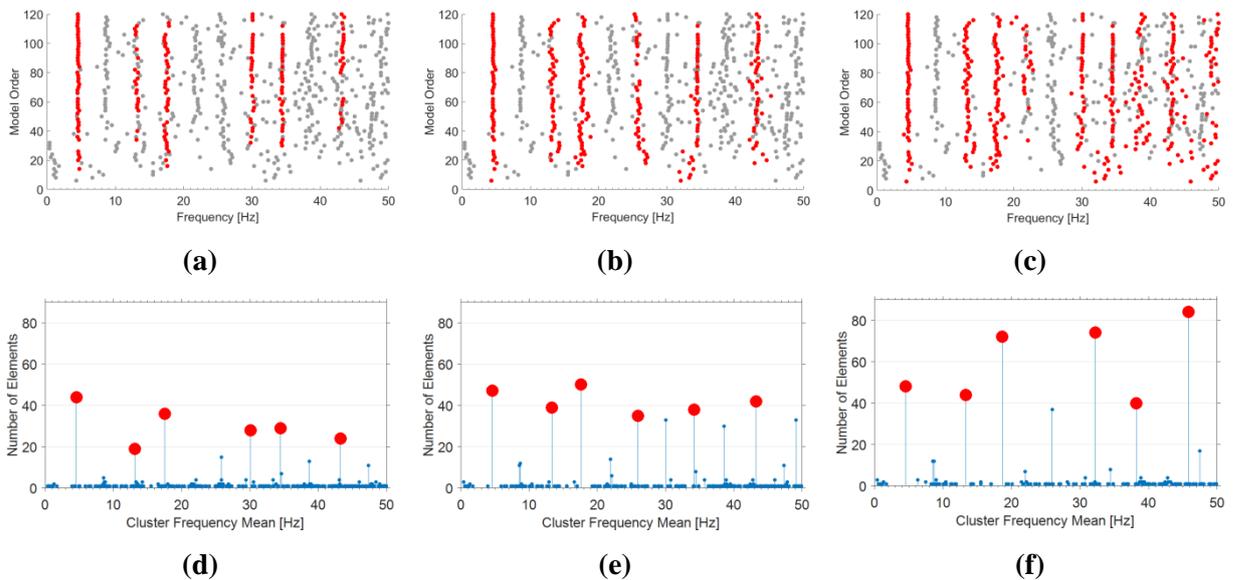


Fig. 3.16 Modes estimates interpretation by reference interpreter, with d_{im} varying: 0.01 (a, d), 0.05 (b, e) and 0.10 (c, f).

For the proposed method, however, all four modes were identified correctly. In Fig. 3.17a, b and c, three realizations are shown in detail. All 21 realizations and the final result can be checked in the CSD presented in Fig. 3.17d.

Figure 3.18 shows the modal shapes evaluated by the proposed methodology. The measured modal amplitudes are marked as dots, whereas interpolated points or supports are represented as hollow squares. The blue line represents modal amplitudes measured by sensors 1, 2, 3 and 4 whereas the red line stands for accelerometers 5, 6, 7 and 8 (see Fig. 3.14).

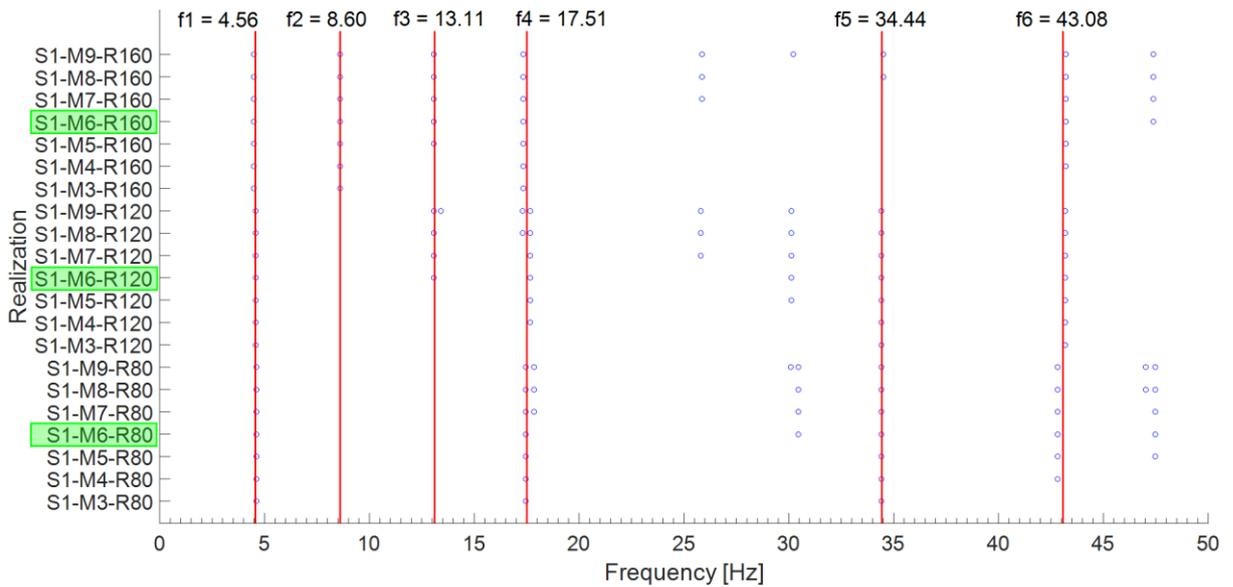
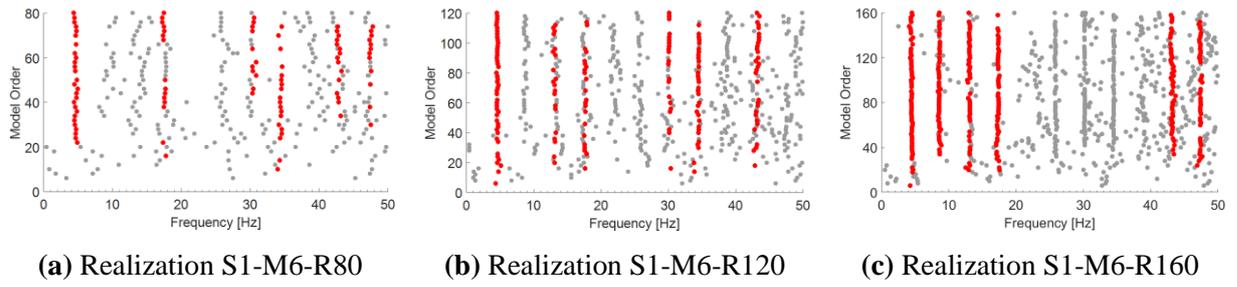


Fig. 3.17 Different realizations results (a, b, c) and CSD of analyzed signal (d).

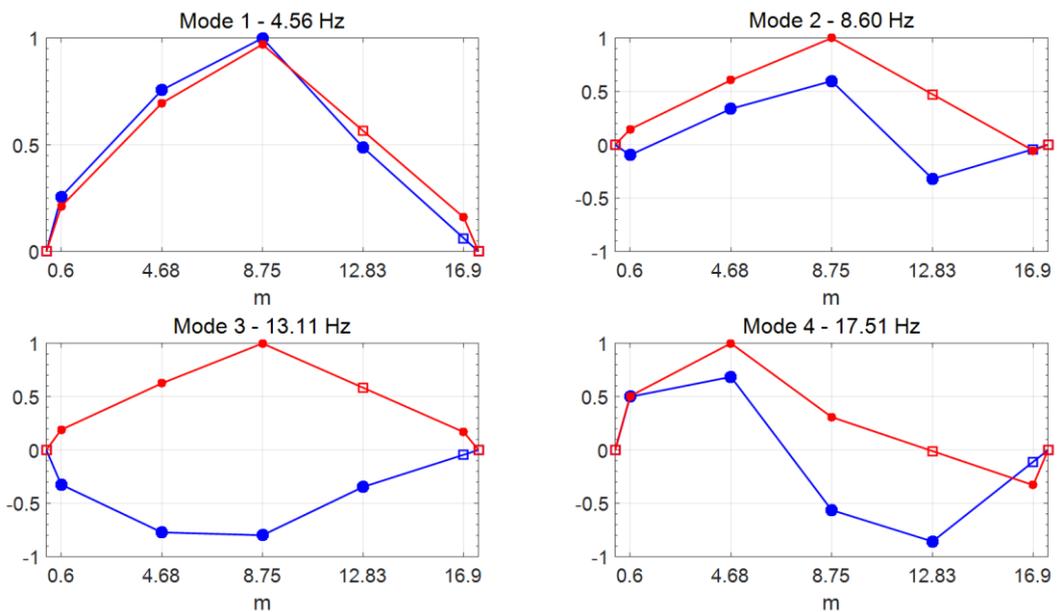


Fig. 3.18 Four identified modal shapes: flexional, flexional-torsional, torsional, and flexional torsional, respectively.

Finally, Table 3.6 gathers the results for the modal identification. One can note that the reference interpreter was unable to identify the second mode regardless of the value of d_{lim} . Once again, the proposed method has proved to be more robust and accurate when compared to the reference approach.

Table 3.6: Results of the automatic modal identification.

Mode	Parameter	Reference Interpreter R120			Proposed
		$d_{lim} = 0.01$	$d_{lim} = 0.05$	$d_{lim} = 0.10$	
01	$f(\text{Hz}) / \sigma (\text{Hz})$	4.579 / 0.085	4.586 / 0.119	4.572 / 0.151	4.563 / 0.056
	$\zeta (\%) / \sigma (\%)$	2.23 / 1.07	2.34 / 1.20	2.36 / 1.19	3.59 / 1.59
	MAC_{min}	0.9981	0.9937	0.9937	0.9999
02	$f(\text{Hz}) / \sigma (\text{Hz})$				8.603 / 0.000
	$\zeta (\%) / \sigma (\%)$	missing	missing	missing	1.19 / 0.00
	MAC_{min}				1.0000
03	$f(\text{Hz}) / \sigma (\text{Hz})$	13.156 / 0.168	13.344 / 0.384	13.311 / 0.392	13.108 / 0.108
	$\zeta (\%) / \sigma (\%)$	4.25 / 0.99	3.58 / 1.82	3.64 / 1.84	2.78 / 1.15
	MAC_{min}	0.9770	0.8787	0.5808	0.9937
04	$f(\text{Hz}) / \sigma (\text{Hz})$	17.520 / 0.208	17.595 / 0.387	18.653 / 1.948	17.511 / 0.178
	$\zeta (\%) / \sigma (\%)$	1.98 / 0.99	2.41 / 1.78	3.30 / 2.72	1.72 / 0.81
	MAC_{min}	0.9131	0.7323	0.5990	0.9778

3.5 Conclusions

This paper presented a novel methodology that uses a parametric time-domain modal identification algorithm and hierarchical clustering techniques to perform automated modal identification. The developed routines used SSI-DATA algorithm to generate modes estimates for several model orders, followed by two steps of clustering analysis. The first one is the clustering analysis of modes estimates coming from a single stabilization diagram, outputting a given number of modes with a high chance of being physical. After many different stabilization diagrams are interpreted, the second step of clustering analysis is carried out to finally select the modes definitely considered physical.

Modes with very closely-spaced frequencies can be correctly detected by the method. The first application of this paper treated numerically generated signals and exposed that, even with the presence of high noise level, the proposed routines were able to successfully identify the natural frequencies.

Experimental applications presented in Sections 3.4.2 and 3.4.3 showed that the enhanced diagram interpreter provided a more robust tool to eliminate spurious modes. In both applications, all physical modes were correctly identified. The laboratory experiment provided a structure with natural frequencies ranging from 9 to 200 Hz. In contrast to what happened with the reference method, it was shown that the proposed method is perfectly suitable to handle either low or high values of natural frequencies. Besides, the railway viaduct test allowed concluding that the developed routines are reliable to be applied to real dynamic tests.

The proposed methodology represents, therefore, a useful tool to modal identification of civil engineering structures subjected to ambient vibrations (Operational Modal Analysis). In this context, the present work offers a suitable method capable of performing, satisfactorily, automated interpretation of modal estimates, which can be obtained by any parametric time-domain identification algorithm.

Acknowledgements

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for the financial support, COPPE/UFRJ and IFSTTAR (former LCPC – Laboratoire Central des Ponts et Chaussées) and SNCF (Société Nationale des Chemins de fer Français) for the data used in this paper (project 01V0527 RGCU “Evaluation dynamique des ponts”).

References

- [1] J.P. Santos, *Smart structural health monitoring techniques for novelty identification in civil engineering structures* PhD thesis by the University of Lisbon, Superior Technical Institute, Portugal, **2014**.
- [2] V. Alves, A. Cury, N. Roitman, C. Magluta, C. Cremona, Novelty detection for SHM using raw acceleration measurements, *Struct. Control Health Monit.* 22 (**2015**) 1193–1207.
- [3] R. Battista, M. Pfeil, Reduction of vortex-induced oscillations of Rio-Niterói Bridge by dynamic control devices, *J. Wind Eng. Ind. Aerodyn.* 84 (3) (**2000**) 273–288.

- [4] Y. Gautier, O. Moretti, C. Cremona, Experimental modal analysis of the Millau Bridge, in: *Proceedings of Experimental Vibration Analysis for Civil Engineering Structures*, **2005**.
- [5] F. Magalhães, A. Cunha, E. Caetano, Online automatic identification of the modal parameters of a long span arch bridge, *Mech. Syst. Signal Process.* 23 (2) (**2009**) 316–329.
- [6] E. Reynders, J. Houbrechts, G. De Roeck, Fully automated (operational) modal analysis, *Mech. Syst. Signal Process.* 29 (**2012**) 228–250.
- [7] F. Ubertini, C. Gentile, A.L. Materazzi, Automated modal identification in operational conditions and its application to bridges, *Eng. Struct.* 46 (**2013**) 264–278.
- [8] A. Cabboi, F. Magalhães, C. Gentile, A. Cunha, Automated modal identification and tracking: application to an iron arch bridge, *Struct. Control Health Monit.* 24 (1) (**2017**).
- [9] B. Peeters, G. De Roeck, Reference-based stochastic subspace identification for output-only modal analysis, *Mech. Syst. Signal Process.* 13 (6) (**1999**) 855–878.
- [10] P. Van Overschee, B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation and Applications*, Kluwer Academic Publishers, **1996**.
- [11] D.J. Ewins, *Modal Testing: Theory, Practice and Application*, Research Studies Press Ltd., USA, **2000**.
- [12] J.N. Juang, R.S. Pappa, An eigensystem realization algorithm for modal parameter-identification and model reduction, *J. Guid. Control Dynam.* 8 (5) (**1985**) 620–627.
- [13] R.S. Pappa, K.B. Elliot, A. Schenk, Consistent-mode indicator for the eigensystem realization algorithm, *J. Guid. Control Dynam.* 16 (5) (**1993**) 852–858.
- [14] P. Vacher, B. Jacquier, A. Bucharles, Extensions of the MAC criterion to complex modes, in: *International Conference on Noise and Vibration Engineering, ISMA 2010*, 2010.
- [15] G.J. Yun, S.G. Lee, S. Shang, An improved mode accuracy indicator for eigensystem realization analysis (ERA) techniques, *KSCE J. Civil Eng.* 16 (3) (**2012**) 377–387.
- [16] R.A. Cardoso, *Development of methodology for automatic modal identification of structures*, MSc dissertation by the Federal University of Ouro Preto, Brazil, **2015**.

PART III

RAW DATA PROCESSING FOR DAMAGE DETECTION

Chapter 4

Paper #3 – Unsupervised real-time SHM technique based on novelty indexes

Both papers presented in the previous chapters provide useful tools for an SHM strategy based on the tracking of modal parameters. Another way of extracting damage-related information from the structural behavior is utilizing direct raw data measurements. Therefore, a new technique was developed. A Symbolic Data Analysis is conducted over collected raw time signals to create an original compact, yet comprehensive representation. Afterward, unsupervised statistical learning algorithms perform pattern recognition on the symbolic representations, with two different damage-related scalar indexes being computed. Such indexes are evaluated upon each new symbolic object acquisition, virtually in real-time. The proposed technique is explained and validated within a paper accepted for publication in the *Journal Structural Control and Health Monitoring*.

Paper Information

Status	Accepted in 2018
Publisher	Structural Control and Health Monitoring ISSN 1545-2263; Qualis A1; Impact Factor (2018): 3.622
Authors	Cardoso R, Cury A, Barbosa F., Gentile, C.
Title	Unsupervised real-time SHM technique based on novelty indexes.

Abstract

Structural Health Monitoring programs play an important role in the field of civil engineering, especially for assessing safety conditions involving large structures such as, viaducts, bridges, tall buildings, towers and old historical buildings. Essentially, a SHM process needs to be based on a trustful strategy for detecting structural novelties or abnormal behaviors. Usually, such a strategy is complemented with human inspection and structural instrumentation routines, where the latter requires proper hardware equipment and software tools. Recently, many advances were achieved regarding the hardware resources, such as wireless communication, remotely configurable sensors and other data management devices. On the other hand, the software counterpart still is in its early development. In fact, many researches are under development to fill this gap. In this context, this paper presents a novel online SHM identification method suitable to unsupervised real-time detection of structural abnormal behaviors. The proposed methodology includes the use of an original representation of raw dynamic signals, i.e., *in situ* measured accelerations. To assess the proposed approach, two experimental applications are studied: a railway viaduct, PK 075+317 in France, and an old masonry tower, in Italy. The results suggest that the proposed detection indexes are suitable to a wide range of SHM applications.

4.1 Introduction

Large structures under continuous operation, such as viaducts, bridges, tall buildings, historical towers, among others, need to have their structural condition monitored permanently. Usually, this task is performed through visual inspection and/or other manual analyses. However, to enhance the structure's safety assessment, it is necessary to develop a continuous monitoring program based on computations capable of looking after the structure 24/7, uninterruptedly. The main purpose of such an approach is to automatically detect damage and/or other structural novelties immediately after their occurrence, ideally at their initial stages.

Recent technological advances allowed the production of reliable, accurate and efficient monitoring equipment that can be remotely configured and operated. That fact encouraged the installation of dynamic monitoring systems in a considerable number of structures abroad. Therefore, especially over the last twenty years, much effort has been made

in order to have computational tools able to detect abnormal structural behaviors from continuously acquired datasets.

In general, the literature presents two different damage identification approaches: inverse and forward. The inverse SHM identification methods are based on model updating or system identification techniques. They consist in finding parametric numerical or analytical models that best fit the structural responses. Due to their computational complexity – this kind of problem is typically nonlinear and non-unique [1] and demands choosing which structural parameters must be analyzed [2] – it is often necessary to use optimization techniques or to have constant human decision to state if there is damage or not. Besides, the literature points out that this type of method requires the use of a significant number of sensors [3,4] or need a considerable amount of identified structural mode shapes [5].

Conversely, the forward SHM approach does not require the development of numerical or analytical models to fit the acquired data [2]. Instead, these techniques aim at extracting sensitive information from the measured time-series, using statistical learning methods [6]. Compared to the inverse approach, forward SHM methods are more flexible and less computationally expensive, since they do not require the development of specific models, which must be usually defined for each structural system [7-9]. Besides their computational simplicity, this kind of technique is more capable of handling structures subjected to operational and environmental effects, since data-based algorithms can reproduce these effects more easily when compared to numerical or analytical models. Due to these characteristics, one can conclude that the forward SHM approach is more suitable for carrying out real-time SHM in large-scale structures [10].

Forward SHM techniques are frequently addressed with four distinct steps: [6,11]

1. Operational evaluation;
2. Data acquisition and normalization;
3. Feature extraction and data fusion;
4. Feature classification.

The first and second steps require the use of structural and sensorial knowledge to define the monitoring plan, including the quantities to be measured and the architecture of the data acquisition system. The third step comprises the application of signal processing techniques to extract representative features from signals and to perform data fusion. Finally,

the fourth step aims at applying statistical learning procedures to classify the features as related to known or novel structural condition.

The statistical learning procedures used in the forward SHM techniques can be classified as supervised or unsupervised. Supervised learning is applied to cases where a prior dataset exist, known as baseline, which is assumed to be the structure's healthy state. The drawback here is evident, since data acquisition during the undamaged baseline period is not always possible. The first reason is that most of SHM applications are based on aged structures, which may have already endured any structural modification or even damage. The second reason is that newly built structures may not present a long enough baseline to allow the establishment of a healthy reference state. Conversely, unsupervised learning methods are more suitable to SHM since they do not need prior knowledge of long periods of the structure's response nor target categories related to an assumed healthy reference state.

Despite its rather impractical aspect, almost all forward SHM methods reported in the literature are based on supervised approaches [12-16]. Even the works addressing clustering methods, which are very well suited for unsupervised methods, report a supervised strategy of pre-defining cluster partitions to describe one or more known structural behaviors and, subsequently, compare them with new ones [17-19]. However, some recently published works [20,21] propose a time-window procedure to extract sensitive features based on a fully unsupervised approach that uses clustering methods for performing feature classification. Such a methodology allows (i) detecting structural novelties without resorting to previous knowledge; (ii) performing real-time identification relying on the detection of novelties upon each new acquired data.

In that way, aiming to turn the feature extraction richer, a novel symbolic data object is proposed in this paper, along with a new strategy for defining indexes able to point out whether there has been any structural change (novelty detection). Thus, the proposed approach is applied to raw measured dynamic signals directly. This paper is organized as follows: sections 2 and 3 detail the methodology, presenting the proposed feature extraction (step 3) and classification (step 4) techniques, respectively. To assess the methodology, section 4 presents the results from two distinct applications of SHM: a railway viaduct, PK 075+317 in France, and an old masonry tower, in Italy. Finally, section 5 presents a brief discussion related to the proposed approach.

