Politecnico di Torino

Politecnico di Torino

Master's Degree in Civil Engineering

Academic Year 2023/2024

Master's Degree Thesis

On the investigation of Statistical Alignment for enhancing damage identification across a population of heterogeneous shear structures

Supervisors

Prof. Cecilia SURACE

Eng. Giulia DELO

Candidate

Sebastian BADARIOTTI

To my family

Abstract

The development of machine learning algorithms for Structural Health Monitoring (SHM) is rapidly advancing. However, their application for real-world structures finds a high number of complications. One is the need for comprehensive data for training the proper algorithms. Thus, Population-Based Health Monitoring (PBSHM) overcomes these challenges by sharing information between different structures.

In this framework, it is necessary to understand to what extent knowledge can be shared, especially for heterogeneous datasets. Therefore, this study implements a simple domain adaptation technique based on Statistical Alignment (SA) on a population of heterogeneous shear structures to investigate how the performance changes due to the variations within the population. The scenarios proposed are solved with normal-condition alignment (NCA) and normal-correlation alignment (NCORAL).

Two case studies are analysed. The first is related to numerical structures. It is created by simulating multiple source and target datasets, containing the features and labels of each data point. The features consist of the natural frequencies of each structure, and the label is a binary vector indicating if the data point corresponds to a damage condition or not. To calculate the natural frequencies, the structure is modelled as a shear-type with chain-like models, and the mass and stiffness matrices are calculated considering the equation of motion. The damage is then introduced with a reduction of the stiffness of a column, leading to reduced values of the related frequencies. It is important to highlight that, in each sample, a variation of the material properties is introduced, trying to simulate the actual variability on measured data. The second case study extends the implementation to an experimental case study of a three-story frame structure to test this methodology for sharing knowledge between real and simulated data.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

The investigation of Statistical Alignment performance for enhancing damage identification across a population of heterogeneous shear structures, is a research topic within the field of Structural Health Monitoring (SHM).

SHM represents a critical discipline in the fields of Civil, Mechanical and Aerospace Engineering, focusing on the development and application of strategies and technologies to assess the condition of existing structures. Its primary aim is to ensure safety, enhance maintenance strategies, and extend the lifespan of structures such as bridges, buildings and other infrastructures. The evolution of SHM has been influenced by two major developments: the growing imperative for infrastructure management that is enduring and eco-friendly, and considerable improvements in sensor technologies, data analytics, and computational techniques. These changes have enhanced the ability to monitor the health of structures effectively, ensuring safety and minimizing the need for repairs [1].

The importance of SHM lies in its ability to provide information about the structural integrity and performance and identify minor damages before they evolve into major failures, thereby preventing catastrophic events and ensuring the safety of users and residents. Furthermore, SHM contributes to cost-effective asset management by optimizing maintenance and repair schedules, reducing unnecessary inspections, and extending the service life of structures through informed decision-making [2].

Damage identification can be enhanced by developing algorithms or models derived from the field of machine learning or pattern recognition. This approach is defined as data-driven SHM and exploits the measured data to identify and classify anomalies. Data-driven models offer bottom-up solutions that include diagnosis and prognosis that include damage detection and remaining life estimation, respectively

[3]. These models are distinguished in *supervised* and *unsupervised*. The first type of model requires labelled data, to learn the relationship between the measurements and the structural behaviour. The latter instead, learn to identify anomalies without requiring labelled data.

Despite its numerous benefits, SHM faces several challenges. These include the handling and interpretation of massive datasets generated by sensors, ensuring the long-term reliability and calibration of sensors in hard environmental conditions, and the development of universally accepted damage detection algorithms capable of accounting for the unique characteristics of different structures.

It is important to enhance that SHM systems designed to classify data into distinct health states mainly depend on supervised statistical models. These models are a category of machine learning algorithms that learn to predict outcomes based on input data that has been labeled with the correct output. This dependence limits their practical use since these models require specific conditions to function effectively. To overcome this limitation, there's a growing interest in utilizing data across a range of structures. This approach has led to the development of a new field known as population-based Structural Health Monitoring (PBSHM) [4, 5, 6, 7].

As a result, PBSHM seeks to enable the sharing of valuable insights among groups of structures, referred to as population. The nature of the population plays a crucial role in this context, as it determines the kind of knowledge that can be exchanged and the methods through which this exchange can occur [4], [5]. PBSHM represents a notable advancement in SHM, with the creation of methodologies that leverage data from a group of structures to learn knowledge that can be shared to the entire group [8].

Understanding the different types of populations that can exist within a population based approach is crucial. One of the primary categories to consider is homogeneous populations. These populations refer to a group of structures or components that are essentially identical in several key aspects. These aspects could include material properties, geometry, boundary conditions and loading conditions. It's essential to acknowledge that absolute homogeneity is rare in real-world scenarios. Variations in manufacturing processes, material inconsistencies, ageing, and environmental exposure can introduce differences over time, even among seemingly identical structures. Therefore, while the homogeneous population concept provides a useful framework for PBSHM, allowances must be made for some level of variation within these populations. Heterogeneous populations represent the second major category in the context of PBSHM, contrasting with homogeneous populations. These populations are characterized by significant differences among the structures they comprise, whether in geometry, materials, topology, or a combination of these and potentially other factors. These differences inherently lead to varied data distributions and, possibly, differences in label classes, making the SHM process

more complex [8].

In the context of heterogeneous structures, enhancing damage identification involves developing algorithms or models that can effectively account for the inherent variability in the structures and still detect damage reliably. Using a combination of statistical models and machine learning it could be possible to develop algorithms capable of identifying damage based on the selected features. This might include supervised learning models trained on labeled data (data where the presence or absence of damage is known) or unsupervised models that identify anomalies without labeled data.

In this field, for damage identification, vibration-based SHM techniques have been investigated over the last decades. In particular, the main changes are presented in the investigation of Hou et al.[9] e Avci et al. [10]. Researchers aimed to develop more robust and reliable SHM systems capable of addressing the evolving challenges of infrastructure maintenance and management. Avci et al. [11], employed a 1D, CNN in conjunction with wireless sensor networks (WSN) for the purpose of detecting damage through vibration data analysis. The effectiveness and reliability of the proposed approach were demonstrated through extensive experiments on a steel frame structure in a laboratory setting. The findings revealed that 1D-CNN was not cost-effective in terms of computational resources but also well-suited for wireless SHM applications. In a subsequently study, Zhang et al. [12] also applied 1D CNN for a similar research purposes.

1.2 The motivation of the investigation

Within the framework of this work, the SHM approach combines statistical methods with structural engineering principles to improve the detection of damage in simplified structures. The primary goal is to identify damage more accurately and efficiently, especially in structures that are not uniform in their construction or material properties.

This investigation proposes that statistical alignment (SA), a technique within DA that aims to directly align lower-order statistical properties, presents a viable approach for domain adaptation in cases where data or expert knowledge is limited. Moreover, it suggests that customized modifications of SA could effectively tackle class imbalance and partial DA challenges. In this context, Statistical Alignment refers to the process of aligning data from different sources or conditions in a way that makes them comparable. Moreover, it will be analyzed if SA could effectively tackle partial DA challenges [13].

Most machine learning algorithms operate effectively under a key assumption: both training and testing data come from identical feature spaces and distributions. However, when these distributions shift, it typically necessitates constructing new

statistical models from the ground up, relying on freshly gathered training data. In numerous practical scenarios, acquiring the necessary training data afresh is either costly or unfeasible. Minimizing the necessity and effort involved in collecting this data again would be beneficial. In these situations, the ability to knowledge transfer or apply transfer learning across different task domains becomes highly valuable [14]. Transfer learning focuses on enhancing the learning process in one domain by leveraging insights from a relevant, but distinct domain. This field encompasses various strategies, each with its unique objective and underlying premises [8]. Domain adaptation, a specialized area within TL, focuses on aligning the feature spaces' distributions between different domains. This is achieved by developing a statistical model capable of performing accurately on a new target structure (target domain) using data from an existing structure (source domain). These methods typically employ distribution distance metrics and nonlinear kernels to facilitate the mapping between domains.

For the analysis done in this work, the application of SA concretely involves aligning natural frequencies, noted as features, from multiple structures to identify patterns or anomalies indicative of damage. They were obtained for the different structures and models. It is important to note, that the damage is manually introduced in a defined number of samples, utilizing a specific crack theory in a determinate column of the structure. The alignment process involves normalization, scaling, or transformation of features data to a common reference domain.

In the present work there is presented an experimental test in which a methodology is applied for the study of PBSHM. The structure is taken from the Bookshelf test done in "Los Alamos National Laboratory".

Evaluating the performance of the damage identification models on unseen data, ideally from structures not included in the training set, to assess their applicability and reliability. It is interesting to analyze how investigating statistical alignment performance for enhancing damage identification in heterogeneous shear structures is a multi-disciplinary effort for structural engineering and data science. The success of such investigations could significantly advance the ability to monitor the health of infrastructure, ensuring safety and efficiency in maintenance practices.

The objective of this investigation is to illustrate the application of Statistical Alignment for enhancing damage detection across a population of structures.

To summarize, there are several reasons that motivates this investigation. First of all, the analyze how the methodology of SA handles with heterogeneous data. Because in real word, datasets come from varied sources, with differences in distributions and in scales. This methodology should help to effectively combined the dataset and improve the prediction. On the other hand, is relatively new in the field of DA, where exists a significant discrepancy between source and target. By aligning the statistical properties of source and target datasets, it becomes easier to transfer knowledge from one domain to another. Statistical Alignment

also solve the problems related to class imbalance and enhance the interpret ability of datasets, understanding the differences between the conditions within the data. Finally it is possible to say that, SA can lead to models that are more robust and generalize better to unseen data, models can achieve improved performance across a variety of scenarios, not just the specific conditions they were trained on.

The layout of the thesis is as follows. Chapter 2 presents in detail the theoretical foundation of the development of population based SHM. Chapter 3 describes the numerical simulation conducted to represent the structures that comprise the entire population. The following chapter, Chapter 4, thoroughly explains the central methodology of the research: Statistical Alignment. In addition, it develops the necessary concepts regarding the machine learning algorithm utilized and the parameters that will be used to analyze the algorithm's performance. Chapter 5 will display the results for the considered cases, focusing on the numerical outcomes and the aforementioned performance. The experimental case will be analyzed in Chapter 6, along with its pertinent conclusions. Chapter 7 will engage in a general discussion of the research to draw conclusion in Chapter 8 and discuss the future of this methodology.

Chapter 2

Population-based Structural Health Monitoring

2.1 Structural Health Monitoring of civil structures

Society grows around many structural and mechanical systems, which are the base of their development. These structures include power generation systems, bridges, underground infrastructures, oil extraction platforms, and buildings. All these systems are subjected to different natures of forces and types of degradation, which lead to short-term and long-term damage during their service life.

Specifically, civil structures present different typologies of damage. The one related to the structural cracking, in concrete, masonry or others structural materials due to the expansion, settlements, or load stresses. The structures revels damage due to the rusting of steel reinforcement, within concrete, reducing its strength and durability. Also, the foundation settlement, leads to cracks, misalignment and structural instability. The damage is caused also by the breaking of pieces of concrete, often due to corrosion of embed steel reinforcement. In specific regions, damage could be caused, in big or small scale, due to seismic actions, caused by an earthquake, generating cracking, structural displacements or collapse. Particularly, in bridges, are present the damage due to localised yielding of reinforcement, cracking or local spalling of concrete, fatigue, which is a long term effect, impact loading, overloading, differential settlement and seismic loads.

It is important to understand that these systems cannot be economically replaced, so different techniques for damage detection are being developed and implemented so that these systems can continue to be safely used when their operation is extended beyond the design basis service life.

The whole investigation done in the present work, is aimed to study in depth the recognition of damage.

Damage will be defined as "intentional or unintentional changes to the material and/or geometric properties of these systems, including changes to the boundary conditions and system connectivity, which adversely affect the current or future performance of these systems" [1].

The term Structural Health Monitoring (SHM) refers to the process that involves several key steps: the acquisition of experimental data, transmission of this data, processing (with algorithms for damage detection, localization, e.g.) and finally, decision-making and management. This is aimed at ensuring the safety and integrity of structures. The general scheme is represented in Figure 2.1. The transition from traditional time-based maintenance to condition-based maintenance is facilitated by SHM technology. Unlike fixed maintenance schedules, which may lead to unnecessary inspections or overlook critical issues, condition-based maintenance relies on real-time data to guide maintenance activities. By equipping structures with sensors that monitor performance indicators, operators can identify emerging problems and intervene before they escalate. This predictive maintenance approach minimizes downtime, reduces costs, and enhances operational efficiency across industries.

Damage identification in mechanical systems can utilize dynamic characteristics such as eigen frequencies, mode shapes, and the Modal Assurance Criterion (MAC), or rely on static measurements like displacement, inclination or strain [15]. In this study, it will be referred to dynamic monitoring (vibration-based).

The fundamental problem of SHM, the question of damage detection, is simply posed [16]. The object is just to identify if and when the system departs from normal condition. This is the most basic question that can be addressed. At a slightly more sophisticated level, the problem of damage identification can be approached. This seeks to determine a much finer diagnosis and can even address issues of prognosis. The broader problem can be regarded as a hierarchy of levels which are as follows [17].

1. **Detection.** The method gives a qualitative indication that damage might be present in the structure.
2. **Localization.** The method gives information about the probable position of the damage.
3. **Assessment.** The method gives an estimate of the extent of the damage.
4. **Prediction.** The method offers information about the safety of the structure, e.g. estimates a residual life.

SHM approaches are broadly divided into the inverse-problem or model-based approach and the data-based approach [18]. The first one, relies on updating

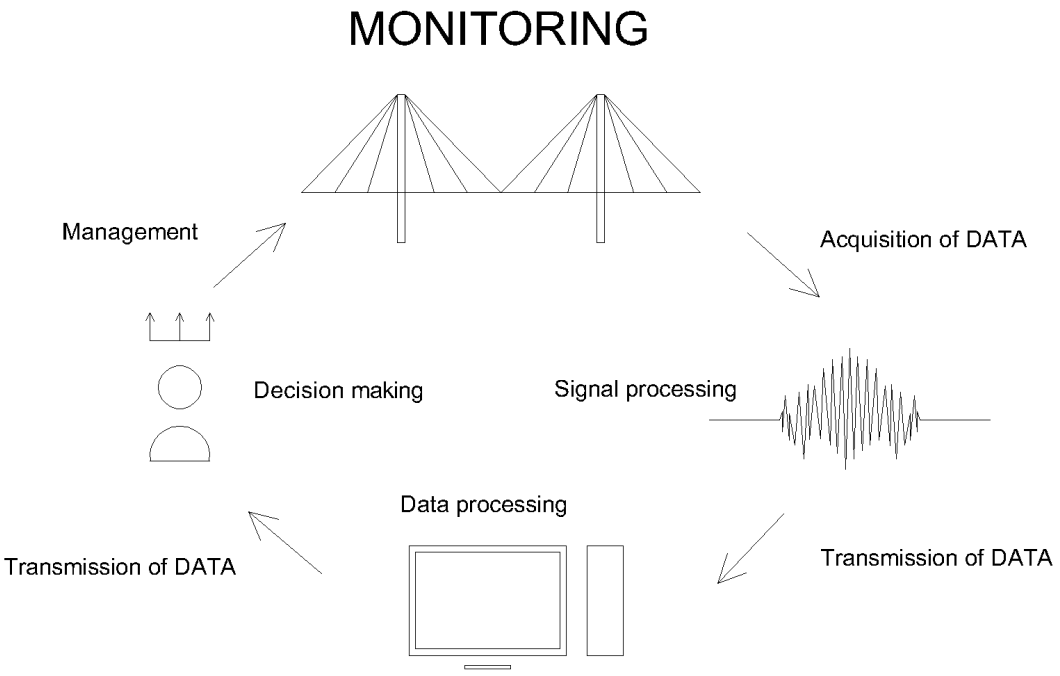

Figure 2.1: Monitoring Scheme

physics-based models, like finite element models, with real structure data to better match its actual condition, assisting in damage diagnosis when deviations are detected. The latter, meanwhile, uses machine learning to classify states based on training data from various health and damage conditions without initially relying on a physics-based model, though such models can still inform feature selection for damage identification.

While both model-based and data-driven approaches have their merits, the choice between them depends on various factors, including the complexity of the structure and the availability of data. Model-based approaches leverage physics-based models to simulate structural behavior and predict potential failure modes. However, these models often require extensive calibration and may struggle to capture the full complexity of real-world scenarios. In contrast, data-driven approaches rely on empirical data to train machine learning models, offering flexibility and scalability in handling diverse structural configurations and operating conditions.

The accuracy of the models in model-driven approaches, depend on the foundations and empirical formulas derived from physical laws to predict behaviour under various scenarios. On the other hand, data-driven approaches leverage the power of data and computational algorithms to identify patterns, correlations, and behaviors directly from the data itself, without necessarily requiring a priory knowledge of the physical principles governing the system. However, this approach requires extensive datasets to train models, especially for complex problems. Acquiring, storing, and processing such large volumes of data can be costly and time-consuming. The performance also depends on the quality and representativeness of the data. Additionally, the data may be scarce or difficult to collect due to logistical challenges. Poor data, biased, or incomplete datasets can lead to inaccurate or misleading results. The way to address these issues, particularly, data availability and model generalization, is to leverage the collective data from a wide range of similar structures and enhance the prediction and detection of damage across an entire population of structures. This methodology is known as population-based SHM 2.2.

Data driven make use of Statistical Pattern Recognition (SPR) framework, encompassing operational evaluation, data acquisition, feature selection, and statistical modeling for feature discrimination. Machine learning within this context aims to uncover the relationship between derived features and the structure's damage state. This process can be categorized into supervised learning, where the relationship is learned from labeled training data, and unsupervised learning, focusing on discovering data's intrinsic patterns, potentially for novelty detection. Operational evaluation involves assessing the structural performance under normal operating conditions to establish baseline parameters. Data acquisition encompasses the process of collecting relevant data through sensors strategically placed on the structure. Feature selection entails identifying key parameters or signatures indicative of structural health status, while statistical modeling enables the interpretation of these features and the prediction of potential failures.

Pattern recognition can be approached through template matching, statistical, neural, and syntactic methods [19], with the statistical approach often deemed most fitting for SHM due to inherent uncertainties in engineering problems. This approach, alongside neural networks—which can also be viewed through a statistical lens—offers a solid foundation for addressing SHM challenges.

Machine learning, a well-established field, employs pattern recognition to mathematically associate measured data with specific class labels. In SHM, the objective is to correlate measured data with certain damage states, crucially distinguishing between 'healthy' and 'damaged' conditions of a structure. The concept of machine learning involves 'learning' the relationship between certain features extracted from measured data and the structural damage state. This entails estimating the function that characterizes this relationship using the training data acquired from the test structure.

2.2 A population-based approach

Utilizing machine learning operations on measured data, data-based SHM can progress through various diagnostic levels, including detection, location, and quantification, given the availability of representative data. Additionally, by including some physics-based models of damage progression, SHM can advance to prognostics, estimating safe residual life.

However, some issues may reduce the applicability of these approaches in real-world scenarios. Specifically, one of the main challenges lies in the scarcity of data, especially concerning damage states. Transitioning to Population-Based Structural Health Monitoring (PBSHM) offers a solution by transferring diagnostic capabilities among structures. PBSHM has been introduced in [4, 5, 6, 7] with the aim of leveraging available damage-state data from a population of structures to enhance diagnostic accuracy for a new target structure. Additionally, for similar structures, PBSHM can improve the relevance of physics-based models across entire populations, ultimately enhancing prognostic capabilities.

In other words, adopting a PBSHM strategy, offers a solution to the problem of incomplete training data. This method groups similar systems or models, enabling the possibility of sharing or transferring information among them. Identifying which systems are sufficiently alike to exchange information is a crucial initial step in PBSHM, with various proposed methods for measuring similarity. Sharing information among similar systems is expected to enhance the accuracy of pattern recognition algorithms significantly. However, it is important to proceed with caution: exchanging information between systems that are not similar can lead to a decrease in predictive accuracy, a phenomenon known as negative transfer. Therefore, PBSHM also investigates different strategies to mitigate this risk. In the case of negative transfer, there are some strategies to mitigate it: implementing stringent criteria for what constitutes similarity within a population, and carefully selecting similarity thresholds. This might involve setting high thresholds for similarity measures or requiring multiple criteria to be met before allowing information sharing. Afterwards, it is possible to focus on domain adaptation techniques, which harmonise the transferred information to better fit the target system.

2.2.1 Structural similarity

PBSHM offers significant benefits when applied to populations of structures sharing analogous characteristics, termed strongly-homogeneous populations. These structures are nearly identical, differing primarily due to manufacturing variances. Knowledge transfer within such populations relies on constructing a model, known as the population form [4], which captures both the generic behavior and characteristic variations across the population. Past research on strongly-homogeneous

populations has exploited the expectation that nominally-identical structures should exhibit similar behavior under similar conditions. However, to facilitate knowledge transfer also between structures that are not identical, it is crucial to quantify their similarities and differences as highlighted in [20, 21]. These populations, where structures can vary significantly from each other, are defined as heterogeneous populations and are the target of the current study.

To achieve this, the degree of structural similarity needs quantification, and mechanisms for transferring inferences through machine learning must be established. The objective is to identify common substructures within non-identical structures, ensuring valid knowledge transfer at a substructure level, as introduced in [5]. PBSHM addresses the similarity assessment phase via abstract representations of the analysed structures, aiding automatised comparisons for generalisation purposes. These representations are called Irreducible Element (IE) models and Attributed Graphs (AG).

An IE model is constructed to decompose structures into components with well-established dynamic behavior, capturing essential elements defining the structure's nature. This model is then converted into an AG, which summarises the same structural information in the form of a graph for comparison. The AGs are widely used in various fields, including chemistry and bio-informatics, and allow the measurement of structural similarity by automatically implementing a predefined distance matrix. The first metrics proposed by Gosliga et al. [5] involves the identification of the maximum common subgraph (MCS), and the evaluation of the Jaccard index. Other approaches are based on kernel-based methods, as shown in [22], and graph neural networks (GNNs) [23, 24], as presented in [25] to highlight the influence of structural attributes, which are not considered in the MCS.

To determine structural similarity efficiently, it is impractical to consider every property or dimension directly. Instead, an abstraction of significant properties and dimensions is achieved through the IE representation, focusing on elements and joints that significantly affect knowledge transfer. Matching properties ensures consistent labels for transfer learning, facilitating efficient comparison and inference.

Once IE models are generated, they are converted into AGs, capturing topology, materials, and geometry information. Common subgraphs are identified using a modular product and clique-finding algorithm, enabling comparison of structural attributes for knowledge transfer. By choosing appropriate information and data structures, AG comparison facilitates selecting the most suitable structures for transfer learning tools, and grouping structures based on their potential for knowledge transfer.

The nature of the populations in PBSHM is crucial as it dictates the type of knowledge that can be shared and the methods used for transfer. Indeed, groups of structures that are composed of the same kind of structures are known as homogeneous populations, and in some instances, a generalized model may be

employed to depict the collective behavior of the group and to detect damage among its members. An example of this kind of population is a population of wind turbines in a wind farm.

In the research conducted by Bull et al. [4], the authors demonstrate that it is feasible to transfer knowledge throughout a population by creating a model – referred to as the form. This model is designed to encapsulate the general behavior of the structure while also accounting for the specific variations seen across the population. Prior investigations into homogeneous populations leveraged the premise that structures deemed nominally identical should manifest similar behaviors under akin boundary conditions. This approach was exemplified by using a model developed for a single wind turbine to anticipate the performance of another turbine within the same wind farm. Such a method yields a model of normal conditions that is resilient to specific environmental influences, as the boundary conditions dictated by wind speed and temperature are expected to remain fairly uniform across the wind farm. Consequently, if a particular turbine begins to display behaviors that deviate from the norm established by its counterparts, this divergence could indicate the presence of damage. Subsequent studies introduced the inaugural version of the form, aiming to symbolize all members within a homogeneous population.

But then it is posed the question, if knowledge transfer is possible between nominally-identical structures, what about structures which are not the same, yet share some degree of similarity? The phase of similarity assessment within PBSHM tries to find this answer for heterogeneous populations. These are groups composed of more diverse members. In the case of the present work are structures with different numbers of floors, materials and geometries. These structures share similarities, however, they are not classified as nominally identical due to their attribute variations.

In summary, PBSHM enables effective knowledge transfer within homogeneous populations and provides methodologies for determining structural similarity and facilitating efficient transfer learning within heterogeneous populations, ultimately enhancing SHM capabilities.

2.2.2 Knowledge sharing

Knowledge sharing is the second phase in PBSHM. Given the findings of the similarity assessment phase, is it possible to define a source structure within the population. The source structure is defined as the most appropriate to inform diagnostic on a certain target structure, for which only few measured data are available. Afterwards, it is necessary to choose between a variety of algorithms to exploit the knowledge acquired from the source dataset.

In this framework, transfer learning is a class of machine learning algorithms

aimed at enhancing the learning process by transferring knowledge from one domain to another. Within this area of machine learning, there are various approaches, each based on different assumptions about the similarities between domains and the nature of the knowledge transferred. This work tries to investigate the transfer of knowledge within the context of PBSHM, with a special emphasis on modal-based features. Alternatively, other approaches try to directly exploit knowledge acquired from the time histories of the measured quantities, as presented in [26]. In the framework of PBSHM, the class of domain adaptation algorithms, a subset of transfer learning, appears particularly effective [8]. Domain adaptation operates under the premise that labeled data from a source domain can be utilized to improve classification in a target domain, which may be unlabeled or partially labeled, by aligning both domains within a shared subspace.

There exists methods that are more or less complex. As an example, it is possible to find in the literature, investigations on Transfer Component Analysis (TCA and TCA+) [27] and [28], also on balanced distribution adaptation (BDA) [29] and on geodesic flow kernel (GFK) [30].

Statistical Alignment, a branch of DA is introduced. Beyond requiring less data, SA, also enhances the visualization of domain-specific data. Unlike several well-known DA techniques, which transforms the data into a latent space using nonlinear mappings, SA preserves the original structure of the feature space through affine transformations. This preservation is particularly beneficial then features are related to a physical phenomena. For instance, a rise in natural frequency is understood as an indication of stiffening, demonstrating the importance of maintaining physically interpretable features [13].

Additionally, knowledge sharing in PBSHM can exploit available data from numerical simulations as well as real-world datasets. Indeed, by combining different types of case studies, the PBSHM methodology can expand the available observations of health states of interest. Moreover, by considering simulated case studies is possible to verify PBSHM performance, and understand the influence of different sources of heterogeneity that may be found also in real-world applications. Therefore, this study explores transfer learning and domain adaptation techniques for damage identification, focusing on their application in a numerical population of heterogeneous shear structures. Some SA techniques, such as Normal Condition Alignment (NCA) and Normal Correlation Alignment (NCORAL), will be used to harmonize the features across the numerical population. These approaches, already presented in [13], are employed on different numerical models of shear structures, to investigate how transfer learning can improve diagnostic performance for changes in the topology and the structural properties. The analysis will be further extended to the experimental case study of a three-story frame structure analyzed by the Los Alamos National Laboratory

Chapter 3

Numerical simulation of a population of structures

3.1 Design of a numerical population

The current Chapter describes the numerical simulation of a population of shear frame structures. This population is adopted to investigate how the performance of the SA algorithm is affected by the possible variations between a source and a target structure, with the aim of understanding to what extent this approach can be used to share knowledge in PBSHM. The population consists of a total of *nine structures*, each featuring different materials, number of floors, and geometries. Specifically, by geometry, it refers to the dimensions of the floor plants and their height. Each of the structures represents a potential source and target domain. There will be examined different simulation models and different cases. Furthermore, the variability proposed in the properties of the structures contributes to ensuring a consistent heterogeneous population.

3.2 Numerical models

The numerical case studies are built with two different types of simulation, considering 2D models and 3D models. The simulated structures consist of three, four and five floors' systems, each floor represented by a mass with a horizontal length "L_x", vertical "L_y" and a thickness *"tm"*. The masses are connected by four columns with a length lb, a width wd and a height lb, which corresponds to the height of the floor.

The numerical population presents a simplification of the models that can be summarized in the following points.

- **Linear Elastic Behavior**: Shear structures are often assumed to behave linearly elastically within the range of expected loads. This assumption simplifies the analysis, as it allows the use of linear elastic material properties.
- **Homogeneous Material**: It is assumed that the material properties (such as Young's Modulus and Shear Modulus) are uniform and consistent throughout the entire structure. In reality, material properties might vary across different sections of a structure.
- **Idealized Supports**: Support conditions are often simplified to fixed, pinned, or roller supports, neglecting the complexities of real-world foundation systems.
- **Neglecting Geometric Nonlinearities**: Geometric nonlinearities, such as *large deformations and buckling effects*, are often neglected in preliminary analyses of shear structures unless they are the primary focus of the study.
- **Neglecting Connection Details**: Detailed modeling of connections between structural elements, which can be complex and involve nonlinear behavior, is often simplified in preliminary analyses.

The structural scheme is presented in Figure 3.1.

For the scheme exhibited, there were considered different geometries. They are shown in Table 3.1, in which every scenario, has a different magnitude for the floor and column's dimensions, and finally, position and dimension of the crack. The consideration of these differences aims to underscore the diversity among the population's structures.

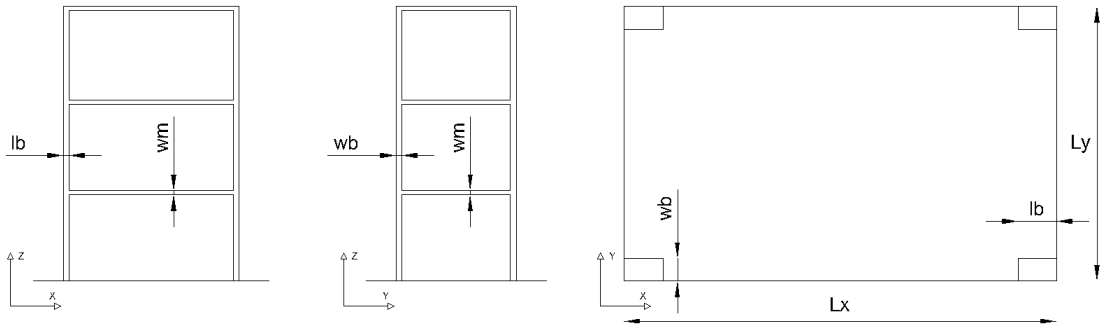

Figure 3.1: Geometrical scheme of the numerical Population.

The structures differ also in the materials. Specifically, all structures are simulated adopting metal materials. However, some structures are assumed to be made of steel and some others are assumed to be made of aluminium, to introduce a source of heterogeneity. Thus, general values are assumed for the mechanical

Geometry				
Dimensions	Unit	Case 1 [C:1]	Case 2 [C:2]	Case 3 [C:3]
lb	cm	40.00	20.00	47.00
wb	cm	5.00	3.00	2.93
tb	cm	0.70	0.80	3.93
Icx	m^4	1.49×10^{-9}	1.28×10^{-9}	1.48×10^{-7}
Icy	m^4	7.29×10^{-8}	1.80×10^{-8}	8.24×10^{-8}
Lx	cm	50.00	35.00	86.80
Ly	cm	50.00	30.00	46.20
tm	cm	3.00	3.00	1.30
l_{cr}	cm	3.00	1.80	1.76
l_{loc}	cm	4.00	2.00	4.70

Table 3.1: Geometry

properties of steel and aluminum. These values can be found in Table 3.2. The numerical population is then created as shown in Table 3.3.

Materials			
	Unit	Steel (St)	Aluminum (Al)
Elastic modulus - E	Pa	210×10^9	71×10^9
Density - μ	Kg/cm^3	7800	2700
Damping ratio - c	Ns/cm^3	8	5
Poisson ratio - ν	-	0.28	0.35

Table 3.2: Materials

The response is simulated for normal health conditions and three damage cases. The simulation of the damage is posed in the next section. This investigation is based on vibration-based SHM, so the features obtained are the natural frequencies. Each sample is a set of natural frequencies obtained by solving the eigenvalue problem. The 3D and 2D models differ in the evaluation of their mass and stiffness matrices. In order to simulate the real acquisition of data and the variations in manufacturing, as performed by Poole et al. [13], a variation in the material properties is introduced. This is performed by defining the elastic modulus, density and damping coefficient as samples of Gaussian distributions. The mean and standard deviations adopted are posed in Table 3.4. Thus, it is possible to simulate the samples' variability that could be found in real datasets. These differences are lower than the difference introduced by damage. Thus, the damage detection is

Population									
	n°1	n°2	n°3	n°4	n°5	n°6	n°7	n°8	n°9
N° of floors	3	3	3	4	4	4	5	5	5
Material	Al	St	Al	Al	St	Al	Al	St	Al
Geometry	C:1	C:2	C:3	C:1	C:2	C:3	C:1	C:2	C:3

Table 3.3: Population

not affected.

Deviation			
	Unit	Steel (St)	Aluminum (Al)
Elastic modulus - E	GPa	$\mathcal{N}(210 \times 10^{-9}$	$\mathcal{N}(71.1 \times 10^{-10}$
Density - μ	Kg/m^3	$\mathcal{N}$(7800,30)	$\mathcal{N}$(2700,10)
Damping ratio - c	Ns/m^3	$\mathcal{N}$(8,08)	$\mathcal{N}$(5,05)

Table 3.4: Deviation materials

The damage cases, represent damage in the first, second and last floor of the source or target structure. The total number of samples corresponds to 800, representing 200 for each class. The label $y = 0$ will be assigned for the healthy state, label $y = 1$ for damage on the first floor, label $y = 2$ for damage on the second floor and label $y = 3$ for damage on the last floor, for both domains, source and target. In addition, the target domain is separated into two different datasets, the first one is used during training and the second one is used for testing.

3.2.1 2D models

The population consists of two shear structures modelled as 3DoF lumped–mass models [8]. The mass of each degree of freedom (DoF) was modelled as a rectangular volume, to represent a floor, characterized by dimensions: length (L_x), width (L_y), thickness (tm), and density (μ). The masses were linked together using four cantilever beams arranged in parallel, leading to a total stiffness, k, calculated as four times the stiffness of a single beam $k = 4k_b$. The stiffness for each beam, denoted as k_b, was determined by the tip stiffness formula for a cantilever beam, $k_b = 3 \cdot EI/b^3$, where E represents the elastic modulus and I is the moment of inertia. As mentioned before, the elastic modulus for each sample was sourced from a Gaussian distribution to introduce variability. Unlike stiffness and mass, damping (c) was not modelled based on a physical equation but was instead directly sampled

from a Gamma distribution. Each four-column system has the same stiffness on the different floors because the material and height remain constant.

Damage is modelled as a reduction in stiffness as in Christides and Barr 1984,[31] $k = k_d + 3k_b$. where k_d is the stiffness of a damaged cantilever beam. Having obtained the parameters of the model, the natural frequencies were calculated by solving the eigenvalue problem. The scheme for the 2D model of the structure is shown in Figure 3.2.