4.2 Feature Extraction and Data Fusion

4.2.1 Literature review

Since raw data (acceleration time-series) are usually unable to provide a clear evidence of structural novelty directly, feature extraction is thus required. The feature extraction procedure can be understood as a transformation of raw data. The idea is to extract its embedded information, resulting in a more compact yet intelligible representation, which is still sensitive to structural novelties. Such a procedure yields sets of multidimensional (or multivariate) data named feature vectors. In general, high dimensionalities inhibit a proper statistical characterization of such features, since long vectors imply higher computational costs. Thus, it is highly desirable that feature vectors are as small as possible (both in length and in dimensionality) but, at the same time, they must preserve all the intrinsic information of the original data about the structure's condition. Such "compressing" process is named data fusion. Therefore, when referring to the process of feature extraction, it should be kept in mind that there is also an inherent data fusion procedure.

The feature extraction procedure is based on the premise that a structural novelty can be revealed by at least one of the five basic data descriptors [13,22,23]:

1. Vibration (raw data, natural frequencies, spectra, etc.);
2. Expected value (mean, median, mode, etc.);
3. Variability (variance, standard deviation, covariance or correlation);
4. Symmetry (skewness);
5. Flatness (kurtosis).

The most commonly used features are those obtained through a dynamic modal analysis, i.e., natural frequencies and/or mode shapes (damping ratios are rarely used due to the imprecision in their identification). Nowadays, dynamic parameters can be accurately identified without the need of human intervention in a fully automatic approach [24]. Nevertheless, even though modal parameters are features of great physical significance and provide a huge reduction in the data volume (multiple time series transformed into just a few values), several studies show that these features are insensitive to small changes in structural behaviors [20,25-29].

Features built over data statistical measures, such as mean values, variance, skewness and kurtosis were also used as damage indicators in the literature. For instance, the works of Torres [30] and Finotti [31] presented reasonable results when those features were used as inputs to unsupervised and supervised classification methods, respectively.

In the literature, some features that reflect only the variability (third data descriptor) were also used, such as those based on Mahalanobis distance [32] and those based on Principal Component Analysis [33]. Another feature reported in the literature [20] is the autocorrelation vector, known by the acronym ACF (Auto-Correlation Function). These objects take into account only the variations in the expected value and in the symmetry (2nd and 4th descriptors) of the raw data distribution.

On the other hand, some authors [20,34] chose to employ Symbolic Data Analysis. A Symbolic Data Object (SDO) is able to perform a great fusion of information present in massive series of classical dynamic data (raw signals), transforming them into a more compact, yet rich representation. Additionally, symbolic objects have the capacity to represent multivariate (multidimensional) data, such as those collected from multiple instrumentation channels in a structural monitoring program. SDO's are statistical objects with great generalization power, such as interquartile ranges and categorical histograms [35]. While the former is more compact, but less representative, the latter is more representative, but less compact.

4.2.2 Proposed symbolic data object

The interquartile range (IQR) has proven to be adequate for forward real-time SHM in an unsupervised framework [21]. However, such a symbolic object includes just one of the non-vibrational data descriptors presented above: variability. To mitigate this limitation, the present work proposes a novel SDO, comprising not only the 1st and 3rd quartiles that define IQR, but also the median (2nd quartile) of the raw data distribution. For this reason, the new symbolic object was named IQRM, where “M” stands for median. This new representation provides a richer, yet compact way to characterize raw dynamic signals.

The IQRM object consists in a hyper-rectangle with a point within. While the facets represent the interquartile range (denoted as IQR) of each property (i.e., acquisition channel), the point locates the median (denoted as M) of the distribution. For the case of just two

properties (two acquisition channels) its geometry collapses into a rectangle with a single point, as shown in Fig. 4.1, where $Q_{k,r}$ is the k -th quartile of the r -th property.

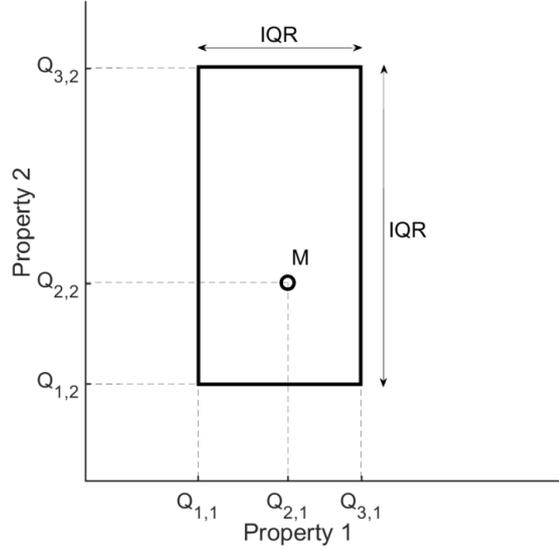


Fig. 4.1 Proposed symbolic data object, named IQRM.

4.2.3 Proposed symbolic distance metric

Aiming to extract a unidimensional feature sensitive to structural novelties, one proposes a new distance metric for unsupervised statistical learning.

Being $Q_{k,r}$ the k -th quartile of property $r=1, 2, 3, \dots, p$, where p is the total number of properties of the raw data, the vectors containing the quartiles information of all properties of object i can be written as shown in Eq. (1):

$$\mathbf{L}_i = \begin{bmatrix} Q_{1,1} \\ Q_{1,2} \\ Q_{1,3} \\ \vdots \\ Q_{1,p} \end{bmatrix}, \mathbf{M}_i = \begin{bmatrix} Q_{2,1} \\ Q_{2,2} \\ Q_{2,3} \\ \vdots \\ Q_{2,p} \end{bmatrix} \text{ and } \mathbf{U}_i = \begin{bmatrix} Q_{3,1} \\ Q_{3,2} \\ Q_{3,3} \\ \vdots \\ Q_{3,p} \end{bmatrix} \quad (1)$$

The p -dimensional vectors \mathbf{L}_i , \mathbf{M}_i and \mathbf{U}_i gather the 1st, 2nd and 3rd quartiles, respectively, of each one of the p properties. Hence, one should notice that such a triplet of vectors defines an IQRM object completely.

Now, as one considers a second object, with index j , it is possible to compute the norm of the vectors difference, as shown in Eq. (2), where Δ is the so-called distance vector, whose

magnitude is the distance $d_{i,j}$ between objects i and j , as shown in Eq. (3), where $\|\bullet\|$ stands for the Euclidean norm of vector \bullet .

$$\Delta_{i,j} = \begin{bmatrix} \|\mathbf{L}_j - \mathbf{L}_i\| \\ \|\mathbf{M}_j - \mathbf{M}_i\| \\ \|\mathbf{U}_j - \mathbf{U}_i\| \end{bmatrix} \quad (2)$$

$$d_{i,j} = \|\Delta_{i,j}\| \quad (3)$$

Figure 4.2 depicts the geometric representation of these vectors in a 2-dimensional space ($p = 2$) regarding the proposed symbolic distance.

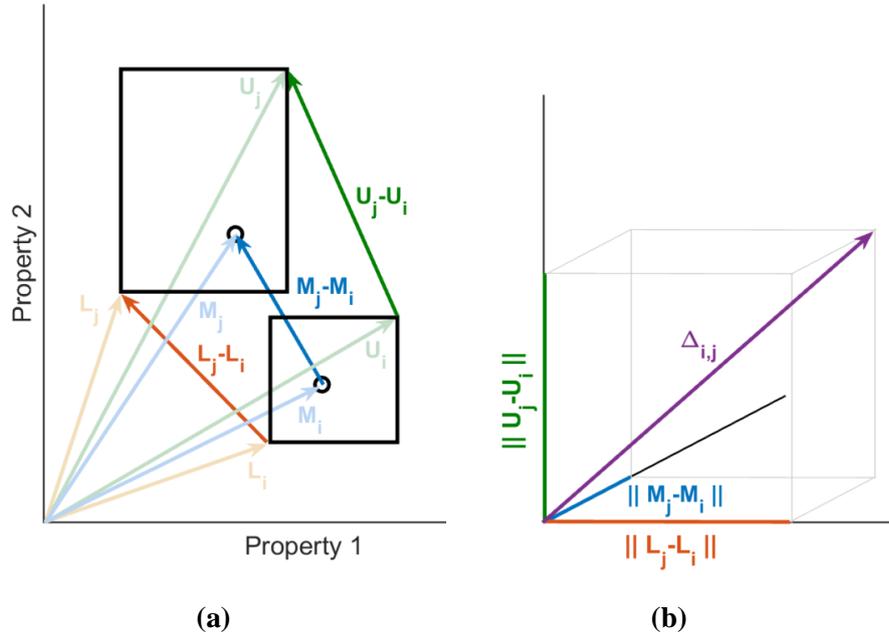


Fig. 4.2 Proposed symbolic distance measurement. (a) Pair of IQRM objects and its relationship; (b) distance vector.

4.2.4 Unsupervised clustering technique

Let us now consider a time-window comprising two generic raw signals containing 2000 samples each, as Fig. 4.3 shows. Each signal can be represented by 20 IQRM objects, each one containing 100 classical objects (samples). Thus, using the proposed metric, one can build a dissimilarity matrix containing the distance values among all pairs of IQRM objects (see Fig. 4.4). Each object is represented by its index both in column and in row of such a matrix. The diagonal elements are always zero (full black in Fig. 4.4), since they represent the

distance values between the same object. Although the direct observation of raw data (Fig. 4.3) does not allow any conclusion about structural novelties, the distance matrix plot, however, clearly suggests the existence of two well-defined behaviors. This conclusion can be reached by observing the dark and light blocks of this matrix, i.e., the group of first ten objects has a high internal similarity (smaller distances or darker colors) and, at the same time, they present a high external dissimilarity from the group of last 10 objects (higher distances or lighter colors).

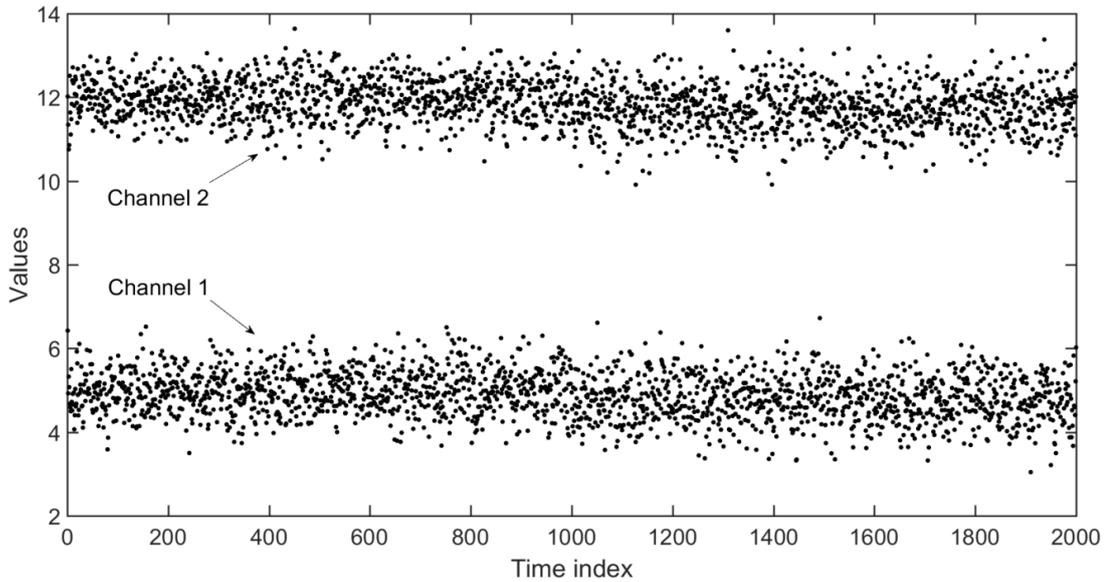


Fig. 4.3 Time-window comprising 2000 bi-dimensional classical objects (samples).

Once the data transformation process is concluded, the k -medoids clustering technique^[36] is applied. In order to conduct a completely unsupervised approach, a novel strategy based on a voting scheme is used to optimally determine the number of clusters k of the dataset. The voting scheme is based on four different clustering validation indexes: Calinski-Harabasz,^[37] Davies-Bouldin,^[38] Gap Statistic^[39] and Silhouette.^[40] These indexes aim at determining the optimal number of clusters within a given dataset. Thus, according to this process, the most “voted” number k is the one to be adopted. In case of a tie, the lowest value of k prevails.

Once the optimal number of clusters k is defined, the k -medoids method provides the results of data partition, pointing out to a single prototype object (medoid) for each one of the clusters. In the current example, the voting scheme elected $k = 2$. Hence, the symbolic data objects were labelled as belonging to cluster 1 (blue) or to cluster 2 (orange), each one with its own prototype (in gray), as Fig. 4.5 and Fig. 4.6 depict. Fig. 4.5a and 5b show the IQRM

objects for each signal in the time domain, whereas Fig. 4.6 shows those objects described in their bi-dimensional form (the prototypes are highlighted in black).

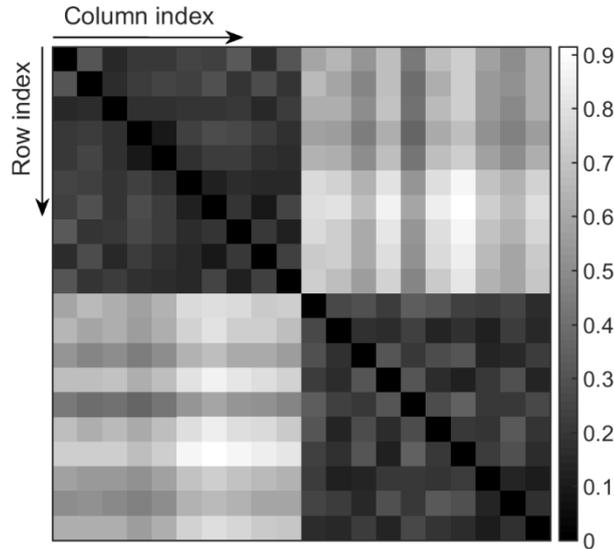


Fig. 4.4 Dissimilarity matrix (20×20), computed using the proposed metric.

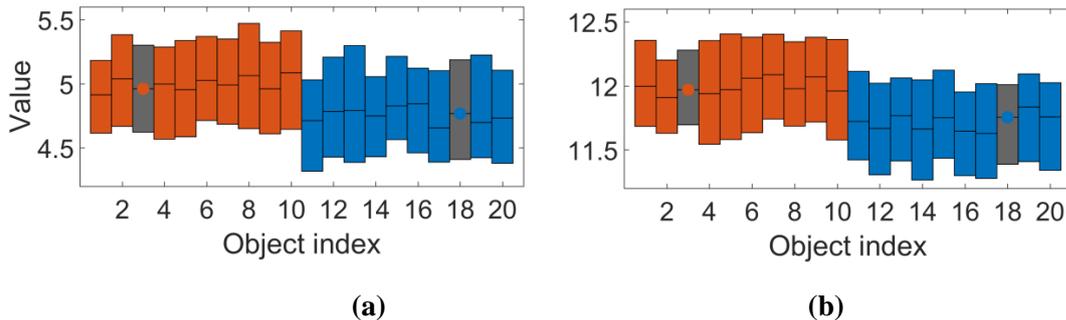


Fig. 4.5 Cluster labels of IQRM objects in time domain: (a) Channel 1 and (b) Channel 2.

4.2.5 Unidimensional feature

Finally, in order to quantify the novelty degree found in the current window, a single-valued index initially proposed by Santos ^[20], named Novelty Index (*NI*), is computed as the mean distance measured among all possible pairs of prototypes, using the same dissimilarity matrix already described. For instance, the time-window containing 20 objects shown from Figs 4.3 to 4.5 would provide a *NI* value equal to 0.627. It is clear that this value alone does not allow concluding if the structure is undergoing a modification or not. However, a series of *NI* values (one per time-window) would suggest if any new structural behavior began to manifest during the real time monitoring. Thus, a time-window containing *S* symbolic objects, each one representing *L* raw data points, moves forward in time, in discrete steps of *L*. Summarizing, as

a new object is acquired, the time-window index TW is increased by one and a new *NI* is computed.

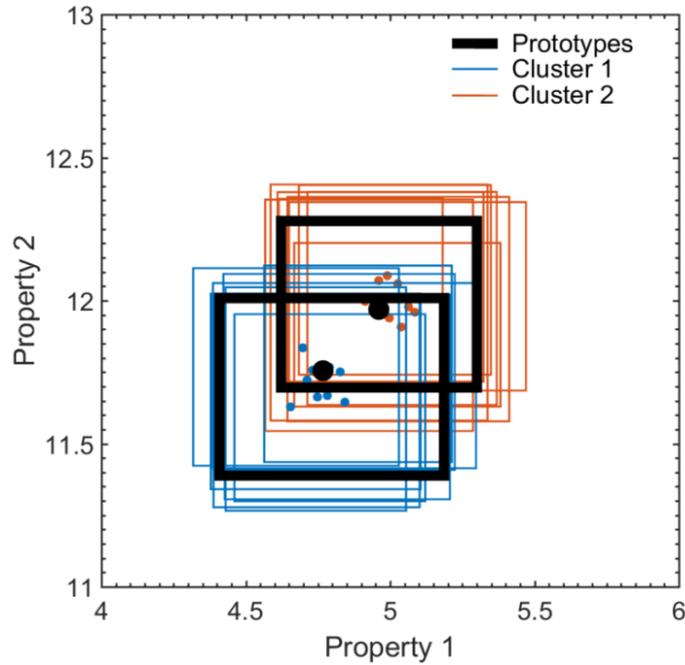


Fig. 4.6 Cluster labels of IQRM objects in their bi-dimensional representation.

Referring to the previous example, one considered a time-window comprising 5 objects ($S = 5$) of 100 points ($L = 100$). Hence, each time-window would contain 500 raw data points ($S \times L = 500$). Figure 4.7 shows the first 12 time-window instants, which are identified by the TW index varying from 5 to 16. On the top right corner of each time-window one can find the *NI* value for each instant. While analyzing the history of these values (Fig. 4.8) it becomes clear that this single-valued feature suggests an increase in the “novelty” degree of the system behavior, which occurs immediately after the acquisition of the 11th symbolic object.

Concerning the choice of the number of symbolic objects in a time-window (S), several tests previously performed showed that the minimum value of 5 is sufficient, provided that an adequate object length (L) is chosen. The object length can be adopted after critically analyzing the trade-off between swiftness and robustness of detection. This means that the shorter the object, the faster the analysis. However, some novelties might be missed.

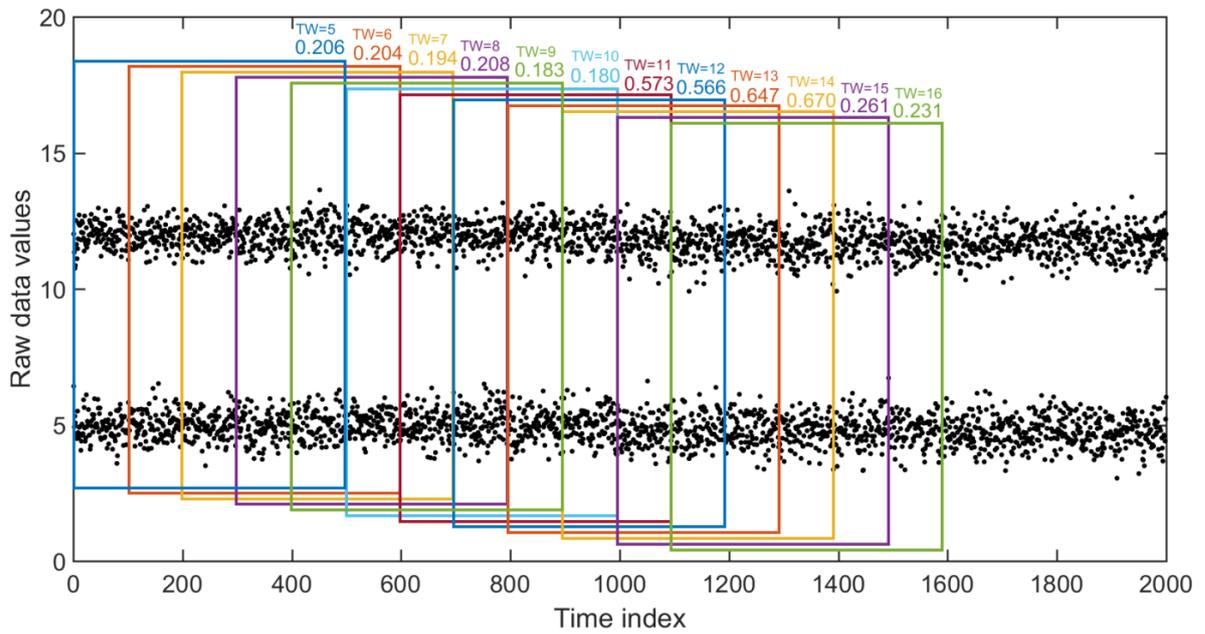


Fig. 4.7 Several instants of the time-window movement, highlighting the respective NI values.