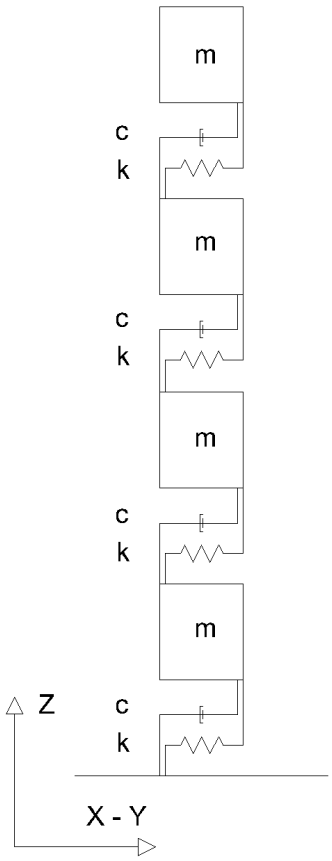

Figure 3.2: Structural Scheme of the 2D models

Mass matrix

The mass matrix of the system is a diagonal mass matrix in which each element is calculated as the density times the dimensions of the floor. The mass of each floor is given by the equation 3.1

$$M_i = lb \cdot wb \cdot tb \cdot \mu \tag{3.1}$$

Being *"i"* the representation of each floor.
The whole system has a mass matrix *NxN* being N the number of floors.

$$M = \begin{bmatrix} M_1 & 0 & 0 & 0 \\ 0 & M_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & M_i \end{bmatrix} \tag{3.2}$$

Stiffness matrix

Each element component of the global stiffness matrix of the system is equal to the stiffness of four cantilever beams, equation 3.3.

$$k_i = 4k_b \tag{3.3}$$

When the numerical population is created, 200 samples will have the undamaged stiffness, so they are not going to present a reduction in the stiffness. Instead, for the other 600 samples of the dataset, there will be a stiffness reduction, as they represent the damage cases. The reduction will be on the first, second and last floor of the structure. To have a correct representation of damage cases. The reduction is set in one column of the floor, so the stiffness for the damaged floor results as 3.4

$$k_i = 3k_b + k_d \tag{3.4}$$

The calculation of k_d is presented at the end of the Chapter. Finally, the stiffness matrix is equal to 3.5.

$$K = \begin{bmatrix} k_1 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & k_i \end{bmatrix} \tag{3.5}$$

Clearly, with this simulation called 2D model, the resulting models are shear-type structures.

Natural frequencies

Performing eigenvalue analysis with mass and stiffness matrices of each structure, the modal parameters are obtained. As mentioned, the natural frequencies obtained according to this mass-lumped model are equal to the total number of floors. In this case, the degree of freedom is also in correspondence with the floor. For the analysis, the first three natural frequencies are, and these features are then standardized with the use of Statistical Alignment.

3.2.2 3D models

Another model employed in the numerical analysis, was the 3D model. This model as the name could anticipate, is based on a three-dimensional analysis of the structure. With respect to the 2D model, the main difference is that there will be three degrees of freedom by floor, two translational in the two principal directions and one rotational of the floor. Thus, the total number of natural frequencies obtained is the product of the total number of floors by three. When the source and target structures differ, the main challenge is the selection of the features to compare. To solve this issue, the MAC criterion is used, to properly compare the translation or rotation modes of the source with the correspondent of the target. The analysis was for the first three natural frequencies. The equilibrium, from which the mass and stiffness will be obtained, is calculated from the free vibrations equation of motion with no damping. Equation 3.6.

$$M \cdot \ddot{X}_t + K \cdot X_t = 0 \tag{3.6}$$

Mass matrix

In this simple three-dimensional model, some general assumptions allow to interpret the behavior simulated. The degrees of freedom are two translations and one rotation by floor, considering a diaphragmatic behavior. Mass is concentrated at the level of the floor, and the columns have no mass. So, to be consistent with displacements of the floor. There are going to be wrote the equilibrium equation, appearing only rotation and translations of the floor. All the members are axial rigid and the axial deformation of the pillars and the beams are not considered. The torsional stiffness is neglected. This corresponds to cylindrical hinges at the intersection between orthogonal frames. Is also considered that torsion is not transmitted. The modal parameters resultant from this model will be compared with the results from the real structure (Bookshelf structure) made in the laboratory in Los Alamos. The inertial forces are given by,

$$F_t(t) = M \cdot \ddot{X}(t) = \begin{bmatrix} M_t & 0 & -S_x \\ 0 & M_t & S_y \\ -S_x & S_y & I_0 \end{bmatrix} \cdot \begin{bmatrix} \ddot{u}_{(t)} \\ \ddot{v}_{(t)} \\ \ddot{\varphi}_{(t)} \end{bmatrix} \tag{3.7}$$

The matrices which compose the total mass matrix are shown in the following.

$$M = \begin{bmatrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & m_N \end{bmatrix}; I_0 = \begin{bmatrix} I_0^1 & 0 & \dots & 0 \\ 0 & I_0^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_0^N \end{bmatrix} \tag{3.8}$$

$$S_x = \begin{bmatrix} S_x^1 & 0 & \dots & 0 \\ 0 & S_x^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_x^N \end{bmatrix}; S_y = \begin{bmatrix} S_y^1 & 0 & \dots & 0 \\ 0 & S_y^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & S_y^N \end{bmatrix} \tag{3.9}$$

"N" is the total number of floors.

Stiffness matrix

This is a simplified model to represent a 3D frame: the procedure starts from the definition of the stiffness matrix for each 2D frame. It is possible to notice that the building, in general, are organized in frames according to main directions, so there are going to be considered only one direction of the stiffness, it should be necessary to determine the 2D matrix, so the 2D frame. This must be done systematically for each frame. Each floor has a rigid motion kinematic behavior.

The elastic force of the system is,

$$F_E(t) = K \cdot X(t) = \begin{bmatrix} K_{xx} & K_{xy} & K_{x\varphi} \\ K_{yx} & K_{yy} & K_{y\varphi} \\ K_{x\varphi} & K_{x\varphi} & K_{\varphi\varphi} \end{bmatrix} \cdot \begin{bmatrix} u_{(t)} \\ v_{(t)} \\ \varphi_{(t)} \end{bmatrix} \tag{3.10}$$

Where the following stiffness values are calculated,

$$K_{xx} = \sum_{j=1}^{4} K_j \tag{3.11}$$

$$K_{yy} = \sum_{j=5}^{8} K_j \tag{3.12}$$

$$K_{xy} = 0 \tag{3.13}$$

$$K_{x\varphi} = -\sum_{j=1}^{4} K_j \cdot d_{jy} \tag{3.14}$$

$$K_{y\varphi} = \sum_{j=5}^{8} K_j \cdot d_{jx} \tag{3.15}$$

$$K_{\varphi\varphi} = \sum_{j=1}^{4} K_j \cdot {d_{jy}}^2 + \sum_{j=5}^{8} K_j \cdot {d_{jx}}^2 \tag{3.16}$$

The distance d_{jx} and d_{jy} are the distances of the frame respect to the origin of the axis, Figure 3.3.

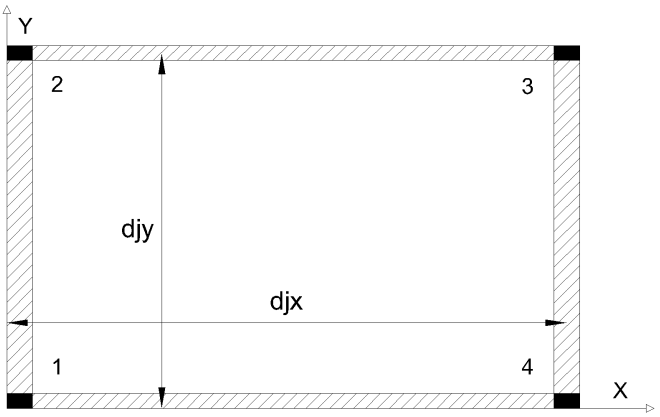

Figure 3.3: Frames Distances d_{jx} and d_{jy}

The matrix K_j is the matrix of one planar frame and it is obtained by the stiffness of each floor. Each floor has a stiffness equal to the sum of two columns, 3.17.

$$K_j = \begin{bmatrix} k_j^1 + k_j^2 & -k_j^2 & 0 & 0 \\ -k_j^2 & k_j^2 + k_j^3 & -k_j^3 & 0 \\ 0 & -k_j^3 & \dots & k_j^N \\ 0 & 0 & -k_j^N & k_j^{-1} + k_j^N \end{bmatrix} \tag{3.17}$$

For the damage case the stiffness changes according to the direction, due to the damage direction, *"X"* or *"Y"* direction.

Natural frequencies

Solving the eigen value problem with the global mass matrix and stiffness matrix, the natural frequencies are obtained. The total number of features by sample, are three by the number of floors. It is created for the source and for the target. Then the label is added according to the damage state.

When the structure from source and target, differ in the number of frequencies, appears the challenge to efficiently take the correspondent frequencies. The error to avoid, is to do not take the first torsional frequency with the first flexural in x direction, for example.

The solution was found in the calculation of the MAC.

MAC

Given the clarification that you're interested in the MAC criterion related to mode selection, it sounds like you're referring to the Modal Assurance Criterion (MAC). The MAC is widely used in structural dynamics and modal analysis to quantify the degree of similarity (or correlation) between two mode shapes.

The Modal Assurance Criterion (MAC) is defined as a value between 0 and 1 that measures the consistency or correlation between two mode shapes. It is computed using the formula 3.18.

$$MAC(\phi_i, \phi_j) = \frac{|{\phi_i}^T \phi_j|^2}{({\phi_i}^T \phi_i) \cdot |{\phi_j}^T \phi_j|^2} \tag{3.18}$$

Where:

- ϕ_i and ϕ_j are the mode shape vectors being compared.
- ${\phi_i}^T$ denotes the transpose of the vector.
- The denominator normalizes the expression, ensuring the MAC value is between 0 and 1.

The MAC value has a clear interpretation. A MAC value close to 1 indicates that the two mode shapes are highly correlated or similar. This is often seen when comparing a mode shape from a theoretical model to the same mode shape obtained from experimental data. A MAC value close to 0 suggests that the mode shapes are different from each other. Intermediate values indicate a degree of similarity but not a perfect match.

The MAC is particularly useful in experimental modal analysis and in validating numerical models against experimental results. It helps in identifying, comparing, and validating the modes of a structure obtained either from two different numerical models or from a numerical model and an experimental test. This is crucial for ensuring that simulations accurately represent real-world behavior and for understanding how modifications to a structure might affect its vibrational characteristics.

Features selection

A further investigation on the use of MAC is made by [32]. The study mentioned and other investigations validate the selection of frequencies, based on the result of the MAC. This parameter is evaluated between the mode shapes of source and target.

In this investigation, there were calculated all the natural frequencies and mode shapes for the source and target domain. Then, there were taken the three natural

frequencies of the source which corresponds to mode shapes that results in a MAC value equal or similar to one, after comparing with the mode shapes of the target domain, in order to proceed with the standardization. For the source is evident that were taken the natural frequencies related to that modes that increase the MAC value. In this way was performed the features selection.

In these operations, it is interesting to mention that, due to a dimension incompatibility between the mode shapes of structures with different number of floors, before calculating the MAC, it was introduced a *mode normalization*.

This standardization is a *novelty* to deal with the problems of mode dimensionality. It essentially consists on the creation of a function of a certain grade, that best represents the mode being analyzed, and then in the evaluation of that function at a certain number of points. The quantity of points is the same for all modes regardless of the number of floors. This ensures that modes of equal dimension are obtained for any structure. The points at which the function is evaluated are points that range from zero to one. An example for normalized mode shapes are presented in Chapter 5, in Figure 5.1 and Figure 5.2.

3.3 Introduction of damage

A critical point of the investigation is the introduction of damage. How to better represent a real damage on a determinate column in order to simulate the possible condition in the reality, was a crucial point. In order to better solve the problematic, it was followed the calculation proposed by Christides and Barr 1984.[31]

For both models the same damage estimation was used. For the case of 2D models is intuitive to consider the columns as elastic beams [13]. As regards the 3D models, the difference is based on the direction in which the crack is considered. It could be define as a crack in the $''X''$ or $''Y''$ direction (see fig.plant). So, the columns were also modelled as beams with different behaviors in the different directions.

Christides: “Elastic beams are three-dimensional continua but for certain classes of their possible motions they can be modelled at a simpler level. The most elementary of these models, for transverse motion, is the well-known Bernoulli-Euler beam which is the simplest one-dimensional theory currently available. The reason for specifying an "open" crack is to avoid at this stage the very interesting complexities which result from the nonlinear characteristics presented by a crack which can open and close.”[31]

The most important points to highlight about the Euler Bernoulli beam are that the beam is initially straight, has a constant cross-sectional area, and is made of a homogeneous, isotropic material (the material properties are the same in all directions). The material of the beam follows Hooke’s law, meaning the stress is

linearly proportional to strain within the elastic limit of the material. Plane Sections Remain Plane: Cross-sections of the beam that are plane and perpendicular to the beam's axis before deformation remain plane and perpendicular to the deformed axis of the beam after bending. This is also known as Bernoulli's hypothesis. It implies that there is no warping or distortion of the cross-sections. The deflections of the beam under load are small compared to the dimensions of the beam. This assumption allows linearizing the curvature of the beam, making the differential equations governing the beam behavior linear. It is assumed that the beam is subject to pure bending (moments without shear forces) in the region of interest, and that the bending moment varies continuously along the length of the beam. This simplifies the analysis by focusing on bending deformations and neglecting the effect of shear deformations. The theory assumes that the shear deformation (the deformation due to shear forces) is much smaller than the deformation due to bending and can be neglected. This is more accurate for beams that are long and slender. Finally, the beam is modeled as a one-dimensional element with its length much greater than its width or depth. The theory does not account for three-dimensional stresses or deformations within the beam.

Continues Christides and Barr: "The introduction of cracks will lead to changes in the stress and strain distributions in the vicinity of the cracked section. It is known that near the crack tip there are large stress concentrations and that over the cracked section of the beam the stress is not linearly distributed and all components of stress are likely to be non-zero. However, since the overall dynamics of the beam is of interest it is assumed that the fine structure of the stress distribution is not particularly significant." [31]

The development of the formulation starts with the extended Hu-Washizu ([33]) variational principle for arbitrary independent variations. Then, the Euler Bernoulli equation was derived. "The change in stress and strain distribution near the crack are brought in by using a function $f(x, z)$ at present unknown, which has its maximum value at the tip of the crack and which decays with distance from the cracked section." With the implementation of strain-displacement, strain-stress, velocity-momentum and dynamic momentum terms, and with the proper boundary conditions it was calculated the function $f(x, z)$.

The scheme for the column is the one in Figure 3.4.

This model was also applied by Gardner et al. [8] In which the elastic modulus times the inertia of the cross section of the beam, according to the length of the beam *"X"*, adopts the function 3.19.

$$EI(x) = \frac{EI_0}{1 + C\exp(\frac{-2\alpha|x-l_{loc}|}{t_b})} \tag{3.19}$$

In this equation α is a coefficient obtained in experimentation equal to 0.667 defined by Christides and Barr,[31] I_0 is the inertia of the undamaged cross section

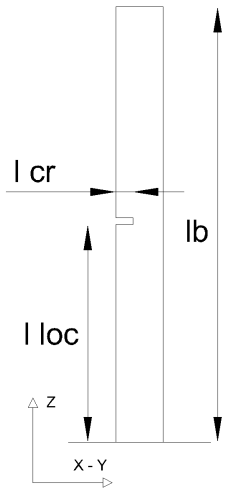

Figure 3.4: Cantilever Damaged Beam

and t_b the thickness of beam.

Highlighting that exists the difference in crack direction that affects to the stiffness of the beam, so that is the reason why is important to calculate separately,

$$I_{x0} = \frac{w_b \cdot t_b^3}{12} \tag{3.20}$$

$$I_{y0} = \frac{t_b \cdot w_b^3}{12} \tag{3.21}$$

The cracked inertia's intuitively can be calculated as 3.22 and 3.23.

$$I_{x,c} = \frac{(w_b - l_{cr}) \cdot t_b^3}{12} \tag{3.22}$$

$$I_{y,c} = \frac{(t_b - l_{cr}) \cdot w_b^3}{12} \tag{3.23}$$

The damaged stiffness k_d is obtained by 3.24.

$$k_d = \frac{-F)}{y_{top}} \tag{3.24}$$

Where F is a given force and y_{top} is the deflection of the top derived from integrating the Euler Bernoulli beam equation,

$$\frac{\partial^2 y}{\partial x^2} = -\frac{M(x)}{EI(x)} \tag{3.25}$$

Solving the differential equation with the boundary conditions of the beam and considering a unitary force, it is obtained k_d. This value will depend on the value of the elastic modulus E, the inertia moment of the cross-section I_0 and both geometries: of the crack and the undamaged area. All the characteristics of the damage geometry were presented in the table and varies for each case. But it was defined as a percentage of the geometry of the column. For the length of the crack corresponds to the 60% of the width (wb) of the cross section and for the position corresponds to the near neighbor around the clamp, which is represented by one tenth of the beam length. This was performed because after several attempts, this location allows the clear distinction of the damage in each sample. For the numerical calculation for the 3D model, it was introduced the damage in *"X"* direction. This distinction is unnecessary for the case of the 2D model.

Chapter 4

Domain adaptation and damage diagnosis

4.1 Statistical Alignment

A significant challenge within the domain of SHM is the scarcity of adequately labeled datasets. The application of TL, and in particular DA, becomes crucial in addressing the aforementioned challenge. DA facilitates the utilization of a vast array of information from a diverse set of structures (cluster), encompassing both physical and simulated environments. DA is instrumental in this context. It seeks to reconcile the distributional disparities in feature spaces through the identification of a mapping mechanism. This facilitates the development of statistical models that are more adaptable and generalizable to specific target structures (target domains) by leveraging data from alternative source structures (source domains). Techniques employed in this approach often hinge on distribution distance metrics, employing either nonlinear kernels or deep neural networks (DNN) to learn the requisite mappings. Within this context, the study concentrates on the analysis of structural natural frequencies, which vary with the structure's model and configuration and are instrumental in assessing its health condition.