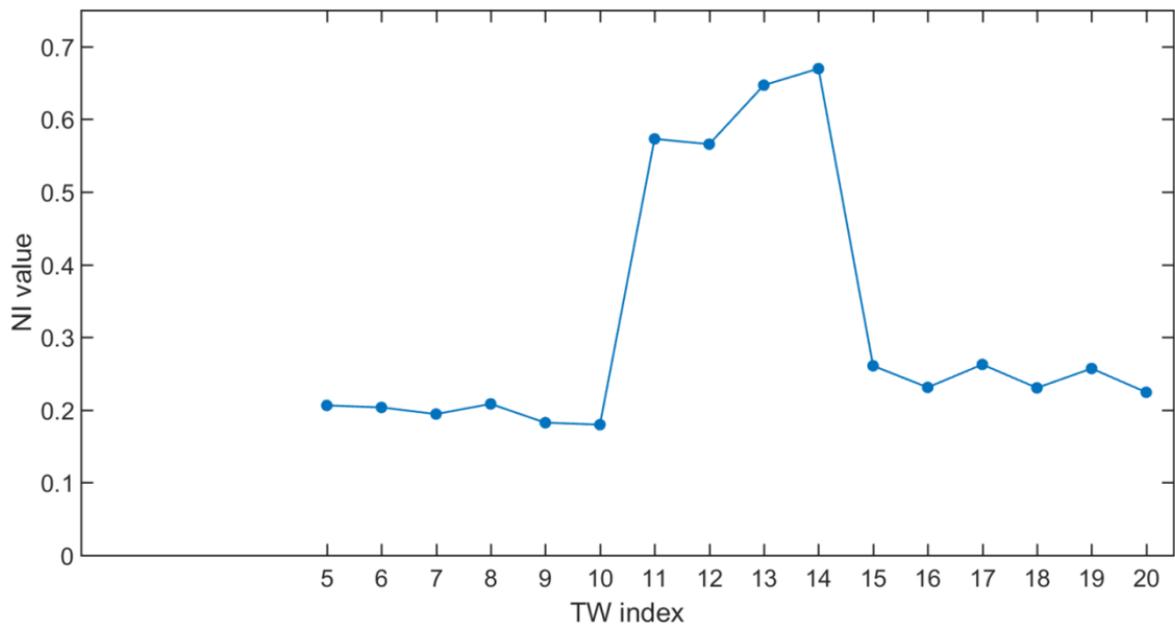


Fig. 4.8 Series of Novelty Indexes obtained by the feature extraction process within each time-window instant.

4.3 Feature Classification

4.3.1 First outlier detection index (FDI)

Despite the interesting comprehension promoted by the NI series (Fig. 4.8), it is still not possible to know whether a newly acquired IQRM object represents a structural damage. Therefore, feature classification is required.

The present paper proposes a technique based on an outlier analysis of NI values within each time-window. A confidence boundary (CB) is defined as being the upper limit of a 99.9% confidence interval, assuming that the data are modelled by a t-student distribution with ‘S-1’ degrees of freedom. To compute CB , one adopts the median and the $S_n^{[41]}$ as the expected value and the variability estimators for the NI values distribution, respectively. Thus, once the NI value exceeds the CB , novelty detection is then characterized.

For the previous example, the moving time-windows would yield the CB series presented in Fig. 4.9. As suspected, a system novelty is detected immediately after the acquisition of the 11th symbolic object.

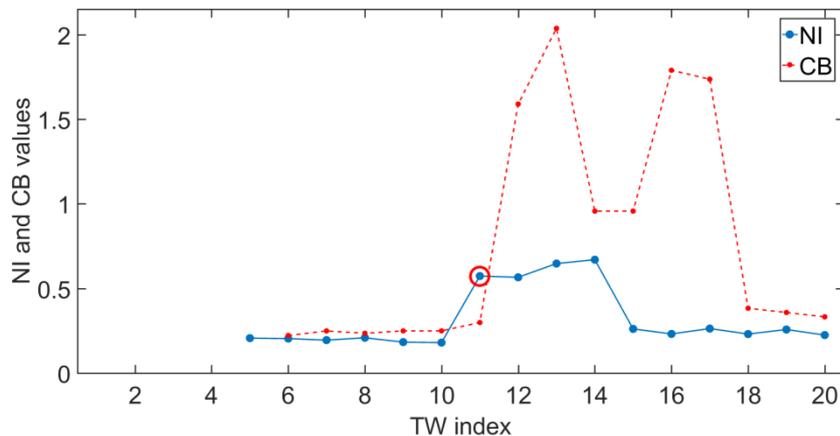


Fig. 4.9 Series of NI and CB values for the illustrative example.

To make the novelty detection even clearer, a First Outlier Detection Index (FDI) is computed in each time-window instant, as defined in Eq. (4). Positive values of FDI indicate a system novelty (see Fig. 4.10). The “First Outlier” term is because the index is computed through a first outlier analysis (over the NI series).

$$FDI_{TW} = NI_{TW} - CB_{TW} \quad (4)$$

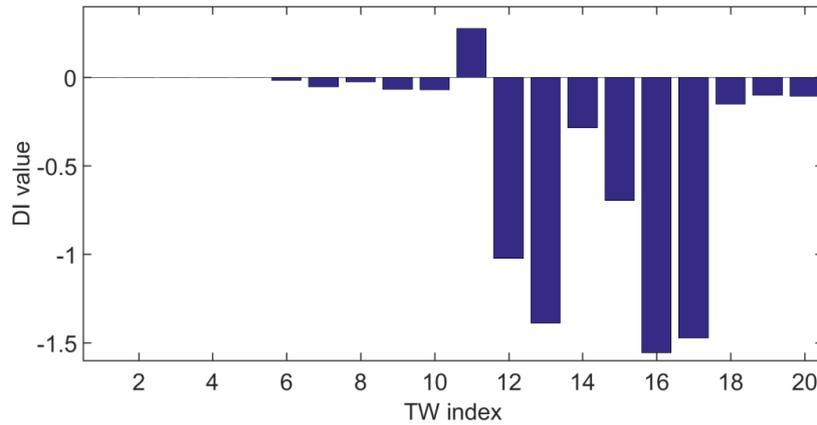


Fig. 4.10 Series of FDI values for the illustrative example.

4.3.2 Second outlier detection index (SDI)

Long-term continuous SHM programs provide massive amounts of data. In most cases, environmental and operational conditions may rise up too many false alarms, especially if no normalization techniques are applied. To solve this problem, many authors suggest performing Principal Component Analysis (PCA) on the raw data, but the vast majority of them point out to a supervised empirical framework. Santos et al.^[21] also proposed a PCA-based normalization strategy, but using the broken-stick rule^[42] for choosing the most appropriate number of principal components related to global effects, such as those caused by environmental and operational actions. This strategy relies on the premise that the higher eigenvalues of PCA are related to environmental and operational effects. Nevertheless, despite the unsupervised content of this approach, there is no absolute certainty that this premise is valid for the many different applications to various SHM scenarios.

For those reasons, the present work proposes the use of an original approach based on a Second Outlier Detection Index. It is named SDI since it consists in a second outlier analysis, which is performed over the FDI series. Therefore, the expression to compute SDI is similar to that of FDI shown in Eq. (4). The difference is that the SDI computation, presented in Eq. (5), considers only the last S positive FDI values as primary data, instead of last S NI values. In this equation, CB_{TW} is the upper confidence boundary defined in the same way shown previously, but considering the last S positive values of FDI.

$$SDI_{TW} = FDI_{TW} - CB_{TW} \quad (5)$$

4.4 Experimental applications

4.4.1 Railway viaduct

This railway viaduct was built in the early eighties and is located in the southeast of France near the cities of Sens and Soucy (Fig. 4.11). It is considered a strategic structure due to its fundamental role in connecting Paris to Lyon via high-speed TGV trains. This structure is 17.5 m long with a single span. Due to resonance manifestations during the train passages, this viaduct has undergone a structural intervention in order to get its flexional stiffness increased.^[19]



Fig. 4.11 Image acquired during the TGV pass.

Eight vertical piezoelectric accelerometers were placed on the viaduct deck and the acquisition rate was set to 4096 Hz. Each time the train passed over the viaduct, 2 seconds of forced vibrations were recorded. This is equivalent to $L = 8192$ classical data points per channel (Fig. 4.12). The monitoring campaign consisted in 13 tests before and 13 tests after the structural intervention. The complete raw data history can be seen in Fig. 4.13, where each time-window instant is demarcated (TW index from 1 to 26).

This application aims at testing if the proposed methodology is suitable to detect changes in the structure's behavior. It is important to emphasize that no modal identification or data normalization procedures are previously performed. The feature extraction and classification are carried out at each time-window directly from the raw signals provided by the eight vertical accelerometers.

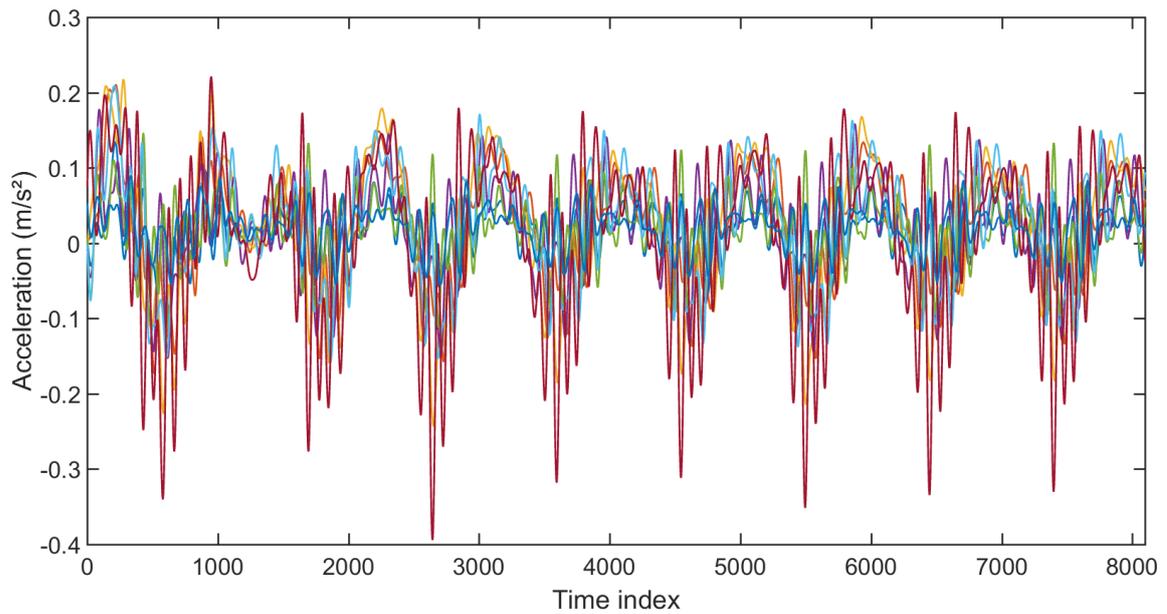


Fig. 4.12 Raw data acquired during a single train passage.

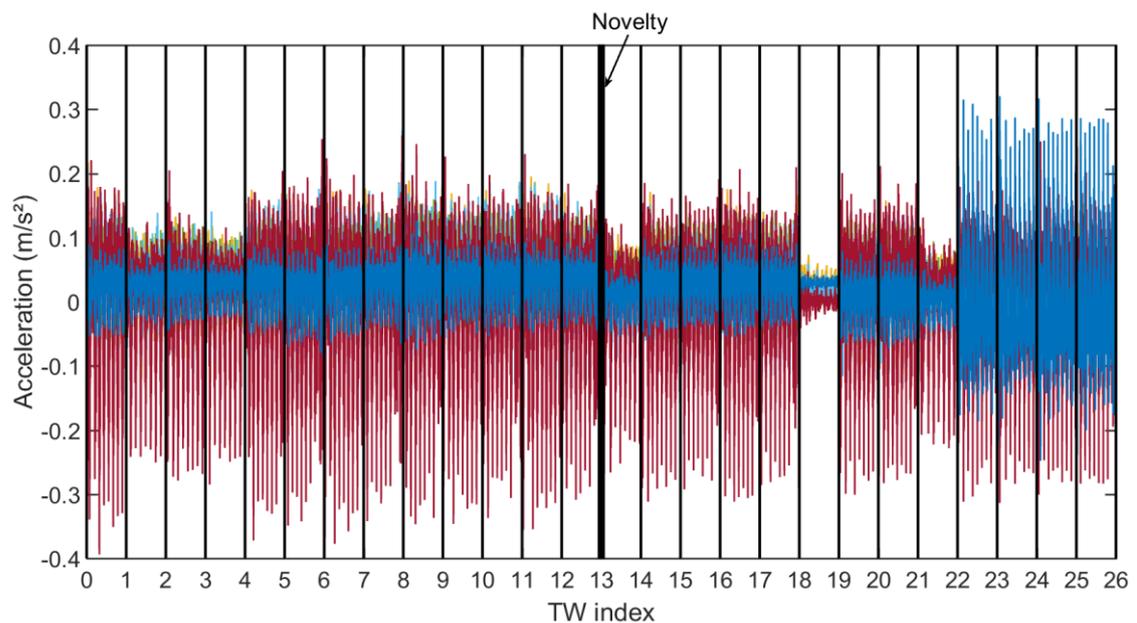


Fig. 4.13 Complete raw data history showing each time-window boundary (one TW per train pass).

The proposed approach considers a single IQRM object per dynamic test. Therefore, each symbolic object consists of only 24 values ($3 \text{ quartiles} \times 8 \text{ channels}$) representing 65,536 raw data values (8,192 values measured by eight channels, each). This symbolic representation yields a “compression” of 2,700 times in terms of data volume. Figure 4.14 shows the variation of the IQRM objects for the first channel over the 26 tests. The blue dashes represent the median values, while the lower and upper limits of the grey rectangles represent the 1st and 3rd quartiles, respectively.

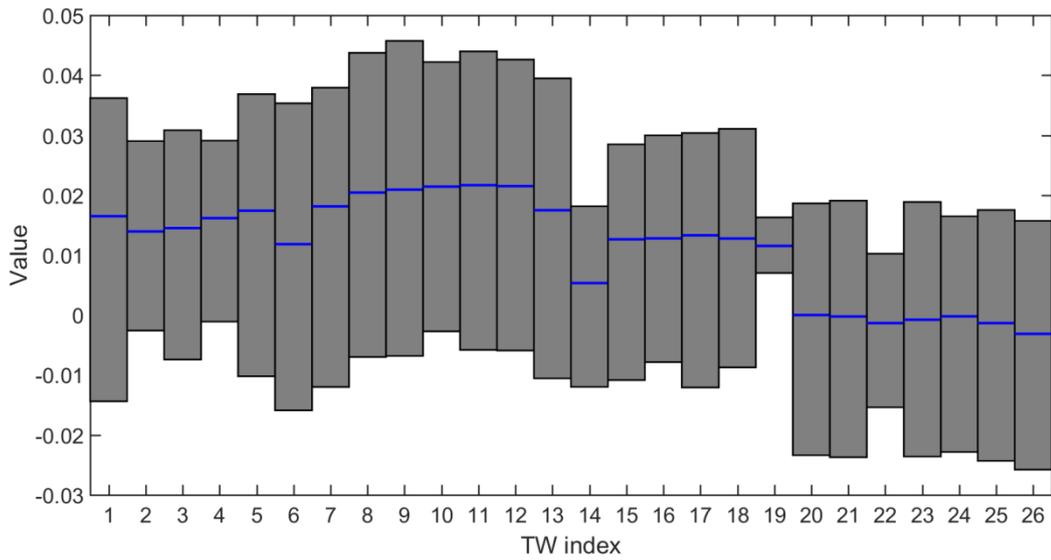


Fig. 4.14 Sequence of IQRM objects, represented here by their first property (channel).

Figure 4.15 shows the evolution of both NI and CB indexes. One observes that the IQRM objects series do not allow concluding whether or when a novelty occurs. However, considering a time-window with five objects ($S = 5$), the NI history highlights the structural “degree of novelty” and, finally, the FDI computation shows that the method detected a single novelty, which occurred immediately after the 13th instant (Fig. 4.16), i.e., right after the structural intervention.

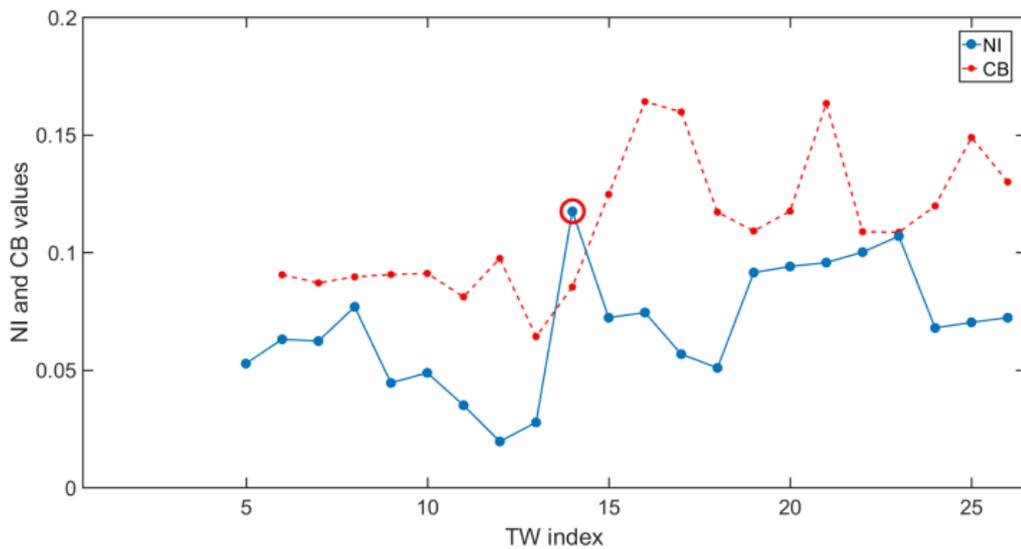


Fig. 4.15 NI and CB values at each instant TW.

These results give an idea of the generalization power provided by the use of IQRM symbolic objects. Besides, for the studied case of forced vibrations, the proposed index FDI appears to be appropriate for detecting structural novelties, even without the application of a

prior normalization procedure. Such detection occurred immediately after the acquisition of the first object containing the data expressing the new behavior (TW = 14), which ensures the real-time aspect of the proposed method. However, one should notice that, despite the unsupervised approach, the first five instants (TW from 1 to 5) had to be used as a short training (or initialization) period so that the method could “learn” about the structure’s behavior.

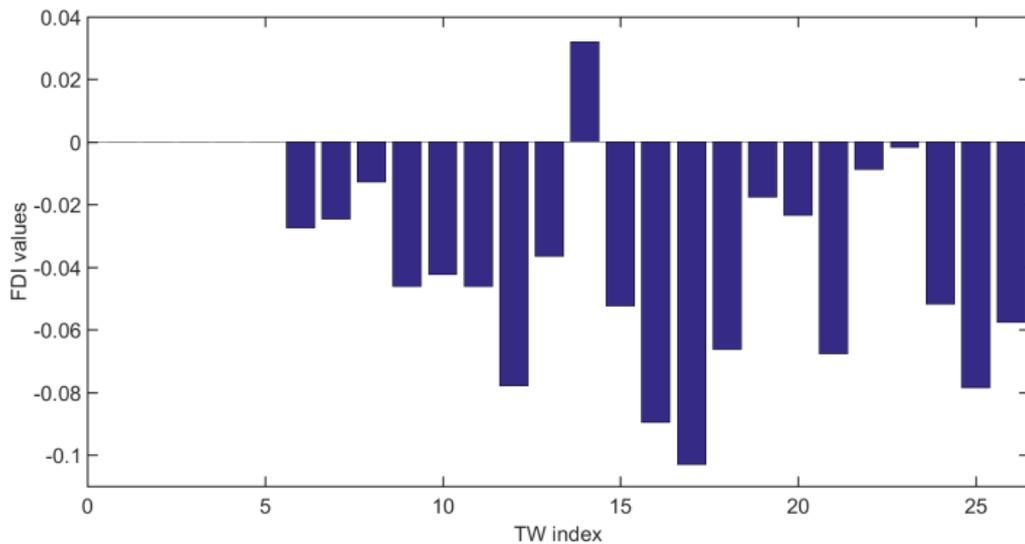


Fig. 4.16 FDI values at each instant TW.

4.4.2 Masonry tower

This application consists in a long-term continuous SHM. For this reason, the SDI series will be used instead of the FDI. The monitored structure is the *Gabbia* Tower, about 54 meters high and the tallest one located in Mantua, Italy. The end of its construction dates to 1227 according to recent researches ^[43]. The tower has a near square plan and the load-bearing walls, built in solid brick masonry are about 2.4 m thick except for the upper levels, where the section decreases to about 0.7 m, as shown in Fig. 4.17.

After preliminary ambient vibration tests, a simple continuous dynamic monitoring system was installed in the tower. It consists of three piezoelectric accelerometers mounted on the cross-section at the crowning level of the tower. A binary file containing three acceleration time series is created every hour. The recorded accelerations were low-pass filtered, using a classic seventh-order Butterworth filter with a cut-off frequency of 20 Hz, and decimated five times, reducing the sampling frequency from 200 Hz to 40 Hz. ^[44]

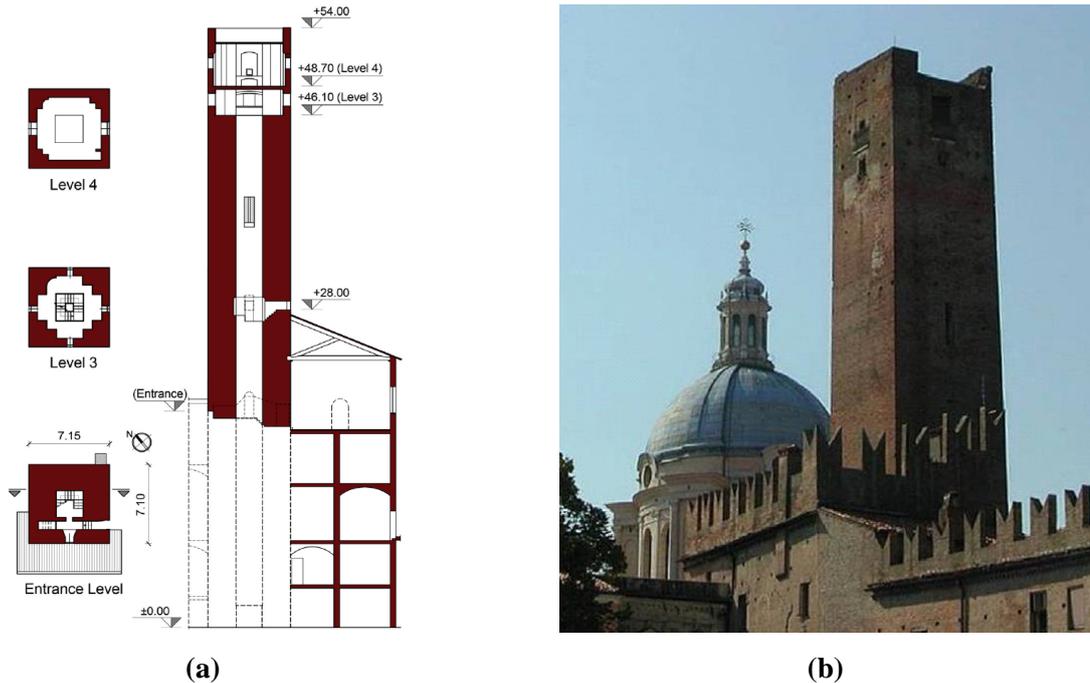


Fig. 4.17 Gabbia tower: (a) floor plans, section and (b) recent photography.

For this application, a 14-day monitoring period (from 15th to 28th of June, 2013) of continuously acquired acceleration time series was used. The purpose is to check if the proposed routines are able to detect the structural novelty caused by a seismic event occurred in 21/06/2013, just by resorting to raw data without signal normalization. Fig. 4.18 shows a typical raw data continuously acquired during a period of 5 days, in June of 2013. In fact, just by checking the three acceleration signals, it is not possible to confirm the existence of an abnormal structural behavior. Thus, again, the method is supposed to extract IQRM objects from raw data.

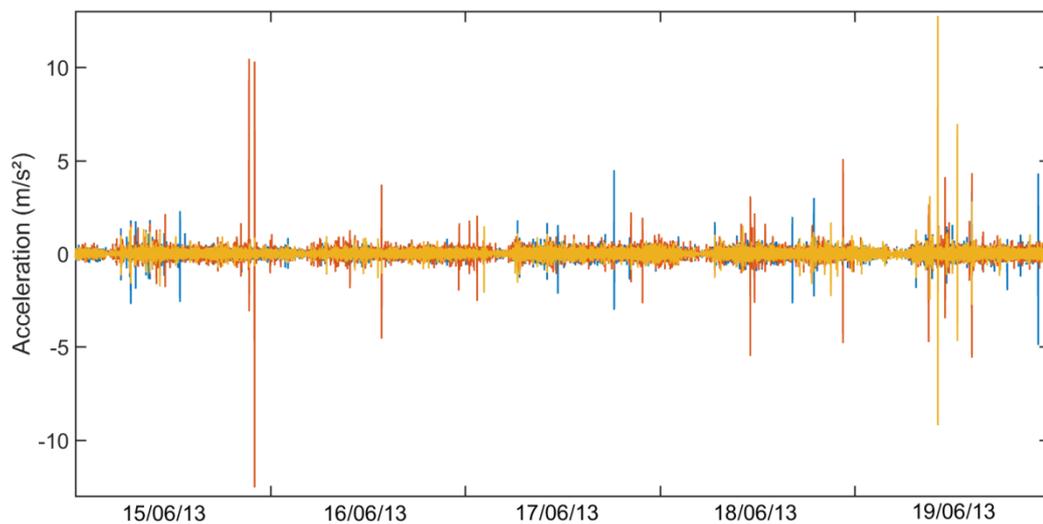


Fig. 4.18 Sample of five days of continuous monitoring (3 channels).

To provide promptness in detection, each IQRM object is composed of $L = 25000$ raw data samples that corresponds to 10 minutes of acquisition, approximately. Fig. 4.19 depicts the IQRM series for the first day of monitoring (15/06/13) considering one property (channel) only. It is interesting to observe how the structure behaves in different intensities during the day: at dawn, the acceleration level is the lowest; in the morning, those levels start to increase until noon, when they keep their intensity until they reach another peak at the beginning of night when, again, it starts to decrease to the minimum level. Traffic-induced vibrations nearby the structure probably explain that pattern.

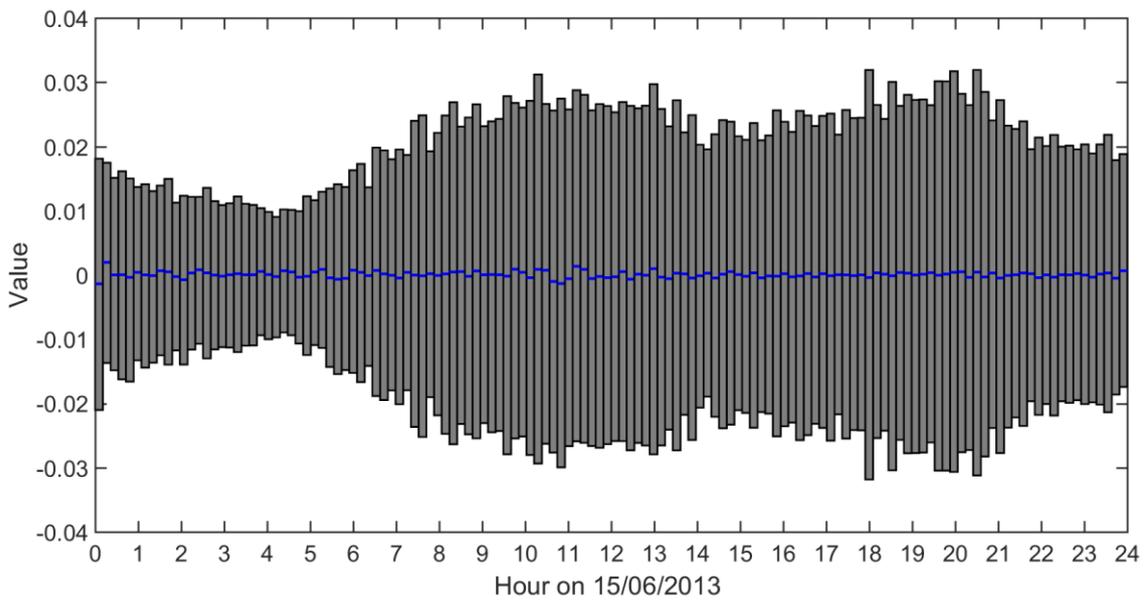


Fig. 4.19 One day of acquired IQRM objects represented by their first property (channel).

Again, each time-window comprises just the minimum of five objects ($S = 5$), which is equivalent to 50 minutes of acquisition, approximately. Figure 4.20 shows the results for both NI and its 99.9% confidence boundary (CB) series over the 14 days.

From Fig. 4.20, it is not possible to observe if any structural novelties have occurred. Hence, in the sequence, the FDI values are computed together with its 99.9% confidence boundary (CB series). The result is presented in Fig. 4.21, where a single novelty becomes evident. Alternatively, this conclusion could also be drawn from Fig. 4.22, through the SDI series. One should notice that the proposed index presented a single positive value, which corresponds exactly to the seismic event that occurred in 21/06/13.^[43,44]

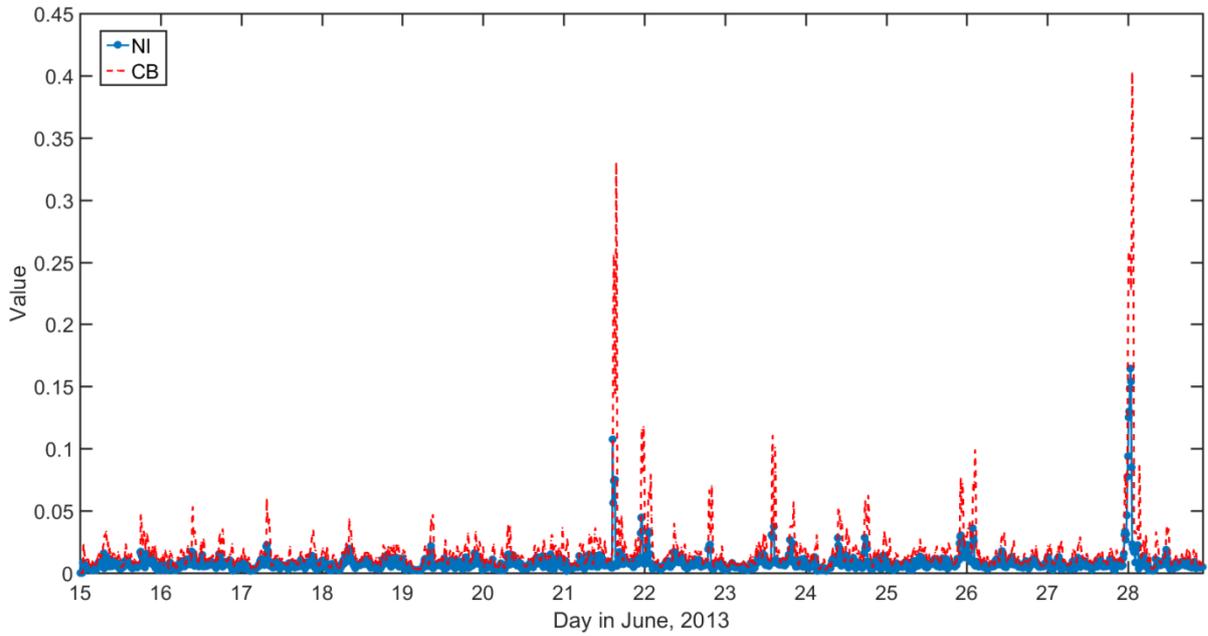


Fig. 4.20 NI and CB series generated in 2 weeks of continuous monitoring.

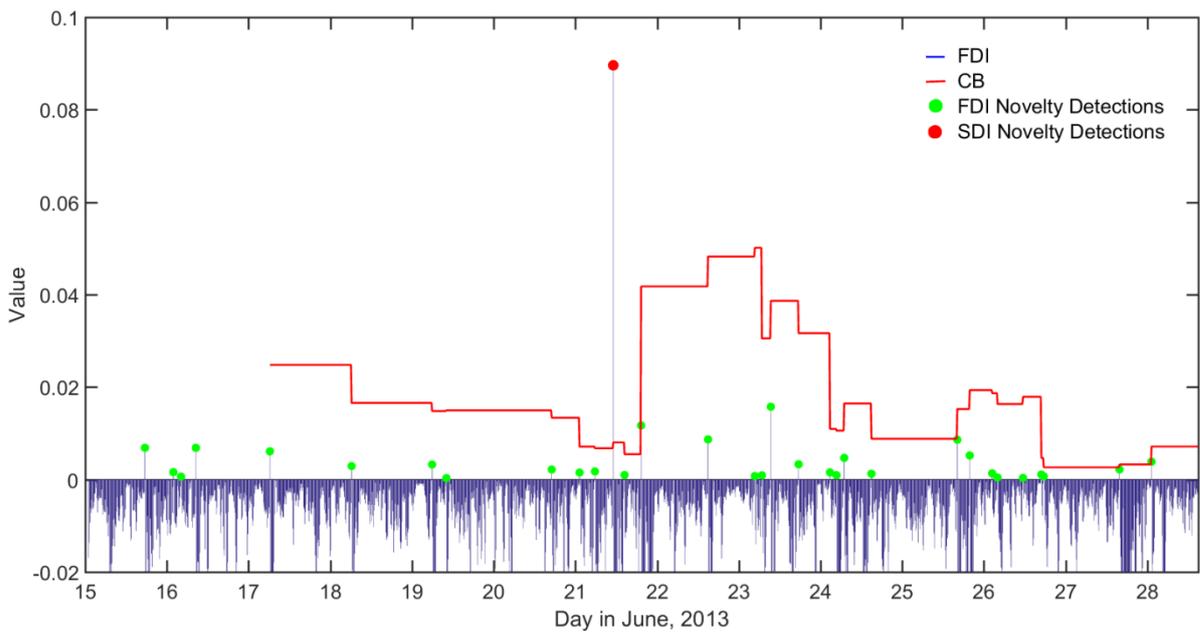


Fig. 4.21 Two-week history of First Order Detection Index (FDI).

From Fig. 4.21, one notices that the FDI pointed out to 31 false alarms. Conversely, the proposed SDI detected just the one positive value, which occurred immediately after the acquisition of the symbolic object containing the novel structural behavior. Additionally, the adaptiveness of the SDI to the new system condition can be observed while the CB values decrease during approximately two days after the novelty occurrence. After such “recovery period” (days 22 and 23), the method is able to detect another abnormal behavior.

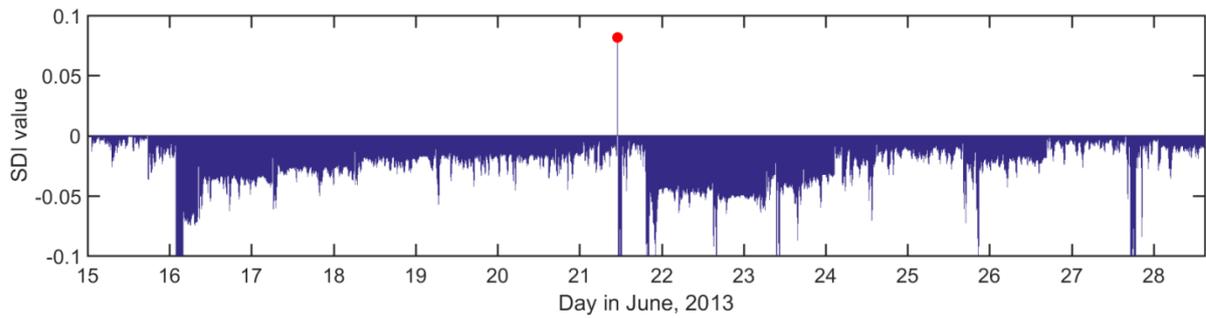


Fig. 4.22 Two-week history of Second Order Detection Index (SDI).

4.5 Conclusion

This paper proposed a fully unsupervised and adaptive methodology capable to detect structural novelties accurately using raw dynamic data directly. To this end, it was employed a new representation for raw dynamic signals, i.e., accelerations, based on an original symbolic data object (SDO). The SDO comprised all four data descriptors, such as expected value, variability, symmetry and flatness. Such a representation provides a richer, yet compact way to characterize raw dynamic measurements. Moreover, a statistical learning procedure based in the k -medoids technique was used to generate the single-valued feature named NI (Novelty Index). From the NI series, the First Outlier Detection Index (FDI) was then obtained, from which the Second Outlier Detection Index (SDI) could be computed.

With the results obtained from the two real SHM application cases, one can infer that the proposed methodology has two variants: the first one defines FDI as the novelty indicator, which seems to be appropriate to short-term monitoring cases, like the one of the studied viaduct. The second one defines SDI as a more capable detection index as it seems to be more robust to continuous long-term SHM, like the masonry tower case. Indeed, the FDI requires a shorter initialization period (Fig. 4.16). On the other hand, it rises up too many false alarms in long-term continuous SHM (Fig. 4.21). The SDI, however, requires a longer initialization period and may not suitable for short-term SHM. At the same time, it seems to be robust enough to perform long-term SHM without revealing too many false alarms or missing real damage scenarios (Fig. 4.22).

Acknowledgements

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico

and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for the financial support, IFSTTAR (former LCPC – Laboratoire Central des Ponts et Chaussées) and SNCF (Société Nationale des Chemins de fer Français) for the data used in this paper (project 01V0527 RGPU “Evaluation dynamique des ponts”).

References

- [1] M. Sahin, R. A. Sheno, *Quantification and localisation of damage in beam-like structures by using artificial neural networks with experimental validation*, *Engineering Structures*, **2003**, 25(14), pp.1785–1802.
- [2] S.D. Glaser, A. Tolman, *Sense of Sensing: From Data to Informed Decisions for the Built Environment*, *Journal of Infrastructure Systems*, **2008**, 14(1), pp.4–14.
- [3] X.G. Hua et al., *Structural Damage Detection of Cable-Stayed Bridges Using Changes in Cable Forces and Model Updating*, *Journal of Structural Engineering*, **2009**, 135(9), pp.1093–1106.
- [4] F.L. Wang et al., *Correlation-Based Damage Detection for Complicated Truss Bridges Using Multi-Layer Genetic Algorithm*, *Advances in Structural Engineering*, **2012**, 15(5), pp.693–706.
- [5] B.H. Koh, S.J. Dyke, *Structural health monitoring for flexible bridge structures using correlation and sensitivity of modal data*, *Computers & Structures*, **2007**, 85(3-4), pp.117–130.
- [6] H. Sohn, C. Farrar et al., *A Review of Structural Health Monitoring Literature: 1996 – 2001*, Los Alamos, USA, **2004**.
- [7] K. Worden, J.M. Dulieu-Barton, *An Overview of Intelligent Fault Detection in Systems and Structures*. *Structural Health Monitoring*, **2004**, 3(1), pp.85–98.
- [8] D.F. Giraldo, *Damage Detection Accommodating Varying Environmental Conditions*, *Structural Health Monitoring*, **2006**, 5(2), pp.155–172.
- [9] T.Y. Hsu, C.H. Loh, *Damage detection accommodating nonlinear environmental effects by nonlinear principal component analysis*, *Structural Control and Health Monitoring*, **2010**, 17(3), pp.338-354.

- [10] D. Posenato, *Model-Free Data Interpretation for Continuous Monitoring of Complex Structures*, École Polytechnique Fédérale de Lausanne, **2009**.
- [11] C.R. Farrar, K. Worden, An introduction to structural health monitoring, *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **2007**, 365(1851), pp.303–315.
- [12] J. Kullaa, Damage Detection of the Z24 Bridge Using Control Charts, *Mechanical Systems and Signal Processing*, **2003**, 17(1), pp.163–170.
- [13] E. Figueiredo, *Damage Identification in Civil Engineering Infrastructure under Operational and Environmental Conditions*. University of Porto, Faculty of Engineering, **2010**.
- [14] D. Posenato et al., Methodologies for model-free data interpretation of civil engineering structures, *Computers & Structures*, **2010**, 88(7-8), pp.467–482.
- [15] H. Sohn, J. Czarnecki, C. Farrar, Structural Health Monitoring Using Statistical Process Control. *Journal of Structural Engineering*, **2000**, 126(11), pp.1356–1363.
- [16] M. Gul, F. Catbas, Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications, *Mechanical Systems and Signal Processing*, **2009**, 23(7), pp.2192–2204.
- [17] K. Nair, & A.S. Kiremidjian, Time Series Based Structural Damage Detection Algorithm Using Gaussian Mixtures Modeling, *Journal of Dynamic Systems, Measurement, and Control*, **2007**, 129(3), pp.285–293.
- [18] S. Silva et al., Structural damage detection by fuzzy clustering, *Mechanical Systems and Signal Processing*, **2008**, 22(7), pp.1636–1649.
- [19] V. Alves, A. Cury, C. Cremona. On the use of symbolic vibration data for robust structural health monitoring, *Proceedings of the Institution of Civil Engineers - Structures and Buildings*, **2016**, 169(9), pp. 715-723.
- [20] J. Santos, *Smart structural health monitoring techniques for novelty identification in civil engineering structures*, Ph.D. thesis, Instituto Superior Técnico, Univ. of Lisbon, Lisbon, Portugal, **2014**.

- [21] J. Santos et al., Early Damage Detection Based on Pattern Recognition and Data Fusion, *Journal of Structural Engineering*, **2017**, 143(2).
- [22] S. Mattson, S. Pandit, Statistical moments of autoregressive mode residuals for damage localization, *Mechanical Systems and Signal Processing*, **2006**, 20(3), pp.627–645.
- [23] J. Santos et al., Baseline-free real-time assessment of structural changes. *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance*, **2014**.
- [24] R. Cardoso, A. Cury, F. Barbosa, A robust methodology for modal parameters estimation applied to SHM. *Mechanical Systems and Signal Processing*, **2017**, 95 (2017), pp.24–41.
- [25] A. Cury, C. Borges, F. Barbosa, A two-step technique for damage assessment using numerical and experimental vibration data, *Structural Health Monitoring*, **2011**, Vol 10, Issue 4, pp. 417 – 428.
- [26] H.F. Zhou et al., Eliminating Temperature Effect in Vibration-Based Structural Damage Detection. *Journal of Structural Engineering*, **2011**, 137(2), pp.785–796.
- [27] X. Hu, H.W. Shenton, Damage Identification Based on Dead Load Redistribution Methodology, *Journal of Structural Engineering*, **2006**, 132(8), pp.1254–1263.
- [28] J.M. Ndambi, J. Vantomme, K. Harri, Damage assessment in reinforced concrete beams using eigenfrequencies and mode shape derivatives, *Engineering Structures*, **2002**, 24(4), pp.501– 515.
- [29] S.W. Doebling et al., *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*, Los Alamos, USA, **1996**.
- [30] A. Torres, V. Alves, A. Cury, F. Barbosa, Advanced statistical techniques applied to raw data for structural damage detection, Advanced Statistical Techniques Applied to Raw Data for Structural Damage Detection, *Experimental Vibration Analysis for Civil Structures (EVACES)*, **2017**, Lecture Notes in Civil Engineering, vol 5, pp 94-103, Springer.