During the analysis of this dataset, another predominant challenge encountered is class imbalance. This phenomenon pertains to the unequal distribution of data categories, where instances belonging to one class (for example, data indicative of a structurally health condition) significantly surpass those of another class (such as data indicative of structural damage). This disproportionate representation can adversely affect the efficacy of machine learning models, leading to a bias towards the more abundantly represented "healthy" condition. As a result, this inclination impairs the models' ability to precisely detect the vital "damaged" signals, which are of critical significance. It is demonstrated how Statistic Alignment (SA) can

avoid this problem of class imbalance in a numerical case study. It is then shown to be capable of aligning the feature spaces of a heterogeneous population. In essence, Statistic Alignment shows promise in not just getting these diverse datasets to align, but also in making our models more balanced. By tackling the class imbalance head-on, ensures that various models have the potential to precisely detect those infrequent yet crucial indications of damage. This approach broadens the horizon for SHM, offering a way to make better use of available data and improve the reliability of the monitoring efforts.

In the tentative to solve the problem of insufficient labelled data, and to facilitate the detection of structural damage, unsupervised machine learning strategies have been employed, leveraging outlier analysis. Nonetheless, SHM frameworks that categorize data into specific health states predominantly depend on supervised statistical models, thus limiting their practical applicability. This limitation underscores the importance of harnessing information from a broad spectrum of structures, paving the way for the emergence of population-based SHM (PBSHM). As it was presented in Chapter 2. A paramount challenge within PBSHM arises from the inherent variability in data distributions among different structures, which go against the conventional machine learning presumption of homogeneity between training and testing datasets. This variability can be interpreted due to numerous factors, such as manufacturing discrepancies and varying operational conditions in nominally identical structures, or intrinsic structural differences in heterogeneous populations. The application of PBSHM in the present investigation is related to the model created and the Bookshelf structure presented as an experimental case. In this way it was created the population and analyzed the performance. Later on, it will be shown.

This study seeks to formulate a DA method that is viable under conditions of limited data availability, a common challenge within engineering fields, especially in SHM where datasets are typically sparse and data acquisition is costly.

DA's utility in a PBSHM context has been studied in different works, Gardner et al [34]. successfully applied DA to facilitate the transfer of localization labels across numerical and experimental structures, including two distinct aircraft wings, as well as across the pre- and post-repair states of aircraft wings. Similarly, Bull et al. [35] demonstrated the effectiveness of DA in a study involving a population of six experimental tail planes, showcasing its capability to transfer a damage detection model.

Statistical Alignment is a branch of DA that directly aligns the lower-order statistics. Therefore, it can be used as a low-risk form of DA, as proposed in [13]. Indeed, these approaches focus on matching the first and second-order statistics and are most appropriate when data can be assumed Gaussian, which is a common assumption in SHM for the linear response of a structure [36]. Instead, the higher-order statistics are assumed to be already similar. Two methods can be

distinguished: normal-condition alignment (NCA) and normal correlation alignment (NCORAL). These methods are shown capable of adapting both numerical (shear structures) and real heterogeneous populations (Bookshelf experimental case). SA maintains the structure of the original feature space, as it is limited to affine transformations being useful for physically-interpretable features, which are common in SHM, because features often correspond to some physical process; for example, an increase in natural frequency can be interpreted as a stiffening effect or decrease as an occurrence of damage, as developed in present work. The limitations of SA will be also discussed. It is important to introduce two key objects in TL, a domain and a task:

- A domain D=X, P(X), defined by of a feature space X and a marginal probability distribution P(X).

- A task for a given domain is defined by $T = Y, f\cdot$, where Y is the label space and $f\cdot$ is a predictive function learnt from a finite sample $\{x_i, y_i\}_{i=1}^{n}$, where $x_i \in X$ and $y_i \in Y$.

In unsupervised DA, a source domain $D_s = \{x_{s,i}, y_{s,i}\}_{i=1}^{n_s}$ with n_s source instances $x_{s,i}$, each with labels $y_{s,i}$ and a target domain $D_t = \{x_{t,j}\}_{j=1}^{n_t}$ with n_t unlabeled target instances, $x_{t,j}$, are used to learn a classifier that generalizes to the target domain. The premise of DA is grounded in the observation that there can be discrepancies between the marginal distributions, $P(X_s) \neq P(X_t)$ and/or the conditional distributions, $P(y_s|X_s) \neq P(y_t|X_t)$, of the source ($D_s$) and target ($D_t$) domains. Consequently, DA focuses on discovering a transformation that harmonizes these data distributions.

It could also be possible to investigate partial DA problems, where the available target data pertains to fewer classes than the source, that is, the target label space is a subset of the source $Y_t \subset Y_s$. For example, in the experimental case analyzed, consider a discrete damage localization problem with three locations: the source may have data for all three locations, whereas the target only has data pertaining to two locations. The aim of using partial DA would be to use the data in the source domain to learn what location the damage in the target relates to.

It is interesting to see the scheme presented by Poole et al. [13] in Figure 4.1, in which a schematic representation of aligning all the available data in the context of partial DA and class imbalance is shown. It is demonstrated how the typical DA approaches naively attempt to align the marginal distributions, not correctly aligning the underlying distributions.

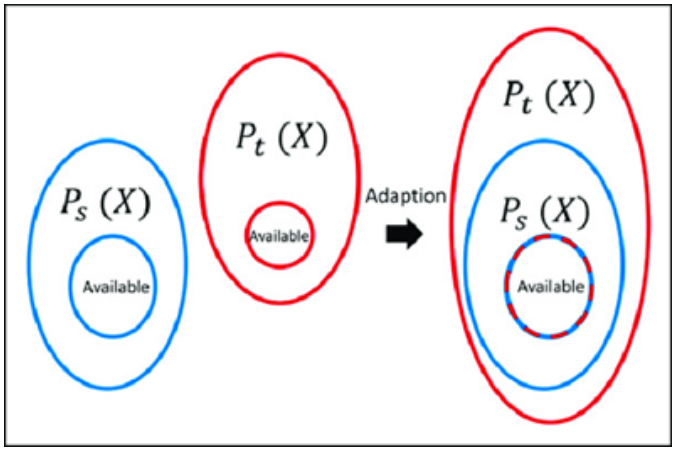

Figure 4.1: Domain Alignment in the context of partial DA with class imbalance or negative transfer, [13].

4.2 Normal Condition Alignment and Normal Correlation Alignment

Standardization, often employed in traditional machine learning, ensures each feature is treated uniformly. Within DA, it offers a means to match the means (μ) and standard deviations (σ) of the marginal distributions $P(X_s)$ and $P(X_t)$, facilitating this alignment through unsupervised transformations as shown in equation 4.1 and equation 4.2,

$$z_s^{(j)} = \frac{x_s^{(j)} - \mu_s}{\sigma_s} \tag{4.1}$$

$$z_t^{(j)} = \frac{x_t^{(j)} - \mu_t}{\sigma_t} \tag{4.2}$$

where μ_s and μ_t are the means of the source and target; σ_s and σ_t are the respective standard deviations.

One of the most popular SA methods is correlation alignment (CORAL), which aligns the source correlation with the target.

Correlation Alignment (CORAL), extends this method to also align the covariance. This is achieved by transforming the source domain via a linear transformation matrix A, see equation 4.3.

$$\min_A \|C_s - C_t\|_F^2 = \min_A \|A^T C_s A - C_t\|_F^2 \tag{4.3}$$

where $\widehat{C_s}$ is the covariance of the transformed source, C_s and C_t denote the covariance matrices of the source and target, respectively, and $\cdot_F$ is the Frobenius

norm. In the application to the model, in order to find the matrix A. Given source domain data D_s and target domain data D_t where n_s and n_t are the numbers of samples in the source and target domains, respectively, the aim is to find a transformation matrix A such that the transformed source domain $\widehat{D_s} = D_s \cdot A$ has similar second-order statistics to the target domain X_t. The steps done are the following:

- Covariance Calculation: Calculate the covariance matrices of the source and target data, C_s and D_t, respectively. To ensure the covariance matrices are positive definite (and thus invertible), a regularization term (identity matrix I) is added: $C_s = \text{cov}(D_s^T) + \lambda I$ and $C_t = \text{cov}(D_t^T) + \lambda I$, where λ is a small positive value (set to 1 for simplicity).
- Square Root of Covariance Matrices: Find the Cholesky decomposition of C_s and C_t, which are the matrices $C_s^{1/2}$ and $C_t^{1/2}$ respectively. The Cholesky decomposition is used here to efficiently compute the square root of the positive definite covariance matrices.
- Finding Transformation Matrix A: The transformation matrix A is then found by solving for A that aligns $C_s^{1/2}$ to $C_t^{1/2}$, which is done by: $A = (C_s^{1/2})^{-1} \cdot C_t^{1/2}$.

4.2.1 Normal-Condition Alignment (NCA)

Structural health monitoring datasets commonly are likely to have class imbalance, because it is not realistic to assume the target structure will contain data from each health-state present in the source structure. This underlines the need for applying partial DA. There are two main reasons: statistics from different domains encapsulate distinct behaviors, and aligning these domains based on statistical moments will not properly align the underlying distributions, potentially causing negative transfer. Furthermore, conventional DA techniques generally focus on minimizing the differences in marginal data distributions, which might lead to improperly matching a subset of target classes with the entire distribution of the source. To address these issues, Normal Condition Alignment (NCA) is introduced. NCA aims to mitigate the risk of aligning data that originates from different processes by leveraging the assumption that data collected at the beginning of a structure's operation are produced under the undamaged state. In NCA, the source domain is first standardized to center the data and give features equal treatment. The normal-condition (health state) of the target domain is then aligned with that of the source as shown in equation (4.4).

$$z_t^{(j)} = \left(\frac{x_t^{(j)} - \mu_{t,n}}{\sigma_{t,n}} \right) \sigma_{s,n} + \mu_{s,n} \tag{4.4}$$

where $\mu_{s,n}$, $\mu_{t,n}$ are the means and $\sigma_{s,n}$, $\sigma_{t,n}$ are standard deviations of the normal-condition data for the source and target, respectively. If variation between the datasets is assumed to be limited to scale vector $\boldsymbol{a}$, and translation vector $\boldsymbol{b}$, the differences can be expressed by the following equation (4.5):

$$X_s = \boldsymbol{a}X_t + \boldsymbol{b} \tag{4.5}$$

In this context, the dataset X is divided into two subsets: X_n, which contains data under normal conditions, and X_d, which contains data representing a damaged state or condition. Affine transformations, denoted as $\boldsymbol{a}$ and $\boldsymbol{b}$, are specific types of linear mappings used for data transformation. Affine transformations include translation, scaling, and rotation. It implies leveraging the normal-condition data to understand and model how data transforms within the correspondent space, assuming that the foundational relationships within the normal data (X_n) hold true or are similar enough to those in the damaged data (X_d).

The principal goal of DA is to adapt the underlying distributions so that a given health-state "c" from each domain follows the same conditional distribution $P_s(Y = c|X) = P_t(Y = c|X)$. The difficulty relies on the fact that, given that the data distributions for a finite sample will contain biases because of class imbalance and differences in the label spaces. Previous SA and typical DA approaches would attempt to align the marginal distributions, which would not correctly align the underlying distributions. Therefore, to align the underlying distributions, a subset of the source label space must be chosen, and the quantity of data in each class should be balanced. Explicitly aligning the marginal distributions of data which are believed to have been generated by the normal condition is a low-risk way of selecting data corresponding to the same label space. The way to solve class imbalance for NCA is by only considering data from the normal-condition.

4.2.2 Normal-correlation alignment (NCORAL)

Correlation alignment may also be prone to negative transfer under class imbalance, so a modification of CORAL is introduced: exploit information in the correlation between the "normal" or "healthy state" data. This method is termed **Normal-correlation alignment (NCORAL)**.

In order to apply **NCORAL** to the features of the dataset, firstly, **NCA** should be applied, as was shown previously. **NCORAL** is then given by the optimization problem shown in equation (4.6):

$$\min_A \|C_{s,n} - C_{t,n}\|_F^2 = \min_A \left\|A^T C_{s,n} A - C_{t,n}\right\|_F^2 \tag{4.6}$$

where $C_{s,n}$ is the correlations of the normal condition of the source, and $C_{t,n}$ is that of the target. **NCORAL** assumes that both domains differ by a translation

and a scale made by **NCA**, and learns the correlation from a subset of the entire dataset.

4.3 KNN algorithm

The k-Nearest Neighbors (KNN) algorithm is a fundamental method used in machine learning for classification and regression tasks. it is a non-parametric method, meaning it does not make any assumptions about the underlying data distribution. KNN works by finding the k closest training examples in the feature space to a given test point, and then making predictions based on the majority vote (for classification).

There is the need for "training data," which corresponds to a labeled dataset with known outcomes. Each data point has features (attributes) and a corresponding class label (for classification). In our models, the features correspond to natural frequencies, and the label to the damage condition. The data point for which you want to make a prediction is called "testing data". It has features but no label.

In the algorithm, the parameter k: The number of nearest neighbors to consider when making predictions, is a hyperparameter that you need to choose before applying the algorithm. Common choices include odd numbers to avoid ties in majority voting. For each data point in the training set, the algorithm calculates the distance between the features of the new data point and the features of each training data point. Common distance metrics include Euclidean distance, Manhattan distance, or other distance measures based on the nature of the data. In this way, the KNN algorithm can select the label for the test data.

Using the toolbox of the software utilized, it was necessary to prepare the data, with the previous normalization through Statistical Alignment, then specifying the number of neighbors k equal to 5, considered enough for the case studied, and considering a Euclidean distance, it was possible to obtain the prediction of the labels of the testing data.

A lot of investigations explore the field of this algorithm [13, 37] and was widely used in the field of machine learning.

There's ongoing research into improving KNN's efficiency and effectiveness. This includes developing better methods for choosing k, investigating more sophisticated distance metrics, exploring the impact of feature selection and transformation, and integrating KNN with other machine learning models in ensemble methods.

In summary, while KNN is a straightforward and versatile algorithm, its performance and efficiency can be significantly influenced by the choice of k, the distance metric, the dataset's dimensionality, and the computational resources available. Continuous research efforts aim to extend its applicability and enhance its performance across various tasks and datasets.

It is possible to emphasize some topics done during the application of SA and KNN to the numerical population, before discussing about the performance and results 5. It was mentioned that the total samples calculated are 800 for the source, 400 for training target and other 400 samples for the test target domain. Before applying the NCA and NCORAL methodologies, it was necessary to calculate the mean value and standard deviation, for the all the samples of the source. So applying equation 4.1 it was obtained the first standardization. The same was performed for the target training samples, using equation 4.2. Finally for the test target, was used same formulation but with mean value and standard deviation correspondent to the test samples, as presented in equation 4.7.

$$z_t est^{(j)} = \frac{x_t^{(j)} - \mu_t}{\sigma_t} \tag{4.7}$$

Then, there were calculated the mean value and standard deviation correspondent to the normal condition (class 0) for source and target training domain. So, only selecting the values with label 0: 200 samples for the source and for 100 samples for target. Practically speaking the values calculated in this step are: $\sigma_{s,n}$ and $\mu_{s,n}$.

Finally, equation 4.4 is applied to the target test data, obtaining the standardize values according to the NCA methodology. After these, the samples already standardize are taken to calculate the value of the matrix "A" with the equation 4.6 and methodology explained. Finally, by simple, multiplying this matrix "A" to all the dataset, the methodology NCORAL is applied, remembering that this matrix permits the "rotation" of data. The predictions of the label are calculated with KNN algorithm, and then compared with the correspondent labels of the target.

4.4 Performance Parameters

An important parameter used to identify the performance of the classification is the confusion matrix. It could be understood as a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. It allows for easy identification of confusion between classes. Most of the time, almost all performance measures are computed from it.

4.4.1 Definitions

Let c_i be any class. Following are the definitions of TP, FP, FN, and TN for c_i (is the actual label for a feature):

- $TP(c_i)$ = number of times when c_i is classified as c_i.
- $FP(c_i)$ = quantity of times when a non-c_i is classified as c_i.

- $FN(c_i)$ = all instances when c_i are not classified as c_i.
- $TN(c_i)$ = all the non c_i cases that are not classified as c_i.

4.4.2 Metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.8}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.9}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.10}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \cdot \frac{TP}{2 \cdot TP + FP + FN} \tag{4.11}$$

The values of the true positive rate and false positive rate for each class ranges from 0 to 1. Obtaining an F1 score equal to zero indicates the worst possible performance of a classification model. An F1 score of 1 represents the best possible performance of a classification model.

This performance criteria are widely used, and present in a lot of investigative works [13, 37, 38].

Chapter 5

Results of the numerical population

The current Chapter presents the numerical results, showing the performance of the SA algorithm on a 2D case study and a population of structures simulated using 3D models.

The bi-dimensional case study includes one source and one target, which are two different structures presenting differences in the number of floors, materials and geometries. This case has been investigated to study the potential of SA in improving damage identification over traditional standardization and pre-processing methods on a simplified simulated pair of source and target systems. The extracted features, consisting of a set of three natural frequencies, are presented in tabular form and in pair-plots, to highlight the differences before and after standardisation. Afterwards, the damage identification performance is estimated adopting the confusion matrix and F1 score, before and after the standardization, to highlight the effect of standardization.