- [31] R. Finotti, A. Bonifacio, F. Barbosa, A. Cury, Evaluation of computational intelligence methods using statistical analysis to detect structural damage, *Mecánica Computacional*, **2016**, Vol XXIV, pp 1389 – 1397.
- [32] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The Mahalanobis distance, *Chemometrics and Intelligent Laboratory Systems*, **2000**, 50(1), pp.1–18.
- [33] I.T. Jolliffe, *Principal Component Analysis* 2nd ed., Aberdeen, UK: Springer, **2002**.
- [34] V. Alves, A. Cury, N. Roitman, C. Magluta, C. Cremona, Novelty detection for SHM using raw acceleration measurements, *Structural Control Health Monitoring*, **2015**, 22, 1193–1207.
- [35] L. Billard; E. Diday. *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. John Wiley & Sons, **2007**.
- [36] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Hoboken, New Jersey: John Wiley & Sons, Inc, **2009**.
- [37] T. Calinski, J. Harabasz,. *A dendrite method for cluster analysis*, *Communications in Statistics*, **1974**, Vol. 3, No. 1, pp.1–27.
- [38] D.L. Davies, D.W. Bouldin, A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **1979**, Vol. PAMI-1, No. 2, pp.224–227.
- [39] R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B*, **2001**, Vol. 63, Part 2, pp. 411–423.
- [40] P.J. Rouseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*. **1987**, Vol. 20, No. 1, pp. 53–65.
- [41] P. Rousseeuw, C. Croux, Alternatives to the Median Absolute Deviaton. *Journal of the American Statistical Association*, **1993**, 88(424), pp.1273–1283.
- [42] D.A. Jackson, Stopping rules for principal component analysis, *Ecology*, **1993**, 74(8), pp.2204–2214.

[43] A. Saisi et al, On Site Investigation and Health Monitoring of a Historic Tower in Mantua, Italy. *Applied Sciences*, **2016**, 6, 173.

[44] A. Saisi et al., Post-earthquake continuous dynamic monitoring of the Gabbia Tower in Mantua, Italy, *Construction and Building Materials*, **2015**, 81, pp.101–112.

Chapter 5

Paper #4 – Automated real-time damage detection strategy using raw dynamic measurements

This chapter presents the most recent paper written during this thesis. In this paper, an original symbolic object is proposed to enrich the signal's representation. Such a description embraces both temporal and spectral aspects of the raw data. Unsupervised pattern recognition procedures are performed to yield an instantaneous damage-sensitive index. The results obtained suggest that the developed technique is accurate and robust to detect even small damage. **With this strategy, the objective related to the second approach of this thesis, i.e., to develop an automatic, unsupervised and real-time damage detection method based on raw measurements is achieved.** This paper was submitted to the journal “*Engineering Structures*”.

Paper Information

Status Submitted in October, 2018

Publisher Engineering Structures
ISSN 0141-0296; Qualis (A1); Impact Factor (2018): 2.755

Authors Cardoso R, Cury A, Barbosa F.

Title Automated real-time damage detection strategy using raw dynamic measurements.

Abstract

Over the last decades, several techniques have been developed in the context of Structural Health Monitoring (SHM) programs. However, when it comes to novelty (or damage) detection, these methods are inevitably based on human decisions. Moreover, most of the strategies already published in this topic mainly focus on modal identification procedures and tracking their outputs i.e., structural modal parameters. Such approaches usually lead to high computational costs and can still be insensitive to minor changes in structural behavior, thus missing crucial damage scenarios in their initial manifestations. To circumvent these drawbacks, recent researches showed that the use of symbolic representations derived directly from raw time-domain data (e.g. acceleration measurements) could provide more damage-sensitive responses with lower computational effort. Indeed, good results have been achieved by representing raw measurements in terms of their statistical distributions over time. Nevertheless, the lack of information regarding the frequency spectrum represents a decisive drawback. Therefore, this paper presents a novel symbolic object, which considers both time and frequency responses of structural dynamic measurements. Then, the proposed methodology employs a k -medoids clustering over such objects within a moving time-window framework and uses a single-valued index to indicate whether a novelty is present in the acquired data. Two practical studies – a 3D frame tested in laboratory and a motorway bridge – show that the proposed approach provide an unsupervised and adaptive scheme for SHM applications.

5.1 Introduction

The existence of a SHM program on strategic structures like viaducts, bridges, ageing buildings and towers can be crucial for preventing disasters. Damage-related information extracted from the structural behavior may save lives and avoid great economical losses. A recent example of disaster occurred in Genoa, Italy (August 14, 2018), where a cable-stayed bridge collapsed leaving 43 people dead. The southern stays (from one tower) that initially went slack are the same where damage was possibly found during dynamic tests performed almost one year earlier ^[1].

This example shows the utmost importance that SHM programs have to complement or even replace human inspections. Ideally, such a monitoring procedure must work

automatically and continuously providing updated sensitive information about the structure's condition. Besides, to reduce maintenance costs and prevent social, physical and economical losses from structural collapses, it is fundamental that such algorithms are able to readily detect abnormal dynamic behavior (novelty) or even damage in its early manifestations.

Recent technological advances allowed the construction of reliable, accurate and efficient monitoring equipment that can be remotely configured and operated. This fact encouraged the development of specific software and the installation of dynamic monitoring systems in a considerable number of structures abroad.

The acquired raw data (usually acceleration measurements) can be processed in many different ways to provide damage-sensitive features. In general, the literature presents two different SHM strategies for damage detection: *inverse* (*model-based* methods) and *forward* (*data-driven* methods).

Model-based methods generally require the user to have a deep knowledge of the structure's behavior, usually including a well-tuned Finite Element model corresponding to the assumed healthy state of the system. Indeed, it is not a simple task to fit an FE model (or an analytic model) to the acquired data by using a limited number of sensors from an actual SHM. Struggling to achieve that, many physical parameters must be adjusted via optimization routines, but yet with considerable human/computational effort and several subjective decisions. Consequently, the problem becomes typically nonlinear and non-unique. Furthermore, the literature points out that this type of method requires the use of a large number of sensors ^[2,3] and usually needs a considerable amount of mode shapes to be identified ^[4], including local ones. Even the state-of-the-art ^[5-7] of this kind of approach is not suitable for automated real time damage detection in SHM.

Conversely, *forward* strategies are capable of handling significant structural complexity with relative low computational cost, depending only on the acquired signals and their subsequent processing, which fundamentally consists in two steps: *feature extraction/data fusion* and *feature classification*. Thus, model construction/fitting and updating for each single structural system being monitored are unnecessary. Moreover, this type of methodology is capable of handling output-only information i.e., structural responses, without requiring the measurement of external excitations, such as operational and environmental effects.

Despite that, most of the *data-driven* methods referred in the literature are still based on *supervised* statistical learning procedures, ^[8-15, 20] since they require long periods of training, known as *baselines*, during which the structure is assumed to be in its healthy state. Such an assumption is not always true or even possible. On one hand, most of SHM applications are performed on aged structures, which may already have endured structural modifications or even damage. On the other hand, newly built structures may not present a long enough baseline period to allow the correct establishment of a healthy reference state.

As an early attempt to detect damage, many of published techniques depended on modal information i.e., natural frequencies, damping ratios and mode shapes. Nevertheless, although modal frequencies and shapes are deeply related to the mechanical behavior of structures (mass and stiffness), they are not sensitive enough to detect small novelties in the system (e.g. early damage). The main reason is that most of these abnormalities are local and, therefore, do not impact significantly on the first vibration modes, which are estimated, in general, with best precision ^[16].

To avoid this limitation, instead of relying on the analysis of modal parameters, some recently published methods propose the use of raw time-domain data (e.g. acceleration series) represented by means of Symbolic Data Objects (SDOs). This type of approach is gaining special attention over the last five years ^[13, 17-20] due to the great power of data fusion provided by SDOs. Indeed, such a representation allows the efficient manipulation of large amounts of data with low computational effort while keeping the original information sufficiently preserved.

Such methods are generally based on machine learning strategies, usually represented by classification techniques. However, some of them are still dependent on long damage-free baseline periods (supervised approach) ^[13, 20]. Besides, they are frequently based on the prior knowledge of the structure's behavior: past, present and future. Therefore, they are not suitable for the SHM purpose as previously mentioned, which is the real-time unsupervised monitoring.

One of the most promising approaches proposed in 2017 ^[18] was based on previous researches published in 2014 ^[17] and 2015 ^[19]. It consisted in an unsupervised machine learning procedure over SDOs. To enforce the "real-time" character to the method, a time-window process was employed to extract updated single-valued damage indexes. Firstly, raw data was written in a Principal Component (PC) basis. In the sequence, the PC series was

truncated using the broken-stick rule (in an attempt to remove environmental related influences). Then, they were fused into interquartile ranges (IQR), which was adopted as the SDO representation. Thus, the SDOs underwent a hierarchical clustering analysis. For that purpose, within each time-window, the distance matrix was computed using the normalized Euclidean Ichino-Yaguchi ^[21] symbolic metric. The number of clusters was automatically chosen by using the Silhouette validity index ^[22]. Finally, an outlier analysis was performed to generate a single-valued index. Nevertheless, it is worth mentioning that even though the method was indeed promising, their study was not validated with applications of actual experimental cases of SHM. Instead, only numerical simulations were used to generate data.

The present paper proposes a fully automated unsupervised real-time SHM method for damage (or novelty) detection. The proposed approach starts from the premise that structural novelties affect both time and frequency domain responses. Thus, an original SDO (richer than IQR) comprising these two information domains is designed. Naturally, due to the new SDO special characteristics, a suitable distance metric is also proposed. The *k*-medoids technique is chosen due to its attested robustness ^[23]. Finally, aiming at detecting damage, a new procedure to obtain a single-valued index is also explained.

This paper is organized as follows: Section 5.2 (*data fusion / feature extraction*) presents the newly proposed symbolic object along with a suitable distance metric and the procedure to obtain a scalar feature, which value is proportional to the structural changes. Section 5.3 (*feature classification*) shows how the time series of such features go through an outlier analysis to generate a single-valued damage index, which points out whether a damage (or novelty) has occurred. The proposed methodology is assessed through two experimental tests performed on real-scaled structures subjected to ambient vibration (Sections 5.4 and 5.5). Finally, Section 5.6 draws conclusions and comments on the overall strategy's performance.

5.2 Feature Extraction and Data Fusion

5.2.1 Extraction of symbolic data objects

A usual SDO applied to Civil Engineering SHM is the Interquartile Range ^[17-19], herein called IQR object. In fact, such a SDO represents the statistical distribution of the amplitudes of the original signal (raw data) in a very compact form by simply using two values per measuring channel: the 1st and 3rd quartiles. Alternatively, a richer way to describe such data is through

histograms, where a higher number of divisions (classes) implies a more detailed (yet more massive) representation ^[13, 20].

Nevertheless, none of these SDOs takes into account the information contained in the frequency spectrum of the original data. Indeed, as the SDOs perform the data fusion exclusively in the time domain, they miss relevant information related to the frequency domain.

This paper proposes a more complete SDO that embraces both *time* and *frequency* information of the original data. The *time* aspect is represented by the three quartiles of the amplitude's distribution of the raw acceleration signals. The *frequency* aspect is characterized by the three quartiles of the vibrational energy distribution expressed in the frequency spectrum (e.g. Power Spectrum Density) of the raw acceleration measurements.

For reference purposes, the new SDO is onwards named Time-Frequency IQRM object (TF-IQRM), where IQR stands for the Interquartile Range, i.e. the difference between the third and the first quartiles (Q_3 and Q_1); and M stands for the median, i.e. the second quartile (Q_2). Moreover, the superscripts T and F indicate whether the quantity refers to time or frequency domain, respectively.

Figure 5.1 illustrates how to obtain the TF-IQRM object from raw data. In this example, the signal consists in a single numerically generated acceleration time history. Let the *number of channels* and the *SDO length* be denoted by p and L , respectively. Thus, for this particular example, $p=1$ and $L=10$ s. The signal is sampled at 102.4 Hz ($f_s=102.4$ Hz), yielding $L=1024$ sampling points.

Thus, two triplets of quartiles compose the novel SDO: one for the time domain $[Q_1^T, Q_2^T, Q_3^T]$ and the other for the frequency domain $[Q_1^F, Q_2^F, Q_3^F]$.

The quartiles related to the time domain are computed over the statistical distribution of the amplitude values of the raw signal. The upper-right plot of Fig. 5.1 shows a histogram of such a distribution along with the representation of the three correspondent quartiles. This exemplifies how data fusion takes place, since only three values represent 1024 points.

Complementarily, the frequency content is evaluated through the Fast Fourier Transform (FFT). This present case has 1024 FFT points ($N=2^{10}$). The information concerning the energy distribution along the spectral range is expressed by its three quartiles

(bottom-left plot of Fig. 5.1). Thus, besides its fusion power (1024 values down to just 3), such quartiles also allow the SDO to be sensitive to frequency shifts and redistributions of the vibration energy. In other words, the quartiles divide the spectrum into four regions with 25% of vibration energy each (analogy to the area under the curve). If the peaks are too sharp, the quartiles tend to fall near them.

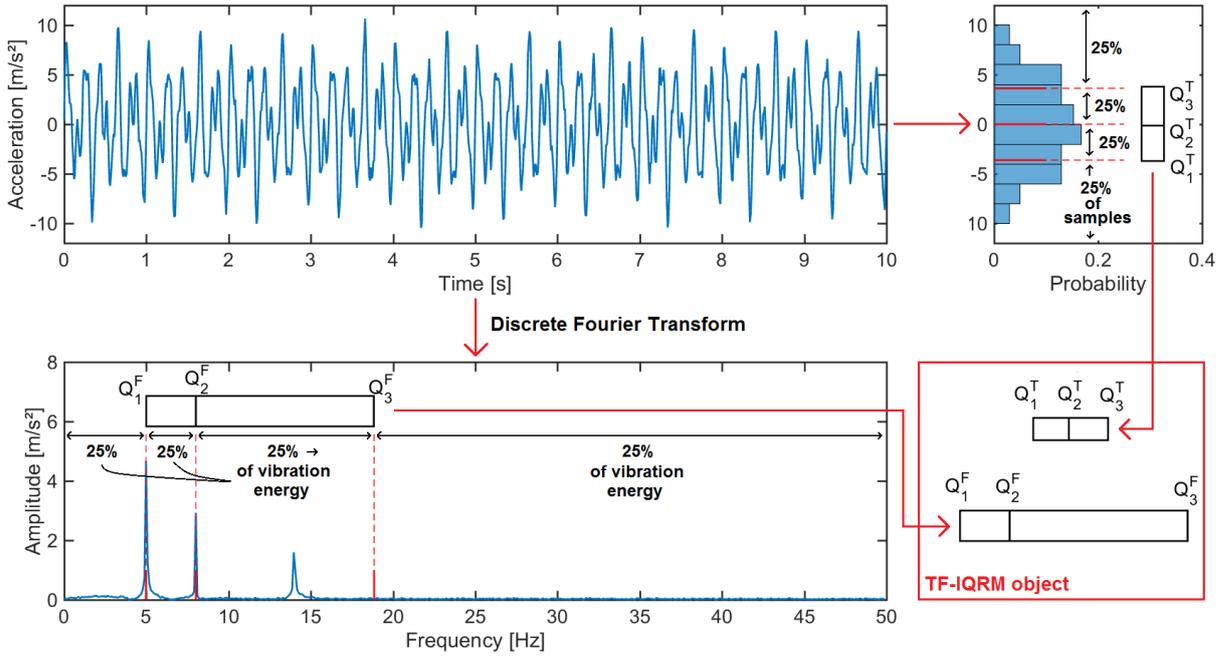


Fig. 5.1 The process to obtain the proposed SDO.

Thus, for this example, the original signal containing 1024 points is symbolically represented by only six values comprising both time and frequency aspects, which constitute the TF-IQRM object.

For the case of a set of signals with an arbitrary number of channels p , the TF-IQRM object is represented by two triplets of p -dimensional vectors: one triplet for the time domain and another for the frequency domain, yielding six p -dimensional vectors. Let $Q_{k,r}$ be the k -th quartile of channel $r=1, 2, 3, \dots, p$, where p is the total number of channels of the raw data. The vectors containing the quartiles information of all channels of test i are expressed in Eqs. (1) and (2). Note that the p -dimensional vectors \mathbf{L}_i , \mathbf{M}_i and \mathbf{U}_i gather the first, second and third quartiles, respectively, of each one of the p channels.

$$\mathbf{L}_i^T = \begin{bmatrix} Q_{1,1}^T \\ Q_{1,2}^T \\ Q_{1,3}^T \\ \vdots \\ Q_{1,p}^T \end{bmatrix}, \mathbf{M}_i^T = \begin{bmatrix} Q_{2,1}^T \\ Q_{2,2}^T \\ Q_{2,3}^T \\ \vdots \\ Q_{2,p}^T \end{bmatrix}, \mathbf{U}_i^T = \begin{bmatrix} Q_{3,1}^T \\ Q_{3,2}^T \\ Q_{3,3}^T \\ \vdots \\ Q_{3,p}^T \end{bmatrix}, \quad (1)$$

$$\mathbf{L}_i^F = \begin{bmatrix} Q_{1,1}^F \\ Q_{1,2}^F \\ Q_{1,3}^F \\ \vdots \\ Q_{1,p}^F \end{bmatrix}, \mathbf{M}_i^F = \begin{bmatrix} Q_{2,1}^F \\ Q_{2,2}^F \\ Q_{2,3}^F \\ \vdots \\ Q_{2,p}^F \end{bmatrix}, \mathbf{U}_i^F = \begin{bmatrix} Q_{3,1}^F \\ Q_{3,2}^F \\ Q_{3,3}^F \\ \vdots \\ Q_{3,p}^F \end{bmatrix}. \quad (2)$$

Geometrically, the TF-IQRM object consists in a pair of hyper-rectangles with a single point inside each one of them. The hyper-rectangles are related to both time and frequency domains. While the facets represent the interquartile range (IQR) of each channel, the point indicates the median (M). For the specific case of just two channels ($p=2$) its geometry collapses into a pair of rectangles with a single point inside, as shown in Fig. 5.2.

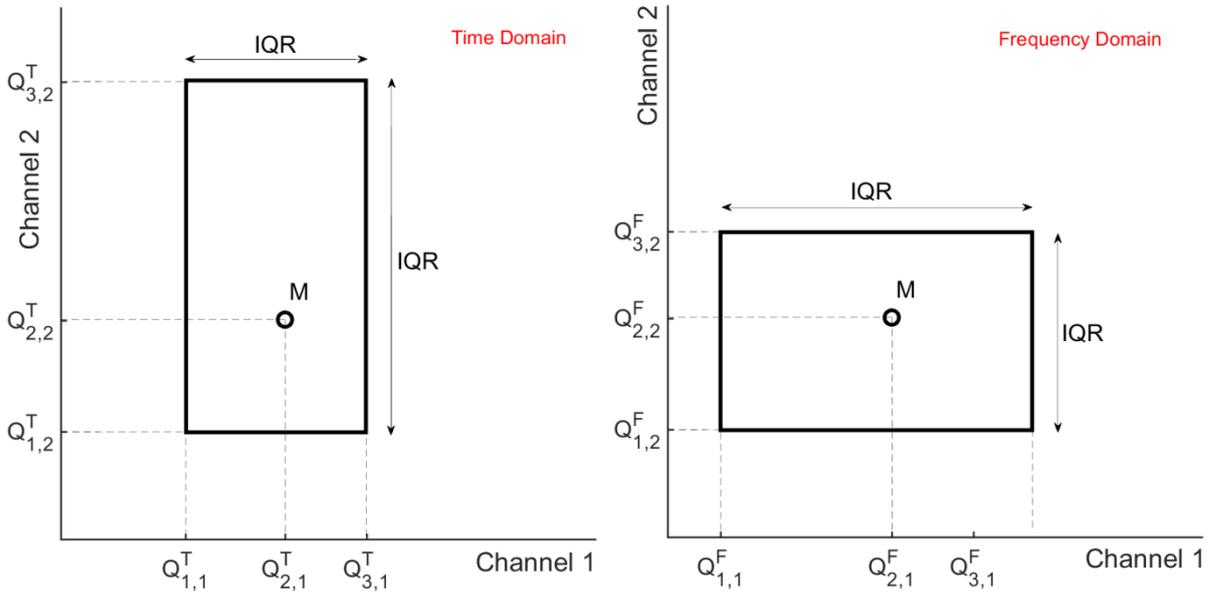


Fig. 5.2 Representation of a TF-IQRM object with two channels.

5.2.2 Dissimilarity between Symbolic Data Objects

The feature extraction procedure relies on the knowledge of the dissimilarity degree of patterns expressed in the SDOs. To do so, the distance measure $d_{i,j}$ between the TF-IQRM objects i and j is defined by an original symbolic metric:

$$d_{i,j} = \frac{\|\Delta_{i,j}^T\|}{\max(RMS_i, RMS_j)} + \frac{\|\Delta_{i,j}^F\|}{N/2}, \quad (3)$$

where N is the number of FFT points ($N/2$ is to account for the half-spectrum); \overline{RMS}_i is the mean of the Root Mean Square values of all p channels of object i ; $\|\bullet\|$ is the Euclidean norm of vector \bullet ; $\Delta_{i,j}^T$ and $\Delta_{i,j}^F$ are the distance vectors, defined as:

$$\Delta_{i,j}^T = \begin{bmatrix} \|\mathbf{L}_j^T - \mathbf{L}_i^T\| \\ \|\mathbf{M}_j^T - \mathbf{M}_i^T\| \\ \|\mathbf{U}_j^T - \mathbf{U}_i^T\| \end{bmatrix}, \quad (4)$$

$$\Delta_{i,j}^F = \begin{bmatrix} \|\mathbf{L}_j^F - \mathbf{L}_i^F\| \\ \|\mathbf{M}_j^F - \mathbf{M}_i^F\| \\ \|\mathbf{U}_j^F - \mathbf{U}_i^F\| \end{bmatrix}. \quad (5)$$

The distance metric shown in Eq. (3) comprehends the sum of two dimensionless fractions. While the first one accounts for the dissimilarity in the time domain, the second one accounts for the dissimilarity in the frequency domain.