Subsequently, the 3D model is used to simulate a population of nine structures, as described in Chapter 3 and Table 3.3. For this case study, the overall results are shown to assess performance trends within the population. In addition, two cases of source-target structures are presented in more detail. Analogously to the previous case study, the extracted natural frequencies are presented in tabular form and in pair-plots to highlight the effect of SA. Tow damage identification tasks are considered: the first one regards damage detection and the second one regards damage localization. All the illustrated natural frequencies are the features selected with the selection procedure explained in section 3.2.2. Thus, they do not corresponds to the first, second or third natural frequencies of the different structures. The performance is evaluated though the mean value of the F1 scores obtained for all the possible source and target combinations within the population.

For both models the standardization consists firstly in the implementation of NCA and secondly in NCORAL in order to calculate the transformation matrix A. The discussion and conclusions will highlight the most important points for the development of this methodology. In addition is important to emphasize that the analysis was performed ten different times, and the performance parameters obtained comes from the mean value of the repeated analysis to reduce the influence of sampling bias.

5.1 Knowledge sharing between two 2D models

The analysis is performed for two structures whose properties are presented in Table 5.1.

Properties		
	Source	Target
Number of floor	4	5
Material	Aluminum	Steel
Geometry	Case 1	Case 2

Table 5.1: 2D:Case of analysis

The natural frequencies and mode shapes of the structures are analyzed. The properties are selected and the eigenvalue analysis is performed.

The features are presented in Table 5.2 for the source structure and in Table 5.3 for the target structure. In these Tables, the mean value for all samples of each damage conditions are presented to show their variability.

Natural Frequencies (Hz)			
			Label
1.685	4.876	7.472	0
1.676	4.856	7.455	1
1.678	4.875	7.443	2
1.684	4.857	7.436	3

Table 5.2: Natural Frequencies of SOURCE (Hz)

Natural Frequencies (Hz)			
			Label
5.798	16.936	26.669	0
5.752	16.825	26.580	1
5.759	16.927	26.635	2
5.794	16.856	26.484	3

Table 5.3: Natural Frequencies of TARGET (Hz)

The mode shapes obtained from the eigenvalue analysis of the structure in the undamaged conditions are reported in Figure 5.1 for the source and Figure 5.2 for the target.

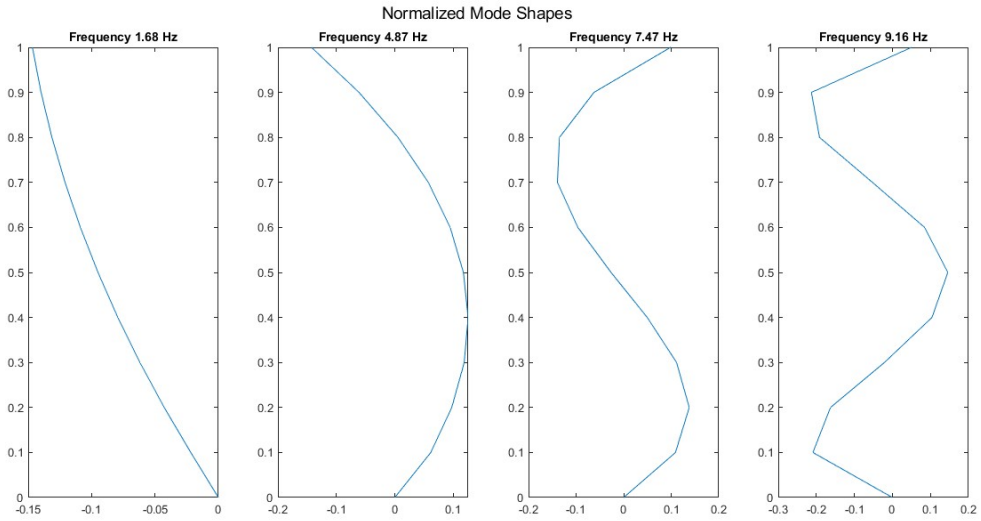

Figure 5.1: Normalized Mode Shapes of Source Structure.

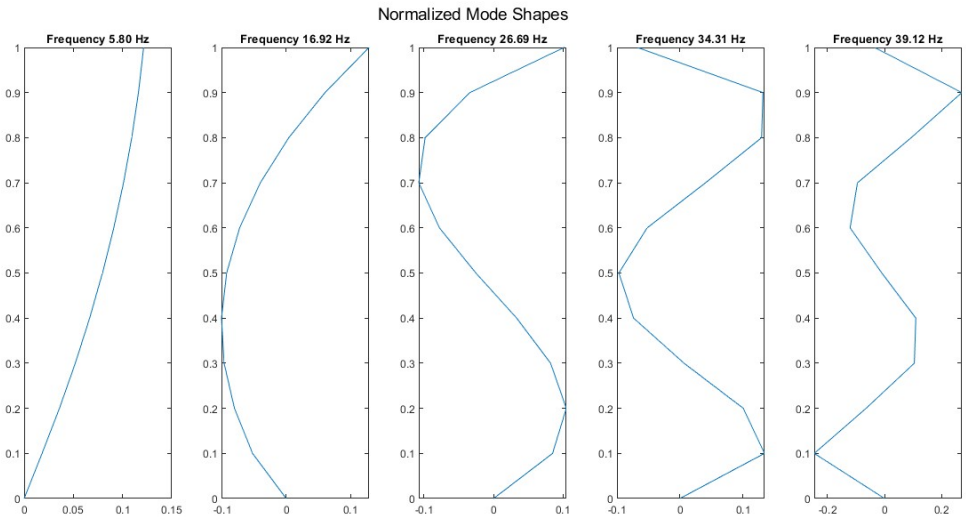

Figure 5.2: Normalized Mode Shapes of Target Structure.

The normal condition simulation presents higher natural frequencies, as the stiffness corresponds to the original set by the numerical construction of the structure. Then, it is possible to see a reduction of the natural frequencies from the normal condition case due the effect of damage in the different degree of freedom for this 2D model. From the observation of the mode shapes, it is possible to describe the behaviour of the structure. From left to right, the quantity of nodes increase from one to four in the case of the source and from one to five in the case of the target, representing the first four and five natural frequencies.

The pair-plots are presented in Figure 5.3 for the source and Figure 5.4 for the target. In the second pair-plot is possible to observe the natural frequencies of the target structure. In comparison with the source, the structure is characterised by a different material, geometry and one more floor. So, the differences with the source are evident.

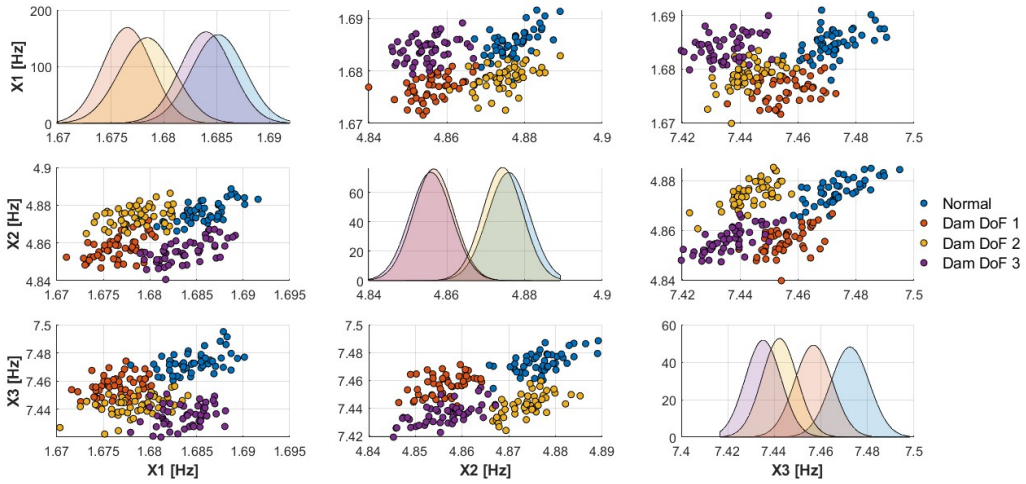

Figure 5.3: Natural Frequencies (Hz) - Source Structure 2D.
Note: For pair-plots a random subset 25% the size of the dataset was used for visualization.

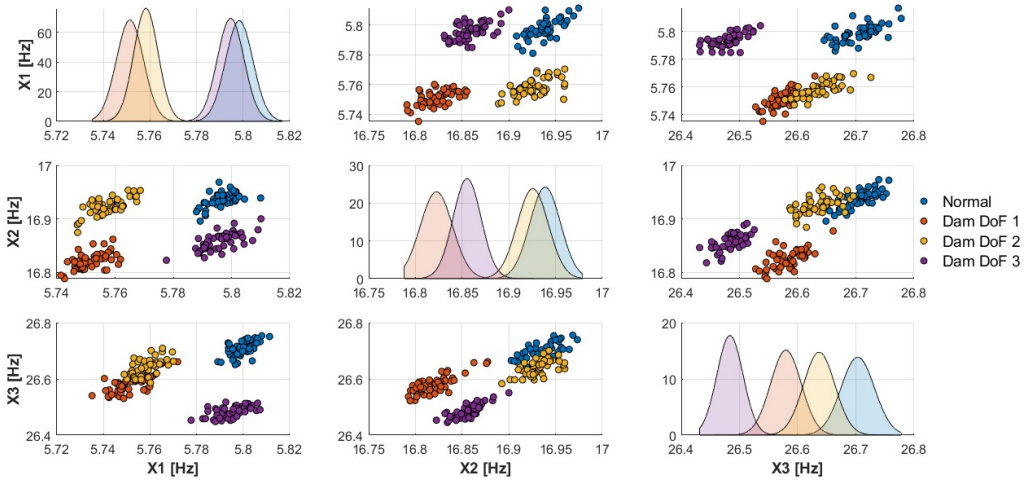

Figure 5.4: Natural Frequencies (Hz) - Target Structure 2D.
Note: For pair-plots a random subset 25% the size of the dataset was used for visualization.

The last pair-plot in Figure 5.5 corresponds to the combination of both datasets after the standardization. In this graph is also possible to see the different groups corresponding to the different labels and damage states.

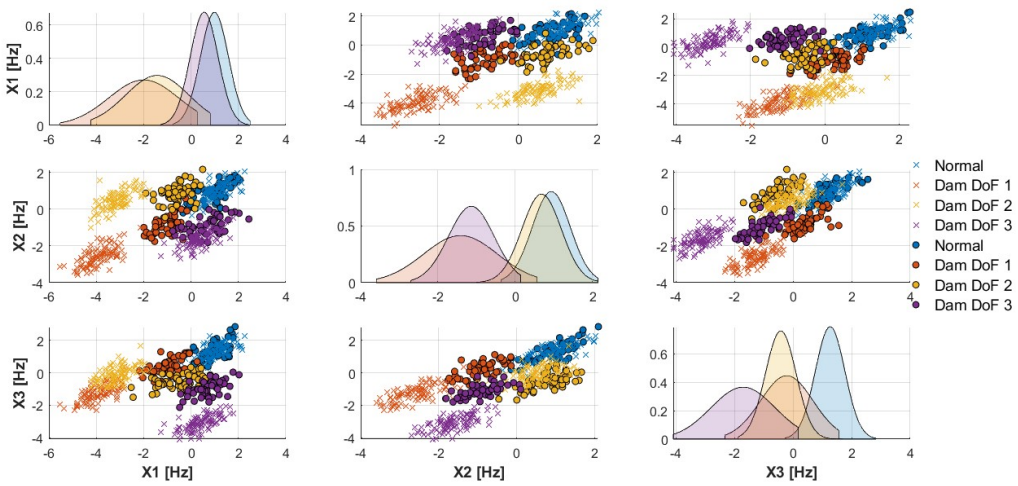

Figure 5.5: Standardized Natural Frequencies (Hz) - Source and Target Structure 2D

Note: indicating with "o": source domain and with "x": target domain

The performance is presented in Table 6.4 through the resultant mean confusion matrix related to localization of damage. Each row of the matrix represents the instances in an actual class, while each column represents the instances in a predicted class. Out of 100 instances of class 1, 96 were correctly identified as class 1 (true positives for class 1) and 4 were incorrectly identified as class 2.

All 100 instances of class 2 were correctly identified as class 2. This indicates perfect classification for class 2. For class 3, 68 out of 100 instances were correctly classified, but 32 instances were incorrectly identified as class 2. None were misclassified as class 1 or class 4. All 100 instances of class 4 were correctly identified as class 4. This indicates perfect classification for class 4.

Confusion Location Matrix			
96	4	0	0
0	100	0	0
0	32	68	0
0	0	0	100

Table 5.4: Confusion matrix for location

The classifier perfectly identifies classes 2 and 4, with no misclassifications. Instead, the classifier struggles slightly with class 3, where a significant number of instances are misclassified as class 2. There is a minor mislabeling between class 1

and classes 2 and 3, but the majority of class 1 instances are correctly classified. In summary, this confusion matrix suggests a satisfactory performance of the KNN classifier after SA.

The value of F1 score obtained is 0.89. This result suggests that the model has a high level of accuracy in terms of both precision and recall. Indeed, this score indicates a good balance between precision and recall. It implies that the model does not significantly favor either avoiding false positives (precision) over capturing all positives (recall) or vice versa.

The damage identification task is additionally performed without SA for comparison purposes. In this case, all 100 instances that are actually of class 1 are correctly predicted as class 1. However, all instances that are actually of class 2, 3, and 4 were incorrectly predicted as class 1, leading to an F1 score of 0.25. This indicates a severe issue with the classifier where it is highly biased towards class 1, to the extent of completely ignoring the features from the damaged conditions. This shows the importance of using a correct pre-processing and alignment strategy.

5.2 Knowledge sharing between a population of 3D models

This section reports the results of the knowledge sharing process applied to the population of nine 3D structures. All the combinations of source and target structures are considered, leading to 72 pairs of source and target structures.

First, two transfer learning case studies are described in detail. Subsequently, the the section reports the overall results of the knowledge sharing process for damage identification within the population.

5.2.1 First case study

The first case analyzed corresponds to the source structure N°4 and the target structure N°2 from Table 3.3. The properties of the structures analyzed as first 3D case are presented in Table 5.5.

Properties		
	Source	Target
Number of floor	4	3
Material	Aluminum	Steel
Geometry	Case 1	Case 2

Table 5.5: 3D:First case of analysis

The extracted natural frequencies from the source and target domain are presented in Table 5.6 for the source and Table 5.7 for the target. The values corresponds to a mean value, in which is possible also to observe the correlated behaviour with the pair-plots shown subsequently.

Natural Frequencies (Hz)			
			Label
1.694	4.878	7.474	0
1.686	4.860	7.459	1
1.688	4.878	7.445	2
1.693	4.860	7.438	3

Table 5.6: Natural Frequencies of SOURCE - 1° case(Hz)

Natural Frequencies (Hz)			
			Label
9.072	25.420	36.733	0
8.964	25.227	36.652	1
9.003	25.364	36.303	2
9.051	25.109	36.462	3

Table 5.7: Natural Frequencies of TARGET - 1° case(Hz)

The resulting pair-plots are presented in Figure 5.6 and 5.7, for source and target respectively.

In both cases, the differences in the features sampled from different damage cases can be observed.

Specifically, these figures highlight how the two structures are affected differently by the presence of simulated damage, due to the variations in number of floors, geometry and material.

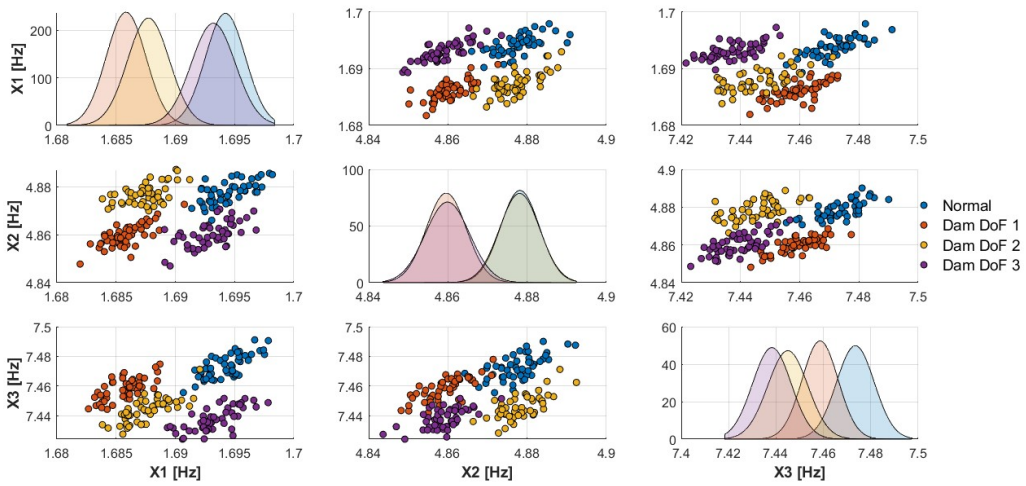

Figure 5.6: Natural Frequencies (Hz) - Source Structure - 1° 3D case study.
Note: For pair-plots a random subset 25% the size of the dataset was used for visualization.