The quartiles of the frequency spectrum (Eq. (2)) are considered in this work as being the indexes of the FFT coefficients and, therefore, are dimensionless. One could alternatively consider such quartiles as actual frequency values (in Hz). However, in this case, the sampling frequency (f_s) should be used instead of N in Eq. (3). Thus, the distance metric would remain compatible in terms of units. Moreover, the proposed distance measure fulfills the requirements for being a valid symbolic metric, which are:

- Symmetry: the metric always generates a symmetric dissimilarity matrix, i.e. $d_{i,j} = d_{j,i}$ for any i and j ;
- Non-negative distances: the metric always provides values equal or greater than zero;
- The metric always provides a null distance measure when a SDO is compared to itself, i.e. $d_{i,j} = 0$ when $i = j$, which generates zeros in all diagonal elements of the distance matrix.

A graphical description of the proposed distance metric is depicted in Fig. 5.3 considering a 2-channel signal. While Fig. 5.3a shows two objects with their related vectors,

Fig. 5.3b shows the graphical interpretation of the distance vector involved. Indeed, the relationship is the same either for the time or the frequency hyper-rectangles (or rectangles in this case), being expressed by Eqs. (1) and (3) – time – and by Eqs. (2) and (4) – frequency.

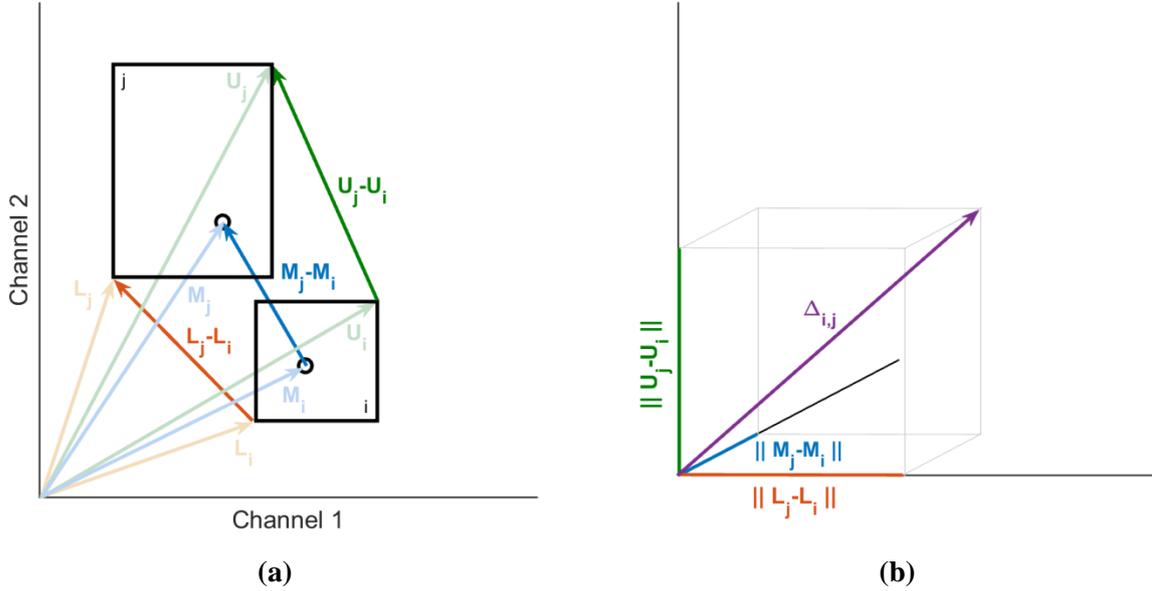


Fig. 5.3 Graphical interpretation of the proposed metric. (a) Pair of TF-IQRM objects and their relationship; (b) distance vector.

5.2.3 Prototypes' choice

The proposed methodology encompasses a procedure of pattern recognition within a time-window, which is based on partitioning the SDOs according to their similarity. Such a procedure is carried out by applying the k -medoids clustering technique^[23], which has shown to be more robust than k -means and hierarchical clustering algorithms^[17].

Within a time-window, the k -medoids algorithm gathers the SDOs into groups with high similarity expressing the same pattern or behavior. Each group is represented by one prototype^[23], which is the cluster's medoid SDO (i.e., the object that has the minimum sum of distances to the rest of other intra-cluster objects).

To perform the prototype's choice, the clustering algorithm needs two input quantities: the dissimilarity matrix and the number of partitions k . The former is obtained from Eq. (3) including all possible pairs of SDOs within a time-window. The latter is automatically chosen by the clustering validity index Calinski-Harabasz^[24].

5.2.4 Extraction of unidimensional feature

The last step for the feature extraction procedure is to obtain a unidimensional quantity capable of measuring the degree of novelty present within a time-window. Such a feature is called *Novelty Index* (NI) and is herein defined as the largest distance between all possible pairs of prototypes by means of the distance matrix already defined. Hence, considering a time-window containing S SDOs, the NI is computed as:

$$\text{NI} = \max(d_{i,j}) \quad | \quad i, j = t_1, t_2, \dots, t_k \quad (6)$$

where $d_{i,j}$ is the distance calculated using Eq. (3) and t_1, t_2, \dots, t_k are the indexes of SDOs assigned as being the prototypes of each one of the k clusters, with $k \in [2, S-1]$.

It is worth mentioning that the NI values are always positive, since they are directly obtained from distance values.

5.2.5 Schematic overview

To better illustrate the entire process to obtain the NI value, let us consider the following example: Figure 5.4 shows a time-window of a two-channel signal containing 10000 raw data points (1). The procedure starts with the extraction of the SDOs, as presented in section 2.1. In this example, the signal is represented by 5 SDOs ($S = 5$), each one with 2000 points ($L = 10000/5 = 2000$) (2). In the sequence, the dissimilarity between the SDOs is computed according to section 2.2 to generate a 5x5 distance matrix (3), depicted as a gray-scaled image. The closer to white, the higher the distance between two SDOs. At this point, the choice of prototypes is performed. To do so, the k -medoids clustering technique is applied as established in section 2.3. In this case, $k=2$ and the chosen prototypes are the second and fourth SDOs (4). Finally, the NI extraction is performed by evaluating the largest distance between the prototypes as explained in section 2.4. In this case, as there are only two prototypes, the distance between them is the NI itself (5).

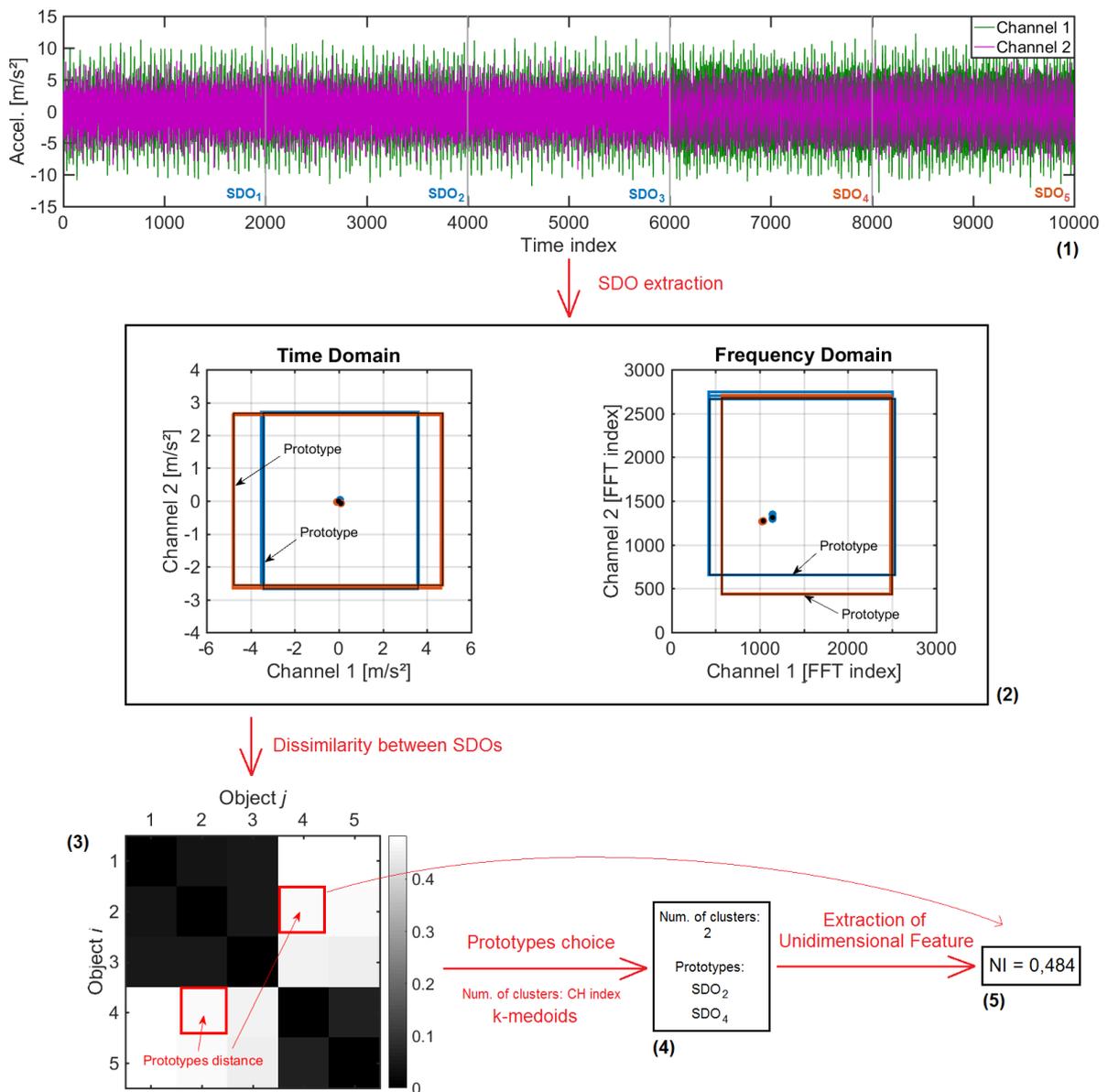


Fig. 5.4 Schematic overview of the unidimensional feature extraction.

5.3 Feature Classification

A powerful data fusion was achieved (as shown in section 2) by extracting the TF-IQRM objects from raw data and proceeding to the NI computation. Thus, a time-window comprising thousands of points is represented by a single value, which is proportional to the novelty degree. However, this isolated value is insufficient to establish precisely whether there is any structural change or damage. For instance, the time-window presented in Fig. 5.4 has a NI value equal to 0,484. Is that feature pointing out to any damage (or novelty) within the analyzed time-window? The answer is provided by what is called *feature classification*.

5.3.1 Confidence boundary

Similar to what was presented in ^[17], this paper suggests the tracking of successive NI values as the time-window advances over time. Hence, as a NI series is generated, an outlier analysis of their values is performed. If the newly evaluated NI is greater than the upper confidence boundary (named CB), a novelty detection is triggered.

To keep the unsupervised premise, such confidence boundary is not defined through the knowledge of a long undamaged baseline period. Instead, this limit is dynamically defined within each time-window by considering the last S values of NI as samples. It is assumed that the undamaged behavior would generate a t-Student distribution with $S-1$ degrees of freedom of NI values, instead of normal distribution, due to the small sampling population.

Thus, the upper confidence boundary defined within the time-window with index TW is defined as:

$$CB_{TW} = E[NI]_{TW} + t_{[S-1, 99.9\%]} \times \frac{E[NI - E[NI]_{TW}]_{TW}}{\sqrt{S}}, \quad (7)$$

where $E[NI]_{TW}$ is the expected value of NI within the time-window TW ; $t_{[S-1, 99.9\%]}$ is the 99.9 percentile of a t-Student distribution with $S-1$ degrees of freedom and $E[NI - E[NI]_{TW}]_{TW}$ is the variability, i.e. the expected value of the variation of NI values from its expected value $E[NI]_{TW}$.

For robustness, the median (or the 50th percentile) estimator is used as the expected value of the NI population once it achieves the maximum possible breakpoint i.e., 50%. In the same way, a robust estimator for the variability is adopted as being the S_n estimator proposed by Rousseuw and Croux ^[25]. Such assumptions can be formally described as:

$$E[NI]_{TW} = \text{med}_i(NI_i) \quad ; \quad i = TW - S + 1, \dots, TW \quad (8)$$

and

$$E[NI - E[NI]_{TW}]_{TW} = S_n = 1.1926 \text{ med}_i \left\{ \text{med}_j |NI_i - NI_j| \right\} \quad ; \quad i, j = TW - S + 1, \dots, TW \quad (9)$$

where $\{ \}$ stands for an array and 1.1926 is a factor defined in [25] with the objective of making this estimator consistent with Gaussian populations.

Hence, by substituting the statements of Eqs. (8) and (9) into Eq. (7), one obtains:

$$CB_{TW} = \overbrace{\text{med}_i(NI_i)}^{\text{Expected Value}} + t_{[S-1, 99.9\%]} \times \overbrace{\frac{1.1926 \text{ med}_i \left\{ \text{med}_j |NI_i - NI_j| \right\}}{\sqrt{S}}}^{\text{Variability}} ; i, j = TW - S + 1, \dots, TW \quad (10)$$

5.3.2 Detection index

Finally, the Detection Index (DI) related to newly acquired SDO with index TW is defined as:

$$DI_{TW} = NI_{TW} - CB_{TW} \quad (11)$$

Therefore, if DI_{TW} is positive, a damage (or structural novelty) is said to have occurred in the SDO with index TW . Thus, the alarm is expected to sound almost in real-time, with the maximum delay equal to L , which is the length of the newly acquired SDO .

The time delay until the first DI computation is called *initialization period*, rather than baseline period. It has the duration of $(S+1)L$ and is further explored in the following section.

5.3.3 Overall scheme of methodology

This section provides an illustrative example to clarify the concept behind both CB and DI values. It also summarizes the entire methodology for damage detection proposed in this paper.

Let us consider the same case showed in section 2.5 under continuous monitoring, generating a sequence of time-windows. This provides, purposely, a series of NI values. At each NI computation described in section 2, a CB value is also computed using the “memory” of the last S NI values, as shown in section 3.1. Whenever NI surpasses CB, damage is immediately characterized (DI positive), as described in section 3.2.

The history of such a continuous monitoring is depicted in Fig. 5.5, where time-windows comprising five 2-channel SDOs of 2000 points ($S=5$ and $L=2000$) are used to perform a clear damage detection, which occurred immediately after the acquisition of the 11th SDO. It is worth noting that the first NI value is computed only when the time-window accumulated S SDOs ($TW=5$, displayed in red). From that moment, after L/f_s seconds, a new SDO is extracted and a new time-window is defined ($TW=6$, displayed in green). Again, a

new value of NI is computed. However, since two values are available, a CB and a DI are also calculated ($TW=6$).

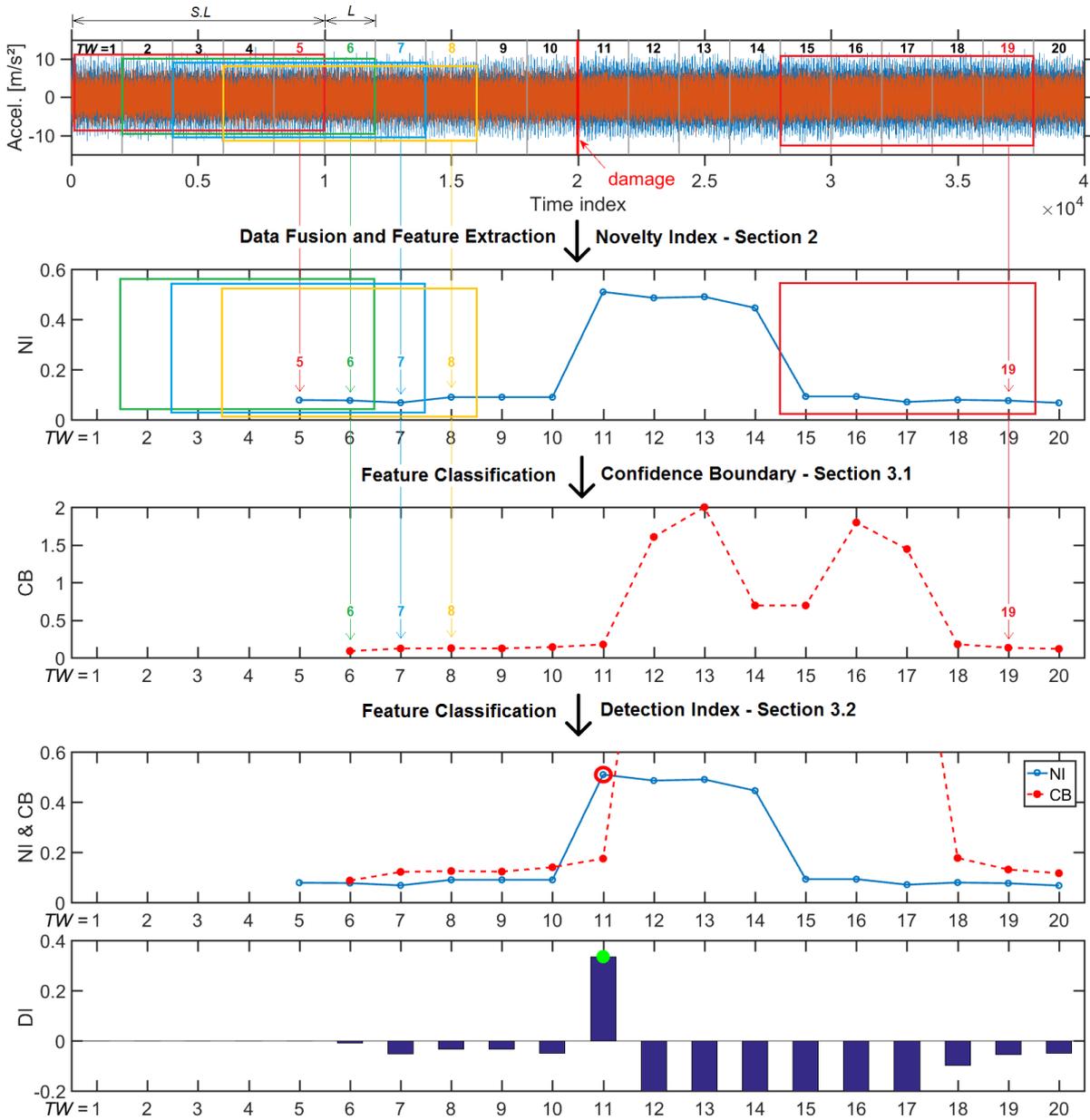


Fig. 5.5 Schematic overview of the entire methodology.

By means of the monitoring history of Fig. 5.5, it is possible to see that the Novelty Index grows so sharply right after the acquisition of the 11th SDO that it falls out of the 99.9% confidence interval, which alarms for a new behavior (delayed L/f_s seconds from the actual occurrence). After $S-1$ new SDOs ($TW=15$), when the old behavior is no longer within the time-window, the NI drops back to its usual values. This shows the adaptiveness and responsiveness of the proposed technique.

5.4 Application 1: Yellow Frame

5.4.1 Test description

The tested structure, named *Yellow Frame*, shown in Fig. 5.6, is a 4-story, 2-bay by 2-bay steel-frame scale-model (1/3) located in the Earthquake Engineering Research Laboratory at the University of British Columbia (UBC), in Canada. The structure was mounted on a concrete slab just outside of the structural testing laboratory on the UBC campus to simulate typical ambient vibration conditions. The nine columns are bolted to a steel base frame encased in the concrete slab. The experimental data described and used in this section are publicly available at [26].



Fig. 5.6 Tested frame structure: (a) braced and (b) unbraced.

The structure is 2.5 m \times 2.5 m in plan and is 3.6 m tall. The members are hot-rolled, grade 300W steel (nominal yield stress 300 MPa). The sections are specifically designed for this scale model test structure. The columns are B100x9 sections and the floor beams are S75x11 sections. Figure 5.7 shows the typical beam-column connection and the bracing system. Figure 5.8 presents the plan view and the east view of the structure.



Fig. 5.7 Zoom at (a) bracing and (b) mass placement.

In each bay, the bracing system consists of two ½-inch diameter steel rods placed parallel along the diagonal. To make the mass distribution reasonably realistic, one floor slab is placed in each bay per floor: four 1000 kg slabs at each of the first, second and third levels, four 750 kg slabs on the fourth floor (see Fig. 5.7b and 8a).

Since the structure’s establishment, several experiments were carried out, including: ambient vibration tests, impact tests and shaker tests. To assess the capability of the proposed approach applied to continuous monitoring, only the ambient vibration tests were used. Therefore, the excitations were due to wind, pedestrians and traffic in the vicinities of the structure. Further details about the experiment setup can be found in [26] and [27].

Fifteen accelerometers (FBA and EPI sensors, 0-50 Hz frequency range, sensitivity of 5 V/g) were placed throughout the frame, three by each floor including the ground level. Such transducers were placed so as they could measure the motions in all directions and also torsional modes (Fig. 5.8a). Anti-aliasing filter cutoff of 50 Hz was used and the data was sampled at 200 Hz.

To create different scenarios, two structural conditions were considered: braced and unbraced (Fig. 5.6). All possible cases are described in Table 5.1. Damage cases (1 to 6) were imposed to the braced structure by removing or placing braces gradually as shown in Fig. 5.9. In order of acquisition, cases 1, 5, 4, 3 and 2 simulate gradual damage on the bracing system. Finally, case 6 corresponds to the repair of several braces and damage of some others in a different face.

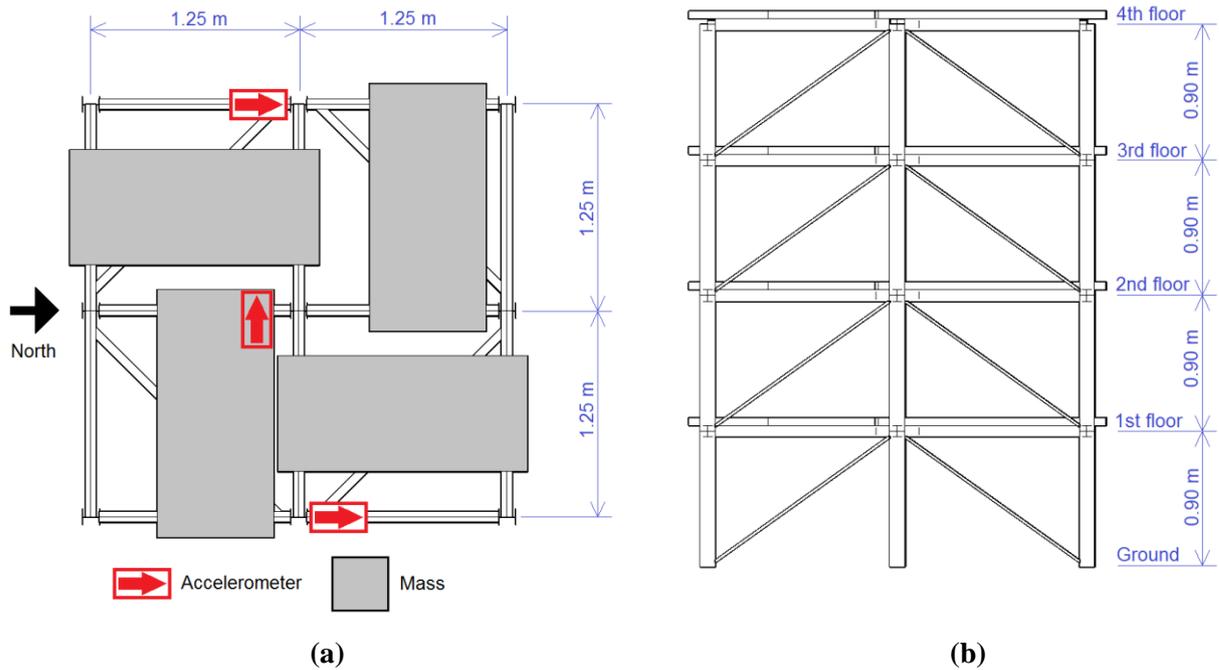


Fig. 5.8 Tested structure: (a) plan view and (b) east view.

Table 5.1: Tested configurations.

Case	Configuration
1	Fully braced configuration.
2	All east side braces removed.
3	Removed braces on all floors in one bay on southeast corner.
4	Removed braces on 1st and 4th floors in one bay on southeast corner.
5	Removed braces on 1st floor in one bay on southeast corner.
6	Removed braces on 2nd floor on north face.
7	All braces removed on all faces.
8	Configuration 7 + loosened bolts on all floors at both ends of beam on east face, north side.
9	Configuration 7 + loosened bolts on floors 1 and 2 at both ends of beam on east face, north side.

The unbraced frame had gradual damage simulated by loosening the bolts at beam-column connections, as depicted in Fig. 5.10 and 5.11. Cases, 7, 9 and 8, in this order, define the progression of damage.

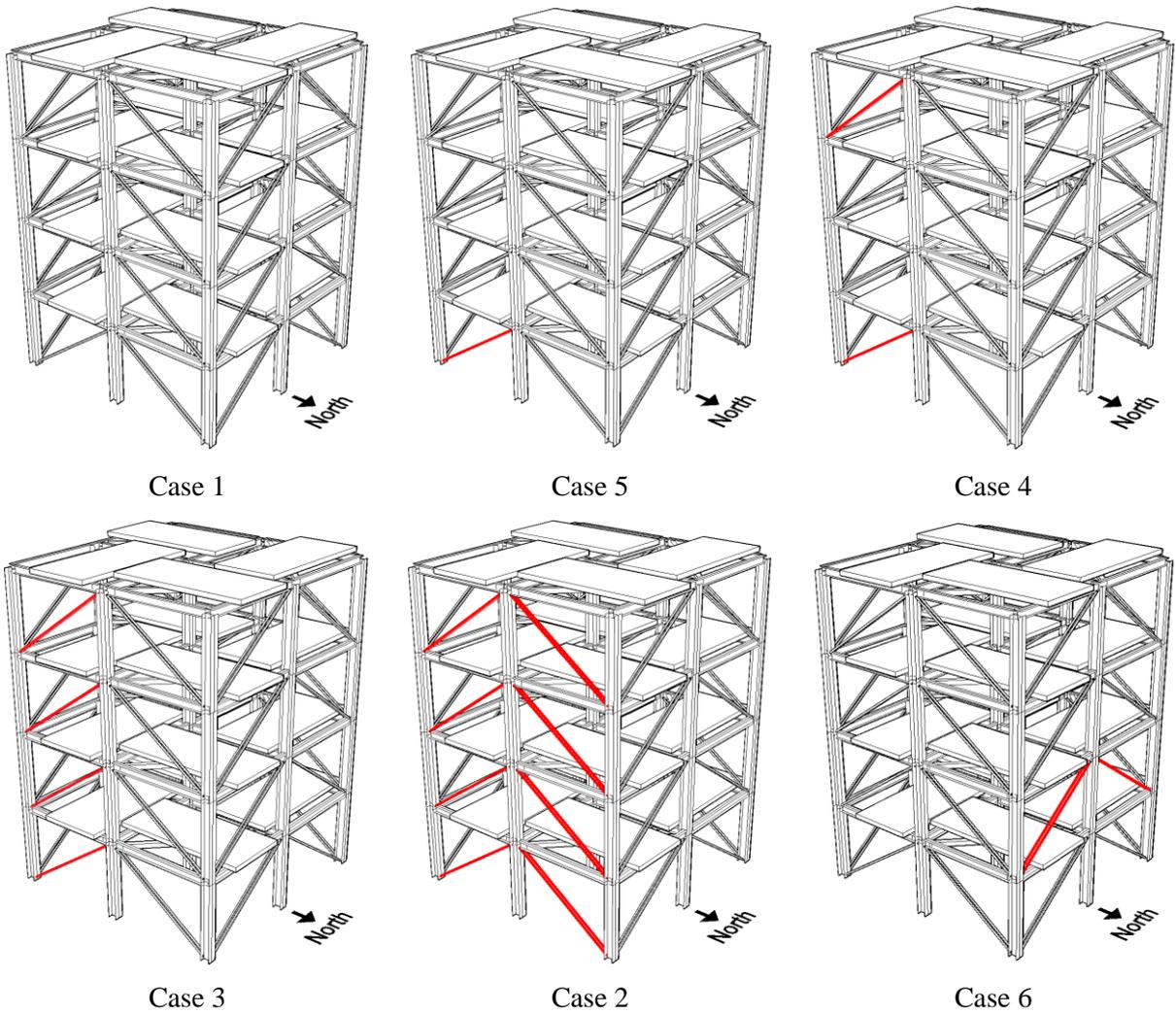


Fig. 5.9 Braced structure: cases 1 to 6.

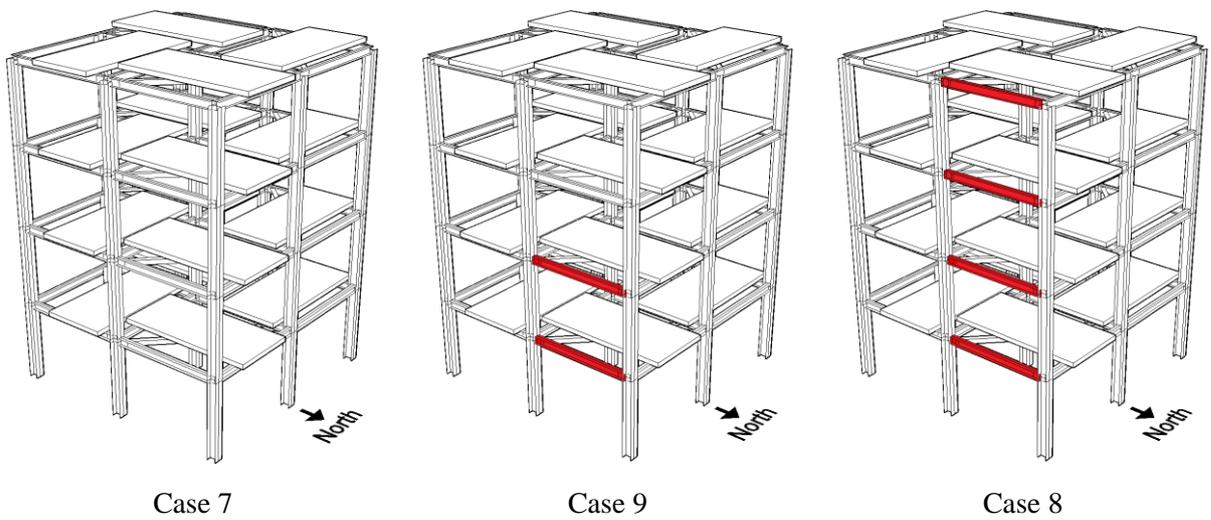


Fig. 5.10 Unbraced structure: cases 7 to 9.



Fig. 5.11 Photograph showing typical loosened bolts at beam-column connection (cases 8 and 9).

5.4.2 Results

In both tested structures, braced and unbraced, the number of SDOs per time-window and the number of terms of FFT is set to $S=5$ and $N=2^{13}$, respectively.

Firstly, the braced structure was subjected to continuous monitoring and had its data processed by the proposed technique. Since each damage case is 60000 points (300 s) long (except for the case 6: 30000 points), a time-window containing half of this length was considered reasonable. Thus, the SDO length was set to $L=6,000$ (30 s), so as to generate a time-window with length $S.L=30000$ points (150 s). The results of such continuous monitoring are depicted in Fig. 5.12.

The first plot of Fig. 5.12 shows the history of the acceleration signals (15 channels each). Magenta vertical lines indicate the limits between cases, i.e. inflicted damage. It is worth mentioning that the wide variation in terms of response amplitudes is mainly due to different levels of ambient excitation of the tested scenarios (more or less wind and traffic).

While examining the second plot, it is interesting to notice the adaptiveness of both NI and CB values by observing their history. For instance, the NI values increase sharply just after the damage occurrences and then quickly decrease after a certain period. Also, note that the damage occurrences are revealed in the NI series as well-defined “peaks”.

The third plot shows the DI series. Again, the method correctly detects four out of five inflicted damage with the time delay of just 30 seconds. Besides, no false alarms were

sounded. Only the first damage was not detected due to the small height of the NI “peak” (TW=11 to 14) in comparison to the large fluctuations in NI values before the damage occurrence, which increased the CB.

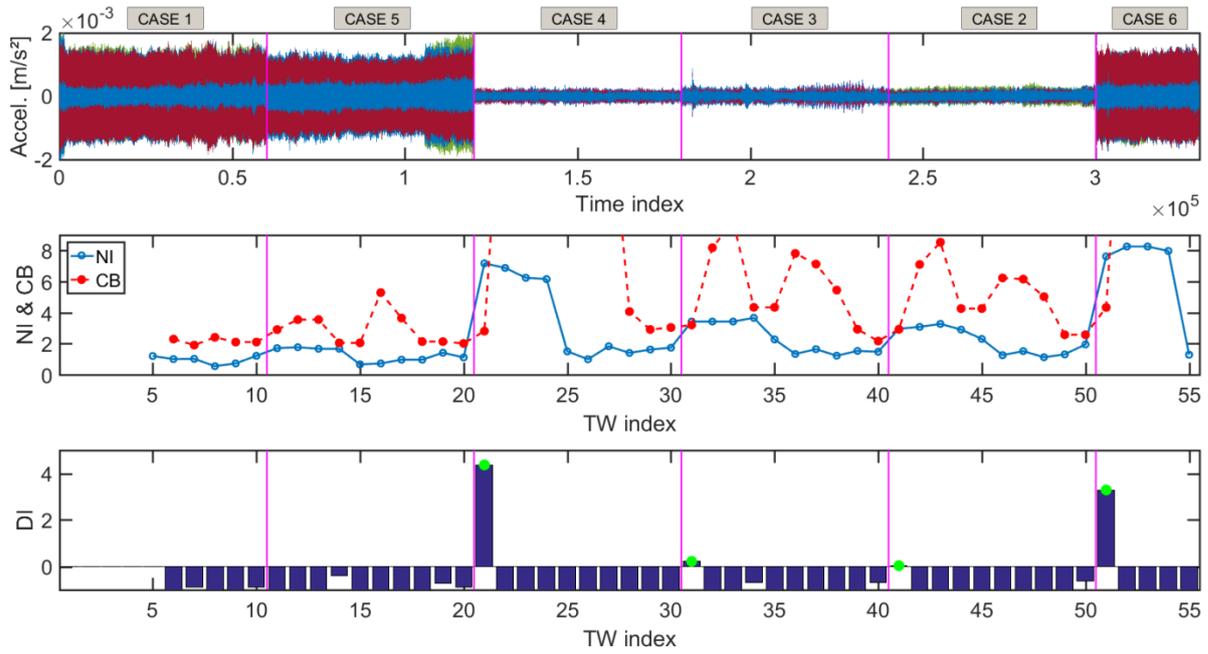


Fig. 5.12 Damage detection of braced structure, using $S=5$ and $L=6000$.

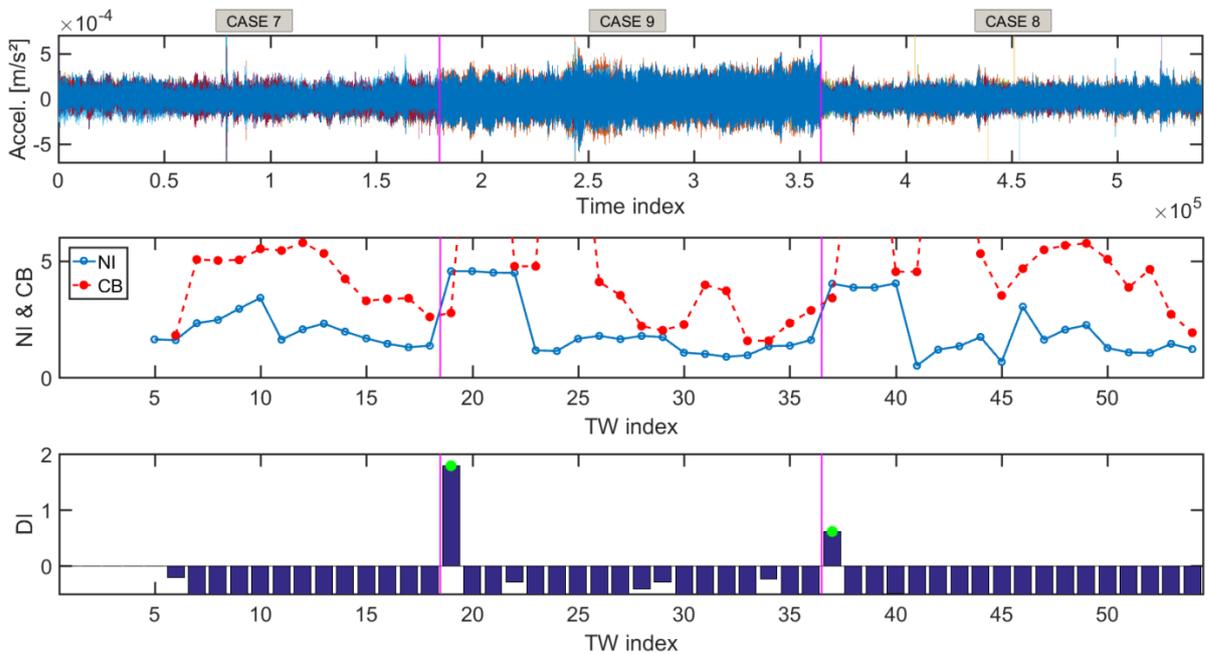


Fig. 5.13 Damage detection of unbraced structure, using $S=5$ and $L=10000$.

Similarly, the unbraced structure was subjected to continuous monitoring, but with the object length set to $L=10,000$. The results can be seen in Fig. 5.13. The two cases of stiffness loss in beam-column connections simulated by loosening bolts were clearly detected, as depicted. No false alarms were sounded.

Despite the encouraging performance of the proposed approach, the choice of the SDO length can still be tuned to improve damage detection. To explore this topic, the following section presents a sensitivity analysis.

5.4.3 Sensitivity analysis

The length of the SDO affects the quality of damage detection. Figures 5.14 and 5.15 show, respectively, the results obtained from the two structures using different object lengths. Complementarily to what was previously presented, the tested lengths are: 1000, 3000, 6000 and 10000 points, corresponding to 5, 15, 30 and 50 seconds, respectively.

It is noticeable that false alarms (red crosses) are more prone to occur as the SDO length decreases. Such a phenomenon is quite natural, since the SDO is carrying little information. As a consequence, the use of short SDOs leads to many alarms (positive DIs), be they real or false. However, this type of SDOs has the advantage of better accusing the occurrence of damage in a shorter time delay, almost instantaneously.

Conversely, the use of longer SDOs leads to more accuracy, i.e. the rate of real alarms compared to false alarms rises significantly. However, as they surpass a certain length, which is difficult to determine, they tend to be prone to miss small damage scenarios.

These explanations highlight the essence of the behavior of the detection method when one adopts different SDO lengths. In some way, this “logic” helps guiding the user. Nevertheless, even if the user has doubts about the definition of such a parameter, fair results can be achieved, as already shown in Figs 5.12 to 5.15.

The other parameters, i.e. the number of SDO per time-window (S) and the number of FFT terms (N) are kept fixed for all method applications and, therefore, do not need to be tuned.

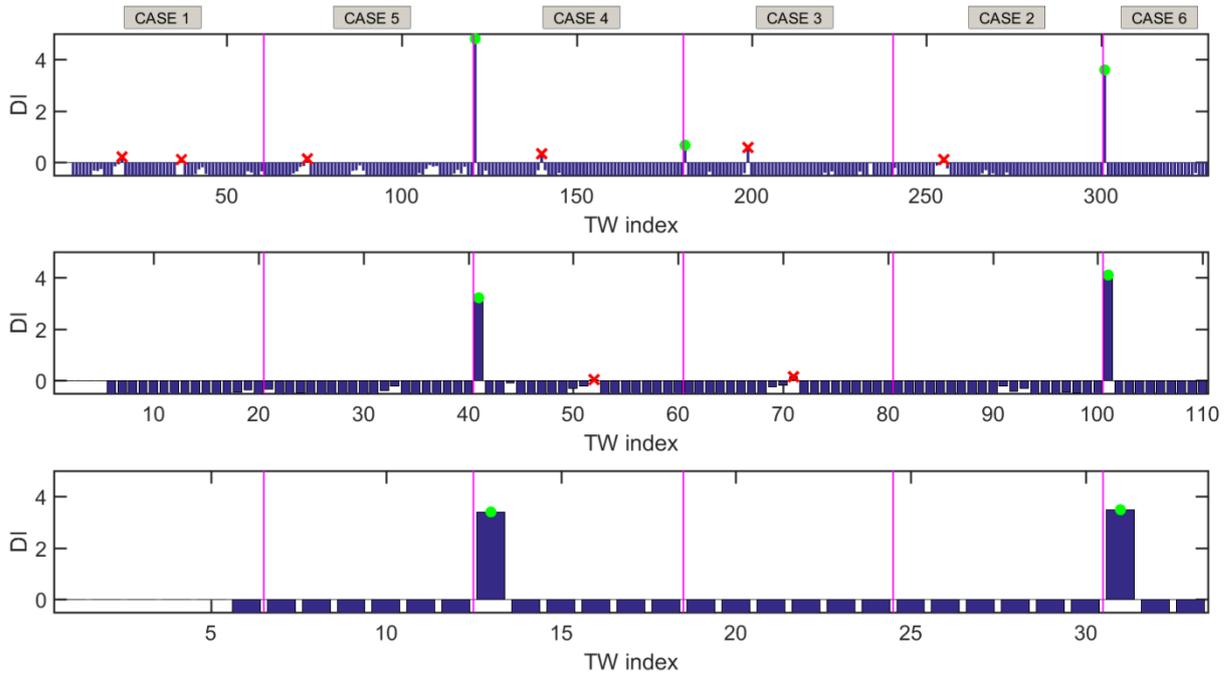


Fig. 5.14 Damage detection of braced structure, using $S=5$ and $L=1000, 3000$ and 10000 , respectively.