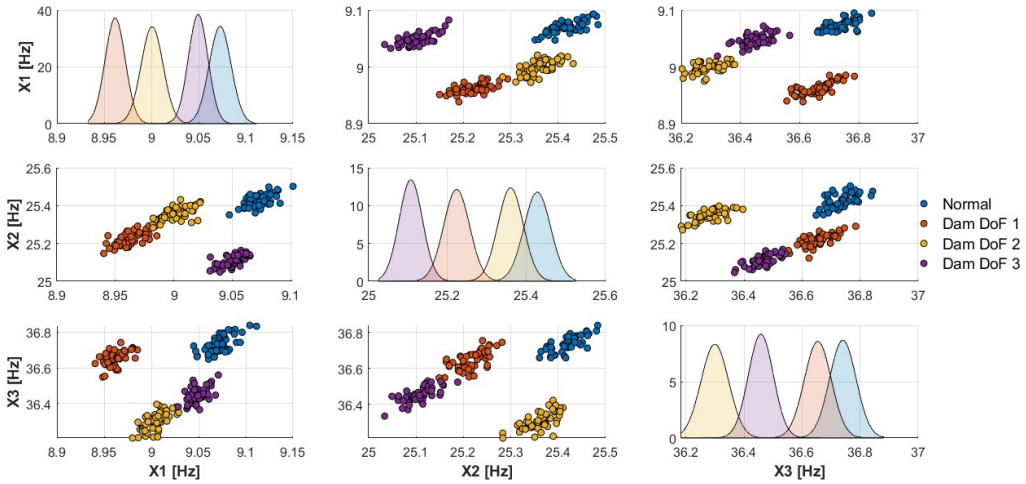

Figure 5.7: Natural Frequencies (Hz) - Target Structure - 1° 3D case study.
Note: For pair-plots a random subset 25% the size of the dataset was used for visualization.

The resultant mean matrix obtained is presented in Table 5.8. This is calculated to analyze the performance of localization of damage, thus, all four labels were considered.

The classifier performs a quasi-optimal classification. Only for label 0 (healthy state) and label 2 (damage on the second floor), there is one sample that is not correctly identified. This leads to an F1 score of 0.99.

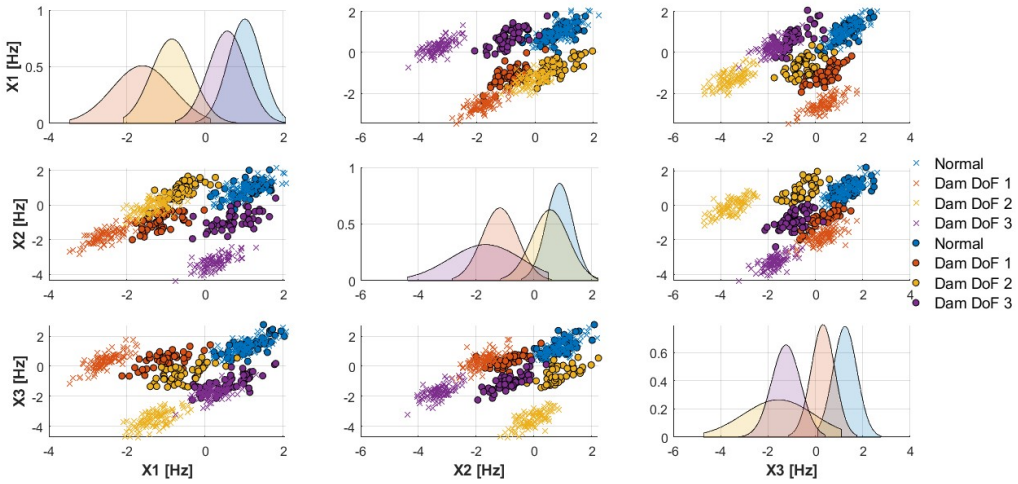

Figure 5.8: Standardized Natural Frequencies - Source and Target Structures - 1° case 3D

Note: indicating with "o": source domain and with "x": target domain

Confusion Location Matrix			
99	1	0	0
0	100	0	0
0	0	99	1
0	0	0	100

Table 5.8: First case: Confusion matrix for location

For the case of damage detection, the samples from classes 1, 2 and 3 are re-labelled using $y = 1$ to create a binary problem, in which $y = 0$ is used for the samples from undamaged conditions and $y = 1$ is used to indicate damage.

The results are presented in Table 5.9. The sum of all elements inside the matrix it is 400, which corresponds to the sum of all samples in the target test domain.

It can be observed that all the 300 samples from the simulated damaged conditions are correctly detected.

A solid performance is achieved also for the 100 samples from the undamaged conditions. Indeed, only 4% of the samples are misclassified as they are labeled as damaged.

Confusion Detection Matrix	
96	4
0	300

Table 5.9: First case: Confusion matrix for detection

The mean value of F1 score obtained for detection is 0.99. This is very high value, meaning a almost perfect detection of damage, as was intuitive from the analysis of the confusion matrix aforementioned.

5.2.2 Second case study

The second case examined corresponds to the source structure N°1 and target N°8 (see Table 3.3). The properties are displayed in the following Table 5.10.

Properties		
	Source	Target
Number of floor	3	5
Material	Aluminum	Steel
Geometry	Case 1	Case 2

Table 5.10: 3D:Second case of analysis

The correspondent results are presented, as another case inside the population analysis.

The mean values for each label are displayed for source in Table 5.11 and for the target 5.12.

By analyzing Figures: 5.9 and 5.10 is possible to observe the features for source and target domains. For a visual analysis regarding the standardize values obtained after the application of NCA and NCORAL, is presented the pair-plot in Figure 5.11.

Natural Frequencies (Hz)			
			Label
2.171	6.083	8.791	0
2.158	6.060	8.780	1
2.163	6.076	8.738	2
2.168	6.045	8.756	3

Table 5.11: Natural Frequencies of SOURCE - 2° case(Hz)

Natural Frequencies (Hz)			
			Label
5.803	16.939	26.702	0
5.756	16.823	26.582	1
5.763	16.928	26.636	2
5.800	16.858	26.487	3

Table 5.12: Natural Frequencies of TARGET - 2° case(Hz)

The distinction between the different labels, in the case of the target is more evident than in the source. In other words, the effect of damage is more visible, even having the same position and intensity respect to the source.

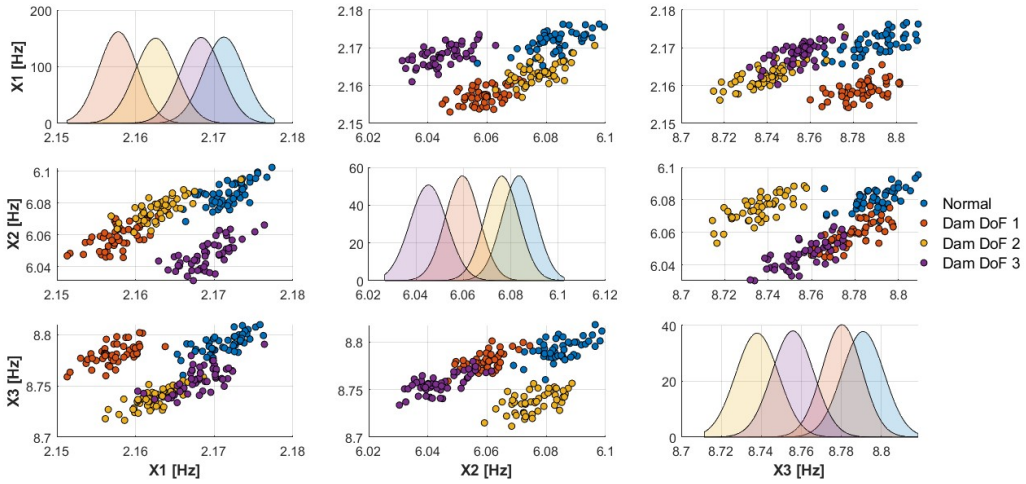

Figure 5.9: Natural Frequencies (Hz) - Source Structure - 2° case 3D
Note: For pair-plots a random subset 25% the size of the dataset was used for visualization.

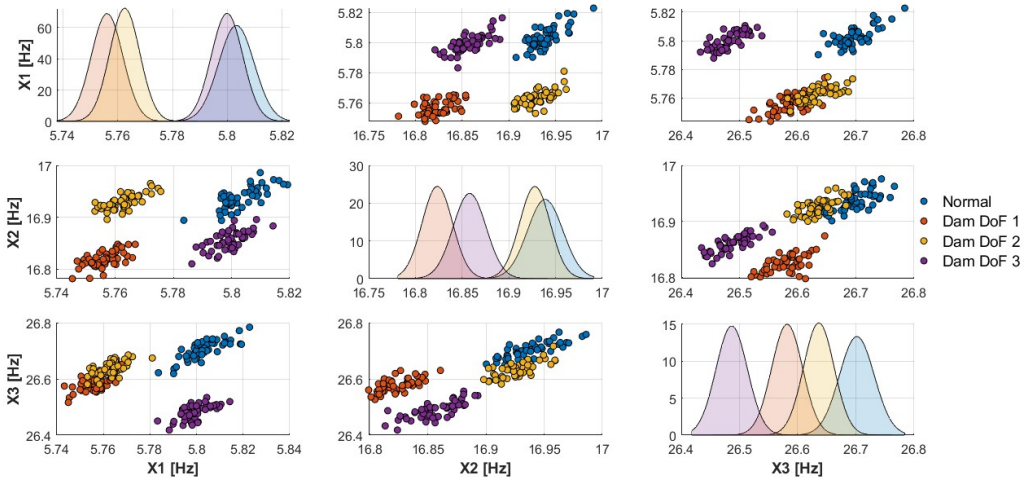

Figure 5.10: Natural Frequencies (Hz) - Target Structure - 2° case 3D
Note: For pair-plots a random subset 25% the size of the dataset was used for visualization.

The localization confusion matrix obtained is a mean value, as performed previously, and in this case it also reveals good results. Is presented in Table 5.13.

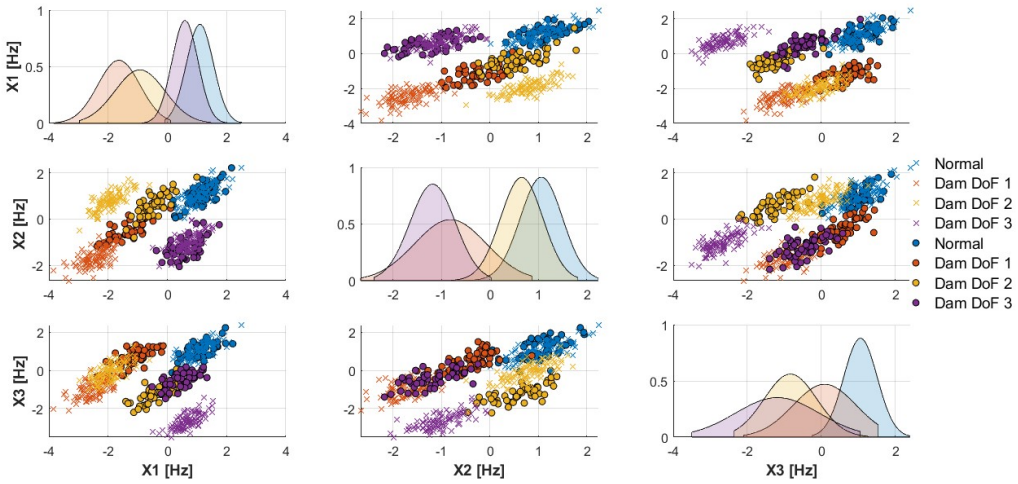

Figure 5.11: Standardized Natural Frequencies - Source and Target Structures - 2° case 3D
Note: indicating with "o": source domain and with "x": target domain

The problems in classifications are for label 2, in which only 10 cases were classified correctly, the others were confused with the label 1 (90 cases). For actual label 0, there were only two incorrect predictions, in which one was predicted as label 1 and the other as label 2. For the label 1 instead, the estimations were all correct. For label 3, 88 features were correctly identified in the correct way, considering just 12 samples with label 2.

Confusion Location Matrix			
98	1	1	0
0	100	0	0
0	90	10	4
0	0	12	88

Table 5.13: Second case: Confusion matrix for location

The mean value obtained for the F1 score for location is 0,76, which indicates a good balance between precision and recall, suggesting that the model is relatively successful at correctly identifying the labels. This value indicates a gap to be improved. But it results in coincidence with the high difference in properties between the structures considered as source and target. In fact, two structures having 2 floor of disparity between them, is a high variation in the structural responses.

In this second case, as well as was done previously, it was estimated the mean confusion matrix for detection. The resultant matrix is very satisfactory. Also in this example, there are 3 undamaged features (with label 0) considered as damaged (label 1, 2 or 3). C matrix for detection is presented in Table 5.14.

Confusion Detection Matrix	
97	3
0	300

Table 5.14: Second case: Confusion matrix for detection

Definitely, the mean value obtained regarding F1 score for detection is 0.99, which put in manifest the almost perfect performance of the SA algorithm and the KNN machine learning algorithm for this specific case.

5.2.3 Population results

The previously analyzed case studies are specific cases exposed to manifest the different outputs of the analysis performed. Thus, this section presents a summary of the overall performance in the population. The two damage identification tasks, detection and localization, are considered. This results allow to acquire a general view and more complete understanding of the performance of SA across an heterogeneous population. Inside these matrices, each row is a structure in the source domain and each column of the target.

F1 score - detection								
1.00	1.00	1.00	0.97	1.00	1.00	0.82	0.98	1.00
0.74	1.00	1.00	0.72	1.00	1.00	0.72	0.87	1.00
0.74	0.75	1.00	0.74	0.75	0.90	0.72	0.75	0.75
0.99	1.00	1.00	0.99	1.00	1.00	0.92	1.00	1.00
0.84	1.00	1.00	0.77	1.00	1.00	0.74	0.99	1.00
0.75	0.99	1.00	0.75	0.75	1.00	0.74	0.75	0.99
0.98	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99
0.87	1.00	1.00	0.93	1.00	1.00	0.83	1.00	1.00
0.81	0.99	1.00	0.75	0.97	0.99	0.75	0.76	0.99

Table 5.15: Population results: F1 score for detection

As an example, the row number one, represents the cases in which the source was considered as the structure number one (N°1) in Table 3.3. And column two,

F1 score - location								
1.00	1.00	1.00	0.95	0.99	1.00	0.61	0.76	0.74
0.49	1.00	0.75	0.47	1.00	0.84	0.47	0.87	0.79
0.25	0.25	1.00	0.25	0.25	0.80	0.25	0.25	0.62
0.98	0.99	0.98	0.99	1.00	1.00	0.84	0.92	0.88
0.61	1.00	0.79	0.53	1.00	1.00	0.50	0.99	0.94
0.25	0.49	1.00	0.25	0.25	0.92	0.25	0.25	0.84
0.74	0.68	0.52	0.96	0.95	0.82	0.96	0.99	0.93
0.65	0.99	0.58	0.72	1.00	0.96	0.58	1.00	0.99
0.32	0.49	0.99	0.25	0.47	0.97	0.25	0.26	0.94

Table 5.16: Population results: F1 score for location

refers to all the cases in which the structure number two (N°2) was considered as target domain. The results for the detection and localization analysis, respectively are presented below in Table 5.15 and in Table 5.16.

For the detection task, the lower value obtained is 0.72. This is a value resulted for the case in which the source is the structure N°2 and the target is structure N°4 and for the case in which the source is also the structure N°2 but the target in this case, is N°7. The main reason is the difference on the number of floors, in which basically the algorithm is treating with data from two structures which are essentially very different. Apart from the number of floors the geometries and material are different. In other words, the structure number 2 finds the greatest difference among the population in reference to its properties, with structures 4 and 7. That is, it differs the most with them, therefore it is been analyzed damage detection between very different structures. Nevertheless, the performance parameter for detection, is not considered as low, in fact, it suggests a good balance between precision and recall in this context. The algorithm performance is reasonably effective, though there is room for improvement. On the other hand, the structure N°3 and structure N°6, considered as target domains in the population, present high performance. Meanwhile the structures N°4 and N°7 set as sources inside the entire population, also leads to strong performance.

For the localization task, the lower cases are presented in three sources: N°3, N°6 and N°9, reaching even values of 0.25. A F1 score of this value, suggests a relatively poor balance between precision and recall, it indicates that either precision, recall or both are low. Or the model is producing many false positives or is missing a significant number of true positives. In other terms, the model is not very effective at correctly identifying true instances of the positive class, and labelling too many instances as positive. It suggests a reconsideration of the approach and an improvement. But this results are not a coincidence. It is due

to the fact that the all the source cases contains the most diverse geometry of the rest of the population, as is Case 3, and this contrast becomes evident when it is analyzed with other geometries which are smaller.

Chapter 6

Experimental Case

6.1 General explanation of the model

Vibration-based SHM and damage assessment involves tracking the dynamic behaviors of structures to identify changes or patterns that indicate whether a structure is in a healthy or damaged state. This process also aids in pinpointing the location and severity of damage. The fundamental stages include selecting appropriate sensors, acquiring data, processing signals, extracting features, and identifying damage. This has led to the emergence of a variety experimental benchmarks, often developed for validating specific SHM strategies. The experimental case proposed in this Chapter aims to process the data obtained from the structure constructed in Los Alamos Laboratory and apply it in the context of Domain Adaptation, analyzing the performance of Statistical Alignment.

The background and time history data are available from the Los Alamos Laboratory [39]. The bookshelf structure was built of unistrut columns and aluminum floor plates with two-bolt connections to the brackets on the unistrut. Twenty-four single-axis piezoelectric accelerometers, with two assigned to each joint, were affixed to the aluminum blocks, distributed as eight per plate. The model is shown in Figure 6.1 and Figure 6.2.

The shaker was positioned at Corner D, connected through a stinger to a tapped hole at the midpoint of the base plate. Additionally, a force transducer was placed between the stinger and the base plate to measure the input from the excitation. The experimental study encompasses various conditions, including healthy and damaged states, featuring both single and multiple points of damage, as well as different degrees of damage severity.

Figure 6.1: Real Bookshelf Structure

6.2 Data available

The data available consists on acceleration time history and taken from the website. For each test scenario, data was gathered in three separate batches, corresponding to shaker input levels of 3, 5, or 7 volts. The frequency range of both the shaker's output and the system's response varied between 800 Hz and 3200 Hz for each test to identify the frequency band most sensitive to damage. In scenarios involving damage, bolts at the specified joint were first loosened and then hand-tightened, allowing for movement between the plate and the column (simulating damage level). The collected time signals consisted of 8192 data points each, sampled at a rate of 1600 Hz.

The data was differentiated in order to obtain three different labels 0, 1 and 3, indicating no damage or healthy state (label 0), damage at first floor (label 1) and damage at third floor (label 3). For each label, the data includes the time signals for the three conditions of the input level of the shaker and different damage levels.

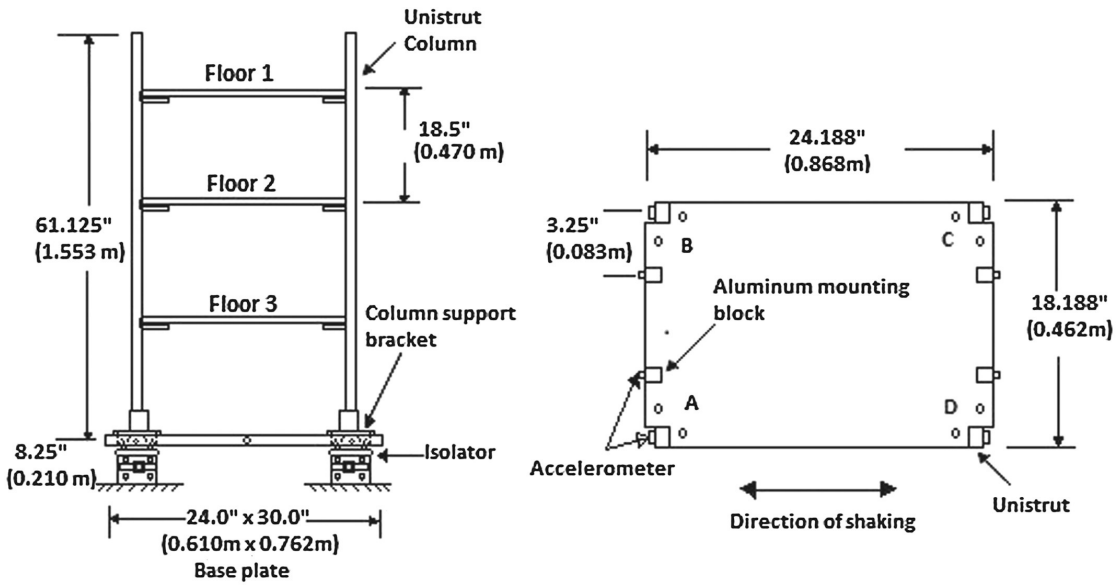

Figure 6.2: Scheme Bookshelf Structure
Note: Figure taken from [39]

6.2.1 Data processing and features selection

The acceleration data from the different scenarios was processed in order to obtain the modal parameters of the system. It was used the Stochastic Subspace Identification – Covariance (SSICOV) method to extract the natural frequencies from the time history data [40].

The function SSICOV set in the processing of the data, aims to automatically identify the natural frequencies, mode shapes and damping ratios of the Bookshelf structure using accelerations time history data. This algorithm of covariance-driven stochastic subspace identification method (SSI-COV) is inspired by the one used by Magalhaes et al. [41]. It has been applied for ambient vibration monitoring of the Lysefjord Bridge [42] and was compared to the frequency domain decomposition technique [43]. Finally, the algorithm was found accurate enough to visualise the evolution of the bridge eigen frequencies with the temperature [44].

At the heart of SSICOV is the Stochastic Subspace Identification method. SSI is a mathematical approach that models the structure's response as a combination of its inherent dynamic properties (such as natural frequencies, mode shapes, and damping ratios) and the stochastic (random) processes that represent the external forces acting on the structure. Essentially, SSI seeks to separate the signal (structural responses) from the noise (random external forces) in the time history data. The "COV" in SSICOV stands for covariance, a statistical measure used in this context to analyze the variability of the structural response data. By calculating

the covariance matrix of the measured data and applying SSI techniques, SSICOV constructs a state-space model. This model describes the dynamic behavior of the structure over time. Using the state-space model, SSICOV then identifies the system matrices that encapsulate the dynamics of the structure. Through mathematical transformations and eigenvalue analysis of these matrices, SSICOV is able to extract the natural frequencies of the structure.

From the identified natural frequencies, three have been selected as features for the knowledge transfer. In this case, the feature selection approach based on MAC comparison was not used due to the lack of complete mode shapes from the available experimental case study. However, the features could be selected based on the physical knowledge on the structure, and the vibration modes obtained from the numerical simulation set with a shear frame 3D model, characterised by three floors, aluminum and case 3 as material and geometry properties. Specifically, the feature selection has been performed based on the expected natural frequency values for the undamaged case (21 Hz, 60 Hz, 114 Hz).

After processing the data and obtaining the natural frequencies, it was possible to consider this case as a new target structure in the domain adaptation field. The total number of samples obtained for the target structure was 240.

6.3 General PBSHM application

The natural frequencies obtained were processed in order to prepare the samples in a matrix form with same format as the population numerical case in 3D model. This allows to just consider the experimental case as a new case in the dataset and perform the same analysis.

The source and target domains utilized for the analysis are presented in Table 6.1. The source domain is obtained by numerically modelling a shear frame 3D structure, meanwhile the target domain is composed by the natural frequencies obtained from the processing of the acceleration data from Bookshelf structure.

The target domain is composed by the experimental natural frequencies. This domain is split in two new sets: testing and training. The training is used with the correspondent label and is used with the source in order to train the KNN algorithm. Before the application of this algorithm, the data have been standardized via the NCA and NCORAL.

Then, KNN has been trained using the data from the source structure and predict the labels of the target test dataset. Finally, the predicted labels have been compared the real labels to assess the acquired performance.

Properties		
	Source	Target
Number of floor	3	
Material	Aluminum	Bookshelf case
Geometry	Case 3	

Table 6.1: Case of analysis

Table 6.2 shows the mean value of the samples for label 0, label 1 and label 3. The total number of samples obtained for healthy state are 150, the total for damage at first floor are 30 and for damage at third floor 60.

Similarly, Table 6.3 shows average for each label of the source structure modelled with the 3D model, as presented in Chapter 3.2.2. The results are obtained as a mean value after repeating the analysis ten different times.

Natural Frequencies-Bookshelf (Hz)			
			Label
20.406	60.104	113.887	0
20.113	58.986	114.112	1
19.925	59.246	111.667	3

Table 6.2: Natural Frequencies - Bookshelf Structure (Hz)

The graphical visualization of the samples in pair-plots helps to understand the quality of the data.

As done in the previous cases, the features from the source and target structures are presented in Figure 6.3 and in Figure 6.4, respectively.

Natural Frequencies (Hz)			
			Label
20.820	58.334	118.059	0
20.383	58.585	117.901	1
20.538	58.094	117.346	2
20.733	57.106	118.597	3

Table 6.3: Natural Frequencies - Source Structure(Hz)

It can be noted how the target dataset presents a more complex behaviour than the simulated data.

In addition, it should be considered that the source dataset is composed of 800 samples and the target is composed of 240 samples. Therefore, this case study allows to test the SA performance with a target domain characterised by higher sparsity and lower data quality.

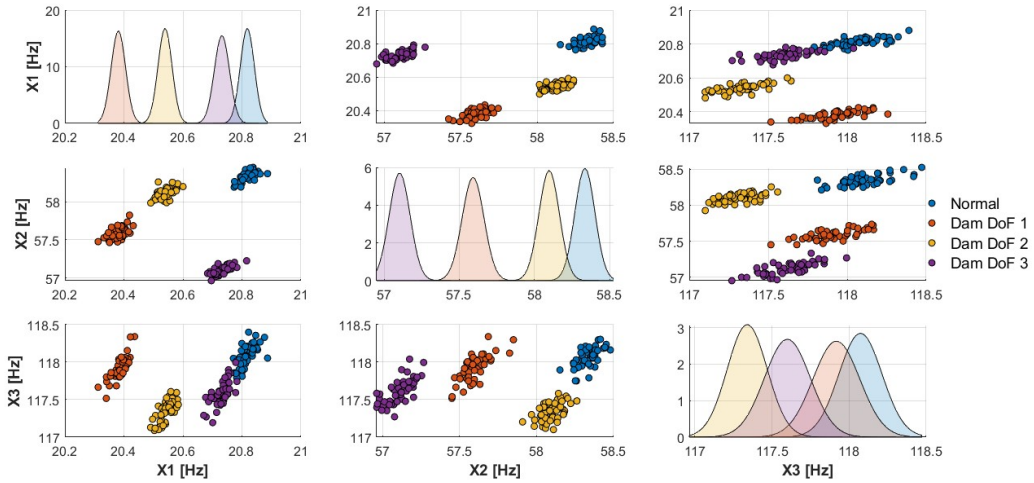

Figure 6.3: Natural Frequencies (Hz) - Source Structure

Note: For pair plots a random subset 25% the size of the dataset was used for visualization.

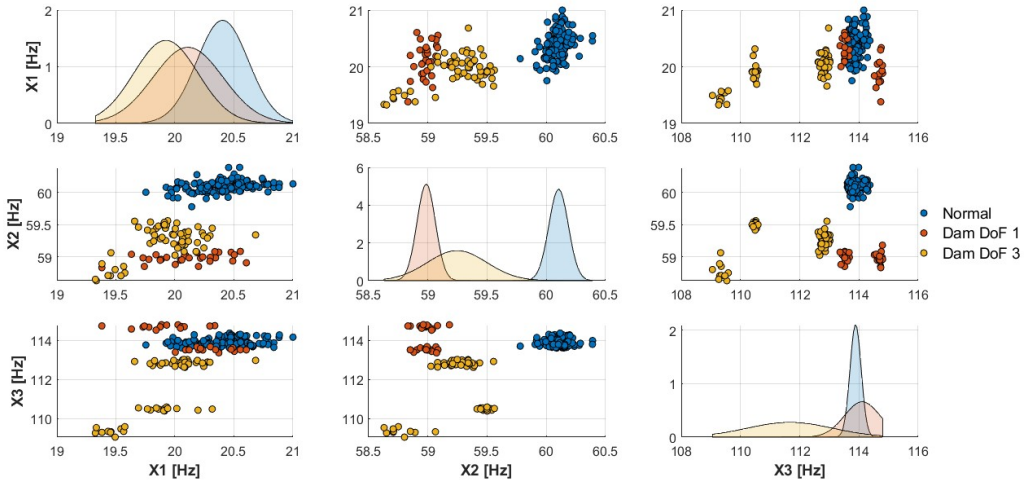

Figure 6.4: Natural Frequencies (Hz) - Target Bookshelf Structure

The standardize values are presented, with source and target domain superimposed. In Figure 6.5 it is possible to see the different groups or labels. For the normal condition, in source and target, the values are comprised in same interval.

Additionally, it is possible to observe how some data corresponding to the label 3 (violet color), are isolated from their respective group.

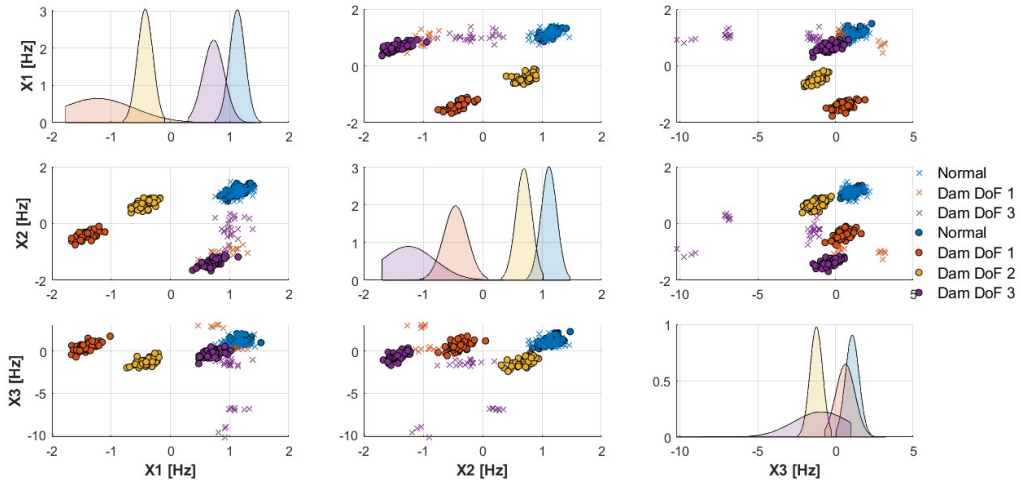

Figure 6.5: Standardized Natural Frequencies - Source and Target Bookshelf Structure

The averaged confusion matrix and F1 score allow to analyze the effectiveness

of SA for the experimental target structure. The results regarding the localization of damage are provided in Table 6.4 and the results regarding damage detection are detailed in Table 6.5.

Confusion Matrix			
75	0	0	0
3	2	0	10
0	0	0	0
0	0	13	17

Table 6.4: Confusion matrix for location

Confusion Matrix	
75	0
5	40

Table 6.5: Confusion matrix for detection

For the localization task, the model correctly predicted 75 instances of class 0. There were no instances of class 0 being incorrectly estimated as any other class. Additionally, the model correctly predicted 2 samples of class 1. However, it incorrectly calculated 3 instances of class 1 as class 0.

There were no predictions of class 1 as class 3. On the other hand, the model did not correctly deduced any instances of class 2 (0 predictions), because target does not present the label 2.

Finally the model correctly predicted 17 instances of class 3 but incorrectly estimated 13 instances of class 3 as class 2.

The F1 score for localization is 0.83. It could be considered as good, indicating that the model has a robust performance in terms of both precision and recall. The model could capture a good proportion of relevant instances without a high rate of false positives. But it also suggests a room for improvement.

For the detection task, the performance is higher.

The model correctly predicted the label 0 (healthy state) 75 times, and correctly predicted the label 1 and 3 (damage) 40 times. Only 5 times the model estimated as undamaged cases, real damaged features.

Model demonstrates excellent performance, showing high accuracy, perfect precision, high recall and very high F1 score for detection, which resulted equal to 0.96. However, the presence of the three cases of false negatives shows that while

the model is highly reliable, it is not entire infallible. Overall, it is possible to conclude that corresponds to a robust model.

Chapter 7

Discussion

Data-driven SHM methods include a variety of strategies for understanding and monitoring structures' behaviour and eventually identify damage. All the information available comes directly from the measurements. Then, if the measured data is correctly elaborated, it can give a high quality, precise and crucial information about the state of the structure. In this way, is possible to perform the different steps of damage identification. The scarcity of data and the low quality of the available one impulse the implementation of transfer learning and domain adaptation approaches as proposed by the PBSHM theory. Understanding the importance of the methodology, the necessity to process correctly the data and improve its condition, it is possible to understand also the development of SA, a special field of DA, as explained in the main text of this study.

On the other hand, the creation of the 2D and 3D models (3.2.1 and 3.2.2) was done trying to catch the principal behavior of the structures in a coherent and simple way. Nevertheless, it took long time to insert the modelling properly in the software, and then create the entire population, considering the insertion of damage and automatically create the labelled healthy and damaged features. By doing so, it was possible to delve much deeper into the methodology and understand the root of the problem beyond the complexity of the models.

The analysis of the application of SA in both 2D and 3D models for damage detection and localization across different source and target domains (varying by the number of floors, materials and geometry) provides insightful results into the effectiveness and adaptability of these models. Synthesizing the findings, it will be possible to draw comprehensive conclusions.

The confusion matrices for both 2D and 3D models reveal high accuracy in damage detection and localization, with very few misclassifications, in the cases showed. Specifically, the 3D models demonstrated high performance in localization tasks, with almost perfect accuracy in distinguishing between the different locations of damage. This high level of accuracy in 3D models, especially with a perfect score

(100) in several instances, underscores the model's robustness in understanding spatial relationships and nuances in structure damage.

For damage detection, the confusion matrices also reflect a high level of accuracy. Particularly, the consistent performance with a low false positive rate (as observed in the minimal number of instances where non-damaged cases were incorrectly classified as damaged) demonstrates the model's precision in identifying actual damage cases.

The variation in performance based on the discrepancy between source and target domains (in terms of the number of floors) is particularly noteworthy. Despite these variations, the models maintained a high level of accuracy, which could suggest a strong capability of the SA method in adapting to different domain specifics without significant loss in performance. This adaptability is crucial for practical applications where the specifics of the source and target domains might not always align perfectly.

The detailed F1 scores for damage detection and localization across different source and target structures offer a nuanced view of model performance.

High F1 scores across the board, especially for detection (with many instances of perfect scores), highlight the models' effectiveness in both recognizing damaged structures and minimizing false positives/negatives.

For localization, while the performance is generally high, the variation in F1 scores suggests that certain configurations of source and target domains are more challenging than others. This is due to the complexity introduced by varying numbers of floors, which affect the spatial features that the model relies on for accurate localization.

By analyzing the cases proposed is possible to affirm that, overall, the application of SA in both 2D and 3D models for damage detection and localization demonstrates high effectiveness, with remarkable accuracy and adaptability across varying domain specifications. The models' ability to maintain high performance despite discrepancies in the source and target domains highlights their potential for wide-ranging applications in SHM and damage assessment.

The analysis of F1 scores for damage detection and localization across a population of structures, offers profound insights into the performance of SA methodologies in SHM. The F1 scores for damage detection are predominantly high across most source-target structure pairings, demonstrating a remarkable capability of the models to accurately identify damage, regardless of the structure.