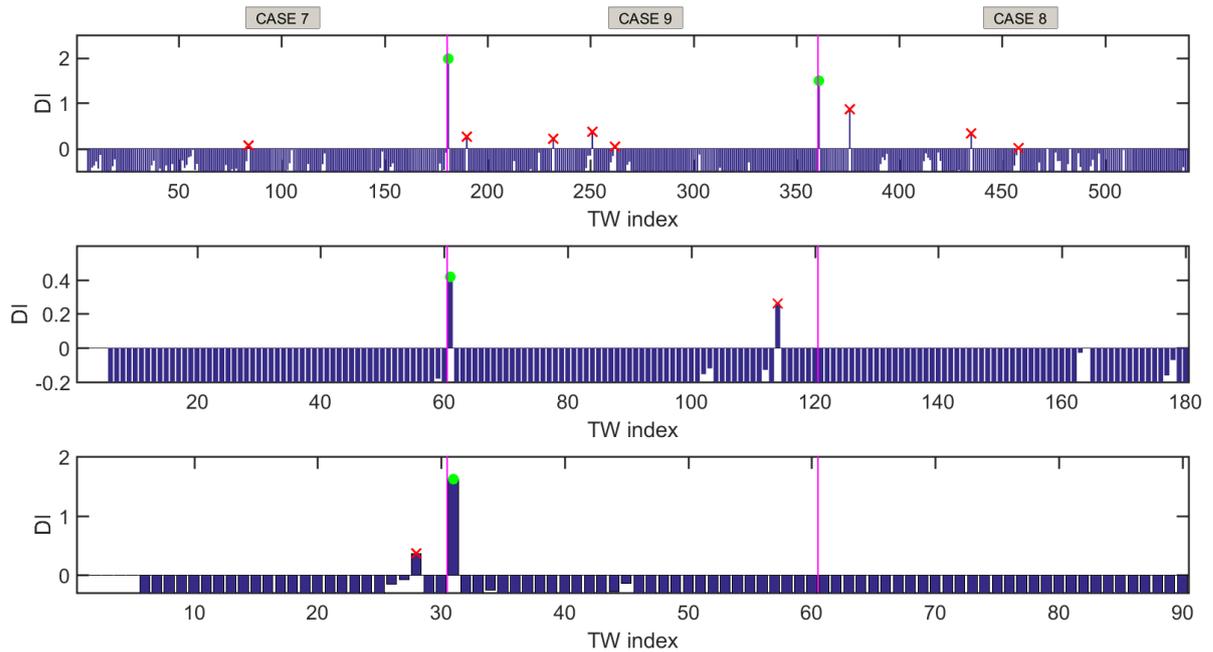


Fig. 5.15 Damage detection of unbraced structure, using $S=5$ and $L=1000, 3000$ and 6000 , respectively.

5.5 Application 2: Concrete Bridge

The PI-57 Bridge is a double-deck bridge located near the town of Senlis in France, crossing the Oise River and carrying the A1 motorway, which connects Paris to Lille (Fig. 5.16). The bridge, a 116.50 m long, cast-in-place, post-tensioned segmental structure built in 1965,

consists of three continuous spans of 18.00, 80.50 and 18.00 m. The two lateral spans play the role of counterweights.



Fig. 5.16 Monitored bridge, over Oise River, France.

This structure encountered several problems during and after construction, causing localized cracks and increasing deflection in the middle of the span. These problems were mainly due to insufficient prestressing caused by the effects of shrinkage, creep and the thermal stresses that were not taken into account during the design and construction (the knowledge of this type of constraint was limited at the time of construction). Due to the potential risk of cracks in the deck, numerical studies have shown that the long-term safety of the structure could have been affected if a corrective action was not taken quickly. Based on these technical assessments and considering the importance of the structure, SANEF (Société des Autoroutes du Nord et de l'Est de la France) decided to reinforce the two decks. Thus, two acquisition campaigns have taken place: one before the reinforcement, between October 2008 and April 2009, and the other after the reinforcement, between October 2009 and April 2010.

The monitoring system counted only on ambient vibrations, due to wind and traffic. Sixteen piezoelectric accelerometers (Bruel & Kjaer 4507B-005 with sensitivity of 1 V/g, frequency range from 0.4 to 6000 Hz, maximum operational level of 75 g, temperature range from -54 to +100 °C) have been installed on the bridge deck (Lille/Paris – Fig. 5.17). For the acceleration recording, a data programmable controller Gantner E-PAC DL was used and connected to an 8 GB USB flash drive. Data were transferred by a TCP/IP modem. Accelerations were filtered within the 0–30 Hz frequency range and sampling was set to 0.004 s (250 Hz). Every hour a 4-minute signal is acquired.

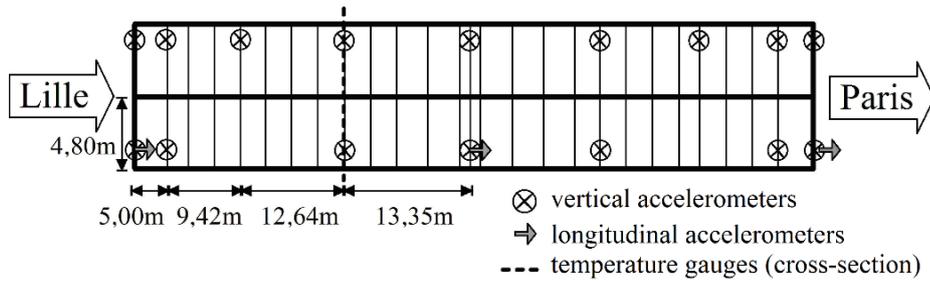


Fig. 5.17 Instrumentation plan.

In this case, the length of the SDO is equal to $L=60000$ (4 min), which corresponds to an entire signal. The number of SDOs per time-window is kept unchanged ($S=5$) and the same is done with the number of FFT points ($N=2^{13}$). After 64 days of monitoring, the obtained time series of Damage Index is plotted in Fig. 5.18.

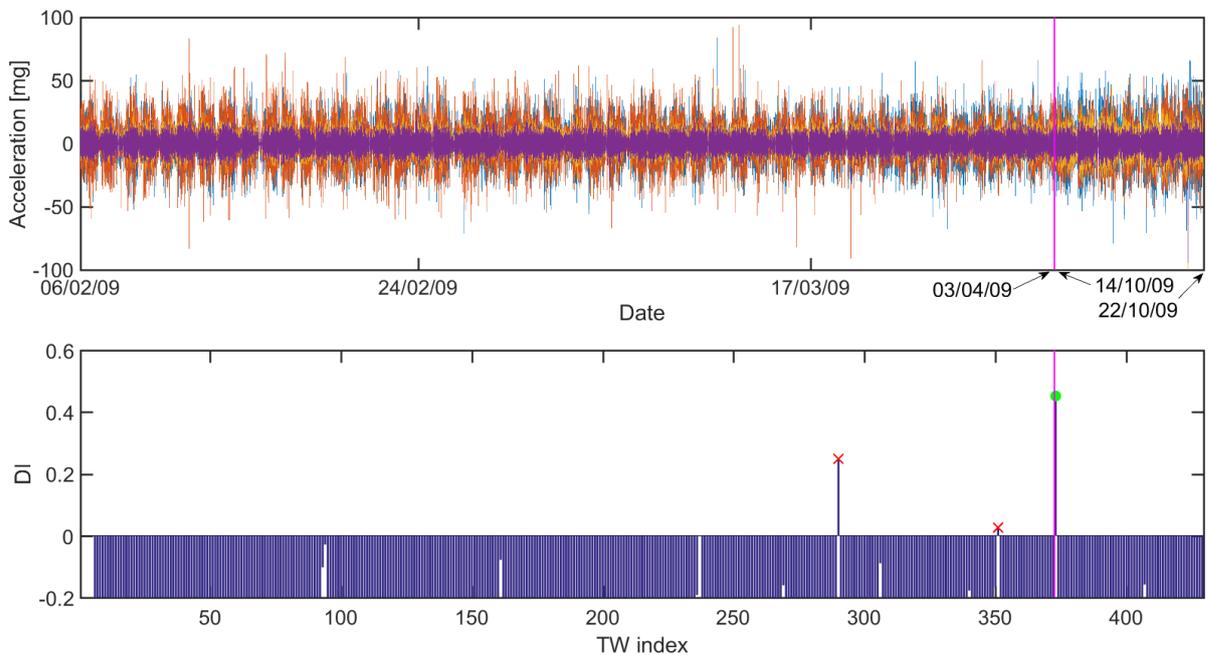


Fig. 5.18 Time series of raw data and Damage Index.

The upper plot of Fig. 5.18 shows the time-history of channels 4 to 7 to illustrate the entire monitoring period. The magenta vertical line separates the periods before and after structural reinforcement, which occurred between April and October, 2009. By observing the DI series, one can notice that the proposed methodology detected the structural novelty just after the acquisition of the first signal from the reinforced bridge (green circle). Furthermore, from February 6 to October 22 (64 days of continuous monitoring), only two false alarms were sounded (red crosses).

Figure 5.19 shows a typical 4-minute acceleration signal registered by the 16 accelerometers. This is the typical package of information, obtained every hour, fused in the TF-IQRM symbolic object. It is important to note that the cycles of day-and-night, which impacts temperature and traffic levels (environmental factors), did not influence the methodology performance. This is a quite interesting result, since no signal normalization procedure is employed by the proposed technique (except for the use of the RMS value, implicit in the symbolic metric).

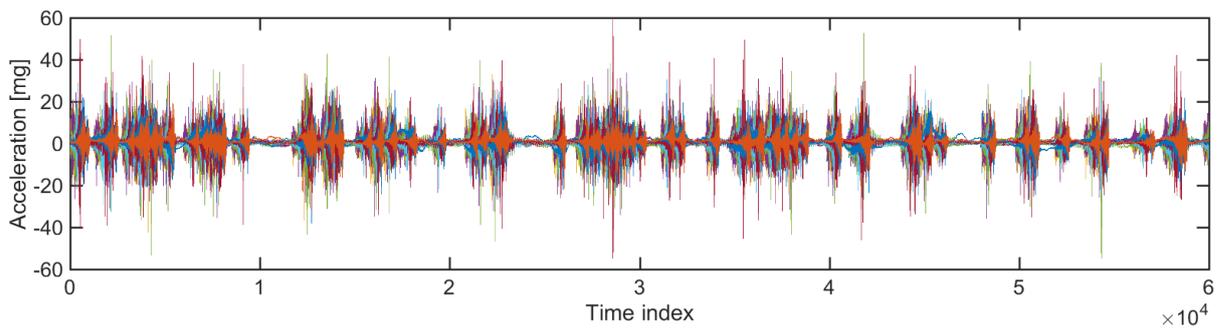


Fig. 5.19 Typical 4-minute signal acquired at every hour.

5.6 Conclusion

This paper presented a methodology able to carry out fully-automated real-time unsupervised damage detection on structures subjected to ambient vibration. The continuous acquisition of signals is transformed into packages of compact information expressed by the proposed novel symbolic object, called TF-IQRM. Such an original representation embraces both time and frequency responses of structural dynamic measurements. Then, these objects are used to feed a procedure of pattern recognition (k -medoids clustering) within a moving time-window to generate single-valued features (NI), which indicates the degree of structural novelty. By tracking the evolution of such an index, an outlier analysis is performed yielding the Detection Index (DI), which indicates the current structure's condition, i.e., damaged or undamaged.

The first application (*Yellow Frame*) assessed the sensitivity of the proposed algorithms as the user adopts different lengths for the symbolic objects. Furthermore, several scenarios of damage are imposed to the structure, some of which with local/small influence. Although the choice of the SDO length is optional, a clear directive instruction is given to the user: while short SDOs produce a more “alarmist” detection (conservative), long ones generate a more silent and assertive detection (lower rate of false alarms). The tuned choice of

this parameter led to impressive results, achieving the detection of 6 out of 7 damage occurrences, with no false alarms (see Figs. 5.12 and 5.13).

The second application presented a real full-scale structure. The method was capable of handling environmental factors, such as temperature, traffic, wind, among others. Even without an explicit normalization step, the algorithms precisely detected the structural novelty. Furthermore, the entire monitoring period of 64 days had only 2 false alarms, which is a quite interesting achievement.

Hence, since the proposed technique is also sensitive to small damage, usually invisible to human inspections, this method represents an important aid to maintenance schedules for preventive actions.

Acknowledgements

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for the financial support; special thanks goes to James Beck, Professor at the California Institute of Technology (USA) and to the IASC-ASCE Structural Health Monitoring Task Group for kindly providing experimental data from the benchmark steel frame at the University of British Columbia (Canada); the authors also thank IFSTTAR (former LCPC - Laboratoire Central des Ponts et Chaussées) and SANEF (Société des Autoroutes du Nord et de l'Est de la France) for the data used in this paper (*project 0560V407*).

References

- [1] J. Glanz et al. Genoa bridge collapse: the road to tragedy. *New York Times*, September 6, 2018.
<https://www.nytimes.com/interactive/2018/09/06/world/europe/genoa-italy-bridge.html>
- [2] X.G. Hua et al., *Structural Damage Detection of Cable-Stayed Bridges Using Changes in Cable Forces and Model Updating*, *Journal of Structural Engineering*, 2009, 135(9), pp.1093–1106.

- [3] F.L. Wang et al., Correlation-Based Damage Detection for Complicated Truss Bridges Using Multi-Layer Genetic Algorithm, *Advances in Structural Engineering*, **2012**, 15(5), pp.693–706.
- [4] B.H. Koh, S.J. Dyke, Structural health monitoring for flexible bridge structures using correlation and sensitivity of modal data, *Computers & Structures*, **2007**, 85(3-4), pp.117–130.
- [5] M. Fallahian,,F. Khoshnoudian, V. Meruane. Ensemble classification method for structural damage assessment under varying temperature. *Structural Health Monitoring*. **2018**, 17(4), pp.747-762.
- [6] A. Entezami and H. Shariatmadar. An unsupervised learning approach by novel damage indices in structural health monitoring for damage localization and quantification. *Structural Health Monitoring*. **2018**, 17(2), pp.325-345.
- [7] M. Vahedi, F. Khoshnoudian, T. Hsu. Application of Bayesian statistical method in sensitivity-based seismic damage identification of structures: Numerical and experimental validation. . *Structural Health Monitoring*. **2018**, 17(5), pp.1255-1276.
- [8] K. Nair, & A.S. Kiremidjian, Time Series Based Structural Damage Detection Algorithm Using Gaussian Mixtures Modeling, *Journal of Dynamic Systems, Measurement, and Control*, **2007**, 129(3), pp.285–293.
- [9] S. Silva et al., Structural damage detection by fuzzy clustering, *Mechanical Systems and Signal Processing*, **2008**, 22(7), pp.1636–1649.
- [10] M. Gul, F. Catbas, Statistical pattern recognition for Structural Health Monitoring using time series modeling: Theory and experimental verifications, *Mechanical Systems and Signal Processing*, **2009**, 23(7), pp.2192–2204.
- [11] E. Figueiredo, *Damage Identification in Civil Engineering Infrastructure under Operational and Environmental Conditions*. University of Porto, Faculty of Engineering, **2010**.
- [12] D. Posenato et al., Methodologies for model-free data interpretation of civil engineering structures, *Computers & Structures*, **2010**, 88(7-8), pp.467–482.

- [13] V. Alves, A. Cury, C. Cremona. On the use of symbolic vibration data for robust structural health monitoring, *Proceedings of the Institution of Civil Engineers - Structures and Buildings*, **2016**, 169(9), pp. 715-723.
- [14] S. Jin, H. Jung. Vibration-based damage detection using online learning algorithm for output-only structural health monitoring. *Structural Health Monitoring*, **2018**, 17(4), pp.727-746.
- [15] Y. Cha and Z. Wang. Unsupervised novelty detection-based structural damage localization using a density peaks-based fast clustering algorithm. *Structural Health Monitoring*, **2018**, 17(2), pp.313-324
- [16] R. Cardoso, A. Cury, F. Barbosa, A robust methodology for modal parameters estimation applied to SHM. *Mechanical Systems and Signal Processing*, **2017**, 95 (2017), pp.24–41.
- [17] J. Santos, *Smart structural health monitoring techniques for novelty identification in civil engineering structures*, Ph.D. thesis, Instituto Superior Técnico, Univ. of Lisbon, Lisbon, Portugal, **2014**.
- [18] J. Santos et al., Early Damage Detection Based on Pattern Recognition and Data Fusion, *Journal of Structural Engineering*, **2017**, 143(2).
- [19] J. Santos et al., Baseline-free real-time assessment of structural changes. *Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance*, **2015**, 11(2), pp.145-161.
- [20] V. Alves, A. Cury, N. Roitman, C. Magluta, C. Cremona. Structural modification assessment using supervised learning methods applied to vibration data. *Engineering Structures*, **2015**, 99, pp.439-448.
- [21] M. Ichino & H. Yaguchi. Generalized Minkowski metrics for mixed feature-type data analysis, *IEEE Trans. Syst. Man Cybern.*, **1994**, 24(4), pp.698–708.
- [22] P.J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, **1987**, (20), pp.53-65.
- [23] L. Kaufman, P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Hoboken, New Jersey: John Wiley & Sons, Inc, **2009**.

- [24] T. Calinski, J. Harabasz. *A dendrite method for cluster analysis*, Communications in Statistics, **1974**, Vol. 3, No. 1, pp.1–27.
- [25] P. Rousseuw & C. Croux. 1993. Alternatives to the Median Absolute Deviaton, *Journal of the American Statistical Association*, **1993**, 88(424), pp.1273–1283.
- [26] J. Dyke et al. *NEES: Database for Structural Control and Monitoring Benchmark Problems*, **2015**, <https://datacenterhub.org/resources/257>.
- [27] J. Dyke, D. Bernal, J. Beck, C. Ventura. Experimental Phase II of the Structural Health Monitoring Benchmark Problem, **2003**.

PART IV

CONCLUSION

Chapter 6

Achieved Milestones

Damage identification in Civil Engineering structures is crucial to avoid social, environmental and economic consequences. Thus, it is of paramount interest to develop methodologies able to detect, locate and quantify damage at early stages. Related studies in the literature show that a lot of research has been done trying to achieve these objectives. For most practical cases, however, the first step – detection – remained an ill-solved problem. Therefore, this thesis sought to provide reliable and robust tools for damage detection that were *unsupervised, automated* and suitable for *real-time* SHM.

Over the last decades, several approaches were proposed to detect damage based on the *identification* and *tracking* of structural modal parameters. Since the latter deeply relies on an accurate and automated estimation of such properties, the development of an automatic modal identification technique was the primary objective of this thesis. Thus, Part II of this thesis started with Paper #1 by presenting a technique for automated modal identification of structures. The developed algorithm depends on a few user-defined parameters that affect very little on the quality of the results. However, uncertainties regarding the choice of the order of the parametric model still exist. To circumvent this situation, additional research provided the development of a second technique, presented in Paper #2, with more automation and enhanced robustness. It was demonstrated that the modal parameters are indeed estimated accurately with no human intervention, including modes with very closely-spaced frequencies. Then, the modal parameters, which are the algorithm's output, can be used as input data for unsupervised real-time damage detection procedures in SHM.

Part III presented two techniques for damage detection based on raw data measurements. Paper #3 showed that damage could be detected in real-time, in an unsupervised and automatic way. Based on unsupervised pattern recognition of a new symbolic representation of raw time-domain data (IQRM object), the technique presents two different damage-sensitive indexes: one for short-term monitoring (FDI) and the other for long-term monitoring (SDI). Alternatively, Paper #4 proposes a novel kind of symbolic object

that represents both time and frequency aspects of raw data. Continuous monitoring of real structures attests the remarkable power of data fusion and representability provided by the TF-IQRM object. In both strategies, damage indexes are computed instantaneously at each object acquisition, in a moving time-window framework. Both approaches have also shown adaptiveness, i.e., after each detection of damage, the algorithm takes few instants to learn about the new structural condition and to be able to detect a new occurrence. A considerably low rate of false alarms was also achieved.

For all that has been presented in this thesis, the developed strategies consist in an essential step for practical SHM programs in civil engineering applications, for three main reasons:

- i. They are all based on output-only algorithms. Therefore, they do not need to have excitations measured or controlled - only Ambient Vibration Tests are required. This is especially the case of structures monitored under regular and continuous operational service;
- ii. They can detect early damage (Level 1 SHM) in real-time, automatically and without needing baseline periods;
- iii. They are insensitive to environmental/operational effects (temperature, wind, traffic, etc.). The occurrence of such phenomena did not significantly affect the performance of the algorithms during the studied practical applications.

For future works, some improvements regarding the detection techniques could be implemented. For instance, the superposition of several time-windows procedures, each one with a different object length running together at the same time would yield various results, which could pass through a further statistical analysis. It is expected that such strategy would reduce the number of false alarms and increase the accuracy and robustness of damage detection.

A convenient idea of improvement regarding the most recent technique (fourth paper) is to use the Power Spectral Density instead of the FFT coefficients for constructing the TF-IQRM objects. The PSD could also be filtered prior to the quartiles extraction. Such procedure could enhance the precision in quantifying the vibrational energy with less influence of noise.

Also, it is worth mentioning that the two modal identification techniques presented in the first and second papers are perfectly suitable to feed the detection techniques presented in the third and fourth papers. Such a combination can be certainly explored in the next step of this research.

Aiming at assessing a deeper knowledge about the detection sensitivity, numerical simulations of tall buildings and bridges could also be explored in the following years. Finally, in order to start applying the developed software into real cases, real structures can be monitored. To accomplish that, the main requirement is to set up an acquisition system able to export registered signals in real-time.

Regarding damage detection (SHM Level 1), other types of symbolic objects could also be proposed as well as different metrics and feature classification techniques. Of course, this subject is still an open line of research. However, since a lot of advances were already achieved by the present work, the bigger challenge relies on the next steps of the Rytter's classification. Therefore, for future works in Civil Engineering SHM, one can establish:

- Damage location (Level 2): through the use of spectral signatures and modal amplitudes. Such quantities provide more “localized” structural data and could help infer about damage location;
- Damage quantification (Level 3): by utilizing modal parameters and damage indicators coupled with unsupervised learning classification algorithms, such as *clustering methods* for determining damage scenarios.

“[...]

As I have said before **CONSCIOUSNESS is LIFE** and **LIFE is CONSCIOUSNESS**.

EVERYTHING in existence is defined by vibrational frequencies of consciousness. Light, Sound, Colour, all physical phenomena both living and inanimate. If you can change the frequencies of vibration of any single thing, you change the appearance of that thing – be it sound, colour, gas, liquid, physical organs.

When physical organs present an appearance of ill-health, it is because the normal frequencies of vibrations of that physical organ have been reduced and the LIFE within the organ has been depleted.

Science presents the universe as being ‘matter’ possessing consciousness but the truth is:

The universe is CONSCIOUSNESS which has taken on the appearance of ‘matter’ as a result of a descent into the lower frequencies of vibration of consciousness.

[...]”

Excerpt from the Christ's Letters - Letter 8