A consistent perfect score (1.00) in several pairings, indicating an exemplary performance in damage detection across these specific configurations. The resilience of the model is evident in scenarios where the source and target structures vary significantly, as seen in the maintenance of high F1 scores (e.g., 0.98-1.00 range) in diverse structural contexts. Certain configurations (e.g., source 3 to target 7 with an F1 score of 0.82) indicate areas where model performance could potentially

be optimized, suggesting that variations in structural features might impact the model's detection capabilities.

The F1 scores for localization present a more nuanced scenario, reflecting both the strengths and limitations of the models in precisely localizing structural damage. High F1 scores (near or equal to 1.00) in many configurations underscore the models' accuracy in correctly localizing damage within structures, particularly when the source and target structures are closely related. Variability in scores, such as the drop to 0.61 or 0.25 in certain configurations, highlights challenges in localizing damage when there are significant discrepancies between the source and target structures. This suggests that spatial and structural complexities could pose difficulties for the models. The broad range of scores (from as low as 0.25 to perfect scores) across different pairings indicates the impact of specific structural characteristics on localization accuracy. It suggests that while the model excels in similar configurations, it encounters difficulties in more complex scenarios.

The analysis reveals a high degree of efficacy in using statistical alignment for damage detection across a varied population of structures. The models demonstrate exceptional accuracy in many scenarios suggesting they are well-suited for practical applications in SHM. However, the variability in localization accuracy underscores the need for further model refinement. Enhancing the models to better handle the complexities introduced by varying structural characteristics could lead to more uniformly high performance.

From the experimental case, it can be observed that there is significant difficulty in data collection, where not only was there dispersion, but there were also many outliers and scattered points. This indicates that the data processing part of the procedure is very important for obtaining high-quality results. However, when the algorithm was used for a three-storey aluminum source with geometry similar to the bookshelf structure, and the target was the available data of the mentioned case, the results were very consistent. IN fact for localization the algorithm showed d strong performance in identifying the healthy state and damage at last floor, but with difficulties in detecting damage at first floor. For detection, the algorithm results effective in identifying the presence or not of damage.

Overall, these results highlight the importance of data processing and the potential effectiveness of the algorithm in complex detection and localization tasks, especially in scenarios with significant data dispersion and outliers. The performance in the three-story aluminum structure case demonstrates the algorithm's applicability in real-world conditions, although there's room for improvement, particularly in localization accuracy across all categories.

One of the main contributions in [13] about SA it is how robust is the methodology by treating class imbalance. In the situations considered by the investigation this problem was not treated but certainly the effectiveness in the partial DA scenarios was proved showing consistent results.

Another aspect is that the machine learning algorithm is prone to diminishing performance, meaning it has its limitations and can restrict effective resolution in predicting various conditions regardless of the standardization developed by SA.

Chapter 8

Conclusions

The goal of this investigation is to analyze the application of Statistical Alignment as a Domain Adaptation approach in the field of PBSHM.

The key motivation for continuously develop SHM systems is to facilitate informed decision-making regarding operation and maintenance. This give arise benefits as cost-optimal operation and maintenance, improved safety and lifetime extension. Statistical Alignment is a possible solution to solve the problems related to the differences in the distributions of the source and target datasets.

Knowing that classical Domain Adaptation domains under this condition, are susceptible to negative transfer due to class imbalance or partial Domain Adaptation, making the mapping difficult to be interpreted.

Through the application of Statistical Alignment in the numerical population was possible to understand how this approach can harmonise heterogeneous domains, and investigate how the differences in materials, geometries and size of the structures were removed by the standardization proposed, demonstrating the robustness of NCA to prevent negative transfer.

In addition, the population was simulated using three-dimensional models, resulting in systems characterized by a different number of degrees of freedom. Besides, these degrees of freedom were related to translations in two directions and rotations. Therefore, it was necessary to implement an additional feature selection step, based on the comparison of normalized modal shapes (see Section 3.2.2, features selection), via MAC criterion.

The normalization of the mode shapes allowed to overcome the issue of different numbers of degrees of freedom, and apply the MAC to compare the features acquired from each pair of source and target structures. Consequently, it was possible to select the most similar features for performing knowledge sharing within the PBSHM framework.

Furthermore, knowledge transfer has been performed involving a benchmark experimental case study. This allowed to extend this investigation to real measured

data, showing how simulated data can be exploited, after Domain Adaptation, to inform diagnostic inferences on real structures.

The analysis of the experimental case study required introducing additional methodologies for data processing and feature extraction. Specifically, the SSICOV, presented in Chapter 6.2.1, has been employed to obtain the desired natural frequencies.

In conclusion, the results affirm the potential of SA methodologies in the field of SHM, with significant success in damage detection and promising, though variable, outcomes in damage localization. Future work aimed at addressing the noted challenges could unlock even broader applications and more reliable performance across an even wider range of structural configurations. Subsequent developments in this study regard be the application of additional techniques and methodologies, such as the Transfer Component Analysis, or Domain-Adversarial Neural Networks, in the field of Transfer Learning and Domain Adaptation approaches. Indeed, these algorithms may be used to perform further harmonization of the features when the source and target structures are more dissimilar, as proposed in [13].

Additionally, the classification and detection of damage may be performed using more complex and accurate approach than KNN, which was chosen to focus on SA rather than classification algorithms.

Bibliography

[1] Charles Farrar and Keith Worden. *Structural Health Monitoring A Machine Learning Perspective*. Jan. 2013. ISBN: 978-1-119-99433-6. DOI: 10.1002/9781118443118 (cit. on pp. 1, 7).

[2] Eloi Figueiredo and Adam Santos. «Machine Learning Algorithms for Damage Detection». In: June 2018, pp. 1–39. ISBN: 978-1-78634-496-0. DOI: 10.1142/9781786344977_0001 (cit. on p. 1).

[3] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X. Gao. «Deep learning and its applications to machine health monitoring». In: *Mechanical Systems and Signal Processing* 115 (2019), pp. 213–237 (cit. on p. 2).

[4] Lawrence Bull et al. «Foundations of population-based SHM, Part I: Homogeneous populations and forms». In: *Mechanical Systems and Signal Processing* 148 (Feb. 2021), p. 107141. DOI: 10.1016/j.ymssp.2020.107141 (cit. on pp. 2, 10, 12).

[5] J. Gosliga, P.A. Gardner, L.A. Bull, N. Dervilis, and K. Worden. «Foundations of Population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities». In: *Mechanical Systems and Signal Processing* 148 (2021), p. 107144. ISSN: 0888-3270 (cit. on pp. 2, 10, 11).

[6] P A Gardner, L A Bull, J Gosliga, N Dervilis, and K Worden. «Foundations of population-based SHM, Part III: heterogeneous populations–mapping and transfer». In: *Mechanical Systems and Signal Processing* 149 (2021), p. 107142 (cit. on pp. 2, 10).

[7] George Tsialiamanis, Charilaos Mylonas, Eleni Chatzi, Nikolaos Dervilis, David J Wagg, and Keith Worden. «Foundations of population-based SHM, Part IV: The geometry of spaces of structures and their feature spaces». In: *Mechanical Systems and Signal Processing* 157 (2021), p. 107692 (cit. on pp. 2, 10).

[8] P. Gardner, X. Liu, and K. Worden. «On the application of domain adaptation in structural health monitoring». In: *Mechanical Systems and Signal Processing* 138 (2020), p. 106550. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2019.106550 (cit. on pp. 2–4, 13, 17, 25).

[9] Rongrong Hou and Yong Xia. «Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019». In: *Journal of Sound and Vibration* 491 (2021), p. 115741. ISSN: 0022-460X. DOI: https://doi.org/10.1016/j.jsv.2020.115741 (cit. on p. 3).

[10] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, Moncef Gabbouj, and Daniel J. Inman. «A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications». In: *Mechanical Systems and Signal Processing* 147 (2021), p. 107077. ISSN: 0888-3270. DOI: https://doi.org/10.1016/j.ymssp.2020.107077 (cit. on p. 3).

[11] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, and Daniel J. Inman. «Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks». In: *Journal of Sound and Vibration* 424 (2018), pp. 158–172. ISSN: 0022-460X. DOI: https://doi.org/10.1016/j.jsv.2018.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0022460X18301792 (cit. on p. 3).

[12] Siyu Zhang, Hua-Liang Wei, and Jinliang Ding. «An effective zero-shot learning approach for intelligent fault detection using 1D CNN». In: *Applied Intelligence* 53 (Dec. 2022). DOI: 10.1007/s10489-022-04342-1 (cit. on p. 3).

[13] Jack Poole, Paul Gardner, Nikolaos Dervilis, Lawrence Bull, and Keith Worden. «On statistic alignment for domain adaptation in structural health monitoring». In: (2022). eprint: 2205.12052 (cs.LG) (cit. on pp. 3, 13, 16, 24, 29–31, 34, 36, 65, 68).

[14] Sinno Jialin Pan and Qiang Yang. «A survey on transfer learning». In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359 (cit. on p. 4).

[15] Sebastian Schommer, Viet Ha Nguyen, Stefan Maas, and Arno Zürbes. «Model updating for structural health monitoring using static and dynamic measurements». In: *Procedia Engineering* 199 (2017). X International Conference on Structural Dynamics, EURODYN 2017, pp. 2146–2153. ISSN: 1877-7058 (cit. on p. 7).

[16] Keith Worden and Graeme Manson. «The application of machine learning to structural health monitoring». In: *Philosophical Transactions of the Royal Society A* 365 (2007), pp. 515–537 (cit. on p. 7).

[17] A. Rytter, Rune Brincker, and L. Pilegaard. «Vibration based inspection of civil engineering structures». In: *Bygningsstatiske Meddelelser* 62 (Jan. 1991), pp. 79–110 (cit. on p. 7).

[18] Xuan Zhang and Luyu Li. «An unsupervised learning damage diagnosis method based on virtual impulse response function and time series models». In: *Measurement* 211 (2023), p. 112635. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2023.112635 (cit. on p. 7).

[19] Anil Jain, Robert Duin, and Jianchang Mao. «Statistical Pattern Recognition: A Review». In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (Jan. 2000), pp. 4–37. DOI: 10.1109/34.824819 (cit. on p. 9).

[20] Keith Worden, D. Hester, A. Bunce, and Julian Gosliga. «When is a Bridge not an Aeroplane?» In: *10th International Conference on Structural Health Monitoring of Intelligent Infrastructure, SHMII 10*. Porto, Portugal, 2021 (cit. on p. 11).

[21] G Delo, A Bunce, EJ Cross, J Gosliga, D Hester, C Surace, K Worden, and DS Brennan. «When is a Bridge Not an Aeroplane? Part II: A Population of Real Structures». In: *European Workshop on Structural Health Monitoring*. Springer. 2022, pp. 965–974 (cit. on p. 11).

[22] Chandula T Wickramarachchi, Julian Gosliga, Elizabeth J Cross, and Keith Worden. «On the Use of Graph Kernels for Assessing Similarity of Structures in Population-Based Structural Health Monitoring». In: *European Workshop on Structural Health Monitoring: EWSHM 2022-Volume 2*. Springer. 2022, pp. 995–1004 (cit. on p. 11).

[23] D. S. Brennan, T. J. Rogers, E. J. Cross, and K. Worden. «On quantifying the similarity of structures via a Graph Neural Network for population-based structural health monitoring». In: *Proceedings of the 30th International Conference on Noise and Vibration Engineering, ISMA 2022*. KU Leuven Department of Mechanical Engineering. 2022 (cit. on p. 11).

[24] Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. «Graph matching networks for learning the similarity of graph structured objects». In: *International Conference on Machine Learning*. PMLR. 2019, pp. 3835–3845 (cit. on p. 11).

[25] G Delo, C Surace, K Worden, and D S Brennan. «On the influence of structural attributes for assessing similarity in population-based Structural Health Monitoring». In: *Proceedings of the 14th International Workshop on Structural Health Monitoring 2023: Designing SHM for Sustainability, Maintainability, and Reliability*. DEStech Publications, Inc., 2023 (cit. on p. 11).

[26] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. «Transfer learning for time series classification». In: *2018 IEEE international conference on big data (Big Data)*. IEEE. 2018, pp. 1367–1376 (cit. on p. 13).

[27] Jaechang Nam, Sinno Pan, and Sunghun Kim. «Transfer defect learning». In: *Proceedings - International Conference on Software Engineering* (May 2013), pp. 382–391. DOI: 10.1109/ICSE.2013.6606584 (cit. on p. 13).

[28] Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang. «Domain Adaptation via Transfer Component Analysis». In: *IEEE Transactions on Neural Networks* 22.2 (2011), pp. 199–210. DOI: 10.1109/TNN.2010.2091281 (cit. on p. 13).

[29] Jindong Wang, Yiqiang Chen, Shuji Hao, Feng Wenjie, and Zhiqi Shen. «Balanced Distribution Adaptation for Transfer Learning». In: (Nov. 2017), pp. 1129–1134. DOI: 10.1109/ICDM.2017.150 (cit. on p. 13).

[30] Boqing Gong, Yuan Shi, Fei Sha, and K. Grauman. «Geodesic flow kernel for unsupervised domain adaptation». In: *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2012), pp. 2066–2073. DOI: 10.1109/CVPR.2012.6247911 (cit. on p. 13).

[31] S. Christides and A.D.S. Barr. «One-dimensional theory of cracked Bernoulli-Euler beams». In: *International Journal of Mechanical Sciences* 26.11 (1984), pp. 639–648. ISSN: 0020-7403 (cit. on pp. 18, 24, 25).

[32] M. Pástor, Michal Binda, and Tomáš Harčarik. «Modal Assurance Criterion». In: *Procedia Engineering* 48 (Dec. 2012), pp. 543–548 (cit. on p. 23).

[33] A. D. S. Barr. «An Extension of the Hu-Washizu Variational Principle in Linear Elasticity for Dynamic Problems». In: *Journal of Applied Mechanics* 33 (1966), pp. 465–465 (cit. on p. 25).

[34] P. Gardner, L.A. Bull, J. Gosliga, J. Poole, N. Dervilis, and K. Worden. «A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings». In: *Mechanical Systems and Signal Processing* 172 (2022), p. 108918 (cit. on p. 29).

[35] L.A. Bull, P.A. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R.S. Mills, and K. Worden. «On the transfer of damage detectors between structures: An experimental case study». In: *Journal of Sound and Vibration* 501 (2021), p. 116072 (cit. on p. 29).

[36] P Gardner, LA Bull, J Gosliga, J Poole, N Dervilis, and K Worden. «A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings». In: *Mechanical Systems and Signal Processing* 172 (2022), p. 108918 (cit. on p. 29).

[37] Valentina Giglioni, Jack Poole, Ilaria Venanzi, Filippo Ubertini, and Keith Worden. «On the use of domain adaptation techniques for bridge damage detection in a changing environment». In: *ce/papers* 6.5 (2023), pp. 975–980 (cit. on pp. 34, 36).

[38] Mahmudul Hasan, Md. Milon Islam, Md Ishrak Islam Zarif, and M.M.A. Hashem. «Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches». In: *Internet of Things* 7 (2019), p. 100059 (cit. on p. 36).

[39] *SHM Data Sets and Software*. `https://www.lanl.gov/projects/national-security-education-center/engineering/software/shm-data-sets-and-software.php`. Los Alamos National Laboratory Engineering Institute (cit. on pp. 54, 56).

[40] E. Cheynet. *Operational modal analysis with automated SSI-COV algorithm*. Version 2.5. Feb. 2021. DOI: `10.5281/zenodo.4501822`. URL: `https://doi.org/10.5281/zenodo.4501822` (cit. on p. 56).

[41] Filipe Magalhães, Álvaro Cunha, and Elsa Caetano. «Online automatic identification of the modal parameters of a long span arch bridge». In: *Mechanical Systems and Signal Processing* 23.2 (2009), pp. 316–329. ISSN: 0888-3270. DOI: `https://doi.org/10.1016/j.ymssp.2008.05.003` (cit. on p. 56).

[42] Etienne Cheynet, Jónas Snæbjörnsson, and Jasna Jakobsen. «Buffeting response of a suspension bridge in complex terrain». In: *Engineering Structures* 128 (Dec. 2016), pp. 474–487. DOI: `10.1016/j.engstruct.2016.09.060` (cit. on p. 56).

[43] Etienne Cheynet, Jasna Bogunović Jakobsen, and Jonas Snæbjörnsson. «Damping estimation of large wind-sensitive structures». In: *Procedia Engineering* 199 (2017). X International Conference on Structural Dynamics, EURODYN 2017, pp. 2047–2053. ISSN: 1877-7058. DOI: `https://doi.org/10.1016/j.proeng.2017.09.471` (cit. on p. 56).

[44] Etienne Cheynet, Jónas Snæbjörnsson, and Jasna Jakobsen. «Temperature Effects on the Modal Properties of a Suspension Bridge». In: June 2017, pp. 87–93. ISBN: 978-3-319-54777-0. DOI: `10.1007/978-3-319-54777-0_12` (cit. on p. 56).

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Cecilia Surace, for giving me this opportunity to complete my studies at Politecnico di Torino.

To Giulia, for her inspiring dedication and unparalleled support during this final phase.

To my parents, Esteban and Gabriela, for teaching me the value of sacrifice and perseverance.

To my sister Josefina, for providing me with a mother's care.

To Hernán, Paulina, and Federico, for their companionship.

To my grandmothers, Marina and Zulema, for the warmth of their affection.

To Juan, Tomás, Ignacio, Agustin, Carme, Cande, and Vicki, for the priceless value of their friendship.

To Rafa, Cesar and Claudio, who through the passion with which they work, have been an example.

To Father Jesus Segura and the Noviciate "Pier Giorgio Frassati", for their close guidance and support during this last period.

And most importantly, to the Virgin Mary for her maternal care, and to God, for blessing me with all of this. Blessed be God.

"And we know that all things work together for good to them that love God" (Romans 8:28).